Discovery of Aminoacyl-tRNA Synthetase Mutants for the Incorporation of Noncanonical Amino Acids into Proteins

Thesis by
İsmet Çağlar Tanrıkulu

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

California Institute of Technology
Pasadena, California
2009

(Defended October 8, 2008)

Acknowledgements

As it is for many, graduate school has been a long, convoluted journey for me as well. Now that this part of my life is coming to a close, I realize more than ever the importance of so many people who helped me reach where I am right now.

First and foremost I would like to thank my “experimental” advisor, Prof. David Tirrell, for his guidance and support for the past six years. His patience in letting me carve my own research path, his willingness to learn about the details of my work, listen to the problems and discuss solutions, his genuine interest in understanding the principles behind the data on the screen all allowed me to learn a great deal on how to do science and how to have fun doing it. My “computational” advisor, Prof. Bill Goddard, was extremely patient with me during the many years of failed binding-site designs, and always had countless ideas for the next step. If it wasn’t for his encouragement for me to move on, I might have still been testing “dead” PheRS mutants until this day.

I would like to thank the current and past members of my thesis committee, Profs. Harry Gray, Linda Hsieh-Wilson, Rich Roberts, and Vaidehi Nagarajan. Our meetings occurred at critical times during my training. Their interest in my work and questions they posed helped me reevaluate my research path in new light and greatly influenced the direction of research I pursued. I very much appreciate their guidance.

Perhaps the most difficult period in my PhD was after my first candidacy exam, which ended less than satisfactorily for most involved. Prof. Judy Campbell, who was the BMB option head at the time, and Alison Ross, our option secretary, were both very supportive during this time and encouraged me in my quest to find a new research lab. My conversations with Prof. Doug Rees, the chair of my first committee, helped me put all events in perspective. Without their help I most likely would not be here today.

My first introduction to both molecular biology and computer programming happened while I was working in the Pierce and Mayo labs. I learned many fundamental molecular biology techniques from Rhonda Diguisto, who is a very

patient and organized teacher. Shannon Marshall introduced me to circular dichroism, and Possu Huang was helpful with anything that went wrong in the lab, which, as some may know, is the norm rather than the exception. As someone who did not have much programming experience I received a lot of help deciphering, modifying, and using the ORBIT code from Geoff Hom, Possu Huang, Dan Bolon, Robert Dirks, Shannon Marshall, Premal Shah, Deepshikha Datta, Eun Jung Choi, Jessica Mao, Kyle Lassila, and Marie Ary.

The Tirrell lab, where I completed the experimental part of my PhD work, has provided me with a wonderful environment, both scientifically and socially. I really doubt I will ever encounter such strong camaraderie again in the future. A diverse set of skills are required for success in the Tirrell lab and I got introduced to many of these through Marissa Mock, Sarah Heilshorn, Jin Montclare, and Jamie Link. However, I became confident with my lab skills only after receiving tremendous amounts of help from Inchan Kwon, with whom I shared both my lab bay and my office. In addition to letting me use his endless knowledge on all types of procedures, Inchan also taught me how to design and pursue experiments, how to judge the factors affecting them, and above all, how to be consistent. I deeply appreciate his help. I was also fortunate enough to have many people around me to troubleshoot problems, to exchange ideas, or just to have a talk with in the lab. Among these I have bothered Rebecca Connor, Tae Hyeon Yoo, Nick (J.D.) Fisk, John Ngo, Julie Champion, Ayae Sugawara, Makoto Ouchi, Soojin Son and Jin Montclare a countless times, and I thank them for their support and friendship. Jamie Link was the person who first developed the screening strategy I describe and use in Chapter 2, and his assistance was very valuable to me as I was starting up. Nick Fisk, Jamie Link, John Ngo and Mark van Eldijk allowed me to use reagents they had synthesized. Inchan Kwon and Tae Hyeon Yoo offered me advice and assistance on flow cytometry and library screening in general, and, with Marissa Mock, helped me with the $^{32}PP_i$-exchange assays. The MetRS-SLL x-ray crystal structure was obtained through a collaboration initiated by Tae Hyeon Yoo, with Yves Mechulam and Emmanuelle Schmitt and made many of the analyses presented here possible.

Finally I would like to thank John Ngo for turning the results of what was originally an intellectual exercise into an exciting tool for cell-specific metabolic labeling.

I performed the computational part of my PhD work in the Goddard lab, which contains an immense amount of expertise. The two people in the group who had the biggest impact on me were Vaidehi Nagarajan and Ravi Abrol. Vaidehi taught me the basics of all the computational work I present here. Her patience while working with me on the difficult problems and her encouragement kept me moving forward even when chances of success seemed bleak. I also learned a lot from Ravi, who has the amazing ability divide large problems into logical, manageable chunks. His guidance enabled me to obtain the results I now present in this work. I was lucky to have access to such great mentors. My work in the group involved the use of many newly introduced computational techniques, which required me to work closely with a number of people. David Zhang, who created the COP strategy for binding site design, helped me greatly as I was getting started in the group. Victor Kam's side chain placement procedure (SCREAM) was integral to my design strategy from the start, and he was always available for my questions and requests. My meetings with Pete Kekenes-Huskey were always intellectually stimulating, whether they were about rotamer generation or binding calculations. I really hope his torsional sampling code finds the audience it deserves. Jonas Oxgaard and Julius Su helped me out whenever my QM calculations got stuck. Andres Jaramillo-Botero was very understanding with my time constraints when I had to work on our ligand docking code and my thesis at the same time. Finally, it has always been very gratifying to see my code being used by the group, and therefore, I also would like to thank the users of various programs I have written through the years.

At Caltech, we are lucky to have an amazing staff to help us get through the day. Alison Ross (BMB), Carol Mann and Anne Horman (Tirrell lab), and Shirley Feng (Goddard lab) have worked with me through many scheduling and administrative problems that came up. Suresh Ghupta (ChE) and Darryl Willick (Goddard lab) have provided solutions to my computer related problems over and over and over again. Haick Issaian, with Karen Baumgartner and other

EHSS staff, answered countless questions on radiation use and helped me locate reagents and working instruments when I needed them. Mona Shahgholi trained me, helped me run MALDI and LC-MS measurements, and helped troubleshoot problems. In addition to research-related staff, I also would like to thank all the people on campus, from Jim Endrizzi and Athena Trentin (ISP) to Tess Legaspi (Registrar's office), from the health center staff to the pizza guys at Chandler, for keeping life at Caltech easy for us.

It would have been impossible for me to live through years of failed experiments if it wasn't for the love and support I received from my friends and family. In addition to ones I've already mentioned, the Turkish community at Caltech, especially Ozgun Konca, Sirin Caliskan, Can Aydiner, Baris Bingol, Arkadas Ozakin, Ercan Gurses, Omer Durak, Ersen Bilgin and Grace Song, and Muruvvet Buyukboyaci have been a great source of support for me, day in and day out. I thank Sarina Mohanty for constantly checking on me during the preparation of my thesis, and to Jose Navarette and Ozgun Konca for playing music with me from time to time. Although I am yet to determine whether she was a blessing or a curse, Eun Jung Choi was there for me during most of my stay at Caltech and deserves credit for her assistance and support.

My parents have always encouraged me to ask questions about anything and everything since I started talking, even if this would cause problems once in a while. They did everything for me to get a good education, and it seems to have worked. I have been away from home for the most part in the last two decades, and during this time they always made sure I knew they were there for me. In this past year, during the preparation of this thesis, I felt their presence even stronger. I am forever grateful for their love, care and support.

At last, I would like to thank Julie Champion, John Ngo, Ravi Abrol, Shelly Tzlil, Grace Song and Prof. David Tirrell for all the time they spent reviewing and editing various parts of this thesis.

Abstract

Efficient *in vivo* incorporation of a noncanonical amino acid into proteins often requires engineering new aminoacyl-tRNA synthetase (AARS) activity into the cell, usually by modifying a natural aaRS. Although experimental methods, relying on mutagenesis and library screening, have identified many successful mutant aaRS-substrate pairs in the recent years, computational approaches have reported only a few successes. Here we compare the results of computational and experimental screens of an *E. coli* methionyl-tRNA synthetase (MetRS) saturation-mutagenesis library for binding (*in silico*), and activation and cell-surface display (*in vivo*) of azidonorleucine (Anl).

Three positions (L13, Y260, and H301) in the methionine binding pocket of *E. coli* MetRS were randomized, and the resulting library was screened for MetRS activity toward Anl, based on a screening strategy previously established in our group. This strategy relies on the introduction of reactive side chains into surface-exposed sites on outer-membrane protein C (OmpC), and their subsequent labeling with reactive, fluorescent probes. We have discovered a large diversity of MetRS mutants that allow the incorporation of Anl into proteins *in vivo*. The extent of OmpC expression and the amount of available Anl during the screen have substantial effects on the outcome of the screens. In addition to displaying improved activities toward Anl, identified mutants also show an improved discrimination against Met. We have shown that the degree of cell-surface labeling *in vivo* correlates well with the measured rates of Anl activation *in vitro*, which reflects the success of the screen design.

Computational analysis of the experimentally identified mutants revealed a good agreement between computed binding energies and *in vitro* activation data. To better test the computation model, we performed an *in silico* screen for Anl binding on a saturation-mutagenesis library comparable to the experimental library. Computational screen predominantly selects mutants that interact with Anl through hydrogen bonds, whereas the hydrophobic residues are selected more often by the experimental screen. We identify that experimental mutants try to optimize packing at the Y260 and H301 sites, but not at the L13 site. We

discuss possible explanations for these results. Combined results from computation and experiments suggest the importance of various factors in ligand recognition and *in vivo* selection of MetRS mutants. We explore the implications of these factors to the future efforts in the engineering of new MetRS activities.

Table of Contents

Page

List of Figures

List of Tables

Chapter 1

Introduction

Role of Aminoacyl-tRNA Synthetases (AARS) in Cellular Protein Synthesis

Cellular protein synthesis is directed by the genetic code and takes place in two major steps. The first step, transcription, is the synthesis of messenger RNA (mRNA) from DNA. This is followed by translation, the synthesis of protein from amino acids as directed by the mRNA. Transcription is in some ways similar to DNA replication: chemical structures and polymer properties of RNA and DNA are similar enough that the ability of complementary sequences to form duplexes through base pairing exists for both DNA and RNA molecules, as well as DNA-RNA hybrids. Hence, DNA is an intrinsic template for RNA polymerization, as it is for its own replication. However, such sequence complementarity does not exist between amino acids that have diverse chemical structures and their polymerization template, the mRNA in the translation step. Therefore a translator is required in order to match every specific code on the mRNA (codon) to its corresponding amino acid.

Transfer RNA (tRNA) molecules act as adapters between amino acids and the codons on the mRNA in translation. Two recognition events characterize this process: codon recognition by tRNA and amino acid–tRNA matching. The recognition of a codon on the mRNA by the complementary anticodon on the tRNA is straightforward due to base pairing. The process of matching the correct amino acid with the correct tRNA is more difficult and is overseen by a class of enzymes called aminoacyl-tRNA synthetases (AARS). These enzymes catalyze the reaction that attaches an amino acid to the 3'-end of its corresponding tRNA molecule. Aminoacyl-tRNA synthesis takes place in two steps:

$$\text{AARS} + \textit{aa} + \text{ATP} \Leftrightarrow \text{AARS}\bullet\textit{aa}\text{-AMP} + \text{PP}_i \,, \qquad (1.1)$$

$$\text{AARS}\bullet\textit{aa}\text{-AMP} + \text{tRNA}^{aa} \Leftrightarrow \textit{aa}\text{-tRNA}^{aa} + \text{AARS} + \text{AMP} \,, \qquad (1.2)$$

where *aa* is an amino acid, and AARS and tRNA^{aa} are the corresponding aminoacyl-tRNA synthetase and tRNA, respectively. In the first step, the acid group on the amino acid is activated by forming a mixed anhydride with the α-phosphate on ATP, resulting in a pyrophosphate leaving group. In the second step, AMP is displaced by a nucleophilic attack from the 2'- or 3'-hydroxyl of the tRNA to give aminoacyl-tRNA [1]. These steps are catalyzed by the

corresponding AARS enzyme. To ensure the fidelity of information transfer from the nucleus to functional cellular proteins, the AARS has to be able to recognize with high accuracy the cognate amino acid out of all the small molecules in the cell, as well as the correct tRNA molecules from the 20 classes present. Upon binding, AARSs make extensive contacts with the tRNA molecules over a large surface area (2,470 to 5,650 $Å^2$) [2, 3]. Such large areas of interaction enable AARSs to distinguish the sequence specific features on the tRNAs [4] and ensure correct recognition. Hence, the error rates in tRNA recognition are usually less than 1 in 10^6 [5]. Amino acids, however, are small molecules and have much fewer features that distinguish one from another. Particularly AARSs that aminoacylate tRNAs with Ile, Thr, and Ala need to discriminate against amino acids that have one fewer methyl group, Val, Ser, and Gly, respectively. Pauling [6] calculated that the binding-energy difference between Ile and Val for the isoleucyl-tRNA synthetase (IleRS) could not result in an error rate better than 1 in 5 for the misactivation of Val, based on the contribution of the extra methyl group on Ile. However, in reality, this rate is only 1 in 3000 for IleRS [7] and around 1 in 10^4 to 10^5 for AARSs in general [8]. This observed high specificity of the AARSs for their cognate amino acid is ensured by additional proofreading or editing activities that hydrolyse the product when noncognate amino acids are activated or charged onto tRNAs.

Recognition of the cognate amino acid by the AARS is critical for the fidelity of the translation process. Theoretical models show that, rather than ribosomal errors, tRNA charging by the AARSs plays a key role in controlling the long-term stability of the cell. Error-prone ribosomes are more likely to produce inactive protein products [9], whereas mistakes in tRNA charging can lead to an increase in error rates by producing “mutant” AARSs in the next generation protein synthesis. A high AARS error rate was predicted to cause an “error catastrophe” [10]. It was recently shown that the viability of cells containing different error-prone valyl-tRNA synthetases (ValRS) is decreased in proportion to ValRS error rates when competitor (α-aminobutyrate) concentrations were increased [11].

Incorporation of Noncanonical Amino Acids into Proteins

The lethality caused by errors from AARS is in part due to the absence of a control mechanism that checks for correct amino acid-tRNA pairing at the ribosome level. The lack of error correction allows amino acids on misacylated tRNAs to be incorporated into proteins. At the same time, this also makes it possible to bypass the recognition and charging steps altogether by chemically charging the tRNAs with the amino acid of choice *in vitro* and providing them directly to the translational machinery. The preparation of chemically misacylated tRNAs was first described in 1984 [12]. In 1989, Schultz and Chamberlin groups independently combined this technique with cell-free translation methods, and successfully incorporated noncanonical amino acids into proteins at amber codon (UAG) sites [13, 14].

The 20 natural amino acids represent a very limited set of structures and functionalities present in organic chemistry. The introduction of noncanonical amino acids expands this set to compounds with diverse electronic and molecular structures. By specifying the exact chemical structure on the amino acid side chain, rather than mutating whole residues, perturbations can be applied to individual atoms in a protein, allowing us to probe protein structure and function more accurately. The ability to introduce noncanonical amino acids also allows unique chemical functionalities and reactions, such as light triggered cross-linking, to be applied to proteins. Due to the enormous potential of noncanonical amino acids, methods for their incorporation drew considerable attention. Since 1989 more than a hundred noncanonical amino acids have been incorporated into proteins using amber suppression and other similar methods [15]. Noncanonical amino acids incorporated this way include structural probes and fluorescent, isotopic and spin labels [16-18]. Successful incorporation was reported both *in vitro* [19] and *in vivo*, through microinjection of misacylated tRNA into *Xenopus* oocytes [20]. In oocytes, mutations to noncanonical amino acids were introduced into the nicotinic acetylcholine receptor (nAChR). Receptors that exhibit constitutive activity at acidic pH were discovered through the

incorporation of modified tyrosines [21]. Effects of backbone hydrogen bonding was probed by mutating the amide linkage on the protein backbone to an ester through hydroxyacid incorporation [22]. Recently, *cis-trans* isomerization at a proline residue in nAChR was modulated by the incorporation of proline analogs with varying preferences for the *cis*-conformer [23]. Despite many successes, the above methods present one major limitation: the yield of engineered protein is intrinsically limited by the amount of chemically aminoacylated tRNA, synthesis of which is laborious. Thus, the yield of expressed protein is usually low at around one milligram or less per reaction [24]. The low protein yield is especially a problem if the engineered protein is to be used in biomaterial applications, where material is required in bulk.

Residue-specific Incorporation

Unlike chemical polymerization, the synthesis of proteins is directed precisely by the genetic code. This control over sequence and size translates into a precise control over the properties of the polymer synthesized. Genetic engineering allows us to modify the genetic code, and therefore, engineer novel protein-based materials using cellular protein synthesis machinery [25, 26]. Tirrell and co-workers have utilized these ideas to prepare artificial proteins that have controllable crystallization properties [27], form liquid crystal phases and ordered films [28, 29], reversible hydrogels [30], and matrix proteins for potential use in vascular reconstruction [31-33]. The utilization of noncanonical amino acids with novel side chains has the potential to provide properties to the peptides not possible with the natural set, and expanding the utility of their applications. When changes in bulk properties are desired, site-specific incorporation is not necessary for this case. However, *in vitro* translation with good efficiency is required in order to get a high protein yield.

Multisite (residue-specific) incorporation of amino-acid analogs in auxotrophic cell strains was observed as early as the 1950s [34, 35] and to a large extent suits the needs described above [36]. This method relies on the construction of a cell strain that is unable to synthesize a particular amino acid,

and the ability of the corresponding AARS to activate its structural analogs. First, auxotrophs to a certain amino acid are grown in media containing 20 natural amino acids. At the onset of the stationary growth phase, this amino acid is depleted and is replaced by one of its structural analogs through a media shift. Upon induction of protein synthesis, the analog taken up by the cell is charged by the cellular AARS onto the corresponding tRNAs, replacing the depleted amino acid. This way, the analog is incorporated into all proteins synthesized by the cell in place of the natural amino acid. Through this strategy, noncanonical amino acids that carry alkene [37], alkyne [38], azido [39], fluoro [40-42] and ketone [43] functionalities on their side chains have so far been incorporated into recombinant proteins in *E. coli* for protein engineering applications.

In the last decade, amino acids that carry noncanonical side chains have become useful protein engineering tools. Introduction of selenomethionine and other amino acids carrying heavy atoms into proteins greatly simplifies determination of x-ray crystal structures through multiwavelength anomalous diffraction [44, 45]. Fluorinated amino acids have been used as ^{19}F-NMR probes in proteins [46]. The hyperhydrophobicity of fluorocarbons have led to the construction of proteins with fluorous cores [40, 47-54]. These proteins exhibit large gains in thermostability compared with their non-fluorinated counterparts. The tryptophan analog 4-amino-tryptophan was used to modify the chromophore of the enhanced cyan fluorescent protein (ECFP) [55]. The resulting protein exhibited a redshift not achieved by laboratory evolution or screening.

The value and versatility of noncanonical amino acids have significantly increased through their use as bioorthogonal reactive handles for conjugation [56, 57], detection [58, 59] and selective isolation [60] of proteins. Upon incorporation into proteins, aryl halides allow palladium-catalyzed Heck and Sonogashira couplings [61, 62], and the acetyl group can be reacted selectively with hydrazides [43]. Among the reactive groups introduced into proteins, azides are perhaps the most significant [63]. Azides can be ligated under physiological conditions to ester-functionalized triaryl phosphines through the Staudinger ligation or to terminal alkynes through Cu(I)-catalyzed [3+2] azide-alkyne

cycloaddition, both in bioorthogonal fashion. Recently a strain-promoted version of the azide-alkyne cycloaddition that does not require the Cu(I) catalyst was reported [64], which allows this chemistry to be applied to live cells. The utility of the introduction of bioorthogonally reactive amino acids into proteins was demonstrated through their application to the study of proteomic changes in cells in response to stimuli [58-60]. Proteins synthesized following a stimulus were metabolically labeled with azide- or alkyne-bearing methionine analogs. Through the use of affinity tags for separation and identification, or fluorescent tags for visualization, the newly synthesized proteins were studied. Photoreactive aryl azides, such as *p*-azidophenylalanine were incorporated into proteins, and used in the construction of ligand-bearing [56] or lithographically patterned surfaces [57].

Modification of AARS Activities

The most prominent limitation for the *in vivo* incorporation of noncanonical amino acids is its reliance on the host's native AARSs to charge the amino acid of interest on to tRNAs. This requirement limits the set of noncanonical amino acids that could be incorporated to only those that are similar in size, shape and electrostatic structure to the natural amino acids they replace. Therefore, the diversity in structural and chemical functionality of analogs obtained using chemical misacylation-based techniques is limited. In cases where the associated AARS shows low activity for the desired analog, it has been possible to increase the overall activity by engineering *E. coli* hosts that overexpress the AARS of choice to achieve incorporation [65]. It is also possible that the analog of choice is activated or charged, but then is hydrolyzed through editing/proofreading. Disrupting the editing activity of the AARSs resulted in successful activation of amino acid analogs in valyl- [66] and leucyl-tRNA synthetases [67]. However, many noncanonical amino acids differ significantly from their natural counterparts in structure and are not substrates for natural AARSs. For these molecules, it is necessary to engineer AARS activity specifically. Point mutations at the active sites of certain AARSs, such as

phenylalanyl- [68] and tyrosyl-tRNA synthetase [69], are known to relax the binding specificity of AARS and allow the incorporation of amino acid analogs that are normally not incorporated. Enlarging the amino acid binding pocket of the *E. coli* phenylalanyl-tRNA synthetase (PheRS) by the A294G mutation made the introduction of various *para*-substituted phenylalanine analogs into proteins [61, 62, 70], as well as the photoreactive amino acid benzofuranyl-alanine [71] possible. An additional mutation (T251G) identified in the PheRS binding pocket by a computational protein design algorithm allowed the introduction of *p*-acetyl-phenylalanine into proteins [43]. Carrying this mutation over to the yeast homolog of this enzyme (T415G) enabled the site-specific incorporation of a series of tryptophan analogs in *E. coli* through the use of this enzyme with a mutant yeast phenylalanine amber suppressor tRNA [72]. A single mutation (C443G) in the proline binding pocket of *E. coli* prolyl-tRNA synthetase (ProRS) also permitted the incorporation of a series of proline surrogates, including *2S*-pipecolic acid, which is a proline analog with a six-membered ring [73].

Schultz and co-workers have employed directed evolution in order to identify mutations on AARSs that lead to specific activity toward a noncanonical amino acid of interest. Their studies focused on the modification of tyrosyl-tRNA synthetase from the archaebacterium *Methanococcus jannaschii* (*Mj*-TyrRS), for its use with its associated $tRNA^{Tyr}$ as a 21^{st} AARS-tRNA pair in *E. coli* in their site-specific *in vivo* incorporation studies [74]. In order to set a clear background for their screen, five residues thought to be in the active site were mutated to alanines. Using oligonucleotide-encoded DNA shuffling, a library of *Mj*-TyrRS mutants was created on these five positions, which was then screened in two steps. Positive selection for active *Mj*-TyrRS mutants was carried out in the first step, where *Mj*-TyrRS activity is required for cells to survive in antibiotic supplemented media. In the second step, negative selection is applied against the *Mj*-TyrRS mutants that retain their activity toward tyrosine. Here, a lethal gene is introduced into the cells, which are then grown in media that contains tyrosine but not its unnatural analog. Cells containing any *Mj*-TyrRS that is active toward tyrosine express the lethal gene and die. After several cycles of

mutagenesis and a two-step selection, the cells surviving the final positive selection are analyzed. This method has produced *Mj*-TyrRS mutants that exhibit high specificities to tyrosine analogs, and allowed the *in vivo* site-specific incorporation of a wide variety of noncanonical amino acids [75-78], including a glucose-modified serine [79]. Application of this methodology to a *E. coli* $tRNA^{Leu}$-leucyl-tRNA synthetase (LeuRS) pair enabled identification of LeuRS variants with diverse activities [80-83] and achieved site-specific incorporation of fluorescent and photocaged amino acids in yeast, among others.

In order to enable the incorporation of translationally silent methionine (Met) analogs into proteins, several screening strategies were developed for methionyl-tRNA synthetase (MetRS) libraries in the Tirrell group. Following their demonstration of the display of unnatural amino acids in *E. coli* cell surface [84], Link and co-workers showed that the display of unnatural amino acids on the cell surface could be used as a reporter for the cellular AARS activity toward an unnatural substrate [85]. Four residues in the *E. coli* MetRS were randomized and the MetRS library was screened for the incorporation of an azide-bearing amino acid (azidonorleucine, Anl) into the surface-accessible Met positions on *E. coli* outer-membrane protein C (OmpC). Only the clones carrying MetRS variants active toward Anl can display azide groups on the cell surface and exhibit fluorescence upon labeling with a fluorescent probe. These cells can be enriched on a fluorescence-activated cell sorter (FACS). Analysis of the enriched clones revealed that a single mutation to glycine (L13G) was sufficient to relax the specificity of MetRS to allow azidonorleucine incorporation. The above methodology requires bioorthogonal derivatization of reactive side chains presented on the cell surface. Yoo and Tirrell reported a more generally applicable strategy that takes advantage of a green fluorescent protein (GFP) variant, which was engineered to permit the incorporation of methionine analogs without any loss of fluorescence [86]. A five-position saturation-mutagenesis library was expressed in cells overproducing this GFP variant in the presence of 6,6,6-trifluoronorleucine (Tfn). Cells exhibiting the highest yields of GFP were isolated by FACS. The enriched population was also subjected to negative

selection, where GFP production was induced in the presence of 19 amino acids (-Met) and cells with low fluorescence were retained. After multiple rounds of selection, the L13S-Y260L-H301L mutant of MetRS was enriched, and quantitative replacement of Met with Tfn was demonstrated through the use of this mutant.

Computational Protein Design and Its Application to AARSs

The success of redirecting AARS activity to noncanonical amino acids through mutations in the binding pocket suggests that changing the binding specificity alone may be sufficient for new AARS activities. Not surprisingly, computational methods utilized for the identification of new AARS activities have focused on optimizing the binding interactions of the new substrate with the binding pocket. Fidelity of phenylalanyl- [87], methionyl- [88], and seryl-tRNA synthetases [89] to binding their natural substrates has been reliably demonstrated through computational methods. The relationship between enzymatic activities of AARSs toward unnatural ligands, and the corresponding computational binding energies was investigated in several studies. Wang et al. reported a good correlation between the translational activities of a set of phenylalanine analogues in *E. coli* and the binding energies of these analogs calculated based on a phenylalanyl-tRNA synthetase crystal structure from *T. thermophilus* (*Tt*-PheRS) [90]. In a similar study, Datta et al. showed that binding energies computed between *E. coli* MetRS and a series of methionine analogs agree with experimental free energies calculated based on kinetic parameters of activation [88]. In each case, the authors presented well-behaved scoring functions that can predict AARS binding, and therefore, the likelihood of incorporation, which suggests the possibility of computationally designing AARS binding sites for each analog of interest.

One such example has been presented by Datta et al. [43], who used a previously established design algorithm (ORBIT) [91, 92], which utilizes an empirical force field, a pairwise-decomposable description of energies and optimization methods based on the dead-end elimination theorem. Eleven

binding site residues were scanned to identify two mutations on the *Tt*-PheRS structure that should assist *p*-acetyl-phenylalanine binding. Corresponding mutations on the *E. coli* PheRS were then shown to allow the incorporation of this analog into proteins in *E. coli*. A different strategy was employed by Zhang et al. [93, 94], who focused on using a full molecular mechanics force field and a scoring function based on binding energies. Unlike ORBIT, this method relies on keeping the combinatorial complexity of the problem low by focusing on the few residues around the ligand. The success of Schultz and co-workers in selecting high fidelity enzymes from libraries of 5-fold mutants verifies this approach. In this method, named "clash-opportunity progressive design" (COP), first the mutations that relieve clashes and then those that improve binding were identified. These mutations were carried out combinatorially, and from this set, the mutants that prefer binding to the analog over any natural amino acids were identified. COP was successful in identifying two key mutations that allow the binding of *O*-methyl-tyrosine to *Mj*-TyrRS, previously discovered by Wang et al. [75]. Even though binding specificity of analogs to AARS is a necessary but not sufficient condition for their activation, the results above show that computational methods may be an alternative to library screening in altering AARS activity.

Discovery of MetRS Mutants Active Toward Azidonorleucine

Although attempts were made to design AARS binding sites, mutational data to verify the computational procedures have been a limiting factor in these studies. Thus, the experimental screens developed for the *E. coli* MetRS system [85, 86] and the encouraging results from computational studies of unnatural AARS activities [88, 93] together present a unique opportunity to evaluate the effectiveness of computational procedures. In this thesis, we take advantage of the methods developed in previous work to a explore a MetRS library both experimentally and computationally.

The work presented here has two main goals: identification of a variety of MetRS mutants that allow the efficient incorporation of azidonorleucine (Anl) into proteins, and evaluation of computational methods using the experimentally

acquired data. In Chapter 2, we build on the work by Link et al. [85] and identify a wide variety of MetRS mutants that enable Anl incorporation through screens of a three-position MetRS saturation mutagenesis library. In Chapter 3, we show a good agreement between computational binding energies and results from *in vitro* activation kinetics experiments. We also report the results from an *in silico* screen of a MetRS saturation mutagenesis library and identify where the computational model succeeds and fails by comparing the *in vivo* and *in silico* results. In Chapter 4, we discuss the implications of our results to future efforts in engineering new MetRS activities.

References

[1] Ibba M and Soll D, "Aminoacyl-tRNA synthesis," *Annu Rev Biochem*. **2000**, *69*:617.

[2] Nadassy K, Wodak SJ and Janin J, "Structural features of protein-nucleic acid recognition sites," *Biochemistry*. **1999**, *38*:1999.

[3] Rich A and Schimmel PR, "Structural organization of complexes of transfer RNAs with aminoacyl transfer RNA synthetases," *Nucleic Acids Res*. **1977**, *4*:1649.

[4] Cusack S, "Aminoacyl-tRNA synthetases," *Curr Opin Struct Biol*. **1997**, *7*:881.

[5] Sankaranarayanan R and Moras D, "The fidelity of the translation of the genetic code," *Acta Biochimica Polonica*. **2001**, *48*:323.

[6] Pauling L, "Festschrift fur Prof. Dr. Arthur Stoll," Birkhauser Verlag: Basel, **1958**; 597

[7] Loftfield RB, "Frequency of errors in protein biosynthesis," *Biochem J*. **1963**, *89*:82.

[8] Loftfield RB and Vanderjagt D, "Frequency of errors in protein-biosynthesis," *Biochem J*. **1972**, *128*:1353.

[9] Kowald A and Kirkwood TBL, "Accuracy of transfer-RNA charging and codon–anticodon recognition — relative importance for cellular-stability," *J Theor Biol*. **1993**, *160*:493.

[10] Freist W, Sternbach H, Pardowitz I and Cramer F, "Accuracy of protein biosynthesis: Quasi-species nature of proteins and possibility of error catastrophes," *J Theor Biol*. **1998**, *193*:19.

[11] Nangle LA, Lagardt VD, Dodring V and Schimmel P, "Genetic code ambiguity — cell viability related to severity of editing defects in mutant tRNA synthetases," *J Biol Chem*. **2002**, *277*:45729.

[12] Heckler TG, Chang LH, Zama Y, Naka T, Chorghade MS and Hecht SM, "T4 RNA ligase mediated preparation of novel chemically misacylated transfer-RNA^{Phe}s," *Biochemistry*. **1984**, *23*:1468.

[13] Bain JD, Glabe CG, Dix TA, Chamberlin AR and Diala ES, "Biosynthetic site-specific incorporation of a non-natural amino-acid into a polypeptide," *J Am Chem Soc*. **1989**, *111*:8013.

[14] Noren CJ, Anthonycahill SJ, Griffith MC and Schultz PG, "A general-method for site-specific incorporation of unnatural amino-acids into proteins," *Science*. **1989**, *244*:182.

[15] Wang P, "Expanding the biosynthetic capacity of the aminoacyl-tRNA synthetases," *PhD Thesis,* California Institute of Technology (Pasadena, CA). **2003**.

[16] Hohsaka T and Sisido M, "Incorporation of non-natural amino acids into proteins," *Curr Opin Chem Biol*. **2002**, *6*:809.

[17] Dougherty DA, "Unnatural amino acids as probes of protein structure and function," *Curr Opin Chem Biol*. **2000**, *4*:645.

[18] Cornish VW, Mendel D and Schultz PG, "Probing protein-structure and function with an expanded genetic-code," *Angew Chem Int Ed Engl*. **1995**, *34*:621.

[19] Gilmore MA, Steward LE and Chamberlin AR, "Incorporation of noncoded amino acids by *in vitro* protein biosynthesis" in "*Implementation and redesign of catalytic function in biopolymers*," Ed., **1999**; 77

[20] Nowak MW, Kearney PC, Sampson JR, Saks ME, Labarca CG, Silverman SK, Zhong W, Thorson J, Abelson JN, Davidson N, Schultz PG, Dougherty DA and Lester HA, "Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells," *Science*. **1995**, *268*:439.

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

[22] England PM, Zhang Y, Dougherty DA and Lester HA, "Backbone mutations in transmembrane domains of a ligand-gated ion channel: Implications for the mechanism of gating," *Cell*. **1999**, *96*:89.

[23] Lummis SC, Beene DL, Lee LW, Lester HA, Broadhurst RW and Dougherty DA, "Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel," *Nature*. **2005**, *438*:248.

[24] Gilmore MA, Steward LE and Chamberlin AR, "Incorporation of noncoded amino acids by *in vitro* protein biosynthesis" in "*Implementation and redesign of catalytic function in biopolymers*," Schmidtchen FP, Ed., Springer: Berlin, **1999**; 77

[25] McGrath KP, Tirrell DA, Kawai M, Mason TL and Fournier MJ, "Chemical and biosynthetic approaches to the production of novel polypeptide materials," *Biotechnol. Prog*. **1990**, *6*:188.

[26] Fournier MJ, Creel HS, Krejchi MT, Mason TL, Tirrell DA, McGrath KP and Atkins EDT, "Genetic synthesis of periodic protein materials," *J Bioact Compat Pol*. **1991**, *6*:326.

[27] Krejchi MT, Atkins ED, Waddon AJ, Fournier MJ, Mason TL and Tirrell DA, "Chemical sequence control of beta-sheet assembly in macromolecular crystals of periodic polypeptides," *Science*. **1994**, *265*:1427.

[28] Yu SJM, Conticello VP, Zhang GH, Kayser C, Fournier MJ, Mason TL and Tirrell DA, "Smectic ordering in solutions and films of a rod-like polymer owing to monodispersity of chain length," *Nature*. **1997**, *389*:167.

[29] Yu SM and Tirrell DA, "Thermal and structural properties of biologically derived monodisperse hairy-rod polymers," *Biomacromolecules*. **2000**, *1*:310.

[30] Petka WA, Harden JL, McGrath KP, Wirtz D and Tirrell DA, "Reversible hydrogels from self-assembling artificial proteins," *Science*. **1998**, *281*:389.

[31] Di Zio K and Tirrell DA, "Mechanical properties of artificial protein matrices engineered for control of cell and tissue behavior," *Macromolecules*. **2003**, *36*:1553.

[32] Welsh ER and Tirrell DA, "Engineering the extracellular matrix: A novel approach to polymeric biomaterials. I. Control of the physical properties of

artificial protein matrices designed to support adhesion of vascular endothelial cells," *Biomacromolecules*. **2000**, *1*:23.

[33] Panitch A, Yamaoka T, Fournier MJ, Mason TL and Tirrell DA, "Design and biosynthesis of elastin-like artificial extracellular matrix proteins containing periodically spaced fibronectin cs5 domains," *Macromolecules*. **1999**, *32*:1701.

[34] Dittmer K, Goering HL, Goodman I and Cristol SJ, "The inhibition of microbiological growth by allylglycine, methallylglycine and crotylglycine," *J Am Chem Soc*. **1948**, *70*:2499.

[35] Rennert OM and Anker HS, "On incorporation of 5',5',5'-trifluoroleucine into proteins of *E. coli*," *Biochemistry*. **1963**, *2*:471.

[36] Budisa N, Minks C, Alefelder S, Wenger W, Dong FM, Moroder L and Huber R, "Toward the experimental codon reassignment *in vivo*: Protein building with an expanded amino acid repertoire," *Faseb Journal*. **1999**, *13*:41.

[37] van Hest JCM and Tirrell DA, "Efficient introduction of alkene functionality into proteins *in vivo*," *FEBS Lett*. **1998**, *428*:68.

[38] van Hest JCM, Kiick KL and Tirrell DA, "Efficient incorporation of unsaturated methionine analogues into proteins *in vivo*," *J Am Chem Soc*. **2000**, *122*:1282.

[39] Kiick KL, Saxon E, Tirrell DA and Bertozzi CR, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proc Natl Acad Sci U S A*. **2002**, *99*:19.

[40] Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF and Tirrell DA, "Fluorinated coiled-coil proteins prepared *in vivo* display enhanced thermal and chemical stability," *Angew Chem Int Ed Engl*. **2001**, *40*:1494.

[41] Dessipri E, Tirrell DA and Atkins EDT, "Fluorinated poly(alpha, l-glutamate)s," *Macromolecules*. **1996**, *29*:3545.

[42] Yoshikawa E, Fournier MJ, Mason TL and Tirrell DA, "Genetically-engineered fluoropolymers — synthesis of repetitive polypeptides containing p-fluorophenylalalanine residues," *Macromolecules*. **1994**, *27*:5471.

[43] Datta D, Wang P, Carrico IS, Mayo SL and Tirrell DA, "A designed phenylalanyl-tRNA synthetase variant allows efficient *in vivo* incorporation of aryl ketone functionality into proteins," *J Am Chem Soc*. **2002**, *124*:5652.

[44] Hendrickson WA, Horton JR and LeMaster DM, "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (mad): A vehicle for direct determination of three-dimensional structure," *EMBO J*. **1990**, *9*:1665.

[45] Bae JH, Alefelder S, Kaiser JT, Friedrich R, Moroder L, Huber R and Budisa N, "Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein x-ray crystallography," *J Mol Biol*. **2001**, *309*:925.

[46] Eichler JF, Cramer JC, Kirk KL and Bann JG, "Biosynthetic incorporation of fluorohistidine into proteins in *E. coli*: A new probe of macromolecular structure," *Chembiochem*. **2005**, *6*:2170.

[47] Wang P, Tang Y and Tirrell DA, "Incorporation of trifluoroisoleucine into proteins *in vivo*," *J Am Chem Soc*. **2003**, *125*:6900.

[48] Tang Y and Tirrell DA, "Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host," *J Am Chem Soc*. **2001**, *123*:11089.

[49] Tang Y, Ghirlanda G, Vaidehi N, Kua J, Mainz DT, Goddard III WA, DeGrado WF and Tirrell DA, "Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine," *Biochemistry*. **2001**, *40*:2790.

[50] Bilgicer B, Xing X and Kumar K, "Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores," *J Am Chem Soc*. **2001**, *123*:11815.

[51] Bilgicer B, Fichera A and Kumar K, "A coiled coil with a fluorous core," *J Am Chem Soc*. **2001**, *123*:4393.

[52] Son S, Tanrikulu IC and Tirrell DA, "Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues," *Chembiochem*. **2006**, *7*:1251.

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

[54] Lee HY, Lee KH, Al-Hashimi HM and Marsh ENG, "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*. **2006**, *128*:337.

[55] Bae JH, Rubini M, Jung G, Wiegand G, Seifert MH, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R and Budisa N, "Expansion of the genetic code enables design of a novel "Gold" Class of green fluorescent proteins," *J Mol Biol*. **2003**, *328*:1071.

[56] Zhang K, Diehl MR and Tirrell DA, "Artificial polypeptide scaffold for protein immobilization," *J Am Chem Soc*. **2005**, *127*:10136.

[57] Carrico IS, Maskarinec SA, Heilshorn SC, Mock ML, Liu JC, Nowatzki PJ, Franck C, Ravichandran G and Tirrell DA, "Lithographic patterning of photoreactive cell-adhesive proteins," *J Am Chem Soc*. **2007**, *129*:4874.

[58] Beatty KE, Liu JC, Xie F, Dieterich DC, Schuman EM, Wang Q and Tirrell DA, "Fluorescence visualization of newly synthesized proteins in mammalian cells," *Angew Chem Int Ed Engl*. **2006**, *45*:7364.

[59] Beatty KE, Xie F, Wang Q and Tirrell DA, "Selective dye-labeling of newly synthesized proteins in bacterial cells," *J Am Chem Soc*. **2005**, *127*:14150.

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

[61] Kirshenbaum K, Carrico IS and Tirrell DA, "Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues," *Chembiochem*. **2002**, *3*:235.

[62] Sharma N, Furter R, Kast P and Tirrell DA, "Efficient introduction of aryl bromide functionality into proteins *in vivo*," *FEBS Lett*. **2000**, *467*:37.

[63] Baskin JM and Bertozzi CR, "Bioorthogonal click chemistry: Covalent labeling in living systems," *Qsar Comb Sci*. **2007**, *26*:1211.

[64] Agard NJ, Prescher JA and Bertozzi CR, "A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of blomolecules in living systems," *J Am Chem Soc*. **2004**, *126*:15046.

[65] Kiick KL, van Hest JCM and Tirrell DA, "Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host," *Angew Chem Int Ed Engl*. **2000**, *39*:2148.

[66] Doring V, Mootz HD, Nangle LA, Hendrickson TL, de Crecy-Lagard V, Schimmel P and Marliere P, "Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway," *Science*. **2001**, *292*:501.

[67] Tang Y and Tirrell DA, "Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins *in vivo*," *Biochemistry*. **2002**, *41*:10635.

[68] Ibba M, Kast P and Hennecke H, "Substrate specificity is determined by amino acid binding pocket size in *Escherichia coli* phenylalanyl transfer RNA synthetase," *Biochemistry*. **1994**, *33*:7107.

[69] Hamano-Takaku F, Iwama T, Saito-Yano S, Takaku K, Monden Y, Kitabatake M, Soll D and Nishimura S, "A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine," *J Biol Chem*. **2000**, *275*:40324.

[70] Ibba M and Hennecke H, "Relaxing the substrate specificity of an aminoacyl transfer RNA synthetase allows in vitro and *in vivo* synthesis of proteins containing unnatural amino acids," *Febs Letters*. **1995**, *364*:272.

[71] Bentin T, Hamzavi R, Salomonsson J, Roy H, Ibba M and Nielsen PE, "Photoreactive bicyclic amino acids as substrates for mutant *Escherichia coli* phenylalanyl-tRNA synthetases," *J Biol Chem*. **2004**, *279*:19839.

[72] Kwon I and Tirrell DA, "Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells," *J Am Chem Soc*. **2007**, *129*:10431.

[73] Kim W, George A, Evans M and Conticello VP, "Cotranslational incorporation of a structurally diverse series of proline analogues in an *Escherichia coli* expression system," *Chembiochem*. **2004**, *5*:928.

[74] Wang L and Schultz PG, "Expanding the genetic code," *Chem. Comm*. **2002**, 1.

[75] Wang L, Brock A, Herberich B and Schultz PG, "Expanding the genetic code of *Escherichia coli*," *Science*. **2001**, *292*:498.

[76] Xie J and Schultz PG, "A chemical toolkit for proteins — an expanded genetic code," *Nat Rev Mol Cell Biol*. **2006**, *7*:775.

[77] Wang L and Schultz PG, "Expanding the genetic code," *Angew Chem Int Ed Engl*. **2005**, *44*:34.

[78] Wang L, Xie J and Schultz PG, "Expanding the genetic code," *Annu Rev Biophys Biomol Struct*. **2006**, *35*:225.

[79] Zhang ZW, Gildersleeve J, Yang YY, Xu R, Loo JA, Uryu S, Wong CH and Schultz PG, "A new strategy for the synthesis of glycoproteins," *Science*. **2004**, *303*:371.

[80] Wu N, Deiters A, Cropp TA, King D and Schultz PG, "A genetically encoded photocaged amino acid," *J Am Chem Soc*. **2004**, *126*:14306.

[81] Tippmann EM and Schultz PG, "A genetically encoded metallocene containing amino acid," *Tetrahedron*. **2007**, *63*:6182.

[82] Lemke EA, Summerer D, Geierstanger BH, Brittain SM and Schultz PG, "Control of protein phosphorylation with a genetically encoded photocaged amino acid," *Nat Chem Biol*. **2007**, *3*:769.

[83] Summerer D, Chen S, Wu N, Deiters A, Chin JW and Schultz PG, "A genetically encoded fluorescent amino acid," *Proc Natl Acad Sci U S A*. **2006**, *103*:9785.

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

[85] Link AJ, Vink MK, Agard NJ, Prescher JA, Bertozzi CR and Tirrell DA, "Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids," *Proc Natl Acad Sci U S A*. **2006**, *103*:10180.

[86] Yoo TH and Tirrell DA, "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 Engl*. **2007**, *46*:5340.

[87] Kekenes-Huskey PM, Vaidehi N, Floriano WB and Goddard WA, "Fidelity of phenylalanyl-tRNA synthetase in binding the natural amino acids," *J Phys Chem B*. **2003**, *107*:11549.

[88] Datta D, Vaidehi N, Zhang DQ and Goddard WA, "Selectivity and specificity of substrate binding in methionyl-tRNA synthetase," *Protein Sci*. **2004**, *13*:2693.

[89] McClendon CL, Vaidehi N, Kam VWT, Zhang DQ and Goddard WA, "Fidelity of seryl-tRNA synthetase to binding of natural amino acids from hierdock first principles computations," *Protein Eng Des Sel*. **2006**, *19*:195.

[90] Wang P, Vaidehi N, Tirrell DA and Goddard WA, "Virtual screening for binding of phenylalanine analogues to phenylalanyl-tRNA synthetase," *J Am Chem Soc*. **2002**, *124*:14442.

[91] Dahiyat BI and Mayo SL, "De novo protein design: Fully automated sequence selection," *Science*. **1997**, *278*:82.

[92] Street AG and Mayo SL, "Computational protein design," *Structure Fold Des*. **1999**, *7*:R105.

[93] Zhang DQ, Vaidehi N, Goddard WA, Danzer JF and Debe D, "Structure-based design of mutant methanococcus jannaschii tyrosyl-tRNA synthetase for incorporation of o-methyl-l-tyrosine," *Proc Natl Acad Sci U S A*. **2002**, *99*:6579.

[94] Zhang DQ, "Structure-based design of mutant proteins," *PhD Thesis,* California Institute of Technology (Pasadena, CA). **2002**.

Chapter 2

Discovery of *E. coli* Methionyl-tRNA Synthetase Mutants for Efficient *In Vivo* Labeling of Proteins with Azidonorleucine

Introduction

Amino acids that carry noncanonical side chains have become useful protein engineering tools in the last decade. Their uses include stabilization of protein-protein interfaces [1-3], modification of spectral properties of proteins [4, 5], and determination of their three-dimensional structures [6-8]. The value and versatility of noncanonical amino acids has significantly increased by their use as bioorthogonal reactive handles for conjugation [9, 10], detection [11, 12] and selective isolation [13] of proteins.

Among the reactive groups introduced into biological molecules, azides are perhaps the most significant [14]. Azides can be ligated under physiological conditions to ester-functionalized triaryl phosphines through the Staudinger ligation or to terminal alkynes through Cu(I)-catalyzed [3+2] azide-alkyne cycloaddition, both in bioorthogonal fashion. Taking advantage of the bioorthogonality of these reactions, azides have been used to selectively label a diverse set of biomolecules including proteins [15], glycans [16] and lipids [17, 18].

Through introduction of the azide-bearing methionine (Met) surrogate azidohomoalanine (Aha) at methionine sites in proteins *in vivo*, Link and co-workers have demonstrated the cell-surface labeling of azide groups displayed on the outer membrane of *E. coli* [19]. This strategy depends on the recognition of the methionine surrogate by the endogenous methionyl-tRNA sytnthetase (MetRS), which catalyzes the charging of the amino acid onto the associated tRNA, $tRNA^{Met}$. Aha is one of the best surrogates for methionine and is known to be activated *in vivo* with only wild-type levels of MetRS activity present [15]. This makes Aha a very attractive chemical handle for labeling proteins *in vivo*. Using bioorthogonal noncanonical amino acid tagging (BONCAT), Dieterich and co-workers identified newly synthesized proteins in mammalian cells through a pulse treatment with Aha [13]. They were able to separate the azide-bearing proteins synthesized during the Aha pulse from the pool of preexisting proteins in the cell through a Cu(I)-catalyzed ligation to alkyne-derivatized biotin and subsequent affinity chromatography.

Although Aha is ideal for identification of newly synthesized proteins in cultures containing a single cell type, because it is such a good substrate of wild-type MetRS, it does not allow directing the labeling to only proteins in certain members of a complex cell mixture. Alternatively, using a reactive amino acid that requires an engineered MetRS for incorporation, labeling can be directed to individual cells that carry this engineered enzyme. In order to allow incorporation of reactive amino acids that cannot be activated by the wild-type cellular machinery, Link and co-workers devised a fluorescence-activated cell sorter (FACS) based high-throughput screen and demonstrated its utility by identifying *E. coli* MetRS mutants that can activate the methionine surrogate, azidonorleucine (Anl) from a four-position saturation-mutagenesis library [20]. (Figure 2.1) Cells carrying MetRS mutants that are active toward Anl can display this residue on their surface, on an *E. coli* outer-membrane protein C (OmpC) variant. (Figure 2.2) Using a strain-promoted version of the [3+2] azide-alkyne cycloaddition [21], which eliminates the copper cytotoxicity associated with the Cu(I)-catalyzed reaction, the azides displayed on cells were ligated to biotin and bound to fluorescent avidin. Clones carrying active MetRS mutants were then isolated with FACS, and screened further. Analysis of enriched clones revealed three MetRS mutants that allowed approximately 50% incorporation of Anl into Met sites when protein expression was carried out in the presence of 8 mM Anl. Recognizing that the L13G mutation was common to all MetRS mutants identified through the screen, the investigators tested this single mutant to discover that this enzyme allows near complete replacement of Met sites at 1 mM Anl.

