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

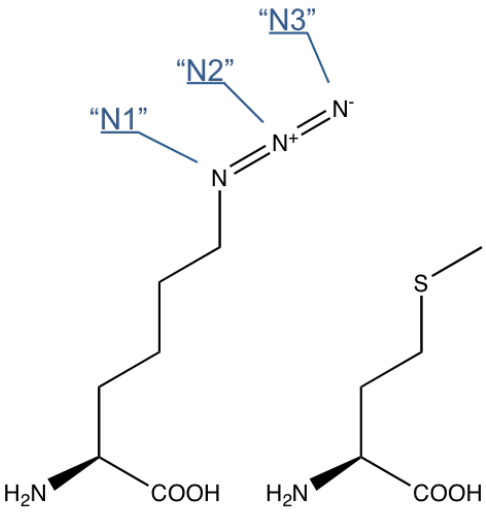
* NLL, the highest ranking experimentally tested mutant, is included here for comparison.

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.

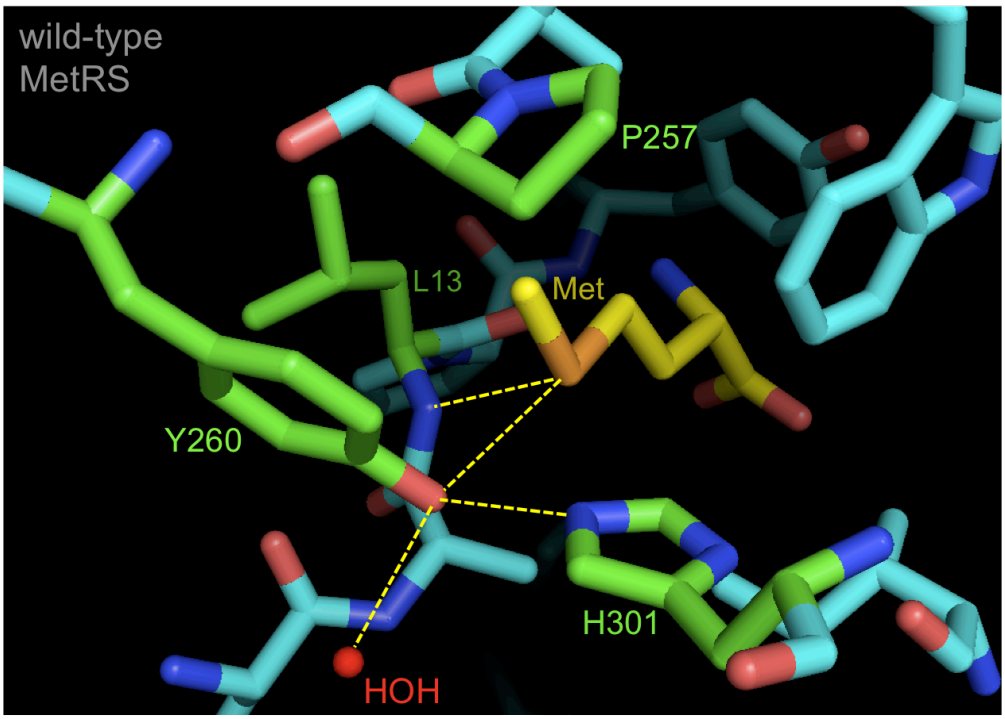
a) Chemical structures of MetRS ligands.



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].

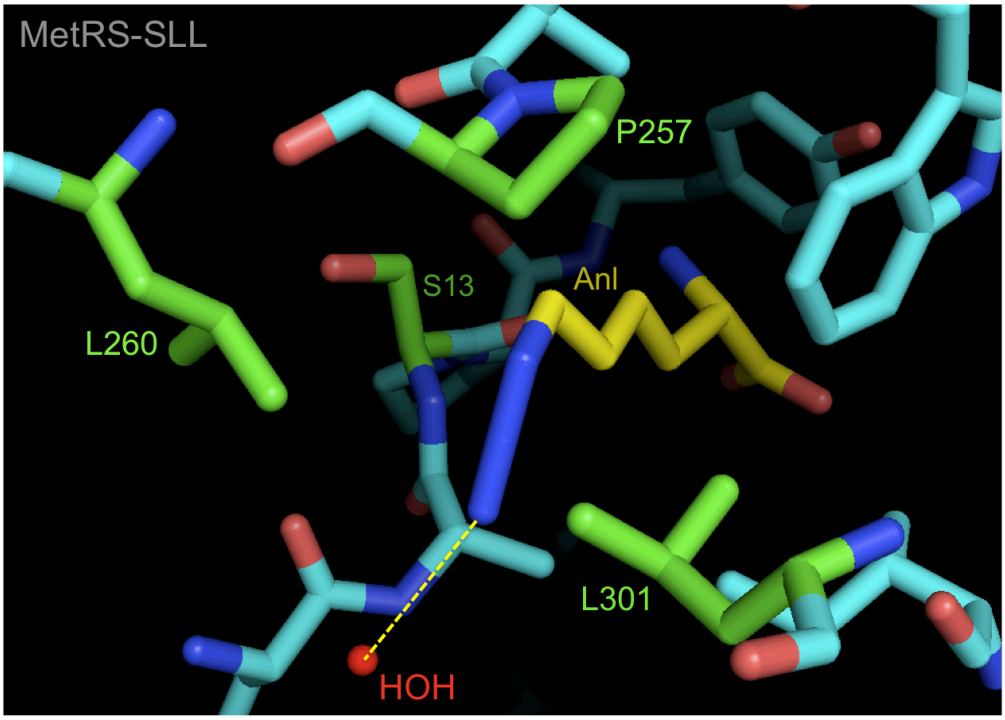


Figure 3.2.

