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

Shona M. Richardson, Peter J. Harrison, Michael A. Herrera, Menglu Wang, Rebecca Verez, Gustavo Perez Ortiz, and Dominic J. Campopiano*
Supplementary Information

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] *

[a] School of Chemistry University of Edinburgh, David Brewster Road, Edinburgh, EH9 3FJ.
[b] Diamond Light Source Ltd, Harwell Science & innovation Campus, Didcot, OX11 0DE

Correspondence to: Dominic.Campopiano@ed.ac.uk
Table of Contents

Figure. S1: Corneybacterium amycolatum SK46 BioWF sequence .................................................. 3
Figure. S2: CaBioWF recombinant protein sequence ................................................................. 4
Figure. S3: Sequence alignment of CaBioWF against BsBioW and AaBioW ............................ 5
Figure. S4: N-terminal sequence alignment of the new CaBioWF construct against BsBioW and AaBioW shown alongside new CaBioWF sequence ......................................................... 6
Figure. S5: New CaBioWF pET-28a construct with N-terminal extension .................................... 7
Figure. S6: New CaBioWF expression tests .................................................................................... 8
Figure. S7: Purification of recombinant CaBioWF from E. coli. ...................................................... 9
Figure. S8: CaBioW pimeloyl-CoA reaction .................................................................................. 10
Figure. S9: CaBioWF kinetic MesG assay ..................................................................................... 11
Figure. S10: Sequence alignment of CaBioF domain alongside EcBioF ........................................ 12
Figure. S11: UV-Vis analysis of full CaBioWF reaction .................................................................. 13
Figure. S12: Analysis of the amino acid substrate scope of the CaBioWF fusion. ......................... 14
Figure. S13: Calibration curve of AON formation catalysed by CaBioWF ...................................... 16
Figure. S14: CaBioWF acyl-CoA formation with a variety of di-carboxylic acid lengths ............... 17
Figure. S15: MS analysis of CaBioWF acyl-CoA di-acid reactions ................................................ 18
Figure. S16: CaBioWF acyl-CoA formation with a variety of mono-carboxylic acid lengths ........ 19
Figure. S17: MS analysis of CaBioWF acyl-CoA mono-acid reactions .......................................... 20
Figure. S18: CaBioWF F192Y purification .................................................................................... 21
Figure. S19: CaBioWF F192Y acyl-CoA reactions ......................................................................... 22
Figure. S20: Full CaBioWF reactions with various di and mono-carboxylic acid chain lengths ...... 23
Figure. S21: CaBioWF unusual carboxylic acid reactions ............................................................. 24
Figure. S22: pLDDT and pTM scores of the predicted CaBioWF domains ........................................ 26
Figure. S23: A closer inspection of the predicted CaBioW domain ................................................ 27
Figure. S24: A closer inspection of the predicted CaBioF domain .................................................. 28
Figure. S25: Visualisation of the CaBioWF simulation .................................................................... 29
Figure. S26: A summary of the CaBioWF MDS ......................................................................... 30
Figure. S27: Fluctuation analysis of the simulated CaBioWF complex .......................................... 31
Figure. S28: The CaBioWF complex sampled midway (t = 5 ns) through the MDS ....................... 32
Figure. S29: A predicted molecular “tunnel” (dashed line) between the CaBioW and CaBioF domains .............................................................................................................................. 33
Supplementary Figures:

ATCGCCTCCTCGCTGATTACGTACATCCTCGTGCAGTTCAAGCGTGCCACCTGCCGAACCTGCGCAACGCATCGCCGGCCCTGACGTTACGCATCTCCGCTGACCTGCTGATAGCTGCTGAGATTCTCTATCTGCTGCTGACGTTACGCATCTCCGCTGACCTGCTGATAGCTGCTGAGATTCTCTATCTGCTGCTGACGTTACGCATCT

