Main Manuscript for

Evolutionary repurposing of a metabolic thiolase complex enables antibiotic biosynthesis

Ge Liao,[1#] Ruolan Sun,[1, 2, 3#] Zilin Shen,[4, 5#] Zhiteng Luo,[4] Cuiping Pang,[6] Zhuanglin Shen,[6] Anfu Wei,[1, 7] Chengneng Mi,[8] Gengfan Wu,[9] Fengfang Li,[1] Yong-Xin Li,[9] Kin Kuan HOI,[1*] Xiaojing Pan,[4*] Xiaoyu Tang[1, 2*]

[1]Institute of Chemical Biology, Shenzhen Bay Laboratory, Shenzhen 518132, China
[2]Shenzhen Medical Academy of Research and Translation, Shenzhen 518107, China
[3]Westlake University, Hangzhou 310030, China.
[4]Institute of Bio-Architecture and Bio-Interactions, Shenzhen Medical Academy of Research and Translation, Shenzhen 518107, China.
[5]Beijing Frontier Research Center for Biological Structure, Tsinghua-Peking Joint Center for Life Sciences, State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China.
[6]Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[7]College of Life Sciences, Northwest University, Xi'an 710069, China
[8]College of Pharmacy, Xiangnan University, Chenzhou 423000, China
[9]Department of Chemistry and The Swire Institute of Marine Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China

[#]These authors contributed equally.

*Xiaoyu Tang
Email: xtang@szbl.ac.cn or xtang@microbechembio.org

*Xiaojing Pan
Email: panxj@smart.org.cn

*Kin Kuan HOI
Email: hughhoi@szbl.ac.cn

Competing Interest Statement: The authors declare no competing interest.

Keywords: thiolase; enzyme complex; Friedel-Crafts acylation; biosynthesis; functional evolution

Abstract

The functional diversification of biosynthetic enzymes underlies the chemical richness of natural products, yet how primary metabolic enzymes evolve to acquire specialized functions in secondary metabolism remains elusive. Here, we report a tripartite enzyme complex from oral *Streptococcus* species—comprising 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase (HMGS), acetyl-CoA acetyltransferase (ACAT), and a DUF35 protein—that catalyzes an unusual Friedel–Crafts *C*-acetylation on a pyrrolidine-2,4-dione scaffold, completing the biosynthesis of the antibiotic reutericyclin A. Cryo-electron microscopy of the *S. macacae*-derived thiolase complex (*Sma*ATase) reveals a conserved architecture resembling the archaeal HMGS/ACAT/DUF35 complex involved in the mevalonate pathway, yet with key catalytic residues rewired to enable novel substrate specificity. Biochemical characterization, molecular modeling, and evolutionary analysis confirmed that the ancestral activity of HMG-CoA synthesis has been lost, while the complex has been repurposed to mediate Friedel–Crafts *C*-acylation of small molecule acceptors. These findings reveal a rare example of thiolase complex neofunctionalization, shedding light on an underexplored trajectory in enzyme evolution and offering a template for engineering C–C bond-forming catalysts in synthetic biology.

Main text

Introduction

Natural products, also known as secondary or specialized metabolites, confer selective advantages to their producing organisms by mediating ecological interactions such as competition, defense, and communication[1]. In contrast to the relatively conserved structures of primary metabolites, natural products display remarkable chemical diversity, largely driven by the evolutionary plasticity of their biosynthetic enzymes[2]. This diversity often arises through the recruitment and functional repurposing of primary metabolic enzymes into specialized metabolic pathways[2, 3]. Beyond the well-known fatty acid synthase–polyketide synthase (FAS–PKS) paradigm[4, 5], notable examples include the evolution of chalcone isomerase from a noncatalytic fatty acid-binding protein[6, 7], and the emergence of β-lactam synthetases from asparagine synthase[8, 9] (Supplementary Fig. 1). While the catalytic repertoire of biosynthetic enzymes continues to expand, the evolutionary trajectories by which primary metabolic enzymes—especially multimeric enzyme complexes—acquire new functions in secondary metabolism remain poorly understood.

The thiolase superfamily comprises highly conserved enzymes present across all domains of life and plays essential roles in central metabolic pathways[10, 11]. For instance, acetyl-CoA acetyltransferase (ACAT) catalyzes the formation of acetoacetyl-CoA, a key intermediate in cholesterol and steroid hormone biosynthesis, while its reverse reaction is critical for fatty acid β-oxidation[12, 13]. In archaea, a tripartite enzyme complex composed of 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase (HMGS), ACAT, and a DUF35 protein (Pfam family PF01796, Domain of Unknown Function) enables mevalonate biosynthesis by coupling a thermodynamically unfavorable thiolase reaction with an favorable HMG-CoA condensation, thereby facilitating intermediate channeling (Fig. 1a)[14]. In previous work, we identified a *HMGS* homolog *mucA*, a *ACAT* homolog *mucB*, and a *DUF35* homolog *mucC* within the *muc* biosynthetic gene cluster (BGC) in *Streptococcus mutans* (Fig. 1a), which directs the production of the antimicrobial compound reutericyclin A (RTC-A) and the immunomodulatory molecule mutanocyclin[15-17]. The *muc* BGC is widely distributed among Firmicutes, particularly in

Streptococcus species[16]. Based on sequence homology and biosynthetic logic, MucABC was hypothesized to catalyze a *C*-acetylation reaction essential for RTC-A biosynthesis[15]. Its divergence from the canonical role in mevalonate biosynthesis indicates a previously unrecognized evolutionary trajectory within the thiolase superfamily.

Here, we report functional and structural characterization of the MucABC enzyme complex from oral *Streptococcus*, exemplified by the *S. macaca*-derived thiolase complex. We show that this enzyme complex catalyzes a non-canonical Friedel–Crafts *C*-acetylation on a pyrrolidine-2,4-dione scaffold during RTC-A biosynthesis. Cryo-electron microscopy (cryo-EM) and biochemical analyses reveal that the neofunctionalization of this complex involves the loss of conserved catalytic motifs—such as the E–C–H–N tetrad in HMGS and the HDxF motif in ACAT—along with active-site remodeling for specialized substrate recognition. Furthermore, we demonstrate that this acetylation step is indispensable for the antimicrobial activity of RTC-A, providing *Streptococcus* a competitive advantage in its native niche. Together, our findings reveal a previously unrecognized evolutionary trajectory in which multimeric enzyme complexes from primary metabolism are repurposed for secondary metabolic functions, offering new opportunities for engineering modular biocatalysts in synthetic biology.

Results

Characterization of Streptococcal acyltransferases (ATases) as non-canonical Friedel-Crafts acylases

A BLASTP search using the MucABC sequence from *S. mutans* as the query showed 31–37% sequence identity with PhlACB from *Pseudomonas protegens* (known as *Pp*ATase), a rare Friedel-Crafts acylase known to mediates the reversible disproportionation of phloroglucinol (PG), monoacetylphloroglucinol (MAPG), and 2,4-diacetylphloroglucinol (DAPG; Supplementary Fig. 2a)[18, 19]. Although their overall sequence similarity is low, the individual components of MucABC retain conserved domains found in both *Pp*ATase and the archaeal HMGS/ACAT/DUF35 complex (Supplementary Fig. 3), suggesting a potential conservation of thiolase-related chemistry. To investigate the catalytic activity of *mucABC*-encoded proteins, we selected three highly similar *mucABC* operons (89–100% identity) from host-adapted oral

streptococci: *S. mutans* B30 (human), *S. macacae* NCTC11558 (monkey), and *S. troglodytae* TKU31 (chimpanzee; Supplementary Fig. 2a). Given that *Pp*ATase is functional only when the entire *phlACB* operon is co-expressed[18-20], we constructed a plasmid expressing the full *mucABC* operon with a C-terminal His_6-tag on MucB to facilitate purification (Supplementary Fig. 2b). Of the three variants, *S. macacae mucABC* exhibited superior expression in *Escherichia coli*, yielding the highest protein amounts and purity after affinity chromatography (Supplementary Fig. 2c), and was therefore chosen for biochemical characterization.

To perform *in vitro* enzymatic assays, we chemically synthesized preRTC-A (Supplementary Fig. 4), the presumed nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS)-derived biosynthetic precursor of RTC-A[16]. Four unnatural acetyl donors previously shown to be accepted by *Pp*ATase—*N*-acetylimidazole (*N*-AcIm), isopropenyl acetate (IPEA), ethyl thioacetate (EtSAc), and phenyl acetate (PA)—were tested with MucABC (Fig. 1b). Ultra-performance liquid chromatography (UPLC) analysis confirmed the formation of a product peak with identical retention time for all four acetyl donors (Fig. 1b). High-resolution mass spectrometry (HRMS) analysis of the enzymatic reaction using PA revealed a 42 Da shift from the substrate ion $[M - H]^-$ at *m*/*z* 306.2087 (Cald. 306.2075) to the product ion at *m*/*z* 348. 2187 (Cald. 348.2180), consistent with an acetylation event (Fig. 1b). Scaled-up enzymatic reactions enabled product isolation and confirmation of its identity as RTC-A by ^1H and ^{13}C NMR spectroscopy (Supplementary Table 1 and Supplementary Fig. 5), with the acetyl group installed at the *C*-3 position of the pyrrolidine-2,4-dione core (Fig. 1b). No *O*-acetylated byproducts were detected in time-course analyses (Supplementary Fig. 6a). Since the chemically synthesized preRTC-A was obtained as a racemic mixture (due to the acidity of the α-proton), we further isolated the individual (*R*)- and (*S*)-preRTC-A via chiral separation. UPLC analysis confirmed that MucABC could utilize both enantiomer as substrate, as RTC-A was produced in both incubations with the same retention times compared to the standard (Supplementary Fig. 6b and 6c). Additionally, MucABC-mediated reverse conversion of RTC-A back to preRTC-A was observed, indicating that the C–C bond formed by MucABC is cleavable by the enzyme, thereby enabling RTC-A to serve as an acetyl donor in a

transacetylation reaction (Supplementary Fig. 7). Collectively, these findings establish that the MucABC from *S. macacae* functions as an acyltransferase (hereafter termed *Sma*ATase) that catalyzes C3-acetylation of preRTC-A, adding a new member of acyltransferases capable of mediating non-canonical Friedel-Crafts acylation.

Acetyl-CoA serves as the potential physiological donor for *Sma*ATase and *Pp*ATase

Although various phenyl esters and thioesters have been employed as acetyl donors in *in vitro* assays of *Pp*ATase[19, 21, 22], its physiological acetyl donor has remained unknown. To identify a biologically relevant donor for *Sma*ATase, we evaluated several endogenous acyl donors, including acetyl-CoA, malonyl-CoA, acetoacetyl-CoA, and acetyl-phosphate. Intriguingly, RTC-A formation was clearly observed in the presence of acetyl-CoA, whereas no signal was observed with acetyl-P and trace signals were detected with either acetoacetyl-CoA or malonyl-CoA (Fig. 1b). To determine whether this feature extends to *Pp*ATase, we expressed and purified PhlACB using the same strategy as employed for *Sma*ATase (Supplementary Fig. 8a and 8b). Consistent with the results for *Sma*ATase, reactions containing *Pp*ATase and PG yielded both MAPG and DAPG when acetyl-CoA was supplied as the donor, with only trace-level signals were detected when acetoacetyl-CoA or malonyl-CoA were used as alternative donors (Supplementary Fig. 8c). The identity of DAPG was confirmed by co-elution with an authentic standard (Supplementary Fig. 8c), while MAPG was verified by HRMS (Supplementary Fig. 8d) and NMR spectroscopy (Supplementary Table 2 and Supplementary Fig. 9). Together, these results support acetyl-CoA as the potential physiological acetyl donor for both *Sma*ATase and *Pp*ATase, aligning their function with that of canonical acetyltransferases[23, 24].

