nature chemical biology

Article

https://doi.org/10.1038/s41589-025-01999-w

Site-selective protein editing by backbone extension acyl rearrangements

Received: 17 March 2025

Accepted: 15 July 2025

Published online: 21 August 2025

Check for updates

Leah T. Roe[1], Isabel M. Piper[1], Carly K. Schissel[1], Taylor L. Dover[2], Bhavana Shah[3], Noah X. Hamlish[4], Arjun M. Garapaty[1], Shuai Zheng[1], Diondra A. Dilworth[2], Nicole Wong[1], Zhongqi Zhang[3], Abhishek Chatterjee[5], Matthew B. Francis[1] ✉, Scott J. Miller[2] ✉ & Alanna Schepartz[1,4,6,7,8] ✉

Protein and polypeptide heteropolymers containing non-α-backbone monomers are highly desirable as potential materials and therapeutics but many remain difficult or impossible to biosynthesize in cells using traditional genetic code expansion. Here we describe a next-generation approach to such materials that relies instead on proximity-guided intramolecular rearrangements that edit the protein backbone post-translationally. This approach relies on orthogonal aminoacyl-tRNA synthetase enzymes that accept α-hydroxy acid monomers whose side chains contain masked nucleophiles. Introduction of such an α-hydroxy acid into a protein translated in vivo, followed by nucleophile unmasking, sets up a thermodynamically favored intramolecular backbone extension acyl rearrangement (BEAR) reaction that edits the protein to install an extended-backbone monomer. In the examples described here, BEAR reactions are used to generate protein heteropolymers containing a β-backbone, γ-backbone or δ-backbone. This report represents a general strategy to install extended backbones into genetically encoded proteins and peptides expressed in cells.

There is widespread current interest in the cellular biosynthesis of proteins and polypeptides whose backbones differ from those in extant proteins[1]. Even single-atom substitutions, such as the introduction of an ester or thioester in place of an amide, can promote new chemistry, facilitate mechanistic studies and generate materials with emergent properties. The simplest α-peptide backbone modification, beyond a single-atom substitution, is the addition of one or more methylene (CH_2) units to generate β-amino acid, γ-amino acid or δ-amino acid monomers (Fig. 1a). Extended backbones containing β-linkages, γ-linkages or δ-linkages are replete in natural products that possess profound biological activity, such as taxol, andrimid, actinoramide A, zorbamycin and pyloricidin D. Oligomers containing one or more extended-backbone linkages can function as protein–protein interaction inhibitors[2,3], antimicrobial agents[4], receptor agonists and antagonists[5], catalysts[6,7] and other useful tools. Extended-backbone-containing oligomers can assemble into structurally defined protein-like materials that fold cooperatively with enhanced thermal stability[8]. Even a small number of judiciously positioned extended-backbone linkages within a protein add value, such as improved protease resistance[9], enhanced cellular permeability[10,11] and altered recognition when presented by major histocompatibility complex receptors[12,13].

[1]Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Chemistry, Yale University, New Haven, CT, USA. [3]Process Development, Attribute Sciences, Amgen, Inc., Thousand Oaks, CA, USA. [4]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. [5]Department of Chemistry, Boston College, Chestnut Hill, MA, USA. [6]California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA, USA. [7]Chan Zuckerberg Biohub, San Francisco, CA, USA. [8]ARC Institute, Palo Alto, CA, USA. ✉e-mail: mfrancis@berkeley.edu; scott.miller@yale.edu; schepartz@berkeley.edu

https://doi.org/10.1038/s41589-025-01999-w

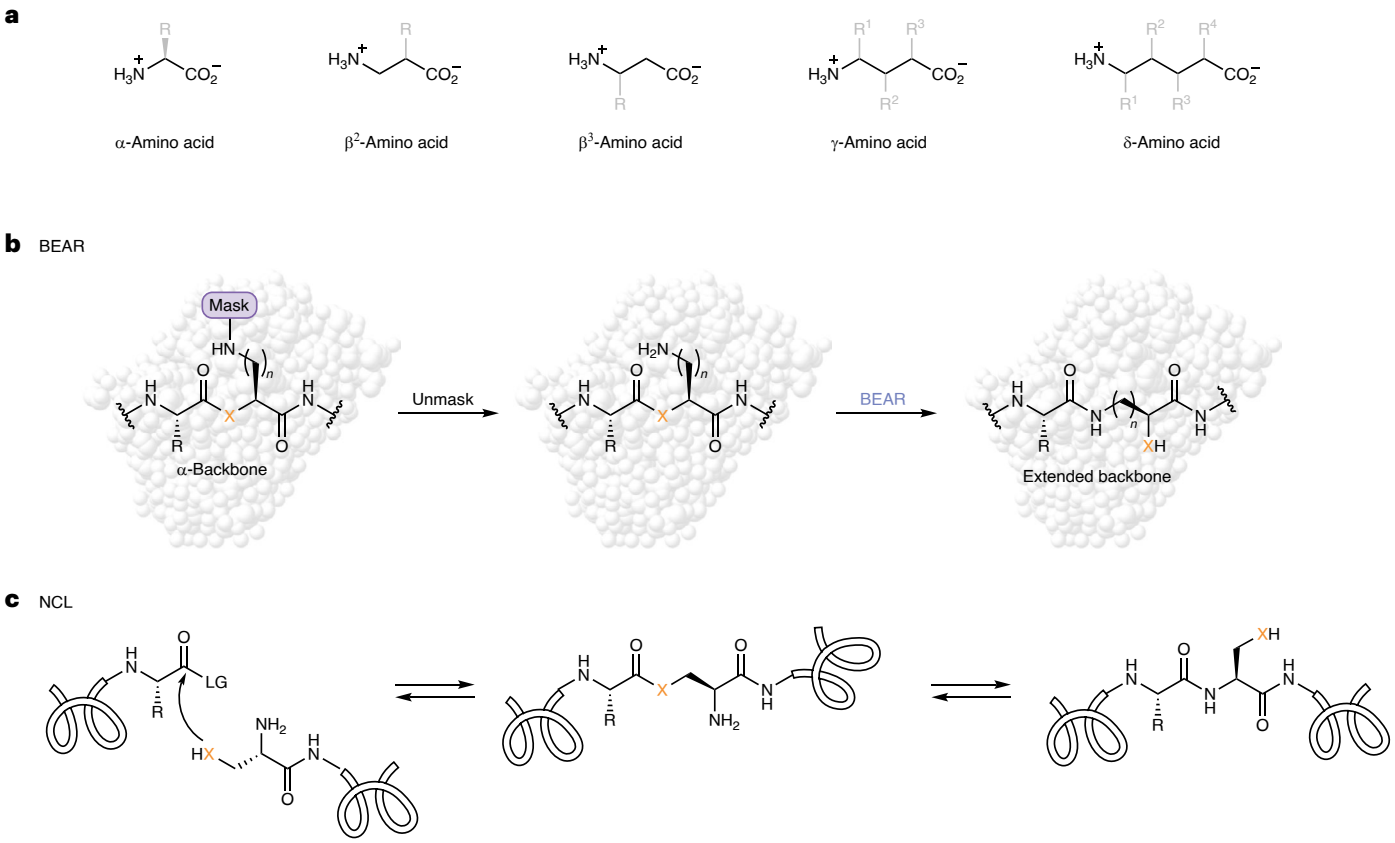


Fig. 1 | BEAR reactions create extended backbones in full-length proteins expressed in vivo. a, Structures of α-amino acid, β^2-amino acid, β^3-amino acid, γ-amino acid and δ-amino acid. **b**, A BEAR reaction results from the intramolecular cyclization of an unmasked nucleophile, in this case, a primary amine, on an N-terminal ester. In the case shown, breakdown of the resultant tetrahedral intermediate generates an extended-backbone linkage embedded within the protein chain. **c**, BEAR reactions are analogous to the second step of the well-known NCL reaction[29]. **d**, Structures of monomers that undergo NCL-like rearrangements within short peptides in vitro. **e**, Structures of BEAR monomers and controls used in this study.

Despite the gain in function provided by extended-backbone linkages, there exist only a few examples[14–17] in which an extended-backbone monomer has been incorporated into a protein or peptide in vivo. In vitro translation (IVT) systems containing wild-type ribosomes support the incorporation of certain extended-backbone monomers into polypeptides[10,18] but IVT methodology is not easily scalable and the incorporation relies on stoichiometric RNA acylation reagents that do not yet function in cells. Indeed, new evidence suggests that complex engineering of EF-Tu, tRNA and/or the ribosome may be needed to efficiently biosynthesize proteins containing even a single extended-backbone linkage in *Escherichia coli*[19–22]. Here, we describe an alternative approach to the programmed cellular synthesis of protein heteropolymers containing extended-backbone β-linkages, γ-linkages or δ-linkages that does not require engineering of EF-Tu, tRNA or the ribosome.

Rather than relying on direct reaction of an extended-backbone monomer within the ribosomal peptidyl-transferase center[14], the strategy reported here relies on a proximity-guided backbone extension acyl rearrangement (BEAR) reaction that edits the protein backbone post-translationally (Fig. 1b). BEAR reactions rely on the fact that α-hydroxy acids are excellent substrates[19,23–25] for enzymes derived from pyrrolysyl-tRNA synthetase (PylRS), the aminoacyl-tRNA synthetase that acylates tRNA with pyrrolysine to introduce this amino acid into proteins in certain archaea and methanogenic bacteria. In *E. coli*, PylRS behaves as an orthogonal enzyme. PylRS variants accept α-amino acids with a variety of side chains[26], including those carrying protected nucleophiles[27,28]. We envisioned that, if an α-hydroxy or α-thiol substrate for such a PylRS variant carried a side chain bearing a masked amine nucleophile, that nucleophile could be unmasked post-translationally to promote a BEAR reaction that is analogous to the second step of the native chemical ligation (NCL) reaction[29] (Fig. 1c). The result of a BEAR reaction would be an internal extended-backbone linkage.