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.

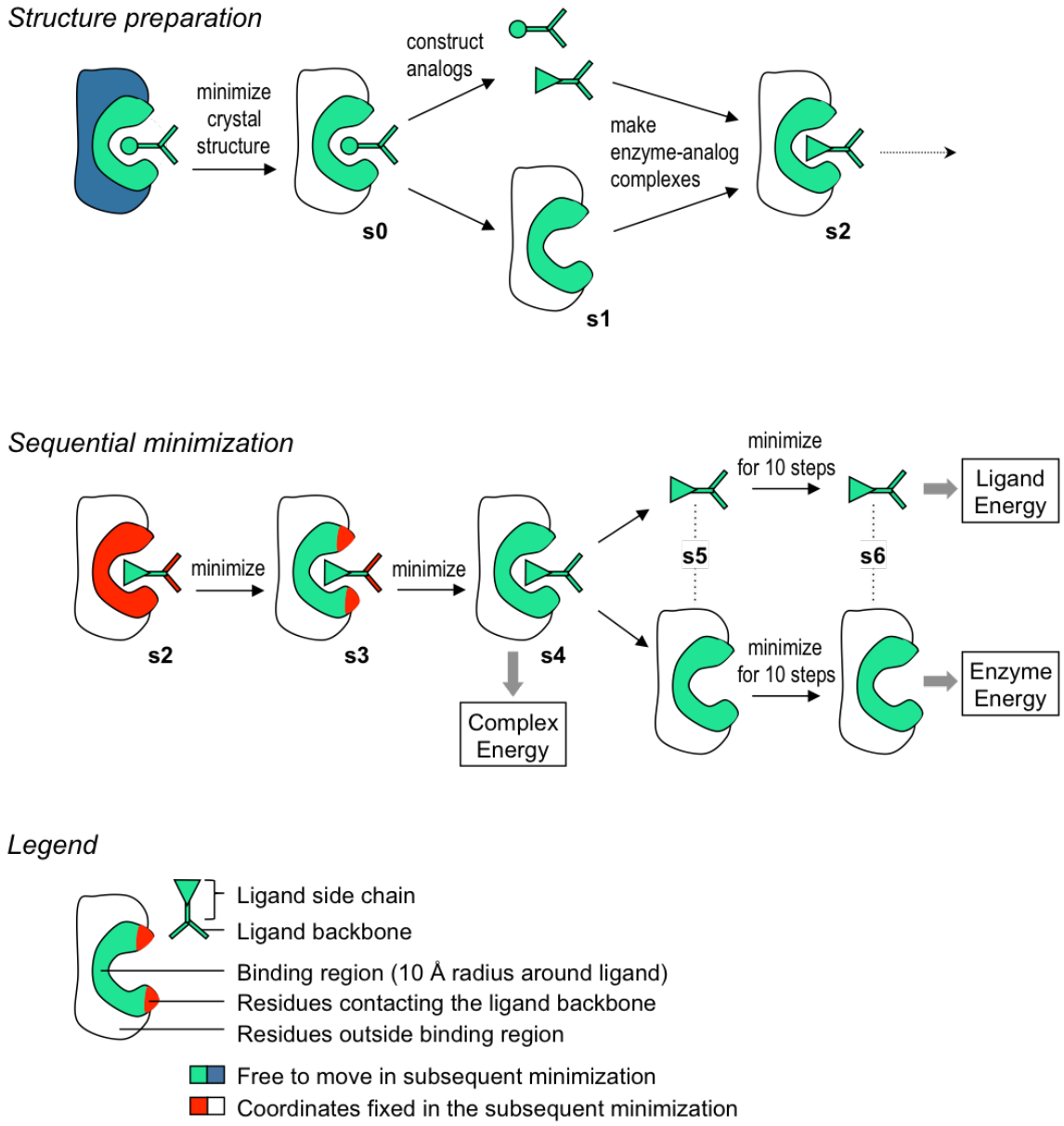


Figure 3.3.

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)

