Structural analysis unravels the functional promiscuity of Quinolone synthase-mediated polyketide biosynthesis in *Aegle marmelos* Correa

Mallika Vijayanathan[1,5], KV Abhinav[2,6], Debashree Bandyopadhyay[3], Kozhinjampara R Mahendran[1], Abdoallah Sharaf[5,7], M Radhakrishna Pillai[4] and EV Soniya [1]*

[1]Transdisciplinary Research Program, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram-14, India.

[2]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka, India

[3]Department of Biological Sciences, BITS Pilani, Hyderabad campus, India.

[4]Cancer Research Program, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, India

[5]Biology Centre, Czech Academy of Sciences, Institute of Plant Molecular Biology, České Budějovice, 370 05, Czech Republic

[6]Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck, University of London, United Kingdom

[7]Genetic Department, Faculty of Agriculture, Ain Shams University, Cairo, 11241, Egypt

[5,6]Present address

*Corresponding author: evsoniya@rgcb.res.in

Abstract:

Quinolone synthase from *Aegle marmelos* (AmQNS) is a type III polyketide synthase that yields therapeutically effective quinolone and acridone compounds. Based on the high-resolution protein structure of AmQNS, this study provided a mechanistic explanation of the structure to synthetic selectivity. Additionally, it displays the comparatively wide active site entry that allows the catalytic pocket to accommodate bulky substrates, which affects the enzyme catalysis. We also develop a model framework for comprehending the structural constraints on ketide insertion, and postulate that AmQNS synthetic diversity is owing to its steric and electrostatic selectivity, which allows it to bind to a variety of core substrates. We further establish that AmQNS is structurally biased toward quinolone synthesis and only synthesizes acridone when malonyl-CoA concentrations are significantly high. In a nutshell, we anticipate that addressing the structural and molecular underpinnings of AmQNS–substrate interaction in terms of its high selectivity and specificity can aid in the development of numerous novel compounds. Besides, the approaches can also be expanded to other potential enzymes, which will help the pharmaceutical sector by expanding the pool of potential medication leads.

Keywords: Quinolone, PKS, Electrostatic, Enzyme, Natural products

Introduction

Polyketides (PKs) are chemically diverse natural products with immense pharmaceutical properties. PKs and their possible derivatives could be used as attractive starting points for the development of new bioactive molecules with clinical applications[1,2]. Polyketide synthases (PKS) are multifunctional enzymes that synthesize PKs in plants, fungi, and bacteria[3]. Due to their remarkable characteristic features like i) wide substrate affinity, ii) alternating condensation steps, and iii) formation of diverge cyclic intermediates[4], the PKS machinery is a good target for producing architecturally diverge natural products by protein engineering and combinatorial biosynthesis. Based on the protein architecture and the mechanism of action, there are three types of PKS, namely type I, II and III. Typically, a polyketide is produced by consecutive addition of 'malonate building blocks' to a starter substrate (acyl thioester), catalysed by type III PKS[5,6]. When compared to type I and type II enzymes, type III PKSs are homodimers and comparatively smaller in size. Each functional unit of type III enzyme contains two ketosynthase (KS) domains ($\sim$ 40–45 kDa, ~350-390 amino acids per monomeric unit)[7–9]. Type III PKSs are further categorized into two subtypes, viz, the i) chalcone-forming (chalcone synthase (CHS)) and ii) non-chalcone-forming (non-CHS), based on the reaction they catalyze[10,11].

Quinolone synthase (AmQNS)[12] from the Indian bael tree (*Aegle marmelos* (L.) Correa. or *Crateva marmelos*) belongs to the non-CHS group of type III PKS. *N*-methyl anthraniloyl-CoA is the natural substrate for AmQNS, and the major products are quinolones and acridones. These anthranilic acid-derived quinolone alkaloids (quinine, chloroquine, etc.) have previously been found to have antibacterial, anticancer, and antiviral properties,[13,14] and could be used as potential pharmacological leads. In general, AmQNS yields diketide 4-hydroxy 1-methyl 2-quinolone (89%) through a one-step condensation reaction between *N*-methyl anthraniloyl-CoA and malonyl-CoA. Acridone (11%) is synthesized in a three-step condensation process that begins with the same substrate and utilizes the same enzyme[12]. When P-coumaroyl CoA is employed as the starting substrate, AmQNS can also produce benzalacetone. The structure and function of two AmQNS-analogue type-III PKSs that use *N*-methyl anthraniloyl-CoA as the starter substrate (*Citrus microcarpa* acridone synthase (CmACS) and quinolone synthase (CmQNS)) have been well characterized[15]. Despite the considerable sequence and structural similarities between AmQNS, CmACS, and CmQNS, their product formation patterns, and catalytic efficiencies are significantly different[12]. i.e., even a minor amino acid substitution can have a significant effect on enzymatic activity. It is remarkable to note that AmQNS and its

69 nearest homolog CmACS both have distinct amino acid variations that favour interaction with

70 the bulky N-methyl anthraniloyl CoA, while hindering the binding of small substrate CoAs. As

71 a result, designing these enzymes to have this mechanical behaviour could improve their

72 biocatalytic characteristics. AmQNS structural and functional features, as well as its reaction

73 mechanism, must be understood in order to optimize its biosynthetic potential for metabolic

74 engineering reprogramming to accelerate natural product discovery.

75

76 Here we provide the high-resolution crystal structures of AmQNS in native and substrate-bound

77 forms, as well as the structural and molecular underpinnings for its synthetic selectivity. We

78 used molecular simulations to identify rate-limiting reaction steps leading to quinolone and

79 acridone structure, as well as quantum chemical transition state calculations to compare the

80 relative kinetic barriers and thermodynamic enthalpies of substrates, clearly demonstrating that

81 AmQNS structurally favours quinolone production. Ultimately, these structural findings,

82 together with its simulation-based reaction studies, uncover the mechanistic behaviour of

83 AMQNS and will eventually assist to engineer and repurpose the enzymatic reaction to expand

84 the natural product reservoir for bioprospecting and drug discovery in the future.

85 **Results and Discussion**

86 **Sequence conservation, evolutionary positioning, and substrate selectivity**

87 In this study, during the homologue search, out of the 283 studied species, type III PKSs were

88 identified in 112 (39.6%) species, and in agreement with previous studies, they are well-

89 conserved in the green lineage (Viridiplantae) and Opisthokonta (Fungi)[3,16](**Supplementary**

90 **Table S1C**). Type III PKS were identified in representatives of Rhizaria, Alveolata, and

91 Stramenopila lineages, and its presence on some marine microalgae was also reported

92 previously[17]. It is noteworthy that type III PKS homologs have only been identified in the

93 biflagellated, unicellular, free-living diplonemid *Diplonema papillatum*, among the members

94 of the Discoba - a lineage currently placed proximal to the root of eukaryotes. Besides, we could

95 not identify type III PKS in any of the studied species of Metamonada, Amoebozoa,

96 Glaucocystophyceae, Rhodophyta, Rhodelphea, Rhodelphea, and Haptophyta

97 (**Supplementary Table S1C**). To explore the eukaryotic and prokaryotic type III PKS

98 evolutionary relationship, we traced back to the enzymes in the prokaryotes. Type III PKS were

99 identified in 36 out of the 136 bacterial species studied (26.5%), representing 26 different

100 taxonomic groups (**Supplementary Table S1D**). We searched type III PKS in Archaea and the

101 first time in the Archaeal-Asgard group. As per the previous report[3], we were unable to identify

102   type III PKS in all archaeal groups but could get homologs in one Asgard species (*Candidatus*

103   *Thorarchaeota archaeon*) (**Supplementary Table S1D, Supplementary Figure S1**). The

104   Asgard (or Asgardarchaeota) group is a separate domain of life representing the closest

105   prokaryotic relatives of eukaryotes[18,19]. These finding emphasizes the Asgardarchaeota group

106   might be the type III PKS enzyme's emergence point in the tree of life. In addition to the

107   aforementioned categories, type III PKS homologs from all Rutaceaen species were included

