# Non-canonical amino acid substrates of *E. coli* aminoacyl-tRNA synthetases

Matthew C. T. Hartman*

Department of Chemistry and Massey Cancer Center, Virginia Commonwealth University, 1001 W Main St., Richmond, VA 23220. mchartman@vcu.edu.

Abstract. In this comprehensive review, I focus on the twenty *E. coli* aminoacyl-tRNA synthetases and their ability to charge non-canonical amino acids (ncAAs) onto tRNAs. The promiscuity of these enzymes has been harnessed for diverse applications including understanding and engineering of protein function, creation of organisms with an expanded genetic code, and the synthesis of diverse peptide libraries for drug discovery. The review catalogs the structures of all known ncAA substrates for each of the 20 *E. coli* aminoacyl-tRNA synthetases, including ncAA substrates for engineered versions of these enzymes. Drawing from the structures in the list, I highlight trends and novel opportunities for further exploitation of these ncAAs in the engineering of protein function, synthetic biology, and in drug discovery.

Biography
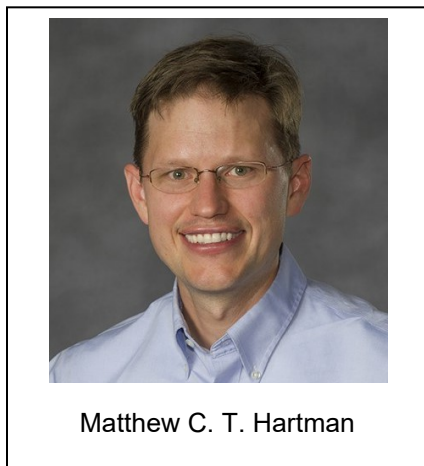
Matthew C. T. Hartman received his Ph.D. degree from the University of Michigan in 2002, studying under Professor James Coward. He was then a HHMI postdoctoral fellow with Professor Jack Szostak at Harvard University (2002-2006). Since 2007 he has been at Virginia Commonwealth University where he currently is a Professor in the Department of Chemistry and the Massey Cancer Center. His research interests focus on the discovery of natural-product like peptide inhibitors of protein-protein interactions and the delivery and activation of chemotherapeutic agents with light.

Matthew C. T. Hartman

1. Introduction.

Researchers have long been interested in the tolerance of aminoacyl-tRNA synthetase (AARS) enzymes to non-canonical amino acid (ncAA) substrates. Decades ago, biochemists like Irving Boime, Paul Berg, and Leslie Fowden investigated AARS activation and protein expression in the presence of non-canonical amino acids [1–3]. Most of these early studies focused on the ability of AARS enzymes to activate

naturally occurring ncAAs, often plant toxins. After the advent of modern molecular biology, this field

has rapidly expanded, and now hundreds of ncAAs have been described as substrates. Although there

are many reviews on AARS and ncAAs, most of these focus on stop codon suppression approaches that

use the *M. jannaschii* tyrosyl-tRNA synthetase and the *Methanosarcina barkeri or Methanosarcina*

masei pyrrolysyl-tRNA synthetases. Truly spectacular engineering feats have been accomplished with

these enzymes [4,5]. Less focus has been given to the *E. coli* aminoacyl-tRNA synthetases and their

tolerance towards ncAAs. Budisa[6,7], Schimmel,[8] and Montclare,[9] and have provided nice reviews

here, but these are now somewhat old and there has been a significant investigation of these enzymes

towards ncAAs in the intervening time. Many other recent reviews of the AARS enzymes have focused

on their biology, structures, and their use in genetic code expansion [10–15]. In this comprehensive

review, I focus specifically on the amino acid substrate specificity of wild-type and mutant *E. coli* AARS

(ecAARS) towards ncAAs and their usefulness in translation.

The promiscuity ecAARS enzymes has been harnessed towards several applications (Fig. 1). The first,

often termed selective pressure incorporation, involves global replacement of a certain amino acid in a

protein with a ncAA (Fig. 1A) [7,16]. In this case, the ncAA or a precursor is added to the growth media of

an auxotrophic strain of *E. coli* in the absence of the natural AA counterpart, prior to induction of

protein expression. *E. coli* auxotrophs for most amino acids are available [17]. The proteins expressed in

the presence of the ncAA often contain high percentages of the ncAA in place of the cognate AA;

however, the high cellular concentrations of some canonical AA competitors can be an issue [18].   A few

researchers have taken this a step further, adapting auxotrophic bacteria to tolerate incorporation of a

ncAA throughout the proteome, through gradual replacement of the proteinogenic AA in the media with

a ncAA analog (Fig. 1B) [19–21]. The evolved organisms can eventually survive on media containing only

the ncAA [21]. EcAARS enzymes have also been used as orthogonal AARS/tRNA pairs introduced into

other organisms (Fig 1C) [22–24]. To achieve this, an ecAARS and a compatible suppressor tRNA are

introduced into a eukaryotic organism. The activation of the ncAA by that AARS will allow its site-specific addition to a protein of interest in the living eukaryotic organism in response to a stop codon. Finally, the promiscuity of the ecAARS enzymes has been harnessed in in vitro selection technologies like mRNA display (Fig. 1D) [25–28]. In this case, the replacement of a proteinogenic AA with a ncAA enables the discovery of peptide ligands containing ncAAs from vast libraries [29].
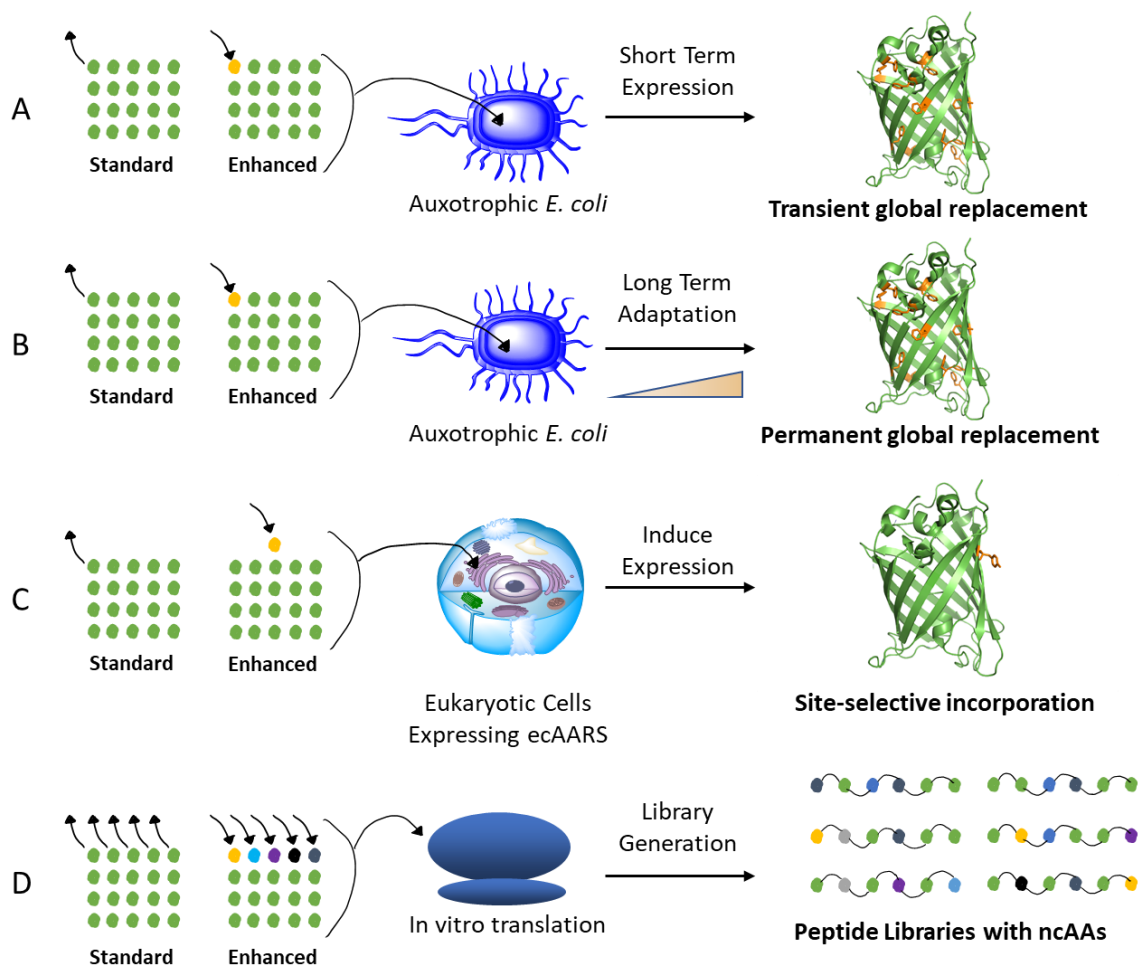
Figure 1. Uses of ncAAs charged by *E. coli* AARS enzymes. (A) Selective pressure incorporation enables protein expression wherein a single AA is substituted by a ncAA in every instance within an expressed protein. The temporary nature of the replacement reduces toxicity to the organism. (B) Permanent global replacement involves gradually allowing an organism to adapt to incorporation of a ncAA within its proteome—eventually the canonical AA can be withheld and replaced with the ncAA, leading to an organism with a permanently altered genetic code. (C) Site-selective incorporation involves incorporating an orthogonal AARS/stop codon suppressor tRNA pair from *E. coli* into another organism. Upon induction of expression, the ncAA can be incorporated selectively in response to a specific stop

codon. (D) Peptide libraries with ncAAs can be developed by using reconstituted in vitro translation systems wherein multiple canonical AAs are withheld and replaced with ncAAs.

Here, I have compiled a comprehensive list of known substrates for ecAARS enzymes, with an emphasis on their usability in translation. For those that have been directly kinetically characterized as AARS substrates, a number of assays have been developed [30]. The most common assay used to test ncAA activation with AARSs is the pyrophosphate exchange assay [31,32] which measures only the formation of the aminoacyl-AMP intermediate, not the acylated tRNA. More recently, Wilson and Uhlenbeck developed an assay that involves using a tRNA labeled at the 3'-terminal internucleotide phosphate [33]. This assay has been adopted for the detection of ncAA-tRNA directly [25,26,34–36]. Hartman and Szostak also developed a MALDI assay useful for screening ncAAs for activation [25]. This assay has the advantage that multiple ncAAs can be screened simultaneously with the full collection of AARS enzymes, but is typically used qualitatively. Finally, many ncAAs that are ecAARS substrates have never been kinetically characterized as substrates [37] but have only been assessed indirectly, through monitoring of the incorporation of the AA into a peptide or protein via translation [38]. Those ncAAs tested in living cells (e.g. via selective pressure incorporation) not only have to be charged onto tRNA, but also must be taken up by the cells [37], metabolically stable, delivered as AA-tRNAs to the ribosome by EF-Tu, and able to serve as competent substrates for the ribosomal peptidyl transfer reaction.

Considering the disparate assays used to evaluate ecAARS substrates, direct comparisons are challenging. Most of the work in this field is driven by the desire to incorporate the ncAAs into proteins or peptides, it is therefore possible to broadly classify them based this ability as I have done in Table 1. Green dots represent good substrates for the translation apparatus. These are analogs that are proven or are likely to work in translation experiments in living *E. coli.* This designation also includes engineered *E. coli* AARS/ncAA pairs that are efficient for site-selective incorporation in eukaryotic organisms. Yellow dots indicate that there may be some challenges in translation with this AARS/ncAA pair when using it in

living cells. This could be because the ncAA is a weak substrate for the AARS (some of these ncAAs have

kcat/Km values >10,000-fold lower than the canonical AA[39]), or because there are high concentrations

of the canonical AA competitor [18]. Alternatively, the AA may be only activated by a highly promiscuous

mutant AARS that would not work for protein expression in cells due to competition from cognate AAs.

These issues can often be solved in vitro by adding high concentrations of the ncAA and ecAARS and

through thorough removal of residual canonical amino acid competitors.   Finally, some ncAAs were

given a yellow dot because they appear to be outliers based on their structure. These may be false

positives, and further validation is recommended.   Red dots denote analogs that are not currently

compatible with translation; however, future engineering of the translation apparatus may enable

incorporation. These rationales are summarized in Table 1.

Table 1. ncAA classification system used in this review.

Designation	Meaning	Reasons
⬤ Green Dot	Efficient charging and incorporation into proteins in living cells.	--Viable substrate for AARS and downstream translation components
⬤ Yellow Dot	Potential issues with incorporation into proteins in living organisms, but likely to work in in vitro experiments.	--Weak ecAARS substrate --Charged by a promiscuous ecAARS --Potential false positive
⬤ Red Dot	Charged onto tRNA, but is not incorporated efficiently even in vitro.	--Not tolerated by downstream translation components

## 2. *E. coli* Aminoacyl-tRNA synthetases and their substrates

2.1 Introduction. In the following paragraphs, I describe all the known ecAARS substrates (as of Dec.

2020), discussing each of the 20 AARS enzymes individually. AsnRS and GlyRS are not included in the list

as, to my knowledge, no ncAA substrates for these enzymes have been found.

2.2 AlaRS. Due to its small binding pocket, very few ncAA substrates for AlaRS have been described (Fig.

2). EcAlaRS can charge 1-aminocyclopropanecarboxylic acid (**A1**), 1-aminocyclobutanecarboxylic acid

(**A2**), and aminoisobutyric acid (Aib) (**A3**) onto tRNA as shown using a MALDI assay [25]. Of the three, **A1** is the best substrate for AlaRS, but all 3 are difficult substrates for the translation apparatus. Provided efficiency can be improved, one could use the unique conformational properties of these amino acids to probe peptide and protein structure.

