C h a p t e r I

NON-CANONICAL PROLINE RESIDUES IN PROTEIN SCIENCE AND ENGINEERING

1.1 Contributions

Stephanie L. Breunig and David A. Tirrell contributed to the writing of this text.

1.2 Abstract

Proline residues are unique in the extent to which they constrain the conformational space available to the protein backbone. Because the conformational preferences of proline cannot be recapitulated by any of the other proteinogenic amino acids, standard mutagenesis approaches that seek to introduce new chemical functionality at proline positions unavoidably perturb backbone flexibility. Here, we describe some unique conformational properties of proline and non-canonical proline variants, and their use in modifying and understanding protein behavior. We further detail the incorporation of proline analogs into recombinant proteins in *Escherichia coli* via a residue-specific mutagenesis strategy. The ability of proline analogs to maintain backbone conformational restraints while introducing new chemistries grant them a useful and intriguing role in protein science and engineering endeavors.

1.3 Main text

1.3.1 The conformational properties of proline

As the sole α-imino acid among the canonical building blocks of proteins, proline is unique. Its pyrrolidine ring imposes conformational constraints on the polypeptide backbone that are essential to protein structure and function (Figure 1.1a,b). The backbone dihedral angle ϕ of proline residues is restricted to 63±15° (Ref. 1), limiting the conformational trajectories that are available to the polypeptide chain. The reduced conformational space sampled by the peptide backbones of proline residues is illustrated by the Ramachandran plots for proline and the acyclic amino acids (Figure 1.1a,b). The pyrrolidine side chain is also limited in its conformational flexibility, and generally adopts one of only two rapidly-interconverting ring puckers: C^{γ}-*endo* or C^{γ}-*exo*[2] (Figure 1.1c-d). The conformation of the five-membered ring can be described quantitatively through the concept of pseudorotation, in which the two defining parameters, phase angle and amplitude, are calculated from the pyrrolidine torsional angles.[3,4] The highly populated phase angles of the C^{γ}-*endo* and -*exo* puckers are approximately 198° and 18°, respectively; intermediate phase angles describe other higher energy envelope and twist conformations.

Because the proline amide linkage cannot serve as a hydrogen-bond donor, the presence of proline residues within α-helical or β-strand structural motifs is disfavored.[5] Proline's *cis* and *trans* conformers are nearly isoenergetic, and the barrier to *cis*-*trans* isomerization is reduced in comparison to the other amino acids. Consequently, *cis* isomers are more common for proline than for any other canonical amino acid,[1] and *cis*-*trans* isomerization at proline residues can play important roles in protein folding[6] and function.[7]

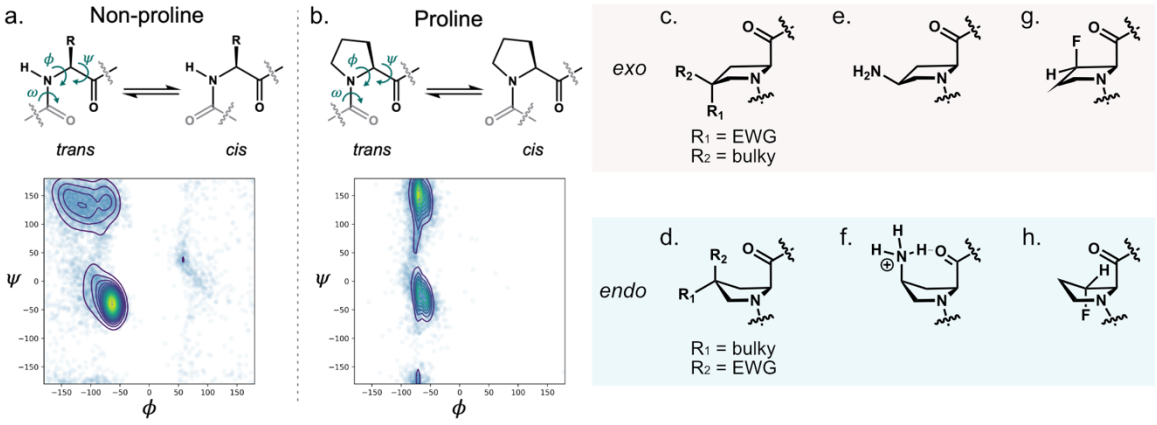


Figure 1.1. **Proline conformational preferences**. **a-b**. Backbone conformations of, and Ramachandran plots for, non-proline (**a**) and proline (**b**) residues found in structural data deposited in the Protein Data Bank (PDB). The torsional angles ϕ, φ, and ω, along with *cis* and *trans* amide isomers are indicated in the amino acid structures. **c-h**. *Exo* and *endo* ring pucker preferences for select proline analogs. ncPro residues with 4*R*- electron-withdrawing groups or bulky 4*S*- substituents prefer the *exo* pucker (c); those with 4*S*- electron-withdrawing groups or bulky 4*R*- substituents prefer *endo* (d). In basic conditions, 4*S*-NH_2 prefers the pseudoequatorial position in the *exo* ring pucker (e). When protonated, the ammonium group functions as an electron-withdrawing substituent; a transannular hydrogen bond further stabilizes the *endo* pucker (f). 3*R*-F (g) and 3*S*-F (h) prefer the *exo* and *endo* puckers, respectively, due to a gauche effect.