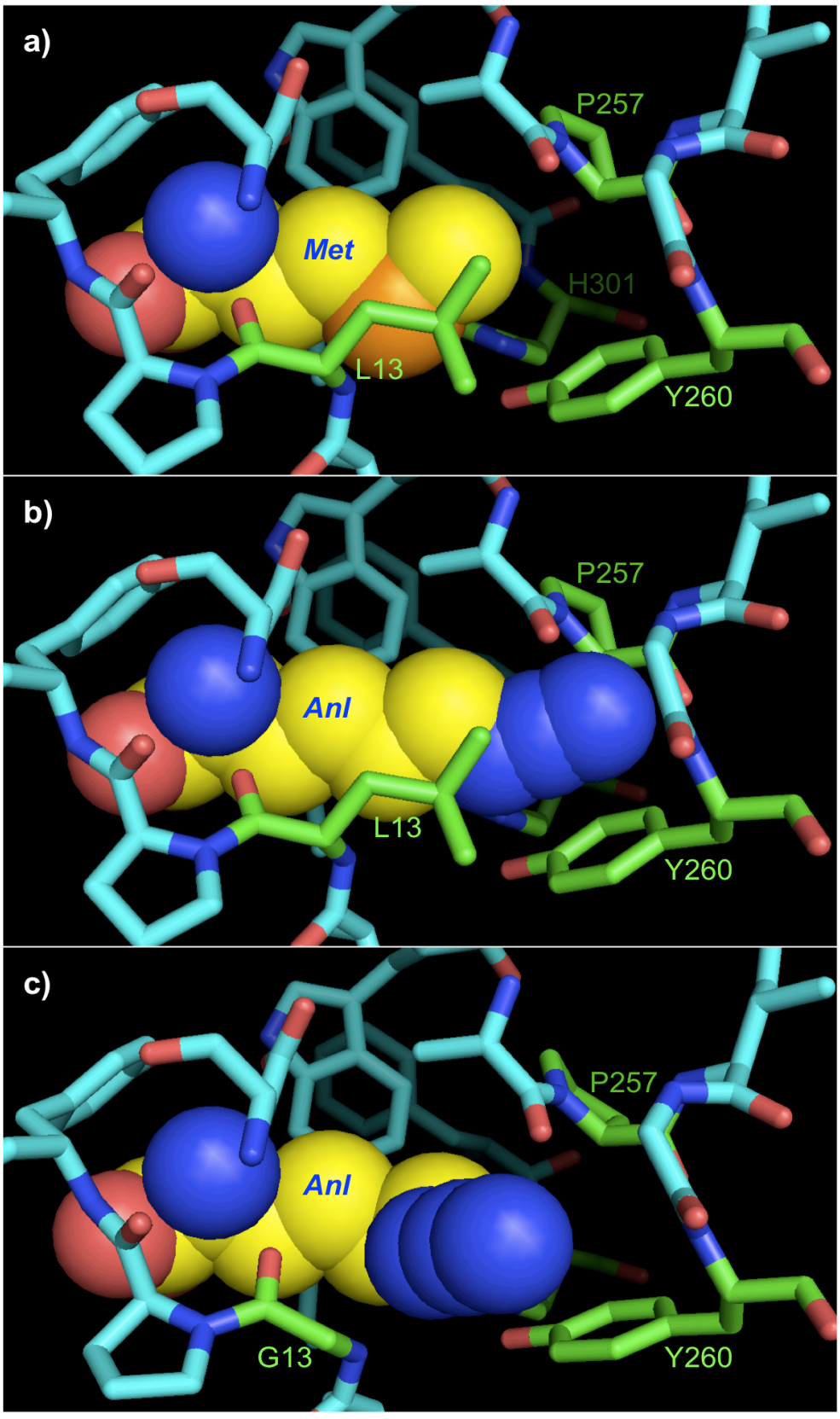


Figure 3.4.

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.

