

Tuning phenylalanine fluorination to assess aromatic contributions to protein function and stability in cells

Received: 13 July 2022

Accepted: 27 December 2022

Published online: 04 January 2023

 Check for updates

Grace D. Galles ^{1,2,4}, Daniel T. Infield^{1,4}, Colin J. Clark¹, Marcus L. Hemshorn², Shivani Manikandan ¹, Frederico Fazan¹, Ali Rasouli ³, Emad Tajkhorshid ³, Jason D. Galpin ¹, Richard B. Cooley², Ryan A. Mehl² & Christopher A. Ahern ¹ ✉

The aromatic side-chains of phenylalanine, tyrosine, and tryptophan interact with their environments via both hydrophobic and electrostatic interactions. Determining the extent to which these contribute to protein function and stability is not possible with conventional mutagenesis. Serial fluorination of a given aromatic is a validated method *in vitro* and *in silico* to specifically alter electrostatic characteristics, but this approach is restricted to a select few experimental systems. Here, we report a group of pyrrolysine-based aminoacyl-tRNA synthetase/tRNA pairs (tRNA/RS pairs) that enable the site-specific encoding of a varied spectrum of fluorinated phenylalanine amino acids in *E. coli* and mammalian (HEK 293T) cells. By allowing the cross-kingdom expression of proteins bearing these unnatural amino acids at biochemical scale, these tools may potentially enable the study of biological mechanisms which utilize aromatic interactions in structural and cellular contexts.

While primarily appreciated for its hydrophobicity, the benzyl side chain of phenylalanine also engages in various types of energetically favorable aromatic interactions (Fig. 1a)^{1–11}. Energetically significant aromatic interactions involving phenylalanine are believed to be widespread in nature^{12,13}. Based on emerging structural data, they have been increasingly proposed to play important mechanistic roles in ligand recognition^{14–19} and protein–protein interactions, both in normal function^{20,21} and as a result of clinical mutation²². They may also play important roles in membrane anchoring, via attraction of the benzyl side chain to choline lipid headgroups²³, and have recently been proposed to mediate conduction in some ion channels²⁴. Testing the specific functional role for aromaticity of the benzyl side chain in an interaction is not possible using conventional mutagenesis since mutation to a non-aromatic natural amino acid also causes concomitant changes in amino acid shape, convoluting data interpretation. Fluorination of the aromatic ring, on the other hand, allows for

redistribution of the electrostatic potential of Phe with only minor steric changes (Fig. 1b); thus allowing the specific testing of the functional relevance of aromaticity. Increasing the number of fluorine substitutions results in a linear decrease in the electrostatic potential for π interactions^{25,26}.

The site-specific installation of fluorinated phenylalanine amino acids can be accomplished via peptide synthesis and by the direct injection of misacylated nonsense suppressor tRNA²⁷, the latter of which has been used to functionally validate and quantify aromatic interactions in some membrane proteins, including ion channels^{28–33}. These methods, while chemically flexible, cannot be easily scaled for biochemical and structural approaches. Additionally, some natural amino acids have been replaced with fluorinated, non-canonical structural analogs via permissive endogenous synthetases, but these methods are not site-specific, resulting in multiple sites within the protein sequence being modified heterogeneously³⁴. On the other

¹Department of Molecular Physiology and Biophysics, University of Iowa College of Medicine, Iowa City, IA, USA. ²The GCE4All Research Center, Department of Biochemistry & Biophysics, Oregon State University, Corvallis, OR, USA. ³Theoretical and Computational Biophysics Group, NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ⁴These authors contributed equally: Grace D. Galles, Daniel T. Infield. ✉e-mail: christopher-ahern@uiowa.edu

hand, nonsense suppression via coevolved tRNA/aminoacyl tRNA-synthetase (tRNA/RS) pairs allows for biochemical-scale production of proteins bearing site-specifically installed non-canonical amino acids (ncAA) in diverse cell systems and even organisms^{27,35}. Previously, one such tRNA/RS pair that encodes select fluorinated phenylalanine analogs into proteins was selected by rational design followed by a screening of a single-residue library of 20 variants. This system was shown to be functional in *Escherichia coli* if grown in minimal media without exogenous phenylalanine³⁶ but the lack of a sufficiently efficient tRNA/RS pair stymies field progress. A method to accurately encode fluorinated Phe ncAAs in complex media and in eukaryotic cells would prove highly valuable for many clinically relevant soluble and membrane protein targets, including the myriad genes that must be expressed and/or studied in mammalian cells¹².

In this work, we describe the identification and validation of a group of synthetases, which we have named Phe_x, from the pyrrolysine (Pyl) system^{37–39}, which are competent for site-specific

encoding of fluorinated di-, tri-, tetra-, and penta-fluoro phenylalanine analogs within proteins in both bacterial and mammalian expression systems. We hypothesized that by screening a greater number of tRNA-synthetase (RS) active site variants ($\sim 10^7$), using a strategy that enabled enrichment of variants able to selectively incorporate fluorinated phenylalanine derivatives but not phenylalanine, we may identify synthetases with the desired utility and versatility. Intact protein mass spectrometry defines the synthetase and ncAA combinations that enable high-fidelity encoding at the amber suppression site without spurious substitution at natural codons. Finally, we demonstrate biochemical scale encoding in two large (>150 kDa) human membrane proteins of high clinical and therapeutic interest.

Results

Encoding fluorinated phenylalanine analogs in *E. coli*

Historical attempts by our groups to screen for robust pyrrolysine (Pyl) synthetases able to incorporate fluorinated phenylalanine analogs were not successful, possibly owing to the close steric resemblance of fluoro-benzene variants. In the present work, we decided to screen a library of Pyl synthetases with para-methyl, tetra-fluoro-phenylalanine (Fig. 2a), in the hope that the additional structural bulk of the methyl at the para position would permit the identification of synthetase's selective against natural amino acids but selective for fluorinated Phe derivatives, perhaps including those lacking the para-methyl group. Two independent rounds (Fig. 2b) of positive and negative screening within a Pyrrolysine-based RS library (see Methods) in *E. coli*⁴⁰ resulted in the identification of several synthetases that enabled amber suppression within superfolder GFP (sfGFP_N150TAG) in response to para-methyl, tetra-fluoro Phe (Fig. 2c). Unique synthetases wherein suppression of sfGFP (from GFP signal) was greater than 10 fold above background. As a preliminary assessment of the quality of the evolved RS pairs, experiments to

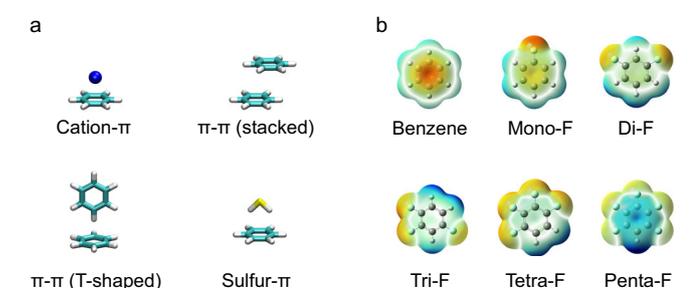


Fig. 1 | Depictions of geometry and electrostatics. **a** Illustrations of types of π interactions in which phenylalanine can participate. **b** Electrostatic views of benzene (as a structural surrogate for phenylalanine) and fluorinated analogs. Coloring is to the convention of Red = negative, Blue = positive.

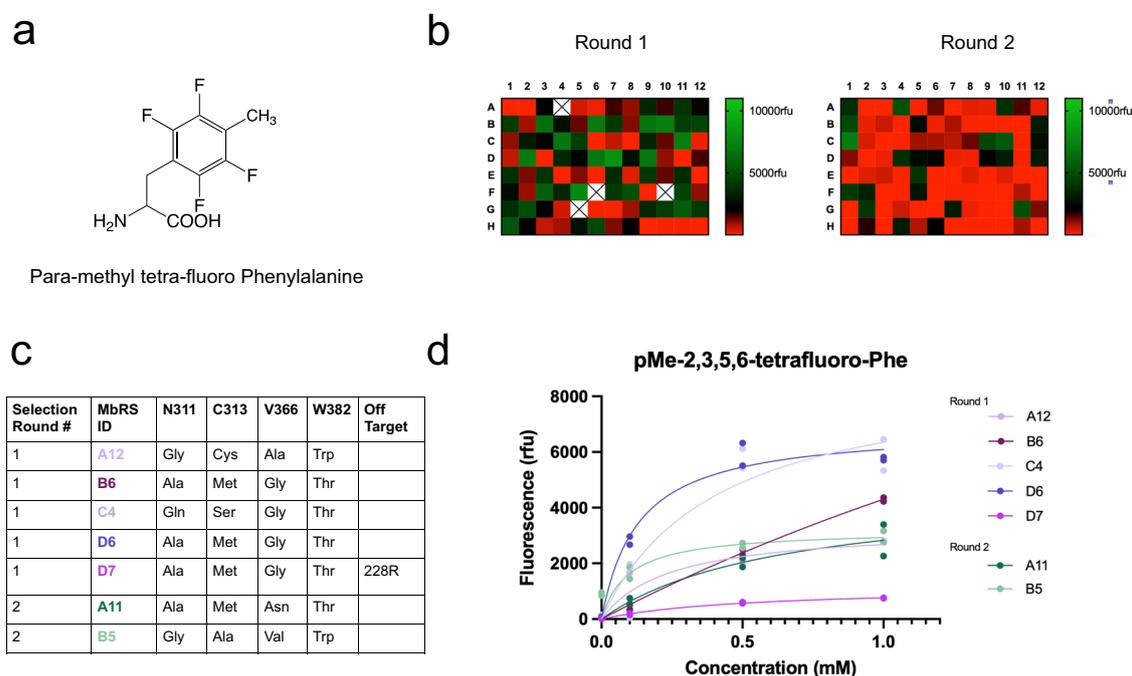


Fig. 2 | Strategy and identification of synthetases. **a** Chemical structure of para-methyl, tetra-fluoro Phe, which was used for screening in a PylRS library in *E. coli*. **b** Heatmaps of synthetases hits and relative GFP N150TAG rescue with para-methyl, tetra-fluoro Phe. Rounds 1 and 2 were independent screens of the library with para-

methyl, tetra-fluoro Phe. **c** Table of the active site and off-target mutations borne by the top performing identified synthetases. MbRS indicates aminoacyl synthetases from the species *Methanosarcina barkeri*. **d** Potency (UP50) curves for the synthetases detailed in panel (c), $n = 2$ independent experiments for all conditions.

determine the concentration of ncAA required to express ncAA-sfGFP at half its maximum expression level, called the Unnatural Protein 50% (UP50)⁴⁰, were performed for the synthetases using para-methyl, tetra-fluoro Phe (Fig. 2D). The UP50 values for para-methyl tetra-fluoro Phe for the selected tRNA/RS pairs validate they will function efficiently at 0.5 mM ncAA⁴⁰.

