1
2

The biosynthesis of the odorant 2-methylisoborneol is compartmentalized inside a protein shell

3

4   Michael P. Andreas[1] and Tobias W. Giessen[1*]

5

6   [1]Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor,

7   MI 48109, USA

8   *correspondence: tgiessen@umich.edu

9
10

11 **Abstract**

12   Terpenoids are the largest class of natural products, found across all domains of
13 life. One of the most abundant bacterial terpenoids is the volatile odorant 2-
14 methylisoborneol (2-MIB), partially responsible for the earthy smell of soil and musty
15 taste of contaminated water. Many bacterial 2-MIB biosynthetic gene clusters were
16 thought to encode a conserved transcription factor, named EshA in the model soil
17 bacterium *Streptomyces griseus.* Here, we revise the function of EshA, now referred to
18 as *Sg* Enc, and show that it is a Family 2B encapsulin shell protein. Using cryo-electron
19 microscopy, we find that *Sg* Enc forms an icosahedral protein shell and encapsulates 2-
20 methylisoborneol synthase (2-MIBS) as a cargo protein. *Sg* Enc contains a cyclic
21 adenosine monophosphate (cAMP) binding domain (CBD)-fold insertion and a unique
22 metal-binding domain, both displayed on the shell exterior. We show that *Sg* Enc CBDs
23 do not bind cAMP. We find that 2-MIBS cargo loading is mediated by an N-terminal
24 disordered cargo-loading domain and that 2-MIBS activity and *Sg* Enc shell structure
25 are not modulated by cAMP. Our work redefines the function of EshA and establishes
26 Family 2B encapsulins as cargo-loaded protein nanocompartments involved in
27 secondary metabolite biosynthesis.

28
29
30

Introduction

Terpenoids, also called isoprenoids, are modified terpenes and represent the largest and most structurally diverse class of natural products[1]. They play essential roles in both primary and secondary metabolism and are found across all domains of life[2,3]. They include vitamins, steroids, fragrances, plant hormones, pigments, and many compounds used as drugs[4]. While the majority of characterized terpenoids have been isolated from plants and fungi, recent genomic studies have revealed a large number of diverse terpenoid biosynthetic gene clusters encoded in bacterial genomes[2]. All terpenoids are assembled from the five-carbon (C5) prenyl diphosphate precursors dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP)[1-3,5]. Through successive additions of IPP building blocks to DMAPP, prenyl synthetases assemble linear terpenoid precursors like geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), and geranylgeranyl pyrophosphate (GGPP). These linear precursors can subsequently undergo a wide range of enzyme-catalyzed modifications to produce diverse and structurally complex terpenoids. Often, linear precursors are modified through carbocation-driven cyclizations catalyzed by terpene cyclases (TCs), yielding (poly)cyclic terpenoids with complex carbon scaffolds[1]. The nomenclature of terpenoids is based on the number of five-carbon units in the parent terpene with the most common bacterial terpenoids being mono- (C10), sesqui- (C15), or diterpenoids (C20)[2].

Two of the most prevalent bacterial terpenoids are the methylated monoterpenoid 2-methylisoborneol (2-MIB) and the sesquiterpenoid geosmin, volatile odiferous compounds responsible for the earthy smell of soil and musty taste of contaminated water[2,6-8]. While the exact biological function of these volatile terpenoids is unclear, roles as insect attractants for bacterial spore dispersal[9] and repellants against bacterial predators[10-12] have recently been suggested. 2-MIB is widely produced by soil-dwelling Actinobacteria, Myxobacteria, and fungi, as well as by Cyanobacteria[2,6-8,13-16]. 2-MIB biosynthesis proceeds in two steps. Initially, an *S*-adenosyl methionine (SAM)-dependent methyltransferase (MT) modifies the linear C10 precursor GPP, yielding 2-methylgeranyl pyrophosphate (2-MeGPP). Next, 2-MeGPP is cyclized by the Mg^{2+}-dependent class 1 terpenoid cyclase 2-methylisoborneol synthase (2-MIBS), resulting in 2-MIB formation[13,16,17]. Both proteins – MT and 2-MIBS – are generally found in co-

regulated operons[13,18,19]. Commonly, a third gene encoding for a cyclic nucleotide-binding domain (CBD)-containing protein is also present in 2-MIB gene clusters[14,15,18]. This third gene was previously hypothesized to represent a transcription factor[20] as CBDs are a common feature of bacterial transcriptional regulators and include well characterized examples like the catabolite activator protein (CAP) from *Escherichia coli*[21]. However, more recently, the CBD-containing proteins found in 2-MIB biosynthetic gene clusters have been suggested to represent Family 2B encapsulin shell proteins instead[22-24].

Encapsulin nanocompartments (encapsulins) are the most recently discovered class of prokaryotic protein compartments able to sequester dedicated cargo enzymes[25]. Cargo encapsulation is mediated by specific targeting sequences found at the N- or C-terminus of all native cargo proteins[22,24,26,27]. Encapsulin shell proteins share the HK97 phage-like fold and likely have a viral origin[22,24,28,29]. Encapsulins self-assemble into icosahedral shells with triangulation numbers of T=1 (60 subunits), T=3 (180 subunits), or T=4 (240 subunits)[30-34]. Encapsulins have recently been categorized into four families based on sequence and genome neighborhood conservation[22]. Family 1 encapsulins are the best characterized and are involved in prokaryotic iron storage[26,31,35], detoxification[36,37], and stress resistance[24,38]. Family 2 encapsulins are subdivided into Family 2A and Family 2B based on the absence (2A) or presence (2B) of a CBD insertion within the encapsulin shell protein[22-24,39]. Some Family 2A encapsulins have recently been shown to encapsulate cysteine desulfurase cargo enzymes and are able to sequester large amounts of elemental sulfur[23,24,39]. Bioinformatic analyses have suggested that Family 2B encapsulins may be involved in terpenoid biosynthesis, with polyprenyl transferase- and terpene cyclase-encoding genes abundantly found in Family 2B encapsulin gene clusters[22,23]. No Family 2B encapsulin system has so far been experimentally characterized or confirmed as a protein compartmentalization system.

As mentioned above, Family 2B encapsulins have historically been annotated as CBD-containing transcription factors. This includes EshA, initially suggested to be involved in the formation of sporogenic hyphae in *Streptomyces griseus* and the regulation of antibiotic production in *Streptomyces coelicolor* A3(2)[20,40]. Purified EshA

93   from *S. griseus* was later shown to form large multimeric complexes[41]. Surface plasmon

94   resonance and targeted mutagenesis further suggested that the CBD of EshA interacts

95   with cyclic adenosine monophosphate (cAMP)[42]. When these early studies were being

96   carried out, encapsulins had not yet been discovered and no further characterization of

97   EshA has been reported since.

98   Here, we revise the function of EshA from *S. griseus* and establish it as a Family

99   2B encapsulin shell protein that sequesters 2-MIBS as its cargo enzyme. Using

100 cryogenic electron microscopy (cryo-EM), we report the structure of a Family 2B

101 encapsulin shell and highlight the presence of a CBD insertion domain and a previously

102 not recognized metal-binding domain within the Family 2B shell protein. We show that

103 2-MIBS cargo-loading is mediated by a flexible proline-rich N-terminal cargo-loading

104 domain (CLD). We find that 2-MIBS encapsulation does not significantly affect enzyme

105 activity or product distribution and that the presence of cAMP does similarly not

106 modulate 2-MIBS activity. Detailed binding studies further suggest that CBDs in Family

107 2B encapsulin shell proteins do not bind cAMP. Our study establishes Family 2B

108 encapsulins as protein nanocompartments with a unique shell architecture and

109 highlights that the biosynthesis of 2-MIB – one of the most prevalent terpenoid natural

110 products – takes place inside an encapsulin shell in the majority of bacteria. Our work

111 lays the groundwork for further elucidating the structure and function of Family 2B

112 encapsulin systems.

113

114 **Results**

115 **The distribution and diversity of 2-MIB biosynthetic gene clusters in**

116 **Actinobacteria**

117 In this study, we focus on the 2-MIB biosynthetic gene cluster found in *S. griseus*,

118 encoding *Sg* Enc (formerly EshA), 2-MIBS, and a methyltransferase (MT) (Fig. 1a). 2-

119 MIB biosynthetic gene clusters are ubiquitous in Actinobacteria, most notably in soil-

120 dwelling *Streptomyces,* which are renowned for their ability to produce diverse

121 secondary metabolites, including volatile terpenoids[2,6,8]. 2-MIBS catalyzes the

122 cyclization of methylated 2-methylgeranyl pyrophosphate (2-MeGPP) and was the first

123 terpene cyclase observed to utilize a non-canonical isoprenoid substrate (Fig. 1b)[13,16,17].

Further, a SAM-dependent MT, responsible for converting GPP to 2-MeGPP, is also widely conserved across 2-MIBS-encoding gene clusters[13,18,19]. Genes encoding for Family 2B encapsulin shell proteins are present in many 2-MIBS-encoding operons, including in *S. griseus*[22]. Most 2-MIBS enzymes contain two main domains, a disordered ca. 120 residue N-terminal domain rich in proline, glycine, and alanine and a C-terminal terpene cyclase domain[17]. The disordered N-terminal domains have recently been proposed to function as cargo loading domains (CLD) in Family 2B encapsulin systems[22]. Similar N-terminal CLDs are found in Family 2A cysteine desulfurase cargo proteins and have already been confirmed as essential for cargo encapsulation[22,23,39,43].

Computational searches and genome neighborhood analysis yielded 723 2-MIB biosynthetic gene clusters in Actinobacteria, with the majority (67.6%) found in the genus *Streptomyces* (Fig. 1c). Family 2B encapsulin-encoding genes were identified in 80% of the operons. The remaining 20% of organisms did not encode any Family 2B encapsulins, even outside the 2-MIB biosynthetic locus, in their genomes. These findings suggest that Family 2B encapsulin association is a common but not essential feature of bacterial 2-MIB biosynthesis.
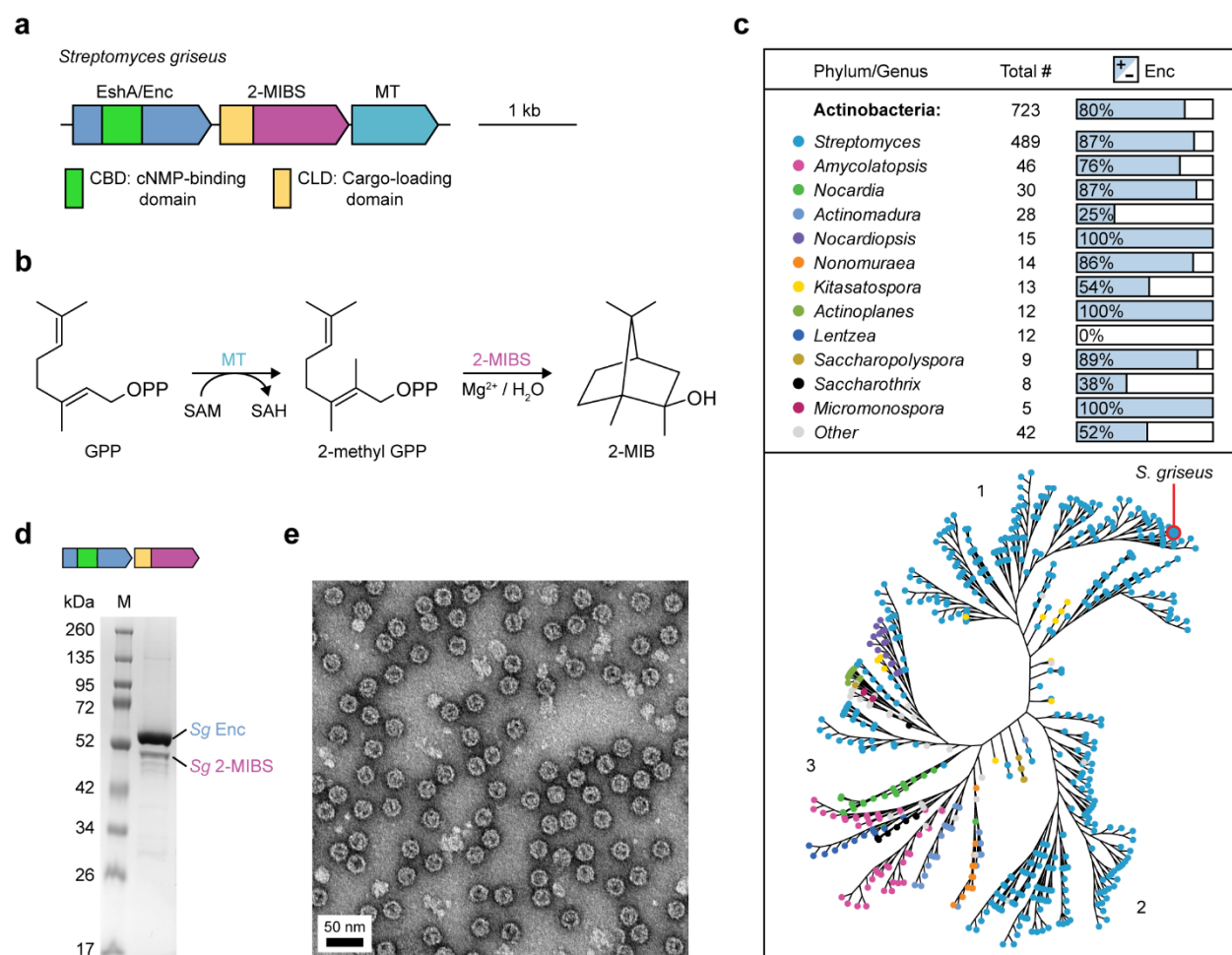
Fig. 1. 2-methylisoborneol biosynthetic gene clusters encoding a Family 2B encapsulin shell protein and its 2-MIBS enzyme cargo are widespread in Actinobacteria. a, The 2-methylisoborneol (2-MIB) biosynthetic gene cluster from *Streptomyces griseus* (*Sg*). The operon encodes a Family 2B encapsulin shell protein (EshA/Enc), 2-methylisoborneol synthase (2-MIBS), and a methyltransferase (MT). kb: kilobases. **b**, Reaction scheme for the biosynthesis of 2-MIB. Geranyl pyrophosphate (GPP) is methylated by the MT, yielding 2-methyl GPP, followed by 2-methyl GPP cyclization catalyzed by 2-MIBS resulting in 2-MIB formation. SAM: *S*-adenosyl methionine. SAH: *S*-adenosyl homocysteine. **c**, Phylogenetic distribution and encapsulin prevalence in actinobacterial 2-MIB biosynthetic gene clusters (top). Phylogenetic analysis of 723 2-MIBSs found in actinobacterial 2-MIB biosynthetic operons (bottom). Branch numbers (1 to 3) are shown. Nodes are colored by bacterial genus. **d**, SDS-PAGE analysis of purified cargo-loaded *Sg* Enc shells resulting from

154   heterologous expression in *Streptomyces coelicolor* A3(2)*. **e**, Negative stain

155   transmission electron microscopy (TEM) images of purified 2-MIBS-loaded *Sg* Enc.

