De novo acyl carrier proteins display structure-independent modification and sequence novelty

Michael A. Herrera[1*], Zoe Ozols[2], Louise K. Charkoudian[2] and Dominic J. Campopiano[1]

[1] School of Chemistry, The University of Edinburgh, Edinburgh EH9 3FJ, UK

[2] Department of Chemistry, Haverford College, Haverford, Pennsylvania 19041, USA

* Corresponding author

Abstract

Acyl carrier proteins (ACPs) are dynamic, structurally-conserved α-helical proteins central to many primary and secondary metabolic processes. Whilst prior engineering efforts have focused on strategic mutagenesis and "helix swaps", much of the ACP sequence design space remains underexplored. Here, we create diverse variants of the archetypal ACP subclass – AcpP – using a bespoke sequence-generating algorithm (ALGO-CP), which utilises a combined evolutionary and physicochemical design approach. Using ALGO-CP, we generated two soluble candidates – ALGO-055 and ALGO-059 – that can undergo full post-translational modification from *apo*→*holo*→acyl forms *in vitro*, using recombinant modifying enzymes. Building on these successful designs, we further adapted ALGO-CP to produce several chimeras, two of which – ^{ch}ALGO-012 and ^{ch}ALGO-024 – also exhibit full modifiability. Despite *in silico* predictions of ACP-like helicity, we reveal that these proteins lack the canonical α-helical fold of an ACP, whilst remaining soluble and readily modifiable. Of note, both ALGO-055 and ALGO-059 harbour several rare amino acid variations across their sequences, whilst preserving many important acidic "hotspots" involved in key protein-protein interactions. By testing the limits of the AcpP design space, our findings suggest that some key aspects of ACP behaviour – specifically, post-translational modification – can be retained independently of the canonical structure. This work establishes a foundation for probing ACP sequence diversity through a hybrid computational-experimental approach. ALGO-CP is available under AGPL3.0 license: https://github.com/MAHerrera-94/ALGO_CP.

Introduction

Acyl carrier proteins (ACPs), and their related peptidyl carrier proteins (PCPs), are small (8-10 kDa) and ubiquitous proteins that drive metabolic flux within cells.[1-4] These highly dynamic proteins work by activating and shuttling metabolic substrates between enzymes within biosynthetic pathways/assemblies. Their function depends on the precise recognition of enzyme partners through transient and dynamic protein–protein interactions (PPIs), with each cognate partner evolving a complementary interface for ACP/PCP docking and substrate transfer. In addition to essential lipids

and vitamins, countless polyketide and non-ribosomal peptide pharmaceuticals originate from ACP/PCP-driven biosynthesis including glycopeptide antibiotics,[5] avermectins[6] and statins.[7] Without the precision and versatility of these proteins, such indispensable molecules are not possible in nature nor in the clinic.

ACPs are remarkably diverse at the sequence level, with many homologues sharing as little as ~20–30% identity. Despite this, ACPs generally conserve a compact helical fold, comprising four antiparallel helices (αI-IV) connected by flexible loops (Fig. 1A). For an ACP to be functionally competent, it must undergo post-translational modification (PTM) to its *holo*- form via the attachment of a 4′-phosphopantetheinyl (4'-PP) prosthetic group, derived from coenzyme A (CoASH, Fig. 1B). This PTM is catalysed by a 4'-phosphopantetheinyl transferase (4'-PPTase), which fixes the 4'-PP group to an invariant serine residue situated before helix αII. In this *holo*- form, substrates are covalently tethered to the terminal thiol group of the 4'-PP; bound lipophilic substrates can be sequestered within the hydrophobic cavity of the ACP, thereby protecting it from premature hydrolysis or undesired side-reactions (Fig. 1C).[8, 9] The surface of the ACP mediates interactions with partner enzymes, with solvent-exposed residues along helix αII (often called the "universal recognition helix") playing a major role in PPIs. The nature and precision of these PPIs is governed by the ACP surface charge distribution, hydropathy and topography, as evidenced by numerous structural and NMR analyses of ACP complexes.[10-17] Like many ancient enzyme families, such as the short-chain dehydrogenase/reductase,[18] it is widely held that ACPs are robust to sequence variation, provided these key structural, functional and biophysical features are retained.

Efforts to engineer ACPs include rational mutagenesis[19, 20] and the creation of chimeras,[21, 22] in which helical segments are exchanged between natural homologues to introduce new functions and/or re-route biosynthetic pathways. Whilst such strategies have produced functional variants, they explore only a narrow corner of ACP design space, leaving many novel ACP designs – and functions – out of reach. A more rigorous exploration of this sequence space could deepen our understanding of ACP evolution and uncover new insights for engineering. Computational biology is well-placed to address this challenge, yet to the best of our knowledge, no such strategy has been applied to the design of new ACPs.

In this work, we developed a simple and tuneable sequence generation algorithm (ALGO-CP) designed to map the sequence space of ACPs. Rather than strictly relying on phylogenetic constraints, ALGO-CP takes a softened design approach that blends evolutionary conservation with local physicochemical information. In theory, this would permit the controlled sampling of amino acids that are rarely observed in natural homologues, thereby pushing the limits of natural sequence diversity. Using the prokaryotic housekeeping ACP (AcpP) subclass as a template, we applied ALGO-CP to generate diverse *de novo* ACP-like sequences, two of which – ALGO-055 and ALGO-059 –

could undergo PTM to their *holo*- and acylated forms *in vitro*. By re-configuring ALGO-CP, we created two soluble chimeric variants of ALGO-055 and ALGO-059 – named ^{ch}ALGO-012 and ^{ch}ALGO-024 – that also undergo complete PTM. We further reveal by circular dichroism (CD) spectroscopy that ALGO-055 and ALGO-059 are modifiable despite lacking the canonical α-helical structure of an ACP. We propose that this may arise from rare amino acid combinations not featured in natural AcpP homologues, alongside the preservation of key acidic regions necessary for productive PPIs.

Results and Discussion

ALGO-CP concept and summary. ALGO-CP is a tuneable sequence generation algorithm designed to explore the boundaries of the ACP sequence design space (Fig. 1D). It operates on a multiple sequence alignment (MSA) of constituent amino acids using two complementary sub-algorithms, route-$\mathbf{A_C}$ and route-$\mathbf{A_P}$, to capture position-specific conservation and physicochemical information, respectively. The physicochemical properties examined – isoelectric point (pI), Kyte–Doolittle hydropathy (H_{KD}), and van der Waals volume (vdW, in Å^3) – are important for ACP structure, solubility and interaction specificity. This information can be combined via a weighting coefficient r, ranging from 0.00 (purely $\mathbf{A_C}$-focused) to 1.00 (purely $\mathbf{A_P}$-focused), to generate a unified amino acid probability profile ($\mathbf{A_U}$) for each amino acid position, from which new residues are sampled stochastically to construct a new ACP-like sequence (see Experimental for full details). By design, ALGO-CP preserves highly conserved residues (such as the essential, modifiable serine), whilst allowing greater diversity at variable positions. Sequences that are generated exclusively via route-$\mathbf{A_C}$ are constrained to observable substitutions within the input MSA. In contrast, route-$\mathbf{A_P}$ promotes novelty by relaxing evolutionary constraints, instead evaluating the physicochemical profile of each alignment block against all 20 proteinogenic amino acids. In theory, this would enable the incorporation of many rare, yet physiochemically compatible, amino acid variations. We reasoned that a calibrated combination of these two design strategies would allow the exploration of ACP sequence space in a controlled and interpretable manner. To test this, we required a reliable and well-characterised template for the design of new ACPs.

Designing AcpP variants using ALGO-CP. The archetypal ACP, known as AcpP, is essential in prokaryotes.[23] With an estimated 350,000 copies per cell, *Escherichia coli* AcpP (*Ec*AcpP) can interact with >25 other proteins to synthesise several life-critical lipids (fatty acids, phospholipids, lipid A, lipoic acid, biotin),[10, 24-26] mediate nutrient stress responses (via communication with SpoT)[27, 28] and promote chromosome organisation (via interaction with the chromosome partition protein MukB).[29] The multi-functional "pocket knife" versatility of AcpP, together with the sheer abundance of sequence homologues, makes it an ideal starting template for *de novo* ACP design using ALGO-CP.

To seed the design process, we first retrieved 2558 AcpP homologues from UniRef90 using 17 validated AcpP sequences as queries (Table S3). This dataset was cleaned to remove duplicate and incomplete sequences, bringing the new total to 2167. These homologues were aligned using MAFFT,[30] and the resulting MSA was used as input for ALGO-CP. We configured ALGO-CP to generate sets of 5000 unique sequences, with each iteration raising the weighting coefficient r from 0.00 to 1.00 in increments of 0.05. For each set, we computed the mean positional values of H_{KD}, pI, and vdW volume, and compared these to the input MSA via pairwise correlation (Fig. S1). As anticipated, the sequences generated by ALGO-CP became more divergent from the input MSA with increasing r, reflecting a shift towards sequence novelty permitted by the softer computational strategy of route-$\mathbf{A_P}$ (Fig. 2A).