Synthetases for para-methyl, tetra-fluoro-Phe encode fluorinated Phe analogs

An important purpose of the screening efforts was to identify synthetases that enable the encoding of fluorinated (but non-methylated) phenylalanine analogs within proteins. Therefore we tested for permissivity⁴¹ of the selected synthetases against a diverse collection

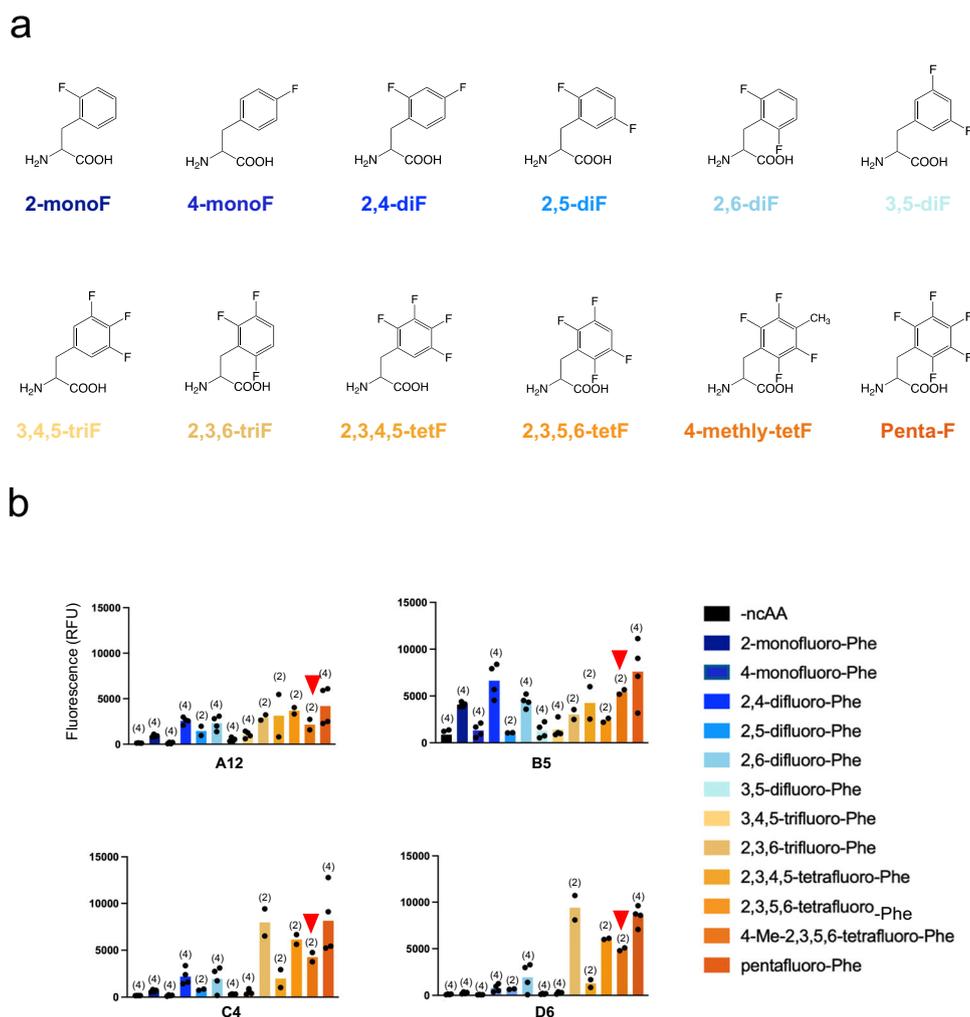


Fig. 3 | Permissivity of the identified synthetases. **a** Chemical structures of the fluorinated phenylalanine residues used in the screen. **b** Green fluorescence from the rescue of sfGFP-N150TAG in *E. coli* using 1 mM of each ncAA after 24-h incubation. N values of independent expressions in parentheses with bar height

indicating the mean value. Red arrowheads indicate the level of GFP fluorescence for para-methyl, tetrafluoro-Phe, which was the Phe analog originally used for screening.

Table 1 | Effects of fluorination on interaction energies with cations

Phe _x substrate	Phe	2F-Phe	2,4F-Phe; 2,6F-Phe	2,5F-Phe	3,4,5F-Phe	2,3,6F-Phe	2,3,4,5F-Phe	2,3,5,6F-Phe	2,3,4,5,6F-Phe
Benzene analog	Benzene	1F	1,3F	1,4F	1,2,3F	1,2,5F	1,2,3,4F	1,2,4,5F	1,2,3,4,5F
	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)	ΔG (kcal/mol) (% PHE)
Sodium	26.7 (100)	22.8 (85.4)	19.1 (71.5)	18.7 (70)	16.2 (60.7)	15.6 (58.4)	12.7 (47.6)	11.9 (44.6)	9.1 (34.1)
Ammonium	19.5 (100)	16.3 (83.6)	13.1 (67.2)	13 (66.7)	10.2 (52.3)	10.1 (51.8)	7.4 (37.9)	6.8 (34.9)	4.6 (23.6)
Guanidinium T-shaped	14.8 (100)	12.3 (83.1)	10.3 (69.6)	9.7 (65.5)	8.1 (54.7)	6.2* (41.9)	3.9* (26.4)	3.9* (26.4)	1.8* (12.2)

*Derived from the single-point calculation.

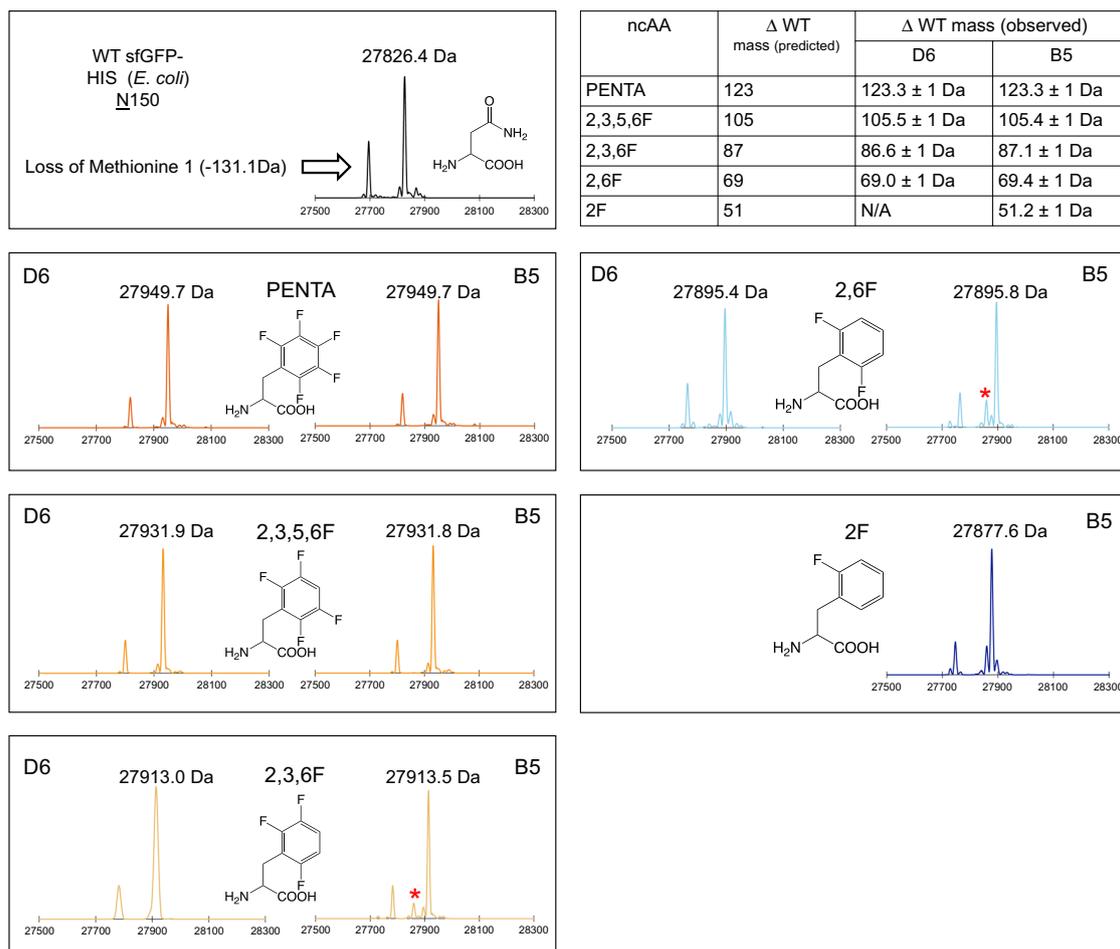


Fig. 4 | ESI mass spectrometry confirmation of site-specific encoding of fluorinated Phe residues within superfolder GFP in *E. coli*. sfGFP was purified via His tag from 50 mL growths. Table at top right: mass differences of sfGFP N150TAG from WT GFP conform to prediction for the molecular mass of each

fluoro Phe species. In some conditions for B5 synthetase, a small fraction of mass consistent with Phe incorporation was observed (asterisks). Mass spectrometry was not pursued for the combination of 2F Phe and D6 due to low expression in the screen.

of phenylalanine analogs (Fig. 3a). In practice, this was an iterative process, with preliminary results from both *E. coli* (Fig. 3) and HEK cells (discussed below) motivating the synthesis and evaluation of additional structural analogs to addition in both systems. Interestingly, synthetases A12, C4, and D6 encoded penta-fluoro Phe and 2,3,5,6 tetra-fluoro Phe at comparable levels to para-methyl, tetra-fluoro Phe in *E. coli*, indicating that the para-methyl group is not required for recognition (Fig. 3b). Each synthetase showed some degree of permissivity; Phe_{X-B5} and Phe_{X-D6} were particularly interesting examples with substantially different recognition profiles (Fig. 3b). Phe_{X-B5} displayed broader utility, but also lower apparent specificity over background than Phe_{X-D6} in *E. coli*.