***Sma*ATase exhibits oligomeric heterogeneity and requires tripartite assembly for catalytic activity**

SDS-PAGE analysis confirmed successful co-purification of all three components of the *Sma*ATase complex purified via affinity chromatography, despite only MucB carrying a C-terminal His_6-tag (Supplementary Fig. 2c). Mass spectrometry further validated the presence of MucA, MucB, and MucC units in the enzyme solution, with experimentally determined masses matching theoretical predictions (Supplementary Fig. 10a). Size-exclusion chromatography indicated

enzyme complex formation by showing a peak corresponding to approximately 400–600 kDa (Supplementary Fig. 10b). To determine the structural basis of *Sma*ATase function, we resolved its cryo-EM structure at 2.32 Å resolution, revealing a dodecameric architecture composed of four MucA, four MucB, and four MucC units (Fig. 2a, Supplementary Table 3, Supplementary Fig. 11–13). This organization closely resembles the reported crystal structures of *Pp*ATase[20] (Supplementary Fig. 2a) and the archaeal HMGS/ACAT/DUF35 complex[14] (Fig. 1a). In addition to the dodecamer, we observed an alternative octameric assembly (2.16 Å resolution), consisting of a $MucA_4C_2$ core and two MucB units (Fig. 2b, Supplementary Table 3, Supplementary Fig. 11–13). Structural superposition of the dodecameric and octameric forms revealed a root-mean-square deviations (RMSD) of 0.160 Å over 2254 superposed Cα atoms, with active sites displaying comparable configuration (Supplementary Fig. 14). Furthermore, native mass spectrometry revealed additional oligomeric species (Fig. 2c), including 11-mer ($MucA_4B_3C_4$), 10-mer ($MucA_4B_2C_4$), 9-mer ($MucA_4B_2C_3$), and 8-mer ($MucA_4BC_3$). These findings were corroborated by native gel electrophoresis (Supplementary Fig. 15a), indicating that *Sma*ATase exists as a heterogeneous mixture of assemblies under the *in vitro* conditions tested.

Similar heterogeneity was observed for *Pp*ATase, which was previously characterized as a 12-mer [$Phl(A_2C_2)_2B_4$] by crystallography[20]. Additional 10-mer ($PhlA_4C_2B_4$) and 8-mer ($PhlA_4B_4$) were detected by native mass spectrometry (Supplementary Fig. 15b), consistent with the three distinct bands observed by native PAGE (Supplementary Fig. 15a). Quantitative mass photometry analysis revealed their relative distribution as ~4:4:2 (8-mer:10-mer:12-mer, Supplementary Fig. 15c).

To assess the requirement for complex formation in catalytic activity, we co-expressed pairwise combinations of MucA, MucB, and MucC in *E. coli*. Co-expression of MucAB or MucBC yielded poorly soluble proteins (Supplementary Fig. 16a and 16b), and purified MucAC lacked detectable *C*-acetylation activity (Supplementary Fig. 16c and 16d). Given that previous studies implicated residues within the PhlC (ACAT) subunit are essential for catalysis in *Pp*ATase[20, 25], we evaluated potential of MucB (the *Sma*ATase ortholog of PhlC) to function independently. However, the individual

expression of MucB in *E. coli* resulted in insoluble aggregates (Supplementary Fig. 17a). Although fusion to an N-terminal trigger factor (TF) improved solubility and enabled purification (Supplementary Fig. 17a), neither TF-tagged nor cleaved MucB displayed detectable *C*-acetyltransferase activity (Supplementary Fig. 17b and 17c). Surface plasmon resonance (SPR) showed that both the full *Sma*ATase complex ($K_D = 2.52 \times 10^{-6}$ M) and the isolated MucB subunit ($K_D = 6.04 \times 10^{-6}$ M) are capable of binding the substrate, with the intact complex exhibiting slightly higher affinity (Supplementary Fig. 18a). However, because the catalytic pocket of MucB is situated at the interface with MucA and MucC (Supplementary Fig. 18b), and given that individual subunits of the related archaeal HMGS/ACAT/DUF35 complex are either insoluble (ACAT) or inactive (HMGS) when expressed alone[14], we hypothesize that MucA and MucC help constrain MucB into a catalytically competent conformation.

Structural comparison of *Sma*ATase, *Pp*ATase, and archaeal HMGS/ACAT/DUF35 complex

Despite low sequence identities and distinct catalytic functions, *Sma*ATase shares striking structural similarity with both *Pp*ATase and the archaeal HMGS/ACAT/DUF35 complex. This structural conservation is evident at multiple levels: (1) **Subunit organization and overall architecture**: In addition to their shared dodecameric architecture, the spatial arrangement of subunits in *Sma*ATase closely resembles that observed in *Pp*ATase and the archaeal complex (Fig. 2d). Structural alignment of the HMGS-ACAT-DUF35 modules between *Sma*ATase and either *Pp*ATase or the archaeal homolog yields RMSD of 1.245 Å (585 Cα atoms) and 1.454 Å (363 Cα atoms), respectively, underscoring the high degree of structural conservation across these systems (Supplementary Fig. 19); (2) **DUF35 subunit architecture and positioning**: MucC, a family member of DUF35 protein, features an N-terminal rubredoxin-like zinc ribbon (Supplementary Fig. 20) and a C-terminal oligonucleotide/oligosaccharide-binding (OB) fold domain. Within *Sma*ATase, MucC is embedded in the multimeric assembly and establishes multiple contacts with neighboring MucA and MucB subunits (Supplementary Fig. 21). This arrangement resembles that of PhlB in *Pp*ATase, which bridges PhlA and PhlC, as well as the DUF35 homolog in the archaeal complex that connects the HMGS and ACAT units (Fig. 2e); (3) **Thiolase fold conservation and conserved**

CoA-binding chamber: Both MucA and MucB adopt a conserved five-layered αβαβα core architecture characteristic of thiolase superfamily enzymes (Supplementary Fig. 22)[11, 26]. This structural signature is also preserved in PhlA and PhlC of *Pp*ATase and the HMGS and ACAT units of the archaeal complex (Supplementary Fig. 22). Additionally, a hallmark of the archaeal HMGS/ACAT/DUF35 complex is the presence of a fused CoA-binding site at the HMGS–ACAT interface, which facilitates shuttling of the pantetheinyl arm between active sites during catalysis (Fig. 2f). Notably, a similar interfacial chamber is retained in both *Sma*ATase and *Pp*ATase (Fig. 2f, supplementary Fig. 23), despite the absence of the HDxF motif in the thiolase domains and the key Claisen condensation residues in the HMGS units (details will be discussed in the following section).

Together, these structural observations indicate that the HMGS/ACAT/DUF35 complex represents a deeply conserved architectural module that may have been evolutionarily repurposed from primary metabolism to secondary metabolism.

Key catalytic residue mutations support functional divergence

Given the striking structural conservation of archaeal and bacterial HMGS/ACAT/DUF35 complexes as acetyl-CoA-utilizing *C*-acetyltransferases, we conducted a phylogenetic analysis comparing their ACAT and HMGS units with representative members of the thiolase superfamily to elucidate the evolutionary origins of *Sma*ATase and *Pp*ATase. As expected, enzymes are clustered according to their functions (Fig. 3a). MucA/PhlA and their homologs formed a clade closely related to archaeal HMGSs, while MucB/PhlC and their homologs grouped near archaeal ACATs (Fig. 3a), supporting their evolutionary relationship with the archaeal HMGS/ACAT/DUF35 complex involved in mevalonate biosynthesis. Beyond the αβαβα structural motif, a hallmark of the thiolase superfamily is a catalytic cysteine (Cys) residue that mediates CoA or acyl carrier protein (ACP) substrate attachment[27]. However, multiple sequence alignment revealed the absence of the typical E-C-H-N tetrad conserved among HMGSs in MucA/PhlA and related sequences (Fig. 3b). Similarly, although MucB/PhlC and their homologs retain the CxS and GHx motifs characteristic of ACATs, they lack the

HDxF motif present in archaeal ACATs (Fig. 3b). These sequence deviations likely contribute to the functional divergence of bacterial HMGS/ACAT/DUF35 complexes from their archaeal counterparts.

To test the functional significance of these divergent residues, we first assessed whether *Sma*ATase and *Pp*ATase could catalyze the Claisen condensation of acetyl-CoA to form HMG-CoA. LC-MS analysis revealed the formation of CoA, but not HMG-CoA, from incubations with either enzyme (Fig. 3c), indicating that both utilize acetyl-CoA solely as a donor. To further validate this, we prepared the archaeal HMGS/ACAT/DUF35 complex using the same purification protocols and generated two variants: one carrying an H284A substitution in the ACAT unit's HDxF motif, and another with substitutions in the HMGS unit's E-C-H-N catalytic tetrad entirely (Supplementary Fig. 24). While the wild-type complex produced HMG-CoA from acetyl-CoA, neither mutant yielded detectable product (Fig. 3c), confirming the essentiality of these residues for canonical HMGS activity.

To gain deeper insights into the sequence diversity of *Sma*ATase-like enzyme complexes, we mined bacterial genomes in the NCBI database and identified 1,030 BGCs containing co-localized *HMGS*, *ACAT*, and *DUF35* genes (Fig. 4a). The vast majority of the identified BGCs (1,001) lacked adjacent biosynthetic machinery for canonical secondary metabolism, only 29 BGCs were predicted to be associated with secondary metabolism (Fig. 4b). These include *muc*-like clusters from *Streptococcus* species, featuring hybrid NRPS-PKS assembly lines, and *phl*-like clusters from *Pseudomonas* species harboring type III PKSs (Fig. 4b).

Given the dominance of this uncharacterized category, and its lack of obvious biosynthetic partners, we next asked whether these minimal HMGS–ACAT–DUF35 loci might encode functionally distinct or evolutionarily intermediate enzyme complexes. To explore this possibility, we selected a representative candidate from *Pseudomonas fluorescens* for further analysis. This *Pf*ACAT/HMGS-DUF35 complex features a fused HMGS-DUF35 protein and, importantly, lacks any neighboring *PKS* or *NRPS* genes (Fig. 4b, Supplementary Fig. 25 and Fig. 26). Sequence alignment showed that its HMGS domain lacks the E-C-H-N tetrad, whereas its ACAT unit retains the catalytic Cys and conserved HDxF and GHx

motifs (Fig. 3b), suggesting that the HMGS unit is inactive whereas the ACAT component remains functionally intact. *In vitro* biochemical assays confirmed this functional divergence: the *Pf*ACAT/HMGS-DUF35 complex did not produce HMG-CoA or acetoacetyl-CoA from acetyl-CoA, resembling *Sma*ATase and *Pp*ATase (Fig. 3c). In contrast, it catalyzed the thiolytic cleavage of acetoacetyl-CoA to generate acetyl-CoA, a reaction not observed for *Sma*ATase and *Pp*ATase (Fig. 4c). Introducing an H294Y mutation in *Pf*ACAT nearly abolished this activity (Fig. 4c and Supplementary Fig. 25), confirming the essential catalytic role of the His residue within the HDxF motif. This finding is consistent with observations in *Geobacter metallireducens* BbsAB, a DUF35-associated sterol-carrier-protein (SCP2)-type thiolase where the HDxF motif mediates reversible benzylsuccinyl-CoA formation from benzoyl-CoA and succinyl-CoA[28].