1.3.2 Non-canonical proline residues

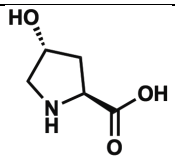
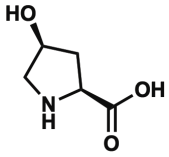
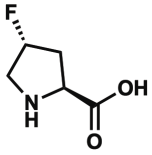
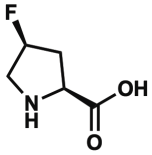
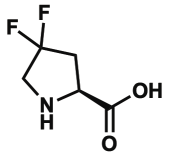
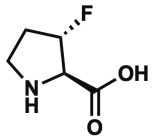
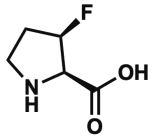
Non-canonical amino acids (ncAAs) have found extensive use in chemical biology and related fields.[8–10] They serve many roles, such as providing chemical handles for protein modification,[11] serving as probes in time-resolved and cell-selective proteomic analyses,[12] interrogating the effects of post-translational modifications,[13] identifying protein-protein interaction partners,[14] tracking protein location in vivo,[15] and probing the importance of non-covalent interactions in protein behavior.[16]

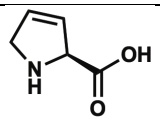
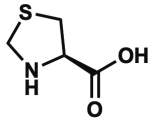
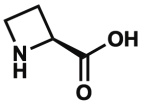
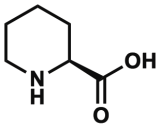
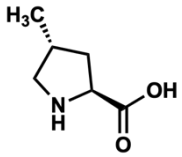
The utility of standard mutagenesis at proline sites is limited by its impact on chain conformation. ncAAs provide a means of addressing this limitation, as replacement of proline by non-canonical analogs allows introduction of new chemical functionality while maintaining conformational constraints. For instance, the hydroxyl groups of the non-canonical proline (ncPro) variants 4*R*-hydroxyproline (4*R*-OH) and 4*S*-hydroxyproline (4*S*-OH, Table 1.1) permit hydrogen-bonding interactions; one such hydrogen bond has been suggested to alter the behavior of an engineered insulin variant.[17] Expanding or contracting the five-membered pyrrolidine ring can add (piperidine-2-carboxylic acid, Pip, Table 1.1) or remove (azetidine-2-carboxylic acid, Aze, Table 1.1) hydrophobic packing interactions.[18] The alkene functionality in 3,4-dehydroproline (Dhp, Table 1.1) has been used as a chemical handle to modify protein-based materials.[19] Fluorinated proline variants can be used as conformational reporters in ^{19}F NMR experiments.[20] For instance, *cis*-*trans* isomerization of the sole 4,4-difluoroproline (44-diF, Table 1.1) residue in a modified β2 microglobulin could be monitored by ^{19}F NMR.[21]

Many ncPro analogs have well-documented conformational biases; we here describe the conformations of a selection of proline analogs, though others are extensively reviewed elsewhere.[2,22] Most widely studied are proline variants containing substituents at the C^{γ} position. The presence of a 4*R*- electron-withdrawing substituent (as in 4*R*-OH and 4*R*-fluoroproline, 4*R*-F; Table 1) stabilizes the C^{γ}-*exo* ring pucker through a gauche effect (Figure 1.1c). The *exo* ring pucker, in turn, pre-organizes the amide in the *trans* conformation. Conversely, ncPro residues with 4*S*- electron-withdrawing groups (such as 4*S*-OH and 4*S*-fluoroproline, 4*S*-F; Table 1.1) favor the *endo* ring pucker (Figure 1.1d),

and have a higher propensity for the *cis* amide isomer compared to canonical proline. The presence of a 4-fluoro substituent (and presumably any electron-withdrawing group) lowers the energy barrier to *cis*-*trans* isomerization, as the substituent decreases the bond order of the preceding amide.[23]

Table 1.1: Proline analogs used for residue-specific incorporation into recombinant proteins

Proline analog	Proline structure	ProRS variant	Reported [NaCl] (M)	Reported level of proline replacement	Reference
4*R*-OH		wt	0.3	88%	17
4*S*-OH		wt	0.3	91%	17
4*R*-F		wt	0.3	96%	24
4*S*-F		wt	0.3	97%	24
44-diF		wt	0.3	86%	24
3*S*-F		–	–	“virtually complete”	25
3*R*-F		–	–	“virtually complete”	25

Proline analog	Proline structure	ProRS variant	Reported [NaCl] (M)	Reported level of proline replacement	Reference
dhp		wt	0.3	≤100%	18
Thz		wt	0.3	90%	18
Aze		wt	0.3	≤100%	18
Pip		C443G	0.3	89%	18
4*R*-Me		C443G	0.6	>60%	26

Sterically-demanding groups at the C^{γ} position instead prefer the pseudoequatorial position, rather than pseudoaxial favored by electron-withdrawing substituents. 4*R*-methylproline (4*R*-Me) thus enforces the *endo* ring pucker and *cis* amide isomer compared to proline (Figure 1.1d), while the 4*S* diastereomer (4*S*-Me) prefers the *exo* pucker and *trans* amide (Figure 1.1c).[27] Similar effects were observed for the 4-mercaptoproline derivatives,[28] though their conformational biases are less pronounced. Other bulky substituents at the C^{γ} position are expected to behave similarly.