Related acyl rearrangements within short peptides are known[30–32]. It was reported in 2004 that a short peptide containing isocysteine (isoC) rearranged in vitro through an NCL-like reaction to introduce a β^2-linkage[33]. More recently, analogous rearrangements involving isoserine (isoS) and the elongated analogs γ-NH_2-OH and δ-NH_2-OH

https://doi.org/10.1038/s41589-025-01999-w

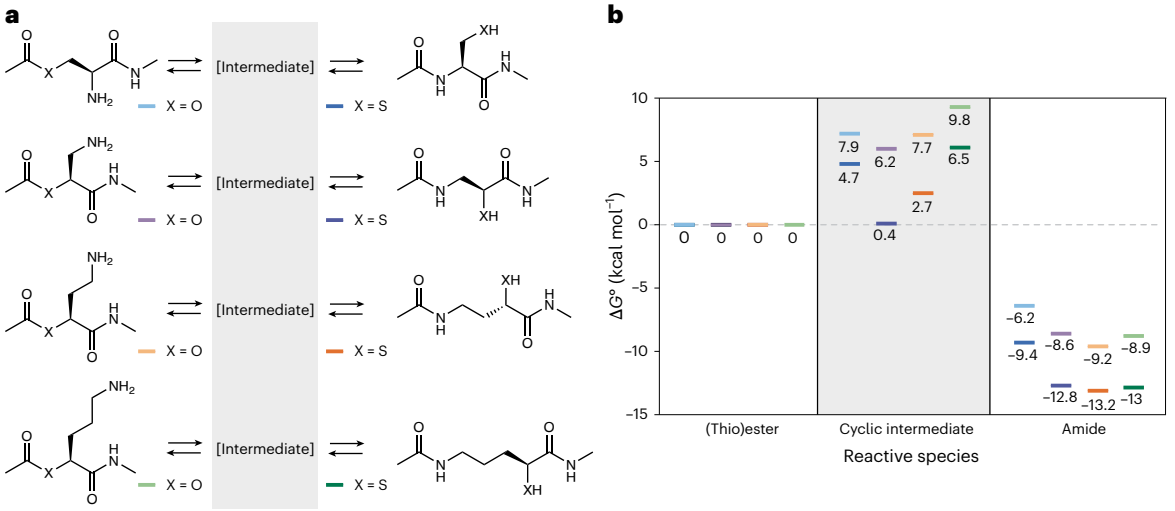


Fig. 2 | Computational analysis provides evidence that BEAR reactions to establish β-linkages, γ-linkages and δ-linkages are thermodynamically favorable. a, Four model systems chosen to assess the thermodynamics of intramolecular cyclization reactions corresponding to the second step of the canonical NCL reaction (top) and BEAR reactions that result in β-linkages, γ-linkages and δ-linkages. All reactions were evaluated using simplified *N*-Me amide derivatives. The species labeled 'intermediate' represents the cyclic intermediate formed during each acyl-transfer step. **b**, Relative thermodynamics (Δ*G*°) of the NCL and BEAR reactions shown in **a** established using DFT (ωB97M-V/def2-TZVPPD, with an SMD solvent model). Each reaction can proceed through two diastereomeric intermediates; only the lowest energy species is listed. All species were assumed to be uncharged (that is, not protonated on basic sites and not deprotonated on acidic sites) for this analysis.

(Fig. 1d) were reported in short peptides prepared during small-scale IVT reactions[34]. Inspired by these findings, we report that proteins containing either the photomasked isoS derivative (pm-isoS **1**) or the chemically masked α-hydroxy acids γ-N_3-OH **2** or δ-N_3-OH **3** (Fig. 1e) can be unmasked post-translationally, either in cells or after protein purification. Unmasking reveals a β-amine, γ-amine or δ-amine nucleophile and triggers a spontaneous, intramolecular BEAR reaction that directly installs a β^2-amino acid, γ-amino acid or δ-amino acid linkage, respectively, at internal protein positions. Notably, installation of a γ-amino acid backbone proceeds with exceptional fidelity and yield. This work represents a generalizable strategy to install extended backbones into genetically encoded proteins and peptides and relies only on established activities of known orthogonal enzymes and the wild-type *E. coli* ribosome. The thus-generated hetero-oligomers greatly expand the diversity of protein-like polypeptides that may be synthesized using the ribosome as a catalyst.

Results

BEAR reactions are thermodynamically favorable

We began by performing computational studies to confirm that BEAR reactions to establish β-linkages, γ-linkages and δ-linkages would be thermodynamically favorable. A previous study made use of density functional theory (DFT) to examine the energetics and mechanistic pathways of the canonical cysteine-based NCL reaction but this study did not include the analogs considered herein[35]. We used DFT to evaluate the relative energetics of acyl rearrangements involving the natural α-amino acids cysteine and serine, their isomers isoC and isoS and analogs capable of installing γ-backbones and δ-backbones in the context of a simplified substrate (Fig. 2a). For each proposed reaction species, molecular mechanics methods using an OPLS4 force field were first used to generate and evaluate 10,000 starting conformers. The geometries of all conformers with energies within 5 kcal mol^{-1} of the global minimum were then reoptimized using DFT methods (r^2SCAN-3c/mTZVPP with a CPCM water model)[36]. The vibrational frequencies of all conformers within 2 kcal mol^{-1} of the new global minimum were determined at the r^2SCAN-3c/mTZVPP level with a CPCM water model. This analysis allowed estimation of the zero-point energy and internal entropy parameters. Finally, single-point electronic energies were determined with greater accuracy using ωB97M-V/def2-TZVPPD (refs. 37,38). These final calculations were performed with a SMD solvent model to represent aqueous solvation[39].

These calculations imply that all three proposed BEAR reactions are thermodynamically favorable. They also revealed expected and unexpected differences with respect to intermediate ring size and heteroatom location and identity. BEAR reactions that generate β-linkages, γ-linkages and δ-linkages are more favorable energetically (by 2–3 kcal mol^{-1}) than the canonical NCL rearrangements of cysteine and serine in terms of both enthalpy and free energy (Fig. 2b and Supplementary Fig. 1). This improvement may be because of increased hydrogen bonding from the nascent α-hydroxy or α-thiol group in the BEAR products and relief of steric congestion. BEAR reactions that generate β-linkages, γ-linkages and δ-linkages are also more disfavored entropically than the canonical NCL rearrangements of cysteine and serine. This difference is especially true for the intermediate species. The differences result in an especially high free energy for the seven-membered ring intermediate that results in a δ-linkage, suggesting that these reactions might proceed more slowly. Nevertheless, because of their high overall energetic favorability, we sought to evaluate the potential of all three BEAR reactions within the context of a translated product.

The validity of DFT calculations was also verified by evaluating the intramolecular cyclization of *O*-to-*N* and *S*-to-*N* BEAR reactions to generate β^2-backbone products in short peptides in vitro under conditions that mimic those found in cells. These studies reveal that the *O*-to-*N* BEAR reaction to establish a β^2-peptide was spontaneous under biomimetic conditions, whereas the analogous *S*-to-*N* BEAR reaction of the thioester was outcompeted by hydrolysis (Supplementary Figs. 2–4). These results suggested that *O*-to-*N* BEAR reactions of oxo-esters might be more straightforward to implement in a cellular context than the *S*-to-*N* reaction of thioesters.

BEAR reactions generate β^2-linkages in vivo

Given the computational data and experimental feasibility of model BEAR reactions to establish β^2-linkages, we next asked whether these reactions could occur within full-length proteins expressed in cells, a more challenging environment than a short, unstructured peptide.

We chose initially to photomask the β^2-amine nucleophile, as photounmasking is compatible with cells and pm-diaminopropionic acid (pm-DAP, **4**) is a known substrate for the PylRS variant DAPRS[27]. We hypothesized that the analogous α-hydroxy acid, pm-isoS (**1**), and α-thio acid, pm-isoC (±)-**5**, would also act as DAPRS substrates and facilitate BEAR reactions in cells or after protein purification. We synthesized pm-isoS **1**, pm-isoC (±)-**5** and the previously reported α-amino acid, pm-DAP **4** (Supplementary Fig. 5). As an initial expression test, we first determined whether pm-isoS **1** or pm-isoC (±)-**5** could be introduced in place of N150 in superfolder green fluorescent protein (sfGFP), a position previously shown to incorporate pm-DAP **4** when encoded by TAG. The overall level of sfGFP fluorescence was, at best, only 1.4-fold over background (Supplementary Fig. 6). We concluded that a more sensitive reporter protein was needed to optimize the incorporation of pm-isoS **1** before evaluating how to best implement a BEAR reaction and establish the desired β^2-linkage. A purified sample of sfGFP containing pm-DAP **4** was used to optimize conditions needed for photounmasking (Supplementary Fig. 7).

We selected nanoluciferase (NanoLuc)[40] as a more sensitive reporter, as its readout (luminescence) is generated catalytically. We prepared a series of reporter plasmids with NanoLuc in place of sfGFP and TAG codons as substitutions or insertions at six different positions. These positions were located either near the NanoLuc N terminus (G2-TAG and G2-TAG-V3)[41], within a loop (G159-TAG-V160 and G103-TAG-V104)[42] or within a β-sheet (T130-TAG and G131-TAG). To validate the plasmids, BL21(DE3) *E. coli* cells were cotransformed separately with each NanoLuc reporter plasmid and pEVOL-PylRS[19], which encodes the *Methanomethylophilus alvus* PylRS/$tRNA^{Pyl}_{CUA}$ pair. In each case, the cells were grown for 18 h in the presence of either 1 mM BocK or α-OH-BocK (positive controls) or lysine or no substrate (negative controls). After 18 h, luminescence was determined using the Nano-Glo luciferase assay system (Promega). These experiments clearly identified the reporter encoding G159-TAG-V160 NanoLuc as possessing the largest dynamic range; the signal generated for growths containing α-OH-BocK was more than three orders of magnitude higher than growths to which no substrate had been added. All other constructs showed a dynamic range of two orders of magnitude or less (Supplementary Fig. 8).

With an optimized detection system in hand, we next probed for in vivo incorporation of pm-isoS **1** or pm-isoC (±)-**5** into NanoLuc G159-TAG-V160. We envisioned that once incorporated, irradiation would unmask a nucleophilic amine to promote the same intramolecular BEAR reaction observed in the model system to generate a β^2-linkage (Fig. 3a). We transformed DH10B *E. coli* with pET15a-NanoLuc(G159-TAG-V160) and pBK-DAPRS encoding DAPRS/$tRNA^{Pyl}_{CUA}$ and grew the cells in the presence of 0.1 to 10 mM of pm-isoS **1** or pm-isoC (±)-**5**. Control cultures contained either 1 mM pm-DAP **4** (positive control) or 1 mM Lys, BocK or no substrate (negative controls). The luminescence intensity of each growth was measured after 18 h as described above (Fig. 3b). Growths supplemented with pm-isoC (±)-**5** showed no evidence of incorporation, with no NanoLuc signal above background at any substrate concentration. However, growths supplemented with pm-isoS **1** led to robust concentration-dependent NanoLuc expression, with up to 5.2-fold enhanced signal above background at a concentration of 2 mM pm-isoS **1** (Fig. 3b). Liquid chromatography (LC)–mass spectrometry (MS) analysis of intact NanoLuc G159-TAG-V160 (0.8 mg l^{-1}) purified from a preparative-scale growth (Fig. 3c) showed a mass corresponding to the introduction of pm-isoS **1** at a single position (Fig. 3d).

We subjected a purified sample of NanoLuc G159-TAG-V160 containing pm-isoS **1** to optimized unmasking conditions established using sfGFP containing pm-DAP **4** and analyzed the products using intact protein LC–MS (Supplementary Fig. 9) Although this analysis provided evidence that pm-isoS **1** was introduced into NanoLuc G159-TAG-V160 and unmasked, intact protein LC–MS alone could not confirm whether the BEAR reaction had occurred, as it is fundamentally an isomerization. To achieve this confirmation, purified NanoLuc G159-TAG-V160 samples from growths supplemented with pm-DAP **4** or pm-isoS **1** were unmasked and evaluated by high-resolution tryptic peptide mapping (Fig. 3e). A total of seven samples were analyzed that differed in supplement (pm-DAP **4** versus pm-isoS **1**) and whether deprotection was induced in whole cells or after protein purification or both.