2.3 ArgRS. L-canavanine (**R1**), is a toxic plant metabolite that is an efficient Arg analog with a surprisingly low pKa (7.0) [6,40]. Its toxicity is mediated through protein incorporation[41], and it is an efficient substrate *E. coli* ArgRS [42,43] (Fig. 2). A few additional analogs have been described. ArgRS also tolerates methylation (**R2**) and hydroxylation (**R3**) at the terminal nitrogen[25]. One amidine analog, vinyl-L-NIO, **R4**, has also been described[25]. **R4** is well-known as an inhibitor of nitric oxide synthases [44].

2.4 AspRS. Three ecAspRS substrates have been described by Hartman and Szostak [25] (Fig. 2). L-threo-3-hydroxyaspartic acid (**D1**) is an excellent substrate. **D1** has is an inhibitor of glutamate transporters[45] and also has antibiotic activity [46]. Two backbone altered substrates, alpha-methyl Asp (**D2**) and N-methyl Asp (**D3**) can be charged [25], but efficient translation with these analogs is difficult.[43]
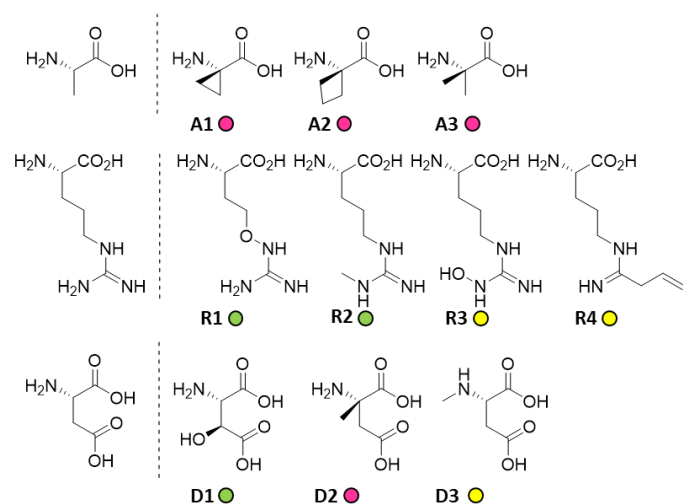

Figure 2. ncAA substrates for ecAlaRS, ecArgRS and ecAspRS.

2.5 CysRS. Only one analog of cysteine has been shown to be activated by ecCysRS, selenocysteine (**C1**) [47,48] (Fig. 3). A point mutant that activates **C1** more efficiently has also been described [47]. **C1** is a good

translation substrate, and opens up opportunities for selective chemical transformations[47] and cyclization[49,50] .

2.6 GluRS. To my knowledge the only paper looking at analogs of GluRS is by Hartman and Szostak [25]. Their MALDI screen showed that ecGluRS is quite tolerant towards substitutions at the gamma position; both fluorination (**E1-E3**) and methylation (**E4-E5**) is allowed at this position, although the methylated analogs are weaker substrates [43] (Fig. 3). Interestingly, two glutamate receptor agonist natural products, quisqualic acid (**E6**) and ibotenic acid (**E7**) also resemble Glu enough to be charged by ecGluRS. However; for any of these analogs, incorporation in living cells will be challenging due to the very high concentration of glutamic acid in the cytoplasm of cells [18], and the lack of a Glu auxotrophic strain for conducting selective pressure incorporation experiments [17].

2.7 GlnRS. To my knowledge, the substrate scope of ecGlnRS has been sparsely studied; however, near-natural analogs containing hydrazide (**Q1**), ethylated amide (**Q2**), and urea (**Q3**) groups have been shown to be substrates [25] (Fig. 3). In addition, Perona created a mutant GlnRS that can activate glutamic acid (**Q4**) [51].
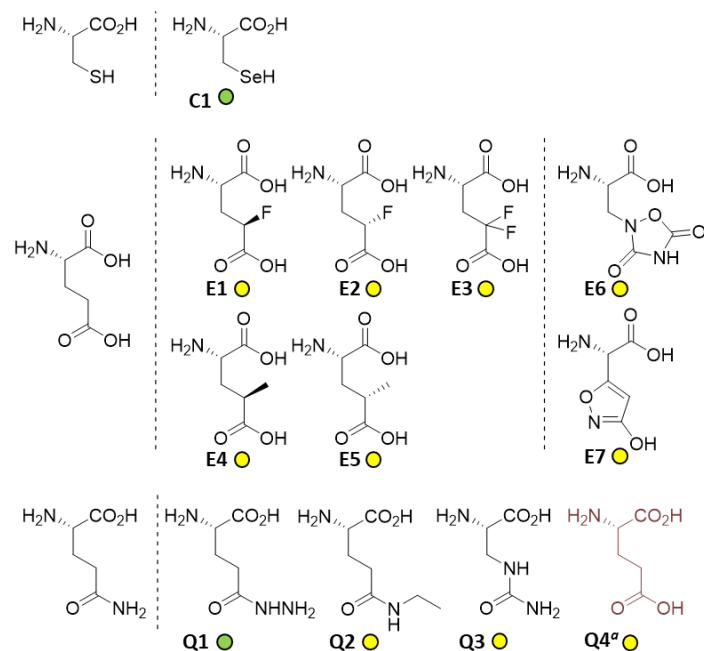
Figure 3. ncAA substrates for ecCysRS, ecGluRs, and ecGlnRS. *a*GlnRS C229R/Q255I/S227A/F233Y.

2.8 HisRS. A decent number of alternate substrates for ecHisRS have been explored (Fig. 4). The 2- and 4-fluoro analogs are excellent substrates (**H1**-**H2**) [52], as are the 1,2,3 and 1,2,4 triazoles, **H3** and **H4** [53–55]. My lab in collaboration with Ashton Cropp has exploited the lower pKa of the imidazole group of **H2** to enable selective capture of peptides containing six **H2** residues on nickel beads under low pH conditions [56]. Both *N*-methylated analogs (**H5** and **H6**) are weak substrates [57,58]. The thiazole analog of His, **H7** [25], and 2-thio-His[58], **H8**, are also substrates. ecHisRS is also able to activate both *N*-Me His (**H9**) and D-His (**H10**), but both are poor substrates for the translation apparatus [25]. Work by Söll and Umehara to create a strain of *E. coli* where ecHisRS made orthogonal through replacement by *C. crescentus* HisRS/tRNA^His [58] bodes well for future engineering of this enzyme.

2.9 IleRS. The substrate scope of IleRS has been explored, and studies have shown some surprising substrate plasticity (Fig. 4). This plasticity can be further extended by blocking the second, editing active site that is responsible for improving fidelity through hydrolysis of non-cognate AA-tRNAs (e.g Val-tRNA^Ile) [34]. All the aliphatic analogs described so far as substrates for the wild-type IleRS have beta branching, while the editing-deficient mutant will also accept some straight chain analogs [34]. A number of groups have looked at fluorinated analogs, and a wide variety of mono and tri-fluoro substituted analogs of Ile, Leu, and Val are substrates (**I1-I5**) [16,34,59,60]. Tirrell has exploited **I1** for the stabilization of β-zip peptides [60]. Koksch has extensively studied the ability of IleRS to charge and edit **I3**, which is activated by IleRS but is also edited off in a post-transfer mechanism [34]. IleRS also tolerates terminal olefin (**I6**) [61,62] and thioether (**I7**) [26] substituted analogs of Ile. Two different alkyl chlorides, **I8** and **I9** have also been described as substrates by Easton [63] and represent interesting electrophiles. Surprisingly, a MALDI screening assay also revealed that IleRS can activate the aryl amino acids phenylglycine (**I10**) and 2-thienyl glycine (**I11**), as well as cyclohexyl glycine (**I12**) [25]. Two novel

substrates include the interesting natural products furanomycin (**I13**)[64] and acevicin (**I14**). [25], which is

also a γ-glutamyl transpeptidase inhibitor.
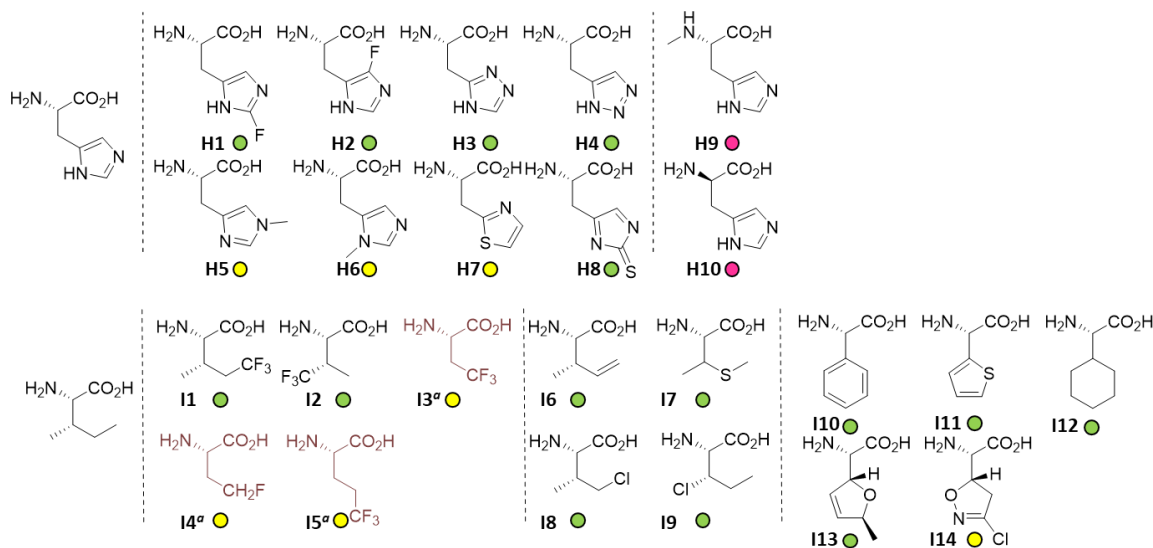
Figure 4. ncAA substrates for ecHisRS and ecIleRS. *a*IleRS T243R, D342A.

2.10 LeuRS. Wild-type and mutant *E. coli* LeuRS have been explored extensively and a huge array of AA

substrates have been uncovered. The ecLeuRS/tRNA^Leu pair has been shown to be orthogonal in

yeast,[65] which has motivated a significant number of mutagenesis studies. This enzyme is particularly

amenable to alterations in substrate specificity because it has a large and malleable binding pocket (Fig.

5A). Amino acids in the binding site surrounding Leu side chain (Fig. 5B) have been extensively mutated

leading to enzymes with remarkable activity towards ncAAs that are very different from Leu. EcLeuRS

also has an editing site, and deactivation of the editing activity has led to an expansion in the ncAA

substrate specificity [66].
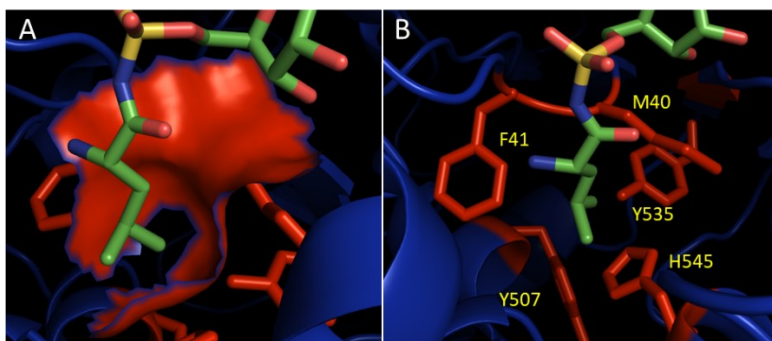
Figure 5. Active site of *E. coli* LeuRS. (A) ecLeuRS binding pocket surface is shown in red. The side chain of Leu sitting in the pocket is shown in green. (B) A sample of residues that form the Leu binding pocket that have been mutated. Images were made using Pymol from pdb 1H3N [67].

A large number of near-cognate amino acids have been explored (Fig. 6). Mono-, Tri- and Hexa-fluoro Leu (**L1-L3**) have been investigated both in in vitro translation and in auxotrophic bacteria [25,26,34,63,68,69]. Coiled, coil proteins containing **L3** are significantly more stable to denaturation than those containing Leu [68]. The wild-type enzyme can also activate t-butyl Leu (**L4**)[25,69], cyclopentyl-alanine (**L5**)[25] and dehydro variants **L6** and **L7** [63,70–72]. An editing deficient version (T252Y) has been exploited by Tirrell to introduce a number of additional alkenes (**L8-L9**), alkynes (**L10-L11**) and straight chain alkanes (**L12-L13**) [66]. Each of these can be incorporated in selective pressure incorporation experiments. An evolved ecLeuRS mutant described by Schultz that can activate amino acids with long-chain hydrocarbon side chains (**L14-L16**) has also been developed [24,73].

EcLeuRS has also been fruitfully studied for the incorporation of a wide variety of interesting functional groups, including a ketone (**L17**) [71], a diazirine photoaffinity label (**L18**)[43] , and a tertiary amine analog where the methine carbon is replaced with a nitrogen (**L19**) [25]. Alkyl chlorides **L21** and **L22** have been shown to be substrates by Easton [61]. Mutant LeuRS's can also activate methionine analogs **L22** and **L23** and even amino acids with long-chain thiols (**L24-L25**) [74].

The specificity of the LeuRS active site has been pushed even further through directed evolution with some spectacular results. A wide variety of aryl amino acids, even those with very large side chains have been uncovered. *p*-Methoxy phenylalanine (**L26**) [73], and photocaged analogs of serine (**L27**) [75], cysteine (**L28-L30**) [65,76,77] and selenocysteine (**L31**)[76,78,79] are substrates for various mutants. Schultz has used the AARS that activates **L27** and its tRNA as an orthogonal system in yeast to control protein serine phosphorylation in a spatiotemporal manner [75]. Mutant ecLeuRS enzymes activating naphthalene (**L32**) [80], naphthalene ketone (**L33**) [80,81] and even the very large dansyl group (**L34**) [73,82,83] have been described; each of these amino acids can serve as a useful fluorescent reporter within proteins. A mutant activating the ferrocene amino acid **L35** has also been reported [84].