To approximate the balance point between the two sub-algorithms, we averaged the pairwise property correlations within each sequence set and analysed their residual proximity to the grand property mean across all sequence sets (Fig. 2B-C). Negative residuals indicate relatively conservative sequences, whereas positive residuals indicate relatively divergent sequences. Based on this proximity analysis, we estimated that sequences generated using r = 0.60, which returned the smallest residual value (8.75 x10^{-4}), may best reflect this strategic balance between conservative vs explorative sequence design. A subset of 100 randomly-sampled, unique sequences generated at r = 0.60 were found to exhibit intermediate overall sequence properties relative to those sampled from pure $\mathbf{A_C}$- and $\mathbf{A_P}$-guided designs (Fig. 2D-E, Tables S4-5, S7-8, S10-11). Furthermore, this subset could be accurately folded using AlphaFold3[31] (pLDDT = 90.1 ± 2.75), suggesting that they may adopt ACP-like folds if expressed in soluble form (Fig. 2F, Tables S6, S9, S12). However, since ALGO-CP does not explicitly predict solubility or biochemical function, we proceeded to characterise a sample of these sequences experimentally.

Expression, purification and PTM of ALGO sequences. The ability to undergo functional PTM is an essential hallmark of all naturally occurring ACPs. For *in vitro* study, recombinant PTM enzymes can efficiently convert *apo*- ACPs to their functionally competent *holo*- and acyl- forms. *Bacillus subtilis* Sfp (*Bs*Sfp)[32] or *E. coli holo*-ACP-synthase (*Ec*AcpS)[33, 34] are the most commonly used 4'-PPTases for *apo*→*holo* PTM. *Holo*- ACPs can be further functionalised to their acyl form using a fatty acyl-ACP synthetase (AasS).[35] AasS activates linear fatty acids as acyl adenylates, which are subsequently loaded onto the ACP via the terminal thiol group of the 4′-PP arm. Amongst this subclass of adenylating enzymes, the *Vibrio harveyi* B392 AasS (*Vh*AasS) is the most extensively characterised and versatile member,[36-39] and it is a remarkably useful tool for *in vitro* ACP acylation. We reasoned that the successful PTM of our *de novo* sequences would serve as a foundational demonstration of

ACP-like behaviour. Thus, we purified and deployed recombinant *Bs*Sfp, *Ec*AcpS and *Vh*AasS in our PTM pathway (Fig. 3A, Fig S4); *Ec*AcpP was used as a positive control throughout (Fig. S11-13).

From the previous subset of r = 0.60 sequences sampled for property analysis, seven candidates – ALGO-(013, 023, 040, 044, 055, 057, 059) – were randomly selected for *in vitro* characterisation (Table S13, Fig. S2). The candidate sequences were codon-optimised for *E. coli*, synthesised and cloned into pET28a with a C-terminal His-tag for ease of purification. Following heat-shock transformation and antibiotic selection, overnight protein expression could be induced with 1 mM IPTG at low temperatures (18 °C). Of the seven candidates, ALGO-(023, 044, 057) were expressed as insoluble inclusion bodies, and were not taken further. The remaining candidates could be captured using Ni^{2+}-affinity resin and further cleaned using a PES centrifugal filtration unit (Fig. 3B and Fig. S5-8).

Following dialysis, the identity and PTM status of each soluble ALGO candidate was confirmed by LC/ESI-MS (Table S14, Fig S14-16, S18). Notably, both ALGO-055 and ALGO-059 were purified partly in *holo*- form, as indicated by a molecular weight (MW) increase of 340 Da consistent with the covalent attachment of 4′-PP. Encouragingly, this implied some *in vivo* interaction with the endogenous *Ec*AcpS during protein expression. Both candidates could be readily converted to their respective *holo* forms *in vitro* using purified *Ec*AcpS, in the presence of Mg^{2+}, CoASH and dithiothreitol (DTT, see Fig. 3C-D and Fig. S17, S19). To our surprise, both ALGO-055 and ALGO-059 could not be fully converted by *Ec*AcpS when immobilised on Ni^{2+}-affinity resin, which is an established method for the reliable and complete PTM of His-tagged ACPs. This observation suggests that immobilisation could impose a significant conformational constraint on these ALGO sequences, hampering productive engagement with *Ec*AcpS (Fig. S20-21). Furthermore, only partial conversion was achieved using *Bs*Sfp, both *in vitro* (using purified recombinant enzyme) and *in vivo* (using the *sfp* knock-in strain *E. coli* BAP1),[40] suggesting that these sequences struggle to establish productive PPIs with this 4'-PPTase (Fig. S22-23). Conversely, neither ALGO-013 nor ALGO-040 could undergo *apo*→*holo* PTM at all, and these candidates were not examined further (Fig. S24-25).

Using our standout candidates ALGO-055 and ALGO-059, we next attempted the covalent attachment of a lauroyl (C_{12}) acyl chain *in vitro* using *Vh*AasS. The acylation of both *holo*-ALGO-055 and *holo*-ALGO-059 proceeded in the presence of Mg^{2+}, ATP, DTT, lauric acid and 1% DMSO. Gratifyingly, we observed >99% acylation of *holo*-ALGO-055 and *holo*-ALGO-059 by LC/ESI-MS, as signified by a MW increase of 182 Da, after only 1 hour of room temperature incubation (Fig. 3C-D and Fig. S26-27). Thus, we identified two unique ACP-like sequences that can not only acquire 4'-PP, but can also fix acyl cargo.

Encouraged by these results, we configured ALGO-CP (via route-$\mathbf{A_C}$) to stochastically generate chimeric variants of ALGO-055 and ALGO-059, mimicking the process of molecular breeding. A

174 sequence alignment of both ALGO candidates was used as input. From 5000 generated sequences,
175 five (^{ch}ALGO-[009, 012, 024, 044, 097]) were sampled from a random subset (n = 100) for
176 experimental testing (Fig. 3E, Table S13). The chimeras ^{ch}ALGO-012 and ^{ch}ALGO-024 (Fig. 3F) were
177 expressed in soluble form and could be further purified by Ni^{2+}-affinity chromatography (Fig. 3G, S9-
178 10). As before, these proteins were also recovered partly in *holo*- form, signalling *in vivo* PTM by
179 endogenous *Ec*AcpS. Like their parental sequences, both candidates could readily undergo complete
180 PTM (*apo*→*holo*→acyl) *in vitro* using *Ec*AcpS and *Vh*AasS (Table S14, Fig. 3H-I, Fig. S28-33).

181 To the best of our knowledge, both ALGO-055 and ALGO-059, and their chimeric variants ^{ch}ALGO-
182 012 and ^{ch}ALGO-024, are the first examples of expressible, soluble and modifiable ACPs designed
183 entirely *de novo* by a computer algorithm. More broadly, this result showcases how ALGO-CP can be
184 tuned to probe the sequence design space and to create further diversity from initial successful
185 designs.

186

187 **Biophysical characterisation**. The canonical *Ec*AcpP is a highly acidic and conformationally
188 dynamic α-helical bundle. This combined electronegativity and structural plasticity is crucial for
189 substrate shuttling and PPIs. To assess whether our ALGO sequences share these characteristics
190 with natural homologues, we used a combination of computational and experimental approaches,
191 starting with structural models generated by AlphaFold3 (pLDDT = 93.74 and 93.52 for ALGO-055
192 and ALGO-059, respectively, Fig. S34). These models were used as input for separate 200 ns
193 molecular dynamics simulations (MDS); as a reference, we also simulated the 1.1 Å crystal structure
194 of *Ec*AcpP (PDB: 1T8K).[41] Analysis of the MDS trajectories showed that both ALGO proteins retained
195 their overall structural integrity throughout the simulation, with gyrational radii and RMSD (1D and 2D)
196 profiles comparable to *Ec*AcpP (Fig. 4A and S35-36). RMSF analysis revealed similar primary chain
197 flexibility across all three ACPs, with well-structured helices and flexible inter-helical loops (Fig. 4B-
198 C). Secondary structure prediction using PSIPRED[42] further supported these observations, assigning
199 both ALGO sequences as predominantly α-helical (Fig. S37-38). DISOPRED3[43] analysis also
200 predicted minimal intrinsic disorder (Fig. S39-40). Furthermore, the solvent-exposed surfaces of these
201 proteins exhibit similar Coulombic electrostatic potential and molecular lipophilicity potential when
202 compared to *Ec*AcpP (Fig. S41-42). Together, these *in silico* analyses suggest that ALGO-055 and
203 ALGO-059 possess high helical propensity and may possess similar biophysical properties to natural
204 AcpP homologues. We reasoned that such predictions align with the experimental evidence collected
205 thus far.