In this study, we repeated the screen carried out for Anl by Link and co-workers, exploring new approaches to improve its performance. Using a smaller, three-position saturation-mutagenesis library and an increased number of clones, the complete coverage of the sequence space by the library was ensured. By optimizing expression conditions to minimize the toxicity associated with OmpC overexpression, we were able to achieve enrichment of a variety of MetRS mutants with greatly enhanced activities toward Anl, and an improved discrimination against Met. Here we report the screening and the

characterization of these MetRS variants, and discuss the structural basis for selection of these mutants.

Materials and Methods

Chemical reagents

Azidonorleucine [22] was received as a gift from A. J. Link and M. B. van Eldijk. Biotin-PEO-cyclooctyne [21] was a gift from J. D. Fisk. Reagents were prepared as described in the corresponding references.

Plasmids, cell strains and cloning reagents

T4 DNA ligase and all restriction enzymes were purchased from New England Biolabs. *E. coli* XL-1 Blue (Stratagene) cells were used for all recombinant DNA manipulations, unless noted otherwise. The methionine auxotroph M15MA [19] was employed for overexpression of proteins in synthetic media containing Anl. ElectroTen (Stratagene) cells were used as the cloning host for library construction. All DNA sequencing was performed by Laragen, Inc., or the Caltech DNA sequencing facility.

The plasmid pAJL-20 was used as a template for the construction of the MetRS library as was previously described [20]. This plasmid, based on pQE-60 (Qiagen), encodes a variant of *E. coli* outer-membrane protein C (OmpC) that carries six surface-exposed methionine mutations [19] under the control of the isopropyl-β-thiogalactoside (IPTG)-inducible T5 promoter. In addition, pAJL-20 carries a cassette encoding a monomeric version of *E. coli* methionyl-tRNA synthetase (MetRS) truncated at residue 548 [23] under the control of its natural promoter. The plasmid pAJL-61 was used to test incorporation of Anl into the N-terminally 6×His-tagged dihydrofolate reductase (DHFR) [20]. This plasmid is based on pQE-80L (Qiagen) and carries the same MetRS-casette found on pAJL-20 as well as a copy of the gene for the *lac* repressor protein $lacI^q$. It also directs the synthesis of an N-terminally 6×His-tagged DHFR under an IPTG-inducible T5 promoter.

The plasmid pMTY21 was constructed by ligating the BamHI/SalI digested

fragment of pQE-80L and a similarly digested MetRS gene PCR-amplified from a pAJL-20 or a pAJL-61 construct using the primers MRS_BamHI and MRS_SalI-r [24]. This plasmid codes for a N-terminally 6×His-tagged MetRS for its purification and characterization

Methionyl-tRNA synthetase (MetRS) library construction

Oligonucleotides encoding degenerate NNK (N = A, T, G, C; K = G, T) codons at the sites corresponding to Leu13, Tyr260, and His301 in *E. coli* MetRS were obtained from Operon. Four separate PCR reactions were performed using PfuUltra HighFidelity polymerase (Stratagene), pAJL-20 as a template and a pair of primers. A series of primer pairs were tested for their PCR product yields, and the following pairs were used to create the final library: lib_fwd2 and L13_lib2-r, L13_lib3 and TYR260_lib3-r, TYR260_lib3 and HIS301_lib3-r, HIS301_lib2 and lib_rev2. Sequences for all primers can be found in Table 2.1. The DNA fragments obtained from these PCR reactions were electrophoresed in agarose gels and purified using Zymo-spin columns (Zymo Research), paying extra attention to limit the UV exposure of the DNA. Approximately equimolar quantities of the fragments were mixed and subjected to 20 to 30 rounds of PCR (95 ºC for 30 sec, 58 ºC for 30 sec, 72 ºC for 2 min) using the PfuUltra HF polymerase. To amplify the assembled product, the primers lib_fwd3 and lib_rev3 were added to the reaction mixture and 20 to 30 more rounds of PCR were carried out using the parameters described above. The resulting 2.0 kb PCR product was digested with *Not* I and *Bsr* GI to reveal a 1.35 kb insert, which was subsequently ligated into pAJL-20 digested with the same enzymes using an insert:vector molar ratio of 3:2. The ligation mixture was transformed into electrocompetent ElectroTen (Stratagene) cells yielding more than 10^7 independent transformants. The plasmid DNA from the pooled transformants was isolated via Miniprep columns (Qiagen), and electroporated into the methionine auxotroph M15MA bearing pREP4 (M15MA[pREP4]) to yield more than 10^7 independent clones, referred to as the LYH.1.0 population. Aliquots of cultures containing 1 mL of the pooled transformants at OD_{600} = 1.0 were stored

at −80 ºC in 50% glycerol until protein expression.

OmpC overexpression

The expression of the OmpC mutant carrying azidonorleucine was performed as outlined previously [20] with few modifications. M15MA[pREP4/pAJL-20] cells were inoculated into M9 medium (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 25 mg/L thiamine) supplemented with 40 mg/L of each of the 20 canonical amino acids, 200 mg/L ampicillin and 35 mg/L kanamycin (referred to as M9+20aa) either 1:100 from cultures grown overnight in 2xYT medium, or from frozen stocks carrying the MetRS library. Cultures of size 4 mL to 20 mL were used for the testing and characterization of mutants. During library screening 30 mL cultures were used per 1 mL aliquot of the library. The cells were grown to an OD_{600} between 0.8 and 1.0, at which point a medium shift was performed to remove methionine from the medium. Cells were pelleted at 5000 × g for 7 min, resuspended in an equal volume of M9 medium supplemented with 19 amino acids (no methionine; referred to as M9+19aa), and incubated with shaking for 15 min at 37 °C to deplete any residual methionine. Following this incubation, the cells were pelleted and resuspended in fresh M9+19aa medium, divided into multiple vessels, and supplemented with methionine, azidonorleucine or no methionine surrogate. Methionine was commonly introduced at a final concentration of 0.3 mM and azidonorleucine at 1.0 mM. Expression of OmpC was initiated with the addition of IPTG to a 1 mM final concentration, and induced for 30 to 45 min. Following expression, cell densities were adjusted to OD_{600} = 1.0, and a 1 mL aliquot of each culture was pelleted and washed with sterile phosphate-buffered saline (PBS, pH 7.4) in preparation for cell-surface labeling.

Cell-surface labeling

Following OmpC expression, the washed cells were treated with 100 μM biotin-PEO-cyclooctyne with agitation for 16 hours at 37 °C. Biotin-labeled cells

were cooled to 4 °C and washed twice with 1 mL of PBS. The cells were then treated with 2.5 µL of a 1 mg/mL solution of an avidin-Alexa Fluor 488 conjugate (Molecular Probes) for 2 hours at 4 ºC with agitation. The cells were washed three more times with PBS to remove the excess fluorophore, and labeling was confirmed using a Safire II (Tecan) plate reader or through flow cytometry.

In order to determine the number of live cells after OmpC expression and labeling, the density of labeled cells was adjusted to $OD_{600} = 0.2$. Cells were further diluted 1 to 10^4 and 1 to 2.5×10^5 by PBS, and 100 µL from each dilution was plated onto 2xYT agar supplemented with 200 mg/L ampicillin and 35 mg/L kanamycin. The plates were incubated at 37 °C overnight, and the number of colonies on each plate was determined.

Flow cytometry and cell sorting

All flow cytometric analyses were carried out on a MoFlo (DakoCytomation) fluorescence activated cell sorter (FACS) equipped with an argon ion laser emitting at 488 nm. The initial sorting on the naïve library was performed in sort purify mode. For all subsequent sorts the sort single mode was used. Sort gates were set and data analysis was performed with Summit software (DakoCytomation). Following incorporation of AnI into OmpC and cell-surface labeling, highly fluorescent cells from each library were sorted by setting a gate in the fluorescence channel corresponding to the top 0.8% to 1.2% of the cells. Additional gates were set in the forward- and side-scatter channels to avoid any events of unusual size. The gate on the fluorescence channel was set especially high for libraries expressing OmpC at low concentrations of AnI, letting as little as the top 0.5% of the events to be saved. In a typical experiment, 10^8 events were interrogated, of which 5×10^5 were taken to the following round. The sorted cells were rescued in 15 mL of SOC medium for 30 min. At this time the medium was supplemented with 200 mg/L ampicillin and 35 mg/L kanamycin, and an aliquot of the rescued cells was plated on 2xYT agar supplemented with the same antibiotics for analysis of individual clones. Cell densities were adjusted to $OD_{600} = 1.0$ and cultures were either stored at −80 ºC in 50% glycerol,

or immediately carried through to the next round of selection.

The dose-response relationship between the azidonorleucine concentrations in media during the OmpC expression and the level of fluorescence observed on FACS after cell-surface labeling was determined as follows. Transformed M15MA[pREP4/pAJL-20] cells bearing a MetRS mutant were prepared for OmpC expression as discussed above. OmpC expression was induced in M9+19aa medium supplemented with azidonorleucine to final concentrations of 0.03, 0.1, 0.3, 1.0 and 3.0 mM. Following cell-surface labeling, FACS histograms were obtained for cells from each culture, and the median fluorescence from each population was determined. Resulting data were fit to the Hill equation using KaleidaGraph (v3.6, Synergy Software) after setting the minimum response to 1.60 fluorescence units. The EC50 values were obtained from the resulting least squares fit.

Recombinant DHFR expression, purification, and analysis

Expression of dihydrofolate reductase (DHFR) containing Anl was performed in M15MA[pREP4/pAJL-61] cells, following the same protocol explained above for OmpC expression, except cells were induced for 3 to 3.5 hours with IPTG. Variants of the pAJL-61 plasmid carrying different MetRS mutants were generated using QuikChange site-directed mutagenesis. (See Table 2.1 for a list of primers) Integrity of all constructs was verified by DNA sequencing. Expressions were carried out in 5 or 10 mL cultures. Following protein expression, aliquots standardized for OD_{600} were saved for SDS-PAGE analysis, and the remainder was pelleted and resuspended in 8 M urea. DHFR was purified under denaturing conditions on Ni-NTA spin columns (Qiagen) according to the manufacturer's suggestions. For MALDI-MS analysis, 15 μL of the eluents in 8 M urea was diluted with 95 μL of 75 mM NH_4HCO_3, and digested with 0.2 μg of porcine trypsin (Promega). After 4 hours, the resulting peptide mixtures were desalted using C18 ZipTip (Millipore) columns. Eluted protein mixtures were diluted 3-fold with a matrix solution (saturated α-cyano hydroxycinnamic acid (Fluka) in 1:1 water:acetonitrile and 0.1% trifluoroacetic

acid) and subjected to MALDI-MS analysis on a Voyager MALDI-TOF mass spectrometer (Applied Biosystems).

MetRS expression, purification, and *in vitro* activation assays

Variants of the pMTY21 plasmid carrying different MetRS mutants were constructed either by the PCR amplification and ligation into pMTY21 of a MetRS gene from a pAJL-20 or a pAJL-61 plasmid (discussed above), or by QuikChange site-directed mutagenesis. (See Table 2.1 for primer sequences.) For every construct, the full length of the MetRS gene was verified to be error free through DNA sequencing. *E. coli* XL-1 Blue cells transformed with pMTY21 were inoculated 1:100 from an overnight culture into 100 mL of Terrific Broth. Cells were grown to OD_{600} = 1.0 at 37 °C, at which point they were induced with IPTG (1 mM final concentration) to produce MetRS for 18 hours at 25 °C. Cells were harvested by centrifugation (6000 × g for 15 min). Histidine-tagged MetRS mutants were isolated under native conditions using the Ni-NTA agarose (Qiagen) resin according to the manufacturer's instructions. Eluted product was buffer exchanged with 100 mM Tris buffer (pH 7.5) containing 2 mM DTT on PD-10 columns (GE Healthcare). The eluent was added to an equal mass of glycerol and mixed thoroughly, and the enzyme stocks prepared were stored at −80 °C until needed. The concentration of enzyme in each stock was determined by measuring the absorbance at 280 nm of the enzyme stock 1:9 in 9 M urea, and assuming an extinction coefficient of 93,280 $M^{-1}cm^{-1}$ for the MetRS mutant as calculated by the ProtParam tool (http://expasy.org/tools/protparam.html).

Amino acid activation assays were carried out as described previously [25]. Radiolabeled sodium ^{32}P-pyrophosphate was purchased from Perkin-Elmer Life Sciences. An enzyme concentration in the range 1–4 μM was used in each reaction, and activation rates were measured at 0.031 to 16 mM Met and 0.25 to 64 mM Anl. All data was fit to a Michaelis-Menten model using KaleidaGraph (v3.6, Synergy Software), and kinetic parameters were determined.

Results and Discussion

To construct a saturation-mutagenesis library, Link et al. randomized four positions in the methionine binding pocket of *E. coli* MetRS: L13, P257, Y260 and H301 [20]. Through screening, three 4-fold mutants of MetRS that enable the *in vivo* incorporation of Anl into proteins were identified: L13G-P257L-Y260T-H301A, L13G-P257S-Y260T-H301L and L13G-P257L-Y260L-H301V. Recognizing that all isolated mutants share the L13G mutation, the authors also examined this single mutant and discovered that it allows near complete replacement of Met by Anl at much lower substrate concentrations (1 mM) than was required for previously identified mutants (~50% incorporation at 8 mM Anl). All mutants identified in this screen, including the L13G mutant, were members of the original saturation-mutagenesis library. It is therefore puzzling that none of the mutants revealed directly by the screen can match the activity of the L13G mutant. In order to identify why mutants of higher activity were not identified by the screen, we focused on two issues: 1) coverage of the saturation mutagenesis library, and 2) cell toxicity due to OmpC expression.

Construction of a MetRS saturation mutagenesis library

The absence of the L13G mutant or any other mutant with similar activity from the screening results can be due a poor library coverage. In order to ensure complete coverage of the saturation mutagenesis library, we first aimed to decrease the number of positions randomized on the MetRS from the four in the original screen (L13, P257, Y260 and H301) to three. (Figure 2.3) Examination of the crystal structure of the methionine-bound *E. coli* MetRS [26] reveals that the methionine sulfur atom is recognized through hydrogen bonds with $O\eta$ of Y260 and the backbone N of Leu13. The residue H301 is also implicated in the recognition of the methionine sulfur atom [27]. Due to their importance in the recognition of methionine, these three residues (L13, Y260 and H301) were selected to be randomized in the library. A previous computational design study (Table 3.4) that included the P257 position resulted in only very conservative mutations at this position. Problems were also recognized during the expression

and purification of 4-fold mutants revealed by the previous screen [28]. We suspected that this could be due to the destabilization of the protein through a mutation away from proline at this position, which prompted us to leave the P257 site unchanged in our library.

The library was constructed through PCR gene assembly using primers containing degenerate NNK (N=A,C,G,T, K=G,T) codons. The use of NNK codons allow all twenty amino acids to be represented at the randomized sites, while excluding two of the three stop codons. The assembled fragments were then ligated into the plasmid pAJL-20. We found that limiting the UV exposure of DNA during visualization and gel purification is of critical importance for the successful ligation of the library. With this method 4.5×10^7 independent transformants of the library were obtained in the cloning host and 1.3×10^7 transformants in the expression host. The library coverage was estimated to be better than 99.9% using the program GLUE [29], which assumes an unbiased library where each member is equally likely to occur, compared with a 61% coverage for the previous four-position saturation-mutagenesis library [20].

OmpC overexpression and cell-surface labeling

The plasmid pAJL-20 directs the synthesis of a variant of *E. coli* OmpC carrying six surface-exposed methionine residues. When this synthesis is induced in conditions where Anl is supplied and Met is depleted through the use of synthetic media and a Met-auxotrophic host, the surface-exposed Met residues on OmpC are replaced with Anl only if the cell harbors a MetRS mutant able to accept Anl as a substrate. Azide groups exposed on OmpC can then be ligated to biotin through a strain-promoted azide-alkyne ligation with biotin-PEO-cyclooctyne [21]. Binding fluorescent avidin to the biotin-labeled cells makes these cells fluorescent, which can be read as a measure of Anl incorporation.

An important limitation to this screening system is that OmpC expression has an adverse effect on cell viability [30]. It has been long known that overexpression of proteins in *E. coli* puts a great amount of stress on the host, inhibiting its growth and triggering a starvation response, which may result in cell

death [31]. Membrane proteins, such as OmpC, are especially toxic to the cells when they are overexpressed since their overproduction blocks protein translocation pathways and greatly diminishes the protein folding capacity of the cell [32]. It was recently shown that overexpression of membrane proteins drastically changes the proteome of the cell, both in the cytoplasm and in the cell envelope, significantly compromises cellular respiration and inhibits cell division [33].

For the screen to work effectively, tight coupling is necessary between Anl incorporation in the cytoplasm and OmpC synthesis and display on the cell surface. This way, a cell carrying a mutant MetRS that is highly active toward Anl will synthesize high levels of OmpC and will display more azide-bearing OmpC on its surface than cells carrying poorer mutants. However, if high levels of OmpC expression inhibit growth of the host, clones carrying highly active MetRS mutants will have a significant disadvantage against inferior clones, and their enrichment through screening will be challenging. Such behavior can explain the poor screening results previously obtained [20], since long induction times for OmpC expression combined with a high concentration of Anl in the media would adversely affect the proliferation of any clones able to utilize Anl efficiently.

In order to determine how the duration of OmpC expression affects cell viability and labeling efficiency, we expressed OmpC in the Met-auxotroph M15MA bearing the MetRS-L13G mutant in the presence of 19 natural amino acids (40 mg/L; –Met) and 1 mM Anl. Aliquots were taken from the culture at 0.5, 1, 2, and 4 hours after induction of OmpC expression. The surface-exposed azide groups on the cells were conjugated to biotin through a 16-hour incubation with biotin-PEO-cyclooctyne at 37 ºC. The cells were labeled with fluorescent avidin, and their fluorescence was measured by a fluorescence plate reader and a flow cytometer. Following labeling, a portion of the labeled cells was grown on agar plates to see how cell viability was affected by OmpC expression.

Expression of OmpC causes a dramatic drop in cell viability, as shown in Figure 2.4. For cells expressing OmpC, the number of viable clones decreases

by an order of magnitude for every 0.5 hours of OmpC expression. This is in stark contrast to cells overexpressing the globular protein DHFR, which experience a similar drop in viable clones over a 4-hour period of protein expression. These results suggest a strong selection against cells overexpressing OmpC.

The extent of cell-surface labeling also does not improve with increasing duration of OmpC expression. (Figure 2.5.a) Most of the gain in fluorescence labeling occurs in the first half hour of OmpC induction and the labeling levels do not change significantly after one hour of induction. However, a steady increase in background signal is observed with increasing OmpC expression from cells that were not treated with Anl. This trend is also observed when the cultures are examined on a flow cytometer. (Figure 2.5.b) Even though there is minimal fluorescence gain in Anl-treated cells between 1 hour and 4 hours of expression, a shift to higher fluorescence is evident in the Met-treated cells. By calculating the fluorescence enhancement due to the Anl treatment, the duration of OmpC expression optimal for fluorescence labeling was determined to be between 30 and 60 min. (Figure 2.5.c) In the light of these results, OmpC expression was limited to 30 to 45 min for all subsequent experiments.

Screening and identification of active mutants

The MetRS library was screened for the ability to efficiently use Anl in protein synthesis as outlined in Figure 2.2. The library in the plasmid pAJL-20 was transformed into M15MA cells (LYH.1.0; Figure 2.6), and grown in minimal medium containing all 20 canonical amino acids (M9+20aa). Before induction of OmpC expression, cells were shifted into minimal medium lacking methionine (M9+19aa) and supplemented with 1.0 mM Anl. In prior experiments cells had been treated with 8 mM Anl at this step, but having the knowledge of the L13G mutant and its activity, we opted for a lower concentration. Cells were induced to produce OmpC for 35 min, ligated to biotin and fluorescently labeled with fluorescent avidin. Labeled cells were analyzed by flow cytometry, and the top 1% of fluorescent events were sorted and regrown in selective rich medium to

reveal the population LYH.1.1a. (Figure 2.6., left column) An aliquot from the sorted cells was plated for identification of individual clones. The population LYH.1.1a was fluorescently labeled following another round of OmpC expression in Anl-supplemented medium. FACS analysis revealed that 2/3 of the cells in this population showed high levels of labeling, confirming that the enrichment was successful. This population was subjected to another round of screening, where the top 1% of the fluorescent events in LYH.1.1a were sorted to create the LYH.1.2 population. Since the fluorescence profile of the population did not change between the last two rounds, we continued on to identify the makeup of the LYH.1.2 population.

The mutations identified in prominent clones from each screen are presented in Table 2.2. Examination of the clones from population LYH.1.2 revealed a great diversity of mutants able to incorporate Anl. From 24 colonies tested, 17 distinct MetRS mutants active toward Anl were identified. Even though the screen was successful in enriching fluorescent clones, enrichment of one or more optimal sequences was not observed. Four of the seventeen clones (GML, GIL, PNL, and GCL; Table 2.2) that exhibited the highest fluorescence from this group as measured on FACS were selected for further examination.

In order to identify mutants that can sustain protein synthesis at low concentrations of Anl, the screen was repeated after treating the cells with a lower Anl concentration (0.3 mM) during OmpC expression. One reason for decreasing the Anl concentration was to increase the stringency of the screen. By decreasing the amount of Anl available to MetRS mutants in culture, we hoped to isolate mutants that bind more strongly to Anl. A second reason was to further limit the synthesis of OmpC in cells by decreasing the concentration of one of its building blocks. By reducing OmpC synthesis, we aimed to further relieve the cells carrying highly active MetRS mutants from the harmful effects of OmpC expression. The top 0.6% of the fluorescent events from LYH.1.0 were sorted into the population LYH.3.1. (Figure 2.6, middle column) A narrower gate on the fluorescence channel was necessary to exclude any background fluorescent events not linked to Anl labeling. Population LYH.3.1, 60% of which

is fluorescent, was further screened to reveal population LYH.3.2, where active clones make up 90% of the population. Subsequent screening did not improve the population fluorescence. Nine distinct mutants were identified from 24 clones examined from LYH.3.2. The NLL mutant, which appeared in 13 selected clones, was shown to be enriched over other mutants. This mutant, and three other highly fluorescent clones (CLL, SLL, and PIL) were selected for further examination. (Table 2.2)

Several mutants, rather than a single clone, were enriched when the screen was applied to LYH.1.0 at the 0.1 mM Anl level. (Figure 2.6, right column) The top 0.5% of fluorescent events in the naïve library were initially sorted using the sort-single mode to minimize sorting non-fluorescent events. Even though there was a 10-fold enrichment of events above 70 fluorescence units in LYH.2.1b, less than 5% of this population indicated Anl activity. Less than 2% of the 0.7 million events sorted at this stage proved to be viable. Further screening resulted in LYH.2.2 and clones identified showed the enrichment of the NLL mutant, as well as the SLL, PLL and PLI mutants. The PLL mutant also carries the H98N mutation at a surface-exposed position. Since PLL clones were isolated together with and without the H98N mutation in other screens (data not shown), we reasoned that H98N is a conservative mutation.

Traditionally, residue-specific incorporation of noncanonical amino acids is done under conditions that minimize the intracellular concentration of the canonical amino acid to be replaced. To replace Met with Anl, auxotrophic cells that are deficient in methionine synthesis are induced to generate proteins in a synthetic medium that is supplemented with Anl but lacks Met. However, if a MetRS mutant can be identified such that it can discriminate against its original substrate, Met, in favor of Anl, residue specific incorporation can also be carried out in rich media, resulting in healthier cells and better protein yields. In order to explore the possibility of such a mutant being present in our library, a screen was devised where an increasing amount of Met was presented to cells expressing OmpC at 1.0 mM Anl. (Figure 2.7) Clones that were able to incorporate Anl despite an abundance of Met in the media were selected based on fluorescent

labeling. The final population obtained (LYH.5.3d; Figure 2.7, right column) shows little sensitivity to the concentration of Met in the media, leading to sufficient labeling even when Met and Anl are each supplied at a concentration of 1.0 mM. Surprisingly, clones isolated from this population show the enrichment of two mutants (NLL and PLL), both already identified through a screen at 0.1 mM Anl. (Table 2.2) These results suggest that among the members of the LYH.1.0 saturation-mutagenesis library, the PLL and NLL mutants excel in both activity and specificity.

Cell-surface labeling on cells bearing MetRS mutants

In order to verify the success of the screening protocol, we continued to characterize the incorporation of Anl *in vivo* using mutants identified in our selection experiments. Mutations featured in the eight selected clones from the LYH 3.2 and LYH 1.2 populations (Table 2.2), as well as the L13G mutant, were introduced into the MetRS gene on pAJL-20, and the extent of fluorescent labeling supported by each mutant was determined. M15MA cells carrying the pAJL-20 plasmid were induced to produce OmpC in the presence of 0.03, 0.1, 0.3, 1.0 and 3.0 mM Anl. The cells were then labeled, and the median fluorescence from the labeled cell population was determined on a flow cytometer. The collected dose-response data was fit to a Hill equation, and EC50 values were determined for each MetRS mutant.

The response of the fluorescence labeling to increasing Anl concentration can be seen in Figure 2.8. Results show a tightly coupled relationship between the concentration of Anl provided to the cells and the extent of fluorescence labeling. For most of the cases explored, the transition from a non-fluorescent to a fluorescent state occurs through a single order of magnitude increase in the Anl concentration. Mutants screened at 0.3 mM Anl show lower EC50 values overall than those screened at 1.0 mM Anl, showing that screening under more stringent conditions returns mutants with better response to Anl. The mutants from LYH.3.2 exhibit high fluorescence levels down to 0.3 mM Anl. The CLL mutant is able to maintain high fluorescence labeling down to 0.1 mM Anl, and has the

lowest EC50 value of all mutants. The fluorescent labeling of cells carrying the L13G mutant show high variability between the two experiments performed, but the response of these cells resembles that of the LYH.1.2 mutants.

Incorporation of Anl into recombinant proteins with MetRS mutants

To determine how well the MetRS mutants tested above can support protein synthesis in the presence of Anl, MetRS variants were transferred to pAJL-61, which codes for a 6×His-tagged dihydrofolate reductase (DHFR) under inducible control. The resulting plasmids were transformed into M15MA, and protein expression was performed for 3.5 hours in M9+19aa media supplemented with 0.1, 0.3 or 1.0 mM Anl, 40 mg/L methionine, or no 20^{th} amino acid. Whole-cell lysates from these cultures were analyzed on SDS-PAGE to determine the expression levels of DHFR in each culture. DHFR was purified from cell lysates through Ni-NTA chromatography. The purified protein was digested with trypsin, and the peptides generated were analyzed using MALDI-mass spectrometry.

The DHFR expression levels follow the trend laid out by the fluorescence study (Figure 2.9, left panel). When DHFR expression is induced at 1.0 mM Anl, all MetRS mutants studied allow levels of expression comparable to each other. However, when expression is induced at 0.3 mM Anl, only the mutants screened at 0.3 mM Anl sustain high levels of protein synthesis, as evidenced by the SDS-PAGE results. The replacement of Met with Anl is readily observed in tryptic peptides analyzed by MALDI-MS, as a 23.05 Da shift in mass (Figure 2.9, right panel). The first set of mutants (GIL, GML, GCL, PNL) show close to complete replacement of Met with Anl at when protein synthesis is carried out at 1.0 mM, but incorporation is not detected at 0.3 mM Anl. The second set (PIL, CLL, SLL, NLL) consistently shows close to 100% replacement of Met with Anl both at 1.0 and 0.3 mM Anl. Although MALDI-MS results are not reliably quantitative, the trends observed are consistent within each group, and the differences between the two groups of mutants are distinct. Tryptic fragments of proteins expressed at low Anl concentrations reveal peaks 15 Da heavier than the fragment

containing Met. Such a shift in mass does not correspond to a mutation to any canonical amino acid. In addition to supporting good protein synthesis levels with Anl, little protein expression is detected when cells are directed to express DHFR in the absence of a 20th amino acid. This indicates that the MetRS mutants do not accept any of the other 19 amino acids as a substrate.

Incorporation of Anl through the MetRS-L13G mutant

Even though its fluorescence characteristics were similar to mutants screened at 1.0 mM Anl, the protein expression levels with the L13G mutant are comparable to mutants that were screened at 0.3 mM Anl. Examination of the MALDI-MS spectra also show that the L13G mutant is able to incorporate Anl into proteins at 1.0 or 0.3 mM Anl concentrations. (Figure 2.9, top row) However, below 1.0 mM Anl, another peak 3 Da lighter than the Met peak appears in the spectra. This peak can be identified at 0, 0.1 and 0.3 mM Anl. Protein synthesis can also be seen on SDS-PAGE in 0 and 0.1 mM Anl lanes. A mutation mass shift of −3 Da corresponds to either lysine (Lys), or glutamine (Gln). This shift is especially visible in Figure 2.10, where a tryptic peptide containing two methionine residues is examined on MALDI-MS. In the absence of Met or Anl, a new peak shifted by −6 Da appears. At 2 mM Anl, a peptide carrying one Anl and one Gln or Lys appears in the spectra. The incorporation of an alternate amino acid by the L13G mutant can explain why the cell-surface labeling through this mutant lags behind most other mutants despite the good expression levels seen on SDS-PAGE for the target protein. Even though the protein-synthesis rate is sufficiently fast, not all the Met sites are replaced by an azide-containing amino acid. As a result, cell-surface labeling proceeds more slowly and the dose-response curve for L13G (Figure 2.8) takes a shallow slope.

In addition to misincorporation of glutamine or lysine, Anl incorporation through the L13G mutant is also very sensitive to the presence of Met in the media. When cells bearing this mutant are directed to synthesize OmpC in the presence of 1.0 mM Anl and varying amounts of Met, fluorescence labeling is not detected above a Met concentration of 0.03 mM. (Figure 2.11) In contrast, a

selection of three mutants identified through screening (GML, PIL, NLL) show much lower sensitivity to the presence of Met in the media. All three mutants show evidence of a distinct labeled population even at 1.0 mM Met. Even though this behavior is expected from the NLL mutant, which was isolated from LYH.5.3d after being screened against Met incorporation, the GML and PIL mutants give responses matching those from the NLL mutant. Labeling and detection of Anl-tagged proteins in bacteria, expressed in the presence of 20 natural amino acids and Anl, confirms this finding [34]. When cells constitutively expressing the L13G or NLL mutants of MetRS are grown in rich media containing 2 mM or more Anl, incorporation of Anl can be easily detected in proteins synthesized in cells expressing MetRS-NLL. In contrast, MetRS-L13G-expressing cells fail to give a strong signal in western-blot analysis.

In vitro activation kinetics for MetRS mutants

MetRS mutants identified were expressed and purified for the determination of their Anl activation kinetics *in vitro*. (Table 2.3) It has been shown in the past that the *in vitro* activation kinetics of a noncanonical amino acid correlates well with its observed efficiency of incorporation *in vivo* [15, 35]. Based on the results of the *in vivo* characterization, we expected faster activation of Anl by mutants identified through screens performed at lower concentrations of Anl. This indeed is true when comparing the activation kinetics of mutants identified at 0.1 and 0.3 mM with 1.0 mM Anl. (Figure 2.12.a) The PLL mutant, which was identified at 0.1 mM Anl, exhibits the highest specific activity (k_{cat}/K_m) for Anl among the MetRS variants studied, whereas mutants isolated at 1.0 mM Anl have the slowest activation kinetics. Both catalytic (k_{cat}) and binding (K_m) components of specific activity are influential in determining the observed differences in k_{cat}/K_m. Mutants screened at higher stringencies support tighter binding (Figure 2.12.b) and faster catalysis (Figure 2.12.c) than those identified through less stringent screens, suggesting that both ligand recognition and catalysis are altered in the screens to achieve optimal activity.

Among mutants tested for Met activation, the NLL and PLL mutants are

the only two that favor Anl over Met, which explains their selection by the screen carried out in the presence of Met. (Table 2.3; Figure 2.12.d–e) Kiick and coworkers have previously reported similar k_{cat} values for the activation of a variety of unnatural substrates by the wild-type MetRS [25, 35], suggesting that the activation rate for a Met analog is determined by how well its side chain is recognized by the MetRS. Our results also agree with this view: For each MetRS mutant tested, similar k_{cat} values are obtained for the activation of Met and Anl, while K_m values may differ greatly. (Figure 2.12.f)

The previously reported parameters for the L13G mutant suggest comparable activation parameters for Anl and Met with this mutant, resulting in k_{cat}/K_m values of 1,600 and 2,000 M^{-1} s^{-1}, respectively [20]. However, these measurements cannot explain the observed sensitivity of the L13G mutant to Met *in vivo*. Re-examination of the activation kinetics reveal that the L13G mutant activates Met an order of magnitude faster than Anl. (Table 2.3) This result agrees with the observations *in vivo*.

Remarkably, the specific activities for Anl activation show a strong correlation with the EC50 values determined *in vivo* through cell-surface labeling. (R^2 = 0.77; Figure 2.13) Specific activities are determined *in vitro* and measure the efficiency of a relatively simple, MetRS-dependent reaction required for the incorporation of Anl into proteins. On the other hand, EC50 values are based on cell-surface labeling, and are rather indirect measures of MetRS activity *in vivo* due to their dependence on many factors such as the stability and expression levels of the mutants in the cell, and the specificity of the mutant for Anl and against the pool of ligands available to it in the cytoplasm. The strong coupling between our reporter for MetRS activity and the directly measured activity of this enzyme demonstrates the success of the screen design.

Distribution of mutations at the randomized sites

When compared to the natural substrate of MetRS. Anl is longer, and carries its hydrogen-bond acceptor atoms further down the side chain. In the binding site of wild-type *E. coli* MetRS, the Sδ atom of the Met is recognized by

hydrogen bonds from aromatic residues on positions 260 and 301, and the backbone nitrogen atom on position 13. An analysis of crystal structure data has revealed that covalently bonded azide groups interact with hydrogen-bond donors through their terminal nitrogen [36]. Thus, in order for a MetRS mutant to be able to recognize Anl better, a hydrogen-bond donor is necessary deeper in the binding site.

In order to gain insight into the structural basis of the selection results, we examined the types of amino-acid substitutions selected at each mutation site in our experiments. Through the analysis of individual clones following various selection experiments, 150 active clones carrying mutations only at the three designated sites were isolated. From this pool, 41 distinct mutants of MetRS active toward Anl were identified. A full list of mutations and the populations they were identified from can be found in Appendix A. Distributions of mutations selected at each site are shown in Figure 2.14.a. The strongest selection takes place at the H301 site, where a mutation to leucine is overwhelmingly selected for. Only three other residues, isoleucine, valine and methionine, were observed at this site and, like leucine, all of these are space-creating substitutions to aliphatic, hydrophobic amino acids. In the wild-type MetRS structure H301 hydrogen bonds to the Y260 side chain. Although the Y260 position allows substitutions to a variety of residues, a majority (88%) of these are small or hydrophobic in nature and not able to exchange hydrogen bonds with the neighboring 301 position. Thus, a hydrophobic mutation at this site is reasonable especially considering the surrounding hydrophobic residues, I293 and I297.

Unlike the H301 position, the L13 and Y260 positions allow a wide variety of amino acids, all compatible with the activation of Anl. When the screens are performed at low stringency (1.0 mM Anl or higher), this diversity is retained. (Figure 2.14.b) When the stringency of the screens increases, however, the choices at these positions narrow down. (Figure 2.14.c) In high-stringency screens the position Y260 is occupied most often by leucine and other aliphatic hydrophobic residues, similar to H301. Proline is the most prevalent residue selected at the L13 position under stringent conditions. This substitution

eliminates the backbone amide nitrogen that recognizes Sδ on Met in the wild-type MetRS, decreasing the affinity of the mutant MetRS for its natural ligand.

X-ray crystal structure of the Anl-bound MetRS-SLL mutant

Although many explanations can be proposed for the selection of observed mutations, the observations fail to identify a mutation that introduces a hydrogen-bond donor in the binding site that can recognize the azide group. The high-resolution (1.5 Å) x-ray crystal structure of *E. coli* MetRS-SLL [37], solved in both Anl-free and bound forms, reveals that an enzyme-bound water molecule is responsible for the recognition of the azide moiety on Anl. (Figure 2.15.a) This water molecule is buried inside the Met binding site, held tightly by hydrogen bonds with the Oγ atom of T10 and the backbone O of residue F292. In fact, this crystal water is present in all ligand-free or ligand-bound crystal structures of both the wild-type MetRS [26, 38] and the SLL mutant [37]. In the wild-type MetRS structure, this water molecule also hydrogen bonds to the Oη atom on the Tyr260 side chain, and is part of the hydrogen-bonding network that keeps this residue oriented to interact with the methionine sulfur atom. (Figure 2.15.b) In the SLL mutant, the terminal azide nitrogen of Anl replaces the Tyr260 Oη atom as the hydrogen bonding partner of this solvent molecule. The fact that a water molecule, instead of a residue side chain, is responsible for recognizing the azide explains the great diversity of mutations at positions 13 and 260 found to be compatible with Anl activation.

Kinetic analysis reveals that mutations at position 13 have a greater effect on the binding and activation of Met than Anl. (Figure 2.12.d–e) It is possible that the identity of the residue at position 13 alters the solvation of the backbone NH on position 13, thereby changing the availability of this hydrogen-bond donor to the Sδ on Met. The rationale behind the selection of mutations at position 13 is further discussed in Chapter 3.

Wild-type MetRS exists in an “open” conformation in the absence of a bound ligand. When methionine binds to the enzyme, a cascade of structural

changes occurs to “close” the binding site around the ligand [26]. These structural changes are thought to initiated by the interaction between the W253 side chain and the bound methionine. Upon ligand binding, this side chain rotates more than 90° in χ_1 to form one side of the binding pocket. (Figure 2.16.a) This movement is concerted with a similar conformational change in the F300 side chain, among others, and puts this residue in a stacked orientation with W253.

The conformational change between “open” and “closed” forms of the enzyme upon ligand binding does not take place in the SLL mutant. (Figure 2.16.b) This mutant is locked in the “closed” configuration even when no ligand is present in the binding site. Analysis of the ligand-bound and unbound conformations of the SLL mutant and wild-type MetRS suggests that the H301L mutation might be responsible for the lack of this conformational switch. The “open” configuration the F300 side chain takes in the wild-type apo-MetRS is not compatible with the orientation assumed by L301 in the MetRS-SLL apo-enzyme. (Figure 2.16.c) Alignment of the protein backbone around residues 300 and 301 from these structures suggests that in an “open”–like configuration the Cδ atoms on the F300 and L301 side chains would come into contact. Since these two residues neighbor each other in the primary sequence, there is limited torsional freedom between the two sites to relieve this clash in the context of the binding site. Possibly, this unfavorable contact is responsible for the “closed” conformation of MetRS-SLL in the absence of ligand.

The selection strategy employed in this study was designed to identify the MetRS variant that best supports the *in vivo* incorporation of Anl into proteins. Considering the structural differences between the Met and Anl, we expected this screen to optimize the recognition of Anl in the binding site by creating space for the bigger ligand and forming polar interactions with the azide group. It is, therefore, interesting that the MetRS characteristic most strongly selected by the screen for Anl incorporation is the H301L mutation and suggests that this mutation may have a critical role for Anl incorporation. The H301L mutation may ensure that the amino-acid binding site is fully formed even when the ligand is

not present which, in turn, could speed up the ligand-binding and activation steps.

MetRS, similar to other class I aminoacyl-tRNA synthetases, edits the methionine-precursor homocysteine [39]. It was postulated that the function of the conformational change in the wild-type MetRS binding site is to distinguish methionine from homocysteine. Detecting the presence of the methyl group on the ligand, W253 residue undergoes a conformational change, triggering the closing of the binding site, which may not happen for homocysteine [26]. This theory predicts an increased rate of homocysteine misincorporation from MetRS carrying the H301L mutation. It is also reported that the effects of ligand binding can be felt 10 Å away from the binding site through the conformational changes triggered by ligand binding, even possibly affecting the recognition of the anticodon on associated tRNA [26]. Such a scheme suggests cooperativity between ligand and tRNA recognition, which could improve the fidelity of Met-tRNA^{Met} synthesis. Breaking such cooperativity can improve the rates of synthesis for Anl-tRNA^{Met} at the expense of its accuracy. Indeed, the structures of the *A. aeolicus* MetRS-tRNA^{Met} complex with and without ligand exhibit limited structural changes upon ligand binding [40]. Additional work is necessary to evaluate these hypotheses.

Conclusions

We have demonstrated successful identification of *E. coli* MetRS mutants that support synthesis of proteins bearing the reactive amino acid Anl *in vivo* through a fast, high-throughput screening strategy. The identified mutants were characterized and a strong correlation was discovered between the *in vitro* activities and the *in vivo* performance of these MetRS mutants. Our screen uses fluorescent labeling of azides displayed on the cell surface as a reporter of MetRS activity toward Anl. We show that this selection of reporter for activity toward Anl does indeed correlate closely with the rates of Anl activation by each MetRS mutant.

The screening strategy can be applied in the presence of competitors as

well. Screening for Anl incorporation in the presence of Met as a competitor returns MetRS variants that activate Anl faster than Met. It is interesting to note that even suboptimal mutants obtained from the screen, such as MetRS-GML, display persistent levels of fluorescent labeling in the presence of high concentrations of Met. (Figure 2.11) The sharp decline of the signal from cells carrying MetRS-L13G when exposed to Met shows that the fluorescent reporter for MetRS activity is very sensitive to "misincorporation" of residues not bearing azides into Met sites. Thus, the screening strategy also implicitly selects for specificity of MetRS activity.

In our screening protocol, cells carrying MetRS mutants have to find a balance between two important factors under the conditions provided to them: first, achieving efficient Anl incorporation and OmpC synthesis and second, avoiding OmpC toxicity. At a given condition the screen will try to pick the poorest mutant that can provide the best fluorescence, leading to a minimal OmpC synthesis for that fluorescence level. Indeed, when populations LYH.3.2 and LYH 2.2 (Figure 2.6) are treated with 1.0 mM Anl and labeled, the fluorescence levels of the active clones in these populations increase to match the levels observed in LYH.1.2. This is despite the very different set of mutants that make up each population. Hence, at 1 mM Anl, highly active mutants like NLL or SLL are not selected by the screen because less active mutants can provide the same fluorescence gain, but with less OmpC toxicity. The interplay between these opposing factors enables us to select for mutants at different activity levels by varying the availability of Anl in the screen.

The results of the library screens point to the importance of the H301L mutation in MetRS. The crystal structure of the MetRS-SLL mutant suggests that this mutation locks the enzyme in the "closed" conformation, thereby aiding the activation of Anl and its subsequent incorporation into proteins *in vivo*. At this point, the effects of this mutation on the specificity and editing activity of MetRS are not known. However, this knowledge will aid future efforts to engineer new MetRS activities.

Table 2.1.

Sequences for primers discussed in this chapter

Primer name	Sequence* (5'-to-3')
lib_fwd2	CTCAGTACCAGTTCGACTTCGGTCTGCGTCCGTCCCTG
L13_lib2-r	GATTGAGCCGTTAGCGTACGGMNNTGCGCACGTCACCAG
L13_lib3	GAAAATTCTGGTGACGTGCGCANNKCCGTACGCTAACGGCTCA
Y260_lib3-r	CTTGAAAGAACCCATMNNGCCAATCGGTGCGTCCAGC
Y260_lib3	GCTGGACGCACCGATTGGCNNKATGGGTTCTTTCAAG
H301_lib3-r	CAGCATGGCAGGCCAGAACAGGCTMNNGAAGTAAACAATATC
H301_lib2	GATATTGTTTACTTCNNKAGCCTGTTCTGGCCTGCCATGC
lib_rev2	CAGTACCGGCTTCAGGTAAGTCATCAGCACGCGGAAC
lib_fwd3	CAGTACCAGTTCGACTTCGGTCTGCGTC
lib_rev3	CTTCAGGTAAGTCATCAGCACGCGGAAC
MRS_BamHI	TTCCGCGGATCCATGACTCAAGTCGCGAAGAAAATTC
MRS_SalI-r	TTTGGGGTCGACTCATTTAGAGGCTTCCACCAGTG
eM_L13G†	CTGGTGACGTGCGCAGGTCCGTACGCTAACGGCTC
eM_G13C†	ATTCTGGTGACGTGCGCATGTCCGTACGCTAAC
eM_G13N†	ATTCTGGTGACGTGCGCAAACCCGTACGCTAAC
eM_G13P†	ATTCTGGTGACGTGCGCACCGCCGTACGCTAAC
eM_G13S†	ATTCTGGTGACGTGCGCAAGCCCGTACGCTAAC
eM_Y260C†	GACGCACCGATTGGCTGCATGGGTTCTTTCAAG
eM_Y260I†	GGACGCACCGATTGGCATCATGGGTTCTTTCAAG
eM_Y260L†	GGACGCACCGATTGGCCTGATGGGTTCTTTCAAG
eM_Y260M†	GGACGCACCGATTGGCATGATGGGTTCTTTCAAG
eM_Y260N†	GGACGCACCGATTGGCAACATGGGTTCTTTCAAG
eM_H301L†	GATATTGTTTACTTCCTGAGCCTGTTCTGGCCTGC

* Degenerate nucleotides: N = A, T, G, C; K = G, T; M= C, A.

† Only the forward sequence is provided for site-directed mutagenesis primers.

Table 2.2.

MetRS mutants identified in clones showing the highest fluorescence labeling library screens

Screen condition* (*Population*)	Name	Clone	Leu 13	Tyr 260	His 301
1.0 mM Anl (*LYH.1.2*)	GML	1.2.14	G	M	L
	GIL	1.2.19	G	I	L
	PNL	1.2.26	P	N	L
	GCL	1.2.23	G	C	L
0.3 mM Anl (*LYH.3.2*)	NLL	3.2.4	N	L	L
	CLL	3.3.15	C	L	L
	SLL	3.2.24	S	L	L
	PIL	3.2.11	P	I	L
0.1 mM Anl (*LYH.2.2*)	NLL	2.2.2	N	L	L
	PLL†	2.2.11	P	L	L
	PLI	2.2.24	P	L	I
	SLL	2.2.4	S	L	L
1.0 mM Anl + Met (*LYH.5.3d*)	NLL	5.3.61	N	L	L
	PLL†	5.3.67	P	L	L

* Library screens were performed using media supplemented with amino acids at shown concentrations.

† The PLL mutants isolated here carry the H98N mutation in addition to the three mutations listed

Table 2.3.

*Kinetic parameters for the activation of Met and Anl by MetRS mutants**

MetRS variant	Amino acid	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	Selectivity†	Relative activity
wt‡	Met	0.024	13.30	550,000	-	1
L13G	Met	0.16 ± 0.06	0.87 ± 0.18	5,600 ± 1,500	0.03	1/92
	Anl	5.1 ± 1.3	0.85 ± 0.07	170 ± 40		1/3,200
NLL	Met	2.6 ± 0.5	0.86 ± 0.07	350 ± 70	1.2	1/1,600
	Anl	2.2 ± 0.8	0.87 ± 0.11	410 ± 80		1/1,400
PLL	Met	5.3 ± 1.9	1.0 ± 0.2	200 ± 50	3.2	1/2,700
	Anl	1.5 ± 0.7	0.92 ± 0.22	650 ± 150		1/850
SLL	Met	1.0 ± 0.3	0.89 ± 0.16	900 ± 80	0.24	1/610
	Anl	4.2 ± 1.1	0.90 ± 0.13	220 ± 20		1/2,600
CLL	Met	0.14 ± 0.04	0.94 ± 0.18	7,100 ± 600	0.07	1/77
	Anl	2.0 ± 0.7	0.98 ± 0.16	520 ± 120		1/1,100
PLI	Anl	2.4 ± 1.0	0.89 ± 0.18	410 ± 140	-	1/1,300
PIL	Anl	6.6 ± 1.8	0.84 ± 0.08	130 ± 40	-	1/4,100
PNL	Anl	5.3 ± 2.2	0.47 ± 0.10	98 ± 37	-	1/5,600
GML	Anl	11 ± 5	0.48 ± 0.11	47 ± 13	-	1/12,000
GIL	Anl	15 ± 3	0.22 ± 0.01	14 ± 3	-	1/38,000
GCL	Anl	12 ± 3	0.12 ± 0.01	11 ± 2	-	1/51,000

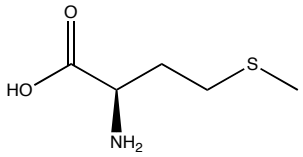
* Data on mutants of lower activity (AQL, SNL, and GVL) are listed in Table 3.5.

† Selectivity is defined as the ratio of the relative activities (k_{cat}/K_m) for Anl to that of Met.

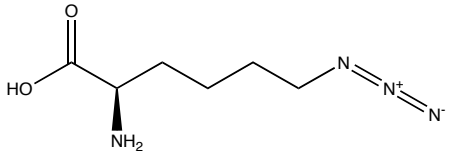
‡ Activation parameters for wild-type MetRS taken from reference [35].

Figure 2.1.

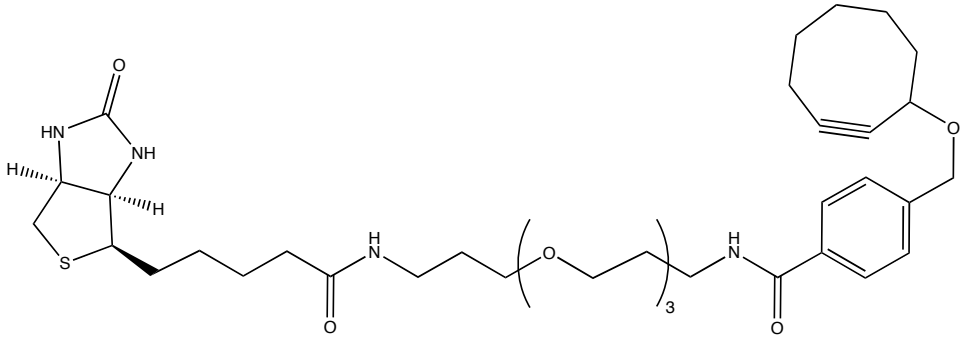
Chemical structures of amino acids and tagging reagents.



methionine (Met)



azidonorleucine (Anl)



biotin-PEO-cyclooctyne

Figure 2.2.

Protocol for screening the MetRS library for activity toward azidonorleucine.