Finally, wild-type LeuRS has been shown to activate various backbone-modified analogs, including the alpha-alpha disubstituted 1-aminocyclohexane and cyclopentane carboxylic acids (**L36-L37**) [25], and alpha-hydroxy substituted compound **L38** [72]. *N*-methyl Leu (**L39**) and β_3-Leu (**L40**) have also been activated, although they are poor substrates in translation [25,43].
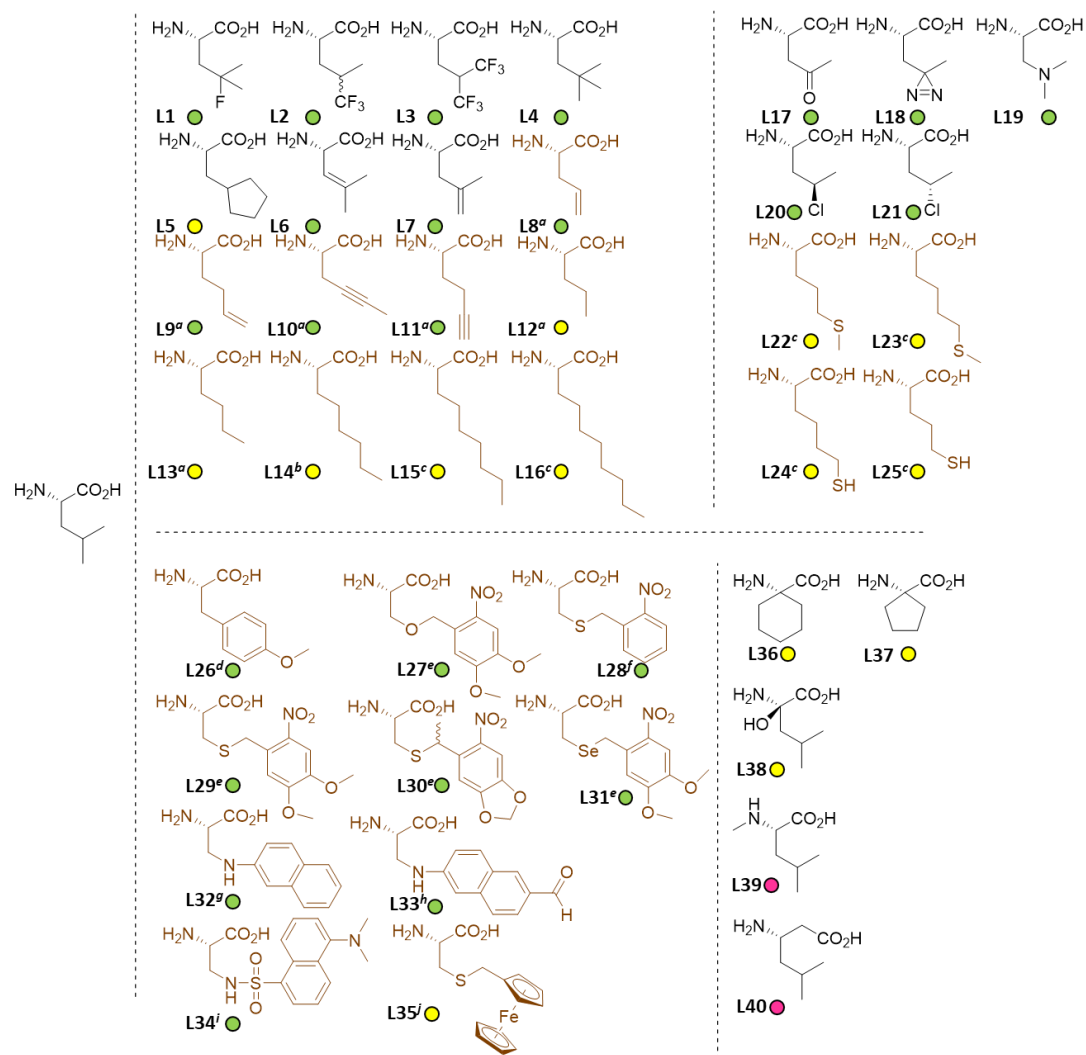
Figure 6. ncAA substrates for ecLeuRS. [a]LeuRS D252A. [b]LeuRS M40V, L41M, Y499L, Y527L, H537G, [c]LeuRS M40I, Y499I, Y527A, H537G. [d]LeuRS M40L, L41E, Y499R, Y527A, H537G. [e]LeuRS M40G, L41Q, Y499L, Y527G, H537F. [f]LeuRS M40W, L41S, Y499I, Y527A, H537G. [g]LeuRS M40G, L41P, Y499G, Y527A, H537T. [h]LeuRS L38F, M40G, L41P, Y499V, Y500L, Y527A, H537G, L538S, F541C, A560V. [i]LeuRS M40A, L41N, T252Y, Y499I, Y527G, and H537T. [j]LeuRS M40G, L41K, Y499S, Y527A, H537G.

2.11 LysRS. EcLysRS has been sparsely explored with regards to ncAAs (Fig. 8, below). The conservative substitutions giving rise to thialysine (**K1**)[25,26] and selenalysine (**K2**)[85] are tolerated. Seebeck and Szostak used analog **K2** as a surrogate for dehydroalanine, leading to lantiboitic-like peptide libraries [86]. LysRS also tolerates introduction of a trans double bond between the gamma and delta carbons (**K3**) [25].

LysRS can activate the β3 analog of lysine (**K4**), N-methyl lysine (**K5**), and D-lysine (**K6**), although each of these are challenging to incorporate into peptides or proteins [25,43].

2.12 MetRS. The substrate scope ecMetRS has been investigated comprehensively, finding analogs consistent with its narrow but deep active site pocket (Fig 7A). Mutagenesis of residues around this pocket (Fig 7B) has led to expanded substrate scope.
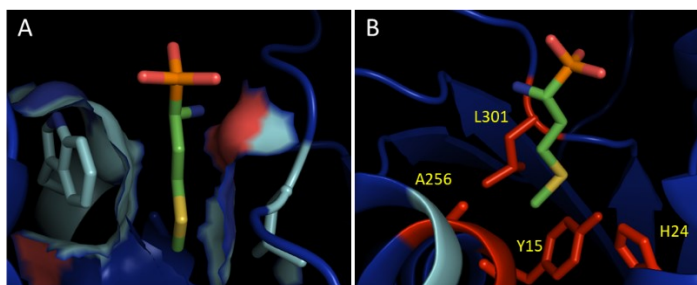

Figure 7. Active site of *E. coli* MetRS. (A) The ecMetRS binding pocket surface is shown in blue. The side chain of Met sitting in the pocket is shown in green and yellow (for sulfur). (B) Residues that form the Met binding pocket that have been mutated. Images were made using Pymol from pdb 1P7P [87].

Wild-type and mutant ecMetRS enzymes tolerate most non-branched amino acids with side chains containing 3-6 atoms (Fig. 8). This includes hydrocarbons (**M1-M3**) [25,39,69,88–92], terminal or *trans*-alkenes (**M4-M6**)[25,39,90], terminal or internal alkynes (**M7-M10**) [39,88,89,91,93–95] , and hydrocarbons terminating in a trifluoromethyl group (**M11-M12**) [25,96,97]. Cyclopropylalanine (**M13**) is also a substrate [25].

Various heteroatom substituted analogs have also been described as substrates. Selenium, oxygen, and even tellurium is tolerated in place of the sulfur (**M14-M16**) [98,99], The sulfur can also be alkylated with fluoromethyl groups (**M17-M18**)[87] or longer chains including the ethyl (**M19**)[25,69,100] or allyl (**M20**) [101] groups. S-allyl cysteine (**M21**) is also a weak substrate for MetRS [101], and this leads to poor incorporation efficiency (it's competence in translation in *E. coli* has been confirmed with an engineered PylRS [102]). S-allyl homocysteine (**M21**) is a better substrate[25,101], and is a powerful reagent for site-

selective labeling via cross metathesis or click chemistry [101]. Azide **M22** has been exploited extensively for labeling, and by Link and my group for the creation of cyclic peptides or proteins [25,26,28,69,88,92,103,104]. Tirrell and Link have investigated mutant MetRS enzymes that can activate the longer-chain azide **M23** [105–107] or shorter chain alkyne **M7** [93] to avoid competition with endogenous Met for cellular labeling [93]. Azide **M24** is a weaker substrate [108]. Surprisingly ecMetRS can also tolerate a few amino acids with polar side chains; the methyl esters of aspartic (**M25**)[25] and glutamic (**M26**)[25,69] acids are substrates. EcMetRS has been shown to activate the β_3-Met analog **M27** although it is a poor substrate [25,109]. The hydroxy acid **M28** is also a substrate[25] and can be incorporated into peptides, but because this analog does not contain an alpha-amino group, it is not processed efficiently by the translation initiation machinery [43].
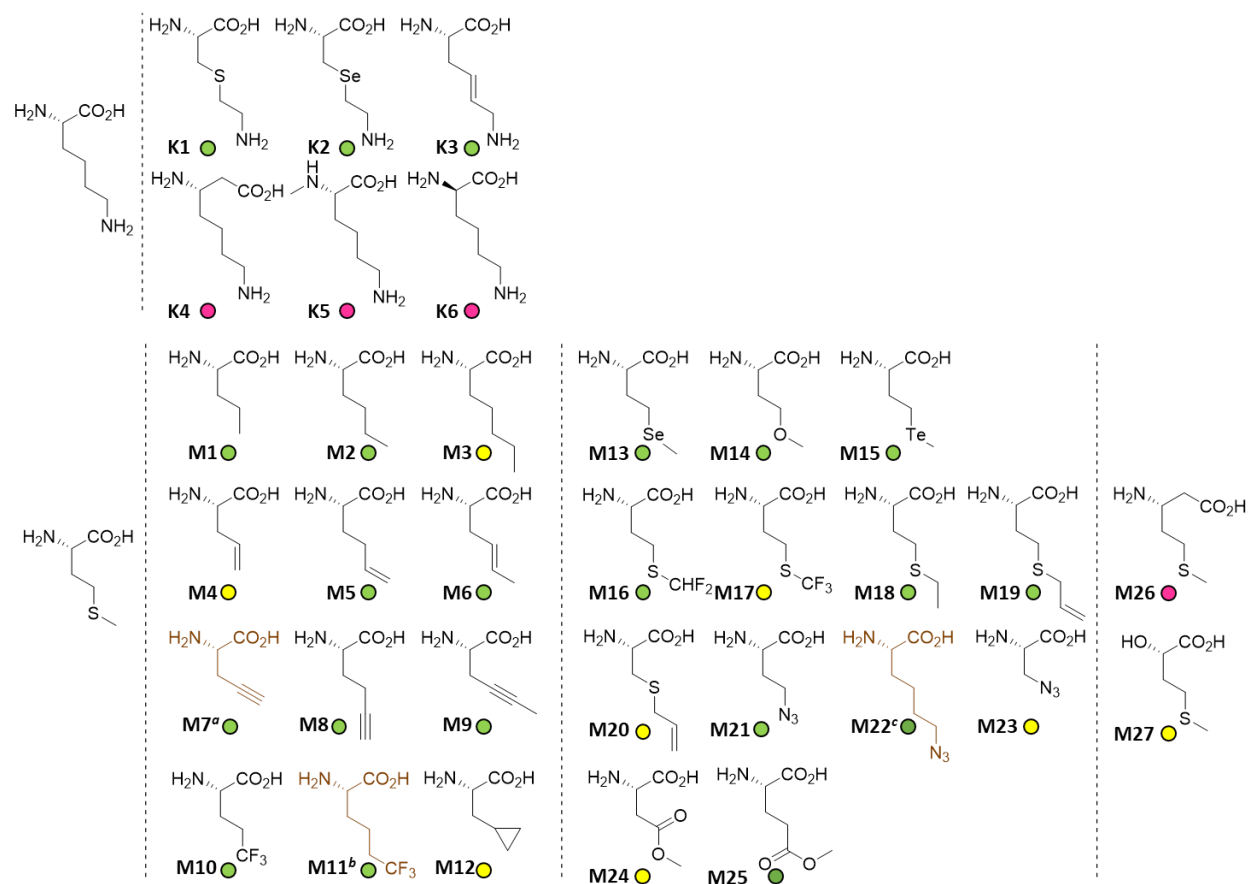
Figure 8. ncAA substrates for ecLysRS and ecMetRS. [a]MetRS L13P, A256G, P257T, Y260Q, H301F. [b]MetRS L13G or MetRS L13P, Y260L, H301L. [c]MetRS L13S, Y250L, H301L. [c]MetRS L13G or MetRS L13P, Y260L, H301L.

2.13 PheRS. Wild-type ecPheRS has been used to incorporate a wide variety of aromatic ncAAs. Early on, Hennecke described a A294G mutant which has an expanded binding pocket (Fig. 9A) near the binding site of para position of the phenyl ring [110] (Fig. 9B). A further mutant described by Tirrell (T251G, A294G), has an even more open binding pocket [111].
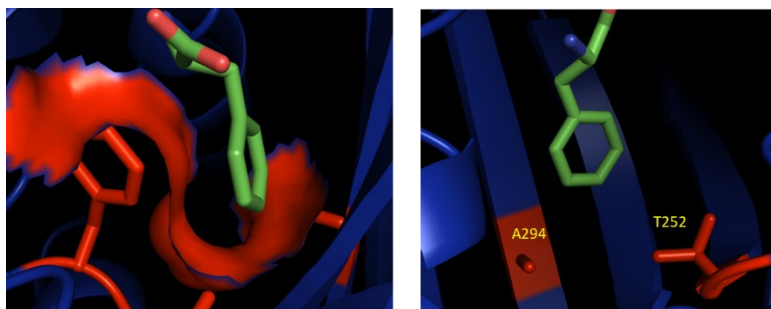
Figure 9. Active site of *E. coli* PheRS. (A) The ecPheRS binding pocket surface is shown in red. The side chain of Phe sitting in the pocket is shown in green. (B) Two residues that form the Phe binding pocket have been mutated. Images were made using Pymol from pdb 3PCO [112].