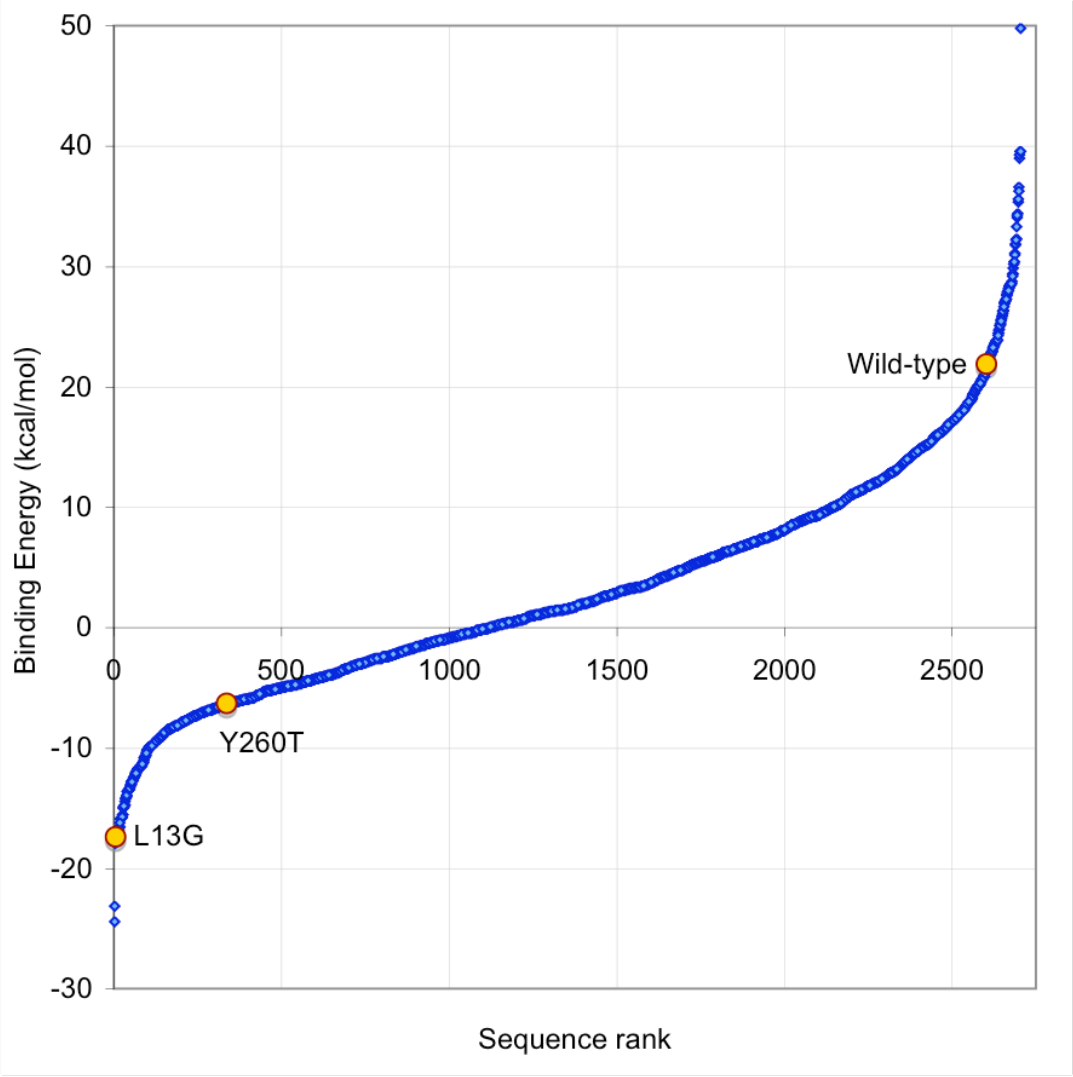


Figure 3.5.

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.

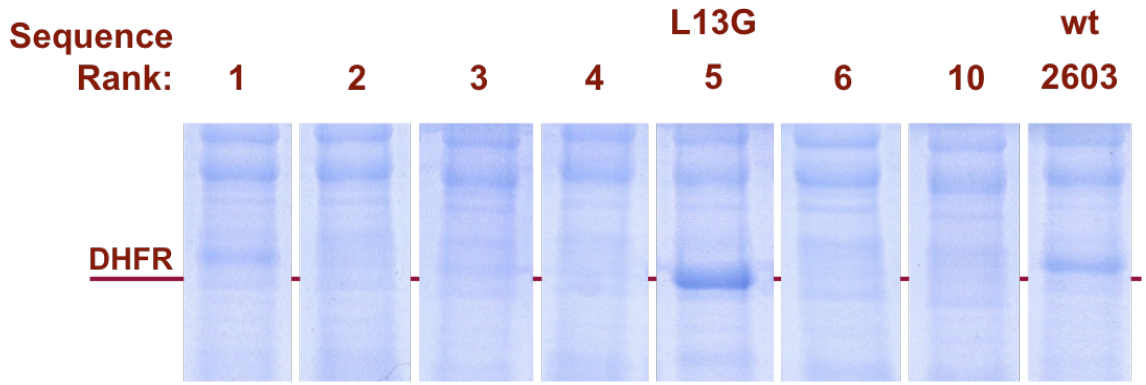


Figure 3.6.

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)

