Gleaning Unexpected Fruits from Hard-Won Synthetases: Probing Principles of Permissivity in Non-canonical Amino Acid–tRNA Synthetases

Citation	Cooley , R. B., Karplus, P. A. and Mehl, R. A. (2014). Gleaning Unexpected Fruits from Hard-Won Synthetases: Probing Principles of Permissivity in Non-canonical Amino Acid-tRNA Synthetases. ChemBioChem, 15(12), 1810-1819. doi:10.1002/cbic.201402180
DOI	10.1002/cbic.201402180
Publisher	John Wiley & Sons, Inc.
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Gleaning unexpected fruits from hard-won synthetases: Probing principles of permissivity in non-canonical amino acid-tRNA synthetases

Richard B. Cooley, P. Andrew Karplus and Ryan A. Mehl*

Department of Biochemistry and Biophysics, 2011 Ag & Life Sciences Bldg, Oregon State University, Corvallis, OR 97331

*corresponding authors: Ryan A. Mehl
Department of Biochemistry and Biophysics
2011 Ag & Life Sciences Bldg
Oregon State University, Corvallis, OR 97331
Phone: (541) 737-4511; Fax (541) 737-0481
E-mail: ryan.mehl@oregonstate.edu

Abstract

The site-specific incorporation of non-canonical amino acids (ncAAs) into proteins is an important tool for understanding biological function. Traditionally, each new ncAA targeted requires a resource-consuming process of generating new ncAA aminoacyl tRNA synthetase/$tRNA_{CUA}$ pairs. However, the discovery that some tRNA synthetases are "permissive," in that they can incorporate multiple ncAAs, means it is no longer always necessary to develop a new synthetase for each newly desired ncAA. Developing a better understanding of what factors make ncAA-synthetases more permissive would increase the utility of this new approach. Here, we characterize two synthetases selected for the same ncAA that have markedly different "permissivity profiles." Remarkably, the more permissive synthetase even incorporates an ncAA for which we had not been able to generate a synthetase using *de novo* selections. Crystal structures reveal that the two synthetases recognize their parent ncAA through a conserved core of interactions, with the more permissive synthetase displaying a greater degree of flexibility in its interaction geometries. We also observe that intra-protein interactions not directly involved in ncAA binding can play a crucial role in synthetase permissivity and suggest that designing such interactions may provide an avenue to engineering synthetases with enhanced permissivity.

Introduction

Genetic code expansion has facilitated the site-specific incorporation of non-canonical amino acids (ncAAs) into proteins by using synthetically evolved, orthogonal aminoacyl-tRNA synthetase/$tRNA_{CUA}$ pairs in prokaryotic cells,[1] eukaryotic cells[2] and animals[3]. The method has allowed the installation of a wide array of chemical functional groups that have provided new approaches to manipulate and study biological systems.[4] For an orthogonal aminoacyl-tRNA synthetase/$tRNA_{CUA}$ pair to efficiently and site-specifically incorporate an ncAA, the active-site of the orthogonal synthetase must bind the ncAA in the proper location and orientation and with the correct enzyme conformation to allow for transfer to the orthogonal tRNA.[5] An effective synthetase must also maintain strict fidelity by discriminating against the canonical amino acids (AAs). Fidelity is typically assessed by measuring the amount of full-length protein expressed from a stop-codon disrupted gene in the absence of ncAA (termed 'absolute' fidelity), however only fidelity in the presence of ncAA (termed 'functional fidelity') is of any consequence as expression of ncAA-containing proteins only occurs in the presence of ncAA.[6]

Currently, the engineering of synthetases that can recognize a desired ncAA is accomplished by one of two methods. The most common is a *de novo* approach consisting of a double sieve selection on a large library of variants with mutations in the synthetase's active-site to identify those variants that recognize the ncAA and not any of the twenty AAs. The final step of this process is the evaluation of "hit" ncAA-synthetases for efficiency (the ability to site-specifically incorporate the ncAA in response to a unique codon) and absolute fidelity. The second, increasingly common method is to screen previously characterized ncAA-synthetases for their ability to incorporate a newly targeted ncAA. This approach takes advantage of the fact that

many ncAA-synthetases identified by the first method have the ability to incorporate a variety of different ncAAs even though they maintain their fidelity against canonical AAs. This quality of ncAA-synthetases has been termed "permissivity"[7, 8] (sometimes also referred to as "promiscuity"[9] and "polyspecificity"[10, 11]) and taking advantage of it has been effective because a surprising number of engineered synthetases have been found to possess broad "permissivity profiles" (i.e. the spectrum of tested ncAAs incorporated by a single synthetase).

This second method, in which known synthetases (or mutants thereof) are screened for permissivity is more widely accessible than the first as it avoids the complexities of generating a large library of orthogonal ncAA-synthetases and performing selections which can require gram quantities of ncAA, many of which are expensive or difficult to synthesize. In contrast, the high-throughput green fluorescent protein (GFP) reporter assay developed[8] to test the fidelity, efficiency, and permissivity profiles of an already selected ncAA-synthetase typically only requires milligram quantities of an ncAA. However, the success of generating permissive synthetases is heavily dependent on understanding the molecular basis of ncAA-synthetase permissivity.

As expected, many different ncAA-tRNA/synthetase pairs have shown permissivity profiles that incorporate ncAAs that are similar to the structure of ncAA for which they were selected [7-10, 12, 13]. While permissivity may seem like a robust attribute for ncAA synthetases, the molecular underpinnings of permissivity have been minimally investigated. In some cases, the limits of permissivity are simply tested by measuring the ability of an ncAA-synthetase to incorporate structurally similar ncAAs without making any alterations to the ncAA-synthetase.[10, 13, 14] In other cases, homology model-guided site-directed mutagenesis of an

evolved synthetase active site is used to explore the source of permissivity and tune or expand its permissivity profile.[7-9]

Thus far, in only one case have structural studies been undertaken to better understand the molecular basis of permissivity.[10] In this case, the crystal structure of the synthetase originally selected for *para*-cyanophenylalanine (pCNF) incorporation was solved in complex with pCNF. Schultz and colleagues noted that although the large substrate binding pocket of the pCNF synthetase could be important for permissivity, the naphthylalanine synthetase has "a significantly larger binding pocket" but "exhibits virtually no permissivity" and that this is also true for "a number of other amino-acyl tRNA syntheses examined with large binding sites." They concluded that there must be "a plasticity unique to pCNF-RS active site." But while noting that the pCNF synthetase active site has a small shift in one helix compared with the wild type synthetase, no specific suggestions were given as to what this "unique plasticity" might involve. Thus, despite this structural study, we still have a poor understanding of why some ncAA-synthetases are widely permissive while others are not.

With the structure of a permissive synthetase, one can identify clues about which components of the ncAA need to be present for recognition and which can be varied. Structural analyses also provide direction for site-directed mutagenesis that tunes or expands synthetase permissivity for additional ncAAs. Using the simple fluorescent reporter assay, we recently reported a fortuitous finding in which a set of synthetases originally selected for incorporation of 4-(2′-bromoisobutyramido)-phenylalanine (BibaF, **2**, Fig. 1) [15] could also incorporate acridon-2-ylalanine (Acd, **4**) and *p*-bromoisobutyryloxymethyl-L-phenylalanine (BiF, **8**). The ability to incorporate these ncAA has solved unique chemical biology problems because Acd (**4**) has distinctive fluorescent properties[16] and BiF can generate hydrolysable polymer junctions from

proteins[17].

Here, we explore in more detail the permissivity limits of these parent BibaF (**2**) synthetases. Of the five unique BibaF (**2**) synthetases identified in the original selection[15], four were very similar in sequence (differing only by a single amino acid) while the fifth synthetase was quite unusual in that it had a different amino acid at five (out of a possible seven) library sites. To see if these variations in sequence manifested differences in activity, we assess the permissivity profile of the more unusual synthetase (G2, Table1) and compare it with that of one of the other synthetases (F9, Table 1). Interestingly, we found the G2 synthetase to be notably more permissive than the F9 synthetase and provide for the first time, to our knowledge, experimental evidence that two synthetases identified from the same selection process can have significantly differing specificities for other ncAA substrates.