A wide variety of 4-substituted Phe analogs have been described (Fig. 10). The wild-type enzyme can only activate 4-fluoro Phe (**F1**)[63,113–116], but the A294G mutant allows for the incorporation of many analogs. The halogen series can be activated (**F2-F4**) as well as the 4-cyano (**F5**), azido (**F6**), and alkyno (**F7**) derivatives [25,94,103,117,118]. The azido Phe, **F6**, can be used as a photoaffinity label [119], and alkyne **F7** has been used along with click chemistry for protein labeling or stapling[94,103] or, by us for the creation of diverse bicyclic peptide libraries [28,120]. 4-Nitro Phe (**F8**) is also a weak substrate for the A294G mutant [25]. The T251G, A294G double-mutant is also able to use the ketone containing **F9** which has been used as a reactive and selective handle for protein derivatization [111]. Wang and Matthews[114] have recently investigated 3-substituted analogs, showing the wild-type enzyme to be quite tolerant to substitution at the 3 position. Analogs include the fluoro (**F10**) [70,92,114,115], chloro (**F11**) [114], methyl (**F12**)[114], hydroxy groups (**F13**)[114] as well as 3,3-disubstituted analogs **F14**, **F15**, and **F16** [114]. 2-substituted analogs, **F17** [25,26,40,70,115] and **F18**[114] have also been shown to be substrates, although substitution at this position is less tolerated than the 3-position [114].

The wild-type enzyme is also able to tolerate aliphatic substitutions on the β-carbon, either methyl or hydroxy substitutions (**F19**-**F21**)[25,121]. Various researchers have also explored changing the type of aromatic ring. Pyridines (**F22-F24**)[69,117], thiophenes (**F25**-**F26**)[25,69,121,122] and even an acyclic

trisubstituted olefin (**F27**)[121] can be used in place of benzene. The A294G and T251G/A294G mutant is also able to activate the interesting benzofuran analog **F28** as shown by Nielsen[123]. Finally, wtPheRS is also able to activate the β-2 analog of phenylalanine (**F29**) with reasonable efficiency, although it is a poor substrate in downstream translation [25].
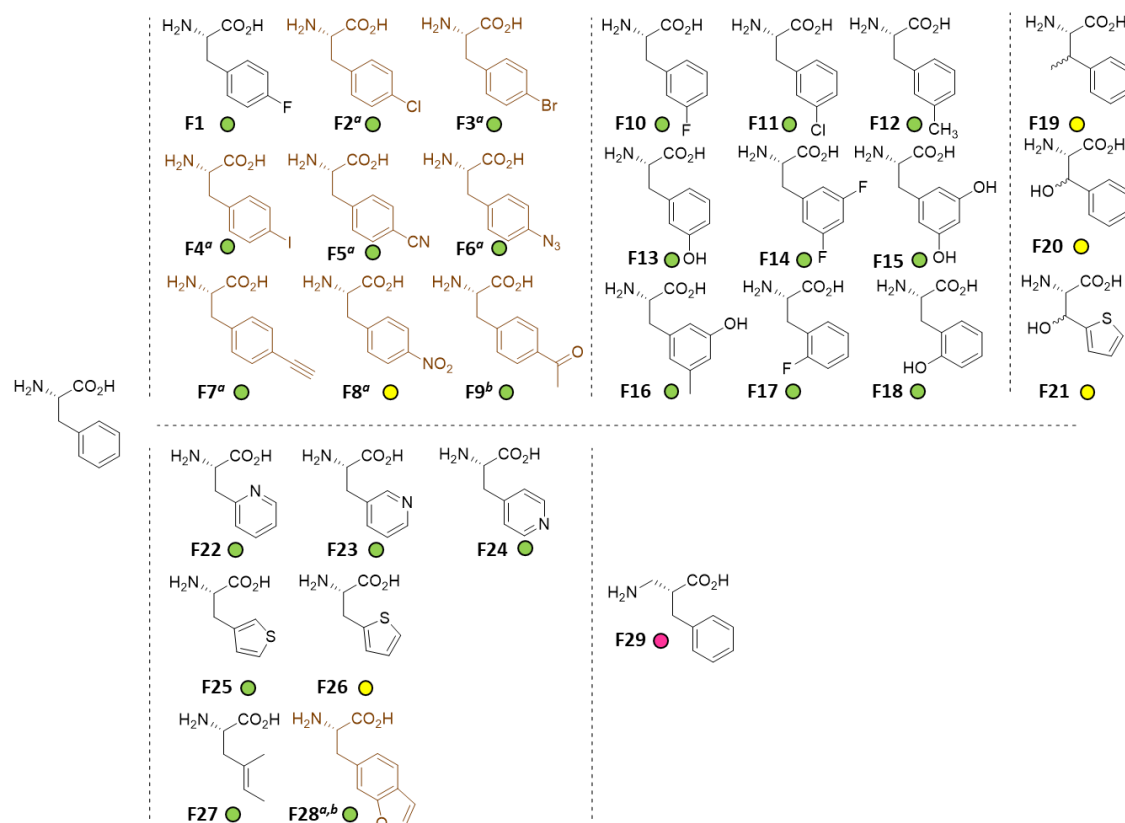
Figure 10. ncAA substrates for ecPheRS. [a]PheRS A294G. [b]PheRS T251G, A294G.

2.14 ProRS. The substrate specificity of ecProRS has been explored (Fig. 11); much of the work has been done by Conticello[124,125]. Most of the analogs consist of conservative substitutions around the ring. The enzyme tolerates fluorination at the 3 and 4 positions (**P1-P4**) [69,89,124–127]; these analogs have different ring pucker characteristics which makes them interesting structural probes [124,126]. Both 4-hydroxy proline diastereomers (**P5-P6**) are also substrates [69,89,92,124,128]. Alpha methyl Pro (**P7**) is also a substrate for ProRS, but downstream translation with this analog is challenging [25]. The 4,4-difluoro

analog, **P8**, has also been studied [124]. Alterations within the ring itself have been explored. 3,4-

dehydro Pro, **P9** [25,124], as well as both thiazolidine ring-analogs **P10** [25,69] and **P11** [124,129] are substrates.

Ring size has also been varied.   The plant toxin azetidine-2-carboxylic acid (**P12**) [130] is a well-known

substrate [25,124]. Conticello created the C443G mutant which is able to activate the ring expanded

piperidine analog **P13** [124]. Both **P12** and **P13** are interesting conformational probes [124].

2.15 SerRS. EcSerRS substrates have not been explored extensively. β-methyl amino alanine (BMAA),

S1, is a toxic amino acid [131] that is known to be a substrate [132] (Fig. 11). Recently Söll uncovered a

couple of putative ecSerRS substrates in a translation screen (**S2** and **S3**) [38]. These new substrates

suggest that the SerRS enzyme can tolerate amino acids with substitutions at the β-carbon, and this

could serve as an interesting avenue for future inquiry.

**2.16 ThrRS.** To my knowledge, the only alternate ThrRS substrate known is β-hydroxynorvaline (**T1**)

[25,26,40,133] (Fig. 11).
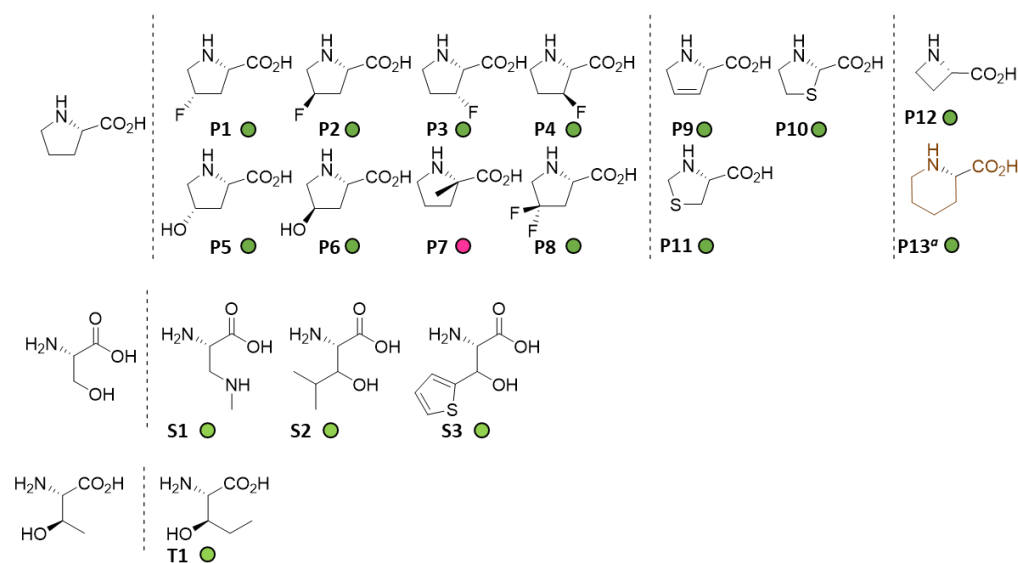
Figure 11. ncAA substrates for ProRS, SerRS, and ThrRS. ᵃProRS C443G.

2.17 TrpRS. A wide variety of Tryptophan analogs are substrates for ecTrpRS, and most are also easily

incorporated into translated peptides or proteins in place of Trp.   These analogs are easily accessed

enzymatically [134], or, if working in auxotrophic bacteria, through feeding the modified indole into the media [135]. ecTrpRS has a large and moldable binding pocket (Fig. 12), and directed evolution of ecTrpRS has been explored by Chatterjee[136] who has established it as an orthogonal synthetase in mammalian systems.
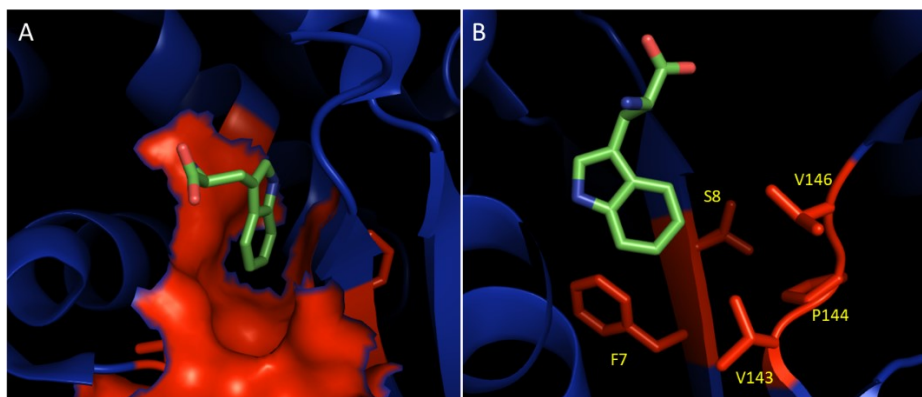
Figure 12.   Active site of *E. coli* TrpRS. (A) The binding pocket surface of TrpRS is shown in red. The side chain of Trp sits in the pocket as shown in green. (B) Residues that form the TrpRS binding pocket thathave been mutated are shown in red with corresponding numbering. Images were made using Pymol from pdb 5V0I.

Much of the work testing Trp analogs has been performed by Budisa. Substitution at the 4-position is tolerated for fluoro[89,137–141], hydroxy[142], methyl[139,143,144] and amino[143,144] groups (**W1**-**W4**) (Fig. 13). Wild-type EcTrpRS tolerates a wide variety of 5-subtitutions including fluoro[25,69,137–141], bromo[25,136], hydroxy [25,89,136,140,145], and methoxy substituents [25,136] (**W5-W8**). Mutagenesis has permitted further substitutions at the 5-position including the azide (**W9**), propargyloxy (**W10**) and amino groups (**W11**)[136]. The 6-position can also be substituted (**W12-W14**) [137,139–141,146]. 7-substitued analogs have not been extensively explored, but the 7-fluoro Trp is a substrate (**W15**)[69,139]. The tetrafluoro analog (**W16**) was also reported as a substrate, but incorporation into proteins has not been established [147]. Substitutions on the indole nitrogen have not been well tolerated [147], although Söll reported that the *N*-formyl and *N*-boc analogs could be incorporated using stop-codon suppression (**W17-W18**) [72]. Since

incorporation was not validated by MS, it is possible that they were metabolized to Trp or were contaminated with a small amount of Trp, giving false positives. To my knowledge, substitutions at the 2-position have not been explored.

The indole ring itself can be extensively altered by substitutions. The 4,5, and 7 positions have been replaced with nitrogen (**W19-W21**) [25,26,140,141,145,146,148,149]. 6-aza incorporation (via addition of the corresponding indole) was tested in auxotrophic bacteria, but incorporation failed [148]. 2-aza tryptophan (**W22**) has conflicting reports about its suitability as a translation substrate [150–152]. Chou and coworkers have recently investigated the 2,6 and 2,7 di-aza substrates (**W23** and **W24**), as fluorescent probes of hydration [153–156]. The benzene ring in the imidazole can be substituted with a 5 member-ring containing a sulfur or selenium atom (**W25-W28**) [139]. The indole nitrogen can also be substituted with a sulfur[25], but this thiophene analog (**W29**) is a mediocre substrate in translation [43]. Surprisingly, S-benzyl cysteine, **W30**, and 2-iodo Phe (**W31**) also are substrates [72].