Figure S1: Corneybacterium amycolatum SK46 BioWF sequence
Original sequence of the Corneybacterium amycolatum SK46 BioWF (Uniprot: E2MUP3).
MGSSHHHHHHSSGLVPRGSHMRSSADYCHISGAERLAPTELPLASAMTSLRALHHDKGRPDTIHITVDKIEESTIS TVPALTPFLESNSSPGDARKLIAQRHLAAGIQADIAAEAMAYSALTGLRGAALIDSSSGERLDNPAPARGVRVSTFD AIS HPSKDCAKDHFHEALILASKVHSAPGIVAEICSLDDPFYTRGYLALDFGFFHRPNKIDHGSTLGTRIFIVEPDTIPEDLID YLNTMPVYIELPPDASSSTTTGSLSSDLAIAAAQRNTAWAGAGLTRLRTFETAGLPHSRLGADYLLFSSSYYGLSLST HPELVSAAATAAIGHFTGSGGSLTHTTGTSHSALEELAQQFFGDADVLFATGQANHSTIAAIATADVEIFSDAAANH ASIIDGCRNARAKVTVPFPHADYQTLDRLLATSSARHKLVISDSVFSMSGEVIDGPALETCRRRNAWLMLDAHG VGVEQGRGTAHLDIRPDPVGVTASKALGVGGVGCLSPVGELELRNQARSFVYSTSMNPGSVAAIRAALKQLEV GDVVKRLQRRNIAVLSLVGAQSDPASAIPIPLVGDTEAMEISQAQLAELGVFIPAIRYPTVPREPGEAMLRTLHTD ADIDQLELARNTGGL

**Figure. S2: CoBioWF recombinant protein sequence**

The recombinant protein sequence of the original pET28a construct of original CoBioWF sequence of *Corynebacterium amycolatum* SK46 BioWF (starting Met shown in bold) with an N-terminal HisTag.
Figure. S3: Sequence alignment of CaBioWF against BsBioW and AoBioW

Sequence alignment of the Genbank predicted BioW domain of CaBioWF against its closest homologues BsBioW (PDB: 5FLL, Uniprot: P53559) and AoBioW (PDB: 5TV6, Uniprot: O67575). The predicted active site residues are highlighted.
Figure S4: N-terminal sequence alignment of the new CaBioWF construct against BsBioW and AaBioW shown alongside new CaBioWF sequence

A) New N-terminal sequence after addition of 7 amino acids (MSTYSIR) to the N-terminus of CaBioWF (Uniprot: E2MUP3), aligned with BsBioW (Bacillus subtilis, Uniprot: P53559) and AaBioW (Aquifex aeolicus, Uniprot: O67575). B) CaBioWF schematic of the didomain protein showing the BioW domain (pink, residues 1-227) and BioF domain (orange, residues 239-619) along with the C) new CaBioWF recombinant protein sequence with the BioW ANL domain (pink), the BioF AON domain (orange) and the linker region (black). The predicted ANL active site residues (Tyr181, Phe192 and Arg194) and key BioF residues including the key PLP binding lysine (Arg271, Ser295, His379, Lys483) are shown in a darker shade and underlined.
**MGSTYSIRMSSADYCHISGARLAPATLPQLASAMTSLHDDKGRPDITIHTVDKIEESTISTVPAALTPFLESNSS PGDARKLIAQLHAAAGIQAADIAAEEMAYSSTGLRGAALIDSSSGERLDPPPARGVRFSTFDAISHPSKDKDSDKHHFHE ALILASKVHSAPGIVAEICLSDDDFFYTRGGLALDGFFFHRIPNIDHGSTLGRIFIVEPDTIDPELIDYENTPVYIELEEP ASSTTTTGLSSDLSAIAAQRNATAWAGLTRLTRTLRTEALQPLPHSRIDGADYLLFFSSSDYLGSLTHPEIVSAATAAIG HFGTSGSGGLTTTGTIIHASELEAQFATGGQANHSTIAAIATADVEIFSDAANHASIIIDGCRNARAK VTVFPHADYQTLDDLATSSARHKLVISDSVFSMSGEVIDGPAERTCRRNWAMLDDAHGVYGIGEQGRGTAA HLDIRPDIVVTASKALGVGENYVLCSPVGELLRNQARSFYVSTSMNPGSVAIRAAALKQLEVGDVVKRLQRNIAR VLSLVGAQSDPASAIPLPVGDETEAMDISQLAELGVFIAPAIYPTVPREGELRLTITALHTDADIDQLELALRNTG LLGENLYQFQGLEHHHHHH**