Trypsin cleaves NanoLuc to generate a high-recovery fragment containing residues V155 through R165 with an additional residue (encoded by TAG) after G159 (V155–R165). When isolated from samples grown in the presence of pm-DAP **4**, the sequence of the V155–R165 fragment was consistent with TAG-encoded incorporation of pm-DAP **4** (44.6% fidelity) along with 11.0% of the amine product (DAP) resulting from spontaneous deprotection (Fig. 3e, sample A). Tryptic digestion of the NanoLuc G159-TAG-V160 sample after purification and irradiation resulted in 57.5% of the amine product, consistent with quantitative deprotection (Fig. 3e, sample B). Analysis of the V155–R165 product generated from NanoLuc samples A and B indicated that both were contaminated with >40% glutamine and other residues (Supplementary Fig. 10).

NanoLuc G159-TAG-V160 isolated from growths supplemented with pm-isoS **1** generated analogous tryptic fragments. Tryptic mapping of sample C revealed that 2.1% of the isolated NanoLuc G159-TAG-V160 contained pm-isoS **1** between G159 and V160 while 1.9% contained the mass of isoS, again indicating some unmasking in ambient light. Tryptic mapping of sample D, however, revealed the isolated NanoLuc G159-TAG-V160 contained 0.0% pm-isoS **1** and 9.5% of a product whose mass corresponded to isoS. We hypothesize that some material in sample C was lost to hydrolysis before peptide mapping, resulting in an artificially decreased percentage of pm-isoS **1** (Fig. 3e and Supplementary Fig. 10).

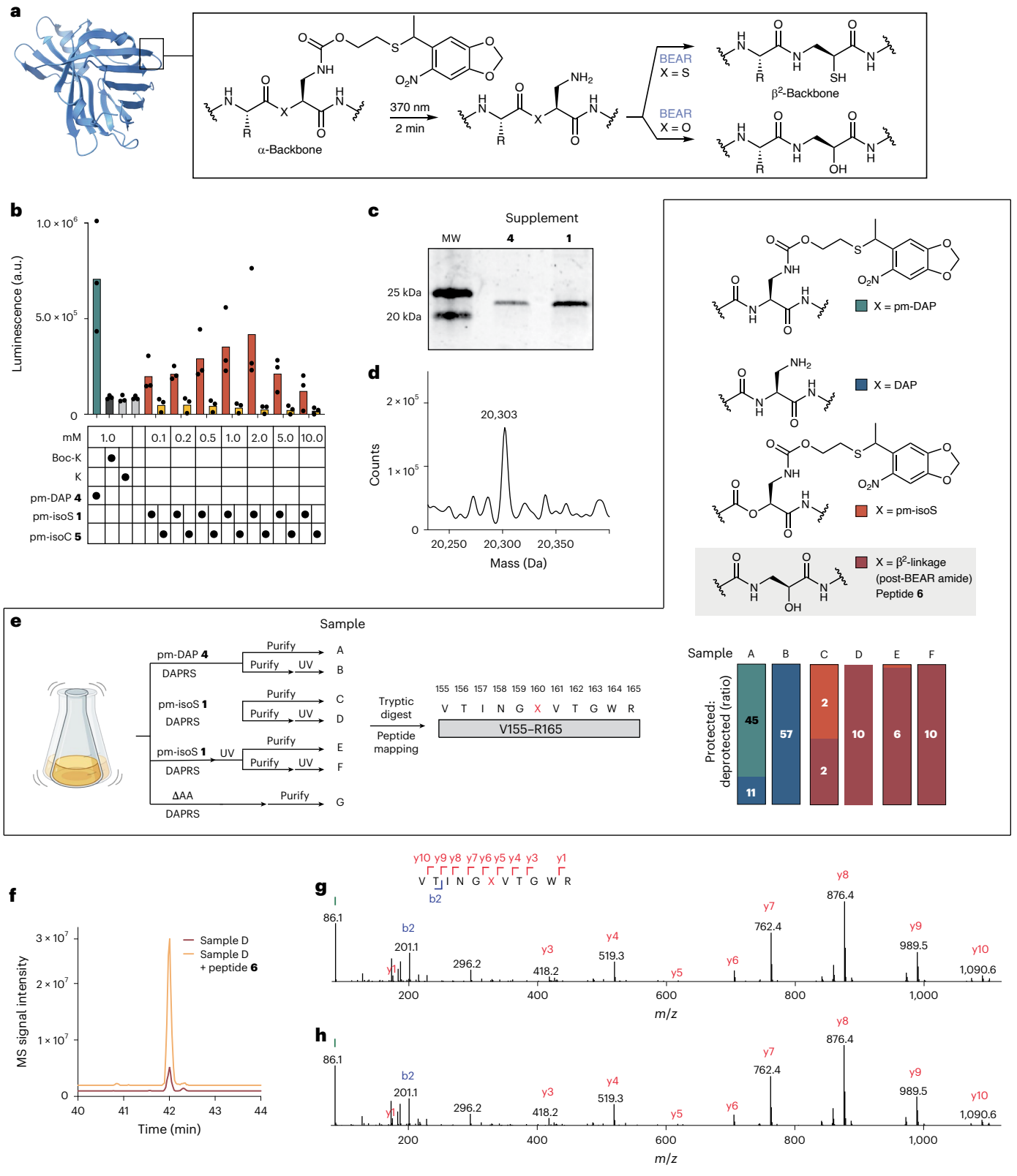
To unequivocally establish the identity of the product generated after photounmasking of NanoLuc G159-TAG-V160 isolated from growths supplemented with pm-isoS **1**, we synthesized an authentic

Fig. 3 | BEAR reactions generate a β^2-linkage in NanoLuc G159-TAG-V160 in vivo. a, Scheme illustrating the photounmasking of pm-isoS **1** and pm-isoC (±)-**5** and their potential BEAR reaction to generate a β^2-linkage within NanoLuc G159-TAG-V160. Cellular incorporation of pm-isoS **1** or pm-isoC (±)-**5** generates a protein containing an internal ester or thioester but retains a natural α-peptide-like backbone. Photounmasking generates an intermediate capable of a BEAR reaction to generate a β^2-linkage (PDB 7SNS). **b**, Luminescence of growths of DH10B *E. coli* expressing DAPRS and NanoLuc G159-TAG-V160 at the 18-h time point as a function of the concentration of pm-DAP **4**, pm-isoS **1** or pm-isoC (±)-**5** and BocK, lysine or the absence of any added substrate. Graphed is the mean of three biological replicates. **c**, Coomassie-stained SDS–PAGE gel showing NanoLuc G159-TAG-V160 affinity purified from growths described in **a** and supplemented with 1 mM pm-DAP **4** or 2 mM pm-isoS **1**. The experiment was performed in three biological replicates. MW, molecular weight. **d**, Deconvoluted intact mass following LC–MS analysis of the NanoLuc G159-TAG-V160 affinity purified from growths described in **a** and supplemented with 1 mM pm-isoS **1**. The mass expected for NanoLuc G159-TAG-V160 containing pm-isoS **1** at a single position is 20,301 Da; the observed mass is 20,303 Da. **e**, Seven samples (samples A–G) of NanoLuc G159-TAG-V160 were isolated from DH10B *E. coli* growths expressing DAPRS and supplemented with pm-DAP **4** or pm-isoS **1** and photounmasked (or not) as indicated. All samples were analyzed by high-resolution tryptic mapping. Analysis of the V155–R165 tryptic product revealed partial photounmasking from ambient light before 370-nm irradiation and full photounmasking upon direct exposure to 370-nm light. UV, ultraviolet. **f**, Trace illustrating the coelution of the V155–R165 tryptic fragment generated from sample D alongside authentic standard peptide **6**. The small later-eluting peak contains serine (Supplementary Fig. 11). **g**,**h**, The MS/MS profile of the V155–R165 tryptic fragment generated from sample D (**g**) is identical to that of authentic peptide **6** (**h**), providing conclusive evidence for a successful BEAR reaction. Note that, for peptide mapping, V160 in the native protein is labeled as V161 and X is at position 160.

https://doi.org/10.1038/s41589-025-01999-w

standard, a peptide whose sequence corresponds to the V155–R165 tryptic product containing the anticipated β^2-linkage between G159 and V160 (peptide **6**). A synthetic peptide containing serine between G159 and V160 (peptide **7**) was also synthesized. When coinjected on LC–MS/MS, peptide **6** coeluted with the V155–R165 tryptic product produced when NanoLuc was isolated from a growth supplemented with pm-isoS **1** and deprotected after purification (Fig. 3f) and the two materials generated identical MS/MS spectra (Fig. 3g,h). This NanoLuc tryptic product did not coelute with the synthetic peptide containing Ser between G159 and V160 (peptide **7**) (Supplementary Fig. 11). Together, these experiments provide conclusive evidence that a protein containing pm-isoS **1** can be unmasked quantitatively and subsequently undergo a spontaneous intramolecular BEAR reaction to install a β^2-linkage at an internal position within NanoLuc.

Analysis of the remaining samples provides insight into an optimal BEAR workflow for establishing β^2-linkages within intact proteins.



The V155–R165 tryptic fragment derived from sample E, where cells were irradiated before lysis and NanoLuc G159-TAG-V160 purification, contained 0.2% of pm-isoS **1** and 5.8% of the BEAR product with the anticipated β^2-linkage between G159 and V160, suggesting that higher levels of the rearranged BEAR product are achieved with postpurification unmasking. Consistent with this notion, sample F, which was irradiated before and after NanoLuc G159-TAG-V160 purification, contained 0.0% of pm-isoS **1** and 9.4% of the β^2-amino acid BEAR product. We note that all NanoLuc samples expressed from growths containing pm-isoS **1** contained notable levels of alanine (~33%) and glutamine (~22%) in place of pm-isoS **1** (Supplementary Figs. 10 and 12). No evidence of pm-DAP **4**, pm-isoS **1** or their deprotected products was found in negative control Sample G (Supplementary Fig. 10). Efforts to increase the fraction of isolated NanoLuc G159-TAG-V160 containing pm-isoS **1** through modification of DAPRS to contain three substitutions previously identified as beneficial for α-hydroxy acid selectivity in PylRS[25] were not successful (Supplementary Figs. 10 and 13). Subsequently, we decided to move to an alternative masking group and higher fidelity orthogonal aaRS–tRNA pair system to achieve purer samples of backbone-modified proteins.

