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 x 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 Anl 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 Anl, 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 x 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

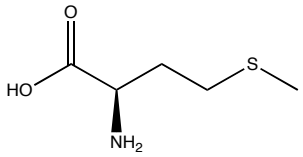
* Data on mutants of lower activity (AQL, SNL, and GVL) are listed in Table 3.5.

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

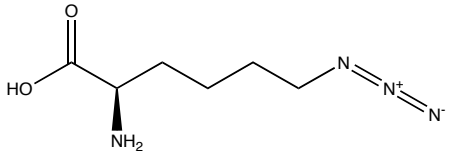
‡ Activation parameters for wild-type MetRS taken from reference [35].

Figure 2.1.

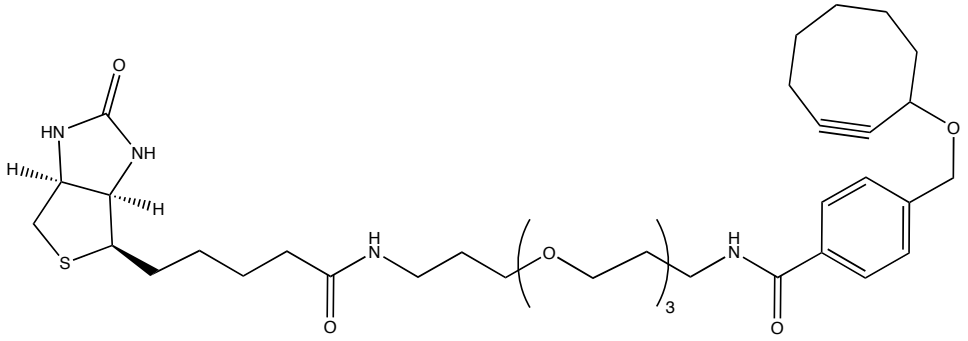
Chemical structures of amino acids and tagging reagents.



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.

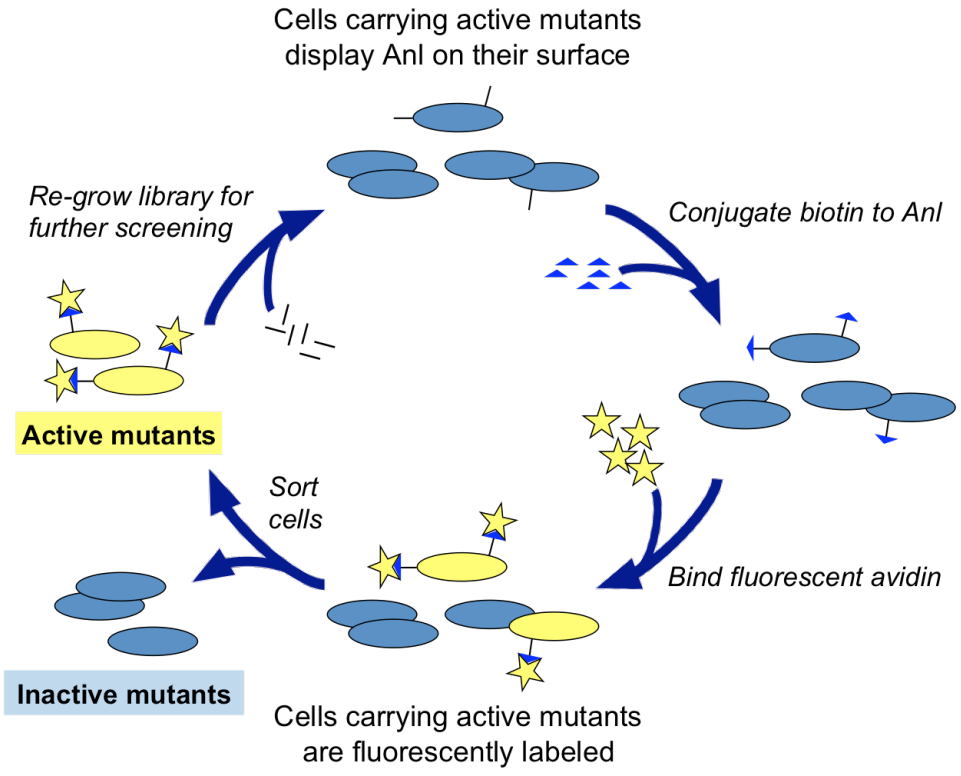


Figure 2.3.

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].

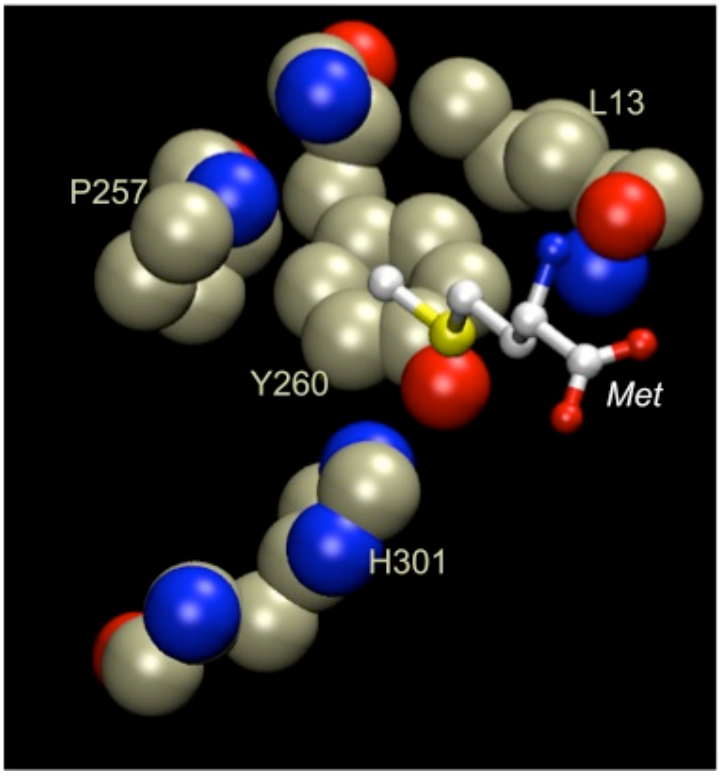


Figure 2.4.

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.

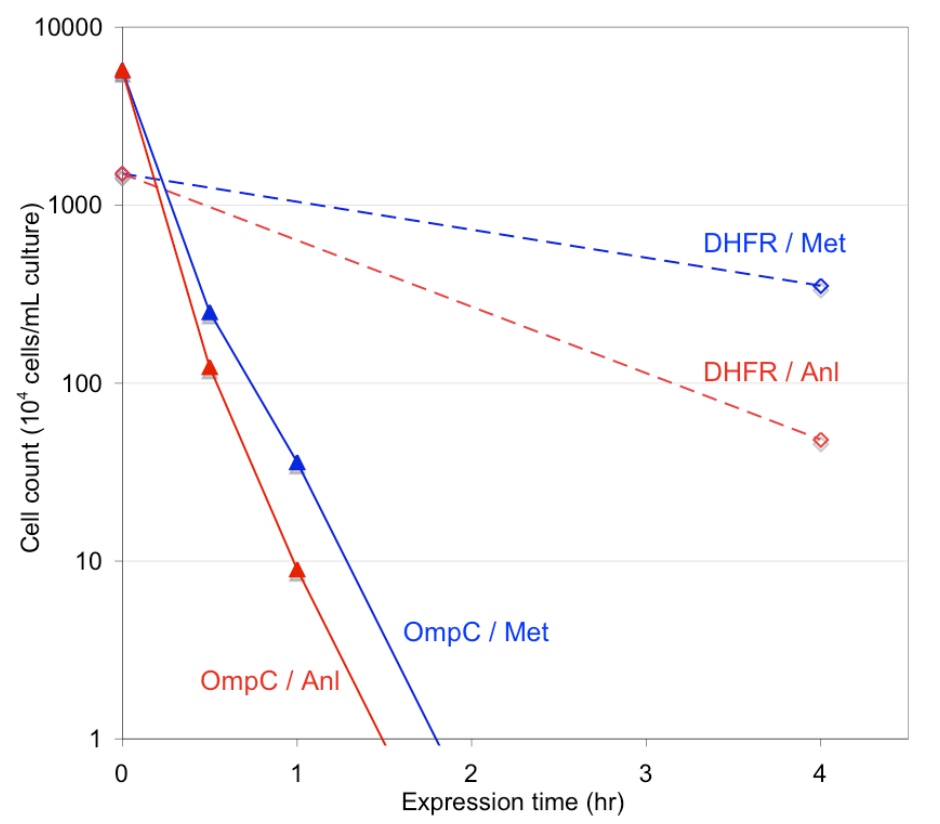


Figure 2.5.

