ComFB, a new widespread family of c-di-NMP receptor proteins

Sherihan Samir¹, Abdalla A. Elshereef¹, Vikram Alva², Jeanette Hahn³, David Dubnau³, Michael Y. Galperin⁴, Khaled A. Selim^{1,5,*}

¹Interfaculty Institute of Microbiology and Infection Medicine, Organismic Interactions
Department, Cluster of Excellence "Controlling Microbes to Fight Infections", Eberhard Karls
University of Tübingen, 72076 Tübingen, Germany

²Department of Protein Evolution, Max Planck Institute for Biology Tübingen, Germany

³Public Health Research Institute and Department of Microbiology, Biochemistry and Molecular Genetics, New Jersey Medical School, Rutgers University, New Jersey, USA

⁴National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

⁵Institute of Phototroph Microbiology, Heinrich-Heine University Düsseldorf, 40225 Düsseldorf, Germany

*For correspondence: Khaled A. Selim (khaled.selim@uni-tuebingen.de)

ABSTRACT

1

2

3

4

5

6 7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

Cyclic dimeric GMP (c-di-GMP) is a widespread bacterial second messenger that controls a variety of cellular functions, including protein and polysaccharide secretion, motility, cell division, cell development, and biofilm formation, and contributes to the virulence of some important bacterial pathogens. While the genes for diguanylate cyclases and c-di-GMP hydrolases (active or mutated) can be easily identified in microbial genomes, the list of c-di-GMP receptor domains is quite limited, and only two of them, PliZ and MshEN, are found across multiple bacterial phyla. Recently, a new c-di-GMP receptor protein, named CdgR or ComFB, has been identified in cyanobacteria and shown to regulate their cell size and, more recently, natural competence. Sequence and structural analysis indicated that CdgR is part of a widespread ComFB protein family, named after the "late competence development protein ComFB" from Bacillus subtilis. This prompted the suggestion that ComFB and ComFB-like proteins could also be c-di-GMP receptors. Indeed, revealed that ComFB proteins from Gram-positive B. subtilis Thermoanaerobacter brockii were able to bind c-di-GMP with high-affinity. The ability to bind c-di-GMP was also demonstrated for the ComFB proteins from clinically relevant Gram-negative bacteria Vibrio cholerae and Treponema denticola. These observations indicate that the ComFB family serves as yet another widespread family of bacterial c-di-GMP receptors. Incidentally, some ComFB proteins were also capable of c-di-AMP binding, identifying them as a unique family of cdi-NMP receptor proteins. The overexpression of *comFB* in *B. subtilis*, combined with an elevated concentration of c-di-GMP, suppressed motility, attesting to the biological relevance of ComFB as a c-di-GMP binding protein.

IMPORTANCE

The cellular content of the bacterial second messenger c-di-GMP is controlled by c-di-GMP synthases (GGDEF domains) and hydrolases (EAL or HD-GYP domains), whose activities, in turn, respond to the signals perceived by their upstream sensory domains. Cyclic-di-GMP transmits the signals to a variety of its targets, which may contain inactivated GGDEF, EAL, or HD-GYP domains, widespread PilZ or MshEN domains, or various lineage-specific c-di-GMP receptors. Many organisms encode multiple GGDEF domains but few c-di-GMP-binding proteins, suggesting the existence of still unidentified c-di-GMP receptors. Here, we demonstrate that the ComFB family proteins, which include the recently characterized cyanobacterial CdgR/ComFB, constitute yet another widespread family of bacterial c-di-NMP receptors. We additionally show that ComFB controls bacillar motility in a c-di-GMP-dependent manner.

INTRODUCTION

50

51

52

53

54 55

56

57

58 59

60

61

62

63

64

65

66 67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

Cyclic dinucleotide (c-di-NMP)-based second messengers are widely used by bacterial, archaeal. and eukaryotic cells to transduce signals from sensor proteins to various cellular receptors (Yoon & Waters, 2021). The first discovered c-di-NMP was 3'-5'-cyclic dimeric GMP (c-di-GMP), initially described in 1987 (Ross et al., 1987) and subsequently shown to be involved in protein and polysaccharide secretion, motility, cell division, cell development, and biofilm formation (Römling et al., 2013, Jenal et al., 2017), as well as contributing to the virulence of some important bacterial pathogens (Römling et al., 2013, Conner et al., 2017, Hall & Lee, 2018, Valentini & Filloux, 2019). Another second messenger, 3'-5'-cyclic dimeric AMP (c-di-AMP), was discovered in bacterial cells in 2008 (Witte et al., 2008, Corrigan & Gründling, 2013). Cyclic-di-AMP controls bacterial response to the changes in osmotic pressure and envelope stress by regulating the transport of potassium ions and glutamate, which makes it essential for cell growth in standard growth media (Commichau & Stülke, 2018, He et al., 2020, Krüger et al., 2021). In Bacillus subtilis and related bacteria, c-di-AMP also controls sporulation and is involved in the response to DNA damage (Commichau et al., 2015, Stülke & Krüger, 2020). In cyanobacteria, c-di-AMP primarily controls day-night metabolism via regulating glycogen synthesis, photosynthesis, redox balance and carbon/nitrogen metabolism (Selim et al., 2021a, Mantovani et al., 2022, Mantovani et al., 2023, Haffner et al., 2023b). Yet another second messenger, cyclic GMP-AMP (cGAMP), has been implicated in Vibrio cholerae virulence (Davies et al., 2012). In the past several years, cyclic dinucleotide second messengers have also been described in eukaryotes, where they are involved in cellular immunity and anti-viral defense (Millman et al., 2020, Yoon & Waters, 2021, Slavik & Kranzusch, 2023).

Cellular levels of c-di-GMP and c-di-AMP are controlled by environmental and intracellular signals that modulate the activities of the respective synthases (diguanylate and diadenylate cyclases, respectively) and hydrolases (phosphodiesterases). Genes encoding c-di-GMP synthetases (containing the GGDEF domain) and hydrolases (containing EAL or HD-GYP domains) are found in members of all bacterial phyla, sequenced so far (see the c-di-GMP census at https://www.ncbi.nlm.nih.gov/Complete Genomes/c-di-GMP.html). In contrast to the well-conserved – and therefore easily identified – sequences of c-di-GMP turnover enzymes, c-di-GMP receptors are much harder to pinpoint (Chou & Galperin, 2016, Khan *et al.*, 2023). So far, only two types of widespread c-di-GMP receptor proteins have been identified to contain PilZ or MshEN c-di-GMP-binding domains (Amikam & Galperin, 2006, Galperin & Chou, 2020, Wang *et al.*, 2016, Junkermeier & Hengge, 2021, Sellner *et al.*, 2021). The list of experimentally

characterized c-di-GMP receptors also includes (i) c-di-GMP-binding riboswitches (Sudarsan et al., 2008); (ii) inactivated enzymes of c-di-GMP turnover, such as PelD, BcsE, or LapD, that contain modified (mutated or truncated) GGDEF, EAL, or HD-GYP domains (Whitney et al., 2012, Whitfield et al., 2020, Fang et al., 2014, Zouhir et al., 2020, Chatterjee et al., 2014), and (iii) a variety of lineage-specific proteins, such as BldD and RsiG (actinobacteria), FleQ (pseudomonads), VpsT and VpsR (vibrios), CLP-like transcriptional regulators (xanthomonads), or CheY-like response regulators with Arg-rich tails in Caulobacter sp. (Tschowri et al., 2014, Gallagher et al., 2020, Hickman & Harwood, 2008, Matsuyama et al., 2016, Waters et al., 2008, Krasteva et al., 2010, Chin et al., 2010, Nesper et al., 2017), reviewed in (Chou & Galperin, 2016, Khan et al., 2023). However, comparing the phylogenetic distribution of c-di-GMP turnover enzymes and c-di-GMP receptors reveals a variety of organisms whose genomes encode multiple GGDEF, EAL, and/or HD-GYP domains but very few, if any, c-di-GMP-binding proteins or riboswitches (see the above-mentioned c-di-GMP census web site). Further, the newly identified c-di-GMP-binding proteins typically have a narrow phylogenetic distribution (Skotnicka et al., 2020, Li et al., 2023, Nie et al., 2024). These discrepancies indicate the existence of previously unrecognized c-di-GMP receptors.

A generally similar picture is seen for c-di-AMP signaling. Genes encoding c-di-AMP-producing diadenylate cyclases are found in most bacteria and many archaea (Galperin, 2023), in keeping with the essentiality of this second messenger (Commichau & Stülke, 2018). Several c-di-AMP-binding domains have been described (Corrigan *et al.*, 2013, He *et al.*, 2020), e.g., various members of P_{II} signaling superfamily are conserved c-di-AMP-receptor proteins (Selim & Alva, 2024, Selim *et al.*, 2023, Selim *et al.*, 2021b, Selim *et al.*, 2021a), but the current list of c-di-AMP-specific receptors is probably still incomplete (He *et al.*, 2020, Stülke & Krüger, 2020, Mantovani *et al.*, 2023).

Recently, a novel c-di-NMP receptor protein has been discovered in cyanobacteria and found to bind both c-di-GMP and c-di-AMP; it was initially named CdgR for "c-di-GMP Receptor" (Zeng *et al.*, 2023). In the multicellular cyanobacterium *Nostoc* sp. PCC 7120 (Alr3277, GenBank accession number BAB74976), this 179-aa protein was shown to bind c-di-GMP with high affinity (Kd = 0.18 μM) and to regulate the bacterial cell size (Zeng *et al.*, 2023). It could also bind c-di-AMP, albeit with a lower affinity (Kd = 2.66 μM). The CdgR homolog Slr1970 from the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 showed weaker c-di-GMP binding (Kd = 1.68 μM) than Alr3277 but allowed solving the structure of its c-di-GMP-bound complex (Protein DataBank (PDB) entry 8HJA). Slr1970 was found to bind also c-di-AMP with comparable affinity to c-di-GMP

(Samir et al., 2023). We showed that both c-di-AMP and its receptor SIr1970 are required for DNA uptake and natural competence in Synechocystis. Since the CdgR proteins possess a close similarity to the late competence factor B domain from Bacillus subtilis (BsComFB; domain PF10719 in the Pfam database), we renamed SIr1970 to ComFB (also annotated as ComFB in the NCBI's RefSeg database) (Samir et al., 2023). However, Zeng and colleagues argued that these proteins were functionally different (Zeng et al., 2023). In B. subtilis, the comFB gene is part of the comF competence operon that also encodes the DNA helicase ComFA and the predicted phosphoribosyltransferase ComFC, which mediate cellular uptake and handling of singlestranded DNA (Londono-Vallejo & Dubnau, 1993, Sysoeva et al., 2015, Damke et al., 2022). However, in contrast to these two neighbors, BsComFB appears to play no role in DNA uptake and its cellular role remains elusive (Sysoeva et al., 2015). Further, comFB genes are found in a variety of bacteria but missing in many comFA- and comFC-encoding organisms, including several Bacillus species (Kovacs et al., 2009, Kovacs et al., 2013). As a result, ComFB was assumed to participate in some kind of regulation. Given the widespread distribution of the ComFB-related proteins among diverse bacteria, we wondered if they could comprise a new family of c-di-GMP receptors that were missing in several bacterial lineages. Here, we present a bioinformatic and biophysical characterization of this protein family, along with binding assays of several of its members. This work shows that ComFB proteins, indeed, represent a widespread family of c-di-GMP (and also c-di-AMP) receptors. Moreover, we show its physiological relevance in controlling B. subtilis motility.