Taken together, these findings suggest that the *Pf*ACAT/HMGS-DUF35 complex may represent an evolutionary intermediate that has lost its primary metabolic function but has not yet been recruited into a defined specialized pathway. In contrast, *Sma*ATase and *Pp*ATase have undergone further evolution, characterized by the loss of the HDxF motif and the acquisition of new active-site architectures within the ACAT domain that enable specific recognition of small-molecule acetyl acceptors (Fig. 5). To elucidate the substrate-binding mode of *Sma*ATase, we performed molecular docking and dynamics simulations (Fig.5 and Supplementary Fig. 27). A comparison with the well-established catalytic mechanism of *Pp*ATase[20, 25] reveals that key catalytic residues are largely conserved: the catalytic nucleophile (Cys88 in PhlC vs Cys87 in MucB); the general base (Glu116 vs Asp115); the proton source residue and proton shuttling residues (Asp137-Thr132-His144 vs Asp136-Thr131-His143, Fig. 5 and Supplementary Fig. 28). Notably, the catalytic cysteine is acetylated in both the *Pp*ATase[20] (Cys88, PDB: 5MG5) and *Sma*ATase (Cys87) structures; this was evidenced by residual electron density in both cases and confirmed for *Sma*ATase by LC-MS/MS analysis (Supplementary Fig. 29).

Despite the conserved catalytic core, clear differences are observed in the substrate-binding region. For example, His56 and Tyr124 in PhlC are replaced by Ala55 and Phe123 in MucB, respectively, and the flexible lid residue Trp211 in PhlC[29] is substituted by Ala210 in MucB (Fig. 5 and Supplementary Fig. 28). Sequence alignment further shows that key

catalytic and substrate-binding residues are well conserved within PhlC or MucB homologs, yet markedly divergent from those in archaeal ACATs (Supplementary Fig. 28). Together, these observations indicate that although *Sma*ATase and *Pp*ATase retain a shared catalytic apparatus inherited from their ancestral thiolase scaffolds, they have undergone functional repurposing through reconfiguration of their substrate-binding sites, enabling them to recognize chemically distinct acetyl acceptors.

Acetylation is essential for the antimicrobial activity of RTC-A and DAPG

RTC-A is a tetramic acid-derived antibiotic with broad-spectrum activity against gram-positive bacteria[15, 30, 31]. To evaluate the biological relevance of the acetylation modification catalyzed by streptococcal ATases, we compared the antimicrobial activities of RTC-A and its unacetylated precursor, preRTC-A, against both oral streptococcal competitors and clinically relevant pathogens. Strikingly, RTC-A exhibited potent antibacterial effects against all tested strains, whereas preRTC-A showed negligible inhibition (Fig. 6 and Supplementary Fig. 30). A similar structure–activity relationship was observed for the DAPG/PG pair, with DAPG displayed markedly stronger antimicrobial activity than PG against most strains (with the exception of *S. gordonii* and S. *oralis*) (Fig. 6 and Supplementary Fig. 30). Together, these results establish acetylation as a critical modification for the antimicrobial potency of both RTC-A and DAPG, underscoring the ecological significance of evolved ATases from *Streptococcus* and *Pseudomonas* species in interspecies competition.

Discussion

Friedel-Crafts acylation represents one of the most powerful strategies in organic chemistry for constructing C-C bonds[32]; however, enzymatic catalysts capable of mediating this reaction remain exceedingly rare. Here, we report the functional and structural characterization of a streptococcal ATase as a non-canonical Friedel-Crafts acylase that, unlike the previously characterized *Pp*ATase, utilizes a pyrrolidine-2,4-dione core as the acetyl acceptor. Notably, a recently identified Friedel-Crafts acylase, SmzB from *Streptomyces* sp., was shown to transfer an acyl chain from an ACP-activated donor onto 5-aminoimidazole ribonucleotide[33]. In contrast, *Sma*ATase exhibits remarkable promiscuity toward acetyl

donors, catalyzing Friedel–Crafts acylation with not only the canonical donor acetyl-CoA but also with unnatural donors such as phenyl acetate.

Beyond their catalytic novelty, *Sma*ATase and *Pp*ATase assemble into atypical multienzyme complexes—an architecture rarely observed among tailoring enzymes in secondary metabolism. While such multimeric assemblies are widespread in primary metabolism, as exemplified by the pyruvate dehydrogenase complex (PDH)[34, 35], the 2-oxoglutarate dehydrogenase complex (ODH)[36], and the fatty acid synthase (FAS) complex[37], enzyme complexes in secondary metabolism are more commonly associated with the primary assembly of carbon skeletons (such as PKS[38] and NRPS[39]) than with the post-assembly tailoring reactions (e.g., methylation, prenylation, hydroxylation). The tripartite organization of *Sma*ATase and *Pp*ATase, therefore, represents a rare instance of complex formation in tailoring enzymes. Multienzyme assemblies often facilitate substrate channeling to improve catalytic efficiency and metabolic flux[40, 41]. For instance, the archaeal HMGS/ACAT/DUF35 complex possesses a shared CoA-binding pocket that enables direct intermediate transfer between ACAT and HMGS[14]. However, *Sma*ATase and *Pp*ATase mediate single reactions and do not require substrate channeling. While the catalytic pocket resides within MucB, our data show that MucB alone is insufficient for catalysis, implying that complex formation may represent an evolutionary vestige, with functional acquisition occurring subsequently.

Although archaeal and bacterial HMGS/ACAT/DUF35 complexes share structural conservation and phylogenetic relatedness, they have diverged functionally. In archaea, the tripartite complex catalyzes sequentially Claisen condensations to support the mevalonate pathway. In contrast, the bacterial homologs *Sma*ATase and *Pp*ATase perform Friedel–Crafts acylation. Both systems retain a common *C*-acetyltransferase activity using acetyl-CoA as the donor but diverge in their acceptor substrates—CoA-derived intermediates in archaea *versus* specialized small molecules in bacteria (Supplementary Fig. 31). This divergence is presumably underpinned by active-site remodeling: loss of the E-C-H-N tetrad in HMGS, absence of the HDxF motif in ACAT, and acquisition of new residues to accommodate noncanonical substrates (Fig. 5). We propose that *Sma*ATase and *Pp*ATase exemplify a functional repurposing of thiolase complexes in secondary

296 metabolism, whereby bacteria innovated new enzymatic functions within conserved structural frameworks through
297 minimal but impactful mutations.

298 The functional divergence of HMGS/ACAT/DUF35 complexes likely reflects distinct metabolic priorities across
299 domains of life. In archaea, these complexes are near-universally conserved due to the strict requirement for isoprenoids
300 as membrane lipids[14]. In contrast, many bacteria rely on the alternative methylerythritol phosphate (MEP) pathway for
301 isoprenoid biosynthesis, reducing their dependence on the mevalonate pathway and freeing HMGS/ACAT/DUF35
302 complexes to acquire new functions. Repurposing existing scaffolds via minimal structural changes represents a cost-
303 effective evolutionary strategy for enzyme specialization[42]. In this study, we show that acetylation is essential for the
304 bioactivity of two structurally distinct antibiotics—RTC-A from oral streptococci and DAPG from plant-associated
305 *Pseudomonas* species. In both organisms, acetylation can largely enhance antimicrobial potency and likely confers a
306 competitive ecological advantage: RTC-A facilitates biofilm dominance in the oral cavity[15, 16], while DAPG suppresses
307 phytopathogens in the rhizosphere[43, 44]. The evolution of dedicated ATases in these bacteria (*Sma*ATase and *Pp*ATase)
308 exemplifies how subtle active-site reconfigurations can endow existing enzymes with new ecological functions.

309 In conclusion, we have elucidated the function and structure of *Sma*ATase and traced its evolutionary origins from a
310 thiolase-based multienzyme complex. Our findings provide a compelling example of how primary metabolic enzyme
311 architectures can be retooled for specialized functions in natural product biosynthesis. Given the abundance of multimeric
312 complexes in primary metabolism, such repurposing may represent a widespread evolutionary strategy. Insights from this
313 work may inspire new avenues in enzyme engineering and synthetic biology, enabling the design of tailored catalysts for
314 biocatalysis and metabolic engineering.

Methods

General methods

All reagents were purchased from Sigma-Aldrich, Thermo Fisher Scientific, Bidepharm, and Aladin. Gene synthesis, plasmid construction, and point mutation were carried out by Tsingke Biotechnology Company (Supplementary Table 4). Luria-Bertani (LB, 1% NaCl, 1% tryptone, and 0.5% yeast extract) medium was used for culturing *E. coli*. Chloramphenicol (25 μg/mL), ampicillin (50 μg/mL), and kanamycin (50 μg/mL) were used for selection of recombinant *E. coli* strains. Antibacterial activity assays were performed using indicator strains cultured in Mueller-Hinton broth (MHB, OXOID), with the exception of *S. oralis* and *S. gordonii*, which were maintained in brain-heart infusion (BHI, OXOID) broth due to their poor growth in MHB.

Protein expression and purification

Codon-optimized genes encoding *Sma*ATase (*mucABC*), *Pp*ATase (*phlACB*), archaeal HMGS/ACAT/DUF35 complex, *pf*ACAT/HMGS-DUF35 complex, and individual combinations of MucABC subunits were cloned into the pACYCDuet-1 and transformed into *E. coli* BL21(DE3) for protein expression. Codon-optimized *mucB* gene was cloned into both pET28a(+) and pCold-TF (TF=trigger factor) for enhanced soluble expression. Cells harboring the expression plasmid were grown overnight in LB supplemented with appropriate antibiotic at 37 °C and 220 r.p.m. to prepare the preculture. Five milliliters of preculture were then inoculated into 500 mL of LB medium containing the same antibiotic and cultured under identical conditions. Protein expression was induced with 0.2 mM isopropyl β-D-thiogalactopyranoside (IPTG) at OD_{600} ~0.6, followed by incubation at 30 °C (21 h, 140 rpm). For the expression of MucB, the culture was incubated at 16 °C, 160 rpm for 16 h after adding 0.2 mM IPTG. Cells were harvested and resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 10 mM imidazole, 10% glycerol, pH 8.0). The suspension was disrupted by ultrasonication (40% amplitude, 10 min, pulse 3 sec, pause 6 sec) and then centrifuged at 13000×g for 30 min to spin down the cell debris. Recombinant proteins were purified using an ÄKTA go system (Cytiva) equipped with HisTrap HP columns (Cytiva). Eluted proteins were buffer-exchanged into 50 mM Kpi buffer (pH 7.5) containing 10% glycerol using PD-10 desalting columns (Cytiva). Protein concentrations were determined after concentration with Amicon Ultra filters (Merck Millipore). Size-exclusion chromatography was performed using an ÄKTA go system equipped with a Superdex 200 Increase 10/300 GL column (Cytiva) pre-equilibrated with 50 mM Kpi buffer (pH 7.5) containing 150 mM NaCl. Purity of proteins was assessed by SDS-PAGE analysis (Supplementary Fig. 32) and imaged using the GelDoc Go system (Bio-Rad). Freshly purified enzyme complexes were used for all *in vitro* assays to ensure activity.