Finally, a few backbone analogs of Trp have been shown to be substrates including N-Me Trp (**W32**) [25,43], the alpha-hydroxy acid analog of Trp (**W33**) [72], and D-Trp (**W34**)[157]. **W33** is a good substrate for the translation apparatus [72].
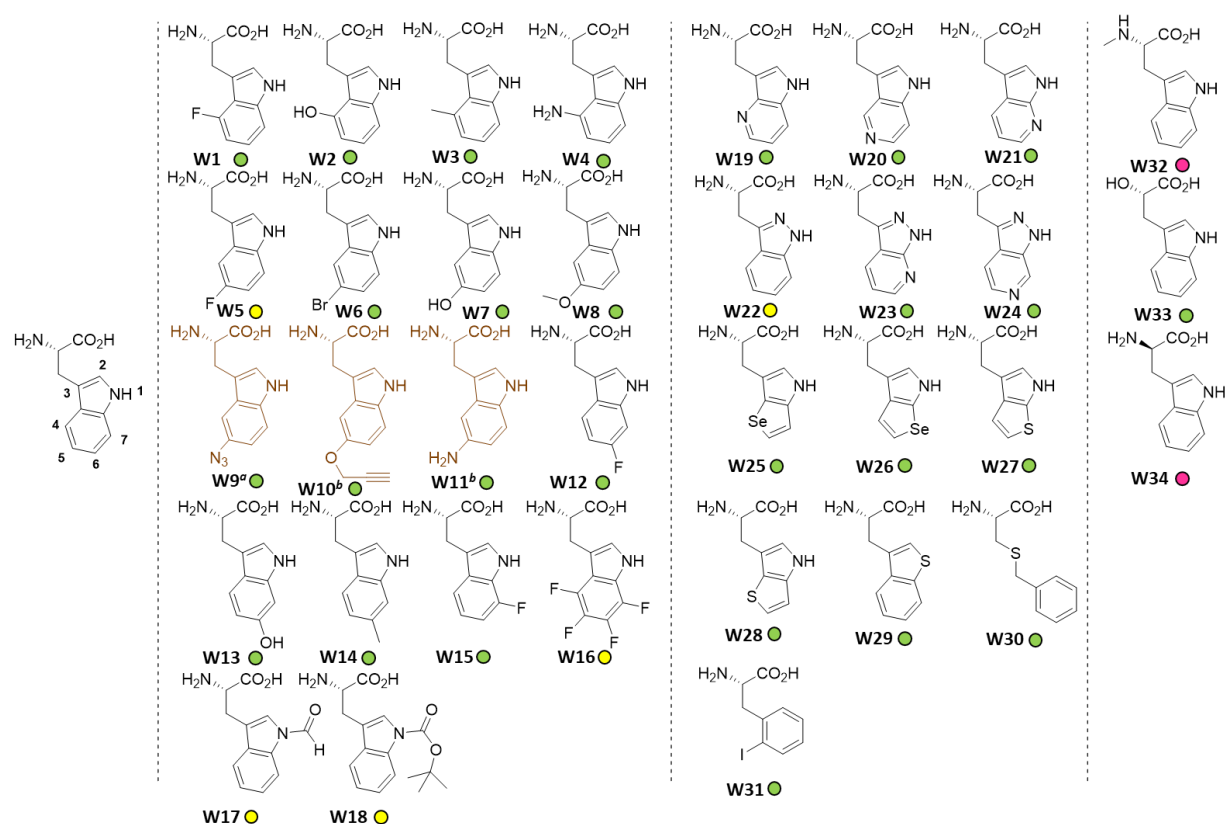
Figure 13. ncAA substrates for ecTrpRS. [a]TrpRS S8A, V144S, V146A. [b]TrpRS S8A, V144G, V146C.

2.18 TyrRS. The plasticity of the ecTyrRS active site (Fig. 14) has been exploited extensively, and a wide variety of AAs, many of which do not significantly resemble tyrosine, can be incorporated. The orthogonality of ecTyrRS/tRNA[Tyr] in yeast and mammalian systems[158] has served to motivate the directed evolution of a wide variety of mutants.
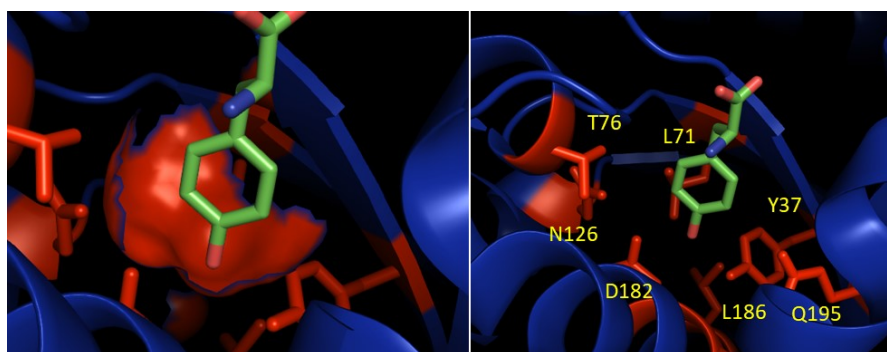
Figure 14. Active site of *E. coli* TyrRS. (A) The ecTyrRS binding pocket surface is shown in red, with the Tyr side chain shown in green. (B) A sampling of residues that form the Tyr binding pocket that have been mutated are shown in red. Images were made using Pymol from pdb 1X8X [159].

The first group of analogs consists of amino acids that retain the 4-hydroxy group but have additional substitutions around the ring (Fig. 15). A wide variety of 3-substituted amino acids (**Y1-Y9**) are substrates[25,40,69,70,92,114,127,160–163], although those with larger substituents (**Y3**, **Y4**, **Y8**) require a mutated version of the AARS (Y37V, Q195C) [162,163]. Several labs have explored 2-fluoro Tyr (**Y10**) [7,92,161], and recently , Wang and Matthews have shown the wild-type enzyme to be fairly permissive to tyrosine analogs substituted at the 2-position (**Y10-Y14**) [114]. Multiple halogenations are tolerated (**Y15-Y16**)[72,161], as well as changing the benzene for a pyridine ring (**Y17**) [164].

A wide variety of mutants have been described that can incorporate analogs that lack the 4-hydroxy group. These include a number of 4-substituted analogs including methoxyl (**Y18**)[22,158,165,166], and nitro (**Y19**)[72] compounds. Various reactive handles have also been introduced including the 4-acetyl (**Y20**) [22,158,165,167], iodo (**Y21**) [22,165,166], and azido (**Y22**) [22,158,165,166,168] compounds. Chatterjee developed a ecTyRS mutant that can activate sulfated tyrosine (**Y22**) and has used it for the overexpression of sulfated proteins both in mammalian cells and *E. coli* [169]. Other O-alkylated compounds include alkyne **Y24** [22,165,166,168], and chloroethyl **Y25** [22]. Mutant synthetases can activate the boronate (**Y26**) [22] and the photoaffinity label trifluoromethyldiazirine (**Y27**) [170]. The wild-type enzyme can also activate the trifluoro and dichloro analogs **Y28** and **Y29** [72]. A TyrRS mutant that can activate benzophenone-containing derivatives has also been developed [158,165–167] for use as an orthogonal AARS/tRNA pair in yeast. Mapp and coworkers have explored a variety of substituted benzophenones (**Y30-Y37**) for this mutant, noting significant differences in photoaffinity labeling efficiencies [171,172]. Brauchi recently exported mutations from a *M. janaschii* enzyme into EcTyrRS, enabling a mutant that can activate

coumarin amino acid **Y38** [173]. Finally, D-Tyrosine (**Y39**) is a surprisingly good substrate [2,32]; however it

is hydrolyzed off of the tRNA by a selective de-acylase to prevent toxicity [157].
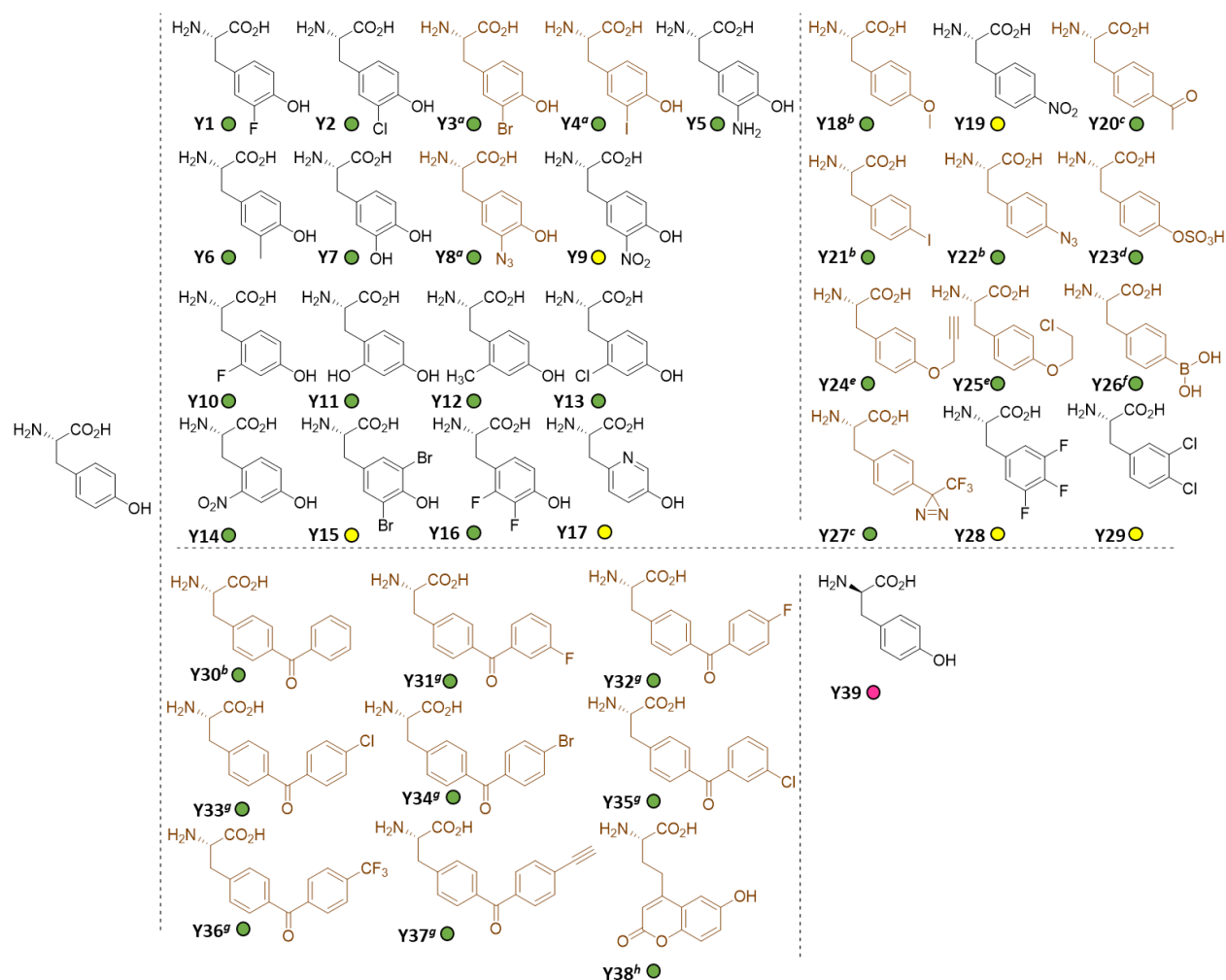
Figure 15. ncAA substrates for ecTyrRS. [a]TyrRS Y37V, Q195C. [b]Multiple mutants reported[158] [c]TyrRS Y37G, D165G, D182G, F183M, L186A. [d]TyrRS L71V, D18G. [e]TyrRS Y37V, D165G, D182S, F183M, L186A. [f]Multiple mutants reported[169]. [g]TyrRS Y37G, D182G, F183Y, L186M. [h]TyRRS Y37G, L56E, L71H, T76G, S120Y, A121H, D182G, F183I, L186G.

2.19 ValRS. EcValRS has been extensively studied for the incorporation of ncAAs (Fig. 16). In particular,

the editing deficient mutant T222P, long known to enable synthesis of Thr-tRNA^{Val} and Cys-tRNA^{Val} [174],

has proven to charge a very diverse range of substrates.

A wide variety of straight and branched chain aliphatics are substrates for wild-type or mutant ValRS (**V1**-**V4**) [25,43,175,176]. The hydrocarbon side chains can also be substituted with cyano groups (**V5**), azide (**V6**), hydroxyl groups (**V7**), a thioether (**V8**), and ethers (**V9**-**V10**) [25,175]. Early experiments by Berg and Dieckmann [176] showed that chlorobutyric acid derivatives **V11** and **V12** are decent substrates; and more recently Easton showed that these interesting AAs could be incorporated into proteins [61].