RESULTS

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

Phylogenetic distribution of ComFB proteins

To investigate the phylogenetic landscape of the ComFB family, we searched for homologs of the ComFB protein from *B. subtilis* (*Bs*ComFB) and CdgR from *Synechocystis* sp. PCC 6803 (Slr1970; *Sy*ComFB) using BLAST against the NCBI non-redundant protein sequence database. The search seeded with Slr1970 predominantly returned matches to other cyanobacterial proteins, while the search with *Bs*ComFB primarily yielded matches to proteins from the Bacillota phylum. Although these two proteins share the same core fold, their sequences are highly divergent, exhibiting less than 14% pairwise identity, with similarity largely limited to the residues involved in protein dimerization and the c-di-GMP-binding motif (Fig. 1A,B). Additionally, the zinc-

binding cysteine residues found in *Bs*ComFB are not conserved in the cyanobacterial sequences (see below; Fig. 1B). To further substantiate the homology between *Bs*ComFB and *Sy*ComFB and to detect additional homologs, we ran HHpred searches against the PDB70 profile HMM database and the profile HMM database of various proteomes. Searches against representative proteomes identified ComFB family members in other phyla, including Actinomycetota, Bdellovibrionota, Myxococcota, Nitrospirota, Pseudomonadota, Spirochaetota, and Thermodesulfobacteriota (Table S1). HHpred also detected a match between *Bs*ComFB and *Sy*ComFB with a probability value greater than 99.5%, indicating their homology.

The structures of BsComFB (PDB: $\underline{4WAI}$) and SIr1970 (PDB: $\underline{8HJA}$) represent stable dimers, with each subunit consisting of four long α -helices and one or two short β -strands (Fig. 1A,B). Additionally, SIr1970 features an N-terminal three-helical subdomain, which is almost exclusively found in its cyanobacterial homologs, as well as in some uncharacterized cyanobacterial proteins, such as *Nostoc punctiforme* Npun_F5121 (UniProt: B2J269). The structures of BsComFB and SIr1970 align with an RMSD of 3.6 Å over 91 C_{α} residues, despite their low (14%) pairwise sequence identity.

Using structure-based sequence alignment of BsComFB and Slr1970 in iterative database searches, we identified members of this protein family in several distinct bacterial lineages. These lineages include, in addition to cyanobacteria, members of the phyla Bacillota (classes Bacilli, Clostridia, Negativicutes, and Tissierellia), Pseudomonadota, Deferribacteriota, Spirochaetota, Thermotogota (Fig. 1B), and several others, as well as various candidate phyla (see a larger alignment in Fig. S1 and a partial list of family members in Table S1). A representative selection of ComFB family proteins is also available in the InterPro database (Paysan-Lafosse et al., 2023) under entry IPR019657, corresponding to the Pfam database (Mistry et al., 2021) domain PF10719. However, we discovered a subfamily of ComFB-related proteins widespread in cyanobacteria (Fig. 1C), but not included in the InterPro database, likely due to the presence of two insertions and changes in several conserved residues, including the c-di-GMP binding motif (Fig. S1 and Table S1). Most members of this new subfamily, such as, Nostoc sp. All3687 (UniProt: Q8YQX4), Anabaena cylindrica Anacy 5104 (UniProt: K9ZNS1), Trichormus variabilis Ava_3600 (UniProt: Q3M730) and CEN44_21400 of Fischerella muscicola (UniProt: A0A2N6JYB2, misannotated as a DUF3349 domain-containing protein), represent stand-alone ComFB-like domains. However, in several proteins within this new subfamily, such as SII1170 from Synechocystis sp. (UniProt: P74197), the N-terminal ComFB-like domain is followed by

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

DUF1816. PAS, and GGDEF domains. This domain architecture suggests potential diguanylate cyclase activity, with the ComFB-like domain presumably acting as a regulatory domain. Hereafter, we refer to these proteins collectively as the ComFB family, as the name CdqR is potentially confusing, having been previously used for the c-di-GMP regulator YdiV, which consists of a stand-alone EAL domain (Hisert et al., 2005, El Mouali et al., 2017). Most bacterial genomes encode a single copy of the comFB gene, if any (Table S1). However, certain lineages, particularly within Clostridia, Cyanobacteria, and Spirochaetota, often encode two or more orthologs of ComFB. Specifically, the cyanobacterial genomes of *Synechococcus* sp. PCC 7502, Pseudanabaena biceps PCC 7429, and Pseudanabaena sp. PCC 7367 encode six ComFB proteins each (Table S1), Synechocystis sp. PCC 6803 possesses four ComFB family members (Slr1970, Slr1505, Sll1739, and Sll1170), while F. muscicola CCMEE 5323 and Nostoc sp. PCC 7120 each has two different ComFB-like proteins. Outside bacterial lineages, ComFB family members have been found in two archaea, Methanocella sp. (GenBank: OPY29888) and Candidatus Methanomethylicus sp. (GenBank: HGS80389, recently suppressed but still available in UniProt and on the NCBI and EBI websites), the fornicate Aduncisulcus paluster (GenBank: GKT31366), and as part of a multidomain protein from the model plant Arabidopsis thaliana (GenBank: OAO89096; see Fig. S1). These instances likely arose from bacterial contamination. In many lineages, ComFB exhibits a patchy distribution, including its previously noted presence in close relatives of B. subtilis but not in other Bacillus species such as B. anthracis, B. cereus, and B. thuringiensis (Kovacs et al., 2009, Kovacs et al., 2013). Within the phylum Pseudomonadota, ComFB is found in the classes Betaproteobacteria and Gammaproteobacteria (Fig. S1), as well as in the former Deltaproteobacteria (recently reclassified as the phylum Thermodesulfobacteriota), but so far not in Alphaproteobacteria or Epsilonproteobacteria (the new phylum Campylobacterota). Furthermore, all betaproteobacterial ComFB sequences come from the single order Burkholderiales, and nearly all deltaproteobacterial sequences are from the order Desulfovibrionales (Table S1). Among gammaproteobacteria, ComFB sequences are found in lineages such as Aeromonadales, Alteromonadales, Pseudomonadales, and Vibrionales, but so far not in enterobacteria or xanthomonads. This patchiness is also evident at lower taxonomic levels. For example, within the order Pseudomonadales, all ComFB sequences are found in the genus Marinobacter and none in Pseudomonas. Similarly, among spirochetes, ComFB is

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238239

240

encoded in oral pathogens Treponema denticola (Fig. 1B) and Treponema socranskii, as well as in the cattle skin pathogen *Treponema brennaborense*, but not in the closely related *Treponema* pallidum, the causative organism of syphilis, or the Lyme disease spirochete Borrelia (Borrelial) burgdorferi (Table S1). **Evolutionary landscape of ComFB superfamily** To gain a comprehensive understanding of the evolutionary relationships among ComFB family proteins, we gathered homologs of several representative ComFB proteins and clustered them using the CLANS tool (Frickey & Lupas, 2004) based on the strength of their all-against-all pairwise sequence similarities. The resulting map (Fig. 1C) revealed four distinct clusters of ComFB sequences: two exclusively composed of proteins from the Cyanobacteriota phylum, one from the Pseudomonadota phylum, and one that includes proteins from a variety of phyla. Despite the low sequence similarity (less than 15%) between these clusters, many residues of the c-di-GMP-binding motif are conserved (Fig. 1B). The B. subtilis ComFB sequence is located in the central cluster (colored orange), which includes sequences from diverse phyla such as Actinomycetota, Bacillota, Bdellovibrionota, Myxococcota, Nitrospirota, Spirochaetota, and Thermodesulfobacteriota, and from which the other three clusters are radiating. The Synechocystis ComFB sequence (Slr1970: SyComFB), along with its additional homologs in Synechocystis (SIr1505 and SII1739), are in the closely connected cyanobacterial cluster (in cyan). This cluster also contains the *Nostoc* CdgR/ComFB (Alr3277), the multiple ComFB homologs found in Synechococcus sp. PCC 7502, Pseudanabaena biceps PCC 7429, and Pseudanabaena sp. PCC 736, as well as the fusions of the ComFB-like domain with phage shock protein A (Synechococcus sp. PCC 7502; UniProt: K9SSQ8) and ABC transporter ATP-binding protein (*Limnothrix* sp. P13C2; UniProt: A0A1C0VII2). The second cyanobacterial cluster (in green), which is weakly connected to the central cluster (in orange), contains the new subfamily of ComFB-like proteins (lower sequence block in Fig. 1B). This cluster consists of stand-alone ComFB-like proteins (e.g., CEN44 21400, All3687, and Ava 3600) and multidomain diguanylate cyclases with an N-terminal ComFB-like domain, such as the Synechocystis diquanylate cyclase SII1170 (UniProt: P74197). The separation of this new cyanobacterial subfamily in the map suggests functional distinction from the other cyanobacterial ComFB proteins. Notably, many cyanobacterial species have representatives in both cyan and

green clusters (e.g., Synechocystis sp. PCC 6803, Nostoc sp. PCC 7120, F. muscicola, and T.

241 *variabilis* ATCC 29413). The fourth cluster (in violet) possesses sequences of Pseudomonadota

phylum, and most sequences of this cluster consist solely of the ComFB domain.

Sequence conservation within the ComFB family

Despite the relatively low level of sequence identity between *Bs*ComFB and *Sy*ComFB, the key functional residues are conserved between these two (Fig. 1B) and within the entire family (Fig. S1). Indeed, the three residues, Asn4, Glu7, and Tyr46 (*Bs*ComFB numbering), that are involved in protein dimerization, are conserved in ComFB proteins from all bacterial phyla (Fig. S1 and Fig. S2). Accordingly, it would be reasonable to assume, that, like *Bs*ComFB and *Sy*ComFB (Zeng *et al.*, 2023, Sysoeva *et al.*, 2015), all family members form dimers (or higher oligomers). In addition, several c-di-GMP-binding residues, identified in the structure of the *Sy*ComFB-c-di-GMP complex (PDB: 8HJA) (Zeng *et al.*, 2023), are widely conserved as well (Fig. 1B and Fig. S1). These residues include (i) Asp33, Asn40, and Arg/Lys41 (*Bs*ComFB numbering) that bind c-di-GMP through electrostatic interactions; (ii) hydrophobic amino acid residues in positions of Ala36, Leu37, and Val47 that form hydrophobic contacts with the c-di-GMP ligand, and (iii) Tyr55 that forms a π-bond with the guanine moiety of c-di-GMP. Several other c-di-GMP-binding residues of *Sy*ComFB are poorly conserved, suggesting greater flexibility in ligand binding within the family with altered affinity (or a loss of binding altogether).

comFB genomic neighborhoods

- 259 Members of the ComFB family are typically annotated as "late competence development protein
- 260 ComFB", based on BsComFB, the founding member of the family (Londono-Vallejo & Dubnau,
- 261 1993, Sysoeva et al., 2015). However, the comFA-comFB-comFC operon organization is only
- seen in B. subtilis (Fig. 2) and its closest relatives, B. amyloliquefaciens, B. atrophaeus,
- 263 B. licheniformis, B. pumilis, B. velezensis, and B. xiamenensis. A comparison of the comF gene
- 264 distribution using the COG database (Galperin et al., 2021) shows that comFB is typically found
- 265 without either comFA or comFC, and often in those organisms that do not carry either of these
- 266 genes (Table S2).

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

- 267 In B. subtilis and several other organisms (e.g. Treponema denticola ATCC 35405), comFB is
- located in the vicinity of one or more flagellar genes (Fig. 2). In many members of the clostridial
- 269 order Thermoanaerobacterales, including Caldicellulosiruptor and Thermoanaerobacterium
- species, comFB is located within the operon that codes for type IV pili (Khan et al., 2020) (see

Fig. 2). In cyanobacteria, *comFB* is frequently found in an operon with *hfq* (Fig. 2), a key component of cyanobacterial type IV-pili machinery that is required for motility and DNA uptake (Samir *et al.*, 2023, Schuergers *et al.*, 2014). This genomic organization suggests a possible involvement of ComFB proteins in the regulation of cell motility or other type IV pili-related functions. In *Vibrio cholerae*, *V. harveyi*, *V. parahaemolyticus* and some other organisms (e.g. *Aliivibrio fischeri*), the *comFB* gene is transcribed divergently from a gene coding for a GGDEF-containing diguanylate cyclase, highlighting a possible functional link to c-di-GMP signaling.

