# Discovery, characterization and engineering of ligases for amide synthesis

https://doi.org/10.1038/s41586-021-03447-w

Received: 21 August 2020

Accepted: 11 March 2021

Published online: 19 May 2021

Check for updates

Michael Winn<sup>1</sup>, Michael Rowlinson<sup>1</sup>, Fanghua Wang<sup>1,2</sup>, Luis Bering<sup>1</sup>, Daniel Francis<sup>1</sup>, Colin Levy<sup>1</sup> & Jason Micklefield<sup>1⊠</sup>

Coronatine and related bacterial phytotoxins are mimics of the hormone jasmonyl-L-isoleucine (JA-Ile), which mediates physiologically important plant signalling pathways<sup>1-4</sup>. Coronatine-like phytotoxins disrupt these essential pathways and have potential in the development of safer, more selective herbicides. Although the biosynthesis of coronatine has been investigated previously, the nature of the enzyme that catalyses the crucial coupling of coronafacic acid to amino acids remains unknown<sup>1,2</sup>. Here we characterize a family of enzymes, coronafacic acid ligases (CfaLs), and resolve their structures. We found that CfaL can also produce JA-Ile, despite low similarity with the Jar1 enzyme that is responsible for ligation of JA and L-lle in plants<sup>5</sup>. This suggests that Jar1 and CfaL evolved independently to catalyse similar reactions–Jar1 producing a compound essential for plant development<sup>4,5</sup>, and the bacterial ligases producing analogues toxic to plants. We further demonstrate how CfaL enzymes can be used to synthesize a diverse array of amides, obviating the need for protecting groups. Highly selective kinetic resolutions of racemic donor or acceptor substrates were achieved, affording homochiral products. We also used structure-guided mutagenesis to engineer improved CfaL variants. Together, these results show that CfaLs can deliver a wide range of amides for agrochemical, pharmaceutical and other applications.

Coronatine 3 (COR) is an important phytotoxin produced by bacterial plant pathogens; it is composed of the polyketide coronafacic acid 1 (CFA), conjugated via an amide bond to coronamic acid 2 (CMA), an unusual cyclopropyl amino acid (Fig. 1a)<sup>1,2</sup>. COR is a structural mimic of IA-Ile (6), a ubiquitous plant hormone that is essential for plant development and defence<sup>3</sup>. The biologically active stereoisomer is (3R,7S)-IA-Ile, but the C7 stereocentre rapidly epimerizes at physiological pH to the more stable but inactive trans (3R,7R) diastereoisomer, which modulates its activity<sup>4</sup>. In contrast, COR is configurationally stable, endowing it with increased potency and longevity. The conjugation of JA and L-Ile in plants is catalysed by the adenosine triphosphate (ATP)-dependent ligase Jar1 (Fig. 1b), a member of the ANL superfamily<sup>5</sup>. ANL enzymes generate acyl-adenylate (acyl-AMP) intermediates that undergo substitution with various nucleophiles. For example, acyl-CoA synthetases (ACSs) are common ANL enzymes generating thioesters, which can be coupled to an amine by a secondary N-acyltransferase. Jar1 is an amide bond synthetase (ABS), a rarer subclass of the ANL superfamily which accept amine nucleophiles directly without requiring an additional partner enzyme (Supplementary Fig. 1)<sup>5,6</sup>.

CFA is assembled by a type I polyketide synthase via the cyclization of a  $\beta$ -ketothioester intermediate (Fig. 1c)<sup>1,7-9</sup>. The biosynthesis of CMA occurs via a nonribosomal peptide synthetase (NRPS)-mediated cryptic chlorination and cyclization (Fig. 1d)<sup>10-12</sup>. A putative ligase (CfaL) within the *Pseudomonas syringae* COR biosynthetic gene cluster is predicted to couple CFA and CMA to form COR (Fig. 1a, Supplementary Fig. 2a and

Supplementary Table 1). Previous attempts to characterize this ligase have been unsuccessful<sup>13</sup>. Other plant pathogens possess a COR-like biosynthetic gene cluster<sup>14,15</sup>, including *Streptomyces scabies* which has a putative CfaL and CFA biosynthetic genes, but no CMA pathway (Supplementary Fig. 2b and Supplementary Table 1)<sup>15</sup>. Consequently, this strain produces predominantly CFA-L-lle **4**, along with smaller quantities of L-Val and L-*allo*-lle adducts, which have also been detected from *P. syringae*<sup>16-18</sup>. As CFA must be coupled to an amino acid to elicit biological activity, we sought to characterize the key CfaL enzymes to enable new routes to COR-like phytotoxins as potential herbicides<sup>19</sup>. We were also interested in exploring the relationship between the bacterial CfaL and the functionally related plant ligase Jar1.

As well as being fundamental in nature, amide formation is one of the most widely used synthetic transformations. Although coupling acids and amines is relatively simple, it often requires three steps, protect–couple– deprotect, to install each amide. Stoichiometric quantities of expensive and deleterious coupling reagents are typically required and purification can be problematic<sup>20</sup>. While some progress has been made in the development of chemocatalytic methods for amide synthesis, these have not been widely adopted<sup>21–27</sup>. Consequently, there is interest in the development of enzymatic alternatives<sup>6,20,28–31</sup>. In this work, we characterize CfaL ligases and demonstrate how they are highly versatile biocatalysts for the synthesis of ubiquitous amides. Additionally, using structure-guided mutagenesis, we generate improved ligases providing more sustainable, alternative routes for production of pharmaceuticals, agrochemicals and other valuable materials.

<sup>1</sup>Department of Chemistry, Manchester Institute of Biotechnology, The University of Manchester, Manchester, UK. <sup>2</sup>Present address: School of Food Science and Engineering, South China University of Technology, Guangzhou, China. <sup>See</sup>-mail: jason.micklefield@manchester.ac.uk



**Fig. 1** | **Biosynthesis of coronatine (COR) and jasmonyl-L-isoleucine (JA-Ile). a**, Bacterial CfaL enzymes are predicted to ligate coronafacic acid (CFA) **1** with coronamic acid (CMA) **2** or L-amino acids to generate phytotoxins, including COR **3** and CFA-Ile **4. b**, In plants, the enzyme Jar1 ligates jasmonic acid (JA) and L-isoleucine to produce JA-Ile epimers (3*R*,7*S*)-**6** and (3*R*,7*S*)-**6. c**, Polyketide synthase (PKS) assembly of **1** from succinic semialdehyde. PKS consists of

acyl-carrier protein (ACP), acyl-transferase (AT), keto-synthase (KS), dehydratase (DH), enoyl-reductase (ER), keto-reductase (KR) and thioesterase (TE) domains. **d**, Nonribosomal peptide synthetase (NRPS)-mediated biosynthesis of **2** in *P. syringae*. NRPS consists of adenylation (A), thiolation (T) and thioesterase (TE) domains.

#### Characterization of the CfaL family

Overproduction of the CfaL from *P. syringae* (PsCfaL) in *Escherichia coli* resulted in only trace amounts of active enzyme. However, assays demonstrated that PsCfaL catalyses the ATP-dependent coupling of L-isoleucine and CFA, obtained from acid hydrolysis of coronatine, confirming the function of CfaL for the first time (Supplementary Figs. 3 and 4). The low quantity of PsCfaL available prevented full characterization, and so alternative CfaL homologues were explored. In addition to the putative *S. scabies* ligase (SsCfaL), other candidates located within putative COR-like clusters were selected from BLAST analysis (Supplementary Fig. 2). Of these, PbCfaL from *Pectobacterium brasiliense* and AlCfaL from *Azospirillum lipoferum* were chosen for characterization, as they are predicted to be more amenable to crystallization (Supplementary Table 2). While *P. brasiliense* is a well-known plant pathogen<sup>32,33</sup>, *A. lipoferum* is a root-dwelling, nitrogen-fixing plant symbiont that is not known to produce coronatine<sup>34</sup>.