Activity assays

Enzymatic assays for functional characterization of *Sma*ATase were carried out in 100 μL reaction mixtures containing 50 mM Kpi buffer (pH 7.5), 12 μM purified enzyme, 1 mM preRTC-A, and various acetyl donors at the following concentrations: 1.5 mM PA, 25 mM *N*-AcIm, 10 mM IPEA (supplemented with 10 mM imidazole), 10 mM EtSAc, 1.5 mM acetyl-CoA, 1.5 mM malonyl-CoA, 1.5 mM acetoacetyl-CoA, 1.5 mM acetyl-P. Reactions were incubated at 30 °C for 16 h and quenched by the addition of 100 μL methanol. After centrifugation at 21,000×g for 30 min, supernatants were analyzed by ultra-performance liquid chromatography (UPLC) using LC method 1 (Supplementary Table 5). UPLC was performed on an Agilent 1290 system equipped with a Phenomenex Kinetex XB-C18 column (4.6 mm × 100 mm, 2.6 μm), and data were processed using Agilent OpenLab CDS ChemStation Edition C.01.07.

For *Pp*ATase activity assays, 100 μL reaction mixtures contained 50 mM Kpi buffer (pH 7.5), 12 μM purified enzyme, 1 mM PG, 1.5 mM of either PA, or acetyl-CoA, malonyl-CoA, acetoacetyl-CoA, acetyl-P. Reactions were incubated at 30 °C for 16 h, and quenched with 100 μL methanol, and centrifuged at 21,000×g for 30 min. Supernatants were analyzed by UPLC using LC method 2 (Supplementary Table 5).

For *in vitro* HMG-CoA formation assays, 50 μL reaction mixtures were prepared containing 50 mM Kpi buffer (pH 7.5), 15 μM purified enzyme, 1 mM acetyl-CoA, and 1mM CoA. After incubation at 30 °C for 2.5 h, reactions were quenched with 2.5 μL formic acid and 50 μL methanol. Samples were centrifuged at 21,000×g for 30 min, and supernatants were analyzed by liquid chromatography-mass spectrometry (LC-MS) using LC method 3 (Supplementary Table 5).

Thiolytic cleavage activity of acetoacetyl-CoA was assessed in 50 μL reaction mixtures containing 50 mM Kpi buffer (pH 7.5), 15 μM purified enzyme, 1 mM acetoacetyl-CoA, and 1mM CoA. After incubation at 30 °C for 30 min, reactions were quenched with 2.5 μL formic acid and 50 μL methanol. Following centrifugation (21,000 × g, 30 min), supernatants were analyzed by LC-MS using LC method 4 (Supplementary Table 5) on a Phenomenex Kinetex XB-C18 column (4.6 mm × 100 mm, 2.6 μm).

All LC-MS analyses were performed on the 1290 infinity UPLC system coupled to an Agilent Technologies 6470 triple quadrupole mass spectrometer. The LC/TQ used a Jet Stream Electrospray Ionization (AJS-ESI) source operated in the negative ionization mode (capillary voltage, 3.5 kV; gas temperature, 300 °C; drying gas, 5 L min^{-1}; fragmentor voltage: 135V; nebulizer, 45 psig; sheath gas flow, 11 L·min^{-1}; acquisition range, 100-1500 m/z). MS data were processed using Agilent MassHunter Qualitative Analysis B.06.00.

Isolation and structural characterization of the enzymatic products

To determine the structure of the enzymatic product of *Sma*ATase, the *in vitro* reaction was scaled up to 50 ml reaction volume containing 50 mM Kpi buffer (pH 7.5), 1 mM preRTC-A, 1.5 mM PA, 15 μM purified enzyme, and incubated at 30 °C for 16 h. For MAPG isolation, the reaction was similarly scaled up to 100 mL using 1 mM PG and the same buffer

and enzyme conditions, followed by incubation at 30 °C for 16 h. Reaction mixtures were extracted three times with equal volume of ethyl acetate (EtOAc), and the organic phase were combined and concentrated under vacuum. RTC-A (t_R = 10.6 min) was purified via semi-preparative HPLC on a Phenomenex Kinetex XB-C_{18} column (250 mm × 10 mm, 5 μm) using isocratic elution with acetonitrile (ACN)/H_2O (95 : 5, v/v) containing 0.1 % TFA at a flow rate of 2.5 mL/min. MAPG (t_R = 6.9 min) was purified under similar conditions using ACN/H_2O (40 : 60, v/v) with 0.1 % TFA. The ^1H and ^{13}C NMR spectra were recorded on a Bruker Advance Neo 400 MHz or 600 MHz spectrometer and processed by MestReNova 6.1.0 (Metrelab). NMR data are presented in Supplementary Tables 1 and 2, and NMR spectra are provided in Supplementary Figures 4, 5, 7, and 9.

Cryo-EM structural determination

For Cryo-EM sample preparation, Holey carbon grids (Au R1.2/1.3, 300 mesh, Quantifoil) were used with 30s glow-discharge using medium power of Plasma Cleaner PDC-32G (Harrick). The Vitrobot Mark IV (Thermo Fisher Scientific) was maintained at the condition of 8 ℃ and 100 % humidity during sample preparation. For each sample, 3.5 μL of protein (3.8 mg/mL of *Sma*ATase in 50 mM potassium phosphate, pH 7.5) were loaded on the grid, blotted for 4.5 s with filter paper, and plunged into liquid ethane. The vitrified grids were subsequently transferred to liquid nitrogen for storage.

Data acquisition was performed on the 300 kV Titan Krios electron microscope (Thermo Fisher Scientific) equipped with a Falcon4i detector and GIF Quantum energy filter. A total of 3,089 micrographs were collected automatically using EPU, with a preset defocus range of −1.2 to −1.8 μm. Each image stack received a total dose of 50 e-/Å², and the final images were binned to a pixel size of 0.927 Å.

Data processing flow shown in Supplementary Fig. 11 was performed on cryoSPARC[45]. Initial particle picking using blob picking yielded 1,155,067 particles. Following 2D classification, 973,419 particles were retained and re-extracted at bin1 for subsequent processing. Ab-initio reconstruction of these particles produced an octameric structure, and the resolution of this structure reaches 2.16 Å after non-uniform refinement. During 2D classification, we observed particles exhibiting higher-order polymeric assemblies. Through iterative ab-initio reconstruction and 3D classification, we isolated 71,520 particles for subsequent processing. Non-uniform refinement followed by local refinement of these particles yielded a dodecameric structure at 2.32 Å resolution.

Structural models for the octameric and dodecameric complexes were built based on their respective cryo-EM maps. The initial monomeric models of MucA, MucB, and MucC, predicted by AlphaFold2, were docked into the density maps using ChimeraX[46] to generate initial complex models. These models were then manually adjusted and refined in Coot[47], followed by automated real-space refinement using phenix.real_space_refine in PHENIX[48]. The stereochemical quality of the final models was validated with MolProbity. Data collection and refinement statistics are summarized in Table S3. The

atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 9VBO (octamer) and 9VBT (dodecamer).

Mass spectrometry

For TOF mass spectrometry, protein samples were buffered to 50 mM ammonium acetate (pH 7.5) before applying on SYNAPT XS (Waters) mass spectrometer equipped with an ACQUITY UPLC I-Class PLUS system. A 2 µL aliquot of protein solution (1 mg/mL) was injected onto a MAbPacTM RP column (3.0 mm × 100 mm, 4 µm, Thermo Fisher Scientific) and separated using LC method 5 (Supplementary Table 5). Mass spectrometry was performed on a Waters SYNAPT XS operating in positive ESI mode over the range of 200−4000 m/z. Chromatography-coupled mass spectrometry was performed at 1 mm scale with the following optimized conditions: Electrospray voltage: 3.0 kV, ion source: 120 °C, desolvation: 450 °C, cone gas: 50 L/h, desolvation gas: 800 L/h, nebulizer pressure: 6.5 bar. Raw spectra were acquired using MassLynx 4.2 (Waters) and processed through Maximum Entropy (MaxEnt) deconvolution to generate the molecular mass spectra of injected proteins.

Native MS analysis was performed using a Q-Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific) as described previously[49]. Briefly, samples were buffer-exchanged into 0.2 M ammonium acetate (pH 7.5) and ionized by a nESI source. Spectra were acquired with 50−200V collisional activation in the HCD cell. Mass and charge assignments were determined using an Excel script kindly provided by the Benesch Laboratory, University of Oxford[50].

For protein posttranslational modification analysis, freshly purified *Sma*ATase was reduced with TCEP [Tris(2-carboxyethyl)phosphine, 10 mM] at 37 °C with agitation (800 rpm) for 60 min, alkylated with iodoacetamide (IAA, 40 mM) in the dark at room temperature for 30 min, and digested in-solution with a combination of trypsin and chymotrypsin, both at a ratio of 1:50 (*w*/*w*) to *Sma*ATase, in 50 mM ammonium bicarbonate at 37 °C for overnight. The digestion was quenched with 0.1% formic acid, and the resulting peptides were desalted prior to LC-MS/MS analysis. Approximately 1 µg of sample was injected into a Vanquish Neo nanoflow LC system equipped with a Orbitrap Exploris 480 Mass Spectrometer (Thermo Fisher Scientific). Peptides were separated on an Acclaim PepMap 100 C18 column (250 mm × 75 µm, 3 µm, Thermo Fisher Scientific) by using LC method 6 (Supplementary Table 5). Full-scan MS spectra (*m*/*z* 350–1500) were acquired in the Orbitrap at a resolution of 120,000 (at *m*/*z* 200), with an RF lens setting of 50%, an automatic gain control (AGC) target of 300%, and a maxIT of 25 ms. MS/MS analysis was performed with the following settings: MS2 resolution of 15,000 (at *m*/*z* 200); the AGC target was set to 200% and the maxIT to 22 ms; dynamic exclusion was set to 30 s; MS2 scan range was set to *m*/*z* 150–2000. Database searching was performed using Peaks Studio software. The mass error tolerances were set to 10 ppm for precursor

ions and 0.02 Da for fragment ions; enzyme specificity was set to semi-specific for both trypsin and chymotrypsin, allowing for up to 2 missed cleavages.

Mass photometry measurements

Protein samples were diluted in phosphate-buffered saline (PBS) to a final concentration of 200 nM for mass photometry (MP) analysis. Measurements were performed on a OneMP instrument (Refeyn) using pre-cleaned glass coverslips. Following the standard protocol[51], 10 μL of PBS was loaded into the sample well for system focusing, followed by addition of 10 μL diluted protein sample (1:1 mixing ratio). Data acquisition commenced immediately after mixing, recording for 60 s at 100 frames/s using AcquireMP software v2.3. The system was calibrated with bovine serum albumin (BSA) standards. Raw data were processed using DiscoverMP software v2.1.