ComFB domain architectures

The vast majority of ComFB domains are found in a stand-alone form (e.g., *Bs*ComFB; PDB: 4WAI) (Fig. 3). Also, most sequences of the Pseudomonadota cluster, e.g. as seen in *V. cholerae*, are found as a sole ComFB domain (see the violet cluster in Fig. 1C). However, canonical cyanobacterial members often contain an additional three-helical N-terminal subdomain, as seen in the structure of *Sy*ComFB (PDB: 8HJA), or other N-terminal and/or C-terminal extensions, as in *Synechocystis* sp. PCC 6803 proteins Slr1505 and Sll1739 (see Fig. 3 and Fig. S3 for domain architectures and GenBank accession numbers). Some cyanobacterial proteins contain two tandem ComFB domains (e.g., Pse7367_0880, UniProt: K9SGA4) (Fig. 3). Remarkably, such tandem-domain ComFB proteins include five out of the six paralogs encoded by the genomes of *Pseudanabaena biceps* PCC 7429, *Pseudanabaena* sp. PCC 7367, and *Synechococcus* sp. PCC 7502 (Table S1 and Fig. S3). Some of these proteins (such as UniProt: K9SSQ8) additionally contain long N-terminal coiled-coil segments that bear limited similarity to phage shock protein A of the PspA/IM30 family (Pfam domain PF04012) (Fig. 3 and Fig. S3).

Several other cyanobacteria, such as *Halothece* sp. PCC 7418, combine ComFB (e.g., UniProt: K9YBB1) with an N-terminal PATAN (DUF4388, PF14332) domain (Fig. S4), which is usually found in PatA-type response regulators that control heterocyst formation in cyanobacteria (Makarova *et al.*, 2006). Remarkably, a large-scale protein interaction screen in *Synechocystis* sp. PCC 6803 revealed an interaction between the second ComFB protein, encoded by *slr1970*, and the Slr1594 protein of the PatA family (Sato *et al.*, 2007); however, the relevance of this interaction is still unclear. Additionally, in the cyanobacterium *Limnothrix* sp. P13C2, the ComFB domain is found in combination with an ATPase subunit of an ABC transporter (UniProt: A0A1C0VII2) (Fig. S3). Also, as mentioned above, in the Sll1170 protein of the cyanobacterium *Synechocystis* sp. PCC 6803 and several other proteins, the N-terminal ComFB domain is followed by DUF1816, PAS, and GGDEF domains (Fig. 3 and Fig. S3).

Other architectures can also be found (Fig. S3), including several clostridial response regulators that combine the ComFB domain and the two-component phosphoacceptor receiver (REC, PF00072) domain (e.g. UniProt: R6WME1) (Fig. S3 and Fig. S4). Similarly, some proteins from the Candidatus Omnitrophota phylum (e.g., UniProt: A0A1G1PT98) couple the ComFB domain with the chemotaxis methyltransferase CheR (PF01739) domain (Fig. S3 and Fig. S4). The ComFB proteins from Treponema denticola (UniProt: Q73MV1) and many other spirochetes combine an N-terminal ComFB domain with a predicted C-terminal immunoglobulin-like (transthyretin-like) domain (Fig. S3 and Fig. S4; Table S1). This diverse array of domain architectures highlights the functional versatility of the ComFB family and suggests that these proteins may play a variety of regulatory roles across different bacterial lineages.

Metal-binding Cys residues in the ComFB family

In the structure of ComFB from B. subtilis, each monomer contains a tightly bound Zn²⁺ ion that is coordinated by four Cys residues: Cys25, Cys27, Cys30, and Cys88, and appears to contribute to the stabilization of the protein (Sysoeva et al., 2015). Examination of ComFB sequences from other members of the phylum Bacillota reveals conservation of the first three Cys residues but not the last one (Fig. 1B). These three Cys residues are also conserved in ComFB proteins from members of the phyla Deferribacteres and Thermotogae and occasionally found in proteins from other lineages, such as Maridesulfovibrio salexigens (Thermodesulfobacteriota) and the cyanobacterium Synechococcus sp. JA-2-3B'a(2-13) (Fig. S1). The same three Cys residues are conserved in some, albeit not all, spirochetes. Some spirochete ComFB sequences, like the one from T. denticola, contain an additional Cys residue that is located in the last turn of the second long α-helix. Others, like ComFB from Spirochaeta africana, carry just two of these Cys residues, and some, like ComFB from Brachyspira murdochii, neither of them (Fig. 1B). Beta- and gammaproteobacterial ComFB sequences typically contain a single conserved Cys residue in the same position, near the end of the second long α -helix. Finally, except for the above-mentioned M. salexigens and Synechococcus sp. JA-2-3B'a(2-13), ComFB proteins from the phyla Cyanobacteriota and Thermodesulfobacteriota do not contain any of these Cys residues. Altogether, the conservation of these Cys residues is seen only in a fraction of ComFB family members (Fig. S2).

Cyclic-di-NMP binding by ComFB proteins

Although c-di-GMP binding by *Sy*ComFB (SIr1970) was suggested to be specific for cyanobacteria (Zeng *et al.*, 2023), the structural similarity between *Sy*ComFB and *Bs*ComFB (Fig. 1A), coupled with the conservation of several c-di-GMP-binding residues within the entire ComFB family (Fig. 1B), suggested that other members of this family might also serve as c-di-GMP and/or c-di-AMP receptors. Moreover, we were able to show that SyComFB can also bind c-di-AMP with comparable affinity to c-di-GMP (Samir *et al.*, 2023). To investigate whether c-di-NMP binding is a common property of diverse members of the ComFB family, we heterologously expressed ComFB proteins from four phylogenetically distinct bacterial lineages and tested their ability to bind c-di-GMP and/or c-di-AMP. The binding assays were performed with ComFB proteins from *B. subtilis* (phylum Bacillota, class Bacilli), *Thermoanaerobacter brockii* (phylum Bacillota, class Clostridia), *Vibrio cholerae* (phylum Pseudomonadota, class Gammaproteobacteria), and *Treponema denticola* (phylum Spirochaetota, class Spirochaetia); their sequences are shown in Fig. 1B. While both *B. subtilis* and *T. brockii* are environmental Gram-positive bacteria, *V. cholerae* and *T. denticola* represent pathogenic Gram-negative bacteria.

First, we used size exclusion chromatography coupled to multiangle light scattering (SEC-MALS) to determine the oligomeric state of all ComFB proteins in solution. As expected, *B. subtilis* ComFB (*Bs*ComFB) protein showed a species of ~ 23.5 kDa (Fig. S5), indicating that *Bs*ComFB protein (the theoretical molecular mass of a monomer with an 8xHis tag is 12.1 kDa) is a dimer in solution, in agreement with the crystal structure of *Bs*ComFB (Sysoeva *et al.*, 2015). Similarly, *T. denticola* ComFB (*Td*ComFB; with the theoretical mass of monomer of 27.8 kDa) behaved as a dimer in solution with a molar mass of 45.5 kDa (Fig. S5). Unexpectedly, ComFB proteins of both *T. brockii* (*Tb*ComFB) and *V. cholerae* (*Vc*ComFB) behaved as monomers in solution with molar masses of 11.4 kDa and 13.4 kDa, respectively (Fig. S5).

Next, we measured the binding affinity of ComFB proteins to either c-di-GMP or c-di-AMP using isothermal titration calorimetry (ITC). The raw isothermal data for titration of c-di-GMP or c-di-AMP were fitted using a one-binding site model for monomeric ComFB proteins to determine the dissociation constant (K_D) (Table 1; Fig. 4 and Fig. 5). For BsComFB, the titration of c-di-GMP yielded an exothermic profile with a high affinity in the very low micromolar range (K_D 0.17 μ M; Fig. 4A). In contrast, the c-di-AMP binding events showed endothermic calorimetric signals with a K_D of 83.3 μ M, indicating very weak binding (Fig. 5), implying that BsComFB preferentially binds to c-di-GMP. Next, we assessed the ability of BsComFB protein to bind to c-di-GMP in the

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

presence of saturating concentrations of c-di-AMP, and *vice versa*. In a competition binding assay, when BsComFB protein was saturated first with 150 µM c-di-AMP, c-di-GMP binding events again showed strong exothermic calorimetric signals, however the binding enthalpy of c-di-GMP was a bit reduced and the K_D value of 0.24 μM was substantially higher than the K_D in the absence of cdi-AMP (compare Fig. 4A with 4B; Table 1). When BsComFB protein was saturated with 150 μM c-di-GMP, the binding of c-di-AMP was further reduced with a K_D of 121.1 µM (Table 1; compare Fig. 5A with 5B). Again, the binding assays for ComFB from T. brockii (TbComFB) showed exothermic signals for c-di-GMP binding and endothermic profile for c-di-AMP binding (Fig. 4 and Fig. 5). However, compared to BsComFB, TbComFB showed weaker affinity towards c-di-GMP (Fig. 4C) with a KD value of 7.9 μM and enhanced affinity towards c-di-AMP (Fig. 5C) with a K_D value of 20.1 μM (Table 1), implying that TbComFB could bind both c-di-GMP/c-di-AMP in vivo, similar to cyanobacterial SyComFB (Samir et al. 2023). The competition binding assays at saturating concentration of 150 µM of c-di-AMP (Fig. 4D) or c-di-GMP (Fig. 5D) also revealed that both molecules could compete with each other for binding to TbComFB. However, the binding affinity was substantially reduced compared to that in the absence of the competitor nucleotide (Table 1; compare isotherm of Fig. 4C with 4D and Fig. 5C with 5D). Finally, we assessed the ability of the ComFB proteins from the gram-negative bacteria V. cholerae (VcComFB) and T. denticola (TdComFB) to bind c-di-GMP or c-di-AMP. Remarkably, the TdComFB possesses in addition to the N-terminal ComFB domain, a C-terminal immunoglobulinlike domain (Fig. S4). As expected, both VcComFB and TdComFB proteins were able to bind to c-di-GMP (Table 1), however TdComFB bound to c-di-GMP endothermically (Fig. 4F) with very low affinity (K_D 284 μM) compared to exothermic binding (Fig. 4E) and high affinity with K_D value of 1.4 µM for VcComFB. Surprisingly, both VcComFB and TdComFB proteins were not able to bind c-di-AMP and c-di-AMP did not compete with c-di-GMP (Table 1 and Fig. 5E,F). To further confirm that VcComFB binds only c-di-GMP, we used nano differential scanning fluorimetry (nanoDSF) compared to TbComFB, which binds both c-di-GMP and c-di-AMP with good affinity (Table 1). Both c-di-GMP and c-di-AMP thermally stabilized TbComFB, but only c-di-GMP stabilized VcComFB (Fig. S6), which further confirms that VcComFB binds only c-di-GMP, while TbComFB binds both c-di-NMPs. As all ComFB proteins showed robust binding of c-di-GMP and, in some cases, c-di-AMP (e.g. TbComFB and SyComFB), these experiments establish the ComFB superfamily as c-di-NMP receptor proteins, with preferential binding of c-di-GMP.