156

157   Phylogenetic analysis of actinobacterial 2-MIBSs yielded three main branches

158   reflecting sequence similarity (Fig. 1c). Branches one and two contained primarily

159   *Streptomyces*-encoded 2-MIBSs*. Sequence alignments of these 2-MIBSs showed poor

160   sequence conservation within CLDs (branch 1 CLDs: 22.5% average pairwise identity;

161   branch 2 CLDs: 41.3% average pairwise identity), but high conservation of the terpen

162   cyclase (TC) domains (branch 1 TCs: 79.9% average pairwise identity; branch 2 TCs:

163   80.8% average pairwise identity). Branch three contained sequences from 36 genera

164   including *Streptomyces, Actinomadura, Nocardia,* and *Amycolatopsis.* Consequently,

165   both the CLD and TC domains within branch three were significantly more divergent

166   than sequences within branches one and two (branch 3 CLDs: 14.9% average pairwise

167   identity; branch 3 TCs: 48.7% average pairwise identity). Sequence similarity network

168   analysis yielded similar results, with two large clusters corresponding to branches one

169   and two, and branch three broken into several smaller clusters (Supplementary Fig. 1, a

170   and b). Sequence alignments of the 2-MIBSs within each cluster revealed short,

171   conserved motifs within CLDs, with most clusters containing a GPxGLGT consensus

172   motif (Supplementary Fig. 1c), which was also highlighted in previous bioinformatic

173   analyses of putative Family 2 encapsulin cargos[22,23]. Only cluster 2 exhibited a

174   divergent CLD consensus motif (LPGPP) instead (Supplementary Fig. 1c). Further

175   analysis of the CLDs within each cluster showed significant variability in length, with

176   cluster 2 containing the shortest CLDs (31 ± 12 residues), and many CLDs exceeding

177   200 residues in length and containing multiple repeats of the abovementioned

178   consensus motifs (Supplementary Fig. 1c).

179

180   **EshA is a Family 2B encapsulin with 2-MIBS as its cargo protein**

181   *Sg* Enc (formerly EshA) has previously been annotated as a CBD-containing

182   transcription factor[41] and has more recently been suggested to represent a Family 2B

183   encapsulin shell protein[22]. To investigate if *Sg* Enc is indeed able to form cargo-loaded

184   protein nanocompartments, we heterologously expressed a two-gene operon encoding

Sg Enc and *Sg* 2-MIBS (*Sg* Enc_2-MIBS) in *Streptomyces coelicolor* A3(2). After ammonium sulfate precipitation and size-exclusion chromatography, clear co-purification of both proteins could be observed (Fig. 1d). Negative-stain transmission electron microscopy (TEM) confirmed the formation of protein shells with a diameter of ca. 28 nm, consistent with a T=1 icosahedral assembly as seen in other T=1 encapsulin shells (Fig. 1e)[23,24,30,39]. Additional densities in the interior of many shells could be observed, likely representing encapsulated 2-MIBS. Regularly spaced protrusions localized on the surface of shells could also be visualized via TEM, which may represent externally displayed CBDs.

Single particle cryo-EM analysis of cargo-loaded *Sg* Enc reveals a unique shell architecture

To investigate the molecular structure and organization of this Family 2B encapsulin system, we carried out single-particle cryo-EM analysis of both the cargo-loaded shell (*Sg* Enc_2-MIBS) and the empty shell without cargo (*Sg* Enc). Icosahedral (I) refinement of *Sg* Enc_2-MIBS and *Sg* Enc yielded cryo-EM maps with global resolutions of 3.02 Å and 2.71 Å (Supplementary Fig. 2 and 3). Consistent with our negative stain TEM analysis, the encapsulin shell was found to possess a diameter of 29.4 nm and consist of 60 protomers of the *Sg* Enc shell protein arranged with T=1 icosahedral symmetry (Fig. 2a). CBDs were clearly visible on the exterior of the shell, albeit at significantly lower local resolution than the HK97-like shell-forming domain of *Sg* Enc (Supplementary Fig. 2d and 3d), arranged around all two-fold axes of symmetry (Fig. 2a). Additional non-*Sg* Enc density was observed lining the shell interior at the three-fold axes of symmetry for the cargo-loaded system (*Sg* Enc_2-MIBS), reminiscent of CLD density previously observed in cargo-loaded *Synechococcus elongatus* and *Acinetobacter baumannii* Family 2A encapsulin shells (Fig 2a)[23,39]. This extra density was not observed in the empty *Sg* Enc shell, further suggesting that it represents part of the 2-MIBS CLD. Apart from the surface-exposed CBDs, the overall architecture of the HK97-like shell is similar to characterized Family 2A shells, including the characteristic turret-like vertices formed at the five-fold symmetry axes (Fig. 2b and Supplementary

215 Fig. 4a,b). In the *Sg* Enc shell, the pentameric vertices are noticeably more flattened
216 compared to
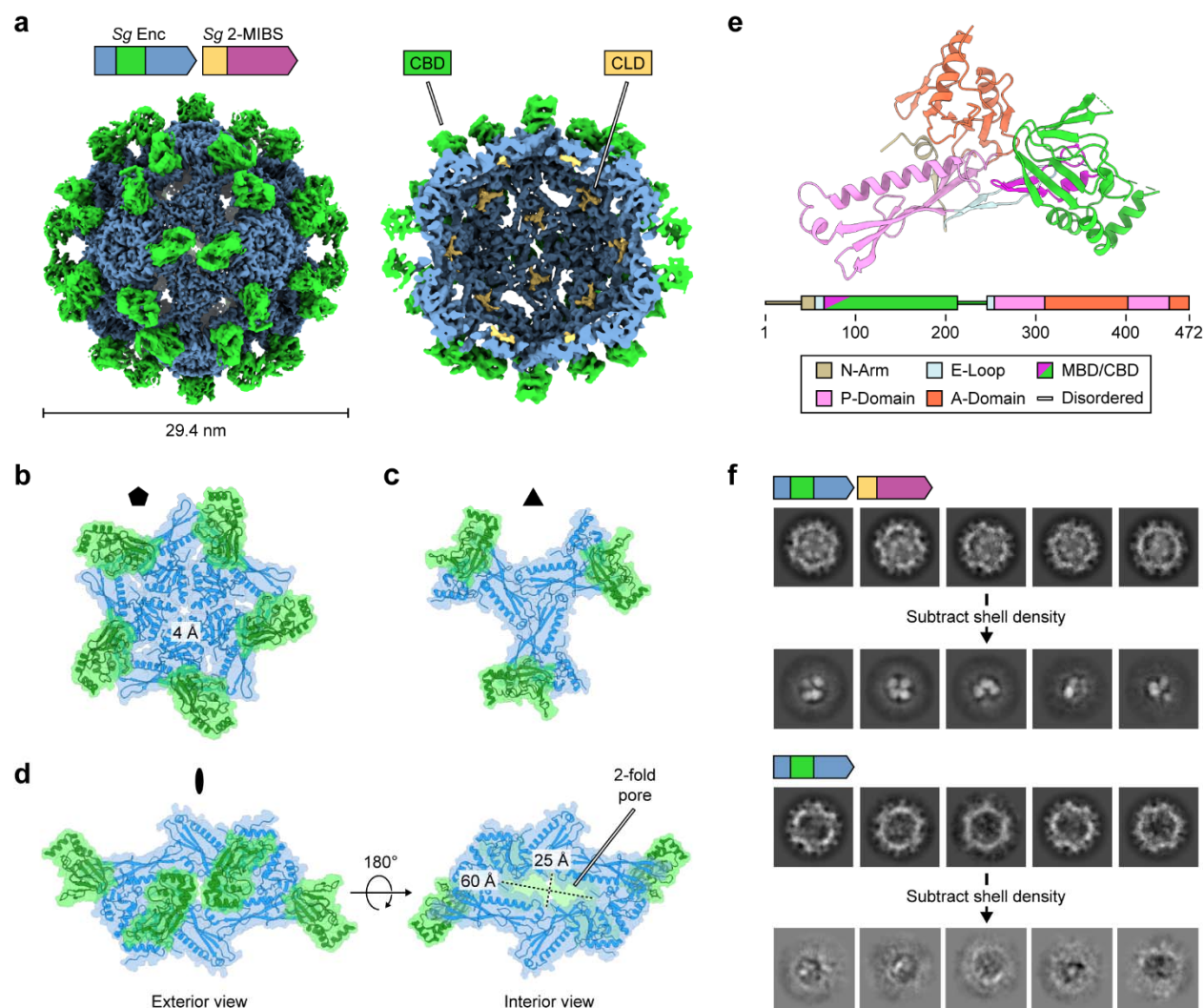
Fig. 2. Cryo-EM analysis of 2-MIBS-loaded *Sg* Enc. a, Cryo-EM map of the cargo-
loaded *Sg* Enc shell viewed from the exterior down the two-fold axis of symmetry (left).
Cut-away view shown on the right. Density corresponding to the HK97-like shell-forming
domain is shown in blue. Externally displayed CBDs are shown in green. Internalized 2-
MIBS CLD densities shown in yellow. **b**, Exterior view of a *Sg* Enc pentamer down the
five-fold axis of symmetry highlighting the five-fold pore. **c**, Exterior view of a *Sg* Enc
trimer down the three-fold axis of symmetry. **d**, Exterior (left) and interior (right) view
down the two-fold axis of symmetry highlighting the large, elongated two-fold pore.
When viewed from the exterior, the CBDs partially occlude the two-fold pore. **e**,
Structural model of the *Sg* Enc shell protomer with characteristic HK97-like domain

228 elements and the CBD highlighted with different colors. MBD: metal-binding domain. **f**,

229 2D class averages of shell-subtracted 2-MIBS-loaded *Sg* Enc particles (top). 2D class

230 averages of shell-subtracted empty *Sg* Enc particles (bottom). While discrete

231 internalized cargo puncta are visible for the cargo-loaded sample, no puncta can be

232 seen in empty shells.

233

234 Family 2A shells due to a shorter C-terminal extension in the *Sg* Enc protomer. Small

235 pores, ca. 4 Å and 2 Å, can be found at the five- and three-fold axes of symmetry,

236 respectively, comparable in size to pores found in Family 2A encapsulin shells (Fig. 2b,c

237 and Supplementary Fig. 4c). Along all two-fold symmetry axes, the *Sg* Enc shell exhibits

238 large, elongated pores, approximately 60 x 25 Å in dimension (Fig. 2d), located beneath

239 the externally displayed CBDs. No encapsulin shell has previously been observed to

240 possess such large pores. In particular, pores at the two-fold axes of symmetry are

241 usually small or non-existent in Family 1 or Family 2A encapsulin shells. Specifically, in

242 the *S. elongatus* and *A. baumannii* Family 2A shells, the N-arm of the HK97-like domain

243 is threaded along the two-fold axes, resulting in essentially closed two-fold pores

244 (Supplementary Fig. 4d)[23,39]. We were able to confidently model most of the *Sg* Enc

245 HK97-like domain, which maintained a nearly identical fold compared to Family 2A

246 encapsulin protomers (Fig. 2e and Supplementary Fig. 4e). However, the first 39

247 residues of the *Sg* Enc N-arm are unresolved, resulting in the observed elongated open

248 two-fold pores (Supplementary Fig. 5a). Closer analysis of our cryo-EM maps revealed

249 a segment of poorly resolved density positioned adjacent to the CBDs near the N-

250 termini (Supplementary Fig. 5b). This density may correspond to parts of the

251 unmodelled N-arm segment of *Sg* Enc or to an unresolved loop connecting the CBDs to

252 the E-loops of the HK97-like domains (Supplementary Fig. 5a).

253   The shells resulting from both I and C1 refinements of cargo-loaded *Sg* Enc_2-

254 MIBS exhibited a significant amount of internal noise which we attribute to encapsulated

255 but fairly mobile 2-MIBS cargo. Shell density subtraction followed by 2D classification of

256 shell-subtracted particles produced classes with three to four distinct puncta in 99% of

257 particles. This observation suggests that approximately three to four dimers of 2-MIBS,

258 previously shown to form dimers[17], are contained within each shell of *Sg* Enc_2-MIBS.

259 In contrast, no internal densities were observed in shell-subtracted particles of the *Sg*

260 Enc shell without cargo (Supplementary Fig. 2 and Supplementary Fig. 6). Contrary to

261 the *A. baumannii* Family 2A cargo-loaded encapsulin, where low resolution lobes of

262 density corresponding to cysteine desulfurase cargo could be clearly observed in the

263 interior of the shell[39], none of our reconstructions were able to resolve the interior

264 orientations of the 2-MIBS cargo. This suggests that 2-MIBS may not be as spatially

265 confined as the cysteine desulfurase cargo in Family 2A encapsulins, exhibiting

266 substantial flexibility instead.

267

268 **The *Sg* Enc shell protein contains unique CBD and metal-binding insertion**

269 **domains**

270 While we were able to confidently dock an AlphaFold model of the CBD into the

271 cryo-EM density resulting from I refinement of the *Sg* Enc shell, the CBD densities

272 exhibited noticeably worse resolution than the HK97-like domain of the shell. Masked

273 3D classification of the CBD densities surrounding the two-fold axis of symmetry,

274 followed by homogeneous C1 reconstruction of the best class, significantly improved

275 the quality of the CBD density (Supplementary Fig. 7 and 8), now enabling confident

276 CBD model building. Similar to other CBDs, the *Sg* Enc CBD is composed of an eight

277 stranded β-barrel domain and an α-helical subdomain composed of three short α-

278 helices (Fig. 3a). This overall fold, and the β-barrel in particular, is strongly conserved as

279 the primary cAMP-binding site in characterized cAMP-binding proteins such as cAMP-

280 dependent protein kinase A, ion-gated channels, and cAMP-responsive transcription

281 factors, most notably the *E. coli* catabolite activator protein (*Ec* CAP) (Fig. 3b)[44,45]. In

282 many cAMP-regulated proteins, binding of cAMP to the β-barrel domain results in a

283 conformational change that is propagated throughout the protein or complex to regulate

284 the activity of a distal functional domain. In the dimeric *Ec* CAP for example, cAMP

285 binding induces a conformational change that stabilizes a central coiled-coil helical

286 domain and re-orients the DNA-binding domains to enable stable interaction with

287 double-stranded DNA[46,47]. Instead of an extended α-helix able to form a coiled-coil, as

288 found in *Ec* CAP, the C-terminus of monomeric *Sg* Enc CBD contains a long disordered

289 loop that is unresolved in our cryo-EM density (Supplementary Fig. 5a). This loop acts