One especially interesting 4-substituted proline analog is 4*S*-aminoproline (4*S*-NH_2).[29] Changes in amine protonation state over physiologically-relevant pH ranges alter the ring puckering preferences of this analog: under basic conditions, sterics dominate, and the *exo* conformation is preferred (Figure 1.1e). The ammonium formed upon protonation behaves

as an electron-withdrawing substituent, and a transannular hydrogen bond with the backbone carbonyl further reinforces the preference for the *endo* pucker (Figure 1.1f). In the case of 4*S*-NH_2, both protonation states prefer the *trans* amide isomer. These properties of 4*S*-NH_2 have led to pH-responsive collagen-like peptides.[29]

Functionalization at the C^{β} position, as in 3*R*-F and 3*S*-F, similarly leads to conformational biases relative to proline. 3*R*-F prefers the C^{γ}-*exo* pucker (Figure 1.1g), and 3*S*-F the *endo* (Figure 1.1h).[25] The conformational preferences of 3*R*-F and 3*S*-F prolines are opposed to those of prolines with 4-fluoro substituents (Figure 1.1c-d). In the context of a polypeptide chain, α-methylproline (2-Me) exists nearly exclusively as the *trans* amide isomer, due to steric clash between the α-substituent and the side chain of the preceding residue.[30] The backbone trajectories of 2-Me are even more constrained than that of proline: this residue overwhelmingly prefers the 3_{10}/α-helical conformation and promotes β-turns at the $i + 1$ corner position.[31]

1.3.3 Proline analogs in modifying protein behavior

The established conformational preferences of proline analogs provide a means of interrogating the effects of conformational properties (such as pyrrolidine ring pucker or amide isomerization) on protein structure and function.[22] For instance, ncPro residues have been used to identify a key proline *cis-trans* isomerization event in ion channel opening,[7] demonstrate the stereoelectronic basis of collagen stability,[2] and modulate the properties of protein–based materials.[32]

In many cases in which the conformational preference of the ncPro residue matches that in the native protein structure, the protein fold is stabilized.[33–36] For instance, in the first example of the biosynthetic incorporation of fluorinated proline residues, barstar Pro46 (*cis* in the native protein) was replaced with 4*R*-F, 4*S*-F, and 44-diF.[36] Pro46 exists in the *cis* conformation in the native protein. Satisfyingly, the stability of the fluorinated barstar variants correlated well with the *cis*-*trans* isomerization preferences of the individual proline analogs.

However, protein behavior is complex, and not all reported examples are straightforward.[37–41] In its native fold, thioredoxin (Trx) contains a proline residue in the *cis* conformation; *trans* to *cis* isomerization is the rate-limiting step in Trx folding, and is essential for Trx function. Despite the opposed *cis*-*trans* preferences of 4*R*-F and 4*S*-F, proline monofluorination destabilized Trx independent of stereochemistry.[39] 4,4-difluorination, which should lower the energy barrier to *cis*-*trans* isomerization,[36] did not accelerate Trx folding.[41] In another context, the two proline residues in the foldon domain were replaced with six proline analogs carrying C^{γ} substituents; changes in stability did not correlate with known *cis*-*trans* isomerization preferences.[38]

β2 microglobulin (β2m) fibrillation was probed by proline mutagenesis in an example that exemplifies the complexity of protein biophysics. Pro32, which exists in the *cis* conformation in the native state, but *trans* in the fibril, was replaced with 4*R*-F, 4*S*-F, and 44-diF.[21] Stability against chemical denaturation correlated well with the established *cis*-*trans* isomerization preferences of the proline analogs, suggesting that the backbone

conformation at Pro32 is important for β2m behavior. Interestingly, 44-diF, which exhibits similar *cis*-*trans* isomer preferences as proline, led to increased structural disorder and the greatest propensity to form amorphous aggregates. This effect was attributed to a lowered energy barrier to *cis*-*trans* isomerization, rather than a thermodynamic preference for one isomer. The authors observed diverse fibril morphologies and fibrillation behaviors after seeding among the fluorinated β2m variants.[21]

In a follow-up study, Pro32 was replaced with 2-Me.[42] This residue's enhanced preference for the *trans* isomer was hypothesized to promote β2m fibril formation. While α-methylation of β2m did lead to greater conformational flexibility of the protein and promoted oligomerization, fibril formation was not observed unless the sample was seeded with pre-formed fibrils. Together, these studies demonstrate that β2m fibrillation is more complex than simple association of *trans*-Pro32 β2m conformers, and highlight the unexpected effects that functionalized proline variants can have on protein behavior.[21,42]

The effect of proline analogs becomes more complex when many proline residues are modified. Global proline mutagenesis has accelerated the folding of EGFP[43] and mRFP1,[44] and stabilized a single-chain Fv fragment.[45] Simultaneous incorporation of 4*S*-F (replacing six proline residues), 4-fluoro-phenylalanine (sixteen phenylalanine residues) and 6-fluorotryptophan (2 tryptophan residues) into lipase did not significantly affect the protein's structure or catalytic activity.[46] Incredibly, replacing 32 proline residues in the KlenTaq DNA polymerase with 4*R*-F led to an active polymerase with no appreciable differences in activity,[47] though thermal stability was modestly affected.[48] In many of these

examples, incorporation of the opposite ncPro diastereomer led to insolubility[43,44] or apparent instability[26,47] during recombinant expression. Due to the number of proline residues modified in these cases, it is difficult to attribute changes in protein behavior to any single effect.