Table 1. c-di-NMP binding by different ComFB proteins

	Organism protein name	Titrant	Competitor	Average K _d	ΔΗ	ΔG
	(Taxonomy)			(μM) ^a	(kcal mol ⁻¹)b	(kcal mol ⁻¹) ^c
Gram-positive		c-di-GMP		0.17 ± 0.07	-8.5 ± 1.3	-9.3 ± 0.25
	Bacillus subtilis ComFB	c-di-GMP	150 μM c-di-AMP	0.24 ± 0.09	-6.0	-8.1 ± 1.8
	(Bacillota, class Bacilli)	c-di-AMP		83.3 ± 73.1	44.8 ± 35.2	-6.2 ± 1.2
		c-di-AMP	150 μM c-di-GMP	121.1 ± 85.9	43.4 ± 36.6	-5.555 ± 0.74
	Thermoanaerobacter brockii	c-di-GMP		7.9 + 1.8	-3.56 ± 0.4	-6.9 ± 0.15
	Thebr_1103 (Bacillota, class	c-di-GMP	150 μM c-di-AMP	12.2 ± 3.6	-3.1 ± 0.1	-6.7 ± 0.3
	Clostridia)	c-di-AMP		20.1 ± 3.4	80.0 ± 0.2	- 5.1 ± 0.1
		c-di-AMP	150 μM c-di-GMP	33.9 ± 2.6	5.4 ± 1.8	- 6.1 ± 0.03
Gram-negative	Vibrio cholerae VC_A0561	c-di-GMP		1.4 ± 1.0	-4.4 ± 1.7	-8.0 ± 0.49
	(Pseudomonadota, class Gammaproteobacteria)	c-di-AMP	150 μM c-di-GMP		No binding	
	Treponema denticola	c-di-GMP		284 ± 97.0	35.9 ± 31.2	- 4.7
	Tde_1406 (Spirochaetota, class Spirochaetia)	c-di-AMP	150 μM c-di-GMP	No binding		

The raw isothermal titration calorimetry (ITC) data were fitted using a one-site binding model for monomeric ComFB proteins. (a) The dissociation constant (K_D) values correspond to the mean of at least 2 independent experiments \pm standard deviation. All titrations were performed in 50 mM Tris-HCl based buffer (pH 8.0); (b) enthalpy; (c) Gibbs free energy. To test potential competition by another c-di-NMP, the protein was preincubated with 150 μ M of the 'competitor' dinucleotide, followed by titration with the other one.

Physiological importance of comFB in Bacillus

Next, to determine the biological significance of *comFB*, we investigated the *in vivo* function of *comFB* in the genetically accessible model bacterium, *B. subtilis*. As mentioned previously, *B. subtilis comFB* is localized in the vicinity of flagellar genes required for motility (Fig. 2), and it binds specifically to c-di-GMP (Fig. 4; Table 1), which is well known to inhibit motility in many bacteria, including *B. subtilis* (Wolfe & Visick, 2008, Römling *et al.*, 2013, Chen *et al.*, 2012, Subramanian *et al.*, 2017). To examine whether ComFB plays a role in motility and if its function depends on c-di-GMP signaling, we expressed it from a constitutive promoter in the ectopic *amyE* locus of *B. subtilis* in the background of both wild-type and Δ*pdeH*, a phosphodiesterase mutant characterized by elevated levels of c-di-GMP (Chen *et al.*, 2012, Gao *et al.*, 2013). In these strains, growing in LB where competence is not expressed, the ComK-driven promoter in front of the *comF* operon is not active, and the main source of ComFB will be from the construct in *amyE*. Along with negative controls carrying empty vectors, all strains were inoculated in the centers of

Petri dishes containing LB Medium with 0.3% agar and permitted to grow overnight. As shown in Fig. 6, ectopic expression of ComFB markedly inhibits swimming, and this inhibition is propagated in the absence of PdeH, the only c-di-GMP phosphodiesterase of *B. subtilis*. Deletion of *pdeH* significantly affects swimming by itself, presumably due to its effect via Motl (Subramanian *et al.*, 2017). To rule out the sensitivity of the motility assay to growth differences, we additionally compared the growth of the relevant strains and detected no significant differences (Fig. S7). These data provide strong evidence that ComFB interferes with motility in the presence of elevated c-di-GMP, confirming the biological relevance of its high affinity for c-di-GMP (Table 1).

DISCUSSION

Proteins of the ComFB family (Pfam domain <u>PF10719</u>) are widespread in bacteria, being encoded, besides *Bacillota* and *Cyanobacteriota*, in the genomes of representatives of at least five other phyla (Fig. 1B and Fig. S1). Further, metagenomic sequencing identified *comFB* genes in members of a dozen more phyla (Table S1) and several candidate phyla. These proteins are usually encoded in a single copy per genome, such as CdgR from *Nostoc* sp. PCC 7120, but are occasionally found in multiple copies, from two genes in *Alkaliphilus metalliredigens* (Amet_2487 and Amet_3088) to six paralogous copies in the genomes of *Synechococcus* sp. PCC 7502 and *Pseudanabaena* PCC 7367 and seven in *Allocoleopsis franciscana* PCC 7113 (Table S1).

With respect to cyanobacterial c-di-GMP signaling, the discovery of CdgR (Zeng *et al.*, 2023) closed an important gap in knowledge: the paucity of known c-di-GMP receptors in cyanobacteria. Indeed, as seen in the c-di-GMP census, *Nostoc* sp. PCC 7120 and *Synechocystis* sp. PCC 6803 encode multiple diguanylate cyclases (Enomoto *et al.*, 2023), but no PilZ-domain proteins and very few MshEN-containing proteins (1 and 3, respectively). The c-di-GMP binding by these MshEN domains has not been tested so far, although c-di-GMP has been shown to control cyanobacterial cell motility and phototaxis (Savakis *et al.*, 2012, Angerer *et al.*, 2017, Enomoto *et al.*, 2023, Wallner *et al.*, 2020). Other cyanobacteria encode PilZ domains as part of their bacterial cellulose synthase (*bcsA*) genes and MshEN domains at the N-termini of their PilB/GspE ATPases, which control twitching motility and type II secretion. No other cyanobacterial c-di-GMP targets had been characterized until last year. Characterization of CdgR, which is widespread in cyanobacteria, and demonstration of its involvement in controlling cell size (Zeng *et al.*, 2023) provided a much-needed rationale for the presence of multiple c-di-GMP turnover enzymes in various cyanobacterial genomes.

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

While Zeng and colleagues assumed that CdgR was specific for cyanobacteria (Zeng et al., 2023), the structural similarity between CdgR and ComFB (Fig. 1A), coupled with the conservation of several c-di-GMP-binding residues within the entire ComFB/CdqR family (Fig. 1B), suggested that other members of this family might also bind c-di-GMP, such that this family would represent a widespread c-di-GMP receptor protein. This suggestion was verified here by demonstrating high-affinity c-di-GMP binding by BsComFB (Table 1), the founding member of the family, whose proposed role in competence had been used for annotating the entire family (PF10719 in Pfam database and IPR019657 in InterPro) (Mistry et al., 2021, Paysan-Lafosse et al., 2023) as "late competence development protein ComFB". BsComFB was shown to bind c-di-GMP with high affinity, Kd = 0.17 μM, which is remarkably close to the value previously reported for the CdgR from Nostoc sp. PCC 7120 (Zeng et al., 2023). Further, three other ComFB family members from diverse bacterial lineages were also found to bind c-di-GMP (Table 1). The shared ability to bind c-di-GMP by ComFB-like proteins from B. subtilis, two cyanobacteria, and these three organisms strongly suggests that the ComFB family represents the missing widespread family of c-di-GMP receptors, the third one after PilZ and MshEN domains. Accordingly, the entire family can now be renamed "ComFB-like c-di-GMP receptors." The new name also makes sense because of the uncertain role of BsComFB in competence. Genome analysis has revealed the absence of the *comFB* gene in several *Bacillus* species that appeared to be transformable (Kovacs et al., 2009, Kovacs et al., 2013). Further, while comFAcomFC operons are fairly widespread, they are commonly found in the genomes that do not encode comFB (Table S2); the comFA-comFB-comFC operons, like the one in B. subtilis are only found in a few closely related Bacillus species. Finally, a B. subtilis strain expressing comFA and comFC in the absence of comFB was reported to be normally transformable (Sysoeva et al., 2015). These data suggested that ComFB may not be required for competence in B. subtilis. Nevertheless, the ability of ComFB proteins to bind c-di-GMP, the master regulator of motilityrelated functions, and the genomic association of comFB with flagella (as seen in B. subtilis and T. denticola) or pili (as seen in cyanobacteria, Caldicellulosiruptor and Thermoanaerobacterium species) genes still suggests a possible involvement of ComFB proteins in regulation of cell motility. However, the role of ComFB might be different in other organisms. In this context, we were able to show that ComFB is participating in a c-di-GMP-mediated inhibition of motility in B. subtilis using swimming assay (Fig. 6). However, it is important to note that the swimming assay

does not distinguish between an effect of ComFB on the production or function of flagella, not does it exclude an effect on chemotaxis. It will require further investigation to distinguish between these alternatives. Nevertheless, *Bs*ComFB joins Motl (Subramanian *et al.*, 2017) as a c-di-GMP binding protein associated with a motility-related phenotype in *B. subtilis*. Moreover, our recent work showed that *Sy*ComFB (encoded by *slr1970*), the CdgR ortholog from the model cyanobacterium *Synechocystis* sp. PCC 6803, is required for pilus biogenesis and natural competence, albeit through its interaction with c-di-AMP, another bacterial second messenger (Samir *et al.*, 2023). These data, along with the recently demonstrated roles of CdgR in the regulation of cell size (Zeng *et al.*, 2023), cyanobacterial motility, and DNA uptake, highlight the unique properties of the ComFB family proteins as dedicated receptors for both c-di-NMP second messengers, c-di-GMP and c-di-AMP.

The demonstration of c-di-GMP-binding by ComFB family proteins leaves many questions to be addressed in future studies. One of them is the ability of (some) ComFB proteins to bind a metal atom, such as Zn²⁺, and, if so, the potential regulatory role(s) of this metal atom. The Zn²⁺-binding Cys residues of *Bs*ComFB are conserved in other members of the phylum *Bacillota*, as well as in Deferribacterota, Thermotogota, and some spirochetes (Fig. S1), suggesting that these proteins may also contain Zn²⁺ atoms. The presence of metal atoms in the ComFB proteins from Betaproteobacteria and Gammaproteobacteria, which contain a single conserved Cys residue (Fig. S1) is uncertain. Some alteromonads, however, have an additional Cys residue in the Cys-Cys motif (Fig. S1). Accordingly, a metalloproteome study of the marine bacterium *Alteromonas* sp. BB2-AT2 found that its ComFB contained Zn²⁺ ions (Mazzotta *et al.*, 2021). Anyway, given the structural similarity and similar c-di-GMP binding properties of Zn²⁺-containing *Bs*ComFB and Zn²⁺-less CdgR (Fig. 1A), the structural role of Zn²⁺ atoms is likely to be dispensable in the ComFB family proteins but might play a role in the regulation of their binding affinities.

In conclusion, this work, along with the recent finding on cyanobacterial *comFB* (Zeng *et al.*, 2023, Samir *et al.*, 2023), shows that diverse members of the ComFB family serve as dedicated receptors for c-di-GMP and/or c-di-AMP and play important regulatory roles in diverse lineages of bacteria. The ability of at least some ComFB proteins to bind two distinct second messengers and the diversity of *comFB*-containing operons (Fig. 2) and ComFB domain architectures (e.g., in combination with PATAN, REC, CheR, PspA/IM30, immunoglobulin-like, or other domains; Fig. 3 and Fig. S3), suggest that the spectrum of these regulatory roles could be quite wide. This establishes the ComFB superfamily as new c-di-NMP receptor proteins widespread in almost all bacterial domains.