Reductive BEAR reactions generate γ-linkages in high yield

Proteins containing γ-amino acid linkages have not been produced in cells or in vitro in isolable quantities. Indeed, tRNAs acylated with γ-amino acids are prone to degradation through intramolecular cyclization reactions[43], complicating their direct incorporation by wild-type or remodeled ribosomes. As DFT calculations suggested that BEAR reactions to generate γ-linkages are energetically favorable, we examined the incorporation of azide-containing γ-N_3-OH **2**, which we envisioned could be unmasked upon reduction (Fig. 4a). The α-amino analog of γ-N_3-OH **2** is a substrate for a *Methanosarcina barkeri* PylRS variant containing two active site substitutions, PylRS-N311M;C313W (ref. 28). We transformed C321.ΔA.exp(DE3) and BL21(DE3) *E. coli* with plasmids encoding SUMO–GFP102TAG and PylRS-N311M;C313W and grew the cells in the presence of increasing concentrations of γ-N_3-OH **2**. Although only modest GFP fluorescence was detected in growths of BL21(DE3) cells after 24 h, growths of C321.ΔA.exp(DE3) cells showed substantial GFP fluorescence over background when supplemented with 5–15 mM γ-N_3-OH **2** (Fig. 4b and Supplementary Fig. 14). A preparative-scale growth of C321.ΔA.exp(DE3) *E. coli* expressing PylRS-N311M;C313W and SUMO–GFP102TAG and supplemented with 10 mM γ-N_3-OH **2** generated 10.2 mg l^{-1} of a protein whose mass corresponds to the introduction of γ-N_3-OH **2** at a single position (Fig. 4c,d and Supplementary Fig. 15).

Next we sought to establish conditions to reductively unmask SUMO–GFP102TAG containing γ-N_3-OH **2** and promote an intramolecular BEAR reaction to install a γ-peptide linkage. Peptide mapping was used to evaluate the fidelity of γ-N_3-OH **2** incorporation and the yield and products generated upon reduction. Two samples of SUMO–GFP102TAG were isolated from C321.ΔA.exp(DE3) *E. coli* expressing PylRS-N311M;C313W supplemented with 10 mM γ-N_3-OH **2**. After protein purification, 50 mM of the reducing agent tris(2-carboxyethyl) phosphine (TCEP) was added or withheld and the reaction was incubated for 16 h at room temperature in a pH 6.8 buffer. The two SUMO–GFP102TAG samples were then digested with GluC and the products analyzed using high-resolution peptide mapping.

GluC cleaves SUMO–GFP102TAG to generate a high-recovery fragment containing residues Q97 through E106 with a TAG codon in place of S102 (Q97–E106) (Fig. 4e and Supplementary Fig. 16). To determine whether the Q97–E106 fragment generated in the presence of TCEP contained the anticipated γ-peptide linkage, we synthesized an authentic standard whose sequence corresponds to the Q97–E106 GluC product containing the anticipated γ-peptide linkage at position 102 (peptide **8**). We also synthesized an analogous peptide containing threonine at position 102 (peptide **9**). SUMO–GFP102TAG samples that were not treated with TCEP generated a Q97–E106 GluC product containing more than 91% γ-N_3-OH **2** at position 102. In contrast, the Q97–E106 GluC products generated after TCEP treatment contained virtually undetectable amounts of γ-N_3-OH **2** at position 102. Instead, they consisted of a 1:2 mixture of two products. The major product corresponded to conversion of the azide side chain at position 102 to an alcohol, a known side product of phosphine-mediated azide reductions[34] (Supplementary Fig. 16a–c). The minor product, whose mass corresponds to either reduction of the side-chain azide to an amine (pre-BEAR) or the anticipated post-BEAR amide, coeluted with the authentic standard containing the anticipated γ-peptide linkage at position 102 (peptide **8**) (Fig. 4f). Moreover, the minor Q97–E106 GluC product and peptide **8** generated identical MS/MS spectra (Fig. 4g,h). Neither of the Q97–E106 GluC products generated after TCEP treatment coeluted with a synthetic peptide containing threonine at position 102 (Supplementary Fig. 17). These experiments provide conclusive evidence that a protein containing γ-N_3-OH **2** can be reduced and undergo a spontaneous intramolecular BEAR reaction to install a γ-peptide linkage at an internal position within SUMO–GFP102TAG.

We next sought conditions to minimize production of the alcohol side product and maximize conversion into the desired BEAR product containing a γ-peptide linkage. We treated samples of SUMO–GFP102TAG containing a single residue of γ-N_3-OH **2** with TCEP, tris(3-hydroxypropyl)phosphine (THPP) or 1,3,5-triaza-7-phosphaadamantane (PTA) under various conditions of time, pH and temperature (Supplementary Fig. 16c). In each case, the products were characterized by monitoring the mass of the Q97–E106 GluC digestion product. Although samples reduced with TCEP or THPP contained substantial alcohol side product at every pH other than 10, samples reduced with PTA contained virtually no alcohol side product under any condition tested. The conditions resulting in the highest BEAR product yield—with no detectable alcohol side product—featured 50 mM PTA at pH 8.5 with incubation a 0 °C for 2 h then 42 °C for 2 h (Fig. 4e). PTA is, thus, the superior reducing agent for cleanly unmasking a side-chain azide into a nucleophilic amine and promoting an intramolecular BEAR rearrangement.

We noted that, when a sample of SUMO–GFP102TAG containing a single residue of γ-N_3-OH **2** was incubated at pH 8.5, 9 and 10 without a reducing agent and analyzed after GluC digestion, the Q97–E106 product appears to have a higher level of natural α-amino acids at position 102. Our interpretation is that, at pH 8.5 or above, ester hydrolysis competes with the intramolecular BEAR reaction, leading to an artificially inflated percentage of background misincorporation in the full-length protein product. The presence of a reducing agent such as TCEP or PTA protects against hydrolysis, as the BEAR reaction replaces the unstable ester with a stable amide bond. Therefore, in the presence of a reducing agent, we observe less background misincorporation.

In an effort to reduce hydrolysis while maintaining a quantitative BEAR reaction, we conducted a final round of optimization. Samples of SUMO–GFP102TAG containing a single residue of γ-N_3-OH **2** were treated with 50 mM PTA at pH 6.8, 7.2, 7.9, 8.2 and 8.5 at 0 °C for 2 h then 42 °C for 2 h and then mapped using GluC as described above (Supplementary Fig. 16d). At pH 6.8, we observe minimal hydrolysis (glutamine: 2.3%) but low BEAR conversion (34.3%). At pH 8.5, we observe near-quantitative BEAR conversion (~74%) but notable hydrolysis (glutamine: ~13%). The optimal pH was pH 8.2, which led to a quantitative BEAR reaction with >90% fidelity and little or no hydrolysis (glutamine: ~6%).

BEAR reactions generate proteins with an internal δ-linkage

We next sought to introduce δ-N_3-OH **3** into SUMO–GFP102TAG in cells, which upon reductive unmasking and subsequent BEAR reaction would generate a protein containing a δ-peptide linkage (Fig. 5a). Cyclization in this case proceeds through a seven-membered ring intermediate whose formation is less favorable enthalpically and entropically than