290 as a linker between the CBD and the HK97-like domain of *Sg* Enc and may also play a
291 role in ligand binding or modulating the conformation of the CBD in analogy to the
292 extended α-helices observed in other CBD containing proteins.
293
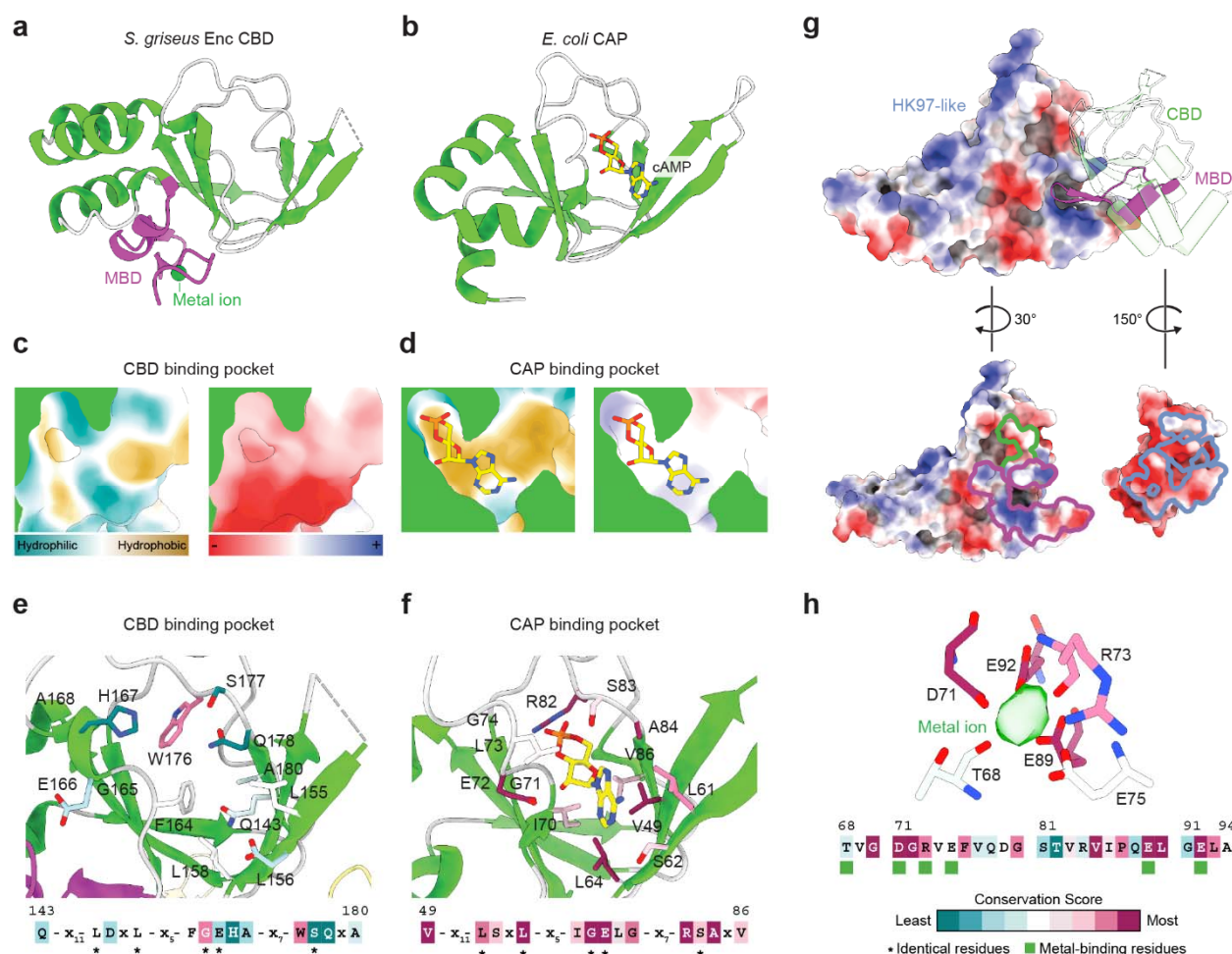
Fig. 3. Structural analysis of the *Sg* Enc CBD. a, An overview of the structure of the *Sg* Enc CBD. The metal-binding subdomain (MDB) is shown in purple and the remainder of the CBD is shown in green and light grey. **b**, The *E. coli* (*Ec*) CAP CBD (residues 1 to 116) bound to cAMP (PDB ID: 1G6N)[21] colored analogously to panel **a**. cAMP is shown as yellow sticks. **c**, Surface representations of the *Sg* Enc CBD binding pocket colored according to hydrophobicity (left) and surface charge (right). **d**, Surface representations of the *Ec* CAP CBD binding pocket colored according to hydrophobicity (left) and surface charge (right). Bound cAMP is shown as yellow sticks. **e**, Key *Sg* Enc CBD residues forming the CBD binding pocket. Individual residues are colored by

304 conservation as predicted by ConSurf analysis[48-50]. ConSurf conservation score key

305 shown in panel **h**. **f**, Key *Ec* CAP CBD residues in the CBD binding pocket involved in

306 cAMP binding. Individual residues are colored by conservation as predicted by ConSurf

307 analysis. **g**, The CBD interacts with the HK97-domain of the *Sg* Enc protomer

308 predominantly via electrostatic interactions mediated by the MBD. The interaction

309 interfaces are outlined on the HK97-domain surface (left) for the CBD (green) and the

310 MBD (purple). Conversely, interaction interfaces of the HK97-domain with the CBD and

311 MBD are outlined in blue (right). **h**, The MBD is a highly conserved feature of Family 2B

312 encapsulin shell proteins. Side chains and backbone atoms coordinating the metal ion

313 are colored according to sequence conservation as predicted by ConSurf analysis.

314 Extracted cryo-EM metal ion density is shown in green.

315

316   No ligands were observed bound to the CBDs for either the cargo-loaded *Sg*

317 Enc_2-MIBS or shell-only *Sg* Enc CBD samples, despite the samples not having been

318 subjected to dialysis or ion exchange chromatography. Closer analysis of the β-barrel

319 binding pocket shows that in *Sg* Enc CBD, a dramatically different hydrophobicity and

320 surface charge pattern can be observed than in *Ec* CAP (Fig. 3, c and d). Furthermore,

321 the usually highly conserved cAMP-interacting residues found in CBDs, including *Ec*

322 CAP, are poorly conserved in *Sg* Enc CBD (Fig. 3, d and e). In particular, a highly

323 conserved arginine residue located at the back of the binding pocket – R82 in *Ec* CAP –

324 is replaced with a tryptophan residue – W176 in *Sg* Enc – in essentially all Family 2B

325 encapsulin CBDs. In canonical cAMP-binding CBDs, the conserved arginine forms part

326 of the phosphate binding cassette (PBC), essential for cAMP binding, and directly

327 coordinates the cyclic phosphate of cAMP[44,45]. Mutations to R82 of *Ec* CAP generally

328 result in a dramatic loss of cAMP binding[51-53]. CBD-like domains lacking this arginine

329 are not responsive to cAMP[45], and have been shown to instead bind a diverse array of

330 other ligands such as heme in the CO-responsive transcription factor CooA[54],

331 chlorophenolacetic acid in the transcriptional regulator CprK[55], 2-oxoglutarate in the

332 transcription factor NtcA[56], and even peptides in the case of KCNH voltage-gated

333 potassium ion channels[57-59]. Because *Sg* Enc, and nearly all identified Family 2B

334   encapsulins, lack this conserved arginine, it is possible that Family 2B encapsulin-

335   associated CBDs may not bind cAMP but alternative ligands instead.

336   Closer analysis of the N-terminal section of the *Sg* Enc CBD revealed a 30

337   residue motif, which we will refer to as the metal-binding domain (MBD), composed of

338   two anti-parallel β-strands and a short α-helix that form a metal binding site, not found in

339   other characterized CBDs (Fig. 3a). This motif appears to be unique to Family 2B

340   encapsulins and similar motifs could not be readily identified in known protein structures

341   via structural homology searches using DALI, HHPRED, or FoldSeek[60-63]. The

342   negatively charged MBD forms an interface with a positively charged surface patch on

343   the exterior of the HK97-like *Sg* Enc domain, which may function to stabilize the

344   orientation of the CBD (Fig. 3g). The CBD, in comparison, maintains a much smaller

345   interface with the HK97-like domain, primarily mediated by hydrophobic interactions. An

346   unknown metal ion, likely Mg^{2+} or Ca^{2+} based on the observed coordination pattern, is

347   bound by conserved residues, primarily glutamates and aspartates, within the metal

348   binding site of the MBD (Fig. 3, a and h), suggesting that metal ion binding is a

349   widespread feature of the MBD throughout Family 2B encapsulins.

350

351   **2-MIBS encapsulation is mediated by a disordered cargo-loading domain**

352   As mentioned above, additional cryo-EM density, likely belonging to a fragment

353   of the 2-MIBS CLD, was observed in the cargo-loaded *Sg* Enc shell interior located at

354   the three-fold axes of symmetry. In I refinements, these additional densities appeared

355   symmetry-averaged and could not be used for model building due to their limited

356   resolution and quality. Asymmetric refinements yielded CLD densities which appeared

357   to partially break the three-fold symmetry (Fig. 4a). Masked 3D classification around the

358   three-fold axis resulted in several classes that also appeared to further break the three-

359   fold symmetry. Still, the quality of the obtained cryo-EM densities was poor, and we

360   were unable to build atomic models or assign amino acid sequences to these extra

361   internal densities (Supplementary Fig. 10). As expected, no CLD density was observed

362   when performing the same analysis on the non-cargo-loaded *Sg* Enc shell

363   (Supplementary Fig. 10), confirming that these densities represent part of the 2-MIBS

364   CLD. The conserved hydrophobic residues F430, Y441, L285, and L264 are positioned

365 below the observed CLD densities, and are nearly identical to the corresponding

366 residues in the *A. baumannii* and *S. elongatus* Familiy 2A encapsulin three-fold CLD

367 binding sites (Fig. 4b)[23,39].

368

369
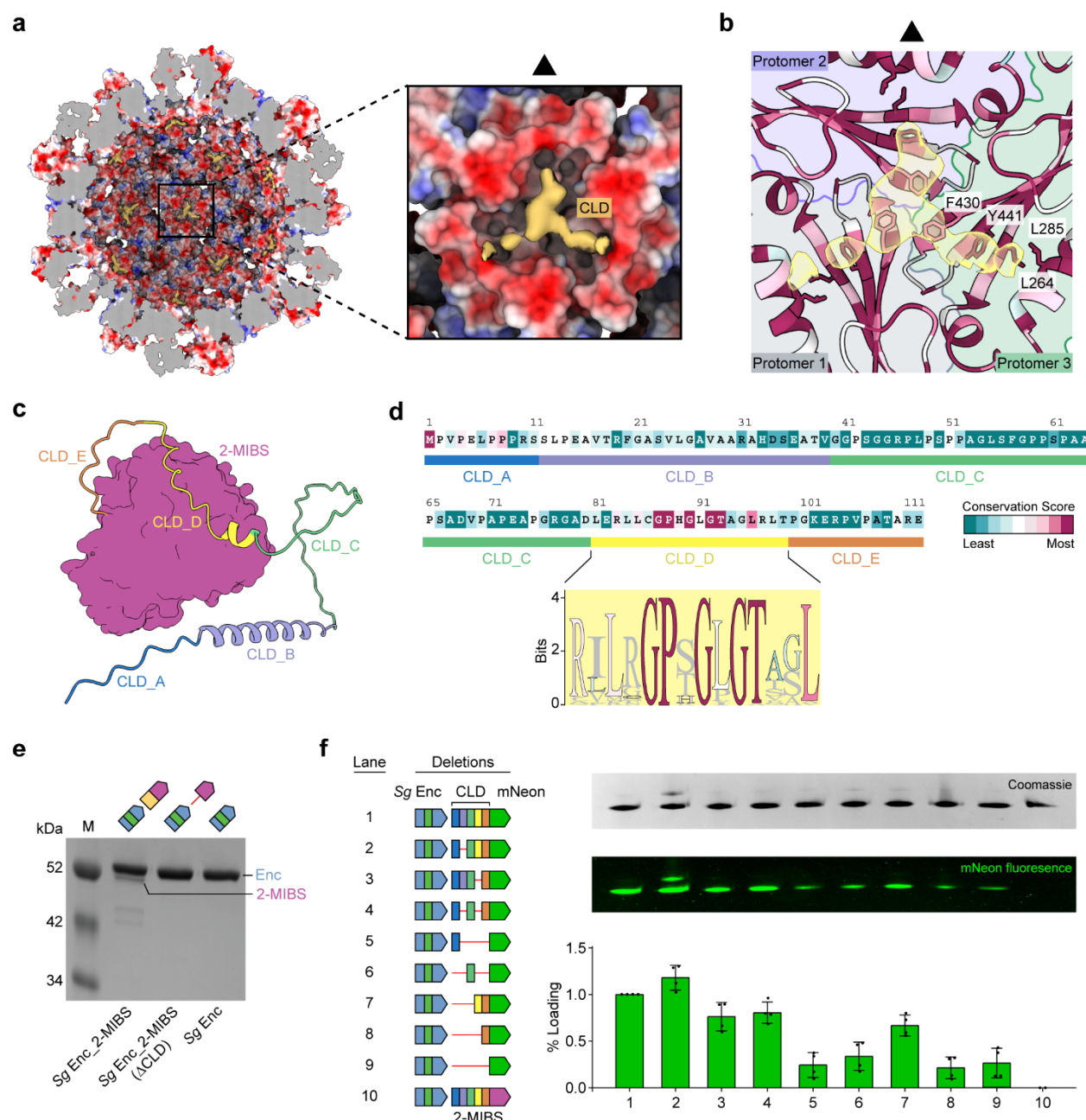
370 **Fig. 4. CLD-mediated cargo-loading of *Sg* Enc. a**, Cut-way view of an *Sg* Enc shell

371 model shown in electrostatic surface representation with extracted cryo-EM CLD

372 densities obtained via asymmetric C1 refinement shown in yellow (left). Zoom-in

highlighting CLD density bound to the interior of the shell at the three-fold axis of symmetry (right). **b**, The interior of *Sg* Enc surrounding the three-fold axis of symmetry shows strong sequence conservation throughout Family 2B encapsulins. Residues are colored based on ConSurf conservation analysis. The three protomers at the three-fold symmetry axis are highlighted by differently colored backgrounds. Conserved hydrophobic residues F430, Y441, L285, and L264 are shown as stick models and are located directly below the observed CLD density (transparent yellow). **c**, AlphaFold model of *Sg* 2-MIBS highlighting predicted secondary structure elements within the disordered N-terminal CLD. Five CLD elements – CLD_A to CLD_E – are shown in different colors. The structured TC domain of *Sg* 2-MIBS is shown in surface representation (purple). **d**, *Sg* 2-MIB CLD sequence colored by ConSurf conservation score. A sequence logo for the conserved motif within CLD_D is shown. **e**, SDS-PAGE analysis of purified *Sg* Enc shells. Deletion of the CLD abolishes cargo loading of *Sg* 2-MIBS when co-expressed with *Sg* Enc. *Sg* Enc_2-MIBS: 2-MIBS-loaded *Sg* Enc. ΔCLD: truncation mutant missing all of the CLD. **f**, Systematic deletion of the five identified CLD regions – CLD_A to CLD_E (left). A co-expression system of *Sg* Enc and CLD-fused mNeonGreen (mNeon), employed as a reporter, was used. Native PAGE analysis of purified truncation mutants shown on the right. Coomassie protein stain (right, top) and mNeonGreen fluorescence (right, middle) are shown. Normalized loading percentages based on gel densitometry and fluorescence analysis are shown (right, bottom). Data are shown as mean values, with error bars representing the standard deviation of four independent experiments.

The CLDs of Family 2 encapsulins are extended and flexible domains rich in proline, glycine, and alanine[22,23,39,43]. Inspection of *Sg* 2-MIBS AlphaFold models suggests two CLD stretches with helical propensity which we used as markers to subdivide the *Sg* 2-MIBS CLD into five regions – CLD_A to CLD_E – for mutational and cargo-loading analysis (Fig. 4c) [64]. CLD_D contained the conserved consensus motif GPxGxGTxxL, mentioned above, which is the only conserved section of the *Sg* 2-MIBS CLD (Fig. 4d)[22,23]. Removal of the entire CLD from *Sg* 2-MIBS completely abolished cargo loading (Fig. 4e). To better understand the specific roles of the five CLD regions

404   (CLD_A to CLD_E), we heterologously expressed a set of two-gene operons consisting

405   of the *Sg* Enc shell protein and different variants of a CLD-mNeonGreen fusion protein

406   containing different combinations of the five CLD regions (Fig. 4f). Analysis of purified

407   *Sg* Enc_CLD-mNeonGreen variants via native PAGE allowed us to quantify cargo

408   loading by comparing CLD-mNeonGreen fluorescence to the total protein content

409   measured by Coomassie staining (Fig. 4f and Supplementary Fig. 11). Similar

410   approaches have previously been successfully utilized to examine cargo loading in

411   Family 1 and Family 2 encapsulin systems[23,65]. Removal of CLD_B resulted in slightly

412   improved cargo-loading, while removal of all other CLD regions reduced cargo loading.