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.

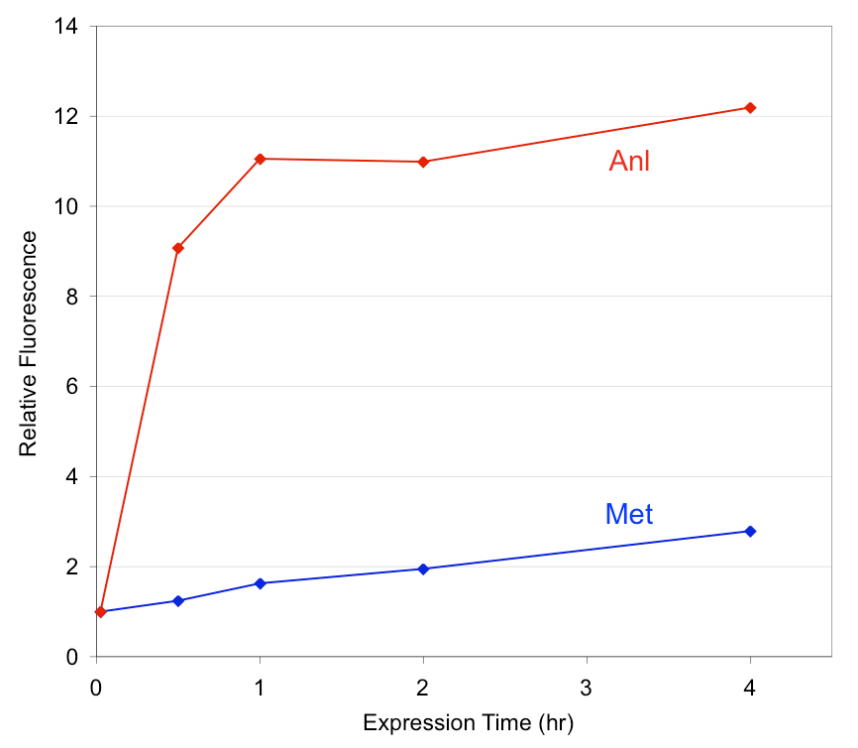


Figure 2.5. (continued)

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).

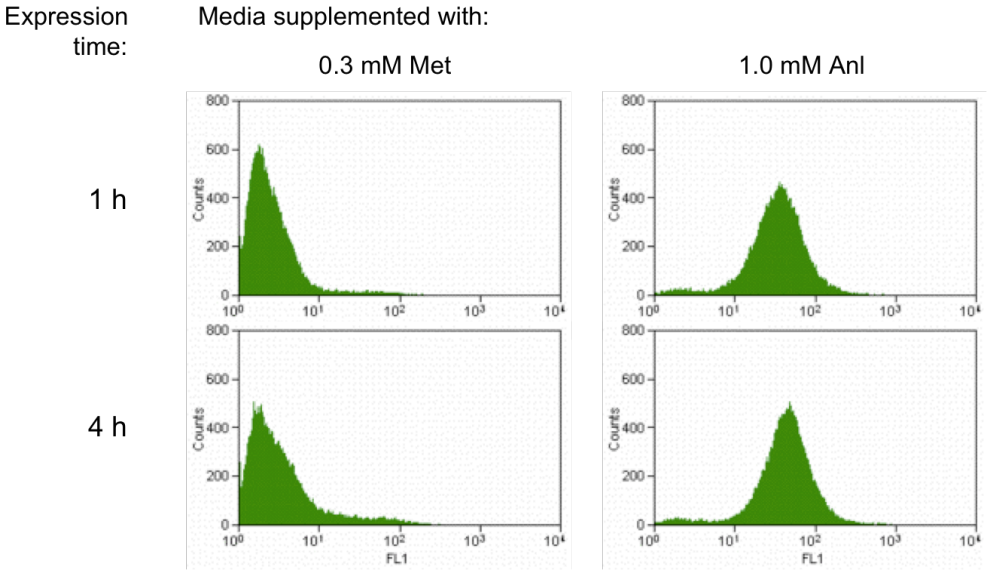


Figure 2.5. (continued)

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.

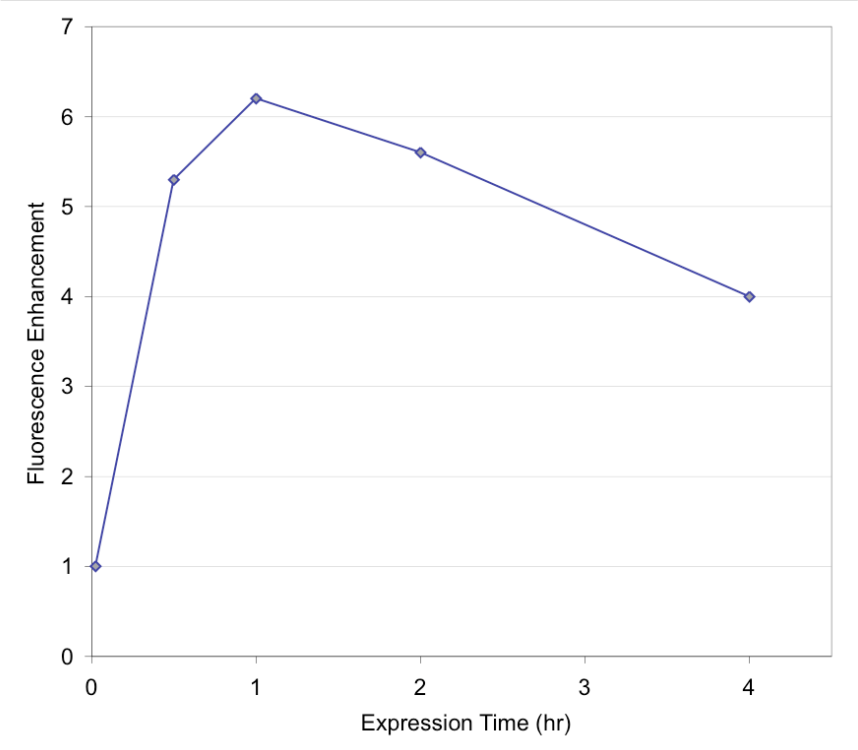


Figure 2.6.

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.