206 However, neither *holo*-ALGO-055 nor *holo*-ALGO-059 demonstrated the diagnostic stalls of an α-
207 helix when studied by CD spectroscopy. In stark contrast to our *in silico* predictions, the CD spectra
208 of *holo*-ALGO-055 and *holo*-ALGO-059 deviate significantly from the characteristic double minima

209 (208 nm and 222 nm), strong maxima (193 nm) and isodichroic point (201 nm) of a well-folded helical
210 protein, as exemplified by *holo-Ec*AcpP (Fig. 4D). Given that some ACPs, such as *V. harveyi* AcpP,
211 are natively unfolded in the absence of Mg^{2+}, we supplemented the buffer with 5 mM $MgCl_2$ to evaluate
212 whether the ALGO variants can reform their helicity with the assistance of divalent cations.
213 Remarkably, this supplement only slightly perturbed the CD spectra and did not clearly restore helicity
214 to either ALGO protein, indicating that they may well exist as minimally-structured, predominantly non-
215 helical ensembles.

216 The discrepancy between our CD spectroscopy results and established findings on *V. harveyi* AcpP
217 prompted a closer inspection of these ALGO sequences. It has been reported that a single $A_{76}H$
218 mutation in *V. harveyi* AcpP restores its helicity via π-stacking interactions with Y_{72}, even in the
219 absence of divalent cations.[44] Whilst ALGO-055 and ALGO-059 retain the equivalent Y_{72} residue,
220 ALGO-055 harbours the "destabilising" A_{76} variant, whereas ALGO-059 carries the "stabilising" H_{76}.
221 Taken together with our Mg^{2+} supplementation experiment, this suggests that ALGO-055 and ALGO-
222 059 may be too divergent from natural AcpP homologues for previously described stabilisation
223 mechanisms to be effective. Alongside our PTM results, this points toward an underlying, hitherto
224 undescribed sequence-structure-function relationship. Accordingly, we performed a deeper post-hoc
225 analysis of our ALGO sequences to explore this relationship further.

226

227 **Primary sequence analysis.** To assess the sequence divergence of our ALGO candidates, we
228 quantified the rarity of all amino acid variations relative to our collection of 2167 AcpP homologues.
229 We then evaluated their evolutionary likelihood and physicochemical dissimilarity relative to *Ec*AcpP
230 using BLOSUM62[45] and Grantham distance[46] matrices, respectively.

231 Our analysis of ALGO-055 and ALGO-059 revealed that both proteins harbour at least 16 amino acid
232 variations that are represented by <5% of the AcpP homologues used in the input MSA. Of these
233 variations, ten (ALGO-055) and seven (ALGO-059) are present in <1% (Table S15-16, Fig. 5A). These
234 variations occur across the sequence length of the proteins, with the regions typically associated with
235 helix αII and αIII being the best-preserved by ALGO-CP. By BLASTp, the closest homologues of
236 ALGO-055 and ALGO-059 belonged to *Providencia heimbachae* (69.23% identity) and
237 *Lacimocrobium alkaliphilum* (65.38% identity), respectively. When compared to *Ec*AcpP, both
238 proteins contain ≥30 amino acid variations, approximately half of which are non-conservative by
239 BLOSUM62 analysis with median Grantham distance scores of 89 (ALGO-055) and 78 (ALGO-059)
240 (Table S15-16, Fig. 5B-C). Remarkably, ALGO-055 and ALGO-059 share only 51.28% sequence
241 identity with each other, with 17 non-conservative variations scoring a median Grantham distance of
242 91 (Table S17).

243 Overall, the Grantham distance scores of many of these variations fall within the lower 50th percentile
244 of all possible amino acid pairings (i.e., <96), suggesting moderate, rather than extreme,
245 physicochemical divergence. Nevertheless, it remains possible that the accumulation of these rare
246 and non-conservative amino acid variations contributes to the observed destabilisation of the
247 canonical α-helical structure, though the precise variants/mechanisms responsible for this effect are
248 difficult to predict. We note, however, that ALGO-055 and ALGO-059 preserve several acidic
249 "hotspots", clustered mostly downstream from the invariant serine, that are known to interface with
250 the highly electropositive docking surfaces of *Ec*AcpS and *Vh*AasS (Fig. S41-S43).[39, 47] This
251 electrostatic complementarity may still provide adequate priming/orientation for productive PPIs, even
252 in the absence of an ordered structure.

253 Similarly, it is difficult to ascertain deleterious amino acid variations (or combinations of mutations)
254 from such a limited set of *de novo* sequences. However, a meta-analysis of our initial ALGO
255 candidates identified 95 total variations that are exclusively found in unsuccessful ALGO sequences
256 (i.e., sequences that failed to express in soluble form or undergo PTM) versus *Ec*AcpP, ALGO-055
257 and ALGO-059 (Table S18, Fig. 5D). Of these, 69 are featured in <10% of all 2167 AcpP homologues
258 studied, including 38 that are found in <1%. In total, we identified 11 non-conservative amino acid
259 variations with Grantham distances >120 when compared to *Ec*AcpP; many of these are clustered
260 around αI and αIV, with none occurring along the crucial recognition helix αII (Table S18, Fig. 1A-C,
261 5E-F). We propose that sequences enriched with these rare and/or unorthodox amino acid variants
262 may be predisposed to poor solubility or functionality, which may explain why such variations were
263 selected against during AcpP evolution.

264 Both ALGO-055 and ALGO-059 occupy a peculiar region of an otherwise fragile sequence design
265 space, in which some biochemical function – specifically, essential PTM – is retained in the absence
266 of a hallmark α-helical structure. In 2005, Christopher Walsh and team used phage display to identify
267 a 11-mer PCP fragment (ybbR-13) that undergoes *apo→holo* PTM via interaction with *Bs*Sfp.[48]
268 Notably, even this highly truncated peptide adopts a clear α-helical conformation in solution, further
269 highlighting the idiosyncrasy of these ALGO designs. The behaviour of these proteins may reflect a
270 "precursor-like" state to the canonical ACP fold, one in which an amorphous configuration of key
271 contact residues is still sufficient for PTM. As biosynthetic assemblies and physiological requirements
272 grew more complex, the emergence of a defined α-helical fold may have conferred advantages in
273 partner selectivity and/or kinetic efficiency through more nuanced PPIs. Based on our data, we
274 postulate that ACP sequences with greater helical propensity – with or without the assistance of
275 chemical chaperones - were gradually selected to meet the demands of biosynthetic precision and
276 function, rather than out of necessity for PTM alone. Assuming that the canonical helical structure of
277 AcpP is essential for cell fitness, we propose that ALGO-055 and ALGO-059 could be evolved, under
278 *in vivo* selective pressure, to determine whether and how this fold can be re-acquired to support

complex physiological functions. To this end, the *E. coli* Δ*acpP* knockout strain CY1877 would make an excellent background for a live/death complementation screen.[22]

Conclusion

Despite their small size, ACPs are deceptively complex and nuanced proteins. Their sheer versatility enables certain subclasses – chiefly, AcpP – to operate at the heart of both cell metabolism, division and physiology. In this study, we developed a tuneable sequence-generating algorithm, ALGO-CP, to create novel AcpP-like sequences that both skirt and surpass the natural design constraints of this ancient protein family. Even from a limited sample of *de novo* sequences, we uncovered surprising insights that both facilitate and test our understanding of ACP engineering. We further observe the tendency for state-of-the-art structural prediction tools to over-predict α-helicity in *our de novo* sequences, perhaps reflecting training biases that make this approach less effective for our specific application. For the first time, we have set the groundwork for computer-led ACP design supported by a simple, yet effective, screen for essential ACP-like behaviour. We propose that ALGO-CP could be adapted to study different ACP subclasses across different kingdoms of life, including those involved in fungal polyketide biosynthesis[49] or mitochondrial function.[50, 51] With further experimental feedback, we envision that ALGO-CP – especially when coupled with machine learning and high-throughput screening – could become a powerful tool for pioneering entirely new ACP lineages, with potential applications in metabolic engineering and beyond.

Experimental

All materials, expression plasmids, media, reagent stock solutions and LC/ESI-MS methods are detailed in the SI.

Retrieval of AcpP homologues. Sequence homologues were retrieved from UniRef90 via ConSurf.[52] In brief, 17 AcpP homologues (table S3) were used as queries to retrieve the top 150 closest sequence homologues each (95-50% sequence identity). These homologous sequences were compiled into a single FASTA file, which was filtered for duplicate/incomplete entries and aligned via MAFFT v7[53] using default parameters.