**Figure. S5: New CaBioWF pET-28a construct with N-terminal extension**

Final *Corynebacterium amycolatum* SK46 (Uniprot: E2MUP3) CaBioWF recombinant protein sequence in a pET-28a plasmid with a C-terminal TEV cleavable HisTag shown in bold. New and old starting Met and also shown in bold, extension added in yellow.
Figure. S6: New CaBioWF expression tests
Expression tests of new pET-28a/CaBioWF construct in BL21 (DE3) cells, with expression tested under the two temperatures of 20 °C o/n and 30 °C for 5 hours at IPTG concentrations of 0.1 mM, 0.5 mM and 1 mM.
Figure S7: Purification of recombinant CaBioWF from E. coli.

Purification of the new CaBioWF construct using A) SDS-PAGE analysis with LMW marker shown alongside eluted HT fractions and Superdex S200 fractions corresponding to the eluted S200 peak at 60.7 mL, B) S200 SEC chromatogram monitored at 280 nm with an elution volume of 60.7 mL and C) LC ESI-MS analysis showing the charge states and D) deconvoluted mass of the obtained protein at 67432.45 ± 0.31 Da.
Figure. S8: CaBioW pimeloyl-CoA reaction
FT-ICR MS analysis of the pimeloyl-CoA intermediate formed by the CaBioW domain shown by the presence of A) the ion m/z = 932.1674 aligned with B) the predicted mass of 932.1674 ([M + Na]^+, C_{28}H_{46}N_{7}O_{19}P_{3}S).
Figure. S9: CaBioWF kinetic MesG assay
Schematic of the coupled pyrophosphate production MesG assay. The pimelic acid is activated by the ANL enzyme BsBioW using ATP, releasing PPI. The PPI is broken into two molecules of Pi by PPase. The Pi is then utilised with MesG by PNP to form 7-methylguanine, which absorbs at 360 nm. Michaelis-Menten analysis of the purified CaBioWF using the coupled MesG assay with a calculated $K_M$ value of 59 ± 3 µM.
Figure. S10: Sequence alignment of CaBioF domain alongside EcBioF

Sequence alignment of the CaBioF domain of CaBioWF against its closest homologues EcBioF (PDB: 1DJ9, Uniprot: P12998) with key active site residues Arg21, Asn47, His133 and Lys236 (EcBioF numbering).
Figure. S11: UV-Vis analysis of full CoBioWF reaction

**A)** UV-Vis spectroscopy scan of CoBioWF PLP-binding, monitoring the changes after the addition of l-Ala (10 mM) followed by pimelic acid and CoASH in the presence of MgCl₂ and ATP, forming pimeloyl-CoA by the CoBioW domain. This intermediate is used by the CoBioF domain, monitoring changes in the absorbance for AON formation. Schematic of **B)** conversion from internal to external aldimine and **C)** condensation of pimeloyl-CoA and l-Ala by CoBioF domain leading to absorbance changes also shown.
Figure. S12: Analysis of the amino acid substrate scope of the CaBioWF fusion.

