Genetic Encoding of Cyanopyridylalanine for In-Cell Protein Macrocyclization by the Nitrile-Aminothiol Click Reaction

Elwy H. Abdelkader, [b]† Haocheng Qianzhu, [a]† Josemon George, [a] Rebecca L. Frkic, [b] Colin J. Jackson, [b] Christoph Nitsche, [a] Gottfried Otting, [b] and Thomas Huber*[a]

[a] Mr Haocheng Qianzhu, Dr Josemon George, Dr Christoph Nitsche, Prof. Thomas Huber
Research School of Chemistry
Australian National University
Canberra, ACT 2601, Australia
E-mail: t.huber@anu.edu.au

[b] Dr Elwy H. Abdelkader, Dr Rebecca L. Frkic, Prof. Colin J. Jackson, Prof. Gottfried Otting
ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry
Australian National University
Canberra, ACT 2601, Australia

† These authors contributed equally to this work.

Supporting information for this article is given via a link at the end of the document.

Abstract: Cyanopyridylalanines are non-canonical amino acids that react with aminothiol compounds under physiological conditions in a biocompatible manner without requiring added catalyst. Here we present newly developed aminoacyl-tRNA synthetases for genetic encoding of *meta*- and *para*-cyanopyridylalanine to enable the site-specific attachment of a wide range of different functionalities. The outstanding utility of the cyanopyridine moiety is demonstrated by examples of (i) post-translational functionalization of proteins, (ii) in-cell macrocyclization of peptides and proteins, and (iii) protein stapling. The biocompatible nature of the protein ligation chemistry enabled by the cyanopyridylalanine amino acid opens a new path to specific *in vivo* protein modifications in complex biological environments.

Bioorthogonal reactions for site-specific protein conjugation with chemical and biochemical tags have a wide range of applications in the material, biological, and health sciences.[1] The site-selective modification of a target protein can be used to confer specific biophysical properties[2] and install labels for spectroscopic imaging and tracking of proteins by fluorescence,[3] nuclear magnetic resonance,[4] or electron paramagnetic resonance spectroscopy techniques.[5]

Site-specific incorporation of noncanonical amino acids (ncAAs) by genetic encoding gives precise control of the sites, where new functional groups are installed in a target protein,[6] but there is only a very small number of established bioorthogonal reactions. To date, the most prominent examples are copper-catalysed azide−alkyne cycloadditions (CuAAC) and inverse-electron demand Diels−Alder reactions (IEDDA).[7] Copper catalysts are barely compatible with physiological conditions, and strain promoted cycloaddition and IEDDA reactions add relatively large non-biological chemical moieties, the synthesis of which can be challenging. In the light of recent progress in genetic code reprogramming to incorporate multiple, distinct ncAAs with different bioorthogonal functionalities into a single protein,[8] there is an unmet demand for additional genetically encoded ncAAs that enable stable conjugations in a reaction that is not only biocompatible but fundamentally different from established bioorthogonal reactions.

The nitrile-aminothiol (NAT) click reaction is a condensation reaction between electrophilic nitriles and 1,2-aminothiols, which proceeds rapidly under biological conditions without the need of any added catalyst and has been shown to be compatible with all canonical amino acids within peptides, except cysteine at the N-terminus of a polypeptide chain.[9] The present work demonstrates its suitability for *in vivo* protein modification.

To enable site-specific NAT click reactions on proteins, we first identified pyrrolysyl-tRNA synthetase (PylRS) mutants specific for *meta*- and *para*-cyanopyridylalanines (mCNP and pCNP, Figure 1). The synthetases were selected from a library of PylRS mutants derived from the methanogenic archaeon ISO4-G1 (G1PylRS), using our previously reported screening approach based on fluorescence-activated cell sorting (FACS),[10] and enabled site-specific incorporation of these amino acids in response to an amber stop codon.

In the absence of a crystal structure for G1PylRS, mutation sites were chosen based on the amino acid sequence alignment between G1PylRS and *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (*Mm*PylRS) (Figure S1). By examining the crystal structure of *Mm*PylRS (PBD ID: 2Q7E),[11] seven residues in *Mm*PylRS (L305, Y306, N346, C348, Y384, V401, W417) were hypothesized to influence substrate recognition (Figure S2) and the corresponding residues were randomized in the G1PylRS library (Supplementary Methods).

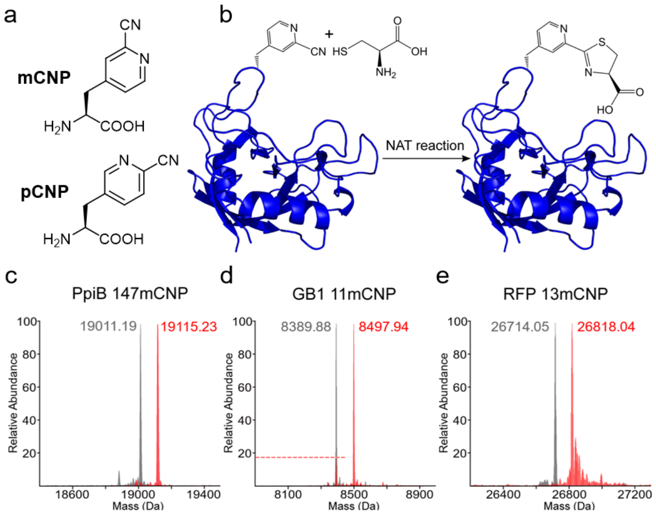


Figure 1. Structures and applications of the cyanopyridylalanines mCNP and pCNP. a, Chemical structures of mCNP and pCNP. b, Example of a NAT reaction between genetically incorporated mCNP and cysteine, which proceeds in aqueous solution at neutral pH and ambient temperature. c-e, Intact protein mass spectrometry analysis of the reaction between genetically incorporated mCNP and cysteine. The grey and red spectra are of samples before and after the NAT reaction, respectively, and the red dashed line indicates the level of unreacted protein educt. Expected masses are reported in Table S4.

The plasmid library encoding G1PylRS mutants and the orthogonal ^{G1Pyl}tRNA$_{CUA}$ (pBK-G1RS) was co-transformed into *E. coli* DH10B cells with the selection plasmid pBAD-H6RFP, which contains the mCherry red fluorescent protein (RFP) gene preceded by a His$_6$ tag with an amber stop codon at position 8 (His6-TAG-RFP). The transformed cells were subjected to multiple rounds of selection under positive (in the presence of the target ncAA) or negative (without the target ncAA) growth conditions. After each selection round, the FACS results showed clear enrichment of the target population. Cells containing synthetase candidates collected from the final round of selection were characterized individually (Figures S3 and S4, and Tables S1 and S2).

The best G1PylRS mutant for mCNP incorporation (G1mCNP34, in the following referred to as G1mCNPRS) featured the mutations L124A, Y125L, V167A, Y204W, and A221S. Despite randomization in the library, N165 and W237 were the same as in the wild-type sequence. The G1PylRS mutant for site-specific incorporation of pCNP (G1pCNP37, in the following referred to as G1pCNPRS) carries the mutations L124A, Y125F, Y204W, A221S, and W237Y, with N165 and V167 remaining unchanged. To gain an atomic level understanding of the mutations, we crystallized G1mCNPRS and solved its structure at a resolution of 2.2 Å (PDB ID: 7R6O; Figure S5; Table S8). As expected, the structure of the amino acid binding domain of the G1mCNP tRNA synthetase is highly conserved, but shows differences in in the loop connecting β-strands 5 and 6, which contains Y204W as one of the key residues of the substrate binding site.

To produce proteins with cyanopyridylalanine residues in high yield, we used our previously developed two-plasmid system for *in vivo* incorporation of ncAAs via amber stop codon suppression.[10b] The gene of G1mCNPRS or G1pCNPRS, together with the orthogonal ^{G1Pyl}tRNA$_{CUA}$, was cloned into a high-copy number pRSF plasmid to obtain the plasmids pRSF-G1mCNPRS and pRSF-G1pCNPRS, respectively, while the gene for the protein of interest was cloned into a low-copy number pCDF plasmid. Using the pCDF/pRSF system, mCNP and pCNP were incorporated with high fidelity and yield, for both single- and double-amber mutants, without any evidence of adduct formation with intracellular metabolites (Figure S6, Table S3). In addition, neither mCNP nor pCNP had any negative effect on *E. coli* cell growth (Figure S7). Next, we tested the reactivity of the incorporated cyanopyridylalanines in the NAT click reaction using L-cysteine as a model 1,2-aminothiol (Figure 1b). Different proteins containing mCNP or pCNP were incubated with 5 mM cysteine at 25 °C. Monitoring the reaction by intact protein mass spectrometry showed formation of the thiazoline product in greater than 90% yield after 4 h (Figure 1c–e), in agreement with the reaction kinetics reported previously for peptides.[9b] The reaction was not impeded by the presence of 10 mM (tris(2-carboxyethyl)phosphine (TCEP) added to prevent oxidation of thiol groups during the reaction.

We previously reported that mCNP installed in peptides by solid-phase peptide synthesis readily undergoes spontaneous cyclization with an N-terminal cysteine residue in aqueous buffers at pH 7.5.[9a] To test the viability of the cyclization reaction in a protein, we installed mCNP in the fusion protein NT-Ubi 7X (X indicating the position of the ncAA in the amino acid sequence), which comprised an N-terminal NT solubility tag[13] followed by a short linker containing the modified TEV protease recognition sequence ENLYFQC and human ubiquitin[14] at the C-terminal end (Figure 2a).[15] The fusion protein was readily expressed in *E. coli* and purified using Ni/nitrilotriacetic acid (Ni-NTA) resin. 256 and 162 mg of purified protein were obtained per 1 L cell culture with mCNP or pCNP, respectively, with high incorporation yield as indicated by intact protein mass spectrometry (Figure S8). The cyclization reaction was triggered by digestion with TEV protease to expose the cysteine residue of the TEV recognition sequence at the N-terminus. The cleaved protein product Cys-Ubi 7X cyclized within 15 minutes, as indicated by mass spectrometry. The TEV protease cleavage was complete after 4 hours, resulting in more than 75% cyclized Cys-Ubi 7X (Figure 2c and d).

Encouraged by these results, we examined the utility of the intramolecular NAT click reaction for *in vivo* peptide cyclization. To generate the required N-terminal cysteine residue, the protein needs to be cleaved inside the bacterial cell. This was achieved by a 3-plasmid system comprising a pCDF plasmid for the expression of NT-Ubi 7X, the pRSF plasmid containing the orthogonal PylRS system, and a pBAD-TEV plasmid for co-expression of TEV protease. The results showed that Cys-Ubi 7X was produced in high yield (134 mg and 60 mg of purified protein per 1 L cell culture with mCNP or pCNP, respectively) and either sample was found to be cyclized quantitatively, regardless of the difference in structural constraints imposed on the 7-residue macrocycle by the two different cyanopyridylalanine residues (Figure 2e and f).

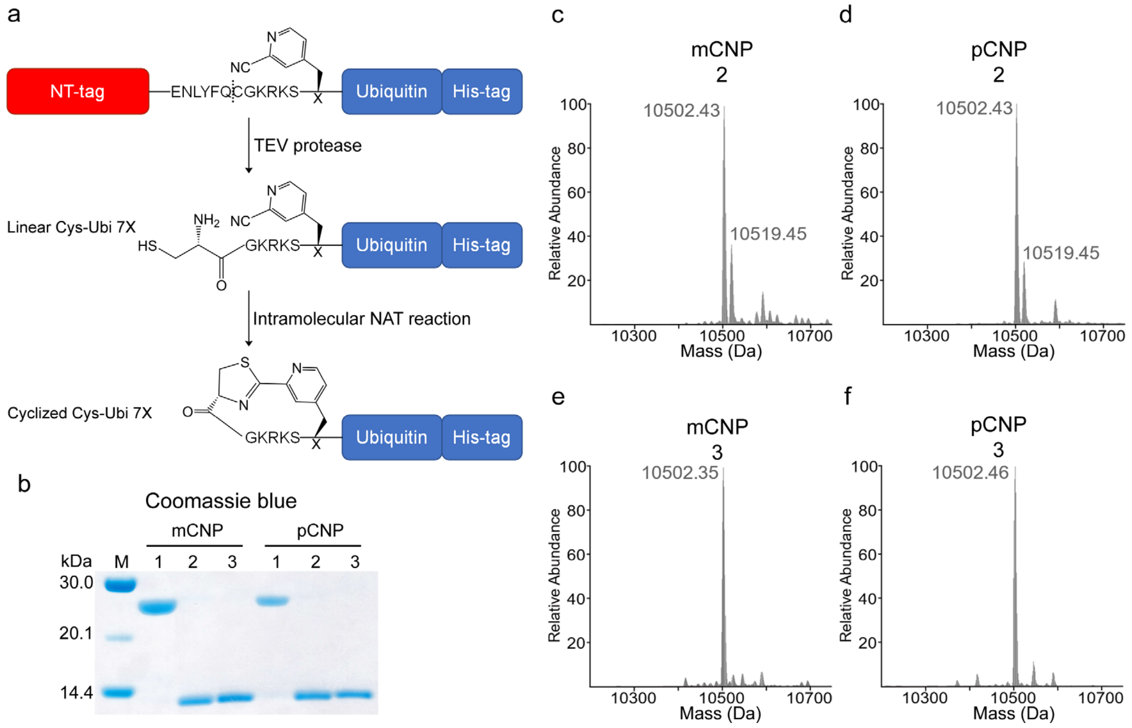


Figure 2. *In vivo* protein macrocyclization via intramolecular NAT click reaction. a, Design of the NT-Ubi 7X fusion protein used in the current study. X indicates the position of the ncAA. b, SDS-PAGE analysis: M, protein molecular weight marker (the size of each band is indicated on the left); lane 1, NT-Ubi 7X before TEV protease cleavage; lane 2, NT-Ubi 7X after TEV protease cleavage for 4 h at 25 °C; lane 3, purified Cys-Ubi X7 expressed using the 3-plasmid system for *in vivo* protein macrocyclization. c–f, Intact protein mass spectra of the Cys-Ubi X7 samples shown in lanes 2 and 3 of b. Expected masses of linear and cyclized Cys-Ubi X7 are 10519.94 and 10502.91 Da, respectively.

An alternative way of polypeptide cyclization is achieved by a tandem NAT click reaction between a polypeptide containing two cyanopyridylalanine residues and a bi-functional aminothiol reagent. We used ethylenediamine dicysteine (EDDC), where the carboxyl groups of two cysteine residues are linked by an ethylenediamine moiety (Figure 3a, Supporting Information). The approach takes advantage of the difference in reaction rate between the slower intermolecular and faster intramolecular NAT reaction, where EDDC first reacts with a single cyanopyridylalanine moiety to form a singly tagged protein and this mono-functionalized intermediate undergoes a fast, spontaneous intramolecular NAT reaction with the second cyanopyridylalanine moiety. As a result, formation of the cyclized product is strongly favoured over doubly tagged uncyclized product even in the presence of a large excess of the di-aminothiol reagent. The scheme was successful with all three proteins tested, which were produced with two mCNP residues (GB1 A24X/K28X, RFP 237X/243X and RFP A204X/237X, where X = mCNP). Following incubation with 5 mM EDDC (10 mM for RFP 237X/243X) at 25 °C, the reactions were complete after 4 h in near-quantitative yields as indicated by intact protein mass spectrometry (Figure 3b–d).