Amino acid and sequence properties. For ALGO-CP, the requisite physicochemical data for each proteinogenic amino acid were retrieved from AAindex.[54] Sequence GRAVY and MW were computed

using Multiple Protein Profiler 1.00.[53] Sequence pI was computed using Isoelectric Point Calculator 2.0.[55]

Predictive structural modelling and molecular dynamics simulation. All structural predictions of were performed using AlphaFold3[31] (via AlphaFold server). Simulations were performed using GROMACS 2021.[56] Protein charges were computed using CHARMM36 all-atom forcefield.[57] The model was solvated in TIP3P water in a rhombic dodecahedron box, and the net protein charge was counterbalanced 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 200 ns (1 x10^9 time steps) production MD, the trajectory was re-centred with additional rotational and translational fitting. Further analysis was performed in GROMACS using gmx gyrate, gmx rms and gmx rmsf. UCSF Chimera v1.18[58] was used for trajectory visualisation and for computing pairwise RMSDs. General model viewing and analysis was performed using ChimeraX v1.8.[59]

ALGO-CP overview. ALGO-CP begins by extracting position-specific data from alignment blocks, including residue frequency and the average H_{KD}, pI, and vdW. Heavily-gapped positions can be excluded from analysis; in this study, only alignment blocks containing ≥50 residues were considered. For any given aligned position i, route-$\mathbf{A_C}$ generates a probability distribution P_i^a based solely on the frequency of each observed amino acid in the alignment block:

$$\mathbf{A_C} = P_i^a = \frac{f_i^a}{N_i}$$

Where f is the frequency of a given amino acid, and N is the total number of aligned residues in the alignment block. For alignment blocks which exhibit ≥90% amino acid identity, the modal amino acid is selected automatically, thus ensuring the preservation of critical functional residues such as the catalytic serine. In contrast, route-$\mathbf{A_p}$ interprets physicochemical descriptions of alignment blocks, rather than explicit conservation information. For any given aligned position i, the mean values of the three physicochemical properties - H_{KD}, pI and vdW - are computed across the alignment block:

$$\vec{\mu}_i = \left[\mu_i^{H_{KD}}, \mu_i^{pI}, \mu_i^{vdW}\right]$$

Each proteinogenic amino acid a is then scored based on the similarity of its physicochemical profile ($\vec{x}_a$) to $\vec{\mu}_i$, where:

$$\vec{x}_a = \left[x_a^{H_{KD}}, x_a^{pI}, x_a^{vdW}\right]$$

The similarity score S_i^a is calculated using the weighted reciprocal difference for each physicochemical property:

$$S_i^a = \sum_k \frac{w_k}{\left|x_a^k - \mu_i^k\right| + \varepsilon}$$

Where k is the physicochemical property, w_k is a user-defined weight assigned to the physicochemical property, and ε is a miniscule constant (10^{-6}) to prevent division by zero. In this study, the weights set for H^{KD}, pI and vdW were 0.22, 0.44 and 0.33, respectively. These scores are subsequently normalised to give a distinct probability distribution based on this physicochemical data:

$$\mathbf{A_P} = \tilde{S}_i^a = \frac{S_i^a}{\sum_{n \in Y} S_i^n}$$

Where $\tilde{S}_i^a$ is the normalised similarity score for a given amino acid, and Y denotes the set of all 20 canonical (proteinogenic) amino acids. This scoring function favours amino acids whose physicochemical properties closely match the local average at each position, enabling route-$\mathbf{A_p}$ to explore substitutions not necessarily observed in natural sequences, but are compatible with the general physicochemical properties of the sequence position.

The distributions generated by route-$\mathbf{A_C}$ and route-$\mathbf{A_P}$ can be weighted and combined into a unified probability profile $\mathbf{A_U}$, allowing the user to bias amino acid selection towards conservation or physicochemical-guided design:

370 $$\mathbf{A_U} = r \cdot \mathbf{A_P} + (1 - r) \cdot \mathbf{A_C}$$

371

372 Where r is a user-defined weight ranging from 0-1. ALGO-CP then generates new sequences end-
373 to-end by stochastically sampling amino acids according to the positional A_U.

374

375 **General procedure for recombinant expression.** For each expression construct used in this work,
376 an exhaustive list of antibiotics, growth media, expression conditions can be found in table S2. The
377 desired expression construct was used to transform chemically-competent *E. coli* BL21 (DE3) cells
378 via the heat-shock method. Colonies were developed overnight on antibiotic VLB-agar plates. A single
379 colony was propagated overnight in antibiotic VLB media (50 mL) by shaking incubation (37 °C). The
380 cells were subcultured (OD_{600} = 0.1, 37 °C) in fresh antibiotic VLB or Autoinduction Media (500 mL)
381 as required, until mid-log phase. Cultures requiring manual induction were cooled to room
382 temperature, and protein expression was induced by the addition of IPTG. All protein expression
383 proceeded with rigorous agitation at the requisite temperature and time. The biomass was harvested
384 by centrifugation using a Fiberlite F14-6 x 250y fixed-angle rotor (7000 rpm, 5 minutes), consolidated
385 into 2-5 g pellets using a Fiberlite F15-8 x 50cy fixed angle rotor (7000 rpm, 10 minutes) and stored
386 at -20 °C.

387

388 **Purification of ALGO-(013, 040, 055, 059) and *Ec*AcpP.**

389 Potassium phosphate (100 mM, pH 7.6) with glycerol (10% v/v) was used as the buffer base
390 throughout, with varying imidazole strengths as described. Cell pellets were defrosted and
391 resuspended (10% w/v) in ice-cold resuspension buffer (10 mM imidazole). Benzamidine
392 hydrochloride (1 mM) was added to the resuspension and the cells were lysed by sonication (10
393 second pulse, 10 second cooldown, 15 cycles) on ice. Cell debris was pelleted by high-speed
394 centrifugation using a Fiberlite F15-8 x 50cy fixed angle rotor (13000 rpm, 45 minutes, 4 °C). The cell-
395 free extract was clarified by filtration (Millex-HP 0.45 µm polyethersulfone, Merck). The recombinant
396 protein was captured from cell-free extract using a Histrap HP (1 mL) column and washed (35 mM
397 imidazole, 15-20 mL). Protein was eluted in fractions (300 mM imidazole, 1 mL) and pooled. ALGO
398 proteins were passed through a centrifugal filtration unit (30-100 kDa MWCO) to remove protein
399 contaminants, and the flowthrough was dialysed in imidazole-free buffer. *Ec*AcpP was polished by
400 size exclusion chromatography using HiLoad 16/600 Superdex 75 pg (120 mL) column. For long-term
401 storage, protein aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

402

403 **Purification of *Bs*Sfp.**

404 HEPES (200 mM, pH 7.5) with glycerol (10% v/v) was used as the buffer base throughout, with varying
405 imidazole strengths as described. Cell pellets were defrosted and resuspended (10% w/v) in ice-cold
406 imidazole-free resuspension buffer. Benzamidine hydrochloride (1 mM) was added to the
407 resuspension and the cells were lysed by sonication (10 second pulse, 10 second cooldown, 15
408 cycles) on ice. Cell debris was pelleted by high-speed centrifugation using a Fiberlite F15-8 x 50cy
409 fixed angle rotor (13000 rpm, 45 minutes, 4 °C). The cell-free extract was clarified by filtration (Millex-
410 HP 0.45 µm polyethersulfone, Merck). The recombinant protein was captured from cell-free extract
411 using a HiTrap TALON Crude (1 mL) column and washed (10 mM imidazole, 15-20 mL). *Bs*Sfp was
412 eluted in fractions (150 mM imidazole, 1 mL), pooled and dialysed in imidazole-free buffer. For long-
413 term storage, protein aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

414

415 ***In vitro apo→holo→*acyl PTM of ALGO candidates and *Ec*AcpP.** An *apo→holo* conversion mixture
416 (100 µL) containing purified ACP (20 µM), *Bs*Sfp/*Ec*AcpS (2 µM), $MgCl_2$ (5 mM), DTT (1 mM) and
417 CoASH (100 µM) was prepared in potassium phosphate buffer (100 mM, pH 7.6) with glycerol (10%
418 v/v). The conversion proceeded at room temperature for 1 hour, after which *Vh*AasS (3 µM), ATP
419 (100 µM) and lauric acid (100 µM, from a 1000X stock prepared in DMSO, see SI Materials and
420 Methods) were added directly to initiate *holo→acyl* conversion (104 µL total volume). *Holo→acyl*
421 conversions proceeded at room temperature for an additional hour prior to LC/ESI-MS analysis.

422

423 **Circular Dichroism Spectroscopy.** CD spectra were collected using an Aviv Model 410A circular
424 dichroism spectropolarimeter. Protein samples (30 µM) were prepared in sodium phosphate buffer
425 (50 mM, pH 7.6, 300 µL). Samples were dispensed into a High Precision Quartz SUPRSIL cuvette
426 with 0.1 cm pathlength (Hellma Analytics). The spectropolarimeter was purged with nitrogen for two
427 hours. The UV lamp was warmed for 30 minutes, after which CD spectra were collected at 25 °C with
428 a range of 190–250 nm using bandwidth of 1 nm, a 0.5 nm step size, averaging time of 3 seconds,
429 and 5 scans. The resulting spectrum was converted to units of mean residue ellipticity (MRE) using
430 the protein amino acid sequence and the sample concentration. Analysis of protein secondary
431 structure characteristics was conducted by uploading normalized data in units of MRE to the web
432 server BeStSel.

433

434 ## Author contributions

435 M.H conceived of the study with conceptual input from D.J.C and L.K.C. M.H designed the algorithm,
436 performed wet and dry lab experiments and analysed the data. Z.O performed wet lab experiments

437 and data analysis under supervision of L.K.C. M.H, D.J.C and L.K.C wrote the manuscript. All authors
438 have given approval to the final version of the manuscript.