108   in our evolutionary analysis (**Supplementary Table S1B**).
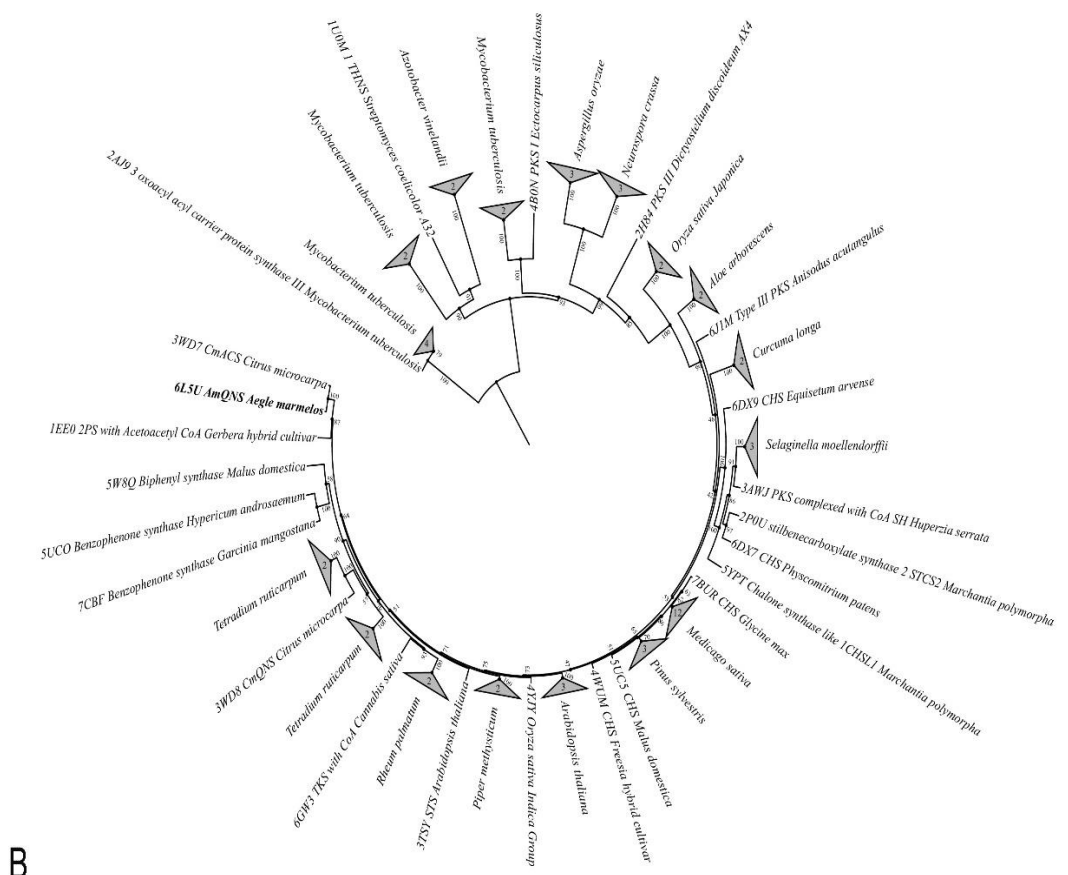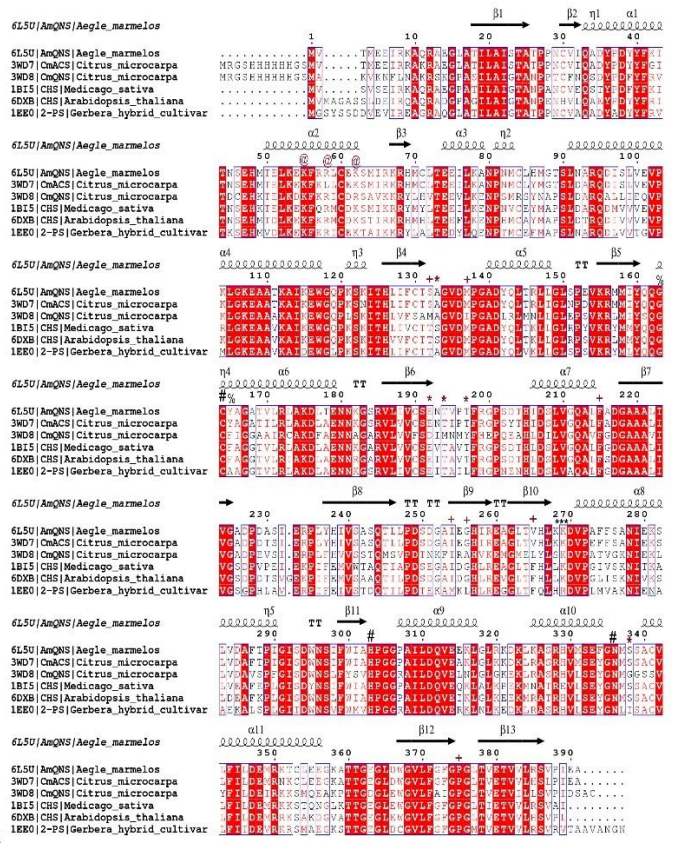
109

110   The evolutionary positioning of AmQNS is depicted in the phylogram of all the available

111   homologous proteins (**Figure 1**). AmQNS, like other type III PKS from members of the

112   Viridiplantae, is well conserved and grouped together with other members of the Rutacean

113   family, according to previous evolutionary studies[20,21]. We detected four interesting horizontal

114   gene transfer (HGT) events (**Figure 1**), two of them are within the bacterial domain and close

115   to the root of the tree. First one, between most of the stramenopiles and the cyanobacterial

116   *Rivularia sp.* (WP_015119976.1) homolog. These results show that this class of enzyme has a

117   cyanobacterial origin. The second HGT was between all the alveolates, except *Durinskia*

118   *baltica* and Chlamydiae homologs. A third HGT was again between Chlamydiae homolog

119   (1444712.BN1013_00142) and the chromerid *Vitrella brassicaformis* homologs; forming a

120   sister group with alveolate *Durinskia baltica*, all fungi, Rhizaria and Discoba homologs and

121   some homologs of stramenopiles (**Figure 1**). A recent report highlighted the contribution of

122   Chlamydiae on the evolution of eukaryotes[22]. Interestingly, the last HGT was in form that all

123   the Planctomycetes and cyanobacterial *Synechococcus sp.* Our analysis shows that this enzyme

124   has a bacterial origin and indicates early origin or even its presence in the Last Eukaryote

125   Common Ancestor (LECA).

126

127   Regarding the conservation pattern among type III PKSs, even though structure-based sequence

128   alignment of AmQNS with its adjacent homologs revealed high-level sequence conservation

129   and high functional conservancy (**Figure 2**), minor amino acid differences, particularly in the

130   CoA binding/substrate binding/cyclization pocket area, have a significant impact on substrate

131   specificity, i.e., will change the product formation profile. Because small changes in the three-

132   dimensional structure might occasionally impact substrate selectivity, amino acid moieties in

133   the above-mentioned locations can govern the biological reactions (i.e., protein-ligand
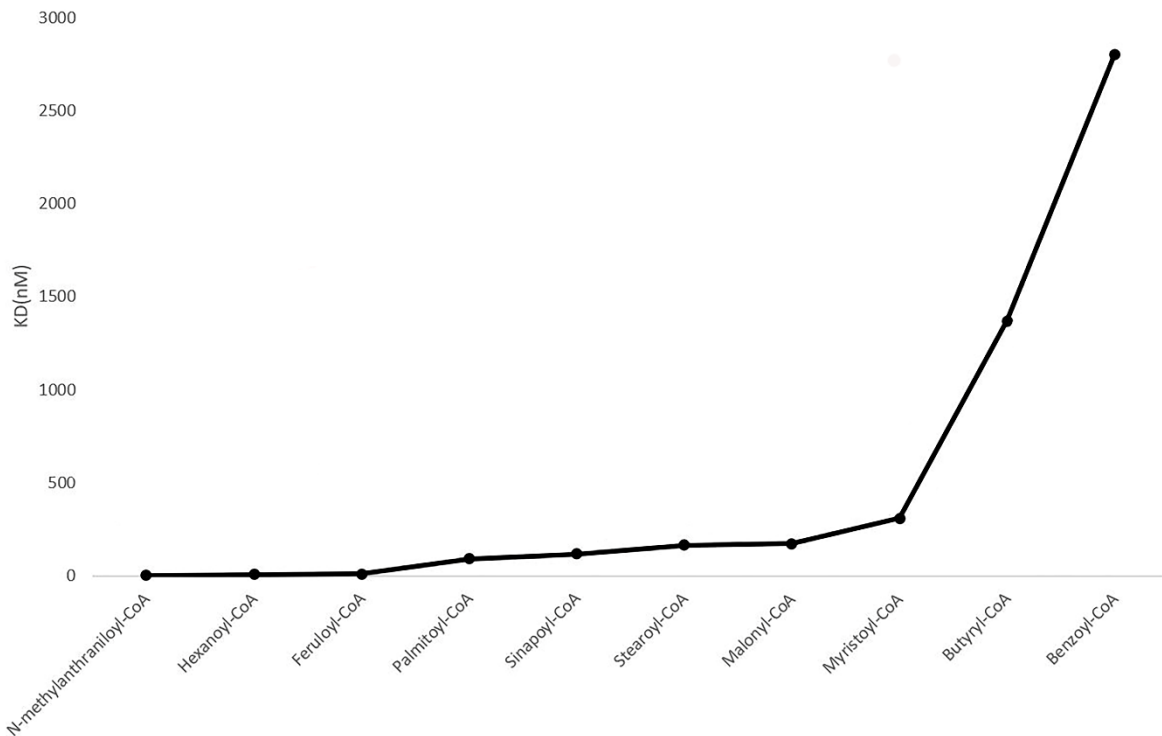
134   interactions).

135

136

137

Figure 1. Maximum likelihood (ML) phylogenetic tree of type III PKS protein homologs, showing the potential evolutionary relationships between the identified prokaryotic and different eukaryotic homologs. Phylogenetic position of AmQNS (6L5U) is highlighted in bold. The tree was rooted using the Asgard homologs, to identify the evolutionary direction of the proteins. The maximum likelihood branch support values are given in % (IQ-TREE/RAxML-NG).

144

145

146

147 **Figure 2: Sequence-structure alignment between different homologs and its evolutionary positions**. **A**)Alignment was prepared using
148 ClustalW [23] and ESPript 3.0 [24]. The PDB IDs are used to represent the sequences. The conservation level is shown by a color gradient (white-poor
149 conservation, red-high conservation). Significant residues are highlighted (#- Catalytic residues; *- Residues in the substrate-binding pocket; +-
150 Residues in the cyclization pocket), @- Residues in the CoA-binding tunnel; %- Residues adjacent to catalytic C164; *- β-turn region of AmQNS
151 **B**) ML phylogenetic tree showing the evolutionary relationship among the structural homologs in RCSB PDB.

152   One of the most essential aspects of enzymes that determines its unique reaction is the specific

153   interaction between proteins and ligands (such as substrates or cofactors). Radio-TLC

154   experiments have previously shown that AmQNS could accept multiple starter CoAs as

155   potential substrates[12]. Moreover, *in-silico* study also indicates that non-physiological substrates

156   could be employed as potential AmQNS ligands[20]. The binding mechanism of several acyl-

157   CoA substrates (small aliphatic to bulky aromatic) with AmQNS was further authenticated

158   using Surface Plasmon Resonance (SPR) based assays, which enable for real-time monitoring

159   of kinetic parameters[25]. The high affinity and reasonable interaction between the AmQNS and

160   small molecule ligands are indicated by the K_D values, which varied from high nanomolar to

161   low millimolar (2 nM-1 mM). AmQNS demonstrated a high affinity for *N*-methylanthraniloyl-

162   CoA, feruloyl-CoA, and hexanoyl-CoA (with K_D of 2.04 nM, 9.83 nM, and 7.30 nM,

163   respectively), and it is worth noting that AmQNS prefers bulkier substrates than short acyl-

164   CoAs (**Figure 3, Supplementary Figure S2**). These affinity parameters complement the

165   previously reported interaction studies using a thin-layer chromatography (TLC)[12], and when

166   comparing the steady state kinetic parameters for AmQNS with different starter substrate

167   CoA's, it's notable that Km values are higher than K_D for the majority of the substrates (for *N*-

168   methylanthraniloyl CoA-2.93 µM; p-coumaroyl CoA- 3.62 µM; Feruloyl CoA- 9.14 µM). This

169   suggests that catalysis is more rapid than dissociation. These findings imply the prospect of

170   utilizing various substrates to create novel chemical scaffolds, and the enzyme can be further