SsCfaL was overproduced in *E. coli* (Supplementary Fig. 5) and assays with synthetic (±)-CFA<sup>19</sup>, L-isoleucine and ATP showed the direct formation of CFA-L-Ile **4** via a CFA-AMP intermediate (Supplementary Fig. 6), confirming CfaL is an ABS enzyme. In addition, SsCfaL also accepted the aromatic CFA variant **7** which when coupled to L-Ile forms coronalone, a simplified synthetic COR analogue with promising herbicidal activity<sup>35</sup>. Given adenylation occurs in the absence of amine substrate, the rate of adenylation can be measured in isolation (Extended Data Table 1 and Supplementary Fig. 7). The rate of adenylation  $(k_{cat})$  was greatest with (±)-CFA but the aromatic analogue 7 was found to have a lower Michaelis constant,  $K_m$ . Both (3R,7R)- and (3S,7S)-enantiomers of trans-JA 5 were also accepted, albeit at a lower level than CFA or 7, with a preference towards the natural (3R,7R)-5 stereoisomer. This indicates that, despite low amino acid sequence similarity (16%), the bacterial SsCfaL and plant Jar1 both catalyse the ligation of JA with L-Ile (Supplementary Fig. 8). Reactions with deactivated enzyme confirmed that the adenylation of 5 and the subsequent reaction with L-IIe are both enzyme-catalysed. Samples of trans-JA (Extended Data Table 1) contain a minor amount of the less stable *cis* epimer, owing to facile C7-epimerization (Fig. 1b)<sup>4</sup>. To explore the stereoselectivity of SsCfaL further, all four stereoisomers of configurationally stable 7-methyl-jasmonic acid were synthesized (Supplementary Fig. 9). Initially the trans and cis diastereoisomers of 7-methyl-jasmonic acid were separated and tested as a racemic mixture. However, these were poor substrates for SsCfaL, hence the resolution of all four stereoisomers was not carried out. The selectivity of SsCfaL was further tested by incubating CFA 1 or (±)-jasmonic acid 5 with 21 proteinogenic amino acids, resulting in a wide range of amino acid conjugates (Supplementary Figs. 10 and 11). Hydrophobic amino acids such as L-isoleucine and L-valine were preferred by SsCfaL which reflects the COR-like metabolites isolated from S. scabies<sup>16</sup>. No activity was seen with D-amino acids, or with primary amines and dipeptides



Fig. 2 | Structure of PbCfaL.a, Main image, X-ray crystal structure of PbCfaL (2Å) in the 'open' or 'adenylation' conformation (PDB ID, 7A9I). PbCfaL has a large N-terminal region (residues 1-403), shown in blue, and a flexible C-terminal region (residues 403-516), shown in red. The boxed region and magnified inset highlight the active site region, which is shown with 7 co-crystallized; labelled residues are conserved between all four CfaLs in this study. 7 lies 3.6 Å from a conserved tryptophan residue (W220), which probably helps to align the substrate via  $\pi$ - $\pi$  stacking interactions. **b**, PbCfaL superimposed onto McbA (gold; PDB ID, 6SQ8)<sup>29</sup>, the closest structural homologue to PbCfaL. In the 'adenylation' state, the two structures show high levels of similarity. In this state the carboxylic acid binding site (shown) is located between the N-terminal and the flexible C-terminal regions and is solvent accessible. c, PbCfaL superimposed on McbA in the 'closed' state (also referred to as the 'thiolation' state in related ACS enzymes). Like all ANLs, the C-terminal region of McbA undergoes a large rotation (direction indicated by dashed red arrow) to lie on top of the carboxylic acid binding site, trapping the adenylated intermediate before amine attack. The more rigid N-terminal region does not substantially change conformation. We would expect the C-terminal region of PbCfaL to undergo a similar rotation during catalysis. Structural alignment was performed with Chimera (version 1.14) MatchMaker.

(Supplementary Fig. 12). AlCfaL and PbCfaL, expressed from codon optimized synthetic genes, were seen to be functionally similar to SsCfaL, although with lower activities (Supplementary Fig. 13).

#### The structure of a CfaL ligase

Crystallography trials revealed that only PbCfaL yielded crystals of sufficient quality for structural studies. A PbCfaL structure in the adenylation conformation was solved to 2 Å resolution (Fig. 2), which is consistent with the ANL superfamily. Despite sharing low sequence identity (<20%), PbCfaL showed substantial structural similarity (>70%) to several other ANL ligases from a variety of organisms, including bacterial benzoate CoA ligases and firefly luciferases (Fig. 2b, Supplementary Table 3). By contrast, PbCfaL shares very little structural similarity with the catalytically equivalent Jar1(30%), suggesting that the two enzymes evolved independently (Supplementary Fig. 14). Plants are known to possess other ACSs that do share high structural homology with PbCfaL (Supplementary Table 3). However, Jar1 and related plant acyl-AMP forming enzymes that conjugate salicylate or indole-6-acetic acid (IAA) with amino acids appear to have evolved separately and specifically for plant hormone signalling<sup>36</sup>.

The structure of PbCfaL is composed of a large N-terminal domain (residues 1-403) and a smaller, flexible C-terminal domain (residues 403-516, Fig. 2). As with other members of the ANL superfamily, it is likely that the C-terminal domain undergoes a large rotation following acyl-adenylate formation to close off the acyl binding pocket and form the amino acid binding site (closed conformation) (Fig. 2). Co-crystallography of PbCfaL with 7 revealed that the extremely solvent-accessible acyl binding pocket lies between the two domains (Fig. 2, Supplementary Fig. 15). Sequence alignment between the CfaLs in this study showed a small number of conserved residues (Supplementary Fig. 16), with only W220 likely to make direct contact with 7, probably aligning the carboxylic acid via  $\pi - \pi$  stacking, explaining the higher binding affinity of 7 versus CFA (Fig. 2). Other conserved residues around 7 probably define the width and depth of the binding pocket. When aligned with the most structurally similar proteins from the Protein Data Bank (PDB; Supplementary Fig. 17) there are few conserved sequences, the most similarity occurring in the ATP binding SSGTTG motif (residues 168–173)<sup>37</sup>. Despite many attempts, determination of a PbCfaL structure in the closed conformation with AMP and amino acid bound could not be achieved.

#### CfaL substrate scope and engineering

We next explored if the synthetic scope of CfaL enzymes could be extended towards other amide targets. The CfaL enzymes were found to possess extremely broad substrate tolerance, accepting a variety of aryl and heteroaryl carboxylic acids 8-29 as well as aliphatic carboxvlic acids **30–46**, including several chiral compounds **39–46** (Fig. 3, Extended Data Table 2). Several acyl-donor substrates possessed other reactive functionalities, such as electrophilic ketones (11, 35, 43, 45), alkenes (33), as well as nucleophilic alcohol (40, 44) or amine groups (13, 17, 18, 19, 46), which would require protecting for traditional coupling chemistries, but do not interfere with the enzymatic ligation to amino acid acceptor substrates. In addition to proteinogenic amino acids (Fig. 4a, Extended Data Table 3), CfaL enzymes also accept a wide range of non-proteinogenic amino acids 47-61, including common pharmaceutical building blocks, with a preference for hydrophobic amino acids (Fig. 4b, c, Extended Data Table 4). Although polar, particularly charged, amino acids are not well accepted by CfaL, both L-2,4-diaminobutyrate 47 and L-ornithine 48 can be selectively acylated at the  $\alpha$ -amino group, obviating the need for protection of the side-chain amino group (Fig. 4c, Supplementary Fig. 17).