Cells harboring the MetRS library are induced to express the outer-membrane protein C (OmpC) in the presence of Anl. Cells carrying a MetRS mutant that is active toward Anl display this residue on their outer membrane. The displayed azide groups are conjugated to biotin through click-chemistry. Subsequent addition of fluorescent avidin enables fluorescent tagging of all cells carrying active MetRS mutants. These mutants are separated from inactive ones through flow cytometry. The sorted population carrying active mutants can be subjected to additional rounds of selection or analyzed.

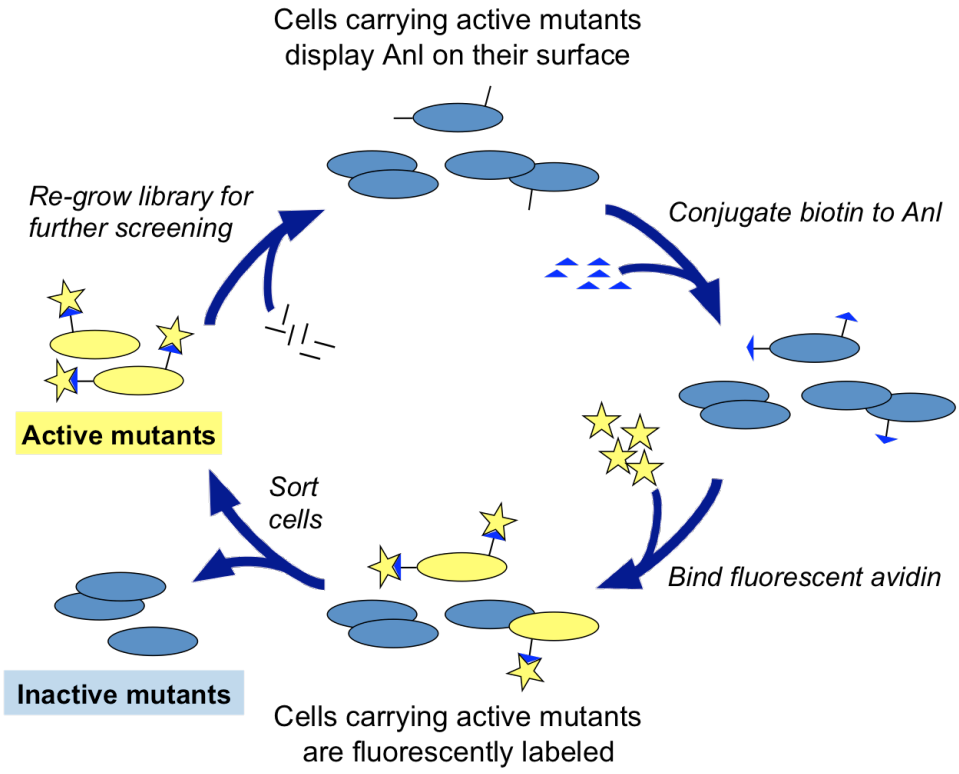


Figure 2.3.

Four residues considered for saturation mutagenesis in the methionine binding pocket of the E. coli *MetRS.*

The ligand, methionine (ball-and-stick model), is shown inside the MetRS active site. From the residues surrounding the Sδ and Cε atoms of methionine (space-filling models), three positions (L13, Y260 and H301) were selected to be randomized in this study. Model was generated using VMD, using coordinates from PDB ID: 1F4L [26].

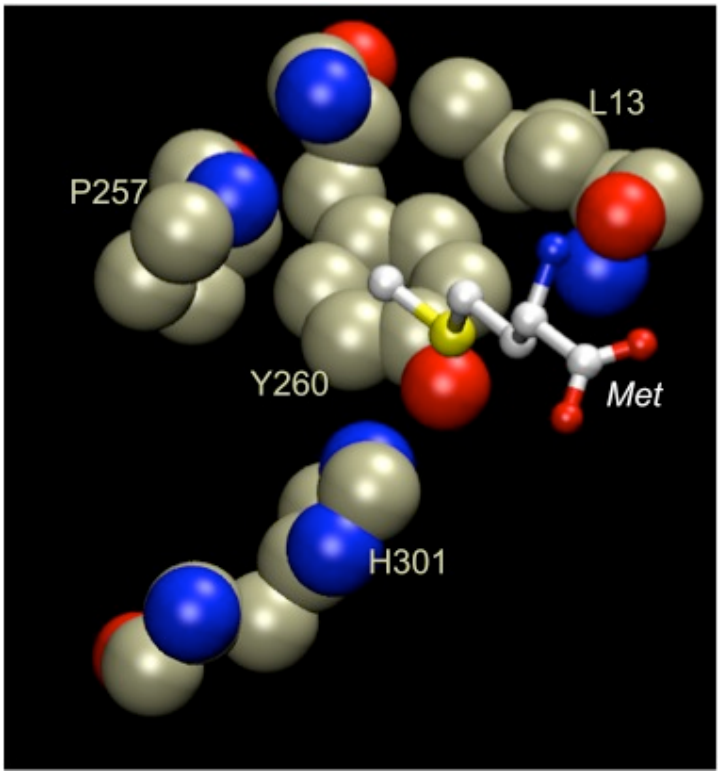


Figure 2.4.

Effect of OmpC overexpression on cell viability.

At different points during expression aliquots were taken from cultures expressing either OmpC or DHFR in media containing either Anl or Met. Following cell-surface labeling, cells were plated on agar plates and the resulting colonies were counted to determine the number of viable cells. Counts for OmpC and DHFR expressing cells are marked with filled triangles (▲) and open diamonds (◇), respectively. Data from Anl-treated cultures are shown in red. No colonies could be observed for cultures that were induced for two or four hours. The data was extrapolated beyond one hour to emphasize this fact.

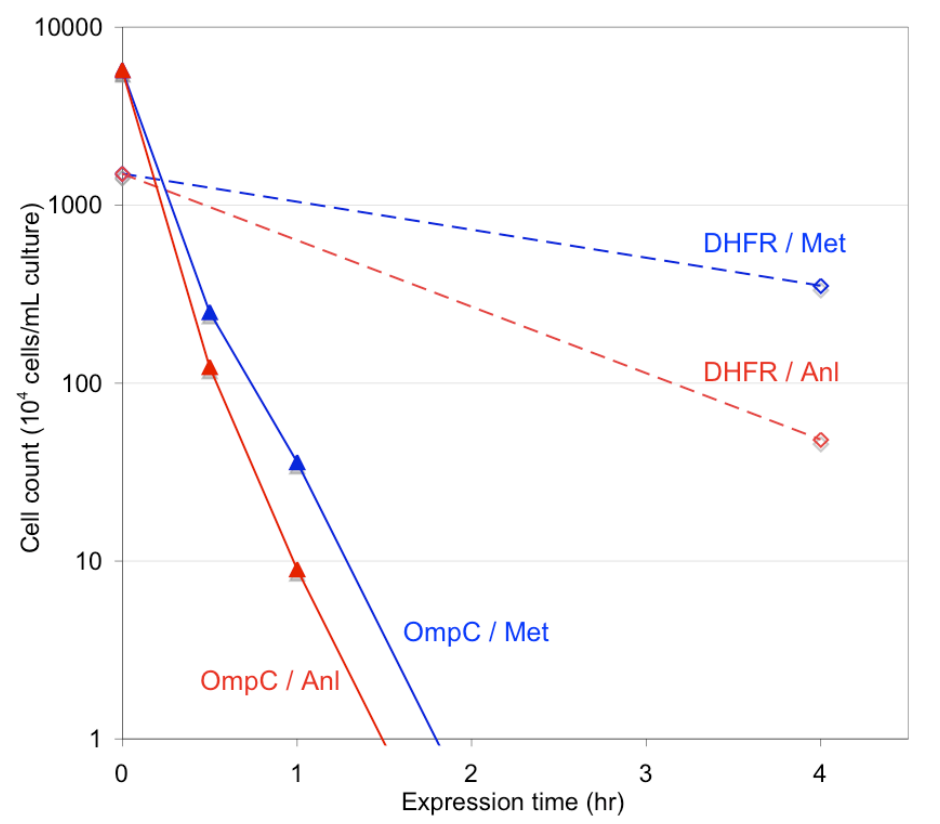


Figure 2.5.

Effect of the duration of OmpC expression on the cell-surface labeling of E. coli *bearing the MetRS L13G mutant.*

Expression of OmpC was induced by the addition of IPTG at 0 hr. Cells were allowed to express OmpC in media supplemented either with 0.3 mM Met, or 1.0 mM Anl. Aliquots were taken at 0.5, 1, 2 and 4 hr from the start of expression, fluorescently labeled, and their fluorescence was measured.

a) Total cell fluorescence as a function of duration of OmpC expression and total cell fluorescence was determined following fluorescent labeling. Data is shown relative to the fluorescence read from uninduced cells.

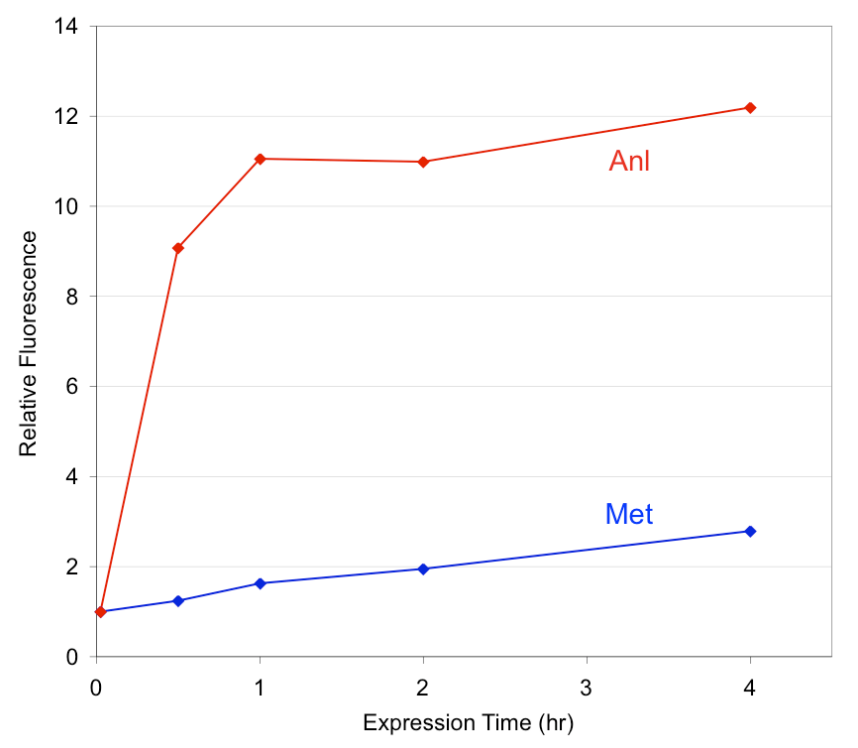


Figure 2.5. (continued)

b) Fluorescence histograms of cells induced for 1 and 4 h to produce OmpC in media containing either Met (negative control) or Anl. The count of interrogated cells is presented against relative cell fluorescence (FL1).

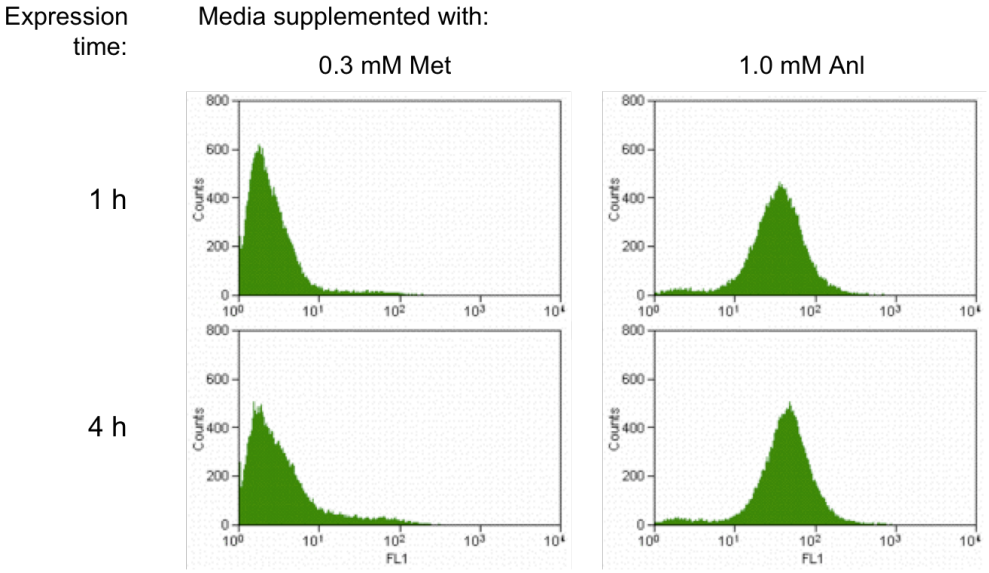


Figure 2.5. (continued)

c) Enhancement of fluorescence through the cell-surface display of AnI after varying durations of OmpC expression. Fluorescence enhancement is measured as the ratio of the fluorescence of cells treated with AnI to that of cells induced in media containing Met.

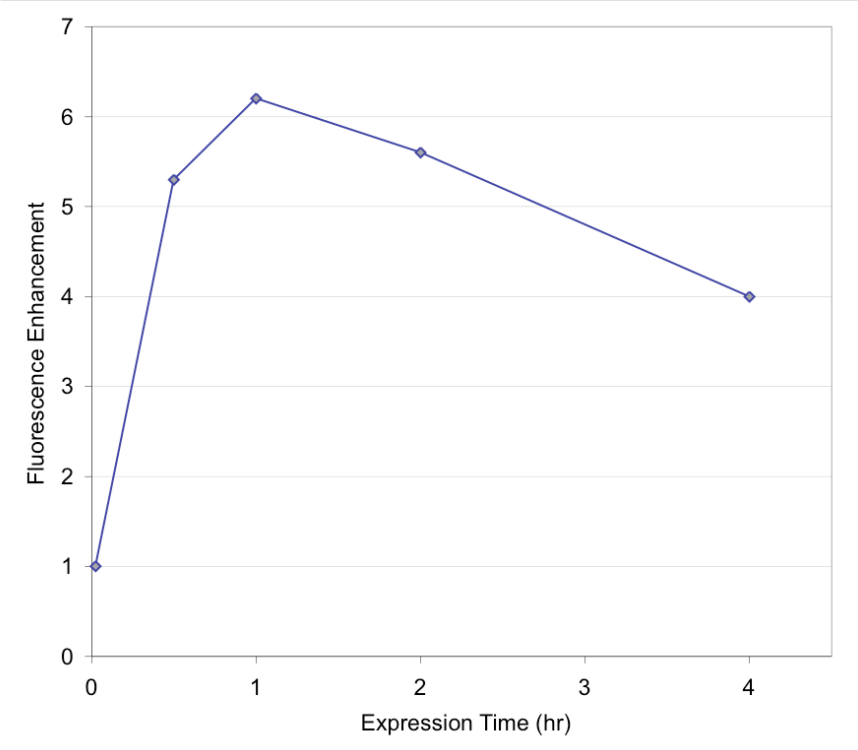


Figure 2.6.

Fluorescence histograms outlining the progression of the library selection.

Selection was carried out on clones carrying the naïve library (LYH.1.0) treated with different concentrations (1.0, 0.3 or 0.1 mM) of AnI during OmpC expression. After cells were fluorescently labeled, the top 0.2% to 1.1% highest fluorescent cells were carried on to the following round.

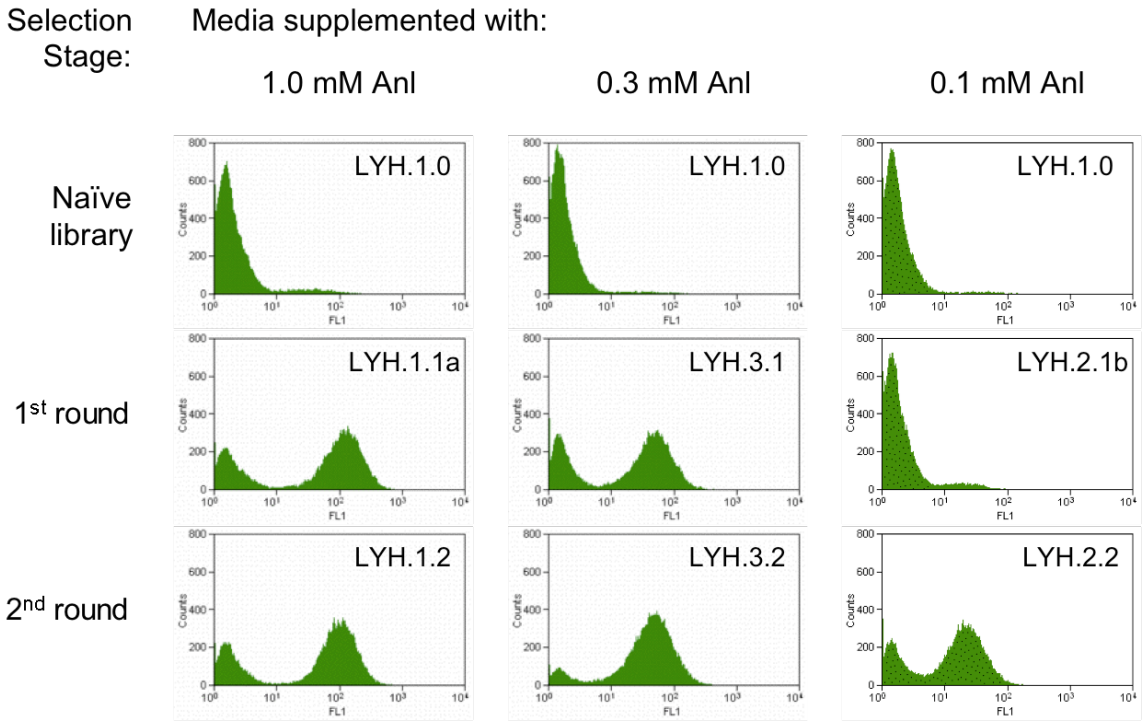


Figure 2.7.

Progression of the library selection where clones were selected for Anl incorporation in the presence of 20 canonical amino acids.

Throughout the screen, OmpC expression was induced in media containing 1.0 mM Anl, and different concentrations of Met. The naïve library (LYH.1.0) was screened following OmpC expression in 1.0 mM Anl and 0.01 mM Met, revealing LYH.5.1d. The concentration of Met in the expression media was increased at each subsequent round of selection, until a final population (LYH.5.3d) not sensitive to the amount of Met in media was reached. For each population, labeling at multiple Met concentrations is shown. Red arrows indicate the conditions at each stage where selections were carried out.

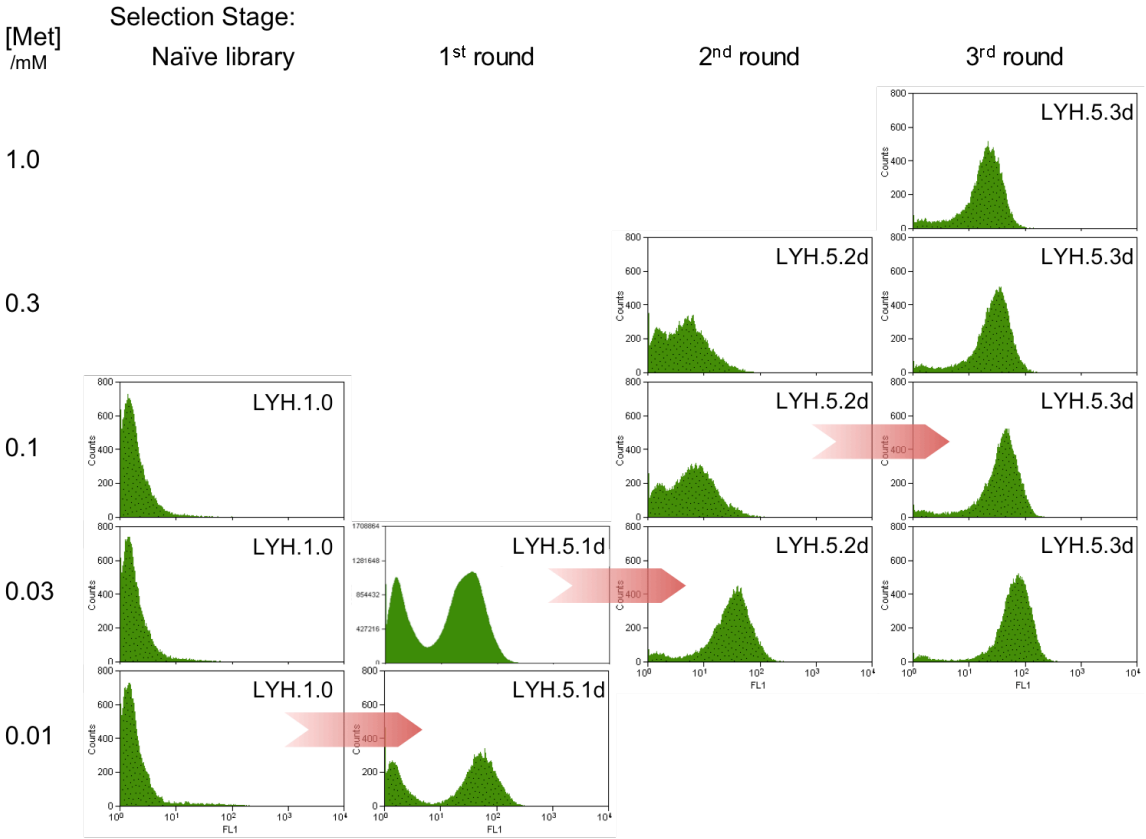


Figure 2.8.

Extent of fluorescence labeling on cells bearing various MetRS mutants at different of Anl concentrations.

OmpC expression was carried out in M9+19aa media supplemented with 0.03, 0.01, 0.3, 1.0 or 3.0 mM Anl. Median fluorescence of the labeled cell population was determined by flow cytometry. The median fluorescence for cells treated at each Anl concentration (red circles), and the Hill equation fit to the data (red lines) are displayed for each MetRS mutant tested. In cases where a fluorescence drop was observed going from 1.0 to 3.0 mM Anl, the final data point was not included in equation fitting. The EC50 value obtained from the Hill equation is indicated on each plot in millimolar units. Mutants isolated through selection at 0.3 mM Anl (labeled in blue) give rise to lower EC50 values than mutants isolated through selection at 1.0 mM Anl (labeled in green).

Figure 2.8. (continued)

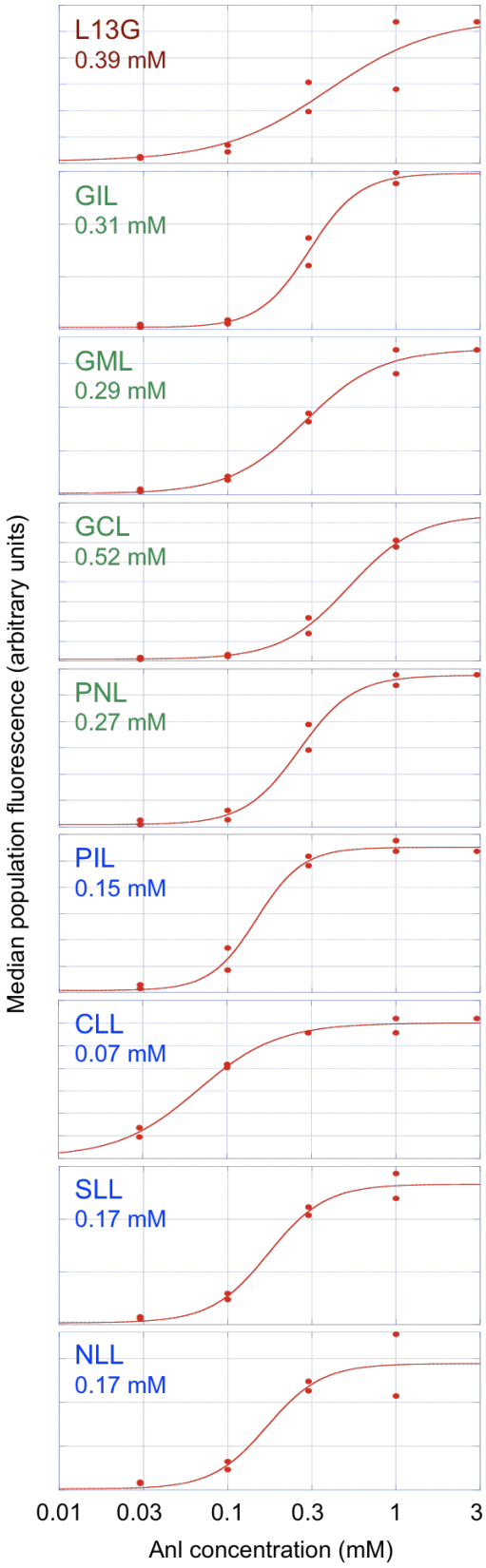


Figure 2.9.

Expression of DHFR at varying concentrations of Anl in cells bearing MetRS mutants.

DHFR was expressed in M15MA[pREP4/pAJL-61] cells encoding the MetRS mutants isolated from screens carried out at 0.3 mM and 1.0 mM Anl, as well as the L13G mutant. Expression was done in M9+19aa media supplemented with 0.1, 0.3 or 1.0 mM Anl, 40 mg/L methionine, or no 20^{th} amino acid. SDS-PAGE analysis of whole-cell lysates show increasing levels of DHFR expression with increasing concentration of Anl in the media. Mutants isolated through selection at 0.3 mM Anl (labeled in blue) can support the synthesis of DHFR well at 0.3 mM Anl in contrast to mutants isolated through a selection at 1.0 mM Anl (labeled in green). The DHFR expressed by the cells was purified, and its tryptic digests were analyzed on MALDI-MS. The peaks belonging to the tryptic fragment IMQEFESDTFFPEIDLGKYK (2437.16 Da) are shown in the range 2425–2475 Da. Replacement of Met by Anl results in a 23.05 Da mass increase. The expected locations of the for Met- and Anl-containing peptides are marked by green and red lines, respectively. A more detailed presentation of the MALDI-MS spectra featured here can be found in Appendix B.

Figure 2.9. (continued)

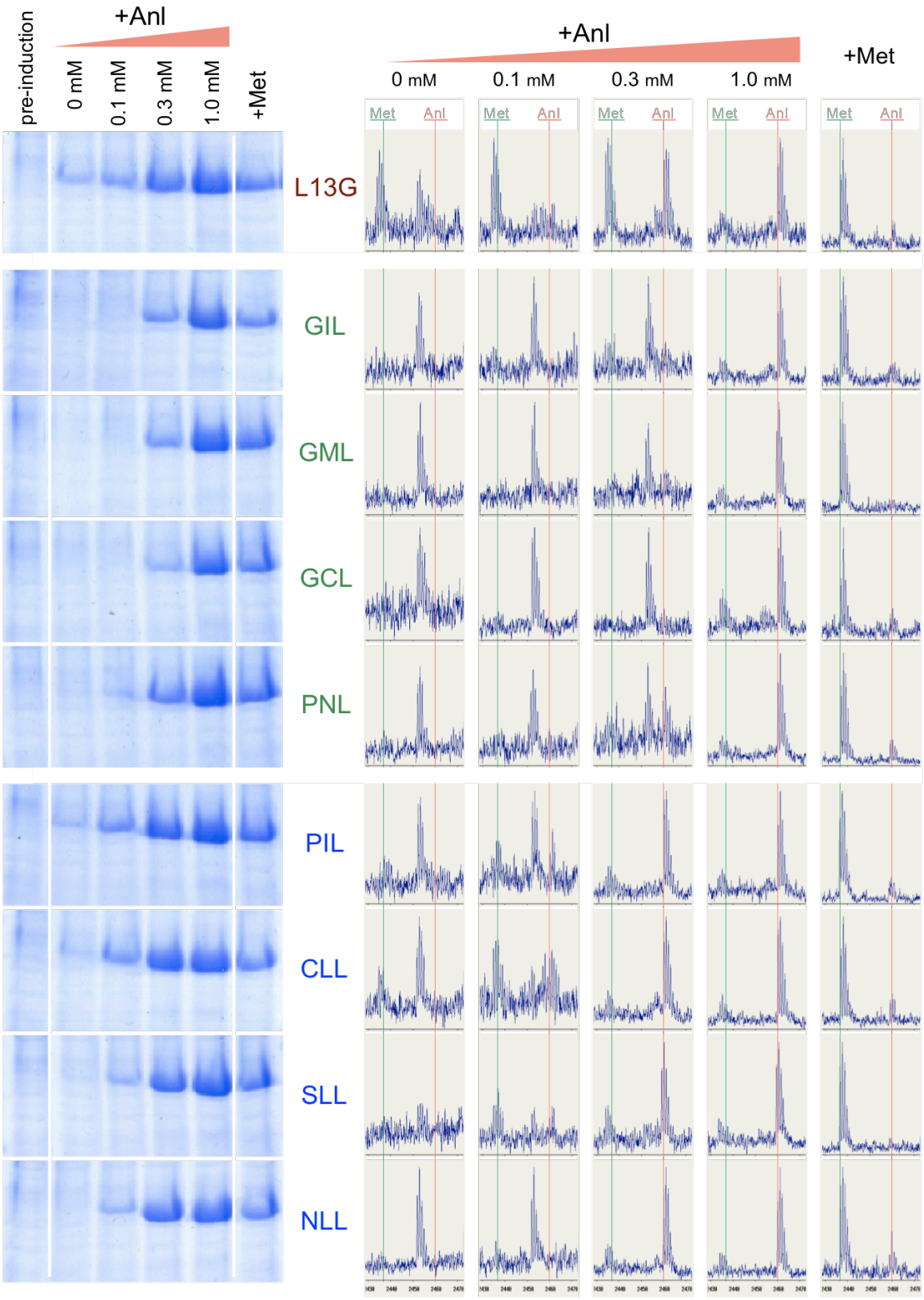


Figure 2.10.

Mass spectra of a tryptic peptides from DHFR expressed in the presence of the MetRS-L13G mutant.

DHFR was expressed in M15MA[pREP4/pAJL-61] cells constitutively expressing the MetRS-L13G mutant in M9+19aa media supplemented with 2.0 mM Anl, 40 mg/L Met, or with no 20th amino acid. Tryptic peptides from purified and digested DHFR were analyzed on MALDI-MS. Peaks belonging to the tryptic fragment VDMVWIVGGSSVYQEAMNQPGHLR (2673.3 Da) containing two methionine residues are shown in the range 2650–2750 Da. Replacement of Met by Anl results in a 23.05 Da mass increase. Expected mass of the peptide containing two Anl residues (marked as +46) is 2719.4 Da.

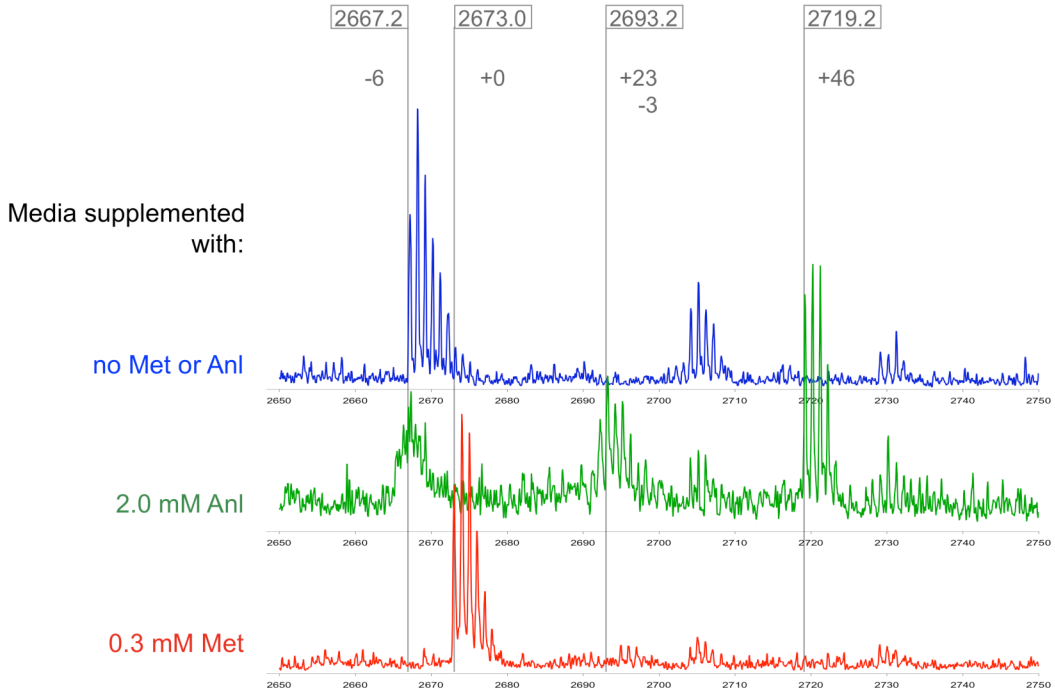


Figure 2.11.

Response of fluorescence labeling to increasing Met concentrations in expression media containing 1.0 mM Anl.

Cells carrying mutants four separate MetRS mutants are tested for fluorescence labeling after OmpC expression in M9+19aa media supplemented with 1.0 mM Anl and varying concentrations of Met. The fluorescence generated by the L13G mutant is very sensitive to the presence of Met in the media, whereas selected mutants isolated from screens are more resistant to the presence of Met.

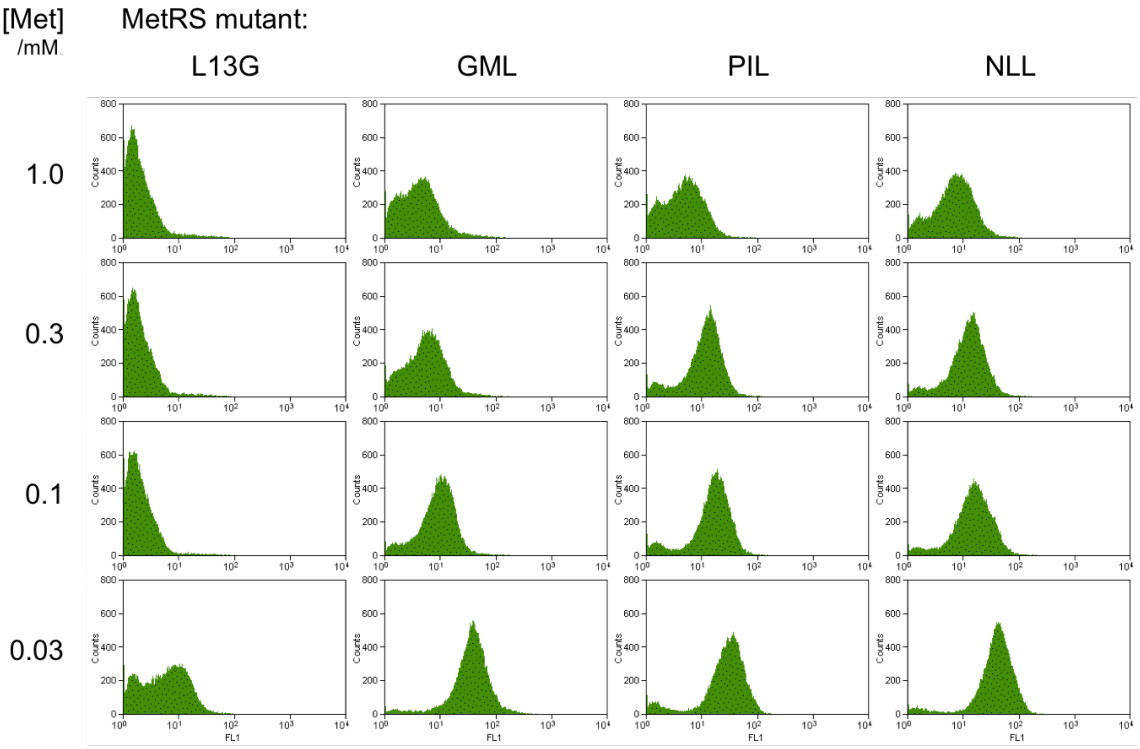


Figure 2.12.

Comparison of in vitro *activation kinetics between MetRS mutants identified through library screening.*

For MetRS mutants tested, a) k_{cat}/K_m, b) K_m, and c) k_{cat} values determined *in vitro* for the activation of Anl are displayed. Mutants identified by screens at 0.1, 0.3 and 1.0 mM Anl are marked in dark blue, light blue, and green, respectively. Data for the L13G mutant is shown in red. Mutants are shown in order of increasing binding or activity. The results indicate that activation of Anl is faster in mutants isolated by screens of higher stringency. The activation of Met was measured only for a subset of the mutants studied. For these mutants, k_{cat}/K_m of d) Anl and e) Met activation, as well as f) the K_m and k_{cat} values measured for the activation of these amino acids are presented. Values for Met are shown in red, while those for Anl are in blue. For a given mutant the both substrates exhibit similar k_{cat} values, and the substrate preference of the enzyme is determined by K_m. All measurements were performed in triplicate. Error bars indicate one standard deviation around the mean.

Figure 2.12. (continued)

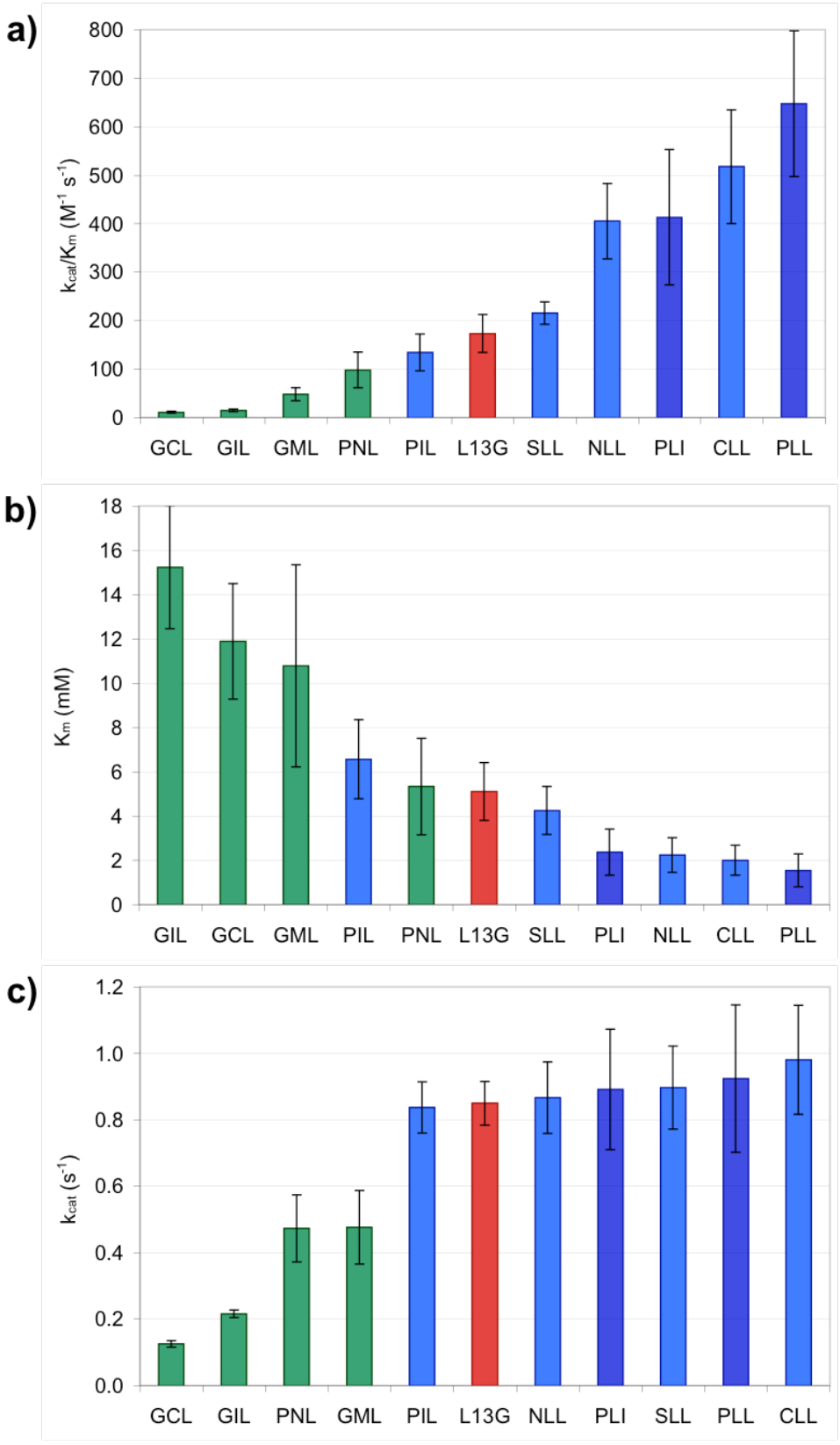


Figure 2.12. (continued)

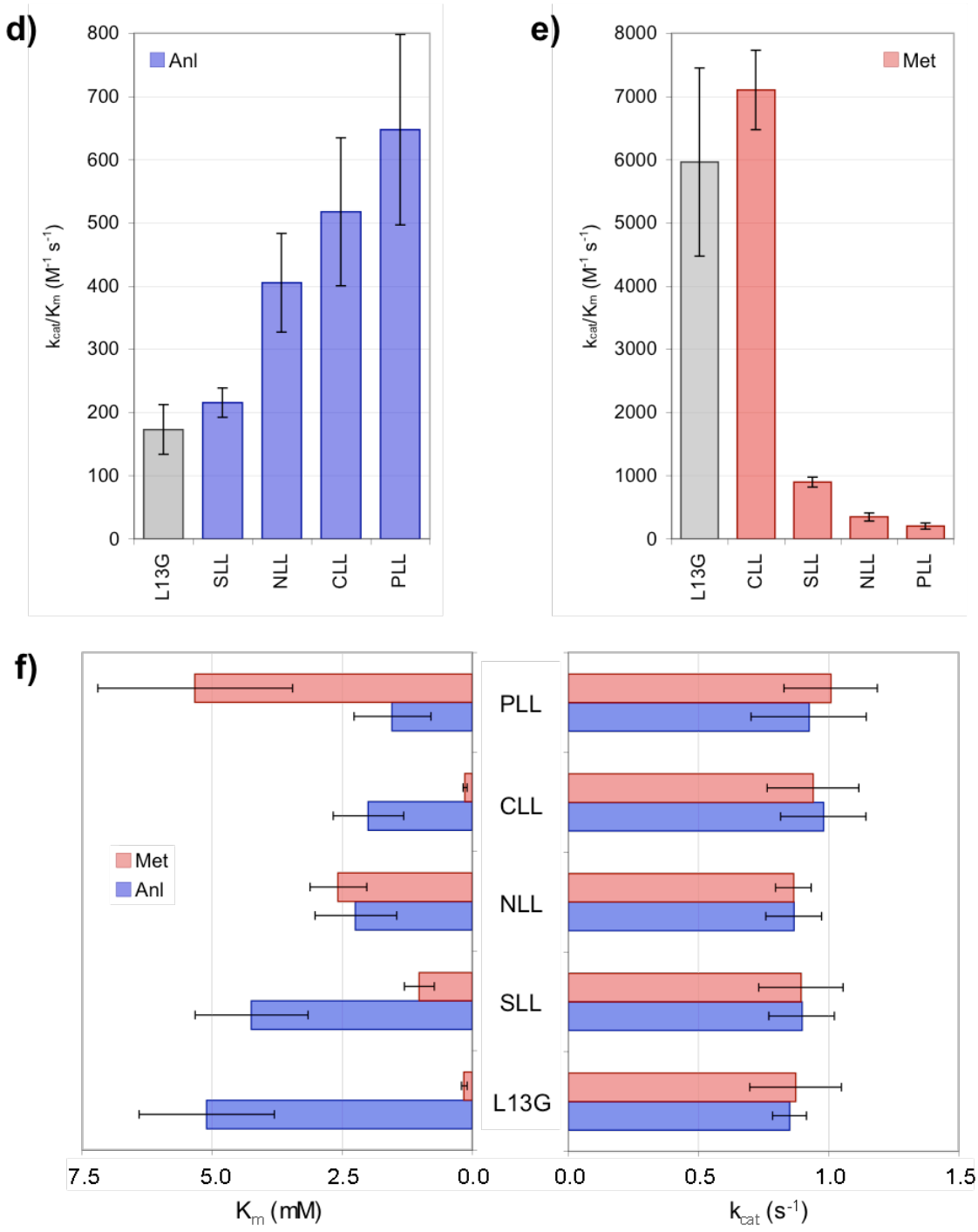


Figure 2.13.

Correlation between MetRS activation parameters for Anl and EC50 values obtained from cell-surface labeling experiments.

The relationship between the logarithm of EC50 values from cell-surface labeling experiments and the logarithm of enzyme activity toward Anl (k_{cat}/K_m) is shown for a series of mutants. This analysis suggests a strongly coupled relationship between fluorescence labeling of cells carrying MetRS mutants and the Anl activation kinetics of these mutants (R^2=0.77).

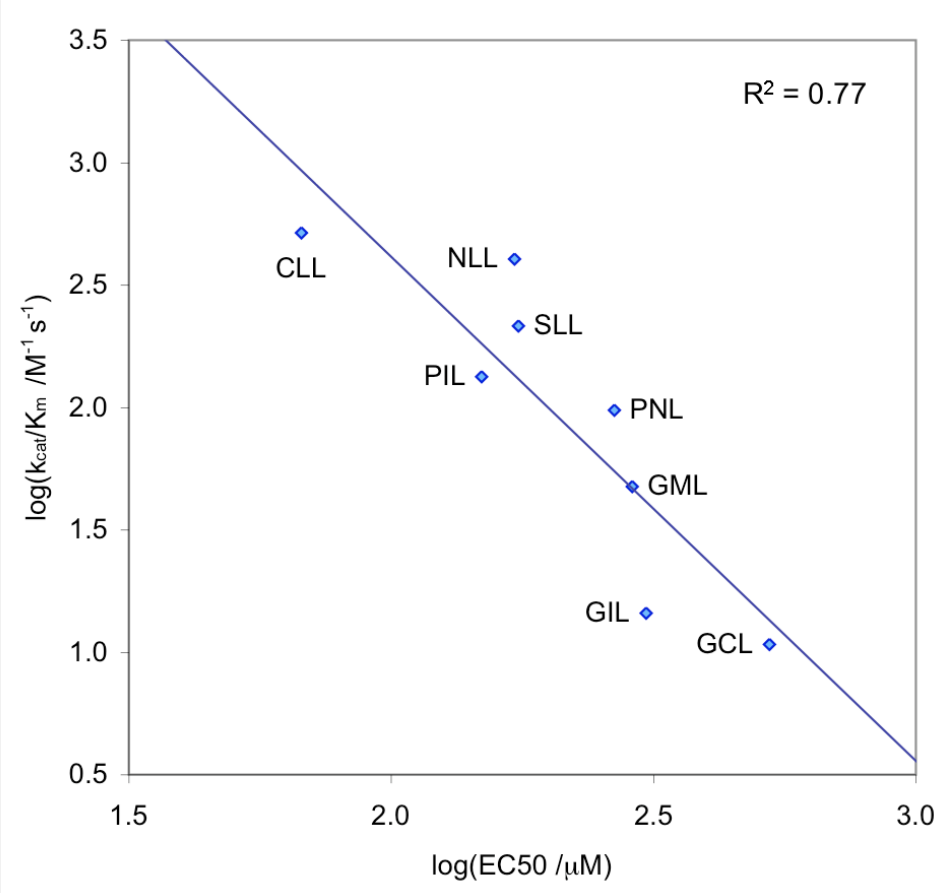


Figure 2.14.

Distribution of mutations selected at each randomized site on MetRS.

From 162 active clones isolated through screening the naïve library, 41 distinct MetRS mutants that show activity towards Anl were identified. (See Appendix A for the complete list.) Distribution of mutations observed at each randomized site among the 41 mutants is displayed**.** Substituted residues at each site are ordered by decreasing hydrophobicity according to the Kyte-Doolittle scale [41]. Based on this scale, residues are grouped as “hydrophobic” (>1.0; shades of yellow) or “hydrophilic” (<–2.0; shades of blue). Residues that rank between these two groups are shown in shades of green. Mutation distributions are compiled from a) all isolated mutants (n=41), b) mutants observed in low-stringency screens (1.0 mM Anl or worse; n=28), and c) mutants observed in high-stringency screens (0.3 mM Anl or better; n=13). Results show that mutations to aliphatic non-polar residues at positions 260 and 301 are compatible with Anl activity. In contrast, small residues with more hydrophilic character are preferred at position 13.

