

VIP Very Important Paper



BioWF: A Naturally-Fused, Di-Domain Biocatalyst from Biotin Biosynthesis Displays an Unexpectedly Broad Substrate Scope

Shona M. Richardson,^[a] Peter J. Harrison,^[a, b] Michael A. Herrera,^[a] Menglu Wang,^[a] Rebecca Verez,^[a] Gustavo Perez Ortiz,^[a] and Dominic J. Campopiano^{*[a]}

The carbon backbone of biotin is constructed from the C₇ diacid pimelate, which is converted to an acyl-CoA thioester by an ATP-dependent, pimeloyl-CoA synthetase (PCAS, encoded by BioW). The acyl-thioester is condensed with L-alanine in a decarboxylative, Claisen-like reaction to form an aminoketone (8-amino-7-oxononanoic acid, AON). This step is catalysed by the pyridoxal 5'-phosphate (PLP)-dependent enzyme (AON synthase, AONS, encoded by BioF). Distinct versions of *Bacillus subtilis* BioW (BsBioW) and *E. coli* BioF (EcBioF) display strict substrate specificity. In contrast, a BioW-BioF fusion from *Corynebacterium amycolatum* (CaBioWF) accepts a wider range of mono- and di-fatty acids. Analysis of the active site of the

BsBioW : pimeloyl-adenylate complex suggested a key role for a Phe (F192) residue in the CaBioW domain; a F192Y mutant restored the substrate specificity to pimelate. This surprising substrate flexibility also extends to the CaBioF domain, which accepts L-alanine, L-serine and glycine. Structural models of the CaBioWF fusion provide insight into how both domains interact with each other and suggest the presence of an intra-domain tunnel. The CaBioWF fusion catalyses conversion of various fatty acids and amino acids to a range of AON derivatives. Such unexpected, natural broad substrate scope suggests that the CaBioWF fusion is a versatile biocatalyst that can be used to prepare a number of aminoketone analogues.

Introduction

Enzyme fusions are bi-/multi-functional biocatalysts that can complete a series of sequential/cascade reactions to form a desired product. Such multi-step reactions are typically found in biosynthetic pathways and can generate natural products of stunning complexity. An increase in enzymatic cascade efficiency results through close proximity of active sites, coupled with a high local concentration of substrates. This also prevents the build-up of intermediates that could inhibit enzymatic activity.^[1] Examples of natural fusions include the fatty acid synthases (FAS), polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS). These multidomain megasynthases adopt modular assembly lines composed of varying enzymatic domains that can produce a myriad of primary metabolites and secondary natural products.^[2–5] Despite their abundance in nature, there is a growing use of synthetically designed fusion

systems, achieved by expressing multiple genes as a single transcript.^[6–8] One major drawback of these “beads-on-a-string” systems is linker design, with few investigations into their structure and function to date.^[9] By exploiting pre-existing natural fusions this need for linker understanding is not required, with the fusion systems having undergone extensive evolutionarily optimisation. Therefore, natural multi-domain fusion systems are ideal candidates for biocatalytic transformations.

Biotin is an unusual bicyclic vitamin constructed through a series of interesting reactions. Biotin production can be divided into two stages. Firstly, formation of an unusual pimeloyl-thioester from the C₇ pimelic di-acid, with the acyl chain linked to either CoASH or an acyl carrier protein (ACP). Two pathways are known for the biosynthesis of this acyl-thioester with each using a pair of coupled enzymes. The more common methyltransferase/hydrolase (BioC/BioH) pathway is found in *E. coli* and hijacks fatty acid biosynthesis to generate pimeloyl-ACP. In contrast, the much rarer BioI and BioW-dependent routes were discovered in *Bacillus subtilis*.^[10–12] The BioI and BioW can act together to produce pimeloyl-thioesters from long chain fatty acid precursors. Surprisingly, the *B. subtilis* BioI is a cytochrome P450 that oxidatively cleaves long chain acyl-ACPs, forming the pimeloyl-ACP intermediate.^[13,14] Alternatively, *B. subtilis* also encodes a pimeloyl-CoA synthetase (PCAS, BioW) that catalyses direct formation of pimeloyl-CoA in an ATP-dependent reaction.^[15,16] The genes encoding these proteins reside in the bio-operon or biosynthetic gene clusters (BGCs); bioABFCD for *E. coli* and bioWAFDBI for *B. subtilis*.^[17]

Following pimeloyl-CoA/ACP production, a conserved four-enzyme pathway ensues. Pimeloyl-CoA/ACP is condensed with

[a] Dr. S. M. Richardson, Dr. P. J. Harrison, M. A. Herrera, Dr. M. Wang, R. Verez, Dr. G. P. Ortiz, Prof. Dr. D. J. Campopiano
School of Chemistry University of Edinburgh
David Brewster Road, Edinburgh, EH9 3FJ (UK)
E-mail: Dominic.Campopiano@ed.ac.uk

[b] Dr. P. J. Harrison
Diamond Light Source Ltd.
Harwell Science & Innovation Campus, Didcot, OX11 0DE (UK)

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cbic.202200171>

 This article is part of a Special Collection dedicated to the Biotrans 2021 conference. Please see our homepage for more articles in the collection.

 © 2022 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

L-Ala to form 8-amino-7-oxononanoic acid (AON). This is catalysed by the pyridoxal 5'-phosphate (PLP)-dependent α -oxoamine synthase (AONS, encoded by BioF), through a decarboxylative, Claisen-like condensation. The AON is then transaminated by the PLP-dependent 7,8-diaminononanoic acid (DAN) synthase (BioA) to form diaminononanoic acid (DAN). The ureido ring of dethiobiotin is formed by insertion of CO₂ between the diamine of DAN, catalysed by the ATP-dependent dethiobiotin (DTB) synthetase (BioD). The final step of biotin formation incorporates sulfur. This step is catalysed by biotin synthase (BioB), a radical SAM enzyme with two Fe-S clusters. The sulfur originates from a deeply buried [2Fe-2S]²⁺ cluster, and generates the final thiophene ring.

To date, only a few natural fusions of biotin biosynthesis enzymes have been identified. For example, investigations into biotin production in *Arabidopsis thaliana* led to the discovery of a natural fusion of the BioA and BioD genes.^[18] The AtBioAD fusion was shown to catalyse consecutive transamination and ring-forming reactions. Similarly, the Cronan group reported a natural fusion of the BioW and BioF genes in *Desulfosporosinus orientis* (Figure 1).^[19] This study showed that the expressed *D. orientis* BioWF fusion was able to complement a biotin auxotrophic *E. coli* BioF mutant when grown on exogenous pimelic acid. Although providing support that both domains of the fusion were active, the recombinant fusion biocatalyst was not isolated.