With permissivity profiles in hand, we performed ab initio quantum mechanical calculations to determine the effect of specific fluorination patterns on the potential of each analog to engage in electrostatic interactions, which has been previously done for a few of these analogs²⁵. We calculated the theoretical ΔG (kcal/mol) for interactions with three different cations and their representative benzene structures (Table 1). As expected, increasing the number of fluorine atoms elicited a linear decrease in the binding energy. Table 1 expresses both the derived ΔG values and how they compare as a percentage of the native benzene of Phe (% Phe). Importantly, the relative positioning of the fluorine atoms did not greatly affect the energetics of binding, suggesting that the synthetases encode analogs with different potential local electrostatic effects but equivalent global perturbations. The encoding of multiple di-, tri-, or tetra-fluoro Phe

analogues may prove useful to better understand phenylalanine side chains involved in complex, or multivalent motifs^{25,42}.

We next verified the fidelity of the top tRNA/RS pairs by analyzing the intact sfGFP proteins by ESI (electrospray ionization)-mass spectrometry to confirm the accurate encoding of the fluorinated phenylalanine derivatives⁴⁰. We used the Phe_{X-D6} and Phe_{X-B5} synthetases to express and purify sfGFP_N150TAG_HIS in *E. coli*. Whole protein sfGFP masses were consistent with the replacement of N150 with each of the respective F-Phe ncAAs. Using Phe_{X-D6}, we observed evidence of high-fidelity incorporation (Fig. 4) for di-, tri-, tetra-, and pentafluoro phenylalanine (see Methods for additional details). When the Phe_{X-B5} synthetase was used to express sfGFP bearing a range of ncAAs at N150, evidence of efficient encoding was clear for a wide range of fluorinated ncAAs (Fig. 4). However, for several ncAAs with lower encoding efficiency, detectable peaks corresponding to the mass of Phe at N150 were observed (Fig. 4, Δ mass of +33 Da from asparagine, asterisks). This background encoding of natural AAs was expected from the higher fluorescence values in the absence of ncAA for Phe_{X-B5} (Fig. 3b), suggesting that Phe_{X-B5} incorporates natural amino acids in *E. coli* when using ncAAs poorly recognized by the synthetase.

Site-specific encoding of fluoro-Phe ncAAs into proteins in HEK cells

To evaluate the encoding of fluorinated analogs in HEK cells, we cloned the Phe_{X-B5} and Phe_{X-D6} synthetases into the pAcBac1⁴³ plasmid and co-transfected them with a plasmid encoding

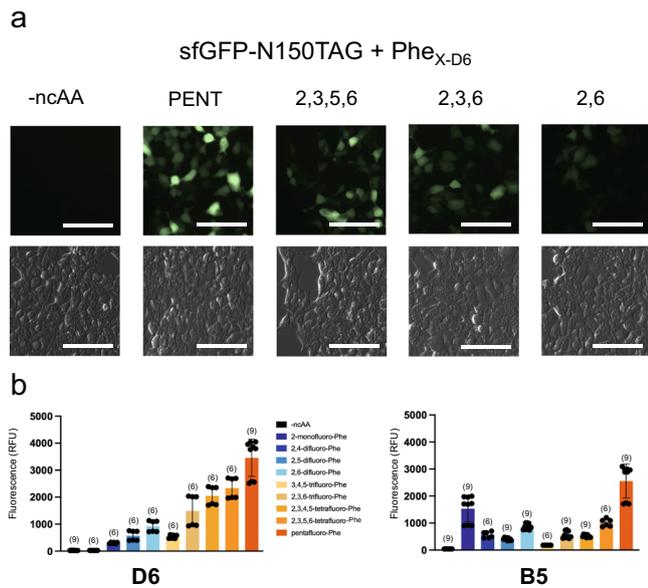


Fig. 5 | Function of Phe_x synthetases in HEKT cells. **a** Micrographs of HEKT cells co-transfected with GFP_{N150TAG} and Phe_{x-D6} in the presence or absence of the given amino acid, GFP (top) and brightfield (bottom). In all cases, images were taken approximately 24 h after transfection, and the unnatural amino acid was used at a 2 mM concentration. Scale bar = 100 μm. Images are representative of four individual fields for no amino acid and penta-fluoro-Phe, and two individual fields for each of the remaining amino acids. **b** Lysate GFP fluorescence for a range of fluorinated phenylalanine residues. *N* values are listed for a number of readings, with bar height indicating mean values ± standard deviation.

sfGFP_{N150TAG} in the presence or absence of fluoro-Phe ncAAs. In HEK cells, both the Phe_{x-D6} and Phe_{x-B5} tRNA/RS pairs produced sfGFP in the presence of fluoro-Phe ncAAs but not appreciably in their absence, yielding green cells which were imaged for fluorescence at 24 h (Fig. 5a). For quantitative comparison of ncAA-protein expression, fluorescence from cell lysate was measured. As with *E. coli* fluoro-Phe encoding, the position and number of fluorine substitutions affected the efficiency of producing ncAA-protein with the Phe_{x-D6} and Phe_{x-B5} tRNA/RS pairs (Fig. 5b). Of note, the relative fidelity of these tRNA/RS pairs was better in HEK cells compared to that in *E. coli*, but the background fluorescence from the Phe_{x-B5} tRNA/RS pair was still higher than Phe_{x-D6} (estimated 0.84% and 0.15% of maximum encoding for Penta-F Phe for Phe_{x-B5} and Phe_{x-D6} pairs, respectively).

To rigorously confirm the efficiency and fidelity of these tRNA/RS pairs in a eukaryotic context, ncAA-sfGFP was expressed in HEKT cells at 2 mM ncAA, purified via c-terminal His₆ tag and subjected to ESI-mass spectroscopy analysis. The Phe_{x-D6} and Phe_{x-B5} pairs encoded multiply fluorinated Phe analogs with high fidelity (Fig. 6). Estimated yield of sfGFP with Penta-fluoro Phe encoded by Phe_{x-D6} was ~34 μg per gram of HEKT cells (Supplementary Fig. 1). From the *E. coli* permissivity evaluation we note a primary feature distinguishing Phe_{x-B5} from Phe_{x-D6} is the ability of the former to recognize 2-mono fluoro Phenylalanine. However, in HEKT cell expression of sfGFP_{N150TAG} with Phe_{x-B5} and 2-mono fluoro Phe, we observed evidence of widespread encoding of 2-mono fluoro Phe at Phe-codons (Supplementary Fig. 2). This observation suggests that the endogenous (human) Phe tRNA/RS pair cannot effectively distinguish between 2-mono F Phe and Phe, preventing site-specific encoding of 2-monofluoro Phe in HEKT cells and likely other human cells as well. Overall, the ESI-mass spectroscopy analysis of the pure ncAA-protein demonstrates the fidelity of the Phe_x RS group to encode 6 of the fluorinated ncAAs; 2 di-fluorinated (2,5/2,6), a tri-fluorinated (2,3,6),

two tetra-fluorinated (2,3,5,6/2,3,4,5), and the penta-fluorinated phenylalanine in HEKT cells with high efficiency and fidelity. Dose-response experiments with ascending concentrations of the ncAAs indicate that a few combinations (such as Phe_{x-D6} + penta-fluoro Phe) are useable at sub-millimolar concentrations of ncAA but that, in general, 1–2 mM is necessary for efficient ncAA-protein expression (Supplementary Fig. 3).

Use of Phe_x synthetases to encode fluoro-Phe ncAAs into human proteins

Finally, we demonstrate the utility of the tRNA/Phe_x pairs for fluoro-Phe ncAA encoding within two large, human ion channels of high clinical significance. Mutations within the ~150 kDa cystic fibrosis (CF) transmembrane conductance regulator (hCFTR) cause the life-shortening disease CF (Fig. 7a)⁴⁴, while mutation and dysregulation of the ~200 kDa cardiac sodium channel (hNa_v 1.5) are associated with various cardiovascular diseases including cardiac arrhythmia and heart failure (Fig. 7b)^{45,46}. These proteins have been studied in a wide variety of eukaryotic cell hosts, but mammalian cell expression is either essential or preferred for many types of functional, biochemical, and structural analyses of channel regulation or pharmacology^{47–60}. We introduced TAG codons at position F508 in hCFTR and at F1486 in hNa_v 1.5 and co-transfected these expression plasmids with Phe_{x-D6} in the presence or absence of 2,3,6F-Phe. We harvested the cells two days post transfection and assessed expression via a western blot of cell lysates. In both cases, expression of full-length channels (as indicated by antibodies specific for C-terminal epitopes) was dependent on the addition of 2,3,6F-Phe in the growth media (Fig. 7a–c), demonstrating specific encoding by Phe_{x-D6}. Independent confirmation of 2,3,6F-Phe encoding within hNa_v 1.5 was accomplished via expression of hNa_v 1.5 F1486 (2,3,6F-Phe) via Phe_{x-D6}, purification via FLAG TAG and in-gel tryptic digestion/mass spectrometry (MS)–determination (Fig. 7d). Finally, WT and hNa_v1.5 F1486 (2,3,6F Phe) channels were expressed in HEK cells and recorded via the whole-cell patch clamp configuration. In agreement with the western blot results, robust macroscopic currents were observed for both WT and mutant channels; these channels appeared to activate and inactivate normally in response to membrane depolarization (Fig. 7e). Analysis of gating parameters revealed that encoding 2,3,6F-Phe at position F1486 caused a subtle enhancement of channel inactivation. This was apparent in a small (~7 mV) but statistically significant left shift in the midpoint of steady-state inactivation (Fig. 7g, Supplementary Table 1). Interestingly, data from exhaustive natural mutation of this critical phenylalanine previously revealed a strong correlation between hydrophobicity and efficiency of inactivation; no natural mutation fully reproduced WT (Phe) behavior⁶¹. The data here suggest that the interaction of this side-chain within the inactivated conformation of channel⁵³ is enhanced via near-isosteric fluorine substitution on the benzene ring of F1486. This is consistent with studies of model hydrophobic cores wherein targeted fluorination enhanced thermodynamic stability over native levels⁴².