The success of our system for genetic encoding of cyanopyridylalanines in high yield is based on (i) our PylRS library derived from the methanogenic archaeon ISO4-G1, which proved exceptionally adaptable for encoding aromatic ncAAs, and (ii) a two-plasmid selection system established previously for a chimeric PylRS variant tailored to the genetic encoding of lysine-based ncAAs.[10b] The versatility of G1PylRS systems has been demonstrated previously in bacteria,[16] mammalian cells,[17] and cell-free protein synthesis.[10b]

The ncAAs mCNP and pCNP present a balanced compromise between biocompatibility and reactivity. They are non-toxic *in vivo* and the activated nitrile functionality does not react with functional groups found in cellular biopolymers. In contrast, the activated nitrile functionality ligates readily and in high yield with 1,2-aminothiol compounds, provided they are present in high local concentration, as achieved in, e.g., intramolecular reactions. In *E. coli*, the cyanopyridylalanine ncAAs appear to be resistant against reaction with metabolic compounds, including cysteine present at natural intracellular concentrations. This is an advantage over ncAAs with a azide group, which are susceptible to chemical reduction in bacterial cells,[4b, 18] or *trans*-cyclooctene (TCO) groups, which have been shown to be prone to isomerization to the non-reactive *cis*-cyclooctene isomer *in vivo*.[19] Previous attempts to genetically encode the 1,2-aminothiol group as a reactive conjugation group proved unsuccessful due to reaction of the aminothiol moiety with pyruvate, which is abundant in cells.[20] A genetic encoding system subsequently developed for a chemically caged 1,2-aminothiol functionality depends on the provision of chemicals for decaging, compromising *in vivo* applications.[21]

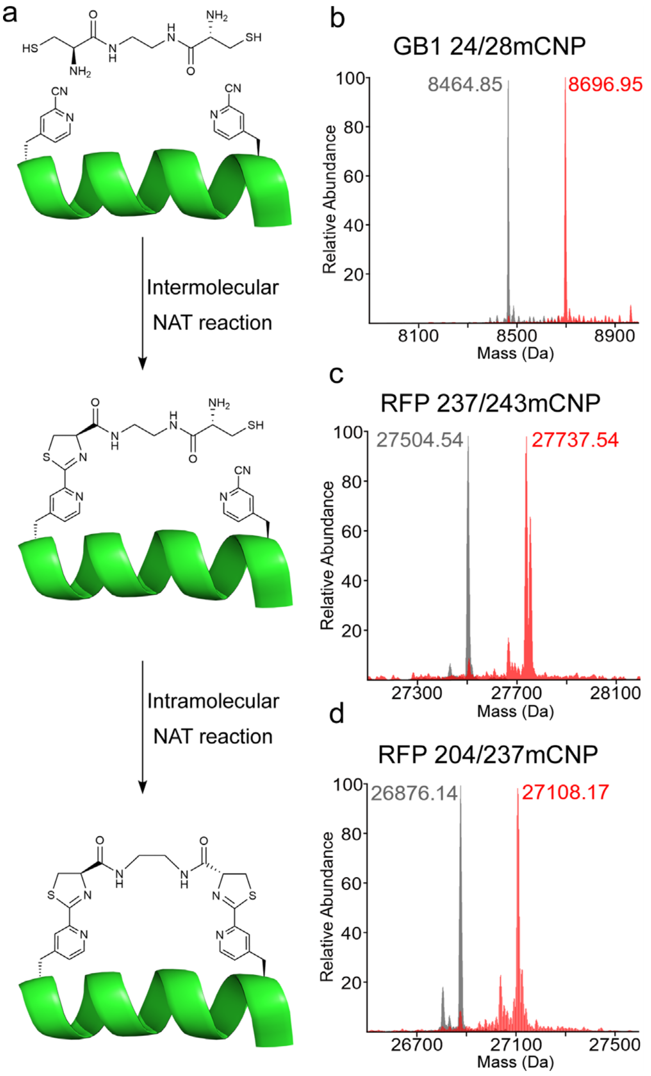


Figure 3. Stapling by tandem intermolecular-intramolecular NAT reaction. a, Intermolecular NAT reaction of EDDC results in the formation of singly tagged protein, which spontaneously undergoes an intramolecular NAT reaction to form the cyclized product. b–d, Intact protein mass spectrum analysis of the reaction between GB1 A24X/K28X, RFP 237X/243X, or RFP A204X/237X after incubation with EDDC for 4 h at 25 °C. Expected masses of the unreacted proteins and the cyclized products are reported in Table S4.

The present work expands the number of genetically encoded noncanonical amino acids by two different versions of cyanopyridylalanine, which are privileged for NAT click reactions owing to a reactive nitrile group, which spontaneously reacts with 1,2-aminothiol compounds. Incorporation of mCNP or pCNP residues in a protein offers a multitude of possibilities for its site-specific conjugation with different tags under physiological conditions, which may contain fluorescent or other reporter groups. As the cyanopyridylalanine residues react with N-terminal cysteine, they afford a new approach for peptide and protein cyclisation. Two cyanopyridylalanine residues in a polypeptide chain provide a means for polypeptide stapling.

The NAT click reaction between cyanopyridylalanine residues and 1,2-aminothiols proceeds spontaneously after simple mixing. Unlike the copper catalyst in CuAAC reactions, which tends to generate reactive oxygen species that degrade proteins and are toxic to cells,[22] the absence of added catalyst in the NAT click reaction is a considerable advantage in modifying proteins within living cells. Beyond protein tagging, cyanopyridylalanines enable efficient protein macrocyclization *in vitro* and in cells, which we envisage to afford a convenient protein engineering tool to enhance the thermal and proteolytic stability of proteins, as well as for the *in vivo* generation of libraries of macrocyclic peptides and proteins.[23]

In conclusion, the biocompatible NAT click reaction between genetically encoded cyanopyridylalanine ncAAs and aminothiols offers an attractive tool for intra- and extra-cellular bioconjugation. The advantages offered by the NAT click reaction over previously reported bioorthogonal reactions makes it a highly valuable alternative to available protein ligation tools.

Acknowledgements

We thank Dr. Harpreet Vohra and Michael Devoy at the John Curtin School of Medical Research, Australian National University for technical support on FACS experiments. This re-search was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. Financial support by the Australian Research Council for a Laureate Fellowship to G.O. (FL170100019), a DECRA Fellowship to C.N. (DE190100015), projects (DP200100348, DP21010088), and through a Centre of Excellence (CE200100012) is gratefully acknowledged.

Keywords: bioorthogonal reaction • cyanopyridylalanine • genetic encoding • noncanonical amino acids • nitrile-aminothiol (NAT) click reaction • protein conjugation

[1] a) R. Rossin, P. Renart Verkerk, S. M. van den Bosch, R. C. M. Vulders, I. Verel, J. Lub, M. S. Robillard, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 3375-3378; b) C. Guo, H. Kim, E. M. Ovadia, C. M. Mourafetis, M. Yang, W. Chen, A. M. Kloxin, *Acta Biomater.* **2017**, *56*, 80-90.

[2] N. K. Devaraj, *ACS Cent. Sci.* **2018**, *4*, 952-959.

[3] a) K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, J. W. Chin, *Nat. Chem.* **2012**, *4*, 298-304; b) H. S. Jang, S. Jana, R. J. Blizzard, J. C. Meeuwsen, R. A. Mehl, *J. Am. Chem. Soc.* **2020**, *142*, 7245-7249.

[4] a) C. T. Loh, K. Ozawa, K. L. Tuck, N. Barlow, T. Huber, G. Otting, B. Graham, *Bioconjug. Chem.* **2013**, *24*, 260-268; b) C.-T. Loh, B. Graham, E. H. Abdelkader, K. L. Tuck, G. Otting, *Chem. Eur. J.* **2015**, *21*, 5084-5092.

[5] a) M. R. Fleissner, E. M. Brustad, T. Kálai, C. Altenbach, D. Cascio, F. B. Peters, K. Hideg, S. Peuker, P. G. Schultz, W. L. Hubbell, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 21637-21642; b) E. H. Abdelkader, A. Feintuch, X. Yao, L. A.

Adams, L. Aurelio, B. Graham, D. Goldfarb, G. Otting, *Chem. Commun.* **2015**, *51*, 15898-15901.

[6] a) R. Brabham, M. A. Fascione, *Chembiochem* **2017**, *18*, 1973-1983; b) J. W. Chin, *Nature* **2017**, *550*, 53-60.

[7] a) E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 6974-6998; b) C. D. Spicer, E. T. Pashuck, M. M. Stevens, *Chem. Rev.* **2018**, *118*, 7702-7743; c) M. L. W. J. Smeenk, J. Agramunt, K. M. Bonger, *Curr. Opin. Chem. Biol.* **2021**, *60*, 79-88.

[8] a) D. L. Dunkelmann, J. C. W. Willis, A. T. Beattie, J. W. Chin, *Nat. Chem.* **2020**, *12*, 535-544; b) W. E. Robertson, L. F. H. Funke, D. de la Torre, J. Fredens, T. S. Elliott, M. Spinck, Y. Christova, D. Cervettini, F. L. Böge, K. C. Liu, S. Buse, S. Maslen, G. P. C. Salmond, J. W. Chin, *Science* **2021**, *372*, 1057-1062.

[9] a) C. Nitsche, H. Onagi, J.-P. Quek, G. Otting, D. Luo, T. Huber, *Org. Lett.* **2019**, *21*, 4709-4712; b) R. Morewood, C. Nitsche, *Chem. Sci.* **2021**, *12*, 669-674; c) N. A. Patil, J.-P. Quek, B. Schroeder, R. Morewood, J. Rademann, D. Luo, C. Nitsche, *ACS Med. Chem. Lett.* **2021**, *12*, 732-737.

[10] a) H. Qianzhu, A. P. Welegedara, H. Williamson, A. E. McGrath, M. C. Mahawaththa, N. E. Dixon, G. Otting, T. Huber, *J. Am. Chem. Soc.* **2020**, *142*, 17277-17281; b) E. H. Abdelkader, H. Qianzhu, Y. J. Tan, L. A. Adams, T. Huber, G. Otting, *J. Am. Chem. Soc.* **2021**, *143*, 1133-1143.

[11] J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Söll, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11268-11273.

[12] Y. Sun, H. Jiang, W. Wu, W. Zeng, X. Wu, *Org. Lett.* **2013**, *15*, 1598-1601.

[13] N. Kronqvist, M. Sarr, A. Lindqvist, K. Nordling, M. Otikovs, L. Venturi, B. Pioselli, P. Purhonen, M. Landreh, H. Biverstål, Z. Toleikis, L. Sjöberg, C. V. Robinson, N. Pelizzi, H. Jörnvall, H. Hebert, K. Jaudzems, T. Curstedt, A. Rising, J. Johansson, *Nat. Commun.* **2017**, *8*, 15504.

[14] M. Békés, K. Okamoto, Sarah B. Crist, Mathew J. Jones, Jessica R. Chapman, Bradley B. Brasher, Francesco D. Melandri, Beatrix M. Ueberheide, E. Lazzerini Denchi, Tony T. Huang, *Cell Rep.* **2013**, *5*, 826-838.

[15] R. B. Kapust, J. Tözsér, T. D. Copeland, D. S. Waugh, *Biochem. Biophys. Res. Commun.* **2002**, *294*, 949-955.

[16] J. C. W. Willis, J. W. Chin, *Nat. Chem.* **2018**, *10*, 831-837.

[17] B. Meineke, J. Heimgärtner, J. Eirich, M. Landreh, S. J. Elsässer, *Cell Rep.* **2020**, *31*, 107811.

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

[19] R. Rossin, S. M. van den Bosch, W. ten Hoeve, M. Carvelli, R. M. Versteegen, J. Lub, M. S. Robillard, *Bioconjugate Chem.* **2013**, *24*, 1210-1217.

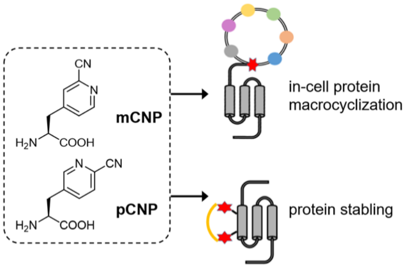
[20] I. E. Gentle, D. P. De Souza, M. Baca, *Bioconjugate Chem.* **2004**, *15*, 658-663.

[21] D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir, J. W. Chin, *J. Am. Chem. Soc.* **2011**, *133*, 11418-11421.

[22] a) D. C. Kennedy, C. S. McKay, M. C. B. Legault, D. C. Danielson, J. A. Blake, A. F. Pegoraro, A. Stolow, Z. Mester, J. P. Pezacki, *J. Am. Chem. Soc.* **2011**, *133*, 17993-18001; b) Craig S. McKay, M. G. Finn, *Chem. Biol.* **2014**, *21*, 1075-1101.

[23] A. Purkayastha, T. J. Kang, *Biotechnol. Bioprocess. Eng.* **2019**, *24*, 702-712.

Entry for the Table of Contents



Genetic encoded cyanopyridylalanines are presented as convenient and highly selective reactive groups for in-vivo functionalization, cyclization and stapling of polypeptides. The non-toxic compounds can be genetically incorporated into proteins with high yields and their biocompatible reaction with amino-thiol moieties proceeds under physiological condition to near completion without requiring a catalyst.

Supporting Information

Genetic Encoding of Cyanopyridylalanine for In-Cell Protein Macrocyclization by the Nitrile-Aminothiol Click Reaction

Elwy H. Abdelkader,[1†] Haocheng Qianzhu,[2†] Josemon George,[2] Rebecca L. Frkic,[1] Colin J. Jackson,[1] Christoph Nitsche,[2] Gottfried Otting,[1] and Thomas Huber[2]*

1. ARC Centre of Excellence for Innovations in Peptide & Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia
2. Australian National University, Research School of Chemistry, Canberra, ACT 2601, Australia

[†] These authors contributed equally to this work.

* e-mail: t.huber@anu.edu.au

Table of Contents:

Supplementary Methods

Screening of functional G1PylRS mutants that recognize cyanopyridylalanine ncAAs by FACS

Construction of the library plasmid and selection plasmid

A pBK-G1RS plasmid was first constructed by introducing a wild-type G1PylRS gene (G1RSwt) under a mutant *GlnS* promoter (*GlnS'*) into a pBK vector.[1-3] An expression cassette for $^{\mathrm{G1Pyl}}$tRNA$_{\mathrm{CUA}}$ under the control of the *lpp* promoter was cloned into the vector in the opposite direction of the synthase gene. From the structure of the G1PylRS's homologues *Mm*PylRS, seven residues were identified to directly interact with the bound amino acid substrate (**Figure S1 and Figure S2**). The equivalent set of residues in G1PylRS are L124, Y125, N165, V167, Y204, A221, and W237. Randomized triplet NNK codons (N = A + G + T + C, K = G + T) were introduced for full randomization at four mutation sites: L124, Y125, A221, and W237, using doped oligonucleotide primers containing these codons. To avoid most of the bulky or charged amino acids at position N165 and V167, which are located in the β-strand closest to the bound substrate, N165 was selectively mutated to G/A/V/S/N/T/I/D through randomized triplet codons RNC (R = A + G); V167 was similarly mutated to G/A/V/S/C/L/F through codon KBM (B = C + G + T, M = A + C). To further reduce the size of the library, residue Y204, which is equivalent to Y384 in *Mm*PylRS, was selectively mutated to aromatic residues (F/Y/W), which we hypothesized to have favorable interactions with the aromatic ring of the substrate via π-stacking. The selective mutation from Y204 to F/Y/W was conducted by mixing three different gene fragments containing either Y204F, Y204Y, or Y204W as one of the PCR templates when conducting overlap PCR for the whole fragment assembly of the G1PylRS library. This library theoretically encodes 2.688 x 10^7 variants of G1PylRS mutants.

The G1PylRS library fragment was ligated back into a pBK vector in the same way as described in our previous work.[2-3] The actual resulting library contained approximately 10^9 individual transformants. DNA sequencing of randomly picked colonies revealed no significant bias. All primers and gene fragments were commercially synthesized (Integrated DNA Technologies, USA).

A pBAD-H6RFP plasmid encoding a His_6-amber(TAG)-mCherry red fluorescent protein (H6RFP) reporter gene under the control of the *araBAD* promoter was used as the selection plasmid.[3]

Selection of active and specific G1PylRSs for mCNP

To carry out the selection, the library plasmid pBK-G1RS was transformed into *E. coli* DH10B cells harboring the selection plasmid pBAD-H6RFP. Following recovery from transformation, the culture was directly inoculated into a flask with 25 mL LB medium containing 100 mg/L carbenicillin and 50 mg/L kanamycin, supplied with 0.4% L-arabinose and 3 mM mCNP, which served as the sample for the first round of positive selection (**1P+**). Overnight expression at 37 °C led to a well-detectable level of RFP expression. Cells were resuspended in 7.5 mL PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) after harvesting. A 100-fold dilution yielded a concentration suitable for cell sorting by FACS (**Figure S3**) on an Aria II high speed cell sorter (BD Biosciences, USA).

The top 1.1% of cells with high RFP levels were selected from **1P+**, collecting 8.0×10^5 cells in 90 min. The cells collected were subjected to the following round of negative selection without the addition of mCNP and regrew as sample **2N–**, from where cells with low RFP expression levels (69.4% of the total) were collected (2.0×10^6 cells in 12 min). These cells were aliquoted to inoculate media with positive (**3P+**) and negative (**3P–**) conditions. The top 13.5% of RFP-positive cells from **3P+** were collected (1.0×10^6 cells in 20 min) and recovered under negative condition to get **4N-**. Following sorting (2.0×10^6 cells in 10 min), 78.0% of the cells showing the lowest level of RFP fluorescence were selected from **4N-**. They were aliquoted to be recovered as **5P+** and **5P–**, respectively. At this stage, the number of RFP fluorescent cells was more than thirty-fold higher in the **5P+** than in the **5P-** sample (33.5% compared to 0.9%). 1.0×10^5 cells collected in 10 minutes from the top 5.0% RFP fluorescent cells of the **5P+** sample were recovered for storage. An aliquot of 5,000 cells were allowed to recover on LB agar plates containing 100 mg/L carbenicillin and 50 mg/L kanamycin, and individual clones were analyzed using 96-well plates. 120 candidates were inoculated into both positive (+3 mM mCNP) and negative (without mCNP) growth conditions. The fluorescence level was measured after expression overnight, using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland) and normalized by the OD_{600} of the cell culture. 7 candidates with the highest RFP level in the positive condition were chosen for sequencing. 6 individually different sequences were found (**Table S1**). One mutation set was identified twice as G1mCNP16 and G1mCNP34 (L124A, Y125L, V167A, Y204W, A221S). G1mCNP34 was used in the subsequent experiments as the best functional G1mCNPRS.