Given that previous analyses of permissive synthetases only focused on a single top performing synthetase, these observations not only highlight the importance of screening permissivity of multiple synthetase hits, but also provided us with a unique opportunity to directly compare two ncAA-synthetases identified from the same original selection with differing permissivity profiles. We report here crystal structures of both of these synthetases with their parent ncAA (BibaF, **2**) bound in the active site in order to identify molecular motifs within the ncAA that are recognized by both synthetases-- an important step in rationally designing synthetases for incorporation of new ncAAs. Comparison of their active site architectures reveals clues about the molecular basis of enhanced permissivity in the G2 synthetase, which is further explored by co-crystallizing this synthetase with a second ncAA, 4-*trans*-cyclooctene-amidophenylalanine (Tco-amF, **3**). This additional crystal structure offers compelling evidence for why this larger ncAA cannot be incorporated by the less permissive F9 synthetase. Lastly,

mutagenesis of active site residues identified in these crystal structures allow us to compare the functional aspects between the G2 and F9 synthetases and identify residues that influence their permissivity profiles. Together this work represents the most comprehensive analysis of permissivity in ncAA-synthetases, and as such lays important groundwork for future expansion of the genetic code.

Results and Discussion

Permissivity Profiles of the G2 and F9 BibaF tRNA synthetases

As noted briefly in the introduction, among five sequence-unique *M. jannaschii* tyrosine tRNA synthetase variants generated to incorporate BibaF (**2**) into proteins[15], all were found to equally incorporate the fluorescent amino acid acridon-2-ylalanine, Acd (**4**), with efficiencies substantially better than those achieved for BibaF (**2**).[16] Given the substantive size, shape and polarity differences between BibaF (**2**) and Acd (**4**), we speculated these might be particularly permissive synthetases and sought to further characterize them. Two of these synthetases, originally identified as the G2 and F9 variants (Table 1), represented the most different sequences among the five, and we chose them as subjects for a carrying out more extensive permissivity profiles. For this study, a set of twelve additional ncAAs with varying degrees of structural similarity (Fig. 1) were tested for their ability to be incorporated into sfGFP (Fig. 2).

Comparisons of the ncAAs and the efficiencies with which they were incorporated into sfGFP identify two notable motifs that impact their ability to be incorporated. First, those with a polar 'linker' group attached at the *para* position of the phenyl ring, particularly N-linked amide groups, were well tolerated (e.g. BibaF (**2**) and AmF (**5**)). In the case of the G2 synthetase, an amine at this position (e.g. Acd (**4**), Tet-F (**6**) and AF (**7**)) was sufficient for incorporation,

though variations such as an ester (e.g. BiF (**8**)) or a carbonyl (e.g. Bpa (**9**)) were tolerated. These observations indicate the synthetase recognizes both hydrogen bond donor and acceptor polar linker groups, though the former with better efficiency. Less polar ether linkages were not tolerated very well (e.g. M-Tyr (**11**) and Thy (**12**)). The second motif contributing to the efficiency of incorporation was the substituent attached to the *para*-positioned polar linker. In general, larger, non-polar and flexibly linked substituents were tolerated. Those amino acids with polar linker groups that did not possess this second criteria such as AQA (**13**), FluA (**14**) and CarF (**15**) were not detectably incorporated.

The ability for the orthogonal tRNA/G2 and tRNA/F9 pairs to efficiently produce ncAA-sfGFP with Acd (**4**) but not at all with AQA (**13**) and FluA (**14**) is surprising considering their structural similarities. However, it must be noted that the lack of success with the latter compounds could be due to a failure at the level of EF-Tu transport of the loaded ncAA-tRNA.[18] Nevertheless, these observations as a whole imply that substrate recognition by these permissive synthetases is largely influenced by recognition of a *para*-positioned polar linker and a larger, flexible hydrophobic moiety.

Permissive synthetase succeeds where de novo *selections failed.* Given its potentially broad application in bioorthogonal ligations with tetrazine-containing compounds[19], our laboratory has attempted on several occasions to isolate through *de novo* selections an ncAA-synthetase for Tco-amF (**3**), but without success (unpublished data). Assessment of Tco-amF (**3**) incorporation at 1 mM concentration by the G2 and F9 synthetases also did not yield convincing evidence for its incorporation. However, Tco-amF (**3**) exists as four different stereoisomers each in approximately equal molar quantities, therefore we supplemented the media with higher concentrations of Tco-amF (**3**) (5mM) than was used in *de novo* selections in the event that only

a single stereoisomer is selected by the synthetase. At a total concentration of 5 mM, the G2 synthetase was able to incorporate Tco-amF (**3**). Though the level of incorporation of Tco-amF (**3**) was modest, this would be expected if the G2 synthetase preferentially incorporates a single stereoisomer— an observation supported by the G2-Tco-amF crystal structure (see below and Supplemental Fig. 1). This suggests that the low efficiency of the G2 synthetase for Tco-amF (**3**) can be overcome with additional ncAA in the media and that this synthetase could very well have been be selected with this ncAA if used at higher concentrations (provided higher ncAA concentrations are not toxic to cells under selection conditions, which was not the case for this particular ncAA due to the high amount of salt present in Tco-amF (**3**) preparations). These observations highlight a particularly important advantage of screening selected ncAA synthetases for permissivity: with permissive synthetases, many more ncAA concentrations, media and growth conditions can be rapidly screened with minimal quantities of reagents and small volumes of media whereas the majority of *de novo* synthetase selections are conducted under a narrow range of well-defined conditions, typically with 1 mM ncAA present in a comparatively large quantity of media. As a result, the limited conditions used in *de novo* selections limits the number of potentially suitable candidate synthetases. To our knowledge, this is the first example in which *de novo* selection approaches to make an ncAA-specific synthetase had failed, but the ncAA incorporation could be achieved by a permissive synthetase.

Structural characterization of the G2 and F9 synthetases

Though both the G2 and F9 synthetases were permissive to some degree, the G2 variant is notably more so than the F9 variant as it incorporated amino acids that the F9 did not, such as Tco-amF (**3**), Tet-F (**6**), AF (**7**) and B-Tyr (**10**). Additionally, for those amino acids that both G2

and F9 could incorporate, the G2 variant was more efficient. Because these two synthetases differ considerably with five of the seven possible library sites having different amino acids, this difference offered a unique opportunity to use structural and mutagenic analyses to investigate the molecular underpinnings of their distinct permissivity profiles, including seeking to solve crystal structures of the same synthetase with different ncAAs bound in its active site.

We solved the structures of both the G2 and F9 synthetases with their parent substrate, Bibaf (**2**), bound to their active site. We also were able to solve the structure of the more permissive G2 synthetase bound to Tco-amF (**3**), an amino acid inefficiently but detectably incorporated by the G2 synthetase, but not incorporated by the F9 synthetase at the concentrations tested. All crystals grew in similar conditions to previously solved *Mj*Tyr-synthetase variants [20, 21] with the G2 synthetase complexes obtained from co-crystallizations with ligand and the F9 synthetase complex derived by soaking crystals of the unliganded enzyme with BibaF (**2**). The three synthetase structures all adopted the same "closed" conformation state as seen for other *Mj*Tyr-synthetase variants with substrate bound to its active site[6, 22, 23] (~0.5 Å Cα RMSD to wild-type for all three structures), and were refined at resolutions near 2.0 Å, sufficient for identifying key protein-substrate interactions.

The G2-synthetase complex with BibaF. Analysis of initial diffraction data sets showed that during data collection, the synchrotron radiation cleaved the bromine atom from the BibaF (**2**) substrate. To minimize this problem, multiple short data sets were merged together to generate a complete data set with minimal radiation damage (see *Materials and Methods*). Using these data, electron density for the substrate, including that of the bromine atom, was well-defined (Fig. 3A). Refinement showed that the bromine atom was best modeled at 0.5 occupancy, implying that some cleavage occurred even during the short periods of data

collection used. The final structure refined at 1.9 Å resolution had R/R_{free} values of 19.4/24.3% (Table 2).

The G2-BibaF complex shows how its six mutations relative to the wild-type Tyr synthetase (Y32G, L65E, F108W, Q109M, D158S and L162K) allow it to accommodate BibaF (**2**) in the active site (Figure 3A). In particular, both the Y32G and D158S mutations open up the active site to better accommodate the larger substrate. The L65E mutation positions the Glu carboxylate within hydrogen bonding distance to the amide nitrogen of BibaF (**2**), and an ordered water molecule hydrogen-bonds to the peptide carbonyl, revealing how the G2 synthetase recognizes this polar 'linker' motif critical for efficient incorporation by this synthetase. The back end of the binding pocket is lined by residues Ile63, Trp108, Ser158, Lys162 and Val164, which creates a large and relatively hydrophobic pocket that accommodates the bromoisobutyryl substituent with room to spare. The bromine atom is positioned too far away from any protein functional groups to make any halogen bonds[24] to the protein. In general, the active site pocket is well suited to accommodate a *para* positioned polar group (e.g. an amide, amine or ester group) that links to the phenyl ring a larger, generally hydrophobic substituent. We infer from this structure that the permissivity of the G2 synthetase combines a strong binding interaction to one part of the ncAA (the amide linkage) with a rather indiscriminant pocket for an additional part of the ncAA.