1.3.4 Incorporation of non-canonical proline residues into polypeptide chains

Virtually any ncAA can be introduced into a polypeptide chain through chemical means (i.e., solid phase peptide synthesis). However, chemical synthesis approaches are typically limited by polypeptide length and scalability, and restrict the ability to use ncAAs *in vivo*. Alternatively, ncAAs can be introduced into ribosomally-synthesized proteins by hijacking a cell's translational machinery. Two general strategies have been developed for *in vivo* introduction of ncAAs into proteins: site-specific and residue-specific replacement. Both approaches rely on the availability of aminoacyl-tRNA synthetases (aaRSs) able to charge their cognate tRNAs with the ncAAs of interest (Figure 1.2a). Site-specific approaches (including nonsense suppression and related techniques) cause minimal perturbation of protein sequence, and important advances in such methods (for example, Ref. 49–51) have been accomplished since the first report of incorporation in *Escherichia coli*.[52] In site-specific ncAA mutagenesis, a re-assigned codon (most often the amber stop codon) is matched with the ncAA of interest by an engineered, orthogonal tRNA/aaRS pair. The resulting ncAA-charged tRNA competes with release factors for amber codon recognition within the ribosomal complex, and successful translational read-through positions the ncAA at the amber site.[8] Challenges in implementing site-specific methods include the formation of truncation products that reduce recombinant protein yields, especially when

ncAAs are incorporated at multiple positions; and the requirement for development of orthogonal tRNA/aaRS pairs, which can be difficult.

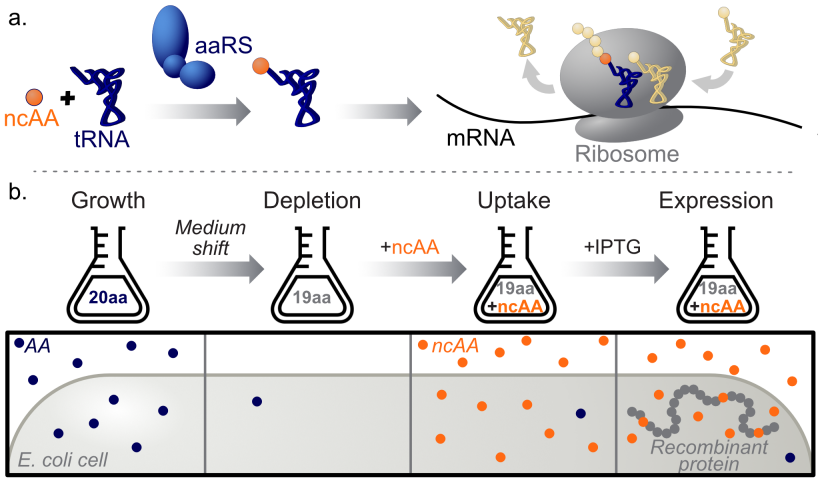


Figure 1.2. **Non-canonical amino acid (ncAA) mutagenesis**. **a**. For both site-specific and residue-specific approaches, incorporation of ncAA residues into recombinant proteins relies on the ability of the translational machinery of the host to accommodate the ncAA of interest. In many cases, the limiting step is aminoacylation by the relevant aminoacyl-tRNA synthetase (aaRS, left). Upon aminoacylation, the charged tRNA is delivered to the ribosome, which adds the ncAA to the growing polypeptide chain (right). **b**. Residue-specific ncAA mutagenesis workflow. After growth in medium that contains the canonical amino acid (AA) to be replaced, a shift to AA-depleted medium is performed. After addition of the ncAA of interest, expression of the recombinant protein of interest is induced. The ncAA replaces the canonical AA residues in all newly synthesized proteins, including the recombinant protein.

In contrast to site-specific methods, residue-specific ncAA mutagenesis[9] results in the global replacement of a canonical amino acid with a non-canonical counterpart. This technique typically uses defined expression media and amino acid auxotrophs (i.e., strains deficient in biosynthesis of the amino acid of interest) as expression hosts. A generalized workflow for residue-specific ncAA mutagenesis is depicted in Figure 1.2b. In some cases,

the endogenous aaRS is promiscuous enough to allow the desired substitution;[12] in other cases, overexpression of the endogenous aaRS[53] or expression of a mutant aaRS is required.[54] Residue-specific replacement often requires less genetic manipulation than site-specific techniques, and can produce high yields of the recombinant protein of interest: typical yields are 50-60% of those obtained from expression in media that contain the canonical amino acid.

Residue-specific ncAA mutagenesis has been used in a variety of contexts. Global incorporation of the methionine analog azidohomoalanine (Aha) enables 'bio-orthogonal non-canonical amino acid tagging' (BONCAT), a proteomic method capable of enriching for newly synthesized proteins.[12] Various methionine analogs have also been incorporated by residue-specific methods to study purified proteins, including elastin-based biomaterials[55] and prion proteins.[56]. Leucine analogs have been used to alter the properties of coiled-coil proteins[57] and introduced into GFP.[58]