439

440 **Abbreviations**

441 ATP: adenosine triphosphate; RMSD: root mean square deviation; RMSF: root mean square
442 fluctuation.

443

444 **Conflict of Interest**

445 There are no conflicts to declare.

446

447 **Data Availability**

448 The data supporting this article have been included as part of the Supplementary Information. The
449 code for ALGO-CP and related data analysis can be found at https://github.com/MAHerrera-
450 94/ALGO_CP. The version of the code employed for this study is version 1.0.

451

452 **Acknowledgements**

453 We thank the Biotechnology and Biological Sciences Research Council (BBSRC) for funding (grant
454 number: BB/Y002210/1). MS data were acquired on instruments funded by the Engineering and
455 Physical Sciences Research Council (EPSRC, EP/K039717/1). We would like to thank Dr.
456 Christopher Wells-Wood, Dr. Antonia Mey and Martyna Kasprzyk (University of Edinburgh) for their
457 invaluable advice and enthusiastic support of this work. We thank Dr. Gustavo Perez-Ortiz (University
458 of Edinburgh) for the expression and purification of *Vh*AasS and *Ec*AcpS. We also thank Prof. Robert
459 Fairman (Haverford College) for his assistance with CD spectroscopy.

460

461 **Footnotes**

462 M.A.H dedicates this manuscript to his wife Claudia.

463

464 **References**

465 1. J. Crosby and M. P. Crump, *Nat Prod Rep*, 2012, **29**, 1111-1137.
466 2. R. M. M. Francois, J. M. Massicard and K. J. Weissman, *Nat Prod Rep*, 2025, **42**, 324-358.
467 3. C. T. Walsh, *Nat Prod Rep*, 2016, **33**, 127-135.

468 4. L. Buyachuihan, F. Stegemann and M. Grininger, *Angew Chem Int Ed Engl*, 2024, **63**,
469 e202312476.
470 5. G. Yim, M. N. Thaker, K. Koteva and G. Wright, *J Antibiot (Tokyo)*, 2014, **67**, 31-41.
471 6. H. Ikeda and S. Omura, *Chem Rev*, 1997, **97**, 2591-2610.
472 7. C. D. Campbell and J. C. Vederas, *Biopolymers*, 2010, **93**, 755-763.
473 8. A. Roujeinikova, W. J. Simon, J. Gilroy, D. W. Rice, J. B. Rafferty and A. R. Slabas, *J Mol Biol*,
474 2007, **365**, 135-145.
475 9. S. C. Epstein, A. R. Huff, E. S. Winesett, C. H. Londergan and L. K. Charkoudian, *Nat*
476 *Commun*, 2019, **10**, 2227.
477 10. V. Agarwal, S. Lin, T. Lukk, S. K. Nair and J. E. Cronan, *Proc Natl Acad Sci U S A*, 2012, **109**,
478 17406-17411.
479 11. M. J. Cryle and I. Schlichting, *Proc Natl Acad Sci U S A*, 2008, **105**, 15696-15701.
480 12. G. W. Heberlig, J. J. La Clair and M. D. Burkart, *Nature*, 2025, **638**, 261-269.
481 13. J. T. Mindrebo, A. Chen, W. E. Kim, R. N. Re, T. D. Davis, J. P. Noel and M. D. Burkart, *ACS*
482 *Catal*, 2021, **11**, 6787-6799.
483 14. L. E. Misson, J. T. Mindrebo, T. D. Davis, A. Patel, J. A. McCammon, J. P. Noel and M. D.
484 Burkart, *Proc Natl Acad Sci U S A*, 2020, **117**, 24224-24233.
485 15. C. Nguyen, R. W. Haushalter, D. J. Lee, P. R. Markwick, J. Bruegger, G. Caldara-Festin, K.
486 Finzel, D. R. Jackson, F. Ishikawa, B. O'Dowd, J. A. McCammon, S. J. Opella, S. C. Tsai and
487 M. D. Burkart, *Nature*, 2014, **505**, 427-431.
488 16. A. Miyanaga, R. Ouchi, F. Kudo and T. Eguchi, *Acta Crystallogr F Struct Biol Commun*, 2021,
489 **77**, 294-302.
490 17. M. Leibundgut, S. Jenni, C. Frick and N. Ban, *Science*, 2007, **316**, 288-290.
491 18. C. Filling, K. D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R. Ladenstein, H.
492 Jornvall and U. Oppermann, *J Biol Chem*, 2002, **277**, 25677-25684.
493 19. K. Li, Y. I. Cho, M. A. Tran, C. Wiedemann, S. Zhang, R. S. Koweek, N. K. Hoang, G. S.
494 Hamrick, M. A. Bowen, B. Kokona, P. Stallforth, J. Beld, U. A. Hellmich and L. K. Charkoudian,
495 *ACS Chem Biol*, 2025, **20**, 197-207.
496 20. X. Liu, W. M. Hicks, P. A. Silver and J. C. Way, *Biotechnol Biofuels*, 2016, **9**, 24.
497 21. Y. I. Cho, C. L. Armstrong, A. Sulpizio, K. K. Acheampong, K. N. Banks, O. Bardhan, S. J.
498 Churchill, A. E. Connolly-Sporing, C. E. W. Crawford, P. L. Cruz Parrilla, S. M. Curtis, L. M.
499 De La Ossa, S. C. Epstein, C. J. Farrehi, G. S. Hamrick, W. J. Hillegas, A. Kang, O. C. Laxton,
500 J. Ling, S. M. Matsumura, V. M. Merino, S. H. Mukhtar, N. J. Shah, C. H. Londergan, C. A.
501 Daly, B. Kokona and L. K. Charkoudian, *Biochemistry*, 2022, **61**, 217-227.
502 22. L. Zhu and J. E. Cronan, *J Biol Chem*, 2015, **290**, 13791-13799.
503 23. J. E. Cronan, *Biochem J*, 2023, **480**, 855-873.

24. A. Chen, R. N. Re, T. D. Davis, K. Tran, Y. W. Moriuchi, S. Wu, J. J. La Clair, G. V. Louie, M. E. Bowman, D. J. Clarke, C. L. Mackay, D. J. Campopiano, J. P. Noel and M. D. Burkart, *J Am Chem Soc*, 2024, **146**, 1388-1395.
25. T. G. Bartholow, T. Sztain, A. Patel, D. J. Lee, M. A. Young, R. Abagyan and M. D. Burkart, *Commun Biol*, 2021, **4**, 340.
26. A. Masoudi, C. R. Raetz, P. Zhou and C. W. t. Pemble, *Nature*, 2014, **505**, 422-426.
27. A. Battesti and E. Bouveret, *Mol Microbiol*, 2006, **62**, 1048-1063.
28. A. Battesti and E. Bouveret, *J Bacteriol*, 2009, **191**, 616-624.
29. J. P. Prince, J. R. Bolla, G. L. M. Fisher, J. Makela, M. Fournier, C. V. Robinson, L. K. Arciszewska and D. J. Sherratt, *Nat Commun*, 2021, **12**, 6721.
30. K. Katoh and D. M. Standley, *Mol Biol Evol*, 2013, **30**, 772-780.
31. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Zemgulyte, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Zidek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis and J. M. Jumper, *Nature*, 2024, **630**, 493-500.
32. L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry*, 1998, **37**, 1585-1595.
33. R. H. Lambalot and C. T. Walsh, *J Biol Chem*, 1995, **270**, 24658-24661.
34. R. S. Flugel, Y. Hwangbo, R. H. Lambalot, J. E. Cronan, Jr. and C. T. Walsh, *J Biol Chem*, 2000, **275**, 959-968.
35. J. Beld, K. Finzel and M. D. Burkart, *Chem Biol*, 2014, **21**, 1293-1299.
36. Y. Jiang, C. H. Chan and J. E. Cronan, *Biochemistry*, 2006, **45**, 10008-10019.
37. Y. Jiang, R. M. Morgan-Kiss, J. W. Campbell, C. H. Chan and J. E. Cronan, *Biochemistry*, 2010, **49**, 718-726.
38. H. Huang, S. Chang, T. Cui, M. Huang, J. Qu, H. Zhang, T. Lu, X. Zhang, C. Zhou and Y. Feng, *PLoS Pathog*, 2024, **20**, e1012376.
39. H. Huang, C. Wang, S. Chang, T. Cui, Y. Xu, M. Huang, H. Zhang, C. Zhou, X. Zhang and Y. Feng, *Nat Struct Mol Biol*, 2025, **32**, 802-817.
40. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science*, 2001, **291**, 1790-1792.
41. X. Qiu and C. A. Janson, *Acta Crystallogr D Biol Crystallogr*, 2004, **60**, 1545-1554.
42. D. W. A. Buchan, L. Moffat, A. Lau, S. M. Kandathil and D. T. Jones, *Nucleic Acids Res*, 2024, **52**, W287-W293.
43. D. T. Jones and D. Cozzetto, *Bioinformatics*, 2015, **31**, 857-863.