413   Most notably, we observed that CLD_D, which contained the GPxGxGTxxL consensus

414   motif, was not required for cargo loading, but did substantially improve cargo loading

415   when present. Together with the observed, relatively low quality CLD cryo-EM densities,

416   these results suggest that there are likely multiple redundant binding modes of CLD and

417   its binding site on the interior of the shell. The CLD of *Sg* 2-MIBS is characterized by

418   repetitive glycine-, alanine-, and proline-rich motifs that may contribute cooperatively to

419   cargo-loading. This amino acid composition may enable multiple heterogeneous modes

420   of binding that are not entirely sequence specific, but instead rely on physicochemical

421   properties like flexibility and hydrophobicity for mediating CLD binding.

422

Sg Enc is not a cAMP-binding protein

424   *Sg* Enc has previously been suggested to be involved in modulating antibiotic

425   production in *Streptomyces coelicolor* A3(2) through binding of cAMP which is a well-

426   known second messenger important for regulating secondary metabolism in

427   *Streptomyces*[42]. As we now recognize *Sg* Enc to be a 2-MIBS-loaded encapsulin

428   nanocompartment, we suspected that if *Sg* Enc is able to bind cAMP via its CBDs, this

429   may result in a conformational change of the two-fold pores which in turn could control

430   2-MIB production by potentially restricting substrate flux across the encapsulin shell. To

431   test this hypothesis, we monitored the activity of encapsulated and free 2-MIBS through

432   a coupled assay with the GPP-methylating *Sg* MT in the presence or absence of

433   cAMP[66-68]. We observed sigmoidal saturation curves (Supplementary Fig. 12), indicative

434   of cooperative substrate binding, as well as substrate inhibition in conditions exceeding

435 90 µM GPP, features also observed for other terpene cyclases and prenyltransferases[66].

436 Apparent k_{cat}/K_m values were consistent between free and encapsulated *Sg* 2-MIBS and

437
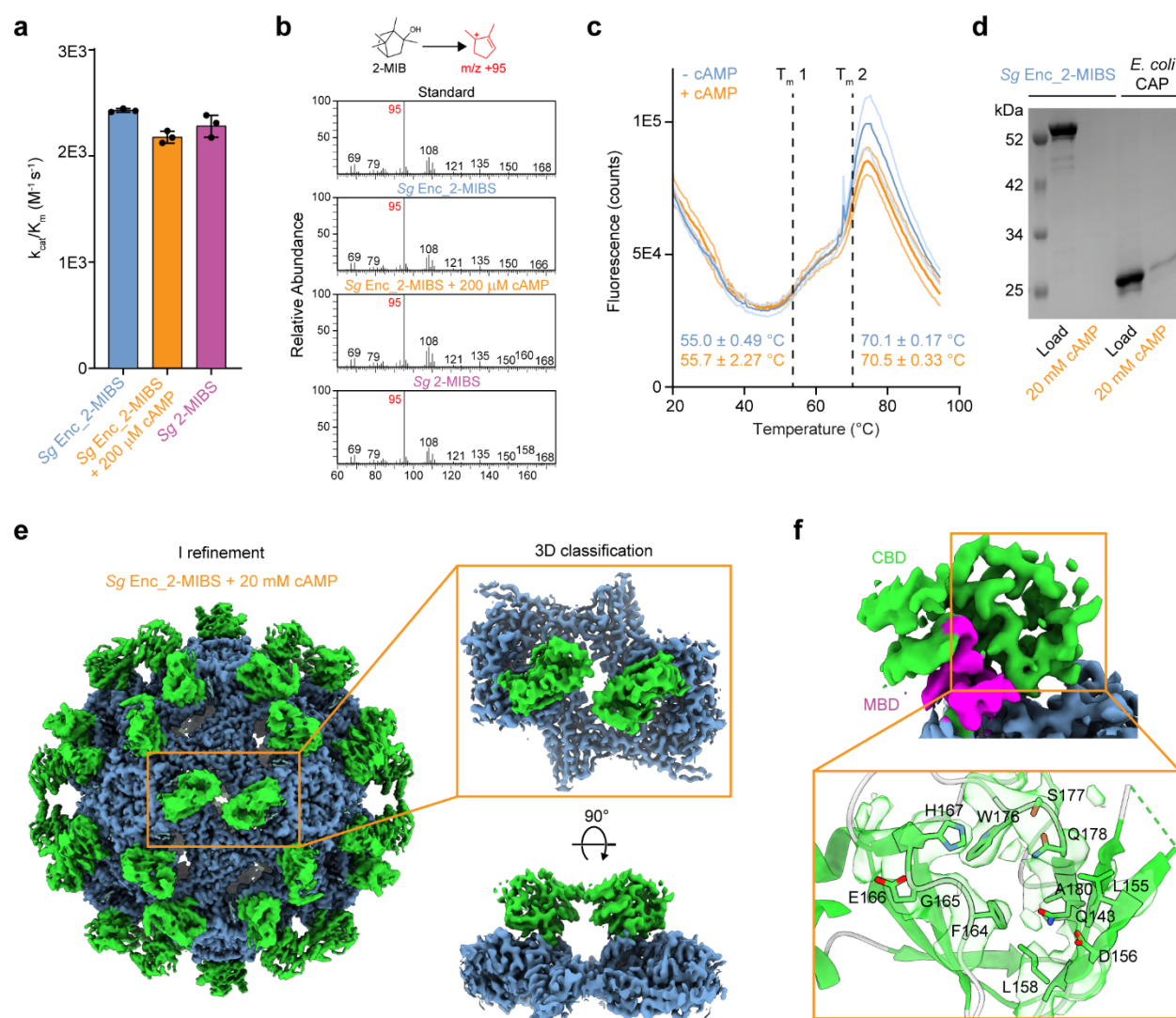
Fig. 5. The effect of cAMP on *Sg* 2-MIBS activity and cAMP binding to *Sg* Enc CBD. a, *In vitro* activity assays of 2-MIBS-loaded *Sg* Enc in the absence and presence of cAMP. Free 2-MIBS activity is shown for reference. k_{cat}/K_m : catalytic efficiency of 2-MIBS. Data are shown as mean values, with error bars representing the standard deviation of three independent experiments. **b,** GC-MS analysis of 2-MIBS activity assays. All assays yielded 2-MIB as the main product, with its diagnostic mass fragment (m/z +95) identified as the most abundant ion. A 2-MIB standard is shown for reference (top). **c,** Differential scanning fluorimetry (DSF) thermal melts of 2-MIBS-loaded Sg Enc in the absence or presence of cAMP. T_m: melting temperature. **d,** 2-MIBS-loaded Sg

447   Enc binding to cAMP-modified agarose resin. A representative SDS-PAGE gel is shown

448   demonstrating that 2-MIBS-loaded *Sg* Enc is unable to bind to cAMP-modified agarose

449   resin. The positive control *Ec* CAP is able to bind to the resin. Load: sample loaded onto

450   the resin. 20 mM cAMP (orange): elution from the resin with 20 mM cAMP. **e,** Cryo-EM

451   analysis of *Sg* Enc in the presence of cAMP. No significant conformational changes of

452   the CBDs or the HK97-domain could be observed when compared to cryo-EM maps

453   obtained in the absence of cAMP. Improved CBD density was obtained via 3D

454   classification followed C1 homogeneous refinement, but no conformational changes

455   could be observed. **f**, Cryo-EM density of the β-barrel CBD binding site is shown (top).

456   No additional density could be observed in the presence of cAMP. The metal-binding

457   subdomain (MBD) is shown in purple. Structural model of the binding site with binding

458   site residues shown as stick (bottom).

459

460   the presence of 500 μM cAMP had no apparent influence on 2-MIBS activity

461   (Supplementary Fig. 12). We suspect that the large two-fold pores of the *Sg* Enc shell

462   allow rapid diffusion of substrates and products across the encapsulin shell causing

463   encapsulation to have little influence on enzyme activity. GC-MS analysis further

464   revealed that the presence of cAMP did not affect the product profile of *Sg* 2-MIBS, with

465   2-MIB as the only detectable product in our assays (Fig. 5b and Supplementary Fig.

466   13). 2-methylenebornane has been observed as a minor product of 2-MIBS in *in vitro*

467   assays[17], yet we did not detect it in any of our assays. These results show that cAMP

468   has no effect on the activity or product profile of encapsulated *Sg* 2-MIBS.

469   We next sought to directly investigate cAMP binding to *Sg* Enc CBDs. Differential

470   scanning fluorimetry (DSF), a method commonly used to measure ligand binding[69], was

471   initially carried out and identified two thermal transitions for cargo-loaded *Sg* Enc shells

472   at melting temperatures (T_m) of 55.0 ± 0.49°C and 70.1 ± 0.17°C (Fig. 5c). Addition of

473   cAMP yielded nearly identical thermal transitions, with T_m values of 55.7 ± 2.27°C and

474   70.5 ± 0.33°C, highlighting that cAMP addition does not result in any observable

475   stabilization of the *Sg* Enc shell. Using an orthogonal cAMP binding assay, we next

476   employed a cAMP-modified agarose resin, another common strategy for purifying and

477   identifying cAMP-binding proteins, to test cargo-loaded *Sg* Enc cAMP binding

478   capacity[70-73]. No binding to cAMP-agarose resin could be observed for *Sg* Enc shells

479   while *Ec* CAP – a known cAMP-binding protein – exhibited clear resin binding (Fig. 5d).

480   Subsequently, cryo-EM analysis of 2-MIBS-loaded *Sg* Enc was carried out in the

481   presence of an excess of 20 mM cAMP to potentially capture a cAMP-bound state

482   (Supplementary Fig. 14). Yet, we observed no apparent conformational changes in cryo-

483   EM maps resulting from I refinements when compared to maps obtained in the absence

484   of cAMP (Fig. 5e). Masked 3D classification and homogeneous C1 refinements of CBDs

485   resulted in cryo-EM maps with improved CBD density but showed no significant

486   conformational changes compared to previous maps (Fig. 5f). Closer analysis of the

487   CBD revealed an empty binding pocket, identical to the binding pockets observed

488   without any cAMP present. In sum, these orthogonal analyses demonstrate that cAMP

489   is likely not a native ligand of *Sg* Enc CBDs and that addition of cAMP has no

490   observable influence on shell structure or CBD conformation.

491

492   **Discussion**

493   In this study, we demonstrate that *Sg* Enc, formerly known as EshA, is not a

494   transcription factor or cAMP-binding protein as previously suggested[20,40-42], but is

495   instead a Family 2B encapsulin shell protein that functions to sequester the terpene

496   cyclase 2-MIBS. Our characterization of *Sg* Enc experimentally confirms that encapsulin

497   nanocompartments are directly involved in secondary metabolite biosynthesis and that

498   the ubiquiteous odorant 2-MIB is biosynthesized inside a protein shell in many soil

499   bacteria. Family 2B encapsulins and their respective gene clusters are widespread

500   among   Actinobacteria,   Proteobacteria,   and   Cyanobacteria   and,   based   on

501   computational searches, have been hypothesized to encapsulate a variety of enzymes

502   associated with secondary metabolism, including diverse terpene cyclases, polyprenyl

503   transferases,   and   xylose-modifying   enzymes[22,23].   Additionally,   many   Family   2B

504   encapsulin-encoding gene clusters encode multiple putative cargo proteins, potentially

505   resulting in protein shells loaded with multiple types of cargo[22]. Further, some Family 2B

506   encapsulin operons seem to encode two distinct shell protein copies, suggesting that

507   these systems may be able to assemble into two-component encapsulin shells[22].

508   Our structural analysis of the 2-MIBS-loaded *Sg* Enc shell reveals that the *Sg*
509   Enc shell protein exhibits an HK97-like fold similar to Family 2A encapsulins[23,39], but
510   contains large and elongated two-fold pores not seen in any other encapsulin shell.
511   These two-fold pores in the *Sg* Enc shell are potentially necessary for substrate and
512   product flux across the shell given the relatively large size of the 2-MIBS substrate – 2-
513   methylGPP – and product – 2-MIB. However, large two-fold pores may not be a
514   feature of all Family 2B encapsulin shells as approximately 9% of Family 2B systems
515   are associated with cysteine desulfurase cargos that utilize the small substrate L-
516   cysteine[22]. In characterized Family 2A encapsulins, lacking CBDs, that encapsulate
517   cysteine desulfurases, the two-fold pores are essentially closed by the structured N-
518   arms of two neighboring shell protein protomers[23,39]. N-arms of Family 2B encapsulins
519   display high sequence and length variability, which may allow different alternate
520   orientations of the N-arms and consequently different two-fold pore states which may be
521   static or dynamic[22]. Dynamic five-fold pores have recently been reported in Family 1
522   encapsulins[74-76]. Furthermore, as mentioned above, some Family 2B gene clusters
523   encode for two distinct Family 2B encapsulin shell proteins, which may form
524   heterogeneous two-component shells with a mixture of open and closed two-fold pores,
525   depending on the respective N-arm states of the constituent shell protomers.

526   Our analysis of the *Sg* 2-MIBS CLD demonstrates its importance for cargo
527   loading, as predicted by previous bioinformatic analyses and related experimental
528   evidence from Family 2A encapsulin systems[22,23,39]. Despite considerable sequence
529   variability, 2-MIBS CLDs are rich in glycine, alanine, and proline, a feature shared with
530   Family 2A CLDs[22,23,39]. Mutational analysis of the *Sg* 2-MIBS CLD suggests that
531   repetitive stretches of glycine, alanine, and proline may provide redundant interaction
532   motifs that mediate CLD-shell binding. Our cryo-EM analysis revealed interactions
533   between the *Sg* 2-MIBS CLD and the interior surface surrounding the three-fold axis of
534   the *Sg* Enc shell. The observed variability and limited resolution of these CLD densities
535   may be a result of multiple heterogenous modes of binding, resulting from redundant
536   CLD interaction motifs[39]. Given the high sequence and length variability observed
537   throughout 2-MIBS CLDs, other 2-MIBSs with longer CLDs may utilize additional
538   binding interfaces extending beyond the three-fold axis. Inversely, 2-MIBSs found in

539 branch 2 of our phylogenetic tree (Fig. 1c) may contain minimal binding motifs with
540 more homogeneous three-fold binding modes.