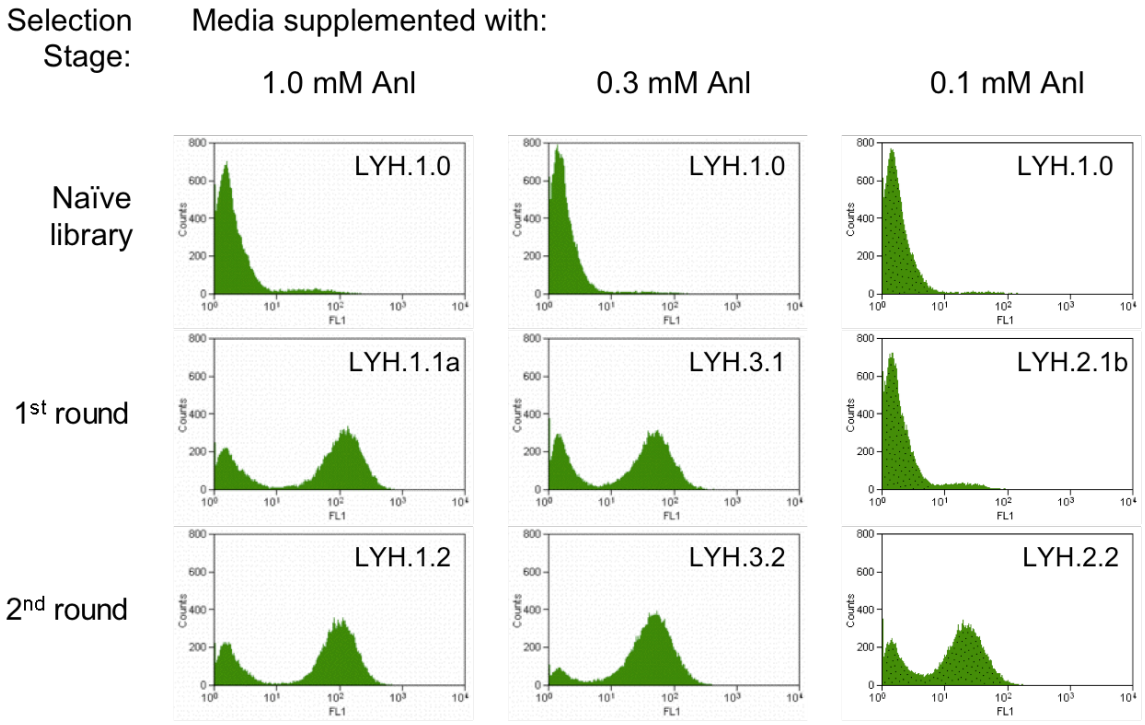


Figure 2.7.

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.

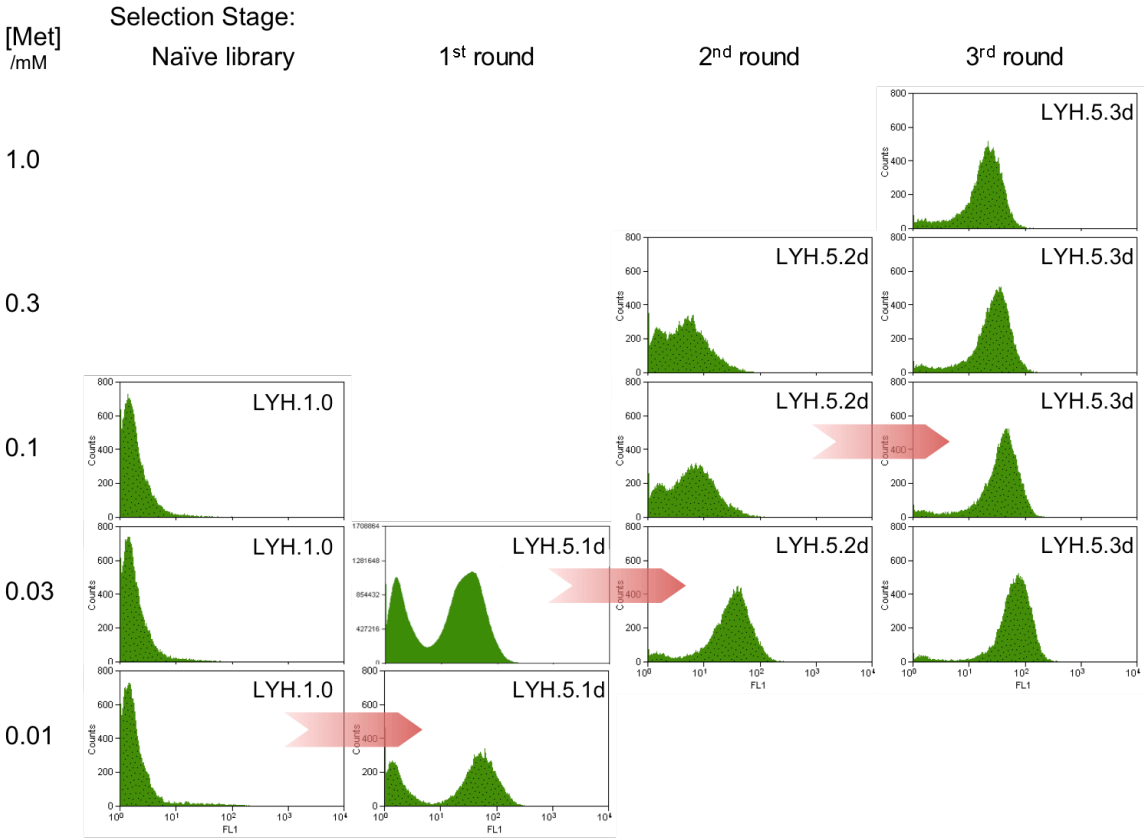


Figure 2.8.

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)

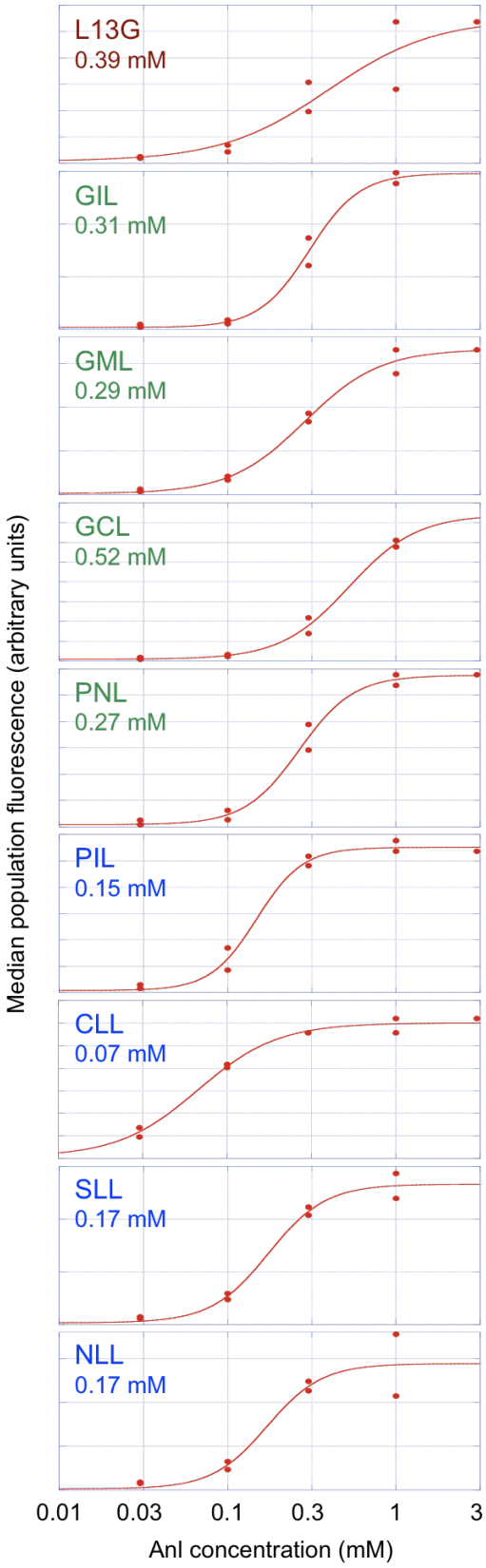


Figure 2.9.