171   engineered to accommodate various substrates to boost the catalytic versatility.
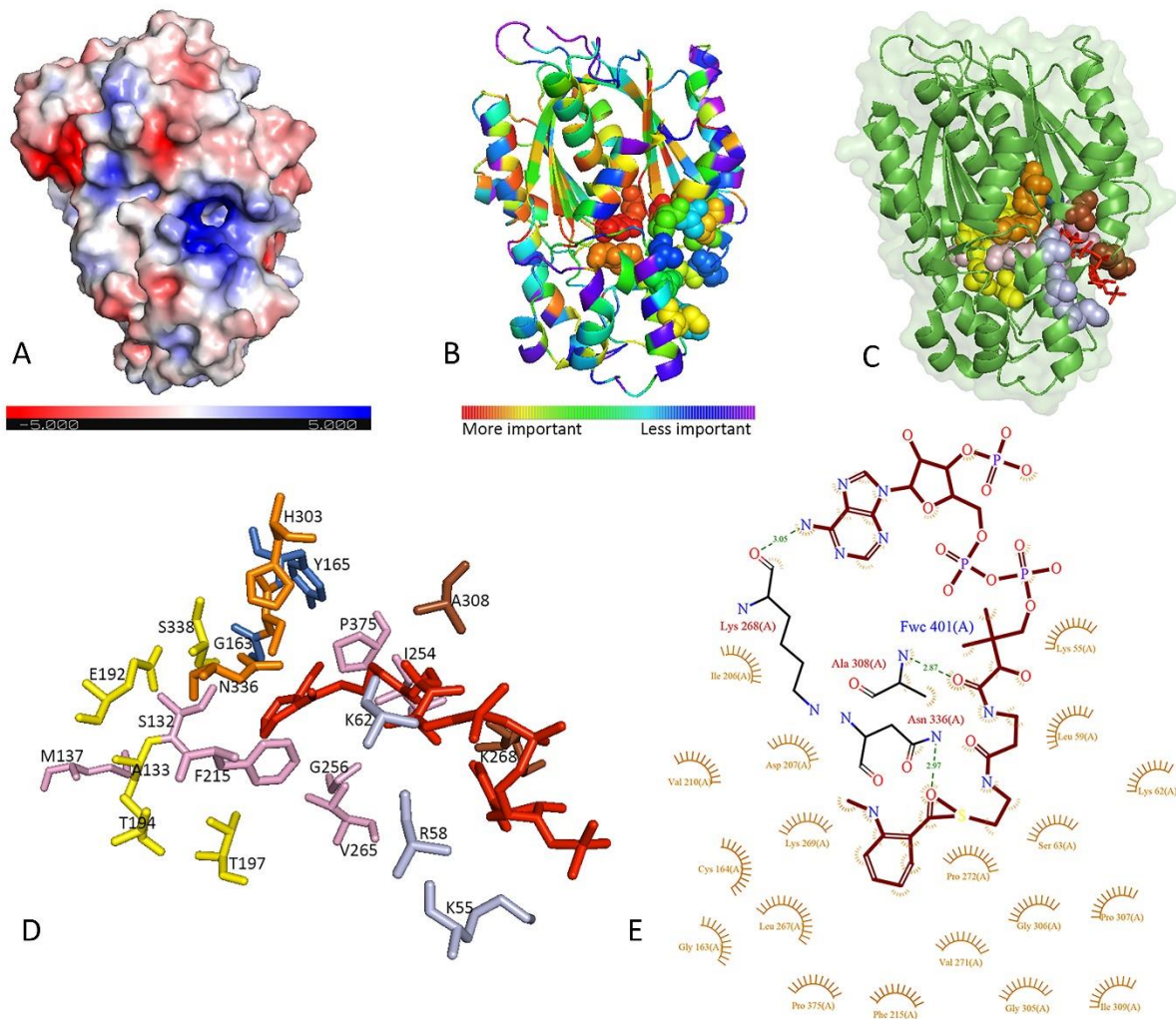
172

Figure 3. AmQNS interactions with various substrate CoAs. The lower the KD value, the better the interaction.

Molecular structure of AmQNS in the native and substrate-bound states.

The high resolution AmQNS crystals provided very precise protein models with well-defined electron density maps and the structural coordinates have been deposited in the RCSB Protein Data Bank[26]. The data collection and refinement statistics are given in **table 1**.

Figure 4: The structure of AmQNS in native and substrate bound form. **A)** AmQNS native structure (surface representation, PDB ID: 6L5U, Resolution 1.85Å) displaying electrostatic charge distribution (positively charged residues in blue and negatively charged residues in red **B)** Evolutionary trace on native AmQNS showing functionally significant residue positions in the structure. **C)** AmQNS- substrate bound form (cartoon and surface representation, PDB ID: 7CCT, Resolution 2.35Å). Red stick represents the substrate *N*-methylanthraniloyl CoA (MANT-CoA). All significant residues are highlighted (catalytic triad (C164, H303, N335)- orange; substrate-binding residues (A133, E192, T194, T197, S338)- yellow; residues in cyclization pocket (S132, M137, F215, I254, G256, V265, P375) -pink; Residues in CoA-binding tunnel (K55, R58, K62)- light blue; residues adjacent to the catalytic C164 (G163, Y165)-marine blue; other significant residues that form polar contacts with substrate (K268, A308)-brown. **D)** Binding environment of MANT-CoA in 7CCT (highlighted the substrate binding area of Figure 4C). MANT-CoA binding pattern (red sticks) and surrounding residues are highlighted. **E)** LIGPLOT of interactions involving ligand MANT-CoA and surrounding residues.

196   AmQNS has a native structure that is similar to other types III PKSs in terms of structural

197   folds (**Figure 4A**), with a unique structural topology that includes a specific upper domain

198   'αβαβα' (ketosynthase domain)[27], that is conserved in all structural homologs and a lower

199   domain that contains the majority of the substrate-binding residues (A133, E192, T194, T197,

200   S338). In AmQNS monomer, these domains are made up of three beta sheets (13 strands

201   (22.4%), 16 helices (37.6%), 3-10 helices (2.7%), and other secondary structures (37.3%-

202   including four beta hairpins, four beta bulges, 30 beta turns, two gamma turns, and other

203   structures) (**Figure 3, supplementary figure S3**). The protein is functionally active in dimeric

204   form, and in each monomer, the β-sheets are organized into two antiparallel β-sheets and one

205   mixed sheet, where the strands are arranged in the AmQNS structure's core, whereas the α-

206   helices are distributed on the surface. The prospective substrate-binding pocket entrance of each

207   AmQNS monomeric unit is bordered by the side chains of the α-helices and β-strands. Despite

208   being structurally comparable even at the active site entrance, AmQNS has a considerably

209   bigger binding pocket (volume wise) than its nearest functional homologs from *Citrus X*

210   *Microcarpa* (PDB IDs: 3WD7 & 3WD8) (**Supplementary figure S4**). This could be the

211   consequence of the single phenylalanine to valine substitution (F265V), where the smaller

212   valine (V) frees up more space in the active site pocket, and the longer tunnel enables the entry

213   of bulky substrates (e.g., *N*-methylanthraniloyl-CoA).

214

215   The orientation and locations of the catalytic residues in AmQNS are comparable to those in

216   the closest homologs (**Supplementary Figure S5**). CmACS did not display many differences

217   when analysing the substrate-binding residues of AmQNS. Nonetheless, CmQNS showed

218   modest variations, as demonstrated by changes in the cavity volume metrics (reduced

219   parameters) (**Supplementary Figure S4 & S6**), and even though these sequence alterations

220   and its orientations are minimal, it will result in potential differences in pocket volume, leads

221   to a change in product formation profile. AmQNS prefers bulkier substrates, as previously

222   stated, and the electrostatic potential surface calculation revealed that the substrate-binding

223   pocket regions of AmQNS have a predominantly positive charge (**Figure 4A**), which facilitates

224   binding with the phosphate groups of the preferred starter substrate CoAs. These positions of

225   positive charges are consistently found across type III PKSs as binding of the CoA portion of

226   substrates is conserved. Salt bridges (~6), hydrogen bonds (~32), and nonbonded interactions

227   (~318) connect the hydrophobic and hydrophilic residues that make up the AmQNS dimeric

228   interface area (2463-2481 Å), and salt bridges at the protein interface help to stabilize the

229   protein. The residues involved in forming prospective salt bridges are D96, D136, D251, H257,

230 R259, and K281 from chain A and R259, H257, R146, D136, D96, and E153 from chain B.

231 The catalytic triad (C164-H303-N336) is located in the upper domain, and are deeply embedded

232 within the entrance cavity like their homologs[15]. The orientation and position of these residues

233 are strikingly similar to those of the homologs to some extent. Furthermore, cysteine in the

234 catalytic triad (C164) is highly nucleophilic (as determined by the pKa, due to the reactivity of

235 its thiol (S-H) group), and that this residue is primarily responsible for thioester exchange

236 reactions[28]. The reduction of the sulfur donor molecule in enzyme catalysis is significant since

237 it binds to the substrate, and it is worth noting that 'thiol groups of cysteines' are typically found

238 at active sites. This is consistent with prior studies[29], where the catalytic cysteine in AmQNS is

239 oxidized to sulfinic acid, showing that it has a higher nucleophilicity and is more vulnerable to

240 oxidation[30,31]. Interestingly, molecular evolution also plays a crucial role in maintaining the

241 active-site environment of type III PKS proteins. According to Liou et al.,[31] CHSs from basal

242 land plants (bryophytes, lycophytes, etc.) have fewer reactive catalytic cysteines than CHSs

243 from higher plants. It is unclear whether these findings regarding the modulation of catalytic

244 cysteine reactivity represents a general pattern in non-chalcone forming PKS family members

245 too. However, AmQNS has a highly nucleophilic cysteine (Cys164) in the catalytic region,

246 indicating that it might have evolved to have a high catalytic potential.

247 The amino acid residues K268, A308, and N336 form polar contacts (the distance of 3.05Å,

248 2.87Å, and 2.97Å, respectively) with the CoA molecule in substrate-bound AmQNS (**Figure**

249 **4B-D**). The interaction with *N*-methyl anthraniloyl CoA (MANT-CoA) was also confirmed by

250 the presence of 76 nonbonded interactions. One of these residues, N336, is a catalytic site,[32]

251 and the interactions between all of these residues imply that the main substrate binding sites are

252 in the phosphate region. Likewise, K55, L267, G305, and A308 establish hydrogen bonds (at

253 distances of 2.51Å, 2.93Å, 3.11Å, and 3.06Å, respectively) in the CoASH (byproduct)-bound

254 form (PDB ID: 6L7J, resolution 1.8, **Supplementary Figure S7**). K55 is located in the CoA-

255 binding tunnel at the entrance, and G305 has previously been reported to play a role in shaping

256 the appropriate geometry of the active site pocket[33] (**Supplementary Figure S7A, B**).