Selection of active and specific G1PylRSs for pCNP

The preparation method for the **1P+** sample of selection for functional pCNPRS was the same as for mCNPRS, except for providing 3mM pCNP (Chem-Space, Latvia (XV6117)) instead of mCNP during overnight expression.

The top 0.3% of cells with high RFP levels were selected from the **1P+** round, collecting 2.0×10^5 cells in 2 hours (**Figure S4**). The cells collected were subjected to the following round of negative selection without the addition of pCNP (**2N–**), from where cells with low RFP expression levels (85.6% of the total) were collected (2.0×10^6 cells in 8 min). These cells were aliquoted to inoculate media with positive (**3P+**) and negative (**3P–**) conditions. The top 0.9% of RFP-positive cells from the **3P+** round were collected (2.0×10^5 cells in 1 hour) and recovered under negative condition to generate **4N-**. Following sorting (2.0×10^6 cells in 8 min), 85.6% of the cells showing the lowest level of RFP fluorescence were selected. They were aliquoted to be recovered for the samples **5P+** and **5P–**. At this stage, the difference in RFP expression level was clear (40.1% compared to 30.6%). The selection process was ended after this round, as sequencing suggested that the remaining sequence diversity encompassed fewer than ten different mutants. 1.0 x 10^5 cells collected in 15 min from the top 5.0% RFP fluorescent cells of the **5P+** sample were recovered for storage. An aliquot of 5,000 cells were allowed to recover on LB agar plates containing 100 mg/L carbenicillin and 50 mg/L kanamycin, and individual clones were analyzed using 96-well plates. 120 candidates were inoculated into both positive (+3 mM pCNP) and negative (without pCNP) growth conditions. The fluorescence level was measured after expression overnight, using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland) and normalized by the OD_{600} of the cell culture. 17 candidates with the highest RFP level in the positive condition were chosen for sequencing (**Table S2**). 3 individually different mutation sets were found: (1) L124A, Y125F, Y204W, W237Y; (2) L124A, Y125F, Y204W, A221S, W237Y; (3) L124S, Y125F, Y204W, A221S, W237Y. G1pCNP37 was used for following experiment as the best functional G1pCNPRS.

In vivo protein expression and purification

For site-specific incorporation of mCNP and pCNP, *E. coli* B-95. ΔA cells[4] were co-transformed with pRSF-G1mCNPRS or pRSF-G1pCNPRS, respectively, and the pCDF plasmid containing the amber codon-interrupted gene of the protein of interest. The cells were grown at 37 °C in LB medium containing 25 mg/L kanamycin and 25 mg/L spectinomycin. An aliquot (0.5 mL) of an overnight culture was used to inoculate 50 mL LB medium supplemented with 25 mg/L kanamycin, 25 mg/L spectinomycin, and 1 mM mCNP or 2 mM pCNP. The cells were grown at 37 °C to an OD_{600} of 0.6–1. At this point, the temperature was reduced to 25 °C and protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

To produce the wild-type proteins, the respective genes were cloned between the *Nde*I and *BamH*I restriction sites of the pET-3a expression plasmid (Novagen, USA). The corresponding pET-3a plasmid was then transformed into *E. coli* BL21(DE3) cells and the cells were recovered at 37 °C in LB medium containing 100 mg/L carbenicillin. An aliquot (0.5 mL) of an overnight culture was used to inoculate 50 mL LB medium supplemented with 100 mg/L carbenicillin. The cells were grown at 37 °C to an OD_{600} of 0.6–1. Afterwards, the temperature was reduced to 25 °C and protein expression was induced by the addition of 1 mM IPTG.

After protein expression for 16 h, the cells were harvested by centrifugation at 4000 *g* for 15 minutes at 4 °C. Following resuspension in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole), the cells were lysed using an Avestin Emulsiflex C5 (Avestin, Canada) using two passes at 10,000–15,000 psi. The cell lysates were centrifuged for 1 h at 30,000 *g* at 4 °C. The supernatant was loaded onto a 1 mL His GraviTrap column (Cytiva, USA). The column was washed with 20 column volumes buffer B (same as buffer A but with 20 mM imidazole) and the protein was eluted with 5 column volumes buffer C (same as buffer A but with 500 mM imidazole). Afterwards, the buffer was exchanged to PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) using an Amicon ultrafiltration centrifugal tube (Merck Millipore, USA) with the appropriate molecular weight cut-off.

Protein tagging with aminothiols

To a 50 μM protein sample in PBS buffer containing 10 mM TCEP, 5 mM cysteine or EDDC was added (from a 100 mM stock solution in PBS). The reaction was incubated with shaking at 25 °C for 4 h. Next, the buffer was exchanged to PBS using an Amicon centrifugal ultrafiltration tube.

Expression, purification, and crystallization of G1mCNPRS

For the expression of G1mCNPRS, high cell-density fermentation was used based on the protocol published by Gossert et al.[5] The gene encoding G1mCNPRS with an N-terminal His_6 tag and TEV cleavage site was cloned between the *Nde*I and *BamH*I restriction sites of pET-3a plasmid (Novagen, USA). pET-G1mCNPRS plasmid was transformed into *E. coli* BL21(DE3) cells. The cells were grown at 37 °C in LB medium containing 100 mg/L carbenicillin. Next, the cells were used to inoculate 50 mL TB medium supplemented with 100 mg/L carbenicillin (pre-culture). Following growth at 37 °C for 16 h, the pre-culture was used to inoculate 950 mL rich fermenter medium containing 100 mg/L carbenicillin in a Labfors 5 bioreactor (Infors-HT, Switzerland). The cells were grown at 37 °C with air flow = 2.5 L/min, minimal pO_2 = 20%, stirring cascade: 500–1200 rpm, and pH = 7.0. When the OD_{600} value of the culture reached 15, temperature was reduced to 18 °C and 1 mM IPTG was added for the induction of protein expression.

After 16 h, the fermentation was stopped, and the cells were harvested by centrifugation at 4000 *g* for 15 minutes.

The harvested cells were resuspended in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole) followed by lysis using an Avestin Emulsiflex C5 (Avestin, Canada; two passes using 10,000–15,000 psi). The clarified cell lysate was loaded onto a 5 mL HisTrap FF column connected to an ÄKTA pure 25 chromatography system (Cytiva, USA). The column was washed with 20 column volumes buffer B (same as buffer A but with 20 mM imidazole) and the protein was eluted with 5 column volumes buffer C (same as buffer A but with 500 mM imidazole). The eluted protein was desalted using a HiPrep Desalting 26/10 column (Cytiva, USA) equilibrated with buffer D (50 mM Tris-HCl pH 7.5, 300 mM sodium chloride, 1 mM dithiothreitol (DTT)) and TEV proteolytic cleavage was done by incubation with TEV protease at 4 °C for 16 h. The cleaved G1mCNPRS was recovered by a reverse IMAC step.

Finally, G1mCNPRS was loaded onto a HiLoad 26/600 Superdex 200 pg column (Cytiva, USA) equilibrated with buffer E (50 mM Tris-HCl pH 7.5, 400 mM sodium chloride). After gel filtration, the fractions containing G1mCNPRS were combined and concentrated to 10 mg/mL using an Amicon ultrafiltration centrifugal tube (Merck Millipore, USA) with a molecular weight cut-off of 10 kDa and buffer-exchanged to buffer F (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 3 mM mCNP).

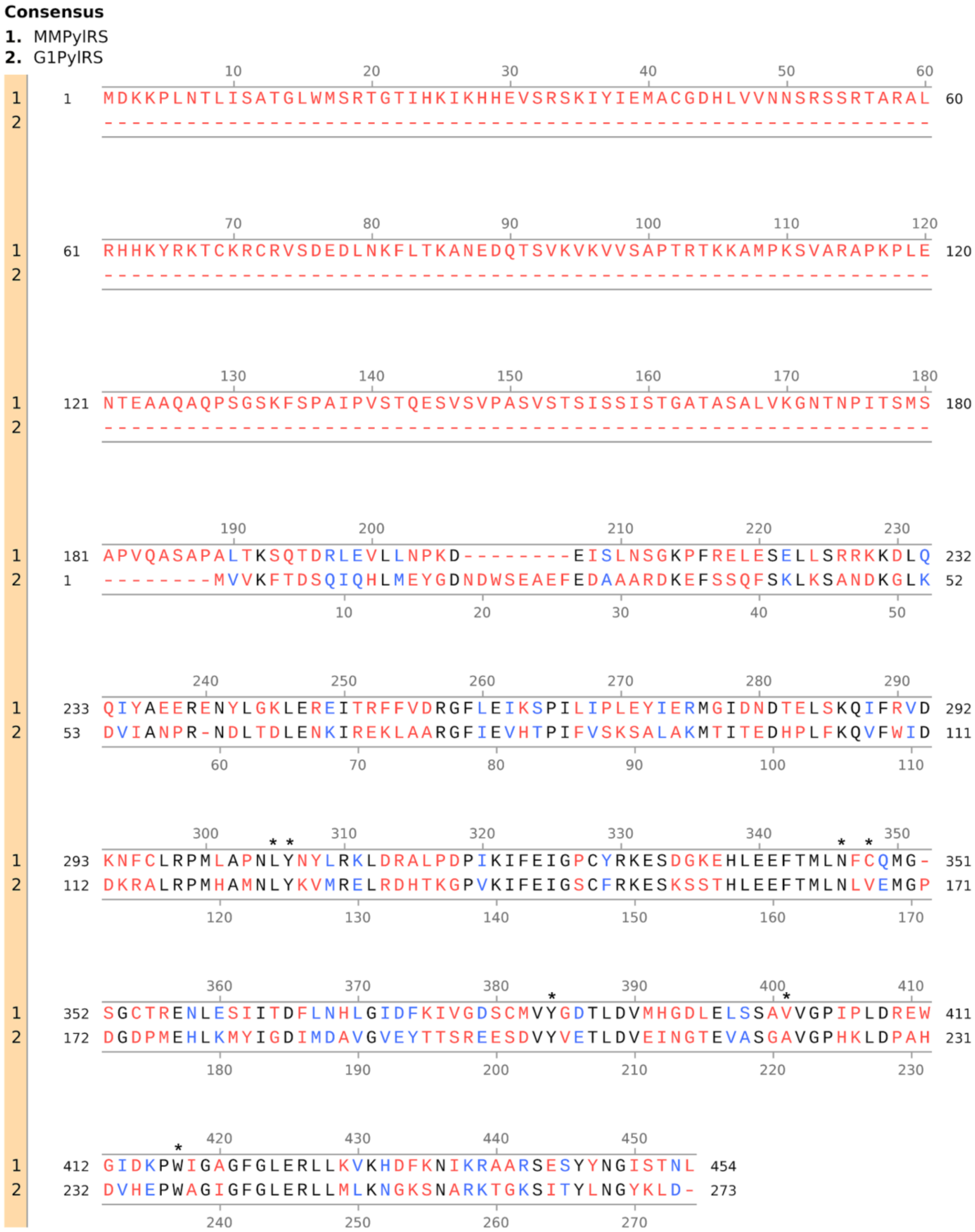
High throughput screening of crystallization conditions was performed using sparse matrix screens Index and Crystal Screen HT (Hampton Research) and SG1 and JCSG (Molecular Dimensions), at 18 °C with 1:1 ratio of protein:reservoir in a total volume of 1 μL, using a protein concentration of 10 mg/ml. Crystallization of G1mCNPRS was optimized by hanging-drop vapor diffusion at 18 °C. Large rod-shaped orthorhombic crystals were obtained by equilibrating drops of 1 μL protein and 1 μL reservoir against a solution of 30% w/v PEG 4000, 100 mM Tris-HCl (pH 9.0), and 200 mM lithium sulfate. Crystals were cryoprotected using Parabar (Hampton Research) prior to flash-freezing in liquid nitrogen.

Diffraction data were collected at 100 K at the MX2 beamline at the Australian Synchrotron.[6] Reflections collected were indexed and integrated using XDS[7] and scaled in Aimless (CCP4).[8] The phase problem was overcome by molecular replacement in Phaser MR (CCP4),[9] using PDB ID 6JP2 with sequence modified in Sculptor[10] as the search model. The structure was refined by iterative rounds of rebuilding in Coot,[11] and refinement in phenix.refine.[12-13] Data collection and refinement statistics are given in **Table S8**. The completed structure was deposited in the Protein Data Bank (PDB: 7R6O).

Intact protein mass spectrometry

Intact protein analysis was performed on an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific, USA) connected to a Thermo Fisher Scientific UltiMate 3000 HPLC system equipped with ZORBAX 300SB-C3, 3.5 µm, 4.6 x 50 mm HPLC column (Agilent Technologies, USA). Approximately 50 pmol of sample was injected using a 500 μL/min linear gradient of solvent A (0.1% (v/v) formic acid in water) and solvent B (0.1% (v/v) formic acid in acetonitrile), ramping solvent B from 5% at the start to 80% after 12 min. Data were collected using an electrospray ionization (ESI) source in positive ion mode. Protein intact mass was determined by deconvolution using the program Xcalibur 3.0.63 (Thermo Fisher Scientific, USA).

Supplementary Figures



Consensus

1. MMPylRS
2. G1PylRS

```
1    1 MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARAL 60
2      ------------------------------------------------------------

1   61 RHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLE 120
2      ------------------------------------------------------------

1  121 NTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMS 180
2      ------------------------------------------------------------

1  181 APVQASAPALTKSQTDRLEVLLNPKD--------EISLNSGKPFRELESELLSRRKKDLQ 232
2    1 --------MVVKFTDSQIQHLMEYGDNDWSEAEFEDAAARDKEFSSQFSKLKSANDKGLK 52

1  233 QIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVD 292
2   53 DVIANPR-NDLTDLENKIREKLAARGFIEVHTPIFVSKSALAKMTITEDHPLFKQVFWID 111

                   **                                       * *
1  293 KNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMG- 351
2  112 DKRALRPMHAMNLYKVMRELRDHTKGPVKIFEIGSCFRKESKSSTHLEEFTMLNLVEMGP 171

                                       *                *
1  352 SGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREW 411
2  172 DGDPMEHLKMYIGDIMDAVGVEYTTSREESDVYVETLDVEINGTEVASGAVGPHKLDPAH 231

            *
1  412 GIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL 454
2  232 DVHEPWAGIGFGLERLLMLKNGKSNARKTGKSITYLNGYKLD- 273
```

Figure S1. Sequence alignment of *Mm*PylRS and G1PylRS. The alignment was done using the semi-global alignment tool in SnapGene® software (from Insightful Science; available at snapgene.com). Residues referenced in the main text are indicated by asterisks.

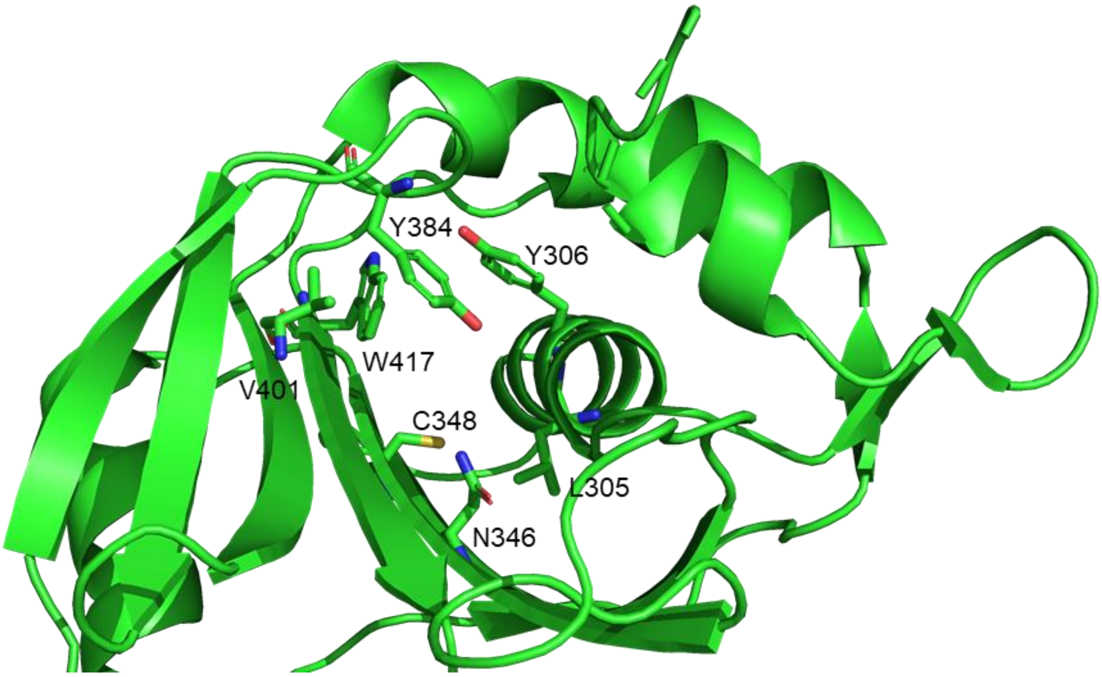


Figure S2. Substrate binding pocket of *Mm*PylRS (PDB: 2Q7E). The residues identified as essential for substrate recognition are shown as sticks.