A wide variety of mono- and tri-fluoro substitutions are also good substrates (**V13**-**V16**) [25,34,40,43,60,177]. Interestingly, **V13**, is a substrate, but not the C-3 epimer [177]. My lab has shown that ValRS and its editing-deficient mutants are able to charge a wide variety of β-amino acids, including β3 analogs of both alanine (**V17**) and valine (**V18**), as well as cyclic beta amino acids containing cyclohexane and cyclopentane rings (**V19**-**V21**) [25,175]. These analogs are known to exert unique conformational constraints into peptides, although their incorporation in translation requires significant engineering of the translation apparatus [178]. Szostak reported that cyclic alpha-alpha disubstituted amino acids **V22** and **V23** were good substrates for the wild-type enzyme [25], and used **V23** within a diverse peptide library [27]. More recently my lab showed that the editing-deficient mutant can activate alpha-methyl analogs of alanine (**V24**)[175], serine (**V25**) [175], and cysteine (**V26**) [179]. The latter when incorporated into a cyclic peptide can serve as a suitable replacement for the standard hydrocarbon staple for promoting alpha-helical peptide conformations [179].   Finally, *N*-Methyl valine (**V27**) can be charged although it is a poor substrate for downstream translation [25,43].
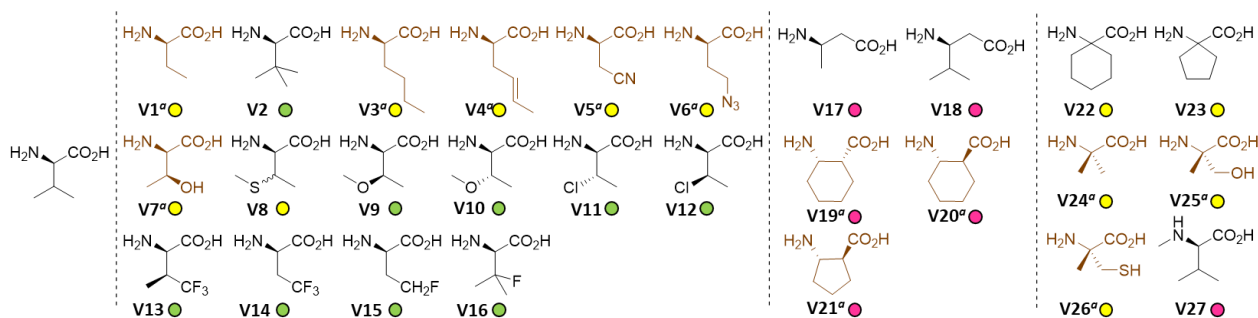
Figure 16. ncAA substrates for ecValRS. ^{a}ValRS T222P.

3. Trends and applications of ncAA substrates

The list of ncAA substrates for ecAARS enzymes can be taken in several productive directions. For each, I highlight a particularly exciting or innovative application.

3.1 Fluorination. A significant fraction of the ncAAs listed in these charts are fluorinated amino acids. The small size of the fluorine atom makes it an excellent substitute for a hydrogen. Budisa's excellent review on global replacement with fluorinated ncAAs in proteins highlights the effects of these molecules on protein structure, catalysis, and their usefulness as NMR and vibrational spectroscopic probes [180]. Pomerantz has pioneered drug screening of protein binding by NMR using proteins globally substituted with fluorinated Trp and Tyr residues [181]. In a notable recent paper, Nash and coworkers investigated the impacts of the presence of trifluoro leucine on the stability of protein interfaces using single molecule force spectroscopy [182].

3.2 Bio-orthogonal chemistry. Many of the amino acids described in the preceding lists have novel functional groups that can be used in bio-orthogonal reactions. Notable functional groups include hydrazides (**Q1**), ketones (**L31, F9/Y19, Y29-Y35**), azides (**M20-M21**, **F6/Y22**, **W9**, **Y8**), terminal alkynes (**L11/M8**, **M7**, **F7**, **W10**, **Y24**), alkyl chlorides (**I9**, **I10**, **L20**, **L21**, **Y25**), aryl iodides (**F4/Y21**, **W31**, **Y4**), an aryl boronate (**Y26**), and an *S*-allyl group (**M18**). Azides and alkynes have, in particular, been exploited using copper-mediated azide-alkyne cyclization or copper free click experiments for selective labeling [94,103,183]. The majority of these experiments have focused on analogs **L11/M8** and **M21** as methionine surrogates. These have been extensively exploited to monitor the production of new proteins inside cells and organisms through adding a pulse of the ncAA. This technology, developed by Tirrell and Schuman is known as BONCAT (Bio-orthogonal non-canonical amino acid tagging) [184]. The more recent development of Trp and Tyr analogs containing azides and alkynes (**W9, W10, Y8, Y22, Y24**) opens up

new opportunities for selective labeling of aryl residues. The other functional groups in this list have not been extensively exploited and offer new avenues for protein labeling and peptide library cyclization.

3.3 Fluorescence. Although native tryptophan is a weak fluorophore, fluorescent amino acids have been widely pursued for their enhanced fluorescent properties, including some ecAARS substrates [185]. Substituted tryptophans have altered wavelengths of emission and excitation, and in some cases, improved quantum yields [142,144,148,186,187]. Due to the lower molar absorptivity of these analogs, they are not as bright in fluorescence as some of the analogs activated by engineered AARS enzymes. Notable fluorescent compounds include Anap containing amino acid **L33** [188,189] and dansyl-derivatized **L34** [73,82,190]. Coumarin-containing **Y38** also has excellent fluorescent properties [173,191]. These amino acids offer a potentially efficient route for the site-selective labeling of fluorescent proteins. The most impressive applications of these amino acids involve Anap (**L33**). Kalstrup and Blunck, incorporated **L33** into several positions of a voltage-gated potassium channel and were able to observe changes in fluorescence under various voltages in patch clamp assays [190]. Sakata and coworkers used a similar strategy to investigate the activity of a voltage sensitive phosphatase [192]. The ability to site-specifically label these channels allowed for a detailed understanding of which parts of these membrane proteins undergo conformational changes as voltage changes. FRET with a membrane-embedded dye also allowed for analysis of how parts of the protein moved in relationship to the membrane [192].

3.4 Photoreactive groups. Several amino acids bearing photoaffinity-labels have been developed, including those containing diazirane **L18**, trifluoromethyl diazirane **Y27**, aryl azide **F6/Y22**, benzofuran **F28** and benzophenones **Y30-Y37**.   Particularly interesting is amino acid **Y37**, developed by Joiner and coworkers. The benzophenone allows for crosslinking, while the alkyne group enables subsequent capture on streptavidin resin after click chemistry with biotin azide [172]. Hino and coworkers also established **Y27** to map the interactions of the SH2 domain of GRB2 using proteomics followed by confirmation by Western blotting [170]. Photocaged serine, cysteine, and selenocysteine-containing

amino acids, **L27**-**L31** have also been developed as LeuRS substrates.   In an impressive example, Kang

and coworkers introduced a mutant ecLeuRS and suppressor tRNA[Leu] into the mouse neocortex. In the

presence of injected **L29**, they were able to control the activity of the Kir2.1 potassium channel with

light in brain slices [77].

3.5 Posttranslational modifications.   Four of the ncAAs on this list correspond to known post-

translational protein modifications. Monomethyl arginine (**R2**) and other arginine methylation patterns

are involved in a variety of cellular pathways including splicing, cell metabolism and the DNA damage

response, and the arginine methyltransferases have become hot drug targets in oncology [193]. The use

of **R2** as a substrate for ecArgRS has, to my knowledge, not been exploited in these studies.

Hydroxylation of Pro-402 or Pro-564 of HIF-1α giving *trans*-4-hydroxyproline (**P6**) is an essential

component of the cellular response to oxygen levels [194]. 3-nitrotyrosine (**Y9**) is a post-translational

chemical modification that is indicative of oxidative stress[195,196]; however, the incorporation of nitro-Tyr

using ecTyrRS is weak [43,114] and this has limited any useful studies.   In an impressive feat of

engineering, Italia et. al. developed a TyrRS that can directly incorporate sulfated tyrosine (**Y23**) into

proteins in mammalian cells and were able to study the effect of tyrosine sulfation at specific sites in a

thrombin inhibitor protein, heparin cofactor II [169].

3.6 Parameterization of local electric fields and vibrational energy transfer. A emerging application of

some of these ncAAs involves probing the electrostatics of enzyme catalysis and binding by the

vibrational stark effect—that is the change in IR stretching frequencies of a given functional group based

on the local electric field [180]. Much of the work in this field has focused on ncAAs and engineered AARS

from other organisms [180,197,198]; however, several of the ncAAs listed above contain functional groups

with sufficient IR-stretching signals and sensitivity to the electronic environment to serve as vibrational

probes. Alkyl and aryl fluorides, cyano groups (**F5** and **V5**), azides (**M20**-**M21**, **F6/Y22**, **W9**, **Y8**), and

nitro groups (**F8**, **Y9**) are all suitable candidates for these investigations [180,199]. Alternatively, one can

introduce the vibrational probe into a bound ligand and use other ncAAs to alter the electronic nature of the protein binding site [180,197,198].

4. Summary and Outlook.

In the preceding paragraphs, I have listed the known substrates for *E. coli* wild-type and mutant enzymes. Some were uncovered out of curiosity for the enzyme substrate specificity or from broad screens. Others were investigated due to their natural origin. Still others were developed with an eye towards protein engineering. By combining research from each of these sources into one location, a full picture of the specificity and engineerability of these enzymes can be highlighted.   The diversity of amino acids activated by these proteins is quite amazing, and due to *E. coli*'s central role in molecular biology, this database of ncAA substrates serves a unique and valuable toolkit of molecular function. As such this review is a valuable resource to inspire new directions in protein engineering, synthetic biology and peptide drug discovery.

5. Acknowledgments.

6. References

[1]   G. Hortin, I. Boime, *Methods Enzymol.* **1983**, *96*, 777–784.

[2]   R. Calendar, P. Berg, *J. Mol. Biol.* **1967**, *26*, 39–54.

[3]   L. Fowden, M. H. Richmond, *Biochim. Biophys. Acta* **1963**, *71*, 459–461.

[4]   R. Brabham, M. A. Fascione, *ChemBioChem* **2017**, *18*, 1973–1983.

[5]   A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, *Chem. Sci.* **2015**, *6*, 50–69.

[6]   N. Budisa, *Angew. Chemie - Int. Ed.* **2004**, *43*, 6426–6463.

[7]   L. Merkel, N. Budisa, *Org. Biomol. Chem.* **2012**, *10*, 7241–7261.

[8]   T. L. Hendrickson, V. De Crécy-Lagard, P. Schimmel, *Annu. Rev. Biochem.* **2004**, *73*, 147–176.

[9]   N. Voloshchuk, J. K. Montclare, *Mol. Biosyst.* **2009**, *6*, 65–80.

[10]   S. V. Melnikov, D. Söll, *Int. J. Mol. Sci.* **2019**, *20*, DOI 10.3390/ijms20081929.

[11]   F. Kaiser, S. Krautwurst, S. Salentin, V. J. Haupt, C. Leberecht, S. Bittrich, D. Labudde, M. Schroeder, *Sci. Rep.* **2020**, *10*, 1–17.

[12]   O. Vargas-Rodriguez, A. Sevostyanova, D. Söll, A. Crnković, *Curr. Opin. Chem. Biol.* **2018**, *46*, 115–122.

[13]   Y. Kim, K. Opron, Z. F. Burton, *Life* **2019**, *9*, 37.

[14]   Y. L. J. Pang, K. Poruri, S. A. Martinis, *Wiley Interdiscip. Rev. RNA* **2014**, *5*, 461–480.

[15]   M. A. R. Gomez, M. Ibba, *RNA* **2020**, *26*, 910–936.

[16]   K. Y. Fang, S. A. Lieblich, D. A. Tirrell, in *Methods Mol. Biol.*, Humana Press Inc., **2018**, pp. 173–186.

[17]   M. T. Lin, R. Fukazawa, Y. Miyajima-Nakano, S. Matsushita, S. K. Choi, T. Iwasaki, R. B. Gennis, in *Isot. Labeling Biomol. - Labeling Methods* (Ed.: Z.B.T.-M. in E. Kelman), Academic Press, **2015**, pp. 45–66.

[18]   B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, J. D. Rabinowitz, *Nat. Chem. Biol.* **2009**, *5*, 593–599.

[19]   J. M. Bacher, A. D. Ellington, *J. Bacteriol.* **2001**, *183*, 5414–5425.

[20]   F. Agostini, L. Sinn, D. Petras, C. J. Schipp, V. Kubyshkin, A. A. Berger, P. C. Dorrestein, J. Rappsilber, N. Budisa, B. Koksch, *ACS Cent. Sci.* **2020**, *7*, 81–92.

[21]   M. G. Hoesl, S. Oehm, P. Durkin, E. Darmon, L. Peil, H. R. Aerni, J. Rappsilber, J. Rinehart, D. Leach, D. Söll, N. Budisa, *Angew. Chemie - Int. Ed.* **2015**, *54*, 10030–10034.

[22]   J. S. Italia, C. Latour, C. J. J. Wrobel, A. Chatterjee, *Cell Chem. Biol.* **2018**, *25*, 1304-1312.e5.

[23]   W. Ding, H. Zhao, Y. Chen, B. Zhang, Y. Yang, J. Zang, J. Wu, S. Lin, *Nat. Commun.* **2020**, *11*, 1–13.

[24]   E. Brustad, M. L. Bushey, A. Brock, J. Chittuluru, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6004–6006.

[25]   M. C. T. T. Hartman, K. Josephson, J. W. Szostak, *Proc. Natl. Acad. Sci. U S A* **2006**, *103*, 4356–4361.

[26]   K. Josephson, M. C. T. T. Hartman, J. W. Szostak, *J. Am. Chem. Soc.* **2005**, *127*, 11727–11735.

[27]   Y. V. Guillen Schlippe, M. C. T. T. Hartman, K. Josephson, J. W. Szostak, Y. V Schlippe, M. C. T. T. Hartman, K. Josephson, J. W. Szostak, *J. Am. Chem. Soc.* **2012**, *134*, 10469–10477.

[28]   D. E. Hacker, J. Hoinka, E. S. Iqbal, T. M. Przytycka, M. C. T. Hartman, *ACS Chem. Biol.* **2017**, *12*, 795–804.

[29]   Y. Huang, M. M. Wiedmann, H. Suga, *Chem. Rev.* **2019**, *119*, 10360–10391.

[30]   C. S. Francklyn, E. A. First, J. J. Perona, Y. M. Hou, *Methods* **2008**, *44*, 100–118.

[31]   A. R. Fersht, J. S. Ashford, C. J. Bruton, R. Jakes, G. L. E. Koch, B. S. Hartley, *Biochemistry* **1975**, *14*, 1–4.

[32]   R. Calendar, P. Berg, *Biochemistry* **1966**, *5*, 1690–1695.

[33]   A. D. Wolfson, O. C. Uhlenbeck, *Proc Natl Acad Sci U S A* **2002**, *99*, 5965–5970.