Global incorporation of proline analogs has been used to modulate the properties of purified, recombinant proteins. Many of the examples of proline mutagenesis in full-length proteins discussed above were generated by recombinant expression. The first example of biosynthetic ncPro incorporation in recombinant proteins expressed in *E. coli* was the replacement of five proline residues in human annexin V with 1,3-thiazolidine-4-carboxylic acid (Thz).[59] Since then, additional proline variants have been reported to show translational activity in *E. coli*.[23,60] The most commonly used analogs are 4*R*-F and 4*S*-F, which have been found to alter the folding and fluorescence properties of fluorescent

proteins[43,44] the melting temperatures of elastin-mimetic peptides.[25,32] Similar approaches have been used to incorporate proline analogs into recombinantly-expressed barstar,[23] ubiquitin,[35] thioredoxin,[26,39] DNA polymerase,[47] and 4-oxalocrotonate tautomerase.[61] We have found that replacement of proline B28 of human insulin by a variety of proline analogs can be used to engineer the therapeutically-relevant biophysical properties of the protein.[17,18] Because no aaRS/tRNA pair capable of site-specific incorporation of proline analogs has been described to date, residue-specific ncAA mutagenesis provides the only option currently available for introducing modified proline residues into recombinant proteins produced in living cells.

The *E. coli* prolyl-tRNA synthetase (ProRS) and downstream translational machinery have been shown to accommodate a variety of structural analogs of proline. For these analogs, simple overexpression of the *E. coli* ProRS (or point mutants thereof), combined with increased expression of proline transporters under hyperosmotic conditions,[62] enables high levels of proline replacement.[60] Proline analogs that have been incorporated into recombinant proteins are detailed in Table 1.1, along with reported expression conditions.

1.4 Overview of thesis chapters

In this thesis, I discuss the residue-specific incorporation of proline analogs into proteins expressed in *E. coli*. Chapters II and III build upon previous work in our lab[17,18,24,63] to engineer the therapeutic protein insulin by proline mutagenesis. In Chapter II, we replace ProB28 of human insulin with three aliphatic proline analogs and evaluate their effect on insulin biophysics; the biosynthetic incorporation of two of these analogs had not

previously been reported. In Chapter III, we expand these proline mutagenesis efforts to the fast-acting insulin lispro, replacing ProB29 with a set of 4-fluorinated proline residues. These two chapters demonstrate the utility of ncAA mutagenesis in probing and modifying the therapeutically-relevant properties of protein drugs.

Chapter IV characterizes the incorporation of photo-proline, a photo-activatable proline analog, into proteins in *E. coli*. We also describe initial attempts in studying the substrates of peptidyl-prolyl isomerases in *E. coli* using this residue. Finally, proline analog incorporation has thus far been fairly limited with respect to the diversity of proline analogs accepted by *E. coli*. In Chapter V, we describe our efforts to engineer the *Ec*ProRS to recognize an expanded set of ncPro residues. Together, this work expands the number of proline analogs incorporated into recombinant proteins, and demonstrates their utility in studying and engineering protein behavior.

1.5 References

(1) MacArthur, M. W.; Thornton, J. M. Influence of Proline Residues on Protein Conformation. *J Mol Biol* **1991**, *218*, 397–412.

(2) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu Rev Biochem* **2009**, *78*, 929–958. https://doi.org/10.1146/annurev.biochem.77.032207.120833.

(3) Park, H. S.; Kang, Y. K. Puckering Transition of the Proline Residue along the Pseudorotational Path: Revisited. *New J Chem* **2021**, *45* (22), 9780–9793. https://doi.org/10.1039/D1NJ01361K.

(4) DeTar, D. L. F.; Luthra, N. P. Conformations of Proline. *J Am Chem Soc* **1977**, *99* (4), 1232–1244. https://doi.org/10.1021/ja00446a040.

(5) Levitt, M. Conformational Preferences of Amino Acids in Globular Proteins. *Biochemistry* **1978**, *17* (20), 4277–4285. https://doi.org/10.1021/BI00613A026.

(6) Lu, K. P.; Finn, G.; Lee, T. H.; Nicholson, L. K. Prolyl *Cis-Trans* Isomerization as a Molecular Timer. *Nat Chem Biol* **2007**, *3* (10), 619–629. https://doi.org/10.1038/nchembio.2007.35.

(7) Lummis, S. C. R.; Beene, D. L.; Lee, L. W.; Lester, H. A.; Broadhurst, R. W.; Dougherty, D. A. *Cis–Trans* Isomerization at a Proline Opens the Pore of a Neurotransmitter-Gated Ion Channel. *Nature* **2005**, *438*, 248–252. https://doi.org/10.1038/nature04130.

(8) Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature* **2017**, *550*, 53–60. https://doi.org/10.1038/nature24031.

(9) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-Specific Incorporation of Non-Canonical Amino Acids into Proteins: Recent Developments and Applications. *Curr Opin Chem Biol* **2010**, *14*, 774–780. https://doi.org/10.1016/j.cbpa.2010.09.013.

(10) Mukai, T.; Lajoie, M. J.; Englert, M.; Soll, D. Rewriting the Genetic Code. *Annu Rev Microbiol* **2017**, *71*, 557–577. https://doi.org/10.1146/annurev-micro-090816.

(11) Link, A. J.; Vink, M. K. S.; Tirrell, D. A. Presentation and Detection of Azide Functionality in Bacterial Cell Surface Proteins. *J Am Chem Soc* **2004**, *126*, 10598–10602. https://doi.org/10.1021/ja047629c.

(12) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective Identification of Newly Synthesized Proteins in Mammalian Cells Using Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT). *Proc Natl Acad Sci* **2006**, *103* (25), 9482–9487. https://doi.org/10.1073/pnas.0601637103.