In general, SsCfaL and AlCfaL both performed better than PbCfaL (Figs. 3 and 4, Extended Data Tables 2–4), which was found to be less





Fig. 3 | Carboxylic acid substrate scope. a, Diverse structures of carboxylic acid (donor) substrates assayed with L-Ile and CfaL enzymes. b, Percentage conversion for ligation of carboxylic acids 8–46 with L-Ile catalysed by CfaL enzymes (column headings). Assays were carried out with wild-type and

thermally stable and frequently precipitated during the reaction timescale (Extended Data Fig. 1). On the basis of our crystallographic studies, we sought to improve the activity and stability of PbCfaL via rational, structure-guided mutagenesis. Sequence comparison between the four CfaL enzymes (Supplementary Fig. 16) identified few obvious distinctions. However, one noticeable difference was found on the flexible hinge-region linking the N- and C-terminal domains (at position 395, Extended Data Fig. 2a). This position is solvent-exposed and likely to be involved in the conformational changes required to shift between the adenylation (open) and amidation (closed) states of CfaL. The large, charged arginine residue that is located in this position in the other CfaLs and ANLs included in the sequence alignment (Supplementary Fig. 16) is occupied by a small and uncharged glycine. A PbCfaL(R395G) mutant showed increased activity against the panel

engineered CfaL enzymes (25  $\mu$ M), carboxylic acids **8**–**46** (2 mM) and L-lle (5 mM). Conversion to amide products was determined by HPLC analysis following 20 h incubation. Actual conversion values and errors can be found in Extended Data Table 2.

of carboxylic acids and some amino acid substrates (Figs. 3 and 4, Extended Data Tables 2–4). An X-ray crystal structure of this mutant was determined, which revealed no overall structural changes (Extended Data Fig. 2a). However, the melting temperature ( $T_m$ ) of PbCfaL(R395G) increased by 5 °C relative to the wild type, suggesting that the replacement of this solvent-accessible, charged R395 is beneficial for stability (Extended Data Fig. 1).

A subsequent double mutant, PbCfaL(R395G/A294P), showed a further increased  $T_m$  and slightly improved activity (Figs. 3 and 4, Extended Data Table 2–4). The location of this second mutation is within a highly conserved ATP binding loop (G289–L297) that is significantly larger in PbCfaL than in other related structures, and which may partially occlude the ATP binding site (Extended Data Fig. 2b). The proline found at this location in SsCfaL, PbCfaL and several other structurally similar ligases may aid in rotating this loop out of the binding site (Supplementary



**Fig. 4** | **Amino acid substrate scope. a**, Percentage conversion for ligation of carboxylic acid **9** with acceptor proteinogenic amino acids (rows). **b**, Percentage conversion for ligation of **9** with acceptor non-proteinogenic amino acids. a-Ile = *allo*-isoleucine. **c**, Structures of non-proteinogenic amino acids. **d**, Reversed-phase (RP)-HPLC trace of ligation product of L-Dab **47** (green) and **9** (*m*-methylbenzoate, red) catalysed by SsCfaL, compared to HPLC traces of synthesized standards of the two possible products **62** and **63**. Product of the

Fig. 16). Using these two PbCfaL mutants, which no longer precipitate during the reaction, we were able to substantially improve the conversions of both the panel of carboxylic acid and amino acid substrates (Figs. 3 and 4), demonstrating that minimal structure-guided mutagenesis can be used to engineer improved CfaL variants.

#### Synthetic applications of CfaL

To demonstrate the synthetic utility of CfaL, we sought to establish preparative-scale ligation reactions. Accordingly, conditions were optimized for the ligation of carboxylic acid **10** and L-lle (Fig. 5a). Reaction of **10** (at 15 mM concentration) with PbCfaL (R395G/A294P) cell free lysate afforded amide **64** in near quantitative conversion as determined by high-performance liquid chromatography (HPLC). The reaction mixture was subjected to a simple solvent extraction, providing 1.48 g of crude **64** from 400 ml of reaction mixture, which would be sufficiently pure (>92% purity by NMR, Extended Data Fig. 3) for further synthetic derivatization. Purification of the extract by column chromatography provided 1.37 g of pure **64** in 87% isolated yield (Fig. 5a). To avoid the use of stoichiometric quantities of the expensive co-factor ATP, we repeated the ligation of **10** and L-lle at the same scale, omitting ATP and instead introducing an ATP recycling system consisting of a polyphosphate

enzymatic reaction (bottom trace) shows selective acylation of the  $\alpha$ -amino group to give amide **62**. All assays were carried out with wild-type and engineered CfaL enzymes (5  $\mu$ M), carboxylic acid **9** (1 mM) and amino acids (2 mM). Conversion to amide products was determined by RP-HPLC analysis following 20 h incubation. Actual conversion values and errors can be found in Extended Data Tables 3 and 4.

(PolyP) kinase enzyme (CHU)<sup>38</sup> and an inexpensive PolyP phosphate donor (Extended Data Fig. 3). Although the isolated yield was lower in this case (52%), there is further scope for optimization. While CfaL cell lysate shows good activity for up to 12 h, we sought to improve enzyme stability/longevity through immobilization of CfaL in the form of a cross-linked enzyme aggregate (CLEA)<sup>39</sup>. PbCfaL(R395G/A294P) CLEAs were shown to retain activity over an extended period of five days, and could be isolated and recycled in five sequential ligation reactions (Extended Data Fig. 4a). Purified PbCfaL(R395G/A294P) was also shown to tolerate several solvents, including MeOH, ethylene glycol and the widely used 'green' solvent 2-methylTHF (Extended Data Fig. 4b).

To further demonstrate the synthetic potential of CfaL, a series of pharmaceutical-relevant scaffolds were prepared in excellent yields (Fig. 5a, b and Supplementary Fig. 18). For example, amides **65** and **67** were prepared in >70% isolated yields at around 100 mg scale. Furthermore, ligations of cinnamic acid **33** and indole carboxylic acids **26** (Fig. 3) with cyclopropyl amino acids **56** and L-Leu, respectively (Fig. 4), produced amides **68** and **69**, which are precursors for the manufacture of promising SARS-CoV-2 protease inhibitors, including PF-07304814 (Pfizer; in phase I clinical trials) (Fig. 5a, b and Supplementary Fig. 18)<sup>40,41</sup>. Similarly, ligation of the thiazole carboxylic acid **29** and *O*-methyl-L-serine **49** provided amide **70**; this is a key component of



**Fig. 5** | **Synthetic potential of CfaL enzymes. a**, Amides including pharmaceutical scaffolds synthesized by CfaL enzymes. **b**, Comparison of percentage conversion for CfaL enzymes in the synthesis of **65**–**70**. **c**, Kinetic resolution of racemic carboxylic acids (donor). Absolute configuration and diastereoisomeric ratios (d.r.) were determined by RP-HPLC using synthetic standards. Inset, the HPLC chromatogram for (S)-**71** formed in the kinetic resolution of racemic ibuprofen (**41**) with L-IIe and AlCfaL (*E* = 94). **d**, Comparison of the enantioselectivities of the different enzymes. Values of *E* = 15–30 are considered moderate–good, *E* > 30 are excellent<sup>43</sup>. For conversions <30% the calculation of *E* is unreliable, so the values were not determined (ND). **e**, Kinetic resolution of racemic amino acid (acceptor) **57** (2 mM) with acid **9** (1 mM) and AlCfaL (5  $\mu$ M) (*E* > 200) following 20 h incubation.

oprozomib, which is in phase II clinical trials for treatment of multiple myeloma<sup>42</sup>. Probing the limits of potential CfaL-reaction scope, we found that the mutant PbCfaL(R395G/A294P) also allowed the generation of amide precursors required for the synthesis of the antiviral telaprevir and the anti-cancer agent bortezomib (Extended Data Fig. 5). These products were only produced in low quantities, but with further engineering it may be possible to synthesize these precursors at higher levels. Overall, these reactions (Fig. 5) clearly demonstrate how structurally diverse carboxylic acids can be combined with proteinogenic and synthetic amino acids to produce pharmaceutically important compounds.