257 Additionally, thermal disorder parameters might indicate conformational flexibility[33], and we

258 observed that ligand binding causes well-defined conformational changes in proteins,

259 particularly in the β-turn region of AmQNS (residues K268-K269-D270). D270 is absolutely

260 conserved in all aligned proteins, and K269 is mostly conserved (**Figure 2A**), however only

261 AmQNS and CmACS maintain the K268. CmQNS has a K268 to S268 alteration, while other

262 homologs have 'K268 to L268'. These regions are more flexible, and comparison studies

263   indicated conformational flexibility at the substrate-binding pocket entrance in AmQNS, which

264   suggested hinge-like movement of the surface loop. This flexibility in the AmQNS enzyme

265   structure provides a larger passageway for a substrate to enter the internal active binding site,

266   which is more evident by the followed simulation experiments.

267   **Structural basis for AmQNS synthetic selectivity**

268   To gain insights into the reaction mechanism, followed by the structural elucidation, molecular

269   simulation studies were used to investigate the mechanistic basis of AmQNS synthetic

270   selectivity. Here we examined if specific ligand-protein interactions can be mapped to

271   characterize the enzyme's relative propensity to select an optimal number of intermediate ketide

272   insertions. We calculated transition states for MANT-CoA binding to AmQNS and defined the

273   three reaction steps (**Figure 5**) required for AmQNS-driven quinolone production. The first

274   step entails a classic Sn2 thiol addition[34,35], through which the MANT-CoA substrate binds to

275   the catalytic C164. The second reaction depicts a ketide unit's concerted process from malonyl-

276   CoA inserts between the cysteine sulfur and the carbonyl carbon of the substrate enzyme

277   complex. The third reaction is then a reverse substitution through which the substrate amine

278   induces product ring closure, which restores the enzymatic cysteine (**Figure 5A**). The chemical

279   structure of MANT-CoA, its derivatives and products (quinolone/ acridones) are given in

280   **Supplementary Figure S8.** The activation energy and enthalpy for each step of the reaction

281   process are provided in **Table 2**.

282

283   The multiple transition states[36–39] for initial complexation between the substrate and enzyme

284   were then demonstrated (**Figure 5B**). Transition states tend to be the portion of any reactive

285   progression, where structural features have the most significant influence on kinetic properties

286   for the subsequent reaction. In the particular case of MANT-CoA binding to the AmQNS active

287   site, it is apparent that the substrate is a good fit for the enzyme, as there are minimal clashes

288   that could either kinetically disfavour or completely abrogate the subsequent reaction. Several

289   observations are made regarding areas on the substrate for which clash is so minimal that

290   (theoretically) increased substrate bulk might reduce the activation barrier (i.e., improve

291   reaction kinetics) through favourable van der Waals (vdW) or electrostatic interactions[40–42] .

292   The observation that the receptor is spacious around the aminomethyl substrate group led to the

293   notion of experimentally investigating whether (computationally) the aminomethyl group could

294   be productively modified as a chloro analog (somewhat bulkier in a potentially favourable

295   manner). Notably, the red and blue receptor patches in the MANT aryl ring region are similarly

296 motivated to explore if a slightly more polar version of the substrate (with a pyridinyl ring,

297 rather than benzyl) might produce kinetically favourable electrostatic complementarity.

298
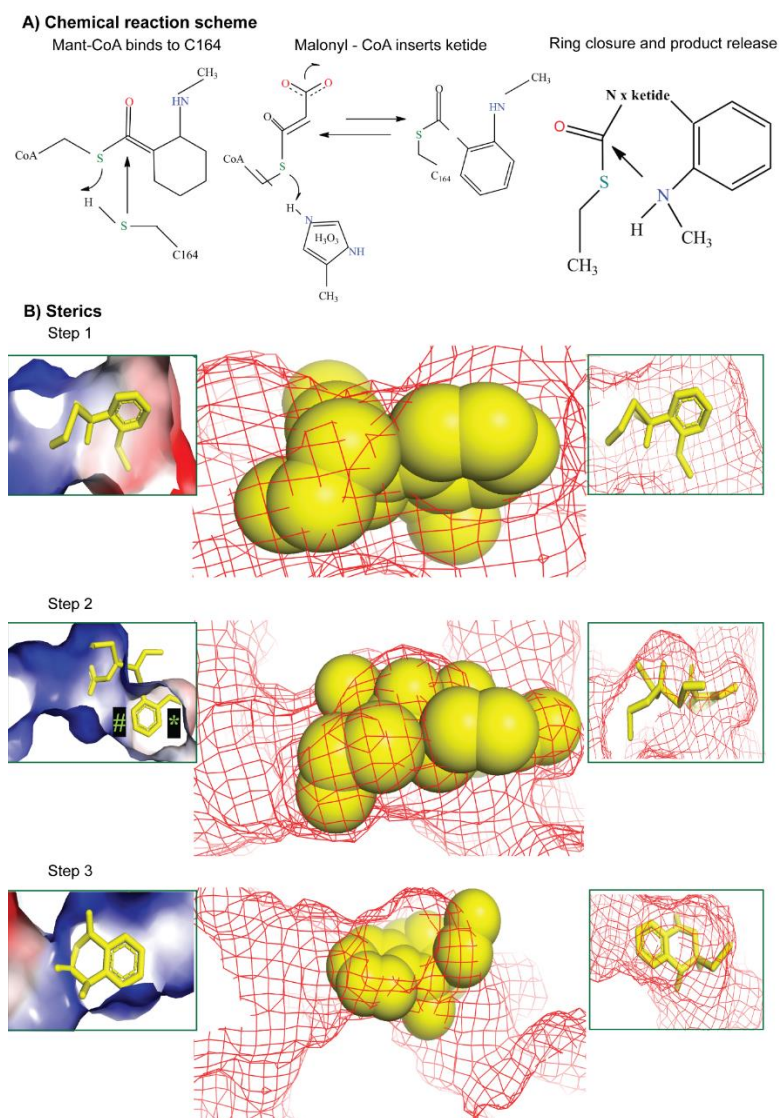
299

300 **Figure 5:** Three steps for AmQNS catalysis. **A)** Chemical reaction scheme showing the

301 substrate binding to the enzyme and subsequent product formation. **B)** Molecular models show

302 different transition state for complexation between the substrate and enzyme. The effect of

303 specific structural features (steric and electrostatics) on kinetic properties for the subsequent

304 reaction (ketide insertion) is represented in the molecular model ('#' and '*' markings convey

305 that the receptor poses negligible clashes with the substrate during the first ketide insertion.

306 Still, it should be noted that there is no significant excess of space available at positions '#' and

307 '*' in the first insertion. Consequently, during the process of second and third ketide insertions,

308 during which the reaction intermediate is growing, clashes would be expected at both positions

309 '#' and '*.'). Transition steps can be better viewed in **supplementary movies (S1-S3)**.

310

311   Notably, step 2 in the reaction sterics (**Figure 5B**) represents prospective receptor structural

312   influences on the first ketide insertion kinetics. In this case, the transition state for this reaction

313   is well accommodated by the receptor, which, in turn, corroborates the prior observation that

314   AmQNS is a viable 'enzymatic engine' for promoting quinolone synthesis. Importantly, we

315   propose that second and third ketide insertions may be somewhat less favored than the first

316   insertion. Nonetheless, during the first insertion, there is no considerable excess of space

317   available. As a result, clashes would be expected (in places '#' and '*') during the second and

318   third ketide insertions, when the reaction intermediate is growing. These clashes may be

319   somewhat defused with a ligand conformational shift that orients the ring slightly out of the

320   plane of this graphic as the aryl ring begins to progress toward the narrow product exit channel,

321   whose position is relatively well marked in the figure ('*'). In step 3, we see transitional

322   interactions between the forming quinolone product and the receptor. It is interesting to note

323   that although the receptor is not hugely antagonistic to product formation, it also does not seem

324   ideally suited, as apparent in the steric clash between the enzymatic surface and the

325   aminomethyl. This clash might be alleviated through a change of conformational twist (to

326   reorient the aryl ring) that is essentially the same factor identified earlier in reaction step 2 as a

327   requisite step for second or third ketide insertions. This has an exciting implication and, this

328   means that although the analysis of step 2 has pointed firmly toward smaller quinolone product

329   formation (compared to a larger acridone product), a kinetic hitch in the final step of quinolone

330   formation may nullify this difference. We suggest that structural modifications to the AmQNS

331   enzyme (e.g., potentially mutating Leu 263 into a smaller valine or alanine or removing the

332   methylamine clash by mutating Ser 132 into a glycine) might favour both the quinolone and

333   acridone product formation, potentially speeding the production of either while not necessarily

334   affecting the relative ratios of quinolone and acridone product. In **Table 2**, we show the

335   computed impacts of the two minor (chloro and pyridinyl) modifications to the MANT-CoA

336   substrate. Our data show that the substrate modifications appear to have only minor influence,

337   and it is difficult to predict if either shift will produce a demonstrable improvement in reactive

338   profile relative to unmodified MANT-CoA. Alternatively, we also suggest that AmQNS may

339   support the catalytic a variety analogs to the standard biologically processed substrates,

340   meaning that their synthetic chemistry can be extended from the production of novel natural

341   product scaffolds to a related display chemical analogs.