(13) Luo, X.; Ju, G.; Wang, R. E.; Zhu, X.; Zambaldo, C.; Liu, R.; Liu, T.; Lyu, X.; Du, J.; Xuan, W.; Yao, A.; Reed, S. A.; Kang, M.; Zhang, Y.; Guo, H.; Huang, C.; Yang, P.-Y.; Wilson, I. A.; Schultz, P. G.; Wang, F. Genetically Encoding Phosphotyrosine and Its Nonhydrolyzable Analog in Bacteria. *Nat Chem Biol* **2017**, *13*, 845–849. https://doi.org/10.1038/nchembio.2405.

(14) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a Photocrosslinking Amino Acid to the Genetic Code of *Escherichia coli*. *Proc Natl Acad Sci* **2002**, *99* (17), 11020–11024. https://doi.org/10.1073/PNAS.172226299.

(15) Wang, J.; Xie, J.; Schultz, P. G. A Genetically Encoded Fluorescent Amino Acid. *J Am Chem Soc* **2006**, *128* (27), 8738–8739. https://doi.org/10.1021/ja062666k.

(16) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Nicotine Binding to Brain Receptors Requires a Strong Cation–Pi Interaction. *Nature* **2009**, *458*, 534–537. https://doi.org/10.1038/nature07768.

(17) Lieblich, S. A.; Fang, K. Y.; Cahn, J. K. B.; Rawson, J.; LeBon, J.; Teresa Ku, H.; Tirrell, D. A. 4*S*-Hydroxylation of Insulin at ProB28 Accelerates Hexamer Dissociation and Delays Fibrillation. *J Am Chem Soc* **2017**, *139*, 8384–8387. https://doi.org/10.1021/jacs.7b00794.

(18) Fang, K. Y.; Lieblich, S. A.; Tirrell, D. A. Replacement of ProB28 by Pipecolic Acid Protects Insulin against Fibrillation and Slows Hexamer Dissociation. *J Polym Sci A Polym Chem* **2019**, *57* (3), 264–267. https://doi.org/10.1002/pola.29225.

(19) Deming, T. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. Biosynthetic Incorporation and Chemical Modification of Alkene Functionality in Genetically Engineered Polymers. *J Macromol Sci A* **1997**, *34* (10), 2143–2150. https://doi.org/10.1080/10601329708010331.

(20) Verhoork, S. J. M.; Killoran, P. M.; Coxon, C. R. Fluorinated Prolines as Conformational Tools and Reporters for Peptide and Protein Chemistry. *Biochemistry* **2018**, *57* (43), 6132–6143. https://doi.org/10.1021/acs.biochem.8b00787.

(21) Torbeev, V. Y.; Hilvert, D. Both the *Cis-Trans* Equilibrium and Isomerization Dynamics of a Single Proline Amide Modulate β2-Microglobulin Amyloid Assembly. *Proc Natl Acad Sci* **2013**, *110* (50), 20051–20056. https://doi.org/10.1073/pnas.1310414110.

(22) Ganguly, H. K.; Basu, G. Conformational Landscape of Substituted Prolines. *Biophys Rev* **2020**, *12*, 25–39. https://doi.org/10.1007/s12551-020-00621-8.

(23) Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. Fluoroprolines as Tools for Protein Design and Engineering. *Angew Chem Int Ed* **2001**, *40* (5), 923–925.

(24) Lieblich, S. Non-Canonical Amino Acid Mutagenesis of Position B28 in Insulin with Proline Analogs, 2017. https://doi.org/10.7907/Z90Z71BD.

(25) Kim, W.; Hardcastle, K. I.; Conticello, V. P. Fluoroproline Flip-Flop: Regiochemical Reversal of a Stereoelectronic Effect on Peptide and Protein Structures. *Angew Chem Int Ed* **2006**, *45* (48), 8141–8145. https://doi.org/10.1002/anie.200603227.

(26) Caporale, A.; Loughlin, J. O.; Ortin, Y.; Rubini, M. A Convenient Synthetic Route to (2*S*,4*S*)-Methylproline and Its Exploration for Protein Engineering of Thioredoxin. *Org Biomol Chem* **2022**. https://doi.org/10.1039/D2OB01011A.

(27) Shoulders, M. D.; Hodges, J. A.; Raines, R. T. Reciprocity of Steric and Stereoelectronic Effects in the Collagen Triple Helix. *J Am Chem Soc* **2006**, *128*, 8112–8113. https://doi.org/10.1021/ja061793d.

(28) Cadamuro, S. A.; Reichold, R.; Kusebauch, U.; Musiol, H. J.; Renner, C.; Tavan, P.; Moroder, L. Conformational Properties of 4-Mercaptoproline and Related Derivatives. *Angew Chem Int Ed* **2008**, *47* (11), 2143–2146. https://doi.org/10.1002/anie.200704310.

(29) Siebler, C.; Erdmann, R. S.; Wennemers, H. Switchable Proline Derivatives: Tuning the Conformational Stability of the Collagen Triple Helix by pH Changes. *Angew Chem Int Ed* **2014**, *53* (39), 10340–10344. https://doi.org/10.1002/anie.201404935.

(30) Torbeev, V. Y.; Fumi, E.; Ebert, M. O.; Schweizer, W. B.; Hilvert, D. Cis-Trans Peptide-Bond Isomerization in α-Methylproline Derivatives. *Helv Chim Acta* **2012**, *95* (12), 2411–2420. https://doi.org/10.1002/hlca.201200483.