The yield reported is based on **9** which equates to a yield of 33% based on **57** (2 equiv. used). Inset, chiral HPLC analysis of the amide product showing a single enantiomer, (*S*)-**77**. Enantiomeric ratio (e.r.) values determined by chiral HPLC. <sup>a</sup>Isolated yield preparative-scale synthesis of **64** with PbCfaL(R395G/A294P) lysate **10** (15 mM), t-IIe (45 mM) and ATP (36 mM) incubated for 24 h. <sup>b</sup>Isolated yields of about 100-mg-scale reactions catalysed by SsCfaL cell lysate with carboxylic acid (5 mM), amine (15 mM) and ATP (15 mM) following 24 h incubation. <sup>c</sup>Conversions determined from HPLC peak area ratios, following assays including CfaL enzymes (25  $\mu$ M), carboxylic acids (2 mM) and amino acid (6 mM) incubated for 20 h. <sup>d</sup>*E* value was calculated from average d.r. or e.r. values, as described previously<sup>43</sup>. Percentage conversions and d.r. values represent means where *n* = 3, error denotes s.d.

The potential of CfaL for use in kinetic resolution of racemic synthetic carboxylic acids was also investigated (Fig. 5c, d). Notably, racemic ibuprofen could be resolved, with excellent enantioselectivity  $(E = 94)^{43}$  leading to the biologically active (*S*)-ibuprofen-L-lle amide **71**. Amide conjugates of ibuprofen and related NSAIDs with amino acids have been explored extensively for applications as prodrugs and/ or hydrogel-based nanomedicine<sup>44,45</sup>. Five other racemic acids were subjected to kinetic resolution, affording amides **72–76**, with modest *E* values (Fig. 5c, d and Supplementary Fig. 19). In the case of amide **73**, the mutant PbCfaL(R395G) was superior to any of the wild-type CfaLs, illustrating how protein engineering could be used to achieve more effective kinetic resolutions. Finally, we sought to exploit the high selectivity of CfaL for L-amino acids to effect the kinetic resolution of racemic amino acids. Amino acid **57** was selected as this is a common pharmaceutical building block, which would normally require multi-step asymmetric synthesis or laborious resolution and protection before acylation or peptide coupling. As anticipated, the reaction between carboxylic acid **9** and racemic amino acid **57** proceeds with excellent enantioselectivity (E > 200), with none of the *R*-configured enantiomer evident in chiral HPLC when SsCfaL or AlCfaL was used (Fig. 5e, Supplementary Fig. 20). This demonstrates how racemic carboxylic acid and racemic amino acids can be resolved during amide bond synthesis, using CfaL, avoiding more laborious asymmetric synthesis or traditional resolution procedures, and the need for protective group manipulations.

#### Discussion

The results presented here demonstrate the role of CfaL enzymes in biosynthesis of the important coronatine family of phytotoxins. BLAST analysis reveals that CfaL-like ligases appear in a large number of distinct COR-like clusters from across a broad range of microorganisms, including bacteria where COR-like phytotoxins have not been observed, suggesting that CfaLs and the biosynthesis of COR-like phytotoxins are widespread. CfaLs can also catalyse ligation of JA with Ile to generate the plant hormone JA-Ile in an identical fashion to the plant ligase Jar1. The lack of sequence and structural similarity between the CfaL and Jar1 suggests that the two enzymes have evolved largely independently in bacteria and in plants to perform very similar reactions. In addition to potential agrochemical applications, the CfaL family of enzymes can be used to produce a wide range of pharmaceutically relevant amides. Used in combination with improving ATP recycling techniques<sup>46,47</sup>, these enzymes could become powerful synthetic tools offering major advantages over other biocatalysts developed for amide synthesis. For example, the combination of ACS and N-acyltransferase enzymes have been investigated for amide synthesis<sup>48</sup>. However, large numbers of ACS and N-acyltransferase enzymes had to be screened to find pairs of enzymes with matching selectivity<sup>48</sup>. In addition to low substrate scope, this system also requires use of two expensive co-factors as well as engineering of two enzymes for further optimization, rather than just one. Other reports describe the use of standalone NRPS adenylation domains to synthesize amides<sup>30,49</sup>. In these examples only the carboxylic acid activation step is directly enzyme catalysed, the subsequent amidation proceeds spontaneously, requiring a large excess (about 100 equiv.) of the amine, which is not viable for many syntheses. CfaLs directly catalyse both steps and can therefore utilize acids and amines with more efficient stoichiometry. Taken together our results show that CfaLs have potential for the synthesis of diverse range of important amide products, offering clear advantages over traditional synthetic methods and other biocatalytic approaches.

#### **Online content**

Any methods, additional references, Nature Research 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/s41586-021-03447-w.

- Rangaswamy, V., Jiralerspong, S., Parry, R. & Bender, C. L. Biosynthesis of the Pseudomonas polyketide coronafacic acid requires monofunctional and multifunctional polyketide synthase proteins. Proc. Natl Acad. Sci. USA 95, 15469–15474 (1998).
- Ullrich, M. & Bender, C. L. The biosynthetic gene cluster for coronamic acid, an ethylcyclopropyl amino acid, contains genes homologous to amino acid-activating enzymes and thioesterases. J. Bacteriol. 176, 7574–7586 (1994).
- Staswick, P. E. & Tiryaki, I. The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis*. *Plant Cell* 16, 2117–2127 (2004).
- Fonseca, S. et al. (+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. Nat. Chem. Biol. 5, 344–350 (2009).