Phylogenetic analysis

Protein sequences of thiolase superfamily members were compiled from published datasets[26]. Homologs of MucA, MucB, PhlA, and PhlC were obtained via BLASTP searches against the NCBI non-redundant protein database (nr). All sequences used in the analysis are provided as a Source Data file. Multiple sequence alignments were performed using MUSCLE with default parameters. Phylogenetic reconstruction was conducted in MEGA11[52] using the Neighbour-Joining method with the Poisson model, incorporating 500 bootstrap replicates to assess node support. The resulting Newick-formatted tree was visualized using the Interactive Tree of Life (iTOL) server tool (https://itol.embl.de/).

Genome mining methods

To systematically identify multi-component enzyme complexes with *C*-acyltransferase activity across the bacterial kingdom, a domain-centric genome mining approach was employed. Based on the biosynthetic gene cluster organization of *muc* and *phl* BGC, hidden Markov model (HMM) search[53] were conducted for Pfam domain[54] corresponding to HMGS (MucA and PhlA), ACAT (MucB and PhlC), DUF35 proteins (MucC and PhlB). Specifically, the domains correspond to PF08541 [3-oxoacyl-(acyl-carrier-protein) synthase III C-terminal], PF00108 (thiolase), and PF01796 (DUF35 OB-fold domain), respectively, were served as probes for further genome mining. The thiolase domain was used as the anchor, and genomic regions encompassing 10 genes upstream and downstream were scanned for the presence of the other two domains. Loci containing all three domains were retained, and redundancy was eliminated using FastANI. The resulting BGCs were analyzed using antiSMASH[55] and visualized with BiG-SCAPE[56].

Molecular docking and molecular dynamics simulation

The ligand structure was first optimized at the B3LYP/def2-TZVP level of theory using the ORCA 5.0.4[57] quantum chemistry package to obtain a reliable equilibrium geometry and electronic structure. The optimized ligand was subsequently docked into the target protein using the DiffDock[58]. The top-ranked docking pose, based on DiffDock's

confidence score and binding-site complementarity, was selected for further molecular dynamics (MD) simulations. Docking poses were visualized using PyMOL v3.0[59] and ProLIF[60].

To investigate the interaction between the ligand and MucA-B-C protein complex, classical molecular dynamics (MD) simulations were carried out. In the constructed systems, the protonation states of all ionizable residues were determined using PROPKA[61], based on empirical pKa predictions under physiological pH conditions. The following residues were modeled in their protonated form: His305 and Asp35 in MucA; Glu57, Asp134, and Glu296 in MucB; and Glu70 in MucC. All remaining residues retained their standard protonation states at neutral pH. All MD simulations were performed with GROMACS[62] (version 2025.03) using an all-atom force field for the protein (AMBER ff14SB)[63] and a compatible ligand topology generated via appropriate parameterization (GAFF2)[64]. RESP[65] charges were generated using Psi4[66] for all ligands as well as for the acetylated Cys87 residue. The protein–ligand complex was solvated in a cubic box with explicit TIP3P water molecules, and counterions (Na^+/Cl^-) were added to neutralize the total charge.

Prior to dynamics, the system underwent energy minimization using the steepest descent algorithm until the maximum force fell below 1000 kJ mol^{-1} nm^{-1}, ensuring removal of steric clashes. Subsequently, two equilibration phases were performed: The system was equilibrated for 1 ns under the canonical ensemble using the V-rescale thermostat[67] to maintain the temperature at 300 K, while restraining the heavy atoms of the protein and ligand with a force constant of 1000 kJ mol^{-1} nm^{-2}. The simulation was continued for 1 ns under the isothermal–isobaric ensemble (NPT), maintaining a pressure of 1 bar using the Parrinello–Rahman barostat[68]. Position restraints were retained to allow solvent and ions to relax around the solute. Finally, to improve statistical sampling, three independent 200 ns production MD simulations were conducted in the NPT ensemble without positional restraints, starting from distinct randomized initial velocity distributions. A 2 fs integration time step was employed, and all covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm[69]. Long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) method[70], and van der Waals interactions were truncated at 1.0 nm. Trajectories were saved every 10 ps for subsequent analyses. Binding free energy between the ligand and protein was calculated using the MM/GBSA method[71].

Antibacterial activity and MIC assays

Stock solution of preRTC-A, RTC-A, PG and DAPG were prepared in DMSO at concentrations of 6.4 mg/mL or 8 mg/mL. Ampicillin sodium salt, dissolved in water at the same concentrations, served as a positive control. The following indicator strains were used: *Streptococcus salivarius* DA547, *Streptococcus oralis* DA1241, *Streptococcus gordonii* ATCC 10558, *Bacillus subtilis* 168, *Staphylococcus aureus* ATCC 29213, *Methicillin-resistant Staphylococcus aureus* 544, *Methicillin-*

resistant Staphylococcus aureus 103 and *Streptococcus pneumonia* ATCC 49619.

For MIC assays, compounds were serially 2-fold diluted in MHB to 50 μL per well in 96-well plates. Bacterial cultures in mid-log phase were diluted to 1×10^6 CFU/mL in MHB, and 50 μL of this suspension was added to each well, yielding a final inoculum of 5×10^5 CFU/mL (validated range: 2×10^5 CFU/mL to 8×10^5 CFU/mL). *Streptococcus oralis* DA1241 and *S. gordonii* ATCC 10558 were cultured in BHI due to poor growth in MHB. Each plate included antimicrobial-free medium as a negative control and untreated bacterial culture as a growth control. Plates were incubated at 37 °C for 18 h, and absorbance was measured at 625 nm using a BioTek Epoch 2 (Agilent). MIC values were defined as the lowest concentration that completely inhibited visible growth ($OD_{625} \leqslant 0.02$ compared to negative control). All assays were performed in duplicate and independently repeated three times.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. Source data are provided with this paper. The cryo-EM structures and corresponding density maps are available in the Protein Data Bank under accession codes 9VBO ($MucA_4B_2C_2$) and 9VBT ($MucA_4B_4C_4$), and in the Electron Microscopy Data Bank database under accession codes EMD-64929 ($MucA_4B_2C_2$) and EMD-64933 ($MucA_4B_4C_4$).

References

1. Wang, Y., Shi, Y.N., Xiang, H. & Shi, Y.M. Exploring nature's battlefield: organismic interactions in the discovery of bioactive natural products. *Nat. Prod. Rep.* **41**, 1630–1651 (2024).
2. Chevrette, M.G. *et al.* Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat. Prod. Rep.* **37**, 566–599 (2020).
3. Moghe, G. & Last, R.L. Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* **169**, 1512–1523 (2015).
4. Jenke-Kodama, H., Müller, R. & Dittmann, E. Evolutionary mechanisms underlying secondary metabolite diversity. *Prog. Drug Res.* **65**, 121–140 (2008).
5. Nivina, A., Yuet, K.P., Hsu, J. & Khosla, C. Evolution and diversity of assembly-line polyketide synthases. *Chem. Rev.* **119**, 12524–12547 (2019).
6. Ngaki, M.N. *et al.* Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* **485**, 530–533 (2012).
7. Kaltenbach, M. *et al.* Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* **14**, 548–555 (2018).
8. Miller, M., Bachmann, B., Townsend, C. & Rosenzweig, A.C. Structure of β-lactam synthetase reveals how to synthesize antibiotics instead of asparagine. *Nat. Struct. Mol. Biol.* **8**, 684–689 (2001).
9. Tahlan, K. & Jensen, S.E. Origins of the β-lactam rings in natural products. *J. Antibiot.* **66**, 401–410 (2013).
10. Harijan, R.K., Dalwani, S., Kiema, T.-R., Venkatesan, R. & Wierenga, R.K. Thiolase: a versatile biocatalyst employing coenzyme A–thioester chemistry for making and breaking C–C bonds. *Annu. Rev. Biochem.* **92**, 351–384 (2023).
11. Haapalainen, A.M., Meriläinen, G. & Wierenga, R.K. The thiolase superfamily: condensing enzymes with diverse

reaction specificities. *Trends Biochem. Sci.* **31**, 64–71 (2006).

12. Kunau, W.H., Dommes, V. & Schulz, H. β-Oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress. *Prog. Lipid Res.* **34**, 267–342 (1995).
13. Heath, R.J. & Rock, C.O. The Claisen condensation in biology. *Nat. Prod. Rep.* **19**, 581–596 (2002).
14. Vögeli, B. *et al.* Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. *Proc. Natl. Acad. Sci. USA* **115**, 3380–3385 (2018).
15. Tang, X. *et al.* Cariogenic *Streptococcus mutans* produces tetramic acid strain-specific antibiotics that impair commensal colonization. *ACS Infect. Dis.* **6**, 563–571 (2020).
16. Zhang, Y. *et al.* Human-associated bacteria adopt an unusual route for synthesizing 3-acetylated tetramates for environmental adaptation. *Microbiome* **11**, 97–112 (2023).
17. Hao, T. *et al.* An anaerobic bacterium host system for heterologous expression of natural product biosynthetic gene clusters. *Nat. Commun.* **10**, 3665–3677 (2019).
18. Hayashi, A. *et al.* Molecular and catalytic properties of monoacetylphloroglucinol acetyltransferase from *Pseudomonas* sp. YGJ3. *Biosci. Biotech. Bioch.* **76**, 559–566 (2012).
19. Schmidt, N.G. *et al.* Biocatalytic Friedel-Crafts acylation and Fries reaction. *Angew. Chem. Int. Ed.* **56**, 7615–7619 (2017).
20. Pavkov-Keller, T., Schmidt, N.G., Żądło-Dobrowolska, A., Kroutil, W. & Gruber, K. Structure and catalytic mechanism of a bacterial Friedel-Crafts acylase. *ChemBioChem* **20**, 88–95 (2019).
21. Schmidt, N.G. & Kroutil, W. Acyl donors and additives for the biocatalytic Friedel-Crafts acylation. *Eur. J. Org. Chem.* **2017**, 5865–5871 (2017).
22. Żądło-Dobrowolska, A., Schmidt, N.G. & Kroutil, W. Thioesters as acyl donors in biocatalytic Friedel-Crafts-type acylation catalyzed by acyltransferase from *Pseudomonas protegens*. *ChemCatChem* **11**, 1064–1068 (2019).
23. Liu, J., Ng, T., Rui, Z., Ad, O. & Zhang, W. Unusual acetylation-dependent reaction cascade in the biosynthesis of the pyrroloindole drug physostigmine. *Angew. Chem. Int. Ed.* **53**, 136–139 (2013).
24. Zhang, S. *et al.* Metabolic engineering for efficient supply of acetyl-CoA from different carbon sources in *Escherichia coli*. *Microb. Cell Fact.* **2019**, 130–140 (2019).
25. Sheng, X., Kazemi, M., Żądło-Dobrowolska, A., Kroutil, W. & Himo, F. Mechanism of biocatalytic Friedel-Crafts acylation by acyltransferase from *Pseudomonas protegens*. *ACS Catal.* **10**, 570–577 (2020).
26. Tan, Z., Clomburg, J.M., Cheong, S., Qian, S. & Gonzalez, R. A polyketoacyl-CoA thiolase-dependent pathway for the synthesis of polyketide backbones. *Nat. Catal.* **3**, 593–603 (2020).
27. Jiang, C., Kim, S. & Suh, D. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol. Phylogenet. Evol.* **49**, 691–701 (2008).
28. Weidenweber, S. *et al. Finis tolueni*: a new type of thiolase with an integrated Zn-finger subunit catalyzes the final step of anaerobic toluene metabolism. *FEBS J.* **289**, 5599–5616 (2022).
29. Żądło-Dobrowolska, A., Hammerer, L., Pavkov-Keller, T., Gruber, K. & Kroutil, W. Rational engineered C-acyltransferase transforms sterically demanding acyl donors. *ACS Catal.* **10**, 1094–1101 (2019).
30. Gänzle, M.G. Reutericyclin: biological activity, mode of action, and potential applications. *Appl. Microbiol. Biotechnol.* **64**, 326–332 (2004).
31. Cherian, P.T. *et al.* Chemical modulation of the biological activity of reutericyclin: a membrane-active antibiotic from *Lactobacillus reuteri*. *Sci. Rep.* **4**, 4721–4729 (2014).
32. Kumar, V., Turnbull, W.B. & Kumar, A. Review on recent developments in biocatalysts for Friedel-Crafts reactions. *ACS Catal.* **12**, 10742–10763 (2022).
33. Xia, Y. *et al.* Biosynthesis of 4-acyl-5-aminoimidazole alkaloids featuring a new Friedel-Crafts acyltransferase. *J. Am. Chem. Soc.* **145**, 26308–26317 (2023).