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)

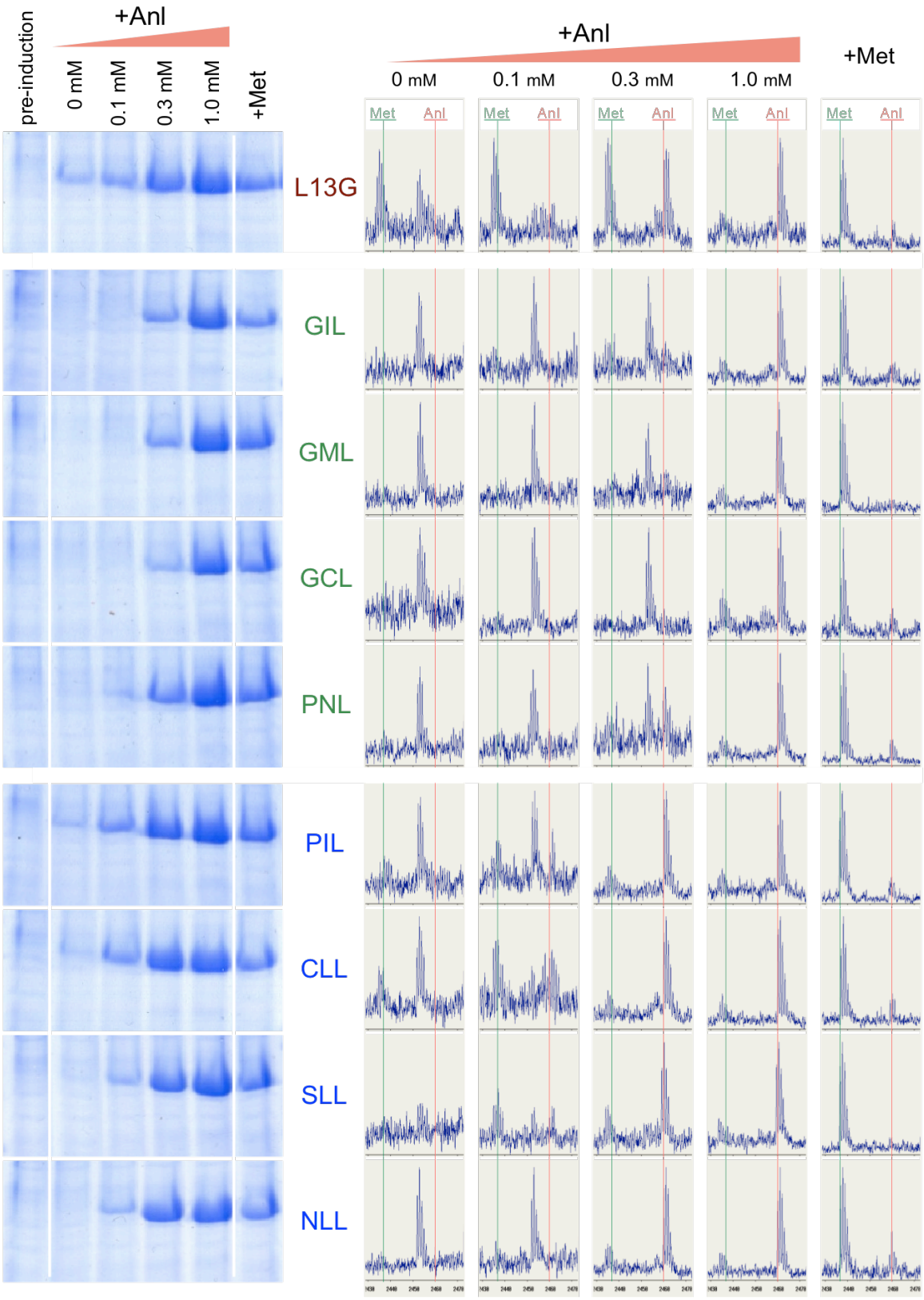


Figure 2.10.

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.

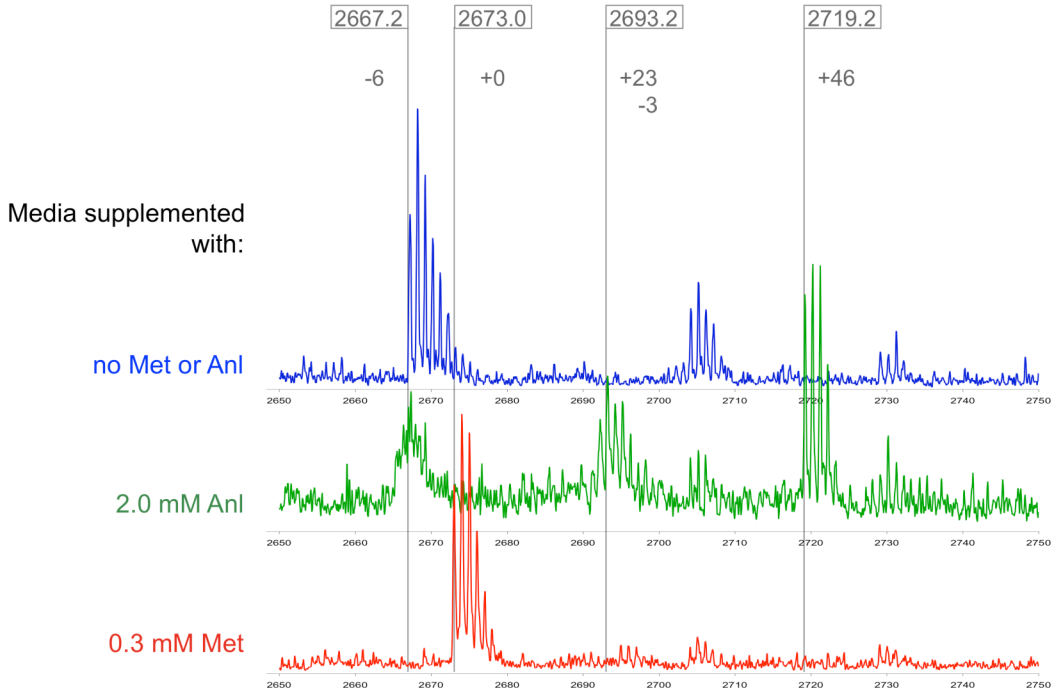


Figure 2.11.

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.

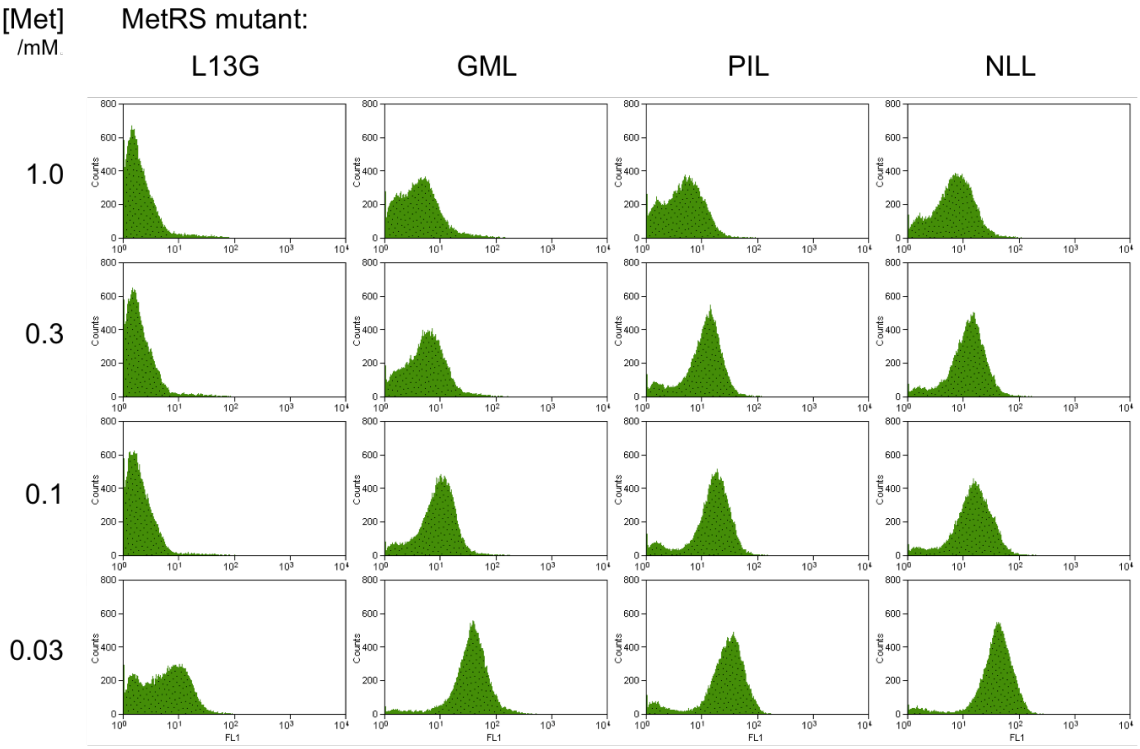


Figure 2.12.