The G2 synthetase complex with Tco-amF. The structure of the G2 synthetase with Tco-amF (**3**) bound in the active site was refined at 1.9 Å resolution to a final R/R_{free} of 20.4/25.4%. The electron density maps revealed unambiguous density for the Tco-amF (**3**) substrate (Fig. 3B), which, due to it possessing two sources of chirality (other than the Cα atom of the amino acid backbone), can exist in four different stereoisomers (Supplemental Figure 1). Further, the

cyclooctene ring of each stereoisomer can adopt two orientations differing by a 180^{o} rotation. Therefore there were 8 potential ways in which the Tco-amF (**3**) substrate could be modeled. Refinements with all eight possibilities provided convincing evidence that one particular stereoisomer and orientation both provided the best fit to the electron density (Supplemental Figure 1) and had the fewest clashes with neighboring atoms. Although the resolution achieved here is not sufficient to unequivocally conclude that no other stereoisomers and orientations are bound in minor populations, that ambiguity has no impact on the insights we derive from the complex. For simplicity all subsequent discussions of the G2-Tco-amF structure will refer to this single stereoisomer that we have modeled (Fig. 3B).

Modeling the Tco-amF (**3**) substrate into the electron density maps of the active site revealed the presence of several protein-substrate interactions that are similar to those seen in the G2-BibaF complex (Fig. 3B). In particular, Glu65 again makes a hydrogen bond to the amide N-H as it did in the G2-BibaF structure while the carbonyl oxygen of the amide makes a similar hydrogen bond to a water molecule. Interestingly, the amide group of Tco-amF (**3**) is rotated out of the plane of the phenyl ring by ~30^{o} while in the BibaF (**2**) structure it is rotated only ~15^{o} out of the plane (Fig. 4A). We suspect that this added rotation indicates some strain present in the complex that is needed to minimize the potential clashes between the bulky cyclooctene ring and the methylene groups of Glu65 and other parts of the pocket. This difference in the binding mode (Fig. 4A) indicates some variation in the position of the polar linker group can be accommodated in the active site of the G2 synthetase, and that in general permissivity can be enabled by variations in the binding details of even well-recognized common parts of a pair of ncAAs.

The accommodation of the very bulky cyclooctene group also involves flexibility in the active site pocket as is observed in the movement of Lys162 and the helix on which it resides.

This helix forms the back end of the binding pocket and has been pushed 1.2 Å further back compared to the G2-BibaF structure (Fig. 4A). In addition to this, it seems that the bulk of the cyclooctene group forces a further adjustment in protein substrate binding pocket that is unusual and also may indicate strain: the phenylalanine part of Tco-amF (**3**) is itself positioned ~0.5 Å less deep into the active site than BibaF (**2**) (Fig. 4A), presumably as an additional means of minimizing protein-substrate clashes. We expect that such a change in the positioning of the substrate backbone would slow catalysis and could thus be part of the reason why this ncAA is less efficiently incorporated. These observations demonstrate that inherent flexibility of the synthetase active site and the substrate itself both play a role of the relatively high permissivity of the G2 synthetase.

The F9-BibaF synthetase structure. Crystals of the F9 synthetase did not grow in the presence of substrate, however crystals grown in the absence of substrate that were incubated with BibaF (**2**) for 5 days diffracted to 2.0 Å. Merging of multiple short-exposure datasets as done for the G2-BibaF structure produced electron density maps with strong evidence for its presence in the F9 active site (Fig. 3C). Also as for the G2-BibaF complex, the BibaF ligand was modeled at 1.0 occupancy except for the bromine atom, which was modeled at 0.5 occupancy due to cleavage by synchrotron radiation. The final F9-BibaF structure had R/R_{free} values of 17.9/23.5% (Table 2).

The F9 synthetase has four mutations relative to the wild-type *Mj*Tyr synthetase (Y32G, L65E, D158G, I159C) and five differences from that of the G2 synthetase (see Table 1). Despite these five differences from G2, the overall architecture of the F9 active site bound to BibaF is quite similar to that of the G2 synthetase (Figs. 3C and 4B). In particular, the N-H group and carbonyl oxygen of the BibaF amide polar linker make the same conserved hydrogen bonds to

Glu65 and a water molecule, respectively. Unlike in the G2 synthetase, Glu65 of the F9 synthetase makes a hydrogen bond to Gln109 (which is a methionine is G2), and this may anchor the conformation of Glu65 more tightly and make it less able (compared with the G2 synthetase) to flex in position to accommodate polar linker groups on different ncAAs (Fig. 3C). Interestingly, the amide group is not rotated out of the plane of phenyl ring unlike the G2-BibaF or G2-Tco-amF structures (Fig. 4B).

Other notable differences between the F9-BibaF and the G2-BibaF structures reside in the back end of the active site pocket. In particular, the relatively inflexible side chain of Leu162 points directly toward the bromine atom, creating a smaller pocket than is present in the G2 active site (Figs. 3C and 4B) and providing a clear rationale for its incompatibility with the larger cyclooctene ring of Tco-amF (**3**) (Fig. 4C).

Also, of interest are the residues Gly158 and Phe108 which line the ligand binding pocket of the F9 synthetase and would appear to open up more space compared with the Ser158 and Trp108 residues present at the equivalent positions in the G2 synthetase pocket (compare Figs. 3C and 3A). While the greater space present in this part of the F9 synthetase pocket should in principle allow for greater permissivity, the fact that the G2 synthetase is more permissive leads us to ask how the Trp and Ser residues might aid permissivity even while making the pocket smaller. With this question in mind, we note that the trio of residues Ser158, Trp108 and Tyr114 form a hydrogen bonding network in the G2 synthetase (Figs. 3A and 3B) that helps stabilize the holo-closed conformation of the active site that is crucial for the catalysis [6, 23]. Such a stabilizing factor could provide a crucial difference that allows the active site to succeed in properly closing even in the presence of a very large ncAA substrate like Tco-amF (**3**) that actually does not fit well. Viewed from the other direction, the loss of this hydrogen bonding network in the F9

synthetase could lessen the intrinsic stability of its closed conformation, such that it must form more ideal protein-substrate interactions in order to successfully adopt the closed conformation needed for catalysis. Regardless of the extent to which this factor plays a role for this pair of synthetases, this concept provides a novel principle to take into account when considering explanations for synthetase permissivity, or lack thereof.

Mutations probing the origins of G2 and F9 synthetases specificity and permissivity

The above structural and functional characterizations of the G2 and F9 synthetases provided important groundwork for understanding the molecular underpinnings of their differing activity and permissivity profiles. As a follow up to this work, we designed several mutations to test specific hypotheses and to sample the sequence space between the two synthetases. The activities of these hybrid synthetases and additional variants we created were evaluated in the same sfGFP reporter system using a representative subset of ncAAs that were tested in the original permissivity profiles.

Mutations of Glu65 of the G2 synthetase to Ser and Asp were made to test our proposal for a conserved role of this side chain in substrate recognition. Interestingly, not only did removal of this carboxylate by mutation to serine abolish all measureable activity of the synthetase (with the tested ncAAs), but shortening this side chain to an aspartate dramatically decreased the permissivity of the synthetase (Fig. 5). These observations are consistent with Glu65 playing a critical role in both substrate recognition and permissivity.

Previous studies have suggested an important factor in synthetase permissivity is possessing a sizeable binding pocket to accommodate different ncAAs even though several synthetases that incorporate large substrates are not permissive.[10] To see if the G2 synthetase

permissivity could be expanded in this manner, mutations were made to increase the size of the back portion of the pocket where the bromoisobutyryl group and *trans*-cyclooctene groups bind. Interestingly, mixed results were observed as mutations V164A and I63G narrowed G2 permissivity while variants K162A and K162G displayed enhanced efficiency and moderately expanded permissivity (Fig. 5). These results confirm what has been seen before that the enlargement of the active site cavity doesn't inherently expand permissivity, but that the substrate binding pocket size can be an important contributing factor to permissivity.