34. Peter W. Stacpoole & McCall, C.E. The pyruvate dehydrogenase complex: Life's essential, vulnerable and druggable energy homeostat. *Mitochondrion* **70**, 59–102 (2023).
35. Chen, M. *et al.* Molecular architecture of mammalian pyruvate dehydrogenase complex. *Protein Cell* **16**, 72–78 (2025).
36. Yang, L. *et al.* High resolution cryo-EM and crystallographic snapshots of the actinobacterial two-in-one 2-oxoglutarate dehydrogenase. *Nat Commun.* **14**, 4851–4865 (2023).
37. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315–1322 (2008).
38. Cogan, D.P. *et al.* Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science* **374**, 729–734 (2021).
39. Pistofidis, A. *et al.* Structures and mechanism of condensation in non-ribosomal peptide synthesis. *Nature* **638**, 270–278 (2024).
40. Castellana, M. *et al.* Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* **32**, 1011–1018 (2014).
41. Wheeldon, I. *et al.* Substrate channelling as an approach to cascade reactions. *Nat. Chem.* **8**, 299–309 (2016).
42. Noda-Garcia, L. & Tawfik, D.S. Enzyme evolution in natural products biosynthesis: target- or diversity-oriented? *Curr. Opin. Chem. Biol.* **59**, 147–154 (2020).
43. Rezzonico, F. *et al.* Is the ability of biocontrol fluorescent pseudomonads to produce the antifungal metabolite 2,4-diacetylphloroglucinol really synonymous with higher plant protection? *New Phytol.* **173**, 861–872 (2006).
44. Gutiérrez-García, K. *et al.* Phylogenomics of 2,4-diacetylphloroglucinol-producing *Pseudomonas* and novel antiglycation endophytes from *Piper auritum*. *J. Nat. Prod.* **80**, 1955–1963 (2017).
45. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
46. Pettersen, E.F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
47. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
48. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
49. Hoi, K.K. *et al.* Detergent-free lipodisq nanoparticles facilitate high-resolution mass spectrometry of folded integral membrane proteins. *Nano Lett.* **21**, 2824–2831 (2021).
50. Mistarz, U.H., Chandler, S.A., Brown, J.M., Benesch, J.L.P. & Rand, K.D. Probing the dissociation of protein complexes by means of gas-phase H/D exchange mass spectrometry. *J. Am. Soc. Mass Spectrom.* **30**, 45–57 (2019).
51. Wu, D. & Piszczek, G. Standard protocol for mass photometry experiments. *Eur. Biophys. J.* **50**, 403–409 (2021).
52. Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
53. Finn, R.D., Clements, J. & Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, 27–39 (2011).
54. Mistry, J. *et al.* Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, 412–419 (2021).
55. Blin, K. *et al.* antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. *Nucleic Acids Res.* **53**, W32–W38 (2025).
56. Navarro-Muñoz, J.C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
57. Neese, F. Software update: The ORCA program system—Version 5.0. *WIREs Comput. Mol. Sci.* **12**, e1606 (2022).

58. Corso, G., Deng, A., Polizzi, N., Barzilay, R. & Jaakkola, T.S. Deep confident steps to new pockets: strategies for docking generalization. *International Conference on Learning Representations* (2024).
59. Yuan, S., Chan, H.C.S. & Hu, Z. Using PyMOL as a platform for computational drug design. *WIREs Comput. Mol. Sci.* **7**, e1298 (2017).
60. Bouysset, C. & Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints. *J. Cheminf.* **13**, 72–80 (2021).
61. Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M. & Jensen, J.H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
62. Páll, S. *et al.* Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **153**, 134110 (2020).
63. Maier, J.A. *et al.* ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
64. He, X., Man, V.H., Yang, W., Lee, T.S. & Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **153**, 114502 (2020).
65. Cornell, W.D., Cieplak, P., Bayly, C.I. & Kollman, P.A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **115**, 9620–9631 (1993).
66. Smith, D.G.A. *et al.* Psi4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **152**, 184108 (2020).
67. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
68. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
69. Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
70. Darden, T., York, D. & Pedersen, L. Particle mesh ewald: An *N*.Log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
71. Weng, G. *et al.* Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein–peptide complexes. *Phys. Chem. Chem. Phys.* **21**, 10135–10145 (2019).

Acknowledgements

This work has been supported by the National Natural Science Foundation of China (32401633 to G.L, 82173719 to X.T., and 32322039 to X.P.), Guangdong Basic and Applied Basic Research Foundation (2023A1515111192 to G.L and 2024A1515010922 to X.T.), Guangdong S&T Program (2024B1111160007 to X.T.), Shenzhen Science and Technology Program (KJZD20240903101104007 to X.T.), Shenzhen Bay Laboratory Startup Fund (21230051 to X.T. and QH28001 to K.K.H.), the Major Program of Shenzhen Bay Laboratory (C1012523006 to X.T.), and Guangdong Province's Pearl River Recruitment Program of Talents (2021QN02Y855 to X.T.). The authors are grateful to T. Zheng and E. Li from the Multi-Omics Mass Spectrometry Core of Shenzhen Bay Laboratory for the help with IMS-TOF mass measurements, H. Yin and J. Zheng from Shenzhen Medical Academy of Research and Translation for the acquisition of Orbitrap mass spectrometry data, and K. Wang from

Refeyn Co., Ltd. for the assistance with MP measurements. We thank L. Dai (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences) for providing *S. salivarius* DA547 and *S. oralis* DA1241.

Author contributions

X.T. and G.L. conceived and designed the project. G.L., R.S., A.W., and C.P. conducted biochemical experiments, metabolic profiling, and antibacterial activity assays. Z.S. and Z.L. acquired the cryo-EM data and performed structural modeling and refinement under the supervision of X.P., G.W. performed bioinformatic analyses and genome mining under the supervision of Y.-X.L., and F.L. carried out native mass spectrometry experiments under the supervision of K.K.H.. Z. S. performed molecular docking and molecular dynamic simulations. C.M. executed chemical synthesis and compound isolation. G.L. and X.T. prepared the manuscript with input from all authors. X.T. supervised the project.

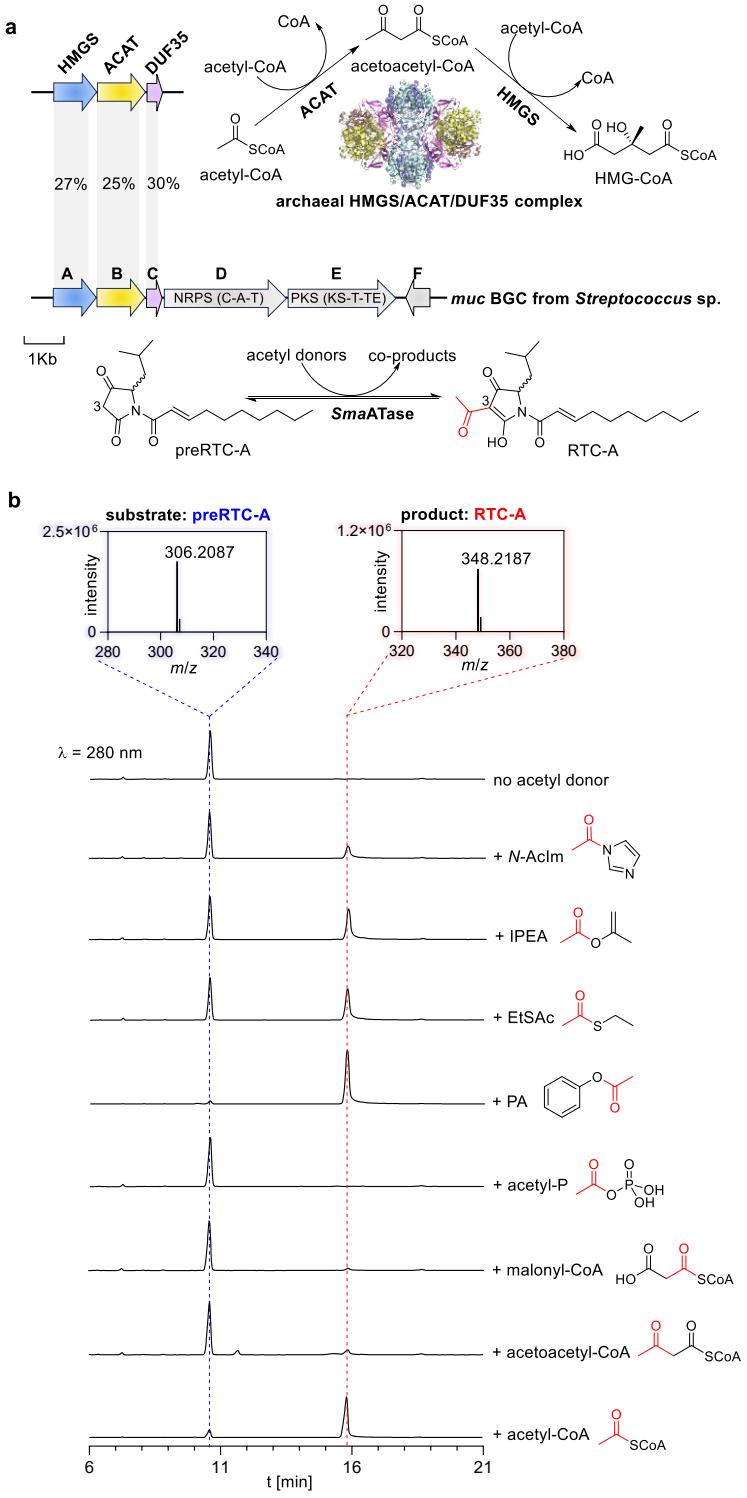

Fig. 1 | *In vitro* functional characterization of *Sma*ATase and comparison with the archaeal HMGS/ACAT/DUF35 complex. a The archaeal HMGS/ACAT/DUF35 complex is a multimeric enzyme assembly that sequentially catalyzes the first two steps of the mevalonate pathway. **b** HR-MS spectra of preRTC-A (substrate) and RTC-A (product), along with UPLC analysis of the *in vitro* enzymatic assays of *Sma*ATase and preRTC-A in the presence of various acetyl donors, including *N*-acetylimidazole (*N*-AcIm), isopropenyl acetate (IPEA), ethyl thioacetate (EtSAc), phenyl acetate (PA), acetyl-phosphate (acetyl-P), malonyl-CoA, acetoacetyl-CoA, and acetyl-CoA. All assays were run in triplicate and representative results are shown.