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)

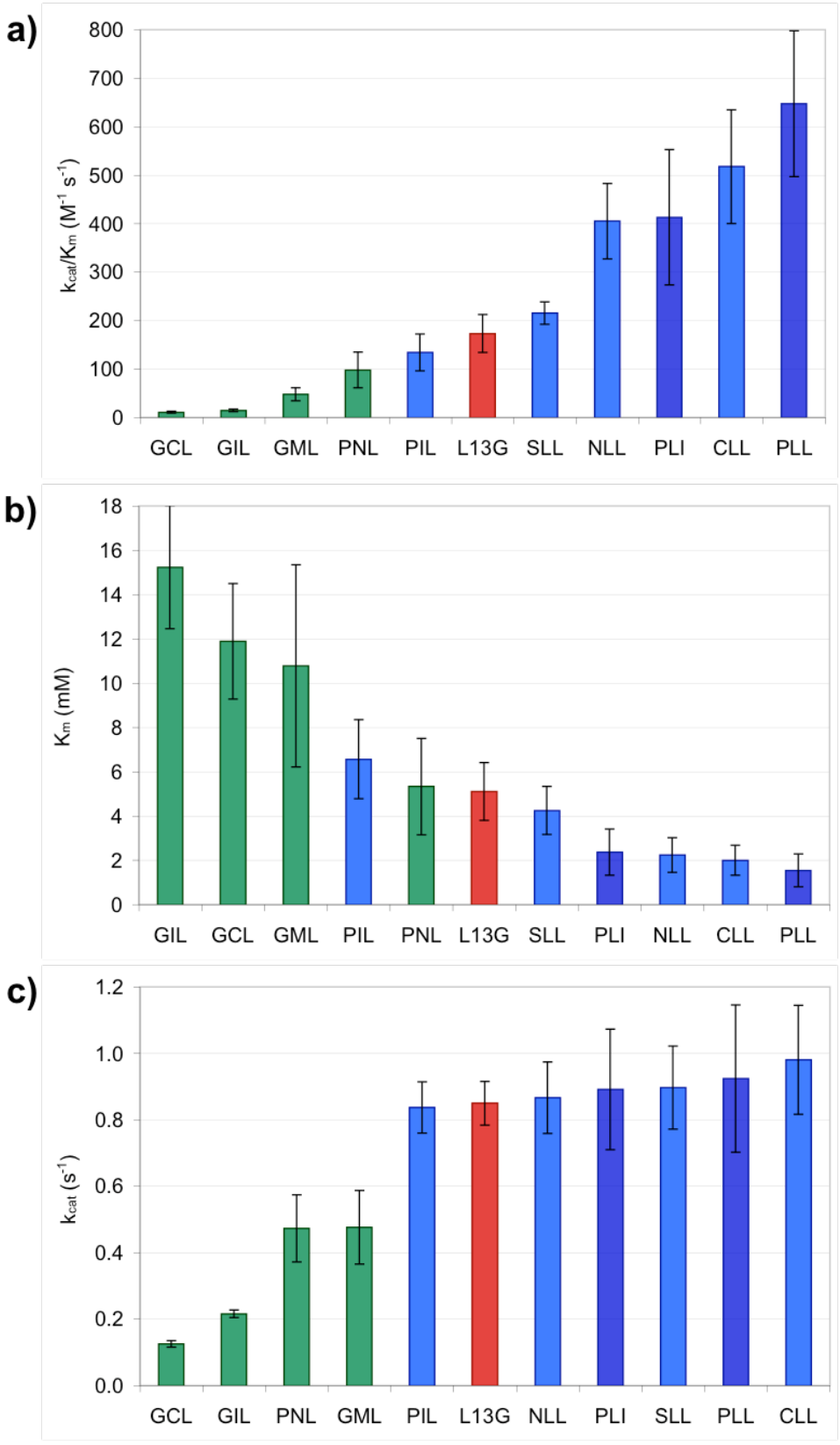


Figure 2.12. (continued)

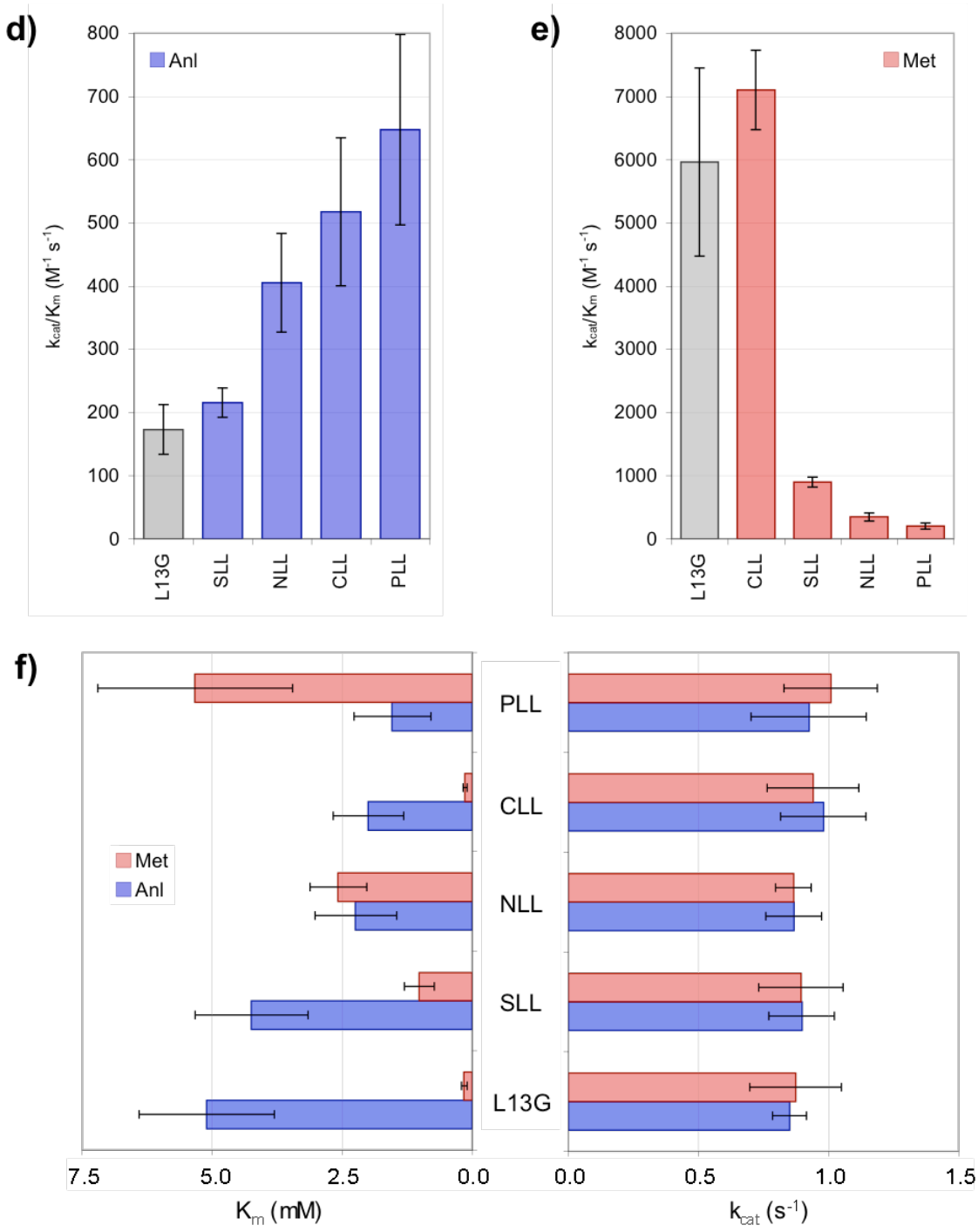


Figure 2.13.