A third set of mutations were designed to expand the relatively restricted permissivity profile of the F9 synthetase by exploring the sequence space between it and the more permissive G2 synthetase. Specifically, two sites of interest were chosen based on the crystal structure. First, position 162 was modified from Leu to Lys as this should create more room through increased flexibility of this side chain in the backend part of the pocket. Although this mutation more than tripled the efficiency of the synthetase toward BibaF (**2**) while maintaining strict fidelity against canonical AAs, its permissivity to all other tested ncAAs was abolished (Fig. 5). This again demonstrates that enlargement of a synthetase active site pocket is alone not sufficient to expand its substrate range.

A second site of interest was the Ser158/Trp108/Tyr114 hydrogen bonding network observed in the G2 synthetase that we have proposed is contributing a novel stability-based factor toward the efficiency and permissivity of the G2 synthetase. The F108W mutation had minimal effect on the F9 synthetase (Fig. 5), and would not be expected by itself to create a stabilizing hydrogen bonding network. On the other hand, the G158S mutation, which would be expected to add an additional hydrogen bond in the active site to Tyr114 (see Fig. 3A), increased the overall synthetase efficiency and expanded its permissivity slightly, but also lowered its

fidelity in that it showed greater activity with canonical AAs. Adding in a second intra-protein hydrogen bond into the active site via the double mutation F108W/G158S enhanced these effects by worsening the synthetase fidelity still further but also further increasing the efficiencies of Acd (**4**) and AmF (**5**) incorporation. These observations with the F108W/G158S F9 variant are consistent with it being a synthetase that is able to bind substrates with less discrimination, and thus supports our hypothesis that Ser158-Trp108-Tyr114 hydrogen bonding network does specifically enhance the stability of the closed state. Future experiments will be aimed at further testing this hypothesis by carrying out *in vitro* studies of enzyme kinetics and thermodynamic stability using purified samples of these variants and of the native F9 and G2 synthetases.

Concluding Remarks

The site-specific incorporation of functionally diverse ncAAs into proteins has proven to be a widely successful tool for studying protein function but it is largely limited by the creation of new and robust ncAA-tRNA/synthetases pairs capable of recognizing these ncAAs. One very effective strategy that helps alleviate this bottleneck is the characterization of the permissivity profiles of existing synthetases and discovering what further ncAAs they are efficiently able to incorporate into proteins. This is nicely demonstrated by our observation that *de novo* selections for a Tco-amF (**3**) synthetase were not successful, yet permissivity analyses of a synthetase identified from the same library but selected for a different parent ncAA showed it was able to incorporate Tco-amF (**3**) under slightly modified conditions compared to those used for the *de novo* selection. Because permissivity is not currently a criterion that is built into the selection strategy of synthetases, the selected synthetases will, in general, not be optimally permissive, and understanding the factors influencing permissivity will allow us the potential ability to engineer

expanded permissivities of the original selected synthetases. In an attempt to elucidate binding motifs and principles that are associated with permissive synthetases, we here have characterized the G2 and F9 synthetases that were selected against the same ncAA but have rather differing permissivity profiles. Two discovered motifs within the synthetase active sites, which are not fully transferrable, are features that (i) recognize a key hydrogen-bond acceptor/donor moiety in the middle of the ncAA and (ii) create a large and rather flexible pocket at the back end of the active site able that can accommodate a diverse set of substituents. A third motif within the permissive G2 synthetase that is of great interest is a hydrogen bond network that does not contribute directly to ligand recognition, but appears to stabilize the closed form of the active site that is required for catalysis. Our mutational studies of this motif show that these amino acid side chains can indeed modulate the fidelity, efficiency and permissivity of the synthetase. We therefore suggest that intra-protein interactions contributing to the overall stability and kinetics of the synthetase as it undergoes conformational changes inherent to the catalytic cycle as a novel factor to consider in engineering permissive ncAA-synthetases.

Materials and Methods

Permissivity profile development for the G2 and F9 synthetases

The ncAAs used in this study were purchased from Peptech (Burlington, MA), Bachem (Torrance, CA) or Sigma (St. Louis, MO), or was synthesized as previously described: tyrosine (Tyr , **1**); 4-(2′-bromoisobutyramido)-phenylalanine, (BibaF **2**)[15]; 4-trans-cycloocten-amidopheylalanine, (Tco-amF, **3**); acridon-2-ylalanine, (Acd, **4**)[16], 4-acetamidopheylalanine, (AmF, **5**); 4-(6-methyl-s-tetrazine-3-yl) aminophenylalanine, (Tet-F, **6**)[19]; 4-aminophenylalanine, (AF, **7**); 4-bromoisobutyryloxymethyl-L-phenylalanine (BiF, **8**)[17], 4-

benzoyl-phenylalanine, (Bpa, **9**); benzyltyrosine, (B-Tyr, **10**), o-methyltyrosine, (M-Tyr, **11**); thyronine, (Thy, **12**); anthraquinonylalanine; (AQA, **13**)[25], 3-(3'-fluorenyl-9'-oxo)alanine, (FluA, **14**)[26]; 4-carbamoylphenylalanine (CarF, **15**). The synthesized Tco-amF (**3**), was provided by Dr. Joe Fox (U. Delaware). Mutations were introduced into the G2 and F9 synthetase genes using established overlap extension PCR protocols[27], and were confirmed by sequencing.

To measure the efficiency and fidelity of ncAA incorporation, a *pBad* expression plasmid containing superfolder GFP (sfGFP) with an amber non-sense (TAG) codon at position 150 (*pBad-sfGFP-150TAG*) was co-transformed with *pDule* plasmids containing the BibaF-G2 or BibaF-F9 synthetase (see Table 1) into DH10B cells[7, 8, 15]. ncAA dissolved at 1 mM (or 5 mM in the case of Tco-amF (**3**)) in 5 ml autoinducing medium (see Supplementary Table 1) was inoculated with 50 μl of overnight culture grown in non-inducing medium. Cells were incubated with shaking at 37 ^{o}C for 40 h, at which time sfGFP fluorescence of the cultures was measured in a 96-well plate (excitation: 488 nm, emission: 510 nm). The amount ncAA-incorporated sfGFP was determined by comparing total fluorescence of each culture with a standard curve of purified sfGFP.[6-8, 15]

Expression, purification and crystallization of the G2 and F9 synthetases

The DNA fragments containing the G2-BibaF and F9-BibaF synthetases were amplified by PCR from their respective *pDule* plasmids and ligated into the NcoI/XhoI sites of the expression vector *pET28a* (forward primer: 5'- CGCGCGCCATGGACGAATTTGAAATG-3', reverse primer: 5'- GGGCGCTCGAGTAATCTCTTTCTAATTGGCTCTAAAATC-3'). These expression vectors were transformed into DH10B cells and purified using a QIAprep kit

(Qiagen). The synthetases were expressed in BL21(DE3) cells and purified as previously described[20] with only minor modifications.[6]

Crystals of BibaF- and Tco-amF-bound G2 synthetase were grown using the hanging-drop vapor diffusion method at room temperature against a reservoir containing 20-22% polyethylene glycol (PEG) 300, 5% PEG 8000, 10% glycerol and 100 mM Tris pH 7.8 – 8.2 by mixing 1.25 μl of reservoir with 1.25 μl of protein (17.5 mg/ml in 20 mM Tris, 50 mM NaCl, 10 mM β-mercaptoethanol, pH 8.5) supplemented with either 5 mM BibaF or Tco-amF. Crystals grew to full size (~200 x 100 x 100 μm) within 3-6 days, at which time they were harvested by direct submersion into liquid nitrogen.

Unlike crystals of the G2 synthetase, crystals of the F9 variant would not grow in the presence of substrate under the conditions tested. To obtain a substrate-bound structure, crystals of the apo F9 enzyme were grown in the absence of substrate as described above for the G2 synthetase except that slightly higher concentrations of PEG 300 were utilized (22-25%). Crystals grew within 4-6 days, at which time crystals were removed from their crystallization drops and incubated in an artificial mother liquor (25% PEG 300, 5% PEG 8000, 10% glycerol and 100 mM Tris pH 8.0) supplemented with 5 mM BibaF. After 5 days of incubation at room temperature, crystals were frozen by directly submerging into liquid nitrogen.

Determination of the G2-BibaF, G2-Tco-amF and F9-BibaF crystal structures

Data were collected from crystals at cryotemperatures (100 K) using beamlines 5.0.2 and 4.2.2 at the Advanced Light Source (Lawrence Berkeley National Laboratory). An initial dataset collected on a single crystal of the BibaF-bound G2 synthetase showed almost no electron density for the bromine atom of BibaF. An F_o (*images 1-20*) – F_o (*images 180-200*) difference

map revealed a large positive peak where the bromine atom was expected to be, demonstrating the bromine atom was indeed present at the start of data collection but was cleaved as a result of radiation damage (data not shown). Subsequently, datasets of BibaF-bound G2 and F9 were collected by merging nine and six datasets of 40 images each, respectively, where each dataset was collected from an individual crystal or, if the crystal was large enough, from distinct volumes at opposite ends of a single crystal. Data were processed using HLK2000 or iMosflm v1.0.040, scaled and merged in space group $P4_32_12$ using SCALA; 5% of the data were randomly flagged for use in R_{free}. Data collection statistics are reported in Table 2.