342

343   Next, we report acridone-specific reaction steps (**Figure 6A**) and the second and third ketide

344   insertions are predicted to be somewhat less favourable kinetically and thermodynamically

345   compared to the first ketide insertion shown in **figure 5**. In contrast, the final acridone ring

346   closure is expected to have a higher activation barrier than the quinolone product formation but

347   a more favourable reaction enthalpy. Finally, we investigated whether AmQNS is better suited

348   for quinolone or acridone production, and we propose that the key difference between the two

349   reactions may be a matter of stoichiometric control, with an excess of malonyl-CoA favoring

350   acridone and tight stoichiometry favoring quinolone. Furthermore, reducing steric bulk by

351   altering Leu 263 or Ser 132 could enhance throughput of both products, indicating that specific

352   amino acid changes could be used to impact enzymatic product selectivity. For instance, the

353   previously studied[12] AmQNS mutants MSD1 (double mutant, S132T/A133S) and MSD2 (triple

354   mutant, S132T/A133S/V265F) had drastically narrowed active site cavities when compared to

355   the wild-type AmQNS. MSD1 demonstrated chalcone-forming activity with p-coumaroyl-CoA

356   like the typical chalcone synthase, whereas MSD2 did not[12]. Since none of the mutants prefer

357   MANT-CoA as starter substrate, the two amino acid alterations S132T and A133S influenced

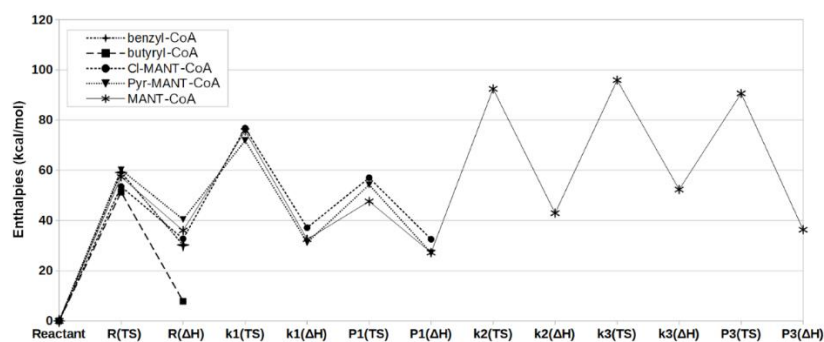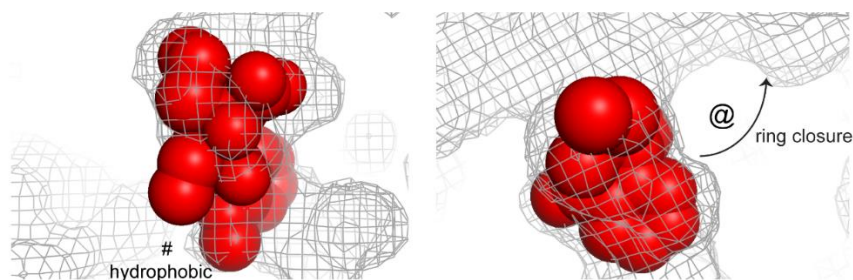358   the enzyme's substrate selectivity.
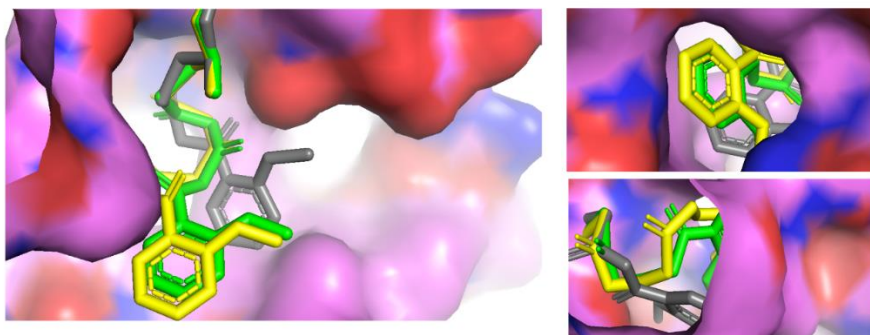
359

360

361

362

A) Transition State (TS) and product enthalpies for various AmQNS reactions

B) Steric and electrostatic interactions between AmQNS and MANT-CoA intermediates

C) Steric and Conformational considerations of mono-ketide, di-ketide and tri-ketide intermediates for the AmQNS catalytic generation of acridone from MANT-CoA

363

Figure 6: AmQNS mediated synthetic diversity based on the structure conformations.
A) Relative transition state (TS) and product enthalpies for various AmQNS reactions, relating to five distinct core substrates named in the legend (upper left; Cl-MANT refers to a chloromethyl analog to MANT; Pyr-MANT refers to a pyridinyl analog to MANT, with the heteroatomic N located para to the methylamine). Specific reaction state enthalpies are quantified for initial binding (R), specific polyketide insertions (k1, k2, k3) and product cleavages after 1st (P1) and third (P3) ketide insertions. **B)** Steric and electrostatic interactions between the AmQNS receptor (mesh) and MANT-CoA-based intermediates (spheres) for the ketide insertion (shown in left) and monoketide product forming (shown in the right) transition states. Steric effects relating to the bulky chlorine atom are represented as '@.' '#' indicating the hydrophobic pocket. **C)** Steric and conformational considerations of mono-ketide (yellow), di-ketide (pink) and tri-ketide (white) intermediates for the AmQNS catalytic generation of acridone from MANT-CoA substrate, as viewed from a cross-section of the whole receptor, the channel through which malonyl-CoA co=substrate enters, and the product release channel.

364 It is inherently challenging to fully characterize how type III polyketide quinolone synthases
365 achieve such impressive synthetic diversity from relatively minor structural variations among
366 different enzyme families. Here, we primarily focus on enzymatic steric and electrostatic
367 selectivity for binding other core substrate units (e.g., relative favorability for specialized
368 binding units such as coumaroyl-CoA, benzyl-CoA, acetyl-CoA, MANT-CoA, versus a
369 universal capacity to malonyl-CoA as a substrate or co-substrate), and the amount of space
370 available to accommodate larger numbers of incrementally inserted ketide units.   We
371 determined quantum chemical transition states to compare the relative kinetic barriers[43] and
372 thermodynamic enthalpies[44] for the initial complexation of MANT-CoA, benzyl-CoA, butyryl-
373 CoA and coumaroyl-CoA. Similar characterization was done, in the case of MANT-based
374 reactions, for the first ketide insertion, the single-ketide quinolone product formation, the
375 second and third ketide insertions, and the triple-ketide acridone product.

377   The initial complexation barrier showed little variation among primary substrates, implying
378 that steric dependencies play a minor impact at this stage (**Figure 6A**). However, steric and
379 electrostatics do appear to discriminate somewhat in the stability of the resulting bound
380 intermediates. Specifically, the pyridinyl analog to MANT-CoA has less stabilization than the
381 others because it places the slightly polar aryl nitrogen directly within a hydrophobic pocket
382 delimited by Ile 254 and Pro 375 ('#' in **Figure 6B**). Simultaneously, the sole flexible substrate
383 (butyryl-CoA) can conformationally adapt to this pocket in a stabilizing manner. Proceeding
384 from the mono-ketide intermediate (k1) to the monoketide product (P1) reflects minimal
385 difference among the three analogs of MANT-CoA, with the exception that the chloromethyl
386 compound has a higher barrier to ring closure, due to steric effects relating to the bulky chlorine
387 atom.

389 Next, in the **Figure 6A**, it showed the quantitative reaction profile of secondary and tertiary
390 ketide insertions by which the monoketide intermediate may progress toward the acridone
391 product (P3). It is worth noting that these latter ketide insertions are predicted to have higher
392 barriers (> 50 kcal/mol) than the first insertion (~40 kcal/mol).   From a computational
393 perspective, this trend is rationalized by higher conformational strain evident in the di-ketide
394 and tri-ketide units relative to the mono-ketide (**Figure 5B**), as opposed to issues relating to the
395 approach of malonyl-CoA co-substrate (for which there is ample space, as shown in Figure 6C)
396 or the situation of the MANT group (**Figure 5B**). Based on this, we propose that AmQNS

397 structurally favours the smaller quinolone product's production and might thus only produce

398 acridone under conditions of significant malonyl-CoA concentration.

399

400 In the future, we will compare analyses for non-MANT substrates. Furthermore, the specific

401 structures available for each of the five (for quinolone) transition states or five (for acridone)

402 reaction steps may be rigorously evaluated to determine which ligand-receptor amino acid

403 contacts are favourable or unfavourable. Also, determine what sorts of mutations could be

404 proposed to significantly alter favourability in a manner that could influence existing AmQNS

405 enzymatic activity and potentially engineer product specificity

406

407

408 **Materials and Methods**

409 **Homolog identification and phylogenetic analysis**.

410 To identify the potential homologs for AmQNS, complete predicted proteome sequences were

411 retrieved from JGI (http://genome.jgi.doe.gov) and NCBI GeneBank

412 (https://ncbi.nlm.nih.gov/). The Marine Microbial Eukaryote Transcriptome Sequencing

413 Project database (MMETSP)[45] were also used to predict the homologs. The *Arabidopsis*

414 *thaliana* chalcone synthase (AT5G13930) reference amino acid sequence was used as a query

415 to search of all potential homologs using the Hidden Markov model (HMM)-based tool

416 jackhammer[46]. Evolutionary genealogy of genes: Non-supervised Orthologous Groups

417 (eggNOG) mapper was used for hierarchical resolution of orthology assignments[47]. Finally, the

418 SMART and Pfam databases were employed to identify conserved domains present in type III

419 PKS from different organisms[48,49], both SMART and Pfam databases were merged, and

420 redundant domains were filtered-out and the Hidden Markov model (HMM)-based tool

421 hmmscan (https://github.com/EddyRivasLab/hmmer). Only sequences with the catalytic or

422 conserved domain of the references were retained. The identified homologs and the respective

423 details are given in the **supplementary table S1**. All identified homologs were aligned using

424 MAFFT[50] and ambiguously aligned regions were excluded for further analysis using trimAl

425 software[51]. Alignments were tested using ProtTest v3[52] to choose an appropriate model for

426 nucleotide substitution. Two separated Maximum likelihood (ML) phylogenetic trees were

427 computed using RAxML-NG[53] and IQ-TREE2[54]. ML analyses were performed using 1000

428 bootstrap replicates. The supporting values from both software were noted on the ML-unrooted

429 tree.