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).

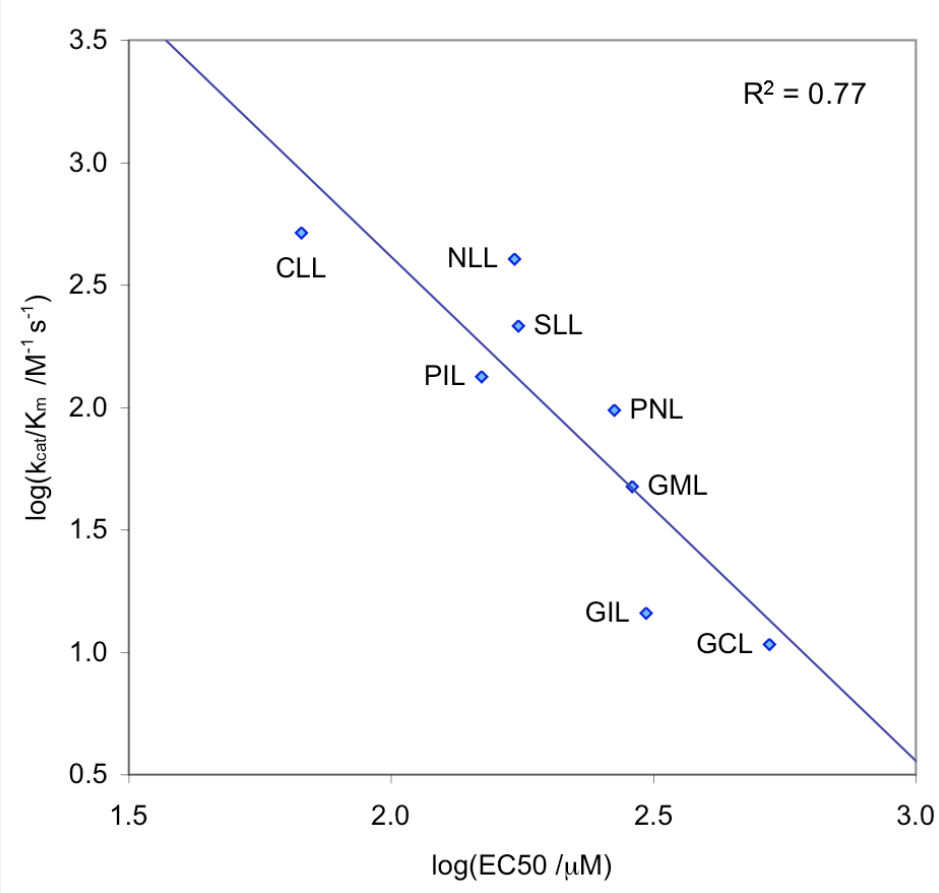


Figure 2.14.

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)

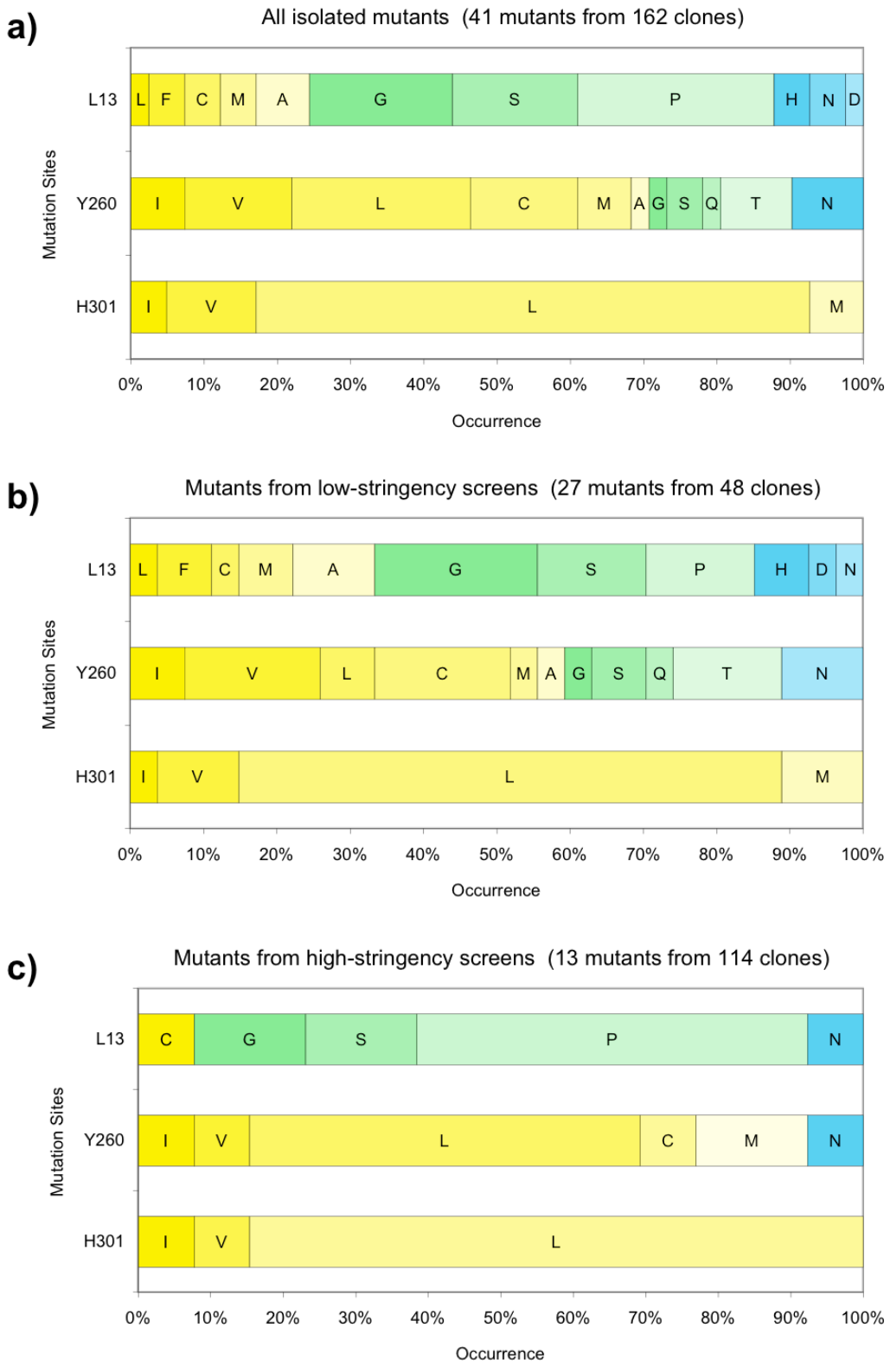


Figure 2.15.

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.