541   The most unique feature of Family 2B encapsulins is the presence of an insertion
542 domain within the encapsulin shell protein, annotated as a CBD. Our structural analysis
543 of the *Sg* Enc CBD reveals that it adopts a fold highly similar to other characterized
544 cAMP-binding CBDs[44]. In addition, we identified a unique and conserved metal-binding
545 subdomain (MBD) as part of the CBD which likely contributes to stabilizing CBD
546 orientation through an extensive interface with the HK97-like domain of the Family 2B
547 encapsulin shell protomer. We were unable to observe any conformational changes of
548 the CBDs or HK97-like domain of *Sg* Enc, regardless of cargo-loading state or the
549 presence of cAMP. Diverse ligand binding and cryo-EM experiments suggest that *Sg*
550 Enc CBDs do not bind cAMP. We hypothesize that this may be partially caused by the
551 presence of a conserved tryptophan (W176) – replacing a conserved arginine found in
552 true cAMP-binding proteins – in the putative binding pocket, a feature widespread
553 among Family 2B encapsulin CBDs. While the conserved arginine of confirmed cAMP-
554 binding proteins is crucial for binding the cyclic phosphate of cAMP, the tryptophan
555 residue found in Family 2B CBDs would not be able to mediate a similar binding
556 interaction with cAMP[22]. A number of examples exist of cAMP-insensitive CBDs, able to
557 bind other ligands[54,56-59,77-79]. Based on our results, we hypothesize that the CBDs of the
558 *Sg* Enc shell may be able to bind to an as yet unidentified, non-cAMP ligand. If the *Sg*
559 Enc CBD behaves like many other CBDs, ligand binding may trigger a conformational
560 change in the *Sg* Enc shell which could function to regulate the activity of internalized
561 cargo in a ligand-dependent fashion.

562   Our results redefine the function of EshA and experimentally establish Family 2B
563 encapsulins as cargo-loaded protein nanocompartments involved in secondary
564 metabolite biosynthesis. A number of key questions remain: What is the molecular logic
565 and advantage of 2-MIBS encapsulation? What is the native ligand of *Sg* Enc CBD and
566 how are CBDs involved in mediating or regulating 2-MIB biosynthesis? Our findings
567 provide a solid foundation for future investigations into these questions, and also
568 provide molecular level detail to aid in adapting Family 2B encapsulin
569 nanocompartments for future synthetic biology and bionanotechnology applications.

570

571

Methods

Computational and phylogenetic analyses

An initial sequence similarity network (SSN) for 2-MIBS was created using the EFI-EST server[80,81] in January of 2023. The amino acid sequence of 2-MIBS from *S. griseus* (Uniprot ID: A0A2X2LWM5) was used as a BLAST query in EFI-EST with an E-value of -5, a maximum of 5,000 sequence hits, and minimum sequence length of 300 amino acids. This initial step returned 990 putative 2-MIBS encoding genes whose genome neighborhoods were then visualized using the EFI-GNT server[80,81]. To distinguish 2-MIBS-encoding genes from other similar single domain terpene cyclases, only genome neighborhoods containing MT-encoding genes (Pfam IDs: PF02353, PF08241, PF13649) within 10 kb of the putative 2-MIBS genes were included in the dataset, resulting in 745 unique 2-MIBS containing gene clusters. The dataset was then manually curated to remove spurious entries, resulting in a final dataset of 734 unique 2-MIBS-encoding gene clusters. 723 gene clusters identified from Actinobacteria were then used to assemble a phylogenetic tree, using the NGphylogeny.fr server,[82] via MAFFT alignment[83] followed by phylogenetic tree assembly using the FastTree algorithm[84,85] with 1,000 bootstraps[86]. The tree was then visualized and annotated using iTOL v6[87].

For SSN analysis of the 2-MIBS proteins contained within the final dataset, the EFI-EST server[80,81] was used to assemble an initial SSN, which was then annotated and modified using Cytoscape v3.1.0[88] to keep only edges with at least 65% sequence identity. Sequence alignments of the CLDs were performed using Clustal Omega on the EMBL-EBI server[89]. Sequence logos were visualized using Geneious Prime version 2023.0.4.

596

Molecular cloning

Plasmids for protein expression in *S. coelicolor* A3(2) were assembled according to previously published methods[90]. Briefly, to construct the *Sg* Enc_2-MIBS expression plasmid, genes encoding for the *Sg* Enc (mmpI_1) and *Sg* 2-MIBS

601 (NCTC13033_02227) were amplified as a single amplicon from *S. griseus* subs. *griseus*
602 (ATCC 23345) genomic DNA and adding NdeI and BamHI overhangs. The amplicon
603 and a pGM1190 vector were cut using NdeI (ThermoScientific, Cat # FD0583) and
604 BamHI (ThermoScientific, Cat # ER0051) restriction enzymes, purified by extraction
605 from an agarose gel. The plasmid was subsequently assembled using T4 ligase (New
606 England Biosciences, Cat # M0202S). *E. coli* BL21 (DE3) cells were then transformed
607 with the assembled plasmid via electroporation and grown on LB agar plates containing
608 50 µg/mL of apramycin at 37°C for 18 h. Plasmid-containing colonies were screened by
609 Sanger sequencing (Eurofins Scientific). The confirmed plasmid was then used as a
610 template for cloning *Sg* Enc, *Sg* Enc_ΔCLD-2-MIBS, and *Sg* Enc_CLD-mNeonGreen
611 constructs, which were cloned by either Gibson assembly or restriction cloning. For the
612 *Sg* Enc, *Sg* 2-MIBS, *Sg* MT, and *Ec* CAP constructs expressed in *E. coli* BL21 (DE3),
613 codon-optimized gBlocks from IDT were used and designed to contain 20 base pair
614 overhangs corresponding to the NdeI and PacI cut sites of the pETDuet-1 vector. The
615 *Sg* MT and *Ec* CAP gBlocks were designed to encode an 8xHi-tag followed by a
616 tobacco etch protease (TEV) cleavage site at the N-terminus, whereas the *Sg* 2-MIBS
617 plasmid was designed to encode a C-terminal TEV site followed by an 8xHis-tag. The
618 pETDuet-1 vector was linearized using NdeI and PacI (ThermoScientific, Cat #
619 FD2204), and purified by extraction from an agarose gel. The plasmids were then
620 assembled via Gibson assembly (New England Biolabs, Cat # E2611L), followed by
621 transformation of *E. coli* BL21 (DE3) cells via electroporation. Cells were plated on LB
622 agar plates containing 100 µg/mL of ampicillin and grown at 37°C for 18 h. Plasmids
623 were then confirmed by Sangar sequencing.

624

625 **Protein expression**

626 *Sg* Enc, *Sg* 2-MIBS, *Sg* MT, and *Ec* CAP were expressed in *E. coli* BL21 (DE3)
627 cells by inoculating 0.5 to 1 L of LB media containing 100 µg/mL of ampicillin with an
628 overnight starter culture, followed by shaking at 200 rpm and 37°C until the cells
629 reached an OD_{600} of 0.5. Protein expression was then induced by the addition of 0.1
630 mM IPTG for *Sg* Enc and 0.5 mM IPTG for *Sg* 2-MIBS, *Sg* MT, and *Ec* CAP, followed by

631   shaking at 200 rpm and 18°C for 18 h. Cells were harvested, and cell pellets were

632   stored at -20°C until use.

633   Heterologous protein expression in *S. coelicolor* A3(2) (ATCC #BAA-471, also

634   known as *Streptomyces violaceoruber* strain John Innes Centre M145) was performed

635   according to previously published protocols[90,91]. *E. coli* ET12567/pUZ8002 cells were

636   transformed with pGM1190 vectors containing the *Sg* Enc_2-MIBS, *Sg* Enc, *Sg*

637   Enc_ΔCLD-2-MIBS, and *Sg* Enc CLD-mNeonGreen constructs via heat shock and

638   maintained on LB agar plates containing 50 µg/mL apramycin, 30 µg/mL

639   chloramphenicol, and 50 µg/mL kanamycin. Plasmid-containing colonies were then

640   used to inoculate 20 mL overnight cultures of LB containing 50 µg/mL apramycin, 30

641   µg/mL chloramphenicol, and 50 µg/mL kanamycin, which were grown overnight at 37°C

642   with shaking at 200 rpm. The overnight cultures of *E. coli* ET12567/pUZ8002 were

643   harvested the following morning by centrifugation at 3,800 x g for 10 min. Cells were

644   washed by resuspending the cell pellets in 20 mL LB followed by centrifugation at 3,800

645   x g for 10 min. This step was repeated again, and the final cell pellet was then

646   resuspended in 2 mL of LB. 200 µL of plasmid-containing resuspended *E. coli*

647   ET12567/pUZ8002 cells were then mixed with 500 µL of freshly prepared *S. coelicolor*

648   A3(2) spore stock in GYM medium which had been heated at 50°C for 10 min

649   immediately prior to mixing with *E. coli* ET12567/pUZ8002 cells. The cell mixtures were

650   then plated on MS agar containing 10 mM $MgCl_2$, and incubated overnight at 30°C. The

651   following morning, the plates were overlaid with 1 mg nalidixic acid and 1 mg of

652   apramycin and incubated at 30°C for 3 to 5 days until spores became visible. Individual

653   colonies with visible spores were then either struck or plated as lawns on fresh plates of

654   MS agar containing 50 µg/mL apramycin and 25 µg/mL nalidixic acid. The plates were

655   then incubated for another 3 to 5 days and 30°C until spores were visible. The spores

656   were scraped from the plates and used to inoculate 500 mL baffled flasks containing

657   YEME media with glass beads and 50 µg/mL apramycin. The cultures were grown by

658   shaking at 200 rpm at 30°C until the OD_{600} reached 0.4 to 0.8, which typically took 24 to

659   36 h. To induce protein expression, 20 µg/mL of thiostrepton was added to the media,

660   and cells were grown with shaking at 200 rpm and 30°C for 72 h. Cells were then

661   harvested by centrifugation and stored at -20°C until further use.

662

Protein purification

664   *Sg* Enc_2-MIBS for cryo-EM analysis was purified according to previously
665   published protocols[90]. 1.5 g of a frozen *S. coelicolor* A3(2) cell pellet was resuspended
666   in 7.5 mL of lysis buffer consisting of 150 mM NaCl, 20 mM Tris pH 7.5, 2 mM $MgCl_2$, 1
667   mg/mL lysozyme, 250 units of Benzonase nuclease (Millipore Sigma), and a protease
668   cocktail consisting of 10 μM leupeptin, 14 μM E-64, 0.4 μM aprotinin, 0.4 mM AEBSF,
669   and 114 μM bestatin. The resuspended cell pellet was then lysed by sonication for 3 min
670   on ice, pulsing for 10 s on and 20 s off, with a power of 96 W using a Model 120 Sonic
671   Dismembrator (Fisher Scientific, Inc). The lysate was then centrifuged at 21,000 x g for
672   10 min to remove cell debris. Ammonium sulfate was added to the clarified supernatant
673   to 50% saturation, and the supernatant was rocked at 4°C for 1 h. Precipitated protein
674   was pelleted by centrifugation at 10,000 x g and 4°C for 10 min. The protein pellet was
675   resuspended in 10 mL of 150 mM NaCl, 2 mM $MgCl_2$, 20 mM Tris pH 7.5, and
676   concentrated to 0.5 mL using an Amicon 100 Ultra-15 centrifugal filter (Millipore) with a
677   molecular weight cutoff of 100 kDa. Approximately 125 units of Benzonase nuclease
678   were added to the concentrated protein and the mixture incubated at 4°C for 30 min.
679   The protein sample was then loaded onto an AKTA Pure FPLC system and purified with
680   a Superose 6 Increase 10/300 GL column using 150 mM NaCl, 20 mM Tris pH 7.5 as
681   the running buffer. *Sg* Enc_2-MIBS containing fractions were pooled, concentrated to
682   approximately 3 mg/mL, and stored at 4°C until use.

683   *E. coli*-expressed *Sg* Enc for cryo-EM analysis was purified by resuspending
684   approximately 3 g of cell pellet in 15 mL of lysis buffer. The resuspended pellet was then
685   lysed by sonication for 3 minutes on ice, pulsing for 10 s on and 20 s off, at a power of
686   75 W. The lysate was clarified by centrifugation at 8,000 x g and 4°C for 10 min. 10%
687   (w/v) PEG 8,000 and 450 mM NaCl were added to the clarified lysate, which was further
688   rocked at 4°C for 1 h to precipitate *Sg* Enc shells. The lysate was then clarified by
689   centrifugation at 8,000 x g and 4°C for 10 min. The supernatant was discarded, and the
690   pellet was resuspended in a minimal volume of 150 mM NaCl, 20 mM Tris pH 7.5 and
691   filtered using a 0.22 μm syringe filter. The clarified protein sample was then loaded onto
692   an AKTA Pure FPLC and purified with a Sephacryl S-500 size exclusion column. *Sg*

693 Enc-containing fractions were then pooled, and subsequently buffer exchanged into 20

694 mM Tris pH 7.5 buffer using an Amicon Ultra-15 100 kDa centrifugal filter, loaded onto

695 an AKTA Pure FPLC system, followed by purification with a HiPrep DEAE FF 16/10

696 column using a linear gradient from 0 to 1 M NaCl, 20 mM Tris pH 7.5. This step was

697 performed twice to better enrich the *Sg* Enc protein sample. *Sg* Enc-containing fractions

698 were pooled, concentrated to 0.5 mL, and loaded onto an AKTA Pure FPLC and purified

699 using a Superose 6 10/300 GL column with 150 mM NaCl, 20 mM Tris pH 7.5 as the

700 running buffer. *Sg* Enc containing fractions were then pooled, concentrated to 3.0

701 mg/mL, and stored at 4°C until further use.

702   *S. coelicolor* A3(2)-expressed *Sg* Enc for cryo-EM analysis was purified using the

703 same method as for *S. coelicolor* A3(2)-expressed *Sg* Enc_2-MIBS outlined above,

704 however, the *Sg* Enc sample was further purified through an additional step by loading

705 onto an AKTA Pure FPLC system and purifying via a HiTrap Q-FF column using a

706 gradient from 20 mM Tris pH 8.0, 5 mM $MgCl_2$ to 20 mM Tris pH 8.0, 5 mM $MgCl_2$, 1 M

707 NH_4Cl. *Sg* Enc-containing fractions were then pooled, concentrated to 3.2 mg/mL,

708 desalted into 150 mM NaCl, 20 mM Tris pH 7.5, and stored at 4°C until use.

709   *S. coelicolor* A3(2)-expressed *Sg* Enc_2-MIBS for cryo-EM analysis in the

710 presence of 20 mM cAMP was purified by resuspending 3 g of frozen *S. coelicolor* A3(2)

711 cell pellet in a lysis buffer consisting of 20 mM tris pH 8.0, 10 mM $MgCl_2$, 20 mM

712 ammonium acetate, 1x protease inhibitor cocktail as above, 250 U of Benzonase

713 nuclease, and 1 mg/mL of lysozyme. The resuspended cell pellet was then lysed by

714 sonication for 3 min on ice, pulsing for 10 s on and 20 s off, with a power of 96 W. The

715 lysate was then centrifuged at 21,000 x g for 10 minutes to remove cell debris.