MATERIALS AND METHODS

Sequence and structure analysis

Structural alignment of the ComFB protein from *B. subtilis* (GenBank: CAB15563; PDB: 4WAI) and *Sy*ComFB (SIr1970) from *Synechocystis* sp. PCC 6803 (GenBank: BAA18199; PDB: 8HJA) was performed with Dali (Holm, 2022). Secondary structure elements were taken from the Dali alignment and adjusted using DSSP outputs of the PDB data. The alignment was visualized with PyMoI (Schrödinger, LLC). Closely and distantly related members of the ComFB/CdgR superfamily were retrieved through iterative searches of the NCBI protein database using PSI-BLAST (Altschul *et al.*, 1997) and of UniProt using HMMer (Potter *et al.*, 2018), respectively. In addition, InterPro (Paysan-Lafosse *et al.*, 2023) was used to retrieve ComFB family members assigned to the Pfam domain PF10719 (InterPro entry IPR019657). Additional proteins, such as *Anabaena cylindrica* Anacy_5104 (UniProt: K9ZNS1) or *Trichormus variabilis* Ava_3600 (UniProt: Q3M730), that were annotated as ComFB in UniProt but not included in these InterPro and Pfam families, were verified with HHpred (Zimmermann *et al.*, 2018) and then used as queries in further PSI-BLAST and HMMer searches. The sequences obtained (see Fig. S1 and Table S1) were added to the ComFB superfamily.

The presence or absence of ComFB family members in certain bacterial lineages, as well as ComFB domain counts in individual bacterial genomes, were evaluated using taxonomy-restricted BLAST searches and taxonomy-based sorting of the BLAST outputs. Sequence alignments of distant family members against *B. subtilis* ComFB and *Sy*ComFB (Slr1970) structures were verified using HHpred (Zimmermann *et al.*, 2018) tool of the MPI Bioinformatics Toolkit (Gabler *et al.*, 2020). Domain architectures of the ComFB proteins were analyzed using the CDD (Wang *et al.*, 2023) and Pfam (Mistry *et al.*, 2021) databases; domain assignments were checked with HHpred (Zimmermann *et al.*, 2018).

Cluster map analysis

To gather protein sequences for cluster analysis, we queried the UniProt database for homologs of the ComFB protein sequences from *B. subtilis* (P39146), *Synechocystis* sp. PCC 6803 (P74113), *Vibrio cholerae* (Q9KM28), and *Fischerella muscicola* (A0A2N6JYB2, A0A2N6JYC8) using BLAST with an E-value threshold of 1e-04 and a "max_target_seqs" parameter of 20,000. We aggregated the full-length sequences of the resulting matches, removing any flagged as "Fragment," and then filtered them using MMseqs2 to retain sequences with a maximum pairwise

identity of 70% and a length coverage of at least 70%. The resulting filtered set, comprising a total of 1,626 sequences, was subjected to clustering analysis using CLANS (CLuster ANalysis of Sequences) tool (Frickey & Lupas, 2004) based on all-against-all pairwise P-values calculated with BLAST. The clustering was performed until equilibrium was reached in a 2D space, applying a P-value cutoff of 1e-06 and using the default settings in CLANS. The clustering was visualized using CLANS.

Protein production and purification

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574575

576

577

The plasmids and primers used in this study are listed in Table S3. The plasmids expressing comFB of Bacillus subtilis (BSU 35460, GenBank: CAB15563) and Vibrio cholerae (VC A0561, GenBank: AAF96463.1) were constructed using polymerase chain reaction (PCR) using a genomic DNA B. subtilis and V. cholerae as a template, while comFB of Thermoanaerobacter brockii (Thebr 1103, GenBank: ADV79679.1) and Treponema denticola (Tde 1406: GenBank: AAS11923.1) were constructed using a designed gBlocks DNA fragments (IDT, USA) as described previously (Selim et al., 2018). The comFB genes were cloned into a pET-28a vector with Gibson assembly, incorporating a C-terminal 8xHis-tag. Positive clones were selected on agar plates supplemented with 50 µg/ml kanamycin. The recombinant proteins were expressed and purified as described previously in E. coli strain LEMO21 (DE) (Selim et al., 2021b, Selim et al., 2019). Briefly, the 8xHis-tagged constructs were expressed by overnight cultivation of E. coli cells at 20°C in the presence of 0.5 mM IPTG and purified by immobilized metal affinity chromatography using Ni²⁺-Sepharose resin (CytivaTM), followed by size exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE HealthCare, Munich, Germany), as described previously (Selim et al., 2020). Protein purity was assessed by Coomassie-stained SDS-PAGE, and protein concentrations were determined using Bradford assay.

Size exclusion chromatography and multiangle light scattering analysis (SEC-MALS)

Analytical size exclusion chromatography was carried out at room temperature, as described previously, using the ÄKTA purifier (GE Healthcare) on Superdex^{TM 200} prep-grade column (GE Healthcare) coupled to a multiangle light scattering (MALS) detector (Selim *et al.*, 2019, 2020; (Walter *et al.*, 2019)). The protein samples diluted 1:4, centrifuged for 5 min at 15000 g, and 250 µl of the supernatant were injected into the column with a flow rate of 0.4 ml/min, after equilibrating the column with the running buffer (50 mM Tris/HCl, pH 8.0, 300 mM NaCl). ASTRA software

(Wyatt) was used for data analysis and molecular mass calculations using the MALS data. The elution volume was plotted against the UV signal and molecular mass (Selim *et al.*, 2019).

Isothermal titration calorimetry (ITC) analysis and nucleotides binding assays

Binding of c-di-GMP or c-di-AMP by recombinantly produced ComFB proteins was analyzed by isothermal titration calorimetry (ITC), as described previously (Haffner et al., 2023a, Mantovani et al., 2024). Protein samples for ITC were dialyzed against the assay buffer (50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 0.55 mM EDTA), then diluted with the same buffer to a working concentration of 30-300 µM, before loading into the ITC cell. C-di-GMP/c-di-AMP was dissolved in the same buffer to the concentration of 1.0 mM and added gradually in 2 µl injections using a microsyringe. ITC measurements were conducted on a MicroCal PEAQ-ITC instrument (Malvern Panalytical, Westborough, MA, USA), at 25 °C, with a reference power of 10 µcal/s. Control experiments to determine the dilution heat of c-di-GMP or c-di-AMP were performed by titrating c-di-GMP or cdi-AMP into a cell filled with assay buffer. Dissociation constants K_D and ΔH values were calculated using the single binding site model with the MicroCal PEAQ-ITC Analysis Software (Malvern Panalytical) after subtracting the dilution heat of the control experiments. In the competition assay of c-di-GMP and c-di-AMP binding to ComFB, the protein was incubated with 150 µM of one of the nucleotides and titrated against 1 mM of the other competing ligand. For reproducibility, different batches of purified ComFB proteins were used for different ITC experiments. The details of ITC measurements are presented in Table 1 and Figures 4 and 5.

Growth of bacteria and strain construction

Constructs were introduced into *B. subtilis* using the transformation of competent cells. Antibiotic selections were on LB-agar plates using spectinomycin (100 μg/ml) or kanamycin (5 μg/ml). All constructs were confirmed by sequencing. The Δ*pdeH::kan* construct and plasmid pKB149, which carries a constitutive promoter and confers spectinomycin resistance, were kind gifts from Dan Kearns (Indiana University) (Subramanian *et al.*, 2017). For overexpression, the coding sequence of *comFB* was isolated by PCR using the P_c-forward and Pc-Reverse primers (Table S3). The resulting DNA fragment was inserted into pKB149 after cutting with *Hind*III and *BamH*I, placing comFB under the control of a constitutive promoter (P_c). The resulting plasmid was linearized using *Sca*I to favor replacement recombination and transformed into *B. subtilis* IS75 for insertion into the ectopic *amyE* locus. Correct insertion was confirmed by testing on starch-plates and the overexpression of ComFB was verified by Western blotting, using an antiserum raised against

purified ComFB. The empty vector (pKB149) was similarly inserted in *amyE* to produce a control strain.

Swim test

The method used was as described in (Hall *et al.*, 2018). Single colonies were picked with toothpicks and inoculated into LB fortified with 0.35% agar. The plates were grown overnight at 30 C in a humidified incubator, then transferred to 37 C for further growth, and photographed when the control strain (BD9422) had nearly reached the outer edge of its Petri dish.

ACKNOWLEDGEMENTS

The KAS laboratory is funded by grants from the German Research Foundation (DFG) as part of the priority research program (SPP2389; SE 3449/1-1) and by the collaborative research center SFB1381 (project number: 403222702). KAS also gratefully acknowledges the infrastructural support and funding by the Cluster of Excellence "Controlling Microbes to Fight Infections (CMFI)" (EXC2124–390838134) and by the Excellence Strategy of the German Federal and Baden-Württemberg State Governments (Projektförderung: PRO-SELIM-2022-14). MYG was supported by the Intramural Research Program of the National Library of Medicine at the U.S. National Institutes of Health. We thank Shan-Ho-Chou (Huazhong Agricultural University) for the initial version of (Fig. 1A), Karl Forchhammer (Tübingen University), Jörg Stülke (Göttingen University) and Annegret Wilde (Freiburg University) for continued support and constructive discussions. We are also grateful to Marcus Hartmann (MPI for Biology Tübingen), and to Heinz Grenzendorf and Filipp Oesterhelt (Tübingen University) for their excellent support and/or assistance. NIH Grant R01GM123487 supported the work in Dubnau lab. DD acknowledges valuable discussions with Dan Kearns, his expert advice, his generous gift of strains, and additional discussions with Louisa Celma and Jeanie Dubnau.

REFERENCES

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.
- Amikam, D., and Galperin, M.Y. (2006) PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **22**: 3-6.

Angerer, V., Schwenk, P., Wallner, T., Kaever, V., Hiltbrunner, A., and Wilde, A. (2017) The protein SIr1143 is an active diguanylate cyclase in *Synechocystis* sp. PCC 6803 and interacts with the photography (Pagding) **163**: 920, 930