LC ESI-MS analysis of the full CaBioWF reaction upon incubation with pimelic acid, MgATP and CoASH leading to first the production of the BioW-catalysed pimeloyl-CoA intermediate and then the BioF-catalysed C-C bond forming reaction with either A) Gly, B) L-Ala or C) L-Ser, each leading to the formation of the corresponding aminoketone AON product.
Figure. S13: Calibration curve of AON formation catalysed by CaBioWF

Calibration curve of AON. **A)** Extracted Ion Chromatograms (EICs) of calibration solutions (0-50 μM) of commercial AON [M+H]^+ m/z = 188.1282 ± 0.01 Da. **B)** Calibration curve of the area under the curve (AUC) of the repeats of each calibration solution. **C)** The EICs of a standard solution of AON, the CaBioWF reaction and a mixture of both the CaBioWF reaction and the AON standard, [M+H]^+ m/z = 188.1282, showing only one peak with the expected mass with the same RT. The reaction was performed using CaBioWF (5 μM), TCEP (0.2 mM), ATP (1 mM), CoASH (0.5 mM), pimelic Acid (1.5 mM) and l-serine (1.5 mM) in buffer (Tris.HCl (25 mM, pH 8), NaCl (50 mM), MgCl₂ (5 mM)) in 10 mL final volume, heated at 30 °C for 5 hrs with 180 rpm agitation. The reaction was quenched using 1.7 % TFA (60 μL per 1 mL of reaction) and centrifuged at 17000 xg for 10 minutes. The supernatant was diluted (1:20, 1:50) for LCMS analysis using the same buffer to complete the volume.
Figure. S14: CoBioWF acyl-CoA formation with a variety of di-carboxylic acid lengths

Demonstration of the substrate promiscuity of the CoBioW domain of CoBioWF fusion. The HPLC assay results (from 10 – 30 mins shown) for CoBioWF enzymatic reactions with DC₆-DC₉ di-acids (schematic shown above the chromatograms) with CoASH leading to the formation of the corresponding acyl-CoA product: DC₆ (16.5 min), DC₇ (17.3 min) DC₈ (18.3 min) and DC₉ (19.1 min).
Figure. S15: MS analysis of CoBioWF acyl-CoA di-acid reactions
FT ICR-MS analysis of the CoBioWF reactions of the CoBioW domain upon incubation of the enzyme with varying acyl chains lengths (DC₆-DC₉) and CoASH (DC₇ shown in Fig.S8) with A) DC₆-CoA B) DC₈-CoA and C) DC₉-CoA.
Figure. S16: CaBioWF acyl-CoA formation with a variety of mono-carboxylic acid lengths

Mono-acid substrate promiscuity of the CaBioW domain of the CaBioWF fusion. HPLC assay results for CaBioWF enzymatic reactions with C$_6$-C$_{10}$ mono-acids (schematic shown above the chromatograms) with CoASH leading to the formation of the corresponding acyl-CoA product (green): C$_6$ (20.1 mins), C$_7$ (21.2 mins), C$_8$ (22.1 min), C$_9$ (23.4 min) and C$_{10}$ (24.0 min).
Figure. S17: MS analysis of CoBioWF acyl-CoA mono-acid reactions
FT ICR-MS analysis of the CoBioWF reactions of the CaBioW domain upon incubation of the enzyme with varying acyl chains lengths (C₆-C₁₀) and CoASH (C₆ was not observed) with A) C₇-CoA B) C₈-CoA, C) D₉-CoA and D) C₁₀-CoA.
Figure S18: CaBioWF F192Y purification

Purification of the CaBioWF F192Y mutant using A) SDS-PAGE analysis with LMW marker, eluted HT fractions and Superdex S200 fractions corresponding to the peak at 60.5 mL, B) S200 SEC chromatogram monitored at 280 nm with an elution volume of 60.5 mL and C) LC-ESI MS analysis showing protein charge states of the D) obtained protein with a deconvoluted mass of 68567 ± 3.19 Da.
Figure. S19: CaBioWF F192Y acyl-CoA reactions