Large-scale AmQNS Expression and Purification

430

431 The QNS gene from 'A. marmelos' was cloned into pET32b as explained by Resmi et al.,

432 2013[12]. Transformed *Escherichia coli* BL21(DE3) cells were incubated at 37°C in Luria-

433 Bertani (LB) medium containing ampicillin (100 g/ml) until they reached the exponential phase

434 of growth (OD600 0.6). Isopropyl 1-thio—D-galactopyranoside (IPTG, 400 mM) was

435 employed to induce AmQNS expression, and the cells were further incubated at 28°C for 5-6

436 hours. At 4°C, all phases of protein purification were carried out. The pellet was resuspended

437 in KPO4 buffer (50 MM, pH 8) containing NaCl (0.1 M), imidazole (40 mM), and lysozyme

438 (750 μg/ml) after centrifugation (5000 g, 30 minutes, 4°C). The lysate was sonicated

439 (amplitude: 35%, 3 seconds on, 5 seconds off, 30 minutes) on ice after being incubated for half

440 an hour on ice. The lysate was then centrifuged (10,000 g, 60 minutes) and the supernatant was

441 then loaded to a Ni-NTA (nickel-nitrilotriacetic acid) affinity column equilibrated with KPO4

442 buffer (50 mM, pH 7.9) containing NaCl (0.5M) and imidazole (40 mM). In the resuspended

443 condition, the system was allowed to bind at 4°C (1-2 hours). The recombinant protein was

444 eluted in 15 mM KPO4 (pH 7.5) buffer containing 500 mM NaCl, 500 MM imidazole, and 10%

445 glycerol after a lengthy wash of the column with the same equilibration buffer (10 column

446 volumes). Purified recombinant AmQNS (61 kDa) fractions (containing an N-terminal Trx-S-

447 His fusion tag) were concentrated (Amicon-Ultra centrifugal filters, 10 kDa cut-off) and

448 overnight enterokinase cleavage was performed to remove the fusion tag1. Size exclusion

449 chromatography on a Superdex 200HR (10/100 GL) column (GE Healthcare) in HEPES-NaOH

450 buffer (20 mM, pH 7.5) containing NaCl (100 mM) and dithiothreitol was used to further purify

451 the AmQNS protein solution to homogeneity (a monomeric molecular weight of 43 kDa) (DTT,

452 2 mM). The purified AmQNS fractions were further concentrated to 20 mg/ml in the same

453 HEPES. SDS-PAGE was used for qualitative analysis, and the quantity was also calculated

454 using the NanoDropTM1000 spectrophotometer (Thermo Scientific, Wilmington, DE) at an

455 optical density (OD) ratio of 260/280 (and default protein absorbance values for 0.1%. i.e., 1

456 mg/mL). MALDI TOF MS analysis was used to determine the protein's homogeneity and mass

457 accuracy. The monomeric molecular weight of the resulting protein was 43 kDa, which was

458 consistent with the expected molecular mass of 43.8 Da (the calculated molecular mass of

459 AmQNS is 42.8 kDa, with the inclusion of approximately 10 amino acid residues from the

460 vector causing the change in molecular mass). The molecular weight of the purified protein

461 obtained by MALDI is 43.9 kDa (data not shown) and corresponds roughly to the expected

462 molecular weight of the full-length polypeptide chain.

463

AmQNS Crystallization by microbatch method.

Diffraction quality crystals were obtained in both native and substrate bound forms and optimized the conditions. AmQNS could produce a good diffraction quality crystal which diffracted up to 1.85 Å (H32 space group), when using 1:1 drop ratio by using 2 μL of protein and 2 μL of precipitant (0.1 M HEPES 7.5, 1.4M Sodium citrate tribasic; Index 20 of Hamptons research screen) with additives (0.1M Magnesium chloride hexahydrate or 0.1M Cadmium chloride hydrate). Co-crystallization trials were performed in presence of its favourable CoA substrates 'N-methyl anthraniloyl CoA (MANT-CoA)'. Solutions containing the substrate was directly added to the concentrated protein solution to a final concentration of ~2mM and incubated in ice for an hour prior to crystallization experiments. Co-crystallization studies were conducted by microbatch method. Diffraction quality of these crystals was also optimized wherever necessary by adding additives, varying drop sizes and protein/precipitant concentrations. The crystals appeared within a span of 2-3 weeks with approximate dimensions of 0.1 mm X 0.1 mm X 0.1 mm. Soaking experiments were also conducted in addition to co-crystallization experiment. Soaking native crystals with substrates is often the method to obtain crystals of protein–ligand complexes. Here, AmQNS native crystal was soaked (30 minutes) in ligand solution, which was prepared in the same crystallization condition. I.e., 2mM *N*-methylanthraniloyl-CoA was used for AmQNS-substrate soaking experiments (Index 20+0.1M Cadmium chloride hydrate). We could also get the byproduct (CoASH) bound AmQNS crystals. After proper incubation, crystal was picked with a nylon loop, flash frozen in liquid nitrogen and data were collected.

X-ray Data Collection, processing, and Refinement

After obtaining diffraction-quality crystals, the crystals were cryoprotected (20% glycerol in condition buffer/crystallization solution) by plunging them into liquid nitrogen using a fine-gauge wire micro loop. Data from the native and CoASH bound crystals belonging to the space group H32 were collected at the Molecular Biophysics Unit (IISc, Bangalore) using a MAR 345 image-plate detector mounted on a Bruker MICROSTAR ULTRA II Cu Kα rotating anode X-ray generator (wavelength of 1.54179 Å). For collecting the high-resolution data, the spacing between the protein crystal and the detector was adjusted to 200 mm. All data were collected at 100 K. Data collection statistics are given in **Table S1**. iMosflm was used to process the diffraction images[55], and data were merged using SCALA[56] in the CCP4[57]. The intensity data were converted into structure-factor amplitudes using TRUNCATE in the CCP4[58,59].

497 Further, structures were solved by the molecular replacement method at 1.85Å, 2.35Å for the

498 H32 space groups. The structure of AmQNS co-crystallized with *N*-methylanthraniloyl-CoA

499 was solved at a resolution of 2.35 Å. PHASER[60] in the CCP4 suite[57] was used for molecular

500 replacement by employing the structure of the acridone synthase from *Citrus microcarpa* (PDB

501 ID: WD7; identity-93%) as the search model[15]. The solutions obtained from molecular

502 replacement were subsequently refined using REFMAC5[61], along with multiple rounds of

503 manual model building using COOT v0.7.1[62,63]. Addition of the ligands and water atoms was

504 performed by PRODRG[64]. The possibility of alternate ligand conformations was also evaluated

505 before finalizing the ligand fitting. The final refinement of the native structures was performed

506 in PHENIX[65]. Images of the protein structures were generated using PyMOL Licenced

507 academic version[66]. The refined models were validated by PROCHECK[67] and the

508 MOLPROBITY[68]. All structural models were manually built, refined, and rebuilt with

509 REFMAC5/PHENIX and COOT.

510 **Structural Analysis**

511 The refined protein structures were evaluated using MolProbity with the Phoenix server and

512 wwPDB server[69]. Structural alignments were performed in ALIGN (Pymol[66]) and mTM-

513 align[70]. The sequence-structure conservation patterns were analysed using ESPript and

514 ENDscript 2.0[24]. The neighbour-joining method[71] was used to construct a structural phylogram.

515 2F0-Fc maps were calculated in CCP4 v7.0 using the 'fft' module, and the maps were visualized

516 in PyMOL using the command line option (contoured at 1.0 sigma around the selection site

517 within 1.6 Å of the selected atoms). The electrostatic properties of AmQNS were calculated

518 using APBS using the PyMOL plugin. PDB2PQR Version 2.0.0[72] was used to convert the PDB

519 files into PQR files. To obtain the detailed characteristic features of the surface pockets and

520 interior voids of AmQNS, CASTp (Computed Atlas of the Surface Topography of Proteins)

521 was used[73]. The default probe radius was used (1.4 Å). The protein secondary structure and

522 protein-ligand interactions (determined using LIGPLOT) were analysed using PDBsum

523 (www.ebi.ac.uk/pdbsum)[74]. The relative position of functional and structural importance

524 among the protein homolog sequence sites was estimated using Evolutionary Trace (ET;

525 http://evolution.lichtargelab.org/). All figures were prepared using PyMOL v2.4.1[66].

526 **Surface Plasmon Resonance (SPR) based assays**

527 *AmQNS immobilization on sensor chip*

528 *ProteOn* 'GLM' chip was used for the SPR interaction studies where AmQNS could give good

529 response (L3 - 9962 RU, L4 - ~7000) on immobilization (Supplementary **Figure S2**). Here,

530 amine coupling works where the amine groups present in the AmQNS covalently bind to
531 chemically activated carboxyl groups of the dextran molecules. Channel L3 and L4 were used
532 for immobilization of protein while L2 was used as reference. The extent of non-specific
533 interactions was eliminated or reduced by optimizing the buffer conditions. Furthermore, ligand
534 stability on the GLM biosensor over time was checked over a period of 30 days and was found
535 that it is active, by resulting in quantifiable interactions. This demonstrates that the AmQNS
536 immobilization onto the GLM sensor surface does not limit the functionality, confirming the
537 use of this SPR label-free technology to study its interaction pattern with different acyl-CoA
538 substrates. N-methylanthraniloyl-CoA was purchased from TransMIT (Plant MetaChem,
539 www.plantmetachem.com), whereas all other substrates were purchased from Sigma-Aldrich
540 (www.sigmaaldrich.com)

541 ### *Binding kinetics studies of AmQNS with different substrates*

542 All binding studies were performed at 30°C in ProteOn XPR array system[75]. The SPR based
543 system measures the changes in refractive index to investigate the direct interaction between
544 AmQNS and different CoA substrates. The analytes (substrates) were injected over the surface
545 of the chip and any binding between the two resulted in the change in surface mass, which is
546 recorded, and measures as a change in refractive index. In our experiments, AmQNS was
547 captured on the surface of the GLM sensor chip and used to screen the preferred substrates
548 (acyl-CoA's) in the presence and absence of malonyl-CoA. We could not find any interaction
549 in the absence of malonyl-CoA, and it was quite interesting to note that the binding modes of
550 these CoA substrates to AmQNS are influenced by the presence of the malonyl-CoA, which is
551 the extender during the biochemical reaction of polyketide formation. Sensogram prepared by
552 processing the data (after subtraction of L2 responses (reference channel). Baseline drift due to
553 the bulk refractive index change, non-specific binding, matrix effects and injection noise were
554 also corrected using the reference spots. Further, the responses obtained from the AmQNS-
555 small molecular interactions at different concentrations were fitted using the Langmuir 1:1
556 biomolecular interaction model using the ProteOn Manager software version 3.1.0.6 (Bio-Rad,
557 USA). Equilibrium dissociation constants (KD) were calculated from the ratio of the association
558 and dissociation rates.