Discussion

The electrostatic contributions of aromatic residues to protein–protein interactions and ligand recognition are becoming increasingly appreciated⁴², but available tools limit the ability to test proposed mechanistic hypotheses. The tRNA/RS pairs reported herein allow for the encoding of a wide spectrum of fluorinated phenylalanine residues, which can be used to tune the electrostatic contributions of the aromatic side chain in ππ- and cation-π interactions (Fig. 1, Table 1). The majority of the fluorinated phenylalanine residues used in this study are commercially available, making this tool set immediately adoptable for those studying aromatic side change contributions in prokaryotes and eukaryotes. As these tRNA/RS pairs were first evolved for a multiply fluorinated para-methyl-Phe analog, it is not surprising that the more highly fluorinated species tend to encode most efficiently into proteins

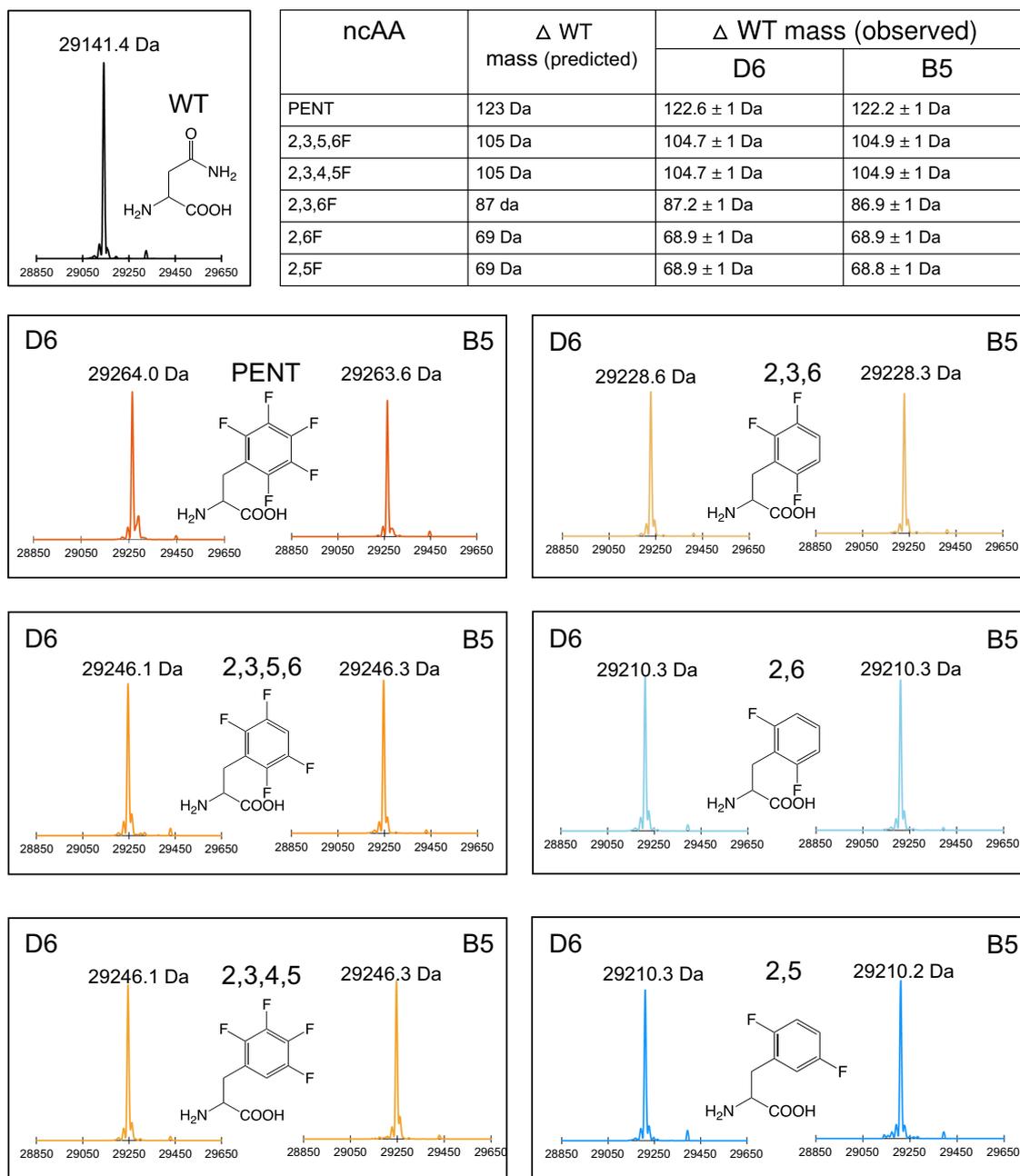


Fig. 6 | ESI mass spec confirmation of encoding of fluorinated Phe analogs within sfGFP expressed in HEKT cells. ncAAs were added at 2 mM concentration and cells were harvested approximately 36 h post transfection. sfGFP-WT-V5-HIS or

N150_TAG-V5-HIS was purified via HIS tag. Table (at top right) reports mass differences of sfGFP N150TAG from WT GFP, which conform to the prediction for the molecular mass of each fluoro Phe species.

(Figs. 3 and 5). We view the fact that penta-fluoro and tetra-fluoro phenylalanine analogs encode so well as fortuitous because it enables the investigator to gain valuable early information on the electrostatic role of Phe residues given the strong electrostatic perturbation imparted by the fluorines. For example, penta-fluoro phenylalanine is estimated to bear only -12–34% of the cation- π binding potential of native Phe (Table 1), whereas tri-fluorinated phenylalanine bears -40–60%, depending on the cation involved.

The efficiency of the synthetases, as demonstrated herein, suggests that they will be useful for the site-specific encoding of non-canonical phenylalanine analogs within proteins in *E. coli* and HEK cells at levels sufficient for biochemical characterization. In HEK cells, the yield of the most robust combination (Phe_{X-D6} and penta-fluoro phenylalanine) was 34 μ g per gram of cell pellet using transient

transfection. This is comfortably within the range necessary for structure determination by cryo-EM, provided the protein is efficiently purified and the growth is sufficiently scaled. It is worth noting that the PyIRS system is useful in other mammalian cell lines beyond HEK⁶². That is, given the broad orthogonality of the *Mb* PyIRS/PyI-tRNA platform, we anticipate the F-Phe incorporation systems described here will be functional across other eukaryotic systems for which Genetic Code Expansion systems have been described. It is noted, however, that such applications will require additional levels of optimization relating to RS, tRNA, and target protein expression levels, as well as ncAA concentrations used in culture media.

The system may also be further optimized via the construction of synthetase-incorporated baculoviruses⁴³ or stable cell lines. The