716 Ammonium sulfate was then added to the clarified supernatant to 50% saturation, and

717 the supernatant was rocked at 4°C for 1 h. Precipitated protein was then pelleted by

718 centrifugation at 10,000 x g for 10 min at 4°C. The protein pellet was resuspended in 10

719 mL of 20 mM Tris pH 8.0, 10 mM $MgCl_2$, 20 mM ammonium acetate, and concentrated

720 to 0.5 mL using an Amicon 100 Ultra-15 100 kDa filter. Approximately 125 units of

721 Benzonase nuclease were added to the concentrated protein followed by incubation at

722 4°C for approximately 30 min. The protein sample was then loaded onto an AKTA Pure

723 FPLC and purified using a Superose 6 Increase 10/300 GL column in 20 mM Tris pH

724 8.0, 10 mM MgCl$_2$, 20 mM ammonium acetate running buffer. *Sg* Enc_2-MIBS-

725 containing fractions were pooled, again loaded onto an AKTA Pure FPLC, and purified

726 via a HiTrap Q-FF column using a gradient from 20 mM Tris pH 8.0, 5 mM MgCl$_2$ to 20

727 mM Tris pH 8.0, 5 mM MgCl$_2$, 1 M NH$_4$Cl. *Sg* Enc_2-MIBS-containing fractions were

728 pooled, desalted into 20 mM Hepes pH 7.5, 5 mM MgCl$_2$ using an Amicon 100 Ultra-15

729 100kDa filter, loaded onto an AKTA Pure FPLC, and purified via a HiTrap S-FF column

730 using a gradient from 20 mM Hepes pH 7.5, 5 mM MgCl$_2$ to 20 mM Hepes pH 7.5, 5

731 mM MgCl$_2$, 1M NH$_4$Cl. *Sg* Enc-containing fractions were then pooled, concentrated to

732 0.5 mL using an Amicon 100 Ultra-15 100 kDa filter, again loaded onto an AKTA Pure

733 FPLC, and purified using a Superose 6 Increase 10/300 GL column equilibrated with 20

734 mM Hepes pH 7.5, 25 mM NaCl, 5 mM MgCl$_2$. *Sg* Enc_2-MIBS-containing fractions

735 were pooled, concentrated, and exchanged into 20 mM Hepes pH 7.5, 5 mM MgCl$_2$.

736 cAMP was added to a final concentration of 20 mM approximately 1 h before cryo-EM

737 grids were prepared with a final protein concentration of 3.5 mg/mL. During this

738 purification, the cation and anion exchange purification steps were performed to remove

739 any non-specifically bound charged ligands.

740 *S. coelicolor* A3(2)-expressed *Sg* Enc_2-MIBS for enzymatic assays and cAMP

741 binding experiments was purified by the same method used to prepare *S. coelicolor*

742 A3(2)-expressed *Sg* Enc_2-MIBS for cryo-EM analysis in the presence of cAMP, except

743 for omitting the Hi-Trap S-FF purification step. Similarly, *S. coelicolor* A3(2)-expressed

744 *Sg* Enc_CLD-mNeonGreen variants were purified using the same methods, except for

745 omitting both the Hi-Trap S-FF and second Superose 6 Increase 10/300 GL purification

746 steps. For enzymatic assays, *Sg* Enc_2-MIBS was stored at 4°C until use. *Sg* Enc_2-

747 MIBS used for product profile experiments, DSF, and cAMP-agarose binding

748 experiments was drop frozen in liquid nitrogen and stored at -80°C until use. *Sg*

749 Enc_CLD-mNeonGreen variants were also drop frozen in liquid nitrogen and stored at -

750 80°C until use.

751 *E. coli*-expressed *Sg* 2-MIBS was purified by resuspending 4 g of frozen cell

752 pellet in 20 mL of 20% glycerol, 300 mM NaCl, 20 mM imidazole pH 8.0, 20 mM Tris pH

753 8.0, 5 mM MgCl$_2$, 1 x protease inhibitor cocktail used in all other purifications, and 0.5

754 mg/mL lysozyme, followed by sonication on ice at 88 W for 3 min, in pulses of 10 s on

755   separated by 20 s off. The lysate was centrifuged for 10 min at 21,000 x g and the

756   clarified supernatant loaded onto an AKTA Pure FPLC system and purified via a His-

757   Trap FF column using a gradient from 20% glycerol, 300 mM NaCl, 20 mM imidazole

758   pH 8.0, 20 mM Tris pH 8.0, 5 mM MgCl$_2$ to the same buffer but containing 1M imidazole.

759   *Sg* 2-MIBS-containing fractions were pooled, concentrated to 0.5 mL in an Amicon Ultra-

760   15 centrifugal filter with a 30 kDa molecular weight cutoff, again loaded on an AKTA

761   pure FPLC, and purified via a Superdex S200 column in 10 mM Hepes pH 7.5, 5 mM

762   MgCl$_2$ running buffer. *Sg* 2-MIBS-containing fractions were pooled, concentrated to

763   approximately 5 mg/mL and stored at 4°C if used for enzymatic activity assays,

764   otherwise, remaining protein was drop frozen in liquid nitrogen and stored at -80°C.

765   *E. coli*-expressed *Sg* 2-MT was purified using the same lysis buffers and initial

766   His-Trap FF step as used for the *Sg* 2-MIBS purification. Following this initial step, 3 mg

767   of recombinantly-expressed tobacco-etch virus (TEV) protease was added to *Sg* MT,

768   which was then dialyzed against 1 L of 20 mM Tris pH 8.0, 300 mM NaCl, 20% glycerol

769   overnight at 4°C to remove the cleaved N-terminal 8xHis-tag. Then, the protein sample

770   was again purified via a HisTrap-FF column, with TEV-cleaved protein collected in the

771   flow-through, exchanged into 20 mM Tris pH 8.0, 300 mM NaCl, 20% glycerol, 1 mM

772   DTT using an Amicon Ultra-15 100 kDa filter unit, and concentrated to 20 mg/mL. The

773   protein was then drop frozen in liquid nitrogen and stored at -80°C until further use.

774   *E. coli*-expressed *Ec* CAP was purified using the same lysis buffers as for *Sg* MT

775   and *Sg* 2-MIBS and initial His-Trap FF step. Following the His-Trap-FF step, *Ec* CAP

776   was dialyzed overnight with stirring against 1 L of 50 mM Na$_2$/KPO$_4$ pH 7.0, 300 mM

777   NaCl, and 0.1 mM TCEP at 4°C. *Ec* CAP was then concentrated to ca. 2.5 mg/mL using

778   an Amicon Ultra-15 centrifugal filter unit with a molecular weight cutoff of 10 kDa, drop

779   frozen in liquid nitrogen, and stored at -80°C until use.

780

781   **Negative stain transmission electron microscopy (TEM)**

782   Protein samples were diluted to 0.1 to 0.15 mg/mL and applied to freshly glow

783   discharged (60 s, 5 mA, PELCO easiGlow) 200 mesh gold formvar/carbon square mesh

784   grids (EMS, Cat # FCF200-AU-EC). After 60 s, the samples were blotted of the grids

785   using filter paper. Grids were then washed with a 5 µL drop of distilled water, followed

786   by two rounds of 5 s and then 30 s staining with 5 µL drops of 0.75% uranyl formate.

787   TEM micrographs were recoreded using an FEI Morgagni TEM operating at 100 kV

788   located at the Michigan University Life Sciences Institute.

789

Native PAGE analysis

791   Approximately 1 µg of purified *Sg* Enc-CLD-mNeonGreen variants, were loaded

792   onto Invitrogen Native PAGE 3-12% Bis-Tris gels. The native PAGE gels were run using

793   1x Invitrogen Native PAGE buffer for 120 min at 150 V at 4°C. In-gel mNeonGreen

794   fluorescence was imaged using a GelDoc Go imaging system (BioRad Laboratories,

795   Inc) equipped with a blue sample tray. Gels were then stained overnight using Ready

796   Blue Stain (Sigma Aldrich), destained briefly in water, and imaged using a ChemiDoc

797   Imaging system (BioRad Laboratories, Inc). Protein bands on Coomassie-stained and

798   fluorescently imaged Native PAGE gels were analyzed via gel densitometry in

799   ImageJ2[92]. Protein amount per lane was normalized to the *Sg* Enc-CLD-mNeonGreen

800   sample. The same approach was taken to obtain a normalized mNeonGreen

801   fluorescence intensities for each sample. Relative cargo-loading was then calculated as

802   a ratio of normalized protein amount and normalized fluorescence intensiy.

803

2-MIBS activity assays

805   2-MIBS activity assays were performed by pre-incubating samples of 10 to 200

806   µM GPP (Sigma Aldrich, Cat # 76532-10MG) in a buffer containing 25 µM purified *Sg*

807   MT, 2 mM SAM (Sigma Aldrich, Cat # A7007-25MG), 10 mM MgCl$_2$, and 20 mM Hepes

808   pH 7.5 for 2 h at 30°C to convert GPP to 2-methyl-GPP. Next, 2-amino-6-mercapto-7-

809   methyl purine ribonucleoside (MESG), purine nucleoside phosphorylase (PNP), and

810   inorganic pyrophosphatase (PPase) were added according to the EnzCheck

811   pyrophosphate assay kit (ThermoFisher, Cat # E-6645). The samples were briefly

812   centrifuged for 3 min at 10,000 x g to remove any precipitated *Sg* MT, and then 30 µL of

813   each sample was aliquoted in triplicate into 384 well plates (Corning, Cat # 3540).

814   Assays were initiated by addition of 10 µL of 2 µM purified free *Sg* 2-MIBS, or 10 µL of

815   20 µM purified encapsulated 2-MIBS (*Sg* Enc_2-MIBS) to each sample well, resulting in

816   final assay volume of 40 µL and final concentrations of 5 to 100 µM GPP/2-methyl-GPP,

6.25 µM *Sg* MT, 0.5 mM SAM/SAH, 5 mM MgCl$_2$, 10 mM Hepes pH 7.5, 200 µM MESG, 0.04 U PNP, and 0.0012 U PPase. Absorbance at 360 nm, corresponding to the conversion of MESG to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine by PNP, were collected every 22 s in sweep mode using a BioTek Synergy H1 microplate reader. Slopes were calculated in the linear reaction periods (8 to 14 min) and used to calculate reaction velocities using a pyrophosphate standard curve. Velocities were plotted and curves were fitted to an allosteric sigmoidal model in GraphPad Prism 9.10.

Gas chromatography mass spectrometry (GC-MS) analysis

Reactions were prepared in sealed vials containing 1 mM SAM, 75 µM GPP, 12.5 µM *Sg* MT, 10 mM Hepes pH 7.5, 5 mM MgCl$_2$, 1 mM DTT and 0.5 µM *Sg* 2-MIBS, 5 µM *Sg* Enc_2-MIBS, or 5 µM *Sg* Enc_2-MIBS + 500 µM cAMP. A 100 µL pentane cushion was pipetted on the top of each reaction, and the reactions were incubated for 20 h at room temperature. Samples were centrifuged for 1 min at 14,000 x g, the aqueous layer was transferred to a tube containing 200 µL of pentane, mixed by vortexing, and spun again for 1 min at 14,000 x g, and the pentane layer then pooled with the other pentane layers. 2 to 3 small crystals of disodium sulfate were added to each pentane extract to remove residual water, incubated for 10 min at room temperature, and the pentane layer then transferred to sample vials for GC-MS analysis. The same extraction procedure was used for the 2-MIB standard with a starting concentration of approximately 100 µM. 1 µL of each sample was injected into a Shimadzu QP-2010 GC-MS instrument equipped with a DB-5 column, run in splitless mode with a column flow rate of 0.97 mL/min and a temperature ramp of 60°C to 280°C at 20°C/min. Data analysis was carried out in OpenChrom v 1.5.

Differential scanning fluorimetry (DSF)

Thermal melts of *Sg* Enc_2-MIBS or *Sg* Enc_2-MIBS + 200 µM cAMP were obtained by DSF using an Uncle instrument (Unchained Labs). DSF samples contained 10x SYPRO orange dye (diluted from 5000x stock in DMSO, Invitrogen, Cat # S6651) , 1 µM *Sg* Enc_2-MIBS, 10 mM Hepes pH 7.5, 0.1 mM TCEP, with or without additional 200 µM cAMP. Thermal ramps progressed from 20°C to 95°C at a rate of 1°C/min, and were monitored by measuring fluorescence emission spectra from 250 to 720 nm with

848 an excitation wavelength of 473 nm. The Uncle Analysis V5.04 software package was

849 used to calculate the area under the curve between 510 and 680 nm, which was used to

850 represent thermal transition curves and calculate differential inflection points

851 representative of thermal transitions (melting temperature, T_m).

852

853

854

855 **cAMP-agarose binding assays**

856 cAMP-agarose binding assays were carried out in batch format. 0.2 mg of

857 purified *Sg* Enc_2-MIBS or *Ec* CAP was exchanged into binding buffer consisting of 100

858 mM NaCl, 10 mM Hepes pH 7.5, and 5 mM $MgCl_2$. 300 µL of *Sg* Enc_2-MIBS or *Ec*

859 CAP was then incubated with 100 µL of cAMP-Agarose (Sigma Aldrich, Cat # A0144) in

860 the same buffer for 15 min at room temperature with gentle rocking. The resin was

861 centrifuged for 3 min at 3,000 x g, the supernatant was discarded, and the resin was

862 resuspended in 500 µL of binding buffer followed by centrifugation at 3,000 x g. The

863 supernatant was discarded, and the wash step was repeated two more times to remove

864 non-specifically bound protein. Bound protein was eluted by resuspending the resin

865 twice in 200 µL of 20 mM cAMP, 100 mM NaCl, 10 mM Hepes pH 7.5, 5 mM $MgCl_2$.

866

867 **Cryo-electron microscopy (cryo-EM)**

868 *Sample preparation*: Purified samples of *Sg* Enc_2-MIBS, *Sg* Enc (*E. coli BL21* (DE)-

869 expressed), *Sg* Enc (*S. coelicolor* A3(2)-expressed), and *Sg* Enc_2-MIBS (20 mM

870 cAMP) were concentrated to 2-3 mg/mL in 150 mM NaCl, 20 mM Tris pH 7.5 for *Sg*

871 Enc_2-MIBS, *Sg* Enc (*E. coli BL21* (DE)-expressed), and *Sg* Enc (*S. coelicolor* A3(2)-

872 expressed). *Sg* Enc_2-MIBS + 20 mM cAMP was prepared in buffer containing 10 mM

873 Hepes pH 7.5, 5 mM $MgCl_2$, and 20 mM cAMP. Samples were applied to freshly glow

874 discharged Quantifoil R1.2/1.3 Cu 200 mesh grids and were immediately plunged into

875 liquid ethane using an FEI Vitrobot Mark IV (100% humidity, 22°C, blot force 20, blot

876 time 4 s). The grids were then clipped and stored in liquid nitrogen until data collection.

877 *Data collection*: Cryo-EM movies of *Sg* Enc_2-MIBS and *Sg* Enc_2-MIBS (20 mM

878 cAMP) were collected using a ThermoFisher Scientific Talos Arctica cryo-electron

879   microscope operating at 200 kV equipped with a Gatan K2 direct electron detector.