559 ### **Molecular simulation studies**

560 The AmQNS polyketide synthase active site structural model was constructed from an AmQNS
561 crystallographic model (PDB ID: 6L5U) in PyMol[76] as the set of all amino acids with at least

562 one atom residing within 12.0 Å of the catalytic Cysteine (C164). Peripheral peptide chain

563 termini were neutralized by simple protonation to neutral amine and aldehyde structures.

564 Specific ligands (MANT-CoA and malonyl-CoA) were constructed *in situ* using PyMol by

565 referring to the -EthSH group of the co-crystallized CoASH ligand. For computational

566 efficiency, the bulk of the conserved CoA moiety (i.e., all except for those mentioned above -

567 EthSH moiety) was removed from each ligand. Transition state calculations were performed

568 using MOPAC 2016 [77], via the PM7 parametrization [78,79]. Due to the exceptional complexity in

569 the potential energy surface (PES)[80,81], it was necessary to manually perform transition states

570 by employing constraints to implement a stepwise approach between reacting atoms.

571 We also calculated transition states for MANT-CoA binding to AmQNS. An initial step size of

572 0.4 Å was used for the initial (distant) ligand approach until the approaching atoms were within

573 1.0 Å of the expected covalent distance. A step size of 0.1 Å was employed to capture subtle

574 structural and energetic effects. All receptor backbone atoms were held rigid to prevent spurious

575 peripheral conformational shifts from quantitatively overwhelming covalent energetics, as were

576 all side chains except those participating directly in enzyme reaction function. In addition, we

577 also determined quantum chemical transition states to compare the relative kinetic barriers and

578 thermodynamic enthalpies for the initial complexation of MANT-CoA, benzyl-CoA, butyryl-

579 CoA and coumaroyl-CoA.

580

581 **Accession codes.** Coordinates and structure factors for the above mentioned AmQNS structures

582 have been deposited in the Protein Data Bank (accession codes: 6L5U, 6L7J and, 7CCT).

Acknowledgments

Author Contribution

592 EVS supervised the study. MV contributed to conceptualization, investigation, methodology,
593 visualization and writing—original draft, review, and editing. AKV contributed to writing,
594 helped in data collection, structure solutions and calculations. DB helped in structure related
595 calculations and critically evaluated the manuscript. KRM assisted in simulation studies and
596 writing. AS helped for phylogeny and contributed to writing. MRP provided scientific advice

597 and provided constant help throughout the studies. All authors have read and agreed to the
598 published version of the manuscript.

599

References

601 1. Shang, S. & Tan, D. S. Advancing chemistry and biology through diversity-oriented
602   synthesis of natural product-like libraries. *Current Opinion in Chemical Biology* **9**, 248–
603   258 (2005).

604 2. Girija, A. *et al.* Harnessing the natural pool of polyketide and non-ribosomal peptide
605   family: A route map towards novel drug development. *Current Molecular Pharmacology*
606   **14**, (2021).

607 3. Shimizu, Y., Ogata, H. & Goto, S. Type III Polyketide Synthases: Functional
608   Classification and Phylogenomics. *ChemBioChem* **18**, 50–65 (2017).

609 4. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide
610   synthases. *Natural Product Reports* **20**, 79–110 (2003).

611 5. Shang, S., Iwadare, H., Macks, D. E., Ambrosini, L. M. & Tan, D. S. A Unified Synthetic
612   Approach to Polyketides Having Both Skeletal and Stereochemical Diversity. *Organic*
613   *Letters* **9**, 1895–1898 (2007).

614 6. Satou, R. *et al.* Structural basis for cyclization specificity of two Azotobacter type III
615   polyketide synthases: a single amino acid substitution reverses their cyclization
616   specificity. *The Journal of biological chemistry* **288**, 34146–34157 (2013).

617 7. Austin, M. B. *et al.* Crystal Structure of a Bacterial Type III Polyketide Synthase and
618   Enzymatic Control of Reactive Polyketide Intermediates. *Journal of Biological Chemistry*
619   **279**, 45162–45174 (2004).

620 8. Pandith, S. A., Ramazan, S., Khan, M. I., Reshi, Z. A. & Shah, M. A. Chalcone synthases
621   (CHSs): the symbolic type III polyketide synthases. *Planta* **251**, 15 (2019).

622 9. Hashimoto, M., Nonaka, T. & Fujii, I. Fungal type III polyketide synthases. *Natural*
623   *Product Reports* **31**, 1306–1317 (2014).

624 10. Schröder, G. & Schröder, G. Stilbene and Chalcone Synthases: Related Enzymes with
625   Key Functions in Plant-Specific Pathways. *Zeitschrift für Naturforschung C* **45**, 1–8
626   (1990).

627 11. Koskela, S., Elomaa, P., Helariutta, Y., Söderholm, P. & Vuorela, P. TWO BIOACTIVE
628   COMPOUNDS AND A NOVEL CHALCONE SYNTHASELIKE ENZYME
629   IDENTIFIED IN GERBERA HYBRIDA. in *Acta Horticulturae* 271–274 (International

630   Society for Horticultural Science (ISHS), Leuven, Belgium, 2001).

631   doi:10.17660/ActaHortic.2001.560.52.

632   12. Resmi, M. S., Verma, P., Gokhale, R. S. & Soniya, E. V. Identification and

633   characterization of a type III polyketide synthase involved in quinolone alkaloid

634   biosynthesis from Aegle marmelos Correa. *The Journal of biological chemistry* **288**,

635   7271–7281 (2013).

636   13. Heeb, S. *et al.* Quinolones: from antibiotics to autoinducers. *FEMS microbiology reviews*

637   **35**, 247–274 (2011).

638   14. Ahmed, A. & Daneshtalab, M. Nonclassical biological activities of quinolone derivatives.

639   *Journal of pharmacy & pharmaceutical sciences : a publication of the Canadian*

640   *Society for Pharmaceutical Sciences, Societe canadienne des sciences pharmaceutiques*

641   **15**, 52–72 (2012).

642   15. Mori, T. *et al.* Cloning and structure-function analyses of quinolone- and acridone-

643   producing novel type III polyketide synthases from Citrus microcarpa. *The Journal of*

644   *biological chemistry* **288**, 28845–28858 (2013).

645   16. Navarro-Muñoz, J. C. & Collemare, J. Evolutionary Histories of Type III Polyketide

646   Synthases in Fungi. *Frontiers in Microbiology* **10**, (2020).

647   17. De Luca, D. & Lauritano, C. In Silico Identification of Type III PKS Chalcone and

648   Stilbene Synthase Homologs in Marine Photosynthetic Organisms. *Biology* **9**, (2020).

649   18. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic

650   cellular complexity. *Nature* **541**, 353–358 (2017).

651   19. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin

652   of eukaryotes. *Nature Reviews Microbiology* **15**, 711–723 (2017).

653   20. Mallika, V., Sivakumar, K. C., Aiswarya, G. & Soniya, E. V. In silico approaches

654   illustrate the evolutionary pattern and protein-small molecule interactions of quinolone

655   synthase from Aegle marmelos Correa. *Journal of Biomolecular Structure and Dynamics*

656   **37**, 195–209 (2019).

657   21. Naake, T., Maeda, H. A., Proost, S., Tohge, T. & Fernie, A. R. Kingdom-wide analysis of

658   the evolution of the plant type III polyketide synthase superfamily. *Plant Physiology* **185**,

659   857–875 (2021).

660   22. Stairs, C. W. *et al.* Chlamydial contribution to anaerobic metabolism during eukaryotic

661   evolution. *Science Advances* **6**, eabb7258.

662   23. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948

663   (2007).

24. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320–W324 (2014).

25. Bronner, V., Bravman, T., Nimri, S. & Lavie, K. *Rapid and efficient determination of kinetic rate constants using the ProteOn XPR36 protein interaction array system. Bio-Rad bulletin 3172* (2006).

26. Goodsell, D. S. *et al.* RCSB Protein Data Bank: Enabling biomedical research and drug discovery. *Protein Science* **29**, 52–65 (2020).

27. Bräuer, A. *et al.* Structural snapshots of the minimal PKS system responsible for octaketide biosynthesis. *Nature Chemistry* **12**, 755–763 (2020).

28. Jez, J. M. & Noel, J. P. Mechanism of Chalcone Synthase: pKa OF THE CATALYTIC CYSTEINE AND THE ROLE OF THE CONSERVED HISTIDINE IN A PLANT POLYKETIDE SYNTHASE*. *Journal of Biological Chemistry* **275**, 39640–39646 (2000).

29. Richau, K. H. *et al.* Subclassification and Biochemical Analysis of Plant Papain-Like Cysteine Proteases Displays Subfamily-Specific Characteristics. *Plant Physiology* **158**, 1583 (2012).

30. Tseng, C. C., McLoughlin, S. M., Kelleher, N. L. & Walsh, C. T. Role of the Active Site Cysteine of DpgA, a Bacterial Type III Polyketide Synthase. *Biochemistry* **43**, 970–980 (2004).

31. Liou, G., Chiang, Y.-C., Wang, Y. & Weng, J.-K. Mechanistic basis for the evolution of chalcone synthase catalytic cysteine reactivity in land plants. *The Journal of biological chemistry* **293**, 18601–18612 (2018).

32. Jez, J. M., Ferrer, J.-L., Bowman, M. E., Dixon, R. A. & Noel, J. P. Dissection of Malonyl-Coenzyme A Decarboxylation from Polyketide Formation in the Reaction Mechanism of a Plant Polyketide Synthase. *Biochemistry* **39**, 890–902 (2000).

33. Wani, T. A. *et al.* Molecular and functional characterization of two isoforms of chalcone synthase and their expression analysis in relation to flavonoid constituents in Grewia asiatica L. *PLOS ONE* **12**, e0179155- (2017).

34. Hamlin, T. A., Swart, M. & Bickelhaupt, F. M. Nucleophilic Substitution (SN2): Dependence on Nucleophile, Leaving Group, Central Atom, Substituents, and Solvent. *ChemPhysChem* **19**, 1315–1330 (2018).

35. Leichert, L. I. & Jakob, U. Protein thiol modifications visualized in vivo. *PLoS biology* **2**, e333–e333 (2004).

36. Schramm, V. L. Enzymatic Transition States and Drug Design. *Chemical reviews* **118**, 11194–11258 (2018).

37. Roston, D. & Cui, Q. QM/MM Analysis of Transition States and Transition State Analogues in Metalloenzymes. *Methods in enzymology* **577**, 213–250 (2016).