541 44. D. I. Chan, B. C. Chu, C. K. Lau, H. N. Hunter, D. M. Byers and H. J. Vogel, *J Biol Chem*,
542 2010, **285**, 30558-30566.
543 45. S. Henikoff and J. G. Henikoff, *Proc Natl Acad Sci U S A*, 1992, **89**, 10915-10919.
544 46. R. Grantham, *Science*, 1974, **185**, 862-864.
545 47. A. M. Marcella, S. J. Culbertson, M. A. Shogren-Knaak and A. W. Barb, *J Mol Biol*, 2017, **429**,
546 3763-3775.
547 48. J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R.
548 Kolter and C. T. Walsh, *Proc Natl Acad Sci U S A*, 2005, **102**, 15815-15820.
549 49. R. J. Cox, *Nat Prod Rep*, 2023, **40**, 9-27.
550 50. V. Schulz, R. Steinhilper, J. Oltmanns, S. A. Freibert, N. Krapoth, U. Linne, S. Welsch, M. H.
551 Hoock, V. Schunemann, B. J. Murphy and R. Lill, *Nat Commun*, 2024, **15**, 3269.
552 51. A. J. Masud, A. J. Kastaniotis, M. T. Rahman, K. J. Autio and J. K. Hiltunen, *Biochim Biophys*
553 *Acta Mol Cell Res*, 2019, **1866**, 118540.
554 52. B. Yariv, E. Yariv, A. Kessel, G. Masrati, A. B. Chorin, E. Martz, I. Mayrose, T. Pupko and N.
555 Ben-Tal, *Protein Sci*, 2023, **32**, e4582.
556 53. G. Sganzerla Martinez, M. Dutt, A. Kumar and D. J. Kelvin, *Protein J*, 2024, **43**, 711-717.
557 54. S. Kawashima and M. Kanehisa, *Nucleic Acids Res*, 2000, **28**, 374.
558 55. L. P. Kozlowski, *Nucleic Acids Res*, 2021, **49**, W285-W292.
559 56. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*,
560 2015, **1-2**, 19-25.
561 57. J. Huang and A. D. MacKerell, Jr., *J Comput Chem*, 2013, **34**, 2135-2145.
562 58. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and
563 T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605-1612.
564 59. E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris and T.
565 E. Ferrin, *Protein Sci*, 2023, **32**, e4792.

566

567

568

Figures

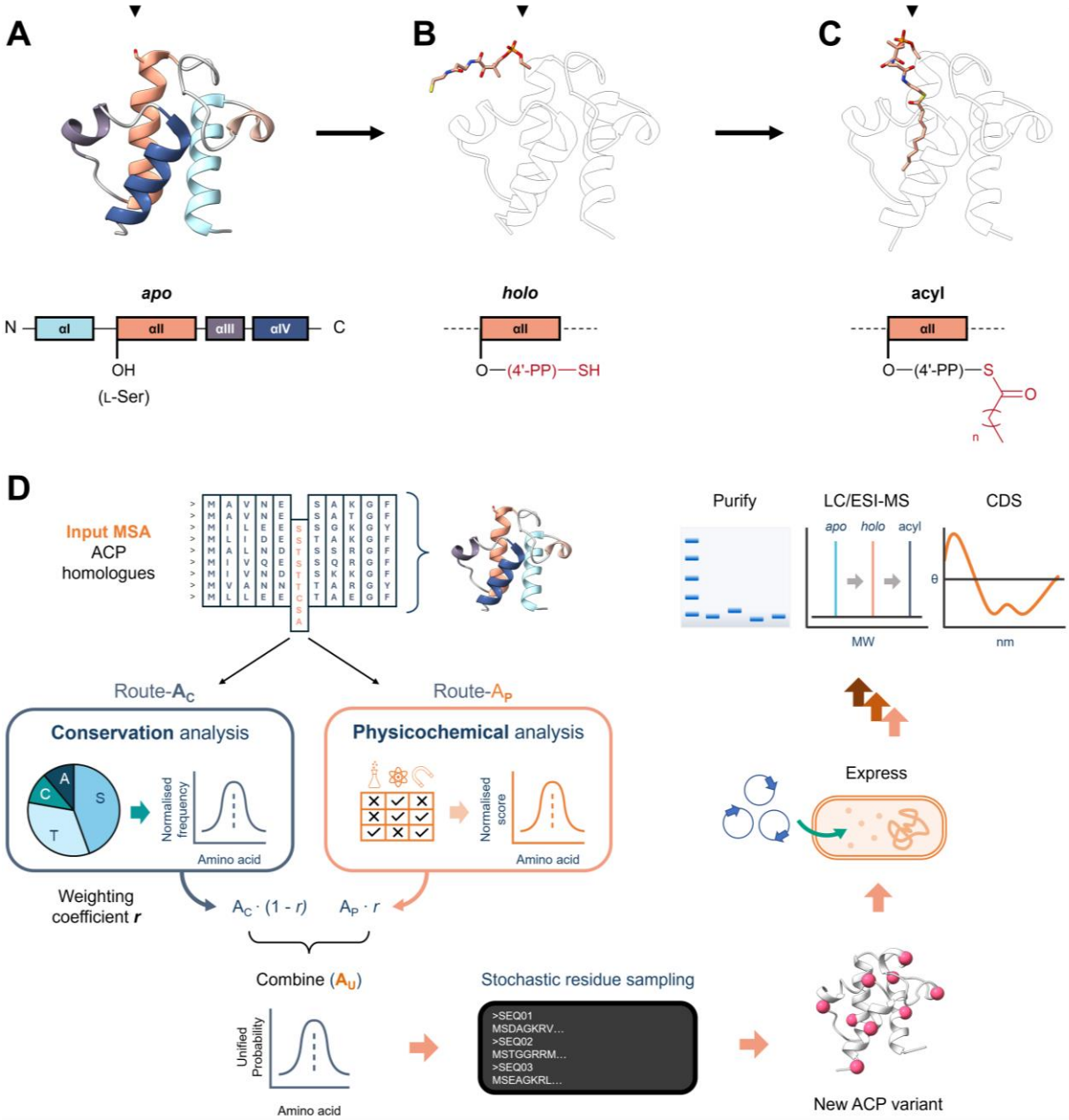

Fig. 1 (A-C) Annotated schematic of an ACP, with colour-coded α-helices, in its *apo* (A), *holo* (B) and acyl (C) state. (PDB: 1T8K, 5H9H and 2FAE). Black triangles indicate the position of the invariant, essential serine. (D) Graphical summary of ALGO-CP and subsequent wet-lab analysis of *de novo* ACP designs, starting with a multiple sequence alignment (MSA) of ACP homologues. Using this MSA as a template, ALGO-CP utilises position-specific conservation and physicochemical analysis to generate new ACP-like sequences. The influence of evolutionary conservation vs. physicochemical analysis can be tuned via a weighting coefficient r, which ranges from 0 to 1. Randomly sampled ACP-like sequences are expressed, purified and examined for PTM from their inactive *apo*- form to functional *holo*- and acyl-forms. Successful candidates are then characterised by circular dichroism (CD) spectroscopy.