HPLC assay results for CaBioWF F192Y mutant with DC₆ – DC₈ and C₆- C₈ (schematic shown above the chromatograms) with CoASH leading to the formation of the corresponding acyl-CoA product (green) for DC₇-CoA (17.3 mins) but no acyl-CoA product for the remaining reactions (green) is visible. ATP (purple) and released AMP (blue) are also shown.
Figure S20: Full CaBioWF reactions with various di and mono-carboxylic acid chain lengths.
CaBioWF reactions transforming a range of carboxylic acids (DC$_6$-DC$_9$ and C$_6$-C$_9$) with L-Ala leading to the production of the corresponding aminoketone. Product formation confirmed by LC ESI-MS analysis.
Figure. S21: CaBioWF unusual carboxylic acid reactions

CaBioWF catalysed formation of unusual AON analogues upon incubation of 7-Bromoheptanoic acid and 6-methylhexanoic acid with L-Ala. Product formation confirmed by LC ESI-MS analysis.
CoBioWF Modelling and Simulation

Initially, the CaBioW (M1-T238) and CaBioF (G239-A620) domains were modelled separately using the accurate deep learning architecture ColabFold (see the Experimental Section in the main text). Both domains were predicted with high confidence (pLDDT >90, pTM >0.85, see figure S22A-B) with homodimeric interfaces comparable to experimentally solved structures including BsBioW (PDB: 5FLL, see figure S23B) and EcBioF (PDB: 1DJ9, see figure S24B). CaBioW was modelled with a subdomain architecture shared by type IV ANL enzymes (see also PDB: 5TV5), wherein the catalytic C-terminal subdomain binds its substrates and the structural N-terminal subdomain comprises a dimer interface. The predicted CaBioF shares strong fold-level similarity with several BioF homologues (PDB: 5JAY, 6ONN, 5VNX, 7SSM), as well as other PLP-dependent enzymes such as serine palmitoyltransferases (SPT, PDB: 3A2B, 2X8U) and 2-amino-3-ketobutyrate CoA ligases (KBL, PDB: 7V58, 3TQX, 7BXP). Furthermore, several highly conserved residues that define the binding pocket of each domain were identified by evolutionary conservation analysis, including Y181 and R194 in CaBioW (figure S23C) as well as H380 and K483 in CaBioF (figure S24C). This initial study provided confidence in the ability of ColabFold to accurately predict the tertiary and quaternary structures of the CaBioWF domains.

The full CaBioWF dimer was subsequently modelled, and the top-ranked output (pLDDT 91.8, pTM = 0.68, see figure S22C) was studied in a 10 ns (5 x 10⁶ time steps) molecular dynamics simulation (MDS, see the Experimental Section in the main text). While the individual domains were confidently predicted on a fold-level, there was some uncertainty regarding the relative orientation of the two CaBioW domains, in part due to the disordered intra-domain linker(s) tethering CaBioW and CaBioF together (figure S22C). The predicted CaBioWF model suggests that both domains contribute towards the dimeric interface, and these interfacial contacts are maintained over the course of the MDS (see figure S25). In particular, the CaBioWF complex is stabilised by an average of 36 ± 7 interfacial hydrogen bonds, the majority of which (52%) occur within 2.72-2.93 Å (figure S26A/B). While the average radius of gyration (Rg = 3.57 ± 0.02 nm, Rg max-min = 0.160 nm) suggests that the CaBioWF complex is stable, pairwise RMSD analysis reveals that the bifunctional enzyme exhibits a moderate amount of conformational flexibility, with RMSDs as high as 6 Å occasionally observed (figure S26C-D). Root Mean Square Fluctuation (RMSF) and B-factor analysis identifies the intra-domain linker and the CaBioW domains as the most mobile regions of the protein (figure S27 A-C). In fact, this linker is flexible enough to allow light orientational adjustment of the CaBioW domains within the first 2 ns of the simulation, with one of the CaBioW domains rotating approximately 18.5° inwards from the start of the trajectory (figure S27D). By the midpoint of the simulation, the CaBioW domains had settled, and both BsBioW and EcBioF could be comfortably superimposed onto the CaBioWF complex (figure S28). Interestingly, the CaBioW and CaBioF binding pockets face each other approximately 4.75 nm apart; the proximity and orientation of these binding pockets suggests that the pimeloyl-CoA product of CaBioW can easily diffuse into active site of CaBioF. Taken all together, this in silico study provides insight into the didomain architecture of CaBioWF, and hints towards both its flexibility and the existence of a potential “tunnel” that can channel products from the CaBioW domain to CaBioF (figure S29). This makes CaBioWF an attractive, curious and potentially challenging target for future crystallographic trials.
Figure S22: pLDDT and pTM scores of the predicted CaBioWF domains