https://doi.org/10.1038/s41589-025-01999-w

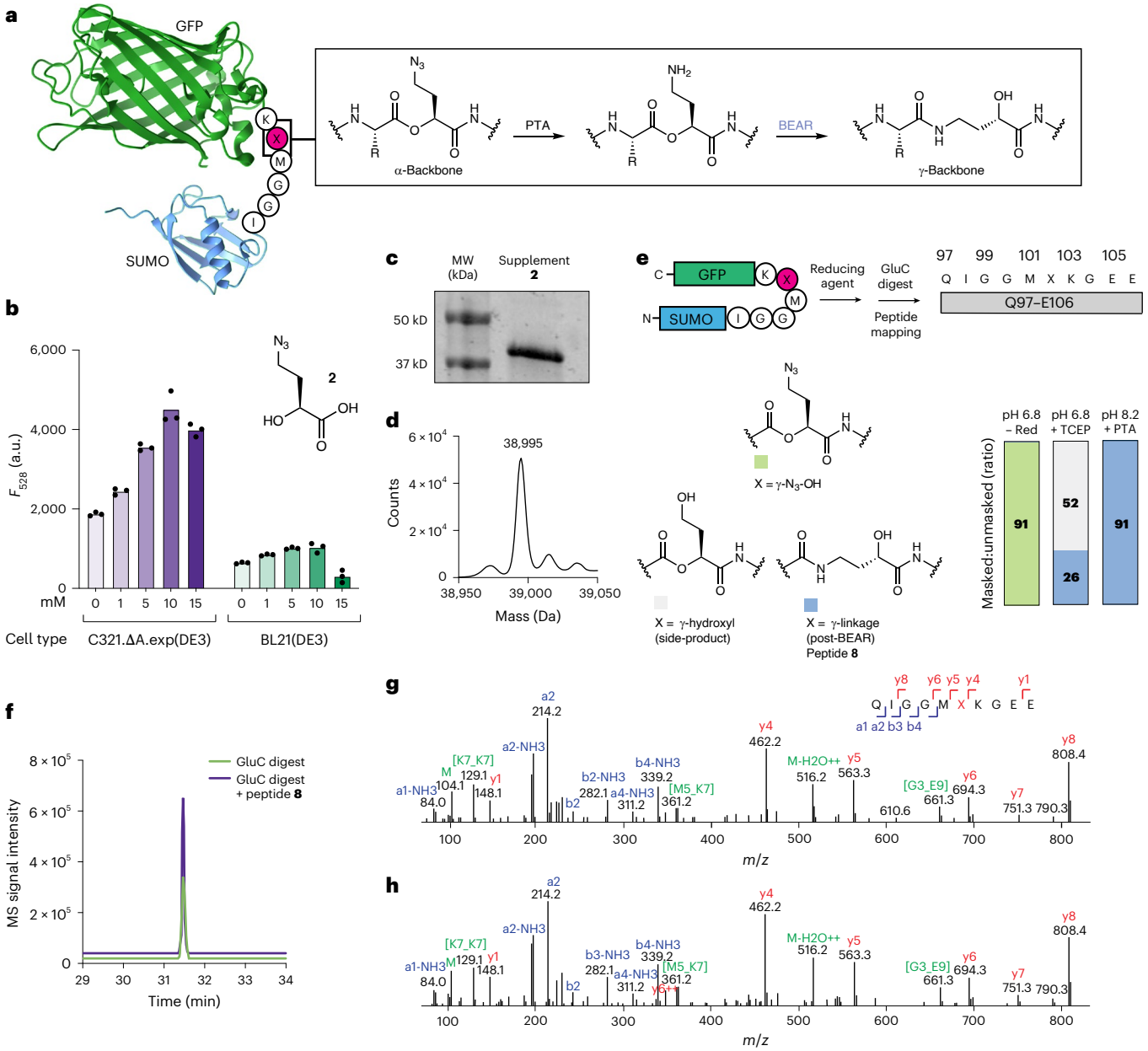


Fig. 4 | BEAR reactions generate a γ-linkage in SUMO–GFP in vivo. a, Scheme illustrating the utility of γ-N_3-OH **2** for intraprotein BEAR reactions. Cellular expression of a SUMO–GFP102TAG in the presence of γ-N_3-OH **2** generates a protein containing an internal α-ester. Chemical unmasking generates an intermediate carrying a side-chain γ-amine that can undergo a BEAR reaction to generate a γ-linkage (PDB 2B3P and 3QHT). **b**, Plot showing F_{528} (emission maximum of GFP) of *E. coli* C321.ΔA.exp(DE3) and BL21(DE3) cell growths expressing PylRS-N311M;C313W and SUMO–GFP102TAG after 24 h as a function of the concentration of γ-N_3-OH **2**. Graphed is the mean of three biological replicates. **c**, Coomassie-stained SDS–PAGE gel loaded with affinity-purified SUMO–GFP102TAG isolated from growths of C321.ΔA.exp(DE3) *E. coli* expressing PylRS-N311M;C313W and SUMO–GFP102TAG and supplemented with 10 mM γ-N_3-OH **2**. The experiment was performed in three biological replicates. **d**, Deconvoluted mass spectrum following LC–MS analysis of the material shown in **c**. The mass expected for SUMO–GFP102TAG containing γ-N_3-OH **2** at a single position is 38,994 Da; the observed mass is 38,995 Da. **e**, The material shown in **c** was treated with or without 50 mM TCEP at pH 6.8 (16 h at room temperature) or with 50 mM PTA at pH 8.2 (2 h at 0 °C followed by 2 h at 42 °C) and subjected to peptide mapping using GluC with analysis of the GluC fragment Q97–E106. The plot shows the relative ratios of Q97–E106 digestion products containing γ-N_3-OH **2**, the γ-hydroxyl (side product) or the anticipated γ-peptide linkage at position 102. **f**, The Q97–E106 GluC digestion product of SUMO–GFP102TAG containing γ-N_3-OH **2** after TCEP or PTA treatment comigrates with the Q97–E106 authentic standard containing the anticipated γ-peptide linkage at position 102 (peptide **8**). **g**,**h**, The MS/MS profile of the Q97–E106 GluC digestion product of SUMO–GFP102TAG containing γ-N_3-OH **2** after TCEP or PTA treatment (**g**) is identical to that of authentic standard peptide **8** (**h**), providing conclusive evidence for a successful BEAR reaction.

formation of the five-membered and six-membered ring intermediates required to form β-peptide and γ-peptide linkages (Supplementary Fig. 1). While Baldwin's rules allow for seven-membered ring closures through an exocyclic pathway, they can be slow in the absence of favorable conformational effects[44]. Indeed, the analogous transformation proceeds in only 40% yield in the context of a short, relatively unstructured peptide[34,45]. One could imagine that a more complex protein context could either inhibit or promote cyclization[44].

We transformed C321.ΔA.exp(DE3) and BL21(DE3) *E. coli* with plasmids encoding SUMO–GFP102TAG and PylRS-N311M;C313W, supplemented the growths with 0–15 mM δ-N_3-OH **3**, and monitored cell optical density at 600 nm (OD_{600}) and GFP fluorescence at 528 nm

https://doi.org/10.1038/s41589-025-01999-w

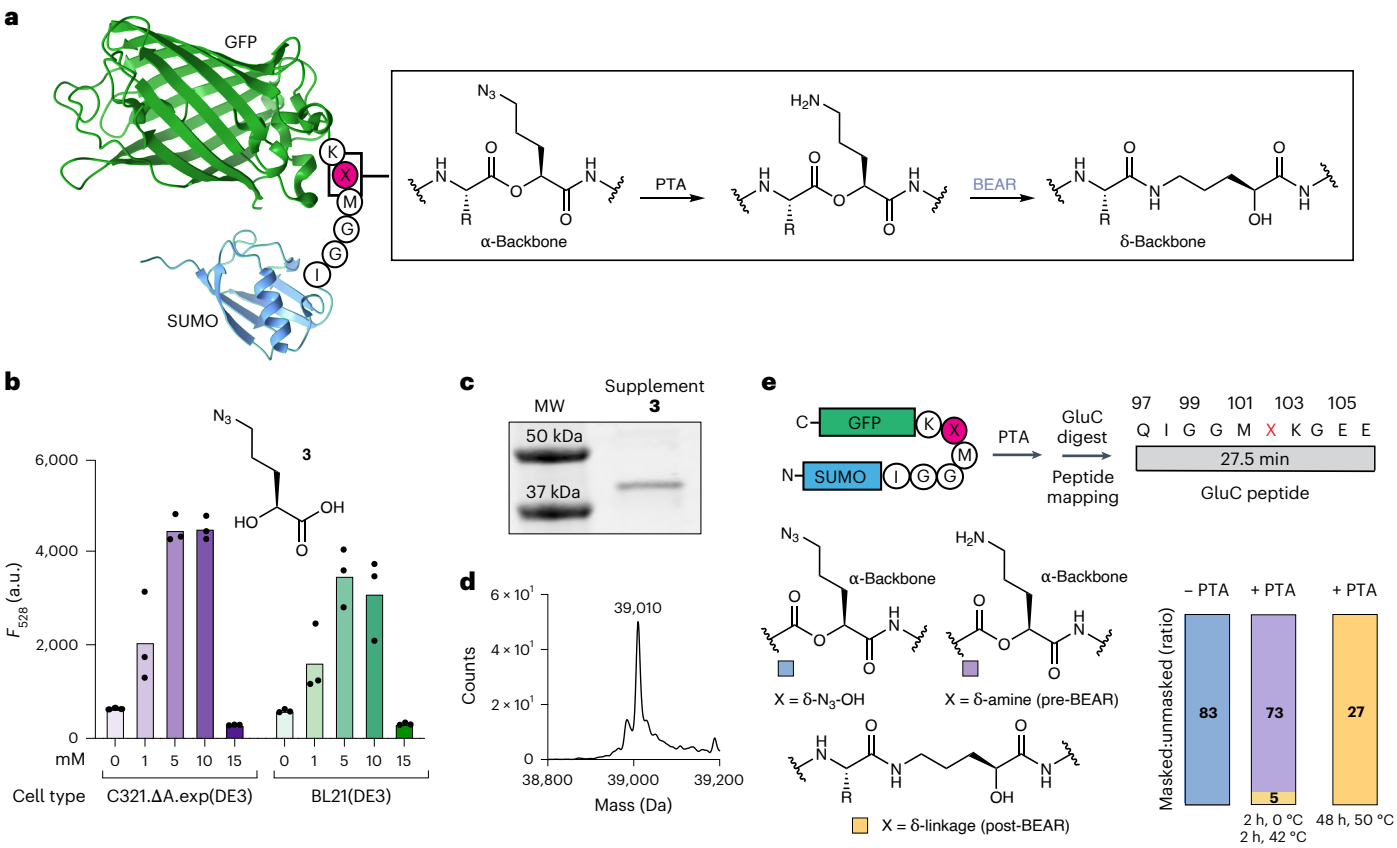


Fig. 5 | BEAR reactions generate one or two δ-linkages in SUMO–GFP in vivo. **a**, Scheme illustrating the utility of δ-N_3-OH **3** for the biosynthesis of proteins containing δ-peptide linkages (PDB 2B3P and 3QHT). **b**, Plot showing F_{528} at the 24 h time point as a function of the concentration of δ-N_3-OH **3** (0–15 mM) in C321.ΔA.exp(DE3) and BL21(DE3) *E. coli*. Graphed is the mean of three biological replicates. **c**, Coomassie-stained SDS–PAGE gel of affinity-purified SUMO–GFP102TAG from growths of BL21(DE3) *E. coli* expressing PylRS-N311M;C313W and SUMO–GFP with a TAG codon at position 102 and supplemented with 5 mM δ-N_3-OH **3**. The experiment was performed in three biological replicates. **d**, Deconvoluted mass spectrum of affinity-purified SUMO–GFP102TAG expressed in the presence of δ-N_3-OH **3**. The mass expected for SUMO–GFP containing δ-N_3-OH **3** at a single position is 39,010 Da; the observed mass is 39,010 Da. **e**, The material shown in **c** was treated with or without 50 mM PTA at pH 8.2 (2 h at 0 °C followed by 2 h at 42 °C or 48 h at 50 °C) and subjected to peptide mapping using GluC with analysis of the GluC fragment Q97–E106. The plot shows the relative ratios of Q97–E106 digestion products containing δ-N_3-OH **3**, the δ-amine (pre-BEAR) or the anticipated δ-peptide linkage at position 102.

(F_{528}) as a function of time. In both cases, increasing concentrations of δ-N_3-OH **3** (to a maximum of 10 mM) resulted in increasing levels of GFP expression with signal to noise ratios of 6.9 and 5.3 for C321.ΔA.exp(DE3) and BL21(DE3) *E. coli*, respectively, relative to a negative control in which no substrate was added (Fig. 5b and Supplementary Fig. 18a,b). To confirm that the expressed SUMO–GFP contained δ-N_3-OH **3**, a preparative-scale growth of BL21(DE3) *E. coli* expressing PylRS-N311M;C313W and SUMO–GFP102TAG was supplemented with 5 mM δ-N_3-OH **3** and the SUMO–GFP102TAG product was purified. This material had a mass of 39,010 Da, corresponding to the introduction of δ-N_3-OH **3** at a single position (Fig. 5c,d and Supplementary Fig. 19b–d). No side products were detected and the product was produced in a yield of 49 mg l^{-1} without further optimization.

Next, we sought to establish conditions to reductively unmask the azide side chain and promote its intramolecular isomerization into a δ-peptide linkage. Initially, we evaluated conditions similar to those used to optimize reduction of SUMO–GFP102TAG containing γ-N_3-OH **2** and its cyclization into a γ-peptide product. Purified SUMO–GFP102TAG containing δ-N_3-OH **3** was incubated with 50 mM PTA at 0 °C for 2 h then at 42 °C for 2 h at a pH of 6.8, 7.2, 7.9, 8.2 or 8.5. Products were evaluated by high-resolution GluC peptide mapping as described above, again focusing on the composition of the Q97–E106 fragment (Supplementary Fig. 20a,b). In the absence of PTA, the Q97–E106 GluC product contained more than 83% δ-N_3-OH **3** at position 102. After treatment with PTA, the presence of δ-N_3-OH **3** decreased to 1% under most conditions (Supplementary Fig. 20b); the majority of the product had a mass consistent with either the δ-amine (pre-BEAR intermediate resulting from only azide reduction) or the desired δ-peptide (post-BEAR). The pre-BEAR and post-BEAR products were differentiated on the basis of sensitivity to base hydrolysis, fragmentation pattern and retention time (Supplementary Methods and Supplementary Figs. 20 and 21). These studies revealed that the desired δ-peptide product was generated upon reduction with PTA at pH 8.2 but in only 5% yield; the majority of the product (73%) was the δ-amine, the pre-BEAR intermediate (Fig. 5e).