- Westfall, C. S. et al. Structural basis for prereceptor modulation of plant hormones by GH3 proteins. Science 336, 1708–1711 (2012).
- Winn, M., Richardson, S. M., Campopiano, D. J. & Micklefield, J. Harnessing and engineering amide bond forming ligases for the synthesis of amides. *Curr. Opin. Chem. Biol.* 55, 77–85 (2020).
- Parry, R. J., Jiralerspong, S., Mhaskar, S., Alemany, L. & Willcott, R. Investigations of coronatine biosynthesis. Elucidation of the mode of incorporation of pyruvate into coronafacic acid. J. Am. Chem. Soc. 118, 703–704 (1996).
- Tao, T. & Parry, R. J. Determination by enantioselective synthesis of the absolute configuration of CPE, a potential intermediate in coronatine biosynthesis. Org. Lett. 3, 3045–3047 (2001).
- Strieter, E. R., Koglin, A., Aron, Z. D. & Walsh, C. T. Cascade reactions during coronafacic acid biosynthesis: elongation, cyclization, and functionalization during Cfa7-catalyzed condensation. J. Am. Chem. Soc. 131, 2113–2115 (2009).
- Parry, R. J., Lin, M. T., Walker, A. E. & Mhaskar, S. Biosynthesis of coronatine: investigations of the biosynthesis of coronamic acid. J. Am. Chem. Soc. 113, 1849–1850 (1991).
- 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).
- Kelly, W. L. et al. Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. *Biochemistry* 46, 359–368 (2007).
- Rangaswamy, V. et al. Expression and analysis of coronafacate ligase, a thermoregulated gene required for production of the phytotoxin coronatine in *Pseudomonas syringae*. *FEMS Microbiol. Lett.* **154**, 65–72 (1997).
- Slawiak, M. & Lojkowska, E. Genes responsible for coronatine synthesis in *Pseudomonas* syringae present in the genome of soft rot bacteria. *Eur. J. Plant Pathol.* **124**, 353–361 (2009).
- Bignell, D. R. D. et al. Streptomyces scabies 87–22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant-microbe interactions. *Mol. Plant Microbe Interact.* 23, 161–175 (2010).
- Fyans, J. K., Altowairish, M. S., Li, Y. & Bignell, D. R. Characterization of the coronatine-like phytotoxins produced by the common scab pathogen *Streptomyces scabies*. *Mol. Plant Microbe Interact.* 28, 443–454 (2015).
- Mitchell, R. E. & Frey, E. J. Production of N-coronafacoyl-L-amino-acid-analogs of coronatine by Pseudomonas syringae pv Atropurpurea in liquid cultures supplemented with L-amino acids. J. Gen. Microbiol. 132, 1503–1507 (1986).
- Mitchell, R. E. & Ford, K. L. Chlorosis-inducing products from Pseudomonas syringae pathovars: new N-coronafacoyl compounds. Phytochemistry 49, 1579–1583 (1998).
- Littleson, M. M. et al. Scalable total synthesis and comprehensive structure-activity relationship studies of the phytotoxin coronatine. *Nat. Commun.* 9, 1105 (2018).
- Sabatini, M. T., Boulton, L. T., Sneddon, H. F. & Sheppard, T. D. A green chemistry perspective on catalytic amide bond formation. *Nat. Catal.* 2, 10–17 (2019).
- Sabatini, M. T., Boulton, L. T. & Sheppard, T. D. Borate esters: simple catalysts for the sustainable synthesis of complex amides. Sci. Adv. 3, e1701028 (2017).
- Krause, T., Baader, S., Erb, B. & Gooßen, L. J. Atom-economic catalytic amide synthesis from amines and carboxylic acids activated *in situ* with acetylenes. *Nat. Commun.* 7, 11732 (2016).
- Stephenson, N. A., Zhu, J., Gellman, S. H. & Stahl, S. S. Catalytic transamidation reactions compatible with tertiary amide metathesis under ambient conditions. J. Am. Chem. Soc. 131, 10003–10008 (2009).
- Allen, C. L., Atkinson, B. N. & Williams, J. M. J. Transamidation of primary amide with amines using hydroxylamine hydrochloride as an inorganic catalyst. *Angew. Chem. Int. Edn* 51, 1383–1386 (2012).
- Al-Zoubi, R. M., Marion, O. & Hall, D. G. Direct and waste-free amidations and cycloadditions by organocatalytic activation of carboxylic acids at room temperature. *Angew. Chem. Int. Edn* 47, 2876–2879 (2008).
- Noda, H., Furutachi, M., Asada, Y., Shibasaki, M. & Kumagai, N. Unique physicochemical and catalytic properties dictated by the B<sub>3</sub>NO<sub>2</sub> ring system. *Nat. Chem.* 9, 571–577 (2017).
- 27. Scheidt, K. Amide bonds made in reverse. Nature 465, 1020–1022 (2010).
- Goswami, A. & Van Lanen, S. G. Enzymatic strategies and biocatalysts for amide bond formation: tricks of the trade outside of the ribosome. *Mol. Biosyst.* 11, 338–353 (2015).
- Petchey, M. et al. The broad aryl acid specificity of the amide bond synthetase McbA suggests potential for the biocatalytic synthesis of amides. Angew. Chem. Int. Edn 57, 11584–11588 (2018).
- Wood, A. J. L. et al. Adenylation activity of carboxylic acid reductases enables the synthesis of amides. Angew. Chem. Int. Edn 56, 14498–14501 (2017).
- Pattabiraman, V. R. & Bode, J. W. Rethinking amide bond synthesis. Nature 480, 471–479 (2011).
- Pérombelon, M. C. M. Potato diseases caused by soft rot *Erwinias*: an overview of pathogenesis. *Plant Pathol.* 51, 1-12 (2002).
- Bell, K. S. et al. Genome sequence of the enterobacterial phytopathogen *Erwinia* carotovora subsp. atroseptica and characterization of virulence factors. *Proc. Natl Acad. Sci. USA* **101**, 11105–11110 (2004).
- Bottini, R., Fulchieri, M., Pearce, D. & Pharis, R. P. Identification of gibberellins A<sub>1</sub>, A<sub>3</sub> and iso-A<sub>3</sub> in cultures of Azospirillum lipoferum. Plant Physiol. **90**, 45–47 (1989).
- Schüler, G. et al. Coronalon: a powerful tool in plant stress physiology. FEBS Lett. 563, 17-22 (2004).
- Shockey, J. M., Fulda, M. S. & Browse, J. Arabidopsis contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. Plant Physiol. 132, 1065–1076 (2003).
- Stuhlsatz-Krouper, S. M., Bennett, N. E. & Schaffer, J. E. Substitution of alanine for serine 250 in the murine fatty acid transport protein inhibits long chain fatty acid transport. J. Biol. Chem. 273, 28642–28650 (1998).
- Nocek, B. P. et al. Structural insights into substrate selectivity and activity of bacterial polyphosphate kinases. ACS Catal. 8, 10746–10760 (2018).
- Sheldon, R. A. Cross-linked enzyme aggregates (CLEAs): stable and recyclable biocatalysts. *Biochem. Soc. Trans.* 35, 1583–1587 (2007).

- Zhang, L. et al. a-Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. J. Med. Chem. 63, 4562–4578 (2020).
- Boras, B. et al. Discovery of a novel inhibitor of coronavirus 3CL protease for the potential treatment of COVID-19. Preprint at https://www.biorxiv.org/content/10.1101/2020.09.12.29 3498v3 (2020).
- 42. Zhou, H.-J. et al. Design and synthesis of an orally bioavailable and selective peptide epoxyketone proteasome inhibitor (PR-047). J. Med. Chem. **52**, 3028–3038 (2009).
- Chen, C.-S., Fujimoto, Y., Girdaukas, G. & Sih, C. J. Quantitative analyses of biochemical kinetic resolutions of enantiomers. J. Am. Chem. Soc. 104, 7294–7299 (1982).
- Jervis, P. J., Amorim, C., Pereira, T., Martins, J. A. & Ferreura, P. M. T. Exploring the properties and potential biomedical applications of NSAID-capped peptide hydrogels. Soft Matter 16, 10001–10012 (2020).
- Tiwari, A. D. et al. Microwave assisted synthesis and QSAR study of novel NSAID acetaminophen conjugates with amino acid linkers. Org. Biomol. Chem. 12, 7238–7249 (2014).