(31) De Poli, M.; Moretto, A.; Crisma, M.; Peggion, C.; Formaggio, F.; Kaptein, B.; Broxterman, Q. B.; Toniolo, C. Is the Backbone Conformation of Cα-Methyl Proline Restricted to a Single Region? *Chem Eur J* **2009**, *15* (32), 8015–8025. https://doi.org/10.1002/CHEM.200900688.

(32) Kim, W.; McMillan, R. A.; Snyder, J. P.; Conticello, V. P. A Stereoelectronic Effect on Turn Formation Due to Proline Substitution in Elastin-Mimetic Polypeptides. *J Am Chem Soc* **2005**, *127* (51), 18121–18132. https://doi.org/10.1021/ja054105j.

(33) Naduthambi, D.; Zondlo, N. J. Stereoelectronic Tuning of the Structure and Stability of the Trp Cage Miniprotein. *J Am Chem Soc* **2006**, *128*, 12430–12431. https://doi.org/10.1021/ja0648458.

(34) Arnold, U.; Raines, R. T. Replacing a Single Atom Accelerates the Folding of a Protein and Increases Its Thermostability. *Org Biomol Chem* **2016**, *14* (28), 6780–6785. https://doi.org/10.1039/c6ob00980h.

(35) Crespo, M. D.; Rubini, M. Rational Design of Protein Stability: Effect of (2S,4R)-4-Fluoroproline on the Stability and Folding Pathway of Ubiquitin. *PLoS One* **2011**, *6* (5), e19425. https://doi.org/10.1371/journal.pone.0019425.

(36) Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. Fluoroprolines as Tools for Protein Design and Engineering. *Angew Chem Int Ed* **2001**, *40* (5), 923–925. https://doi.org/10.1002/1521-3773(20010302)40:5<923::AID-ANIE923>3.0.CO;2-#.

(37) Zheng, T. Y.; Lin, Y. J.; Horng, J. C. Thermodynamic Consequences of Incorporating 4-Substituted Proline Derivatives into a Small Helical Protein. *Biochemistry* **2010**, *49* (19), 4255–4263. https://doi.org/10.1021/bi100323v.

(38) Dietz, D.; Kubyshkin, V.; Budisa, N. Applying γ-Substituted Prolines in the Foldon Peptide: Polarity Contradicts Preorganization. *ChemBioChem* **2015**, *16* (3), 403–406. https://doi.org/10.1002/cbic.201402654.

(39) Rubini, M.; Schärer, M. A.; Capitani, G.; Glockshuber, R. (4*R*)- and (4*S*)-Fluoroproline in the Conserved *Cis*-Prolyl Peptide Bond of the Thioredoxin Fold: Tertiary Structure

Context Dictates Ring Puckering. *ChemBioChem* **2013**, *14* (9), 1053–1057. https://doi.org/10.1002/cbic.201300178.

(40) Roderer, D.; Lockshuber, R.; Rubini, M. Acceleration of the Rate-Limiting Step of Thioredoxin Folding by Replacement of Its Conserved *Cis*-Proline with (4*S*)-Fluoroproline. *ChemBioChem* **2015**, *16*, 2162–2166. https://doi.org/10.1002/cbic.201500342.

(41) O' Loughlin, J.; Napolitano, S.; Rubini, M. Protein Design with Fluoroprolines: 4,4-Difluoroproline Does Not Eliminate the Rate-Limiting Step of Thioredoxin Folding. *ChemBioChem* **2021**, *22* (23), 3326–3332. https://doi.org/10.1002/CBIC.202100418.

(42) Torbeev, V.; Ebert, M. O.; Dolenc, J.; Hilvert, D. Substitution of Proline32 by α-Methylproline Preorganizes β2-Microglobulin for Oligomerization but Not for Aggregation into Amyloids. *J Am Chem Soc* **2015**, *137* (7), 2524–2535. https://doi.org/10.1021/ja510109p.

(43) Steiner, T.; Hess, P.; Bae, J. H.; Wiltschi, B.; Moroder, L.; Budisa, N. Synthetic Biology of Proteins: Tuning GFPs Folding and Stability with Fluoroproline. *PLoS One* **2008**, *3* (2), e1680. https://doi.org/10.1371/journal.pone.0001680.

(44) Deepankumar, K.; Nadarajan, S. P.; Ayyadurai, N.; Yun, H. Enhancing the Biophysical Properties of MRFP1 through Incorporation of Fluoroproline. *Biochem Biophys Res Commun* **2013**, *440* (4), 509–514. https://doi.org/10.1016/j.bbrc.2013.09.062.

(45) Edwardraja, S.; Sriram, S.; Govindan, R.; Budisa, N.; Lee, S.-G. Enhancing the Thermal Stability of a Single-Chain Fv Fragment by in Vivo Global Fluorination of the Proline Residues. *Mol BioSyst* **2011**, *7*, 258–265. https://doi.org/10.1039/c0mb00154f.

(46) Merkel, L.; Schauer, M.; Antranikian, G.; Budisa, N. Parallel Incorporation of Different Fluorinated Amino Acids: On the Way to "Teflon" Proteins. *ChemBioChem* **2010**, *11* (11), 1505–1507. https://doi.org/10.1002/cbic.201000295.