- interacts with the photoreceptor Cph2. *Microbiology (Reading)* **163**: 920-930.
- 643 Chatterjee, D., Cooley, R.B., Boyd, C.D., Mehl, R.A., O'Toole, G.A., and Sondermann, H. (2014)
 644 Mechanistic insight into the conserved allosteric regulation of periplasmic proteolysis by
 645 the signaling molecule cyclic-di-GMP. *eLife* **3**: e03650.
- 646 Chen, Y., Chai, Y., Guo, J.H., and Losick, R. (2012) Evidence for cyclic di-GMP-mediated signaling in *Bacillus subtilis*. *J Bacteriol* **194**: 5080-5090.
- 648 Chin, K.H., Lee, Y.C., Tu, Z.L., Chen, C.H., Tseng, Y.H., Yang, J.M., Ryan, R.P., McCarthy, Y.,
 649 Dow, J.M., Wang, A.H., and Chou, S.H. (2010) The cAMP receptor-like protein CLP is a
 650 novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in
 651 *Xanthomonas campestris. J Mol Biol* **396**: 646-662.
- 652 Chou, S.H., and Galperin, M.Y. (2016) Diversity of c-di-GMP-binding proteins and mechanisms.
 653 *J. Bacteriol.* **198**: 32-46.
- 654 Commichau, F.M., Dickmanns, A., Gundlach, J., Ficner, R., and Stülke, J. (2015) A jack of all 655 trades: The multiple roles of the unique essential second messenger cyclic di-AMP. *Mol.* 656 *Microbiol.* **97**: 189-204.
- 657 Commichau, F.M., and Stülke, J. (2018) Coping with an essential poison: a genetic suppressor 658 analysis corroborates a key function of c-di-AMP in controlling potassium ion 659 homeostasis in Gram-positive bacteria. *J. Bacteriol.* **200**: e00166-00118.
- 660 Conner, J.G., Zamorano-Sanchez, D., Park, J.H., Sondermann, H., and Yildiz, F.H. (2017) The 661 ins and outs of cyclic di-GMP signaling in *Vibrio cholerae*. *Curr Opin Microbiol* **36**: 20-29.
- Corrigan, R.M., Campeotto, I., Jeganathan, T., Roelofs, K.G., Lee, V.T., and Gründling, A. (2013)
 Systematic identification of conserved bacterial c-di-AMP receptor proteins. *Proc. Natl.* Acad. Sci. USA 110: 9084-9089.
- 665 Corrigan, R.M., and Gründling, A. (2013) Cyclic di-AMP: another second messenger enters the fray. *Nat. Rev. Microbiol.* **11**: 513-524.
- Damke, P.P., Celma, L., Kondekar, S.M., Di Guilmi, A.M., Marsin, S., Depagne, J., Veaute, X., Legrand, P., Walbott, H., Vercruyssen, J., Guerois, R., Quevillon-Cheruel, S., and Radicella, J.P. (2022) ComFC mediates transport and handling of single-stranded DNA during natural transformation. *Nat Commun* **13**: 1961.
- Davies, B.W., Bogard, R.W., Young, T.S., and Mekalanos, J.J. (2012) Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* **149**: 358-370.
- 674 El Mouali, Y., Kim, H., Ahmad, I., Brauner, A., Liu, Y., Skurnik, M., Galperin, M.Y., and Römling, U. (2017) Stand-alone EAL domain proteins form a distinct subclass of EAL proteins

involved in regulation of cell motility and biofilm formation in enterobacteria. *J. Bacteriol.* **199**: e00179-00117.

- Enomoto, G., Wallner, T., and Wilde, A. (2023) Control of light-dependent behaviour in cyanobacteria by the second messenger cyclic di-GMP. *microLife* **4**: uqad019.
- Fang, X., Ahmad, I., Blanka, A., Schottkowski, M., Cimdins, A., Galperin, M.Y., Römling, U., and Gomelsky, M. (2014) GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. *Mol. Microbiol.* **93**: 439-452.
- Frickey, T., and Lupas, A. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**: 3702-3704.
- Gabler, F., Nam, S.Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A.N., and Alva, V. (2020) Protein sequence analysis using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* **72**: e108.
- Gallagher, K.A., Schumacher, M.A., Bush, M.J., Bibb, M.J., Chandra, G., Holmes, N.A., Zeng, W., Henderson, M., Zhang, H., Findlay, K.C., Brennan, R.G., and Buttner, M.J. (2020) c-di-GMP arms an anti-sigma to control progression of multicellular differentiation in *Streptomyces. Mol. Cell* **77**: 586-599 e586.
- 692 Galperin, M.Y. (2023) All DACs in a row: Domain architectures of bacterial and archaeal diadenylate cyclases. *J. Bacteriol.* **205**: e0002323.
- Galperin, M.Y., and Chou, S.H. (2020) Structural conservation and diversity of PilZ-related domains. *J. Bacteriol.* **202**: e00664-00619.
- Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D., and Koonin, E.V.
 (2021) COG database update: focus on microbial diversity, model organisms, and
 widespread pathogens. *Nucleic Acids Res* 49: D274-D281.
- Gao, X., Mukherjee, S., Matthews, P.M., Hammad, L.A., Kearns, D.B., and Dann, C.E., 3rd (2013) Functional characterization of core components of the *Bacillus subtilis* cyclic-di-GMP signaling pathway. *J Bacteriol* **195**: 4782-4792.
- Haffner, M., Hou, V.-T., Mantovani, O., Walke, P.R., Hauf, K., Borisova, M., Hagemann, M., Zhou, C.-Z., Forchhammer, K., and Selim, K.A. (2023a) PII signal transduction superfamily acts as a valve plug to control bicarbonate and ammonia homeostasis among different bacterial phyla. *BioRxiv*: doi: 10.1101/2023.1108.1110.552651.
- Haffner, M., Mantovani, O., Spät, P., Maček, B., Hagemann, M., Forchhammer, K., and Selim, K.A. (2023b) Diurnal rhythm causes metabolic crises in the cyanobacterial mutants of c-di-AMP signalling cascade. *BioRxiv*: doi: 10.1101/2023.1111.1114.567006.
- Hall, A.N., Subramanian, S., Oshiro, R.T., Canzoneri, A.K., and Kearns, D.B. (2018) SwrD (YIzI)
 promotes swarming in *Bacillus subtilis* by increasing power to flagellar motors. *J Bacteriol* 200.

- Hall, C.L., and Lee, V.T. (2018) Cyclic-di-GMP regulation of virulence in bacterial pathogens.

 Wiley Interdiscip Rev RNA 9: wrna.1454.
- He, J., Yin, W., Galperin, M.Y., and Chou, S.H. (2020) Cyclic di-AMP, a second messenger of primary importance: tertiary structures and binding mechanisms. *Nucleic Acids Res* **48**: 2807-2829.
- Hickman, J.W., and Harwood, C.S. (2008) Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor. *Mol. Microbiol.* **69**: 376-389.
- Hisert, K.B., MacCoss, M., Shiloh, M.U., Darwin, K.H., Singh, S., Jones, R.A., Ehrt, S., Zhang,
 Z., Gaffney, B.L., Gandotra, S., Holden, D.W., Murray, D., and Nathan, C. (2005) A
 glutamate-alanine-leucine (EAL) domain protein of *Salmonella* controls bacterial survival
 in mice, antioxidant defence and killing of macrophages: role of cyclic diGMP. *Mol. Microbiol.* 56: 1234-1245.
- Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Res* **50**: W210-W215.
- Jenal, U., Reinders, A., and Lori, C. (2017) Cyclic di-GMP: second messenger extraordinaire.

 Nat. Rev. Microbiol. **15**: 271-284.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool,
 K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard,
 A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen,
 S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T.,
 Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., and
 Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*596: 583-589.
- Junkermeier, E.H., and Hengge, R. (2021) A novel locally c-di-GMP-controlled exopolysaccharide synthase required for bacteriophage N4 infection of *Escherichia coli*. mBio **12**: e0324921.
- 738 Khan, A., Hauk, V.J., Ibrahim, M., Raffel, T.R., and Blumer-Schuette, S.E. (2020)
 739 *Caldicellulosiruptor bescii* adheres to polysaccharides via a type IV pilin-dependent
 740 mechanism. *Appl Environ Microbiol* **86**: e00200-00220.
- 741 Khan, F., Jeong, G.J., Tabassum, N., and Kim, Y.M. (2023) Functional diversity of c-di-GMP receptors in prokaryotic and eukaryotic systems. *Cell Commun Signal* **21**: 259.
- Kovacs, A.T., Eckhardt, T.H., van Hartskamp, M., van Kranenburg, R., and Kuipers, O.P. (2013) Functional analysis of the ComK protein of *Bacillus coagulans. PLoS One* **8**: e53471.
- Kovacs, A.T., Smits, W.K., Mironczuk, A.M., and Kuipers, O.P. (2009) Ubiquitous late
 competence genes in *Bacillus* species indicate the presence of functional DNA uptake
 machineries. *Environ. Microbiol.* 11: 1911-1922.

- Krasteva, P.V., Fong, J.C., Shikuma, N.J., Beyhan, S., Navarro, M.V., Yildiz, F.H., and
 Sondermann, H. (2010) *Vibrio cholerae* VpsT regulates matrix production and motility by
 directly sensing cyclic di-GMP. *Science* 327: 866-868.
- Krüger, L., Herzberg, C., Rath, H., Pedreira, T., Ischebeck, T., Poehlein, A., Gundlach, J., Daniel,
 R., Volker, U., Mader, U., and Stülke, J. (2021) Essentiality of c-di-AMP in *Bacillus* subtilis: Bypassing mutations converge in potassium and glutamate homeostasis. *PLoS* Genet 17: e1009092.
- Li, S., Liu, Q., Duan, C., Li, J., Sun, H., Xu, L., Yang, Q., Wang, Y., Shen, X., and Zhang, L. (2023) c-di-GMP inhibits the DNA binding activity of H-NS in *Salmonella*. *Nat Commun* **14**: 7502.
- Londono-Vallejo, J.A., and Dubnau, D. (1993) *comF*, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. *Mol. Microbiol.* **9**: 119-131.
- Makarova, K.S., Koonin, E.V., Haselkorn, R., and Galperin, M.Y. (2006) Cyanobacterial response regulator PatA contains a conserved N-terminal domain (PATAN) with an alpha-helical insertion. *Bioinformatics* **22**: 1297-1301.
- Mantovani, O., Haffner, M., Selim, K.A., Hagemann, M., and Forchhammer, K. (2023) Roles of second messengers in the regulation of cyanobacterial physiology: the carbon-concentrating mechanism and beyond. *microLife* **4**: ugad008.
- Mantovani, O., Haffner, M., Walke, P., Elshereef, A.A., Wagner, B., Petras, D., Forchhammer, K., Selim, K.A., and Hagemann, M. (2024) The redox-sensitive R-loop of the carbon control protein SbtB contributes to the regulation of the cyanobacterial CCM. *Sci Rep* **14**: 7885.
- 770 Mantovani, O., Reimann, V., Haffner, M., Herrmann, F.P., Selim, K.A., Forchhammer, K., Hess, 771 W.R., and Hagemann, M. (2022) The impact of the cyanobacterial carbon-regulator 772 protein SbtB and of the second messengers cAMP and c-di-AMP on CO₂-dependent 773 gene expression. *New Phytol* **234**: 1801-1816.
- 774 Matsuyama, B.Y., Krasteva, P.V., Baraquet, C., Harwood, C.S., Sondermann, H., and Navarro,
 775 M.V. (2016) Mechanistic insights into c-di-GMP-dependent control of the biofilm regulator
 776 FleQ from *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **113**: E209-E218.
- 777 Mazzotta, M.G., McIlvin, M.R., Moran, D.M., Wang, D.T., Bidle, K.D., Lamborg, C.H., and Saito,
 778 M.A. (2021) Characterization of the metalloproteome of *Pseudoalteromonas* (BB2-AT2):
 779 biogeochemical underpinnings for zinc, manganese, cobalt, and nickel cycling in a
 780 ubiquitous marine heterotroph. *Metallomics* **13**: mfab060.
- Millman, A., Melamed, S., Amitai, G., and Sorek, R. (2020) Diversity and classification of cyclicoligonucleotide-based anti-phage signalling systems. *Nat Microbiol* **5**: 1608-1615.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L.,
 Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., and Bateman, A.
 (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res* 49: D412-D419.