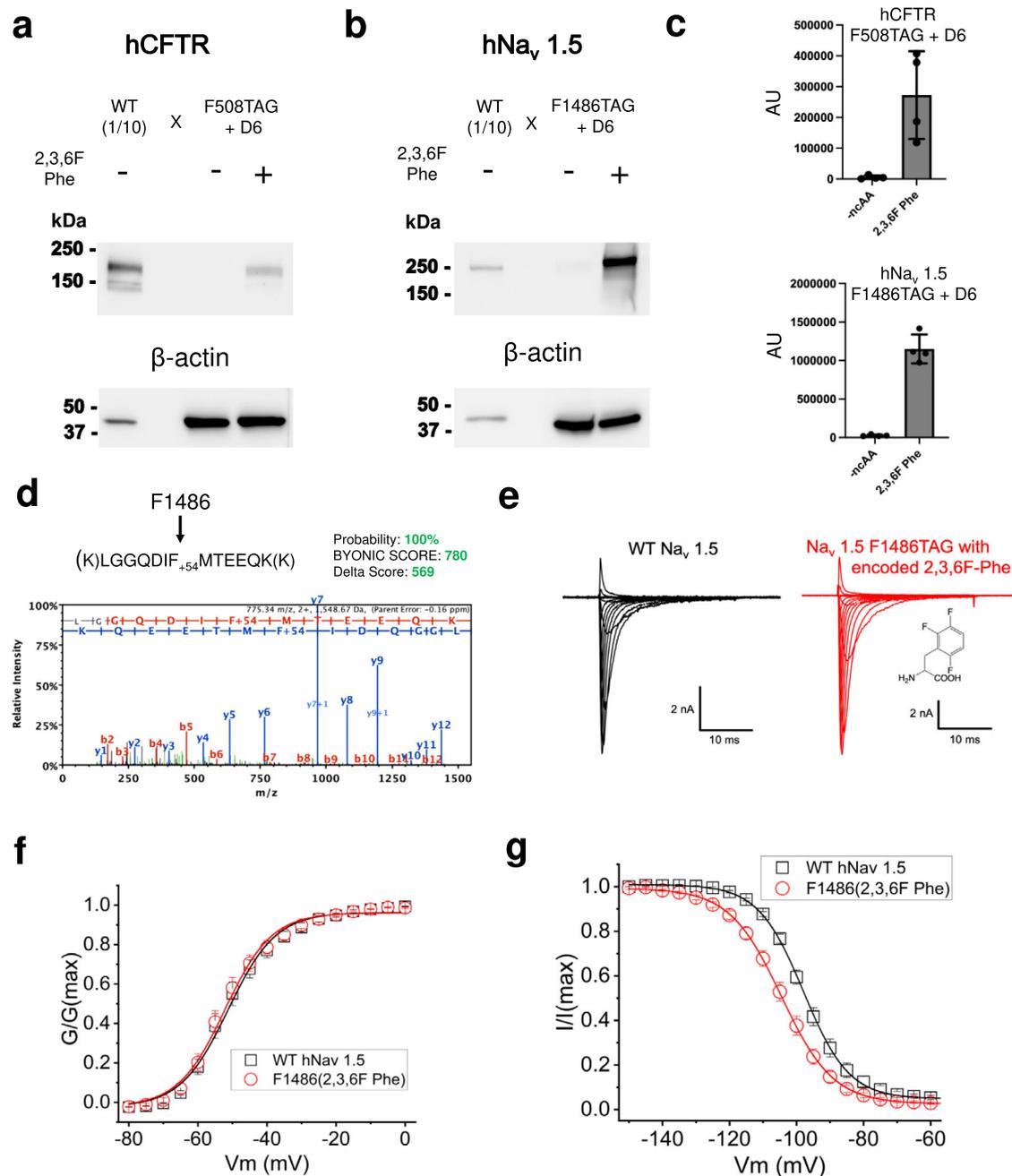


Fig. 7 | Site-specific encoding of tri-fluoro-Phe within two large, human ion channels. **a** Western blot of hCFTR using C terminal antibody (ab596) of WT CFTR or F508TAG-CFTR co-transfected with Phe_{X-D6} in the presence or absence of 2 mM 2,3,6F Phe. WT lysates are loaded at 1:10 of TAG conditions. Uncropped blots are shown in the Supplementary Information file. **b** Western blot of hNav_v1.5 using C terminal antibody (D9J7S) of WT hNav_v1.5 or F1486TAG-hNav_v1.5 co-transfected with Phe_{X-D6} in the presence or absence of 2 mM 2,3,6F Phe. Images are representative of four individual blots. Uncropped blots are shown in the Supplementary Information file. **c** Quantification (densitometry) of signal from 60 μg of the loaded lysate

($n = 4$ blots for each condition). **d** Mass spectra of a tryptic fragment from expressed and purified hNav_v1.5 F1486 (2,3,6F Phe) via D6. **e** Example current families for whole (HEK) cells expressing WT or mutant channels. Cells held at -140 mV and activated via depolarization in 5 mV increments. **f** Activation (G/V) curves for WT and mutant channels, $n = 5$ independent cells for wild-type and eight independent cells for hNav_v1.5 F1486 2,3,6F Phe. **g** Steady-state inactivation curves for WT and mutant channels, $n = 5$ independent cells for wild-type and eight independent cells for hNav_v1.5 F1486 2,3,6F Phe.

function of the synthetases in mammalian cells enables their users to answer questions that were previously intractable with other systems (prokaryotic, cell-free, *Xenopus* oocyte) and, in principle, can be extended to express and characterize myriad membrane and soluble proteins. Beyond mechanistic work, the synthetases may be used to site-specifically incorporate fluoro-phenylalanine nuclear magnetic resonance probes⁵³, and an optimized system utilizing these synthetases may enable the biotechnological production of therapeutic

proteins wherein fluorinated phenylalanine residues are encoded for pharmacologically advantageous reasons⁵⁴.

Methods

Amino acids

All solvents and reagents were supplied by Sigma Aldrich and were used as-is unless explicitly stated. Dry nitrogen was supplied by Praxair and passed through two moisture-scrubbing columns

of dry calcium sulfate (Drierite) prior to use. High-performance liquid chromatography (HPLC) analyses were performed on a Waters 1525 Binary HPLC pump equipped with a Waters 2998 Photodiode Array Detector, employing Sunfire C18 analytical (3.5 μm , 4.6 mm \times 150 mm, 0.8 ml/min) or preparative (5.0 μm , 19 mm \times 150 mm, 10 ml/min) columns and Empower software. Buffers were drawn in linear gradients from 100% A (50 mM ammonium acetate) to 100% B (acetonitrile) over 30 min. Mass spectra were recorded on a Waters QToF Premier Quadrupole instrument, in both positive and negative modes. 2,3,5,6 tetrafluoro Phe amino acid was synthesized according to a published procedure⁴² with the following small modification: in lieu of ethanol extraction, following saponification of the methyl amide, the neutralized and lyophilized salt was purified directly via HPLC. The product was lyophilized to a white powder. (Calculated mass for $\text{C}_9\text{H}_7\text{F}_4\text{NO}_2$: 237.1 Da, found 238.0 Da (M + 1)/236.0 Da (M - 1) for MS in positive and negative modes). Synthesis of 2,3,5,6 tetrafluoro, paramethyl Phe was carried out according to the report of Mikaye-Stoner et al.⁶⁵ MS revealed a mass of 252.05 in positive mode and 250.06 in negative mode, which agreed with the predicted mass of 251.06.

The other Phe analogs were acquired commercially as follows: **2F** Phe -Astatech, cat # 73308; **4F** Phe- Chem-impex cat # 02572; **2,4F** Phe-1ClickChemistry, cat #5C96757; **2,5F** Phe- Astatech, cat # 60350; **2,6F**-Phe HCl salt- Chem-impex, cat #24171; **3,5F** Phe Chem-impex, cat # 04123; **2,3,6F**-Phe- Astatech, A50355; **3,4,5F** Phe- Chem-impex, cat # 07394; **2,3,4,5F** Phe- Enamine, cat # en300-27751009; **2,3,4,5,6F** Phe- Chem-impex, cat # 07183.

PylRS library screen in *E. coli*

A plasmid library composed of the 20 canonical amino acids randomized at 5 sites (Asn 311, Cys 313, Val 366, Trp 382, and Gly 386) was used alongside positive (pREP-pylT) and negative selection (pYOB2-pylT) plasmids to select for aminoacyl-tRNA synthetases (RS) unique to pMe-2,3,5,6-tetrafluoro-Phe. The specific library used here has been used previously by our group⁶⁶. This library was built off the wild-type *M. barkeri* pyrrolysine synthetase with five sites (N311, C313, V366, W382, and G386) randomized to all 20 amino acids using NNK degenerate codons. In most of our selections to date, including those described in this manuscript, G386 remains unchanged, indicating it that is not tolerant to alteration while still maintaining synthetase function. Indeed, for all our described hits here, residue 386 is conserved to Gly. Because of this invariance, we chose to leave it out of the Table in Fig. 2. All incubations for selection are carried out at 37 °C for 16 h unless otherwise noted. All recovery steps last 1 h and are shaken at 250 rpm. Cultures are grown at 37 °C for 16 h and shaken at 250 rpm.

For positive selection, 500 μL of DH10B electrocompetent cells containing the positive pREP-pylT selection plasmid were transformed by electroporation with 1 μg of library plasmid. Cells were recovered in SOC media at 37 °C for 1 h. Recovered cells were serially plated from 10 to 10⁻⁶, and grown for 16 h at 37 °C to insure proper library coverage. Coverage was calculated to $\sim 150\times$, representing >99% of possible library members. The pooled recovery was used to inoculate 500 mL of LB media with 50 $\mu\text{g}/\text{mL}$ of Kanamycin (Kan) and 25 $\mu\text{g}/\text{mL}$ of tetracycline (Tet). The culture was grown overnight to saturation. This saturated culture was used to inoculate 500 mL of fresh LB media. This was grown to an OD₆₀₀ of 4.1, and 100 μL was plated on each of ten 15 cm LB-agar plates containing 50 $\mu\text{g}/\text{mL}$ Kan, 25 $\mu\text{g}/\text{mL}$ Tet, 40 $\mu\text{g}/\text{mL}$ chloramphenicol (Cm), and 1 mM nCAA. Plates were incubated overnight. To harvest the resulting cells, 5 mL of LB media was added to each plate. Colonies were then scraped from plates, pooled, and recovered for 1 h. Cells were then pelleted, and plasmid DNA was extracted using a Macherey-Nagel miniprep kit. The resulting DNA was then plasmid separated by isolating the library plasmid on a 0.8%

agarose gel via agarose gel electrophoresis and extracting using the Thermo Scientific GeneJET gel extraction kit.

For negative selection, 100 ng of plasmid DNA from the positive selection was transformed into 50 μL of DH10B cells containing PYOBB2-pylT using electroporation and recovered in 1 mL SOC media. Following recovery, 100 μL was plated on each of 3 LB-agar plates containing 50 $\mu\text{g}/\text{mL}$ Kan, 25 $\mu\text{g}/\text{mL}$ Cm, and 0.2% arabinose. After 16 h, cells were scraped and DNA prepped as described above.

To identify functional synthetase variants, 100 ng of the pBK DNA from negative selection was transformed into 25 μL of electrocompetent DH10B cells containing the reporter plasmid, pALS-pylT. This plasmid contains the sfGFP reporter with a TAG site at residue 150. In the presence of the selected nCAA, the TAG site will be suppressed, and the resulting colonies should appear green. Transformed cells were recovered in 1 mL SOC media. After recovery, cells were diluted by a factor of 100, and 100 μL of this dilution was plated on three 15 cm auto-inducing plates⁴⁰. All plates contain 50 $\mu\text{g}/\text{mL}$ Kan and 25 $\mu\text{g}/\text{mL}$ Tet; two contain 1 mM nCAA and the other is left without nCAA as a control. Plates were incubated overnight. After 16 h, plates were kept at RT for 24 h to further develop. After fully matured, colonies that appeared green were individually selected and used to inoculate 500 μL of non-inducing media⁶⁷ containing 50 $\mu\text{g}/\text{mL}$ Kan and 25 $\mu\text{g}/\text{mL}$ Tet in a 96-well block. This block was incubated for 20 h at 37 °C shaking at 300 rpm. After adequate growth, 20 μL of each well was used to inoculate two 96-well blocks with 500 μL of auto-inducing media^{40,67} with 50 $\mu\text{g}/\text{mL}$ Kan and 25 $\mu\text{g}/\text{mL}$ Tet. One block contained 1 mM nCAA while the other did not. After 24 h of incubation (37 °C, 300 rpm), sfGFP fluorescence was measured. The 30 highest-performing synthetases were sequenced, and 17 unique sequences were identified.