880   Cryo-EM movies of *Sg* Enc (*E. coli BL21* (DE)-expressed) and *Sg* Enc (*S. coelicolor*

881   A3(2)-expressed) were collected using a ThermoFisher Scientific Glacios cryo-electron

882   microscope operating at 200 kV equipped with a Gatan K2 direct electron detector. The

883   Leginon[93] software package was used to collect all datasets at a 45,000x magnification

884   and pixel size of 0.91 Å. Further data collection statistics can be found in

885   Supplementary Table 1.

886   *Data processing*: cryoSPARC v 4.4.0+231114[94] was used to process the *Sg* Enc_2-

887   MIBS dataset. 1,454 movies were imported, motion corrected by patch motion

888   correction, and the contrast transfer function (CTF) fit was estimated using patch CTF

889   estimation. Movies with a CTF fit worse than 5 Å were excluded from the dataset,

890   resulting in 1,302 movies. Manual picker was used to select 296 particles, which were

891   used to create templates for particle picking. 54,206 particles with a box size of 440

892   pixels were selected and extracted from the dataset using template picker. 2 rounds of

893   2D classification were then used to further sort the particles, resulting in 51,501

894   remaining particles. Initial volumes were created using ab-initio reconstruction with 5

895   classes and I symmetry, resulting in a major class containing 51,173 particles. These

896   particles were then used for homogeneous refinement against the initial ab-initio

897   reconstruction map with I symmetry imposed, per particle defocus optimization, per-

898   group CTF parameterization, and Ewald sphere correction enabled with a positive

899   curvature sign, resulting in a 3.02 Å map. The particles were then symmetry-expanded

900   according to I symmetry and used for downstream 3D classification. For analysis of the

901   CBDs, 3,070,380 symmetry-expanded particles were sorted by 3D classification without

902   alignment into 10 classes using a mask containing two CBDs arranged around the two-

903   fold axis of symmetry with a target resolution of 3.5 Å, an O-EM learning rate init of

904   0.75, class similarity of 0.25, and force hard classification enabled. The classes were

905   then manually inspected, and the class with the best-resolved CBD density was

906   reconstructed using homogeneous reconstruction only, resulting in a 3.49 Å map. The

907   same 3D classification parameters were used for analysis of the CLD, except a map

908   encompassing the three-fold axis of the shell and CLD density was used for

909   classification. Particle subtraction was performed using a mask of the *Sg* Enc shell to

910 generate shell-subtracted particles. 2D classification was then used to classify and sort

911 shell-subtracted particles.

912 cryoSPARC v 4.4.0+231114[94] was used to process the *Sg* Enc (*E. coli BL21*

913 (DE)-expressed) dataset. 1,930 movies were imported, motion corrected with patch

914 motion correction, and CTF fit was estimated using patch CTF estimation. Movies with

915 CTF fit values worse than 5 Å were removed from the dataset, resulting in 1,898

916 remaining movies. 454 particles were manually selected from micrographs and used to

917 generate templates for particle picking. 195,618 particles were selected using template

918 picker and extracted with a box size of 440 pixels. Two rounds of 2D classification were

919 then used to sort the particles, resulting in 135,962 remaining particles. An initial map

920 was then created using ab-initio reconstruction with 3 classes and I symmetry imposed,

921 resulting in a majority class containing 135,222 particles. These particles were then

922 used for homogeneous refinement against the ab-initio map with I symmetry imposed,

923 per-particle defocus optimization, per-group CTF parameterization, spherical aberration

924 fit enabled, tetrafoil fit enabled, Ewald sphere correction enabled with a negative

925 curvature sign, and minimize over per-particle scale enabled, resulting in a 2.75 Å

926 resolution map. The particles were then polished using reference based motion

927 correction, resulting in 135,027 particles which were used to perform a final

928 homogeneous refinement with I symmetry imposed, per-particle defocus optimization,

929 per-group CTF parameterization, spherical aberration fit enabled, tetrafoil fit enabled,

930 Ewald sphere correction with negative curvature, and minimize over per-particle scale

931 enabled, resulting in a 2.71 Å resolution map. The particles were then symmetry-

932 expanded and used for subsequent 3D classification without alignment of the CBD and

933 two-fold of the shell. For analysis of the CBDs, 8,101,620 symmetry expanded particles

934 were sorted by 3D classification into 10 classes using a mask containing two CBDs

935 along the two-fold axis with a target resolution of 3.5 Å, an O-EM learning rate init of

936 0.75, class similarity of 0.25, and force hard classification enabled. The 3D classes were

937 then manually inspected and the class with the best resolved CBD containing 828,041

938 symmetry-expanded particles was then reconstructed using homogeneous

939 reconstruction only, resulting in a 3.10 Å map. The same 3D classification parameters

940 were used for analysis of the three-fold, using a map encompassing the three-fold shell

941 and CLD density. Particle subtraction was performed using a mask of the *Sg* Enc shell
942 to generate shell-subtracted particles. 2D classification was then used to classify and
943 sort shell-subtracted particles.

944 cryoSPARC v 4.4.0+231114[94] was used to process the *Sg* Enc (*S. coelicolor*
945 A3(2)-expressed) dataset. 156 movies were imported, motion corrected by patch motion
946 correction, and the CTF fit was estimated using patch CTF estimation. Manual picker
947 was used to select 174 particles, which were used to create templates for particle
948 picking. 7,003 particles were identified using template picker and extracted with a box
949 size of 440 pixels. 2 rounds of 2D classification were used to further sort the particles,
950 resulting in 4,119 particles. Ab-initio reconstruction was performed with I symmetry and
951 4 classes, resulting in a majority class containing 4,091 particles. These particles were
952 then used to for homogeneous refinement against the ab-initio volume with I symmetry
953 imposed, per-particle defocus optimization, per-group CTF parameterization, tetrafoil fit
954 enabled, anisotropic magnification fit enabled, and Ewald sphere correction with
955 negative curvature, resulting in a 3.47 Å reconstructed map. Particle subtraction was
956 performed using a mask of the *Sg* Enc shell to generate shell-subtracted particles. 2D
957 classification was then used to classify and sort shell-subtracted particles.

958 cryoSPARC v 4.4.0+231114[94] was used to process the *Sg* Enc_2-MIBS + 20 mM
959 cAMP dataset. 86 movies were imported, motion corrected using patch motion
960 correction, and the CTF fit was estimated using patch CTF. Exposures with CTF fits
961 worse than 5 Å were removed from the dataset, resulting in 83 remaining exposures.
962 315 particles were manually selected and used to establish templates for particle
963 picking. 9,265 particles were identified using template picker and extracted with a box
964 size of 440 pixels. The particles were sorted by two rounds of 2D classification, resulting
965 in 9,181 particles. Ab-initio reconstruction was used with these particles to generate
966 starting volumes with I symmetry imposed and two classes, resulting in a majority class
967 containing 9,170 particles. These particles were then used for homogeneous refinement
968 with I symmetry, per-particle defocus optimization, per-group CTF parameterization,
969 spherical aberration fit enabled, tetrafoil fit enabled, anisotropic magnification fit
970 enabled, Ewald sphere correction with negative curvature, resulting in a 2.92 Å map.
971 Particles were polished using reference based motion correction, and then used for an

972 additional homogenous refinement with I symmetry, per-particle defocus optimization,

973 per-group CTF parameterization, spherical aberration fit, tetrafoil fit, anisotropic

974 magnification fit, and Ewald sphere correction with a positive curvature sign, resulting in

975 a 2.93 Å reconstructed map. The particles were then symmetry-expanded according to I

976 symmetry and used for 3D classification. For analysis of the CBDs, 550,200 symmetry-

977 expanded particles were sorted into by 3D classification without alignment using a mask

978 containing two CBDs at the two-fold axis into 4 classes with a target resolution of 3.5 Å,

979 O-EM learning rate init of 0.75, force hard classification enabled, and a class similarity

980 of 0.25. The 3D classes were manually inspected, and the class with the best resolved

981 CBD containing 142,420 symmetry-expanded particles was then reconstructed using

982 homogeneous reconstruction only, resulting in a 3.43 Å reconstructed map.

983 *Model building*: For building the *Sg* Enc model, an AlphaFold[95] generated model of *Sg*

984 Enc was obtained from UniProt[96] and manually placed in the I- refined cryo-EM map

985 using ChimeraX v 1.5[97], and the fit was refined using the Fit in Map command. The

986 model was then refined iteratively by alternating rounds of manual refinement in Coot v

987 8.9.1[98,99] and real-space refinement using Phenix v 1.20.1-4487[100,101]. Symmetry

988 operators were identified from the map using the map_symmetry command and applied

989 using the apply_ncs command to generate an icosahedral shell with 60 copies of the *Sg*

990 Enc protomer. The NCS-expanded shell was then refined against the I-refined cryo-EM

991 map using real-space refinement with NCS restraints. A single protomer of the NCS-

992 refined model was then refined against the C1 map of the CBDs generated by 3D

993 classification and homogenous reconstruction using real space refinement to improve

994 the model of the CBD. BioMT operators were identified using the find_NCS command

995 and manually applied to the headers of the model. Model building of *Sg* Enc_2-MIBS

996 and *Sg* Enc_2-MIBS + 20 mM cAMP followed a nearly identical strategy, except that the

997 *Sg* Enc model was used as a starting model.

998

999 **Data availability**

1000 The cryo-EM maps and structural models of *Sg* Enc_2-MIBS, *Sg* Enc_2-MIBS +

1001 cAMP, and *Sg* Enc have been deposited and are publicly available in the Electron

1002 Microscopy Data Bank (EMDB-44554, EMDB-44557, and EMDB-44553) and Protein

1003 Data Bank (PDB ID: 9BHV, PDB ID: 9BI0, and PDB ID: 9BHU). Sequence similarity

1004 networks and sequence alignments are supplied as supplementary data.

1005

1006 **References**

1007 1 Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews*
1008   **117**, 11570-11648 (2017). https://doi.org/10.1021/acs.chemrev.7b00287
1009 2 Rudolf, J. D., Alsup, T. A., Xu, B. & Li, Z. Bacterial terpenome. *Nat Prod Rep* **38**, 905-980 (2021).
1010   https://doi.org/10.1039/d0np00066c
1011 3 Tholl, D. in *Biotechnology of Isoprenoids* (eds Jens Schrader & Jörg Bohlmann) 63-106 (Springer
1012   International Publishing, 2015).
1013 4 Harvey, A. L. Natural products in drug discovery. *Drug Discovery Today* **13**, 894-901 (2008).
1014   https://doi.org/https://doi.org/10.1016/j.drudis.2008.07.004
1015 5 Cane, D. E. & Ikeda, H. Exploration and mining of the bacterial terpenome. *Acc Chem Res* **45**,
1016   463-472 (2012). https://doi.org/10.1021/ar200198d
1017 6 Dickschat, J. S. Bacterial terpene cyclases. *Nat Prod Rep* **33**, 87-110 (2016).
1018   https://doi.org/10.1039/c5np00102a
1019 7 Juttner, F. & Watson, S. B. Biochemical and ecological control of geosmin and 2-methylisoborneol
1020   in source waters. *Appl Environ Microbiol* **73**, 4395-4406 (2007).
1021   https://doi.org/10.1128/AEM.02250-06
1022 8 Cane, D. E. & Ikeda, H. Exploration and Mining of the Bacterial Terpenome. *Accounts of Chemical*
1023   *Research* **45**, 463-472 (2012). https://doi.org/10.1021/ar200198d
1024 9 Becher, P. G. *et al.* Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a
1025   soil arthropod to Streptomyces bacteria promoting spore dispersal. *Nat Microbiol* **5**, 821-829
1026   (2020). https://doi.org/10.1038/s41564-020-0697-x
1027 10 Zaroubi, L. *et al.* The Ubiquitous Soil Terpene Geosmin Acts as a Warning Chemical. *Appl Environ*
1028   *Microbiol* **88**, e0009322 (2022). https://doi.org/10.1128/aem.00093-22
1029 11 Garbeva, P., Avalos, M., Ulanova, D., van Wezel, G. P. & Dickschat, J. S. Volatile sensation: The
1030   chemical ecology of the earthy odorant geosmin. *Environ Microbiol* **25**, 1565-1574 (2023).
1031   https://doi.org/10.1111/1462-2920.16381
1032 12 Saha, M. & Fink, P. Algal volatiles - the overlooked chemical language of aquatic primary
1033   producers. *Biol Rev Camb Philos Soc* **97**, 2162-2173 (2022). https://doi.org/10.1111/brv.12887
1034 13 Wang, C. M. & Cane, D. E. Biochemistry and molecular genetics of the biosynthesis of the earthy
1035   odorant methylisoborneol in Streptomyces coelicolor. *J Am Chem Soc* **130**, 8908-8909 (2008).
1036   https://doi.org/10.1021/ja803639g
1037 14 Giglio, S., Chou, W. K., Ikeda, H., Cane, D. E. & Monis, P. T. Biosynthesis of 2-methylisoborneol in
1038   cyanobacteria. *Environ Sci Technol* **45**, 992-998 (2011). https://doi.org/10.1021/es102992p
1039 15 Wang, Z., Xu, Y., Shao, J., Wang, J. & Li, R. Genes associated with 2-methylisoborneol biosynthesis
1040   in cyanobacteria: isolation, characterization, and expression in response to light. *PLoS One* **6**,
1041   e18665 (2011). https://doi.org/10.1371/journal.pone.0018665
1042 16 Dickschat, J. S. *et al.* Biosynthesis of the off-flavor 2-methylisoborneol by the myxobacterium
1043   Nannocystis exedens. *Angew Chem Int Ed Engl* **46**, 8287-8290 (2007).
1044   https://doi.org/10.1002/anie.200702496
1045 17 Koksal, M., Chou, W. K., Cane, D. E. & Christianson, D. W. Structure of 2-methylisoborneol
1046   synthase from Streptomyces coelicolor and implications for the cyclization of a noncanonical C-

1047   methylated monoterpenoid substrate. *Biochemistry* **51**, 3011-3020 (2012).
1048   https://doi.org/10.1021/bi201827a

1049 18 Komatsu, M., Tsuda, M., Omura, S., Oikawa, H. & Ikeda, H. Identification and functional analysis
1050   of genes controlling biosynthesis of 2-methylisoborneol. *Proc Natl Acad Sci U S A* **105**, 7422-7427
1051   (2008). https://doi.org/10.1073/pnas.0802312105

1052 19 Koksal, M., Chou, W. K., Cane, D. E. & Christianson, D. W. Structure of geranyl diphosphate C-
1053   methyltransferase from Streptomyces coelicolor and implications for the mechanism of
1054   isoprenoid modification. *Biochemistry* **51**, 3003-3010 (2012). https://doi.org/10.1021/bi300109c

1055 20 Kawamoto, S. *et al.* Molecular and functional analyses of the gene (eshA) encoding the 52-
1056   kilodalton protein of Streptomyces coelicolor A3(2) required for antibiotic production. *J Bacteriol*
1057   **183**, 6009-6016 (2001). https://doi.org/10.1128/JB.183.20.6009-6016.2001

1058 21 Passner, J. M., Schultz, S. C. & Steitz, T. A. Modeling the cAMP-induced Allosteric Transition Using
1059   the Crystal Structure of CAP-cAMP at 2.1Å Resolution. *Journal of Molecular Biology* **304**, 847-859
1060   (2000). https://doi.org/https://doi.org/10.1006/jmbi.2000.4231

1061 22 Andreas, M. P. & Giessen, T. W. Large-scale computational discovery and analysis of virus-derived
1062   microbial nanocompartments. *Nat Commun* **12**, 4748 (2021). https://doi.org/10.1038/s41467-
1063   021-25071-y

1064 23 Nichols, R. J. *et al.* Discovery and characterization of a novel family of prokaryotic
1065   nanocompartments involved in sulfur metabolism. *Elife* **10** (2021).
1066   https://doi.org/10.7554/eLife.59288

1067 24 Giessen, T. W. Encapsulins. *Annu Rev Biochem* (2022). https://doi.org/10.1146/annurev-
1068   biochem-040320-102858

1069 25 Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. & Savage, D. F. Encapsulins: molecular
1070   biology of the shell. *Crit Rev Biochem Mol Biol* **52**, 583-594 (2017).
1071   https://doi.org/10.1080/10409238.2017.1337709