Recognizing that formation of the desired δ-linkage could proceed through higher-energy intermediates (Fig. 2), we explored additional reduction and cyclization conditions that varied time (8–48 h) and temperature (42 °C versus 50 °C) at pH 8.2; we also examined an exceptionally long-term reaction (264 or 456 h) at room temperature or 4 °C, respectively. Increases in both time and temperature improved the yield of the desired δ-linkage product within SUMO–GFP102TAG (Supplementary Fig. 20b). When SUMO–GFP102TAG containing δ-N_3-OH **3** was reduced at pH 8.2 and 50 °C for 48 h, we observed a 27% yield of the desired δ-peptide product with no δ-amine pre-BEAR intermediate remaining (Fig. 5e). Under these conditions we observed high levels of leucine and methionine at the TAG-encoded position (Supplementary Fig. 20b), suggesting that hydrolysis was competing with a relatively

slow BEAR reaction. Further structural and computational work will be required to understand the underlying conformational effects that promote higher δ-BEAR reactivity.

Discussion

There is great interest in strategies to achieve the programmed cellular synthesis of biopolymers whose monomers are not simply α-amino acids[1]. The availability of such molecules would provide otherwise nonexistent opportunities to expand and evolve protein and polypeptide structure and function and develop improved therapeutic agents[4–6]. However, despite the availability of multiple orthogonal aminoacyl-tRNA synthetase enzymes for noncanonical α-amino acids[47,48], since the initial report in 2016 (ref. 14), there have been only three additional examples in which an extended-backbone monomer has been introduced directly into a protein in cells[14,16,17,49]. Although certain PylRS enzymes process monomers with β^2-backbones and β^3-backbones, three challenges limit applying the classic form of genetic code expansion[47] to γ-monomers, δ-monomers or even more complex building blocks. One challenge is the ribosome, whose ability to promote highly efficient bond-forming reactions in vivo to and from anything other than an α-amino, β-amino or hydroxy acids remains unestablished[20]. The second is EF-Tu, whose ability to deliver tRNAs acylated with even modestly extended-backbone monomers (β^2-amino acids and β^3-amino acids) is severely challenged[22]. The third challenge is the availability of ribosome substrates—acylated tRNAs. Although recent work has expanded the diversity of monomers accepted by certain archaeal aaRS enzymes[19,23], there is still only one report of an orthogonal enzyme that acylates tRNA with anything other than a β^2-hydroxy or β^3-amino acid[49].

Here we bypass these challenges by reframing the problem of cellular hetero-oligomer synthesis in the language of chemistry. Rather than relying on direct reactions of extended-backbone monomers within the ribosome active site, the strategy reported here relies on post-translational, proximity-guided intramolecular rearrangements of ribosomal products that effectively edit the protein backbone into a heteropolymer. In the three cases reported here, the intramolecular BEAR reaction converts a cellular protein containing only an α-backbone into one containing a site-specific β^2-backbone, γ-backbone or δ-backbone. Although we demonstrate this concept using relatively unadorned side-chain nucleophiles, the strategy is easily expanded to alternative nucleophiles, side-chain motifs and unmasking strategies capable of installing even more diverse backbones into proteins expressed in cells and with temporal control. Although increasingly complex BEAR reactions may demand aaRS engineering, they should not require modifications to the ribosome or EF-Tu. We anticipate that BEAR reactions, such as those reported here, will enable the straightforward cellular synthesis and ultimately the evolution of extended-backbone protein hybrids for research and development of next-generation biomaterials and protein therapeutics.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-025-01999-w.

References

1. Kwon, D. How scientists are hacking the genetic code to give proteins new powers. *Nature* **618**, 874–876 (2023).
2. Kritzer, J. A., Lear, J. D., Hodsdon, M. E. & Schepartz, A. Helical β-peptide inhibitors of the p53–hDM2 interaction. *J. Am. Chem. Soc.* **126**, 9468–9469 (2004).
3. Sadowsky, J. D. et al. Chimeric (α/β+α)-peptide ligands for the BH3-recognition cleft of Bcl-xL: critical role of the molecular scaffold in protein surface recognition. *J. Am. Chem. Soc.* **127**, 11966–11968 (2005).
4. Hamuro, Y., Schneider, J. P. & DeGrado, W. F. De novo design of antibacterial β-peptides. *J. Am. Chem. Soc.* **121**, 12200–12201 (1999).
5. Cheloha, R. W., Maeda, A., Dean, T., Gardella, T. J. & Gellman, S. H. Backbone modification of a polypeptide drug alters duration of action in vivo. *Nat. Biotechnol.* **32**, 653–655 (2014).
6. Wang, P. S. P., Nguyen, J. B. & Schepartz, A. Design and high-resolution structure of a β^3-peptide bundle catalyst. *J. Am. Chem. Soc.* **136**, 6810–6813 (2014).
7. Girvin, Z. C., Andrews, M. K., Liu, X. & Gellman, S. H. Foldamer-templated catalysis of macrocycle formation. *Science* **366**, 1528–1531 (2019).
8. Daniels, D. S., Petersson, E. J., Qiu, J. X. & Schepartz, A. High-resolution structure of a β-peptide bundle. *J. Am. Chem. Soc.* **129**, 1532–1533 (2007).
9. Seebach, D. et al. Biological and pharmacokinetic studies with β-peptides. *Chimia* **52**, 734–734 (1998).
10. Adaligil, E., Song, A., Hallenbeck, K. K., Cunningham, C. N. & Fairbrother, W. J. Ribosomal synthesis of macrocyclic peptides with β^2- and $\beta^{2,3}$-homo-amino acids for the development of natural product-like combinatorial libraries. *ACS Chem. Biol.* **16**, 1011–1018 (2021).
11. Salveson, P. J. et al. Expansive discovery of chemically diverse structured macrocyclic oligoamides. *Science* **384**, 420–428 (2024).
12. Cheloha, R. W. et al. Consequences of periodic α-to-β^3 residue replacement for immunological recognition of peptide epitopes. *ACS Chem. Biol.* **10**, 844–854 (2015).
13. Cheloha, R. W. et al. Recognition of class II MHC peptide ligands that contain β-amino acids. *J. Immunol.* **203**, 1619–1628 (2019).
14. Melo Czekster, C., Robertson, W. E., Walker, A. S., Söll, D. & Schepartz, A. In vivo biosynthesis of a β-amino acid-containing protein. *J. Am. Chem. Soc.* **138**, 5194–5197 (2016).
15. Scott, T. A. et al. Widespread microbial utilization of ribosomal β-amino acid-containing peptides and proteins. *Chem* **8**, 2659–2677 (2022).
16. Hamlish, N. X., Abramyan, A. M., Shah, B., Zhang, Z. & Schepartz, A. Incorporation of multiple β^2-hydroxy acids into a protein in vivo using an orthogonal aminoacyl-tRNA synthetase. *ACS Cent. Sci.* **10**, 1044–1053 (2024).
17. Dunkelmann, D. L. et al. Adding α,α-disubstituted and β-linked monomers to the genetic code of an organism. *Nature* **625**, 603–610 (2024).
18. Katoh, T. & Suga, H. Ribosomal incorporation of consecutive β-amino acids. *J. Am. Chem. Soc.* **140**, 12159–12167 (2018).
19. Fricke, R. et al. Expanding the substrate scope of pyrrolysyl-transfer RNA synthetase enzymes to include non-α-amino acids in vitro and in vivo. *Nat. Chem.* **15**, 960–971 (2023).
20. Watson, Z. L. et al. Atomistic simulations of the *Escherichia coli* ribosome provide selection criteria for translationally active substrates. *Nat. Chem.* **15**, 913–921 (2023).
21. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
22. Cruz-Navarrete, F. A. et al. β-Amino acids reduce ternary complex stability and alter the translation elongation mechanism. *ACS Cent. Sci.* **10**, 1262–1275 (2024).
23. Kobayashi, T., Yanagisawa, T., Sakamoto, K. & Yokoyama, S. Recognition of non-α-amino substrates by pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* **385**, 1352–1360 (2009).

24. Li, Y.-M. et al. Ligation of expressed protein α-hydrazides via genetic incorporation of an α-hydroxy acid. *ACS Chem. Biol.* **7**, 1015–1022 (2012).
25. Spinck, M. et al. Genetically programmed cell-based synthesis of non-natural peptide and depsipeptide macrocycles. *Nat. Chem.* **15**, 61–69 (2023).
26. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **1844**, 1059–1070 (2014).
27. Huguenin-Dezot, N. et al. Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **565**, 112–117 (2019).
28. Koch, N. G., Goettig, P., Rappsilber, J. & Budisa, N. Engineering pyrrolysyl-tRNA synthetase for the incorporation of non-canonical amino acids with smaller side chains. *Int. J. Mol. Sci.* **22**, 11194 (2021).
29. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
30. Sohma, Y., Sasaki, M., Hayashi, Y., Kimura, T. & Kiso, Y. Novel and efficient synthesis of difficult sequence-containing peptides through *O–N* intramolecular acyl migration reaction of *O*-acyl isopeptides. *Chem. Commun.* 124–125 (2004).
31. Dos Santos, S. et al. Switch-peptides: controlling self-assembly of amyloid β-derived peptides in vitro by consecutive triggering of acyl migrations. *J. Am. Chem. Soc.* **127**, 11888–11889 (2005).
32. Schissel, C. K. et al. Ribosomal synthesis of ketone-containing peptide backbone via O to C acyl shift. *J. Am. Chem. Soc.* **147**, 6503–6513 (2025).
33. Dose, C. & Seitz, O. New isocysteine building blocks and chemoselective peptide ligation. *Org. Biomol. Chem.* **2**, 59–65 (2004).
34. Kuroda, T., Huang, Y., Nishio, S., Goto, Y. & Suga, H. Post-translational backbone-acyl shift yields natural product-like peptides bearing hydroxyhydrocarbon units. *Nat. Chem.* **14**, 1413–1420 (2022).
35. Wang, C., Guo, Q.-X. & Fu, Y. Theoretical analysis of the detailed mechanism of native chemical ligation reactions. *Chem. Asian J.* **6**, 1241–1251 (2011).
36. Grimme, S., Hansen, A., Ehlert, S., & Mewes, J.-M. r^{2}SCAN-3c: a "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **154**, 064103 (2021).
37. Mardirossian, N. & Head-Gordon, M. ωB97M-V: a combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **144**, 214110 (2016).
38. Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overviewand extensive assessment of 200 density functionals. *Mol. Phys.* **115**, 2315–2372 (2017).
39. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).
40. Hall, M. P. et al. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).
41. Volkwein, W., Maier, C., Krafczyk, R., Jung, K. & Lassak, J. A versatile toolbox for the control of protein levels using N^{ε}-acetyl-L-lysine dependent amber suppression. *ACS Synth. Biol.* **6**, 1892–1902 (2017).
42. Leisle, L. et al. Cellular encoding of Cy dyes for single-molecule imaging. *eLife* **5**, e19088 (2016).
43. Lee, J., Schwarz, K. J., Kim, D. S., Moore, J. S. & Jewett, M. C. Ribosome-mediated polymerization of long chain carbon and cyclic amino acids into peptides in vitro. *Nat. Commun.* **11**, 4304 (2020).
44. Agouridas, V. et al. Native chemical ligation and extended methods: mechanisms, catalysis, scope, and limitations. *Chem. Rev.* **119**, 7328–7443 (2019).
45. Baldwin, J. E. Rules for ring closure. *J. Chem. Soc. Chem. Commun.* **18**, 734–736 (1976).
46. Lajoie, M. J. et al. Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
47. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
48. Young, T. S. & Schultz, P. G. Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J. Biol. Chem.* **285**, 11039–11044 (2010).
49. Chen, S., Ji, X., Gao, M., Dedkova, L. M. & Hecht, S. M. In cellulo synthesis of proteins containing a fluorescent oxazole amino acid. *J. Am. Chem. Soc.* **141**, 5597–5601 (2019).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