38. Lundberg, M., Kawatsu, T., Vreven, T., Frisch, M. J. & Morokuma, K. Transition States in a Protein Environment − ONIOM QM:MM Modeling of Isopenicillin N Synthesis. *Journal of Chemical Theory and Computation* **5**, 222–234 (2009).

39. Grambow, C. A., Pattanaik, L. & Green, W. H. Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry. *Scientific Data* **7**, 137 (2020).

40. Rifai, E. A., van Dijk, M., Vermeulen, N. P. E., Yanuar, A. & Geerke, D. P. A Comparative Linear Interaction Energy and MM/PBSA Study on SIRT1–Ligand Binding Free Energy Calculation. *Journal of Chemical Information and Modeling* **59**, 4018–4033 (2019).

41. Ren, P. *et al.* Biomolecular electrostatics and solvation: a computational perspective. *Quarterly reviews of biophysics* **45**, 427–491 (2012).

42. Cruz, J. N. *et al.* Molecular dynamics simulation and binding free energy studies of novel leads belonging to the benzofuran class inhibitors of Mycobacterium tuberculosis Polyketide Synthase 13. *Journal of Biomolecular Structure and Dynamics* **37**, 1616–1627 (2019).

43. Bernetti, M., Cavalli, A. & Mollica, L. Protein-ligand (un)binding kinetics as a new paradigm for drug discovery at the crossroad between experiments and modelling. *MedChemComm* **8**, 534–550 (2017).

44. Du, X. *et al.* Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *International journal of molecular sciences* **17**, 144 (2016).

45. Keeling, P. J. *et al.* The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biology* (2014) doi:10.1371/journal.pbio.1001889.

46. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).

47. Huerta-Cepas, J. *et al.* EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* (2019) doi:10.1093/nar/gky1085.

731 48. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Research*
732   (2019) doi:10.1093/nar/gky995.

733 49. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource.
734   *Nucleic Acids Research* (2018) doi:10.1093/nar/gkx922.

735 50. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
736   Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–
737   780 (2013).

738 51. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated
739   alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* (2009)
740   doi:10.1093/bioinformatics/btp348.

741 52. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest-HPC: Fast selection of
742   best-fit models of protein evolution. in *Lecture Notes in Computer Science (including*
743   *subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*
744   vol. 6586 LNCS 177–184 (2011).

745 53. Kozlov, A. M. *et al.* RAxML-NG: A fast, scalable and user-friendly tool for maximum
746   likelihood phylogenetic inference. *Bioinformatics* (2019)
747   doi:10.1093/bioinformatics/btz305.

748 54. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
749   Inference in the Genomic Era. *Molecular Biology and Evolution* (2020)
750   doi:10.1093/molbev/msaa015.

751 55. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W.
752   iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.
753   *Acta Crystallogr D Biol Crystallogr* **67**, 271–281 (2011).

754 56. Evans, P. R. An introduction to data reduction: space-group determination, scaling and
755   intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**, 282–292 (2011).

756 57. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta*
757   *Crystallogr D Biol Crystallogr* **67**, 235–242 (2011).

758 58. French, S. & Wilson, K. On the treatment of negative intensity observations. *Acta*
759   *Crystallographica Section A* **34**, 517–525 (1978).

760 59. Potterton, L. *et al.* CCP4i2: the new graphical user interface to the CCP4 program suite.
761   *Acta Crystallogr D Struct Biol* **74**, 68–84 (2018).

762 60. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674
763   (2007).

764   61. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal
765   structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355–367 (2011).

766   62. Emsley, P. & Cowtan, K. ıt Coot: model-building tools for molecular graphics. *Acta*
767   *Crystallographica Section D* **60**, 2126–2132 (2004).

768   63. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot.
769   *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).

770   64. Schüttelkopf, A. W. & van Aalten, D. M. F. ıt PRODRG: a tool for high-throughput
771   crystallography of protein–ligand complexes. *Acta Crystallographica Section D* **60**,
772   1355–1363 (2004).

773   65. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and
774   electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877
775   (2019).

776   66. The PyMOL Molecular Graphics System, Version 2.4.1 Schrödinger, LLC.

777   67. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. ıt PROCHECK: a
778   program to check the stereochemical quality of protein structures. *Journal of Applied*
779   *Crystallography* **26**, 283–291 (1993).

780   68. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and
781   nucleic acids. *Nucleic Acids Res* **35**, W375–W383 (2007).

782   69. Young, J. Y. *et al.* Worldwide Protein Data Bank biocuration supporting open access to
783   high-quality 3D structural biology data. *Database (Oxford)* **2018**, bay002 (2018).

784   70. Dong, R., Pan, S., Peng, Z., Zhang, Y. & Yang, J. mTM-align: a server for fast protein
785   structure database search and multiple protein structure alignment. *Nucleic Acids Res* **46**,
786   W380–W386 (2018).

787   71. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing
788   phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425 (1987).

789   72. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an
790   automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic*
791   *Acids Res* **32**, W665–W667 (2004).

792   73. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface
793   topography of proteins. *Nucleic Acids Res* **46**, W363–W367 (2018).

794   74. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S. & Thornton, J. M. PDBsum:
795   Structural summaries of PDB entries. *Protein Sci* **27**, 129–134 (2018).

796   75. Bronner, V et al., Rapid and efficient determination of kinetic rate constants using the
797   ProteOn XPR36 protein interaction array system, Bio-Rad bulletin 3172 (2006).

798   76. PyMol 2.1. https://github.com/schrodinger/pymol-open-source (2019).

799   77. Stewart, J. J. P. MOPAC2016-Stewart Computational Chemistry, Colorado Springs, CO,
800   USA. *2016*.

801   78. Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more
802   modifications to the NDDO approximations and re-optimization of parameters. *Journal of*
803   *molecular modeling* **19**, 1–32 (2013).

804   79. Mato, J. & Guidez, E. B. Accuracy of the PM6 and PM7 Methods on Bare and Thiolate-
805   Protected Gold Nanoclusters. *The Journal of Physical Chemistry A* **124**, 2601–2615
806   (2020).

807   80. Unke, O. T., Koner, D., Patra, S., Käser, S. & Meuwly, M. High-dimensional potential
808   energy surfaces for molecular simulations: from empiricism to machine learning. *Machine*
809   *Learning: Science and Technology* **1**, 013001 (2020).

810   81. Bushnell, E. A. C., Huang, W. & Gauld, J. W. Applications of Potential Energy Surfaces
811   in the Study of Enzymatic Reactions. *Advances in Physical Chemistry* **2012**, 867409
812   (2012).

813

814

815

Table 1. Crystallization, data collection and refinement statistics of different crystal forms obtained for native, substrate and CoASH bound AmQNS.

	Native (6L5U)	CoASH (6L7J)	Substrate bound (7CCT)
Condition	0.1 HEPES (pH-7.5), 1.4 M sodium citrate tribasic dihydrate	0.1 HEPES (pH-7.5), 1.4 M sodium citrate tribasic dihydrate	0.1 HEPES (pH-7.5), 1.4 M sodium citrate tribasic dihydrate +2mM *N*-methylanthraniloyl-CoA (soacked)
Additive	0.1M magnesium chloride hexahydrate	0.1M $BaCl_2$	0.1M Cadmium chloride hydrate
Protein concentration (mg/ml)	10	10	10
Cryoprotectant	20% glycerol	20% glycerol	-
Space group	H32	H32	H32
Number of molecules/ASU	1 monomer	1 monomer	1 monomer
Unit cell dimensions a (in Å) b (in Å) c (in Å) α (in degrees) β (in degrees) γ (in degrees)	149.84 149.84 105.49 90 90 120	150.86 150.86 105.61 90 90 120	148.84 148.84 105.34 90.00 90.00 120.00
Resolution (Last shell)	40.93-1.85 (1.95-1.85)	35.23-1.80 (1.92-1.82)	48.76 -2.35 (2.41-2.35)
Completeness (%)	99.1 (100)	99.3 (100)	96.24
$\langle I/\sigma (I)\rangle$	9.9 (1.79)	14.0 (3.42)	8.5 (2.00)
R_{merge} (%)	13.3 (100)	14.0 (91.7)	-
Multiplicity	11.1 (10.3)	8.0 (7.4)	-
R-factor (%)	18.1	16.3	16.2
R-free (%)	21.6	18.9	23.9
RMS deviations from ideal values Bond length (Å) Bond angle (°)	0.37 0.56	0.77 0.85	0.29 0.40
Residues (%) in Ramachandran plot Most favoured regions [A,B,L] Additional allowed regions [a,b,l,p] Generously allowed regions[~a,~b,~l,~p] Disallowed regions [XX]	92.6% 7.1% 0.3% 0	92% 7.6% 0.3% 0	94.0% 5.0% - 1%

Table 2: MANT quinolone and acridone kinetics

Structural dependency for MANT quinolone kinetics and thermodynamics

	1. Mant-CoA binds C164	2. Malonyl-CoA inserts ketide	3. Quinolone closure & product release
MANT	Enthalpy: 36.0 kcal/mol Activation: 57.3 kcal/mol	Enthalpy: -3.5 kcal/mol Activation: 39.6 kcal/mol	Enthalpy: -5.3 kcal/mol Activation: 15.1 kcal/mol
MANT-pyridine	Enthalpy: 40.4 kcal/mol Activation: 60.2 kcal/mol	Enthalpy: -9.2 kcal/mol Activation: 31.4 kcal/mol	Enthalpy -4.1 kcal/mol Activation: 23.0 kcal/mol
MANT-chloro	Enthalpy: 32.7 kcal/mol Activation: 53.5 kcal/mol	Enthalpy: 4.4 kcal/mol Activation: 44.1 kcal/mol	Enthalpy: -4.6 kcal/mol Activation: 19.9 kcal/mol

Structural dependency for MANT acridone kinetics and thermodynamics

	1. 2nd Malonyl-CoA inserts ketide	2. 3rd Malonyl-CoA inserts ketide	3. Acridone closure and product release
	Enthalpy: 10.4 kcal/mol Activation: 59.9 kcal/mol	Enthalpy: 9.4 kcal/mol Activation: 53 kcal/mol	Enthalpy: -16.0 kcal/mol Activation: 38.3 kcal/mol