Figure 2.14. (continued)

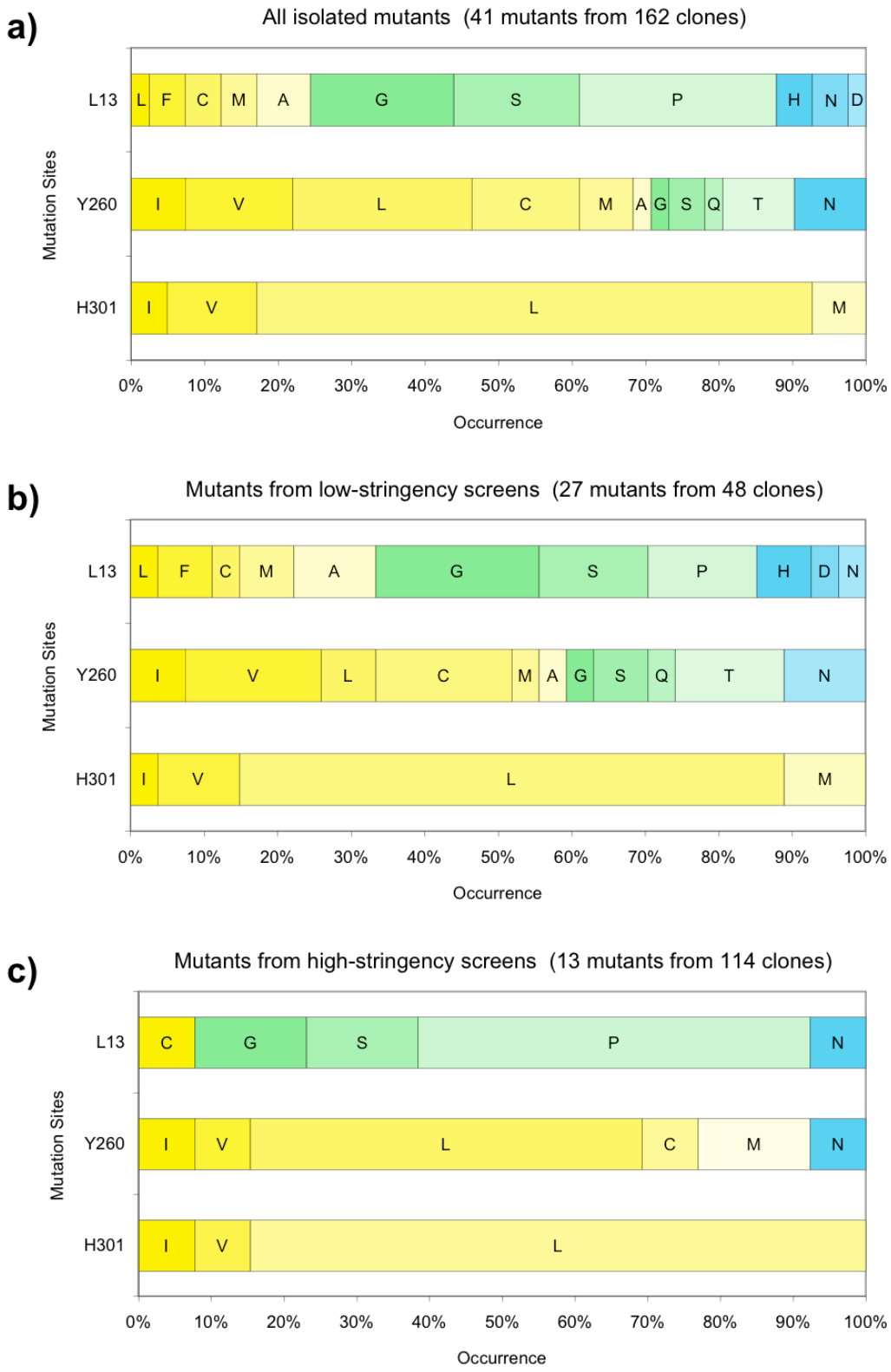


Figure 2.15.

Comparison of the amino-acid binding sites of the ligand-bound crystal structures of MetRS-SLL and wild-type MetRS.

The ligand, Anl or Met (yellow), is shown inside the MetRS active site. The positions 13, 260 and 301, which were randomized in this study, are highlighted in cyan. Crystal waters are shown as red spheres. Models were generated using PyMol.

a) The Anl-bound crystal structure of the SLL mutant at 1.5 Å resolution [37], focusing on the hydrogen bond between the azide and a conserved water molecule.

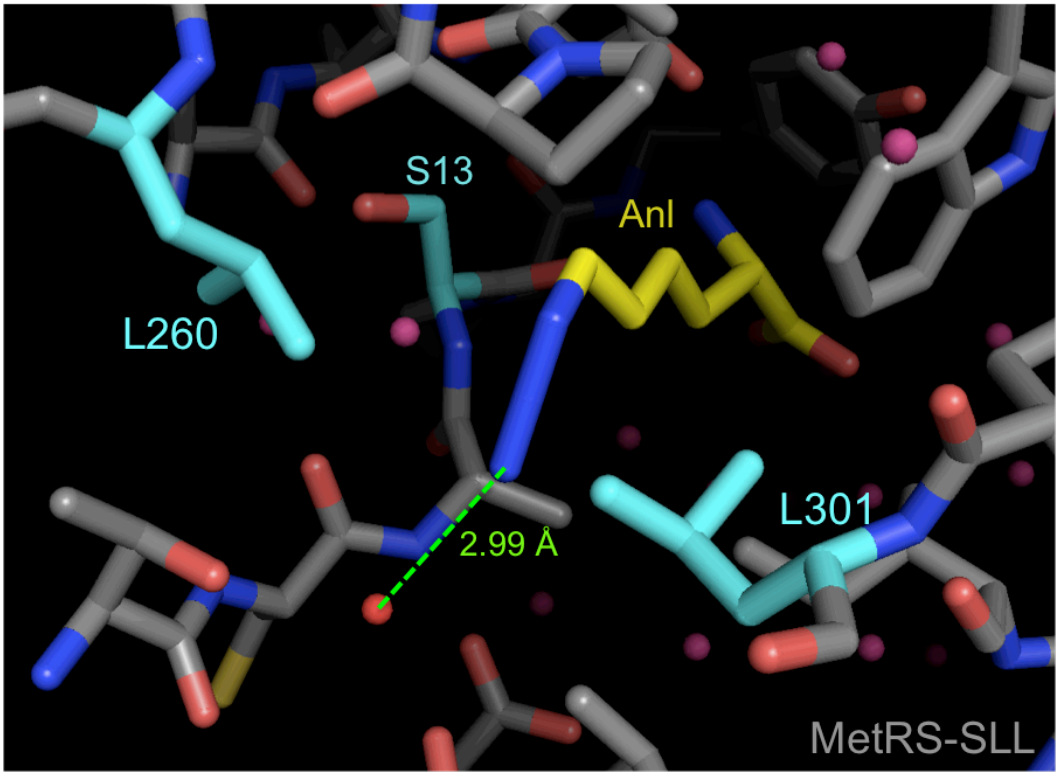


Figure 2.15. (continued)

b) The Met-bound crystal structure of the wild-type MetRS at 1.85 Å resolution [26], highlighting the interaction of the same conserved water molecule with the Y260 residue.

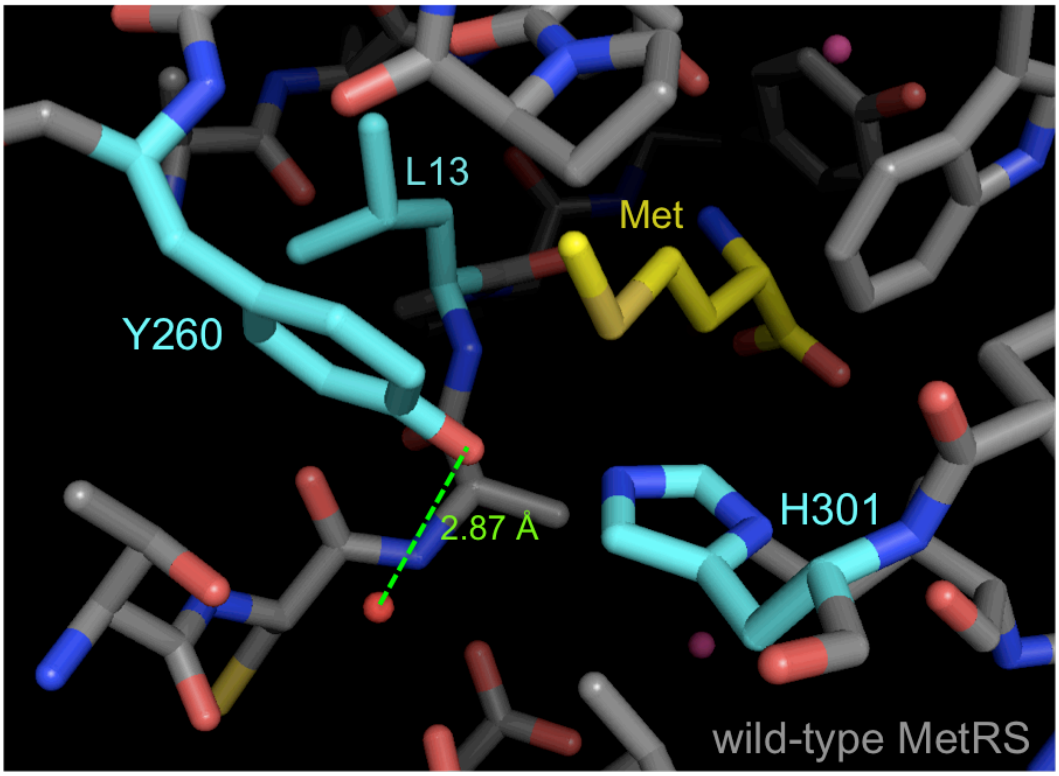


Figure 2.16.

Comparison of the ligand-bound and unbound structures of wild-type MetRS and the MetRS-SLL mutant.

The amino acid recognition pocket from four crystal structures of MetRS variants are shown: wild-type MetRS (PDB ID: 1QQT [38]), wild-type MetRS bound to Met (PDB ID: 1F4L [26]), ligand-free MetRS-SLL and MetRS-SLL bound to Anl [37]. The ligand, Anl or Met, is shown in yellow inside the MetRS binding site. On each panel, the model of the ligand-free enzyme is shown on the top, whereas the bottom model is that of the ligand-bound enzyme. The positions 13, 260 and 301, which were randomized in this study, are highlighted in cyan. The positions W253 and F300, side chains of which undergo a conformational change upon ligand binding in the wild-type structure, but not in the MetRS-SLL mutant, are displayed in green. Crystal waters are omitted in these models. All models were generated using PyMol.

Figure 2.16. (continued)

a) Wild-type MetRS binding pocket, with and without the ligand, methionine.

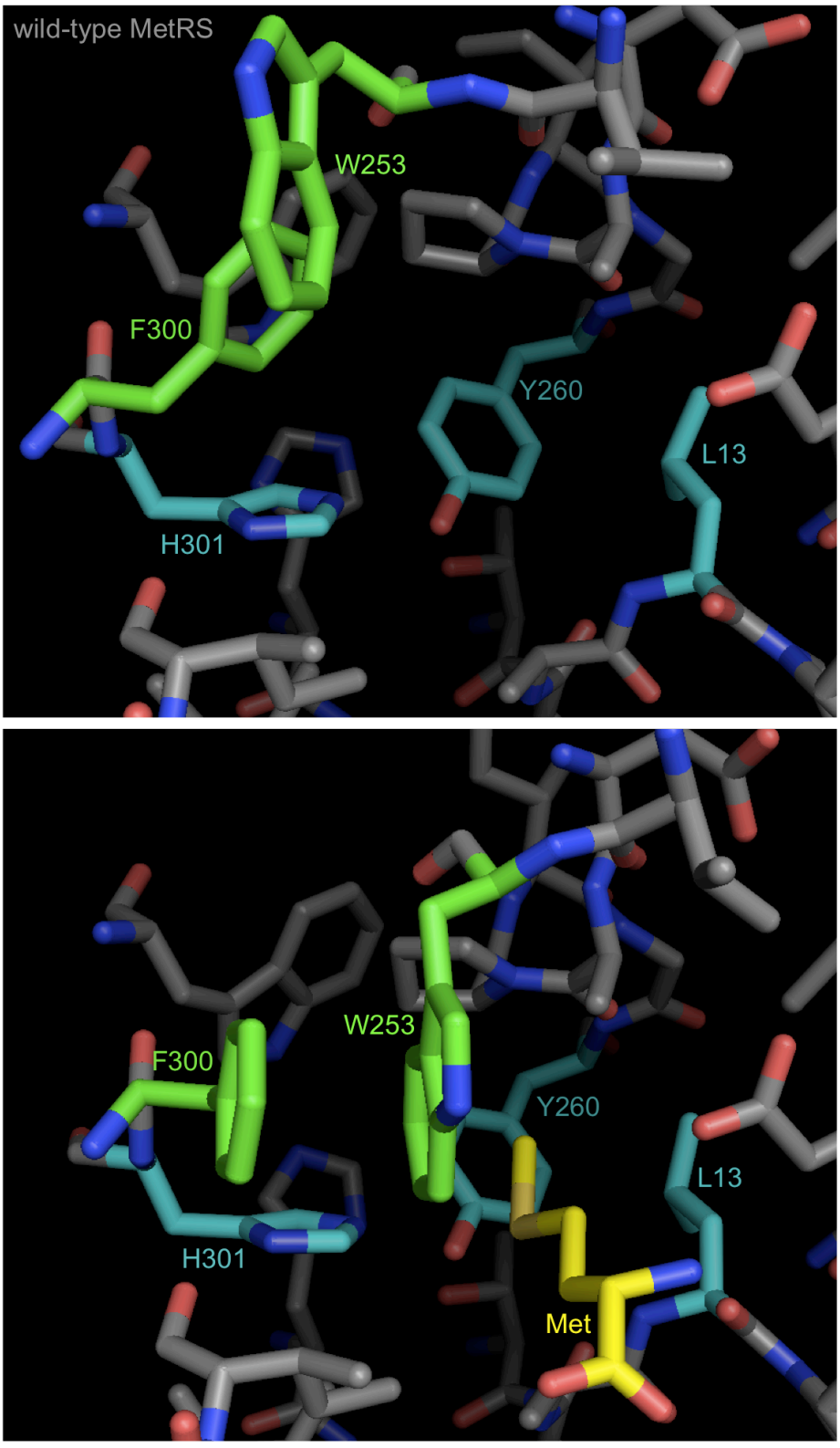


Figure 2.16. (continued)

b) MetRS-SLL binding pocket, with and without the ligand, azidonorleucine.

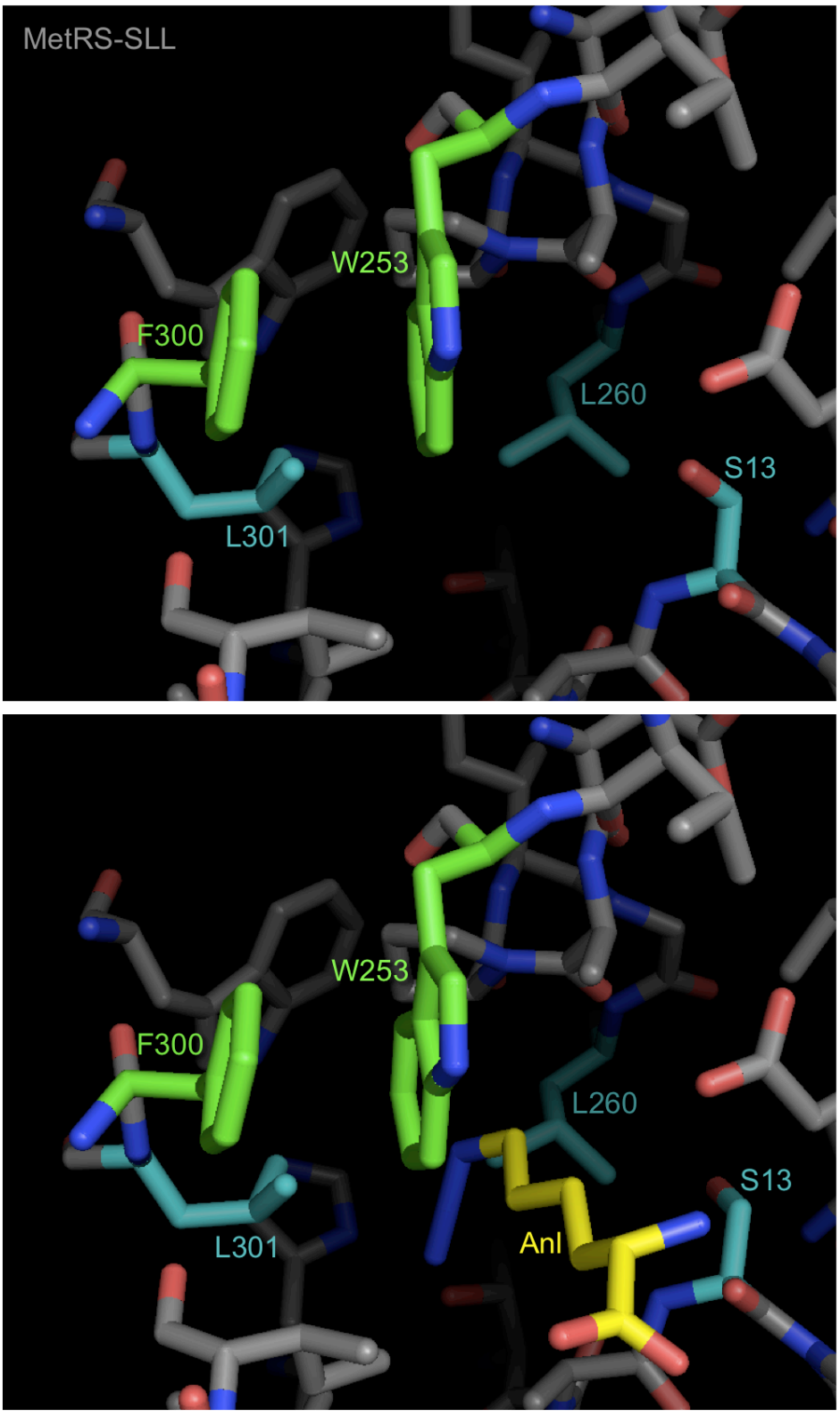
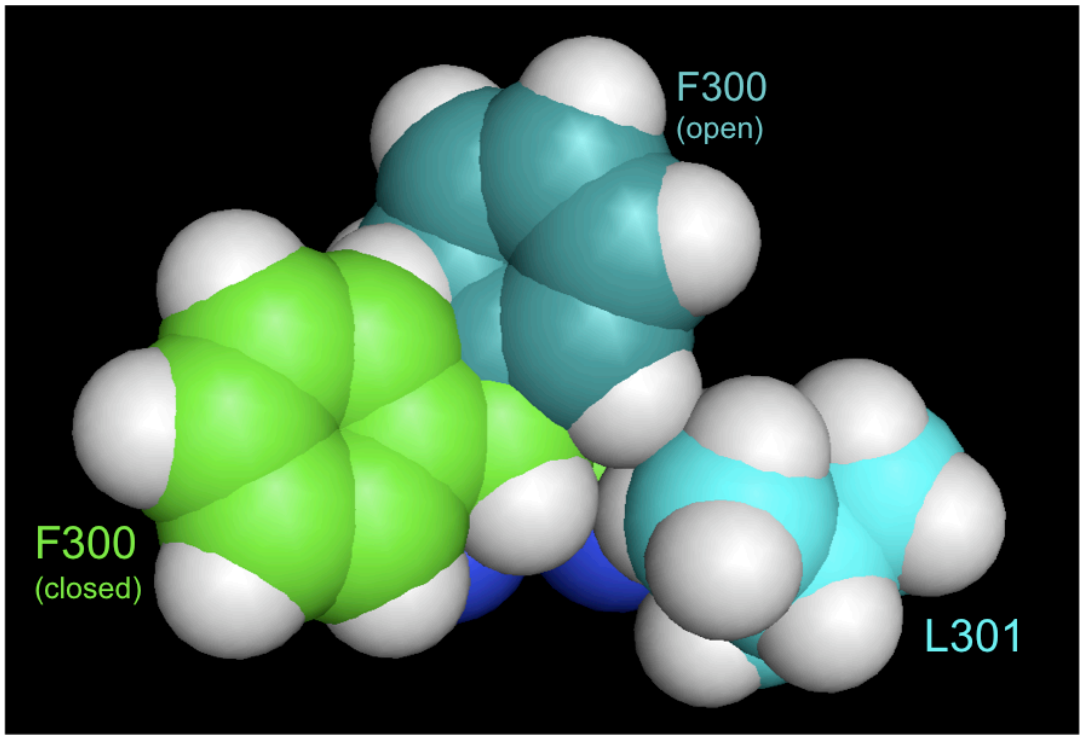


Figure 2.16. (continued)

c) The orientation of the F300 residue (teal) in the open conformation of the wild-type MetRS is shown superimposed with the conformations of the F300 (green) and L301 (cyan) residues observed in the crystal structure of the MetRS-SLL apo-enzyme, which takes the closed conformation. The backbone coordinates of the residues 300 and 301 from the crystal structures of the MetRS wild-type and SLL mutant were structurally aligned using Swiss-PDB Viewer. The open conformation of F300 is seen to be incompatible with the orientation of the L301 residue, due to clashes between their side chain Cδ atoms.



References

[1] Bilgicer B, Xing X and Kumar K, "Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores," *J Am Chem Soc*. **2001**, *123*:11815.

[2] Tang Y and Tirrell DA, "Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host," *J Am Chem Soc*. **2001**, *123*:11089.

[3] Son S, Tanrikulu IC and Tirrell DA, "Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues," *Chembiochem*. **2006**, *7*:1251.

[4] Bae JH, Rubini M, Jung G, Wiegand G, Seifert MH, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R and Budisa N, "Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins," *J Mol Biol*. **2003**, *328*:1071.

[5] Budisa N, Rubini M, Bae JH, Weyher E, Wenger W, Golbik R, Huber R and Moroder L, "Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical ph sensors," *Angew Chem Int Ed Engl*. **2002**, *41*:4066.

[6] Eichler JF, Cramer JC, Kirk KL and Bann JG, "Biosynthetic incorporation of fluorohistidine into proteins in *E. coli*: A new probe of macromolecular structure," *Chembiochem*. **2005**, *6*:2170.

[7] Bae JH, Alefelder S, Kaiser JT, Friedrich R, Moroder L, Huber R and Budisa N, "Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein x-ray crystallography," *J Mol Biol*. **2001**, *309*:925.

[8] Hendrickson WA, Horton JR and LeMaster DM, "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure," *EMBO J*. **1990**, *9*:1665.

[9] Zhang K, Diehl MR and Tirrell DA, "Artificial polypeptide scaffold for protein immobilization," *J Am Chem Soc*. **2005**, *127*:10136.

[10] Carrico IS, Maskarinec SA, Heilshorn SC, Mock ML, Liu JC, Nowatzki PJ, Franck C, Ravichandran G and Tirrell DA, "Lithographic patterning of photoreactive cell-adhesive proteins," *J Am Chem Soc*. **2007**, *129*:4874.

[11] Beatty KE, Liu JC, Xie F, Dieterich DC, Schuman EM, Wang Q and Tirrell DA, "Fluorescence visualization of newly synthesized proteins in mammalian cells," *Angew Chem Int Ed Engl*. **2006**, *45*:7364.

[12] Beatty KE, Xie F, Wang Q and Tirrell DA, "Selective dye-labeling of newly synthesized proteins in bacterial cells," *J Am Chem Soc*. **2005**, *127*:14150.

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

[14] Baskin JM and Bertozzi CR, "Bioorthogonal click chemistry: Covalent labeling in living systems," *Qsar Comb Sci*. **2007**, *26*:1211.

[15] Kiick KL, Saxon E, Tirrell DA and Bertozzi CR, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proc Natl Acad Sci U S A*. **2002**, *99*:19.

[16] Saxon E and Bertozzi CR, "Cell surface engineering by a modified staudinger reaction," *Science*. **2000**, *287*:2007.

[17] Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng J, Jaunbergs J, Weinbaum C, Tamanoi F, Falck J and Zhao Y, "A tagging-via-substrate technology for detection and proteomics of farnesylated proteins," *Proc Natl Acad Sci U S A*. **2004**, *101*:12479.

[18] Hang HC, Geutjes EJ, Grotenbreg G, Pollington AM, Bijlmakers MJ and Ploegh HL, "Chemical probes for the rapid detection of fatty-acylated proteins in mammalian cells," *J Am Chem Soc*. **2007**, *129*:2744.

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

[20] Link AJ, Vink MK, Agard NJ, Prescher JA, Bertozzi CR and Tirrell DA, "Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids," *Proc Natl Acad Sci U S A*. **2006**, *103*:10180.

[21] Agard NJ, Prescher JA and Bertozzi CR, "A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of blomolecules in living systems," *J Am Chem Soc*. **2004**, *126*:15046.

[22] Link AJ, Vink MK and Tirrell DA, "Presentation and detection of azide functionality in bacterial cell surface proteins," *J Am Chem Soc*. **2004**, *126*:10598.

[23] Ghosh G, Brunie S and Schulman LH, "Transition state stabilization by a phylogenetically conserved tyrosine residue in methionyl-tRNA synthetase," *J Biol Chem*. **1991**, *266*:17136.

[24] Yoo TH and Tirrell DA, "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 Engl*. **2007**, *46*:5340.

[25] Kiick KL, Weberskirch R and Tirrell DA, "Identification of an expanded set of translationally active methionine analogues in *Escherichia coli*," *FEBS Lett*. **2001**, *502*:25.

[26] Serre L, Verdon G, Choinowski T, Hervouet N, Risler JL and Zelwer C, "How methionyl-tRNA synthetase creates its amino acid recognition pocket upon l-methionine binding," *J Mol Biol*. **2001**, *306*:863.

[27] Crepin T, Schmitt E, Mechulam Y, Sampson PB, Vaughan MD, Honek JF and Blanquet S, "Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase," *J Mol Biol*. **2003**, *332*:59.

[28] Link AJ, Unpublished results, **2006**

[29] Firth AE and Patrick WM, "Statistics of protein library construction," *Bioinformatics*. **2005**, *21*:3314.

[30] Alba BM and Gross CA, "Regulation of the *Escherichia coli* sigma-dependent envelope stress response," *Mol Microbiol*. **2004**, *52*:613.
[31] Kurland CG and Dong H, "Bacterial growth inhibition by overproduction of protein," *Mol Microbiol*. **1996**, *21*:1.
[32] Wagner S, Bader ML, Drew D and de Gier JW, "Rationalizing membrane protein overexpression," *Trends Biotechnol*. **2006**, *24*:364.
[33] Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, Nygren PA, van Wijk KJ and de Gier JW, "Consequences of membrane protein overexpression in *Escherichia coli*," *Mol Cell Proteomics*. **2007**, *6*:1527.
[34] Ngo JT, Champion JA, Mahdavi A, Tanrikulu IC, Beatty KE, Connor RE, Yoo TH, Dieterich DC, Schuman EM and Tirrell DA, "Cell-selective metabolic labeling of proteins," *Unpublished manuscript*. **2009**,
[35] Kiick KL and Tirrell DA, "Protein engineering by *in vivo* incorporation of non-natural amino acids: Control of incorporation of methionine analogues by methionyl-tRNA synthetase," *Tetrahedron*. **2000**, *56*:9487.
[36] Tchertanov L, "Structural metrics relationships in covalently bonded organic azides," *Acta Crystallogr B*. **1999**, *55*:807.
[37] Schmidt E, Tanrikulu IC, Yoo TH, Panvert M, Tirrell DA and Mechulam Y, "Switching from an induced fit to a lock and key mechanism in an aminoacyl-tRNA synthetase with modified specificity," *Unpublished manuscript*. **2009**,
[38] Mechulam Y, Schmitt E, Maveyraud L, Zelwer C, Nureki O, Yokoyama S, Konno M and Blanquet S, "Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features," *J Mol Biol*. **1999**, *294*:1287.
[39] Ibba M and Soll D, "Aminoacyl-tRNA synthesis," *Annu Rev Biochem*. **2000**, *69*:617.
[40] Nakanishi K, Ogiso Y, Nakama T, Fukai S and Nureki O, "Structural basis for anticodon recognition by methionyl-tRNA synthetase," *Nat Struct Mol Biol*. **2005**, *12*:931.
[41] Kyte J and Doolittle RF, "A simple method for displaying the hydropathic character of a protein," *J Mol Biol*. **1982**, *157*:105.

Chapter 3

Computational Modeling and Design of Methionyl-tRNA Synthetase Mutants for Activity Toward Azidonorleucine

Introduction

Aminoacyl-tRNA synthetases (AARS) with altered substrate specificities have enabled the introduction of many noncanonical amino acids that are translationally inactive the wild-type enzymes into proteins [1-3]. The use of combinatorial selection and evolution techniques for the optimization of amino acid binding pockets for new substrates has especially been successful in engineering new AARS activities. Schultz and coworkers have developed an effective strategy for identifying AARS variants selective for a non-natural substrate and applied this to tyrosyl- [4, 5], and leucyl-tRNA synthetases [6, 7] to achieve site-specific incorporation. Tirrell and coworkers have reported two complementary screening strategies for the discovery of methionyl-tRNA synthetase (MetRS) mutants for the global replacement of methionine (Met). [8, 9] By screening a saturation mutagenesis library, Link et al. have discovered a series of *E. coli* MetRS mutants that allow the incorporation of azidonorleucine (Anl) [8, 10]. (Figure 3.1.a; Table 3.1) Further optimization of this screening strategy has revealed a diverse set of MetRS mutants that enable efficient incorporation of Anl *in vivo*. (See Chapter 2.)

Computational methods utilized for the identification of new AARS activities have focused on optimizing the binding interactions of the new substrate with the binding pocket. Fidelity of phenylalanyl- [11], methionyl- [12], and seryl-tRNA synthetases [13] to binding their natural substrates was demonstrated computationally, and *in vivo* activities of phenylalanyl- [14] and methionyl-tRNA synthetases [12] toward noncanonical amino acids were correctly predicted by computed binding energies for the non-natural substrates. Mutations in a *M. jannaschii* tyrosyl-tRNA synthetase variant selected for *O*-methyl-tyrosine activity were correctly identified in the immediate vicinity of the ligand through the comparative evaluation of binding interactions for the natural and non-natural substrates [15]. A computationally designed variant of *E. coli* phenylalanyl-tRNA synthetase was reported to enable the *in vivo* incorporation of *p*-acetylphenylalanine into proteins [16].

Although attempts were made to design AARS binding sites, mutational

data to verify the computational procedures has been the limiting factor in these studies. Therefore, the wealth of mutational data available for *E. coli* MetRS provides a unique opportunity for verifying the accuracy of our computational capabilities, and designing new mutants for Anl incorporation. Multiple high-resolution x-ray crystal structures are available for this enzyme [17-19] (Figure 3.1.b) and for a mutant MetRS complexed with Anl [20]. (Figure 3.1.c) In this report, we demonstrate a good correlation between computed binding energies and *in vitro* activation data for a series of MetRS mutants. We also compare the results from a saturation-mutagenesis library screen performed *in silico* with the results obtained with the same library *in vivo*. Modeling and design of the MetRS binding site were initiated in 2005, before the discovery of the MetRS mutants discussed in Chapter 2. This work was revisited in 2008 after additional mutation data became available. Initial studies focused on data reported by Link et al. [8, 10] Validation of the computational protocol was followed by the computational design of three positions (L13, P257 and Y260) in the MetRS binding pocket. In the text, this part of the work will be referred to as the "2005 STUDY" and the accompanying design study as the "LPY design." Following the availability of new data from screens of a saturation-mutagenesis library on positions L13, Y260, and H301 for Anl incorporation, we showed that our binding predictions for Anl agree with new data. We later carried out the design of the MetRS binding site at residues L13, Y260 and H301. This second part of this study will be referred to in the text as the "2008 STUDY," and the design work as the "LYH design." The triple mutants of MetRS discussed in the 2008 STUDY are named based on the mutations they carry at positions L13, Y260 and H301, so that the MetRS variant bearing the L13N-Y260L-H301L mutations is referred to as the "NLL" mutant.

Materials and Methods

Simulation parameters

All minimizations in the 2005 STUDY were carried out using MPSIM [21] through conjugate gradient minimization. Non-bond interactions were calculated

using the cell-multipole method [22]. Minimizations were assumed to have converged when a 0.5 kcal/mol/Å RMS-force was reached. Coulombic forces were calculated using a dielectric constant of $\varepsilon = 2.5$ for the interior of the protein, and minimizations were carried out under solvent-generalized Born (SGB) continuum solvation. The forces on the protein were described by the DREIDING force field [23] and CHARMM22 charges [24] were used for the protein atoms. The charges on the native ligand, as well as its analogs, were obtained from the Mulliken charges based on the molecular orbitals calculated by quantum mechanics (QM). All QM calculations were run using JAGUAR 4.0 (Schrodinger) at the Hartree-Fock level with a 6-31G** basis set. The charges were obtained after the geometry of the ligand was optimized with forces calculated under Poisson-Boltzmann (PB) continuum solvation [25].

Simulations in the 2008 STUDY were carried out as was described for the 2005 STUDY above, with few differences. For the 2008 STUDY, a more recent version of the DREIDING force field, with newly parameterized hydrogen-bond terms, was employed [26]. This force field uses new hydrogen bond equilibrium distances and well depths that were fit to reproduce QM geometries and energies obtained at the X3LYP [27] level with the aug-cc-pVTZ-(-f) basis set. Minimizations were carried out without a solvation potential, and coordinates were assumed to have converged when a 0.2 kcal/mol/Å RMS-force was reached. Charges on the native ligand, or ligand analogs were obtained from Mulliken charges revealed by QM calculations using B3LYP with the 6-31G** basis set, determined after geometry optimization under PB solvation. All QM calculations were carried out using JAGUAR 7.0 (Schrodinger).

Calculations for the modeling and design of AARS for noncanonical amino acids, and the analysis of the results were accomplished with the aid of a series PERL and PYTHON scripts. A list of the software written for and used in this study is provided in Appendix D.

Structure preparation

The crystal structures of the free *E. coli* MetRS at 2.03 Å resolution, and

the methionine-bound enzyme at 1.85 Å resolution were obtained from PDB database, using the PDB IDs 1QQT [18] and 1F4L [17], respectively. In the 2005 STUDY, BIOGRAF (Molecular Simulations) was used to add hydrogen atoms to the structures. Annealing dynamics were performed on the supplemented hydrogens in order to optimize their coordinates. This was followed by an all-atom minimization of the structures under SGB solvation. In the resulting ligand-bound structure, the residues that have atoms within 10 Å of the native ligand, methionine, were identified to be in the "binding region." We focused exclusively on this region of the enzyme in the remainder of this study. Crystal structures were also optimized in the absence of any heteroatoms (crystal waters and ions) through the same procedure described above. Minimized structures in the absence and presence of heteroatoms showed minimal differences in their coordinates (less than 0.2 Å RMSD over all atoms, as well as over side chain atoms within 6 Å of the ligand). Therefore, the structure without any solvent molecules was used in modeling. Application of mutations and determination of binding energies were performed starting with this minimized structure of the enzyme. (**s0** in Figure 3.2)

In addition to the two aforementioned crystal structures of the *E. coli* MetRS [17, 18], two new MetRS structures were used in the 2008 STUDY: Crystal structures of the SLL mutant of *E. coli* MetRS (L13S-Y260L-H301L) were obtained from Y. Mechulam [20] in free and Anl-bound form at 1.7 and 1.5 Å resolution, respectively. Hydrogen atoms were added and rotamer orientations in the crystal structure were corrected using WHAT IF [28] prior to any minimization. Structures were minimized in vacuum without any constraints on the protein or the heteroatoms. Residues in the final ligand-bound structure that are within 10 Å of the Anl ligand were marked as the "binding region," and, as before, this structure was used as the base model (**s0**; Figure 3.2) for any subsequent calculations.

Preparation of the Anl-MetRS complexes

Before the availability of an Anl-bound crystal structure of MetRS, the

conformation of Anl inside the MetRS binding site was generated based on the coordinates of the enzyme-bound Met structure by mutating, adding or deleting the appropriate atoms on the Met side chain. This way, both the native ligand and the analog share the same coordinates for the ligand backbone and the same general binding mode. This is done to ensure that the catalytic portion of the enzyme is not affected by changes on the ligand side chain. Enzyme-analog complexes were prepared by replacing the native ligand in the binding site of minimized enzyme with the new ligand.

Due to the close structural similarity between Met and Anl, both are expected to bind the enzyme in similar orientations overall. However, the three extra heavy atoms on the azide group need to be placed carefully to optimize packing and hydrogen-bonding interactions, and getting accurate binding energies. Different conformations of the enzyme-bound Anl are explored using rotamer libraries. Anl rotamers are prepared inside the MetRS binding pocket. To allow rotamers to take advantage of available space in the binding site after any mutations, positions without a wild-type identity are mutated to glycine prior to rotamer generation.

Rotamers were generated using MOLECULEGL [29] in the 2005 STUDY. This program grows the ligand in the binding site from a specified anchor (Cα atom of the ligand) to generate a diverse set of rotamers, and reduces the diversity based on a diversity cutoff. In this study 10 rotamers were generated at 0.3 Å RMSD diversity to represent different Anl conformations. The rotamer generation procedure was modified in the 2008 STUDY. Anl conformations were sampled exhaustively using BIOGRAF (Molecular Simulations) inside the MetRS binding site carrying the L13G, Y260G, H301G mutations. Trans and gauche dihedrals were allowed for C-C bonds, whereas the C-N_3 bond was sampled at every 30 degrees. Rotamers were ranked by energy using MPSIM and the best 20 rotamers were used.

Mutations were introduced into MetRS using MODULASIM-SCREAM [30] in the 2005 STUDY, and by SCREAM [26] in the 2008 STUDY. Rotamer libraries at 1.0 Å and 0.2 Å RMSD diversity were used for the placement of 20 natural

amino-acid rotamers in these studies, respectively. When placing side chains using SCREAM, the flat-bottom van der Waals (VdW) potential it employs was disabled to allow accurate placements while using a very high-resolution rotamer library. Mutations were applied to MetRS after the Anl rotamer was placed into the binding site. The energy of the ligand was analyzed for all Anl rotamers, either after minimization, or immediately, and the best ligand-enzyme complex was selected.

Binding-energy calculations

Several minimization steps were carried out for each enzyme-ligand complex (**s2**) before the binding energies were calculated as outlined in Figure 3.2. In the first step, the potential energy of the ligand side chain is optimized with the positions of the ligand backbone and the binding region fixed during minimization, to produce **s3**. This step serves to resolve any large clashes between the ligand and the binding pocket and is performed to prevent the ligand from flipping out of the binding site in the subsequent steps. Next, the binding region and the ligand side chain atoms on **s3** are minimized, keeping the positions of the ligand backbone and the residues on the enzyme that recognize the ligand backbone fixed. This step allows reorganization in the binding pocket while keeping the interactions of the ligand backbone with the enzyme intact to ensure that the reaction center for the adenylation reaction is not perturbed. The ligand binding energy is calculated on the resulting structure, **s4**. This binding energy represents the binding energy of the ligand in a conformation that allows its activation by the MetRS.

The strength of binding for each ligand was evaluated based on the vertical binding energies (VBEs) calculated according to the equation:

$$\Delta\Delta H_{binding} = \Delta H_{complex} - (\Delta H_{enzyme} + \Delta H_{ligand}) . \qquad (3.1)$$

The complex structure, **s4**, was used to calculate $\Delta H_{complex}$. For determining ΔH_{enzyme} and ΔH_{ligand}, **s4** was spilt into its components, and the energy of the components were determined to calculate the direct vertical binding-energy (dVBE). Alternatively, the components (**s5**) were minimized for 10 steps and the

resulting structures (**s6**) were used to calculate the relaxed vertical binding energy (rVBE). Any strain imposed on the enzyme or the ligand in the ligand-bound conformation is reflected in the rVBEs calculated. In the 2005 STUDY, where comparisons were made across different ligands, complexes evaluated based on rVBE. In the 2008 STUDY, binding energies for only the non-natural analog was determined. To avoid the uncertainty of a ligand reference state, only "two-point" dVBEs that omit ΔH_{ligand} from the binding-energy expression were calculated.

When the construction of the ligand-enzyme complex requires mutations in the binding site, mutations were applied to the **s2** structure separately in the presence of each ligand rotamer to obtain a binding site compatible with the ligand orientation. In the 2005 STUDY, the best rotamer for the ligand was selected by the rVBEs calculated on **s4** and **s6**. In the 2008 STUDY, non-bond interactions of the ligand side chain with the binding region was determined for each ligand rotamer inside the **s4** structure. The structure showing the best interactions with the ligand rotamer was selected for binding-energy analysis.

For a given structure, ΔH was calculated as:

$$\Delta H = \Delta H_{bonding} + \Delta H_{non\text{-}bond} + \Delta H_{solvation} \ . \qquad (3.2)$$

The $\Delta H_{bonding}$ and $\Delta H_{non\text{-}bond}$ terms are calculated based on the DREIDING force field. The $\Delta H_{bonding}$ is made up of terms describing bond stretching and bending, torsions, as well as an inversion term. The non-bond component, $\Delta H_{non\text{-}bond}$, which is also referred to as "interaction energy", was calculated as the sum of the Van der Waals (E_{vdw}; VdW), Coulomb (E_{coul}), and hydrogen bonding (E_{hb}; H-bond) components:

$$\Delta H_{non\text{-}bond} = \Delta E_{vdw} + \Delta E_{coul} + \Delta E_{hb} \ . \qquad (3.3)$$

The interaction energy terms between two groups of atoms, such as the residue *K* and ligand *L*, are calculated as:

$$E_{vdw}(K,L) = \sum_{i \in K, j \in L} D_0 \left[\left(\frac{r_0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_0}{r_{ij}} \right)^{6} \right], \qquad (3.4)$$

$$E_{coul}(K,L) = \sum_{i \in K, j \in L} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} , \quad (3.5)$$

$$E_{hb}(K,L) = \sum_{i \in K, j \in L} D_{HB}\left[5\left(\frac{r_{HB}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{HB}}{r_{ij}}\right)^{10}\right]\cos^4\theta , \quad (3.6)$$

where each value is calculated as a sum over all atoms *i* on *K* and *j* on *L*, using the interatomic distance, r_{ij}, partial charges on *i* and *j*, q_i and q_j, and in the case of hydrogen bonds, the donor-hydrogen-acceptor angle, θ. Other parameters used are D_0 (the geometric mean of equilibrium well depth for *i* and *j*), r_0 (the geometric mean of the VdW radii for *i* and *j*), ε (the dielectric constant), D_{HB} (the equilibrium hydrogen-bond well depth), and r_{HB} (the equilibrium hydrogen-bond distance).

Different solvation methods were employed for the calculation of the solvation component of binding energies. In the 2005 STUDY, all minimizations were carried out under solvent generalized Born (SGB) solvation [31], whereas the binding energies were determined using either analytical-volume generalized Born (AVGB) [32], or SGB solvation. Solvation was not applied during minimizations in the 2008 STUDY, but PB solvation calculated using DELPHI 4.1 [33] was included in the binding energies.

Design of MetRS binding sites for AnI

The sites in the MetRS binding site were selected based on the positions randomized in the library screening experiments of Link et al. [8] From the four positions (L13, P257, Y260 and H301) randomized three positions were selected for the *in silico* library in both the 2005 STUDY (L13, P257 and Y260), and the 2008 STUDY (L13, Y260 and H301). These computational "saturation mutagenesis" libraries were called LPY- and LYH-designs, respectively. Computational design through the randomization of three positions in the binding site requires a maximum of 20^3 = 8000 mutants to be evaluated. This number is feasible for both a computational and an experimental study.

Both design studies employed the methods described above in sequence: Starting from a fully minimized enzyme structure ligand rotamers were inserted

into the binding site. Side chains for the selected positions were placed in the presence of the ligand rotamer. First the ligand side chain, then the whole binding region was minimized. Using ligand interaction of binding energies the best ensemble of conformations for the binding site residues and ligand was identified and binding energies were analyzed. All calculations were run on Beowolf clusters featuring 2.2 GHz to 3.06 GHz Intel Xeon processors with 1 GB memory installed per processor. The calculations were run using the **s1** structure as the base structure. The 2005 STUDY was initialized using the fully minimized wild-type MetRS structure [17] after removing all solvent from the model. The 2008 STUDY was started from a fully minimized model of the MetRS-SLL [20]. From the 1161 water molecules resolved in this structure, only the 92 water molecules that interact best (stronger than −10 kcal) with the enzyme were included in the design calculation. During dVBE calculation these water molecules were assumed to be part of the enzyme.

Identification of low-fluorescence clones through library screening

The materials used and experimental procedures employed in this study were described previously in Chapter 2, and are explained only briefly here. A three-position saturation mutagenesis library designated LYH.1.0 was created by randomizing the L13, Y260 and H301 positions in the *metS* gene, which codes for MetRS, on the pAJL-20 plasmid [8]. In addition to constitutively expressing MetRS, the pAJL-20 plasmid also codes for an *E. coli* OmpC variant which carries surface-exposed mutations to methionine . The expression of OmpC was induced in the absence of Met and presence of Anl in the methionine-auxotrophic expression of host, M15MA [34]. Surface-exposed azide-groups were covalently linked to biotin through strain-catalyzed [3+2] azide-alkyne cycloaddition. The biotin-labeled cells were bound to fluorescently labeled avidin and analyzed on a fluorescence-activated cell sorter (FACS).

The LYH.1.0 library was screened after the expression of OmpC in the presence of 1.0 mM Anl to reveal the population LYH.1.1a. This population was grown and induced to express OmpC at 1 mM Anl for 30 min. After biotin

conjugation and fluorescent labeling of these cells, their analysis was carried out using flow cytometry. In order to isolate clones that display low fluorescent labeling levels, the gate on the fluorescence channel was set on the low-fluorescence shoulder of the labeled population. Cells between 31 and 37 fluorescence units comprising 2.1% of the total population was sorted. Clones from this population, designated LYH.6.2, was analyzed on FACS individually, and the identity of the MetRS mutants carried by these clones were determined through sequencing.

Characterization of the activities of MetRS variants toward Anl

The response of the cell-surface labeling of the LYH.6.2 population to the amount of Anl present in the expression medium was also determined. Extent of cell-surface labeling was determined after the LYH.6.2 population was induced to express OmpC in the presence of 0.3, 1.0 and 3.0 mM Anl, in the absence of Met. FACS histograms were obtained for each culture, and the median fluorescence of the fluorescent part of the population was determined. Resulting data was fit to the Hill equation using KaleidaGraph (v3.6, Synergy Software) after setting the minimum response to 1.60 fluorescence units. The EC50 values were obtained from the resulting least squares fit.

Tests for expression of protein in the presence of Anl were carried out using the medium-shift procedure described in Chapter 2. Using QuikChange site-directed mutagenesis, MetRS variants were cloned into the plasmid pAJL-61 (see Table 3.2 for primer sequences), which was used to test the incorporation of Anl into the N-terminally 6×His-tagged dihydrofolate reductase (DHFR) [8]. Protein expression was induced for 3.5 hours in synthetic media containing 2.0 mM Anl, but no Met, using M15MA as the host. DHFR expression was determined by SDS-PAGE analysis of whole cell lysates.

In vitro activation kinetics for three mutants identified from the population LYH.6.2 was measured, as was described in Chapter 2. Variants of the pMTY21 plasmid [9] carrying the 6×His-tagged MetRS mutants AQL, SNL and GVL were constructed using PCR with the MRS_BamHI and MRS_SalI-r primers and a

subsequent ligation. MetRS variants were expressed and purified, and their concentrations were determined. Amino acid activation assays were carried out as described previously [35, 36], Radiolabeled sodium ^{32}P-pyrophosphate was purchased from Perkin-Elmer Life Sciences. Enzymes were added to reactions at 6 to 8 μM concentration. Anl concentrations tested ranged from 2.0 to 64 mM. All data was fit to a Michaelis-Menten model using KaleidaGraph (v3.6, Synergy Software), and kinetic parameters were determined.

Results and Discussion

Discrimination of active MetRS variants among the Link et al. mutants: The 2005 STUDY

A variety of MetRS mutants were identified by Link et al. [8, 10] showing varying *in vivo* activities toward Anl. (Table 3.1) Using the methods outlined above, models of these mutants were constructed starting from the crystal structure of methionine-bound *E. coli* MetRS at 1.85 Å resolution. In order to determine how well the constructed models capture the experimental activities of these mutants, vertical binding energies were calculated for each complex. Energies with AVGB solvation obtained after optimization of the binding region through annealing dynamics are reported in Table 3.3. Although the binding energies fail to show a distinction between the L13G mutant, which shows the highest activity toward Anl, and other active mutants, the inactive mutants are clearly separated from the active set. These trends persist regardless of the solvation method of choice (AVGB or SGB) or the use of annealing dynamics.

Examination of the energies and the models show that hydrogen bonds made to the azide group on Anl distinguish the mutant enzymes from the wild-type enzyme. All mutants that carry the L13G mutation interact with Anl through a hydrogen bond between the backbone nitrogen on position 13 and the azide group. Upon methionine binding this backbone nitrogen aids in the recognition of the sulfur atom in the wild-type enzyme. (Figure 3.3.a) In the L13G mutant, this hydrogen bond is satisfied by an interaction with the first nitrogen of the azide group (Figure 3.3.c). For the Y260T mutant, formation of this hydrogen bond

increases the strain in the model. The VdW contribution to the binding energies shows that the wild-type enzyme provides more VdW stabilization than any of the mutants. This trend is unexpected since Anl is larger by three heavy atoms than the native ligand, Met. When Anl is placed into the binding site in an extended configuration, clashes with the P257 and Y260 residues are observed. (Figure 3.3.b) The extent of the VdW stabilization by the wild-type MetRS suggests that the wild-type active site has ample space for binding ligands as large as Anl. This observation agrees with previous findings, where trace amounts of Anl incorporation was detected on the *E. coli* cell surface through the wild-type MetRS [37].

Design of the MetRS binding site at three positions: The LPY design study

In order to determine if the discrimination observed between active and inactive mutants can be applied to predict new active mutants, the L13, P257, and Y260 positions were randomized *in silico* and the resulting mutants were evaluated for Anl binding. These three positions make up three of the four positions originally randomized by Link et al. [8], and were selected because their relative positions make cooperative interactions between these positions more likely. To simplify the sequence space to be searched, the set of amino acid mutations allowed at each site was restricted. Because Anl does not have a net charge, charged residues were not allowed at any site, with the exception of histidine. In addition, aromatic residues (Phe, Tyr, and Trp) were only allowed at position 260, since this site has ample room to fit aromatic residues. These restrictions reduce the number of mutants to be explored from 8,000 to 2,704. Using the L13G-P257G-Y260G mutant as the background, 10 Anl rotamers were generated using MOLECULEGL. Each rotamer was tested against the whole mutation set, and the best binding conformation of Anl was recorded for each mutant.

The distribution of binding energies obtained for the mutants tested is shown in Figure 3.4, each sequence ranked by its binding energy. From the three previously characterized MetRS variants present in this mutation set, the

wild-type MetRS and the Y260T mutants are known to be inactive toward Anl, whereas the L13G mutant can efficiently activate Anl. Indeed, the L13G mutant was ranked 5th by binding energy, at the top of the list. In contrast, the inactive MetRS variants, Y260T and wild-type, were listed more than 10 kcal/mol lower in binding energy than the L13G mutant. The top 15 mutants with the best binding energies are shown in Table 3.4. The top mutants cover a very narrow range of sequences from the available set, carrying small and polar mutations at positions 13 and 260 while maintaining wild-type identity at position 257. Mutants with the Y260N mutation display good hydrogen-bonding energies, counterbalanced by relatively poor bonding and VdW components.