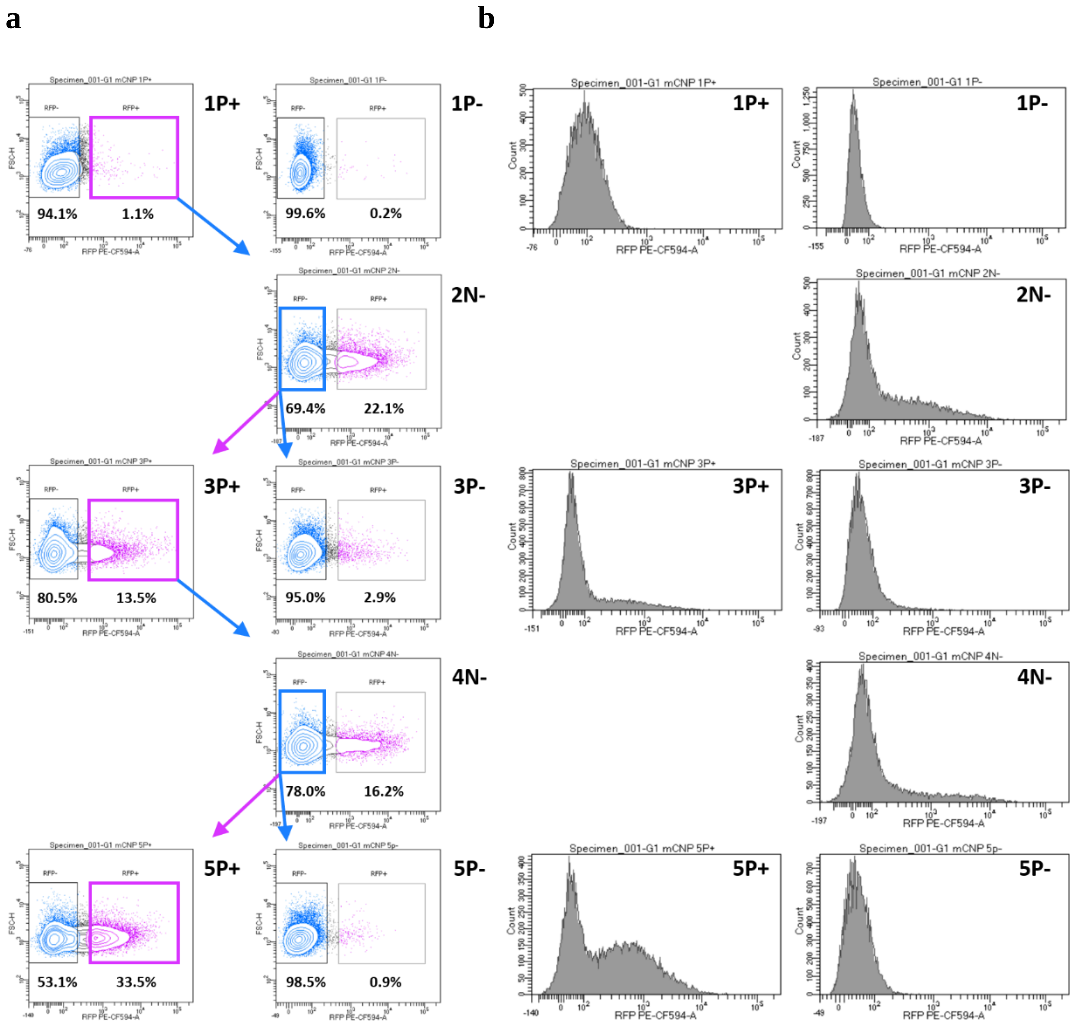


Figure S3. a, FACS enrichment of *E. coli* cells with active and specific G1PylRS enzymes that recognize mCNP. **P** and **N** labels refer to positive and negative selection rounds, respectively. Violet/blue frames identify the fractions collected from each positive/negative round. **+/-** labels and violet/blue arrows refer to cell growth in the presence or absence of target ncAA (herein, 3 mM of mCNP). The vertical axis plots the forward scatter that reports on cell size and the horizontal axis indicates the level of RFP fluorescence. 33.5% of the **5P+** sample with the brightest RFP fluorescence level were collected and further characterized. b, Histograms of cell counts in all five rounds of mCNPRS selection. At the fifth round of positive selection, the ratio of cells with high RFP fluorescence was much greater in the presence of mCNP (**5P+**) than in the absence of mCNP (**5P-**).

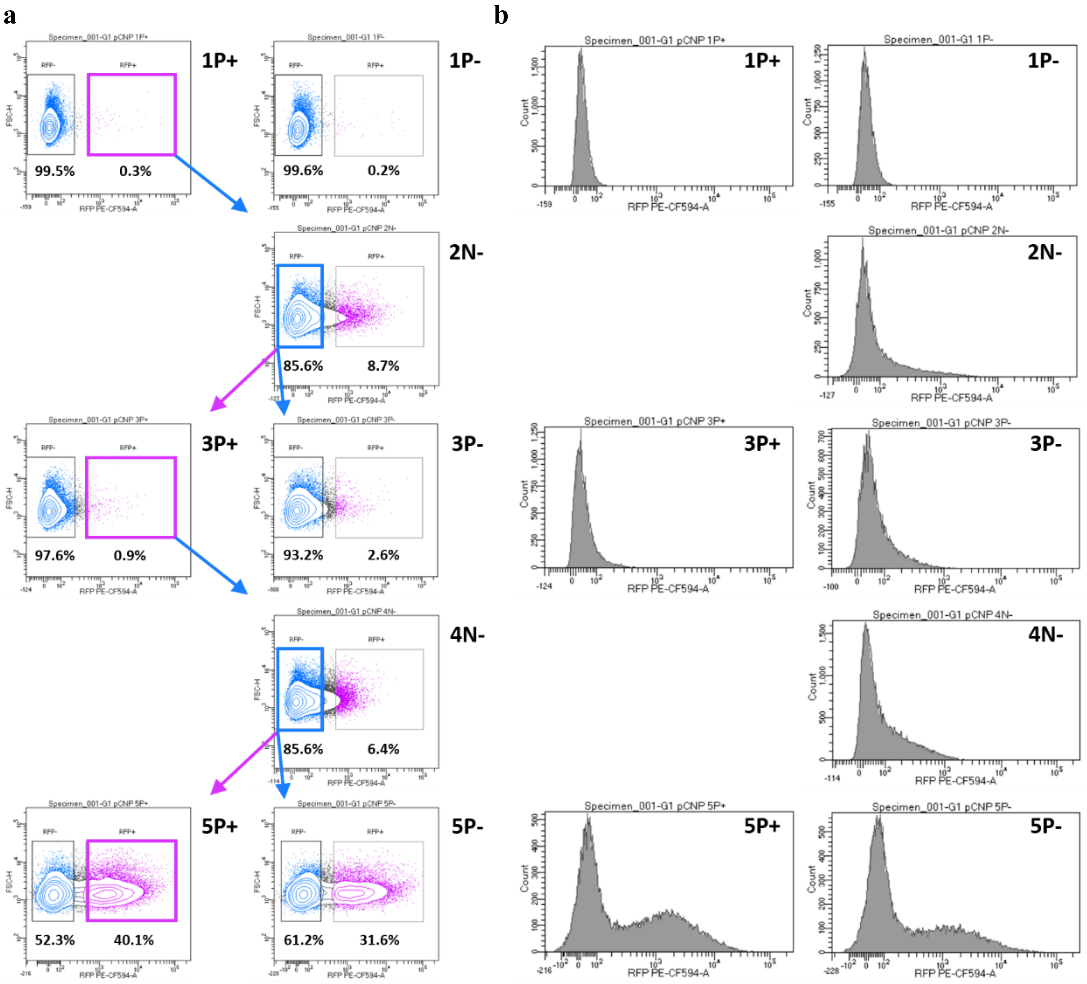


Figure S4. a, FACS enrichment of *E. coli* cells with G1PylRS enzymes specific for pCNP. **P**/**N** and +/- labels, violet/blue frames and arrows have the same indication as in Figure S3. b, Histograms of cell counts in all five rounds of selection for active and specific G1pCNPRS. The ratio of cells with high RFP fluorescence in the **5P+** sample (40.1%) was 1.27 times higher than in the **5P-** sample (31.6%), which indicates a good chance of selecting true-positive candidates during further characterization.

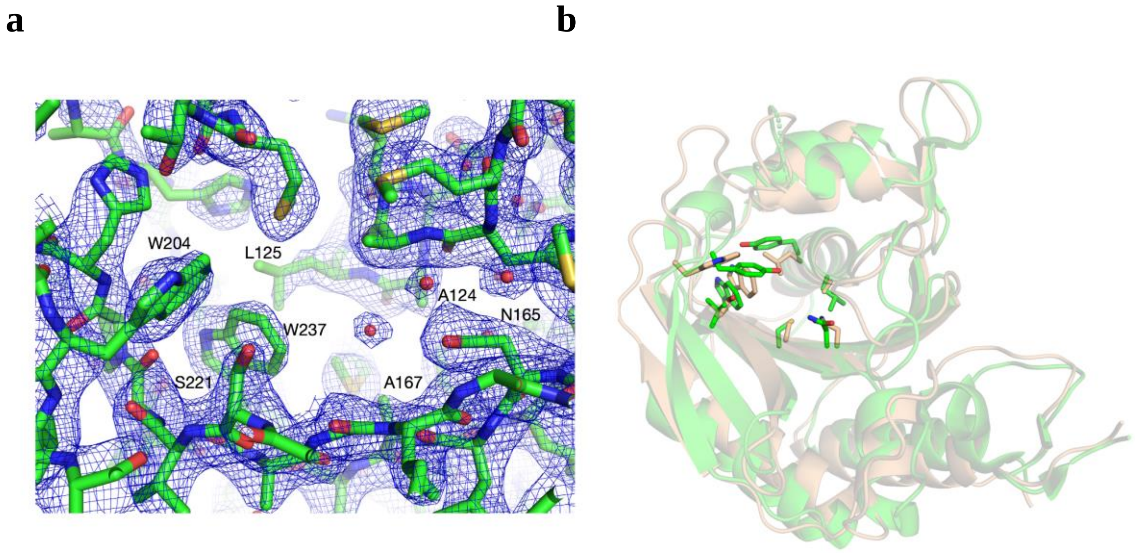


Figure S5. The crystal structure of G1mCNPRS (PDB: 7R6O). a, The amino acid binding site of G1mCNPRS, with $2mF_o$-DF_c electron density, showing the binding site is well resolved in the crystal structure but no significant electron density for the m-CNP substrate is observed. Residues that were mutated (A124, L125, N165, A167, W204, W237, A167) are labelled. b, The structure of the amino acid binding domain of the G1mCNP tRNA synthetase (beige) is highly conserved, and superimposes with 1.63 Å Cα-RMSD over 186 residues on the C-terminal domain of the pyrrolysyl-tRNA synthetase from *M. mazei* (PDB: 2Q7E; green).[14] Residues that were mutated (A124, L125, N165, A167, W204, W237, A167) are shown as sticks.

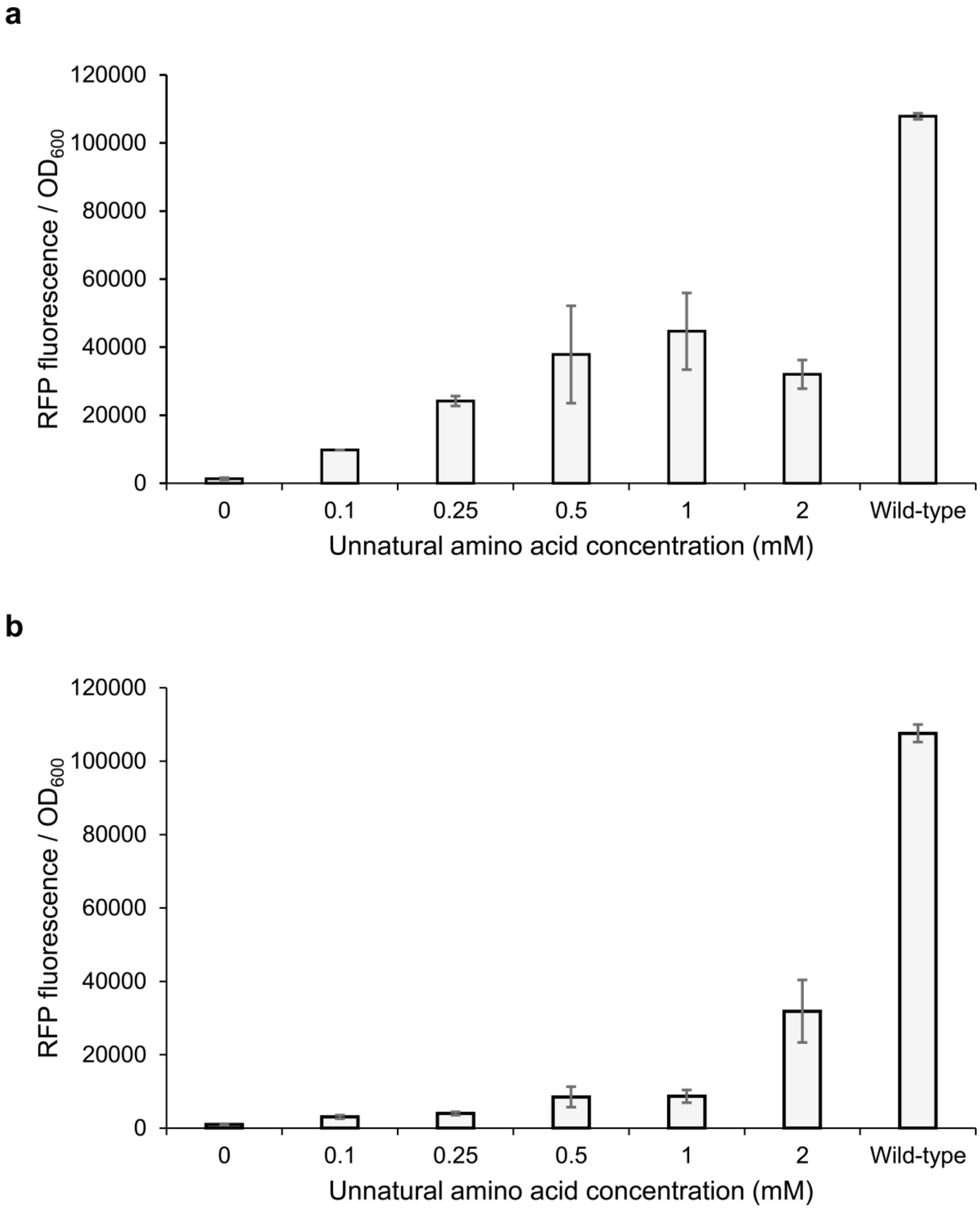


Figure S6. Effect of ncAA concentration on the expression yield of RFP 13X with different cyanopyridylalanine ncAAs. a, mCNP. b, pCNP. RFP fluorescence is normalized by the optical density (OD_{600}) and reported as the mean of three biological replicates ± standard error. Relative to wild-type RFP, 41 and 30 % suppression efficiency were achieved at 1 mM mCNP and 2 mM pCNP, respectively.