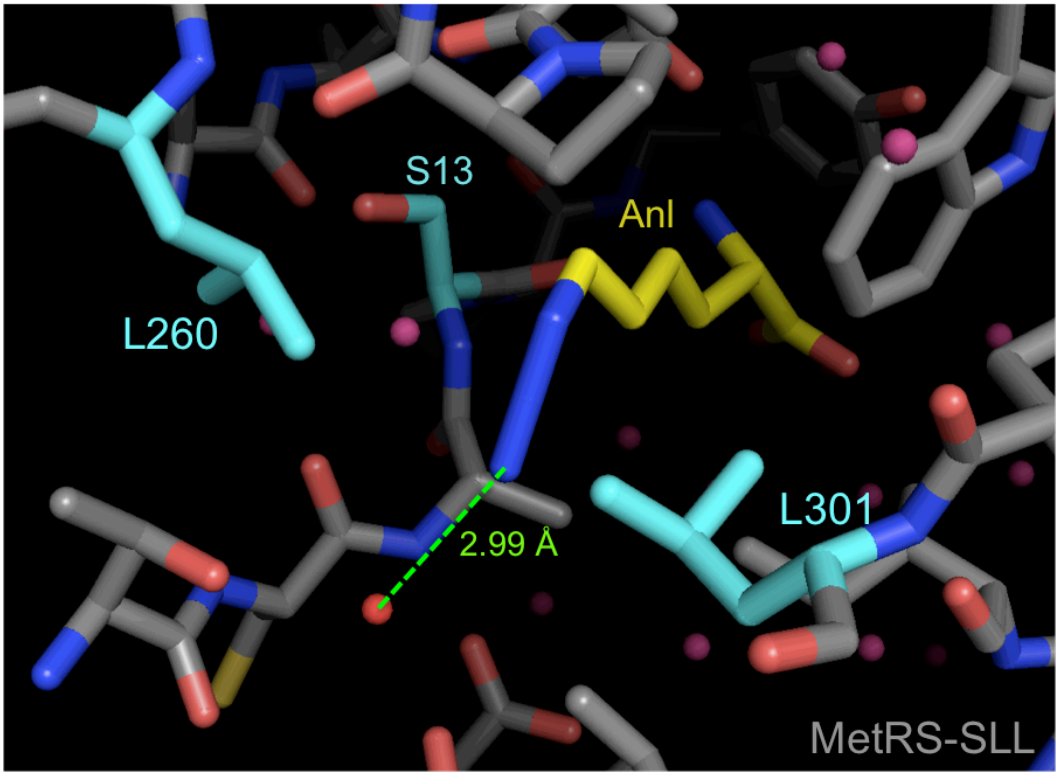


Figure 2.15. (continued)

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.

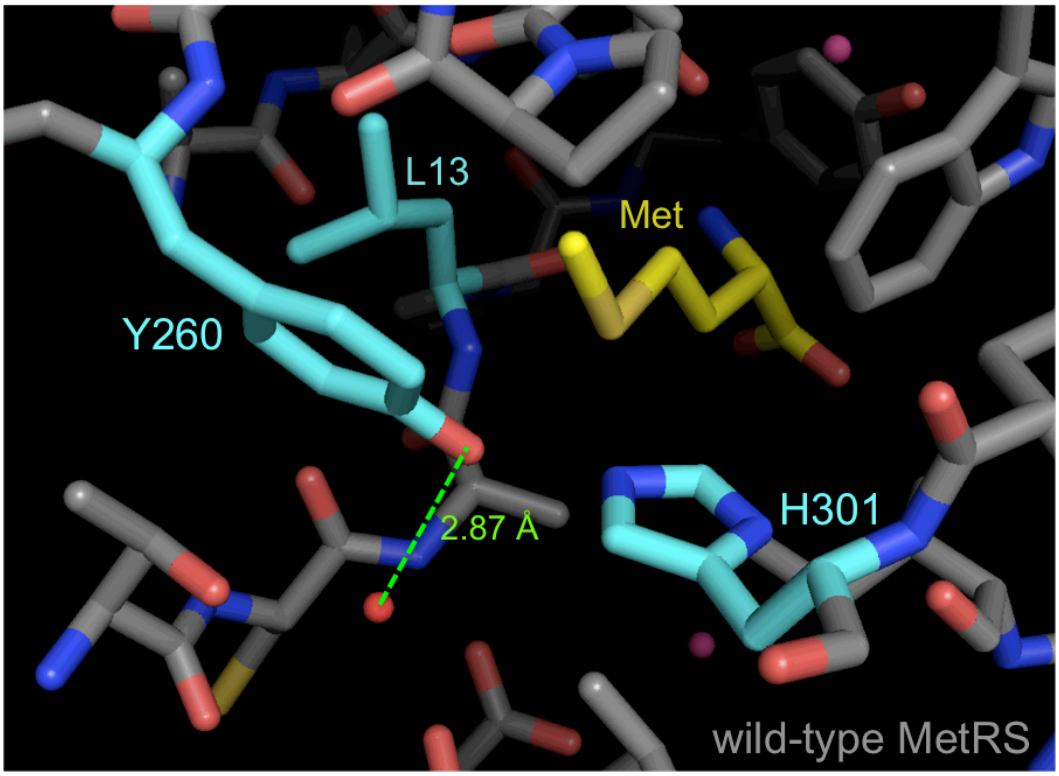


Figure 2.16.

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.