1072 26 Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat*
1073   *Struct Mol Biol* **15**, 939-947 (2008). https://doi.org/10.1038/nsmb.1473

1074 27 Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals
1075   functional diversity. *Nat Microbiol* **2**, 17029 (2017). https://doi.org/10.1038/nmicrobiol.2017.29

1076 28 Heinemann, J. *et al.* Fossil record of an archaeal HK97-like provirus. *Virology* **417**, 362-368
1077   (2011). https://doi.org/10.1016/j.virol.2011.06.019

1078 29 Altenburg, W. J., Rollins, N., Silver, P. A. & Giessen, T. W. Exploring targeting peptide-shell
1079   interactions in encapsulin nanocompartments. *Sci Rep* **11**, 4951 (2021).
1080   https://doi.org/10.1038/s41598-021-84329-z

1081 30 Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat*
1082   *Struct Mol Biol* **15**, 939-947 (2008). https://doi.org/10.1038/nsmb.1473

1083 31 McHugh, C. A. *et al.* A virus capsid-like nanocompartment that stores iron and protects bacteria
1084   from oxidative stress. *EMBO J* **33**, 1896-1911 (2014). https://doi.org/10.15252/embj.201488566

1085 32 Giessen, T. W. *et al.* Large protein organelles form a new iron sequestration system with high
1086   storage capacity. *Elife* **8** (2019). https://doi.org/10.7554/eLife.46070

1087 33 Akita, F. *et al.* The crystal structure of a virus-like particle from the hyperthermophilic archaeon
1088   Pyrococcus furiosus provides insight into the evolution of viruses. *J Mol Biol* **368**, 1469-1483
1089   (2007). https://doi.org/10.1016/j.jmb.2007.02.075

1090 34 Kwon, S., Andreas, M. P. & Giessen, T. W. Structure and heterogeneity of a highly cargo-loaded
1091   encapsulin shell. *J Struct Biol* **215**, 108022 (2023). https://doi.org/10.1016/j.jsb.2023.108022

1092 35 Eren, E. *et al.* Structural characterization of the Myxococcus xanthus encapsulin and ferritin-like
1093   cargo system gives insight into its iron storage mechanism. *Structure* **30**, 551-563 e554 (2022).
1094   https://doi.org/10.1016/j.str.2022.01.008

36   Jones, J. A., Andreas, M. P. & Giessen, T. W. Structural basis for peroxidase encapsulation inside the encapsulin from the Gram-negative pathogen Klebsiella pneumoniae. *Nat Commun* **15**, 2558 (2024). https://doi.org/10.1038/s41467-024-46880-x

37   Sugano, Y. & Yoshida, T. DyP-Type Peroxidases: Recent Advances and Perspectives. *Int J Mol Sci* **22** (2021). https://doi.org/10.3390/ijms22115556

38   Almeida, A. V. *et al.* Condensation and Protection of DNA by the Myxococcus xanthus Encapsulin: A Novel Function. *Int J Mol Sci* **23** (2022). https://doi.org/10.3390/ijms23147829

39   Benisch, R., Andreas, M. P. & Giessen, T. W. A widespread bacterial protein compartment sequesters and stores elemental sulfur. *Science Advances* **10**, eadk9345 (2024). https://doi.org/doi:10.1126/sciadv.adk9345

40   Kwak, J., McCue, L. A., Trczianka, K. & Kendrick, K. E. Identification and characterization of a developmentally regulated protein, EshA, required for sporogenic hyphal branches in Streptomyces griseus. *J Bacteriol* **183**, 3004-3015 (2001). https://doi.org/10.1128/JB.183.10.3004-3015.2001

41   Saito, N., Matsubara, K., Watanabe, M., Kato, F. & Ochi, K. Genetic and biochemical characterization of EshA, a protein that forms large multimers and affects developmental processes in Streptomyces griseus. *J Biol Chem* **278**, 5902-5911 (2003). https://doi.org/10.1074/jbc.M208564200

42   Saito, N. *et al.* EshA accentuates ppGpp accumulation and is conditionally required for antibiotic production in Streptomyces coelicolor A3(2). *J Bacteriol* **188**, 4952-4961 (2006). https://doi.org/10.1128/JB.00343-06

43   Jones, J. A., Benisch, R. & Giessen, T. W. Encapsulin cargo loading: progress and potential. *J Mater Chem B* **11**, 4377-4388 (2023). https://doi.org/10.1039/d3tb00288h

44   Berman, H. M. *et al.* The cAMP binding domain: an ancient signaling module. *Proc Natl Acad Sci U S A* **102**, 45-50 (2005). https://doi.org/10.1073/pnas.0408579102

45   Kannan, N. *et al.* Evolution of allostery in the cyclic nucleotide binding module. *Genome Biol* **8**, R264 (2007). https://doi.org/10.1186/gb-2007-8-12-r264

46   Krol, E., Werel, L., Essen, L. O. & Becker, A. Structural and functional diversity of bacterial cyclic nucleotide perception by CRP proteins. *microLife* **4** (2023). https://doi.org/10.1093/femsml/uqad024

47   Popovych, N., Tzeng, S.-R., Tonelli, M., Ebright, R. H. & Kalodimos, C. G. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proceedings of the National Academy of Sciences* **106**, 6927-6932 (2009). https://doi.org/doi:10.1073/pnas.0900595106

48   Yariv, B. *et al.* Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. *Protein Sci* **32**, e4582 (2023). https://doi.org/10.1002/pro.4582

49   Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350 (2016). https://doi.org/10.1093/nar/gkw408

50   Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-533 (2010). https://doi.org/10.1093/nar/gkq399

51   Belduz, A. O., Lee, E. J. & Harman, J. G. Mutagenesis of the cyclic AMP receptor protein of Escherichia coli: targeting positions 72 and 82 of the cyclic nucleotide binding pocket. *Nucleic Acids Res* **21**, 1827-1835 (1993). https://doi.org/10.1093/nar/21.8.1827

52   Gronenborn, A. M., Sandulache, R., Gärtner, S. & Clore, G. M. Mutations in the cyclic AMP binding site of the cyclic AMP receptor protein of Escherichia coli. *Biochem J* **253**, 801-807 (1988). https://doi.org/10.1042/bj2530801

1142 53 Moore, J., Kantorow, M., Vanderzwaag, D. & McKenney, K. Escherichia coli cyclic AMP receptor
1143 protein mutants provide evidence for ligand contacts important in activation. *J Bacteriol* **174**,
1144 8030-8035 (1992). https://doi.org/10.1128/jb.174.24.8030-8035.1992
1145 54 Lanzilotta, W. N. *et al.* Structure of the CO sensing transcription activator CooA. *Nature*
1146 *Structural Biology* **7**, 876-880 (2000). https://doi.org/10.1038/82820
1147 55 Joyce, M. G. *et al.* CprK Crystal Structures Reveal Mechanism for Transcriptional Control of
1148 Halorespiration *. *Journal of Biological Chemistry* **281**, 28318-28325 (2006).
1149 https://doi.org/10.1074/jbc.M602654200
1150 56 Zhao, M.-X. *et al.* Structural basis for the allosteric control of the global transcription factor NtcA
1151 by the nitrogen starvation signal 2-oxoglutarate. *Proceedings of the National Academy of*
1152 *Sciences* **107**, 12487-12492 (2010). https://doi.org/doi:10.1073/pnas.1001556107
1153 57 Brelidze, T. I., Carlson, A. E., Sankaran, B. & Zagotta, W. N. Structure of the carboxy-terminal
1154 region of a KCNH channel. *Nature* **481**, 530-533 (2012). https://doi.org/10.1038/nature10735
1155 58 Marques-Carvalho, M. J. *et al.* Structural, Biochemical, and Functional Characterization of the
1156 Cyclic Nucleotide Binding Homology Domain from the Mouse EAG1 Potassium Channel. *Journal*
1157 *of Molecular Biology* **423**, 34-46 (2012).
1158 https://doi.org/https://doi.org/10.1016/j.jmb.2012.06.025
1159 59 Haitin, Y., Carlson, A. E. & Zagotta, W. N. The structural mechanism of KCNH-channel regulation
1160 by the eag domain. *Nature* **501**, 444-448 (2013). https://doi.org/10.1038/nature12487
1161 60 van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nature*
1162 *Biotechnology* (2023). https://doi.org/10.1038/s41587-023-01773-0
1163 61 Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One
1164 hundred discoveries. *Protein Science* **32**, e4519 (2023).
1165 https://doi.org/https://doi.org/10.1002/pro.4519
1166 62 Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc*
1167 *Bioinformatics* **72**, e108 (2020). https://doi.org/10.1002/cpbi.108
1168 63 Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New
1169 HHpred Server at its Core. *J Mol Biol* **430**, 2237-2243 (2018).
1170 https://doi.org/10.1016/j.jmb.2017.12.007
1171 64 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods* **19**, 679-
1172 682 (2022). https://doi.org/10.1038/s41592-022-01488-1
1173 65 Cassidy-Amstutz, C. *et al.* Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading
1174 of Encapsulin. *Biochemistry* **55**, 3461-3468 (2016).
1175 https://doi.org/10.1021/acs.biochem.6b00294
1176 66 Faylo, J. L. *et al.* Structural insight on assembly-line catalysis in terpene biosynthesis. *Nat*
1177 *Commun* **12**, 3487 (2021). https://doi.org/10.1038/s41467-021-23589-9
1178 67 Ronnebaum, T. A., Eaton, S. A., Brackhahn, E. A. E. & Christianson, D. W. Engineering the
1179 Prenyltransferase Domain of a Bifunctional Assembly-Line Terpene Synthase. *Biochemistry* **60**,
1180 3162-3172 (2021). https://doi.org/10.1021/acs.biochem.1c00600
1181 68 Ronnebaum, T. A., Gupta, K. & Christianson, D. W. Higher-order oligomerization of a chimeric
1182 αβγ bifunctional diterpene synthase with prenyltransferase and class II cyclase activities is
1183 concentration-dependent. *Journal of Structural Biology* **210**, 107463 (2020).
1184 https://doi.org/https://doi.org/10.1016/j.jsb.2020.107463
1185 69 Gao, K., Oerlemans, R. & Groves, M. R. Theory and applications of differential scanning
1186 fluorimetry in early-stage drug discovery. *Biophys Rev* **12**, 85-104 (2020).
1187 https://doi.org/10.1007/s12551-020-00619-2

1188	70	Dremier, S., Kopperud, R., Doskeland, S. O., Dumont, J. E. & Maenhaut, C. Search for new cyclic AMP-binding proteins. *FEBS Letters* **546**, 103-107 (2003). https://doi.org/https://doi.org/10.1016/S0014-5793(03)00561-1
1191	71	Bertinetti, D. *et al.* Chemical tools selectively target components of the PKA system. *BMC Chemical Biology* **9**, 3 (2009). https://doi.org/10.1186/1472-6769-9-3
1193	72	Endoh, T. & Engel Joanne, N. CbpA: a Polarly Localized Novel Cyclic AMP-Binding Protein in Pseudomonas aeruginosa. *Journal of Bacteriology* **191**, 7193-7205 (2009). https://doi.org/10.1128/jb.00970-09
1196	73	Jäger, A. V. *et al.* Identification of novel cyclic nucleotide binding proteins in Trypanosoma cruzi. *Molecular and Biochemical Parasitology* **198**, 104-112 (2014). https://doi.org/https://doi.org/10.1016/j.molbiopara.2015.02.002
1199	74	Jones, J. A., Andreas, M. P. & Giessen, T. W. Exploring the Extreme Acid Tolerance of a Dynamic Protein Nanocage. *Biomacromolecules* **24**, 1388-1399 (2023). https://doi.org/10.1021/acs.biomac.2c01424
1202	75	Ross, J. *et al.* Pore dynamics and asymmetric cargo loading in an encapsulin nanocompartment. *Sci Adv* **8**, eabj4461 (2022). https://doi.org/10.1126/sciadv.abj4461
1204	76	Adamson, L. S. R. *et al.* Pore structure controls stability and molecular flux in engineered protein cages. *Sci Adv* **8**, eabl7346 (2022). https://doi.org/10.1126/sciadv.abl7346
1206	77	Joyce, M. G. CprK Crystal Structures Reveal Mechanism for Transcriptional Control of Halorespiration*. (2006).
1208	78	Ober, V. T. *et al.* Purine nucleosides replace cAMP in allosteric regulation of PKA in trypanosomatid pathogens. *eLife* **12**, RP91040 (2024). https://doi.org/10.7554/eLife.91040
1210	79	Bachmaier, S. *et al.* Nucleoside analogue activators of cyclic AMP-independent protein kinase A of Trypanosoma. *Nature Communications* **10**, 1421 (2019). https://doi.org/10.1038/s41467-019-09338-z
1213	80	Zallot, R., Oberg, N. & Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**, 4169-4182 (2019). https://doi.org/10.1021/acs.biochem.9b00735
1217	81	Oberg, N., Zallot, R. & Gerlt, J. A. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *Journal of Molecular Biology* (2023). https://doi.org/10.1016/j.jmb.2023.168018
1220	82	Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* **47**, W260-W265 (2019). https://doi.org/10.1093/nar/gkz303
1222	83	Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-780 (2013). https://doi.org/10.1093/molbev/mst010
1225	84	Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490 (2010). https://doi.org/10.1371/journal.pone.0009490
1227	85	Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* **26**, 1641-1650 (2009). https://doi.org/10.1093/molbev/msp077
1230	86	Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452-456 (2018). https://doi.org/10.1038/s41586-018-0043-0
1232	87	Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296 (2021). https://doi.org/10.1093/nar/gkab301
1234	88	Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003). https://doi.org/10.1101/gr.1239303

, , ND.0 International license.

preprint doi: https://doi.org/10.1101/2024.04.23.590730; this version posted April 23, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

89 Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic acids research* **50**, W276-W279 (2022). https://doi.org/10.1093/nar/gkac240

90 Andreas, M. P. & Giessen, T. W. Heterologous expression and purification of encapsulins in Streptomyces coelicolor. *MethodsX* **9**, 101787 (2022). https://doi.org/10.1016/j.mex.2022.101787

91 Andreas, M. P. & Giessen, T. W. Cyclodipeptide oxidase is an enzyme filament. *bioRxiv*, 2023.2009.2025.559410 (2023). https://doi.org/10.1101/2023.09.25.559410

92 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).

93 Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41-60 (2005). https://doi.org/10.1016/j.jsb.2005.03.010

94 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). https://doi.org/10.1038/nmeth.4169

95 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2

96 UniProt, C. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531 (2023). https://doi.org/10.1093/nar/gkac1052

97 Pettersen, E. F. *et al.* UCSF ChimeraX : Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70-82 (2021). https://doi.org/10.1002/pro.3943

98 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486-501 (2010). https://doi.org/10.1107/S0907444910007493

99 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004). https://doi.org/10.1107/S0907444904019158

100 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019). https://doi.org/10.1107/S2059798319011471

101 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010). https://doi.org/10.1107/S0907444909052925

Acknowledgements

Author contributions

M.P.A. and T.W.G. designed the project. M.P.A. and T.W.G conducted the bioinformatic and phylogenetic analyses. M.P.A. conducted all laboratory experiments. M.P.A. collected cryo-EM data, and M.P.A. and T.W.G. processed and analyzed the cryo-EM data. M.P.A. built the structural models. M.P.A. and T.W.G. wrote the manuscript. T.W.G. oversaw the project in its entirety.

Competing interests

The authors declare no competing interests.