UP50 measurements

Electrocompetent DH10B cells containing the pALS reporter plasmid were transformed with isolated pBK plasmids containing the Phe_x synthetases. The transformants of eight synthetases, A11, A12, B5, B6, C4, C10, D6, and D7, were used to inoculate five 500 μL cultures each, consisting of auto-induction media with 50 $\mu\text{g}/\text{mL}$ Kan and 25 $\mu\text{g}/\text{mL}$ Tet containing no nCAA, 0.1 mM, 0.2 mM, 0.5 mM, or 1 mM nCAA. Note that we subsequently realized that C10 and D6 had the same sequence and thus further characterized D6. These were performed in duplicate, and fluorescence and OD₆₀₀ were measured after 24 h.

Permissivity screen (*E. coli*)

Electrocompetent DH10B cells containing the pALS reporter plasmid were transformed with isolated pBK plasmids containing the synthetases. Cultures of 5 mL were inoculated with single colonies of each synthetase in non-inducing media⁴⁰ with 50 $\mu\text{g}/\text{mL}$ kanamycin and 25 $\mu\text{g}/\text{mL}$ tetracycline and grown for 16 h. Noncanonical amino acid stocks were made at 100 \times their final concentration. Ninety-six-well blocks were prepared with 500 μL of auto-induction media containing antibiotics, and nCAAs with a final working concentration of 1 mM was added to their respective wells. Each well was then inoculated with a corresponding synthetase, yielding a final 96-well block where each synthetase was tested against each amino acid of interest. Cultures were grown in duplicate and fluorescence measurements were taken at 24 h. Previous work has shown that fluorination of aromatic residues encoded at position N150 does not substantially affect intrinsic sfGFP fluorescence⁶⁵.

Expression and purification of sfGFP (*E. coli*)

Protein sequences of sfGFP variants used in this study are given in Supplementary Fig. 5. Previously existing *E. coli* containing sfGFP150TAG and a given RS in non-inducing media were used to inoculate 5 mL cultures of non-inducing media containing 50 $\mu\text{g}/\text{mL}$ Kan and 25 $\mu\text{g}/\text{mL}$ Tet. Cultures were incubated for 20 h at 37 °C,

shaking at 250 rpm. After adequate growth, 500 μL of the 5 mL cultures were used to inoculate 50 mL cultures of auto-inducing media containing 1 mM ncAA. Cultures were incubated for 24 h (37 °C, 250 rpm), after which cultures were then spun down at 5000 rcf for 10 min and resuspended in 10 mL of a Tris buffer solution containing 100 mM Tris, 0.5 M NaCl, and 5 mM imidazole. Cells were lysed by micro-fluidization at 18,000 psi and centrifuged at 20,000 rcf for 30 min. sfGFP proteins in cell lysate were bound to TALON Metal Affinity Resin (bed volume of 100 μL) at 4 °C for 1 h. Lysate and bound resin were transferred to a gravity-flow column and the resin was washed with 30-bed volumes of buffer. Protein was eluted using the Tris buffer solution previously described supplemented with 200 mM imidazole. Protein concentration was assessed by comparing sfGFP fluorescence to a standard curve and submitted for ESI-MS.

As is apparent in the traces shown in Fig. 4, the deconvoluted mass spectra showed multiple masses for WT as well as mutant sfGFP. The spectra were analyzed in-house by Novatia (Promass) to yield relative intensities at given masses. In addition to the base (main) peak, both WT and mutant sfGFP displayed minor masses of -131 Da (lacking Methionine 1), -1378 Da, and -4165 Da. For the purpose of estimating encoding fidelity, these were considered part of the expected mass, and all other signals were considered non-expected mass. Note that the first time that sfGFP-N150TAG was expressed via Phe_{X-D6} + 2,3,6F Phe, a minor +36.9 Da peak was present in the spectra. This was classified as a K⁺ adduct of the expected mass by Promass software. It was suggested to us that this mass may have, in fact, been due to contamination of this sample by sfGFP with penta-fluoro Phe encoded (as a result of technical error). We re-expressed the sfGFP-N150TAG via Phe_{X-D6} + 2,3,6F Phe, and in this second purification, we did not observe the +36.9 Da peak. The trace for this sample is shown in Fig. 4. For Phe_{X-D6}, encoding fidelity for Penta-, 2,3,5,6-, 2,3,6-, and 2,6-fluoro Phe was 98.2%, 98.7%, 100.0%, and 95.0%, respectively. For Phe_{X-B5}, encoding fidelity for Penta-, 2,3,5,6-, 2,3,6-, 2,6-, and 2-fluoro Phe was 97.5%, 97.9%, 88.1%, 81.8%, and 95.6%, respectively.

Mammalian cell culture

HEK 293T cells (CRL 3216) were maintained in DMEM high glucose medium supplemented with Pen/Strep, L glutamine, and 10% FBS (Sigma). Cells were used in passages 5–35.

Construction of pAcBac1-Phe_X Plasmid

The pAcBac1.tR4-MbPyl, used to express MbPylRS in mammalian cells, was a gift from Peter Schultz (Addgene plasmid # 50832). pAcBac1.tR4-MbPyl and relevant human codon optimized RS DNA fragments were digested with NheI/EcoRI restriction enzymes, ligated and transformed into NEB Stable cells. Plasmid integrity from individual clones was initially checked by analytical restriction digestion, and then confirmed by sequencing.

Permissivity screen (mammalian)

HEK293T cells were seeded in 60 mm dishes so that they would be approximately 60–80% confluent the next day. We transfected the cells using 1.75 μg of the PacBac1-based synthetase plasmid (Phe_{X-D6} or Phe_{X-B5}) and 0.75 μg of sfGFP_N150TAG per dish. Polyjet was used (7.5 μL per dish, 2.5 mL media volume). Media were exchanged to contain ncAAs at a given concentration before transfection and approximately 16 h after transfection. Amino acids were solubilized either in NaOH or synthesized as free acids (penta-fluoro, 2,3,6 trifluoro, 3,4,5 trifluoro, 2,5 difluoro, 2,4 difluoro, 2 monofluoro) or in HCl for those synthesized as HCl salts (2,3,5,6 tetrafluoro, 2,3,4,5 tetrafluoro, 2,6 difluoro). Cells were imaged and harvested approximately 24 h after transfection. Images were taken with 20 \times objective using a Leica DFC9000GT camera with 700 ms exposure, without binning. Excitation was through an X-Cite 120 LED (Lumen Dynamics) set to

100% intensity. Images were taken in monochromatic mode using a GFP filter (Chroma, Inc. filter #41017) and pseudocolored to green. For bright field images, the light was adjusted manually. To harvest, cells were washed twice with ice-cold DPBS, then sloughed off the dishes in 1 mL ice-cold DPBS supplemented with Roche protease inhibitors. Cells were collected in microcentrifuge tubes and pelleted by centrifugation. The supernatant was removed and the cell pellets were flash-frozen in liquid nitrogen, then stored at -80 °C. To lyse, 350 μL of RIPA buffer (Sigma) plus Roche protease inhibitor tablets were added to each pellet. After cell debris and nuclei were cleared via centrifugation, GFP fluorescence was read from supernatants in 96-well plates; for each concentration of amino acid, at least six readings were made from at least two transfections. The level of fluorescence from lysates from untransfected cells averaged 21.7 ± 2.1 RFU. By comparison, cells transfected with sfGFP_N150TAG and Phe_{X-D6} averaged 26.9 ± 5.6 RFU, and Phe_{X-B5} averaged 43.1 ± 2.5 . The maximum background signal thus attributable to Phe_{X-B5} and Phe_{X-D6} was thus -5 RFU and -21.4 RFU, respectively.

Expression for purification and MS (mammalian)