Methods

Computational studies

For each compound studied, molecular mechanics methods (Macromodel, OPLS4 force field) were used to minimize a starting population of 10,000 conformers that was initially generated using the conformational search tool in the Schrödinger Maestro 2022-2 environment. The geometries of all conformers that were found to be within 5 kcal mol^{-1} of the global minimum were then optimized again using DFT in Orca (version 6.0) with r^2SCAN-3c/mTZVPP and a CPCM water model (with Bondi radii)[36]. After this step, all of the unique conformations within 2 kcal mol^{-1} of the global minimum were subjected to a second geometry optimization and vibrational frequency calculation in Orca (version 6.0, again using r^2SCAN-3c/mTZVPP and a CPCM water model) to determine the zero-point energies, enthalpy corrections and the entropy values at 298.15 K. All compounds were confirmed to have no imaginary vibrational frequency modes. For all of these species, single-point electronic energy calculations were performed on the optimized geometries using an improved functional and an expanded basis set (ωB97M-V/TZVPPD)[37,38]. An SMD solvent model (for water) was used in these calculations[39]. The final enthalpy, entropy and free energy values were calculated by combining the results of the r^2SCAN-3c/mTZVPP/CPCM frequency calculations with the ωB97M-V/TZVPPD/SMD electronic energy values. These results were used to identify the global energy minima for each intermediate and these species were compared to determine the relative energy changes for each reaction under study.

Expression test of NanoLuc from pET15a-NanoLuc(G159-TAG-V160 insertion)-6×His

Starter cultures of *E. coli* picked from a single colony were grown overnight in 10 ml of Miller's Luria–Bertani broth (AmericanBio, AB01201) in a 15-ml culture tube supplemented with antibiotics at 37 °C. Saturated overnight cultures were used to induce a 50-ml culture (1:10) of each construct and grown in 37 °C with shaking at 220 rpm and the OD was monitored over the next several hours. Before cultures reaching the desired OD, appropriate amounts of monomers (stored as 200 mM stocks) were transferred into a black, clear-bottom 96-well plate (1 mM final concentration unless otherwise indicated). Once cultures reached an OD_{600} of 0.6, protein expression was induced by addition of 1 mM IPTG or 0.2% arabinose (for pET32a and pET15a, respectively) and culture was transferred into the plate for a total volume of 200 µl in each well. A Breathe-Easy sealing membrane (Sigma, Z380059) was placed over the 96-well plate and placed into a 37 °C shaker while shielding the plate from light. After 16 h, 50-µl samples from each condition were transferred onto a white-walled 96-well plate and the Nano-Glo luciferase assay system (Promega N1110) was added according to manufacturer instructions. The luminescence was measured in BioTek Synergy H1 microplate reader as an endpoint assay. Points represented in Fig. 3 are biological replicates where three random colonies were picked on a plate from a single transformation. Points represented in Supplementary Fig. 8 are biological replicates where three random colonies were picked and the same experiment was performed on each of the three cultures.

Large-scale expression and purification of NanoLuc G159-pm-DAP-V160 and G159-pm-isoS-V160

Starter cultures of 10 ml of Miller's LB broth (AmericanBio, AB01201) supplemented with tetracycline and kanamycin were inoculated with a single colony of *E. coli* MegaX DH10B T1R Electrocomp cells harboring pBK-DAPRS and pET15a-NanoLuc(G159-TAG-V160 insertion)-6×His and grown for 24 h at 37 °C with shaking at 220 rpm until the culture was saturated. The starter culture (10 ml) was used to inoculate a 200-ml expression culture of Miller's LB broth supplemented with 1 mM pm-DAP **4** or 2 mM pm-isoS **1**, in addition to tetracycline and kanamycin. The expression culture was grown at 37 °C with shaking at 220 rpm to an OD_{600} of 0.6 (shielded from light), at which point it was induced with 0.2% arabinose and grown for 16 h under the same conditions. The expression culture was harvested by centrifugation at 4,300*g* at 4 °C for 30 min. The resulting cell pellet was suspended in 10 ml of lysis buffer (20 mM HEPES, 50 mM KCl, 10% glycerol and 10 mM imidazole, pH 8.6) containing one tablet of cOmplete, mini EDTA-free ULTRA protease inhibitor cocktail (Sigma-Aldrich). The cell suspension was disrupted by sonication on ice (Branson Sonifier 250; five cycles of 30-s pulses at 50% duty cycle and a microtip limit of 5 followed by a 30-s pause). The cell lysate was cleared by centrifugation at 23,000*g* at 4 °C for 20 min. TALON metal affinity resin (2 ml; Takara Biosciences, 635504) was equilibrated with lysis buffer, added to the cleared cell lysate and incubated on a rotisserie at 4 °C for 1 h. The TALON resin–lysate mixture was then passed through a gravity flow Poly-Prep chromatography column (Bio-Rad Laboratories). Nonspecifically bound proteins were removed by washing the TALON resin with 10 ml of lysis buffer. The 6×His-tagged protein was eluted by washing the TALON resin with 2 ml of elution buffer (20 mM HEPES, 50 mM KCl, 10% glycerol and 1 M imidazole, pH 8.6). The purified protein was loaded on to a PD-10 column (Cytiva Life Sciences, 17085101) and exchanged into 3.5 ml of storage buffer (20 mM HEPES, 50 mM KCl and 10% glycerol, pH 8.6) according to manufacturer instructions. The purified protein was quantified using absorbance at 280 nm, snap-frozen as single-use aliquots and stored at −80 °C. The typical expression yield of NanoLuc constructs using the above protocol was 0.4 mg l^{-1} of *E. coli* culture.

Trypsin digestion for β²-linkages in NanoLuc G159-TAG-V160

The initial concentration of the nine NanoLuc samples varied from 0.1 to 0.8 mg ml^{-1} in storage buffer (20 mM HEPES, 50 mM KCl and 10% glycerol, pH 8.6). Each sample was concentrated, if necessary, to ~0.8 mg ml^{-1} using a Microcon 10-kDa filter. Then, 50 µl of each sample (~40 µg) was then added to 150 µl of denaturing buffer (8 M guanidine HCl and 0.2 M Tris, pH 7.5) to achieve a final concentration of 6 M guanidine and 0.15 M Tris. Disulfides were reduced by adding 400 mM dithiothreitol (DTT) stock solution in deionized water to achieve a final DTT concentration of 8 mM, followed by incubating at 37 °C for 30 min. The reduced and denatured product was then alkylated by adding 400 mM iodoacetamide to achieve a final concentration of 4 mM and incubating at 25 °C for 25 min, followed by the addition of DTT to a final concentration of 6 mM to quench the reaction. The reduced and alkylated protein was exchanged into ~45 µl of 0.1 M Tris buffer at pH 7.5 using a Microcon 10-kDa membrane, followed by addition of 4 µg trypsin (in a 1 µg $µl^{-1}$ solution) directly to the membrane to achieve an enzyme-to-substrate ratio of at least 1:10. After 2 h at 37 °C, the digestion was quenched with an equal volume of 0.25 M acetate buffer (pH 4.8) containing 6 M guanidine. Peptide fragments were collected by spinning down through the membrane and then subjected to LC–MS/MS analysis.

Peptide mapping data collection for β²-linkages

LC–MS/MS analysis was performed on an Agilent 1290-II high-performance LC (HPLC) instrument directly connected to a Thermo Fisher Q Exactive HF high-resolution MS instrument. Peptides were separated on a Waters HSS T3 reverse-phase column (2.1 × 150 mm) at 50 °C with a 70-min acetonitrile gradient (0.5% to 35% acetonitrile in H_2O) containing 0.1% formic acid in the mobile phase and a total flow rate of 0.25 ml min^{-1}. The MS data were collected at 120,000 resolution, followed by data-dependent higher-energy collision dissociation (HCD) MS/MS at a normalized collision energy of 25%.

Peptide mapping data analysis for β²-linkages

Proteolytic peptides were identified and quantified on MassAnalyzer, an Amgen in-house developed program (available in Biopharma Finder from Thermo Fisher). The program performs feature extraction, peptide identification, retention time alignment and peak integration in

an automated fashion. Searched modifications include those shown in Supplementary Table 2 and amino acid substitutions.

Solid-phase peptide synthesis for β²-linkages. Peptides **6** and **7** were synthesized on a 50-µmol scale using Fmoc chemistry and Fmoc-Arg(Pbf)-Wang resin (0.35 mmol g^{-1} substitution, 100–200 mesh) on a PurePep Chorus from Gyros Protein Technologies. Resin was first swollen in 6 ml of dimethylformamide (DMF) for 20 min at room temperature. The Fmoc protecting group was removed by treating the resin twice (3 min each, at room temperature) with 5 ml of deprotection solution (20% (v/v) piperidine in DMF). Excess deprotection solution was removed by washing the resin four times with 6 ml of DMF. Coupling reactions were performed by adding five equivalents of amino acid, hydroxybenzotriazole (HOBt) and benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) (from 200 mM stocks in DMF) and ten equivalents of *N*,*N*-diisopropylethylamine (DIPEA) (from a 2 M stock in *N*-methyl-2-pyrrolidone (NMP)) to the deprotected resin for a final resin-bound peptide concentration of 12.5 mM. In general, each amino acid coupling reaction was performed at 70 °C for 5 min. For peptide **6**, the coupling reaction that added isoS was modified to contain only one equivalent of Fmoc-isoS(*t*Bu)-OH. Excess coupling reagents were removed by washing four times with 6 ml of DMF. Once the synthesis was complete, the resin-bound peptide was dried under nitrogen and removed from the resin by treatment with 2 ml of cleavage cocktail (95% trifluoroacetic acid (TFA), 2.5% water and 2.5% triisopropyl silane (TIPS)) stirred for 1 h at room temperature. Cleaved peptide was dried under reduced pressure before purification by HPLC. Peptide **32** was synthesized as described above for peptides A and B with some modifications. Coupling reactions were performed by adding five equivalents of amino acid, HOBt and hexafluorophosphate benzotriazole tetramethyl uronium (from 200 mM stocks in DMF) and ten equivalents of DIPEA (from a 2 M stock in NMP) to the deprotected resin for a final resin-bound peptide concentration of 12.5 mM.