Structures were determined by molecular substitution[28] using a previously solved synthetase structure (PDB 2zp1 with substrate omitted). All model building and refinements were conducted using Coot v0.6.1[29] and Phenix v1.8[30], respectively. Standard criteria were used for modeling water molecules (>1 ρ_{rms} intensity in the $2F_o - F_c$ map, >2.4 Å distance from nearest contact). Translation/libration/screw (TLS)[31] refinement of B-factors was performed for each structure using two groups which roughly correspond to the N- and C-terminal domains (residues 1-195 and 196-310). For all structures, the electron density for the substrate was not interpreted until the final rounds of refinement. Molecular restraints for the BibaF (**2**) and Tco-amF (**3**) substrates were created using Phenix-Elbow[30]. All substrates were modeled at an occupancy of 1.0, however the bromine atom of BibaF was modeled at ~50% occupancy. Molprobility[32] was used to monitor model geometry. The G2-BibaF (1.9 Å resolution), G2-Tco-amF (1.9 Å resolution) and F9-BibaF (2.0 Å) were refined to final R/R_{free} values of 19.4/24.3, 20.4/25.4 and 17.9/23.5%, respectively. Additional refinement statistics are reported in Table 2.

Accession Numbers.

Coordinates and structure factors for G2-BibaF, G2-Tco-amF and F9-BibaF synthetases have been deposited in the Protein Data Bank under accession numbers 4PBR, 4PBT and 4PBS, respectively.

Acknowledgements.

We thank Professor Joe Fox of University of Delaware for the gift of 4-trans-cycloocten-amidopheylalanine, (Tco-amF, **3**). This work was supported by NSF-MCB-0448297, and the services provided by the Cell Imaging and Analysis Facilities and Services Core of the Environmental Health Sciences Center, Oregon State University, grant number P30 ES00210, National Institute of Environmental Health Sciences, National Institutes of Health.

References

[1] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498-500.

[2] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science* **2003**, *301*, 964-967; K. Sakamoto, A. Hayashi, A. Sakamoto, D. Kiga, H. Nakayama, A. Soma, T. Kobayashi, M. Kitabatake, K. Takio, K. Saito, M. Shirouzu, I. Hirao, S. Yokoyama, *Nucleic Acids Res* **2002**, *30*, 4692-4699.

[3] S. Greiss, J. W. Chin, *Journal of the American Chemical Society* **2011**, *133*, 14196-14199; A. Bianco, F. M. Townsley, S. Greiss, K. Lang, J. W. Chin, *Nat Chem Biol* **2012**, *8*, 748-750.

[4] L. Davis, J. W. Chin, *Nature Rev Mol Cell Biol* **2012**, *13*, 168-182; C. C. Liu, P. G. Schultz, *Annu Rev Biochem* **2010**, *79*, 413-444.

[5] K. Lang, J. W. Chin, *Chem Rev* **2014**.

[6] R. B. Cooley, J. L. Feldman, C. M. Driggers, T. A. Bundy, A. L. Stokes, P. A. Karplus, R. A. Mehl, *Biochemistry* **2014**, *53*, 1916-1924.

[7] S. J. Miyake-Stoner, C. A. Refakis, J. T. Hammill, H. Lusic, J. L. Hazen, A. Deiters, R. A. Mehl, *Biochemistry* **2010**, *49*, 1667-1677.

[8] A. L. Stokes, S. J. Miyake-Stoner, J. C. Peeler, D. P. Nguyen, R. P. Hammer, R. A. Mehl, *Mol Biosyst* **2009**, *5*, 1032-1038.

[9] Y. S. Wang, X. Fang, A. L. Wallace, B. Wu, W. R. Liu, *Journal of the American Chemical Society* **2012**, *134*, 2950-2953.

[10] D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon, P. G. Schultz, *Biochemistry* **2011**, *50*, 1894-1900.

[11] D. D. Young, S. Jockush, N. J. Turro, P. G. Schultz, *Bioorg Med Chem Lett* **2011**, *21*, 7502-7504.

[12] E. Brustad, M. L. Bushey, A. Brock, J. Chittuluru, P. G. Schultz, *Bioorg Med Chem Lett* **2008**, *18*, 6004-6006; T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chemistry & biology* **2008**, *15*, 1187-1197.

[13] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *Journal of the American Chemical Society* **2009**, *131*, 8720-8721.

[14] J. C. Jackson, J. T. Hammill, R. A. Mehl, *Journal of the American Chemical Society* **2007**, *129*, 1160-1166.

[15] J. C. Peeler, B. F. Woodman, S. Averick, S. J. Miyake-Stoner, A. L. Stokes, K. R. Hess, K. Matyjaszewski, R. A. Mehl, *Journal of the American Chemical Society* **2010**, *132*, 13575-13577.

[16] L. C. Speight, A. K. Muthusamy, J. M. Goldberg, J. B. Warner, R. F. Wissner, T. S. Willi, B. F. Woodman, R. A. Mehl, E. J. Petersson, *Journal of the American Chemical Society* **2013**, *135*, 18806-18814.

[17] S. Averick, C. G. Bazewicz, B. F. Woodman, A. Simakova, R. A. Mehl, K. Matyjaszewski, *European Polymer Journal* **2013**, *49*, 2919-2924.

[18] Y. Doi, T. Ohtsuki, Y. Shimizu, T. Ueda, M. Sisido, *Journal of the American Chemical Society* **2007**, *129*, 14458-14462.

[19] J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox, R. A. Mehl, *Journal of the American Chemical Society* **2012**, *134*, 2898-2901.

[20] J. M. Turner, J. Graziano, G. Spraggon, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 6483-6488.

[21] W. Liu, L. Alfonta, A. V. Mack, P. G. Schultz, *Angew Chem Int Ed Engl* **2007**, *46*, 6073-6075.

[22] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, *Nature structural biology* **2003**, *10*, 425-432.

[23] Y. Zhang, L. Wang, P. G. Schultz, I. A. Wilson, *Protein Sci* **2005**, *14*, 1340-1349.

[24] P. Auffinger, F. A. Hays, E. Westhof, P. S. Ho, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 16789-16794.

[25] T. Matsubara, H. Shinohara, M. Sisido, *Macromolecules* **1997**, *30*, 2651-2656.

[26] Y. Han, A. Bisello, C. Nakamoto, M. Rosenblatt, M. Chorev, *J Pept Res* **2000**, *55*, 230-239.

[27] K. L. Heckman, L. R. Pease, *Nat Protoc* **2007**, *2*, 924-932.

[28] D. E. Tronrud, *Methods in enzymology* **1997**, *277*, 306-319.

[29] P. Emsley, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 2126-2132.

[30] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213-221.

[31] J. Painter, E. A. Merritt, *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 439-450.

[32] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res* **2007**, *35*, W375-383.

[33] K. Diederichs, P. A. Karplus, *Nat Struct Biol* **1997**, *4*, 269-275.

[34] P. A. Karplus, K. Diederichs, *Science* **2012**, *336*, 1030-1033.

Tables

Table 1. Sequences for G2 and F9 BibaF-synthetases at the positions that were allowed to vary in the library used for screening. Residues highlighted in gray indicate mutations relative to wild-type. Sequences of the remaining three BibaF (**2**) synthetases identified in the same selection process as the G2 and F9 synthetases, which are not discussed here, can be found in ref [15].