A) Confidence metrics for the predicted CaBioW dimer. B) Confidence metrics for the predicted CaBioF dimer. C) Confidence metrics for the predicted CaBioWF dimer.
Figure. S23: A closer inspection of the predicted CaBioW domain

A) CaBioW was predicted to form dimeric contacts between the N-terminal subdomains. B) The crystal structure of BsBioW superimposed on the predicted CaBioW homodimer (RMSD: 1.06 Å, 174 pruned atom pairs). C) Highly conserved pocket residues identified in the predicted CaBioWF. The pimeloyl adenylate was extracted from the BsBioW crystal structure (grey, PDB: 5FLL) and is displayed here for reference.
Figure. S24: A closer inspection of the predicted CaBioF domain
A) CaBioF was predicted to form dimeric contacts, as commonly observed with BioF homologues and other PLP-dependent enzymes. B) The crystal structure of EcBioF superimposed on the predicted CaBioF homodimer (RMSD: 0.99 Å, 289 pruned atom pairs). C) Highly conserved pocket residues identified in the predicted CaBioF model. The AONS:PLP-AON product external aldamine was extracted from the EcBioF crystal structure (grey, PDB: 1DJ9) and is displayed here for reference.
Figure. S25: Visualisation of the CaBioWF simulation
The displayed structures and RMSD plot (vs. t = 0 ns) depict the CaBioWF morphology through time. Chains A and B are coloured lime and cyan respectively.
Figure. S26: A summary of the CaBioWF MDS
A) An examination of the number of inter-chain hydrogen bonds over time. B) A distance distribution of the inter-chain hydrogen bonded contacts. C) The radius of gyration (Rg) of the CaBioWF complex over time. D) A pairwise (2D) RMSD map computed between structures at every time point of the simulation.
Figure. S27: Fluctuation analysis of the simulated CaBioWF complex
A) A per-residue root mean square fluctuation (RMSF) plot calculated for the Cα atoms of the protein. B) A per-residue B-factor plot. C) The per-residue B-factors computed in B mapped onto the CaBioWF structure (pre-simulation). D) Superimposition of the t = 0 ns (lime) and t = 10 ns (cyan) structures, highlighting the mobility of the chain A CaBioW domain; angles were calculated in PyMOL.
Figure S28: The CaBioWF complex sampled midway (t = 5 ns) through the MDS
At this timepoint, the CaBioW domains have relaxed into a stable orientation. 
A) Annotation of the relaxed CaBioWF complex. B) Superimposition of BsBioW (magenta, RMSD: 1.21 Å, 112 pruned atom pairs) and EcBioF (blue, RMSD: 1.19 Å, 252 pruned atom pairs).
Figure S29: A predicted molecular “tunnel” (dashed line) between the CaBioW and CaBioF domains. The structure was sampled at the midpoint ($t = 5$ ns) of the MDS. The ligands displayed (grey) were extracted from the BsBioW and EcBioF crystal structures and are displayed here for reference. A) The BioW and BioF binding pockets face towards each other, providing an easy diffusion path from CaBioW domain to the active site of CaBioF. B) The predicted “tunnel” is estimated to be 4.5-5 nm in length.