[34]   J. S. Völler, M. Dulic, U. I. M. Gerling-Driessen, H. Biava, T. Baumann, N. Budisa, I. Gruic-Sovulj, B. Koksch, *ACS Cent. Sci.* **2017**, *3*, 73–80.

[35]   J. M. Ho, N. M. Reynolds, K. Rivera, M. Connolly, L. T. Guo, J. Ling, D. J. Pappin, G. M. Church, D. Söll, *ACS Synth. Biol.* **2016**, *5*, 163–171.

[36]   M. Bilus, M. Semanjski, M. Mocibob, I. Zivkovic, N. Cvetesic, D. S. Tawfik, A. Toth-Petroczy, B. Macek, I. Gruic-Sovulj, *J. Mol. Biol.* **2019**, *431*, 1284–1297.

[37]   S. Nehring, N. Budisa, B. Wiltschi, *PLoS One* **2012**, *7*, e31992.

[38]   C. Fan, J. M. L. Ho, N. Chirathivat, D. Söll, Y.-S. Wang, *ChemBioChem* **2014**, *15*, 1805–1809.

[39]   K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Lett.* **2001**, *502*, 25–30.

[40]   S. J. Oh, K. H. Lee, H. C. Kim, C. Catherine, H. Yun, D. M. Kim, *Biotechnol. Bioprocess Eng.* **2014**, *19*, 426–432.

[41]   C. F. Schachtele, P. Rogers, *J. Mol. Biol.* **1965**, *14*, 474–489.

[42]   E. G. Worst, M. P. Exner, A. De Simone, M. Schenkelberger, V. Noireaux, N. Budisa, A. Ott, *Bioorganic Med. Chem. Lett.* **2015**, *25*, 3658–3660.

[43]   M. C. T. Hartman, K. Josephson, C. W. Lin, J. W. Szostak, *PLoS One* **2007**, *2*, e972.

[44]   B. R. Babu, O. W. Griffith, *J. Biol. Chem.* **1998**, *273*, 8882–8889.

[45]   V. J. Balcar, G. A. Johnston, B. Twitchin, *J. Neurochem.* **1977**, *28*, 1145–1146.

[46]   T. Ishiyama, T. Furuta, M. Takai, Y. Okimoto, *J. Antibiot. (Tokyo).* **1975**, *28*, 821–823.

[47]   K. S. Hoffman, O. Vargas-Rodriguez, D. W. Bak, T. Mukai, L. K. Woodward, E. Weerapana, D. Söll, N. M. Reynolds, X. Noah, M. Reynolds, N. M. Reynolds, *J. Biol. Chem.* **2019**, *294*, 12855–12865.

[48]   I. Kaiser, *Arch Biochem Biophys* **1975**, *171*, 483–489.

[49]   A. Dantas de Araujo, S. R. Perry, D. P. Fairlie, *Org. Lett.* **2018**, *20*, 1453–1456.

[50]   Y. Yin, Q. Fei, W. Liu, Z. Li, H. Suga, C. Wu, *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 4880–4885.

[51]   T. L. Bullock, A. Rodríguez-Hernández, E. M. Corigliano, J. J. Perona, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7428–7433.

[52]   J. F. Eichler, J. C. Cramer, K. L. Kirk, J. G. Bann, *ChemBioChem* **2005**, *6*, 2170–2173.

[53]   Y. Ikeda, S. I. Kawahara, M. Taki, A. Kuno, T. Hasegawa, K. Taira, *Protein Eng.* **2003**, *16*, 699–706.

[54] M. J. Schlesinger, *J. Biol. Chem.* **1967**, *242*, 3369–3372.

[55] P. Soumillion, J. Fastrez, *Protein Eng.* **1998**, *11*, 213–217.

[56] C. M. Ring, E. S. Iqbal, D. E. Hacker, M. C. T. Hartman, T. A. Cropp, *Org. Biomol. Chem.* **2017**, *15*, 4536–4539.

[57] S. Schlesinger, M. J. Schlesinger, *J. Biol. Chem.* **1969**, *244*, 3803–3809.

[58] M. Englert, O. Vargas-Rodriguez, N. M. Reynolds, Y. S. Wang, D. Söll, T. Umehara, *Biochim. Biophys. Acta - Gen. Subj.* **2017**, *1861*, 3009–3015.

[59] P. Wang, Y. Tang, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 6900–6906.

[60] S. Son, I. C. Tanrikulu, D. A. Tirrell, *ChemBioChem* **2006**, *7*, 1251–1257.

[61] D. J. Stigers, Z. I. Watts, J. E. Hennessy, H. K. Kim, R. Martini, M. C. Taylor, K. Ozawa, J. W. Keillor, N. E. Dixon, C. J. Easton, *Chem. Commun.* **2011**, *47*, 1839–1841.

[62] M. L. Mock, T. Michon, J. C. M. Van Hest, D. A. Tirrell, *ChemBioChem* **2006**, *7*, 83–87.

[63] I. N. Arthur, J. E. Hennessy, D. Padmakshan, D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, G. Otting, J. G. Oakeshott, C. J. Easton, *Chem. - A Eur. J.* **2013**, *19*, 6824–6830.

[64] T. Kohno, D. Kohda, M. Haruki, S. Yokoyama, T. Miyazawa, *J. Biol. Chem.* **1990**, *265*, 6931–6935.

[65] N. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 14306–14307.

[66] Y. Tang, D. A. Tirrell, *Biochemistry* **2002**, *41*, 10635–10645.

[67] S. Cusack, A. Yaremchuk, M. Tukalo, *EMBO J.* **2000**, *19*, 2351–2361.

[68] Y. Tang, D. A. Tirrell, *J. Am. Chem. Soc.* **2001**, *123*, 11089–11090.

[69] N. Canu, P. Belin, R. Thai, I. Correia, O. Lequin, J. J. J. J. Seguin, M. Moutiez, M. Gondry, *Angew. Chemie - Int. Ed.* **2018**, *130*, 3172–3176.

[70] M. Liutkus, S. A. Fraser, K. Caron, D. J. Stigers, C. J. Easton, *ChemBioChem* **2016**, *17*, 908–912.

[71] Y. Tang, P. Wang, J. A. Van Deventer, A. J. Link, D. A. Tirrell, *ChemBioChem* **2009**, *10*, 2188–2190.

[72] C. Fan, J. M. L. Ho, N. Chirathivat, D. Söll, Y. S. Wang, *ChemBioChem* **2014**, *15*, 1805–1809.

[73] J. Wang, J. Xie, P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 8738–8739.

[74] E. Brustad, M. L. Bushey, A. Brock, J. Chittuluru, P. G. Schultz, *Bioorganic Med. Chem. Lett.* **2008**, *18*, 6004–6006.

[75] E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, *Nat. Chem. Biol.* **2007**, *3*, 769–772.

[76] R. Rakauskaitė, G. Urbanavičiūtė, M. Simanavičius, R. Lasickienė, A. Vaitiekaitė, G. Petraitytė, V. Masevičius, A. Žvirblienė, S. Klimašauskas, *iScience* **2020**, *23*, DOI 10.1016/j.isci.2020.101833.

[77] J. Y. Kang, D. Kawaguchi, I. Coin, Z. Xiang, D. D. M. O'Leary, P. A. Slesinger, L. Wang, *Neuron* **2013**, *80*, 358–370.

[78] J. C. Peeler, J. A. Falco, R. E. Kelemen, M. Abo, B. V. Chartier, L. C. Edinger, J. Chen, A. Chatterjee, E. Weerapana, *ACS Chem. Biol.* **2020**, *15*, 1535–1540.

[79] R. Rakauskaite, G. Urbanavičiute, A. Rukšenaite, Z. Liutkevičiute, R. Juškenas, V. Masevičius, S. Klimašauskas, *Chem. Commun.* **2015**, *51*, 8245–8248.

[80] H. S. Lee, J. Guo, E. A. Lemke, R. D. Dimla, P. G. Schultz, S. L. Hyun, J. Guo, E. A. Lemke, R. D. Dimla, P. G. Schultz, *J. Am. Chem. Soc.* **2009**, *131*, 12921–12923.

[81] A. Chatterjee, J. Guo, H. S. Lee, P. G. Schultz, *J. Am. Chem. Soc.* **2013**, *135*, 12540–12543.

[82] W. Wang, J. K. Takimoto, G. V. Louie, T. J. Baiga, J. P. Noel, K. F. Lee, P. A. Slesinger, L. Wang, *Nat. Neurosci.* **2007**, *10*, 1063–1072.

[83] Q. Wang, L. Wang, *J. Am. Chem. Soc.* **2008**, *130*, 6066–6067.

[84] E. M. Tippmann, P. G. Schultz, *Tetrahedron* **2007**, *63*, 6182–6184.

[85] F. P. Seebeck, J. W. Szostak, *J. Am. Chem. Soc.* **2006**, *128*, 7150–7151.

[86] F. T. Hofmann, J. W. Szostak, F. P. Seebeck, *J. Am. Chem. Soc.* **2012**, *134*, 8038–8041.

[87] T. Crepin, E. Schmitt, Y. Mechulam, P. B. Sampson, M. D. Vaughan, J. F. Honek, S. Blanquet, *J. Mol. Biol.* **2003**, *332*, 59–72.

[88] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19–24.

[89] F. Oldach, R. Altoma, A. Kuthning, T. Caetano, S. Mendo, N. Budisa, R. D. Süssmuth, *Angew. Chemie - Int. Ed.* **2012**, *51*, 415–418.

[90] K. L. Kiick, J. C. M. Van Hest, D. A. Tirrell, *Angew. Chemie - Int. Ed.* **2000**, *39*, 2148–2152.

[91] J. C. M. M. Van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282–1288.

[92] M. G. Hoesl, C. G. Acevedo-Rocha, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, N. Budisa, G. Antranikian, *ChemCatChem* **2011**, *3*, 213–221.

[93] F. Truong, T. Hyeon Yoo, T. J. Lampo, D. A. Tirrell, *J. Am. Chem. Soc* **2012**, *134*, 8551–8556.

[94] K. E. Beatty, F. Xie, Q. Wang, D. A. Tirrell, *J. Am. Chem. Soc.* **2005**, *127*, 14150–14151.

[95] M. Grammel, M. M. Zhang, H. C. Hang, *Angew. Chemie - Int. Ed.* **2010**, *122*, 6106–6110.

[96] T. H. Yoo, D. A. Tirrell, *Angew. Chemie - Int. Ed.* **2007**, *119*, 5436–5439.

[97] D. B. Cowie, G. N. Cohen, *BBA - Biochim. Biophys. Acta* **1957**, *26*, 252–261.

[98] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, *230*, 788–796.

[99] C. Wolschner, A. Giese, H. A. Kretzschmar, R. Huber, L. Moroder, N. Budisa, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7756–7761.

[100] J. Spížek, J. Janeček, *Biochem. Biophys. Res. Commun.* **1969**, *34*, 17–21.

[101] B. Bhushan, Y. A. Lin, M. Bak, A. Phanumartwiwath, N. Yang, M. K. Bilyard, T. Tanaka, K. L. Hudson, L. Lercher, M. Stegmann, S. Mohammed, B. G. Davis, *J. Am. Chem. Soc.* **2018**, *140*,

14599–14603.

[102] M. P. Exner, T. Kuenzl, T. M. T. To, Z. Ouyang, S. Schwagerus, M. G. Hoesl, C. P. R. Hackenberger, M. C. Lensen, S. Panke, N. Budisa, *ChemBioChem* **2017**, *18*, 85–90.

[103] D. M. Abdeljabbar, F. J. Piscotta, S. Zhang, A. J. Link, *Chem. Commun.* **2014**, *50*, 14900–14903.

[104] A. J. Link, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 11164–11165.

[105] I. C. Tanrikulu, E. Schmitt, Y. Mechulam, W. A. Goddard, D. A. Tirrell, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15285–15290.

[106] E. Schmitt, I. C. Tanrikulu, T. H. Yoo, M. Panvert, D. A. Tirrell, Y. Mechulam, *J. Mol. Biol.* **2009**, *394*, 843–851.

[107] A. J. Link, M. K. S. Vink, N. J. Agard, J. A. Prescher, C. R. Bertozzi, D. A. Tirrell, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10180–10185.

[108] A. J. Link, M. K. S. Vink, D. A. Tirrell, *J. Am. Chem. Soc.* **2004**, *126*, 10598–10602.

[109] G. Nigro, S. Bourcier, C. Lazennec-Schurdevin, E. Schmitt, P. Marlière, Y. Mechulam, *J. Struct. Biol.* **2020**, *209*, 107435.

[110] P. Kast, H. Hennecke, *J. Mol. Biol.* **1991**, *222*, 99–124.

[111] D. Datta, P. Wang, I. S. Carrico, S. L. Mayo, D. A. Tirrell, *J. Am. Chem. Soc.* **2002**, *124*, 5652–5653.

[112] I. Mermershtain, I. Finarov, L. Klipcan, N. Kessler, H. Rozenberg, M. G. Safro, *Protein Sci.* **2011**, *20*, 160–167.

[113] E. R. White, L. Sun, Z. Ma, J. M. Beckta, B. A. Danzig, D. E. Hacker, M. Huie, D. C. Williams, R. A. Edwards, K. Valerie, J. N. M. M. Glover, M. C. T. T. Hartman, *ACS Chem. Biol.* **2015**, *10*, 1198–1208.

[114] Z. Wang, H. Matthews, *RSC Adv.* **2020**, *10*, 11013–11023.

[115] C. Minks, R. Huber, L. Moroder, N. Budisa, *Anal. Biochem.* **2000**, *284*, 29–34.

[116] S. J. Hughes, T. W. Glover, X. X. Zhu, R. Kuick, D. Thoraval, M. B. Orringer, D. G. Beer, S. Hanash, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12410–12415.

[117] K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, *3*, 235–237.

[118] N. Sharma, R. Furter, P. Kast, D. A. Tirrell, *FEBS Lett.* **2000**, *467*, 37–40.

[119] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020–11024.

[120] D. E. Hacker, N. A. Abrigo, J. Hoinka, S. L. Richardson, T. M. Przytycka, M. C. T. Hartman, *ACS Comb. Sci.* **2020**, *22*, 306–310.

[121] H. J. Gabius, F. von der Haar, F. Cramer, *Biochemistry* **1983**, *22*, 2331–2339.

[122] S. Kothakota, T. L. Mason, D. A. Tirrell, M. J. Fournier, *J. Am. Chem. Soc.* **1995**, *117*, 536–537.

[123] T. Bentin, R. Hamzavi, J. Salomonsson, H. Roy, M. Ibba, P. E. Nielsen, *J. Biol. Chem.* **2004**, *279*, 19839–19845.

[124] W. Kim, A. George, M. Evans, V. P. Conticello, *ChemBioChem* **2004**, *5*, 928–936.

[125] W. Kim, K. I. Hardcastle, V. P. Conticello, *Angew. Chemie - Int. Ed.* **2006**, *45*, 8141–8145.

[126] R. W. Newberry, R. T. Raines, *Top. Heterocycl. Chem.* **2017**, *48*, 1–25.

[127] K. Deepankumar, S. P. Nadarajan, S. Mathew, S.-G. G. Lee, T. H. Yoo, E. Y. Hong, B.-G. G. Kim, H. Yun, *ChemCatChem* **2015**, *7*, 417–421.

[128] D. D. Buechter, D. N. Paolella, B. S. Leslie, M. S. Brown, K. A. Mehos, E. A. Gruskin, *J. Biol. Chem.* **2003**, *278*, 645–650.

[129] N. Budisa, C. Minks, F. J. Medrano, J. Lutz, R. Huber, L. Moroder, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 455–459.

[130] L. Fowden, *Biochem. J.* **1956**, *64*, 323–332.

[131] R. A. Dunlop, P. A. Cox, S. A. Banack, K. J. Rodgers, *PLoS One* **2013**, *8*, e75376.

[132] Z. Liu, Characterization of the in Vitro and in Vivo Specificity of Trans-Editing Proteins and Interacting Aminoacyl-TRNA Synthetases, Ohio State University, **2014**.

[133] A. Minajigi, B. Deng, C. S. Francklyn, *Biochemistry* **2011**, *50*, 1101–1109.

[134] M. Lee, R. Phillips, *Bioorg Med Chem Lett* **1992**, *2*, 1563–1564.

[135] P. B. Crowley, C. Kyne, W. B. Monteith, *Chem. Commun.* **2012**, *48*, 10681–10683.

[136] J. S. Italia, P. S. Addy, C. J. J. Wrobel, L. A. Crawford, M. J. Lajoie, Y. Zheng, A. Chatterjee, *Nat. Chem. Biol.* **2017**, *13*, 446–450.

[137] C. Minks, R. Huber, L. Moroder, N. Budisa, *Biochemistry* **1999**, *38*, 10649–10659.

[138] E. A. Pratt, H. Chien, *Biochemistry* **1975**, *14*, 3035–3040.

[139] N. Budisa, P. P. Pal, S. Alefelder, P. Birle, T. Krywcun, M. Rubini, W. Wenger, J. H. Bae, T. Steiner, *Biol. Chem.* **2004**, *385*, 191–202.

[140] C.-Y. Wong, M. R. Eftink, *Biochemistry* **1998**, *37*, 8947–8953.

[141] C. Y. Wong, M. R. Eftink, *Biochemistry* **1998**, *37*, 8938–8946.

[142] N. Budisa, M. Rubini, J. H. Bae, E. Weyher, W. Wenger, R. Golbik, R. Huber, L. Moroder, *Angew. Chemie - Int. Ed.* **2002**, *41*, 4066–4069.

[143] M. Rubini, S. Lepthien, R. Golbik, N. Budisa, *Biochim. Biophys. Acta - Proteins Proteomics* **2006**, *1764*, 1147–1158.

[144] J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, J. S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber, N. Budisa, *J. Mol. Biol.* **2003**, *328*, 1071–1081.

[145] C. Y. Wong, M. R. Eftink, *Protein Sci.* **1997**, *6*, 689–697.

[146] A. B. Pardee, L. S. Prestidge, *BBA - Biochim. Biophys. Acta* **1958**, *27*, 330–344.

[147] M. K. Azim, N. Budisa, *Biol. Chem.* **2008**, *389*, 1173–1182.

[148] S. Lepthien, M. G. Hoesl, L. Merkel, N. Budisa, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16095–16100.

[149] P. Soumillion, L. Jespers, J. Vervoort, J. Fastrez, *Protein Eng. Des. Sel.* **1995**, *8*, 451–456.

[150] J. Attias, M. J. Schlesinger, S. Schlesinger, M. J. Schlesinger, *J. Biol. Chem.* **1969**, *244*, 3803–3809.

[151] G. Brawerman, M. Ycas, *Arch. Biochem. Biophys.* **1957**, *68*, 112–117.

[152] J. B. A. Ross, A. G. Szabo, C. W. V. Hogue, *Methods Enzymol.* **1997**, *278*, 151–190.

[153] J. Y. Shen, W. C. Chao, C. Liu, H. A. Pan, H. C. Yang, C. L. Chen, Y. K. Lan, L. J. Lin, J. S. Wang, J. F. Lu, S. Chun-Wei Chou, K. C. Tang, P. T. Chou, *Nat. Commun.* **2013**, *4*, 1–7.

[154] W. C. Chao, J. Y. Shen, C. H. Yang, Y. K. Lan, J. H. Yuan, L. J. Lin, H. C. Yang, J. F. Lu, J. S. Wang, K. Wee, Y. H. Chen, P. T. Chou, *Biophys. J.* **2016**, *110*, 1732–1743.

[155] W. C. Chao, J. F. Lu, J. S. Wang, T. H. Chiang, L. J. Lin, Y. L. Lee, P. T. Chou, *Biochim. Biophys. Acta - Gen. Subj.* **2020**, *1864*, 129483.

[156] W. C. Chao, L. J. Lin, J. F. Lu, J. S. Wang, T. C. Lin, Y. H. T. H. Chen, Y. H. T. H. Chen, H. C. Yang, P. T. Chou, *Biochim. Biophys. Acta - Gen. Subj.* **2018**, *1862*, 451–459.

[157] J. Soutourina, P. Plateau, S. Blanquet, *J. Biol. Chem.* **2000**, *275*, 32535–32542.

[158] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science (80-.).* **2003**, *301*, 964–967.

[159] T. Kobayashi, T. Takimura, R. Sekine, K. Vincent, K. Kamata, K. Sakamoto, S. Nishimura, S. Yokoyama, *J. Mol. Biol.* **2005**, *346*, 105–117.

[160] A. Goulding, S. Shrestha, K. Dria, E. Hunt, S. K. Deo, *Protein Eng. Des. Sel.* **2008**, *21*, 101–106.

[161] B. Brooks, R. S. Phillips, W. F. Benisek, *Biochemistry* **1998**, *37*, 9738–9742.

[162] D. Kiga, K. Sakamoto, K. Kodama, T. Kigawa, T. Matsuda, T. Yabuki, M. Shirouzu, Y. Harada, H. Nakayama, K. Takio, Y. Hasegawa, Y. Endo, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9715–9720.

[163] F. Iraha, K. Oki, T. Kobayashi, S. Ohno, T. Yokogawa, K. Nishikawa, S. Yokoyama, K. Sakamoto, *Nucleic Acids Res.* **2010**, *38*, 3682–3691.

[164] F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Söll, S. Nishimura, *J. Biol. Chem.* **2000**, *275*, 40324–40328.

[165] W. Liu, A. Brock, S. S. Chen, S. S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239–244.

[166] T. S. Young, I. Ahmad, A. Brock, P. G. Schultz, *Biochemistry* **2009**, *48*, 2643–2653.

[167] S. Ye, C. Köhrer, T. Huber, M. Kazmi, P. Sachdev, E. C. Y. Yan, A. Bhagat, U. L. RajBhandary, T. P. Sakmar, *J. Biol. Chem.* **2008**, *283*, 1525–1533.

[168] A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782–11783.

[169] J. S. Italia, J. C. Peeler, C. M. Hillenbrand, C. Latour, E. Weerapana, A. Chatterjee, *Nat. Chem. Biol.*

2020, *16*, 379–382.

[170] N. Hino, M. Oyama, A. Sato, T. Mukai, F. Iraha, A. Hayashi, H. Kozuka-Hata, T. Yamamoto, S. Yokoyama, K. Sakamoto, *J. Mol. Biol.* **2011**, *406*, 343–353.

[171] C. M. Joiner, M. E. Breen, A. K. Mapp, *Protein Sci.* **2019**, *28*, 1163–1170.

[172] C. M. Joiner, M. E. Breen, J. Clayton, A. K. Mapp, *ChemBioChem* **2017**, *18*, 181–184.

[173] X. Steinberg, J. Galpin, G. Nasir, R. V. Sepúlveda, E. Ladron de Guevara, F. Gonzalez-Nilo, L. D. Islas, C. A. Ahern, S. E. Brauchi, *Heliyon* **2020**, *6*, e05140.

[174] V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel, P. Marliere, *Science* **2001**, *292*, 501–504.

[175] E. S. Iqbal, K. K. Dods, M. C. T. Hartman, *Org. Biomol. Chem.* **2018**, *16*, 1073–1078.

[176] F. H. Bergmann, P. Berg, M. Dieckmann, *J. Biol. Chem.* **1961**, *236*, 1735–1740.

[177] P. Wang, A. Fichera, K. Kumar, D. A. Tirrell, *Angew. Chemie - Int. Ed.* **2004**, *43*, 3664–3666.

[178] T. Katoh, T. Sengoku, K. Hirata, K. Ogata, H. Suga, *Nat. Chem.* **2020**, *12*, 1081–1088.

[179] E. S. Iqbal, S. L. Richardson, N. Abrigo, K. K. Dods, H. E. Osorio-Franco, H. S. Gerrish, H. K. Kotapati, I. M. Morgan, D. S. Masterson, M. C. T. Hartman, *Chem. Commun.* **2019**, *55*, 8959–8962.

[180] J. S. Völler, H. Biava, P. Hildebrandt, N. Budisa, *Biochim. Biophys. Acta - Gen. Subj.* **2017**, *1861*, 3053–3059.

[181] A. Divakaran, S. E. Kirberger, W. C. K. Pomerantz, *Acc. Chem. Res.* **2019**, *52*, 3407–3418.

[182] B. Yang, H. Liu, Z. Liu, R. Doenen, M. A. Nash, *Nano Lett.* **2020**, *20*, 8940–8950.

[183] I. C. Tanrikulu, E. Schmitt, Y. Mechulam, W. A. Goddard, D. A. Tirrell, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15285–15290.

[184] D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, E. M. Schuman, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9482–9487.

[185] Z. Cheng, E. Kuru, A. Sachdeva, M. Vendrell, *Nat. Rev. Chem.* **2020**, *4*, 275–290.

[186] N. Budisa, P. P. Pal, *Biol. Chem.* **2004**, *385*, 893–904.

[187] S. Veettil, N. Budisa, G. Jung, *Biophys. Chem.* **2008**, *136*, 38–43.

[188] S. L. Hyun, J. Guo, E. A. Lemke, R. D. Dimla, P. G. Schultz, *J. Am. Chem. Soc.* **2009**, *131*, 12921–12923.

[189] A. Chatterjee, J. Guo, H. S. Lee, P. G. Schultz, *J. Am. Chem. Soc.* **2013**, *135*, 12540–12543.

[190] T. Kalstrup, R. Blunck, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8272–8277.

[191] J. Wang, J. Xie, P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 8738–8739.

[192] S. Sakata, Y. Jinno, A. Kawanabe, Y. Okamura, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E7866.

[193] Q. Wu, M. Schapira, C. H. Arrowsmith, D. Barsyte-Lovejoy, *Nat. Rev. Drug Discov.* **2021**, DOI

10.1038/s41573-021-00159-8.

[194] M. A. McDonough, V. Li, E. Flashman, R. Chowdhury, C. Mohr, B. M. R. Liénard, J. Zondlo, N. J. Oldham, I. J. Clifton, J. Lewis, L. A. McNeill, R. J. M. Kurzeja, K. S. Hewitson, E. Yang, S. Jordan, R. S. Syed, C. J. Schofield, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9814–9819.

[195] N. Campolo, F. M. Issoglio, D. A. Estrin, S. Bartesaghi, R. Radi, *Essays Biochem.* **2020**, *64*, 111–133.

[196] H. Ahsan, *Hum. Immunol.* **2013**, *74*, 1392–1399.

[197] Y. Wu, S. G. Boxer, *J. Am. Chem. Soc.* **2016**, *138*, 11890–11895.

[198] T. Baumann, M. Hauf, F. Schildhauer, K. B. Eberl, P. M. Durkin, E. Deniz, J. G. Löffler, C. G. Acevedo-Rocha, J. Jaric, B. M. Martins, H. Dobbek, J. Bredenbeck, N. Budisa, *Angew. Chemie - Int. Ed.* **2019**, *58*, 2899–2903.

[199] E. Deniz, L. Valiño-Borau, J. G. Löffler, K. B. Eberl, A. Gulzar, S. Wolf, P. M. Durkin, R. Kaml, N. Budisa, G. Stock, J. Bredenbeck, *Nat. Commun.* **2021**, *12*, 1–8.

Surprising diversity: The aminoacyl-tRNA synthetases from *E. coli* have been extensively explored, and a surprising number of amino acid substrates have been described. This comprehensive review catalogs all known substrates for these enzymes and highlights trends and novel applications.