	Residue number								
Variant	**32**	**65**	**70**	**108**	**109**	**155**	**158**	**159**	**162**
Wild-type	Y	L	H	F	Q	Q	D	I	L
BibaF-synthetases									
G2	G	E	H	W	M	Q	S	I	K
F9	G	E	H	F	Q	Q	G	C	L

Table 2. Data Collection and Refinement Statistics

	G2-BibaF	G2-Tco-amF	F9-BibaF
Data collection [a]			
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit cell axes *a*, *b*, *c* (Å)	101.26, 101.26, 71.55	101.64, 101.640, 72.24	101.47, 101.47, 71.67
Resolution Limits (Å)	45.3-1.9 (1.97-1.90)	71.9 - 1.90 (1.97-1.90)	24.6-2.01 (2.07-2.01)
Unique Observations	29,723 (2,866)	30,390 (2,967)	24,581 (2,323)
Completeness	99.5 (97.7)	100.0 (100.0)	94.8 (91.5)
Multiplicity	15.7 (11.3)	24.5 (24)	7.8 (6.4)
Average I/σ	9.6 (1.6)	11.6 (0.8)	9.6 (1.0)
R_{meas} [b] (%)	0.04 (0.62)	0.07(1.35)	0.097(1.20)
$CC^{1/2}$ [c]	1.0 (0.70)	1.0 (0.27)	1.0 (0.2)
Refinement			
R_{cryst} / R_{free} (%)	19.4/24.3	20.4/25.4	17.9/23.5
No. protein molecules	1	1	1
No. protein residues	310	310	310
No. water molecules	160	210	214
Total number atoms	2735	2873	2753
rmsd bond angles (°)	0.836	0.433	0.804
rmsd bond lengths (Å)	0.007	0.003	0.006
<B> protein (Å^2)	45.08	31.3	36.39
<B> water (Å^2)	47.7	35.4	41.86
<B> ncAA (Å^2)	40.0	35.2	25.5
Ramachandran Plot (%) [c]			
Favored	99.4	98.8	98.4
Outliers	0.3	0.3	0.0
PDB code	4PBR	4PBT	4PBS

[a] Numbers in parentheses correspond to values in the highest resolution bin
[b] R_{meas} is the multiplicity-weighted merging *R*-factor [33]
[c] Correlation coefficient between two randomly chosen subsets containing the average intensities of each unique reflection[34]
[c] Ramachandran plot generated using Molprobity [32]

Figure Legends

Figure 1. Structures of the non-canonical amino acids tested for incorporation by the G2 and F9 synthetases: tyrosine (Tyr , **1**), 4-(2′-bromoisobutyramido)-phenylalanine (BibaF, **2**), 4-*trans*-cyclooctene-amidopheylalanine (Tco-amF, **3**), acridon-2-ylalanine (Acd, **4**), 4-acetamidopheylalanine (AmF, **5**), 4-(6-methyl-s-tetrazine-3-yl) aminophenylalanine (Tet-F, **6**), 4-aminophenylalanine (AF, **7**), 4-bromoisobutyryloxymethyl-L-phenylalanine (BiF, **8**), 4-benzoyl-phenylalanine (Bpa, **9**), benzyltyrosine (B-Tyr, **10**), o-methyltyrosine (M-Tyr, **11**), thyronine (Thy, **12**), anthraquinonylalanine (AQA, **13**), 3-(3'-fluorenyl-9'-oxo)alanine (FluA, **14**) and 4-carbamoylphenylalanine (CarF, **15**).

Figure 2. Permissivity profiles of the G2 and F9 synthetases. Reported is the amount of sfGFP with each of the ncAAs listed in Figure 1 incorporated at position 150. For each expression, ncAA concentration was 1 mM, except for Tco-amF (**3**) which was 5 mM because preparations of it consisted of an approximately equal mixture of four stereoisomers and only at these elevated concentrations did sfGFP production became significantly higher than background. Fidelity of each synthetase variant was determined by the amount of sfGFP produced from the same amber stop codon disrupted expression construct but in the absence of any ncAA in the media (indicated by the '-ncAA' bar). For reference, native sfGFP without an amber stop codon expresses at ~800-1000 mg per liter culture under identical conditions. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures.

Figure 3. Electron density evidence for and active site environment of the synthetase-ligand complexes. Shown are the (A) G2-BibaF, (B) G2-Tco-amF and (C) F9-BibaF complexes. In each

panel, blue mesh represents $2F_o$-F_c density contoured to 1.5 ρ_{rms}. Carbon atoms for the G2-BibaF, G2-Tco-amF and F9-BibaF structures are colored in gray, salmon and green, respectively. In all three panels, nitrogen (blue), oxygen (red), sulfur (yellow) and bromine (brown) atoms are indicated. Black dashed lines represent hydrogen bonds. Synthetase side chains lining the binding pocket are shown and are identified by their amino acid side chain type and residue number. For clarity purposes, Gly32 is omitted as it is obscured in this view by being positioned behind and underneath the substrate.

Figure 4. Comparison of the active sites of the G2 and F9 synthetases. (A) Overlay of the G2-BibaF (gray carbon atoms) and G2-Tco-amF (salmon carbon atoms) depicting the conformational differences between the two crystal structures. (B) Overlay of the G2-BibaF (gray carbon atoms) and F9-BibaF (green carbon atoms) active sites. (C) Modeling of the Tco-amF substrate (salmon carbon atoms) into the active site of the F9 synthetase (green carbon atoms) demonstrates a general incompatibility of this amino acid in the F9 active site (clash represented by red asterisks). In all three panels, coloring is identical to that of Fig. 3. Black arrows in panels A and B represent conformational differences between the compared structures.

Figure 5. Permissivity profiles of site-directed mutants of the G2 and F9 synthetases. For each variant, the amount of sfGFP produced containing the indicated ncAA at position 150 is reported. Fidelity of each synthetase variant was determined by the amount of sfGFP produced from the same amber disrupted expression construct but in the absence of any ncAA in the media (indicated by the '-ncAA' bar). Error bars represent the standard deviation of sfGFP produced from three independent side-by-side expression cultures. Subtle differences in expression yield

(like those seen between Fig. 3 and Fig. 5) reflect experimental variation resulting from slight differences in media and ncAA preparation.

Figures

Figure 1.

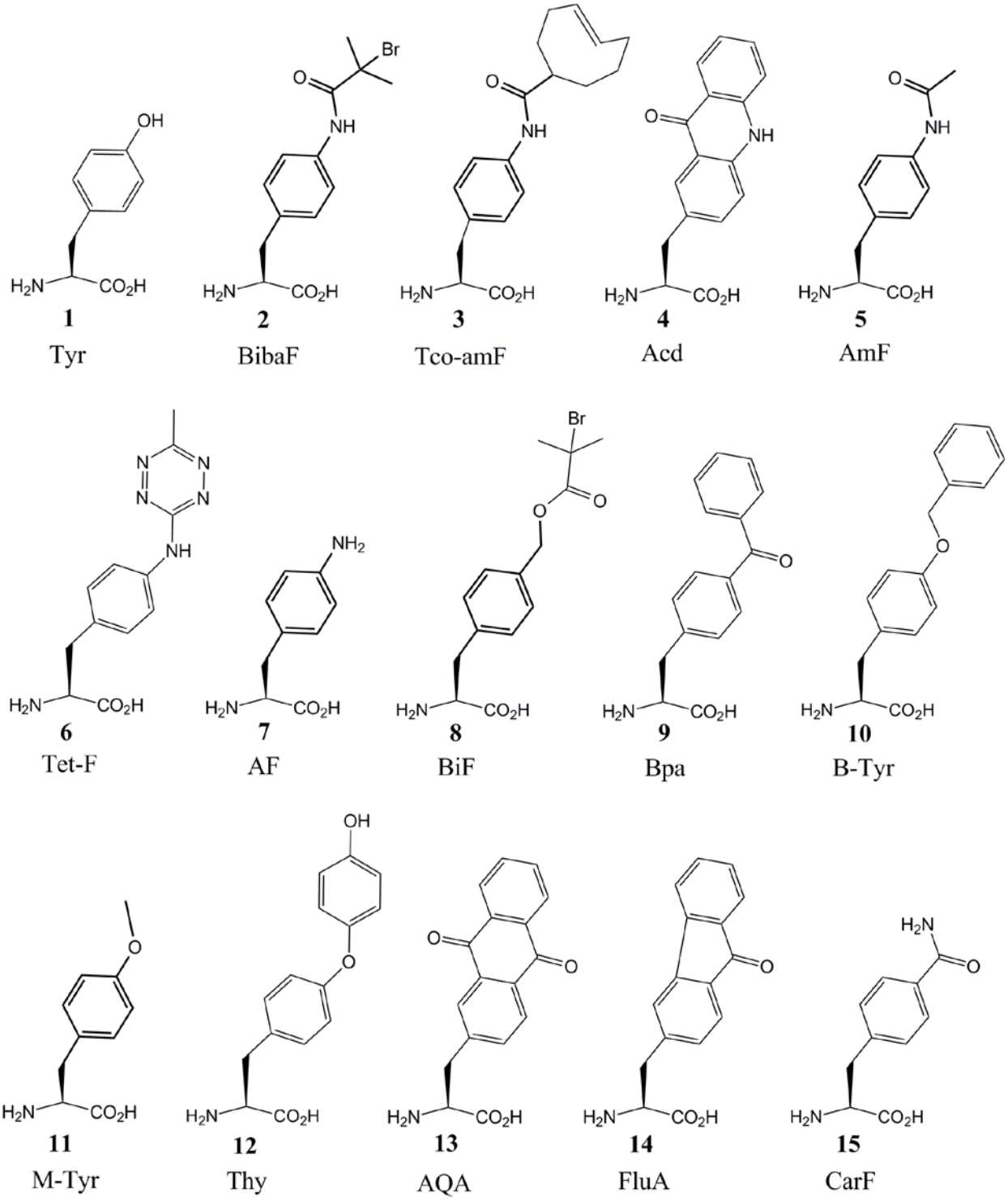


Figure 2.

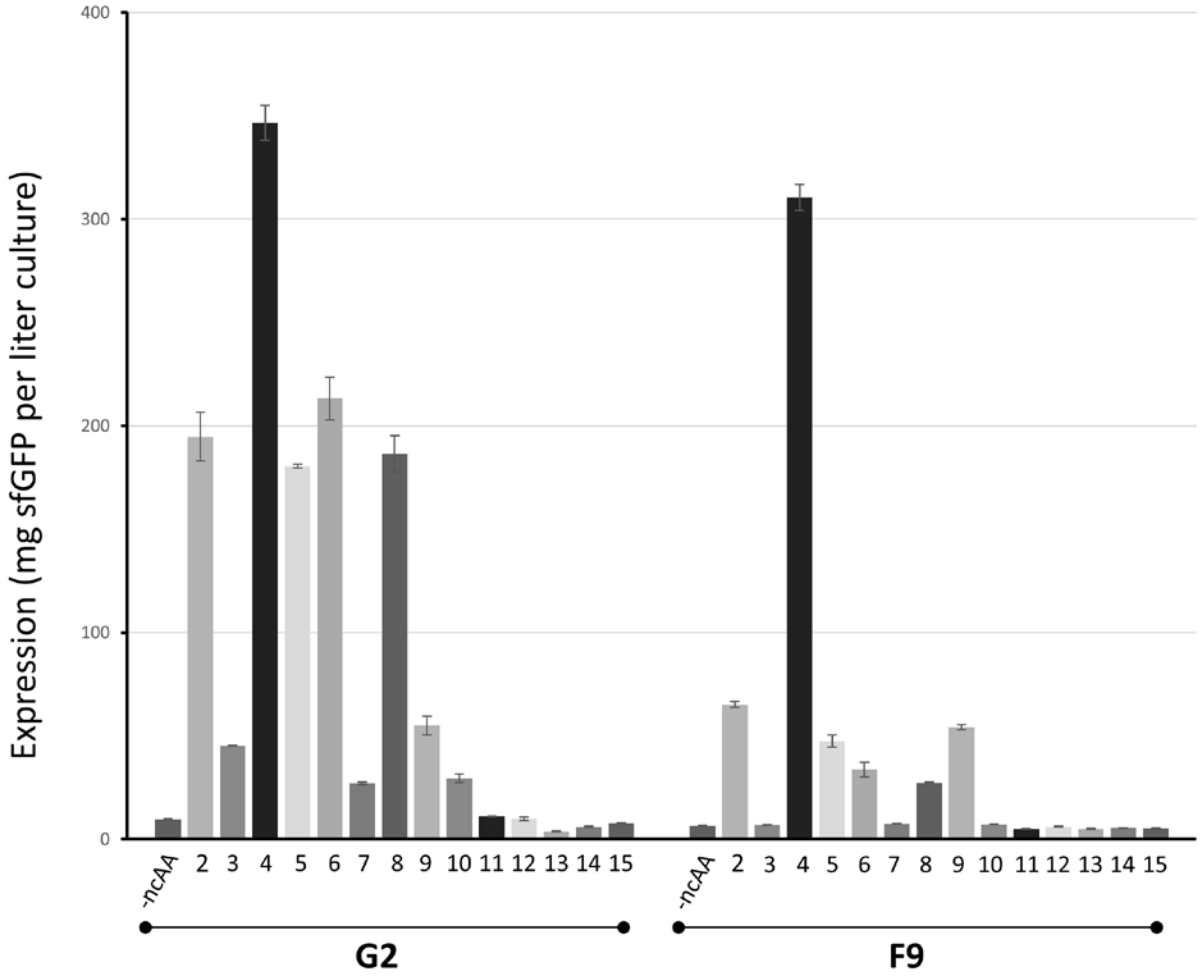

Expression (mg sfGFP per liter culture)
400
300
200
100
0
-ncAA
2 3 4 5 6 7 8 9 10 11 12 13 14 15
G2
-ncAA
2 3 4 5 6 7 8 9 10 11 12 13 14 15
F9

Figure 3.

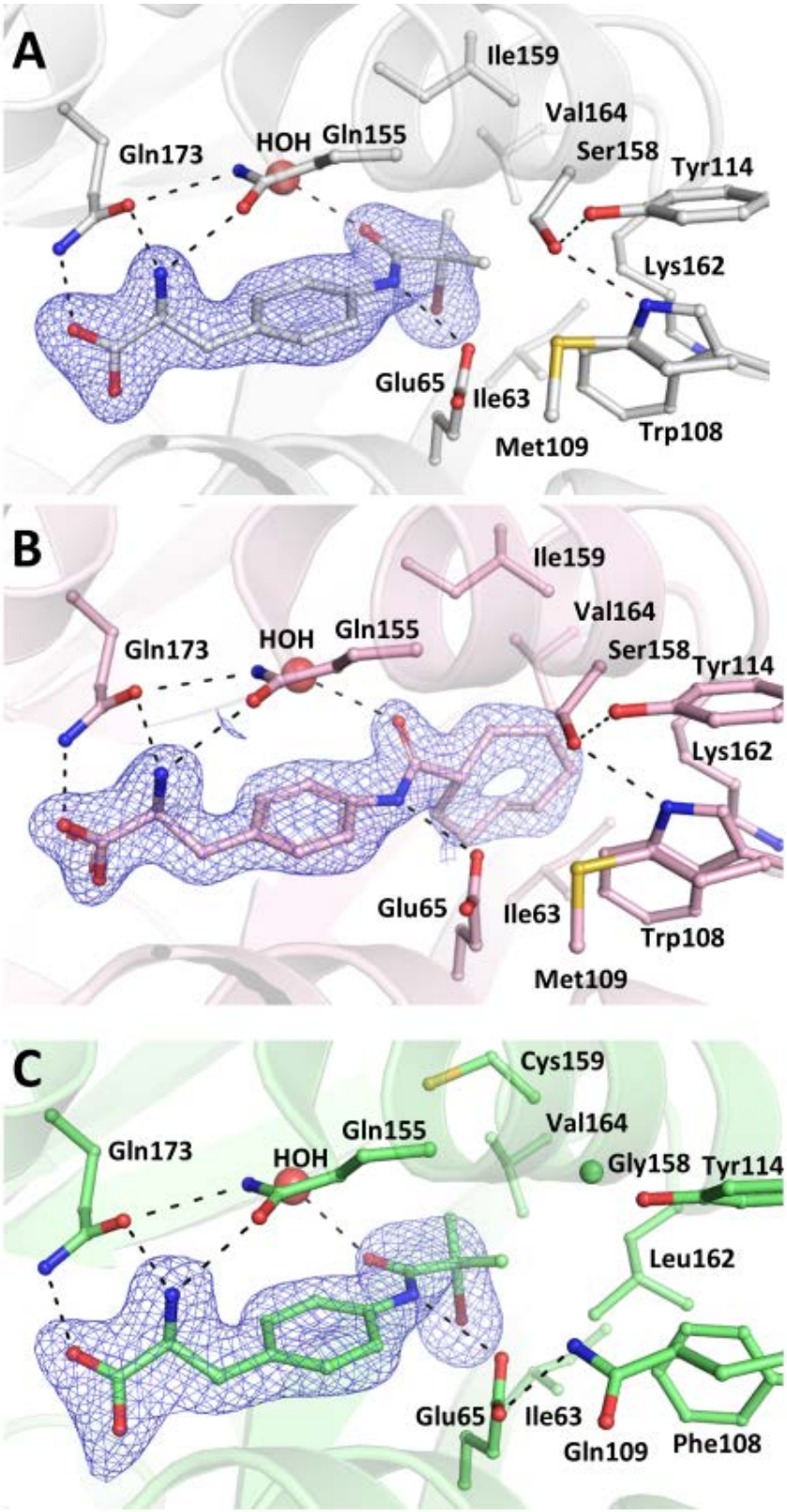

A
Ile159
Val164
Gln173
HOH
Gln155
Ser158
Tyr114
Lys162
Glu65
Ile63
Trp108
Met109
B
Ile159
Val164
Gln173
HOH
Gln155
Ser158
Tyr114
Lys162
Glu65
Ile63
Trp108
Met109
C
Cys159
Gln173
Gln155
Val164
HOH
Gly158
Tyr114
Leu162
Glu65
Ile63
Gln109
Phe108

Figure 4.