HEK 293 T cells were expanded in 10 mm dishes. For a given synthetase/amino acid combination, the size of the transfection (the number of plates) was estimated from differences in relative sfGFP_N150TAG yield in the multi-specificity screens, with poorer encoders requiring more cell mass. Overall, 7–40 dishes were transfected with master mixes such that each dish received 3.5 μg of pACBAC1-Phe_X, 1.5 μg sfGFP_N150TAG, and 15 μL of PolyJET according to the manufacturer's instructions. Media was changed before transfection and approximately 18 hrs after transfection. Cells were harvested approximately 36 h after transfection. To harvest, each dish was washed twice with DPBS. Cells were harvested via scraping in ice-cold DPBS plus Roche protease inhibitors, on ice, pelleted by centrifugation, and flash-frozen in liquid nitrogen. Cells were lysed in hypotonic lysis buffer (10 mM Tris/HCl pH 8, Roche protease inhibitors, 0.1% Triton X 100 Sigma) via dounce homogenization on ice. Cell debris and nuclei were cleared via centrifugation. Clarified lysate was diluted 1:2 in Wash buffer (25 mM Tris/HCl pH 8, 20 mM Imidazole, 150 mM NaCl) and applied to pre-equilibrated Ni-NTA resin (Qiagen). The column was washed in at least 30 volumes of wash buffer, then the protein was eluted in elution buffer (25 mM Tris/HCl pH 8, 250 mM Imidazole, 150 mM NaCl). Elution was exchanged into 100 mM TEAB, prepared via 1:10 dilution of 1 M Stock (Thermo Fischer) in ultrapure water. Finally, the protein was concentrated to ~ 0.5 mg/ml, flash frozen, and submitted to Novatia Inc for ESI mass spec of the intact protein. The LC-UV traces from Novatia indicated that purified WT and mutant sfGFP existed as three predominant isoforms, a dominant peak, and two minor peaks representing $+532 \pm 1$ Da and -42 ± 1 Da mass differences (Supplementary Fig. 4). Note that the WT condition was co-transfected with Phe_{X-B5} synthetase, which did not affect the mass (within <1 Da of the same WT sfGFP expressed alone⁴⁰). As was also the case when expressed in *E.coli*, the deconvoluted mass spectra for the dominant LC peak showed multiple masses for WT, as well as mutant sfGFP, from HEK cells. The spectra were analyzed in-house by Novatia (Promass software) to yield relative intensities of the dominant mass and any other apparent peaks. WT and mutant sfGFP shared minor peaks of -42 Da (appx. 2% of integration), -1288 Da (appx. 2% of integration) -183 Da (appx. 3% of integration), and $+53$ Da (appx. 1% of integration) in common and thus, for the sake of estimation of encoding fidelity, these were counted among the expected mass. To be conservative, all other peaks were considered unexpected mass, even though they may not represent actual misincorporation. Fidelity for Phe_{X-D6} and Penta-, 2,3,4,5-, 2,3,5,6-, 2,3,6-, 2,6-, and 2,5-fluoro phenylalanine were estimated to be 97.5%, 100%, 98.2%, 98.5%, 97.6%, and 95.6%, respectively. For Phe_{X-B5}, fidelity was 96.2%, 97.4%, 100%, 100%, 100%, and 91.3% for the same species.

To accurately estimate sfGFP yield in a scenario wherein loss during purification is minimized, we transfected and harvested 31-10 cm dishes of HEKT cells with Phe_{X-D6} and sfGFP_{N150TAG} in the presence of 2 mM penta-fluoro Phenylalanine. sfGFP was purified as described above. Percent sfGFP was quantified using band detection and densitometry in Biorad Imagelab software (68% of lane signal).

Expression of ion channel constructs with the site-specific encoding of 2,3,6F Phe

pCMV-CFTR was a gift from Paul McCray (University of Iowa). WT CFTR-TAA (the existing TAG stop codon changed to TAA), F508 TAG CFTR (TAG amber codon introduced at F508 in the above WT construct), as well as WT and F1486TAG hNav_{1.5}strep_{HIS}strep were produced using standard methods and sequenced through the open reading frame. For expression of WT CFTR, 10 cm dishes of HEK 293 T cells were transfected with 1.5 µg WT CFTR-TAA and 0.25 µg WT GFP. For expression of WT hNav_{1.5}, 10 cm dishes of HEK 293 T cells were transfected with 1.5 µg of WT hNav_{1.5}strep_{HIS}strep and 0.25 µg WT GFP. For encoding of 2,3,6 F Phe, 10 cm dishes of HEK 293 T cells were transfected from master mixes of DNA such that each dish received 3.5 µg of Phe_{X-D6} and 0.25 µg of sfGFP_{N150TAG}. Depending on condition, 2 µg of F508TAG-CFTR or 1.5 µg of F1486TAG-Nav_{1.5}SHS were added. Media was changed twice (-16 h and -32 h post transfection), and cells were harvested -44 hrs post transfection and flash frozen in aliquots. Cell pellets were resuspended in RIPA (Sigma Aldrich) plus Roche protease inhibitors. Totally, 60 µg of protein from cleared, unconcentrated lysate (or a dilution as indicated) was loaded on 4–20% sodium dodecyl sulfate (SDS) gels. For Western blotting, gels were transferred to nitrocellulose membranes and probed with anti-CFTR (ab596 from Cystic Fibrosis Foundation, 1:1750) or anti-Nav_{1.5} (D9J7S, Cell Signaling, 1:5000) overnight in 5% milk in TBST. Blots were washed, probed with HRP secondaries (1:10,000), and imaged using Clarity ECL (Biorad). Blots were subsequently probed for beta-actin using HRP-conjugated primary (AC-15, Novus Biologicals 1:2000) appx 3 h in 5% milk in TBST. Blots were washed and imaged using Clarity ECL (Biorad). Densitometry (background subtracted integrated density) was done in ImageJ.

For patch clamp recording of hNav_{1.5}, in parallel, we co-transfected 10 cm dishes with 3.5 µg of Phe_{X-D6}, 0.25 µg of sfGFP_{N150TAG}, and 1.5 µg of either WT hNav_{1.5}strep_{HIS}strep or hNav_{1.5} F1486TAGstrep_{HIS}strep. Both conditions were cultured in 2 mM 2,3,6 trifluoro Phe throughout. Media was changed at 16 hrs and again when seeding onto 35 mm dishes for recording. Recordings were done approximately -42 h post transfection. The pipette solution contained 105 CsF, 35 NaCl, 10 EGTA, and 10 HEPES (pH 7.4). The bath contained (in mmol/L) 150 NaCl, 2 KCl, 1.5 CaCl₂, 1 MgCl₂, and 10 Hepes (pH 7.4). Pipette resistances were around 2 mΩ and series resistance was compensated ≥85%. For a generation of *I/V* curves, cells were held at -140 mV and pulsed in 5 mV depolarizing increments. For measurement of steady-state inactivation, cells were held at -140 mV and conditioned for 500 ms in 5 mV depolarizing increments. The test pulse was at -30 mV. Normalized *G/V* and SSI curves were fit to Boltzmann functions using Origin software.

For purification of hNav_{1.5} F1486(2,3,6 F Phe), ten 10 cm dishes of HEKT cells were co-transfected so that each dish received 1.5 µg pCDNA hNav_{1.5} F1486TAG_{FLAG}, 3.5 µg Phe_{X-D6}, and 0.25 µg sfGFP_{N150TAG} plasmids. Media changes were as described for the western blots. Channels were solubilized and purified similar to the preparation used previously to obtain the cryo-EM structure of the cardiac channel⁴⁷, with the following modifications. First, we used DDM/CHS as a detergent throughout rather than exchanging for GDN during washes. Secondly, the elution was concentrated and run directly on SDS page gel instead of being subjected to size exclusion chromatography. Gel band excision, tryptic digestion, and MS/MS were performed by the Iowa Proteomics Core using standard methods.

Quantum mechanical calculations of cation-π binding potential

All the models were prepared using GaussView 6 and all quantum chemical calculations were performed using Gaussian 16⁶⁸. Full geometry optimizations were carried out at the M06/6-31G(d,p) level of theory. In some fluorinated aromatic cases where the optimization resulted in geometries that were not considered cation-π interactions, the binding energy was obtained using the single-point method, as done previously²⁵. In this approach, full geometry optimization of the cation and the aromatic system is performed without the fluorination at the position which caused incompatible geometries upon optimization. Then fluorine is appended to the aromatic ring with a bond distance calculated from the aromatic system with fluorination at the desired position, which is optimized in isolation. The energy of this system is then calculated and used to find the binding energy in the single-point calculations.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data from this study are included in this published article, associated source data file, and supplementary information document. Source data are provided in this paper.

References

1. Dougherty, D. A. Cation-π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163–168 (1996).
2. Schottel, B. L., Chifotides, H. T. & Dunbar, K. R. Anion-π interactions. *Chem. Soc. Rev.* **37**, 68–83 (2008).
3. Dougherty, D. A. The cation-π interaction. *Acc. Chem. Res.* **46**, 885–893 (2013).
4. Weber, D. S. & Warren, J. J. The interaction between methionine and two aromatic amino acids is an abundant and multifunctional motif in proteins. *Arch. Biochem. Biophys.* **672**, 108053 (2019).
5. Ringer, A. L., Senenko, A. & Sherrill, C. D. Models of S/π interactions in protein structures: comparison of the H₂S benzene complex with PDB data. *Protein Sci.* **16**, 2216–2223 (2007).
6. Daeffler, K. N., Lester, H. A. & Dougherty, D. A. Functionally important aromatic-aromatic and sulfur-π interactions in the D₂ dopamine receptor. *J. Am. Chem. Soc.* **134**, 14890–14896 (2012).
7. Burley, S. K. & Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **203**, 139–143 (1986).
8. Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W. & Huang, R. B. The multiple roles of histidine in protein interactions. *Chem. Cent. J.* **7**, 44 (2013).
9. Levitt, M. & Perutz, M. F. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **201**, 751–754 (1988).
10. Pace, C. J. & Gao, J. Exploring and exploiting polar-π interactions with fluorinated aromatic amino acids. *Acc. Chem. Res.* **46**, 907–915 (2013).
11. Lucas, X., Bauza, A., Frontera, A. & Quinonero, D. A thorough anion-π interaction study in biomolecules: on the importance of cooperativity effects. *Chem. Sci.* **7**, 1038–1050 (2016).
12. Infield, D. T. et al. Cation-π interactions and their functional roles in membrane proteins. *J. Mol. Biol.* **433**, 167035 (2021).
13. Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
14. Yen, T. J., Lolicato, M., Thomas-Tran, R., Du Bois, J. & Minor, D. L. Jr. Structure of the saxiphilin:saxitoxin (STX) complex reveals a convergent molecular recognition strategy for paralytic toxins. *Sci Adv.* **5**, eaax2650, (2019).
15. Rahman, M. M. et al. Structure of the native muscle-type nicotinic receptor and inhibition by snake venom toxins. *Neuron* **106**, 952–962 e955 (2020).