To explore the use of fused domains as viable biocatalysts, we identified a BioWF fusion in *Corynebacterium amycolatum* SK46. Here we describe the isolation and characterization of the CaBioWF fusion and show that the recombinant biocatalyst can catalyse the sequential conversion of pimelic acid to pimeloyl-CoA, followed by formation of the predicted AON product. Furthermore, we discovered that the CaBioWF fusion displays a surprisingly broad substrate scope. The versatile biocatalyst accepts a range of mono- and di-carboxylic acids, activates them as acyl-CoA thioesters, and then catalyses formation of a range of α -oxoamines with three different amino acids. Structural and sequence analysis of the CaBioWF fusion suggested that a phenylalanine residue in the predicted BioW active site is the key to its broad activity. The BioF domain of the fusion displays features common to other members of the α -oxoamine synthase (AOS) family which can accept a range of amino acids. A structural model of the CaBioWF fusion provides insight into how the two domains interact. Our work suggests

that other natural fusion enzymes should be explored as potentially useful biocatalysts.

Results and Discussion

Cloning an unusual BioWF Fusion from *C. amycolatum* (CaBioWF)

An investigation into bacterial BioW homologues led to the discovery of a predicted natural fusion of the BioW and BioF domains in the organism *C. amycolatum* SK46 (Uniprot: E2MUP3, Figures S1 and S2), a non-sporulating Gram positive bacterium.^[20] The annotated CaBioWF gene was purchased as a synthetic clone from Genscript, cloned in a pET-28a expression plasmid. Recombinant protein expression and isolation was optimized in *E. coli* BL21 (DE3); however incubation of the purified, recombinant CaBioWF with the reaction components (pimelate, ATP, CoASH, L-alanine) yielded no product formation (data not shown). However, the addition of purified, active BsBioW to this reaction mixture led to AON production, detected by LC ESI-MS analysis.^[15] This suggested that the CaBioW domain of the fusion was inactive, whereas the CaBioF domain was able to catalyse conversion of the pimeloyl-CoA (formed by the additional BsBioW) to AON. The N-terminus was re-examined using information gained through detailed analysis of BsBioW (Uniprot: P53559, PDB: 5FLL) and the *Aquifex aeolicus* homologue (AaBioW, Uniprot: O67575, PDB: 5TV6) sequences and structures (Figures S3 and S4). This revealed a number of conserved residues preceding the annotated starting Met were missing from the initial CaBioWF (equivalent to Met12 in BsBioW).^[15,16] The predicted N-terminus also starts in the middle of key secondary structure (β -sheet) of BsBioW and therefore is likely to interfere with protein folding. Therefore, the N-terminal domain was extended and 7 amino acids were inserted through Gibson assembly cloning (Figure S5).

Characterisation of a fully active CaBioWF fusion

The expression of this new CaBioWF construct was first tested under a series of standard conditions, with protein production observed for each one tested (Figure S6). The protein was subsequently extracted and purified via nickel immobilised affinity chromatography (IMAC), followed by HisTag removal by

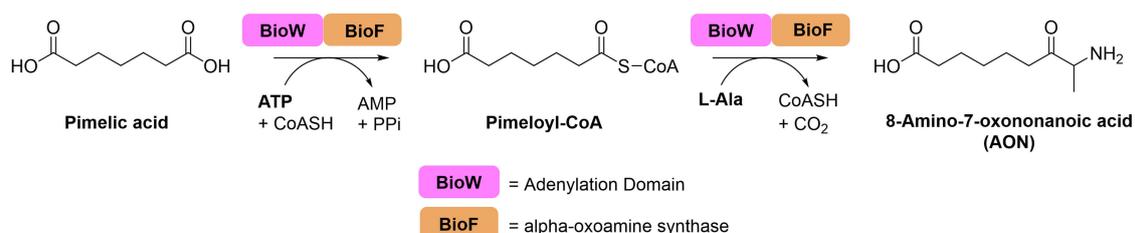


Figure 1. CaBioWF enzymatic reaction taking pimelic acid through the pimeloyl-CoA intermediate to form the final AON product.

TEV-protease and size exclusion chromatography (SEC), characterising the protein as a homodimer. A mass of 67432.45 ± 0.31 was obtained via LC ESI-MS analysis, corresponding to the untagged predicted mass of 67431.93 Da (Figure S7). After isolation of this new construct with a re-designed N-terminus, the activity of the *CaBioW* domain was once again tested.

Activity of the BioW domain of the *CaBioWF* fusion

This N-terminal BioW adenylation domain should catalyse the formation of pimeloyl-CoA in an ATP-dependent reaction. Product formation can be monitored using HPLC which can detect the acyl-CoA thioester at 260 nm. This assay was previously reported by Wang *et al.*^[15] The appearance of a peak at 17.3 min in the chromatogram signified the formation of pimeloyl-CoA, which was later confirmed by FT-ICR MS, with detection of an ion with $m/z = 932.16912$ Da, matching the predicted mass of 932.1674 ($[M + Na]^+$, $C_{28}H_{46}N_7O_{19}P_3S$) (Figure 2 and Figure S8). This confirmed that the re-engineered N-terminus recovered the *CaBioW* activity in the *CaBioWF* fusion.

The kinetics of the newly “re-activated” domain was studied using the coupled MesG assay originally reported by Aldrich *et al.* and Webb *et al.*^[21,22] When *CaBioWF* (0.1 μ M) was incubated with varying amounts of pimelic acid (0 to 1000 μ M), ATP

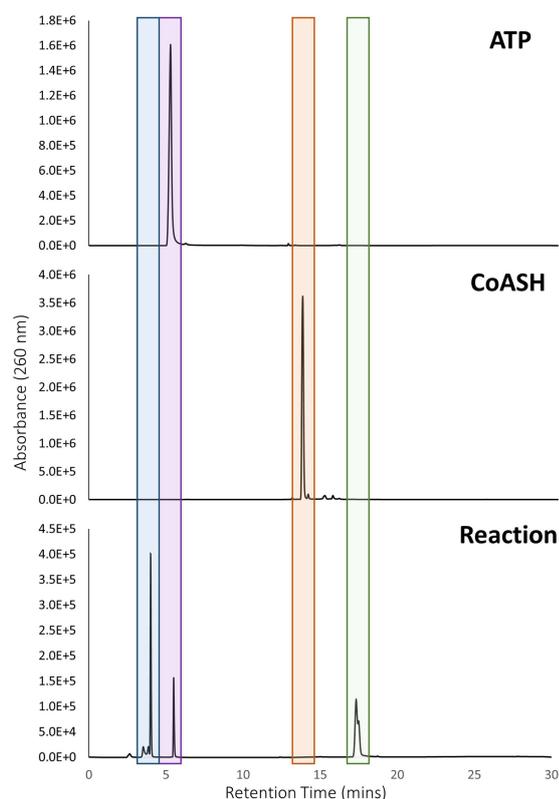


Figure 2. HPLC assay results for *CaBioWF* enzymatic reaction with DC₇ pimelic acid, ATP and CoASH leading to the formation of pimeloyl-CoA (green, 17.3 min) and released AMP (blue, 4.0 min). This is shown with the corresponding ATP (purple, 5.2 min) and CoASH (orange, 13.7 min) standards.

and CoASH, an increase in the absorbance at 360 nm was monitored. The kinetic data was fitted with a Michaelis-Menten curve to give a K_M value of $58.6 \pm 3 \mu$ M, and a k_{cat} of $0.21 s^{-1}$ for pimelic acid (Figure S9). This analysis confirms that the activity of the *CaBioW* domain of the fused system is comparable to the *BsBioW* homologue.