- Nesper, J., Hug, I., Kato, S., Hee, C.S., Habazettl, J.M., Manfredi, P., Grzesiek, S., Schirmer, T., Emonet, T., and Jenal, U. (2017) Cyclic di-GMP differentially tunes a bacterial flagellar motor through a novel class of CheY-like regulators. *eLife* **6**: e28842.
- Nie, L., Xiao, Y., Zhou, T., Feng, H., He, M., Liang, Q., Mu, K., Nie, H., Huang, Q., and Chen, W. (2024) Cyclic di-GMP inhibits nitrate assimilation by impairing the antitermination function of NasT in *Pseudomonas putida*. *Nucleic Acids Res* **52**: 186-203.
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi,
 M.L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer,
 A., Mi, H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A.,
 Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., and Bateman, A. (2023)
 InterPro in 2022. *Nucleic Acids Res* 51: D418-D427.
- Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res* **46**: W200-W204.
- Römling, U., Galperin, M.Y., and Gomelsky, M. (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77**: 1-52.
- 801 Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., de 802 Vroom, E., van der Marel, G.A., van Boom, J.H., and Benziman, M. (1987) Regulation of 803 cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**: 279-804 281.
- Samir, S., Doello, S., Zimmer, E., Haffner, M., Enkerlin, A.M., Mueller, T., Dengler, L., Lambidis, S.P., Sivabalasarma, S., Albers, S.V., and Selim, K.A. (2023) The second messenger c-di-AMP controls natural competence via ComFB signaling protein. *bioRxiv*: doi: 2023.2011.2027.568819.
- Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y., and Tabata, S. (2007) A large-scale protein protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res* **14**: 207-216.
- Savakis, P., De Causmaecker, S., Angerer, V., Ruppert, U., Anders, K., Essen, L.O., and Wilde,
 A. (2012) Light-induced alteration of c-di-GMP level controls motility of *Synechocystis*sp. PCC 6803. *Mol. Microbiol.* **85**: 239-251.
- Schuergers, N., Ruppert, U., Watanabe, S., Nürnberg, D.J., Lochnit, G., Dienst, D., Mullineaux, C.W., and Wilde, A. (2014) Binding of the RNA chaperone Hfq to the type IV pilus base is crucial for its function in *Synechocystis* sp. PCC 6803. *Mol Microbiol* **92**: 840-852.
- 818 Selim, K.A., and Alva, V. (2024) PII-like signaling proteins: a new paradigm in orchestrating cellular homeostasis. *Curr Opin Microbiol* **79**: 102453.
- Selim, K.A., Haase, F., Hartmann, M.D., Hagemann, M., and Forchhammer, K. (2018) P_{II}-like signaling protein SbtB links cAMP sensing with cyanobacterial inorganic carbon response. *Proc Natl Acad Sci U S A* **115**: E4861-E4869.

Selim, K.A., Haffner, M., Burkhardt, M., Mantovani, O., Neumann, N., Albrecht, R., Seifert, R.,
 Kruger, L., Stülke, J., Hartmann, M.D., Hagemann, M., and Forchhammer, K. (2021a)
 Diurnal metabolic control in cyanobacteria requires perception of second messenger

- signaling molecule c-di-AMP by the carbon control protein SbtB. *Sci Adv* **7**: eabk0568.
- Selim, K.A., Haffner, M., Mantovani, O., Albrecht, R., Zhu, H., Hagemann, M., Forchhammer, K., and Hartmann, M.D. (2023) Carbon signaling protein SbtB possesses atypical redox-regulated apyrase activity to facilitate regulation of bicarbonate transporter SbtA. *Proc Natl Acad Sci U S A* **120**: e2205882120.
- Selim, K.A., Haffner, M., Watzer, B., and Forchhammer, K. (2019) Tuning the in vitro sensing and signaling properties of cyanobacterial P_{II} protein by mutation of key residues. *Sci Rep* **9**: 18985.
- Selim, K.A., Lapina, T., Forchhammer, K., and Ermilova, E. (2020) Interaction of N-acetyl-l-glutamate kinase with the P_{II} signal transducer in the non-photosynthetic alga *Polytomella parva*: Co-evolution towards a hetero-oligomeric enzyme. *FEBS J* **287**: 465-482.
- Selim, K.A., Tremiño, L., Marco-Marín, C., Alva, V., Espinosa, J., Contreras, A., Hartmann, M.D., Forchhammer, K., and Rubio, V. (2021b) Functional and structural characterization of PII-like protein CutA does not support involvement in heavy metal tolerance and hints at a small-molecule carrying/signaling role. *FEBS J* **288**: 1142-1162.
- Sellner, B., Prakapaite, R., van Berkum, M., Heinemann, M., Harms, A., and Jenal, U. (2021) A new sugar for an old phage: A c-di-GMP-dependent polysaccharide pathway sensitizes Escherichia coli for bacteriophage infection. *mBio* **12**: e03246-03221.
- Skotnicka, D., Steinchen, W., Szadkowski, D., Cadby, I.T., Lovering, A.L., Bange, G., and Sogaard-Andersen, L. (2020) CdbA is a DNA-binding protein and c-di-GMP receptor important for nucleoid organization and segregation in *Myxococcus xanthus*. *Nat Commun* **11**: 1791.
- Slavik, K.M., and Kranzusch, P.J. (2023) CBASS to cGAS-STING: The origins and mechanisms of nucleotide second messenger immune signaling. *Annu. Rev. Virol.* **10**: 423-453.
- Stülke, J., and Krüger, L. (2020) Cyclic di-AMP signaling in bacteria. *Annu Rev Microbiol* **74**: 159-179.
- Subramanian, S., Gao, X., Dann, C.E., 3rd, and Kearns, D.B. (2017) Motl (DgrA) acts as a molecular clutch on the flagellar stator protein MotA in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **114**: 13537-13542.
- Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., and Breaker, R.R. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321: 411-413.
- Sysoeva, T.A., Bane, L.B., Xiao, D.Y., Bose, B., Chilton, S.S., Gaudet, R., and Burton, B.M. (2015) Structural characterization of the late competence protein ComFB from *Bacillus* subtilis. *Biosci Rep* **35**: e00183.

- Tschowri, N., Schumacher, M.A., Schlimpert, S., Chinnam, N.B., Findlay, K.C., Brennan, R.G., and Buttner, M.J. (2014) Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell* **158**: 1136-1147.
- Valentini, M., and Filloux, A. (2019) Multiple roles of c-di-GMP signaling in bacterial pathogenesis. *Annu Rev Microbiol* **73**: 387-406.
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe,
 O., Wood, G., Laydon, A., Zidek, A., Green, T., Tunyasuvunakool, K., Petersen, S.,
 Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N.,
 Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022) AlphaFold
 Protein Structure Database: massively expanding the structural coverage of proteinsequence space with high-accuracy models. *Nucleic Acids Res* **50**: D439-D444.
- Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov,
 M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A.,
 Zidek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney,
 E., Steinegger, M., Hassabis, D., and Velankar, S. (2024) AlphaFold Protein Structure
 Database in 2024: providing structure coverage for over 214 million protein sequences.
 Nucleic Acids Res **52**: D368-D375.
- Wallner, T., Pedroza, L., Voigt, K., Kaever, V., and Wilde, A. (2020) The cyanobacterial phytochrome 2 regulates the expression of motility-related genes through the second messenger cyclic di-GMP. *Photochem Photobiol Sci* **19**: 631-643.
- Walter, J., Selim, K.A., Leganés, F., Fernández-Piñas, F., Vothknecht, U.C., Forchhammer, K.,
 Aro, E.M., and Gollan, P.J. (2019) A novel Ca²⁺-binding protein influences photosynthetic
 electron transport in *Anabaena* sp. PCC 7120. *Biochim Biophys Acta Bioenerg* **1860**:
 519-532.
- Wang, J., Chitsaz, F., Derbyshire, M.K., Gonzales, N.R., Gwadz, M., Lu, S., Marchler, G.H.,
 Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki,
 C.J., and Marchler-Bauer, A. (2023) The conserved domain database in 2023. *Nucleic Acids Res* 51: D384-D388.
- Wang, Y.C., Chin, K.H., Tu, Z.L., He, J., Jones, C.J., Sanchez, D.Z., Yildiz, F.H., Galperin, M.Y.,
 and Chou, S.H. (2016) Nucleotide binding by the widespread high-affinity cyclic di-GMP
 receptor MshEN domain. *Nat Commun* 7: 12481.
- Waters, C.M., Lu, W., Rabinowitz, J.D., and Bassler, B.L. (2008) Quorum sensing controls biofilm formation in *Vibrio cholerae* through modulation of cyclic di-GMP levels and repression of *vpsT. J. Bacteriol*. **190**: 2527-2536.
- Whitfield, G.B., Marmont, L.S., Bundalovic-Torma, C., Razvi, E., Roach, E.J., Khursigara, C.M.,
 Parkinson, J., and Howell, P.L. (2020) Discovery and characterization of a Gram-positive
 Pel polysaccharide biosynthetic gene cluster. *PLoS Pathog* **16**: e1008281.
- Whitney, J.C., Colvin, K.M., Marmont, L.S., Robinson, H., Parsek, M.R., and Howell, P.L. (2012)
 Structure of the cytoplasmic region of PelD, a degenerate diguanylate cyclase receptor

901 that regulates exopolysaccharide production in Pseudomonas aeruginosa. J. Biol. 902 Chem. 287: 23582-23593. 903 Witte, G., Hartung, S., Buttner, K., and Hopfner, K.P. (2008) Structural biochemistry of a 904 bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA 905 recombination intermediates. Mol. Cell 30: 167-178. 906 Wolfe, A.J., and Visick, K.L. (2008) Get the message out: cyclic-di-GMP regulates multiple 907 levels of flagellum-based motility. J Bacteriol 190: 463-475. 908 Yoon, S.H., and Waters, C.M. (2021) The ever-expanding world of bacterial cyclic 909 oligonucleotide second messengers. Curr Opin Microbiol 60: 96-103. Zeng, X., Huang, M., Sun, Q.X., Peng, Y.J., Xu, X., Tang, Y.B., Zhang, J.Y., Yang, Y., and Zhang, 910 911 C.C. (2023) A c-di-GMP binding effector controls cell size in a cyanobacterium. Proc Natl 912 Acad Sci U S A 120: e2221874120. 913 Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, 914 J., Lupas, A.N., and Alva, V. (2018) A completely reimplemented MPI Bioinformatics 915 toolkit with a new HHpred server at its core. J Mol Biol 430: 2237-2243. 916 Zouhir, S., Abidi, W., Caleechurn, M., and Krasteva, P.V. (2020) Structure and multitasking of the 917 c-di-GMP-sensing cellulose secretion regulator BcsE. mBio 11: e01303-01320. 918

FIGURE LEGENDS

919

920

924

926

927

928929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

Figure 1. Sequence and structural conservation within the ComFB superfamily.

921 (A) Structural alignment of the dimeric forms of *Bacillus subtilis* ComFB (*Bs*ComFB, PDB: <u>4WAI</u>,

yellow and teal) and CdgR from Synechocystis sp. PCC 6803 (PDB: 8HJA, orange and red). The

923 CdgR-bound c-di-GMP molecules are shown in stick mode with carbon atoms in blue. The c-di-

GMP binding residues D53, N100, R101 and Y115 of CdgR are shown in stick mode with carbon

925 atoms in green.

(B) Sequence alignment of representative members of the ComFB superfamily. Proteins are shown under their UniProt identifiers, and secondary structure assignments (H, α-helix, Ε, βstrand) of ComFB and CdgR are shown with their PDB codes. The numbers indicate the positions of the start and end of the alignment and the lengths of the gaps between the aligned blocks. Conserved negatively (D, E) and positively (K, R) charged residues are shown in red and blue. respectively; nonpolar hydrophilic residues (N, Q, S, T) are in purple. Conserved hydrophobic residues are indicated with yellow shading, and conserved turn residues (G, P, S, A) are shaded green. Zinc-binding Cys residues of ComFB and the conserved Cys residues in other proteins are shown on a light blue background. The last sequence in the upper block represents the Pfam entry PF10719. The symbols in the "Function" line indicate (as specified by Zeng et al., 2023): d, residues responsible for protein dimerization; asterisks, residues involved in binding c-di-GMP; h and π, residues involved in hydrophobic interactions with the c-di-GMP ligand. The lower block represents ComFB-related sequences that are not recognized by the PF10719 sequence model; its top two lines display secondary structure predictions by AlphaFold (Varadi et al., 2022) and HHpred (Zimmermann et al., 2018). The last three lines show sequences of the N-terminal ComFB domains of four-domain diguanylate cyclases. The sequences in the upper block are from the following organisms: ComFB, Bacillus subtilis; Q8YS15, Nostoc sp. PCC 7120; P74113, Synechocystis sp. PCC 6803 (both, cyanobacteria); E4S4A5, Caldicellulosiruptor acetigenus; Q0AV46, Syntrophomonas wolfei; E8USF0, Thermoanaerobacter brockii (all three, Clostridia); D3P9N1, Deferribacter desulfuricans (Deferribacterota); H2J3K4, Marinitoga piezophila (Thermotogota); Q73MV1, Treponema denticola; H9UJ27, Spirochaeta africana; D5U3J8, Brachyspira murdochii (all three, Spirochaetota); Q21X05, Albidiferax ferrireducens; A2SG25, Methylibium petroleiphilum (both, Betaproteobacteria); Q9KM28, Vibrio cholerae; Q8EG00, Shewanella oneidensis (both, Gammaproteobacteria); Q728W3, Desulfovibrio vulgaris; Q313N7, Oleidesulfovibrio alaskensis (both, Thermodesulfobacteriota). All sequences in the lower block are from cyanobacteria.