16. Plenge, P. et al. The antidepressant drug vilazodone is an allosteric inhibitor of the serotonin transporter. *Nat. Commun.* **12**, 5063 (2021).
17. Silberberg, J. M. et al. Deciphering ion transport and ATPase coupling in the intersubunit tunnel of KdpFABC. *Nat. Commun.* **12**, 5098 (2021).
18. Yang, D. & Gouaux, E. Illumination of serotonin transporter mechanism and role of the allosteric site. *Sci. Adv.* **7**, eabl3857 (2021).
19. Kumar, A. et al. Mechanisms of activation and desensitization of full-length glycine receptor in lipid nanodiscs. *Nat. Commun.* **11**, 3752 (2020).
20. McDaniel M. J. et al. NMDA receptor channel gating control by the pre-M1 helix. *J. Gen. Physiol.* **152**, 4 (2020).
21. Kise, Y. et al. Structural basis of gating modulation of Kv4 channel complexes. *Nature* **599**, 158–164 (2021).
22. Diwanji, D. et al. Structures of the HER2-HER3-NRG1beta complex reveal a dynamic dimer interface. *Nature* **600**, 339–343 (2021).
23. Waheed, Q. et al. Interfacial aromatics mediating cation-pi interactions with choline-containing lipids can contribute as much to peripheral protein affinity for membranes as aromatics inserted below the phosphates. *J. Phys. Chem. Lett.* **10**, 3972–3977 (2019).
24. Last NB, Sun S, Pham MC, Miller C. Molecular determinants of permeation in a fluoride-specific ion channel. *Elife* **6**, e31259 (2017).
25. Davis, M. R. & Dougherty, D. A. Cation-pi interactions: computational analyses of the aromatic box motif and the fluorination strategy for experimental evaluation. *Phys. Chem. Chem. Phys.* **17**, 29262–29270 (2015).
26. Mecozzi, S., West, A. P. & Dougherty, D. A. Cation-pi interactions in simple aromatics: electrostatics provide a predictive tool. *J. Am. Chem. Soc.* **118**, 2307–2308 (1996).
27. Leisle, L., Valiyaveetil, F., Mehl, R. A. & Ahern, C. A. Incorporation of non-canonical amino acids. *Adv. Exp. Med. Biol.* **869**, 119–151 (2015).
28. Santarelli, V. P., Eastwood, A. L., Dougherty, D. A., Horn, R. & Ahern, C. A. A cation-pi interaction discriminates among sodium channels that are either sensitive or resistant to tetrodotoxin block. *J. Biol. Chem.* **282**, 8044–8051 (2007).
29. Santarelli, V. P., Eastwood, A. L., Dougherty, D. A., Ahern, C. A. & Horn, R. Calcium block of single sodium channels: role of a pore-lining aromatic residue. *Biophys. J.* **93**, 2341–2349 (2007).
30. Pless, S. A., Galpin, J. D., Frankel, A. & Ahern, C. A. Molecular basis for class Ib anti-arrhythmic inhibition of cardiac sodium channels. *Nat. Commun.* **2**, 351 (2011).
31. Ahern, C. A., Eastwood, A. L., Dougherty, D. A. & Horn, R. Electrostatic contributions of aromatic residues in the local anesthetic receptor of voltage-gated sodium channels. *Circ. Res.* **102**, 86–94 (2008).
32. Zhong, W. et al. From ab initio quantum mechanics to molecular neurobiology: a cation-pi binding site in the nicotinic receptor. *Proc. Natl Acad. Sci. USA* **95**, 12088–12093 (1998).
33. Padgett, C. L., Hanek, A. P., Lester, H. A., Dougherty, D. A. & Lummis, S. C. Unnatural amino acid mutagenesis of the GABA(A) receptor binding site residues reveals a novel cation-pi interaction between GABA and beta 2Tyr97. *J. Neurosci.* **27**, 886–892 (2007).
34. Budisa N. *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins.* (Wiley-VCH, 2006).
35. Brown, W., Liu, J. & Deiters, A. Genetic code expansion in animals. *ACS Chem. Biol.* **13**, 2375–2386 (2018).
36. Lee, Y. J. et al. Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader. *Chem. Commun.* **52**, 12606–12609 (2016).
37. Hao, B. et al. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**, 1462–1466 (2002).
38. Arbely, E., Torres-Kolbus, J., Deiters, A. & Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134**, 11912–11915 (2012).
39. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 (2002).
40. Galles, G. D., Infield, D. T., Mehl, R. A. & Ahern, C. A. Selection and validation of orthogonal tRNA/synthetase pairs for the encoding of unnatural amino acids across kingdoms. *Methods Enzymol.* **654**, 3–18 (2021).
41. Cooley, R. B., Karplus, P. A. & Mehl, R. A. Gleaning unexpected fruits from hard-won synthetases: probing principles of permissivity in non-canonical amino acid-tRNA synthetases. *ChemBiochem* **15**, 1810–1819 (2014).
42. Zheng, H., Comeforo, K. & Gao, J. Expanding the fluorine arsenal: tetrafluorinated phenylalanines for protein design. *J. Am. Chem. Soc.* **131**, 18–19 (2009).
43. Chatterjee, A., Xiao, H., Bollong, M., Ai, H. W. & Schultz, P. G. Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl Acad. Sci. USA* **110**, 11803–11808 (2013).
44. Rowe, S. M., Miller, S. & Sorscher, E. J. Cystic fibrosis. *N. Engl. J. Med.* **352**, 1992–2001 (2005).
45. Rivaud, M. R., Delmar, M. & Remme, C. A. Heritable arrhythmia syndromes associated with abnormal cardiac sodium channel function: ionic and non-ionic mechanisms. *Cardiovasc. Res.* **116**, 1557–1570 (2020).
46. Remme, C. A. & Bezzina, C. R. Sodium channel (dys)function and cardiac arrhythmias. *Cardiovasc. Ther.* **28**, 287–294 (2010).
47. Jiang, D. et al. Structure of the cardiac sodium channel. *Cell* **180**, 122–134 e110 (2020).
48. Cai, Z., Sohma, Y., Bompadre, S. G., Sheppard, D. N. & Hwang, T. C. Application of high-resolution single-channel recording to functional studies of cystic fibrosis mutants. *Methods Mol. Biol.* **741**, 419–441 (2011).
49. Gunderson, K. L. & Kopito, R. R. Conformational states of CFTR associated with channel gating: the role ATP binding and hydrolysis. *Cell* **82**, 231–239 (1995).
50. Denning, G. M. et al. Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive. *Nature* **358**, 761–764 (1992).
51. Chakrabarti, S. et al. MOG1 rescues defective trafficking of Na(v)1.5 mutations in Brugada syndrome and sick sinus syndrome. *Circ. Arrhythm. Electrophysiol.* **6**, 392–401 (2013).
52. Liu, F. et al. Structural identification of a hotspot on CFTR for potentiation. *Science* **364**, 1184–1188 (2019).
53. Pan X. et al. Structure of the human voltage-gated sodium channel Nav1.4 in complex with beta1. *Science* **362**, eaau2486 (2018).
54. Shen H. et al. Structural basis for the modulation of voltage-gated sodium channels by animal toxins. *Science* **362**, eaau2596 (2018).
55. Liu, F., Zhang, Z., Csanady, L., Gadsby, D. C. & Chen, J. Molecular structure of the human CFTR ion channel. *Cell* **169**, 85–95 e88 (2017).
56. Zhang, Z. & Chen, J. Atomic structure of the cystic fibrosis transmembrane conductance regulator. *Cell* **167**, 1586–1597 e1589 (2016).
57. Laselva, O. et al. Identification of binding sites for ivacaftor on the cystic fibrosis transmembrane conductance regulator. *iScience* **24**, 102542 (2021).
58. Veit G, et al. Allosteric folding correction of F508del and rare CFTR mutants by elxacaftor-tezacaftor-ivacaftor (Trikafta) combination. *JCI Insight* **5**, e139983 (2020).
59. Serohijos, A. W. et al. Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. *Proc. Natl Acad. Sci. USA* **105**, 3256–3261 (2008).
60. Fiedorczuk, K. & Chen, J. Mechanism of CFTR correction by type I folding correctors. *Cell* **185**, 158–168 e111 (2022).

61. Kellenberger, S., West, J. W., Scheuer, T. & Catterall, W. A. Molecular analysis of the putative inactivation particle in the inactivation gate of brain type IIA Na⁺ channels. *J. Gen. Physiol.* **109**, 589–605 (1997).
62. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **1844**, 1059–1070 (2014).
63. Danielson, M. A. & Falke, J. J. Use of 19F NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 163–195 (1996).
64. Awad, L. F. & Ayoub, M. S. Fluorinated phenylalanines: synthesis and pharmaceutical applications. *Beilstein J. Org. Chem.* **16**, 1022–1050 (2020).
65. Miyake-Stoner, S. J. et al. Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry* **49**, 1667–1677 (2010).
66. Jang, H. S., Jana, S., Blizzard, R. J., Meeuwse, J. C. & Mehl, R. A. Access to faster eukaryotic cell labeling with encoded tetrazine amino acids. *J. Am. Chem. Soc.* **142**, 7245–7249 (2020).
67. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
68. Frisch M.J., et al. *Gaussian 16 Rev. C.01*. (2016).

Acknowledgements

The authors thank members of the Ahern and Mehl lab for helpful discussions. The work was supported by National Institutes of Health (NIH) grants R01-GM128420 and NINDS R24 NS104617 (to C.A.A.), R01-GM123455 and P41-GM104601 (to E.T.). R.B.C. and R.A.M. were supported in part by the GCE4All Biomedical Technology Development and Dissemination Center supported by the National Institute of General Medical Science grant RM1-GM144227.

Author contributions

D.T.I., E.T., R.B.C., R.A.M., and C.A.A. designed experiments, G.D.G., D.T.I., C.J.C., M.L.H., S.M., F.F., A.R., and J.D.G., performed experiments and analyzed the data; D.T.I. and C.A.A. prepared an initial draft and all authors edited the paper.

Competing interests

C.A.A., D.T.I., R.B.C., and R.A.M. are co-founders of Halide Biologics. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-35761-w>.

Correspondence and requests for materials should be addressed to Christopher A. Ahern.

Peer review information *Nature Communications* thanks Chloe Martens and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023