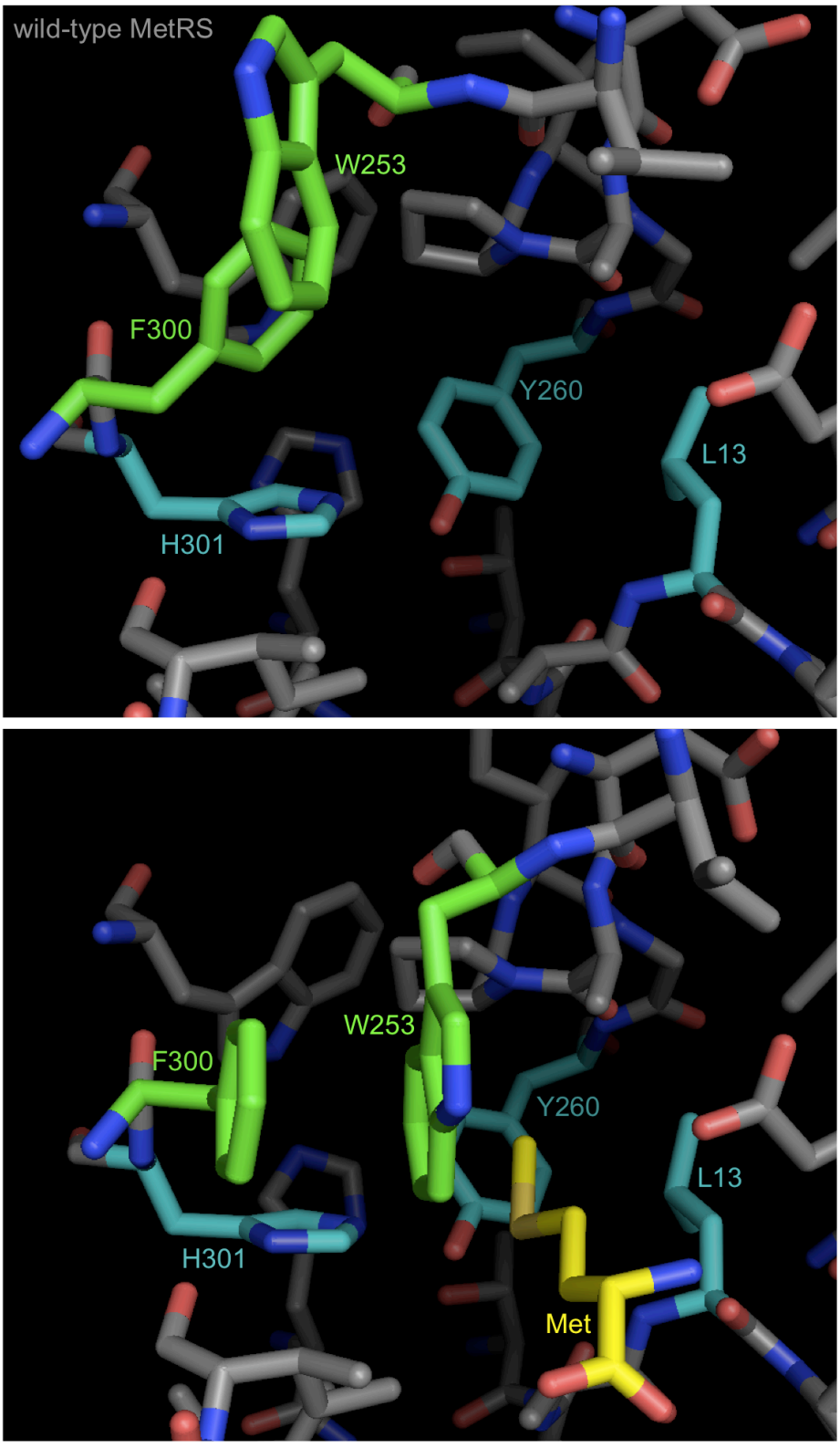


Figure 2.16. (continued)

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

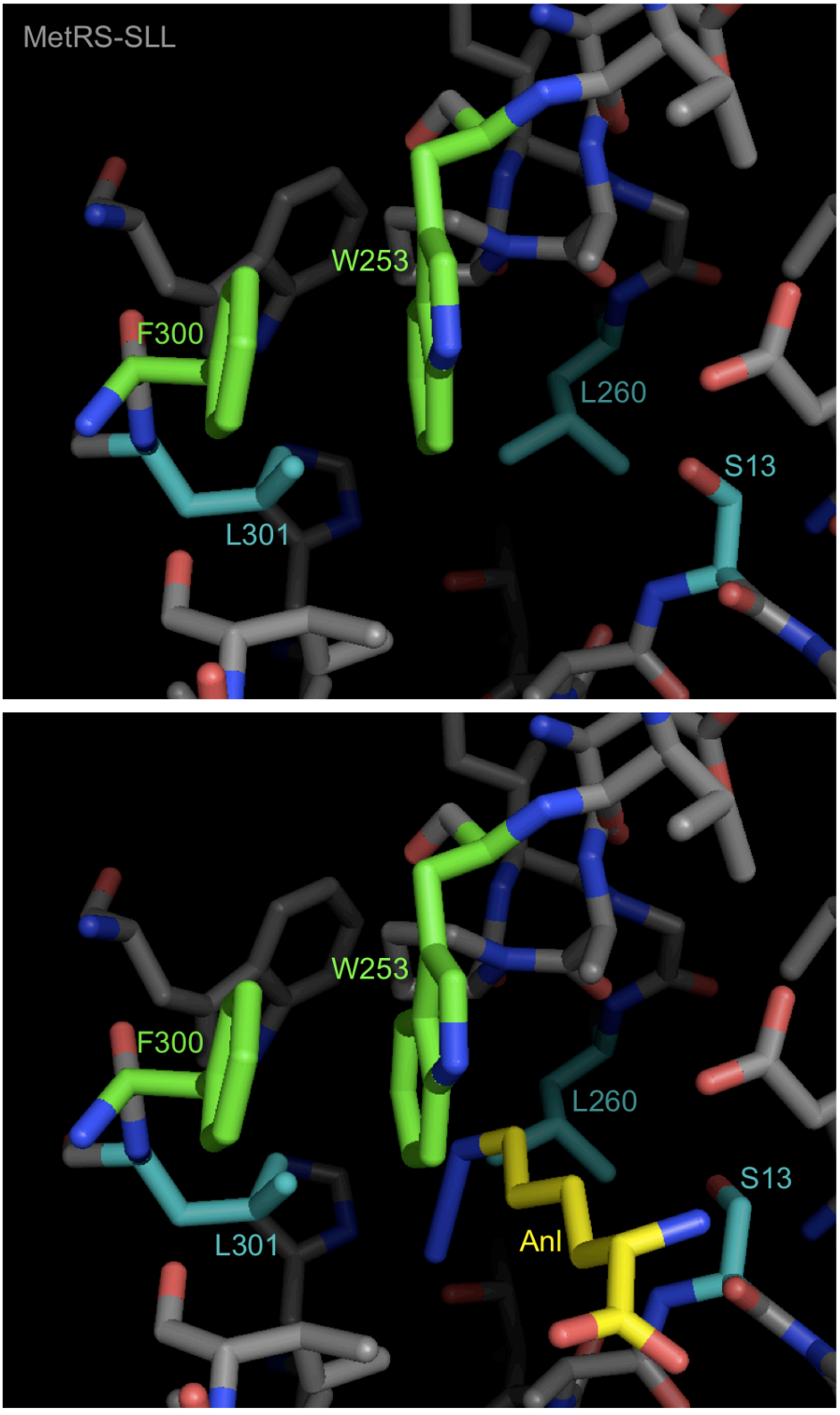
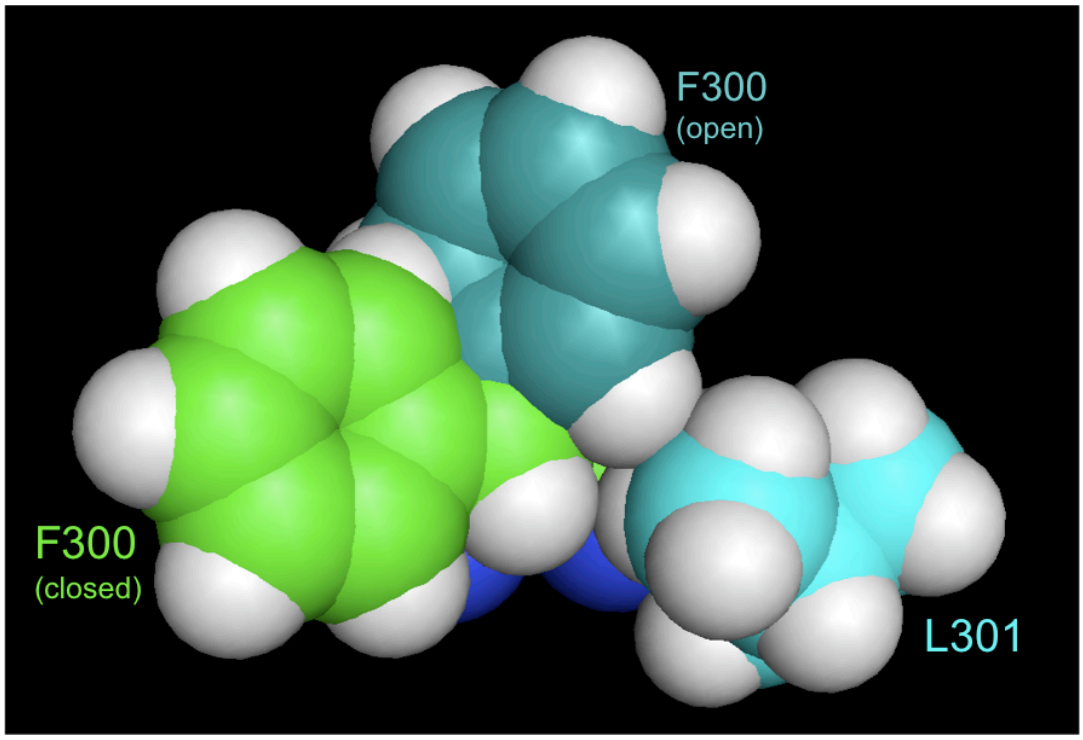


Figure 2.16. (continued)

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.