Activity of the BioF domain of the *CaBioWF* fusion

Having confirmed activity of the N-terminal BioW domain, we next tested the C-terminal *CaBioF* domain for its predicted AONS activity. This domain should catalyse the formation of AON from L-Ala and pimeloyl-CoA in a PLP-dependent reaction.^[23] Sequence alignments to the previously studied *EcBioF* identified the conserved, key PLP-binding residue as Lys484 (equivalent to *EcBioF* residue Lys236, Figure S10). PLP-binding can be monitored by changes in the characteristic UV-Vis spectrum.^[19,23] Upon titration of L-Ala (0–40 mM) into the PLP-bound *CaBioWF* enzyme we noted a change in absorbance from 250 nm to 500 nm, indicating the formation of the PLP:L-Ala external aldimine (Figure 3A). In order to monitor the changes as the reaction proceeds, pimelic acid and CoASH were then titrated into the mixture. With the addition of MgCl₂ and ATP, the *CaBioW* domain should produce the pimeloyl-CoA product, which can then be immediately utilised by the *CaBioF* domain. This is apparent in the UV-Vis spectrum, with changes in absorbance being observed up to 5 h after the addition (Figure S11). The binding of the amino acid can be quantified through the calculation of a dissociation constant (K_d). Again, L-Ala was titrated into the *CaBioWF* PLP-bound mixture at varying concentrations (0–40 mM). The changes in absorbance were measured after each addition and the absorbance at 425 nm used to calculate the K_d . This gave rise to a K_d of 1.83 ± 0.25 mM (Figure 3A).

In addition to the natural substrate L-Ala, binding of two other amino acids (L-Ser and Gly) was also observed and binding was quantified with K_d values of 9.03 ± 0.22 mM for L-Ser (Figure 3B) and 16.12 ± 1.95 mM for Gly (Figure 3C). Although the affinity for both of these amino acids is weaker than L-Ala, both amino acids are still accepted.

The full *CaBioWF* reaction

Having observed that the *CaBioF* domain could bind different amino acids, we next determined if the *CaBioWF* fusion is able to generate a number of AON analogues. Therefore, a series of full *CaBioWF* reactions, incubating pimelic acid with L-Ala, L-Ser or Gly, were conducted. To study the complete reaction, *CaBioWF* was incubated with the *CaBioW* reaction components and L-Ala at 30 °C for 5 h. The mixture was quenched by the addition of 1:1 (v:v) acetonitrile (MeCN) (0.01 % TFA) and the final AON product was analysed by LC ESI-MS analysis. An ion with $m/z = 188.1282$ was observed in the LC ESI-MS and corresponds to the predicted mass of AON at 188.1280 ($[M +$

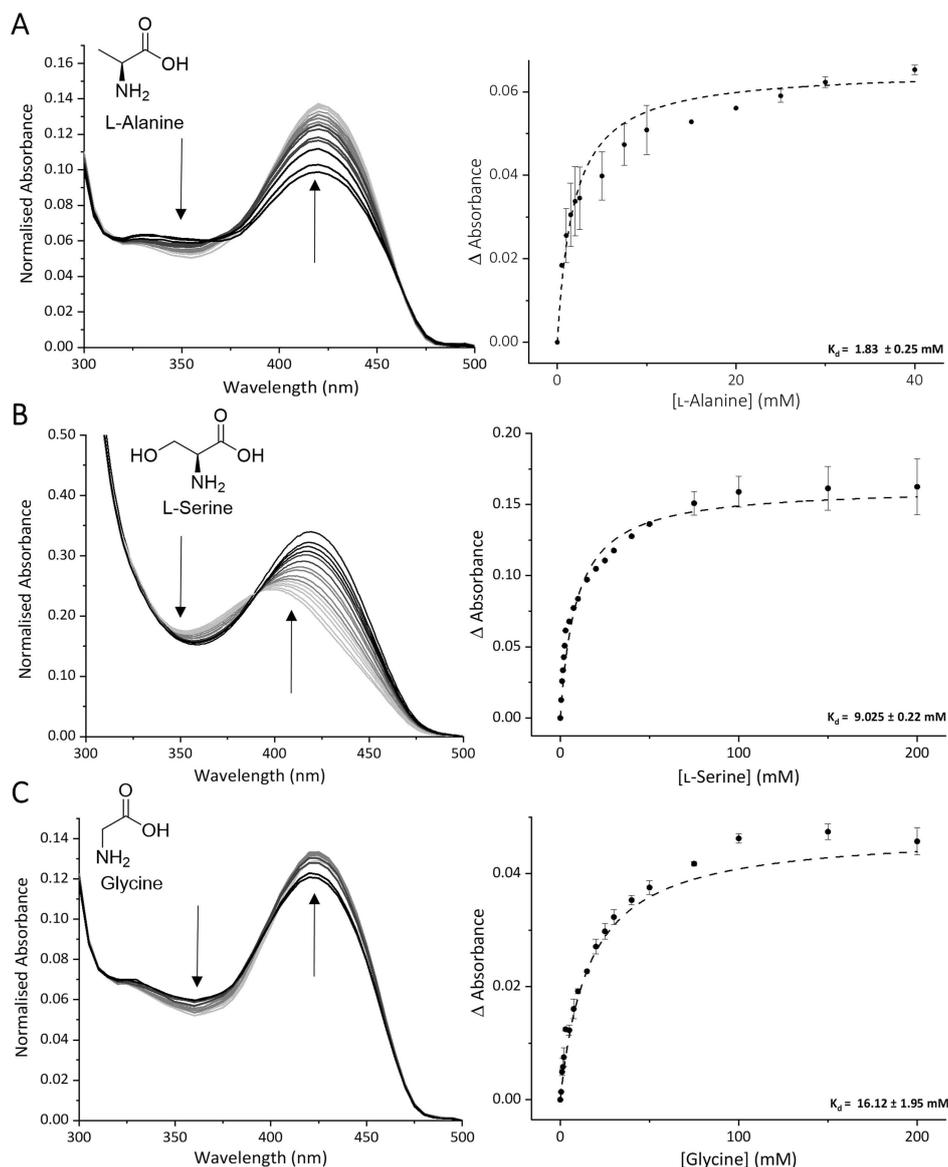


Figure 3. UV-Vis spectroscopy scan of the CaBioF domain PLP binding monitoring the changes after the addition of A) L-Ala (0–40 mM) and using the changes in the absorbance at 425 nm, a K_d curve was plotted and the dissociation constant of 1.83 ± 0.25 mM calculated. Also, after the addition of B) L-Ser (0–200 mM) a K_d of 9.03 ± 0.22 mM was calculated and for C) Gly (0–200 mM) a K_d of 16.12 ± 1.95 mM.

H^+ , $\text{C}_9\text{H}_{18}\text{NO}_3$, Figure S12), thereby confirming the cooperation of the two domains in the formation of the AON product.

Using a commercially-available AON standard, a calibration curve was constructed using HPLC and LC-MS analysis (Figure S13A and B). This allowed the amount of AON to be determined initially in a small-scale CaBioWF reaction. Two reactions, using either 1 mM or 0.5 mM CoASH, were performed, leading to the production of AON at concentrations of 0.76 mM and 0.79 mM respectively. These values correspond to a conversion 50.6 and 52.6% (based on pimelic acid). The negligible difference between the two reactions emphasizes the ability of CaBioWF to recycle CoASH. We also explored the utility of CaBioWF at a larger scale. A 10 mL reaction, using 0.5 mM CoASH and 5 μM biocatalyst, resulted in an AON

conversion of 51.3% (0.77 mM) over 5 h, similar to those obtained on a smaller scale (Figure S13C). CaBioWF is therefore a promising biocatalyst with an inbuilt CoASH recycling system.,

We then repeated this analysis using pimelic acid and both L-Ser and Gly. In both cases the corresponding aminoketone product can be identified, with $m/z = 204.1267$ and $m/z = 174.1132$ matching the predicted mass of 204.1231 ($[\text{M} + \text{H}]^+$, $\text{C}_9\text{H}_{18}\text{NO}_4$) and 174.1125 ($[\text{M} + \text{H}]^+$, $\text{C}_8\text{H}_{16}\text{NO}_3$) respectively for the L-Ser and Gly derived AON derivatives (Figure 4, S12A and C).

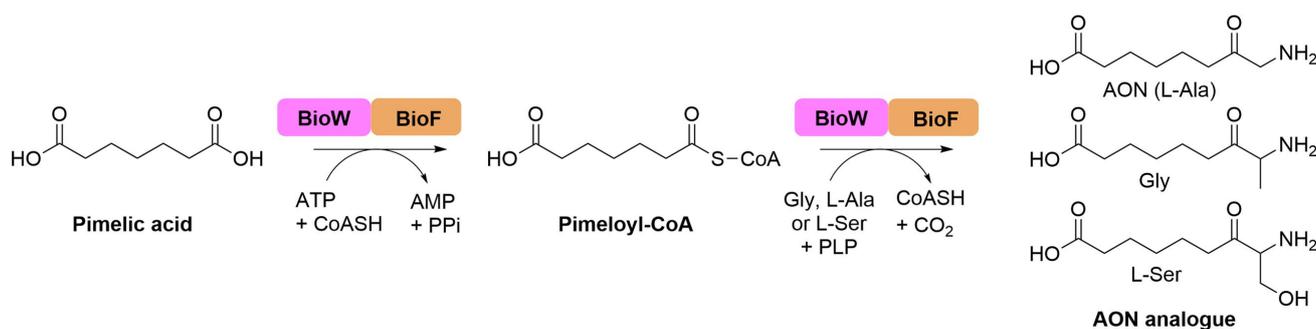


Figure 4. LC ESI-MS analysis of the full CaBioWF reaction upon incubation with pimelic acid and CoASH leading to first the production of the pimeloyl-CoA intermediate and then the C–C bond forming reaction either A) Gly, B) L-Ala leading to the production of AON or C) L-Ser, each leading to the formation of the corresponding aminoketone product. Product formation was confirmed by LC ESI-MS analysis.

The full CaBioWF substrate scope

Since BioW is the key enzyme that controls the acyl-chain length of biotin, we expected the equivalent domain of the CaBioWF fusion would only accept pimelic acid as a substrate. This was previously observed for both BsBioW and AaBioW.^[15,16,24] Having found that the enzyme is active with the native substrate, the CaBioWF was incubated with various diacids (DC₆–DC₉) and mono acids (C₆–C₁₀), using the same reaction conditions and analysis as before. A series of small peaks between 15 and 25 minutes was observed by HPLC, and we expected them to correspond to the acyl-CoA products - DC₆ (16.5 min), DC₈ (18.3 min), DC₉ (19.1 min), C₆ (20.1 min), C₇ (21.2 min), C₈ (22.1 min), C₉ (23.4 min) and C₁₀ (24.0 min) (Figures S14 and S16). MS analysis of each reaction confirmed the formation of the expected product, with the correct [M + H]⁺ ion observed for each reaction (Figures S15 and S17, except for C₆-CoA for which no product ion was observed despite the presence of a visible peak in the HPLC).

The broad substrate scope of CaBioWF was unexpected. Estrada *et al.* noted acyl-adenylate formation was observed when AaBioW was incubated with DC₆ and DC₈ acids, however no acyl-CoA was formed upon incubation with CoASH, but rather hydrolysis of the acyl-adenylate was observed leading to acid reformation and AMP release.^[24] Therefore, a multiple sequence alignment (CaBioWF, BsBioW and AaBioW) enabled the identification of the corresponding active site residues in CaBioWF (Figure S3).

Three active site residues (Tyr199, Tyr211 and Arg213 in BsBioW) were shown to form key H-bond and electrostatic interactions with the terminal carboxyl group of the bound pimeloyl-CoA. By aligning CaBioWF with BsBioW and AaBioW, one of the key Tyr residues is a Phe (Phe192), which corresponds to the Tyr199 in BsBioW and Tyr188 in AaBioW (Figure 5). Rational engineering generated a BsBioW Y199F mutant that was able to accept a range of mono-acids and diacids.^[15] Therefore we hypothesised that the equivalent Phe192 residue in the CaBioWF fusion is the reason for the observed broad substrate range.

To test this hypothesis, a CaBioWF F192Y fusion mutant was prepared. The purified fusion mutant biocatalyst was incubated

under the same conditions as the wild type CaBioWF version. LC ESI-MS analysis of the enzyme gave a mass of 68567 ± 3.19 Da, which corresponds to the predicted mass of 68682.23 Da, equivalent to a mass difference of +16 Da (F - Y mutation) and -131 Da (loss of N-terminal Met) compared to the wild type CaBioWF (Figure S18). The CaBioWF F192Y mutant was then incubated with the series of acid substrates (DC₆–DC₉ and C₆–C₉). In contrast to the wild type enzyme, acyl-CoA product formation was only observed for the natural DC₇ pimelic acid substrate, with no other products detected via HPLC analysis (Figure S19). This result provides evidence to support the key role that the Phe192 residue plays in allowing the CaBioWF fusion to accept a range of acid substrates.

The full CaBioWF coupled biocatalytic reaction

Having established the broad substrate scope of the CaBioWF fusion, we examined the synthetic potential of this biocatalyst by screening combinations of various fatty acid and amino acid substrates. The CaBioWF was incubated with PLP, L-Ala and acids (DC₆–DC₉ and C₆–C₉). We also included one example with L-Ser and Gly. Gratifyingly, all of the predicted AON products were observed by LC ESI-MS (Figure S20, Table 1).

Table 1. Tabulation of the predicted and observed *m/z* values for the AON analogues, formed upon incubation of CaBioWF with various combinations of fatty acids and amino acids.

Compound ^[a]	<i>m/z</i> Predicted	<i>m/z</i> Observed ^[b]
DC ₆ +L-Ala	174.1125	174.0664
DC ₇ +L-Ala	188.1281	188.1282
DC ₈ +L-Ala	202.1438	202.1433
DC ₉ +L-Ala	216.1594	216.1598
C ₆ +L-Ala	144.1387	144.1383
C ₇ +L-Ala	158.1539	158.1527
C ₈ +L-Ala	172.1696	172.1693
C ₉ +L-Ala	186.1852	186.1856
DC ₇ +L-Ser	204.1230	204.1267
DC ₇ +Gly	174.1125	174.0664
C ₇ -Br+L-Ala	236.0686	236.0664
	238.0664	238.06625
C ₆ -Me+L-Ala	172.1696	172.1675

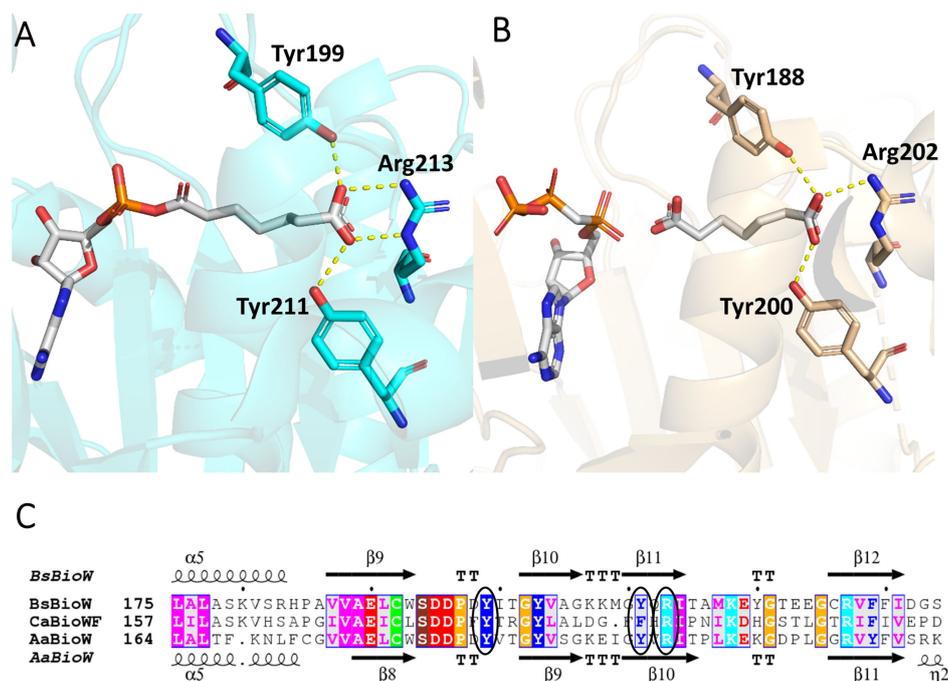


Figure 5. Sequence and structural analysis of the BioW domains A) *BsBioW* structure (PDB:5FLL) with pimeloyl-CoA bound, showing interactions with key active site residues Tyr199, Tyr211 and Arg213 and B) *AaBioW* structure (PDB: 5TV8) with CoASH and pimelate bound showing interactions with active site residues Tyr188, Tyr200 and Arg202. C) Sequence alignment of *BsBioW*, *CaBioWF* and *AaBioW*, with the corresponding predicted active site residues circled (Phe192 corresponds to *BsBioW* Tyr211 and *AaBioW* Tyr200).

Encouraged by this result, a pair of unusual carboxylic acids (7-bromoheptanoic acid and 6-methylhexanoic acid) were incubated with the enzyme. In both cases the corresponding aminoketone product was formed with L-Ala and was observed in the mass spectrum. In particular the formation of the AON derived from 7-bromoheptanoic highlights the synthetic utility of this novel enzyme, as the C–Br can be used as a handle to introduce further diversity (Figure S21, Table 1).

We conclude that the unexpected broad substrate scope extends between both domains of the natural *CaBioWF* fusion and thus this biocatalyst can be used to prepare a number of AON analogues.

Structural analysis of the *CaBioWF* fusion

CaBioWF was accurately modelled as a homomeric complex using the deep learning architecture ColabFold,^[25,26] this complex was subsequently relaxed and studied in a 10 ns molecular dynamics simulation (MDS, see Supporting Information for detailed analysis (Figures S22–29)). In summary, both *CaBioW* and *CaBioF* domains are expected to contribute towards homodimerisation; the predicted complex could maintain its structural integrity when scrutinised by MDS. *CaBioWF* also demonstrates moderate conformational flexibility, owing partly to the disordered linker(s) connecting the *CaBioW* and *CaBioF* domains. Intriguingly, the relative orientation and proximity of the *CaBioW* and *CaBioF* active sites implies the existence of a ~5 nm molecular “tunnel”, permitting the facile diffusion of

pimeloyl-CoA from *CaBioW* to *CaBioF*. Taken together, this *in silico* study afforded valuable insight into the didomain, homomeric architecture of *CaBioWF*, making it an attractive target for further structural studies.

Conclusion

Here we demonstrate that fusion enzymes can be useful biocatalytic tools with the ability to perform multiple, consecutive transformations in one-pot. This work has explored a novel bifunctional fusion enzyme (*CaBioWF*) that displays an inherently broad substrate scope for the formation of a range of synthetically useful aminoketone building blocks.

The recombinant *CaBioWF* has been shown to catalyse the consecutive steps of pimeloyl-CoA formation, followed by condensation with L-Ala to form AON. The individual *CaBioW* and *CaBioF* domains within the fusion biocatalyst were assayed. Analysis of the *CaBioW* domain revealed a natural substrate promiscuity that is unprecedented when compared with other BioW homologues. The domain was able to convert a series of mono- and di-acid (C₆–C₁₀ and DC₆–DC₉) substrates, forming the corresponding acyl-CoA products. A divergent Phe residue in the predicted *CaBioW* active site was proposed to be responsible for this broad substrate scope. Conversion of this Phe to Tyr restored the narrow substrate specificity exhibited by single domain BioW homologues.

In contrast to the single domain BioF homologues, the PLP-dependent *CaBioF* domain of the fusion also displayed an

expanded substrate scope. It was able to bind L-Ala, L-Ser and Gly and catalyse formation of the corresponding AON derivatives. By combining both activities, the CaBioWF fusion biocatalyst was able to condense a range of fatty acid substrates (C₆-C₉ and DC₆-DC₉) with L-Ala to generate eight aminoketone derivatives. Furthermore, the head-group could be diversified by condensation with L-Ser and Gly. A further two aminoketones were generated from unusual fatty acid substrates. In total, twelve aminoketone products were produced, but the inherent properties of this fusion suggest that it has even greater synthetic potential. Structural insights into this unusual CaBioWF fusion biocatalyst were obtained by predictive modelling and simulation. This revealed a putative tunnel that may enable facile diffusion of the pimeloyl-CoA product from the BioW to the BioF domain. The characterisation of this unique di-domain biocatalyst should encourage the discovery of other natural fusions embedded in the BGCs of other genomes. This will be facilitated by the continued advances in sequencing and metagenomics.

Experimental Section

Cloning and expression of CaBioWF in *E. coli*: The originally purchased synthetic CaBioWF gene was re-cloned into pET28 α , adding 7 amino acids (MSTYSIR, Figure S4) to the N-terminus using the following primers. BioWF For (NcoI) – GAGATATACC ATGGG-CAGTA CCTACAGCA, BioWF Rev (BamHI) – TTTCGGATCC CAG-CAGGCC pET28 α For (BamHI) – CTGCTGGGAT CCGAAAACCT GTATTTTC, pET28 α Rev (NcoI) – GTAGGTACTG CCCATGGTAT ATCTCCTTCT. After PCR, both PCR products were incubated together with the Gibson master mix at 50 °C before being transformed into C2987 cells. This generated the plasmid pET-28a/CaBioWF which was used to transform BL21 (DE3) cells. Expression from this construct was optimised (Figure S6) and suitable conditions were chosen for a larger scale (typically 2 litres of LB broth supplemented with kanamycin). The recombinant CaBioW was expressed using 0.5 mM isopropyl β -D-thiogalactoside (IPTG) at 20 °C for 18 h. Cells were harvested by centrifugation and stored at –20 °C.

Purification of CaBioWF variants from *E. coli*: Cell pellets were defrosted on ice and then resuspended in the corresponding lysis buffer (sodium phosphate pH 7.5 (20 mM), NaCl (500 mM), imidazole (10 mM), PLP (25 μ M)). Around 10 mL of resuspended pellets was then sonicated for 10 min in 30 s intervals to lyse the cells. Cell free extract was obtained by centrifugation at 10,000 rpm for 45 min and the supernatant was filtered (0.45 μ M filter). The protein was purified from this cell lysate. Each purification step was completed at 4 °C or on ice. The first purification step consisted of nickel IMAC. The protein was loaded onto a 5 mL HisTrap HP column (GE Healthcare) attached to an ÄKTA Purifier system and protein elution monitored at 280 nm. The filtered cell lysate was loaded onto the column at 3 mL/min and washed with the corresponding lysis buffer for 3 column volumes (CVs) to remove any unbound protein. The protein eluted with an increasing imidazole gradient (10 mM–300 mM) over 20 CVs using an increasing gradient of the elution buffer (Sodium phosphate pH 7.5 (20 mM), NaCl (100 mM), imidazole (300 mM), PLP (25 μ M)). The eluted protein was concentrated down to ~1–5 mL using a molecular weight cut-off concentrator (MWCO, satorius). The HisTagged protein was incubated with pure TEV protease (1:10 ratio based off protein concentration (mg/mL)) and dialysed against

the size exclusion chromatography (SEC) buffer 2 L (HEPES pH 8 (20 mM), NaCl (100 mM), PLP (25 μ M), 10% glycerol) for 2 h at 4 °C. Following dialysis, the protein was injected onto a 5 mL HisTrap column (Nickel IMAC) and the untagged enzyme was collected with the flow through over 5 CVs. This removed the TEV and any remaining tagged protein which remained bound to the columns and then eluted off the column using the elution buffer (300 mM imidazole). The collected untagged protein was concentrated down using a MWCO concentrator to a final volume of 1–5 mL. This was then loaded onto a pre-equilibrated (SEC buffer, 1 CV) HiLoad 16/60 Superdex S200 column (120 mL) from GE healthcare. The protein was eluted over 1 CV at a flow rate of 0.5–1 mL/min with elution monitored at 280 nm. The fractions containing CaBioWF (Figure S7) were combined, concentrated and flash frozen before being stored at –80 °C.

MesG Assay of BioW domain: The activity of BioWF was determined using the 7-methyl-6-thioguanosine (MesG) assay. Each well contained a final concentration of Tris HCl pH 8 (50 μ M), NaCl (100 mM), MgCl₂ (10 mM), triscarboxyethylphosphine (TCEP, 1 mM), CoASH (1 mM), ATP (1 mM), 7-methyl-6-thioguanosine (MesG, Berry & Associates, 0.5 mM), inorganic pyrophosphatase (PPase from baker's yeast, Sigma-Aldrich, 0.03 U), bacterial purine nucleoside phosphorylase (PNP, Sigma-Aldrich, 1 U) and CaBioWF (0.1 μ M) enzyme. After the addition of the components the plate was pre-incubated at 30 °C for 15 min. The reaction was then initiated by the addition of pimelic acid (0 to 1000 μ M). The increase in absorbance at 360 nm resulting from the enzymatic conversion of MesG to 7-methyl-6-thioguanine was monitored over 1 hr on a BioTek Synergy HT plate reader with Costar 96-well UV-transparent plates. The data from the first 10 min was analysed using the Michaelis-Menten model and a nonlinear regression fit on origin gave values of K_M and k_{cat}.

BioWF acyl-CoA formation reaction: The CaBioW domain was analysed by acyl-CoA formation which can be monitored by HPLC. Reactions contained CaBioWF (5 μ M), TCEP (0.2 mM), ATP (1 mM), CoASH (1 mM) and mono- or di-acids (DC₆-DC₉ and C₆-C₁₀, 1.5 mM) in buffer (Tris·HCl (25 mM, pH 8), NaCl (50 mM), MgCl₂ (5 mM)) in a final volume of 1 mL and heated at 30 °C for 5 h.

Acyl-CoA HPLC assay: Reactions were quenched with a 1:1 (v/v) ratio of MeCN. 10 μ L of the sample was then injected onto Luna 5 μ m C18 RP-HPLC column (100 Å, 250 \times 4.60 mm, Phenomenex), and eluted with 95% water (0.1% TFA (trifluoroacetic acid), v/v) for 5 min followed by a 25 min gradient from 5% to 55% acetonitrile (0.1% TFA, v/v)/water (0.1% TFA, v/v) which was maintained for 2 min at 260 nm.

Acyl-CoA MS analysis: Reactions were quenched using a 1:1 (v/v) of MeCN with 0.01% formic acid and centrifuged at 17000 \times g. 5 μ L of the supernatant was injected into a ProSwift C4-RP-5H (Thermo) column coupled to a FT-ICR MS (Daltonics 12T SolariX). The LC gradient (acetonitrile/water) ran from 0% to 60% acetonitrile with 0.1% formic acid over 25 min before being increased to 90% acetonitrile for 3 min. EICs and masses were determined on DataAnalysis V4.3 software.

CaBioWF full reaction: Reactions contained CaBioWF (5 μ M), TCEP (0.2 mM), ATP (1 mM), CoASH (1 mM) and mono- or di-acids (DC₆-DC₉ and C₆-C₁₀, 1.5 mM) and amino acid (L-Ala, L-Ser or Gly) in buffer (Tris·HCl (25 mM, pH 8), NaCl (50 mM), MgCl₂ (5 mM)) in a final volume of 1 mL and heated at 30 °C for 5 h.

CaBioWF full reaction MS analysis: Reactions were quenched using a 1:1 (v/v) of MeCN with 0.01% TFA and centrifuged at 17000 \times g. 5 μ L of the supernatant was injected into a Phenomenex Kinetex 5 μ m C18 100 Å column coupled to a micrOTOF II (Bruker). The LC gradient ran from 5% acetonitrile and 95% water with 0.1% formic

acid to 95% acetonitrile over 10 min. EICs and masses were determined on DataAnalysis V4.3 software.

PLP binding UV-Vis spectroscopy: The UV-Vis analysis was carried out on a Varian Cary® UV-Vis spectrophotometer. Excess PLP was removed through desalting using a PD10 column (GE healthcare) and the enzyme exchanged into the storage buffer minus PLP. Assays were carried out in 1 cm pathlength quartz cuvettes and baseline correction was carried out before acquiring spectra.

CaBioWF changes in PLP binding: The protein was diluted to ~20 μM and an initial PLP bound enzyme curve was obtained by scanning from 250 nm to 500 nm. L-Ala was titrated from a 1 M stock to a final concentration of 10 mM. The mixture was left to allow time for binding and after 30 min a second scan was taken. Next pimelic acid and CoASH (0.5 mM) were added, alongside MgCl_2 (10 mM) and ATP (1 mM). The changes in absorbance were analysed over 5 h, showing the changes in absorbance as the CaBioW domain formed pimeloyl-CoA which was then utilized by the CaBioF domain.

PLP enzyme-amino acid dissociation constants (K_d): The protein was diluted to ~20 μM and the amino acid substrates were titrated from a 1 M stock solution to varying final concentrations (0–100 mM). The spectra were normalized against the 280 nm peak to account for dilution of the sample with addition of substrate solution. Changes in the absorbance maximum of the ketoenamine peak (~425 nm) were plotted and fitted with a hyperbolic saturation curve using Origin software.

AON calibration curve: To estimate the relative abundance of AON produced by CaBioWF, a calibration curve was created using AON hydrochloride (Cayman Chemical) and analysed by LCMS, monitoring the mass of 188.1280 ($[\text{M} + \text{H}]^+$, $\text{C}_9\text{H}_{18}\text{NO}_3$). A 10 mg/mL stock solution of AON (44.7 mM) was made in DMSO. This was diluted to 1 mM in buffer (Tris-HCl (25 mM, pH 8), NaCl (50 mM), MgCl_2 (5 mM)). Further dilutions made calibration solutions of 0, 2, 5, 10, 20, and 50 μM in triplicate. The LC-MS used has an Agilent 1200 Series HPLC and a Bruker Daltonics micrOTOF-II High performance Time-of-flight mass spectrometer (TOF-MS). 5 μL of each calibration solution was injected onto a Phenomenex Kinetex C18 reverse phase column (50 \times 2.1 mm, 2.6 μm particle size) at a flow rate of 0.2 mL/min. The elution was conducted as follows (5% MeCN for 0.5 mins, 100% MeCN for 5 mins, then maintained at 100% for 2 mins, then at 7 mins return to 5% MeCN, followed by re-equilibration for 2.5 mins). The retention time for AON using this method is 1.67 mins. EICs and masses were determined on DataAnalysis V4.3 software. The area under the curve (AUC) was used to construct a calibration curve of concentration of AON (Figure S13).

CaBioWF full reaction for quantitative analysis: Reactions contained CaBioWF (5 μM), TCEP (0.2 mM), ATP (1 mM), CoASH (1 mM or 0.5 mM), pimelic acid (1.5 mM) and L-alanine (1.5 mM) in buffer (Tris-HCl (25 mM, pH 8), NaCl (50 mM), MgCl_2 (5 mM)) in final volumes of 1000 and 500 μL respectively and heated at 30 °C for 5 h. Reactions were quenched using 1.7% TFA (60 μL per 1 mL of reaction) and centrifuged at 17000 \times g for 10 minutes. The supernatant was diluted (1:10, 1:20, 1:50) for LCMS analysis using the same buffer and loaded onto the LCMS, treating the samples in the same way as the calibration standards.

CaBioWF preparative scale reaction and quantitative analysis: Reactions contained CaBioWF (5 μM), TCEP (0.2 mM), ATP (1 mM), CoASH (0.5 mM), Pimelic Acid (1.5 mM) and L-alanine (1.5 mM) in buffer (Tris-HCl (25 mM, pH 8), NaCl (50 mM), MgCl_2 (5 mM)) in 10 mL final volume, were heated at 30 °C for 5 h with 180 rpm agitation. Reactions were quenched using 1.7% TFA (60 μL per 1 mL of reaction) and centrifuged at 17000 \times g for 10 minutes. The

supernatant was diluted (1:10, 1:20, 1:50) for LCMS analysis using the same buffer and loaded onto the LCMS, treating the samples in the same way as the calibration standards. The rest of the reaction was stored at –80 °C.