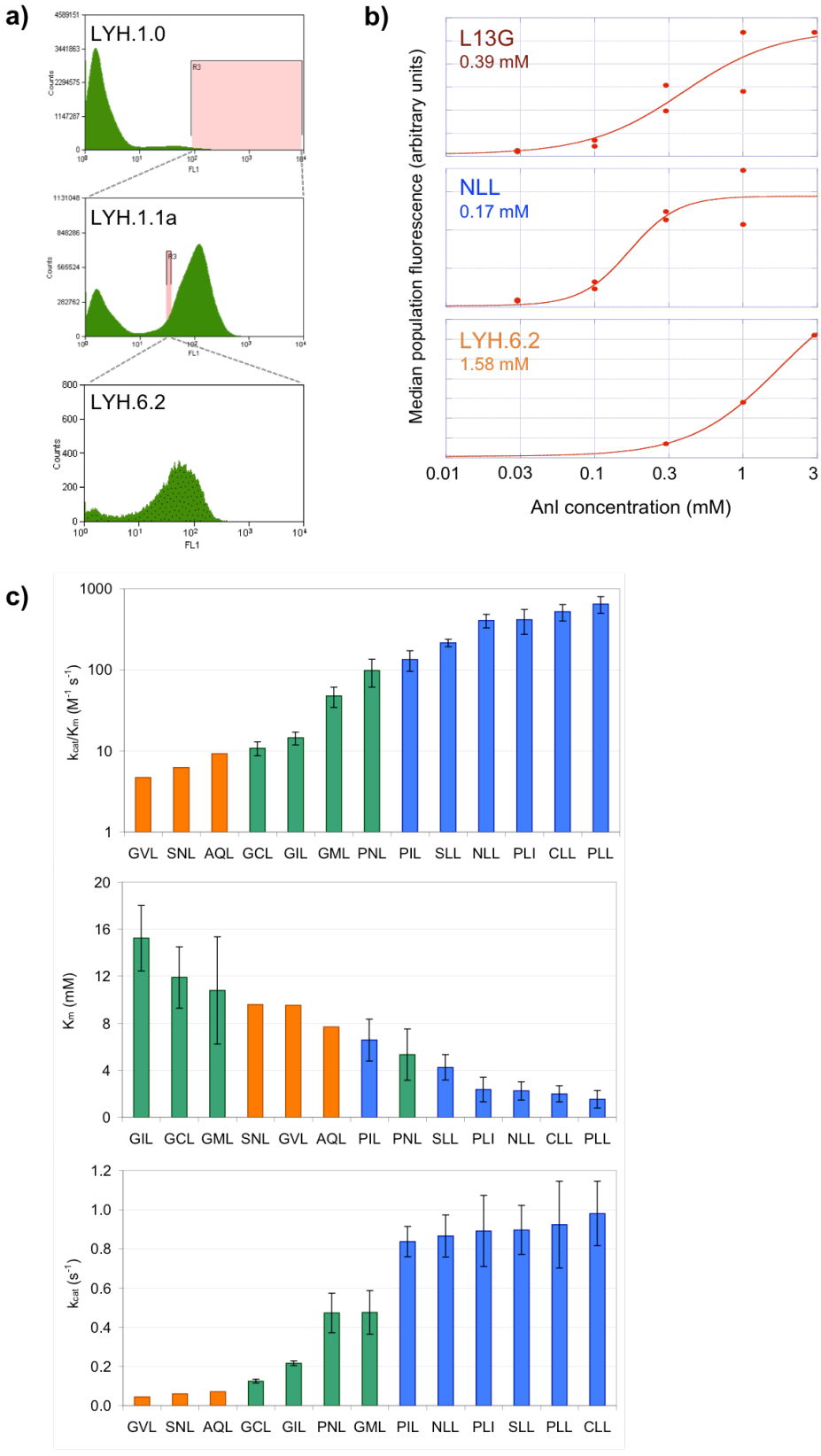


Figure 3.7.

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.

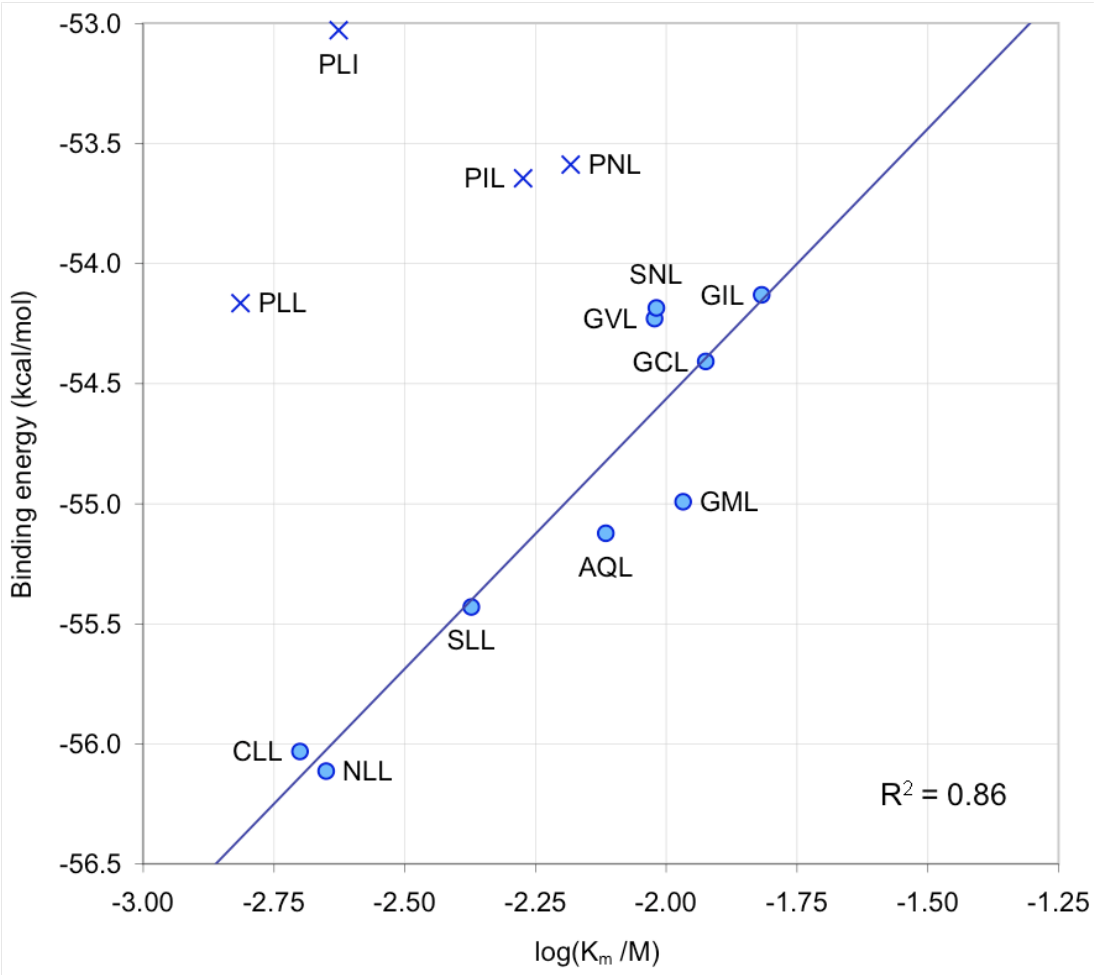


Figure 3.7. (continued)