Peptide HPLC purification. HPLC purification was performed on an Waters LC Prep 150 system equipped with a Waters 2998 ultraviolet photodiode array detector, a Waters 2707 autosampler and a Waters Fraction Collector III using a preparative reverse-phase C18 column (CSH C18 19 × 150 mm OBD column, 5 µm). The mobile phase for HPLC was water with 0.1% (v/v) trifluoroacetic acid (solvent A) and acetonitrile with 0.1% (v/v) trifluoroacetic acid (solvent B). Peptides were eluted at a flow rate of 20 ml min^{-1} using a linear solvent gradient from 5% to 50% solvent B in solvent A over 30 min. Peptides were collected on the basis of their absorbance at 280 nm.

MS of purified peptides. LC–MS analysis of each peptide was performed on an Agilent 1290 Infinity II HPLC instrument connected to an Agilent 6530B quadrupole time-of-flight AJS-ESI MS instrument. The mobile phase for LC–MS was water and acetonitrile with 0.1% (v/v) formic acid at a flow rate of 0.7 ml min^{-1}. Each peptide was injected onto an Eclipse XDB C18 column (2.1 × 50 mm, 1.8 µm, room temperature; Agilent) and separated using a linear gradient from 5% to 95% acetonitrile over 4.5 min after an initial hold at 5% acetonitrile for 0.5 min. The following parameters were used during acquisition: fragmentor voltage, 175 V; gas temperature, 300 °C; gas flow, 8 L min^{-1}; sheath gas temperature, 350 °C; sheath gas flow, 11 L min^{-1}; nebulizer pressure, 35 psi; skimmer voltage, 65 V; V_{cap}, 3,500 V; one spectrum per second.

Expression assays of SUMO–GFP102TAG–His

Starter *E. coli* cultures were grown overnight in 10 ml of Miller's LB broth (AmericanBio, AB01201) in 15-ml culture tubes supplemented with antibiotics at 37 °C. The next morning, 500 µl of the saturated overnight culture was added to 50 ml of LB in a 250-ml baffled flask with antibiotics and grown at 37 °C to an OD_{600} of 0.6. Before cultures reaching an OD_{600} of 0.6, the appropriate amount of each monomer (stored as 100 mM stocks) was allotted into the wells of a black, clear-bottom 96-well plate. Once cultures reached an OD_{600} of 0.6 (roughly 3 h), protein expression was induced by addition of 0.2% arabinose and the culture was transferred into the wells of the plate to ensure a total volume of 200 µl in each well. A Breathe-Easy sealing membrane (Sigma, Z380059) was placed over the 96-well plate and the plate was loaded into a BioTek Synergy HTX microplate reader with no lid. OD_{600} and F_{528} values (λ_{ex} = 485 nm) were measured every 10 min for 24 h. The plate was maintained at 37 °C and was shaken during this time (Supplementary Figs. 14 and 18a,b). Points represented in Figs. 4 and 5 are biological replicates where three random colonies were picked on a plate from a single transformation.

Expression and purification of SUMO–GFP102TAG–His

The expression and purification of SUMO–GFP102γ-N_3-OH–His and SUMO–GFP102δ-N_3-OH–His followed prior established protocols[27,28] (Supplementary Figs. 15b–d and 19b–d).

GluC digestion for γ-linkages and δ-linkages in SUMO–GFP102TAG

Each GFP sample (~33 µg) was denatured with 6 M guanidine in a 0.15 M Tris buffer at pH 7.5, followed by disulfide reduction with 8 mM DTT at 37 °C for 30 min. The reduced GFP sample was alkylated in the presence of 14 mM iodoacetamide at 25 °C for 25 min, followed by quenching using 6 mM DTT. The reduced and alkylated protein was exchanged into ~90 µl of 0.1 M Tris buffer at pH 7.5 using a Microcon 10-kDa membrane. To half of the sample (45 µl or 16 µg), 2 µg of GluC (in a 0.5 µg $µl^{-1}$ solution) was added directly to the membrane to achieve an enzyme-to-substrate ratio of at least 1:8. After ~2.5 h at 37 °C, the digestion was quenched with an equal volume of 0.25 M acetate buffer (pH 4.8) containing 6 M guanidine. Peptide fragments were collected by spinning down through the membrane and subjected to LC–MS/MS analysis.

Peptide mapping data collection for γ-linkages and δ-linkages

LC–MS/MS analysis was performed on an Agilent 1290-II HPLC directly connected to a Thermo Fisher Q Exactive HF high-resolution MS instrument. Peptides (~10 µg) were separated on a Waters peptide HSS T3 reverse-phase column (2.1 × 150 mm) at 50 °C with a 70-min acetonitrile gradient (0.5% to 35%) containing 0.025% TFA in the mobile phase and a total flow rate of 0.25 ml min^{-1}. The MS data were collected at 120,000 resolution setting, followed by data-dependent higher-energy collision dissociation MS/MS at a normalized collision energy of 25%.

Peptide mapping data analysis for γ-linkages and δ-linkages

Proteolytic peptides were identified and quantified on MassAnalyzer, an Amgen in-house developed program (available in Biopharma FinderTM from Thermo Fisher). The program performs feature extraction, peptide identification, retention time alignment and peak integration in an automated fashion. Searched modifications include those shown in Supplementary Tables 3 and 4, amino acid substitutions and unrestricted searches of any mass changes.

Solid-phase peptide synthesis for γ-linkages. Peptides **33** and **8** were synthesized on a 50-µmol scale using Fmoc chemistry and Fmoc-Glu(OAll)-wang resin (0.4 mmol g^{-1} substitution, 100–200-mesh) on a PurePep Chorus from Gyros Protein Technologies. Resin was first swollen in 6 ml of DMF for 20 min at room temperature. The Fmoc protecting group was removed by treating the resin twice (3 min each, at room temperature) with 5 ml of deprotection solution (20% (v/v) piperidine in DMF). Excess deprotection solution was removed by washing the resin four times with 6 ml of DMF. Coupling reactions were performed by adding five equivalents of amino acid, HOBt and PyBOP (from 200 mM stocks in DMF) and ten equivalents of DIPEA (from a 2 M stock in NMP) to the deprotected resin for a final resin-bound peptide

concentration of 12.5 mM. In general, each amino acid coupling reaction was performed at 70 °C for 5 min. For peptide **33**, the coupling reaction that added isothreonine (isoT) was modified to contain only one equivalent of Fmoc-isoT(*t*Bu)-OH. Excess coupling reagents were removed by washing four times with 6 ml of DMF. Once the synthesis was complete, the resin-bound peptide was dried under nitrogen and removed from the resin by treatment with 2 ml of cleavage cocktail (95% TFA, 2.5% water and 2.5% TIPS) stirred for 1 h at room temperature. Cleaved peptide was dried under reduced pressure before purification by HPLC.

Reporting summary

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

Data availability

Data that support the main findings of this study are available within the article and Supplementary Information. All primary data for figures included as Supplementary Information are provided in Supplementary Data 1. Alternatively, data included in this study are also available from the corresponding authors upon request. Source data are provided with this paper.

Acknowledgements

We are grateful to members of the A.S., S.J.M. and A.C. labs for comments and suggestions, and to J. Chin (Medical Research Council) and N. Budisa (University of Manitoba) for sharing materials. We thank H. Celik for assistance with nuclear magnetic resonance spectroscopy, National Institutes of Health grant S10OD024998 for spectrometer support and the Chemical and Biophysical Instrumentation Center at Yale University (RRID:SCR_021738). This work was supported by the National Science Foundation (NSF) Center for Genetically Encoded Materials (CHE 2002182 to A.S.). L.T.R., T.L.D. and D.A.D. were supported by the NSF Graduate Research Fellowship Program (DGE-1752814, DGE-2139841 and DGE-1122492, respectively). C.K.S. was supported by the Miller Institute for Basic Research in Science (University of California, Berkeley). A.S. is a Chan-Zuckerberg Biohub San Francisco Investigator and an Arc Innovation Investigator.

Author contributions

Study conceptualization and design, L.T.R., A.C., M.B.F., S.J.M. and A.S. Preparation of materials, L.T.R., I.M.P., C.K.S., T.L.D., N.X.H., A.M.G., S.Z., D.A.D. and N.W. Data collection, L.T.R., I.M.P., C.K.S., T.L.D., B.S., N.X.H., A.M.G., S.Z., N.W. and M.B.F. Analysis and interpretation of results, L.T.R., I.M.P., C.K.S., T.L.D., B.S., Z.Z., M.B.F., S.J.M. and A.S. Paper preparation, L.T.R., C.K.S., T.L.D., M.B.F., S.J.M. and A.S.

Competing interests

L.T.R. and A.S. have submitted a patent application related to this work through The Regents of the University of California (provisional application 63/587,179 filed October 2, 2023; PCT Application PCT/US2024/048032 filed 9/23/2024; aspect covered: methods for synthesis of protein heteropolymers). The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-025-01999-w.

Correspondence and requests for materials should be addressed to Matthew B. Francis, Scott J. Miller or Alanna Schepartz.

Reprints and permissions information is available at www.nature.com/reprints.

nature portfolio

Corresponding author(s): Matthew Francis, Scott Miller, Alanna Schepartz

Last updated by author(s): May 21, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	A BioTek Synergy H1 and HTX microplate reader was used to collect F528 and luminescence expression data using the Gen5 3.08 software. Peptide mapping using LC-MS/MS analysis was performed on an Agilent 1290-II HPLC directly connected to a Thermo Fisher Q Exactive HF high-resolution mass spectrometer.
Data analysis	DFT analysis was performed in the Schrödinger Maestro 2022-2. Expression data plots was created using GraphPad Prism 10.4.2. Proteolytic peptides were identified and quantified using MassAnalyzer 6.08. NMR spectra were processed using MestReNova 14.2.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All primary data for main text figures are provided as Source Data. All primary data for figures included as Supplementary Information are provided in Supplementary Data 1. Any additional data included in this study are available from the corresponding author upon request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Based on previous reports (doi: 10.1038/s41586-023-06897-6), three biological replicates were performed for all in vivo incorporation experiments (Fig. 3bc, 4bc, and 5bc). Three technical replicates were performed for supplementary figs. 8 and 13. There are no animal studies described, so statistical methods to determine sample size were not performed.
Data exclusions	Data was only excluded in cases where expired reagents were discovered to have been used accidentally.
Replication	For all experiments described in the main text, a minimum of three biological replicates were performed. For supplementary fig. 6 and 8, three technical replicates were performed. Individual values are reported in the figures.
Randomization	For biological replicates, colonies were selected from a plate at random to avoid bias in colony selection. Colonies were expected to be genetically identical (a clonal population was transformed with purified plasmid and selected via antibiotic resistance on the same plate), so covariates such as treatment condition or genotype were not relevant.

Blinding	Blinding was not possible for these experiments which were all performed by individual authors.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging