LETTER

https://doi.org/10.1038/s41586-019-1020-y

Discovery of a pathway for terminal-alkyne amino acid biosynthesis

J. A. Marchand[1], M. E. Neugebauer[1], M. C. Ing[2], C.-I. Lin[3], J. G. Pelton[4] & M. C. Y. Chang[2,3]*

Living systems can generate an enormous range of cellular functions, from mechanical infrastructure and signalling networks to enzymatic catalysis and information storage, using a notably limited set of chemical functional groups. This observation is especially notable when compared to the breadth of functional groups used as the basis for similar functions in synthetically derived small molecules and materials. The relatively small cross-section between biological and synthetic reactivity space forms the foundation for the development of bioorthogonal chemistry, in which the absence of a pair of reactive functional groups within the cell allows for a selective in situ reaction[1–4]. However, biologically 'rare' functional groups, such as the fluoro[5], chloro[6,7], bromo[7,8], phosphonate[9], enediyne[10,11], cyano[12], diazo[13], alkene[14] and alkyne[15–17] groups, continue to be discovered in natural products made by plants, fungi and microorganisms, which offers a potential route to genetically encode the endogenous biosynthesis of bioorthogonal reagents within living organisms. In particular, the terminal alkyne has found broad utility via the Cu(I)-catalysed azide-alkyne cycloaddition 'click' reaction[18]. Here we report the discovery and characterization of a unique pathway to produce a terminal alkyne-containing amino acid in the bacterium *Streptomyces cattleya*. We found that L-lysine undergoes an unexpected reaction sequence that includes halogenation, oxidative C–C bond cleavage and triple bond formation through a putative allene intermediate. This pathway offers the potential for de novo cellular production of halo-, alkene- and alkyne-labelled proteins and natural products from glucose for a variety of downstream applications.

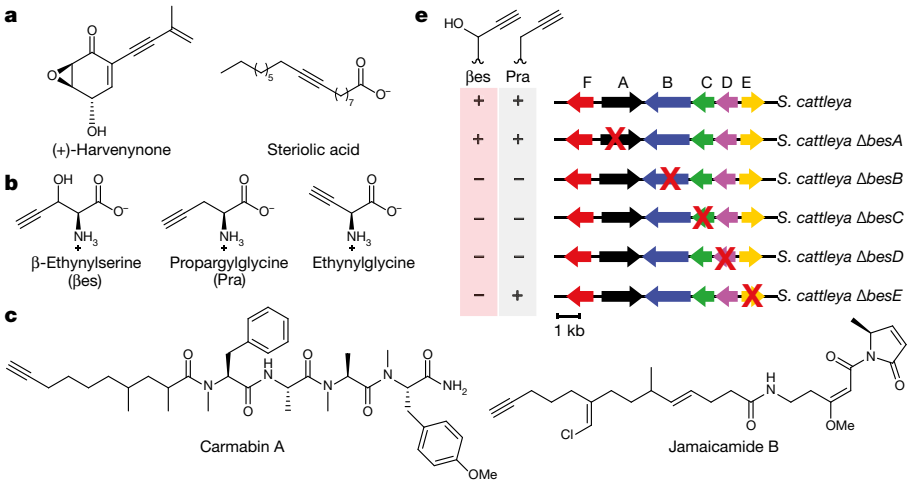
Although acetylenic natural products form a relatively large class of compounds (approximately 2,000)[17,19], very few of these contain a terminal alkyne that can be used in the Cu(i)-catalysed azide-alkyne cycloaddition (CuAAC) reaction (Fig. 1a). Notably, a set of terminal alkyne-containing amino acids has been reported to be produced by a select group of fungi and bacteria[20–22] (Fig. 1b). Both residue-specific[23] and site-specific[24] labelling of proteins and proteomes has been achieved using synthetic terminal-alkyne amino acids. The identification of the biosynthetic pathway for these amino acids could potentially provide a general route to encode the in situ production of terminal alkyne-containing proteins directly from glucose.

Early isotope-labelling experiments showed that many terminal-alkyne compounds originate from acetate-derived fatty acid or polyketide biosynthetic pathways. It has previously been suggested that these pathways use a fatty acid desaturase-like enzyme to form the triple bond, similar to internal acetylenic compounds[19]. More recently, fatty acid desaturases have been shown to participate in the formation of the terminal alkyne of the polyketide natural products jamaicamide B and carmabin A[15,25] (Fig. 1c). However, the few terminal-alkyne amino acids that have been described appear to be structurally different from products that originate from fatty acids[26]. We thus focused on identifying the genes for production of β-ethynylserine (βes) in a genetically tractable soil bacterium, *S. cattleya*. We initially targeted all four fatty acid desaturases found in *S. cattleya* for gene disruption. However, knockout of the three non-essential fatty acid desaturases (SCAT_4823, SCAT_p1525 and SCAT_0184) did not disrupt βes production, which suggests that βes is not produced by a canonical desaturation (Extended Data Fig. 1a, Supplementary Table 1). We then shifted our attention to comparative genomic approaches to identify uniquely shared gene clusters among *Streptomyces* spp. that produce terminal-alkyne amino acids. Of the two other known producers, only *Streptomyces catenulae* had an available genome sequence and was therefore validated for terminal-alkyne amino acid production in cell culture (Extended Data Fig. 1b). By comparison of the 2 genomes, we were able to reduce the roughly 650 total gene clusters conserved between them to just 4 unique clusters not found in the other 26 control *Streptomyces* spp. genomes (Extended Data Fig. 1c). Among these four, we focused on one particular cluster that contained gene functions predicted to be associated with amino acid metabolism.

The putative βes biosynthetic cluster features six proteins (Fig. 1d). BesA possesses a highly conserved ATP-grasp domain and shares homology with carboxylate-amine ligases. BesB is homologous to the pyridoxal-5′-phosphate (PLP)-dependent cystathionine-β-lyase/cystathionine-γ-synthase family. BesC displays homology to enzymes in the HemeO non-haem iron oxidase superfamily, and both BesD and BesE are predicted to be non-haem Fe/α-ketoglutarate (αKG)-dependent oxidases. Finally, BesF is a putative EamA-like transporter, homologues of which have been implicated in amino acid efflux. Searching the non-redundant protein database for additional instances of these genes, we identified five additional *Streptomyces* spp. that had not been reported to produce terminal-alkyne amino acids but that also possessed the genes that encode BesABCD (Extended Data Fig. 2a, Supplementary Table 2). Three of these newly identified strains were found to produce either L-propargylglycine (Pra) or βes in cell culture (Extended Data Fig. 2b). On the basis of these results and the comparison of gene clusters, we propose that BesABCD are responsible for Pra biosynthesis and form the core of the cluster and that BesE is related to the conversion of Pra to βes by hydroxylation at the β-position.

The biosynthetic genes *besA–besE* were then targeted for deletion from the genome to generate the *S. cattleya* Δ*besA*, Δ*besB*, Δ*besC*, Δ*besD* and Δ*besE* knockout strains. Amino acid analysis of the supernatant and intracellular metabolites of these knockouts showed that the Δ*besB*, Δ*besC* and Δ*besD* strains no longer produced detectable amounts of Pra or βes (Fig. 1e). Complementation studies further support the assignment of terminal-alkyne amino acid production to the bes gene cluster (Extended Data Fig. 2c). By comparison, the Δ*besA* strain showed lowered yields of both products, and the Δ*besE* strain still produced Pra but not βes. These results support the hypothesis that *besB*, *besC* and *besD* are crucial for formation of Pra, and suggest that BesE is the hydroxylase that converts Pra to βes. To identify additional intermediates in this pathway, we performed an untargeted comparative metabolomics experiment on the intracellular metabolite extracts of *S. cattleya* wild-type, Δ*besA*, Δ*besB*, Δ*besC*, Δ*besD* and Δ*besE* strains using high-resolution liquid chromatography with mass spectrometry (LC–MS). The resulting mass spectra were processed

[1]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [3]Department of Chemistry, University of California, Berkeley, CA, USA. [4]QB3 Institute, University of California, Berkeley, CA, USA. *e-mail: mcchang@berkeley.edu



d

Gene	aa	Homologue	%ID/%S	Proposed function
besF	297	EamA family transporter (*Streptomyces xylophagus*)	69/78	Amino acid transporter
besA	511	Carboxylate-amine ligase (*Cylindrospermopsis raciborskii*)	32/50	Amino acid ligase
besB	465	Cystathionine beta-lyase (*Opitutaceae* bacterium TSB47)	42/54	Gamma-lyase
besC	252	Pyrroloquinoline-quinone synthase (*Marinobacter pelagius*)	44/65	Oxidase
besD	268	ArpA protein (*Pseudomonas azotoformans*)	56/71	Halogenase
besE	257	Proline hydroxylase (*Streptomyces misionensis*)	54/66	Hydroxylase

Fig. 1 | Identification of the β-ethynylserine biosynthetic gene cluster. **a**, Examples of natural products with internal alkynes from terpenoid or fatty acid sources. **b**, Terminal-alkyne amino acids from *Streptomyces* spp. **c**, Natural products that contain a terminal alkyne formed by fatty acid desaturation, carmabin A and jamaicamide B. **d**, Genes in the βes biosynthetic gene cluster with nearest homologue, per cent identity and similarity to nearest homologue (%ID/%S), and proposed function. **e**, Summary of terminal alkyne production phenotypes of knockouts of each gene in the βes biosynthetic gene cluster from *S. cattleya*. aa, amino acids.

using XCMS[27] for feature detection to generate an initial set of about 1,000 lead compounds with high statistical significance ($P < 0.05$) and high fold change (>2) between wild-type and knockout strains. By focusing on the metabolites that accounted for the highest degree of variance, we were able to identify a set of putative intermediates and precursors for βes biosynthesis (Fig. 2a, Extended Data Fig. 3a). On the basis of these findings, we hypothesized that βes biosynthesis is initiated by chlorination of L-lysine (**1**) at the C_γ position by the BesD halogenase (Fig. 2b). The resulting 4-Cl-lysine (**2**) is then subject to an unusual oxidative cleavage by the BesC oxidase to form 4-Cl-allylglycine (**3**). The PLP-dependent enzyme BesB can next catalyse γ-elimination of chloride from this intermediate followed by isomerization to form the first terminal-alkyne product, Pra (**4**). We propose that Pra is ligated to glutamate by BesA to form a γ-glutamyl-dipeptide (**5**), which serves as the substrate for the BesE hydroxylase to produce the corresponding βes dipeptide (**6**) in the last step. As no other conserved gene was found in these clusters, it is possible that a non-pathway-specific peptidyl-transglutaminase or amide hydrolase is responsible for the release of free Pra and βes.

To reconstitute the biosynthetic pathway in vitro, we first characterized each individual enzyme of the βes pathway (Extended Data Fig. 3b, c). BesD appears to contain the characteristic HXG motif that is required for halide coordination to the active-site iron in the Fe/αKG-dependent halogenase family, rather than the HXD/E motif of the related family of hydroxylases[28] (Extended Data Fig. 4a). Indeed, when L-lysine was incubated with purified BesD (Extended Data Fig. 4b–d), the major product of the reaction exhibited a characteristic Cl isotopic pattern and could be assigned as Cl-lysine (Fig. 3a). Using nuclear magentic resonance (NMR) analysis of the methyl esterified product, we found that the chlorination occurred at the C_γ (Extended Data Fig. 4e–h). Owing to the inherent instability of 4-Cl-lysine (**2**) towards intramolecular cyclization, the lactam, lactone and 4-OH-lysine shunt products were also observed (Extended Data Fig. 5).

In addition to 4-Cl-lysine, we identified a second chlorinated metabolite from comparative metabolomics studies, which had a mass consistent with a loss of $-CH_5N$ from 4-Cl-lysine. This compound accumulated in the Δ*besB* strain but was not observed in the Δ*besC* strain (Fig. 2a), which suggests that BesC—a putative non-haem dinuclear Fe enzyme (Extended Data Fig. 6a)—is the most likely candidate for its production. We thus propose that BesC catalyses an unusual oxidative C–C bond-cleaving reaction to form a five-carbon terminal-alkene amino acid, 4-Cl-allylglycine (**3**) with eventual release of the methylamine fragment as ammonia and formaldehyde (Fig. 3b). To further characterize BesC, we tested the ability of the purified enzyme to produce 4-Cl-allylglycine (**3**) in vitro from purified 4-Cl-lysine or a coupled reaction with BesD and L-lysine (Extended Data Fig. 6b). The BesC product was confirmed to be **3** by comparison to a commercial standard, and exhibited the characteristic Cl isotope pattern (Fig. 3c, Extended Data Fig. 6b). Both release of ammonia (Extended Data Fig. 6c–h) and formaldehyde (Extended Data Fig. 7) were found to be stoichiometric with terminal-alkene amino acid formation. Together, the data show that BesC catalyses alkene formation with release of formaldehyde and ammonia as the final co-products of the reaction.

We next turned our attention to formation of the terminal-alkyne functional group of Pra. On the basis of the accumulation of **3** upon deletion of the *besB* gene (Fig. 2a), we assigned the PLP-dependent BesB enzyme as the acetylenase that catalyses the γ-elimination of chloride from **3** to achieve substrate desaturation. In vitro assays with purified BesB show both that Pra can be produced either directly from a 4-Cl-allylglycine standard or by a coupled assay with BesC and BesD, and that the presence of a γ-halogen as a leaving group is required (Extended Data Fig. 8a, b). As an initial probe of BesB mechanism, we used high-resolution mass spectroscopy to measure deuterium exchange along the backbone of the substrate. Using 4-Cl-allylglycine (**3**) as a substrate in D_2O resulted in formation of $[M+2D]^+$ Pra, which suggests that at least two deprotonation–protonation events occur during the course of BesB catalysis (Fig. 3d). This model is consistent with a canonical cystathionine γ-synthase mechanism in which C_α-deprotonation leads to eventual formation of a 4-Cl-allylglycine-ketimine intermediate[29] (Fig. 3e, Extended Data Fig. 8c). A second deprotonation on the amino acid at C_β would initiate the elimination of chloride to form an allene intermediate, which could subsequently isomerize to the terminal alkyne before eventual conversion back to aldimine and release from PLP. This proposed mechanism follows from previous studies of Pra as a mechanism-based inhibitor of PLP enzymes, based on its ability to isomerize to the corresponding allene[30–32].

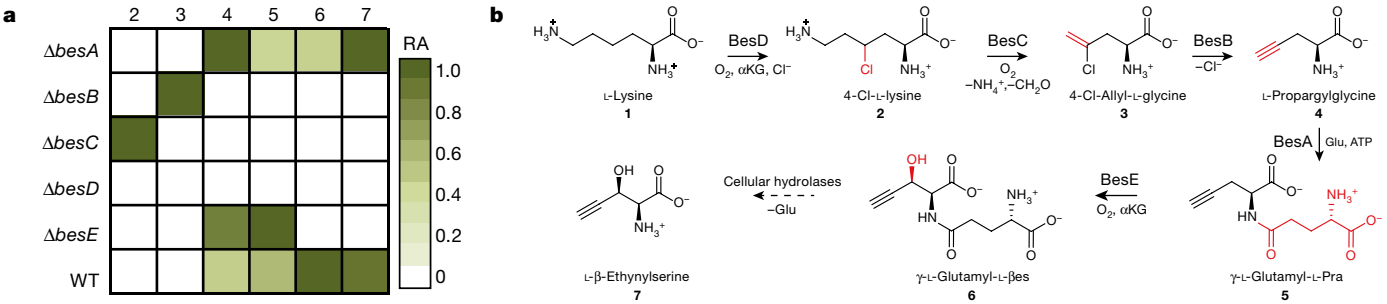


Fig. 2 | Biosynthetic pathway for production of βes from L-lysine. **a**, Relative abundance of pathway amino acids found in *S. cattleya* knockout strains and wild type (WT), by average integrated peak area of each compound ($n = 5$ biological replicates) and normalized by relative abundance (RA) within each species (green, more abundant; white, less abundant). **b**, Putative pathway for biosynthesis of βes from L-lysine.

From the terminal-alkyne amino acid profile of the different streptomycete producers along with the Δ*besE* knockout strain (Fig. 2a, Extended Data Fig. 2), it seemed reasonable to believe that hydroxylation of Pra by BesE yields βes (**7**). However, no βes was observed when purified BesE was incubated with Pra in the presence of Fe, ascorbate and αKG. We therefore looked for an alternative substrate for BesE, which we were able to identify as a glutamate-Pra dipeptide (**5**) through comparative metabolomics of *S. cattleya* Δ*besE* and wild-type strains (Extended Data Fig. 9a). This compound was also present at a low concentration in the Δ*besA* knockout. Upon fragmentation, characteristic fragments consistent with amide bond cleavage were detected (Extended Data Fig. 9b). Comparison of the products of purified BesA and a commercial γ-glutamyl transpeptidase further support assignment of this product as γ-Glu-Pra (Extended Data Fig. 9c). Additional biochemical characterization indicates that BesA is selective for Pra over norvaline, allylglycine and the standard proteinogenic amino acids (Extended Data Fig. 9d, e). Notably, *besA* is conserved among these Pra biosynthetic gene clusters, which suggests that Glu-Pra formation has a physiological function—perhaps to reduce side reactions of Pra[31,32] or to facilitate transport. Glu-Pra was then shown to be the substrate for purified BesE, which was competent for its hydroxylation to yield γ-L-glutamyl-L-βes (Glu-βes) (**6**) (Extended Data Fig. 9f), giving rise to the question of how this enzyme can create a stable radical so close to a terminal-alkyne without rearrangement. Elucidation of the final

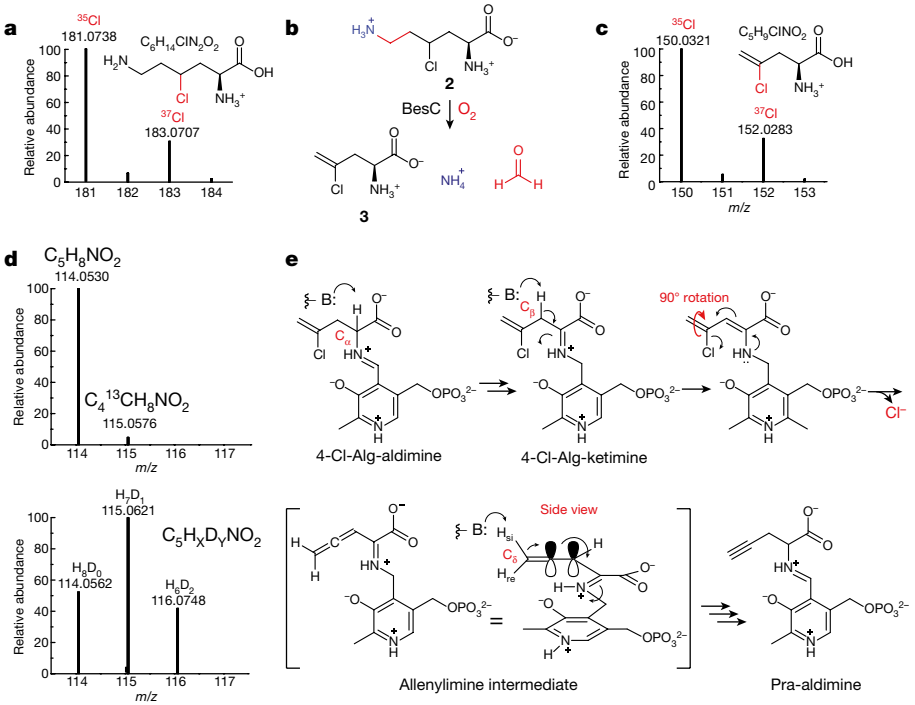


Fig. 3 | Characterization of BesD, BesC and BesB. a, Mass spectrum of product formed when BesD is incubated with L-lysine and chloride showing the characteristic chlorine isotope pattern. Representative of at least three independent experimental replicates. **b**, Reaction catalysed by BesC, forming ammonia and formaldehyde as co-products from oxidative cleavage between C_δ and C_ε. **c**, Mass spectrum of product formed when BesC is incubated with 4-Cl-lysine. Representative of at least three independent experimental replicates. **d**, Mass spectra showing deuterium incorporation at two positions of Pra (**4**) when BesB reaction is carried out in D_2O compared to H_2O. Top, extracted ions for the BesB reaction run in H_2O. Expected $m/z = C_5H_8NO_2$ 114.0550 (100%), $C_4{}^{13}CH_8NO_2$ 115.0580 (5.9%); observed m/z: 114.0530 (100%), 115.0576 (4.8%). Bottom, extracted ions for the BesB reaction run in D_2O. Observed m/z = 114.0562 (52.3%), 115.0576 (3.2%, not labelled for clarity), 115.0621 (100%), 116.0748 (41.8%). Representative of at least three independent experimental replicates. **e**, Putative mechanism for BesB-catalysed formation of Pra (**4**) from 4-Cl-allylglycine (**3**). Formation of a covalent external aldimine adduct between PLP and 4-Cl-allylglycine followed by C_α deprotonation and protonation of the quinonoid intermediate (not shown) is proposed to lead to formation of the initial ketimine intermediate. C_β deprotonation results in elimination of chloride upon 90° bond rotation to form an allene intermediate. Subsequent C_δ deprotonation could then initiate isomerization of the allene to form the terminal alkyne. Following deprotonation of the PLP-bound alkyne to regenerate the quinonoid form (not shown), reprotonation of C_α reforms the external aldimine, from which Pra can be released as the free amino acid (not shown).

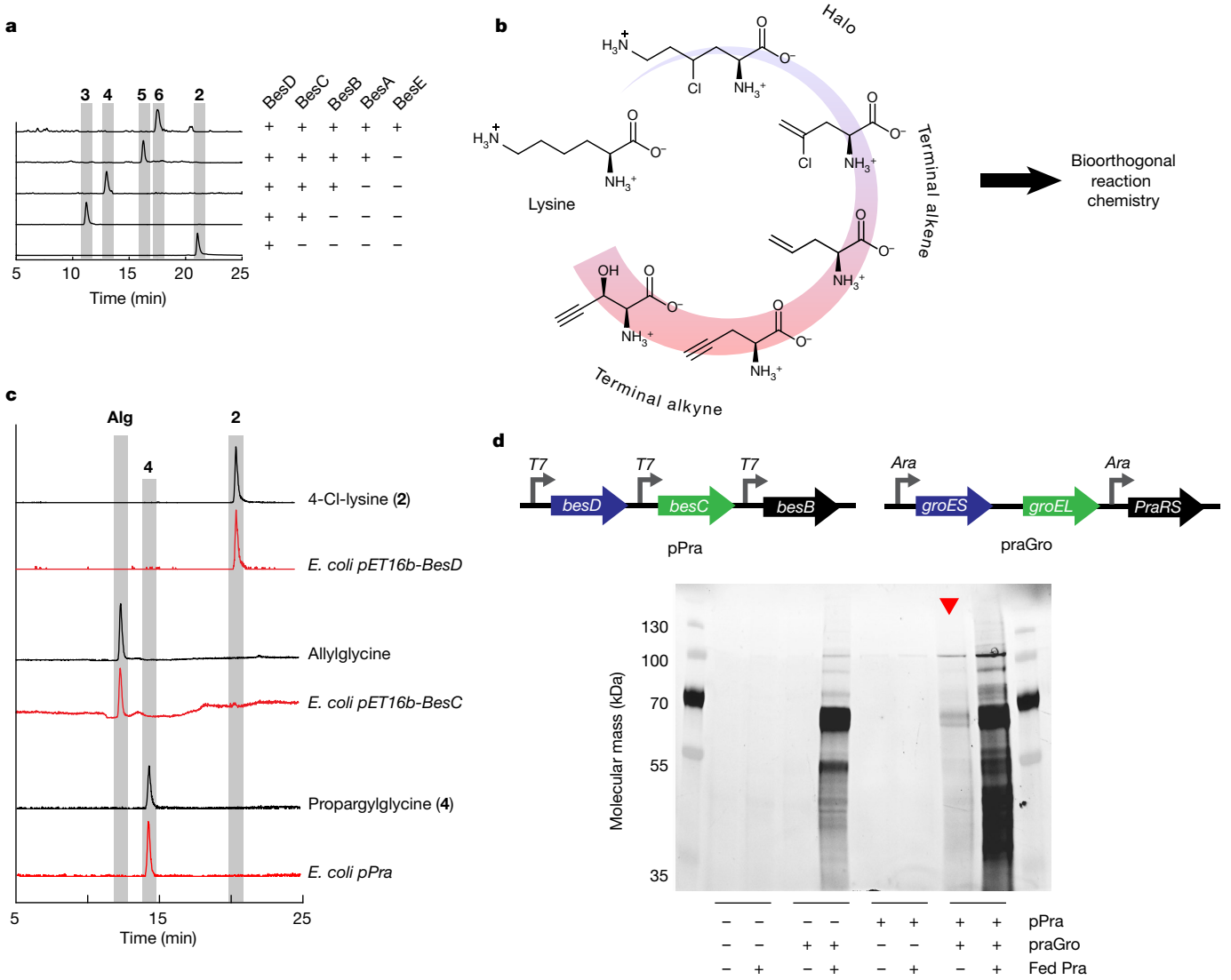


Fig. 4 | In vitro and in vivo reconstitution of βes biosynthetic pathway. **a**, In vitro reconstitution of βes biosynthesis. Extracted ion chromatogram for each pathway intermediate in the presence of corresponding enzymes required to catalyse each step (**2**, $m/z = 181.0738$; **3**, $m/z = 150.0316$; **4**, $m/z = 114.550$; **5**, $m/z = 243.0975$; **6**, $m/z = 259.0925$). Peaks are normalized for each trace to the most-abundant ion. Chromatograms shown are representative of at least three independent experimental replicates. **b**, The βes pathway enables the enzymatic synthesis of amino acids with a range of bioorthogonal handles that can be used for downstream reaction chemistry. **c**, Extracted ion chromatograms showing in vivo production of 4-Cl-lysine (**2**, $m/z = 181.0738$), allylglycine (Alg, $m/z = 116.0706$), and Pra (**4**, $m/z = 114.550$) in *E. coli* BL-21 Star (DE3). Amino acid yields of 50–100 μM are reached within 48 h. Chromatograms shown are representative of at least three independent experimental replicates. **d**, Co-expression of Pra biosynthetic genes (pPra) along with PraRS, GroEL and GroES (praGro) in a heterologous host leads to proteome incorporation of Pra. *E. coli* B834(DE3) pPra praGro was cultured for 2 d along with the respective empty vector controls and Pra feeding controls. Cell lysate was then reacted with Tamra-azide dye by CuAAC before analysis by SDS–PAGE in fluorescence mode ($\lambda_{Ex} = 546$ nm, $\lambda_{Em} = 565$ nm). Red arrowhead indicates *E. coli* B834(DE3) pPra praGro, showing incorporation of endogenously synthesized Pra. Ladder consists of 70-kDa marker that can be visualized by fluorescence overlaid with the Coomassie-stained ladder from the same gel for reference. Gel shown is representative of at least five independent experimental replicates.

committed step enabled us to fully reconstitute the biosynthesis of Glu-βes (**6**) in vitro (Fig. 4a).

The identification of the βes biosynthetic pathway yields the opportunity to produce several non-standard amino acids with potential for downstream bioorthogonal reactions (Fig. 4b). Using *Escherichia coli* as a heterologous host, we were able to produce 4-Cl-lysine (**2**), allylglycine, and Pra (**4**) with the expression of one to three βes cluster genes (Fig. 4c). To show that these endogenously produced amino acids could be translated into proteins, we adapted a method for the residue-specific replacement of Met residues with Pra using an engineered aminoacyl-tRNA synthase (PraRS)[23] that uses a two-plasmid system for the co-expression of the BesB, BesC and BesD proteins for Pra production (pPra) with PraRS and GroEL–GroES (praGro) (Extended Data Fig. 10a, b). *E. coli* B834(DE3) pPra praGro cells were cultured and their soluble protein extracts were derivatized with a Tamra-azide fluorescent dye using the CuAAC reaction. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE), visualized by fluorescence imaging, showed that alkyne was incorporated into the proteome through endogenous Pra biosynthesis, and was supported by proteomic analysis of the cell extracts (Fig. 4d, Extended Data Fig. 10c–e).

In conclusion, we have identified and characterized a biosynthetic pathway for the production of halo, terminal-alkene and terminal-alkyne amino acids, which includes several unusual chemical transformations and intermediates. The subsequent introduction of these functional groups into peptides and proteins either site- or residue-specifically opens opportunities to use a variety of downstream

bioorthogonal chemistries, including alkyne-mediated CuAAC click reactions, olefin metathesis or polymerization, and halogen-dependent reactions for manipulating the structure, function and physical properties of products.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1020-y.

Received: 9 February 2018; Accepted: 6 February 2019;
Published online 13 March 2019.

1. Prescher, J. A. & Bertozzi, C. R. Chemistry in living systems. *Nat. Chem. Biol.* **1**, 13–21 (2005).
2. Li, J. & Chen, P. R. Development and application of bond cleavage reactions in bioorthogonal chemistry. *Nat. Chem. Biol.* **12**, 129–137 (2016).
3. Chin, J. W. et al. An expanded eukaryotic genetic code. *Science* **301**, 964–967 (2003).
4. Wright, T. H. et al. Posttranslational mutagenesis: a chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).
5. O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T. G. & Murphy, C. D. Biochemistry: biosynthesis of an organofluorine molecule. *Nature* **416**, 279 (2002).
6. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194 (2005).
7. Agarwal, V. et al. Enzymatic halogenation and dehalogenation reactions: Pervasive and mechanistically diverse. *Chem. Rev.* **117**, 5619–5674 (2017).
8. Agarwal, V. et al. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol.* **10**, 640–647 (2014).
9. Cicchillo, R. M. et al. An unusual carbon–carbon bond cleavage reaction during phosphinothricin biosynthesis. *Nature* **459**, 871–874 (2009).
10. Liu, W., Christenson, S. D., Standage, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297**, 1170–1173 (2002).
11. Ahlert, J. et al. The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**, 1173–1176 (2002).
12. Jensen, N. B. et al. Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects. *Nat. Commun.* **2**, 273 (2011).
13. Sugai, Y., Katsuyama, Y. & Ohnishi, Y. A nitrous acid biosynthetic pathway for diazo group formation in bacteria. *Nat. Chem. Biol.* **12**, 73–75 (2016).
14. Rui, Z. et al. Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proc. Natl Acad. Sci. USA* **111**, 18237–18242 (2014).
15. Zhu, X., Liu, J. & Zhang, W. *De novo* biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol.* **11**, 115–120 (2015).
16. Haritos, V. S. et al. The convergent evolution of defensive polyacetylenic fatty acid biosynthesis genes in soldier beetles. *Nat. Commun.* **3**, 1150 (2012).
17. Scrimgeour, C. M. Natural acetylenic and olefinic compounds, excluding marine natural products. *Aliphatic Relat. Nat. Prod. Chem.* **2**, 1–19 (1979).
18. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).
19. Minto, R. E. & Blacklock, B. J. Biosynthesis and function of polyacetylenes and allied natural products. *Prog. Lipid Res.* **47**, 233–306 (2008).
20. Potgieter, H. C., Vermeulen, N. M. J., Potgieter, D. J. J. & Strauss, H. F. A toxic amino acid, 2(*S*)3(*R*)-2-amino-3-hydroxypent-4-ynoic acid from the fungus *Sclerotium rolfsii*. *Phytochemistry* **16**, 1757–1759 (1977).
21. Sanada, M., Miyano, T. & Iwadare, S. β-Ethynylserine, an antimetabolite of L-threonine, from *Streptomyces cattleya*. *J. Antibiot. (Tokyo)* **39**, 304–305 (1986).
22. Scannell, J. P., Pruess, D. L., Demny, T. C., Weiss, F. & Williams, T. Antimetabolites produced by microorganisms. II. L-2-amino-4-pentynoic acid. *J. Antibiot. (Tokyo)* **24**, 239–244 (1971).
23. Truong, F., Yoo, T. H., Lampo, T. J. & Tirrell, D. A. Two-strain, cell-selective protein labeling in mixed bacterial cultures. *J. Am. Chem. Soc.* **134**, 8551–8556 (2012).
24. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114**, 4764–4806 (2014).
25. Zhu, X., Su, M., Manickam, K. & Zhang, W. Bacterial genome mining of enzymatic tools for alkyne biosynthesis. *ACS Chem. Biol.* **10**, 2785–2793 (2015).
26. Shin-Ichi, H. Amino acids from mushrooms. *Prog. Chem. Org. Nat. Prod.* **59**,117–140 (1992).
27. Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* **84**, 5035–5039 (2012).
28. Wong, S. D. et al. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320–323 (2013).
29. Brzović, P., Holbrook, E. L., Greene, R. C. & Dunn, M. F. Reaction mechanism of *Escherichia coli* cystathionine γ-synthase: direct evidence for a pyridoxamine derivative of vinylglyoxylate as a key intermediate in pyridoxal phosphate dependent γ-elimination and γ-replacement reactions. *Biochemistry* **29**, 442–451 (1990).
30. Sun, Q. et al. Structural basis for the inhibition mechanism of human cystathionine γ-lyase, an enzyme responsible for the production of H_2S. *J. Biol. Chem.* **284**, 3076–3085 (2009).
31. Marcotte, P. & Walsh, C. Vinylglycine and propargylglycine: complementary suicide substrates for L-amino acid oxidase and D-amino acid oxidase. *Biochemistry* **15**, 3070–3076 (1976).
32. Abeles, R. H. & Walsh, C. T. Acetylenic enzyme inactivators. Inactivation of γ-cystathionase, *in vitro* and *in vivo*, by propargylglycine. *J. Am. Chem. Soc.* **95**, 6124–6125 (1973).

Acknowledgements We thank W. Zhang, D. Nomura and D. Berkowitz for discussions and advice. J.A.M. acknowledges the support of a UC Berkeley Chancellor's Fellowship, Howard Hughes Medical Institute Gilliam Fellowship, and National Institutes of Health NRSA Training Grant (1 T32 GMO66698). M.E.N. acknowledges the support of a National Science Foundation Graduate Research Fellowship. This work was funded by generous support from the National Science Foundation (CHE-1710588). The College of Chemistry NMR Facility at U.C. Berkeley is supported in part by the National Institutes of Health (1S10RR023679-01 and S10 RR16634-01).

Reviewer information *Nature* thanks Gonçalo Bernardes, Rebecca Goss and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.A.M. carried out gene disruption and in vivo reconstitution experiments with assistance from M.C.I. J.A.M. and M.E.N. carried out experiments for comparative metabolomics and enzyme characterization. J.G.P. planned and carried out NMR experiments. J.A.M., M.E.N., C.-I.L. and M.C.Y.C. planned experiments. J.A.M., M.E.N. and M.C.Y.C. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1020-y.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1020-y.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.C.Y.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Commercial materials. Commercial materials used and their sources are provided in Supplementary Information.

Bacterial strains. *S. cattleya* NRRL 8057 (ATCC 35852) was purchased from the American Tissue Type Collection. *Streptomyces lavenduligriseus* (NRRL B-3173), *Streptomyces catenulae* (NRRL B-2342), *Streptomyces achromogenes* (NRRL B-2120), *Streptomyces* sp. (NRRL S-1448) and *Streptomyces* sp. (NRRL S-31) were provided by Agricultural Research Service Culture Collection: Bacterial Foodborne Pathogens and Mycology Research Unit (National Center for Agricultural Utilization Research). *E. coli* DH10B-T1^{R} was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production of BesA, BesB, BesC, BesD and BesE. *E. coli* GM272 containing the non-transmissible, *oriT*-mobilizing plasmid pUZ8002 was used for conjugative plasmid transfer into *S. cattleya*. *E. coli* B834(DE3) was used for in vivo studies involving incorporation of Pra into proteins using the engineered pathway.

Comparative genomics and clustering analysis. Protein genome sequences and associated protein lists for 27 *Streptomyces* spp. were downloaded from the NCBI database (NIH). Three species of Streptomycetes have been reported to produce terminal-alkyne amino acids: *S. cattleya, S. catenulae* and *Streptomyces* sp. HLR-599A. However, *S.* sp. HLR-599A is held in a private collection and its genome has not been sequenced; therefore, it was not included in this study. An all-by-all protein pBLAST was performed between *S. cattleya* (chromosome: NC_015876, plasmid: NC_015875.1) and *S. catenulae* (NZ_JODY00000000.1) as the in-group and *S. cattleya* + *Streptomyces hygroscopicus* (NZ_CP018627.1), *Streptomyces rimosus* (NZ_JNYR00000000.1), *Streptomyces griseus* (NC_010572.1), *Streptomyces albulus* (chromosome: NZ_CP007574.1, plasmid: NZ_CP007575.1), *Streptomyces filamentosus* (NZ_ABYX00000000.2), *Streptomyces viridochromogenes* (NZ_ACEZ00000000.1), *Streptomyces lavendulae* (chromosome: NZ_CP024985.1, plasmid: NZ_CP024986.1), *Streptomyces chartreusis* (NZ_LT963352.1), *Streptomyces ipomoeae* (NZ_AEJC00000000.1), *Streptomyces sviceus* (NZ_CM000951.1), *Streptomyces pristinaespiralis* (NZ_CP011340.1), *Streptomyces bingchenggensis* (NC_016582.1), *Streptomyces ghanaensis* (NZ_ABYA00000000.1), *Streptomyces coelicolor* (NC_003888.3), *Streptomyces clavuligerus* (chromosome: NZ_CP016559.1, plasmid: NZ_CP016560.1), *Streptomyces ambofaciens* (NZ_CP012949.1), *Streptomyces griseoplanus* (NZ_LIQR00000000.1), *Streptomyces scabiei* (NC_013929.1), *Streptomyces avermitilis* (chromosome: NC_003155.5, plasmid: NC_004719.1), *Streptomyces canus* (NZ_LMWO00000000.1), *Streptomyces antibioticus* (NZ_CM007717.1), *Streptomyces roseochromogenus* (NZ_CM002285.1), *Streptomyces cyaneogriseus* (NC_CP010849.1), *Streptomyces prunicolor* (NZ_BARF00000000.1) and *Streptomyces lydicus* (NZ_CP017157.1) as the out-group. Strict adjacency criteria, within two coding sequences of the best homologue, were used to determine whether proteins were clustered in both organisms. Proteins were clustered and visualized using Cytoscape. Proteins that had a lower *E*-value in the in-group were marked unique, whereas proteins with a lower *E*-value in the out-group were marked common. Of the initial ~650 clusters of proteins that were greater than three proteins in size, only four were unique between *S. cattleya* and *S. catenulae*. The pipeline and tutorial for comparative clustering analysis are available online as a php script on Github (https://github.com/jamarchand/io).

Cell culture of *Streptomyces* spp. for metabolite analysis. Single colonies of *S. catenulae*, *S.* sp. NRRL S-1448, *S.* sp. NRRL S-31, *S. achromogenes*, *S. lavenduligriseus*, *S. cattleya*, *S. cattleya* Δ*SCAT_0184*::Am^{R}, *S. cattleya* Δ*SCAT_4823*::Am^{R}, *S. cattleya* Δ*SCAT_p1525*::Am^{R}, *S. cattleya* Δ*besA*::Am^{R}, *S. cattleya* Δ*besB*::Am^{R}, *S. cattleya* Δ*besC*::Am^{R}, *S. cattleya* Δ*besD*::Am^{R}, *S. cattleya* Δ*besE*::Am^{R}, *S. cattleya* Δ*besB*::Am^{R} attB$^{\Phi C31}$::ermEp*-*besB*, *S. cattleya* Δ*besC*::Am^{R} attB$^{\Phi C31}$::ermEp*-*besC*, and *S. cattleya* Δ*besD*::Am^{R} attB$^{\Phi C31}$::ermEp*-*besD* were picked from soy-mannitol plates and inoculated into GYM pH 7 (25 ml). After 3 d of growth, cultures were sub-cultured into GYM pH 7 (25 ml), and grown for 7–14 d.

Derivatization and characterization of βes and propargylglycine production in cell culture. Culture supernatants were sampled by removing 1 ml of culture broth and centrifuging at 9,800*g* for 10 min at room temperature. One hundred microlitres of supernatant was added to a solution containing 3-azido-7-hydroxycoumarin (20 μM), THPTA (100 μM) and sodium ascorbate (10 mM), with and without $CuSO_4$ (50 μM) in 400 μl PBS (pH 7). After being incubated for 10 min, aliquots were analysed using an Agilent 1290 ultra-performance liquid chromatography (UPLC) system on a Poroshell 120 SB-Aq column (2.7 μm, 2.1 × 50 mm; Agilent) using a linear gradient from 5% to 95% acetonitrile over 5 min at a flow rate of 0.6 ml/min with 0.1% (v/v) formic acid as the mobile phase. Mass spectra were acquired using an Agilent 6530 quadrupole time of flight (QTOF) system with the following source and acquisition parameters: gas temperature 325 °C; drying gas 10 l/min; nebulizer 45 psig; capillary voltage 3,500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp 750 V; acquisition rate 3 spectra/s; acquisition time 333.3 ms/spectrum.

Extraction of polar metabolites from lysates and supernatants of *Streptomyces* spp. cultures. An aliquot of culture broth (1 ml) was removed and centrifuged at 9,800*g* for 10 min at room temperature to pellet cells. The supernatant was then diluted 1:1 in methanol with 1% (v/v) formic acid and centrifuged at 9,800*g* for 10 min to remove precipitate for LC–MS analysis. Culture pellets were then collected after removal of the supernatant and resuspended by vigorous vortexing in a 50% (v/v) solution of methanol (100 μl methanol and 100 mg wet cell pellet). These solutions were subjected to five freeze-thaw cycles from −80 °C (30 min) to −20 °C (30 min) with vortexing in between. Solutions were then centrifuged at 9,800*g* for 10 min to remove remaining cell debris before LC–MS analysis. Comparative metabolic experiments with *S. cattleya* wild-type, Δ*besA*, Δ*besB*, Δ*besC*, Δ*besD*, and Δ*besE* strains were carried out on cell pellet extracts.

General procedure for high-resolution high-performance LC–MS analysis of polar metabolites. Samples containing polar metabolites were analysed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm, 2.1 × 100 mm; EMD-Millipore) using the following buffers: buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% buffer A over 17 min followed by a linear gradient from 60% to 33% buffer A over 8 min was then applied at a flow rate of 0.2 ml/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: gas temperature 325 °C; drying gas 10 l/min; nebulizer 45 psig; capillary voltage 3,500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp 750 V; acquisition rate 3 spectra/s; acquisition time 333.3 ms/spectrum.

Comparative metabolite profiling of *S. cattleya* wild-type, Δ*besA*, Δ*besB*, Δ*besC*, Δ*besD* and Δ*besE* strains. Raw mass spectra data were converted to mzXML format using Mzconvert. Peak detection was processed on XCMS platform. Data-collection parameters were set to default high-performance liquid chromatography (HPLC) Orbitrap values (centwave feature detection, 0.5 minimum fraction of samples in one group to be a valid group, *P* value thresholds for significant feature 0.05, isotopic p.p.m. error 5, *m/z* absolute error 0.015) except the following: p.p.m. = 20, signal/noise threshold = 2, mzwid = 0.025, and with loess nonlinear retention time alignment (default settings). Lists of *m/z* values, retention times, peak intensities and peak areas were exported to Excel for processing.

In vitro assays of BesD and BesD(G139D). Reactions (50 μl) contained L-lysine·HCl (4 mM), sodium αKG (5 mM), sodium ascorbate (5 mM), and $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) in 100 mM MOPS buffer (pH 7). Chloride was introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Various assay components were omitted for control reactions as noted in the data. Reactions were initiated by addition of wild type or BesD(G139D) (17.6 μM final concentration) and allowed to proceed for 1 h at room temperature before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above. To analyse 4-Cl-lysine instability, samples were quenched at $t = 0$ and $t = 1$ h after removal of the enzyme with Pall Nanosep spin filters (10 kDa molecular weight cut-off (MWCO)).

In vitro assay for BesC. BesC was assayed by either coupling to BesD to produce 4-Cl-lysine in situ or by using purified 4-Cl-lysine. For coupled reactions, reactions (50 μl) contained BesD (2.9 μM), L-lysine·HCl (4 mM), sodium αKG (5 mM), sodium ascorbate (5 mM) and $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) in 100 mM HEPES (pH 7.5). For reactions using purified 4-Cl-lysine, reactions (50 μl) contained 4-Cl-lysine (0.5 mM), sodium ascorbate (5 mM) and $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) in 100 mM HEPES (pH 7.5). Chloride was introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. For production of allylglycine, the reaction was run without addition of BesD and sodium αKG. Various assay components were omitted for control reactions as noted in the data. Reactions were initiated by addition of BesC (15.7 μM final concentration) and incubated for 1 h at room temperature before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid and being centrifuged at 9,800*g* for 10 min. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

In vitro assay for BesB. Reactions (100 μl) contained 4-Cl-allylglycine, 4-Br-allylglycine or allylglycine (1 mM) and PLP (10 μM) in 100 mM HEPES (pH 7.5). Reactions were initiated by the addition of BesB (2.4 μM final concentration) and incubated at room temperature for 2 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

D_2O exchange assay for BesB. BesB (1.2 mg/ml) was exchanged into 50 mM ammonium bicarbonate buffer in 90% D_2O (pH 7.5). 4-Cl-allylglycine was prepared as a 100 mM stock solution in >99% D_2O. Reactions (1 ml) contained 4-Cl-allylglycine (10 mM) and PLP (20 μM) in 50 mM ammonium bicarbonate (pH 7.5 in D_2O). Reactions were initiated by the addition of BesB (4.8 μM final concentration) and incubated for 30 min at room temperature before being

quenched in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Steady-state kinetic characterization of BesA. Reactions (100 μl) contained varying concentrations of L-Pra (0.4–5.0 mM) or L-cysteine (1.25–20 mM), with L-glutamate (10 mM), ATP (2.5 mM), $MgCl_2$ (5 mM), phosphoenol-pyruvate (PEP, 10 mM), NADH (0.3 mM), lactate dehydrogenase (LDH, 10 U/ml) and pyruvate kinase (PK, 10 U/ml) in 100 mM HEPES (pH 7.5). Reactions were initiated by addition of BesA (6.4 μM final concentration). Initial rates were measured by monitoring λ_{340} using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature. Kinetic parameters (k_{cat}, K_M) were determined by fitting the data using Microcal Origin to the equation: $v_o = v_{max}[S]/(K_M + [S])$, in which v is the initial rate and $[S]$ is the substrate concentration. Error bars on graphs represent mean ± s.d. ($n = 3$ technical replicates). Error in k_{cat}/K_M was calculated by propagation of error from the individual kinetic parameters.

In vitro assay for BesA. Reactions (50 μl) contained L-Pra (2 mM), L-glutamate (5 mM), ATP (2.5 mM) and $MgCl_2$ (5 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of BesA (6.4 μM) and incubated at room temperature for 1 h before being quenched in 1.5 volumes of methanol with 1% (v/v) formic acid. Reactions with BesA were compared with the production of γ-Glu-Pra from commercial γ-glutamyl-transpeptidase (γ-GT). γ-GT reactions (50 μl) contained L-Pra (2 mM) and γ-glutamyl-*p*-nitroaniline (5 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of γ-GT (1 U) and allowed to proceed for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

In vitro assay for BesE. BesE was assayed by coupling to either BesA or γ-GT to produce Glu-Pra in situ. With BesA coupling, reactions (50 μl) contained BesA (6.4 μM), L-Pra (2 mM), L-glutamate (5 mM), ATP (2.5 mM), $MgCl_2$ (5 mM), sodium αKG (5 mM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) and sodium ascorbate (1 mM) in 100 mM MOPS buffer (pH 7). With γ-GT coupling, BesA and L-glutamate were replaced with γ-GT (1 U) and γ-glutamyl-*p*-nitroaniline (5 mM). Reactions were initiated by addition of BesE (4.4 μM) and incubated at room temperature for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

In vitro reconstitution of Pra biosynthesis from L-lysine with BesB, BesC and BesD. Reactions (50 μl) contained BesD (2.9 μM), BesC (15.7 μM), BesB (2.4 μM), sodium αKG (5 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) and PLP (10 μM) in 100 mM HEPES (pH 7.5). Chloride was introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Reactions were initiated by addition of L-lysine·HCl (4 mM) and incubated at room temperature for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

In vitro reconstitution of Glu-βes biosynthesis with BesA, BesB, BesC, BesD and BesE. Reactions (50 μl) contained BesD (2.9 μM), BesC (15.7 μM), BesB (2.4 μM), BesA (6.4 μM), BesE (4.4 μM), sodium αKG (5 mM), sodium ascorbate (5 mM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM), PLP (10 μM), L-glutamate (5 mM), ATP (2.5 mM) and $MgCl_2$ (5 mM) in 100 mM HEPES (pH 7.5). Chloride was introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Reactions were initiated by addition of L-lysine·HCl (4 mM) and allowed to proceed for 16 h before being quenched in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis.

In vivo reconstitution of Cl-Lys, allylglycine and Pra biosynthesis in *E. coli*. Single colonies of *E. coli* BL21 Star (DE3) transformed with pET16b-BesD (from *Pseudomonas fluorescens*, WP_016975823), pET16b-BesC (from *P. fluorescens*, WP_080628534) or pPra were inoculated into TB (10 ml, pH 6.5) containing the appropriate antibiotics and grown overnight at 37 °C with shaking at 200 r.p.m. Overnight cultures were inoculated into TB (25 ml, pH 6.5) in a 250-ml baffled shake flask to $OD_{600} = 0.05$ and grown at 37 °C and 200 r.p.m. until the OD_{600} reached 0.6–0.8. Cultures were cooled on ice for 20 min before being induced with isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.2 mM) and then cultured for 72 h at 16 °C and 200 r.p.m. Supernatant samples were then analysed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis.

In vivo incorporation of Pra into the *E. coli* proteome. Single colonies of *E. coli* B834(DE3) transformed with appropriate plasmids (empty plasmid control: pETDUET-1/pGro7, no endogenous Pra control: pETDUET-1/praGro, no PraRS control: pPra/pGro7, or pPra/praGro) were inoculated into LB with appropriate antibiotics (10 ml) and grown overnight at 37 °C with shaking at 200 r.p.m. Overnight cultures were inoculated into LB with the appropriate antibiotics (25 ml) with 0.2% (w/v) L-arabinose in a 250-ml baffled shake flask to $OD_{600} = 0.05$ and grown at 37 °C and 200 r.p.m. until the OD_{600} reached 0.6–0.8. Cultures were cooled on for 20 min before being induced with IPTG (0.2 mM) and then cultured for 48 h at 16 °C and 200 r.p.m. Cells were washed twice with 0.9% (w/v) NaCl by pelleting and resuspension. The cell pellet was then resuspended in M9 medium containing 33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM NH_4Cl, 0.4% glucose, 100 mg/l of the 19 proteinogenic amino acids (−Met), 50 mg/l Met, 1 mM $MgSO_4$, 300 μM $CaCl_2$, and 1 mg/ml thiamine to $OD_{600} = 0.4–0.6$ and cultured for 24–48 h at 16 °C and 200 r.p.m. For positive controls, Pra (0.2 mM) was also added.

Derivatization and visualization of alkyne-labelled proteins with Tamra-azide. Cell cultures (1 ml) were collected and centrifuged at 9,800*g* for 7 min at 4 °C. The resulting cell pellet was resuspended in SoluLyse (150 μl) and incubated for 1 h at room temperature. Samples were centrifuged at 9,800*g* for 5 min and the supernatant (50 μl) was mixed with an assay solution (50 μl) containing Tamra-azide (40 μM), THPTA (100 μM), and sodium ascorbate (10 mM) in PBS (pH 7) with or without $CuSO_4$ (50 μM). The labelling reaction was incubated for 1 h at room temperature. Laemmli buffer was added and samples were heated at 95 °C for 5 min before being analysed by SDS–PAGE (MiniProtean 8–16% TGX gels; BioRad) using a constant current (30 A) and maximum voltage (200 V) for 1.5 h. Gels were imaged for fluorescence using a ChemiDoc MP Imager (Biorad; $\lambda_{ex} = 520–545$ nm, $\lambda_{em} = 577–613$ nm, exposure = 150 s).

Proteomic analysis of *E. coli* with endogenous Pra production. Cell pellets for *E. coli* B834(DE3) pPra praGro and B834(DE3) pETDUET-1 pGro7 were submitted to the UC Davis Proteomics Core at the UC Davis Genome Center for bottom-up proteomic analysis as positive and negative controls, respectively. Total protein was extracted and subjected to tryptic digestion and iodoacetamide treatment before bottom-up proteomic analysis using a Thermo LC-Q Exactive. Following extraction of tandem mass spectra, charge state deconvolution, and deisotoping, MS/MS spectra were analysed using X! Tandem (The GPM; Tandem Alanine 2017.2.1.4). X! Tandem was searched with a fragment-ion mass tolerance of 20 p.p.m. and a parent-ion tolerance of 20 p.p.m. The following modifications were specified as fixed: carbamidomethyl of Cys and Se-Cys (iodoacetamide cap). The following modifications were specified as variable: Met-36 (Met → Pra) of Met and the N terminus, Glu → Pyro-Glu of the N terminus, ammonia loss from the N terminus, Gln → Pyro-Glu of the N terminus, deamidation of Asn and Gln, oxidation of Met and Trp, and dioxidation of Met and Trp. Scaffold 4.8.7 (Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 99.0% probability by the Scaffold Local FDR algorithm. Peptide identifications were also required to exceed specific database search engine thresholds. Protein identifications were accepted if they could be established at greater than 99.0% probability to achieve a false discovery rate of less than 5.0% and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The pipeline and tutorial for comparative clustering analysis used in this study are available online as a php script on Github (https://github.com/jamarchand/io).

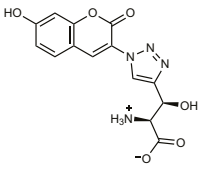
Data availability

Accession codes for genes and proteins in this study are provided in Supplementary Tables. Source data files for figures are provided. Datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

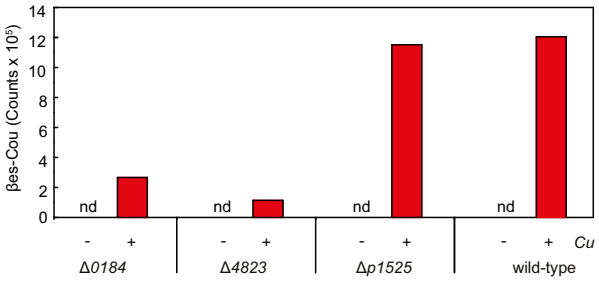
33. Cline, M. S. et al. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols* **2**, 2366–2382 (2007).
34. Merkx, R. et al. Scalable synthesis of γ-thiolysine starting from lysine and a side by side comparison with δ-thiolysine in non-enzymatic ubiquitination. *Chem. Sci. (Camb.)* **4**, 4494 (2013).
35. Schwarzenbacher, R. et al. Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. *J. Biol. Chem.* **279**, 29320–29324 (2004).
36. Guy, J. E. et al. Remote control of regioselectivity in acyl-acyl carrier protein-desaturases. *Proc. Natl Acad. Sci. USA* **108**, 16594–16599 (2011).
37. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).
38. McCune, C. D. et al. Synthesis and deployment of an elusive fluorovinyl cation equivalent: Access to quaternary α-(1′-fluoro)vinyl amino acids as potential PLP enzyme inactivators. *J. Am. Chem. Soc.* **139**, 14077–14089 (2017).
39. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).

A

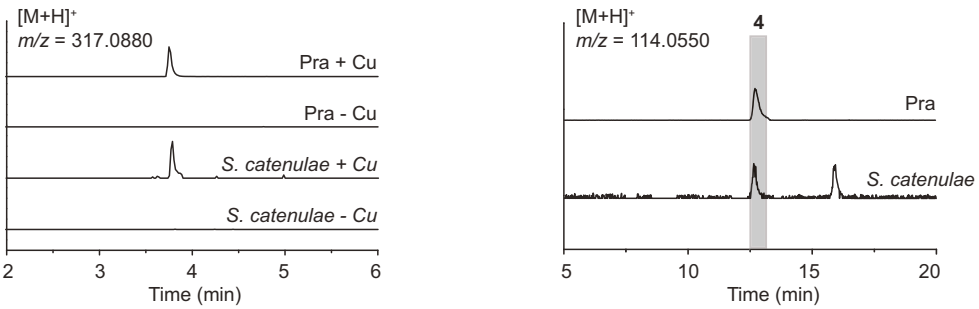
Gene	Homologue	%ID/%S	aa
SCAT_0184	Fatty acid desaturase [*Streptomyces uncialis*]	50/64	346
SCAT_2136	Fatty acid desaturase [*Streptomyces megasporus*]	83/91	332
SCAT_4823	Acyl-CoA desaturase [*Streptomyces* sp. NRRL S-813]	98/99	372
SCAT_p1525	ΔFatty acid desaturase [*Streptomyces* sp. IMTB 2501]	86/90	361



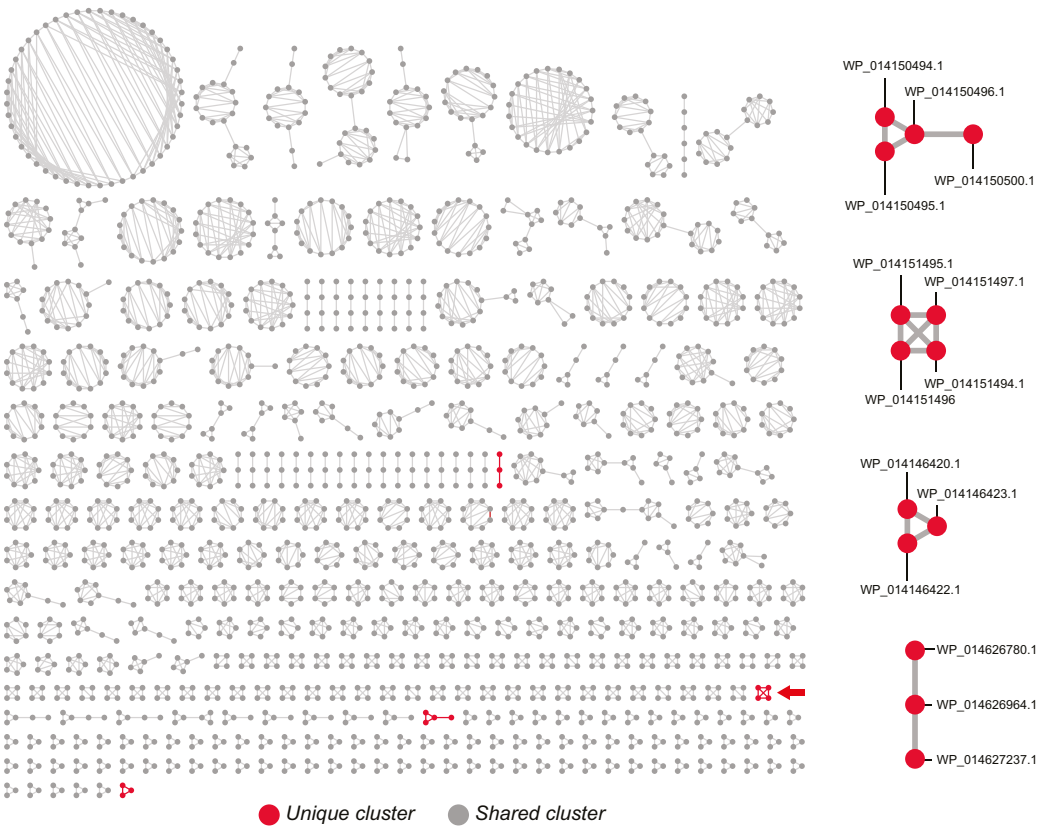
βes-Cou
$[M+H]^+$
m/z = 333.0831



B

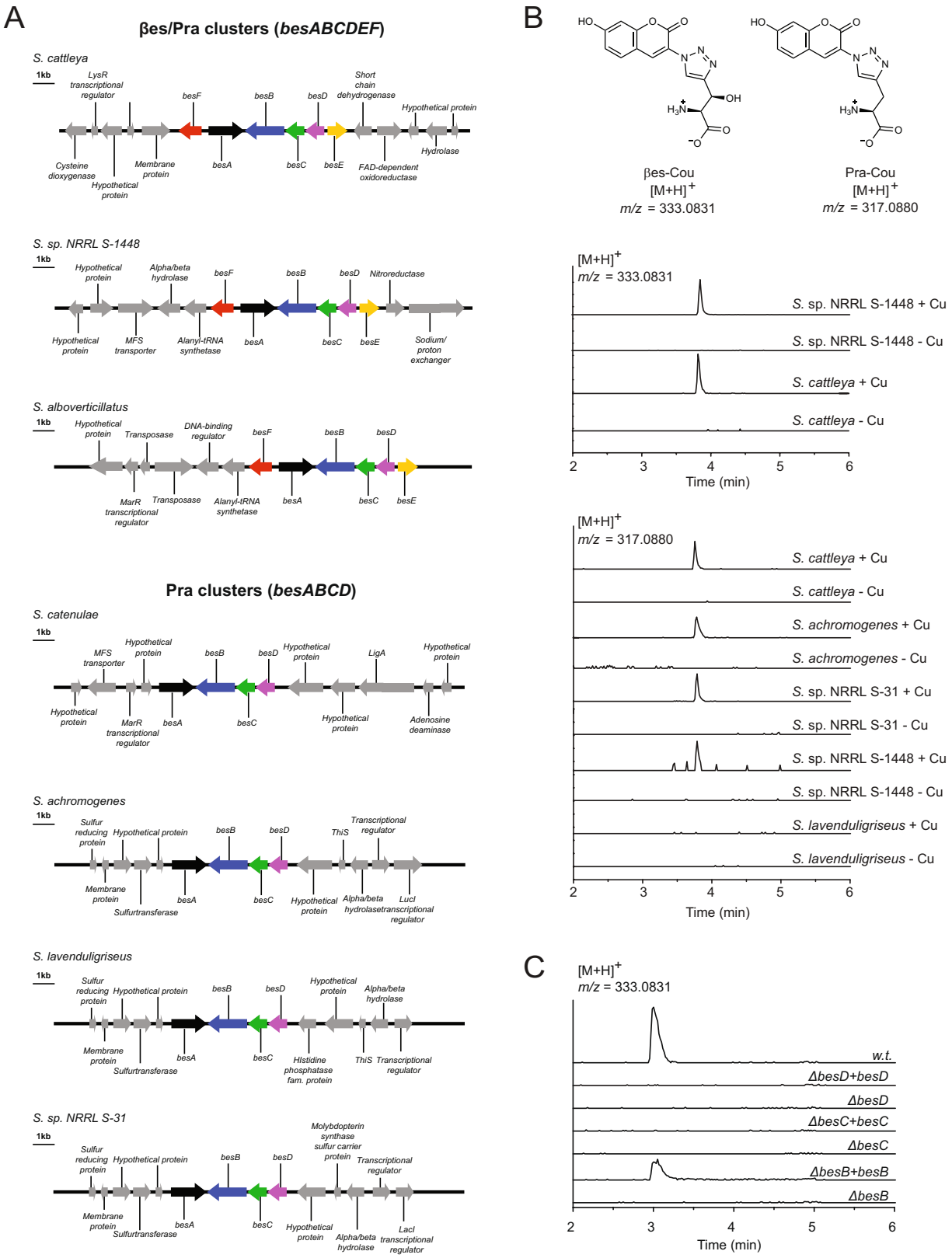


C



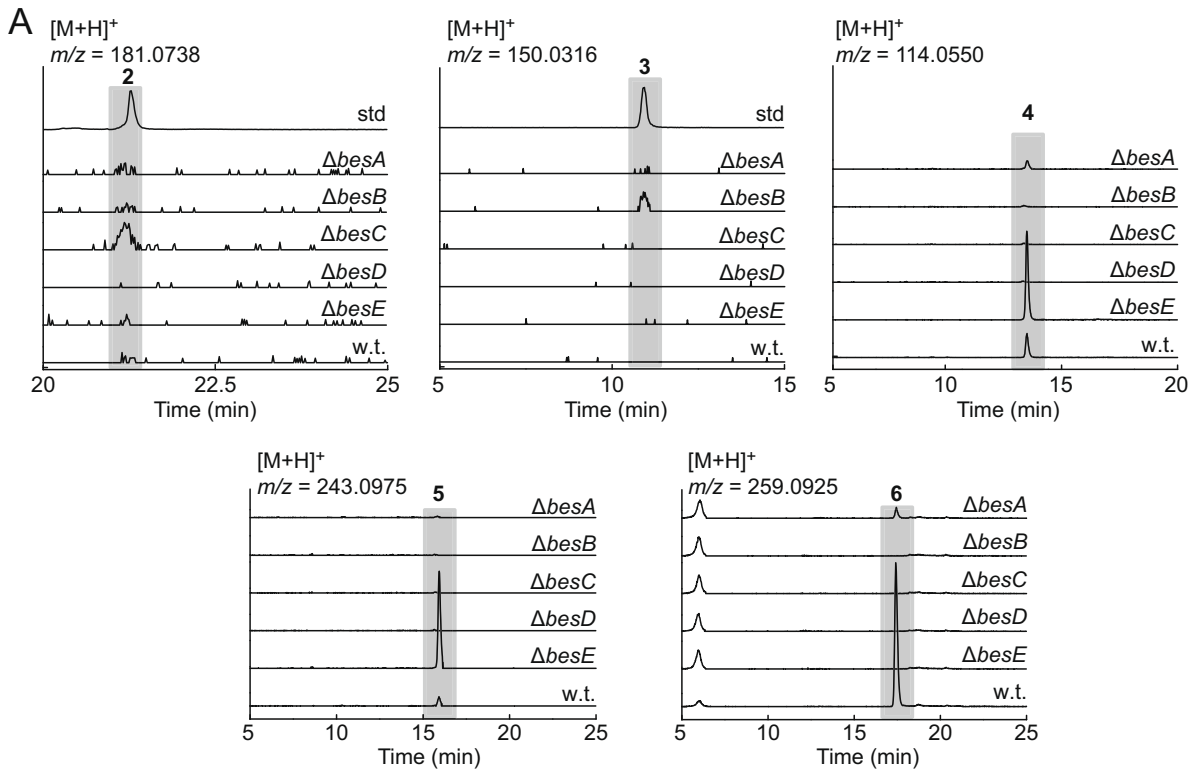
Extended Data Fig. 1 | See next page for caption.

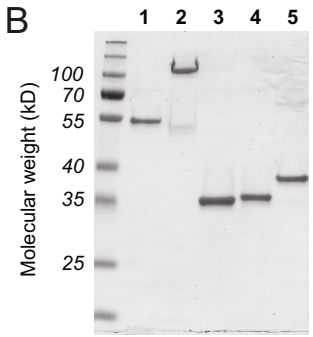
Extended Data Fig. 1 | Genetic and bioinformatic analysis of terminal-alkyne amino acid production in *S. cattleya*. **a**, Four fatty acid desaturases were identified in the genome of *S. cattleya* and targeted for deletion. Per cent identity (%ID) and per cent similarity (%S) compared to the closest sequenced homologue are shown with the host in brackets. Number of amino acids in protein also tabulated. *SCAT_0184* is adjacent to type I polyketide synthase and may be part of a biosynthetic cluster. *SCAT_p1525* may also have a metabolic role, although its genomic context is not highly informative. *SCAT_2136* and *SCAT_4823* are clustered with proteins thought to be involved in primary metabolism and rRNA transcription. We were unable to identify any deletion strains of *SCAT_2136*, suggesting that *SCAT_2136* may be an essential fatty acid desaturase. Knockouts of *SCAT_0184*, *SCAT_4823*, and *SCAT_p1525* were found to still produce βes by CuAAC derivatization with 3-azido-7-hydroxy-coumarin followed by liquid chromatography and quadrupole time of flight (LC–QTOF) monitoring of the βes-coumarin CuAAC product (βes-Cou). The reaction in which Cu(I) is omitted provides a negative control because Cu(I) is required for the cycloaddition to occur. **b**, Extracted ion chromatograms of the culture supernatant of *S. catenulae* with and without CuAAC derivatization, monitoring either the Pra-coumarin adduct (Pra-Cou; $m/z = 317.0880$) or Pra ($m/z = 114.0550$) compared to a synthetic Pra standard using LC–QTOF. Chromatograms shown are representative of at least three independent experimental replicates. **c**, We generated a list of approximately 650 gene clusters (>2 genes) that were conserved between both species. Clusters were generated using Cytoscape[33]. Nodes are individual genes and edges connect genes that are within two protein-coding regions of one another. Gene clusters comprise of connected nodes. Connected grey nodes are gene clusters also found in any of the other 26 *Streptomyces* spp. used in the analysis. Connected red nodes are gene clusters found to be unique between *S. cattleya* and *S. catenulae*. We then curated this list by removing gene clusters that are more closely related between *S. cattleya* and 26 other *Streptomyces* spp., narrowing down the number of conserved unique clusters to just 4. One of these unique clusters contains member nodes WP_014151494.1, WP_014151495.1, WP_014151496.1 and WP_014151497.1, which composes the putative bes gene cluster (indicated with a red arrow).

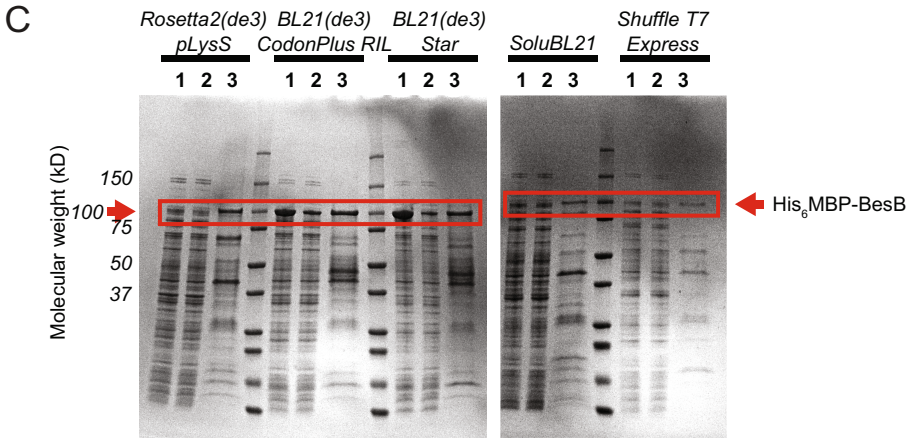


Extended Data Fig. 2 | See next page for caption.

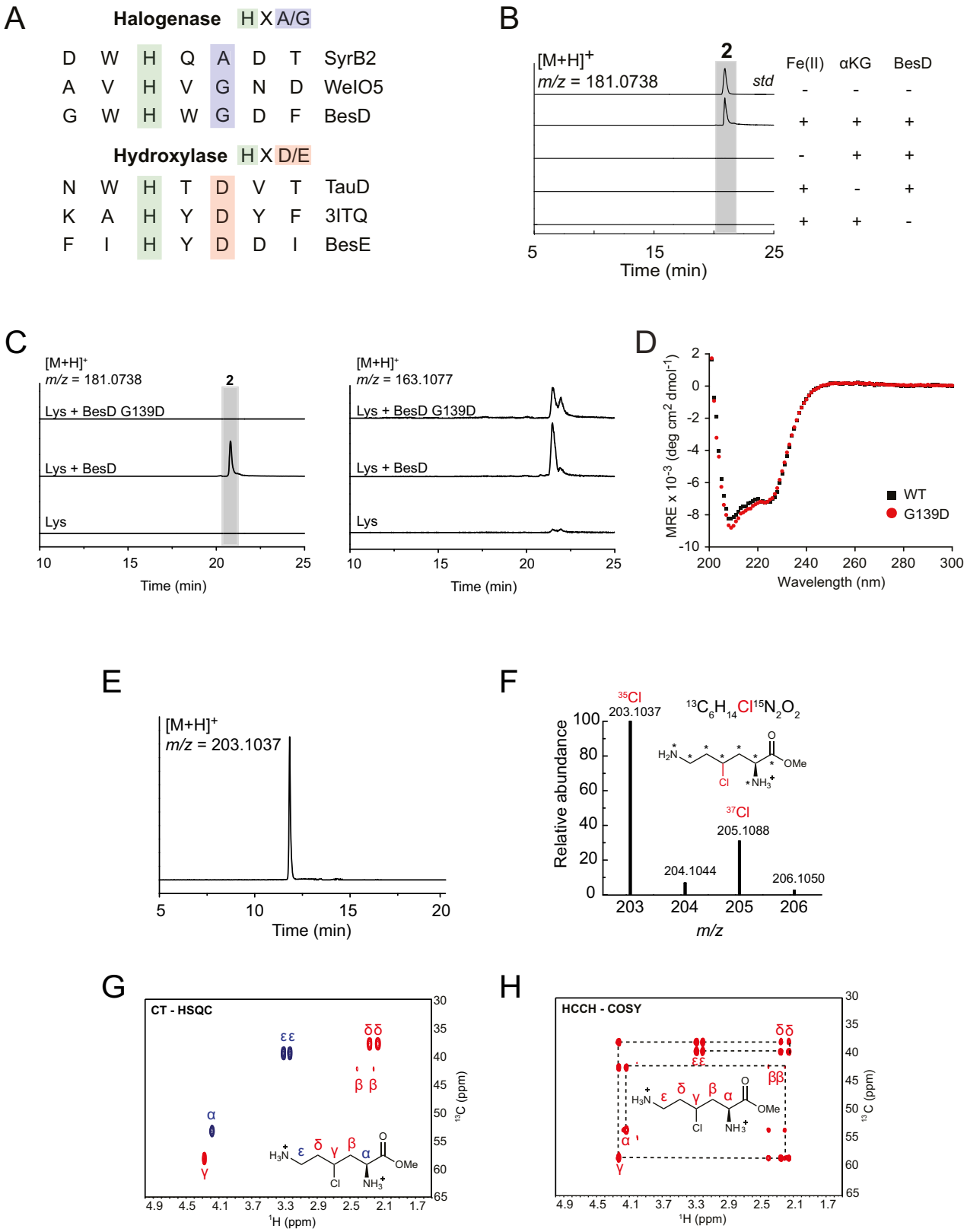
Extended Data Fig. 2 | βes biosynthetic clusters in *Streptomyces* spp. and production phenotypes of wild-type and knockout strains. **a**, Genome clusters and genomic context of βes/Pra and Pra biosynthetic clusters (coloured) in *S. cattleya*, *S.* sp. NRRL S-1448, *S. alboverticillatus*, *S. catenulae*, *S. achromogenes*, *S. lavenduligriseus*, and *S.* sp. NRRL S-31. The βes/Pra core cluster contains *besE* and *besF* in addition to *besA–besD*, whereas the Pra core cluster contains only *besA–besD*. The *S. alboverticillatus besF* is predicted on the basis of partial sequencing of the gene from the incomplete genome assembly. **b**, Characterization of *Streptomyces* spp. containing the βes/Pra and Pra biosynthetic clusters for their ability to produce terminal-alkyne amino acids. The presence of βes and Pra were analysed by derivatization with 3-azido-7-hydroxy-coumarin by CuAAC and monitoring of either the βes-coumarin adduct (βes-Cou; $m/z = 333.0831$) or the Pra-coumarin adduct (Pra-Cou; $m/z = 317.0880$) in the presence and absence of Cu(I) using LC–QTOF. The reaction in which Cu(I) was omitted provided a negative control because it is required for the cycloaddition to occur. Extracted ion chromatograms of *S. cattleya* and *S.* sp. NRRL S-1448 supernatants showing the production of βes (βes-Cou; $m/z = 333.0831$) (top). Extracted ion chromatograms of *S. cattleya*, *S. achromogenes*, *S.* sp. NRRL S-31, *S.* sp. NRRL S-1448 and *S. lavenduligriseus* supernatants show the production of Pra (Pra-Cou; $m/z = 317.0880$). *S. lavenduligriseus* was cultured but not found to produce Pra. Attempts to amplify genes in the βes cluster failed, which suggests that this particular strain may have lost the bes genes. Chromatograms shown are representative of at least three independent experimental replicates. **c**, *S. cattleya* knockout strains were complemented with the corresponding wild-type gene expressed from the ErmEp∗ promoter by insertion into the phage attachment site in the genome. Knockout and complementation strains were cultured for 7 d. The supernatants were derivatized with 3-azido-7-hydroxy-coumarin by CuAAC and analysed for the βes-coumarin adduct (βes-Cou; $m/z = 333.0831$) by LC–QTOF. Only the *besB* complementation strain appeared to recover βes production, possibly owing to polar effects related to ordering of the genes in the putative operon (*besD–besC–besB*). Chromatograms shown are representative of at least three independent experimental replicates.





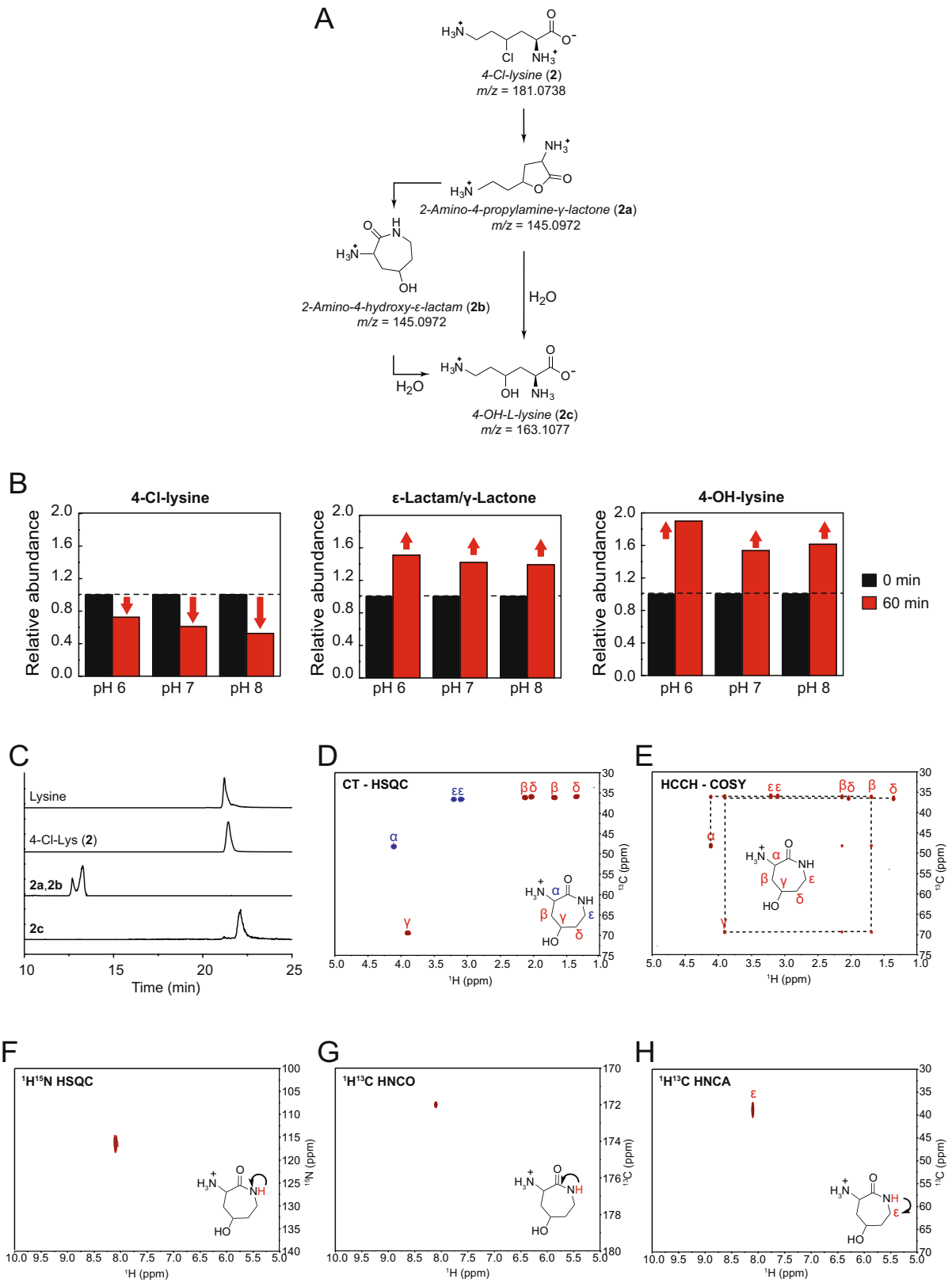


Extended Data Fig. 3 | Identification of pathway intermediates using comparative metabolomics and Bes proteins used for in vitro biochemical experiments. a, Extracted ion chromatograms of pathway intermediates identified using comparative metabolomics from *S. cattleya* pellet extracts for 4-Cl-lysine (**2**, $m/z = 181.0738$), 4-Cl-allylglycine (**3**, $m/z = 150.0316$), propargylglycine (**4**, $m/z = 114.0550$), γ-Glu-Pra (**5**, $m/z = 243.0975$) and γ-Glu-Bes (**6**, $m/z = 259.0925$). Chromatograms shown are representative of at least five independent experimental replicates. **b**, SDS–PAGE of purified enzymes used for biochemical assays and in vitro reconstitution of pathway. **1**, His_{10}–BesA; **2**, His_6MBP–BesB; **3**, BesC; **4**, BesD; **5**, His_{10}–BesE. All proteins are from *S. cattleya* except for BesB, which is from *S.* sp. NRRL S-1448. **c**, SDS–PAGE of His_6MBP–BesB solubility screen from *S.* sp. NRRL S-1448. BesB from *S. cattleya* was found to be highly insoluble, therefore the orthologue *S.* sp. NRRL S-1448 was cloned and purified. Lanes correspond to each cell line tested (1, pre-induction; 2, post-induction; 3, elution from Ni-NTA column). The highest soluble expression was observed in *E. coli* BL21(DE3)-Star.



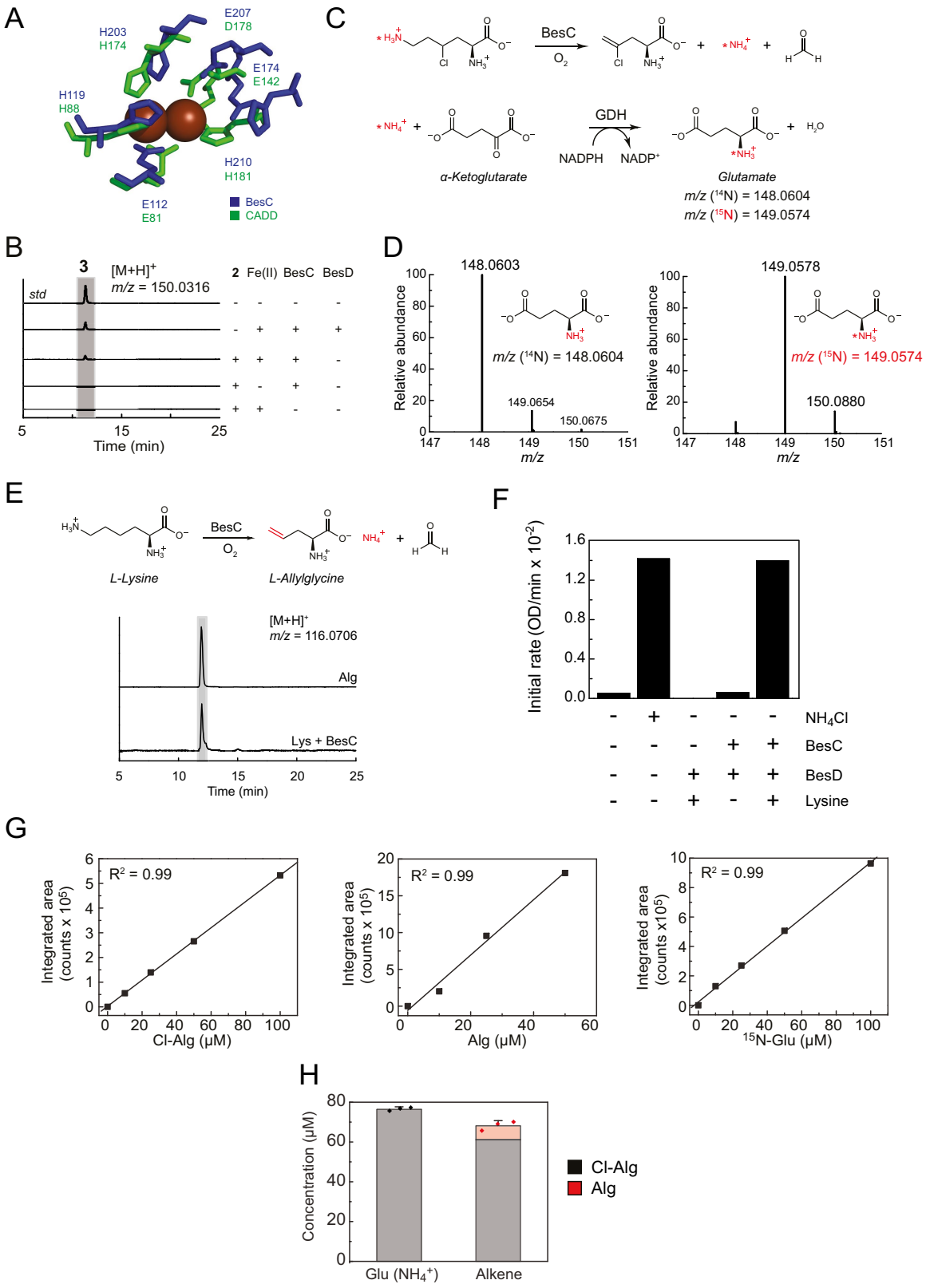
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Characterization of the BesD halogenase product and side products. a, BesD is a putative Fe/αKG-dependent halogenase with a conserved HXG motif rather than the expected HXD/E motif for hydroxylases. **b**, BesD activity requires L-lysine, Cl^-, Fe(II) and αKG (4-Cl-lysine, **2**; $m/z = 181.0738$). Chromatograms shown are representative of at least three independent experimental replicates. **c**, LC–QTOF traces monitoring production of 4-Cl-lysine (**2**, $m/z = 181.0738$) and 4-OH-lysine ($m/z = 163.1077$) in reactions with no enzyme, BesD or the BesD(G139D) mutant. Chromatograms shown are representative of three independent experimental replicates. As previously discussed, a low level of 4-OH-lysine is expected from direct hydroxylation of L-lysine by wild-type BesD, as has been commonly observed in other members of the αKG/Fe-dependent halogenase family. **d**, Circular dichroism (CD) spectra of wild-type BesD compared to the G139D mutant suggest that BesD(G139D) retains a similar fold or secondary structure content (MRE, molar residue ellipticity). Spectra shown are representative of three independent experimental replicates. **e**, BesD was incubated with fully ^{15}N- and ^{13}C-labelled L-lysine, Fe and αKG for 20 min before being quenched with methanolic HCl (3M), to yield the [$^{15}N_2$,$^{13}C_6$]-4-Cl-lysine methyl ester product. Chloride was provided as NaCl in the enzyme storage buffer. We found that derivatization was necessary to avoid intramolecular lactam formation. The derivatized BesD product was then isolated by HPLC and characterized. Extracted ion chromatograms of the [$^{15}N_2$, $^{13}C_6$]-4-Cl-lysine ($m/z = 203.1037$) produced by BesD are shown. Chromatograms shown are representative of three independent experimental replicates. **f**, The mass spectrum shows the characteristic Cl isotope pattern for [$^{15}N_2$,$^{13}C_6$]-4-Cl-lysine methyl ester produced by BesD and is representative of three independent experimental replicates. **g**, Two-dimensional ^1H-^{13}C constant-time heteronuclear single quantum correlation (CT-HSQC) was used to confirm assignment of C_α and C_ε of the [$^{15}N_2$,$^{13}C_6$]-4-Cl-lysine methyl ester produced by BesD. In this experiment carbons with two neighbours (red) have the opposite phase of carbons with one or three neighbours (blue). **h**, Two-dimensional ^1H-^{13}C HCCH correlation spectroscopy (COSY) was used to assign connectivity of carbons of the [$^{15}N_2$,$^{13}C_6$]-4-Cl-lysine methyl ester produced by BesD.



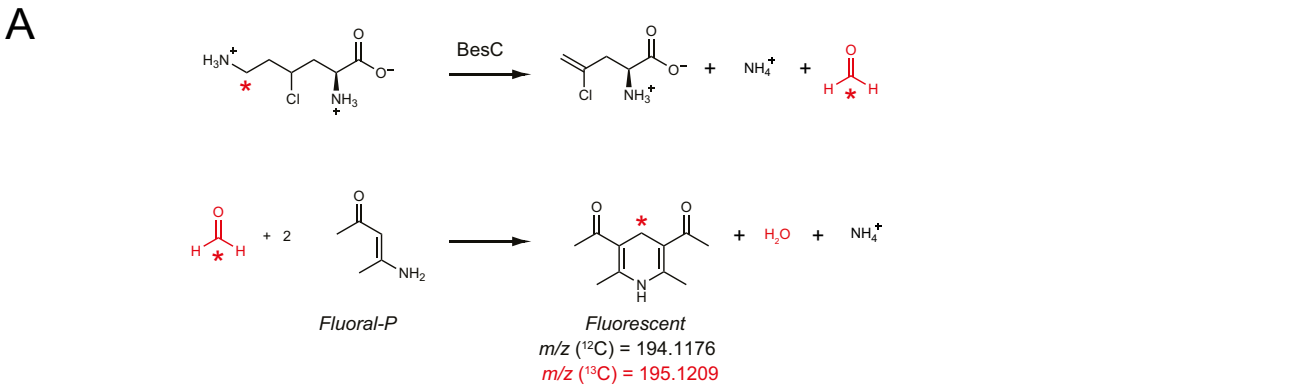
Extended Data Fig. 5 | See next page for caption.

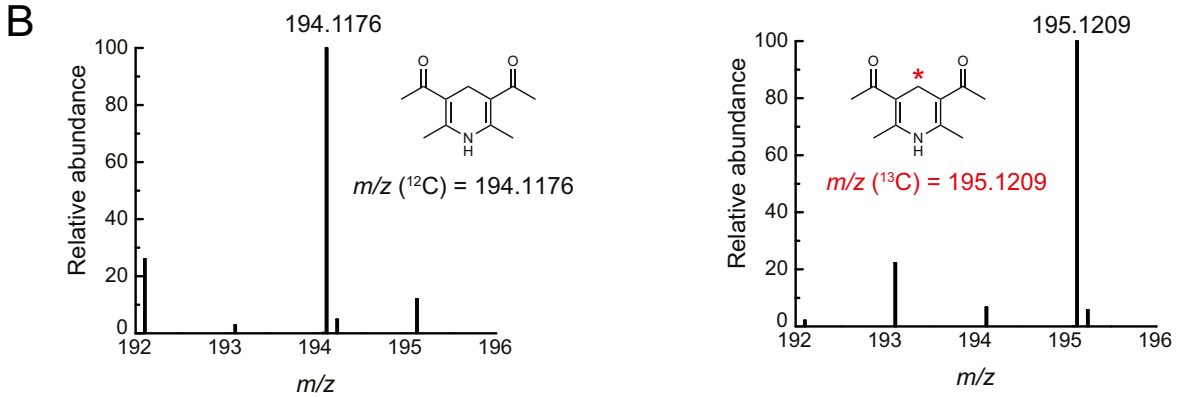
Extended Data Fig. 5 | Characterization of the 4-Cl-lysine (2) shunt pathway. a, 4-Cl-lysine (**2**) is proposed to undergo an intramolecular nucleophilic attack to form 2-amino-4-propylamine-γ-lactone with an estimated $t_{1/2} = 1$ h[34]. The γ-lactone is also unstable and can form either the 2-amino-4-hydroxy-ε-lactam or 4-OH-lysine. 4-OH-lysine could also arise from direct hydroxylation of L-lysine by BesD, as a low level of hydroxylation is observed in other members of the Fe/αKG-dependent halogenase family. **b**, Integrated extracted ion chromatograms for each of the observed products at pH 6, 7 and 8, following removal of BesD by filtration (Pall Nanosep filter, 10 kDa MWCO). Data are plotted relative to lysine, which is taken to be constant following enzyme removal. After 1 h, 4-Cl-lysine decreases, whereas the lactam, lactone and 4-OH-lysine products increase in abundance. **c**, Extracted ion chromatograms of the decomposition products observed upon halogenation of L-lysine by BesD by LC–QTOF. Lysine, $m/z = 147.1128$; 4-Cl-lysine (**2**), $m/z = 181.0738$; ε-lactam (**2a**)/γ-lactone (**2b**), $m/z = 145.0972$; 4-OH-lysine (**2c**), $m/z = 163.1077$. Chromatograms shown are representative of at least three independent experimental replicates. **d**, To obtain the lactam structure, BesD was incubated with fully ^{15}N- and ^{13}C-labelled L-lysine, Fe, and αKG for 16 h before being quenched with 1% (v/v) formic acid in methanol to yield the [$^{15}N_2$,$^{13}C_6$]-lactam as the major product. Chloride was provided as NaCl in the enzyme storage buffer. NMR analysis of the ε-lactam was carried out after HPLC purification. Two-dimensional ^{1}H-^{13}C CT-HSQC with a 13.6-ms ^{13}C evolution period was used to confirm assignment of C_α and C_ε. In this experiment, carbons with two neighbours (red) have the opposite phase of carbons with one or three neighbours (blue). **e**, ^{1}H-^{13}C HCCH COSY was used to assign backbone connectivity of carbons for the lactam. **f**, The ^{1}H-^{15}N HSQC spectrum confirms the presence of the amide in the lactam structure, with a reported chemical shift typical of amides (δ 6–9 p.p.m.). **g**, The 2D ^{1}H-^{13}C HNCO spectrum confirms that the epsilon N is adjacent to a carbonyl species. **h**, The 2D ^{1}H-^{13}C HNCA spectrum confirms that the epsilon N is coupled to C_ε, forming a cyclic structure.

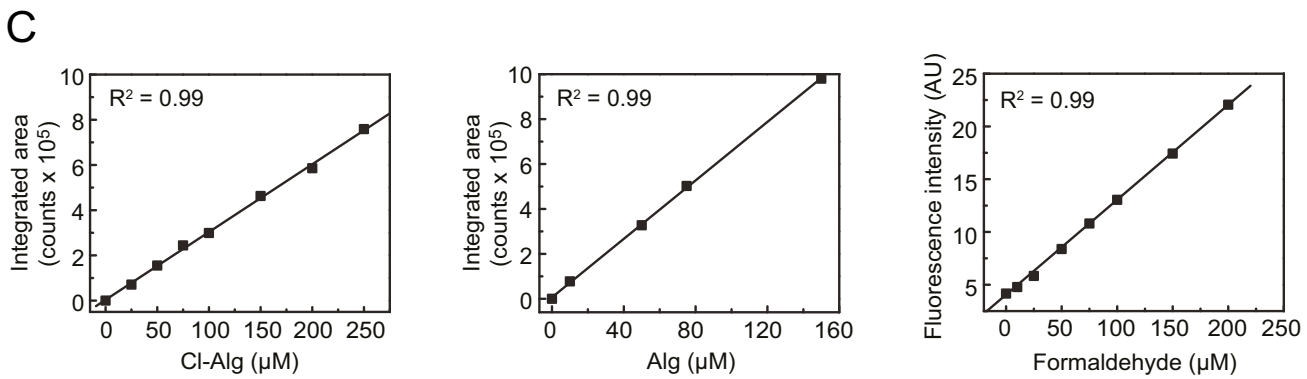


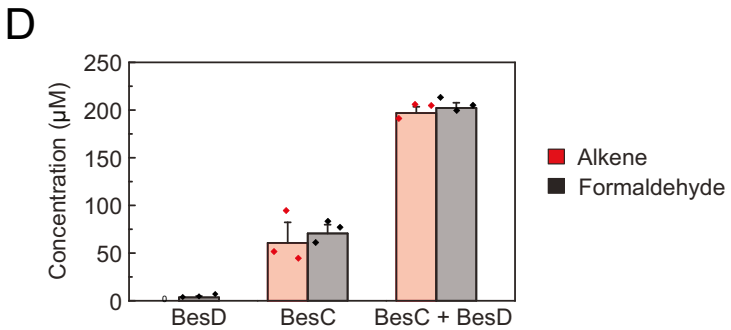
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Detection of ammonia as a co-product of the BesC reaction. a, BesC shows homology to non-haem diiron proteins such as the *Chlamydia* protein associated with death domains (CADD, 34% sequence similarity[35]) and stearoyl-ACP desaturase (26% sequence similarity[36]). Overlay of the predicted homology model of BesC (blue) with the CADD crystal structure (green)[37] reveals six conserved putative Fe-binding residues in the active site. Putative iron ligands shown in brown. Homology model was generated using Phyre 2. **b**, BesC catalyses formation of 4-Cl-allylglycine from 4-Cl-lysine and requires Fe(II) as a cofactor (4-Cl-allylglycine, **3**; $m/z = 150.0316$). Chromatograms shown are representative of at least three independent experimental replicates. **c**, The BesC reaction is run as a coupled reaction with BesD, starting from L-lysine, αKG and chloride. To detect the ammonia co-product, we used glutamate dehydrogenase (GDH), which produces glutamate from ammonia and αKG with concurrent oxidation of NADPH. **d**, The release of ammonia from 4-Cl-lysine was monitored through incorporation of the nitrogen from the ^{15}N-labelled lysine substrate into glutamate using LC–QTOF. Positive ionization mass spectrum showing formation of ^{14}N-glutamate or ^{15}N-L-glutamate when unlabelled L-lysine (left) or [$^{15}N_2$,$^{13}C_6$]-L-lysine (right) are used as the substrate, respectively. Spectra shown are representative of three independent experimental replicates. **e**, BesC is also competent to react with L-lysine directly to produce allylglycine in vitro. LC–QTOF traces show allylglycine produced by BesC compared to a synthetic standard (Alg, $m/z = 116.0706$). Chromatograms shown are representative of at least three independent experimental replicates. **f**, Consumption of NADPH was observed spectrophotometrically. Reactions contained BesC, BesD, lysine, αKG, sodium ascorbate, Fe(II) and chloride, as indicated. After 1 h, NADPH and glutamate dehydrogenase were added and NADPH consumption was monitored by A_{340}. **g**, Standard curves constructed by integrating extracted ion counts for the relevant species by LC–QTOF. R^2 value shown was determined by ordinary least squares. **h**, Quantification by LC–QTOF showing stoichiometric ammonia release measured as glutamate (Glu) compared total alkene product (4-Cl-allylglycine, Cl-Alg; allylglycine, Alg) in a coupled assay of BesD and BesC with L-lysine using L-glutamate dehydrogenase. Data are mean ± s.d. ($n = 3$ technical replicates).