Structure prediction: All structural predictions were performed using ColabFold via AlphaFold2_advanced.ipynb. In brief, a deep multiple sequence alignment (MSA) was generated using MMSeqs2 prior to structure prediction using AlphaFold 2 (structural templates were not utilised for prediction). ColabFold was configured to perform homodimeric prediction, and the output of the AlphaFold 2 structure module was recycled up to 3 times for refinement. For each sequence, a total of 5 models were generated and ranked by Predicted Template Model score (pTM); Predicted Local Distance Difference Test (pLDDT) scores were also computed for each model to evaluate fold-level confidence. The best ranked model was subsequently relaxed to eliminate steric clashes. Visual inspection was performed in UCSF ChimeraX 1.3^[27] and PyMOL 2.4.

Evolutionary conservation analysis: Evolutionary conservation analysis was performed using the ConSurf^[28,29] server configured to build MSAs using MAFFT. 150 homologous sequences with identities ranging from 30–95% were compiled from UNIREF90 using the HMMER search algorithm. Conservation scores were calculated via the Bayesian method and visualised using UCSF ChimeraX 1.3.

Molecular dynamics simulation: Simulations were performed using GROMACS 2021.4.^[30] Protein charges were computed using CHARMM36 all-atom forcefield.^[31] The model was solvated in TIP3P water in a cubic box, and the net protein charge was counter-balanced using simulated sodium ions. The system was energy-minimised by sequential steepest descent/conjugate gradient descent and equilibrated to 300 K and 1 bar using V-Rescale thermostat/Berendsen barostat. Following a 10 ns (5×10^6 time steps) production MD, the trajectory was recentred with additional rotational and translational fitting. Further analysis was performed in GROMACS using *gmx gyrate*, *gmx hbond*, *gmx rms* and *gmx rmsf*. UCSF Chimera 1.16^[32] was used for trajectory visualisation and for computing pairwise RMSDs.

Acknowledgements

The authors thank the Biotechnology and Biological Sciences Research Council (BBSRC) for funding (BB/M003493/1, BB/S506953/1 and BB/R013993/1 to D.J.C. We also thank the Derek Stewart Charitable Trust (DSCT) for PhD funding (to S.M.H.) and the BBSRC for PhD funding to M.A.H. We thank Evgenii Lobzaev with access to super computing resources and running our MDS script.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: aminoketones · adenylate · biocatalysis · fusion · pyridoxal 5'-phosphate

- [1] M. B. Quin, K. K. Wallin, G. Zhang, C. Schmidt-Dannert, *Org. Biomol. Chem.* **2017**, *15*, 4260–4271.
- [2] G. J. Williams, *Curr. Opin. Struct. Biol.* **2013**, *23*, 603–612.
- [3] J. Beld, D. J. Lee, M. D. Burkart, *Mol. BioSyst.* **2014**, *11*, 38–59.
- [4] E. Rubinstein, Y. Keynan, *Front. Public Heal.* **2014**, *2*, 217.
- [5] J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature* **1990**, *348*, 176–178.
- [6] F. S. Aalbers, M. W. Fraaije, *ChemBioChem* **2019**, *20*, 20–28.
- [7] K. Hölsch, D. Weuster-Botz, *Biotechnol. Appl. Biochem.* **2010**, *56*, 131–140.
- [8] D. E. Torres Pazmiño, R. Snajdrova, B. J. Baas, M. Ghobrial, M. D. Mihovilovic, M. W. Fraaije, *Angew. Chem. Int. Ed.* **2008**, *47*, 2275–2278; *Angew. Chem.* **2008**, *120*, 2307–2310.
- [9] X. Chen, J. L. Zaro, W. C. Shen, *Adv. Drug Delivery Rev.* **2013**, *65*, 1357–1369.
- [10] S. Lin, J. E. Cronan, *J. Biol. Chem.* **2012**, *287*, 37010–37020.
- [11] S. Lin, J. E. Cronan, *Mol. BioSyst.* **2011**, *7*, 1811–1821.
- [12] S. Lin, R. E. Hanson, J. E. Cronan, *Nat. Chem. Biol.* **2010**, *6*, 682–688.
- [13] J. E. Stok, J. J. De Voss, *Arch. Biochem. Biophys.* **2000**, *384*, 351–360.
- [14] M. J. Crayle, I. Schlichting, *Proc. Natl. Acad. Sci. USA.* **2008**, *105*, 15696–15701.
- [15] M. Wang, L. Moynié, P. J. Harrison, V. Kelly, A. Piper, J. H. Naismith, D. J. Campopiano, *Nat. Chem. Biol.* **2017**, *136*, 660–667.
- [16] P. Estrada, M. Manandhar, S. H. Dong, J. Deveryshetty, V. Agarwal, J. E. Cronan, S. K. Nair, *Nat. Chem. Biol.* **2017**, *136*, 668–674.
- [17] G. Schneider, Y. Lindqvist, *FEBS Lett.* **2001**, *495*, 7–11.
- [18] D. Cobessi, R. Dumas, V. Pautre, C. Meinguet, J. L. Ferrer, C. Alban, *Plant Cell* **2012**, *24*, 1608–1625.
- [19] M. Manandhar, J. E. Cronan, *Appl. Environ. Microbiol.* **2018**, *84*, e02084–17.
- [20] N. Soltan Mohammadi, S. Mafakheri, N. Abdali, I. Bárcena-Uribarri, A. Tauch, R. Benz, *Biochim. Biophys. Acta Biomembr.* **2013**, *1828*, 2574–2582.
- [21] D. J. Wilson, C. C. Aldrich, *Anal. Biochem.* **2010**, *404*, 56.
- [22] M. R. Webb, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4884–4887.
- [23] S. P. Webster, D. Alexeev, D. J. Campopiano, R. M. Watt, M. Alexeeva, L. Sawyer, R. L. Baxter, *Biochemistry* **1999**, *39*, 516–528.
- [24] M. Manandhar, J. E. Cronan, *Chem. Biol.* **2013**, *20*, 1441–1446.
- [25] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohli, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
- [26] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, *bioRxiv* **2022**, <https://doi.org/10.1101/2021.08.15.456425>.
- [27] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, *Protein Sci.* **2021**, *30*, 70–82.
- [28] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, *Nucleic Acids Res.* **2016**, *44*, W344–W350.
- [29] M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, N. Ben-Tal, *Nucleic Acids Res.* **2005**, *33*, W299–W302.
- [30] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1–2*, 19–25.
- [31] J. Huang, A. D. Mackerell, *J. Comput. Chem.* **2013**, *34*, 2135–2145.
- [32] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.

Manuscript received: March 28, 2022

Revised manuscript received: June 10, 2022

Accepted manuscript online: June 13, 2022

Version of record online: July 13, 2022