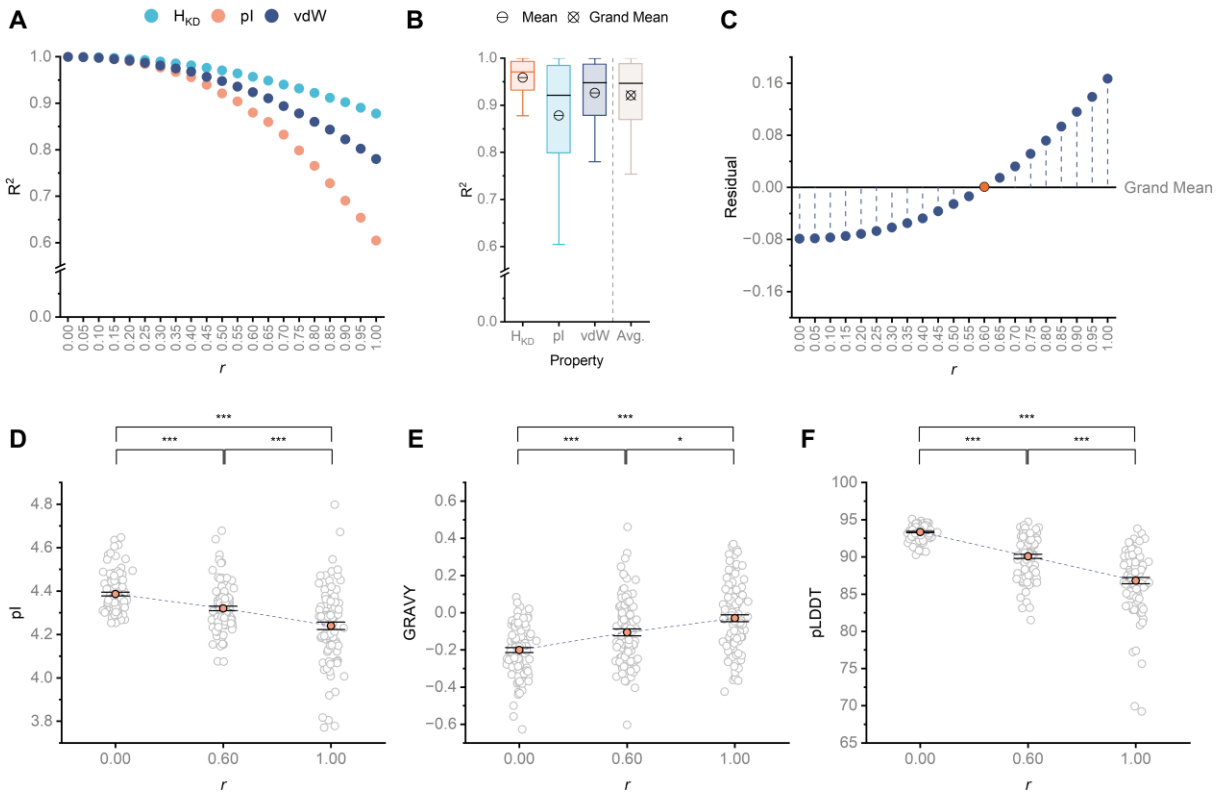

Fig. 2 (A) Positional, pairwise property correlations (R^2) between ALGO-CP sequences and the input MSA of natural AcpP homologues, with varying weighting coefficient r. Lower R^2 indicates greater overall physicochemical deviation from natural AcpP homologues. (B) The spread and central tendency of individual (H_{KD}, pI, vdW volume) and averaged (Avg.) pairwise property correlations across all ALGO-CP sequence sets (r = 0-1, n = 20). Median values are shown as black bands. Boxes represents the interquartile range (IQR). Whiskers extend to data points up to 1.5 x the IQR from the lower and upper quartiles. The crossed circle denotes the grand mean across all sets and properties. (C) Residual analysis of averaged pairwise property correlations relative to the grand mean. Negative residuals reflect sequences with more conservative physicochemical profiles, whereas positive residuals indicate more divergent profiles. Based on its proximity to the grand mean, sequences generated using r = 0.60 (orange circle) were taken to represent the ideal balance between sequence conservation vs. exploration. (D) Mean pIs of ALGO-designed sequences generated using r = 0.00, 0.60 or 1.00. Each datapoint (grey circle) represents a unique sequence. Error bars represent 95% confidence interval (n = 100). (***) = p ≤ 0.001 (Welch's ANOVA with Games-Howell post-hoc test). (E) Mean Grand Average of Hydropathy (GRAVY) values of ALGO-designed sequences generated using r = 0.00, 0.60 or 1.00. Each datapoint (grey circle) represents a unique sequence. Error bars represent 95% confidence interval (n = 100). (*) = p ≤ 0.05; (***) = p ≤ 0.001 (Welch's ANOVA with Games-Howell post-hoc test). (F) Mean Predicted Local Distance Difference Test (pLDDT) of ALGO-designed sequences, folded using AlphaFold3, generated at r = 0.00, 0.60 or 1.00. Each datapoint (grey circle) represents a unique sequence folded by AlphaFold3. Error bars represent 95% confidence interval (n = 100). (***) = p ≤ 0.001 (Welch's ANOVA with Games-Howell post-hoc test).

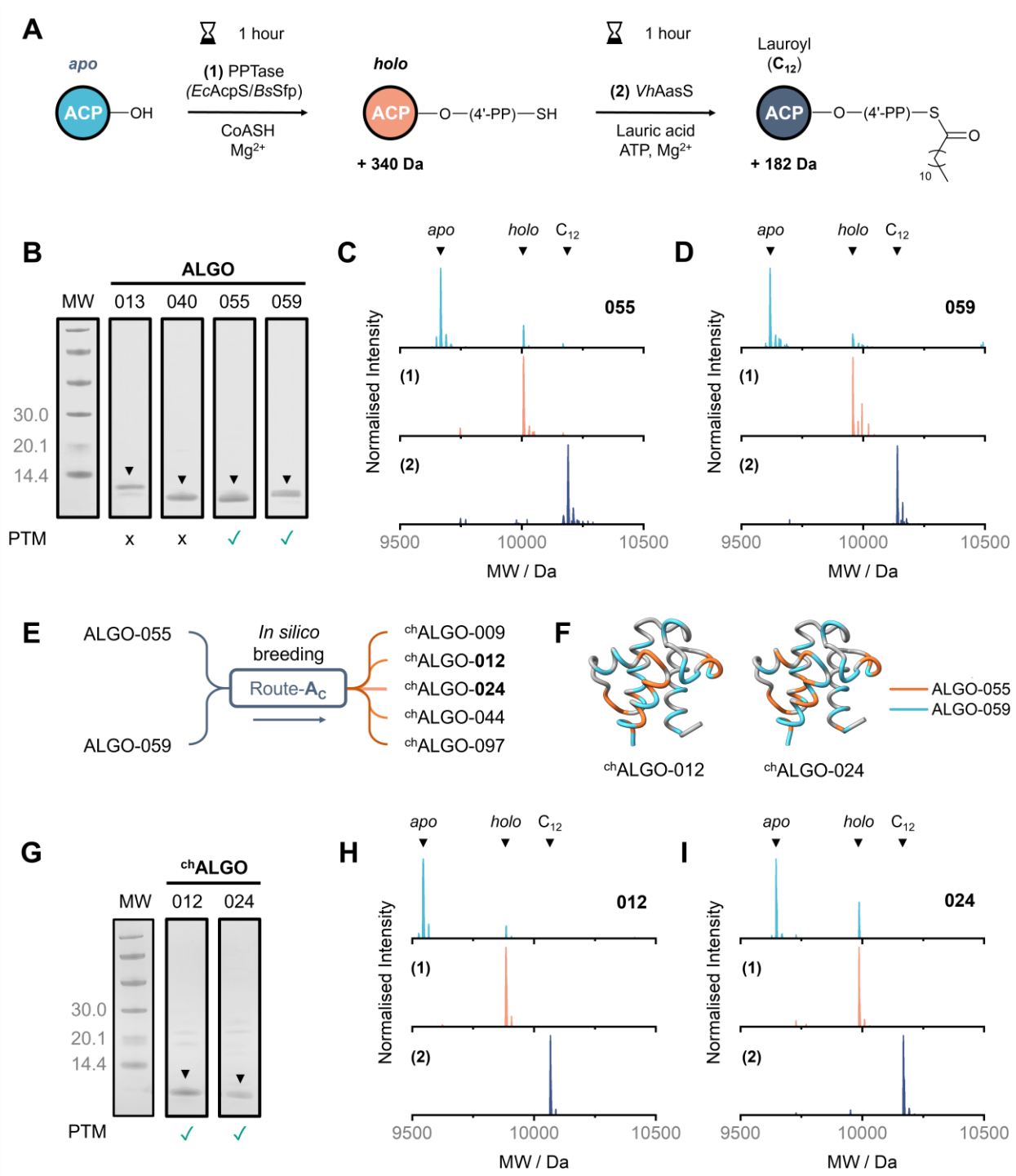

Fig. 3 Expression, purification and PTM of ALGO-CP candidate sequences. (A) General scheme for the one-pot, two-step PTM of ACPs *in vitro*, via the sequential PPTase-catalysed attachment of 4'-PP **(1)** and AasS-catalysed acylation using lauric acid **(2)**. Anticipated MW increases are shown. (B) Annotated SDS-PAGE of purified ALGO-013, 040, 055 and 059. Candidates which undergo PTM are indicated with a green tickmark. An Amersham low molecular weight calibration kit was used as the protein ladder (lane MW). The figure is a composite of two 4-12% Bis-Tris gels that were run, stained and imaged under identical conditions (see also SI) (C) Deconvoluted mass spectra of ALGO-055, tracking PTM from *apo*- (top) to *holo*- (middle) to C_{12}- (bottom) as previously outlined. *Ec*AcpS was used as the 4'-PPTase for reaction **(1)**. (D) Deconvoluted mass spectra of ALGO-059, tracking PTM from *apo*- (top) to *holo*- (middle) to C_{12}- (bottom) as previously outlined. *Ec*AcpS was used as the 4'-PPTase for reaction **(1)**. (E) Generation of randomised

chimeric variants ^{ch}ALGO-009, 012, 024, 044 and 097 from parental candidates ALGO-055 and ALGO-059, by repurposing ALGO-CP route-**A**$_{\mathbf{C}}$. (F) AlphaFold3 models of ^{ch}ALGO-012 (pLDDT = 94.48) and ^{ch}ALGO-024 (pLDDT = 94.27). Positions/regions inherited from ALGO-055 or ALGO-059 are highlighted in orange or blue, respectively. Identical positions/regions are shown in grey. (G) Annotated SDS-PAGE of purified ^{ch}ALGO-012 and ^{ch}ALGO-024. Green tickmarks indicate successful PTM. An Amersham low molecular weight calibration kit was used as the protein ladder (lane MW). See also SI. (H) Deconvoluted mass spectra of ^{ch}ALGO-012, tracking PTM from *apo*- (top) to *holo*- (middle) to C_{12}- (bottom) as previously outlined. *Ec*AcpS was used as the 4'-PPTase for reaction **(1)**. (I) Deconvoluted mass spectra of ^{ch}ALGO-024, tracking PTM from *apo*- (top) to *holo*- (middle) to C_{12}- (bottom) as previously outlined. *Ec*AcpS was used as the 4'-PPTase for reaction **(1)**.