A set of mutants that were predicted to bind Anl best were selected to be tested experimentally for *in vivo* Anl activation. The mutants ranked 1st, 2nd, 3rd, 4th, 6th, and 10th were constructed in the laboratory and introduced into pAJL-61 plasmid, which also encodes for DHFR. DHFR synthesis was induced in the methionine-auxotroph M15MA, in methionine-free medium supplemented with 2.0 mM Anl. The SDS-PAGE analysis of cell lysates following DHFR expression is shown in Figure 3.5. The L13G mutant is known to allow close to quantitative replacement of Met sites with Anl in the presence of 1.0 mM Anl [8]. This mutant is the only MetRS variant that supports DHFR synthesis at the conditions tested. DHFR production was not detected with other top ranking MetRS mutants. The appearance of high-ranking false positives may be due to the limited scope of the computational model. Some aspects of the biological system, such as binding site rearrangements linked to ligand binding, are not accounted for in the model and can lead to false positives. Alternatively, it is possible that the resolution of computational binding energies that are used to distinguish between mutants is lower than the resolution observed in experiments, such that two mutants that cannot be distinguished by binding energies might vary widely when their activities are determined in the laboratory. In order to better understand the limitations of the computational strategy on the MetRS system, additional mutation data was obtained through *in vivo* selection experiments. (Chapter 2)

Identification of MetRS mutants active toward Anl through library screening *in vivo*

New MetRS mutants that are active toward Anl were identified through the screening of a saturation mutagenesis library, constructed by randomizing the identities of residues 13, 260, and 301 on MetRS, following methods outlined by Link and coworkers [8]. In the wild-type enzyme, the residues Y260 and H301 are part of a hydrogen-bond network that is implicated in the recognition of the sulfur atom on methionine [17]. The residue L13 also donates a hydrogen bond to the methionine sulfur from the amide group on the backbone, and its mutation to glycine allows the recognition of Anl as a substrate by MetRS. Randomizing only three positions on the MetRS facilitates the complete coverage of the saturation mutagenesis library *in vivo* and *in silico*: It is easily possible to evaluate 20^3 = 8,000 mutations computationally, as well as constructing a library with a size sufficiently larger than the 32^3 = 32,768 genetic variants of MetRS in the laboratory. In theory, this allows for a one-to-one comparison of the results from the computation and experiments.

Details on the construction and the screening of the experimental saturation mutagenesis library, and the mutants identified and characterized through these efforts are discussed in detail in Chapter 2. The kinetic parameters for Anl activation by the MetRS mutants identified through this work can be found in Table 2.3. These studies focused on identifying MetRS mutants that can activate Anl when it is present at low concentrations in the media. In order to increase the spread in the activation data available for computation, a screen was carried out to identify MetRS variants that show lower activities than the mutants identified before. (Figure 3.6.a) The population LYH.1.1a was obtained by screening the top 1% most fluorescent clones from the naïve library. This screen, which enriches the population in the active clones, was followed by a second step, where a portion of the LYH.1.1a that displays suboptimal fluorescence labeling characteristics was isolated (LYH.6.2). The response of the fluorescence labeling on the LYH.6.2 population to increasing Anl concentrations in the media was determined, and an EC50 value of 1.58 mM was

obtained for this population. (Figure 3.6.b) This value is well above the EC50 values previously determined for any individual clone listed in Figure 2.8. The fluorescent labeling of the LYH.6.2 population following treatments by 0.3, 1.0, and 3.0 mM Anl all result in a single, uniform fluorescent population. Thus, similar Anl activation characteristics are expected from the members of this population.

The kinetic parameters of Anl activation by three of the MetRS mutants identified (AQL, GVL, and SNL) from the LYH.6.2 population were determined *in vitro*. (Figure 3.6.c; Table 3.5) When compared with past data, the activation kinetics for these mutants are at or below the activity levels exhibited by low-activity MetRS variants previously identified. Interestingly, the low k_{cat}/K_m were due to the poor k_{cat} values exhibited by these mutants. The Michaelis constants for the new mutants were comparable with or better than those obtained for mutants isolated by the 1.0 mM Anl screen, and only 3 to 5-fold worse than the mutants with the highest activities. This might indicate that improvements to binding by mutations are more gradual, whereas the negative effects of the mutations to catalysis are more sharp for this system. (Figure 3.6.c) Nevertheless, these results provide us with 13 mutants with known activation kinetics that can be modeled computationally.

Implications of results from the experimental characterization to the computational modeling of MetRS mutants

The azide group contains two electron-rich nitrogen atoms (N1 and N3) that can act as hydrogen bond acceptors. The relative strength of the hydrogen-bond interaction among the various contributors to the energy of the complex makes the identification of a hydrogen-bond donor to the azide group of primary importance for the accurate prediction of the mutant MetRS binding sites. The examination of mutations that appear in the most stringent screens for Anl incorporation reveal that mostly hydrophobic residues are selected for at the positions 260 and 301. (Figure 2.14.c) It is therefore unlikely that residues introduced at these positions are hydrogen bond donors to the azide. The crystal

structure of the SLL mutant complexed with AnI reveals that the N3 atom of the azide group on AnI accepts a hydrogen bond from a structural water molecule in the binding site [20]. The position of this water molecule is unlikely to be affected by the mutations studied here, and therefore, it is also unlikely that any mutations applied to MetRS will be the primary hydrogen bonding partner for AnI. Based on this model, the residues introduced to positions 13, 260, and 301 will mostly affect ligand packing, and effects due to their mutation will be small.

Comparison of the crystal structures of MetRS-SLL in AnI-free and AnI-bound forms show that a conformational change that accompanies ligand-binding in the wild-type MetRS is not observed with the SLL mutant. Analysis of the crystal structures reveal that the H301L mutation may have a role in blocking this conformational change in the SLL mutant. (Figure 2.16.c) Since the computational model does not take into account the rearrangement of atoms in the binding site, the absence of such conformational changes will help in obtaining more accurate binding energies. However, although H301L mutation might be implicated, the exact set of mutations that abolish this conformational change is not known. Thus, factors other than AnI recognition might be influential in the selection of certain mutants in the *in vivo* experiments.

Correlation between computational binding energies and activation parameters of MetRS mutants for AnI: The 2008 STUDY

The crystal structure of the SLL mutant was used as a starting point for the construction of the models for the MetRS mutants with available AnI activation kinetics data. Models were evaluated based on "two-point" dVBEs, which also include a PB solvation component calculated by DELPHI. (Table 3.6) Even though the spread in the computed binding energies is very narrow (2.6 kcal/mol), their comparison to kinetic parameters of AnI activation indicates a strong correlation between these parameters. ($R^2 > 0.8$; Figure 3.7) However, mutants that bear the L13P mutation fall outside the general trend. Examination of these models reveal unfavorable van der Waals interactions between the proline ring and the Cε atom on the ligand, as well as strain around the P13

residue. It is well known that proline residues can disrupt secondary structure. It is likely that the introduction of proline at position 13 forces the local backbone to take a conformation different from that displayed in the SLL crystal structure. The presence of another proline residue at position 14 might exacerbate this effect. Due to these uncertainties, mutants bearing the L13P mutation were not included in the regression analysis.

The correlations observed between binding energies and activation parameters are sensitive to several factors. The solvation contribution to the binding energy is responsible for much of the discrimination between high and low activity mutants. In fact, a strong correlation exists between $\log(K_m)$ and the solvation contribution to VBE alone ($R^2 = 0.69$; including all 13 mutants). The observed correlations are also affected by the choice of ligand rotamer. When alternate configurations of the Anl side chain in the MetRS binding site are represented by a library of rotamers, the orientation that results in the best binding interactions is within 0.2 Å RMSD of the crystal coordinates of the ligand in a great majority of the cases. Such a choice is reasonable, since there are very few orientations for the Anl side chain that allow a water-azide hydrogen bond to be established. However, for the SNL and AQL mutants, where position 260 is occupied by a polar side chain, alternate rotamers that form hydrogen bonds with these side chains may be selected, disrupting the correlation. However, position 260 is solvent accessible, especially when mutations to small amino acids are present at position 13. Because hydrogen bonds that likely form between the solvent and position 260 are not modeled and the fluctuations of the coordinates for solvent exposed side chains are not considered in our simulations, the contribution of such hydrogen-bonds may be overestimated. Selection of alternate Anl rotamers is also extremely sensitive to the exact minimization and side chain placement procedures used. To avoid such uncertainties, we assumed that the binding mode for Anl presented in the MetRS-SLL crystal structure is shared with other mutants studied here.

Differentiation of highly active MetRS mutants from mutants of poor activity

In library screens performed by Link et al., five MetRS mutants that can attain up to 50% incorporation of Anl into methionine sites were identified [8, 10]. However, treatment with 8.0 mM Anl was required to achieve such incorporation levels. The *in vivo* activities of these MetRS variants are considerably lower than the mutants studied above, most of which achieve near 100% incorporation at 1.0 mM Anl. The computed binding energies also reflect this trend. (Table 3.7) Mutants in the Link et al. set can be differentiated from mutants with higher activities based on binding energies, with the exception of the GSTL mutant. Examination of the model for the GSTL mutant reveals a second hydrogen bond to the azide group from the S257 residue, and leads this mutant to be an outlier to the general trends observed above. A discussion of the effects of such hydrogen bonds to the azide group will be addressed in the following section.

Screening a saturation mutagenesis library *in silico*: The LYH design study

A wealth of mutation data was obtained by screening the saturation mutagenesis library LYH.1.0, where residues 13, 260 and 301 are randomized, for Anl incorporation *in vivo*. In order to see the how well the experimental results are represented in the computational model we randomized the same three positions on MetRS *in silico*. All residue types, except for polar mutations (Asp, Glu, Lys, and Arg), were allowed at each of the three randomized positions, generating 16^3 = 4,096 mutants to be evaluated. Energy of Anl binding to each MetRS mutant was determined following the procedure established above. The NLL mutant, which exhibits the highest Anl activation kinetics among experimentally characterized mutants, ranks 728th in binding energy, 9.7 kcal/mol behind the top sequence, WYN. (Figure 3.8.a)

Through experiments, 41 different MetRS mutants were identified to activate Anl (see Appendix A for a full list). The binding energies of MetRS variants in this list (excluding those that bear the L13P mutation) fall in a very narrow range, within 2.6 kcal/mol of each other. It is estimated that the K_m

values for the set of mutants examined here are within two orders of magnitude, between 1 and 100 mM. This range is likely indicative of the resolution of the current computational model in distinguishing mutants of different activity.

Mutant sequences that were selected as optimal binders of Anl by the computational screen are presented in Table 3.8, as well as binding energies. Analysis of the binding-energy components indicate that it is the hydrogen-bonding component of the binding energy that sets these mutants apart from the experimental top mutant, NLL. The H301N mutation, as well as a mutation to an aromatic residue bearing polar contacts at position 260, are strongly selected for Anl binding among the top 100 mutants. Residues at these two sites interact with the azide group through hydrogen bonds, as can be observed in the model for the top ranking WYN mutant. (Figure 3.9.a)

Of the sequences evaluated for binding to Anl, more than a third exhibit stronger hydrogen bonding interactions to Anl than any mutant experimentally identified. Hydrogen bonds are perhaps the most important determinant of molecular recognition. It is, therefore, surprising that the mutants determined to be most active *in vivo* do not interact with the azide group through hydrogen bonds, while computational analysis indicates the possibility of many possible hydrogen bonding interactions between the ligand and the binding site. Such a result may be caused by an overestimation of the strength of hydrogen bonds to the azide group by the computational model. Despite the strong tendency of the azide anion to be involved in hydrogen bonds, organic azides do not form strong hydrogen bonding interactions [38]. An examination of the DREIDING force field shows that the strength of hydrogen bonds to azide nitrogens are overestimated by 1.5 to 2-fold compared with energies from QM [39]. However, this does not fully explain the discrepancy between the computational and experimental results, since the contribution from hydrogen bonds would still be very significant even at 50% strength. The crystal structure of the MetRS-SLL mutant shows multiple water molecules in the close vicinity of positions 13, 257, and 260. (Figure 3.10) Therefore, polar mutations at these positions may interact with solvent molecules, choosing the energetically stronger hydrogen bonds to water

over interactions with the azide. This would further augment the discrepancy between computation and experiment. In addition, vertical binding energies do not include the effects of rearrangements in the binding site in the unbound state. Introduction of polar groups into the binding site may strengthen the interactions of the binding site residues with each other and with the solvent in the absence of the ligand, as well. If these interactions need to be broken for ligand binding, negative consequences of this will not be represented in VBEs, and may result in stronger VBEs for polar binding sites. An alternate, but complementary explanation for the absence of azide hydrogen bonds in experimental mutants may be the absence of competing ligands in the computation model. Although the introduction of hydrogen bond donors into the MetRS binding site may improve Anl binding, it may also help recruit other natural amino acids to the binding site. Incorporation of polar amino-acids into methionine sites have been observed with both the L13G mutant (Figure 2.10) and the wild-type MetRS. (Figure B.3, first panel) The experimental screen is sensitive toward the presence of competitors (Figure 2.11), and therefore such mutants would be selected against when screened *in vivo*. Experimental testing of computationally identified mutants, and simulations of MetRS mutants in explicit solvent may shed more light on the factors influential in the selections of mutants *in vivo*.

Even though MetRS sequences that allow hydrogen bonds to the azide rank high among the 4,096 mutants evaluated, a significant number of these mutants do not recognize Anl through hydrogen bonding. In order to see where the experimentally selected sequences rank against this set of MetRS variants, mutants that show hydrogen bonding interactions weaker than −0.5 kcal/mol with the azide group were analyzed separately. Binding-energy distribution in this set of 3,222 mutants, as well as the binding-energy components for top mutants are presented in Figure 3.8.b and Table 3.9, respectively. The NLL mutant ranks 145th in binding energy, only 1.2 kcal/mol behind the top sequence, MML. The model for the MML mutant displays a structure very similar to the NLL model. (Figure 3.9.b−c) The results show that experimental mutations rank competitively among this set of mutants.

Comparison of computed mutation distributions at the randomized sites with experimental observations

In order to better understand the basis for the selection of the experimental mutants *in vivo*, the distribution of mutations selected at the randomized sites among the top 100 MetRS mutants in the set of 3,222 was compared with the distributions previously observed in experiments. (Figure 3.11) This set of sequences does not include any experimentally identified MetRS mutant. Similar to the experimental observations, mutations on H301 are strongly selected for, and almost exclusively include substitutions to Leu, Ile, Met, and Val. The H301L mutation is implicated in abolishing the conformational change wild-type MetRS undergoes upon ligand binding, and its prominence at this position was previously attributed to this effect. (Figure 2.16.c) The prediction of this mutation through a computational model that does not account for conformational changes implies that this mutation provides favorable interactions with Anl as well. The fact that all experimental mutations are correctly identified at this position suggest that the inhibition of the conformational change might not be the primary reason for the selection of these mutations in experiments. A greater diversity of mutations appear at the two remaining positions both *in vivo* and *in silico*. Although the relative frequencies of mutations differ between the distributions at position 260, the types of mutations selected at this position are surprisingly similar between experiments and computation. Mutations at this site common to both computation and experiments are observed in more than 85% of the mutants surveyed in both distributions.

Such an agreement is not observed at position 13. The computational procedure favors aromatic side chains (57%) whereas most experimentally identified mutants carry small residues (76%) at this position. Position 13 is the most solvent exposed of the three mutation sites, and there is ample space available at this site for the introduction of aromatic side chains. A recent molecular dynamics study of the wild-type MetRS has identified multiple communication pathways that run through the enzyme, initiated by movements at the L13 residue. [40] Introduction of large hydrophobic side chains at this

position may inhibit movements of the backbone at this position and reduce the tRNA aminoacylation efficiency of the enzyme, which may account for the low prevalence of aromatic residues observed at this site. In addition, the crystal structure of the MetRS-SLL mutant reveals the presence of many water molecules around position 13. Interaction of a strongly held water molecule at this site with both the backbone amide and the side chain hydroxyl groups of S13 is especially noteworthy. (Figure 3.10) The backbone amide hydrogen (NH) participates in a hydrogen bond the sulfur atom on the Met ligand in the wild-type enzyme. (Figure 3.1.b) However, a hydrogen-bonding partner for this group is not available on the ligand when Anl occupies the binding site. (Figure 3.1.c) In the Anl-bound MetRS-SLL structure, the S13 side chain helps recruit a water molecule near the backbone, which becomes a hydrogen-bonding partner for the backbone amide when Anl is bound. It is possible that small and/or polar residues at position 13 are selected *in vivo* to allow sufficient hydration for this amide group in the ligand-bound state.

The role suggested above for position 13 can explain the data obtained for the activation of Anl and Met by a series of related MetRS mutants (The XLL set: SLL, CLL, NLL and PLL). Position 13 plays a direct role in Met recognition, and mutations at this site have a greater effect on the activation rates of Met than Anl. (Figure 2.12.d–e) The 1.4 Å crystal structure of the Met-bound MetRS-SLL [20] shows that the backbone NH of S13 is shared between two hydrogen-bond acceptors, the sulfur atom on Met and the water molecule recruited by the S13 side chain. This suggests that better stabilization of this water molecule might weaken the interaction between the backbone NH and the ligand, Met. Among the mutants tested, MetRS-CLL has the highest activation rate for Met. Compared with the hydroxyl group on the serine side chain, the cysteine thiol is a poor hydrogen-bond donor, and would be limited in its ability to recruit a water molecule near the backbone NH group. In agreement with this, switching the hydroxyl with a thiol by a C13S mutation results in a significant drop in the rate of Met activation. The NLL mutant activates Met 2.5-fold slower than MetRS-SLL suggesting that the S13N substitution better stabilizes the backbone NH. In fact,

asparagine side chains are known to associate with polar atoms on the local backbone, and both side chain and backbone functional groups on the same residue can participate in simultaneous interactions with other polar groups nearby [41]. It is, therefore, plausible for the asparagine side chain to participate in either a direct or a water-mediated hydrogen bond with the backbone NH, rendering this group less available for interacting with the sulfide group on Met. The lowest activation rate for Met belongs to the PLL mutant. The mutation to proline at position 13 removes the amide hydrogen from the backbone, creating a binding site free of any specific polar contacts for the Met side chain.

Effect of mutations at position 13 on Anl activation is more modest. Although, based on their size and orientation it is unlikely for side chains at this position to form direct, polar contacts with Anl, when they recruit solvent molecules near the binding site, they may alter the desolvation cost associated with Anl binding. Computational analysis had shown a strong correlation between the observed K_m values and the PB solvation energies, identifying solvation as one of the most prominent factors in Anl binding. Consistent with these views, the highest rates of Anl activation in the XLL set are observed with PLL and CLL, the mutants that provide the most hydrophobic side chains at position 13. (Figure 2.12.d) Interestingly, the NLL mutant has 2-fold higher activity toward Anl than MetRS-SLL, even though asparagine carries more polar groups than serine. This seems to conflict with the proposed model. However, the 1.7 Å crystal structure of the apo-MetRS-SLL [20] reveals an alternate conformation for the S13 side chain when no ligand is present, pointing inside the binding site and hydrogen bonding with water molecules therein. This necessitates the reorganization of the S13 conformation and the associated water molecules upon Anl binding, resulting in weaker Anl binding. On the other hand, the asparagine side chain is similar in size to the wild-type residue (leucine) and is unlikely to experience a similar conformational change. This might explain the higher activity observed for the NLL mutant.

Conclusions

We have presented a detailed comparison between results from screens of a saturation mutagenesis library *in vivo* and an *in silico*. Analysis of results reveal a reasonable agreement between experimental data and the computational model. MetRS mutants can be differentiated according to their activities toward Anl based on computed binding energies. Although the resolution of computation energies do not match the resolution observed in experiments, binding energies can help identify active AARS variants among poorly active or inactive mutants. It is also remarkable that mutations that can donate hydrogen bonds to the azide group are not selected by the experimental screen, even though such mutations are possible, as indicated by the computational results. Factors not represented in the computational model, such as explicit solvent in and around the binding pocket and the presence of competitors in the cell, may account for this discrepancy. Nevertheless, the analysis of mutants that do not rely on hydrogen bonds for Anl recognition reveals that the types of mutations selected at randomized sites by the *in silico* screen is in remarkable agreement with the experimental data at positions 260 and 301. Such an agreement is not observed at position 13, the most solvent exposed of the three randomized sites. This position may have a structural role in communicating the binding of ligand to other locations on the enzyme. It may also be necessary for mutations at this position to allow sufficient hydration of the backbone in the ligand-bound state.

Structural design of AARS binding sites for non-natural substrates is not a simple problem. AARSs are large enzymes with multiple domains, and catalyze a multi-step reaction essential to the viability of the cell and involves three different substrates. They are dynamic entities, functioning in the complex environment of the cell, where potential competitors are abundant. Moreover, there is increasing evidence on the involvement of AARSs in a variety of cellular functions beyond protein synthesis [42]. It is, therefore, difficult to fully represent the complex nature of AARS activity *in vivo* by computed binding energy. Even so, this work demonstrates that many aspects of AARS activity can be explained

based on ligand binding. Observations made here will guide future efforts to engineer AARSs for non-natural substrates.

Table 3.1.

MetRS mutants tested for the incorporation of Anl into proteins by Link et al.

Name	L13	P257	Y260	H301	Relative activity
L13G*	G	P	Y	H	++
GSTL†	G	S	T	L	+
GLLV†	G	L	L	V	+
GLTA†	G	L	T	A	+
GLAA†	G	L	A	A	+
GMGV†	G	M	G	V	+
Y260T*	L	P	T	H	–
wt‡	L	P	Y	H	–

* The L13G (++) and the Y260T (–) single mutants were tested to isolate the effects of two recurring mutations observed on mutants identified by library screening. Relative activities of these were determined based on protein expression levels in the presence Anl and absence of Met.

† Relative activities of the five 4-fold MetRS mutants (+) identified by library screens were determined based on the extent of cell-surface labeling measured by flow cytometry.

‡ Positions with wild-type identities are indicated in the table in light-gray.

Table 3.2.

Sequences of primers discussed in this chapter

Primer name	Sequence (5'-to-3')
MRS_BamHI	TTCCGCGGATCCATGACTCAAGTCGCGAAGAAAATTC
MRS_SalI-r	TTTGGGGTCGACTCATTTAGAGGCTTCCACCAGTG
eM_G13C*	ATTCTGGTGACGTGCGCATGTCCGTACGCTAAC
eM_G13N*	ATTCTGGTGACGTGCGCAAACCCGTACGCTAAC
eM_G13S*	ATTCTGGTGACGTGCGCAAGCCCGTACGCTAAC
eM_L13G*	CTGGTGACGTGCGCAGGTCCGTACGCTAACGGCTC
eM_Y260A*	GACGCACCGATTGGCGCGATGGGTTCTTTCAAG
eM_Y260N*	GGACGCACCGATTGGCAACATGGGTTCTTTCAAG
eM_P257S_Y260N*	CTGGCTGGACGCATCTATTGGCAACATGGGTTCTTTC
eM_P257T_Y260G*	CTGGCTGGACGCAACTATTGGCGGTATGGGTTCTTTC

* Only the forward sequence is provided for QuikChange primers.

Table 3.3.

Anl binding energies calculated for the Link et al. mutants in the 2005 STUDY

MetRS	Relaxed vertical binding-energy components* (kcal/mol)						Relative activity
	VdW	Coulomb	H-bond	Bonding	Solvation	Total	
L13G	4.4	-0.1	-11.1	0.3	-0.5	-7.0	++
GSTL	6.3	-1.9	-12.7	-0.2	2.3	-6.4	+
GLLV	3.6	1.0	-6.6	-0.4	-2.9	-5.4	+
GLTA	4.5	-1.4	-8.8	0.6	0.5	-4.7	+
GLAA	5.5	0.9	-10.8	-0.2	-3.2	-7.8	+
GMGV	4.5	-3.0	-12.7	0.6	3.3	-7.4	+
Y260T	6.4	-2.2	-9.9	1.0	2.9	-1.8	-
wt	0.0	0.0	0.0	0.0	0.0	0.0	-

* All energies are reported relative to the binding energy of the wild-type enzyme to Anl.

Table 3.4.

Binding energies for the top 15 sequences from the LPY design study

Rank	L13	P257	Y260	Relaxed vertical binding-energy components (kcal/mol)					
				VdW	Coulomb	H-bond	Bonding	Solvation	Total
1	L	P	N	-8.4	-28.9	-24.1	6.1	30.9	-24.4
2	G	P	N	-8.8	-29.2	-23.7	5.7	32.9	-23.1
3	L	T	G	-9.7	-28.9	-15.8	5.7	30.8	-17.9
4	S	P	A	-10.5	-28.3	-13.7	2.7	32.2	-17.6
5*	G	P	Y	-7.9	-27.3	-16.4	2.6	31.6	-17.4
6	C	S	N	-5.4	-27.8	-21.9	6.7	31.0	-17.4
7	S	P	G	-10.0	-28.4	-13.9	2.7	32.2	-17.4
8	A	S	N	-6.9	-28.5	-20.7	6.7	32.1	-17.3
9	C	P	N	-6.8	-26.8	-20.8	6.4	30.9	-17.1
10	N	P	A	-10.7	-29.1	-14.4	4.7	32.5	-17.0
11	G	P	A	-9.6	-27.6	-16.0	4.7	31.5	-17.0
12	G	P	G	-5.3	-28.3	-16.6	2.4	30.9	-16.9
13	N	P	G	-10.3	-29.1	-14.6	4.8	32.4	-16.8
14	S	P	C	-10.2	-28.1	-13.2	2.5	32.2	-16.8
15	G	T	A	-10.1	-29.2	-15.1	5.2	32.6	-16.6

* The L13G mutant is known to activate Anl, and is highlighted here to indicate this.

Table 3.5.

Kinetic parameters for the activation of Anl by mutants identified in screens performed to isolate mutants with low activities

MetRS variant	Amino acid	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1} s^{-1}$)	Relative activity
wt*	Met	0.024	13.30	550,000	1
L13G†	Anl	1.5	0.57	170	1/3,200
NLL†	Anl	2.2	0.62	410	1/1,400
AQL	Anl	7.7	0.071	9.3	1/59,000
SNL	Anl	9.6	0.060	6.2	1/88,000
GVL	Anl	9.5	0.044	4.7	1/120,000

* Activation parameters for wild-type MetRS taken from reference [43].

† Activation parameters for the L13G and NLL mutants were reported in Chapter 2.

Table 3.6.

Binding energies calculated for MetRS mutants with known Anl-activation characteristics

MetRS	Direct vertical binding-energy components (kcal/mol)					
	VdW	Coulomb	H-bond	Bonding	Solvation	Total
NLL	-15.6	-27.9	-26.1	3.1	10.3	-56.1
CLL	-15.5	-28.8	-26.1	3.1	11.2	-56.0
SLL	-15.1	-28.4	-26.1	3.1	11.0	-55.4
AQL	-15.0	-29.2	-26.0	3.1	12.0	-55.1
GML	-14.9	-29.5	-25.9	3.1	12.3	-55.0
GCL	-14.3	-29.7	-26.0	3.1	12.4	-54.4
GVL	-14.2	-29.7	-25.9	3.1	12.4	-54.2
SNL	-14.5	-28.2	-25.9	3.0	11.4	-54.2
GIL	-14.2	-29.7	-25.9	3.1	12.5	-54.1
PLL*	-13.8	-29.6	-25.7	3.9	11.0	-54.2
PNL*	-13.5	-29.6	-25.9	4.1	11.2	-53.6
PIL*	-13.7	-29.6	-25.6	4.0	11.2	-53.6
PLI*	-13.0	-29.5	-25.8	4.4	10.9	-53.0

* MetRS variants that bear the L13P mutation do not follow the trends set by other mutations as shown in Figure 3.7) and are, therefore, listed separately.

Table 3.7.

Revised Anl binding energies for Link et al. mutants in the 2008 STUDY

MetRS	Direct vertical binding-energy components (kcal/mol)					
	VdW	Coulomb	H-bond	Bonding	Solvation	Total
GSTL	-13.7	-31.2	-31.4	3.6	13.0	-59.7
GLLV	-13.6	-29.9	-26.1	3.0	12.7	-53.9
GLLA	-13.0	-30.0	-26.1	3.0	12.8	-53.2
GLTA	-12.6	-29.8	-26.1	3.0	12.8	-52.7
GMGV	-11.4	-29.8	-26.3	2.8	12.5	-52.2

Table 3.8.

Binding energies for the top 12 sequences out of 4,096 evaluated in the LYH design study

Rank	L13	Y260	H301	Direct vertical binding-energy components (kcal/mol)					
				VdW	Coulomb	H-bond	Bonding	Solvation	Total
1	W	Y	N	-10.3	-31.3	-39.1	3.8	11.0	-65.9
2	L	H	N	-13.4	-30.5	-36.7	3.3	11.4	-65.9
3	F	H	N	-13.4	-30.6	-36.7	3.2	12.1	-65.3
4	M	W	N	-15.2	-29.3	-34.8	4.2	10.7	-64.4
5	Y	H	N	-10.8	-29.5	-37.8	3.6	10.4	-64.2
6	L	W	N	-15.4	-28.5	-34.5	4.1	10.1	-64.1
7	A	W	N	-14.1	-29.9	-35.2	4.1	11.1	-64.0
8	F	W	N	-15.2	-28.8	-34.9	4.4	10.4	-64.0
9	Y	W	N	-15.2	-28.5	-34.4	4.3	10.1	-63.6
10	Q	W	N	-15.0	-30.4	-34.9	4.4	12.4	-63.5
11	H	W	N	-14.6	-27.3	-35.3	4.4	9.3	-63.5
12	T	W	N	-13.7	-29.4	-34.7	4.2	10.2	-63.4
728*	N	L	L	-15.6	-27.8	-26.1	3.1	10.2	-56.2

* NLL, the highest ranking experimentally tested mutant, is included here for comparison.

Table 3.9.

Binding energies for the top 12 sequences out of 3,222 evaluated in the LYH design study that do not interact with Anl through hydrogen bonds

Rank	L13	Y260	H301	Direct vertical binding-energy components (kcal/mol)					
				VdW	Coulomb	H-bond	Bonding	Solvation	Total
1	M	M	L	-16.8	-29.3	-26.2	3.2	11.7	-57.4
2	F	M	L	-16.6	-29.1	-26.2	3.2	11.4	-57.4
3	M	M	I	-16.9	-29.3	-26.2	3.2	11.9	-57.3
4	Y	L	M	-16.5	-28.9	-26.1	3.1	11.2	-57.3
5	Y	L	L	-16.5	-28.9	-26.1	3.2	11.0	-57.3
6	F	M	I	-16.7	-29.1	-26.2	3.2	11.5	-57.3
7	H	H	I	-16.2	-27.9	-26.9	3.0	10.7	-57.2
8	M	M	M	-16.8	-29.3	-26.3	3.4	11.8	-57.2
9	M	L	M	-16.6	-29.5	-26.2	3.0	12.0	-57.2
10	Y	M	I	-15.8	-28.8	-26.2	3.1	10.5	-57.2
11	N	H	M	-16.3	-28.0	-26.7	2.9	10.9	-57.1
12	Y	M	L	-15.6	-28.8	-26.2	3.0	10.5	-57.1
145*	N	L	L	-15.6	-27.8	-26.1	3.1	10.2	-56.2

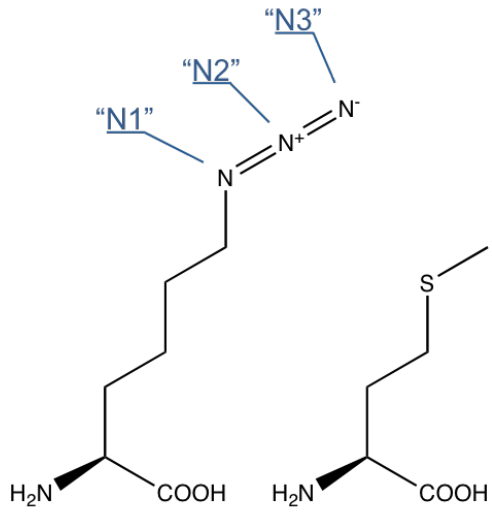
* NLL, the highest ranking experimentally tested mutant, is included here for comparison.

Figure 3.1.

Structures of the E. coli *methionyl-tRNA synthetase (MetRS) active site and the ligands studied.*

The chemical structures of azidonorleucine (Anl) and methionine (Met), as well as the structures of the ligand-bound binding sites of two *E. coli* MetRS variants are presented. The nomenclature used in the text to identify nitrogen atoms of the azide group (N1, N2, and N3) are indicated on the Anl structure. In the binding site models the side chains of the four residues that were computationally designed for the recognition of Anl are highlighted in green, and the ligand in yellow. The crystal water molecule conserved in both crystal structures is labeled “HOH” and dashed lines mark important hydrogen bonds between the ligand and residues in the binding site.

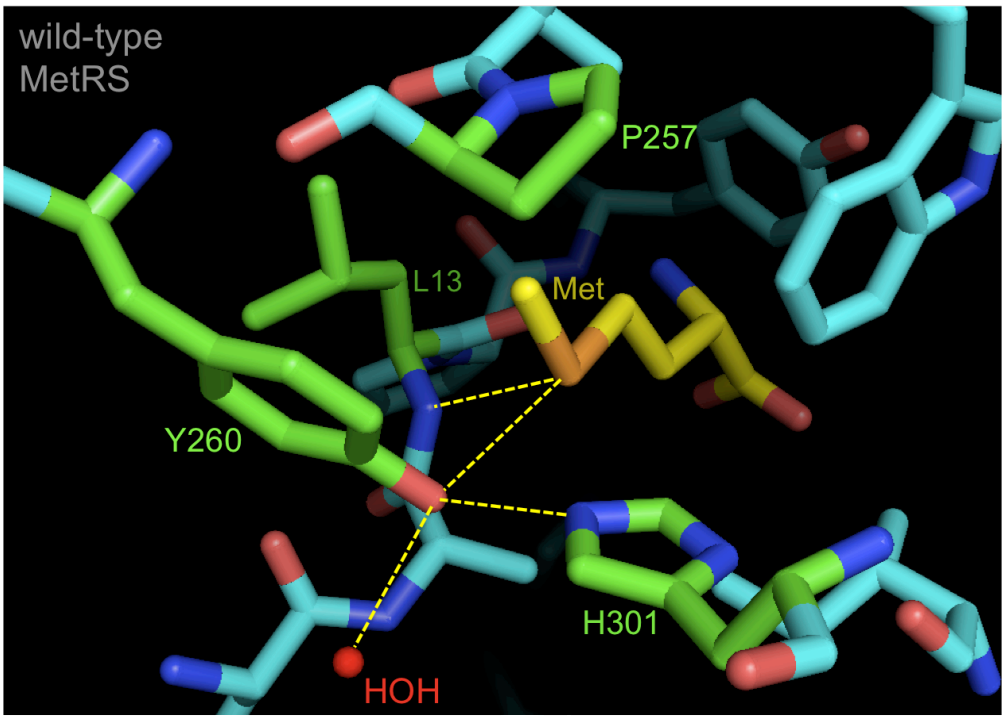
a) Chemical structures of MetRS ligands.



azidonorleucine (Anl) | methionine (Met)

Figure 3.1. (continued)

b) Structure of the wild-type MetRS binding site with methionine ligand [17].



c) Structure of the MetRS-SLL binding site with azidonorleucine ligand [20].

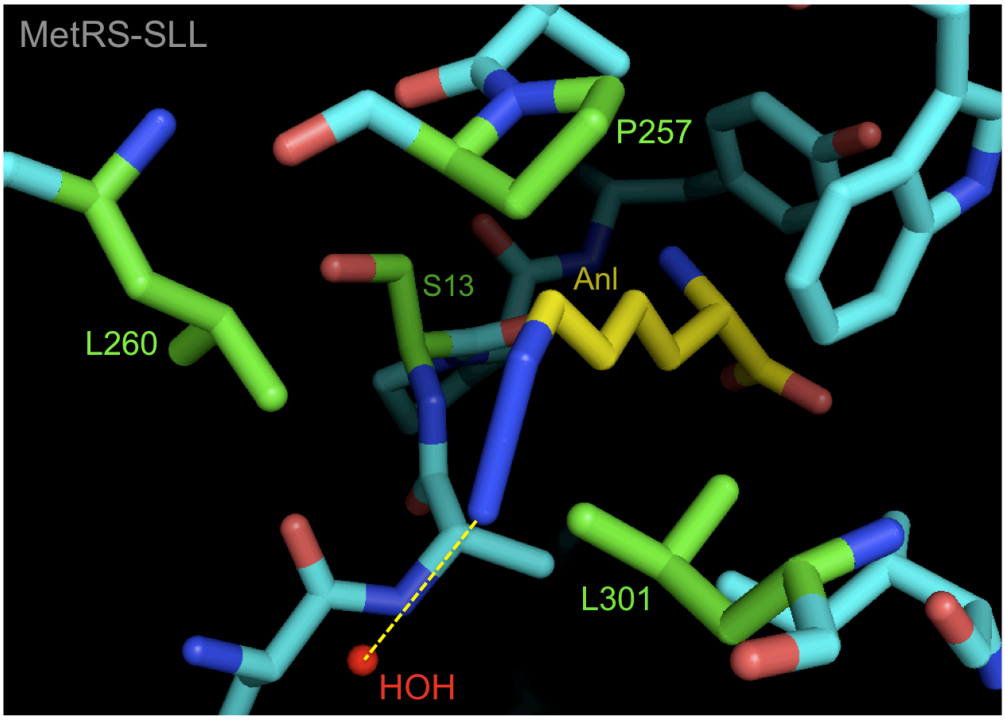


Figure 3.2.

Complex preparation and optimization scheme.

An outline of the steps taken for the preparation and sequential minimization of each ligand-enzyme complex are presented. Energies obtained for the complex, ligand and enzyme are used to calculate vertical binding energies.

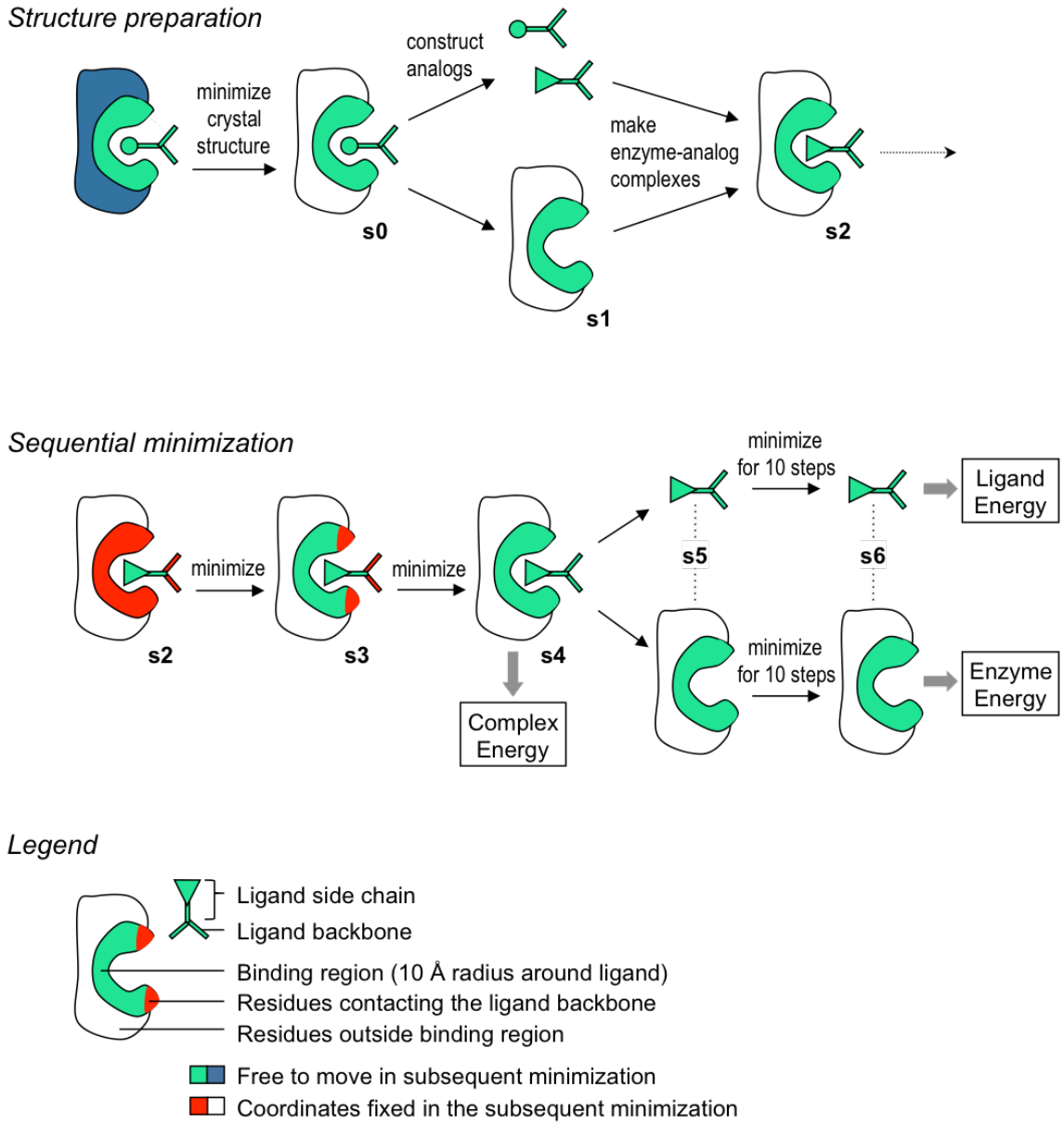


Figure 3.3.

The model for the Anl-bound MetRS-L13G mutant.

A comparison of three models of ligand-bound MetRS binding sites are presented: a) Met in wild-type MetRS binding site, b) Anl in wild-type MetRS binding site, and c) Anl in MetRS L13G mutant binding site. Four residues randomized in the saturation mutagenesis study by Link et al. are highlighted in green, whereas the ligands, Met and Anl, are shown as van der Waals spheres with yellow carbon atoms. The model of wild-type MetRS with Anl in b) was obtained by inserting Anl in an extended conformation into the wild-type MetRS binding site. Clashes with the residues P257 and Y260 are apparent in this model. The L13G mutation, shown in c), allows an alternate configuration for Anl which allows the azide group to hydrogen bond to the backbone N atom on position 13.

Figure 3.3. (continued)

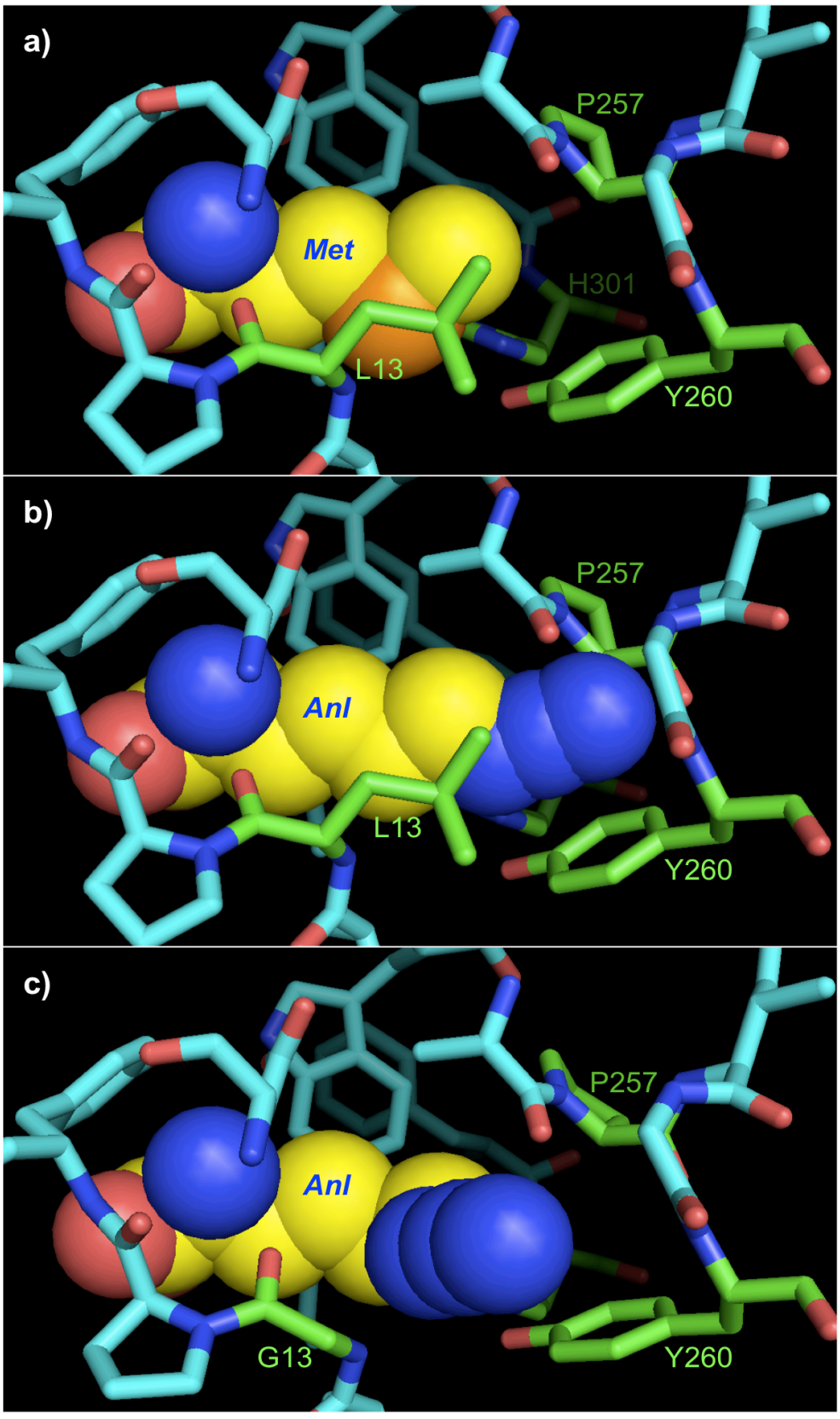


Figure 3.4.

Distribution of binding energies for mutants evaluated in the LPY design study. The binding energy (rVBE) for each mutant explored in the LPY design study is plotted against the rank of that sequence. Mutations are spread over a range of 75 kcal/mol in binding energy. The L13G mutant ranks 5th, below other sequences with known activities, wild-type MetRS and the Y260T mutant.

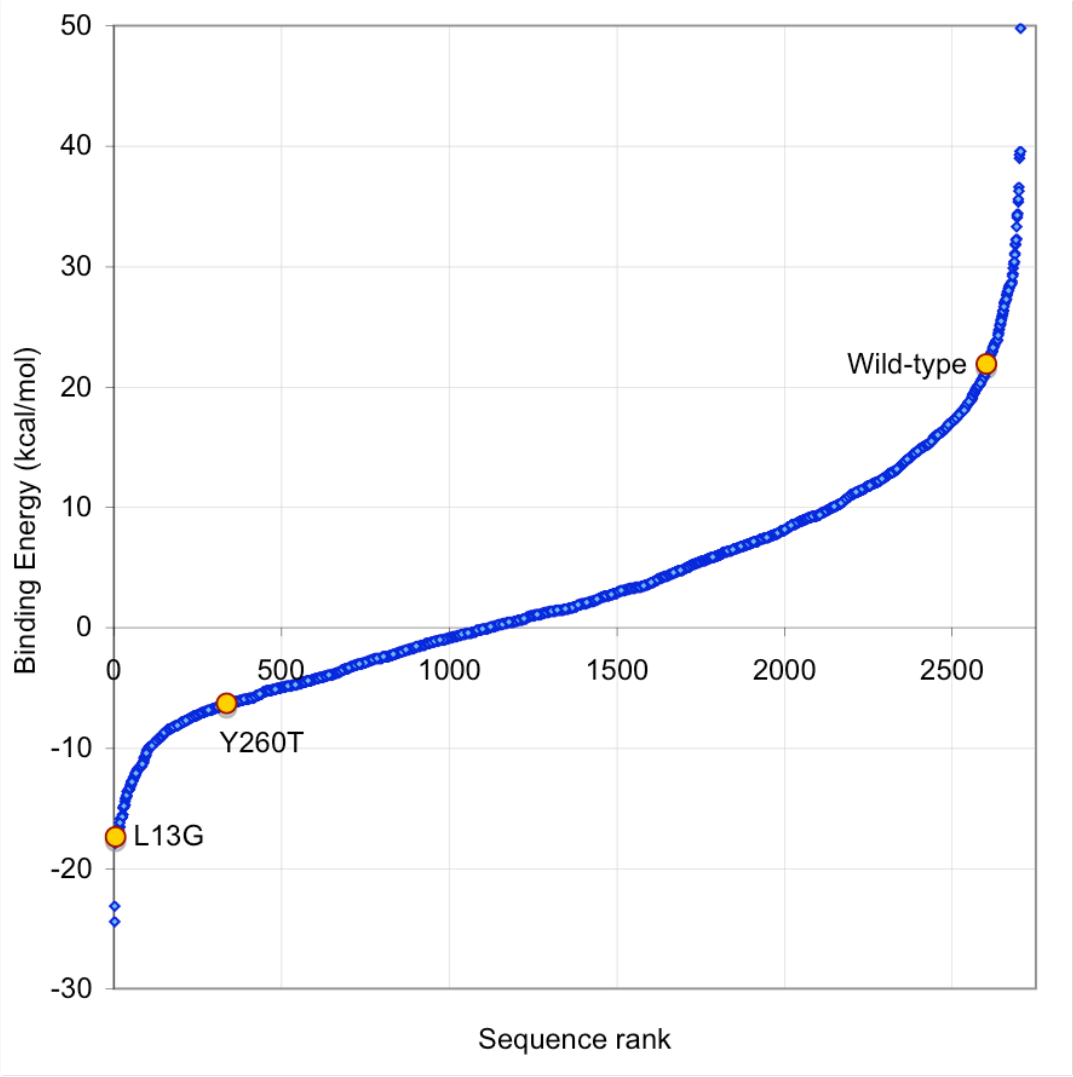


Figure 3.5.

Expression of DHFR in the presence of Anl in cells bearing MetRS mutants identified in the LPY design study.

DHFR was expressed in methionine-auxotrophic M15MA cells encoding the MetRS variants in the presence of 2.0 mM Anl and absence of Met. SDS-PAGE analysis of whole-cell lysates indicate expression of DHFR only in the presence of the L13G mutant. DHFR is a 24 kDa protein, and the expected location of a DHFR band is marked by a red line. The rank in binding energy of each MetRS sequence tested is indicated above each lane.

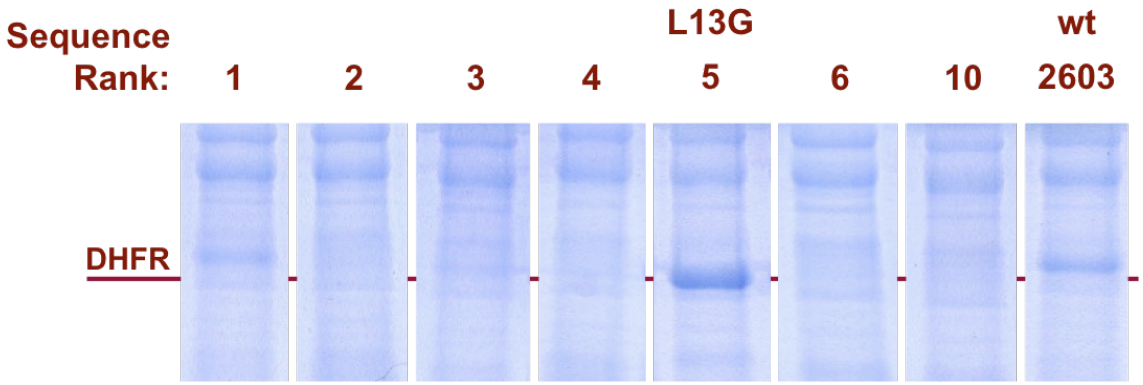


Figure 3.6.

Screening for and characterization of a population of clones displaying low cell-surface fluorescence.

a) The screen for active clones that carry MetRS variants providing only low-levels of cell-surface fluorescence is outlined. The LYH.1.0 library is screened for two rounds using flow cytometry in the presence of 1.0 mM Anl, revealing the population LYH.6.2. b) The LYH.6.2 clones were induced to express OmpC in media supplemented with 0.3, 1.0 or 3.0 mM Anl, and the median fluorescence of the labeled population was determined on a flow cytometer. The median fluorescence for cells (red circles), and the Hill equation fit to the data (red lines) are displayed for the L13G and NLL mutants as well as the LYH.6.2 population. The EC50 value obtained from the Hill equation differs by an order of magnitude between the NLL mutant and the LYH.6.2 population. c) Kinetic parameters for the activation of Anl were determined for three MetRS mutants, AQL, SNL and GVL, identified from the LYH.6.2 population. A comparison of the activation parameters, K_m and k_{cat}/K_m, is presented between the newly identified mutants (shown in orange), and mutants characterized previously from screens carried out at 1.0 mM Anl (green) and 0.3 or 0.1 mM Anl (blue).

Figure 3.6. (continued)

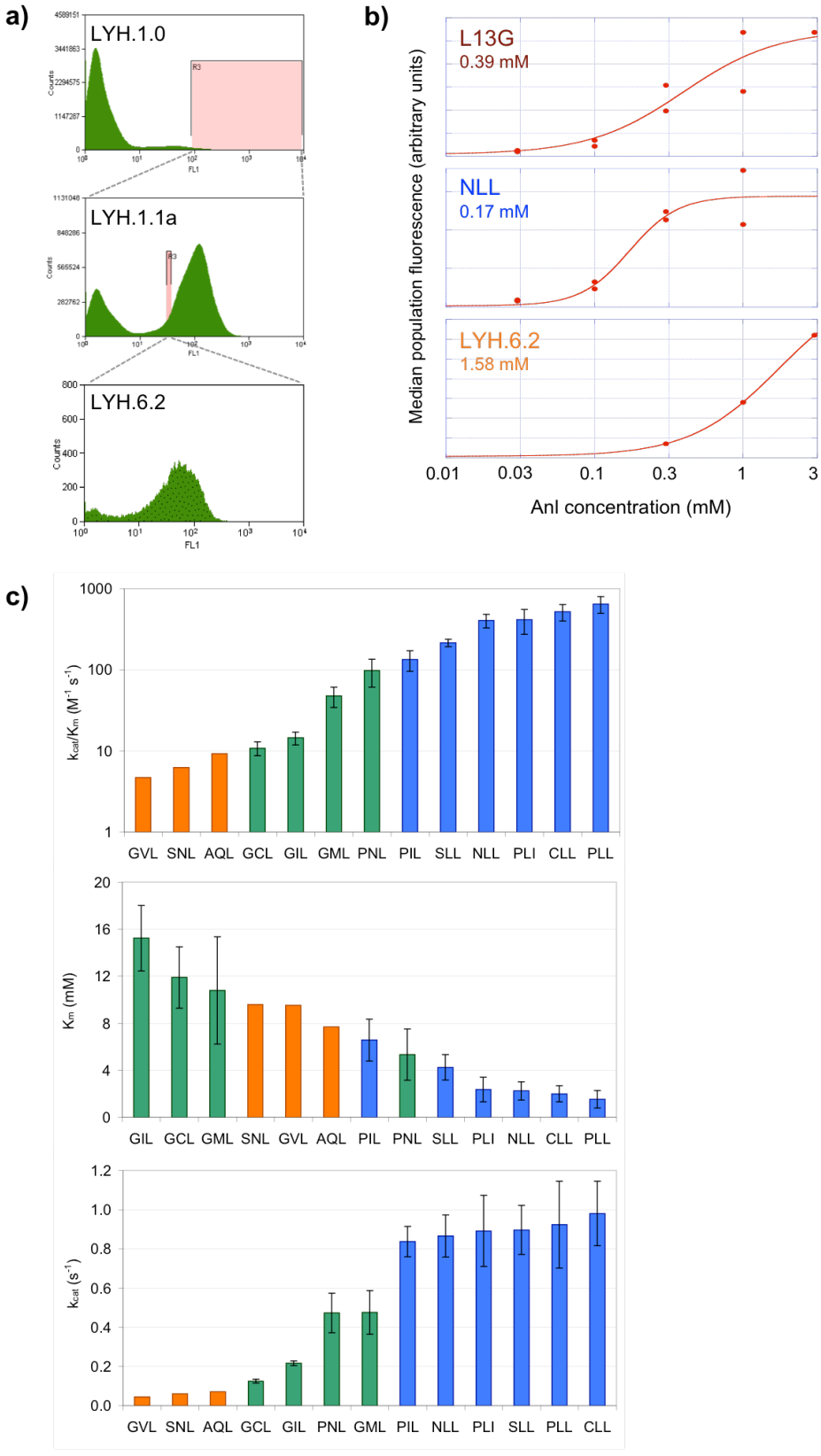


Figure 3.7.

Correlation between Anl activation parameters and computed Anl-binding energies for MetRS mutants.

Anl binding energies (dVBE) calculated for each mutant are plotted against Anl activation parameters, in logarithmic scale. Linear least-squares fits and correlation coefficients were determined on data (blue circles) excluding mutants that contain the L13P mutation (blue crosses).

a) Correlation between binding energies and K_m values.

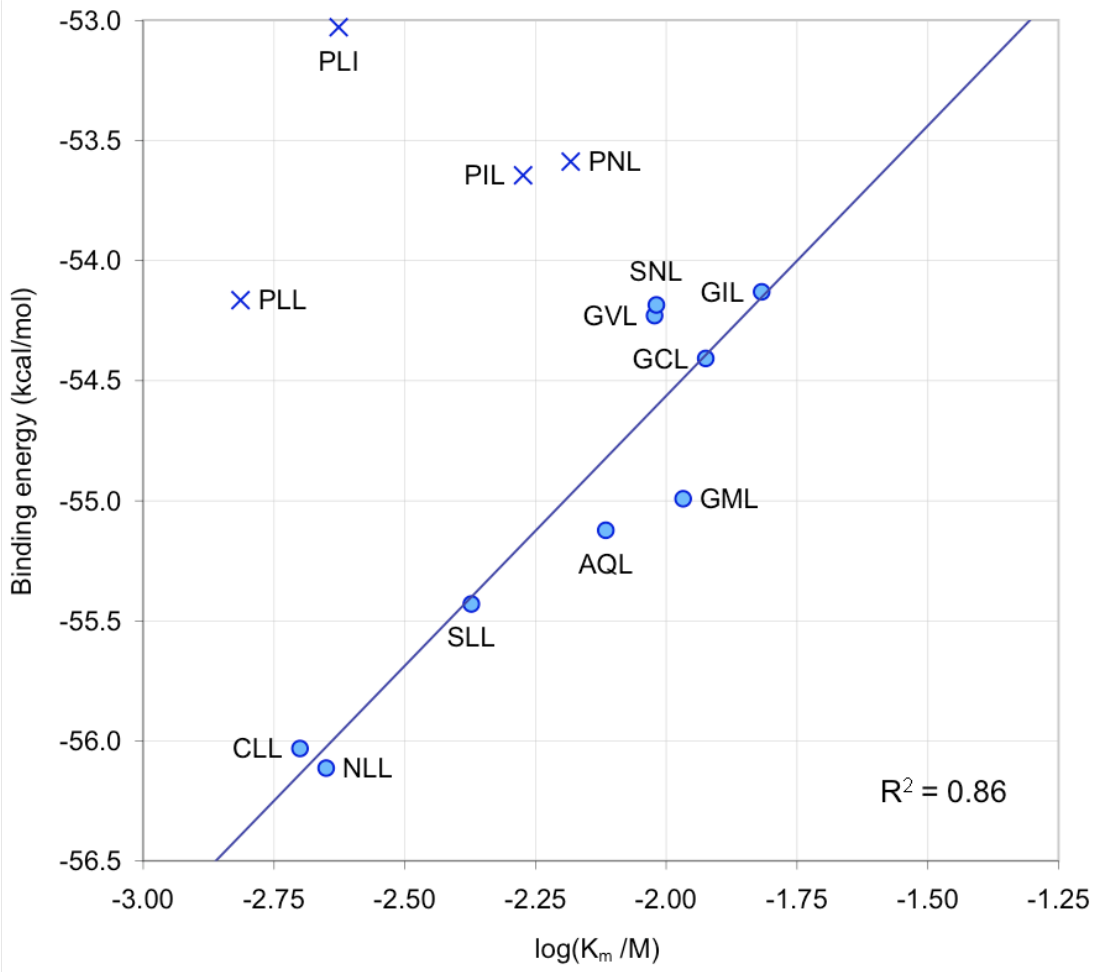


Figure 3.7. (continued)

b) Correlation between binding energies and k_{cat}/K_m values.

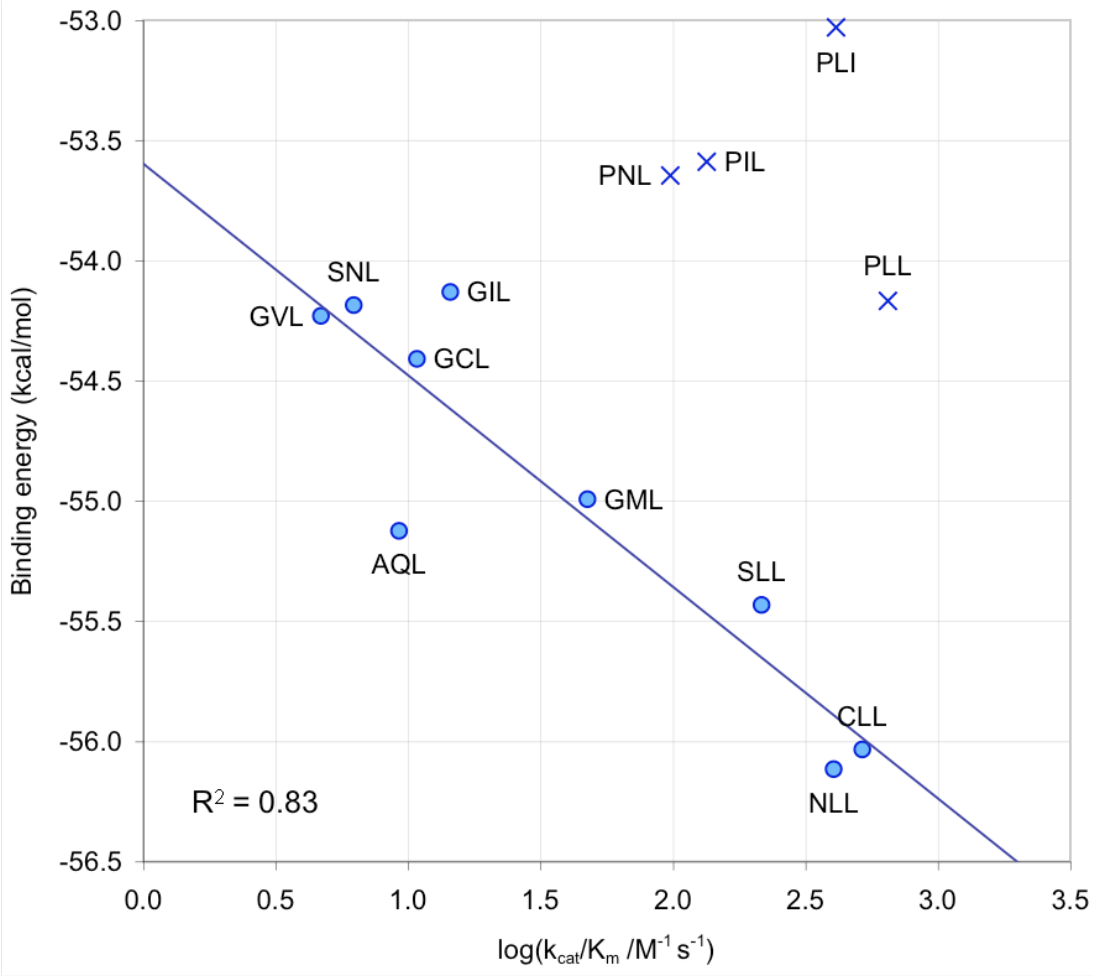


Figure 3.8.

Distribution of binding energies for mutants evaluated in the LYH design study.
The Anl binding energy (two-point dVBE) for each mutant explored in the LYH design study is plotted against the rank of that sequence in the list of binding energies (blue diamonds). Twenty-nine MetRS mutants that were previously identified in the experimental studies to activate Anl are shown as circles outlined in red. Experimental mutants that bear the L13P mutation are omitted from the plots. Mutants that were previously used to establish the correlation between experimental and computational binding parameters are highlighted in yellow, and the names and sequence ranks of these mutants are indicated next to each data point. Three plots show the ranks of the experimentally identified mutations among: a) the whole set of 4,096 mutants and b) the set of 3,222 mutants that do not directly interact with the azide group through hydrogen bonds.

Figure 3.8. (continued)

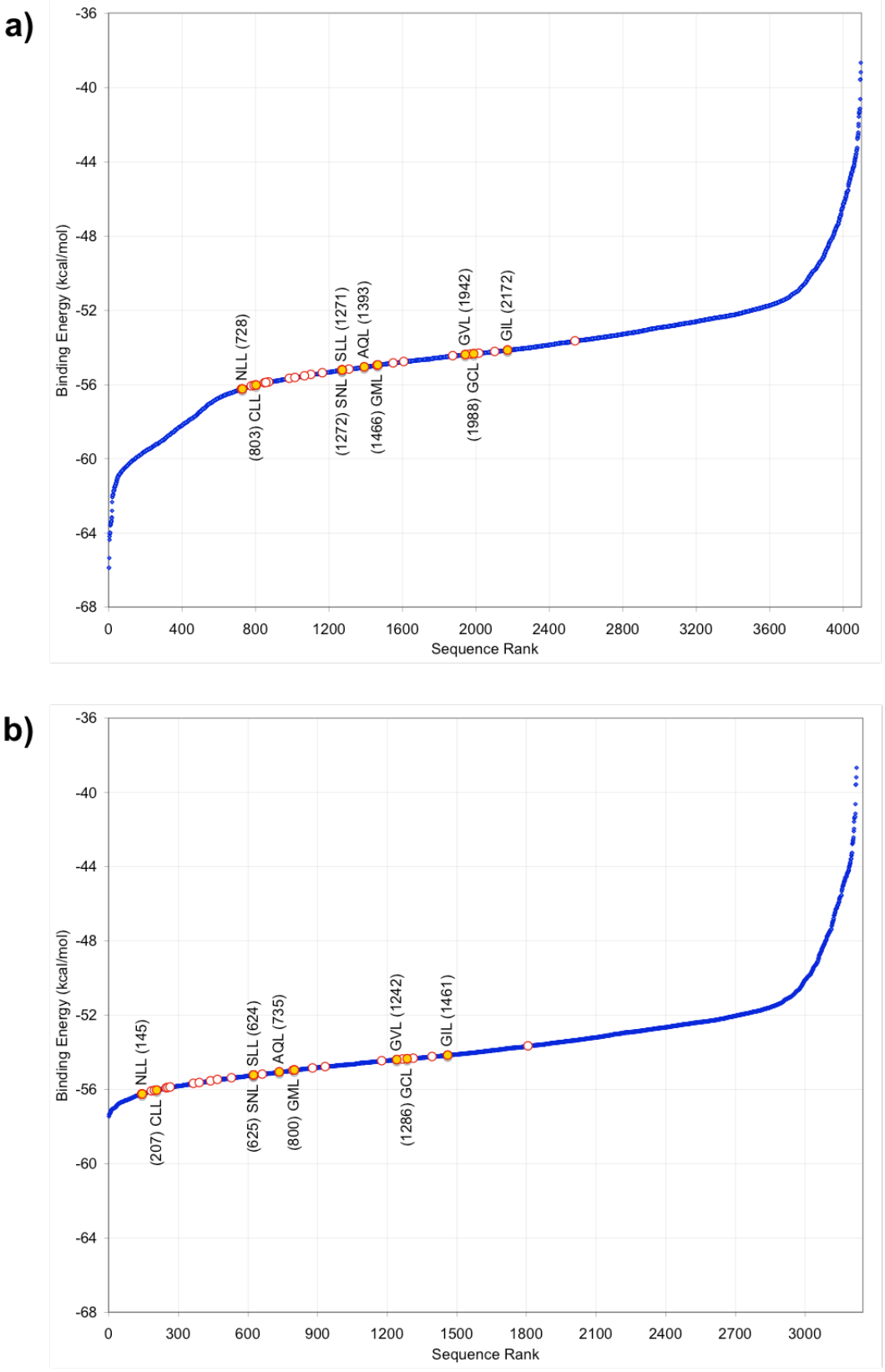


Figure 3.9.

Models for the best MetRS mutants in the in vivo *and* in silico *screens complexed with Anl.*

Models for mutants that show the best binding interactions among: a) the complete set of mutants (WYN), b) set of mutants that do not make hydrogen bonds with the azide (MML), and c) experimental Anl-activation rates (NLL) are shown. Side chains of the residues at the positions randomized in the *in vivo* and *in silico* screens (13, 260, and 301) are highlighted in green, whereas the Anl ligand is shown in yellow. The conserved water molecule in the binding site is labeled “HOH” and dashed lines mark the hydrogen bonds donated to the azide group. In the WYN mutant, the azide group on Anl receives two hydrogen bonds from Y260 (to N1) and N301 (to N3).

Figure 3.9. (continued)

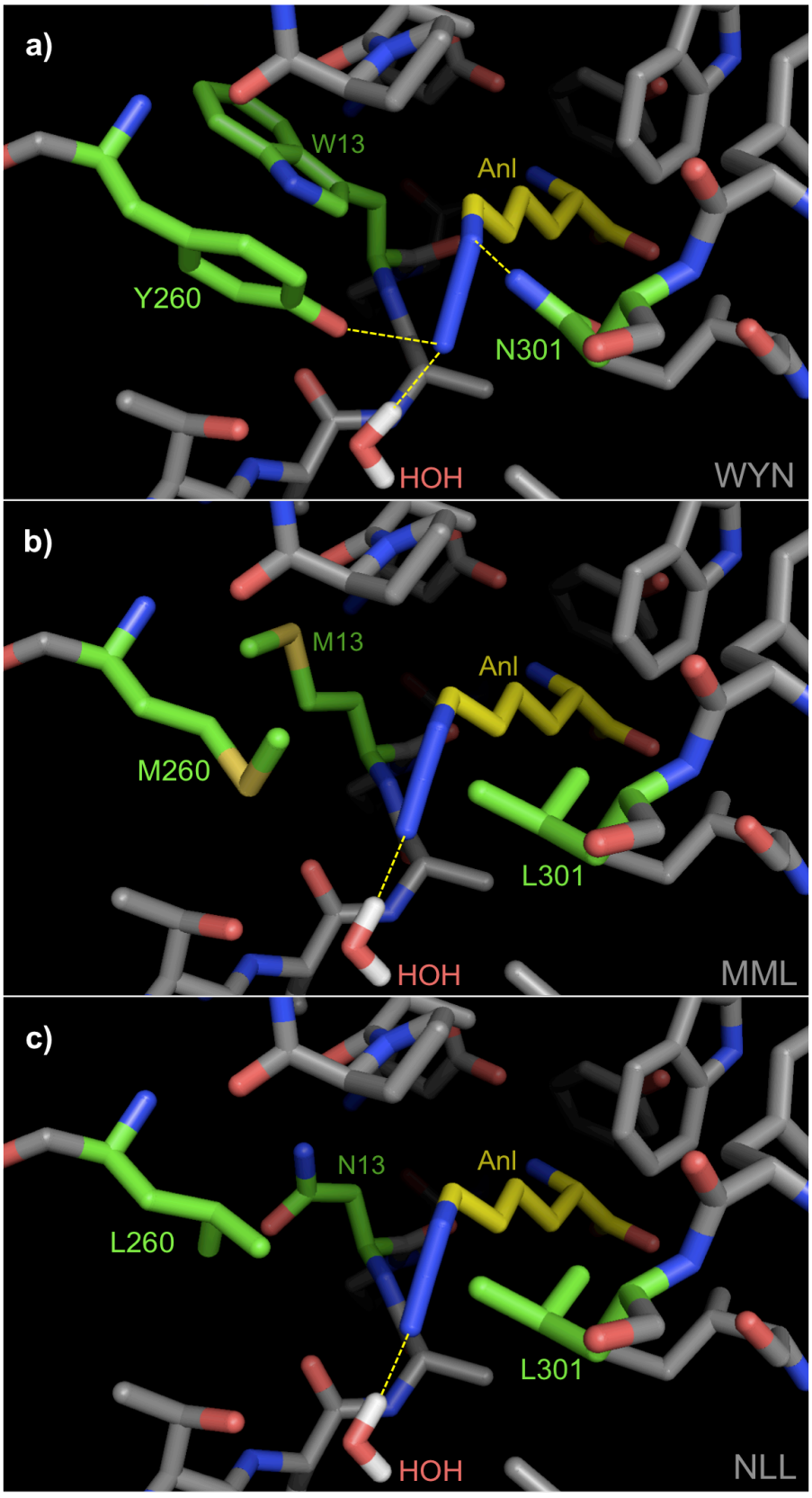


Figure 3.10.

Organization of water molecules around residue S13.

Position S13 and residues and water molecules surrounding it are presented based on the MetRS-SLL crystal structure [20]. The side chains of the computationally designed positions, S13 and L260, are highlighted in green, and the ligand in yellow. A crystal water molecule that forms multiple hydrogen bonds with S13 is labeled “HOH,” and dashed lines mark the important hydrogen bonds between this water molecule and MetRS.

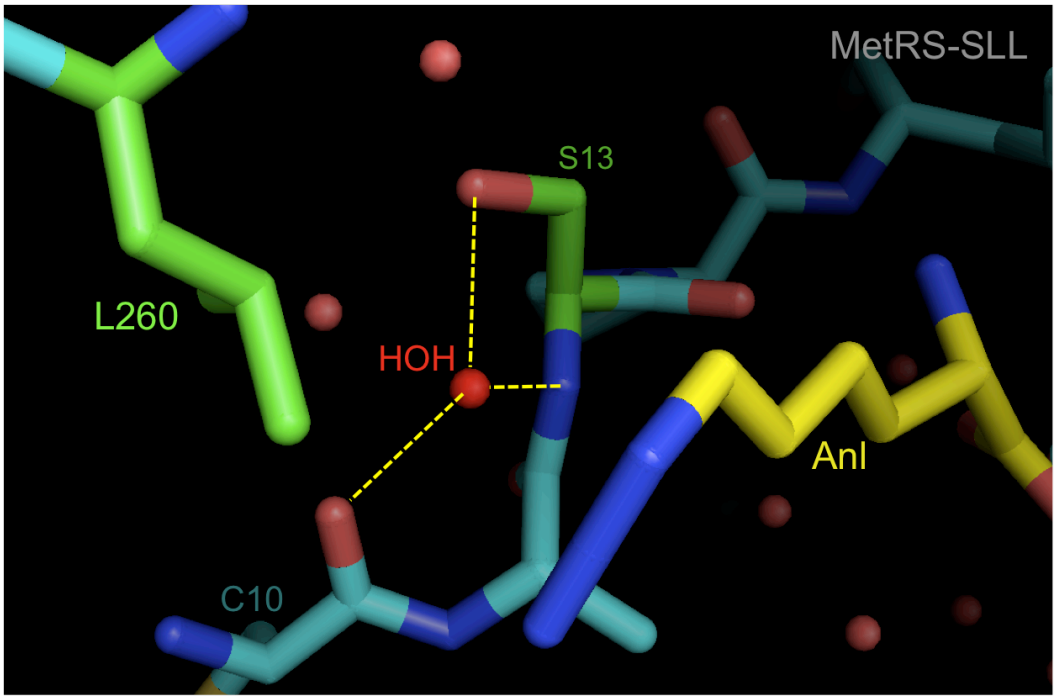
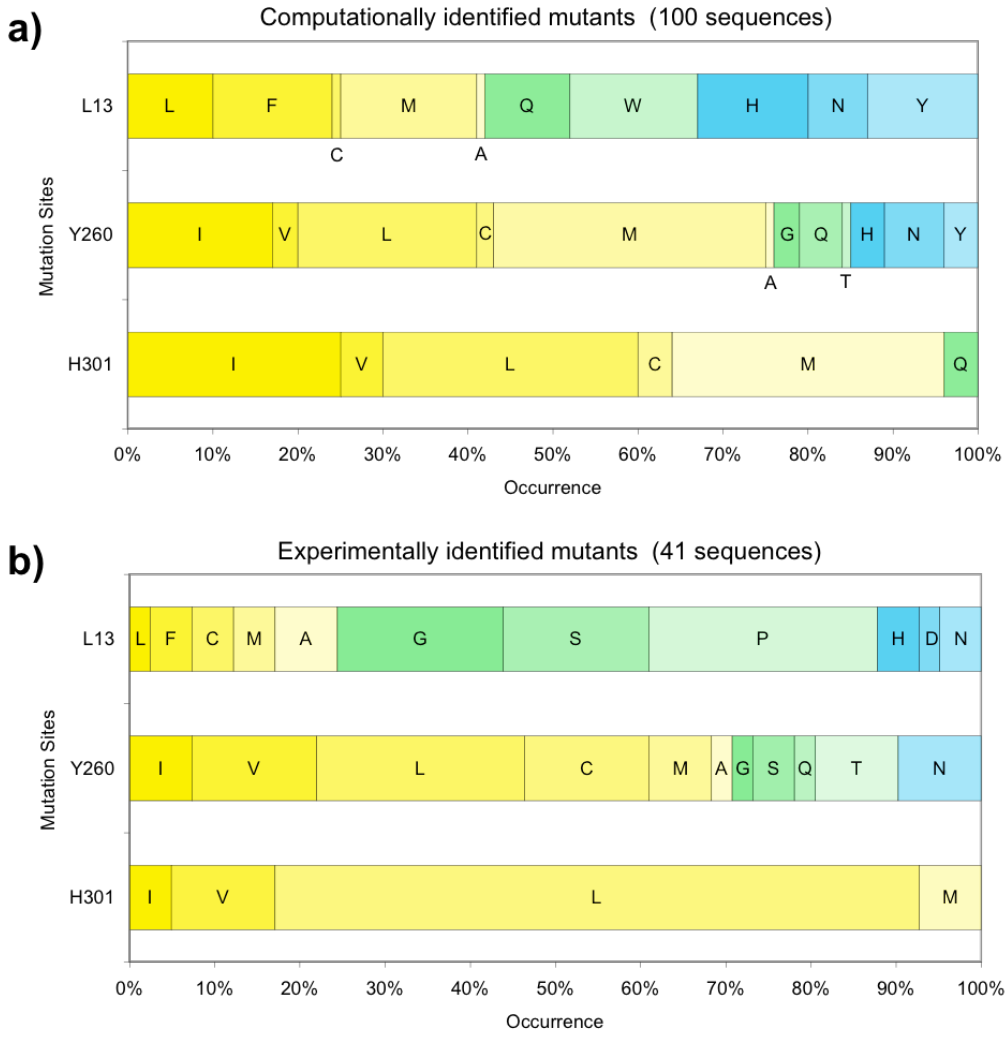


Figure 3.11.

Comparison of mutation distributions at each randomized site between the in vivo *and* in silico *screens.*

Distribution of mutation types by residue position is shown for a) top 100 sequences out of 3,222 evaluated in the LYH design study that do not interact with Anl through hydrogen bonds, and b) 41 sequences identified in experimental screens. (Figure 2.14) Substitutions at each site are ordered in decreasing hydrophobicity according to the Kyte-Doolittle scale [44]. Based on this scale, residues are grouped as "hydrophobic" (>1.0; shades of yellow) or "hydrophilic" (<–2.0; shades of blue). Residues that rank between these groups are shown in shades of green.



References

[1] Wang L and Schultz PG, "Expanding the genetic code," *Angew Chem Int Ed Engl*. **2005**, *44*:34.

[2] Link AJ, Mock ML and Tirrell DA, "Non-canonical amino acids in protein engineering," *Curr Opin Biotechnol*. **2003**, *14*:603.

[3] Budisa N, "Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire," *Angew Chem Int Ed Engl*. **2004**, *43*:6426.

[4] Santoro SW, Wang L, Herberich B, King DS and Schultz PG, "An efficient system for the evolution of aminoacyl-tRNA synthetase specificity," *Nat Biotechnol*. **2002**, *20*:1044.

[5] Wang L, Brock A, Herberich B and Schultz PG, "Expanding the genetic code of *Escherichia coli*," *Science*. **2001**, *292*:498.

[6] Summerer D, Chen S, Wu N, Deiters A, Chin JW and Schultz PG, "A genetically encoded fluorescent amino acid," *Proc Natl Acad Sci U S A*. **2006**, *103*:9785.

[7] Wu N, Deiters A, Cropp TA, King D and Schultz PG, "A genetically encoded photocaged amino acid," *J Am Chem Soc*. **2004**, *126*:14306.

[8] Link AJ, Vink MK, Agard NJ, Prescher JA, Bertozzi CR and Tirrell DA, "Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids," *Proc Natl Acad Sci U S A*. **2006**, *103*:10180.

[9] Yoo TH and Tirrell DA, "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 Engl*. **2007**, *46*:5340.

[10] Link AJ, Unpublished results, **2006**

[11] Kekenes-Huskey PM, Vaidehi N, Floriano WB and Goddard WA, "Fidelity of phenylalanyl-tRNA synthetase in binding the natural amino acids," *J Phys Chem B*. **2003**, *107*:11549.

[12] Datta D, Vaidehi N, Zhang DQ and Goddard WA, "Selectivity and specificity of substrate binding in methionyl-tRNA synthetase," *Protein Sci*. **2004**, *13*:2693.

[13] McClendon CL, Vaidehi N, Kam VWT, Zhang DQ and Goddard WA, "Fidelity of seryl-tRNA synthetase to binding of natural amino acids from HierDock first principles computations," *Protein Eng Des Sel*. **2006**, *19*:195.

[14] Wang P, Vaidehi N, Tirrell DA and Goddard WA, "Virtual screening for binding of phenylalanine analogues to phenylalanyl-tRNA synthetase," *J Am Chem Soc*. **2002**, *124*:14442.

[15] Zhang DQ, Vaidehi N, Goddard WA, Danzer JF and Debe D, "Structure-based design of mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase for incorporation of o-methyl-l-tyrosine," *Proc Natl Acad Sci U S A*. **2002**, *99*:6579.

[16] Datta D, Wang P, Carrico IS, Mayo SL and Tirrell DA, "A designed phenylalanyl-tRNA synthetase variant allows efficient *in vivo* incorporation

of aryl ketone functionality into proteins," *J Am Chem Soc*. **2002**, *124*:5652.

[17] Serre L, Verdon G, Choinowski T, Hervouet N, Risler JL and Zelwer C, "How methionyl-tRNA synthetase creates its amino acid recognition pocket upon l-methionine binding," *J Mol Biol*. **2001**, *306*:863.

[18] Mechulam Y, Schmitt E, Maveyraud L, Zelwer C, Nureki O, Yokoyama S, Konno M and Blanquet S, "Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features," *J Mol Biol*. **1999**, *294*:1287.

[19] Crepin T, Schmitt E, Mechulam Y, Sampson PB, Vaughan MD, Honek JF and Blanquet S, "Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in escherichia coli methionyl-tRNA synthetase," *J Mol Biol*. **2003**, *332*:59.

[20] Schmidt E, Tanrikulu IC, Yoo TH, Panvert M, Tirrell DA and Mechulam Y, "Switching from an induced fit to a lock and key mechanism in an aminoacyl-tRNA synthetase with modified specificity," *Unpublished manuscript*. **2009**,

[21] Lim KT, Brunett S, Iotov M, McClurg RB, Vaidehi N, Dasgupta S, Taylor S and Goddard WA, "Molecular dynamics for very large systems on massively parallel computers: The MPSim program," *Journal of Computational Chemistry*. **1997**, *18*:501.

[22] Ding HQ, Karasawa N and Goddard WA, "Atomic level simulations on a million particles - the cell multipole method for coulomb and nonbond interactions," *J Chem Phys*. **1992**, *97*:4309.

[23] Mayo SL, Olafson BD and Goddard WA, "Dreiding - a generic force-field for molecular simulations," *Journal of Physical Chemistry*. **1990**, *94*:8897.

[24] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D and Karplus M, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J Phys Chem B*. **1998**, *102*:3586.

[25] Tannor DJ, Marten B, Murphy R, Friesner RA, Sitkoff D, Nicholls A, Ringnalda M, Goddard WA and Honig B, "Accurate first principles calculation of molecular charge distributions and solvation energies from *ab initio* quantum mechanics and continuum dielectric theory," *J Am Chem Soc*. **1994**, *116*:11875.

[26] Kam VWT, "Methods in computational protein design," *PhD Thesis,* California Institute of Technology (Pasadena, CA). **2008**.

[27] Xu X and Goddard WA, "The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties," *Proc Natl Acad Sci U S A*. **2004**, *101*:2673.

[28] Vriend G, "What if: A molecular modeling and drug design program," *J Mol Graph*. **1990**, *8*:52.

[29] Kekenes-Huskey PM, "A monte carlo-based torsion construction algorithm for ligand design," *PhD Thesis,* California Institute of Technology (Pasadena, CA). **2009**.
[30] Kam VWT, Unpublished results, **2005**
[31] Ghosh G, Brunie S and Schulman LH, "Transition state stabilization by a phylogenetically conserved tyrosine residue in methionyl-tRNA synthetase," *J Biol Chem*. **1991**, *266*:17136.
[32] Zamanakos G, "A fast and accurate analytical method for the computation of solvent effects in molecular simulations," *PhD Thesis,* California Institute of Technology (Pasadena, CA). **2001**.
[33] Nicholls A and Honig B, "A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poission-Boltzmann equation," *Journal of Computational Chemistry*. **1991**, *12*:435.
[34] Link AJ and Tirrell DA, "Cell surface labeling of *Escherichia coli* via copper(I)-catalyzed [3+2] cycloaddition," *J Am Chem Soc*. **2003**, *125*:11164.
[35] Kiick KL, Weberskirch R and Tirrell DA, "Identification of an expanded set of translationally active methionine analogues in *Escherichia coli*," *FEBS Lett*. **2001**, *502*:25.
[36] Kiick KL, Saxon E, Tirrell DA and Bertozzi CR, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proc Natl Acad Sci U S A*. **2002**, *99*:19.
[37] Link AJ, Vink MK and Tirrell DA, "Presentation and detection of azide functionality in bacterial cell surface proteins," *J Am Chem Soc*. **2004**, *126*:10598.
[38] Tchertanov L, "Structural metrics relationships in covalently bonded organic azides," *Acta Crystallogr B*. **1999**, *55*:807.
[39] Tanrikulu IC, Unpublished results, **2008**
[40] Ghosh A and Vishveshwara S, "A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis," *Proc Natl Acad Sci U S A*. **2007**, *104*:15711.
[41] Pal TK and Sankararamakrishnan R, "Self-contacts in Asx and Glx residues of high-resolution protein structures: Role of local environment and tertiary interactions," *Journal of Molecular Graphics & Modelling*. **2008**, *27*:20.
[42] Hausmann CD and Ibba M, "Aminoacyl-tRNA synthetase complexes: Molecular multitasking revealed," *FEMS Microbiol Rev*. **2008**, *32*:705.
[43] Kiick KL and Tirrell DA, "Protein engineering by *in vivo* incorporation of non-natural amino acids: Control of incorporation of methionine analogues by methionyl-tRNA synthetase," *Tetrahedron*. **2000**, *56*:9487.
[44] Kyte J and Doolittle RF, "A simple method for displaying the hydropathic character of a protein," *J Mol Biol*. **1982**, *157*:105.

Chapter 4

Future Directions for the Discovery of New Methionyl-tRNA Synthetase Activities

Ligand Recognition by Methionyl-tRNA Synthetase and Its Mutants

The methionine (Met) side chain is commonly accepted to be hydrophobic. However, unlike the related valyl-, isoleucyl- and leucyl-tRNA synthetases, the amino acid binding pocket of the methionyl-tRNA synthetase (MetRS) contains a series of polar residues forming an intricate network of hydrogen bonds. (Figures 3.1.b; Figure C.1) Upon Met binding, these polar groups recognize the ligand side chain by donating hydrogen bonds to the Sδ atom on Met. It is foreseeable that in the absence of these polar contacts ligand selectivity would be compromised and other hydrophobic moieties in the environment might compete with Met for the binding site. In addition, the MetRS binding pocket experiences a large conformational change upon ligand binding. To enable this conformational change, the W253 residue, which rotates more than 90° around χ1 and forms one side of the binding pocket, has to be labile. (Figure 2.16.a) Because of this, small incompatibilities that arise when binding a non-cognate ligand might be better tolerated by MetRS than other similar synthetases, like isoleucyl- or valyl-tRNA synthetase. Indeed, in a computational study of the fidelity of MetRS to ligand binding, Datta et al. identified leucine (Leu) as the primary competitor of Met [1]. (Figure 4.1) Thus, having ligand-specific hydrogen bonding interactions likely allows MetRS to correctly identify Met among other hydrophobic amino acids.

Although hydrogen bonds can help in ligand recognition, hydrogen bonds to organic sulfur atoms are relatively weak and hydrogen bonds are not very prevalent for methionine residues in proteins [2, 3]. Therefore, while the availability of hydrogen bond donors in the binding site will improve discrimination against hydrophobic amino acids, it may also create new competitors by allowing polar amino acids to bind MetRS. Although the interaction of any polar competitor with the MetRS binding pocket will be suboptimal, the competitor may compensate for that by forming strong hydrogen bonds with the donor groups in the binding site. The amino acids glutamic acid (Glu) and glutamine (Gln) were identified by Datta et al. as the best competitors for the MetRS binding site after Met and Leu [1]. Interestingly, when methionine-auxotrophic *E. coli* with elevated

levels of the wild-type MetRS is induced to express protein in a methionine-deficient medium, expression of protein is observed. (Figure B.1) Mass spectrometry analysis of tryptic fragments from the expressed protein indicate the replacement of Met by Glu. (Figure B.3, first panel) This observation supports the notion that MetRS needs a fine balance between hydrophobic and polar elements in the binding site to be able to recognize the Met side chain, which is hydrophobic but also is a weak hydrogen bond acceptor.

In our studies, we aimed to discover mutations in the MetRS binding site that permit the activation and the subsequent charging of the noncanonical amino acid azidonorleucine (Anl) onto $tRNA^{Met}$. By screening a MetRS saturation mutagenesis library, we identified a diverse set of mutants that allow azidonorleucine (Anl) incorporation *in vivo*. Quantum mechanical (QM) calculations on the azide group indicate a high extent of charge separation between the azide nitrogens, making the central nitrogen atom electron poor (δ = +0.5) with respect to the terminal and base nitrogen atoms (δ = -0.3 and -0.5, respectively). Thus, mutations that recognize the polar character of the azide group were anticipated from the screens *in vivo*. However, the results of the screens revealed hydrophobic substitutions for Y260 and H301, the two sites best positioned for polar interactions with Anl. While the experimentally obtained MetRS mutants preferred non-polar side chains, the computational models predicted that many mutants in the saturation mutagenesis library could hydrogen bond with the azide group. If hydrogen bonding interactions are available for the ligand, it is puzzling why such interactions are not utilized.

A survey of crystal structures of organic azides concluded that hydrogen bonds to organic azides are weak based on their frequency and the donor-acceptor distance [4]. It is possible, then, that the same problem is encountered with Anl, as with MetRS and Met. Placement of polar groups into the binding site may promote polar amino acids competing for this site. Unlike in the wild-type MetRS, the screen might be choosing hydrophobic residues compatible with binding Anl to avoid such a competition. The L13G mutant of MetRS, which displays good activation kinetics for Anl, is an example of a mutant that suffers

from a competition with polar amino acids. In the absence of Met and Anl, the L13G mutant allows the incorporation of glutamine (Gln) or lysine (Lys) into Met sites in proteins, based on the MALDI-MS analysis of tryptic fragments. (Figure 2.10) The screening strategy we employ for selecting MetRS mutants active toward Anl is also sensitive toward the presence of any competitors. Met is a strong competitor against Anl for the MetRS-L13G binding site and the cell-surface fluorescence labeling drops sharply in the presence of Met for cells equipped with this mutant. (Figure 2.11) Gln and/or Lys are weaker competitors than Met and their incorporation through MetRS-L13G allows them to replace Anl in about 50% of the Met sites in the presence of 0.3 mM Anl, and 0.27 mM Gln and Lys. This likely is responsible for the lower cell-surface fluorescence levels achieved through this mutant at 0.3 mM Anl (Figure 2.8) when compared with other MetRS mutants that have similar Anl activation characteristics, such as the NLL and SLL mutants (Table 2.3), and explains why this mutant was not encountered in the screens performed.

Selected MetRS Mutants May Display Broad Activities

The selection of hydrophobic side chains in the MetRS binding site opens the possibility for other aliphatic amino acids with largely hydrophobic side chains to be incorporated by the selected enzymes. The SLL mutant was previously identified by Yoo and Tirrell in an independent screen of a MetRS library for the incorporation of 6,6,6-trifluoronorleucine (Tfn) [5]. Ethionylnorleucine (Enl), which bears the alkyne functionality and can be ligated to terminal azides through a Cu(I)-catalyzed [3+2] azide-alkyne ligation, is another ligand of interest for MetRS. When a series of MetRS mutants previously characterized with Anl (Figure 2.9) were tested for protein expression in the presence of Enl and absence of Met, protein expression was achieved through all MetRS mutants tested, strongly suggesting the incorporation of Enl into proteins. (Figure 4.2.)

The screen by Yoo and Tirrell employed both negative (against incorporation of 20 canonical amino acids) and positive (for incorporation of Tfn) screening strategies to reach the MetRS-SLL mutant that allows Tfn

incorporation. Screens we have carried out for Anl incorporation in the presence of 20 canonical amino acids have returned the NLL and PLL mutants for incorporating Anl, while discriminating against canonical amino acids. (Figure 2.7; Table 2.2) These results suggest that Y260L-H301L mutants of MetRS that also carry a small and/or polar residue at position 13 (XLL) may be less sensitive to competition by Met and other canonical amino acids. These mutants may, therefore, be good candidates for exploring new noncanonical amino acids for incorporation into proteins *in vivo*.

Creating a More Hydrophobic MetRS Binding Pocket

MetRS mutants of the XLL-type present a more hydrophobic binding pocket than wild-type MetRS, but they still carry a water molecule deep in the binding pocket. In the MetRS-SLL crystal structure, this water molecule interacts with the terminal azide nitrogen of Anl through a hydrogen bond [6] (Figures 3.1.c) and its presence in the binding site is detected in various ligand-bound and ligand-free structures of MetRS [6-9]. The presence of a hydrogen-bond donor deep in the binding pocket may allow better recognition of ligands that carry a terminal hydrogen-bond acceptor. Nζ-acetyllysine (NaK) may be such a ligand. This amino acid is produced *in vivo* through the enzymatic acetylation of lysine, and is implicated in histone reorganization [10]. The terminal amide oxygen might enable this ligand to be efficiently recognized by the XLL-type mutants. However, the presence of this water molecule may interfere with the activation of amino acids such as 2-aminooctanoic acid (Aoc), which do not have terminal polar groups. In fact, the MetRS-SLL mutant allows only very limited levels of protein expression in the presence of this ligand [11]. For efficient incorporation of such amino acids, it may be necessary to remodel the binding pocket to exclude this water molecule.

In the MetRS binding pocket, the water molecule is held tightly by a hydrogen bond from the side chain of residue T10, and a hydrogen bond to the backbone carbonyl group of F292. (Figure 4.3) Although the second hydrogen of the water molecule forms a hydrogen bond with the ligand in the Anl-bound

MetRS-SLL crystal structure, in the wild-type enzyme this hydrogen bond is formed with the Oη atom on the Y260 side chain. (Figure 3.1.b.) Because the Y260 Oη atom also engages in a hydrogen bond with the ligand (Met) in the wild-type MetRS crystal structure, it was suggested that this water molecule was important for enzyme function [8]. However, because many different types of mutations are permitted at position 260, we believe that the removal of the water molecule through mutations may also be permitted. In addition to its hydrogen bond to the water, the T10 residue is stabilized by a hydrogen bond from the H291 side chain. It may be possible to engineer positions 10 and 291, together with the nearby position 260, and determine if it is possible to replace this hydrophilic region in the binding site with hydrophobic residues. Such a modification may create a MetRS binding pocket with more hydrophobic character, and enable better incorporation of long-chain hydrophobic residues such as Aoc.

Proteomic Applications of MetRS-NLL

Azidohomoalanine (Aha) is an azide-bearing amino acid, readily incorporated into proteins through the wild-type MetRS in bacterial and eukaryotic cells [12, 13]. Taking advantage of this and the bio-orthogonal nature of the azide group, Dieterich and co-workers showed that proteins synthesized in cells during an Aha ‘pulse’ can be tagged with the azide group. This enabled the researchers to selectively enrich proteins synthesized in neurons in response to stimuli, and analyze their composition [13]. Called bio-orthogonal noncanonical amino acid tagging (BONCAT), this method promises to significantly decrease sample complexity during proteomic analyses by restricting the sample composition to proteins synthesized at events of interest in a cell’s life.

However, because Aha is taken up by a large variety of cells, this method has limited applications to situations where the response of a single organism, or cell type is to be investigated in a mixture of cells. In investigations of neurons in the presence of glia, or of microbes infecting a eukaryotic host, where proteins from the subject make up a small fraction of the total protein mass, gathering

useful proteomic data from the subject can be very challenging. In these cases, it is not only necessary to restrict metabolic labeling temporally, but also to a specific cell type in the culture. This can be accomplished by giving the subject the ability to incorporate an azide-bearing amino acid like Anl that is not incorporated by the wild-type protein synthesis machinery. By equipping only the subject with a mutant MetRS (MetRS-NLL), metabolic labeling can be directed to only the cells of interest. Other cells in the culture lacking an Anl-active mutant MetRS will be inert to this amino acid. Proteomic responses from the subject can then be examined through the use of a variety of reactive probes for isolation or visualization. Ngo, Champion and co-workers have have demonstrated the utility of this strategy in bacterial co-cultures, as well as in co-cultures of bacteria and murine alveolar macrophages [14]. By infecting macrophages with *E. coli* equipped with MetRS-NLL in the presence of Anl, the researchers were successful in restricting metabolic labeling to proteins in the bacteria, almost exclusively.

The success of this study establishes the use of mutant AARSs to direct metabolic labeling. This is a brand new role for the AARS. The above concept is applicable to other reactive amino acids rejected by the wild-type metabolism, such as Enl. It is easy to envision simultaneous labeling of interacting cells that make up an organism, through the use of multiple AARS mutants and ligands. Such studies may better expose intricate connections in complex cell communities. Therefore, the current effort proves a starting point for many future biological and structure-based design studies.

Figure 4.1.

Structures of natural and noncanonical methionine analogs.

The chemical structure of methionine is shown with other natural amino acids that may be substrates for the wild-type or mutant MetRS. Structures for several noncanonical amino acids that are or may be substrates for MetRS mutants are also displayed.

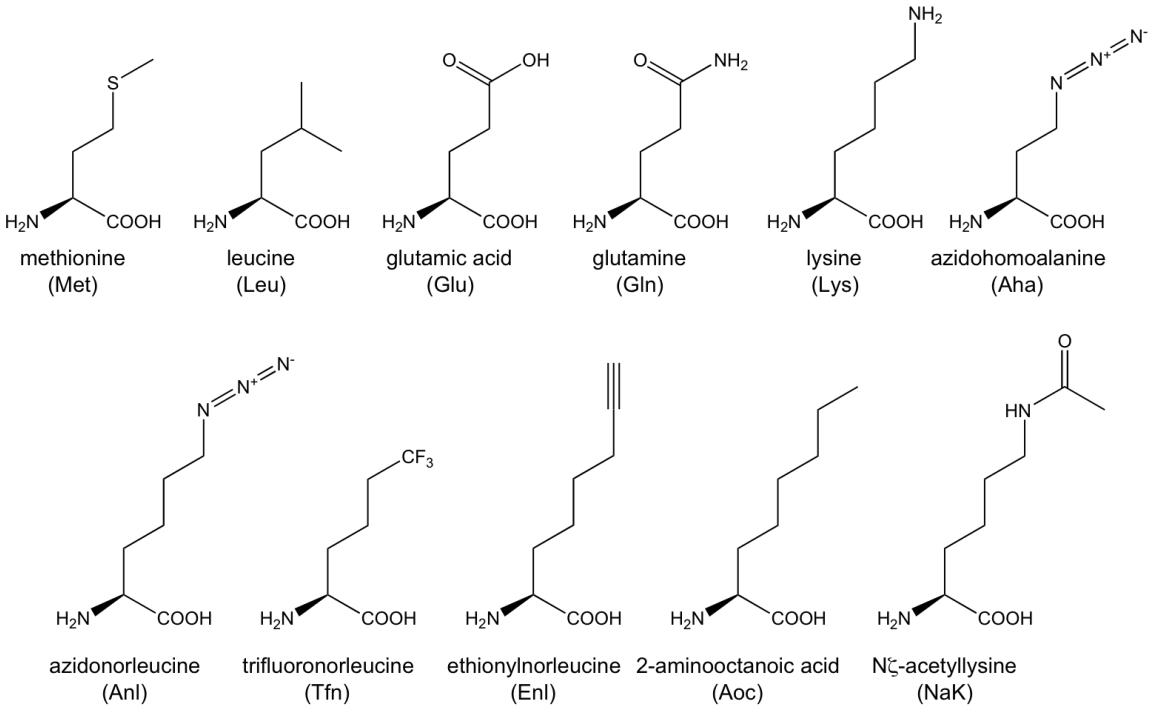


Figure 4.2.

Expression of DHFR in the presence of Enl in cells bearing various MetRS mutants.

DHFR was expressed in M15MA[pREP4/pAJL-61] cells encoding the MetRS mutants isolated from screens carried out at 0.3 mM (blue) and 1.0 mM Anl (green), as well as the L13G mutant and wild-type MetRS. Expression was done in M9+19aa media supplemented with 2.0 mM *dl*-Enl (gift from John T. Ngo; labeled "+Enl"), 40 mg/L Met ("+Met"), or no 20th amino acid ("19aa"). An aliquot taken before induction is labeled "pre-ind." SDS-PAGE analysis of whole-cell lysates show DHFR expression in the presence of Enl for all MetRS mutants.

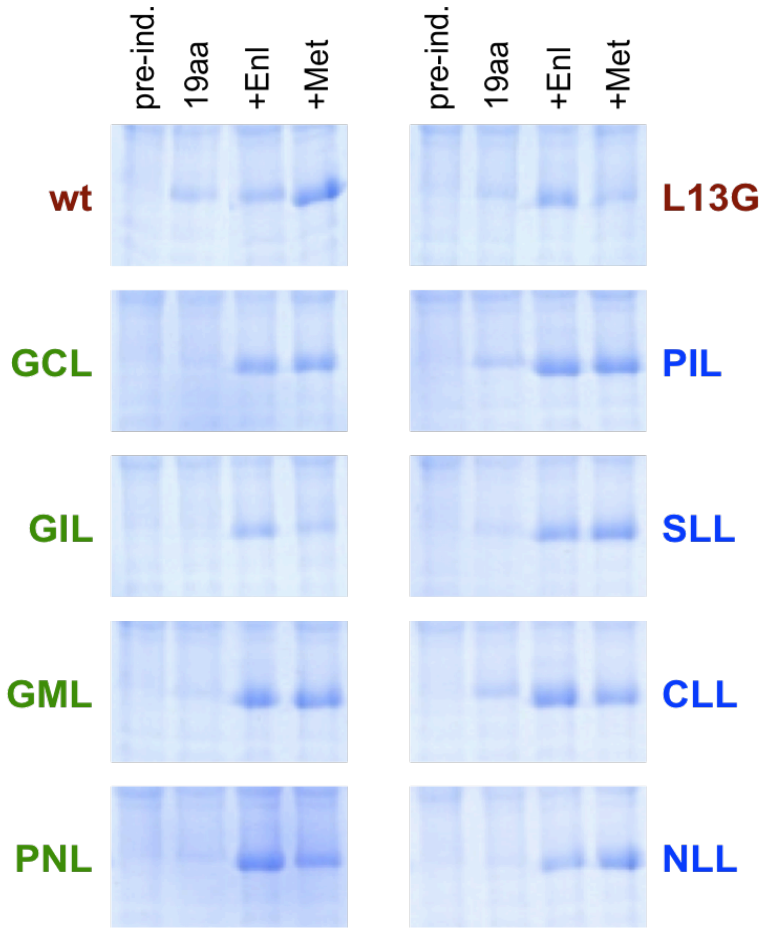
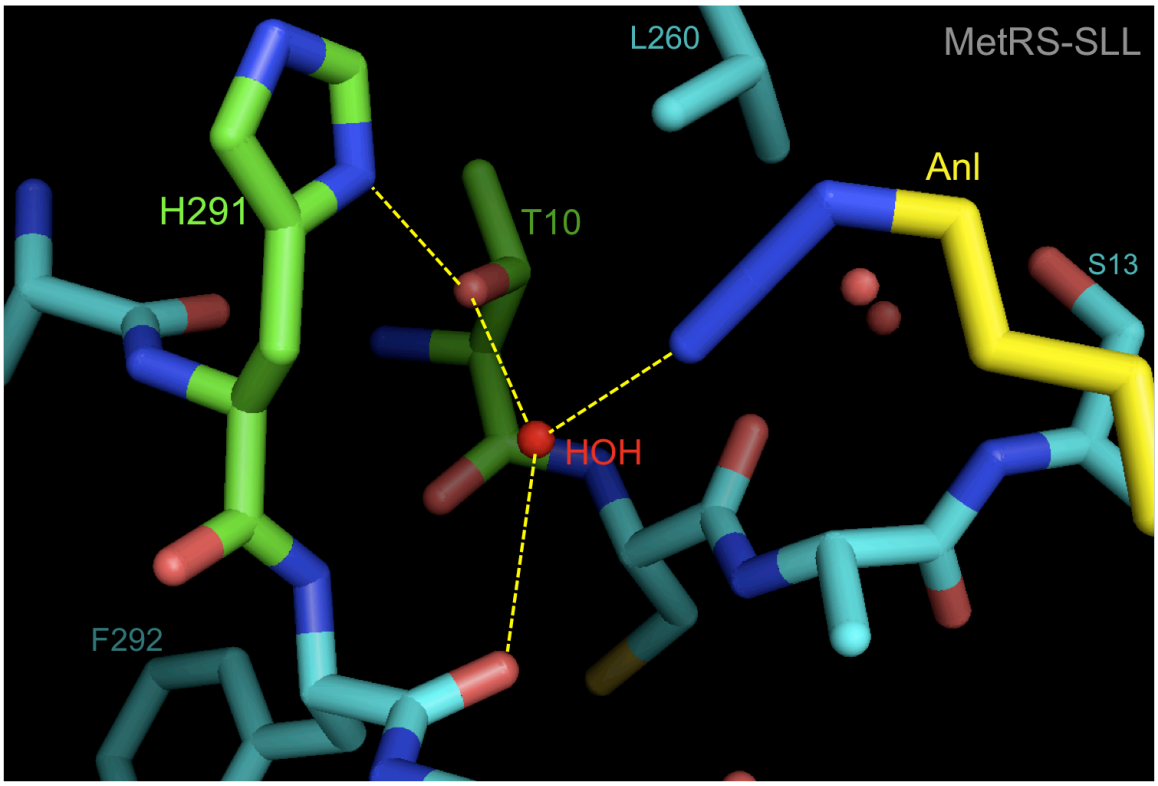


Figure 4.3.

Interactions of a conserved water molecule in the MetRS binding site.

The interactions of the conserved water molecule with the MetRS-SLL binding site and the ligand Anl are shown. The water molecule is labeled “HOH” and dashed lines mark important hydrogen bonds between the ligand and residues in the binding site. The residues in the binding site that facilitate the association of the water molecule are shown in green, and the ligand in yellow. This representation was prepared using PyMol, based on coordinates from [6].



References

[1] Datta D, Vaidehi N, Zhang DQ and Goddard WA, "Selectivity and specificity of substrate binding in methionyl-tRNA synthetase," *Protein Sci*. **2004**, *13*:2693.

[2] Panigrahi SK and Desiraju GR, "Strong and weak hydrogen bonds in the protein-ligand interface," *Proteins*. **2007**, *67*:128.

[3] Gregoret LM, Rader SD, Fletterick RJ and Cohen FE, "Hydrogen bonds involving sulfur atoms in proteins," *Proteins*. **1991**, *9*:99.

[4] Tchertanov L, "Structural metrics relationships in covalently bonded organic azides," *Acta Crystallogr B*. **1999**, *55*:807.

[5] Yoo TH and Tirrell DA, "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 Engl*. **2007**, *46*:5340.

[6] Schmidt E, Tanrikulu IC, Yoo TH, Panvert M, Tirrell DA and Mechulam Y, "Switching from an induced fit to a lock and key mechanism in an aminoacyl-tRNA synthetase with modified specificity," *Unpublished manuscript*. **2009**,

[7] Crepin T, Schmitt E, Mechulam Y, Sampson PB, Vaughan MD, Honek JF and Blanquet S, "Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase," *J Mol Biol*. **2003**, *332*:59.

[8] Serre L, Verdon G, Choinowski T, Hervouet N, Risler JL and Zelwer C, "How methionyl-tRNA synthetase creates its amino acid recognition pocket upon l-methionine binding," *J Mol Biol*. **2001**, *306*:863.

[9] Mechulam Y, Schmitt E, Maveyraud L, Zelwer C, Nureki O, Yokoyama S, Konno M and Blanquet S, "Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features," *J Mol Biol*. **1999**, *294*:1287.

[10] Owen DJ, Ornaghi P, Yang JC, Lowe N, Evans PR, Ballario P, Neuhaus D, Filetici P and Travers AA, "The structural basis for the recognition of acetylated histone h4 by the bromodomain of histone acetyltransferase gcn5p," *EMBO J*. **2000**, *19*:6141.

[11] Yoo TH, Unpublished results, **2007**

[12] Kiick KL, Saxon E, Tirrell DA and Bertozzi CR, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proc Natl Acad Sci U S A*. **2002**, *99*:19.

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

[14] Ngo JT, Champion JA, Mahdavi A, Tanrikulu IC, Beatty KE, Connor RE, Yoo TH, Dieterich DC, Schuman EM and Tirrell DA, "Cell-selective metabolic labeling of proteins," *Unpublished manuscript*. **2009**,

Appendix A.

List of E. coli *Methionyl-tRNA Synthetase Mutants Identified Through Library Screening to Show Activity Toward Azidonorleucine* In Vivo

A list of all methionyl-tRNA synthetase (MetRS) mutants identified from clones isolated through various screens of the LYH.1.0 saturation mutagenesis library for activity toward azidonorleucine (Anl) is presented here. Only variants carrying mutations at the three mutagenised sites (L13, Y260, and H301) are listed. The populations each mutant has appeared in are also indicated. These populations are explained below. From 150 active clones, a total of 41 distinct MetRS mutants were identified. A complete list of these mutants are provided in the following page.

Table A.1

List of various populations that were obtained by screening the LYH.1.0 library

Population	Information
1.2	See LYH.1.2 in Figure 2.6.
2.2	See LYH.2.2 in Figure 2.6.
2.3	Revealed by screening 2.2 at 0.1 mM Anl
3.2	See LYH.3.2 in Figure 2.6.
3.3	Revealed by screening 3.2 at 0.3 mM Anl
4.3	Revealed by screening 3.2 at 1.0 mM Anl
5.3c	The mixture of LYH.1.1, LYH.2.1 and LYH.3.1 (See Figure 2.6.) screened two rounds with 1.0 mM Anl and Met
5.3d	See LYH.5.3d in Figure 2.7.
5.4b	The mixture of LYH.1.1, LYH.2.1 and LYH.3.1 (See Figure 2.6.) screened three rounds with 0.3 mM Anl and Met
6.2	See LYH.6.2 in Figure 3.?

Table A.2

List of MetRS mutants identified through library screening to activate Anl in vivo

L13	Y260	H301	Occurrence	Populations clones are isolated from:							
A	C	L	1	1.2							
A	Q	L	2	1.2	6.2						
A	V	L	3	6.2							
C	L	L	2	3.2							
C	V	L	1	1.2							
D	L	I	1	1.2							
F	G	L	1	6.2							
F	V	L	1	3.3							
G	C	L	2	1.2							
G	I	L	3	1.2							
G	L	L	2	3.2	5.3d						
G	L	V	1	6.2							
G	M	L	4	1.2							
G	N	L	3	3.2	3.3	4.3					
G	T	L	4	1.2	3.3	6.2					
G	V	L	4	6.2							
H	C	L	1	1.2							
H	T	M	1	1.2							
L	C	L	1	6.2							
M	A	L	1	1.2							
M	S	L	1	6.2							
N	L	L	62	2.2	2.3	3.2	3.3	4.3	5.3c	5.3d	5.4b
N	V	L	1	1.2							
P	C	L	1	3.2							

L13	Y260	H301	Occurrence	Populations clones are isolated from:
P	I	L	1	3.2
P	I	M	1	1.2
P	L	I	9	2.2 5.3c 5.4b
P	L	L	2	5.3c
P	L	V	1	3.2
P	M	L	1	2.2
P	N	L	5	1.2 3.3 4.3
P	N	V	2	6.2
P	T	V	1	1.2
P	V	L	1	5.3d
S	C	M	1	6.2
S	L	L	12	2.2 2.3 3.2 5.4b
S	L	V	1	4.3
S	M	L	4	3.2 3.3
S	N	L	1	6.2
S	S	L	2	6.2
S	T	L	1	1.2
Total			**150**	

Appendix B.

MALDI-MS Data Demonstrating the Incorporation of Azidonorleucine into Methionine Positions on Tryptic Fragments of Murine Dihydrofolate Reductase

Murine dihydrofolate reductase (DHFR) expression profiles of cells containing different MetRS mutants, and a brief description of the MALDI-MS data on tryptic digests of DHFR were previously presented in Figure 2.9. Here, this data is presented in greater detail.

DHFR was expressed in M15MA[pREP4/pAJL-61] cells encoding various methionyl-tRNA synthetase (MetRS) mutants. These included mutants isolated from screens carried out at 0.3 mM and 1.0 mM azidonorleucine (Anl), as well as the L13G mutant and the wild-type MetRS. Expression was done in five different conditions: 1) in unsupplemented M9+19aa medium (labeled as "19aa" or "0 mM Anl"), 2) in M9+19aa media supplemented with 0.1 mM, 3) 0.3 mM, or 4) 1.0 mM Anl, or 5) with 40 mg/L methionine. The cell lysates standardized for cell densities show the differing amounts of DHFR expressed at each of these conditions. (Figure B.1) DHFR was purified using Ni-NTA affinity chromatography, and the purified protein was analyzed by SDS-PAGE. (Figure B.2)

Purified DHFR was subjected to digestion by trypsin, and the resulting fragments were examined by MALDI-MS. For each sample, the region of the mass spectrum containing the tryptic fragments IMQEFESDTFFPEIDLGKYK (2437.16 Da) and QNLVIMGR (930.52 Da), both of which contain a single methionine position. Replacement of Met by Anl results in a 23.05 Da mass shift, which is shown on each panel. The expected locations of the for Met- and Anl-containing peptides, as well as other major peaks are labeled by their molecular weights. (Figure B.3)

It is apparent in the SDS-PAGE analysis that the wild-type MetRS (labeled "wt") supports protein synthesis even in the absence of methionine. This expression is not affected by the presence or absence of 1 mM Anl in the media.

Peptides from these samples indicate the incorporation of another amino acid 2 Da smaller than methionine, which matches the mass of glutamic acid. Similarly, in the absence of methionine or azidonorleucine, the MetRS-L13G mutant incorporates another amino acid 3 Da smaller than methionine, which corresponds to either lysine or glutamine.

Figure B.1

Expression levels of DHFR at different media compositions.

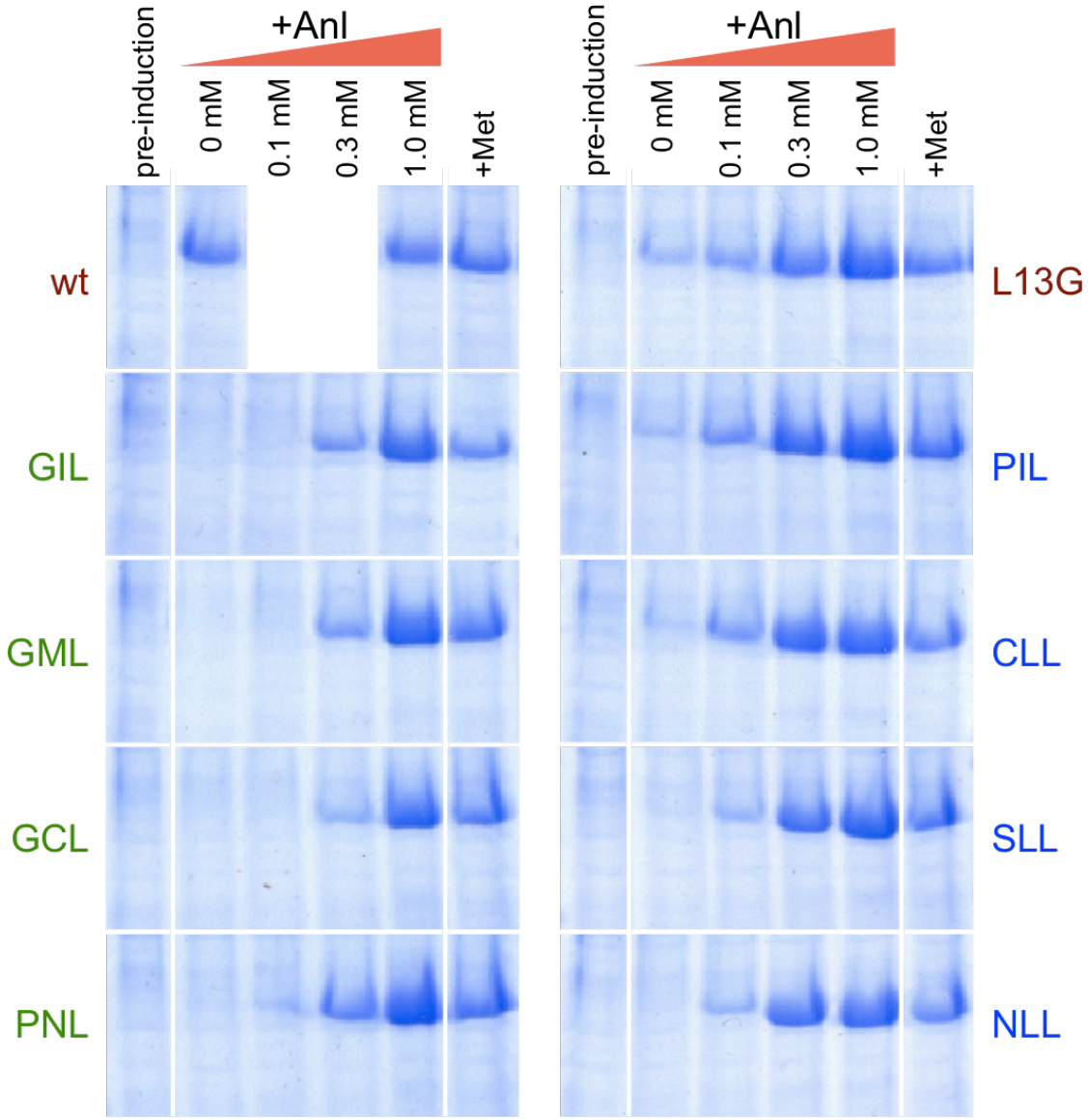


Figure B.2

SDS-PAGE analysis of DHFR induced at different media compositions and purified.

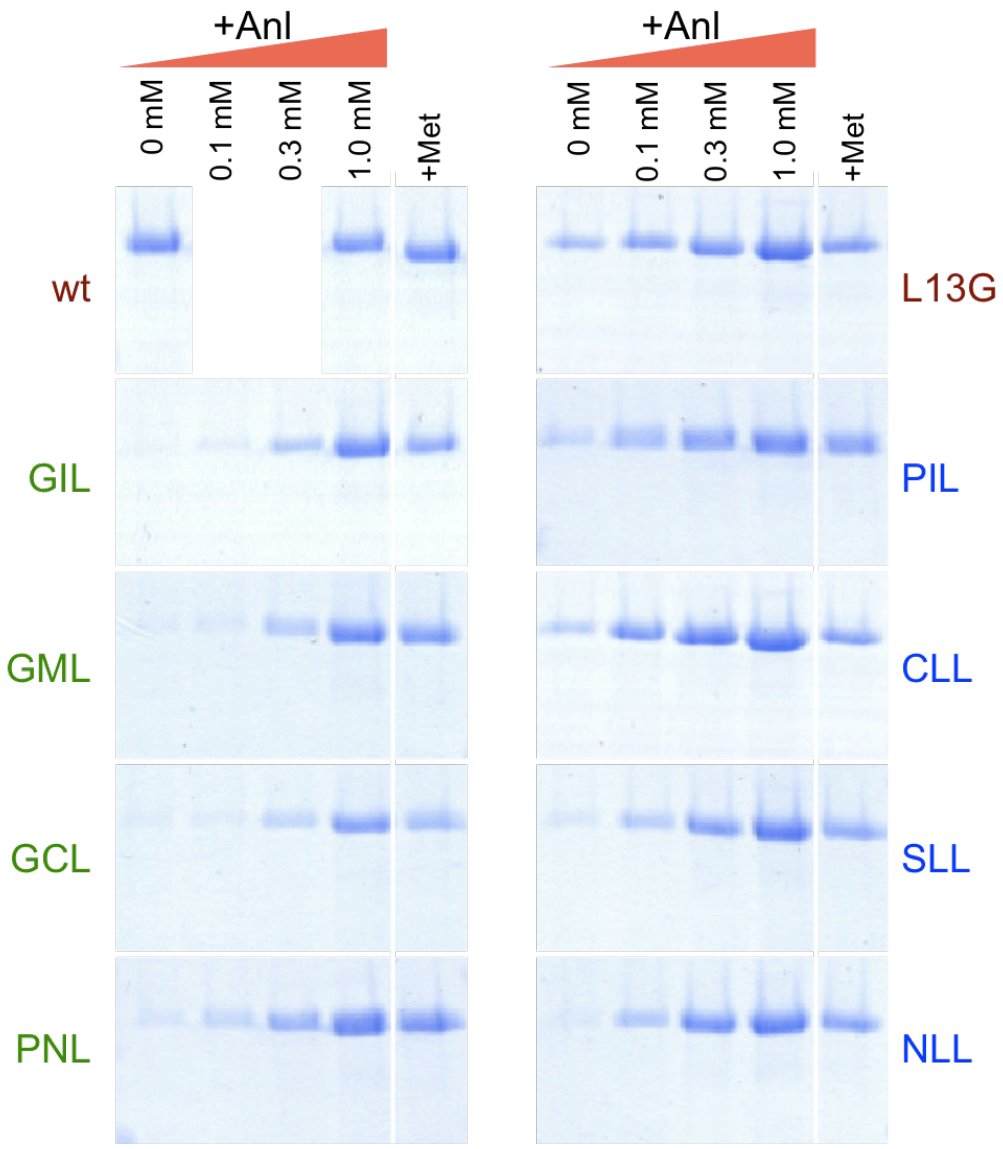


Figure B.3

Mass spectrometric analysis of tryptic fragments of DHFR purified from cells bearing various MetRS variants.

Purified DHFR (Figure B.2) was subjected to digestion by trypsin and the digested peptides were analyzed with MALDI-MS. With the exception of the first panel, each of the following panels show data pertaining to each MetRS variant tested. The initial panel contains data from wild-type MetRS and accompanying controls, showing the locations of Met- and Anl-containing peptides. In addition, data from a digest where DHFR was omitted was also presented (labeled as “Trypsin only”) in order to gauge the contribution of the self-digestion of trypsin to the signal observed. Replacement of Met by Anl increases the mass of each peptide by 23.05 Da, and this shift is presented on each panel by a black arrow, for each peptide analyzed. Red arrows show the differences in mass from the Met-containing peptide to other major peaks that appear nearby, in the absence of Met or Anl.

Figure B.3 (continued)

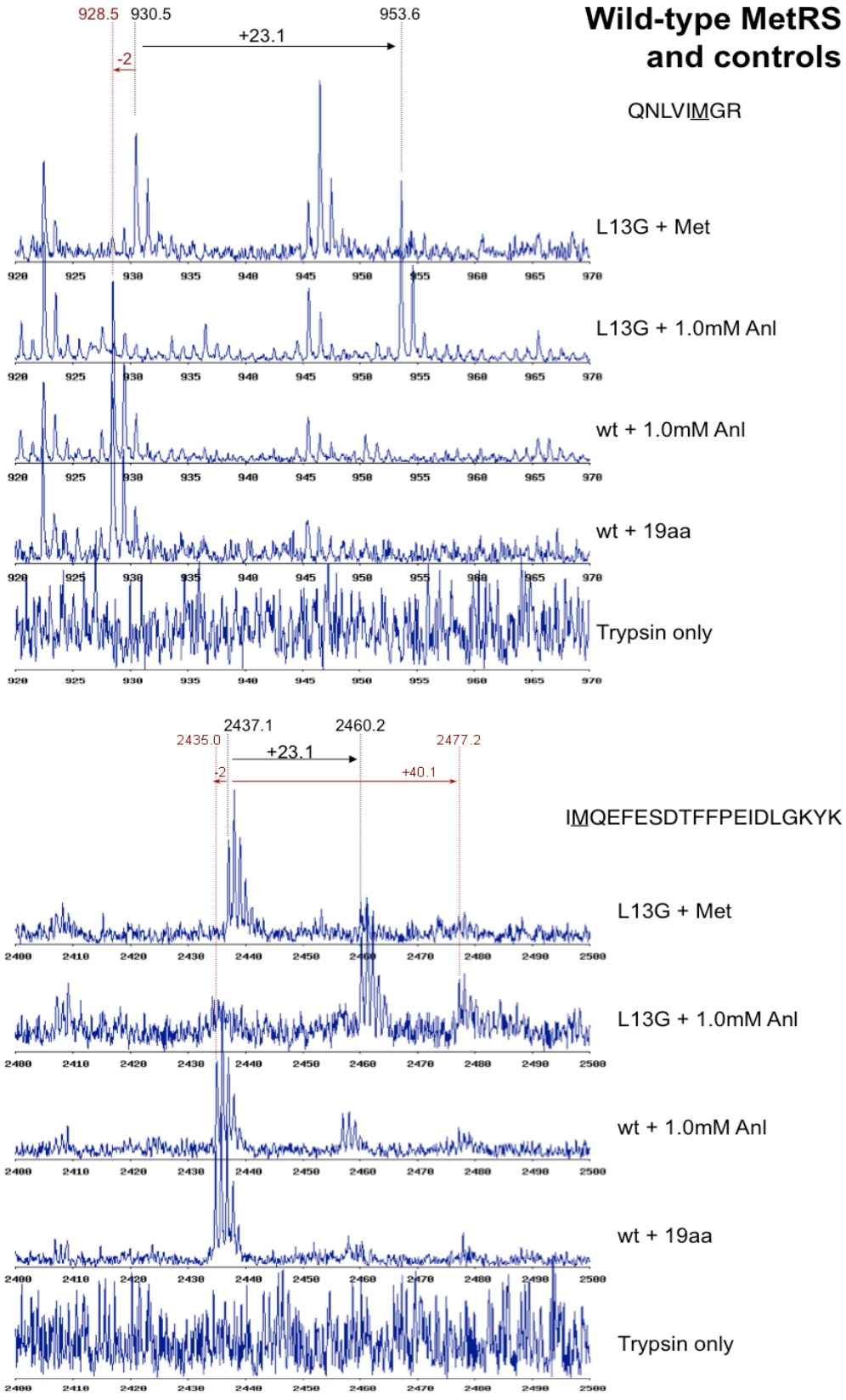


Figure B.3 (continued)

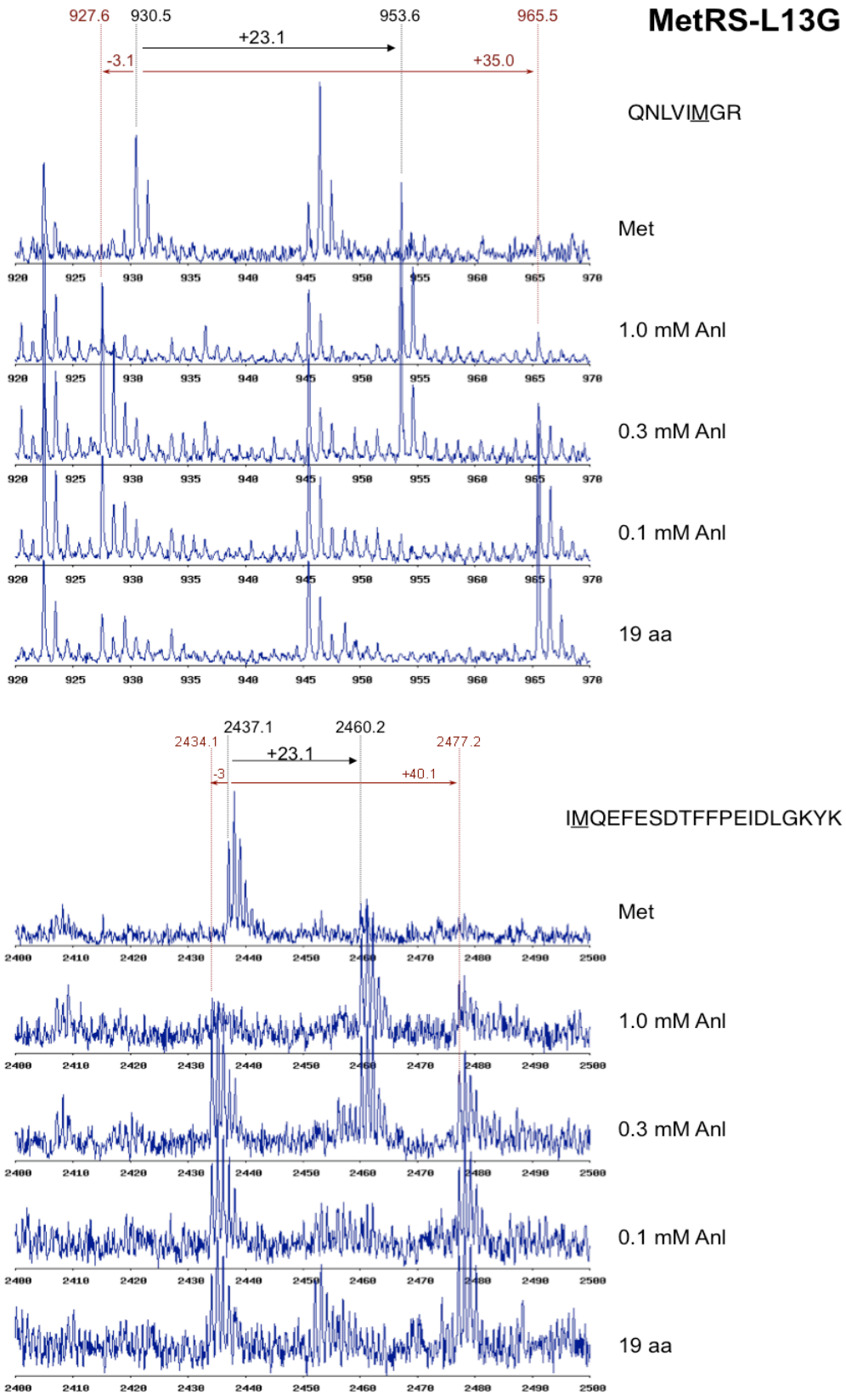


Figure B.3 (continued)

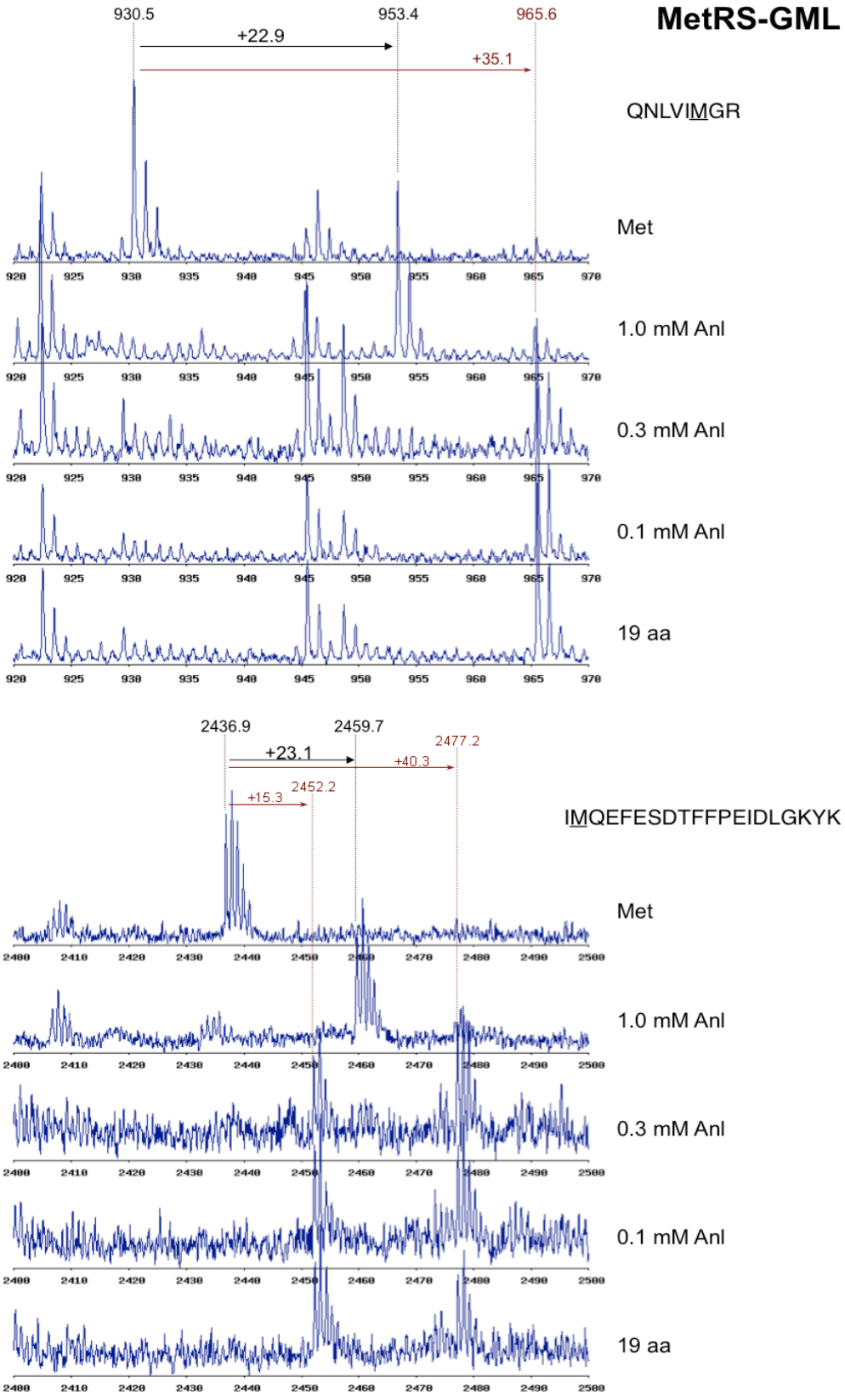


Figure B.3 (continued)

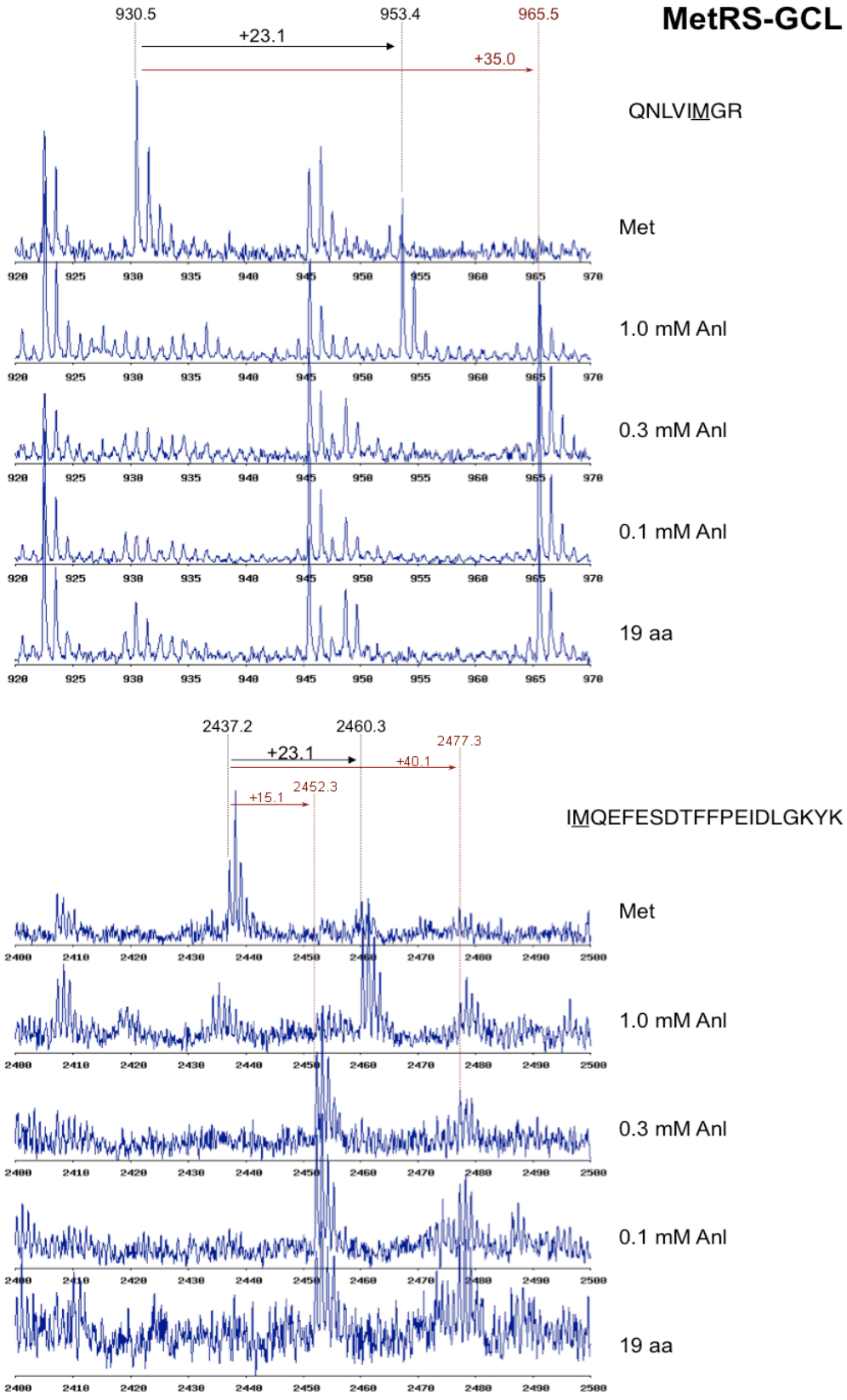


Figure B.3 (continued)

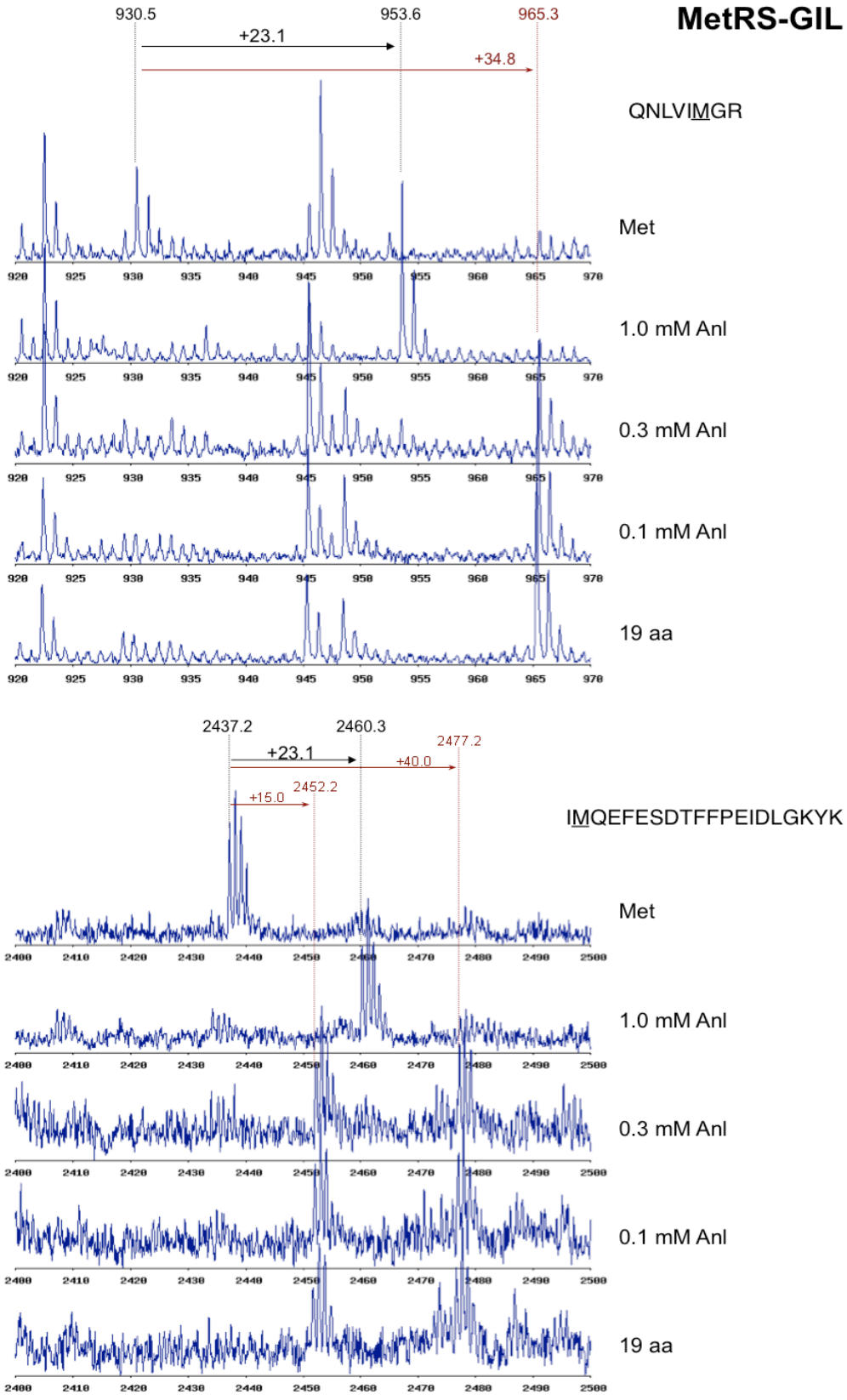


Figure B.3 (continued)

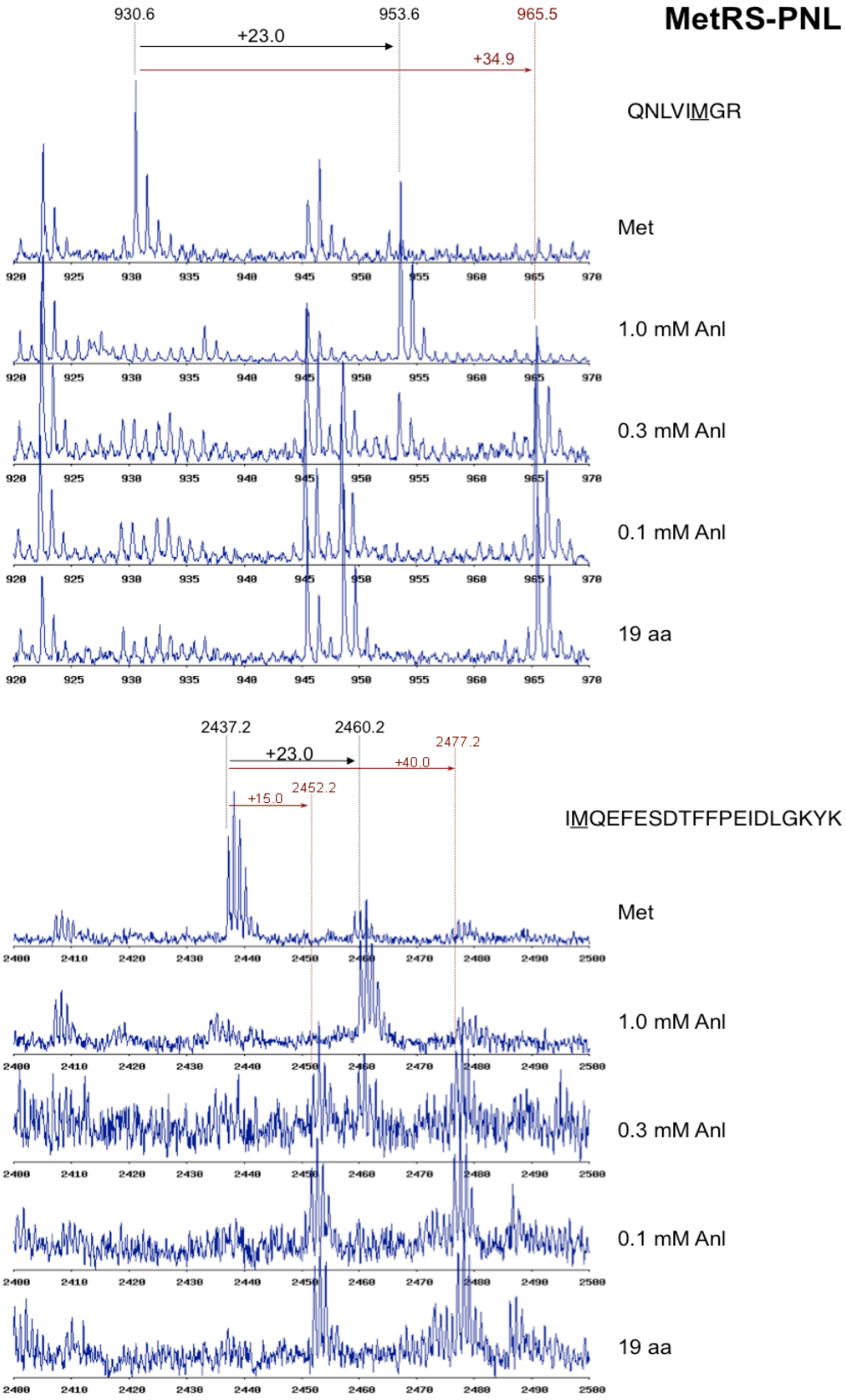


Figure B.3 (continued)

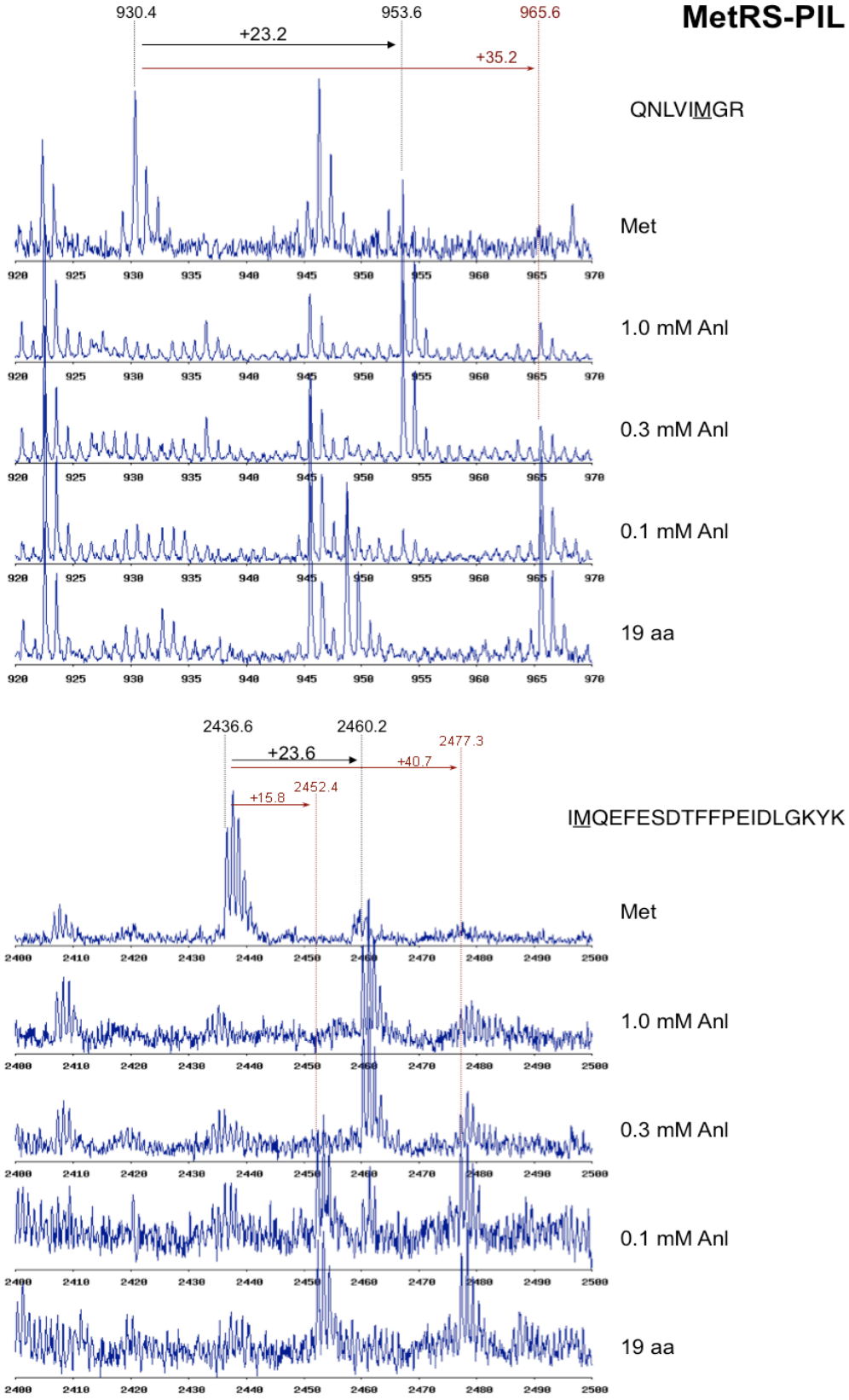


Figure B.3 (continued)

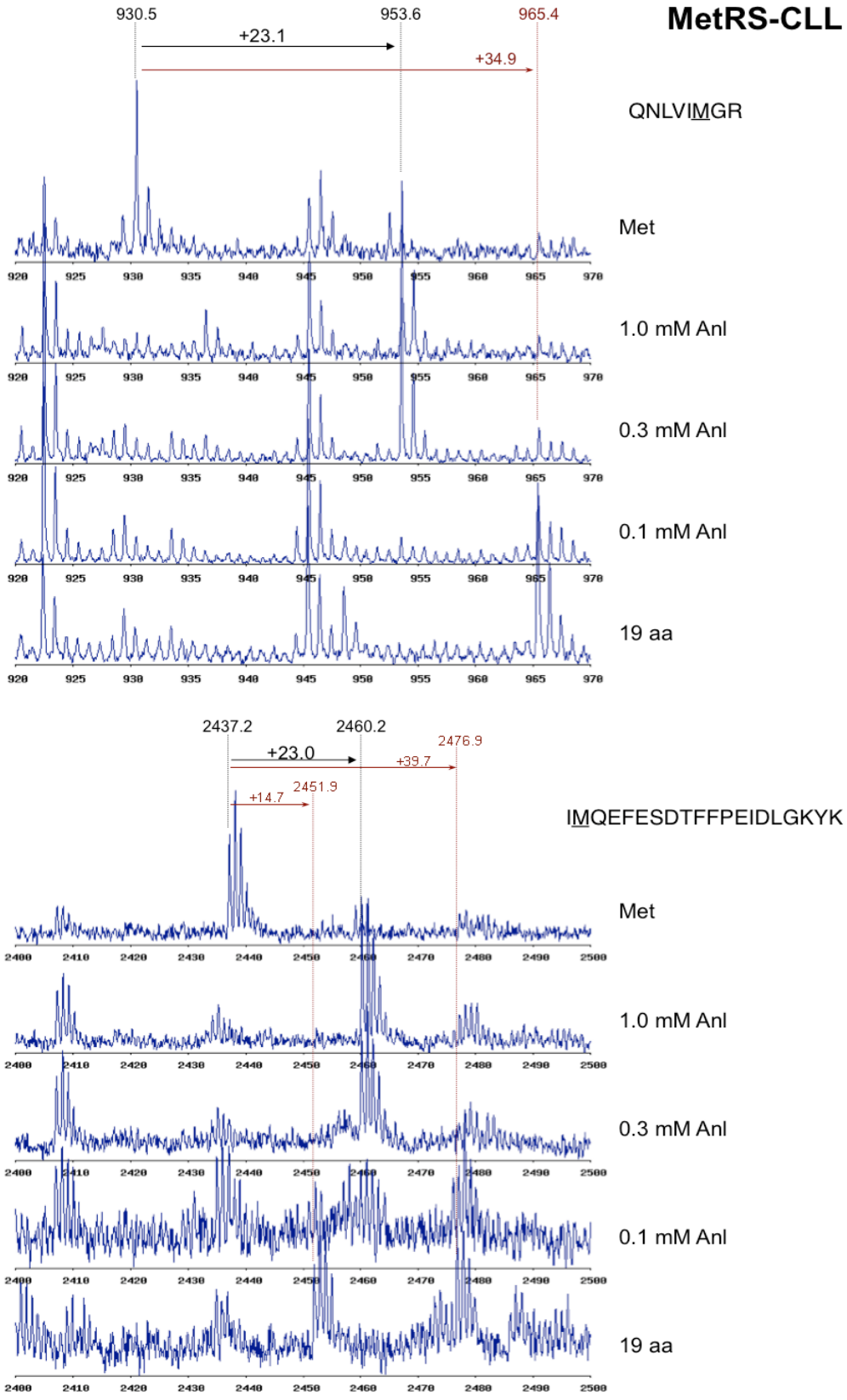


Figure B.3 (continued)

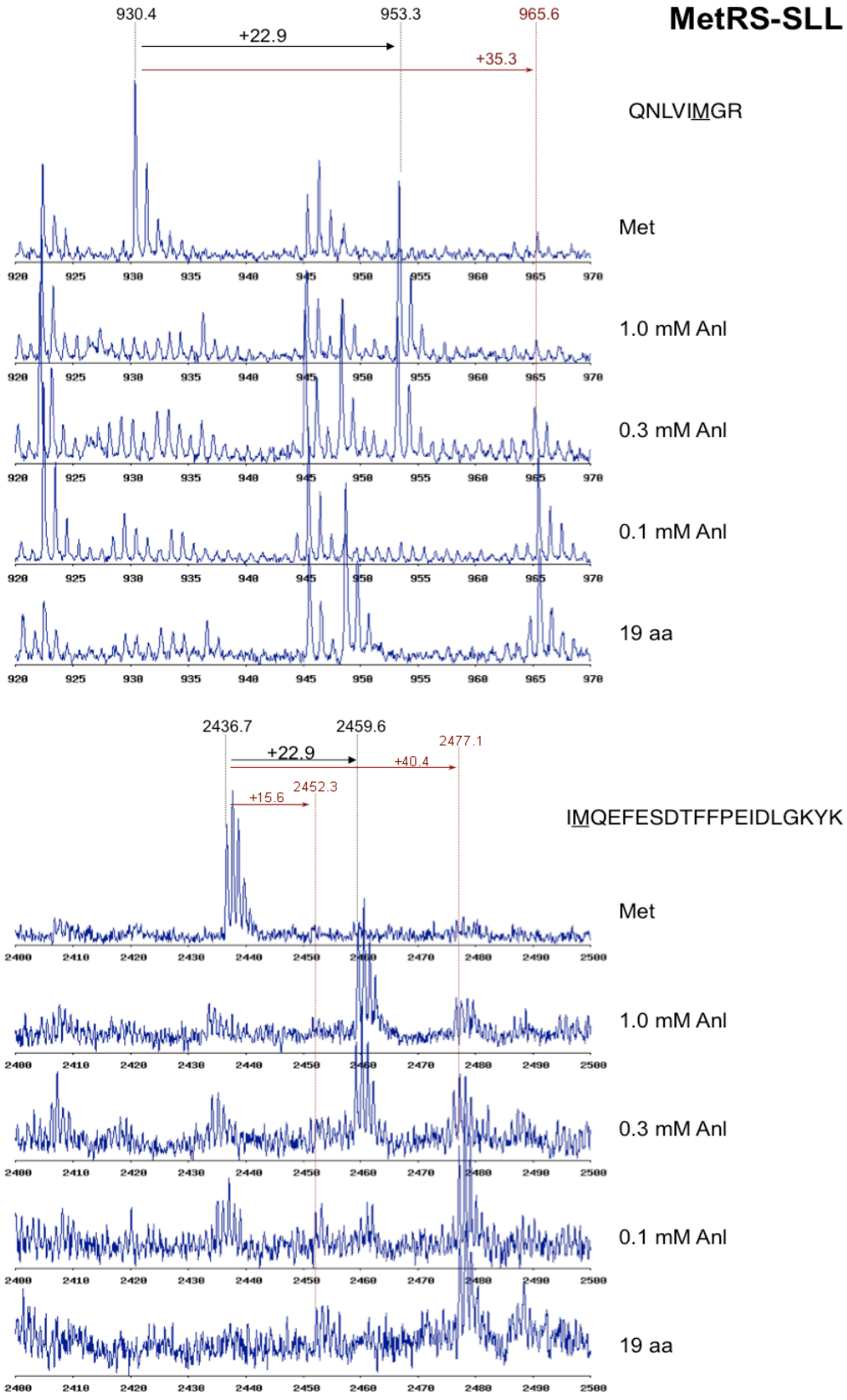
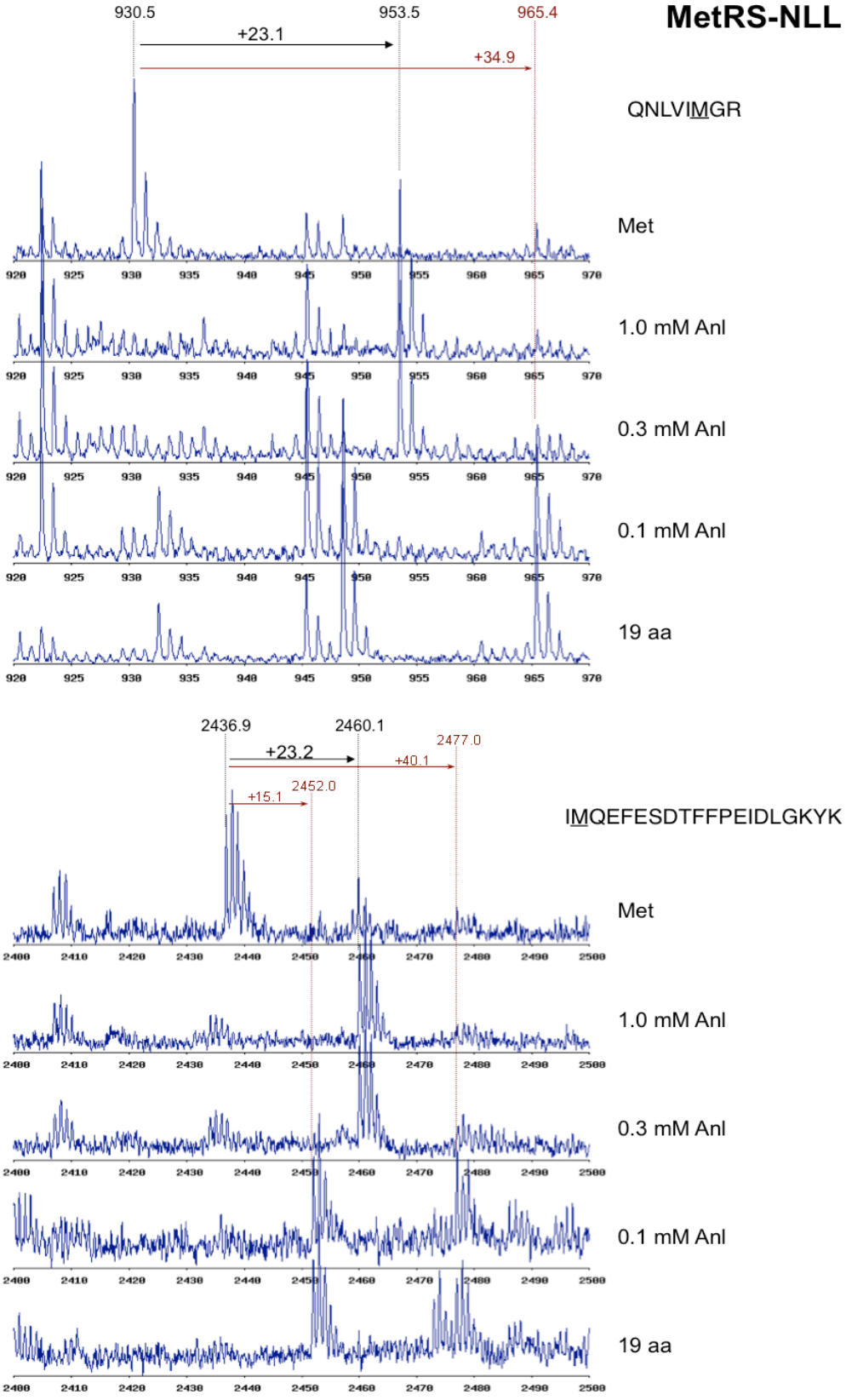


Figure B.3 (continued)



Appendix C.

Summary of Computational Design and Mutation Studies Performed on Aminoacyl-tRNA Synthetases Other than the Methionyl-tRNA Synthetase

Studies discussed in this dissertation all relate to the *E. coli* methionyl-tRNA synthetase (MetRS) and the identification of variants of this enzyme that can activate azidonorleucine. This section presents a summary of work done on aminoacyl-tRNA synthetases, prior to the MetRS study. A description of each system and the results of each study are reported briefly.

Design of the *E. coli* isoleucyl-tRNA synthetase for the incorporation of hexafluorovaline into proteins *in vivo*

- *Hexafluorovaline (Hfv):* Hfv is a highly fluorinated amino acid, which can be used in the modulation of hydrophobicity of proteins, and stabilization of protein cores.
- *Isoleucyl-tRNA synthetase (IleRS):* Because the crystal structure of the *E. coli* IleRS is not available, the *Thermus thermophilus* IleRS structure at 2.8 Å resolution was used [1]. The *T. thermophilus* enzyme shares the same binding site residues with the *E. coli* IleRS. (Figure C.1)
- *Design:* The Clash-Opportunity Progressive (COP) design procedure [2] was carried out on three positions to generate a series of one-, two-, and three-fold mutants. Three mutants which show improved VdW interactions with the ligand were chosen for experimental characterization (A, AF and F; Table C.1.a). The F518 was chosen based on the analysis of the IleRS structure.
- *Results:* Hfv incorporation was not observed *in vivo* for any of the mutants, and its activation was not detected *in vitro* for A, F and AF. Isoleucine was activated *in vitro* by A and AF, but at lower rates than the activation by the wild-type IleRS.

Design of the *E. coli* valyl-tRNA synthetase for the incorporation of hexafluorovaline into proteins *in vivo*

- *Valyl-tRNA synthetase (ValRS):* Because the crystal structure of the *E. coli* ValRS is not available, the *Thermus thermophilus* ValRS structure at 2.9 Å resolution was used [3]. As it is for IleRS, the *T. thermophilus* ValRS also shows extensive similarities to the *E. coli* enzyme, sharing the same binding site residues.
- *Design:* COP was used on positions P41 and I491 and eight double mutants were chosen for experimental characterization. In addition to these, eight additional single ValRS mutants were chosen based on their differential interaction energies between Val and Hfv. (Table C.1.b) Expression profiles for clones carrying these mutants and the MALDI-MS for the expressed proteins were obtained.
- *Results:* None of the 16 ValRS mutants constructed showed activity toward Hfv, although the incorporation of isoleucine or leucine was commonly observed. This indicates the importance of especially the residues P42 and I491 in discriminating between Ile and Val.

Designing the *M. jannaschii* tyrosyl-tRNA synthetase binding site for tyrosine analogs

- *Ligands:* Following the selection experiments of the laboratory of Peter Schultz on the *in vivo* incorporation of *O*-methyl-tyrosine through a mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*-TyrRS) [4], Zhang et al. showed that it was possible to correctly identify the mutations in the immediate vicinity of the ligand, at positions Y32 and D158, through computational means using the COP procedure [2]. We continued these efforts to verify the success of the COP strategy using other ligands activated by *Mj*-TyrRS mutants identified by the Schultz laboratory: *p*-acetyl-phenylalanine (Acf) [5], *p*-azido-phenylalanine (Azf) [6], *p*-amino-phenylalanine (Amf) [7], 2-naphthylalanine (Nal) [8]. These ligands were selected for the low degrees of freedom they have in their side chain

coordinates. (Figure C.2)

- Mj-*TyrRS*: A 1.95 Å resolution crystal structure of this enzyme complexed with its native ligand, tyrosine, was used in this study [9]. The structure of the mutant *Mj*-TyrRS specific for *O*-methyl-tyrosine [10] was not available at the time of this study.
- *Design:* The *Mj*-TyrRS binding site was designed over the two positions (Y32 and D158) identified as clashing with the selected ligands. The positions Y32 and D158 are in the 5 positions randomized in the selection experiments. (Figure C.2)
- *Results:* Almost all *Mj*-TyrRS mutants identified by the Schultz laboratory contain four or more mutations. For each mutant-analog pair we introduced the full set of mutations into *Mj*-TyrRS, and determined the binding energy of each mutant with the analog, and with tyrosine as a competitor. We verified that the full mutants show good discrimination for their respective analogs and against tyrosine.

  For Acf, the two double mutants containing the experimental mutations at the positions 32 and 158 mutations at positions were ranked 3rd and 12th in the design. A comparison of the model of the LW1 mutant (Y32L-D158G-I159C-L162R) [5] with the Y32L-D158G mutant retrieved from COP can be seen in Figure C.3.a.

  For the remaining ligands, it was not possible to retrieve the experimental mutations via COP. Comparing the COP results with models containing the full set of experimental mutations showed that desired sequences cannot be selected at positions 32 and 158 in the absence of more distant secondary mutations, outside the scope of COP. One example of this is the Amf case. Santoro et al. showed that Amf is activated by a *Mj*-TyrRS variant carrying the mutations Y32T-E107T-D158P-I159L-L162A [7]. COP cannot identify the Y32T-D158P mutant for this ligand in the absence of the L162A mutation. The L162A mutation creates space to allow the movement of the backbone around position 158, which in turn allows proline to be placed at that position without any clashes with the ligand. (Figure C.3.b)

A similar observation is made in the designs for Nal. The *Mj*-TyrRS mutant Y32L-D158P-I159A-L162Q-A167V can activate Nal [8]. COP does not consider the Y32L mutation since introducing this mutation in the absence of the L162Q mutation results in L32 to clash with the ligand. The conformation of the L32 residue in the presence and absence of the L162Q mutation can be seen in Figure C.3.c.

- This study suggest a strong coupling between the L162 position and the positions Y32 and D158 in *Mj*-TyrRS. The identity of the residues at such secondary sites may have strong influences determining what is selected at the immediate vicinity of the ligand. Because this behavior is context dependent, COP may or may not agree with the selection experiments, depending on the system considered.

Designing the *E. coli* phenylalanyl-tRNA synthetase for the incorporation of *p*-acetyl phenylalanine into proteins *in vivo*

- p-*Acetyl phenylalanine (Acf)*: Using the computational design program ORBIT, Datta et al. showed that the T251G-A294G double mutant of the *E. coli* phenylalanyl-tRNA synthetase (*Ec*-PheRS) is able to activate Acf [11]. In this study we tested whether if we can capture these mutations using COP. Unlike the *Mj*-TyrRS designs, there are no secondary mutations in this system.
- Ec-*PheRS:* Because a crystal structure of the *E. coli* PheRS is not available, a homology model of the *E. coli* PheRS was constructed [12] using the available structures from *Thermus thermophilus* [13, 14].
- *Design:* COP correctly identified the two positions, T251 and A294, clashing with the ligand. These mutations were designed for the recognition of Acf.
- *Results:* We were able to correctly identify the T251G-A294G mutations in *Ec*-PheRS required to activate Acf from all possible double mutants at these positions.

Designing the *S. cerevisiae* phenylalanyl-tRNA synthetase for the incorporation of cyano-phenylalanine into proteins *in vivo*

- para- *and* meta-*cyano phenylalanine (pCnf and mCnf)*: Cyano-phenylalanine is a probe for the internal electric field of proteins for vibrational Stark spectroscopy. (Figure C.4) Although this analog shows about 10-15% incorporation through the yeast PheRS variants, more than 50% incorporation is desired for spectroscopy experiments.
- S. cerevisiae *phenylalanyl-tRNA synthetase (*Sc*-PheRS):* Because a crystal structure of *Sc*-PheRS is not available, a homology model of this enzyme was constructed [12] using the available structures from *Thermus thermophilus* [13, 14]. (Figure C.4.)
- *Design:* No mutation sites could be identified for pCnf via COP. However, for mCnf three mutants were proposed (mutants 1-3; Table C.1.c.). Because all these mutants contain a glycine substitution at position 412, we also tested mutants that have alanine at this site (mutants 4-6).

  With no success from mutations 1-6, we expanded the sites considered for mutation, by including position S418 in this set. The combinatorial set of all triple mutants at the N412, T415 and S418 positions were evaluated for their interactions with mCnf. From this study four 3-fold and one 2-fold mutant was selected for experimental testing (mutants 7-11).
- *Results: In vivo* incorporation of mCnf was not observed for any of the mutants. We concluded that our homology model of *Sc*-PheRS may not be sufficiently accurate for a design study.

Introduction of *E. coli*–like mutations to the *S. cerevisiae* PheRS to achieve sufficient *in vivo* incorporation of 2-quinoxalinylalanine (Qxa)

- *2-quinoxalinylalanine (Qxa)*: Structurally similar to Nal, Qxa is a reporter of slow time-scale solvation of charges on proteins it is incorporated into.
- *Background:* Experiments [15] suggest a nitrogen at the ortho position of phenylalanine is not accepted well by the *Sc*-PheRS, but is better permitted by the *Ec*-PheRS:

- 2-Pyridylalanine is accepted by the *Ec*-PheRS [16], but not the *Sc*-PheRS.
- Nal is accepted both by mutant *S. cerevisiae* (T415G) and *E.coli* PheRS (T251G-A294G).
- 2-Quinolinylalanine (Qla) and Qxa are not accepted by *Sc*-PheRS (T415G).

- Four positions in the *Sc*-PheRS binding site (336, 412, 414, 436) that differ from the corresponding residues in the *Ec*-PheRS were identified. To see if these changes have a role in the differences observed between the two enzymes, residues in *Sc*-PheRS were mutated to their *E. coli* identities The T415G mutation was included in all mutants constructed since this mutant creates the space required to bind the napthyl ring.
- Qla, which is commercially available, was used instead of Qxa when evaluating the mutants. Seven *Sc*-PheRS variants containing *E. coli*–like mutations were tested with Qla. The E376Q mutation was also tested to see if it can help decrease tryptophan misincorporation. (Table C.1.c)
- *Results:* Low levels of protein expression was observed with most mutants. Mutants with higher expression levels, which provided good MALDI-MS signals, showed equal or lower levels of Qla incorporation as the T415G mutant. Higher expression rates are needed to establish the behavior of other mutants.

Table C.1

List of aminoacyl-tRNA synthetase mutants tested in design studies in this section

a) IleRS mutants experimentally tested with Hfv

IleRS	A	AF	F	F518
P 46	A	A	P	P
W 518	W	W	W	F
W 558	W	F	F	W

b) ValRS mutants identified by COP and experimentally tested with Hfv

ValRS	SD	SV	AV	GV	CV	AF	AN	AD
P 42	S	S	A	G	C	A	A	A
I 491	D	V	V	V	V	F	N	D

ValRS single-mutants experimentally tested with Hfv

ValRS	1	2	3	4	5	6	7	8
P 42	G	A	C	P	P	P	P	P
W 456	W	W	W	F	W	W	W	W
I 491	I	I	I	I	V	I	I	I
W 495	W	W	W	W	W	F	Y	H

Table C.1 (continued)

c) *Sc*-PheRS mutants identified by COP and experimentally tested with mCnf

PheRS	1	2	3	4	5	6	7	8	9	10	11
N 412	G	G	G	A	A	A	G	G	G	G	G
T 415	T	A	G	T	A	G	A	T	C	Q	M
S 418	S	S	S	S	S	S	A	A	A	A	A

E. coli–like mutations introduced into the *Sc*-PheRS in this study

PheRS	336	412	414	436	2ble	3ple	4ple	E376Q
S 336	Q	S	S	S	S	S	Q	S
N 412	N	F	N	N	N	F	F	N
Y 414	Y	Y	F	Y	F	F	F	Y
T 415	G	G	G	G	G	G	G	G
N 436	N	N	N	G	G	G	G	N
E 376	E	E	E	E	E	E	E	Q

Figure C.1

Comparison of the T. thermophilus *IleRS and ValRS binding sites.*

A comparison of the binding sites for ValRS (yellow) and IleRS (purple) from *T. thermophilus* is shown to highlight the differences and similarities between the two enzymes. The ligands used in the studies of these enzymes are shown on the right-hand side.

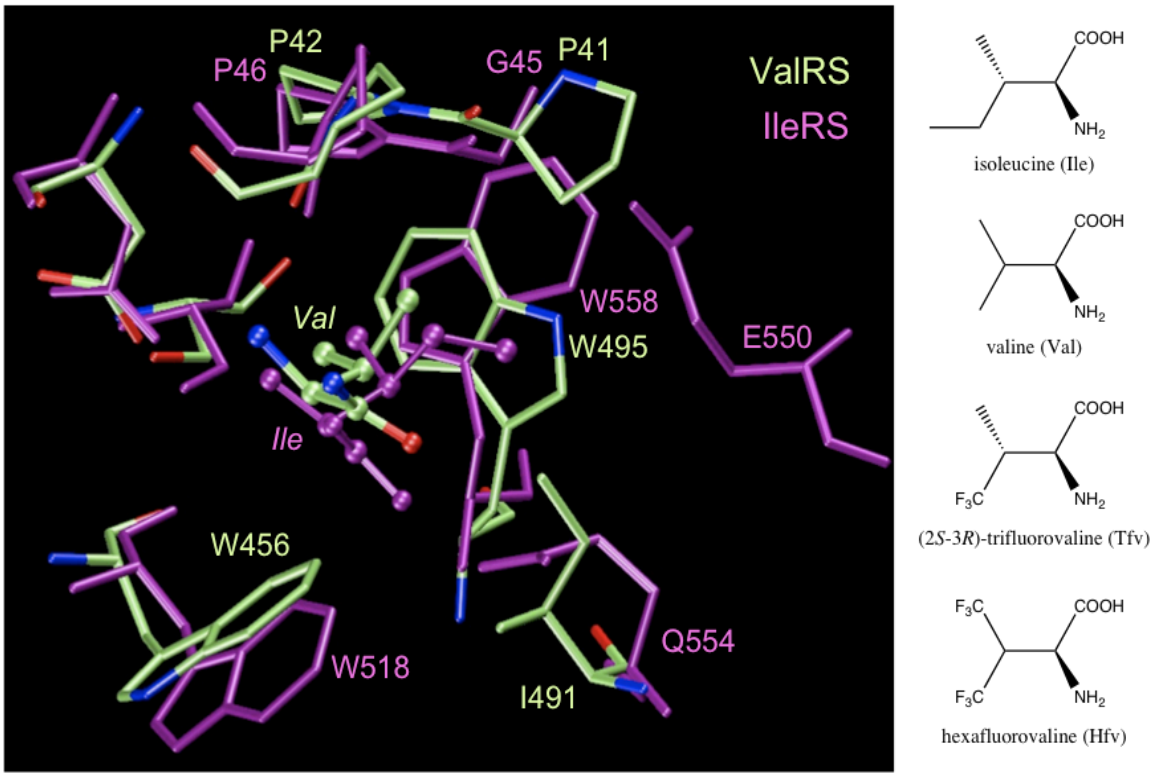
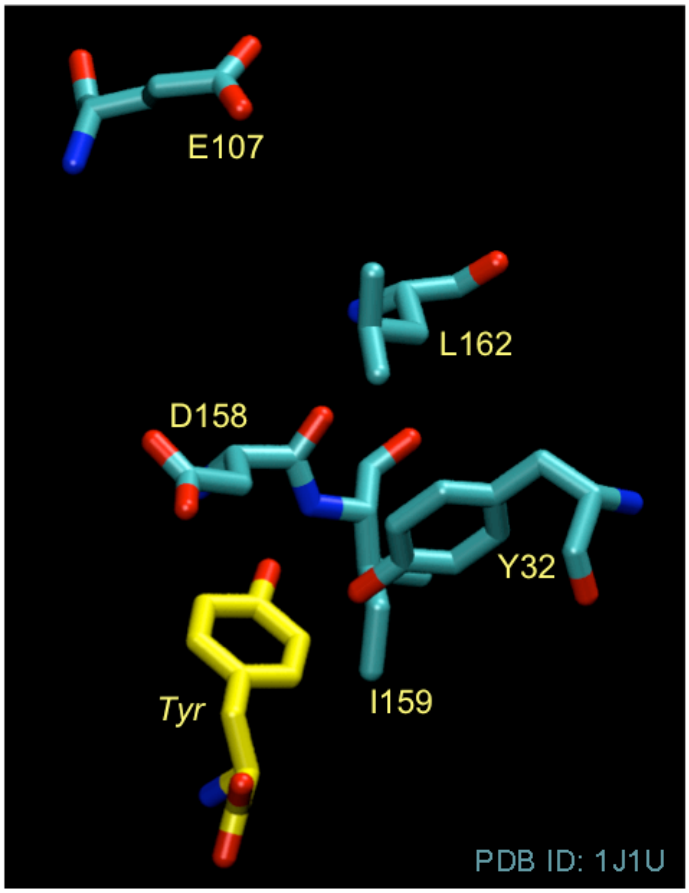


Figure C.2

Randomized positions in the Mj-*TyrRS binding site and the ligands studied.* Positions randomized in the *Mj*-TyrRS saturation mutagenesis library used in the selection experiments performed by the Schultz laboratory (cyan) are shown in the presence of the native ligand, tyrosine (yellow). The ligands discussed in this study are presented on the right-hand side.



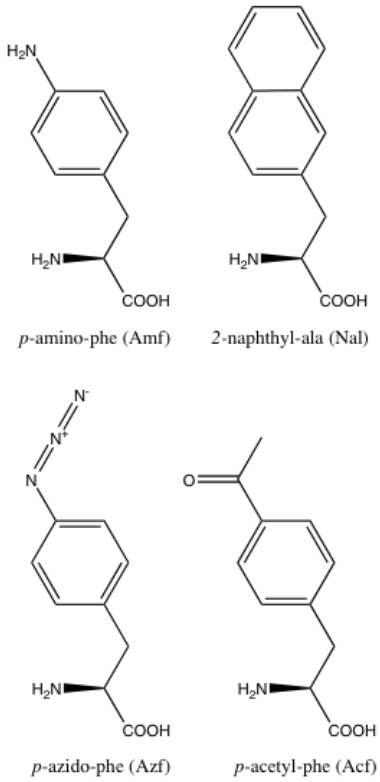


Figure C.3

Comparison of Mj-*TyrRS models bearing mutations identified by experiments with double mutants received from the COP design procedure.*

Mj-TyrRS models containing the full set of experimental mutations (cyan) were shown to prefer the non-natural ligand over the tyrosine in the binding-energy analyses performed. When residues identified as clashing with the ligand by the COP design procedure (Y32 and D158) were mutated to their experimental identities, the purple models were obtained. These two models compare favorably when mutants for Acf are analyzed. However, the double mutants for Nal and Amf show significant differences in their binding characteristics from the experimental mutants identified for these ligands.

a) Comparison of the Acf-bound models of the L32-G158 mutant from COP and the four-fold mutant LW1.

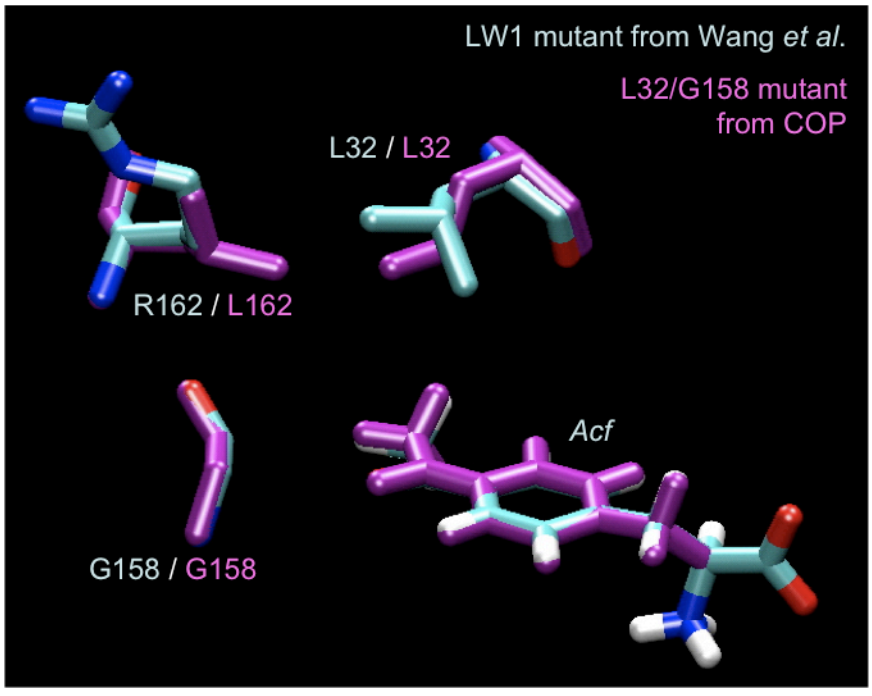
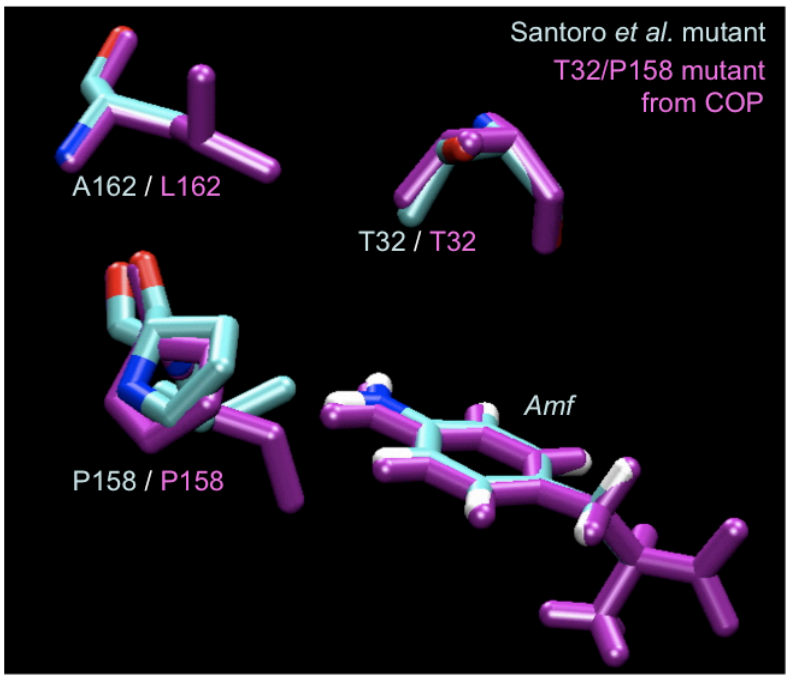


Figure C.3 (continued)

b) Comparison of the Amf-bound models of the T32-P158 mutant from COP and the five-fold mutant from Santoro et al.



c) Comparison of the Nal-bound models of the L32-P158 mutant from COP and the five-fold mutant from Wang et al.

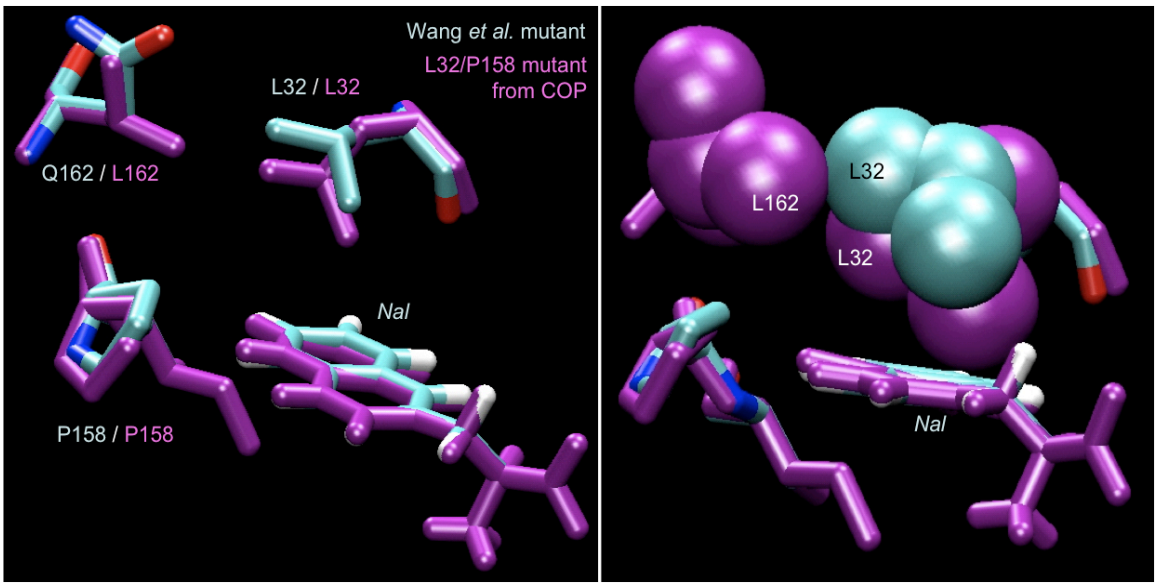
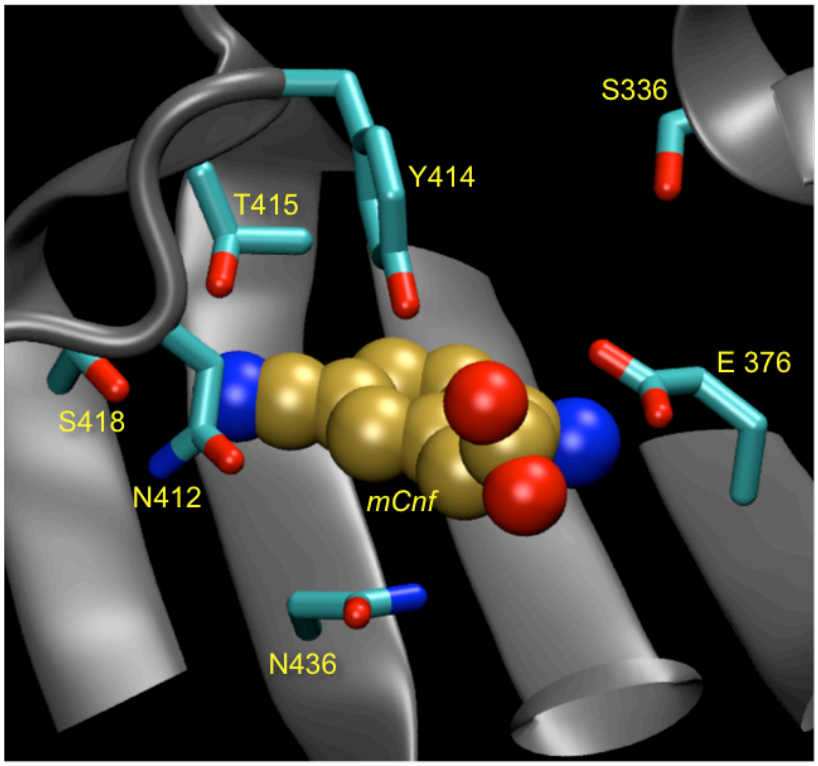
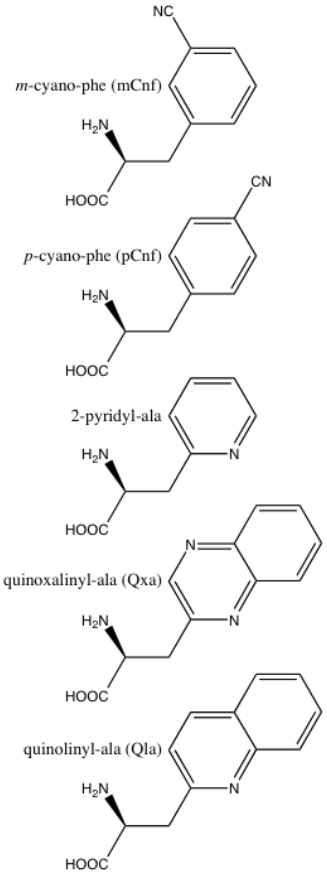


Figure C.4

The homology model of the S. cerevisiae *PheRS.*

The homology model of the *S. cerevisiae* PheRS is shown complexed with the ligand mCnf. Residues interacting with the ligand are shown in cyan. The ligands discussed in this study are shown on the right-hand side.





References

[1] Nureki O, Vassylyev DG, Tateno M, Shimada A, Nakama T, Fukai S, Konno M, Hendrickson TL, Schimmel P and Yokoyama S, "Enzyme structure with two catalytic sites for double-sieve selection of substrate," *Science*. **1998**, *280*:578.

[2] Zhang DQ, Vaidehi N, Goddard WA, Danzer JF and Debe D, "Structure-based design of mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase for incorporation of *O*-methyl-l-tyrosine," *Proc Natl Acad Sci U S A*. **2002**, *99*:6579.

[3] Fukai S, Nureki O, Sekine S, Shimada A, Tao JS, Vassylyev DG and Yokoyama S, "Structural basis for double-sieve discrimination of l-valine from l-isoleucine and l-threonine by the complex of tRNA(val) and valyl-tRNA synthetase," *Cell*. **2000**, *103*:793.

[4] Wang L, Brock A, Herberich B and Schultz PG, "Expanding the genetic code of *Escherichia coli*," *Science*. **2001**, *292*:498.

[5] Wang L, Zhang ZW, Brock A and Schultz PG, "Addition of the keto functional group to the genetic code of *Escherichia coli*," *Proc Natl Acad Sci U S A*. **2003**, *100*:56.

[6] Chin JW, Santoro SW, Martin AB, King DS, Wang L and Schultz PG, "Addition of p-azido-l-phenylaianine to the genetic code of *Escherichia coli*," *J Am Chem Soc*. **2002**, *124*:9026.

[7] Santoro SW, Wang L, Herberich B, King DS and Schultz PG, "An efficient system for the evolution of aminoacyl-tRNA synthetase specificity," *Nat Biotechnol*. **2002**, *20*:1044.

[8] Wang L, Brock A and Schultz PG, "Adding l-3-(2-naphthyl)alanine to the genetic code of e-coli," *J Am Chem Soc*. **2002**, *124*:1836.

[9] Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K and Yokoyama S, "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion," *Nat Struct Biol*. **2003**, *10*:425.

[10] Zhang Y, Wang L, Schultz PG and Wilson IA, "Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for *O*-methyl-l-tyrosine," *Protein Sci*. **2005**, *14*:1340.

[11] Datta D, Wang P, Carrico IS, Mayo SL and Tirrell DA, "A designed phenylalanyl-tRNA synthetase variant allows efficient *in vivo* incorporation of aryl ketone functionality into proteins," *J Am Chem Soc*. **2002**, *124*:5652.

[12] Kekenes-Huskey PM, Unpublished results, **2005**

[13] Fishman R, Ankilova V, Moor N and Safro M, "Structure at 2.6 Å resolution of phenylalanyl-tRNA synthetase complexed with phenylalanyl-adenylate in the presence of manganese," *Acta Crystallogr D Biol Crystallogr*. **2001**, *57*:1534.

[14] Reshetnikova L, Moor N, Lavrik O and Vassylyev DG, "Crystal structures of phenylalanyl-tRNA synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue," *J Mol Biol*. **1999**, *287*:555.

[15] Kwon I, Unpublished results, **2005**
[16] Kirshenbaum K, Carrico IS and Tirrell DA, "Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues," *Chembiochem*. **2002**, *3*:235.

Appendix D.

List of Software Written for and Used in the Modeling, Design, and Study of Methionyl-tRNA Synthetase Variants

A list of software written to facilitate the study and design of methionyl-tRNA synthetase (MetRS) mutants is presented here. The list contains software written in Perl and Python, distributed into multiple categories. A brief description is provided for each entry.

Table D.1

Software written for the study of MetRS mutants.

The programs listed here can be found on hulk.wag.caltech.edu, inside the directories "/ul/caglar/Python", "/ul/caglar/Perl". Most scripts will print usage information when executed without arguments.

a) List of utilities written to aid in the construction and experimental characterization of MetRS mutants

Name	Description
formatCodingSeq.pl	Reads in a DNA sequence, translates genes and reformats sequences, finds DNA/protein sequences, as well as ORFs, within the entered sequence
getMSrange.pl	Reads in a ABI Voyager mass spectrum data file as text, and saves selected data ranges, detects peaks, and plots spectra using Gnuplot
degenerateCodons.pl	Given a set of desired and/or undesired amino acids, searches for degenerate codons that code the amino acids specified, and lists any that are found
mutationMassShifts.pl	Provides suggestions for the identities for the incorporated amino acids in a residue-specific incorporation experiment to compare with mass spectrum data
primerReverse.pl	Prints the complementary stand and the reverse complement of the entered DNA sequence

Table D.1 (continued)

b) List of utilities for modifying three dimensional models of molecules in BGF or PDB formats

Name	Description
playWithBGF	Makes selections in a PDB or a BGF file and performs operations on the selection
addChainDsg2BGF.pl	Assigns the given chain designator to an atom range specified in a BGF file
addCharges2bgf	Copies charges between two similar BGF files, or allows you to enter them interactively
bgf2jagCoords.pl	Reads atom coordinates from a BGF file and reformats them to be pasted into a JAGUAR input file
copyCoords_bgf2bgf.pl	Copies the atom coordinates between two BGF files that have the same number of atoms
setBGF_DESCRP.pl	Adds or sets the DESCRP line in a BGF file
supplementBGFrecs.pl	Supplements missing fields in a BGF file based on a reference BGF, user-defined defaults or a conversion file
toggleMovableRecs.pl	Sets atoms in a BGF file movable or fixed based on a playWithBGF atom-list file
translateBGF.pl	Translates all coordinates in a BGF file by the given vector
atomicRuler.py	Measures distances, angles or dihedrals between atoms in a BGF file
checkForMissing-AtomsInBGF.py	Reads a BGF file and checks for the number of atoms in each residue against a pre-defined dictionary to see if any residues are missing or have extra atoms

Name	Description
listClashesWith-RotLib.py	Determines clashes between a binding site, and a SCREAM rotamer library generated by MOLECULEGL
mergeBGFs.py	Merges BGF files into a single file, in the order they are entered
reorderBackbone-Atoms.py	Reorders atoms in each residue of a BGF file so that backbone atoms are followed by side chain atoms
resListFromBGF.py	Reads a BGF file and writes a playWithBGF residue-list file (for use with option -R)
sumCharge.py	Sums charges and reports them by residue, chain, or structure
zeroCharge.py	Sets all charges in a BGF file to zero

Table D.1 (continued)

c) List of programs that combine multiple types of functionalities (usually through wrappers, or other scripts) in order to achieve a certain calculation, e.g. a binding-energy calculation

Name	Description
getVerticalBE.pl	Calculates vertical binding energies given a BGF file with a ligand marked by its chain identifier.
determineSASAon-Ligand.pl	Calculates solvent accessible surface area using the MPSIM final energy output files for the selected atoms
seqMin.pl	Runs sequential MPSIM minimizations setting different sets of atoms movable at each stage.
compileLibFromBGF.py	Prepares a SCREAM-compatible rotamer library file for a given set of BGF structures of a certain molecule, and lists MPSIM energies for each rotamer
getEnerDistPerAA.py	Tabulates energy contribution of selected residues to the total energy using MPSIM energy lists by atom

Table D.1 (continued)

d) List of programs that parse and act on the output of other programs

Name	Description
diffCavan.pl	Calculates the difference of interaction energies reported by cavity_analysis.pl between two files
diffFinEner.pl	Given two MPSIM final energy lists, calculates the energy differences between the two files
diffPWB_iaTables.pl	Given two playWithBGF output files containing nonbond interaction analysis tables based on the same selection on two different structures, this script calculates the difference between the two outputs
finEnerBE.pl	Given MPSIM final energy lists for the enzyme, ligand and complex this script calculates the binding-energy contribution for each atom
get-tvnE.pl	Calculates average energy from an MPSIM dynamics run output (adapted from David Zhang)
mergePWBatomLists.pl	Merges multiple atom-list files written by playWithBGF (option -l) into a single list
jaguarBindingE.py	Calculates and tabulates the binding energy of a complex (AB - A - B) from JAGUAR output files
jaguarOutputToBGF.py	Given a BGF file and a corresponding JAGUAR output file, writes a new BGF file containing the JAGUAR coordinates and Mulliken charges
tabulateMPSimEout.py	Tabulates the final energies from MPSIM output files
tabulateVBEout.py	Tabulates binding-energy components from getVerticalBE.pl output directories

Table D.1 (continued)

e) List of programs used in the binding site design calculations

Name	Description
cbMuts.py	Customizable program for generating combinatorial mutants, optimizing structures, and determining binding energies through the use of SCREAM and MPSIM. (Used in the 2005 STUDY)
evaluateLYHmutant.py evaluateMutant.py	For a specified mutation, runs MPSIM minimizations, interaction energy calculations with playWithBGF, and binding-energy calculations with DELPHI solvation. (Used in the 2008 STUDY)
foreachMutn.pl	Given the amino-acid pool at each mutation site, enumerates all combinatorial mutations and executes the given commands for each mutation.
compileMutationData.py compileMutationVBE.py	Each script reads the specified evaluateMutant.py data output files and tabulates the results from each calculation
shelveCompleted-Mutns.py	Given a calculation prefix, checks the for errors in each evaluateMutant.py output and for completed calculations, files the output file into appropriate directories

Table D.1 (continued)

f) List of wrappers, which provide easy access to various programs

Name	Description
runMPSim.pl	A customizable wrapper for MPSIM for running minimizations, simulated annealing, dynamics and energy evaluations.
runScream.py	A wrapper for SCREAM to perform side chain placement and mutations on BGF structures
convertTov400.pl	A wrapper for John Wendel's bgf2fsm, which converts a BGF file into the v400 format
generateLigand-RotLib.py	Uses MOLECULEGL (version 1) to generate a ligand rotamer library readable by SCREAM
run_delphi.py	A wrapper for DELPHI which calculates the contribution of PB solvation to the energy of a given BGF structure. (Written by John Wendel)

Table D.1 (continued)

g) List of Python modules utilized in Python scripts listed here

Name	Description
BGF.py	Module containing the BgfAtom and BgfFile classes for reading, writing and manipulating BGF files
crunners.py	Module containing adaptor functions or wrappers that run common Perl scripts or standalone programs that are employed for structure optimization, analysis or design
cutils_getopt.py	An extension on the Python's own getopt module to allow it to accept options of type +X
cutils.py	Module containing general utilities for data conversion, file input-output, program execution

Table D.1 (continued)

h) List of general utilities

Name	Description
addAvg.pl	Calculates the sums and averages of entered or numbers ignoring any non-numerical characters
changeFileNames.pl	Substitutes a given text for a certain text in the file names of the specified files
execInDirs.pl	Script to execute the same command in multiple directories, or recursively in all subdirectories
foreachFile.pl	Script to execute a shell command for each of the files specified in the command line, repetitively.
foreachWord.pl	Script to execute a shell command for each word entered in the command line, repetitively.
memlist.pl	Sorts and tabulates disk usage on the specified files and directories
r3D.pl	Calculates the distance between two points in space
runViaPBS.pl	Writes a PBS file for the entered command and submits it to the job queue
secs2time.pl	Translates seconds into common time
sortTable.pl	Interactive script to sort an arbitrary-size table containing numerical data
tarTextFiles.pl tarTextFilesInDirs.pl	Creates a tar archive containing all the text files found in the directories or files specified.
zipTextFiles.pl	Compresses any text files specified using gzip