- Andexer, J. N. & Richter, M. Emerging enzymes for ATP regeneration in biocatalytic processes. ChemBioChem 16, 380–386 (2015).
- Strohmeier, G. A., Eiteljorg, I. C., Schwarz, A. & Winkler, M. Enzymatic one-step reduction of carboxylates to aldehydes with cell-free regeneration of ATP and NADPH. *Chem. Eur. J.* 25, 6119–6123 (2019).
- Philpott, H. K., Thomas, P. J., Tew, D., Fuerst, D. E. & Lovelock, S. L. A versatile biosynthetic approach to amide bond formation. *Green Chem.* 20, 3426–3431 (2018).
- Hara, R., Hirai, K., Suzuki, S. & Kino, K. A chemoenzymatic process for amide bond formation by an adenylating enzyme-mediated mechanism. Sci. Rep. 8, 2950 (2018).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

#### **Reporting summary**

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

#### **Data availability**

Nucleotide sequences for the mutants generated as part of this study are available in Supplementary Information. Other nucleotide sequences for the enzymes used in this study were obtained from GenBank, and their accession numbers are provided within the paper or in Supplementary Information. The original materials and data that support the findings of this study are either available within the paper or are available from the corresponding author upon reasonable request. Crystallographic coordinates of wild type and mutant PbCfaL have been deposited in the Protein Data Bank as 7A91 (wild type) and 7A9J (R395G mutant). Acknowledgements We thank the BBSRC (grants BB/K002341/1 and BB/N023536/1) and Syngenta for funding. F.W. was supported by the China Scholarship Council (grant no. 201806155100) and L.B. was funded by the Deutsche Forschungsgemeinschaft (DFG, grant BE 7054/1). The Michael Barber Centre for Collaborative Mass Spectrometry provided access to MS instrumentation. We also thank J. Vincent and N. Mulholland (Syngenta) for helpful discussions in the early stages of the project, and N. J. Turner (University of Manchester) for kindly providing the CHU plasmid. We also thank Diamond Light Source for beamtime access on i03 and i04-1 (proposal mx17773-56 and 76).

Author contributions M.W. and J.M. designed experiments; M.W., M.R., F.W., L.B. and D.F. carried out the experiments and provided additional experiment design; C.L. performed crystallographic studies. M.W. and J.M. wrote the manuscript. J.M. led the study.

Competing interests The authors declare no competing interests.

#### Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03447-w.

Correspondence and requests for materials should be addressed to J.M. Peer review information Nature thanks Francesca Paradisi and the other, anonymous,

reviewer(s) for their contribution to the peer review of this work.

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



**Extended Data Fig. 1** | **Melting point determination of CfaL enzymes.** Melting point temperatures ( $T_m$ ) of the CfaLs in this study, obtained using a fluorescence-based assay conducted in a Bio-Rad CFX Connect qPCR machine. Higher  $T_m$  indicates improved thermal stability. The  $T_m$  is calculated as the lowest point when plotting the negative derivative of RFU (relative fluorescence units) as a function of temperature (dT), versus the temperature (degrees Celsius).





Extended Data Fig. 2 | Rational mutagenesis of PbCfaL. a, Structural comparison between PbCfaL (left) and the mutant PbCfaL(R395G) (right). R395 (circled) of PbCfaL (PDB ID 7A9I) is in the hinge region between the N-terminal domain (blue) and the flexible C-terminal domain (red). In PbCfaL(R395G) (PDB ID 7A9J) this large arginine residue is replaced by a much smaller glycine (circled) that is found in the other members of the CfaL family and many other similar ANL ligases. The overall structure of this mutant exhibits no other substantial structural difference from that of the wild type. **b**, Overlay of PbCfaL with three published ATP-dependent ligase structures (in ellipse) showing the conserved ATP binding location. When superimposed, PbCfaL (PDB ID 7A9I, blue), McbA (AMP bound, PDB ID 6SQ8, red), GrsA (ATP bound, PDB ID 1AMU, green) and AuaEII (anthranoyl-AMP bound, PDB ID 4WV3, light brown) show the conserved location of ATP binding. The corresponding loop in PbCfaL (inset, arrowed) is larger than in the other structures which may affect ATP binding. The location of this region within the structure of PbCfaL (grey) is also shown for reference. Structural alignment was performed using Chimera (version 1.14) MatchMaker.



#### Extended Data Fig. 3 | Synthesis and NMR spectra of crude product 64. a, Preparative-scale synthesis of 64 from 10 and L-lle catalysed by PbCfaL(R395G/A294P) (lysate), with either the addition of ATP (87% isolated

yield) or recycling the endogenous ATP present in the lysate using the kinase (CHU) and polyphosphate (polyP) (52% isolated yield). **b**, <sup>1</sup>H-NMR spectrum of crude product **64.** 



Extended Data Fig. 4 | Catalysed conversion of 10 and t-lle to give 64. a, Catalysed by PbCfaL (R395G/A294P) CLEAs. Reactions (100 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 10 mM ATP, 1 mM 10, 3 mM L-isoleucine, 50 ml total volume) were run for 24 h; the CLEA (cross-linked enzyme aggregate) was then removed, washed, and reintroduced to an identical reaction. While activity was seen to reduce over the 5 days, the CLEA still retained high levels of productivity even after 5 recycles over 5 days, whereas cell lysates generally precipitated and lost all activity within 12 h. Although CfaL undergoes extensive conformational changes during catalysis, encapsulating it within CLEAs shows the potential of immobilization to extend the functional lifespan of the CfaL. More sophisticated immobilization techniques may have the potential to further retain activity. Conversion values were calculated from HPLC peak area ratios of product and starting materials, and represent means where n = 5, error bars denote s.d. **b**, Catalysed by purified PbCfaL(R395G/A294P), showing percentage conversions of **10** and L-lle in the presence of various solvents and at different concentrations. Conversion values were calculated from HPLC peak area ratios of product and starting materials and represent means where n = 3, error bars denote s.d.



Extended Data Fig. 5 | LCMS analysis and extracted-ion chromatograms (EICs) of fragments of telaprevir and bortezomib synthesized by PbCfaL(R395G/A294P). a, Proposed route towards the antiviral agent telaprevir by CfaL (see reaction at top). The expected product of the reaction, 110, was detected by LCMS (top trace, expected *m*/z 262.1197, observed 262.1193 [M-H]<sup>-</sup>). Additional peaks consistent with dipeptide 111, formed from condensation of two cyclohexylglycines, **58**, were also detected (bottom trace, **111a** and **111b**, expected *m*/z 295.2027, observed 295.2036 [M-H]<sup>-</sup>). Although CfaL are highly selective for L-amino acid substrates, the appearance of two products of the same mass suggests formation of diastereomers, which may be due to a lack of enantioselectivity in the adenylation step forming the acyl donor when racemic cyclohexylglycine (**58**) is used. This indicates that **58** can function as both a carboxylic acid and an amine donor. **b**, Proposed route towards anti-cancer agent bortezomib via the synthesis of **112** by CfaL (see reaction at top). The expected product of the reaction was detected by LCMS (top trace, expected *m/z* 270.0884, observed 270.0878 [M-H]<sup>-</sup>). An additional peak consistent with an L-Phe dipeptide (**113**) was also detected (bottom trace, expected *m/z* 311.1401). This indicates that L-Phe can function as both acyl donor and amine acceptor. **c**, The reaction between carboxylic acid substrate (**9**), which is a good substrate for the enzyme, and cyclohexylglycine (**58**) gives only the desired product (**114**, top trace, expected *m/z* 274.1449, observed 274.1460 [M-H]<sup>-</sup>). No cyclohexylglycine homodimer (dipeptide **111**) was evident in this case, indicating that homocoupling of **58** only takes place when carboxylic acid (acyl donor) substrates that are not well accepted by CfaL are used.