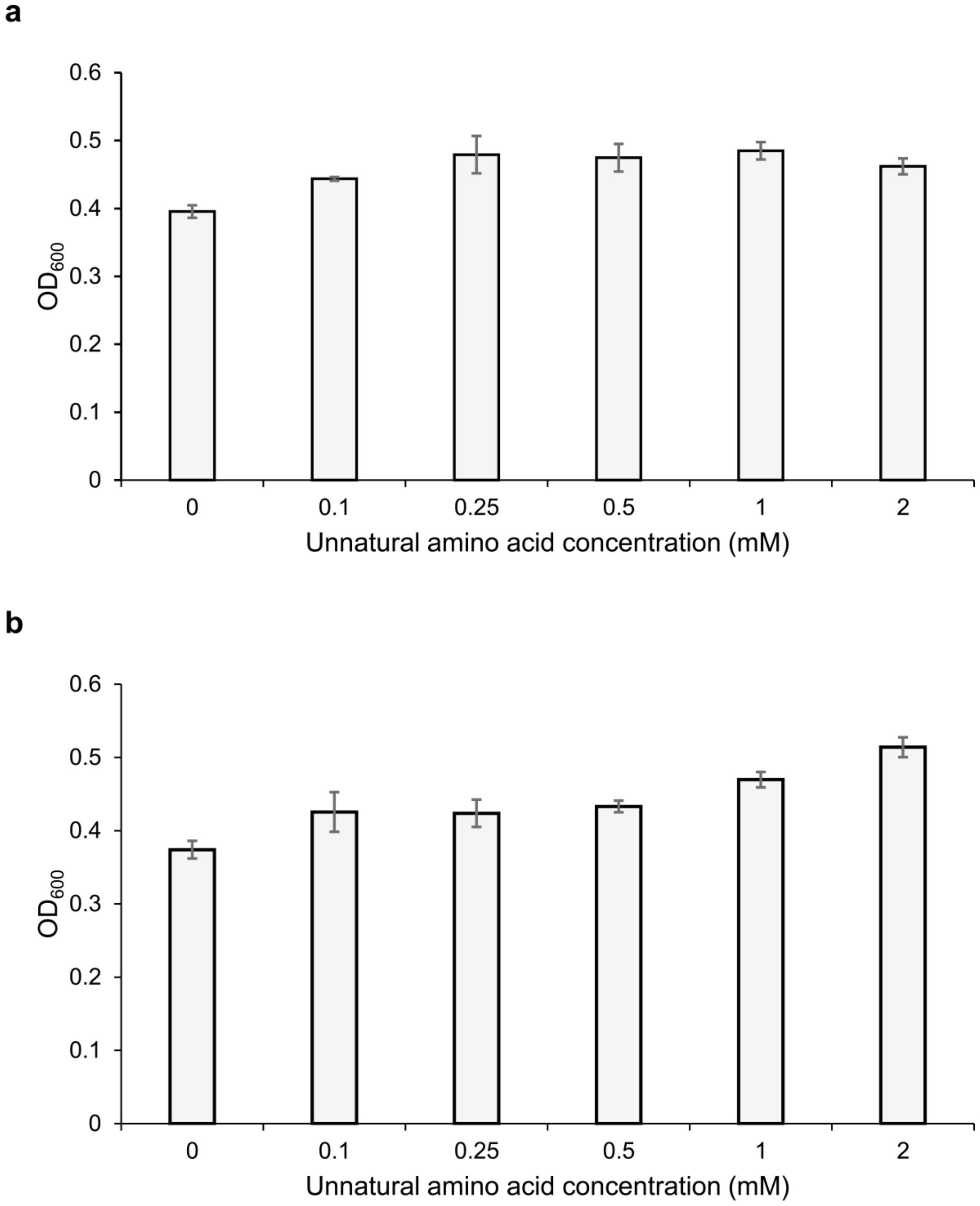


Figure S7. Effect of ncAA concentration on *E. coli* cell growth. Both mCNP and pCNP are well tolerated by the cells, without any negative effects of cell growth. OD$_{600}$ values are reported as the mean of three biological replicates ± standard error measured after incubation with the corresponding ncAA for 16 h at 25 °C.

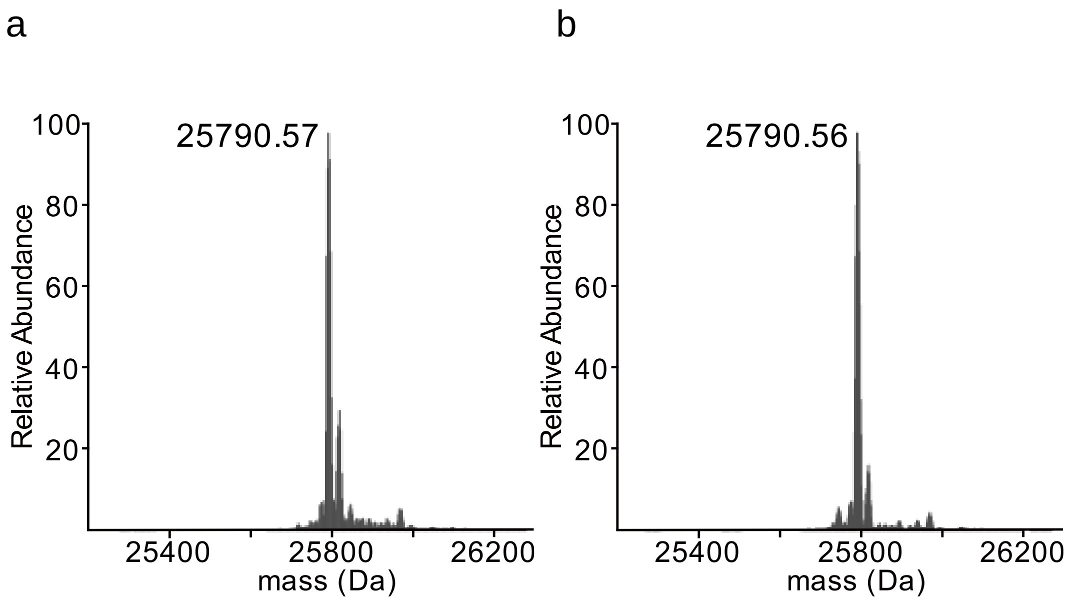


Figure S8. Intact protein mass spectra of the fusion protein NT-Ubi 7X (X indicating the position of the ncAA in the amino acid sequence) showing the high incorporation yield of cyanopyridylalanines. a) NT-Ubi 7mCNP, b) NT-Ubi 7pCNP. The calculated mass minus the N-terminal methionine is 25791.91 Da. The minor peak at 25817.57 Da is likely from formylation in the N-terminal NT domain, because after TEV protease cleavage this modification is not observed in the mass spectra of Cys-Ubi-X7.

Supplementary Tables

Table S1. Mutations found in the 7 selected G1PylRS variants that recognize mCNP. The selected mutation set in grey indicates the variant used for later applications.

Mutant	Site					
*Mm*PylRS wt	L305	Y306	N346	C348	Y384	V401	W417
G1PylRS wt	L124	Y125	N165	V167	Y204	A221	W237
G1mCNP11	R	C	S	C	W	A	W
G1mCNP13	G	R	N	V	W	C	W
G1mCNP16	A	L	N	A	W	S	W
G1mCNP34	A	L	N	A	W	S	W
G1mCNP43	A	F	N	V	F	S	Y
G1mCNP58	A	L	N	C	W	S	W
G1mCNP60	A	M	N	S	W	C	W

Table S2. Mutations found in the 17 selected G1PylRS variants that recognize pCNP. The selected mutation set in grey indicates the variant used for later applications.

Mutant			Site				
*Mm*PylRS wt	L305	Y306	N346	C348	Y384	V401	W417
G1PylRS wt	L124	Y125	N165	V167	Y204	A221	W237
G1pCNP01	A	F	N	V	W	A	Y
G1pCNP02	S	F	N	V	W	S	Y
G1pCNP03	A	F	N	V	W	S	Y
G1pCNP04	A	F	N	V	W	S	Y
G1pCNP05	A	F	N	V	W	S	Y
G1pCNP06	A	F	N	V	W	A	Y
G1pCNP07	A	F	N	V	W	A	Y
G1pCNP08	A	F	N	V	W	A	Y
G1pCNP09	A	F	N	V	W	S	Y
G1pCNP11	A	F	N	V	W	A	Y
G1pCNP12	A	F	N	V	W	S	Y
G1pCNP16	A	F	N	V	W	A	Y
G1pCNP26	A	F	N	V	W	A	Y
G1pCNP37	A	F	N	V	W	S	Y
G1pCNP42	A	F	N	V	W	S	Y
G1pCNP43	A	F	N	V	W	A	Y
G1pCNP58	A	F	N	V	W	S	Y

Table S3. Yields of proteins expressed in the present study reported as mg per 1 L cell culture[a].

Protein[b]	mCNP	pCNP
RFP 13X	115	75
RFP 237X/243X	69	N/A
RFP 204X/237X	70	N/A
GB1 11X	92	25
GB1 24X/28X	68	16
Ppib 147X	100	23
NT-Ubi 7X	256	162
Cys-Ubi 7X	134	60

[a] Yields are calculated based on the expression of 50 mL cell culture.

[b] X indicates the position of the ncAA (either mCNP or pCNP).

Table S4. Expected mass of the proteins reported in the present study.

Protein	Expected mass (Da)
RFP 13X[a] (-Met[b])	26714.06
RFP 13X + cysteine	26818.19
RFP 237X/243X (-Met)	27504.01
RFP 237X/243X + EDDC	27736.33[c]
RFP 204X/237X (-Met)	26876.26
RFP 204X/237X + EDDC	27108.58[c]
GB1 11X (-Met)	8390.17
GB1 11X + cysteine	8494.30
GB1 24X/28X (-Met)	8465.19
GB1 24X/28X + EDDC	8697.51[c]
Ppib 147X	19012.40
Ppib 147X + cysteine	19116.53
Linear Cys-Ubi 7X	10519.94
Cyclized Cys-Ubi 7X	10502.91

[a] X indicates the position of the ncAA (either mCNP or pCNP).

[b] -Met indicates loss of the N-terminal methionine during protein expression.

[c] Expected mass of the cyclization product from the reaction with EDDC.

Table S5. Complete nucleotide sequences of the plasmids used in the current study.*

Plasmid	DNA sequence
pBK-G1RS	ATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGAT TTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTAT GGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACA GATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATC CGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTA GAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCAT TCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCA CGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAA CAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGAT TTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTC GGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCA TTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCAT TTGATGCTCGATGAGTTTTTCTAACACTGGCAGAGCATTACGCTGACTTGGAGGATCTAGGTGAAG ATCCTTGGTACGGCCGCCAAGCTTAAAAAAAATCCTTAGCTTTCGCTAAGGATCTGCAGTGGCGAA AGGCCGGGGGGTCGAACCCCGCTTACAGGTTTTAGAGACCCGTTTGCTCGCCGGAGCGCCCTCCGA ATTCAGCGTTACAAGTATTACACAAAGTTTTTTATGTTGAGAATATTTTTTTGATGTCCTCGGGTT GTCAGCCTGTCCCGCTTTAATATCATACGCCGTTATACGTTGTTTACGCTTTGAGGAGGATCCATA TGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATATGGTGATAATGATTGGAGCG AAGCCGAATTTGAAGATGCAGCAGCACGTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAA GCGCCAATGATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGATCTGACCGATCTGGAAA ACAAAATTCGCGAAAAACTGGCAGCCCGTGGTTTTATTGAAGTTCATACCCCGATTTTTGTGAGCA AAAGCGCACTGGCAAAAATGACCATTACCGAAGATCATCCGCTGTTCAAACAGGTGTTTTGGATTG ATGATAAACGTGCACTGCGTCCGATGCATGCAATGAATCTGTATAAAGTTATGCGTGAACTGCGCG ATCATACCAAAGGTCCGGTTAAAATCTTTGAAATTGGTAGCTGCTTTCGCAAAGAAAGCAAAAGCA GTACCCATCTGGAAGAATTTACCATGCTGAACCTGGTTGAAATGGGTCCTGATGGTGATCCGATGG AACATCTGAAAATGTATATTGGCGATATCATGGATGCCGTTGGTGTTGAATATACCACCAGTCGTG AAGAATCAGATGTTTATGTTGAAACCCTGGACGTGGAAATTAATGGCACCGAAGTTGCAAGCGGTG CAGTTGGTCCGCATAAACTGGATCCGGCACATGATGTGCATGAACCGTGGGCAGGTATTGGTTTTG GTCTGGAACGTCTGCTGATGCTGAAAAATGGTAAAAGCAATGCACGCAAAACCGGCAAAAGTATTA CCTATCTGAATGGCTACAAACTGGATTAACTGCGTCGACGTCTTGGTTTAGAAACAGCAAACAATC CAAAACGCCGCGTTCAGCGGCGTTTTTTCTGCTTTTCTTCGCGAATTAATTCCGCTTCGCACATGT GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC TCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGC TTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTA GGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAG CCAGTTACCTTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAGCCACCGCTGGTAGCGG TGGTTTTTTTGTTTGCAGCAGCAGATTACACGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCT TTTCTACGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAA AACTGTCTGCTTACATAAACAGTAATACAAGGGGTGT
pRSF-G1mCNPRS	TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAG AAGGAGATATACATATGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATATGGTG ATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCAGCACGTGATAAAGAATTTAGCAGCCAGT TTAGCAAACTGAAAAGCGCCAATGATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGATC TGACCGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCCCGTGGTTTTATTGAAGTTCATACCC CGATTTTTGTGAGCAAAAGCGCACTGGCAAAAATGACCATTACCGAAGATCATCCGCTGTTCAAAC AGGTGTTTTGGATTGATGATAAACGTGCACTGCGTCCGATGCATGCAATGAATGCTTTGAAAGTTA TGCGTGAACTGCGCGATCATACCAAAGGTCCGGTTAAAATCTTTGAAATTGGTAGCTGCTTTCGCA AAGAAAGCAAAAGCAGTACCCATCTGGAAGAATTTACCATGCTGAACCTGGCAGAAATGGGTCCTG ATGGTGATCCGATGGAACATCTGAAAATGTATATTGGCGATATCATGGATGCCGTTGGTGTTGAAT ATACCACCAGTCGTGAAGAATCAGATGTTTGGGTTGAAACCCTGGACGTGGAAATTAATGGCACCG

	AAGTTGCAAGCGGTTCTGTTGGTCCGCATAAACTGGATCCGGCACATGATGTGCATGAACCGTGGGCAGGTATTGGTTTTGGTCTGGAACGTCTGCTGATGCTGAAAAATGGTAAAAGCAATGCACGCAAAACCGGCAAAAGTATTACCTATCTGAATGGCTACAAACTGGATTAAGAATTCGAGCTCCCGGGTACCAGATAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGTTTGTGAGCTCCCGTAATTCCGCTTCGCAACATGTGAGCACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTGGAGGGCGCTCCGGCGAGCAAACGGGTCTCTAAAACCTGTAAGCGGGGTTCGACCCCCCGGCCTTTCGCCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAACTTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGTGTCAGCTCACTCAAAAGCGGTAATACGGTTATCCACAGAATCAGGGGATAAAGCCGGAAAGAACATGTGAGCAAAAAGCAAAGCACCGGAAGAAGCCAACGCCGCAGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGCCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTTGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCATTGGTAACTGATTTAGAGGACTTTGTCTTGAAGTTATGCACCTGTTAAGGCTAAACTGAAAGAACAGATTTTGGTGAGTGCGGTCCTCCAACCCACTTACCTTGGTTCAAAGAGTTGGTAGCTCAGCGAACCTTGAGAAAACCACCGTTGGTAGCGGTGGTTTTTCTTTATTTATGAGATGATGAATCAATCGGTCTATCAAGTCAACGAACAGCTATTCCGTTGAACTGTGGTACGCGTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGGTCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGCATGCAGCGCTTGACTTGACGGGACGGCGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAA
pRSF-G1pCNPRS	TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATATGGTGATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCAGCACGTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAAGCGCCAATGATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGATCTGACCGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCCCGTGGTTTTATTGAAGTTCATACCCCGATTTTTGTGAGCAAAAGCGCACTGGCAAAAATGACCATTACCGAAGATCATCCGCTGTTCAAACAGGTGTTTTGGATTGATGATAAACGTGCACTGCGTCCGATGCATGCAATGAATGCTTTTAAAGTTATGCGTGAACTGCGCGATCATACCAAAGGTCCGGTTAAAATCTTTGAAATTGGTAGCTGCTTTCGCAAAGAAAGCAAAAGCAGTACCCATCTGGAAGAATTTACCATGCTGAACCTGGTAGAAATGGGTCCTGATGGTGATCCGATGGAACATCTGAAAATGTATATTGGCGATATCATGGATGCCGTTGGTGTTGAATATACCACCAGTCGTGAAGAATCAGATGTTTGGGTTGAAACCCTGGACGTGGAAATTAATGGCACCGAAGTTGCAAGCGGTTCTGTTGGTCCGCATAAACTGGATCCGGCACATGATGTGCATGAACCGTATGCAGGTATTGGTTTTGGTCTGGAACGTCTGCTGATGCTGAAAAATGGTAAAAGCAATGCACGCAAAACCGGCAAAAGTATTACCTATCTGAATGGCTACAAACTGGATTAAGAATTCGAGCTCCCGGGTACCAGATAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGTTTGTGAGCTCCCGTAATTCCGCTTCGCAACATGTGAGCACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTGGAGGGCGCTCCGGCGAGCAAACGGGTCTCTAAAACCTGTAAGCGGGGTTCGACCCCCCGGCCTTTCGCCAAATTCGAAAAGCCTGCTCAACGA

	GCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAACTTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGTGTCAGCTCACTCAAAAGCGGTAATACGGTTATCCACAGAATCAGGGGATAAAGCCGGAAAGAACATGTGAGCAAAAAGCAAAGCACCGGAAGAAGCCAACGCCGCAGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGCCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTTGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCATTGGTAACTGATTTAGAGGACTTTGTCTTGAAGTTATGCACCTGTTAAGGCTAAACTGAAAGAACAGATTTTGGTGAGTGCGGTCCTCCAACCCACTTACCTTGGTTCAAAGAGTTGGTAGCTCAGCGAACCTTGAGAAAACCACCGTTGGTAGCGGTGGTTTTTCTTTATTTATGAGATGATGAATCAATCGGTCTATCAAGTCAACGAACAGCTATTCCGTTGAACTGTGGTACGCGTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGGTCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGCATGCAGCGCTTGACTTGACGGGACGGCGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAA
pCDF-PpiB147TAG	TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGTTACTTTCCACACCAATCACGGCGATATTGTCATCAAAACTTTTGACGATAAAGCACCTGAAACAGTTAAAAACTTCCTGGACTACTGCCGCGAAGGTTTTTACAACAACACCATTTTCCACCGTGTTATCAACGGCTTTATGATTCAGGGCGGCGGTTTTGAACCGGGCATGAAACAAAAAGCCACCAAAGAACCGATCAAAAACGAAGCCAACAACGGCCTGAAAAATACCCGTGGTACGCTGGCAATGGCACGTACTCAGGCTCCGCACTCTGCAACTGCACAGTTCTTCATCAACGTGGTTGATAACGACTTCCTGAACTTCTCTGGCGAAAGCCTGCAAGGTTGGGGCTACTGCGTGTTTGCTGAAGTGGTTGACGGCATGGACGTGGTAGACAAAATCAAAGGTGTTGCAACCGGTCGTAGCGGTATGTAGCAGGACGTGCCAAAAGAAGACGTTATCATTGAAAGCGTGACCGTTAGCGAGCACCACCATCATCACCACTAATAAAGAATTCGAGCTCCCGGGTACCATGGCATGCATCGATAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTACAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTAC

	GGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAAT
pBAD-TEV	TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGGAGAAAGCTTGTTTAAGGGGCCGCGTGATTACAACCCGATATCGAGCAGCATTTGTCATTTGACGAATGAATCTGATGGGCACACAACATCGTTGTATGGTATTGGATTTGGTCCCTTCATCATTACAAACAAGCACTTGTTTAGAAGAAATAATGGAACACTGTTGGTCCAATCACTACATGGTGTATTCAAGGTCAAGGATACCACGACTTTGCAACAACACCTCGTGGATGGGAGGGACATGATAATTATTCGCATGCCTAAGGATTTCCCACCATTTCCTCAAAAGCTGAAATTTAGAGAGCCACAAAGGGAAGAGCGCATATGTCTTGTGACAACCAACTTCCAAACTAAGAGCATGTCTAGCATGGTGTCAGACACTAGTTGCACATTCCCTTCATCTGATGGCATATTCTGGAAGCATTGGATTCAAACCAAGGATGGGCAGTGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTCATTGTTGGTATACACTCAGCATCGAATTTCACCAACACAAACAATTATTTCACAAGCGTGCCGAAAAACTTCATGGAATTGTTGACAAATCAGGAGGCGCAGCAGTGGGTTAGTGGTTGGCGATTAAATGCTGACTCAGTATTGTGGGGGGCCATAAAGTTTTCATGGTGAAACCTGAAGAGCCTTTTCAGCCAGTTAAGGAAGCGACTCAACTCATGAATTAAGAATTCGAGCTCCCGGGTACCATGGCATGCATCGATAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGTGCGTTTCTACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTACGCGCCC

```
TGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGC
GCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGT
CAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAA
AAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG
ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTTGAACAACACTCAACCCTATC
TCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTG
ATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTAAAAGGATCTAGGT
GAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTT
GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTT
TCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTT
AGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
CAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAG
ACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATTTC
TGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAATAACT
GCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCT
GCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTC
GCCTTGCGTATAATATTTGCCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGG
TGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGA
TACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCG
GAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTG
AGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGAC
GCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCT
CCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTT
TGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCAC
GAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAA
AGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTC
ATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACC
TCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTT
TTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAA
ATATTTGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGC
GCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCTGCTC
ATGTTTGACAGCTTATCATCGATCAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACC
ATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTG
ACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCC
AACTTTTGGCGAAAATGAGACGTCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGCCCGCG
AAAT
```

* The sequence encoding the protein of interest is underlined.

Table S6. DNA and corresponding amino acid sequences of the proteins used in the current study.

Protein	DNA sequence	Amino acid sequence[a]
G1RSwt	ATGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATATGGTGATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCAGCACGTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAAGCGCCAATGATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGATCTGACCGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCCCGTGGTTTTATTGAAGTTCATACCCCGATTTTTGTGAGCAAAAGCGCACTGGCAAAAATGACCATTACCGAAGATCATCCGCTGTTCAAACAGGTGTTTTGGATTGATGATAAACGTGCACTGCGTCCGATGCATGCAATGAATCTGTATAAAGTTATGCGTGAACTGCGCGATCATACCAAAGGTCCGGTTAAAATCTTTGAAATTGGTAGCTGCTTTCGCAAAGAAAGCAAAAGCAGTACCCATCTGGAAGAATTTACCATGCTGAACCTGGTTGAAATGGGTCCTGATGGTGATCCGATGGAACATCTGAAAATGTATATTGGCGATATCATGGATGCCGTTGGTGTTGAATATACCACCAGTCGTGAAGAATCAGATGTTTATGTTGAAACCCTGGACGTGGAAATTAATGGCACCGAAGTTGCAAGCGGTGCAGTTGGTCCGCATAAACTGGATCCGGCACATGATGTGCATGAACCGTGGGCAGGTATTGGTTTTGGTCTGGAACGTCTGCTGATGCTGAAAAATGGTAAAAGCAATGCACGCAAAACCGGCAAAAGTATTACCTATCTGAATGGCTACAAACTGGATTAA	MVVKFTDSQIQHLMEYGDNDWSEAEFEDAAARDKEFSSQFSKLKSANDKGLKDVIANPRNDLTDLENKIREKLAARGFIEVHTPIFVSKSALAKMTITEDHPLFKQVFWIDDKRALRPMHAMNLYKVMRELRDHTKGPVKIFEIGSCFRKESKSSTHLEEFTMLNLVEMGPDGDPMEHLKMYIGDIMDAVGVEYTTSREESDVYVETLDVEINGTEVASGAVGPHKLDPAHDVHEPWAGIGFGLERLLMLKNGKSNARKTGKSITYLNGYKLD
RFP 13X	ATGGCTTCTATGACCGGTCATCACCATCACCATCACTAGGCCAGTAGTGAAGACGTTATCAAGGAGTTTATGCGTTTCAAAGTACGTATGGAGGGTAGTGTTAACGGACACGAATTTGAGATCGAGGGAGAGGGGGAAGGTCGTCCTTACGAGGGAACTCAAACGGCCAAATTAAAGGTGACCAAAGGTGGGCCCTTGCCATTCGCGTGGGACATCTTGTCACCCCAGTTCCAGTACGGGTCGAAGGCATACGTAAAACACCCAGCGGACATTCCTGACTATCTTAAGTTATCTTTCCCGGAAGGTTTTAAATGGGAACGCGTGATGAACTTTGAGGATGGGGGGGTTGTTACGGTGACACAAGACTCCTCATTGCAAGATGGAGAGTTTATCTATAAAGTCAAACTTCGCGGCACCAATTTTCCATCTGACGGTCCTGTAATGCAGAAAAAAACAATGGGCTGGGAAGCCTCCACAGAACGTATGTACCCCGAAGATGGAGCTTTAAAGGGCGAAATTAAAATGCGCTTAAAACTTAAAGACGGCGGCCATTACGACGCCGAAGTGAAAACGACGTATATGGCTAAGAAACCCGTCCAGCTTCCGGGAGCCTATAAAACTGACATCAAACTGGATATTACATCACACAACGAAGATTATACTATTGTCGAACAGTACGAACGCGCCGAAGGCCGCCATTCAACGGGAGCATAA	MASMTGHHHHHHXASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRHSTGA
RFP 237X/243X	ATGGCTTCTATGACCGGTCATCACCATCACCATCACATGGCCAGTAGTGAAGACGTTATCAAGGAGTTTATGCGTTTCAAAGTACGTATGGAGGGTAGTGTTAACGGACACGAATTTGAGATCGAGGGAGAGGGGGAAGGTCGTCCTTACGAGGGAACTCAAACGGCCAAATTAAAGGTGACCAAAGGTGGGCCCTTGCCATTCGCGTGGGACATCTTGTCACCCCAGTTCCAGTACGGGTCGAAGGCATACGTAAAACACCCAGCGGACATTCCTGACTATCTTAAGTTATCTTTCCCGGAAGGTTTTAAATGGGAACGCGTGATGAACTTTGAGGATGGGGGGGTTGTTACGGTGACACAAGACTCCTCATTGCAAGATGGAGAGTTTATCTATAAAGTCAAACTTCGCGGCACCAATTTTCCATCTGACGGTCCTGTAATGCAGAAAAAAACAATGGGCTGGGAAGCCTCCACAGAACGTATGTACCCCGAAGATGGAGCTTTAAAGGGCGAAATTAAAATGCGCTTAAAACTTAAAGACGGCGGCCATTACGACGCCGAAGTGAAAACGACGTATATGGCTAAGAAACCCGTCCAGCTTCCGGGAGCCTATAAAACTGACATCAAACTGGATATTACATCACACAACGAAGATTATACTATTGTCGAACAGTACGAACGCGCCGAAGGCCGCCATTCAACGGGATAGGGCAAACGCAAAAGCTAGTAA	MASMTGHHHHHHMASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRHSTGXGKRKSX

RFP 204X/237X	ATGGCTTCTATGACCGGTCATCACCATCACCATCACATGGCCAGT AGTGAAGACGTTATCAAGGAGTTTATGCGTTTCAAAGTACGTATG GAGGGTAGTGTTAACGGACACGAATTTGAGATCGAGGGAGAGGGG GAAGGTCGTCCTTACGAGGGAACTCAAACGGCCAAATTAAAGGTG ACCAAAGGTGGGCCCTTGCCATTCGCGTGGGACATCTTGTCACCC CAGTTCCAGTACGGGTCGAAGGCATACGTAAAACACCCAGCGGAC ATTCCTGACTATCTTAAGTTATCTTTCCCGGAAGGTTTTAAATGG GAACGCGTGATGAACTTTGAGGATGGGGGGGTTGTTACGGTGACA CAAGACTCCTCATTGCAAGATGGAGAGTTTATCTATAAAGTCAAA CTTCGCGGCACCAATTTTCCATCTGACGGTCCTGTAATGCAGAAA AAAACAATGGGCTGGGAAGCCTCCACAGAACGTATGTACCCCGAA GATGGAGCTTTAAAGGGCGAAATTAAAATGCGCTTAAAACTTAAA GACGGCGGCCATTACGACGCCGAAGTGAAAACGACGTATATGGCT AAGAAACCCGTCCAGCTTCCGGGATAGTATAAAACTGACATCAAA CTGGATATTACATCACACAACGAAGATTATACTATTGTCGAACAG TACGAACGCGCCGAAGGCCGCCATTCAACGGGATAGTAA	MASMTGHHHHHHMASSEDV IKEFMRFKVRMEGSVNGHE FEIEGEGEGRPYEGTQTAK LKVTKGGPLPFAWDILSPQ FQYGSKAYVKHPADIPDYL KLSFPEGFKWERVMNFEDG GVVTVTQDSSLQDGEFIYK VKLRGTNFPSDGPVMQKKT MGWEASTERMYPEDGALKG EIKMRLKLKDGGHYDAEVK TTYMAKKPVQLPGXYKTDI KLDITSHNEDYTIVEQYER AEGRHSTGX
GB1 11X	ATGGCTTCTATGACCGGTATGACCTACAAACTGATCCTGAACGGT AAATAGCTGAAAGGTGAAACCACCACCGAAGCGGTTGACGCGGCG ACCGCGGAAAAAGTTTTCAAACAGTACGCGAACGACAACGGTGTT GACGGTGAATGGACCTACGACGACGCGACCAAAACCTTCACCGTT ACCGAAGAAAACCTGTATTTTCAGGGCCATCATCATCACCATCAC TAA	MASMTGMTYKLILNGKXLK GETTTEAVDAATAEKVFKQ YANDNGVDGEWTYDDATKT FTVTEENLYFQGHHHHHH
GB1 24X/28X	ATGGCTTCTATGACCGGTATGACCTACAAACTGATCCTGAACGGT AAAACCCTGAAAGGTGAAACCACCACCGAAGCGGTTGACGCGTAG ACCGCGGAATAGGTTTTCAAACAGTACGCGAACGACAACGGTGTT GACGGTGAATGGACCTACGACGACGCGACCAAAACCTTCACCGTT ACCGAAGAAAACCTGTATTTTCAGGGCCATCATCATCACCATCAC TAA	MASMTGMTYKLILNGKTLK GETTTEAVDAXTAEXVFKQ YANDNGVDGEWTYDDATKT FTVTEENLYFQGHHHHHH
Ppib 147X	ATGGTTACTTTCCACACCAATCACGGCGATATTGTCATCAAAACT TTTGACGATAAAGCACCTGAAACAGTTAAAAACTTCCTGGACTAC TGCCGCGAAGGTTTTTACAACAACACCATTTTCCACCGTGTTATC AACGGCTTTATGATTCAGGGCGGCGGTTTTGAACCGGGCATGAAA CAAAAAGCCACCAAAGAACCGATCAAAAACGAAGCCAACAACGGC CTGAAAAATACCCGTGGTACGCTGGCAATGGCACGTACTCAGGCT CCGCACTCTGCAACTGCACAGTTCTTCATCAACGTGGTTGATAAC GACTTCCTGAACTTCTCTGGCGAAAGCCTGCAAGGTTGGGGCTAC TGCGTGTTTGCTGAAGTGGTTGACGGCATGGACGTGGTAGACAAA ATCAAAGGTGTTGCAACCGGTCGTAGCGGTATGTAGCAGGACGTG CCAAAAGAAGACGTTATCATTGAAAGCGTGACCGTTAGCGAGCAC CACCATCATCACCACTAA	MVTFHTNHGDIVIKTFDDK APETVKNFLDYCREGFYNN TIFHRVINGFMIQGGGFEP GMKQKATKEPIKNEANNGL KNTRGTLAMARTQAPHSAT AQFFINVVDNDFLNFSGES LQGWGYCVFAEVVDGMDVV DKIKGVATGRSGMXQDVPK EDVIIESVTVSEHHHHHH
NT-Ubi 7X[b]	ATGAGCCATACCACACCGTGGACCAATCCTGGTCTGGCAGAAAAC TTTATGAATAGCTTTATGCAGGGTCTGAGCAGCATGCCTGGTTTT ACCGCAAGCCAGCTGGACAAAATGAGCACCATTGCACAGAGCATG GTTCAGAGCATTCAGAGCCTGGCAGCACAGGGTCGTACCAGTCCG AATGATCTGCAGGCACTGAATATGGCATTTGCAAGCAGCATGGCA GAAATTGCAGCAAGCGAAGAAGGTGGCGGTAGCCTGAGCACCAAA ACCAGCAGCATTGCAAGCGCAATGAGCAATGCATTTCTGCAGACA ACCGGTGTTGTTAATCAGCCGTTTATTAACGAAATTACCCAGCTG GTTAGCATGTTTGCACAGGCAGGTATGAATGATGTTAGCGCAGGT AATAGCGGTGGTGGTGGTAGCGGTGAAAACCTGTATTTTCAGTGC GGCAAACGCAAAAGCTAGGGCGGTGGTGGCAGCATGCAGATCTTC GTGAAGACTCTGACTGGTAAGACCATCACCCTCGAGGTTGAGCCC AGTGACACCATTGAGAATGTCAAGGCAAAGATCCAAGATAAGGAA GGCATCCCTCCTGACCAGCAGAGGCTGATCTTTGCTGGAAAACAG CTGGAAGATGGGCGCACCCTGTCTGACTACAACATCCAGAAAGAG TCCACCCTGCACCTGGTACTCCGTCCGAGAGGTGGACACCATCAT CACCACCATTAA	MSHTTPWTNPGLAENFMNS FMQGLSSMPGFTASQLDKM STIAQSMVQSIQSLAAQGR TSPNDLQALNMAFASSMAE IAASEEGGGSLSTKTSSIA SAMSNAFLQTTGVVNQPFI NEITQLVSMFAQAGMNDVS AGNSGGGGSGENLYFQCGK RKSXGGGGSMQIFVKTLTG KTITLEVEPSDTIENVKAK IQDKEGIPPDQQRLIFAGK QLEDGRTLSDYNIQKESTL HLVLRPRGGHHHHHH