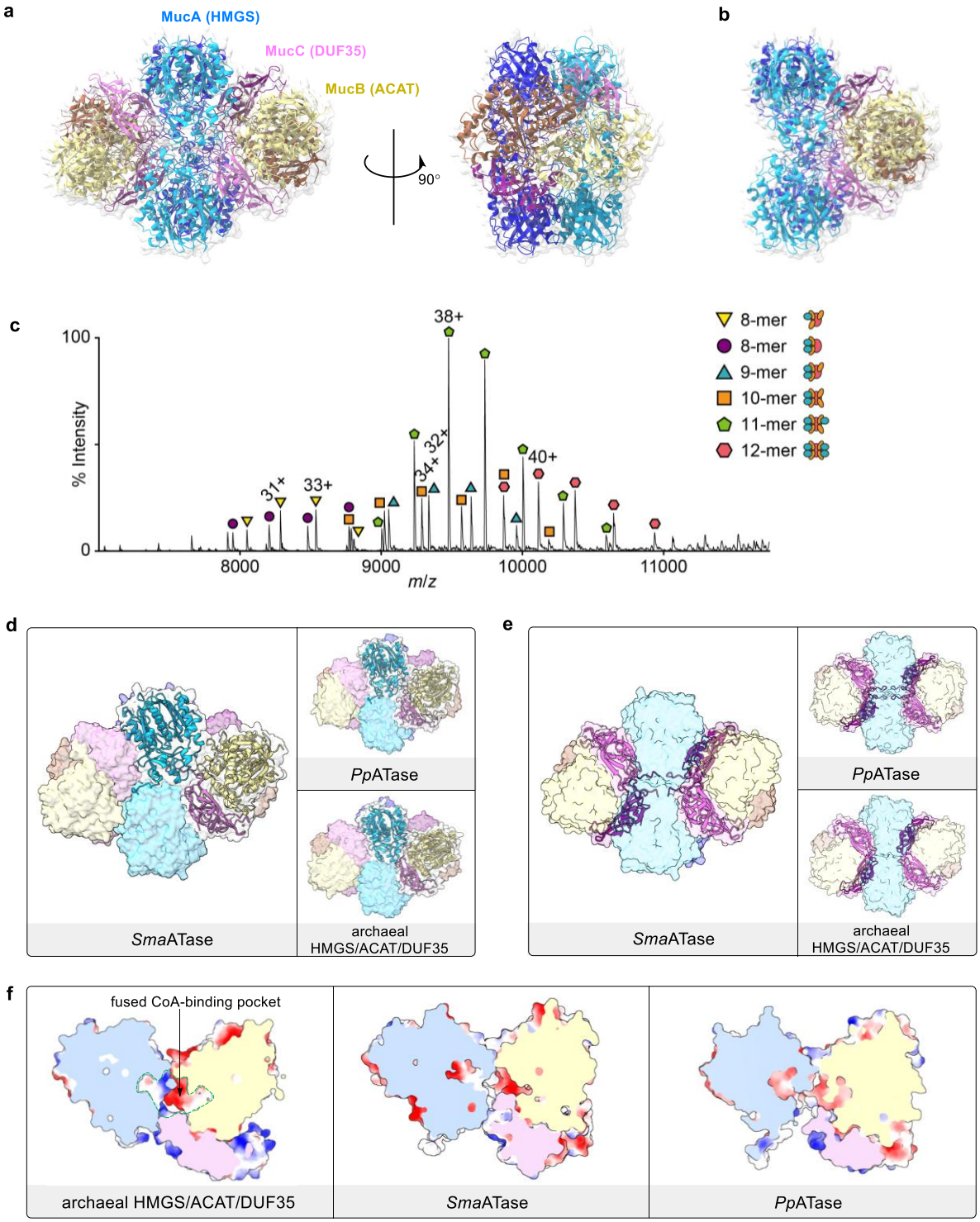

Fig. 2 | Structural analysis of *Sma*ATase. a Cryo-EM structural analysis of *Sma*ATase revealed a dodecameric complex with $MucA_4B_4C_4$ composition. **b** A second octameric state contains two MucB units bound to the $MucA_4C_2$ core. Subunit coloring: MucA = deep sky blue and blue; MucB = khaki and sienna; MucC = orchid and dark magenta. **c** Native mass spectra of *Sma*ATase shows peaks corresponding to multiple oligomeric states (the most abundant charge states of each oligomeric states are denoted). **d** Cartoon representations of HMGS/ACAT/DUF35 complexes illustrate a conserved tripartite architectural module. **e** Surface representations of HMGS/ACAT/DUF35 complexes, with DUF35 units highlighted in cartoon format. **f** Side view of the cut-open electrostatic surface of the HMGS (light steel blue), ACAT (wheat), DUF35 (thistle) subunits, showing fused CoA-binding sites.

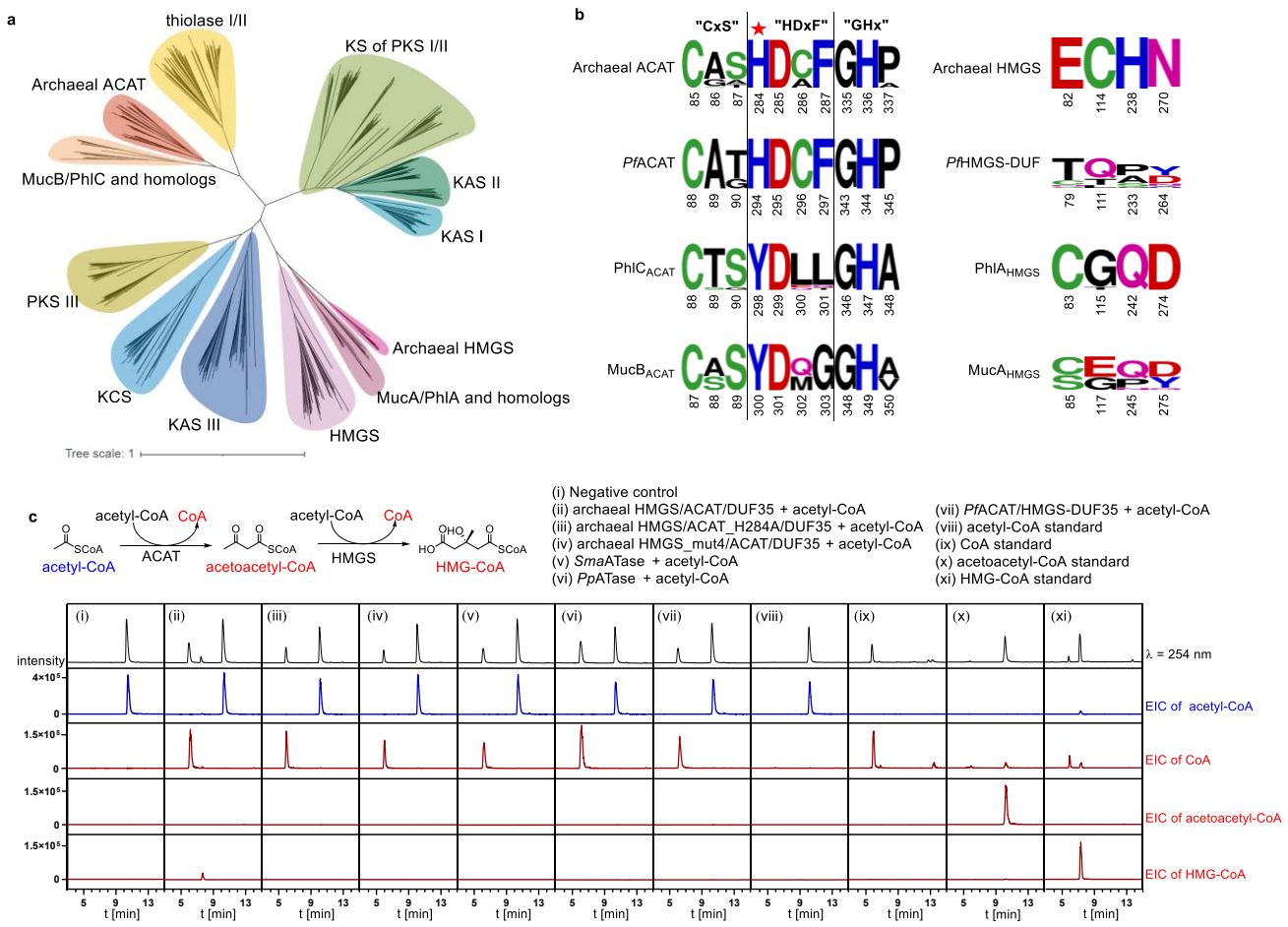

Fig. 3 | Comparison of conserved residues and biochemical activities of HMGS/ACAT/DUF35 complexes from archaea and bacteria. **a** Phylogenetic analysis of the thiolase superfamily, including ketoacyl-CoA thiolase (thiolase I), ACAT (also known as acetoacetyl-CoA thiolase, thiolase II), β-ketoacyl-acyl carrier protein synthase I, II, and III (KAS I, II, and III), β-ketoacyl-CoA synthase (KCS), ketosynthase (KS) domain from type I/II PKS systems, and type III PKSs, HMGS, HMGS-like sequences (e.g., MucA, PhlA and their homologs), and ACAT-like sequences (e.g., PhlC, MucB and their homologs). **b** Sequence alignment comparing conserved residues in archaeal and bacterial ACATs and HMGSs. Asterisks indicate key residues that are essential for catalysis in archaeal ACAT and *Pf*ACAT but mutated in bacterial counterparts (PhlC and MucB), potentially facilitating alternative substrate binding. **c** LC-MS analysis of *in vitro* biochemical assays using acetyl-CoA as substrate to evaluate the activity of archaeal and bacterial HMGS/ACAT/DUF35 complexes. EIC for acetyl-CoA (*blue*) represents the substrate, while EICs in red denote potential product ions. UV chromatogram monitored at 254 nm are displayed in black. The "mut4" variant corresponds to the quadruple mutant E82C/C114G/H238Q/N270D. All assays were run in triplicate and representative results are shown.