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

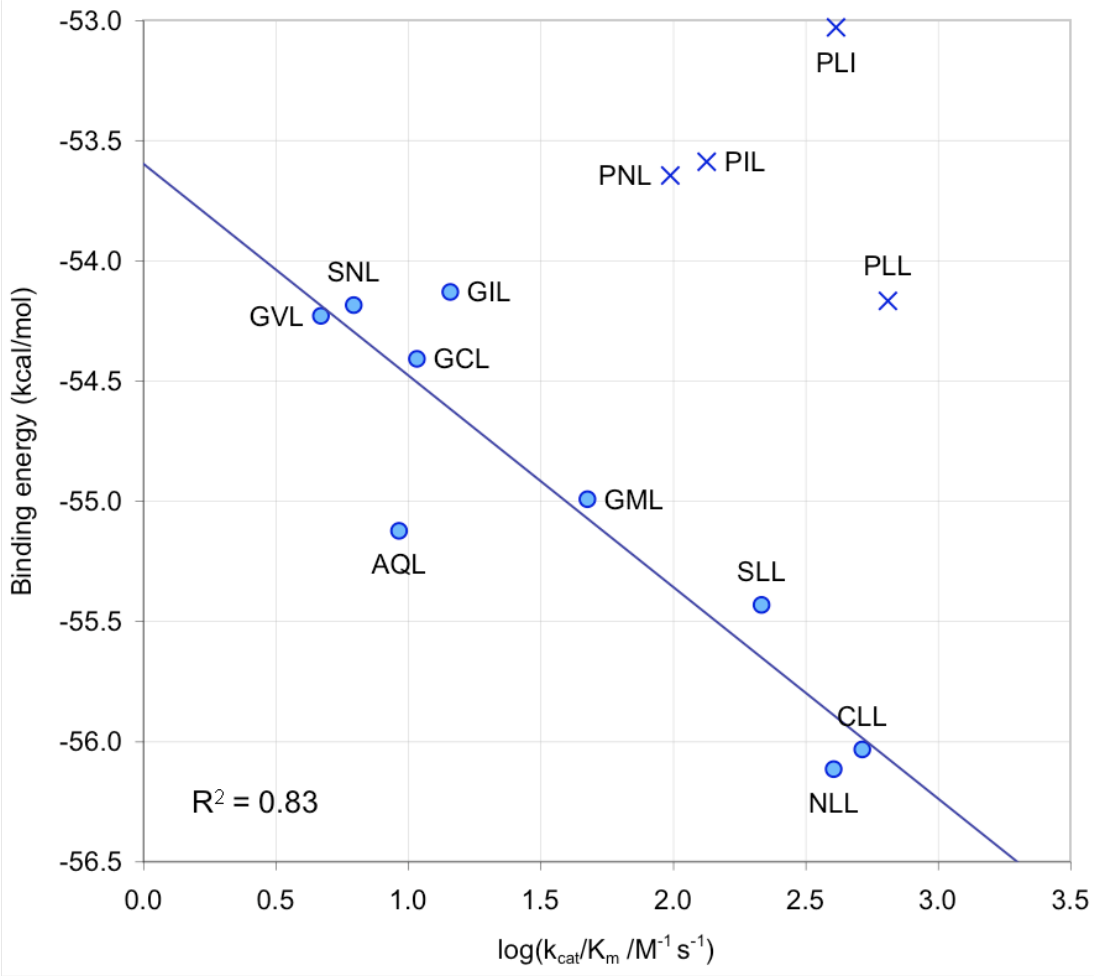


Figure 3.8.

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)