G1mCNPRS[c]	ATGCACCACCACCACCACCATGAGAATCTGTATTTCCAGGGTATG GTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAATAT GGTGATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCAGCA CGTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAAGCGCC AATGATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAATGAT CTGACCGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCCCGT GGTTTTATTGAAGTTCATACCCCGATTTTTGTGAGCAAAAGCGCA CTGGCAAAAATGACCATTACCGAAGATCATCCGCTGTTCAAACAG GTGTTTTGGATTGATGATAAACGTGCACTGCGTCCGATGCATGCA ATGAATGCTTTGAAAGTTATGCGTGAACTGCGCGATCATACCAAA GGTCCGGTTAAAATCTTTGAAATTGGTAGCTGCTTTCGCAAAGAA AGCAAAAGCAGTACCCATCTGGAAGAATTTACCATGCTGAACCTG GCAGAAATGGGTCCTGATGGTGATCCGATGGAACATCTGAAAATG TATATTGGCGATATCATGGATGCCGTTGGTGTTGAATATACCACC AGTCGTGAAGAATCAGATGTTTGGGTTGAAACCCTGGACGTGGAA ATTAATGGCACCGAAGTTGCAAGCGGTTCTGTTGGTCCGCATAAA CTGGATCCGGCACATGATGTGCATGAACCGTGGGCAGGTATTGGT TTTGGTCTGGAACGTCTGCTGATGCTGAAAAATGGTAAAAGCAAT GCACGCAAAACCGGCAAAAGTATTACCTATCTGAATGGCTACAAA CTGGATTAA	MHHHHHHENLYFQGMVVKF TDSQIQHLMEYGDNDWSEA EFEDAAARDKEFSSQFSKL KSANDKGLKDVIANPRNDL TDLENKIREKLAARGFIEV HTPIFVSKSALAKMTITED HPLFKQVFWIDDKRALRPM HAMNALKVMRELRDHTKGP VKIFEIGSCFRKESKSSTH LEEFTMLNLAEMGPDGDPM EHLKMYIGDIMDAVGVEYT TSREESDVWVETLDVEING TEVASGSVGPHKLDPAHDV HEPWAGIGFGLERLLMLKN GKSNARKTGKSITYLNGYK LD

[a] X identifies the positions of the ncAA.

[b] The amino acid sequence of Cys-Ubi 7X is underlined.

[c] Sequence of G1mCNPRS with the N-terminal His$_6$ tag and TEV recognition sequence. The amino acid sequence of G1mCNPRS after TEV cleavage is underlined.

Table S7. List of primers used in the present study.

Primer	DNA sequence
T7 Promotor Forward Primer	TAATACGACTCACTATAGGG
T7 Promotor Reverse Primer	CCCTATAGTGAGTCGTATTA
T7 Terminator Forward Primer	CCGCTGAGCAATAACTAGC
T7 Terminator Reverse Primer	GCTAGTTATTGCTCAGCGG
G1RSlib-frag1 Forward Primer	TCATACGCCGTTATACGTTGT
G1RSlib-frag1 Reverse Primer	TTCATTGCATGCATCGGACG
G1RSlib-frag2 Forward Primer	CGATGCATGCAATGAATNNKNNKAAAGTTATGCGTGAACTGCGC
G1RSlib-frag2 Reverse Primer	CCATCAGGACCCATTTCKVMCAGGNYCAGCATGGTAAATTCTTCCAGATG
G1RSlib-frag3 Forward Primer	AAATGGGTCCTGATGGTGATC
G1RSlib-frag3 Reverse Primer	CGCTTGCAACTTCGGTG
G1RSlib-frag4 Forward Primer	ACCGAAGTTGCAAGCGGTNNKGTTGGTCCGCATAAACTGG
G1RSlib-frag4 Reverse Primer	GACCAAAACCAATACCTGCMNNCGGTTCATGCACATCATGTG
G1RSlib-frag5 Forward Primer	CAGGTATTGGTTTTGGTCTGGAAC
G1RSlib-frag5 Reverse Primer	GCGAAGAAAAGCAGAAAAAACG
G1RSY204F Forward Primer	TCAGATGTTTTTGTTGAAACCCTGGACGTGG
G1RSY204F Reverse Primer	GTTTCAACAAAAACATCTGATTCTTCACGACTG
G1RSY204W Forward Primer	TCAGATGTTTGGGTTGAAACCCTGGACGTGG
G1RSY204W Reverse Primer	GTTTCAACCCAAACATCTGATTCTTCACGACTG

Table S8. Structural data of the single-crystal X-ray structure determined of G1mCNPRS.

G1mCNPRS (G1PylRS mutant - G1mCNP34)	
PDB ID	7R6O
Data collection	
Space group	P 1 2_1 1
Cell dimensions	
a, b, c (Å)	58.9, 81.4, 121.9
α, β, γ (°)	90, 102.1, 90
Resolution *(Å)*	48.06–2.2 (2.279–2.2)*
R_{merge}	0.16 (2.31)
R_{pim}	0.063 (0.93)
I/σI	6.7 (0.9)
$\text{CC}_{1/2}$	0.998 (0.341)
Completeness (%)	99.9 (99.9)
Multiplicity	7.0 (7.1)
Refinement	
Resolution (Å)	46.56–2.2 (2.279–2.2)
No. reflections	57135
$R_{\text{work}}/R_{\text{free}}$	0.2187/ 0.2684 (0.3473/0.3102)
No. atoms	9301
Protein	8660
Ligand/ion	17
Water	624
B-factors (overall)	56.06
Protein	56.38
Ligand/ion	77.34
Water	51.04
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.47

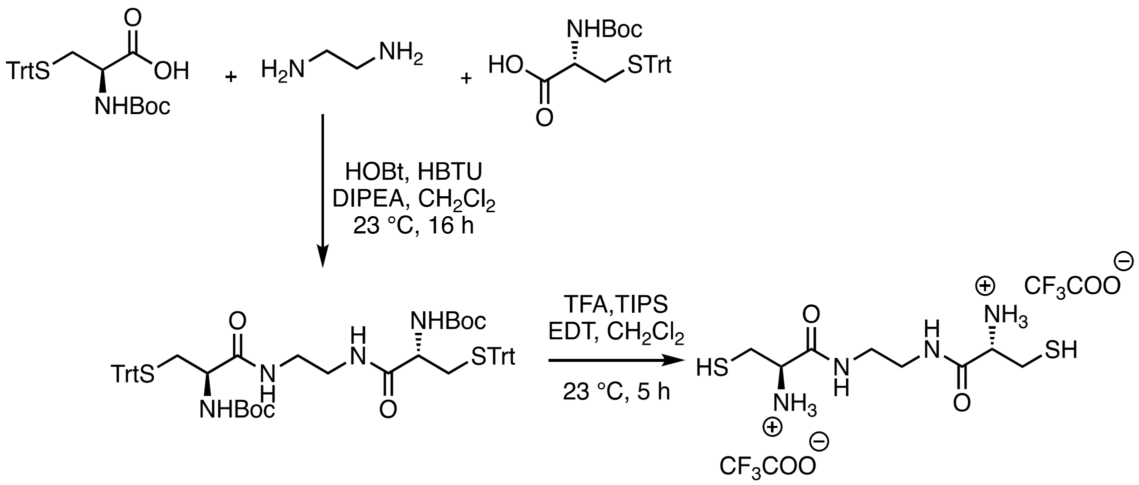
* Statistics for the highest resolution shell are shown in parentheses.

Synthesis of EDDC, and mCNP

Reactions were conducted under a positive pressure of dry nitrogen in oven-dried glassware and at room temperature, unless specified otherwise. Starting materials, solvents, and reagents were purchased from commercial sources and used as such for the reactions. Analytical thin-layer chromatography was conducted with aluminium-backed silica gel 60 F_{254} (0.2 mm) plates supplied by Merck and visualized using UV fluorescence (λ_{max} = 254 nm), or developed using basic $KMnO_4$ solution followed by heating. For filtration column chromatography, Merck Kiesegel 60 silica gel (230 – 400 mesh) was used. Flash chromatography was conducted on a Biotage Isolera One.

^{1}H-NMR spectra were recorded at 400 MHz using a Bruker Avance 400 MHz NMR spectrometer. Residual solvent peaks were used as an internal reference for ^{1}H-NMR spectra (D_2O δ = 4.79). Coupling constants (*J*) are quoted to the nearest 0.1 Hz. The assignment of proton signals was assisted by COSY, HSQC, and HMBC experiments. ^{13}C-NMR spectra were recorded at 100 MHz using a Bruker Avance 400 MHz NMR spectrometer. The following abbreviations are used to denote ^{1}H-NMR multiplicities: s = singlet, d = doublet, t = triplet, m = multiplet. Low-resolution ESI mass spectra were recorded on a ZMD Micromass spectrometer with Waters Alliance 2690 HPLC. High resolution ESI mass spectra were recorded on a Waters LCT Premier time-of-flight (TOF) mass spectrometer.

Synthesis of EDDC (ethylenediamine dicysteine)

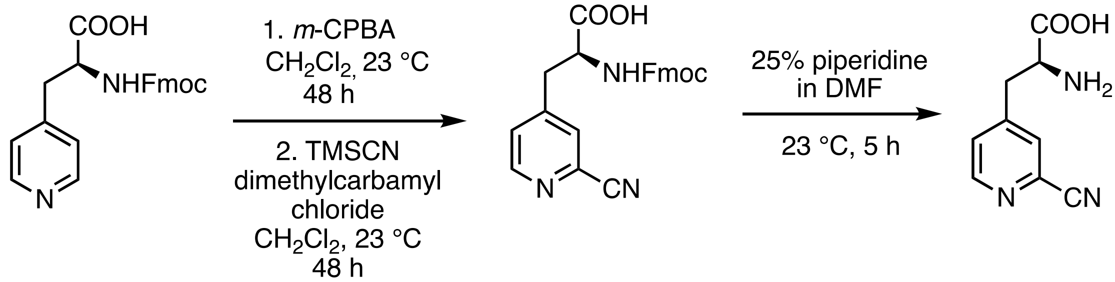


To a solution of Boc-Cys(Trt)-OH (3.86 g, 8.32 mmol, 2.0 mol equiv), HBTU (4.70 g , 12.5 mmol, 3.0 mol equiv), HOBt (562 mg, 4.15 mol equiv, 1.0 mol equiv), and DIPEA (2.20 mL, 12.5 mmol, 3.0 mol equiv) in CH_2Cl_2 (45 mL) was added a solution of ethylenediamine (250 mg, 4.16 mmol, 1.0 mol equiv) in CH_2Cl_2 (5 mL) at 0 °C. The reaction mixture was brought to 25 °C and stirred for 16 h. Reaction completion was confirmed by TLC and mass spectrometric analysis. The reaction mixture was quenched with ice-cold water (100 mL) and layers were separated. The aqueous layer was extracted with CH_2Cl_2 (30 mL). The combined organic layer was washed with brine solution (20 mL), separated, dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The crude product was dissolved in CH_2Cl_2 (20 mL), passed through a plug of silica to remove the polar impurities, and washed with 10% MeOH in CH_2Cl_2 (50 mL). The solvent was evaporated to obtain the protected ethylendiamine dicysteine as a viscous liquid and used as such for the next step.

The crude product from the coupling reaction was dissolved in CH_2Cl_2 (20 mL) and cooled to 0 °C before a solution of TFA: TIPS: EDT: water (10 mL: 0.5 mL: 0.5 mL: 0.5 mL) was added. The reaction mixture was allowed to warm to 23 °C and stirred for 5 h. Reaction completion was confirmed by mass spectrometric analysis. The solvent and excess reagents were removed by evaporation under reduced pressure. The crude waxy liquid was then triturated with hexane (2×5 mL) to remove the non-polar impurities and further trituration with Et_2O (2×5 mL) yielded the TFA salt of the product as a white amorphous solid (1.45 g, 2.93 mmol, 70% over two steps).

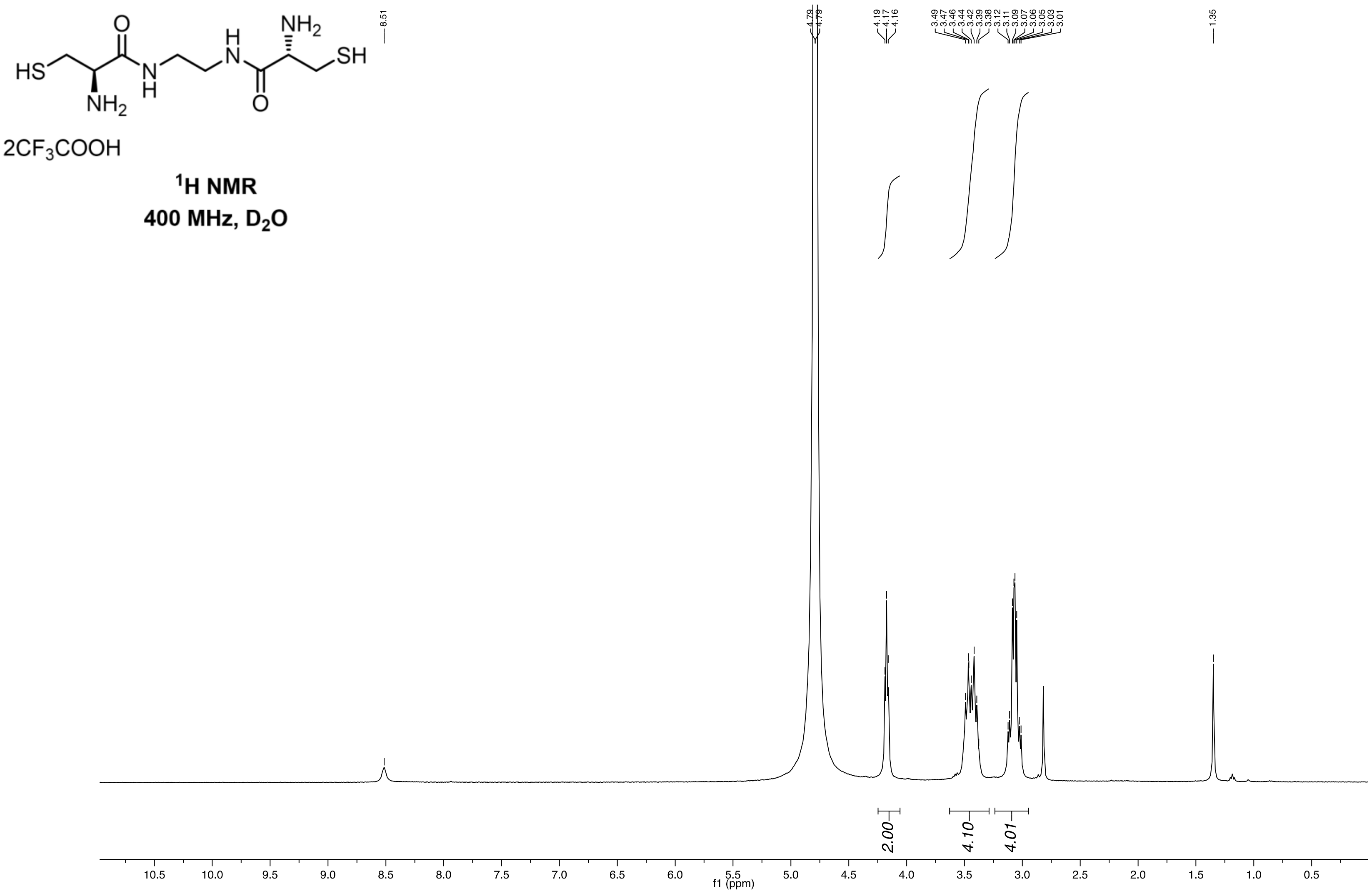
^{1}H NMR (400 MHz, D_2O) δ 8.51 (brs, 2×CONH), 4.17 (t, J = 5.7 Hz, 2H), 3.55 – 3.33 (m, 4H), 3.15 – 2.99 (m, 4H) ppm, 1.35 (s, 2×SH) ppm; **^{13}C NMR** (100 MHz, D_2O) δ 168.1 (2×Cq), 54.5 (2×CH), 38.8 (2×CH_2), 24.8 (2×CH_2) ppm; **LRMS** (ESI+): m/z (%): 267 ([M+H], 85), 533 ([2M+H], 80), 555 ([2M+Na], 100); **HR-MS** (ESI+): m/z calculated for $C_8H_{19}N_4O_2S_2$: 267.0949; found: 267.0952.