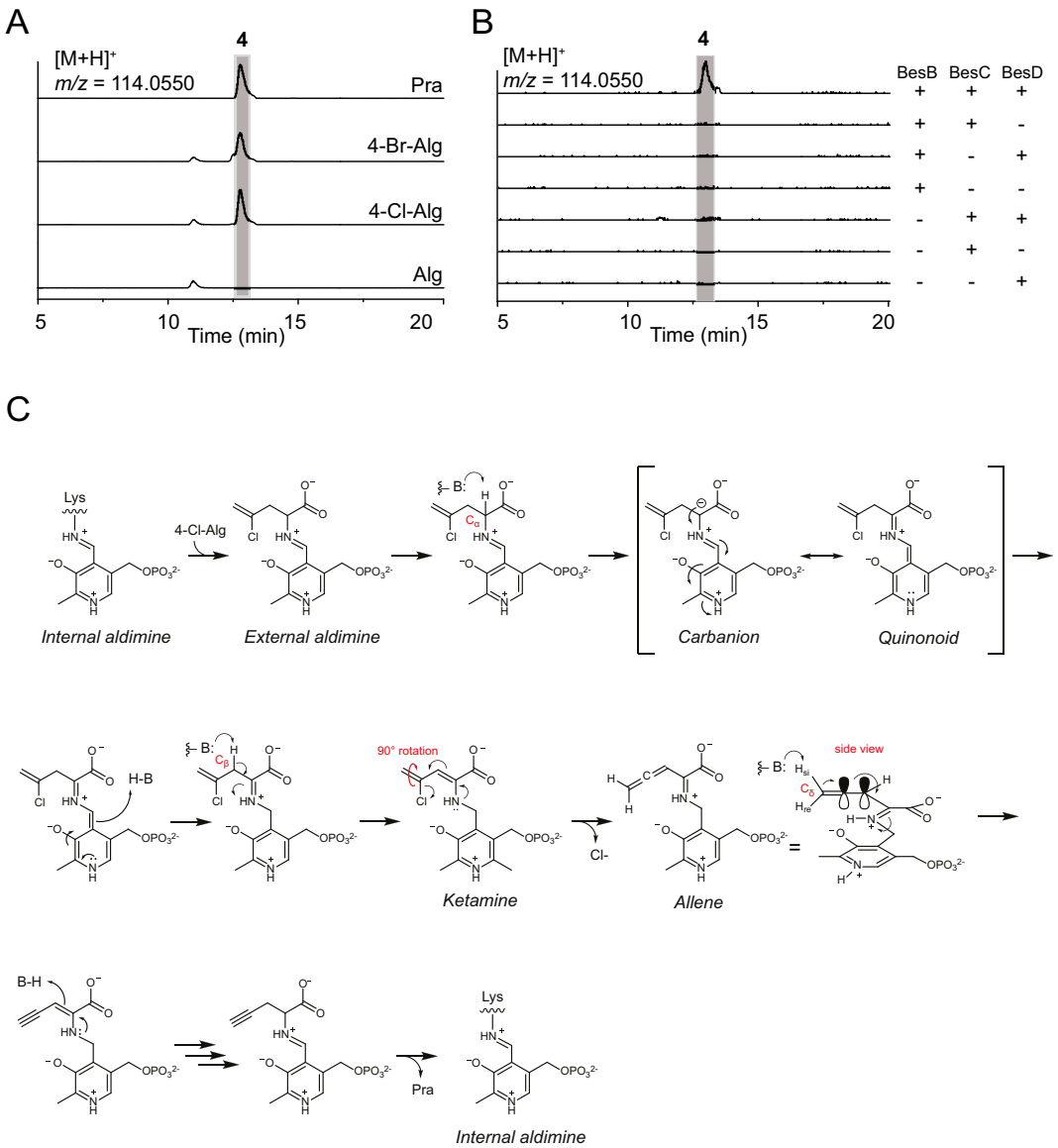




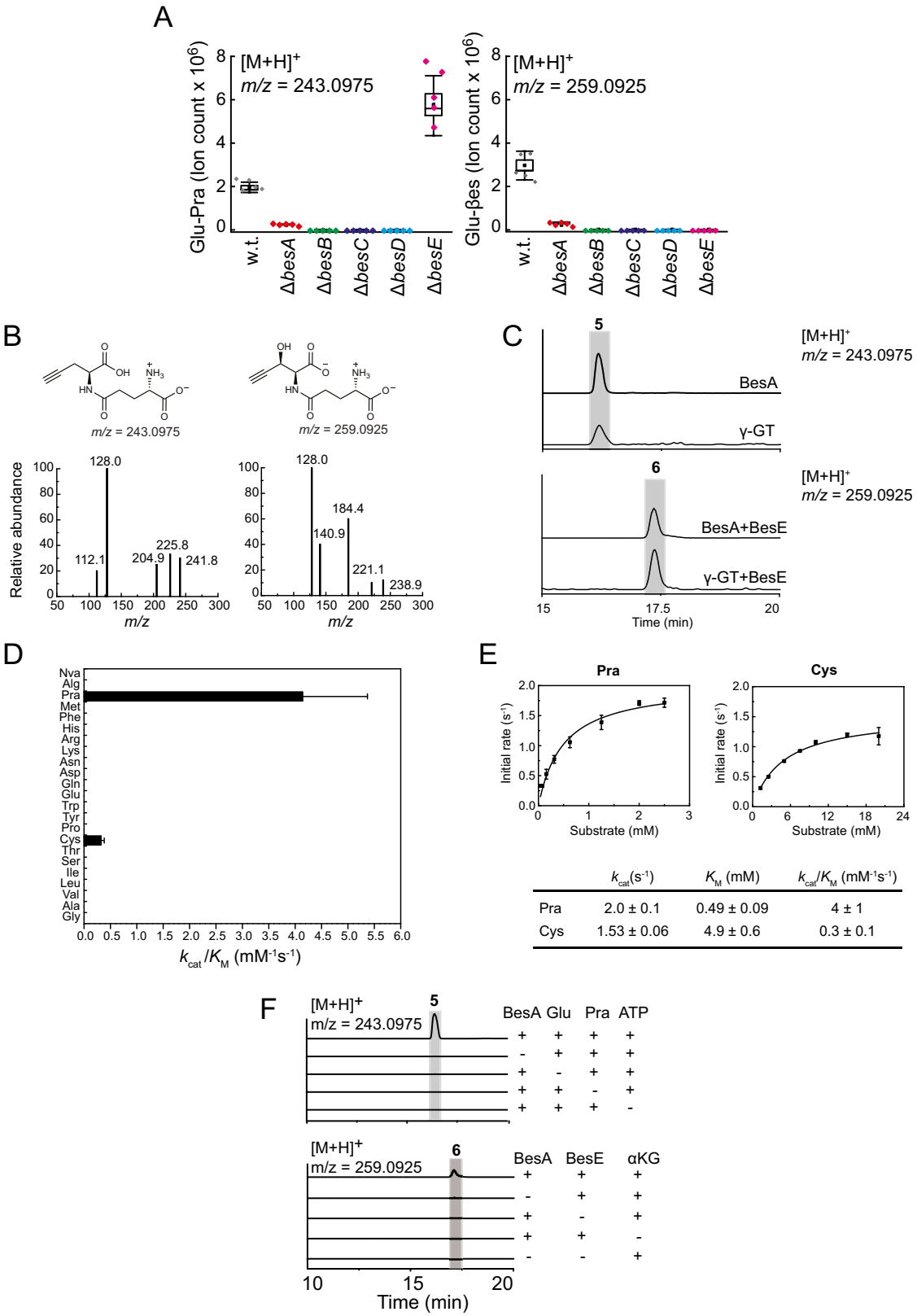


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Detection of formaldehyde as a co-product of the BesC reaction. The BesC reaction was run as a coupled reaction with BesD starting from L-lysine, αKG and chloride. **a**, To derivatize the putative formaldehyde co-product, we performed the enzymatic reaction in the presence of Fluoral-P. The release of formaldehyde from 4-Cl-lysine was monitored through incorporation of the carbon from ^{13}C-labelled lysine substrate into 3,5-diacetyl-1,4-dihydro-2,6-lutidine (DDL). **b**, Positive ionization mass spectrum showing formation of ^{12}C-DDL or ^{13}C-DDL when unlabelled L-lysine (left) or [$^{15}N_2$,$^{13}C_6$]-L-lysine (right) is used as the substrate, respectively. Spectra shown are representative of three independent experimental replicates. **c**, Standard curves for the relevant species were generated by integration of the extracted ion chromatogram from LC–TOF (Cl-Alg, Alg) or by fluorescence quantification (formaldehyde). R^2 value shown was determined by ordinary least squares. **d**, Quantification of formaldehyde produced from Fluoral-P reaction with 5 mM L-lysine and purified BesD, BesC, or both BesC and BesD. The resulting product was quantified by fluorescence and compared to alkene product formation measured by LC–QTOF and found to be stoichiometric. Data are mean ± s.d. ($n = 3$ technical replicates).

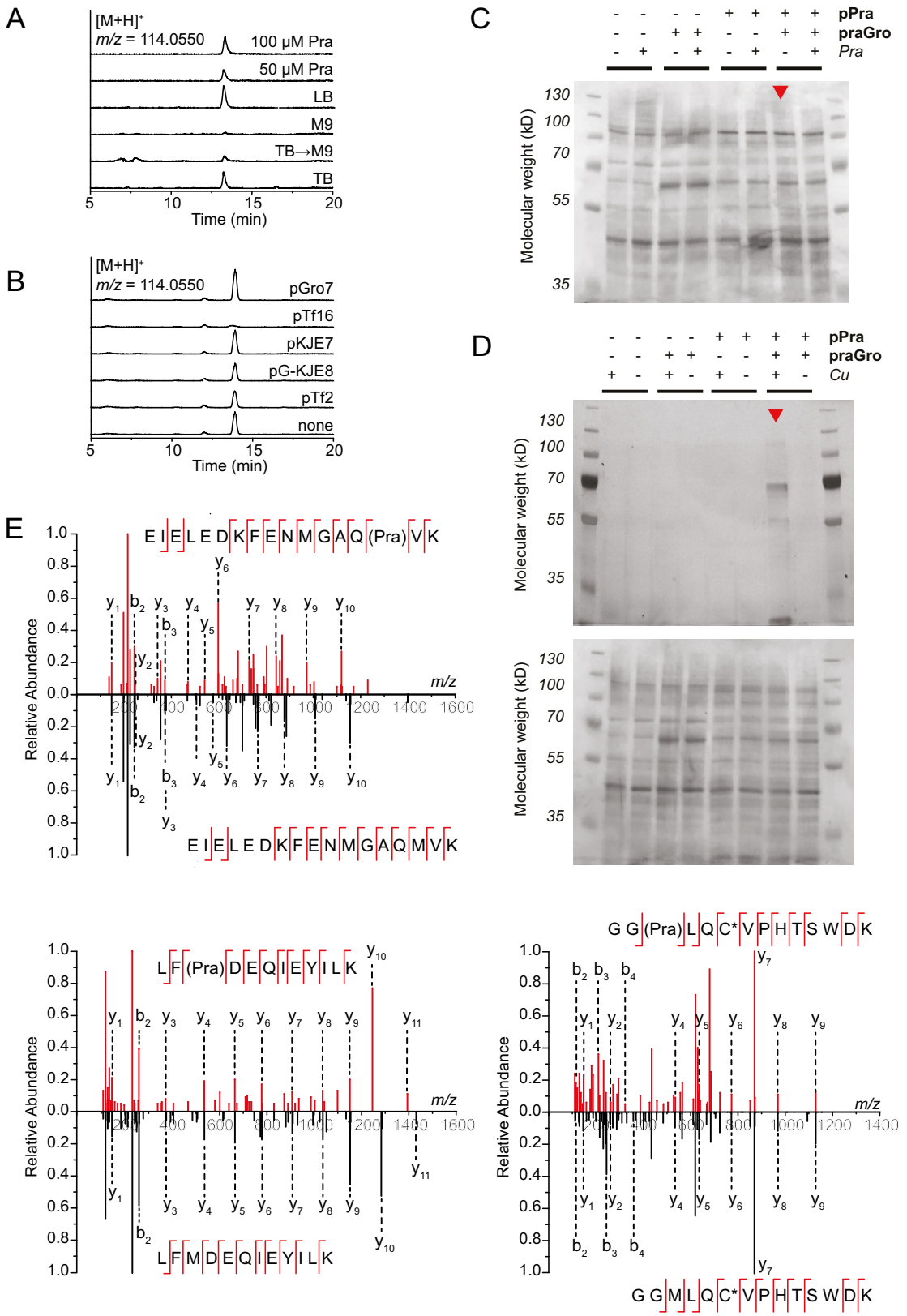


Extended Data Fig. 8 | In vitro characterization of BesB and proposed mechanism of terminal alkyne formation. a, LC–QTOF analysis of in vitro reactions containing BesB, PLP and substrate. When the substrate was 4-Cl-Alg or 4-Br-Alg, Pra (**4**, $m/z = 114.0550$) was observed. However, no Pra is observed if allylglycine is used as a substrate. Chromatograms shown are representative of at least three independent experimental replicates. **b**, In vitro reconstitution of Pra (**4**, $m/z = 114.0550$) formation from L-lysine with purified BesB, BesC, and BesD and Fe(II), αKG, chloride, PLP and ascorbate, monitored by LC–QTOF. Chromatograms shown are representative of at least three independent experimental replicates. **c**, A mechanistic model for BesB is proposed in analogy to cystathionine γ-synthase and related PLP mechanisms that form allene-intermediates[38]. In the first step, the external 4-Cl-allylglycine-aldimine is formed. Deprotonation at the C_α-position yields the 4-Cl-allylglycine-carbanion or quinonoid or intermediate. Deprotonation at the C_β-position produces the 4-Cl-allylglycine-ketamine intermediate. Elimination of chloride can then occur upon a 90° rotation to break planarity or aromatization to form a proposed terminal-allene intermediate. Work using Pra as a mechanistic inhibitor of PLP enzymes shows that isomerization between the terminal alkyne and allene can occur[30–32]. At this time, we suggest that the reverse can occur to convert the PLP-bound allene to the terminal alkyne, yielding a PLP-bound Pra intermediate. The C_δ deprotonation to initiate this step is facilitated by overlap of the C_δ–H bond with the conjugated C_β–C_γ and PMP–imine π-bonds. Subsequent steps then allow regeneration of the external aldimine with Pra, which is then released from the cofactor.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | In vitro and in vivo characterization of BesA and BesE. a, Relative abundance of the γ-Glu-Pra (**5**, $m/z = 243.0975$) and γ-Glu-βes (**6**; $m/z = 259.0925$) dipeptides in *S. cattleya* wild-type and pathway knockout strains ($n = 5$ biological replicates; dot shows mean; line shows median; box edges show 25th and 75th percentiles; error bars show 95th percentile). **b**, Fragmentation of γ-Glu-Pra (left) and γ-Glu-βes (right) in negative ionization mode by liquid chromatography with triple quadrupole mass spectrometry shows fragments characteristic of γ-glutamyl-dipeptides ($m/z = 128.0$). Spectra shown are representative of at least three independent experimental replicates. **c**, LC–MS traces for enzymatically synthesized γ-Glu-Pra (**5**, $m/z = 243.0975$) by either γ-GT or BesA and LC–QTOF traces for enzymatically synthesized γ-Glu-βes (**6**, $m/z = 259.0925$) by either γ-GT plus BesE or BesA plus BesE. Chromatograms shown are representative of at least three independent experimental replicates. **d**, Steady-state kinetic analysis for BesA reactions between L-glutamate and various L-amino acid partners (Nva, norvaline). Each bar represents the k_{cat}/K_M value for the corresponding amino acid calculated from the individual kinetic terms (k_{cat} and K_M) obtained from nonlinear curve fitting of data to the Michaelis–Menten equation. Data shown are mean ± s.e.m. ($n = 3$ technical replicates) obtained by propagation. Only reactions between L-glutamate and L-propargylglycine or L-cysteine showed any detectable activity, which suggests that BesA is relatively selective and that the residual Cys activity may be a product of its evolutionary history. **e**, Steady-state kinetic analysis of BesA with 10 mM glutamate and L-propargylglycine or L-cysteine as substrates. Data are mean ± s.d. ($n = 3$ technical replicates). Table contains k_{cat}, K_M, and k_{cat}/K_M calculated by nonlinear curve fitting to the Michaelis–Menten equation. Data are mean ± s.e. Error in k_{cat}/K_M is obtained by propagation from the individual kinetic terms. **f**, Top, extracted ion chromatograms for γ-Glu-Pra in reactions containing a combination of BesA, BesE and cofactors (γ-Glu-Pra, **5**; $m/z = 243.0975$). Bottom, extracted ion chromatograms for γ-Glu-βes in reactions containing a combination of BesA, BesE and cofactors (γ-Glu-βes, **6**; $m/z = 259.0925$). Chromatograms shown are representative of at least three independent experimental replicates.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | In vivo production of amino acids and alkyne-labelled proteins using genes from the bes gene cluster in *E. coli*. **a**, BesB functional expression was tested in different medium conditions as it is the limiting factor in the biosynthesis of Pra. Expression was tested by analysing production of Pra (**4**, $m/z = 114.550$) by LC–QTOF in *E. coli* BL21 Star (DE3) pSV272.1-MBP-BesB cultures fed with substrate (1 mM 4-Cl-allylglycine, **3**). Cells were collected after 24 h of growth. Chromatograms shown are representative of three independent experimental replicates. **b**, Optimization of BesB functional expression with co-expression of various *E. coli* protein chaperones. Production of Pra (**4**, $m/z = 114.550$) was monitored by LC–QTOF in *E. coli* BL21 Star (DE3) pSV272.1-MBP-BesB cultures fed with substrate (1 mM 4-Cl-allylglycine, **3**). Cells were collected after 24 h of growth in TB. Chromatograms shown are representative of three independent experimental replicates. **c**, SDS–PAGE gel from Fig. 4d showing Pra incorporation into the *E. coli* proteome by CuAAC derivatization, stained with Coomasie dye to visualize total protein loading in each lane. Gel shown is representative of six independent experimental replicates. **d**, Additional controls for the SDS–PAGE gel from Fig. 4d showing that successful derivatization with Tamra-azide requires endogenous Pra production, PraRS and Cu. In this experiment, *E. coli* B834(DE3) pPra praGro was cultured for 2 d along with the respective empty vector controls and no-Cu controls. The cell lysates were then derivatized with Tamra-azide dye by CuAAC before analysis by SDS–PAGE under fluorescence mode. Reactions without Cu act a negative control because Cu is required for CuAAC. Ladder consists of 70-kDa marker that can be visualized by fluorescence overlaid with the Coomassie-stained ladder from the same gel for reference. Gel shown is representative of six independent experimental replicates. **e**, Total protein was extracted from cell pellets of *E. coli* B834(DE3) pPra PraGro. The extracted protein was subject to tryptic digest and bottom-up proteomic analysis using an using a Thermo LC-Q Exactive and spectra were analysed using X! Tandem[39]. Representative high-resolution MS/MS spectra of peptides (GroEL, showing Pra replacing Met from *E. coli* B834(DE3) pPra PraGro are shown (EIELEDKFENMGAQ(Pra)VK, GroL, WP_000729117; LF(Pra)DEQIEYILK, BesC, WP_080628534; GG(Pra) LQC∗VPHTSWDK, BesD, WP_016975823). The spectrum above the axis gives peptide-spectrum matches to sequences with Pra substituted for Met; the spectrum below the axis shows the corresponding Met-containing native peptide (C∗ = carbamidomethyl derivative from iodoacetamide capping). Spectra shown are representative of two independent experimental replicates.

nature research

Corresponding author(s): Michelle C. Y. Chang

Last updated by author(s): Dec 14, 2018

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	MassHunter LC/MS Data Acquisition (B0.06.01) was used for acquisition of mass spectra on an Agilent 6530 Q-TOF. NMR spectra were aquired on Bruker 900 MHz instrument and collected with accompanying software. Proteomics was performed using a Thermo LC-Q Exactive and collected using accompanying software.
Data analysis	XCMS (3.5, Scripps) was used for comparitive metabolomics. Cytoscape (3.4.0, Agilent) was used to visualize results from comparitive genomics. Origin (6.0) was used for non-linear curve fitting. Mass Hunter Qualitative Analysis (B.07.00, Agilent) was used for integration of chromatograms. Comparative genomics was performed using custom software deposited on Github (https://github.com/jamarchand/ioblast). NMR spectra were analyzed using Mnova (Mestrelab). Proteomics analysis was performed using Scaffold 4. Details of data analysis are provided with paper.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Accession codes for genes and proteins in this study are provided in supplementary tables. Source data files for all figures are provided. Datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

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	For comparative metabolomics, 5 technical replicates were chosen for each sample type. Recommendation on sample size came from a peer and has been suggested by publications. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4215863/pdf/ac500734c.pdf). For enzyme assays with quantitative data, 3 technical replicates were used to determine mean and standard deviation and is stated when relevant. Kinetic parameters (kcat and Km) were calculated using non-linear curve fitting. For all chromatograms and mass spectra shown, data shown is representative from at least 3 biological replicates.
Data exclusions	No data was excluded from the analysis.
Replication	At least 3 biological replicates were performed for samples that show representative traces or mass spectra. For fluorescent PAGE, experiments were replicated 6 times, all successfully. For proteomics, two experimental replicates were submitted for analysis and one representative set is shown.
Randomization	No data was randomized since it was not applicable for our set of experiments.
Blinding	No data was blinded since it was not applicable for our set of experiments.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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