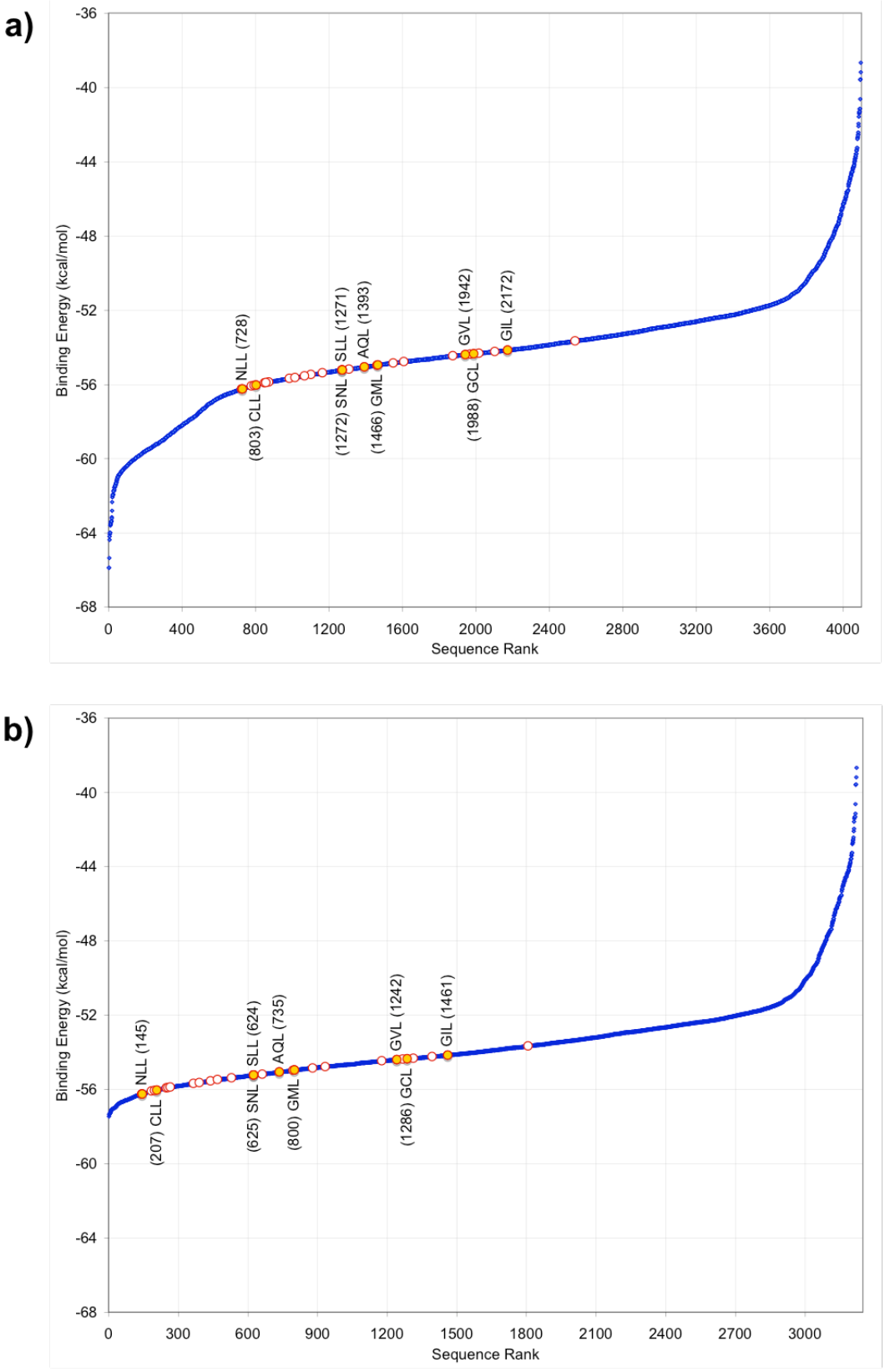


Figure 3.9.

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)

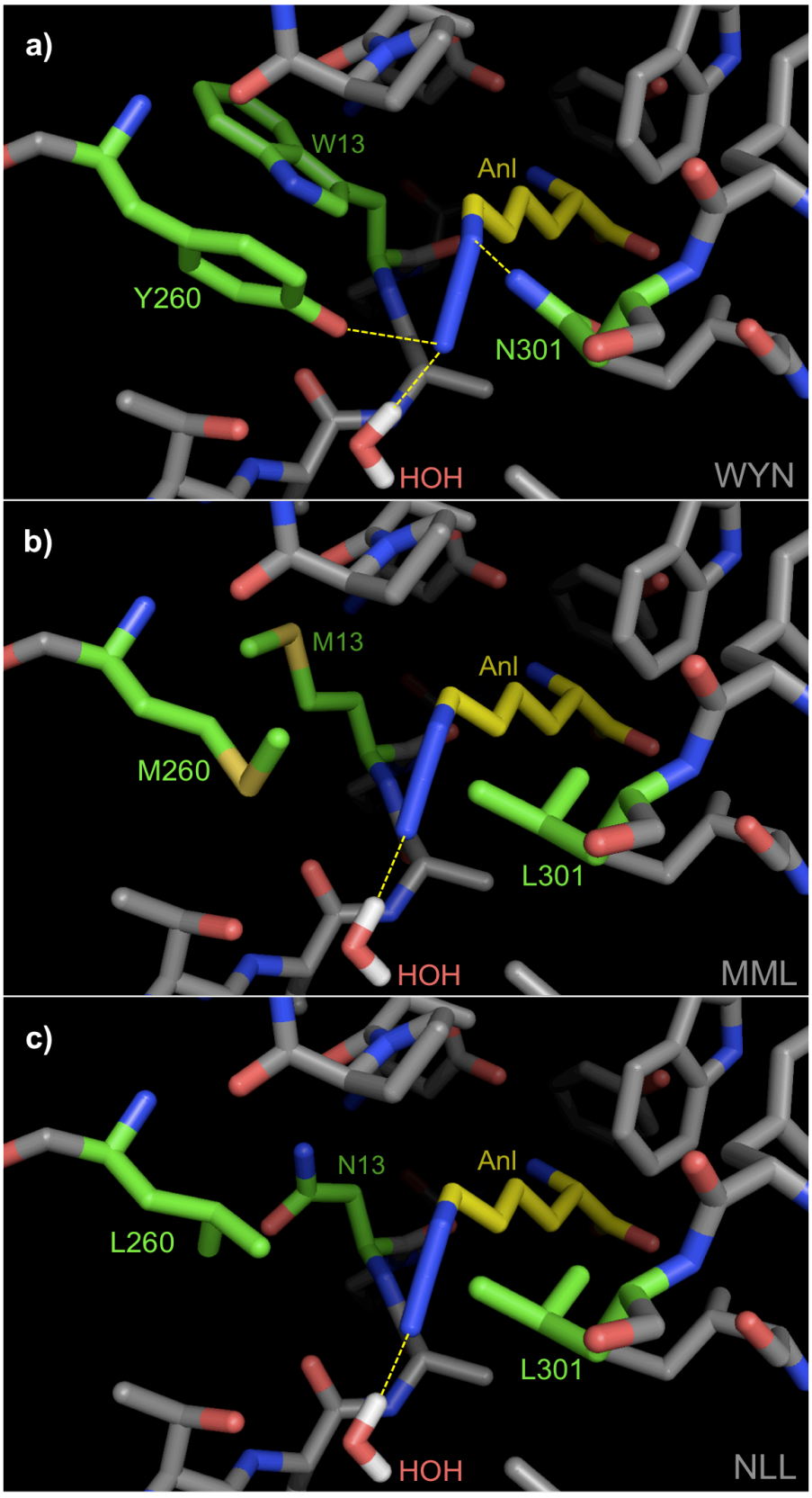


Figure 3.10.

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.

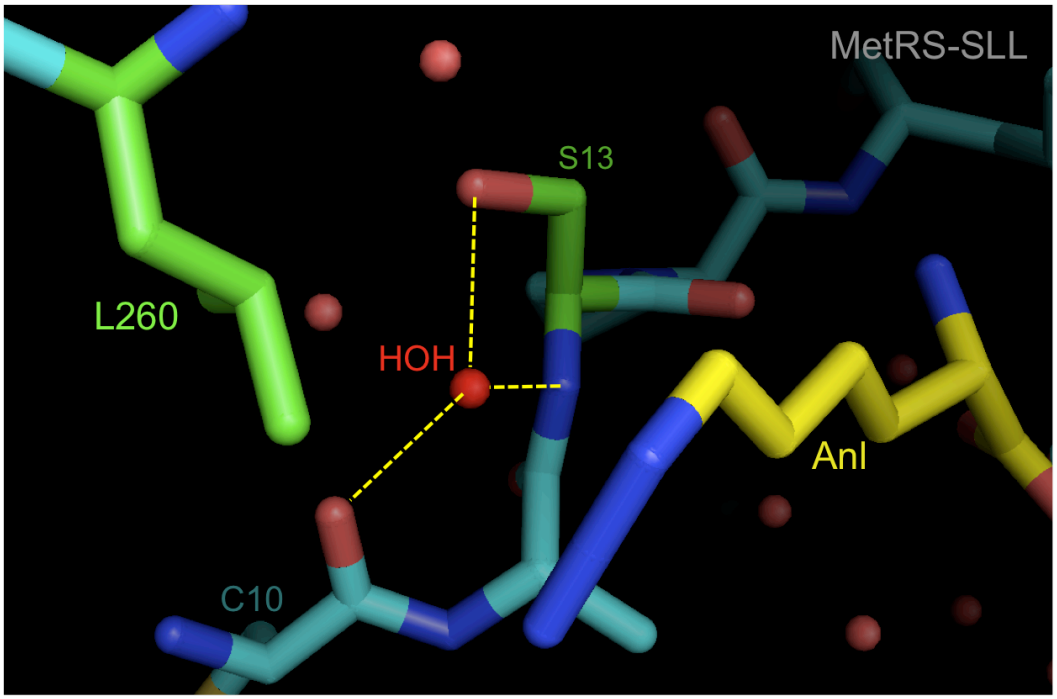
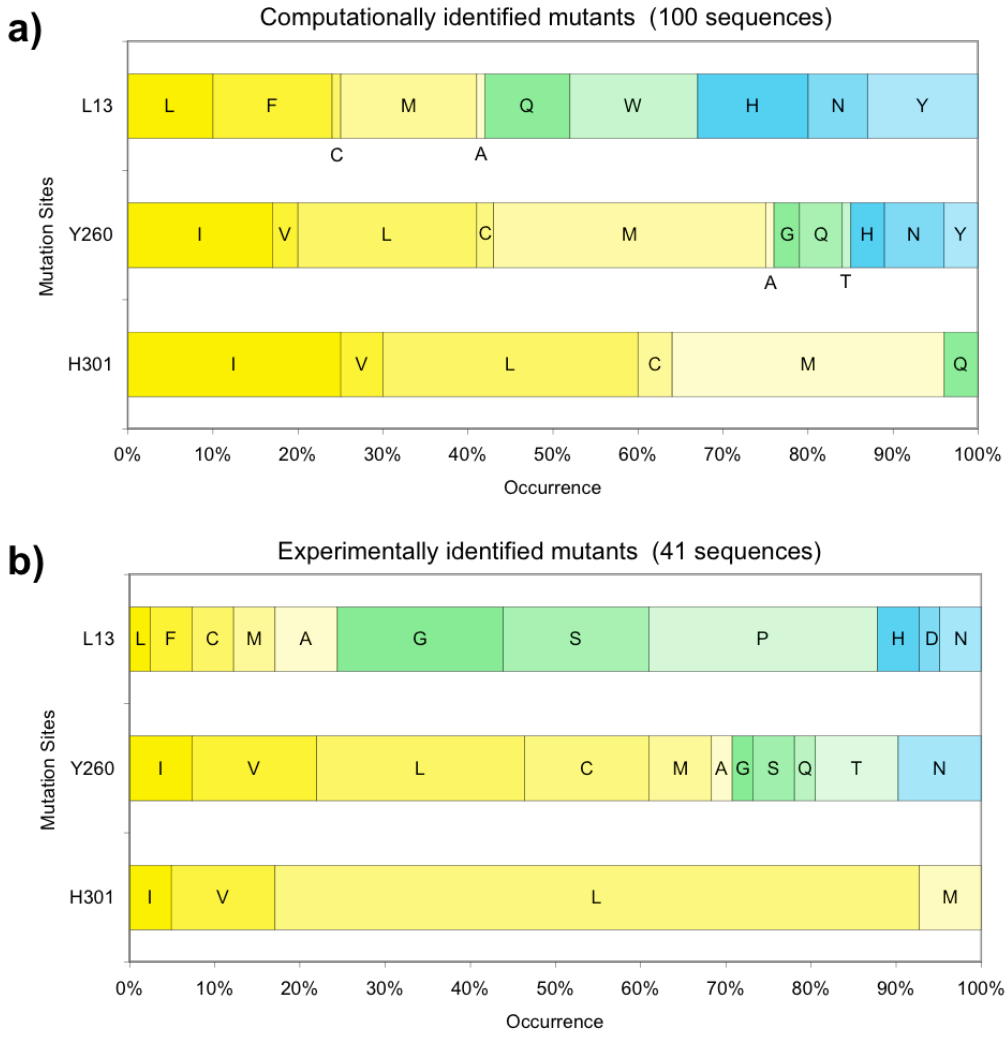


Figure 3.11.

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.