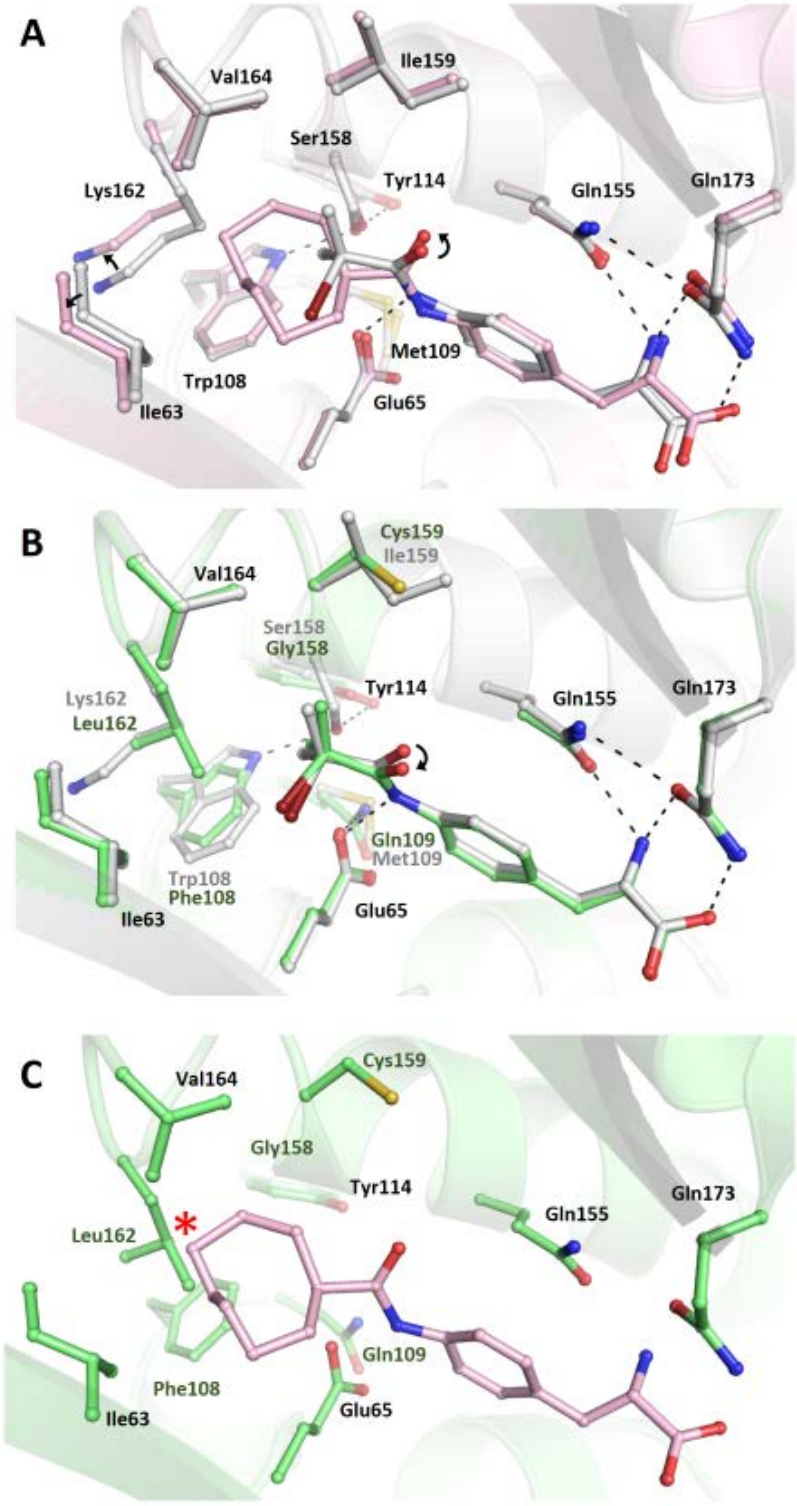
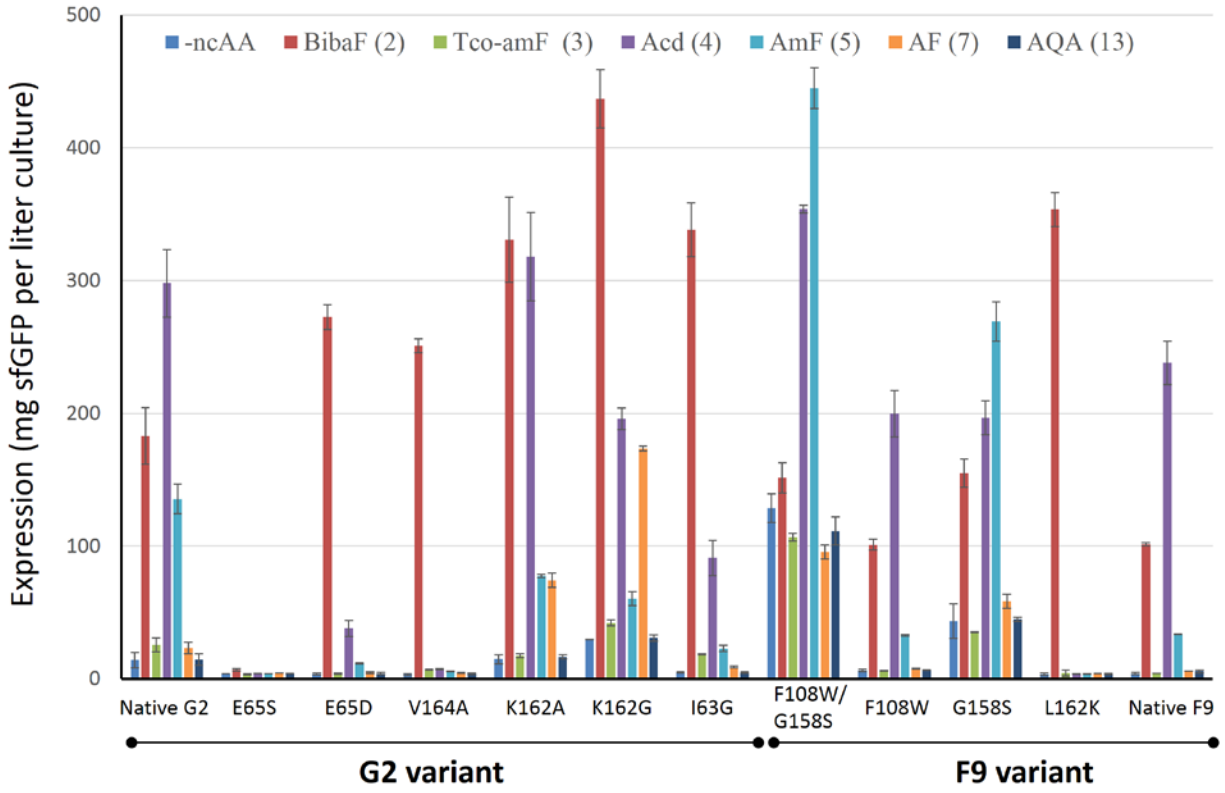

A
Val164
Ile159
Ser158
Tyr114
Gln155
Gln173
Lys162
Met109
Trp108
Glu65
Ile63
B
Cys159
Ile159
Val164
Ser158
Gly158
Tyr114
Gln155
Gln173
Lys162
Leu162
Gln109
Met109
Trp108
Phe108
Glu65
Ile63
C
Val164
Cys159
Gly158
Tyr114
Gln155
Gln173
Leu162
*
Gln109
Phe108
Glu65
Ile63

Figure 5.


500
400
300
200
100
0
Expression (mg sfGFP per liter culture)
-ncAA
BibaF (2)
Tco-amF (3)
Acd (4)
AmF (5)
AF (7)
AQA (13)
Native G2
E65S
E65D
V164A
K162A
K162G
I63G
F108W/ G158S
F108W
G158S
L162K
Native F9
G2 variant
F9 variant