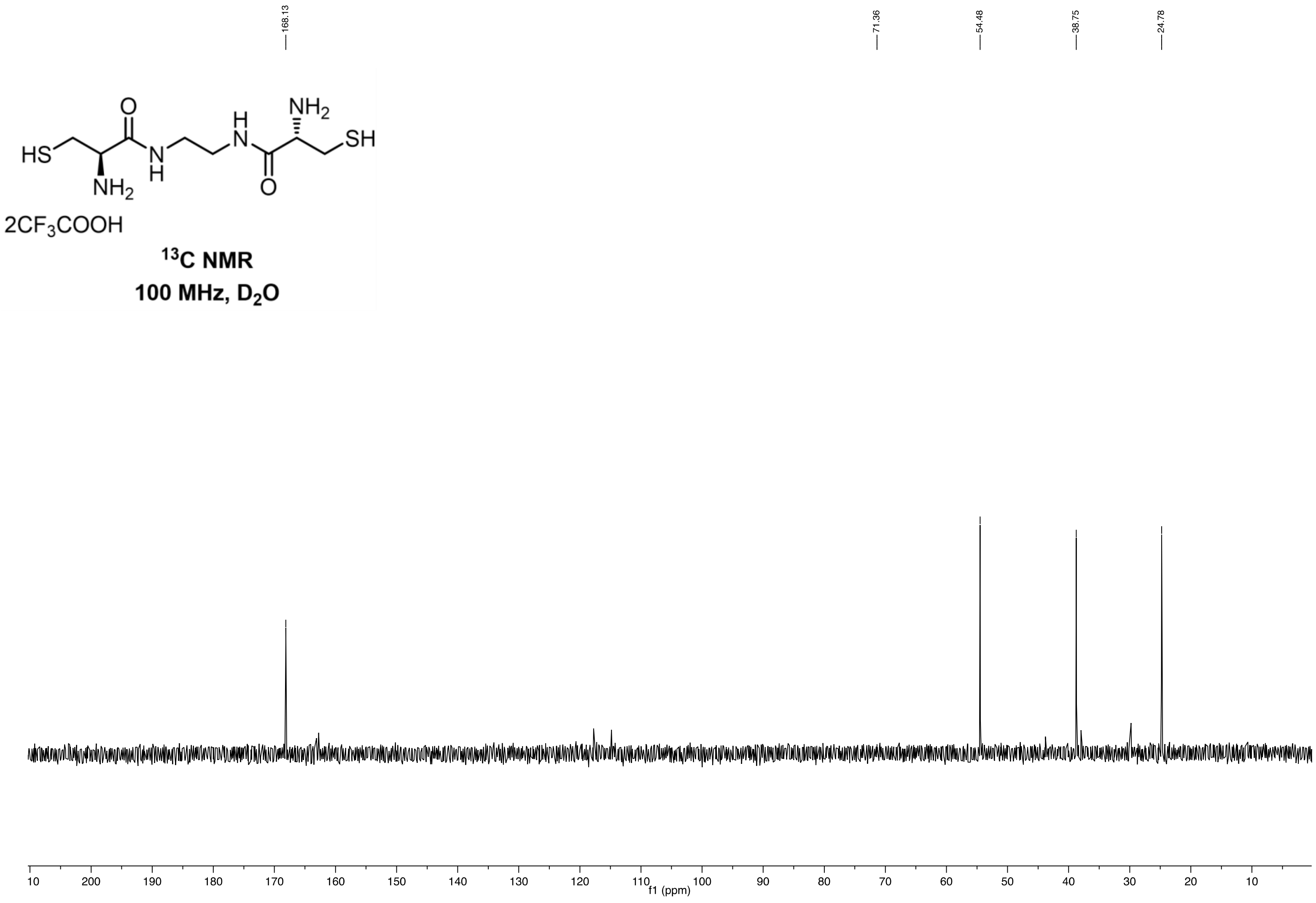
Synthesis of mCNP (L-3-(2-cyano-4-pyridyl)alanine)



Fmoc-L-3-(2-cyano-4-pyridyl)alanine was synthesized according to a reported procedure.[15] Fmoc-L-3-(2-cyano-4-pyridyl)alanine (1.5 g, 3.63 mmol) was solved in 15 mL of 25% piperidine solution in DMF and stirred at 23 °C for 5 h. Reaction completion was confirmed by TLC and mass spectrometric analysis. Excess reagents were removed under reduced pressure and the remaining viscous liquid was stirred with hexane (2×20 mL) to remove the non-polar impurities. Upon trituration with diethyl ether (2 × 15 mL), pure L-3-(2-cyano-4-pyridyl)alanine was obtained as a white solid (550 mg, 2.88 mmol, 79%).

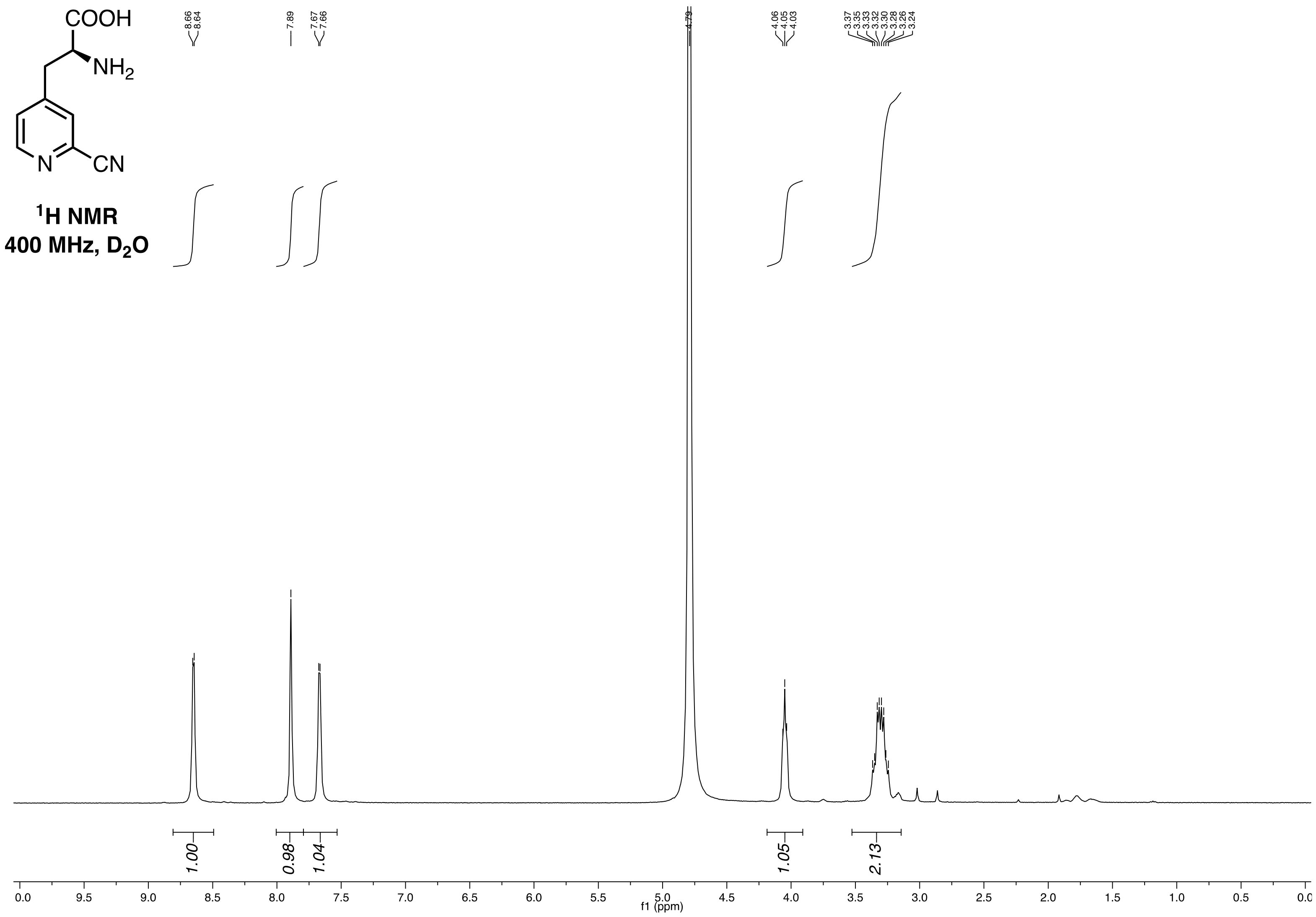
^{1}H NMR (400 MHz, D_2O) *δ* 8.65 (d, *J* = 4.6 Hz, 1H), 7.89 (s, 1H), 7.67 (d, *J* = 3.9 Hz, 1H), 4.05 (t, *J* = 5.9 Hz, 1H), 3.45 – 3.17 (m, 2H) ppm; **^{13}C NMR** (100 MHz, D_2O) *δ* 172.9 (Cq), 150.9 (CH), 147.8 (Cq), 132.3 (Cq), 130.2 (CH), 129.0 (CH), 117.0 (Cq), 54.8 (CH), 35.7 (CH_2) ppm; **LRMS** (ESI–): m/z (%): 190 ([M–H], 100), 381 ([2M–H], 80); **HR-MS** (ESI–): m/z calculated for $C_9H_8N_3O_2$: 190.0617; found: 190.0618.

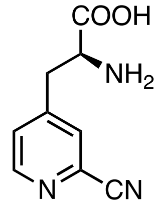

HS
NH2
O
N
H
H
N
O
NH2
SH
2CF3COOH
1H NMR
400 MHz, D2O
8.51
4.79
4.79
4.19
4.17
4.16
3.49
3.47
3.46
3.44
3.42
3.39
3.38
3.12
3.11
3.09
3.07
3.06
3.05
3.03
3.01
1.35
2.00
4.10
4.01
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)



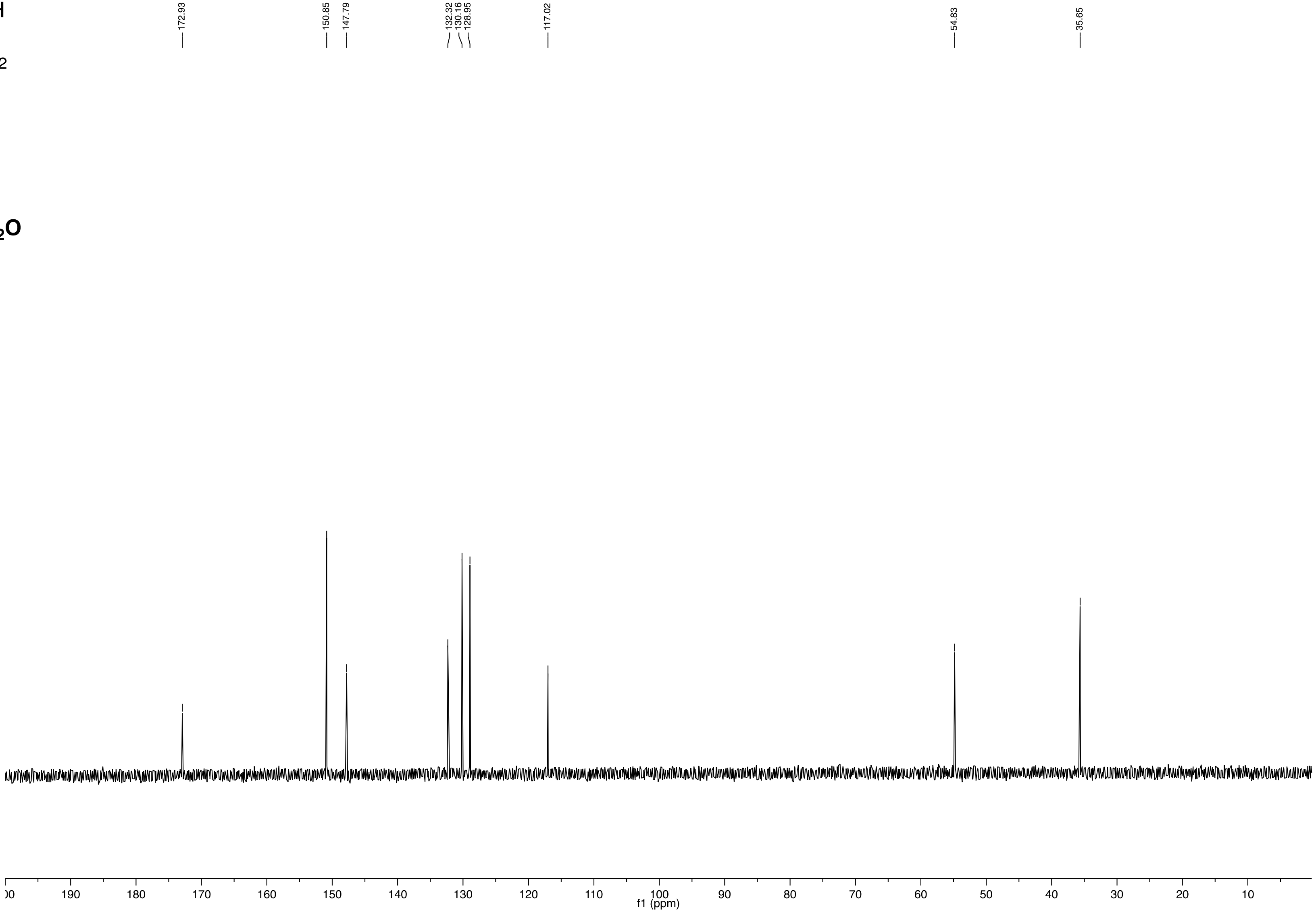
^{13}C NMR

100 MHz, D$_2$O


COOH
NH_2
N
CN
^{1}H NMR
400 MHz, D_2O
8.66
8.64
7.89
7.67
7.66
4.79
4.06
4.05
4.03
3.37
3.35
3.33
3.32
3.30
3.28
3.26
3.24
1.00
0.98
1.04
1.05
2.13
0.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)



^{13}C NMR
400 MHz, D_2O



References

1. Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **2010**, *395*, 361–374.

2. Qianzhu, H.; Welegedara, A. P.; Williamson, H.; McGrath, A. E.; Mahawaththa, M. C.; Dixon, N. E.; Otting, G.; Huber, T. Genetic encoding of *para*-pentafluorosulfanyl phenylalanine: a highly hydrophobic and strongly electronegative group for stable protein interactions. *J. Am. Chem. Soc.* **2020**, *142*, 17277–17281.

3. Abdelkader, E. H.; Qianzhu, H.; Tan, Y. J.; Adams, L. A.; Huber, T.; Otting, G. Genetic Encoding of N^6-(((Trimethylsilyl)methoxy)carbonyl)-L-lysine for NMR studies of protein–protein and protein–ligand interactions. *J. Am. Chem. Soc.* **2021**, *143*, 1133–1143.

4. Mukai, T.; Hoshi, H.; Ohtake, K.; Takahashi, M.; Yamaguchi, A.; Hayashi, A.; Yokoyama, S.; Sakamoto, K. Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* **2015**, *5*, 1–9.

5. Klopp, J.; Winterhalter, A.; Gébleux, R.; Scherer-Becker, D.; Ostermeier, C.; Gossert, A. D. Cost-effective large-scale expression of proteins for NMR studies. *J. Biomol. NMR* **2018**, *71*, 247–262.

6. Aragao, D.; Aishima, J.; Cherukuvada, H.; Clarken, R.; Clift, M.; Cowieson, N. P.; Ericsson, D. J.; Gee, C. L.; Macedo, S.; Mudie, N.; Panjikar, S.; Price, J. R.; Riboldi-Tunnicliffe, A.; Rostan, R.; Williamson, R.; Caradoc-Davies, T. T. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.* **2018**, *25*, 885–891.

7. Kabsch, W. XDS. *ACTA Crystallogr. D* **2010**, *66*, 125–132.

8. Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *ACTA Crystallogr. D* **2013**, *69*, 1204–1214.

9. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

10. Bunkoczi, G.; Read, R. J. Improvement of molecular-replacement models with *Sculptor*. *ACTA Crystallogr. D* **2011**, *67*, 303–312.

11. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of *Coot*. *ACTA Crystallogr. D* **2010**, *66*, 486–501.

12. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards automated crystallographic structure refinement with *phenix.refine*. *ACTA Crystallogr. D* **2012**, *68*, 352–367.

13. Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C., Automated structure solution with the PHENIX suite. In *Structural Proteomics*, Springer: 2008; pp 419–435.

14. Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Söll, D.; Steitz, T. A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11268–11273.

15. Nitsche, C.; Onagi, H.; Quek, J.-P.; Otting, G.; Luo, D.; Huber, T. Biocompatible macrocyclization between cysteine and 2-cyanopyridine generates stable peptide inhibitors. *Org. Lett.* **2019**, *21*, 4709–4712.