(47) Holzberger, B.; Marx, A. Replacing 32 Proline Residues by a Noncanonical Amino Acid Results in a Highly Active DNA Polymerase. *J Am Chem Soc* **2010**, *132* (44), 15708–15713. https://doi.org/10.1021/ja106525y.

(48) Holzberger, B.; Obeid, S.; Welte, W.; Diederichs, K.; Marx, A. Structural Insights into the Potential of 4-Fluoroproline to Modulate Biophysical Properties of Proteins. *Chem Sci* **2012**, *3* (10), 2924–2931. https://doi.org/10.1039/c2sc20545a.

(49) Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. Genomically Recoded Organisms Expand Biological Functions. *Science (1979)* **2013**, *342* (October), 357–360. https://doi.org/10.1126/science.1241459.

(50) Dunkelmann, D. L.; Willis, J. C. W.; Beattie, A. T.; Chin, J. W. Engineered Triply Orthogonal Pyrrolysyl–tRNA Synthetase/tRNA Pairs Enable the Genetic Encoding of Three Distinct Non-Canonical Amino Acids. *Nat Chem 2020 12:6* **2020**, *12* (6), 535–544. https://doi.org/10.1038/s41557-020-0472-x.

(51) Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. Site-Specific Incorporation of an Unnatural Amino Acid into Proteins in Mammalian Cells. *Nucleic Acids Res* **2002**, *30* (21), 4692–4699. https://doi.org/10.1093/NAR/GKF589.

(52) Furter, R. Expansion of the Genetic Code: Site-Directed p-Fluoro-Phenylalanine Incorporation in *Escherichia coli*. *Protein Sci* **1998**, 7419–7426. https://doi.org/10.1002/pro.5560070223.

(53) Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host. *Angew Chem Int Ed* **2000**, *39* (12), 2148–2152. https://doi.org/10.1002/1521-3757(20000616)112:12<2232::AID-ANGE2232>3.0.CO;2-0.

(54) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. Discovery of *Escherichia coli* Methionyl-tRNA Synthetase Mutants for Efficient Labeling of Proteins with Azidonorleucine in Vivo. *Proc Nat Acad Sci* **2009**, *106* (36), 15285–15290. https://doi.org/10.1073/pnas.0905735106.

(55) Teeuwen, R. L. M.; Van Berkel, S. S.; Van Dulmen, T. H. H.; Schoffelen, S.; Meeuwissen, S. A.; Zuilhof, H.; De Wolf, F. A.; Van Hest, J. C. M. "Clickable" Elastins: Elastin-like Polypeptides Functionalized with Azide or Alkyne Groups. *Chem Commun* **2009**, No. 27, 4022–4024. https://doi.org/10.1039/B903903A.

(56) Wolschner, C.; Giese, A.; Kretzschmar, H. A.; Huber, R.; Moroder, L.; Budisa, N. Design of Anti- and pro-Aggregation Variants to Assess the Effects of Methionine Oxidation in Human Prion Protein. *Proc Natl Acad Sci* **2009**, *106* (19), 7756–7761. https://doi.org/10.1073/PNAS.0902688106.

(57) Montclare, J. K.; Son, S.; Clark, G. A.; Kumar, K.; Tirrell, D. A. Biosynthesis and Stability of Coiled-Coil Peptides Containing (2*S*,4*R*)-5,5,5-Trifluoroleucine and (2*S*,4*S*)-5,5,5-Trifluoroleucine. *ChemBioChem* **2009**, *10* (1), 84–86. https://doi.org/10.1002/CBIC.200800164.

(58) Yoo, T. H.; Link, A. J.; Tirrell, D. A. Evolution of a Fluorinated Green Fluorescent Protein. *Proc Natl Acad Sci* **2007**, *104* (35), 13887–13890. https://doi.org/10.1073/pnas.0701904104.

(59) Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L. Residue-Specific Bioincorporation of Non-Natural, Biologically Active Amino Acids into Proteins as Possible Drug Carriers: Structure and Stability of the per-Thiaproline Mutant of Annexin V. *Proc Natl Acad Sci* **1998**, *95* (2), 455–459. https://doi.org/10.1073/pnas.95.2.455.

(60) Kim, W.; George, A.; Evans, M.; Conticello, V. P. Cotranslational Incorporation of a Structurally Diverse Series of Proline Analogues in an *Escherichia coli* Expression System. *ChemBioChem* **2004**, *5* (7), 928–936. https://doi.org/10.1002/cbic.200400052.

(61) Lukesch, M. S.; Pavkov-Keller, T.; Gruber, K.; Zangger, K.; Wiltschi, B. Substituting the Catalytic Proline of 4-Oxalocrotonate Tautomerase with Non-Canonical Analogues Reveals a Finely Tuned Catalytic System. *Sci Rep* **2019**, *9*, 2697. https://doi.org/10.1038/s41598-019-39484-9.

(62) Grothe, S.; Krogsrud, R. L.; Mcclellan, D. J.; Milner, J. L.; Wood, J. M. Proline Transport and Osmotic Stress Response in *Escherichia coli* K-12. *J Bacteriol* **1986**, *166* (1), 253–259.

(63) Fang, K. Y. Modulating Biophysical Properties of Insulin with Non-Canonical Mutagenesis at Position B28, California Institute of Technology, 2017. https://doi.org/10.7907/Z9V40S6J.