### Extended Data Table 1 | Table of kinetic parameters for adenylation reactions

	O CO <sub>2</sub> H	CO-H	0 7 3 CO <sub>2</sub> H	0 	
Substrate	1	7	(3 <i>R</i> ,7 <i>R</i> )- <b>5</b>	(3S,7S)- <b>5</b>	(±)- <b>5</b>
$V_{max}$ (µM <sup>-1</sup> min <sup>-1</sup> )	10.1 (± 0.4)	5.6 (± 0.1)	2.16 (± 0.1)	1.75 (± 0.06)	1.9 (± 0.04)
K <sub>m</sub> (mM)	0.2 (± 0.02)	0.07 (± 0.01)	0.08 (± 0.02)	0.12 (± 0.02)	0.18 (± 0.01)
<u>k<sub>cat</sub></u> (min⁻¹)	40.3 (±0.4)	22.6 (± 0.1)	8.63 (± 0.1)	7.0 (± 0.06)	7.55 (± 0.04)
$k_{cat}/K_m (min^{-1} mM^{-1})$	199 (±0.4)	343 (± 0.1)	112.1 (± 0.1)	58.9 (± 0.07)	41.4 (± 0.04)

These reactions were catalysed by SsCfaL and used substrates CFA (1), aromatic variant 7, enantiomers of *trans*-jasmonic acid 5 ((3R,7R)-5 and (3S,7S)-5) and racemic ( $\pm$ )-5. The JA samples also contain a small amount of *cis*-epimer. Michaelis–Menten plots can be found in Supplementary Fig. 7. Values represent means, error denotes s.d., with  $n \ge 3$ .

#### Extended Data Table 2 | Percentage yield of ligation reactions of carboxylic acids

			Percentage Yield			
Carboxylic Acid	SsCfaL	AlCfaL	PbCfaL	PbCfaL (R395G)	PbCfaL (R395G/A294P)	Product
8	75 (± 2)	>99 (± 1)	30 (± 1)	88 (± 0.2)	98 (± 2)	78
9	>99 (± 5)	>99 (± 4)	>99 (± 4)	>99 (± 1)	>99 (± 6)	79
10	>99 (± 8)	>99 (± 5)	>99 (± 16)	>99 (± 4)	>99 (± 5)	64
11	>99 (± 1)	>99 (± 4)	>99 (± 2)	>99 (± 3)	>99 (± 1)	80
12	28 (± 0.2)	44 (± 0.4)	4 (± 0.01)	21 (± 0.2)	23 (± 0.1)	81
13	41 (± 0.3)	10 (± 0.3)	3 (± 0)	5 (± 0.1)	5 (± 0.1)	82
14	>99 (± 1)	>99 (± 0.3)	3 (± 0)	13 (± 1)	10 (± 0.5)	83
15	27 (± 0.4)	3 (± 0.2)	ND	ND	0.4 (± 0.04)	84
16	28 (± 8)	63 (± 5)	43 (± 8)	41 (± 6)	59 (± 2)	85*
17	>99 (± 4)	96 (± 12)	64 (± 0.2)	>99 (± 2)	>99 (± 8)	86
18	>99 (± 9)	>99 (± 2)	>99 (± 1)	>99 (± 1)	>99 (± 4)	87
19	43 (± 1)	33 (± 1)	4 (± 0.1)	22 (± 1)	43 (± 1)	88
20	>99 (± 2)	>99 (± 1)	>99 (± 14)	>99 (± 1)	>99 (± 7)	89
21	>99 (± 3)	85 (± 2)	81 (± 2)	>99 (± 5)	>99 (± 10)	90
22	>99 (± 13)	94 (± 2)	>99 (± 2)	>99 (± 3)	>99 (± 3)	91
23	81 (± 5)	67 (± 3)	57 (± 0.4)	79 (± 2)	77 (± 6)	92
24	>99 (± 8)	56 (± 7)	18 (± 3)	58 (± 2)	72 (± 1)	93
25	95 (± 5)	91 (± 4)	51 (± 5)	>99 (± 4)	99 (± 3)	94
26	63 (± 5)	21 (± 4)	1 (± 0.4)	10 (± 1)	20 (± 6)	95
27	82 (± 14)	31 (± 0.1)	30 (± 1)	76 (± 1)	80 (± 6)	96
28	0.2 (± 0)	5 (± 0.1)	ND	1 (± 0.03)	3 (± 2)	97
29	27 (± 1)	21 (± 0.4)	4 (± 0.2)	21 (± 0.4)	19 (± 0.1)	98
30	32 (± 0.2)	81 (± 15)	17 (± 0.1)	60 (± 3)	62 (± 1)	99
31	50 (± 3)	53 (± 10)	1 (± 0.1)	9 (± 2)	19 (± 8)	100
32	8 (± 1)	44 (± 8)	ND	2 (± 1)	2 (± 1)	101
33	74 (± 2)	66 (± 9)	4 (± 0.3)	26 (± 1)	39 (± 1)	102
34	3 (± 2)	53 (± 3)	43 (± 3)	>99 (± 7)	>99 (± 8)	103
35	8 (± 0.5)	58 (± 12)	0.2 (± 0)	10 (± 2)	14 (± 1)	104
36	73 (± 8)	39 (± 6)	1 (± 0.2)	8 (± 0.4)	10 (± 3)	105
37	<1	23 (± 1)	5 (± 0.2)	74 (± 1)	48 (± 8)	106
38	5 (± 4)	64 (± 1)	51 (± 2)	>99 (± 8)	>99 (±7)	107
39	24 (± 6)	25 (± 2)	17 (± 3)	44 (± 1)	37 (± 3)	72
40	4 (± 2)	3 (± 1)	4 (± 1)	25 (± 8)	34 (± 9)	74
41	ND	54 (± 1)	ND	ND	0.1 (± 0.01)	71
42	69 (± 5)	>99 (± 12)	54 (± 9)	77 (± 8)	59 (± 7)	76
43	39 (± 2)	65 (± 6)	1 (± 0)	1 (± 0.03)	2 (± 0.02)	75
44	18 (± 4)	12 (± 9)	6 (± 1)	47 (± 2)	53 (± 7)	73
45	<1	12 (± 1)	0.1 (± 0.1)	13 (± 1)	2 (± 2)	108
46	0.3 (± 0.02)	2 (± 0.1)	0.2 (±0)	3 (± 0.1)	7 (± 1)	109

Table shows yields of ligation of carboxylic acids (8–46) with L-lle to give the corresponding amides (64, 71–76, 78–109) catalysed by CfaLs. Assays were carried out with wild-type and engineered CfaL enzymes ( $25 \mu$ M), carboxylic acids (8–46) (2 mM) and L-lle (5 mM). Conversions were almost all calculated by HPLC or liquid chromatography mass spectrometry (LCMS) using standard curves of synthesized standards following 20 h incubation (see footnote a for the sole exception). Values represent means, error denotes s.d., with n = 3.

ND, not detected.

<sup>a</sup>Yield of compound **85** was calculated using HPLC peak area ratios of product and starting material.