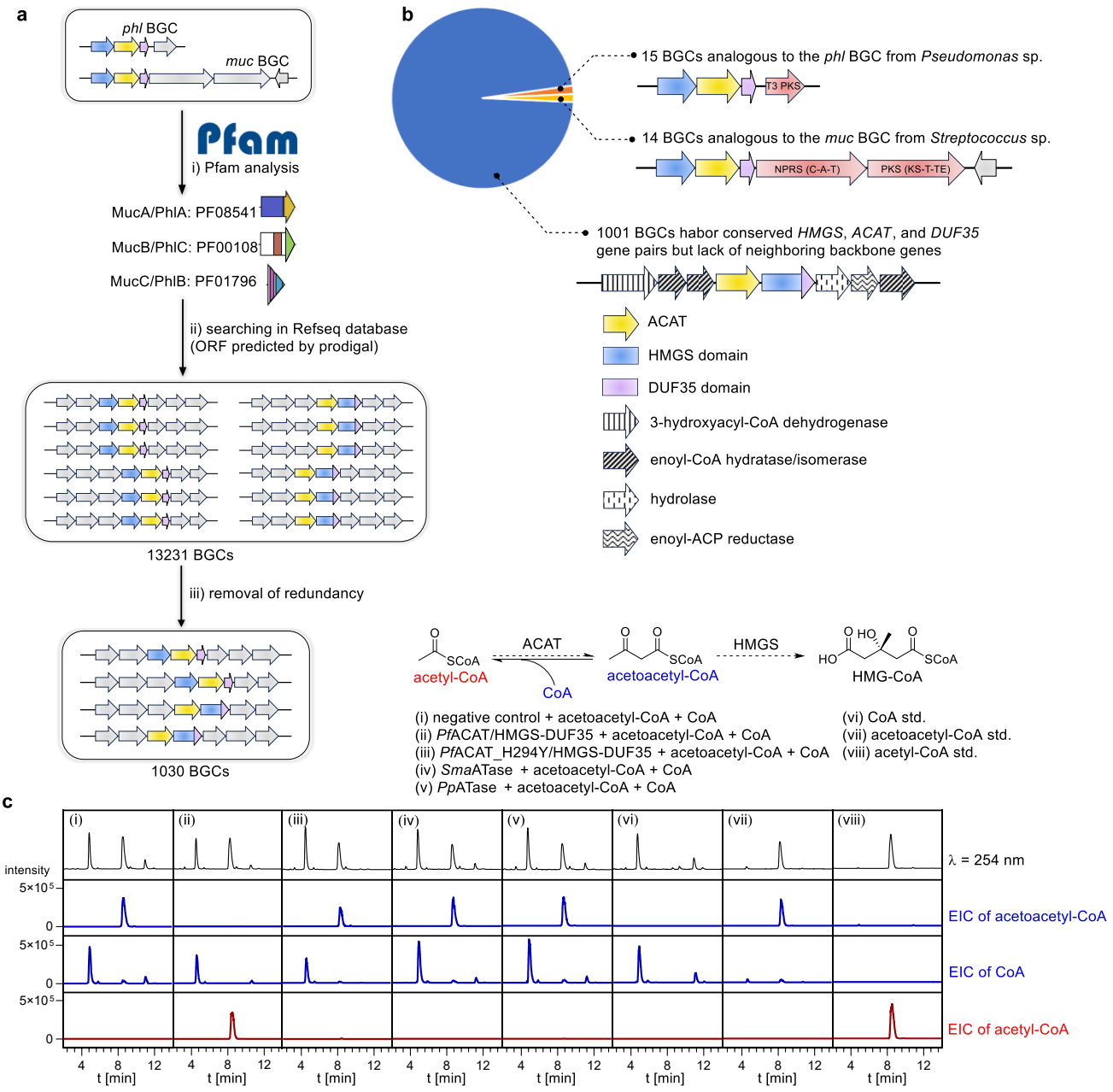

Fig. 4 | Genome-mining and functional characterization of the ACAT/HMGS-DUF35 complex from *P. fluorescens*. **a** Schematic overview of the genome mining workflow used to identify bacterial gene clusters encoding co-occurring ACAT, HMGS, and DUF35 domains. **b** Representative gene cluster architectures identified from genome mining: the *phl* BGC, *muc* BGC, and *ACAT-HMGS-DUF35* gene clusters from *P. fluorescens*. **c** LC-MS analysis of acetoacetyl-CoA thiolytic cleavage activity of *Sma*ATase, *Pp*ATase, *Pf*ACAT/HMGS-DUF35, and *Pf*ACAT_H294Y/HMGS-DUF35, respectively. EICs for acetoacetyl-CoA and CoA are colored in blue to represent the substrates, while EIC in red represents acetyl-CoA as the potential product. UV chromatogram monitored at 254 nm are displayed in black. All assays were run in triplicate and representative results are shown.

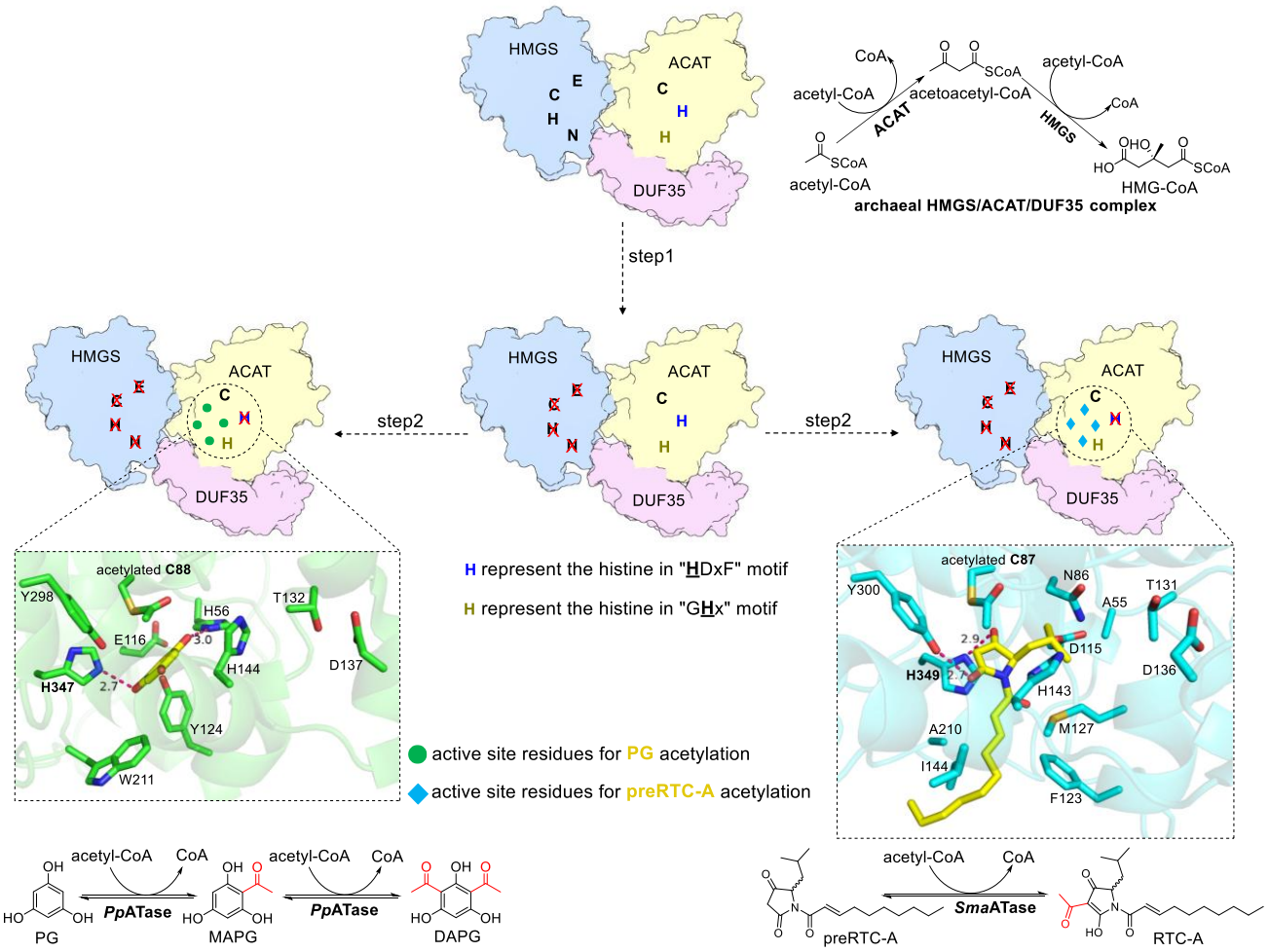

Fig. 5 | Proposed evolutionary trajectory for the neofunctionalization of HMGS/ACAT/DUF35 complexes. **Step1**: Mutations of the conserved E-C-H-N tetrad lead to inactivation of the HMGS unit. **Step2**: Subsequent mutations of the His residue in the HDxF motif and acquisition of specialized active-site residues enable the use of PG or preRTC-A as acetyl acceptors.

Hydrogen bonds are depicted as dashed lines, representing interactions measured between the donor and acceptor heavy atoms.

The reliability of the simulated substrate-binding mode is supported by the identification of specific protein-ligand interactions, alongside with the convergence of the protein-ligand RMSD, the low residual flexibility (RMSF) in the binding pocket, and the stable compactness (Rg) and solvent accessibility (SASA) (Supplementary Fig. 27).

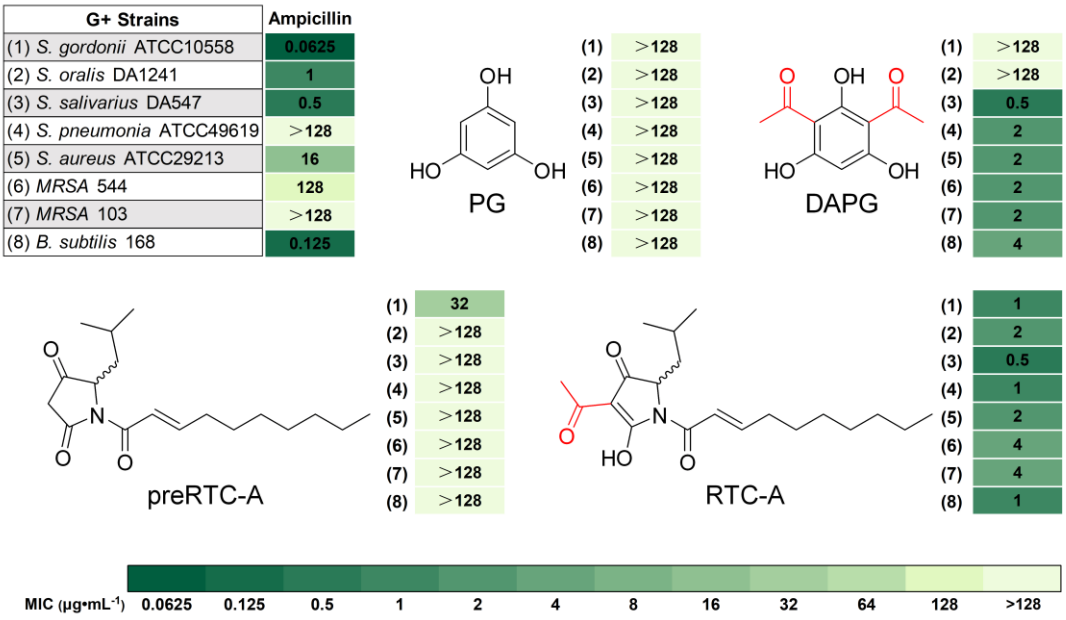

Fig. 6 | Antimicrobial activities of RTC-A and DAPG compared to their unacetylated precursors, preRTC-A and PG. Minimum inhibitory concentrations (MIC, μg·ml^{-1}) were tested against a panel of strains, including *Streptococcus salivarius* DA547, *Streptococcus oralis* DA1241, *Streptococcus gordonii* ATCC10558, *Bacillus subtilis* 168, *Staphylococcus aureus* ATCC29213, methicillin-resistant *Staphylococcus aureus* 544, methicillin-resistant *Staphylococcus aureus* 103, *Streptococcus pneumonia* ATCC49619. Acetylated products (RTC-A and DAPG) consistently exhibited enhanced antimicrobial activity compared to their unacetylated precursors (preRTC-A and PG).