575

576

577

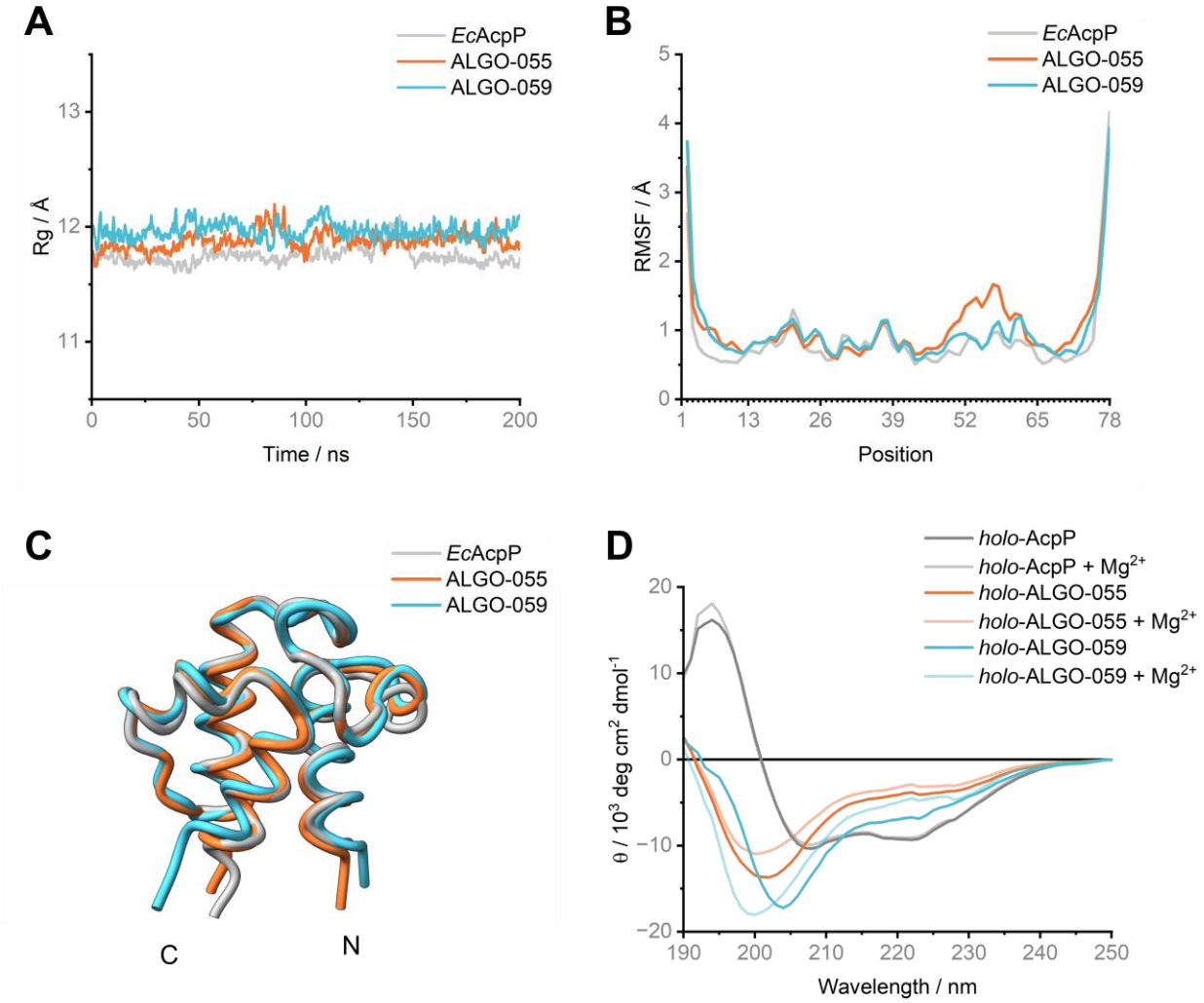

Fig. 4 (A) Gyrational radii (Rg) of ALGO-055 and ALGO-059 AlphaFold3 models during a 200 ns MDS, with *E. coli* AcpP (PDB: 1T8K) included for reference. (B) RMSF profiles of simulated ALGO-055, ALGO-059 and *Ec*AcpP. Higher RMSF values indicate greater Cα mobility. (C) Superimposed frames of ALGO-055, ALGO-059 and *Ec*AcpP, extracted from the endpoint of their respective MDS trajectories. (D) CD spectra of ALGO-055, ALGO-059 and *Ec*AcpP in the presence and absence of 5 mM $MgCl_2$, revealing the minimally-structured nature of the *de novo* ACPs.

578

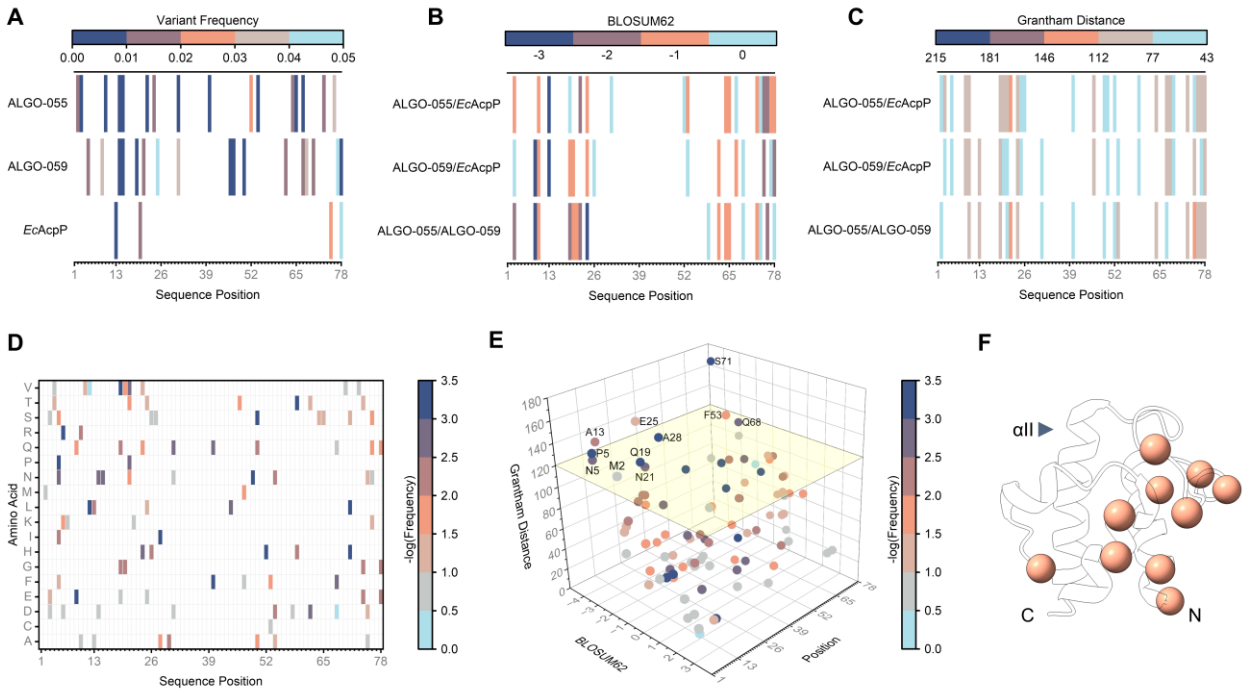

Fig. 5 (A) Distribution of rare amino acid variations (≤5% occurrence) in ALGO-055 and ALGO-059, relative to the input multiple sequence alignment (MSA) of AcpP homologues (n = 2167) used for ALGO-CP. *Ec*AcpP was also studied for comparison. (B) BLOSUM62 analysis of non-conservative amino acid variations identified by sequence alignment. BLOSUM62 scores >0 are not coloured. (B) Grantham distance analysis of physicochemically-divergent amino acid variations observed by sequence alignment. Grantham distances <43 are not coloured. (D) Frequency and distribution of amino acid variations found exclusively in unsuccessful ALGO sequences. Neither tryptophan nor tyrosine were identified in this analysis and are therefore not featured in the heatmap. (E) Grantham distance and BLOSUM62 scores of amino acid variations found exclusively in unsuccessful ALGO sequences, determined by sequence alignment with *Ec*AcpP (see also Table S18). Datapoints are colour coded by -$\log_{10}$(frequency). Non-conservative variants with Grantham distances ≥120 (above yellow plane) are labelled. (F) Positions of non-conservative, physiochemically-divergent variations mapped onto *Ec*AcpP (PDB: 1T8K). The position of helix αII is indicated. N- and C- termini are labelled.