### Extended Data Table 3 | Ligation of 9 to a range of proteinogenic amino acids

			Percentage Yield		
Compound no	SsCfaL	AlCfaL	PbCfaL	PbCfaL (R395G)	PbCfaL (R395G/A294P)
L-Isoleucine	100 (± 0)	100 (± 0)	92 (± 1)	100 (± 0)	100 (± 0)
D-Isoleucine	ND	ND	ND	ND	ND
L-Valine	100 (± 0)	100 (± 0)	53 (± 0.1)	63 (± 2)	55 (± 0.5)
L-leucine	58 (± 0.2)	91 (± 1)	40 (± 0.5)	49 (± 1)	40 (± 1)
L-Methionine	63 (± 0.5)	100 (± 0)	20 (± 0.5)	25 (± 0.5)	20 (± 4)
L-Alanine	13 (± 2)	27 (± 13)	ND	ND	ND
L-Threonine	23 (± 1)	70 (± 1)	7 (± 0.2)	7 (± 0.2)	5 (± 0.2)
L-Cysteine	39 (± 1)	46 (± 3)	10 (± 0)	9 (± 0.4)	10 (± 0.2)
D-Cysteine	5 (± 0.4)	5 (± 0.4)	10 (± 0)	6 (± 0.1)	7 (± 0.4)
L-Serine	10 (± 6)	20 (± 2)	ND	ND	ND
L-Asparagine	16 (± 4)	42 (± 1)	<1	<1	<1
L-Glutamine	4 (± 0.4)	ND	ND	ND	ND
L-Tryptophan	34 (± 4)	<1	18 (± 3)	22 (± 4)	17 (± 3)
L-Tyrosine	8 (± 3)	11 (± 0.5)	5 (± 0.4)	8 (± 2)	6 (± 2)
L-Phenylalanine	5 (± 2)	14 (± 0.4)	4 (± 0.2)	6 (± 2)	3 (± 0.2)
L-Arginine	14 (± 0.6)	ND	ND	ND	ND
L-Histidine	14 (± 1)	ND	ND	ND	ND
L-Lysine	ND	ND	ND	ND	ND
L-Aspartic Acid	ND	ND	ND	ND	ND
L-Glutamic Acid	ND	ND	ND	ND	ND
Glycine	ND	ND	ND	ND	ND
L-Proline	ND	ND	ND	ND	ND

Table shows yields of ligation of carboxylic acid  $\mathbf{9}$  to a range of proteinogenic amino acids catalysed by CfaLs. Assays were carried out with wild-type and engineered CfaL enzymes (5  $\mu$ M), carboxylic acid  $\mathbf{9}$  (1 mM) and amino acids (2 mM). Conversion of  $\mathbf{9}$  to amide products was calculated using HPLC peak area ratios of product and starting material following 20 h of incubation. Low level of activity with D-cysteine possibly occurs via native chemical ligation mechanisms. Values represent means, error denotes s.d., with n = 3. ND, not detected.

#### Extended Data Table 4 | Ligation of 9 to a range of non-proteinogenic amino acids

				Percentage Yield		
Compound n	0	SsCfaL	AlCfaL	PbCfaL	PbCfaL (R395G)	PbCfaL (R395G/A294P)
L-alloIsoleuci	ne	100 (± 0)	100 (± 0)	56 (± 1)	67 (± 0.5)	59 (± 2)
D-alloIsoleuc	ine	ND	ND	ND	ND	ND
L-Dab (47)		15 (± 0.8)	12 (± 0.6)	6 (± 0.3)	4 (± 0.1)	3 (± 0.1)
L-Orn (48)		0.9 (± 0.2)	11 (± 0.2)	7 (± 1)	4 (± 0.01)	3 (± 0.05)
49 \\\	Сон	28 (± 0.2)	100 (± 0)	2 (± 0.02)	4 (± 0.04)	14 (± 0.03)
50 \s	Сон	22 (± 0.1)	100 (± 0)	1 (± 0.4)	1 (± 0.4)	49 (± 0.6)
51 V	Сон	65 (± 2)	100 (± 0)	15 (± 0.05)	38 (± 0.1)	76 (± 0.4)
52 ~~~~	он Н	31 (± 0.5)	100 (± 0)	11 (± 0.2)	30 (± 0.02)	61 (± 0.1)
53 ×	`он	79 (± 0.4)	100 (± 0.1)	11 (± 0.1)	33 (± 0.2)	39 (± 0.4)
54 A	рн	20 (± 3)	69 (± 3)	ND	ND	ND
55 A	ЮН	44 (± 0.4)	96 (± 0.3)	4 (± 0)	13 (± 0.2)	39 (± 0.2)
56 V	он t	40 (± 0.2)	100 (± 0)	12 (± 0.7)	36 (± 0.2)	72 (± 0.5)
57 J	OH 2	12 (± 0.5)	75 (± 0.9)	13 (± 0.3)	31 (± 0.3)	61 (± 0.3)
58 Q	Сон	31 (± 0.2)	32 (± 0.5)	19 (± 0.2)	44 (± 0.2)	62 (± 1)
59 () <sup>N</sup>	ОН На	4 (± 0.1)	81 (± 8)	2 ± (0.04)	4 (± 0.1)	16 (± 0.5)
60 Q	ОН	9 (± 0.2)	19 (± 0.1)	0.7 (± 0.02)	2 (± 0.04)	3 (± 0.04)
61	OH NH <sub>2</sub> OH	7 (± 1)	36 (± 0.5)	2 (± 0.1)	5 (± 0.2)	15 (± 1)

Table shows yields of ligation of carboxylic acid 9 to a range of non-proteinogenic amino acids catalysed by CfaLs. Assays were carried out with wild-type and engineered CfaL enzymes (5  $\mu$ M), carboxylic acid 9 (1 mM) and amino acids (2 mM). Conversions of 9 to amide products were calculated using HPLC peak area ratios of product and starting material following 20 h of incubation. Values represent means, error denotes s.d., with n = 3. ND, not detected.

## nature research

Corresponding author(s): Jason Micklefield

Last updated by author(s): Jan 20, 2021

### **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 our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

### **Statistics**

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Сог	nfirmed
	$\boxtimes$	The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
	$\boxtimes$	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
$\boxtimes$		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.
$\boxtimes$		A description of all covariates tested
$\boxtimes$		A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
$\boxtimes$		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)
$\boxtimes$		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.
$\boxtimes$		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
$\boxtimes$		For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
$\boxtimes$		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 <u>availability of computer code</u>				
Data collection	Shimadzu LabSolutions, BioTek Gen5 v2, Thermo Xcalibur, Water UNIFI, GE healthcare UNICORN, BioRad CFX connect			
Data analysis	Thermo Xcalibur, Shimadzu LabSolutions, BioTek Gen5 v2, Clustal OMEGA, MestReNova, ChemDraw Professional, Chimera, GraphPad Prism, XtalPred-RF, GE Healthcare UNICORN, SigmaPlot, BioRad CFX connect			

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 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

Data Availability: Nucleotide sequences for the mutants generated as part of this study are available within supplementary information. Other nucleotide sequences for the enzymes used in this study were obtained from GenBank and their accession numbers are provided within the paper or its supplementary information files. The original materials and data that support the findings of this study are either available within the paper or are available from the corresponding author upon reasonable request. Crystallographic coordinates have been deposited in the Protein Data Bank as 7A9I and 7A9J.

### 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 <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

### Life sciences study design

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

Sample size	No sample size calculation was used. Quantitative assays were performed in triplicate and mean and standard deviation values calculated. For some kinetic experiments more than 3 replicates were taken to ensure accuracy, and mean and standard deviation values calculated. All replicates were included in the analysis.
Data exclusions	no data was excluded
Replication	Experiments were carried out in at least triplicate and standard deviation provided. Each measurement is from a separate experiment but carried out at the same time.
Randomization	No randomization was performed during this study as it was not applicable for our experiments
Blinding	no blinding was involved in this study as it does not involve animal or human subjects or group allocation.

### Reporting for specific materials, systems and methods

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		
$\boxtimes$	Antibodies		
$\boxtimes$	Eukaryotic cell lines		
$\boxtimes$	Palaeontology and archaeology		
	Animals and other organisms		
$\boxtimes$	Human research participants		
$\boxtimes$	Clinical data		
$\boxtimes$	Dual use research of concern		

### n/a Involved in the study

- ChIP-seq
- MRI-based neuroimaging

### Animals and other organisms

 Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

 Laboratory animals
 no laboratory animals were used for this study

 Wild animals
 No wild animals were used for this study

 Field-collected samples
 No field collected samples were used for this study

 Ethics oversight
 No oversight is required for use of E. coli

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