(C) Cluster map of ComFB homologs. A set of 1,626 representative ComFB sequences (≤ 70% pairwise identity and ≥ 70% length coverage) was clustered using the CLANS tool (Frickey & Lupas, 2004) based on pairwise BLAST P-values. Dots represent individual sequences, colored according to their group. Line color intensity reflects sequence similarity, with darker lines indicating higher similarity. The analysis revealed four clusters: two within Cyanobacteriota, one comprising diverse phyla (e.g., Actinomycetota, Bacillota), and a distinct Pseudomonadota cluster, highlighting conserved c-di-GMP-binding residues across these diverse groups.

Figure 2. Genomic neighborhoods of selected ComFB/CdgR family proteins.

Genomic fragments are listed with the organism names, GenBank accession numbers, and genomic coordinates. Gene sizes are drawn approximately to scale, and gene names are from GenBank, RefSeq, and/or the COG database. *ComFB* genes are in red, other competence-related genes are in pink, flagella-related genes are in orange, pili-related genes are in green, signal transduction genes are in yellow, metabolic genes are in various shades of blue, poorly characterized genes are in grey or white. The graph displays fragments of the following genomes: A, *Bacillus subtilis* 168, GenBank accession AL009126: 3,643,558..3,637,338; B, *Synehocystis* sp. PCC 6803, BA000022: 1,776,983..1,783,355; C, *Thermoanaerobacter tengcongensis* MB4, AE008691: 1,261,742..1,271,495; D, *Allochromatium vinosum* DSM 180, CP001896: 3,319,557..3,327,133; E, *Desulfohalobium retbaense* DSM 5692, CP001734: 792,659..799,553; F, *Treponema denticola* ATCC 35405, AE017226: 1,452,441..1,444,590; G, *Vibrio cholerae* O1 biovar El Tor str. N16961 chromosome II, AE003853: 495,055..503,153. The genomic fragments for *B. subtilis* and *T. denticola* are shown in reverse complement.

Figure 3. Structural gallery of representative ComFB domain-containing proteins from various species. α-helices in the ComFB domain are colored red, β-strands are in yellow, and the remainder of the protein in grey. For proteins with two ComFB domains, one domain is shown in lighter shades. The structures are AlphaFold2 (Jumper et al., 2021) predictions from UniProt/AlphaFold DB (Varadi et al., 2024), except for Bacillus subtilis (PDB 4WAI), The species represented include Synechocystis sp. PCC 6803 (Slr1970, Sll1170, Slr1505, and Sll1739, UniProt accessions P74113, P74197, P73943, and P73385, respectively), Vibrio cholerae variabilis (Q9KM28), Fischerella muscicola (A0A2N6JYB2), Trichormus (Q3M730),Synechococcus sp. PCC 7502 (K9SSQ8), and Pseudanabaena sp. PCC 7367 (K9SGA4). Several of these proteins also contain additional features, such as N- or C-terminal extensions

(e.g., SIr1505), coiled-coil segments (e.g., PspA), or other domains: SII1739 and SIr1970 have uncharacterized α -helical bundle domains, and SII1170 contains DUF1816, PAS, and GGDEF domains.

Figure 4. Isothermal titration calorimetry (ITC) analysis of c-di-GMP binding to phylogenetically different ComFB proteins. Upper panels show the raw ITC data in the form of heat produced during the titration of c-di-GMP on different ComFB proteins; lower panels show the binding isotherms and the best-fit curves according to the one binding site model. (A-D) ITC analysis of c-di-GMP binding to *B. subtilis* or *T. brockii* ComFB proteins in the absence (A,C) or presence of 150 μM c-di-AMP (B,D). (E,F) ITC analysis of c-di-GMP binding to *V. cholerae* (E) or *T. denticola* (F) ComFB proteins.

Figure 5. Isothermal titration calorimetry (ITC) analysis of c-di-AMP binding to phylogenetically different ComFB proteins. Upper panels show the raw ITC data in the form of heat produced during the titration of c-di-GMP on different ComFB proteins; lower panels show the binding isotherms and the best-fit curves according to the one binding site model. (A-D) ITC analysis of c-di-AMP binding to *B. subtilis* or *T. brockii* ComFB proteins in the absence (A,C) or presence of 150 μM c-di-GMP (B,D). (E,F) ITC analysis of c-di-AMP binding to *V. cholerae* (E) or *T. denticola* (F) ComFB proteins in the presence of 150 μM c-di-GMP.

Figure 6. ComFB inhibits swimming. The swimming assay was conducted as described in Methods by inoculating cells into 0.3% agar LB plates. Plasmid vectors carrying *comFB* under the control of a constitutive promoter or the same vector without comFB (empty vector) were inserted separately at amyE in wild type and $\Delta pdeH$ backgrounds. The image was acquired after 20 hours of growth at 30 C in a humidified chamber, followed by a further 5 hours at 37 C.

1010

1011

1012

1013

10141015

1016

1017

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

Supplementary Materials Table S1. A list of representative ComFB family members in various bacterial phyla. Table S2. Phylogenetic distribution of ComFB, ComFA and ComFC proteins, according to the COG database. **Table \$3**. Plasmids and primers used in this study. Figure S1. Sequence alignment of representative ComFB/CdqR family members Figure S2. Sequence logo of ComFB/CdgR family, based on an alignment of 180 distinct family members. Figure S3. Domain architectures of various ComFB proteins. Figure S4. Additional representative structures of ComFB domain-containing proteins from various species. The structures are AlphaFold2 predictions obtained from the UniProt/AlphaFold DB, with the exception of Treponema denticola, which was predicted independently using AlphaFold2. The species represented include *Thermoanaerobacter brockii* (UniProt E8USF0), Halothece sp. PCC 7418 (K9YBB1), Roseburia sp. CAG:380 (R6WME1), and Omnitrophica WOR 2 bacterium (A0A1G1PT98). Four of these proteins feature additional domains: T. denticola contains an Ig-like domain, *Halothece* sp. PCC 7418 a PATAN domain, *Roseburia* sp. CAG:380 a REC domain, and *Omnitrophica* WOR 2 bacterium a CheR domain. Figure S5. The oligomeric state of different ComFB proteins was determined by SEC-MALS. Figure S6. NanoDSF of c-di-NMP binding to ComFB proteins from T. brockii and T. denticola. Light scattering thermograms were obtained by measuring the attenuation of the back-reflected light intensity passing through the protein sample with or without 0.5 mM c-di-AMP or c-di-GMP as a function of temperature. Figure S7. BD9422 (amyE::pDR511), BD9385 (amyE::pDR511 pdeH::kan), BD9389 (Pc-comFB), BD9398 (Pc-comFB pdeH::kan) and the wild-type strain IS75, were grown in liquid LB medium with vigorous shaking at 37 C. Growth was followed as turbidity with a Klett colorimeter.

Figure 1, A,C

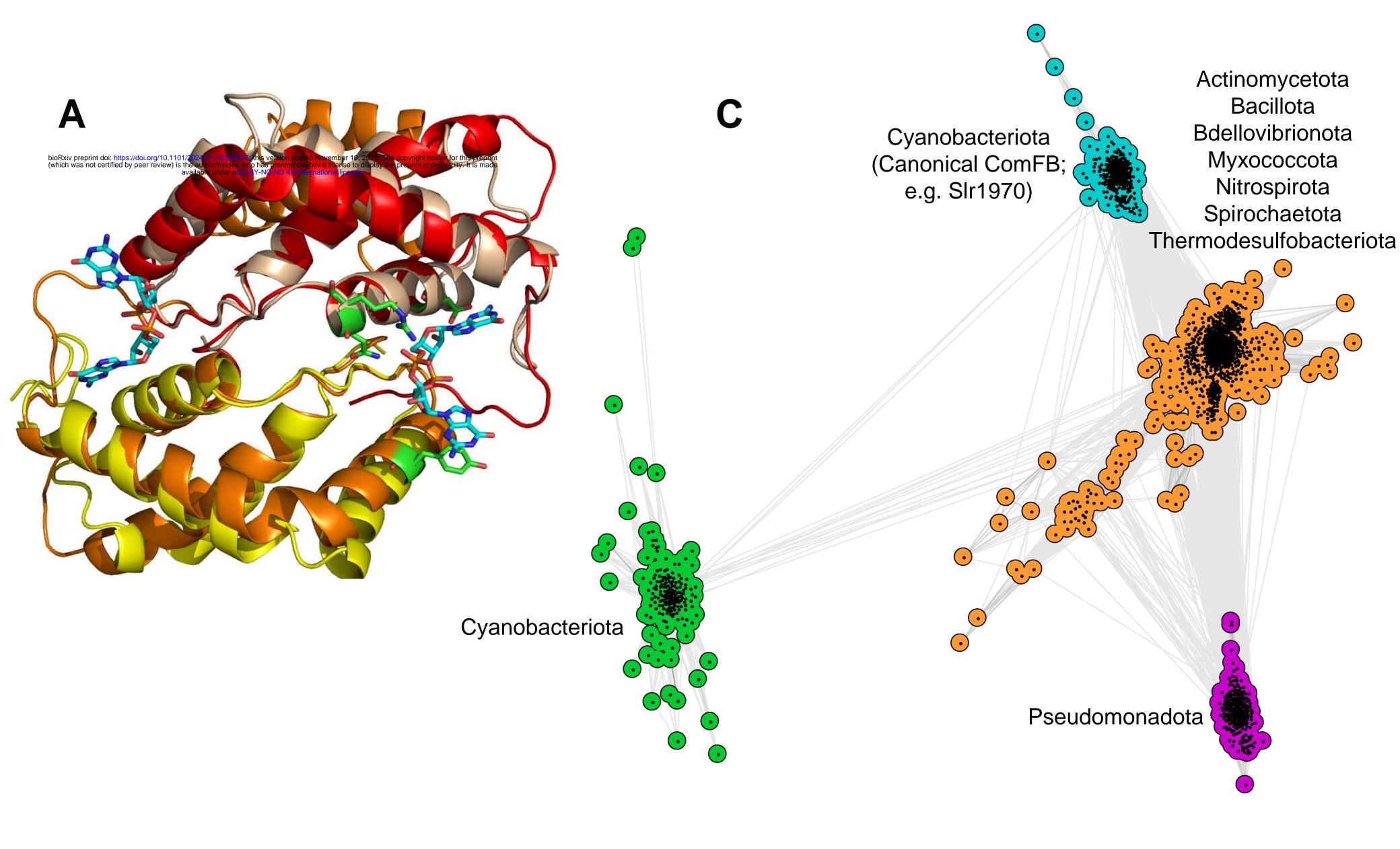
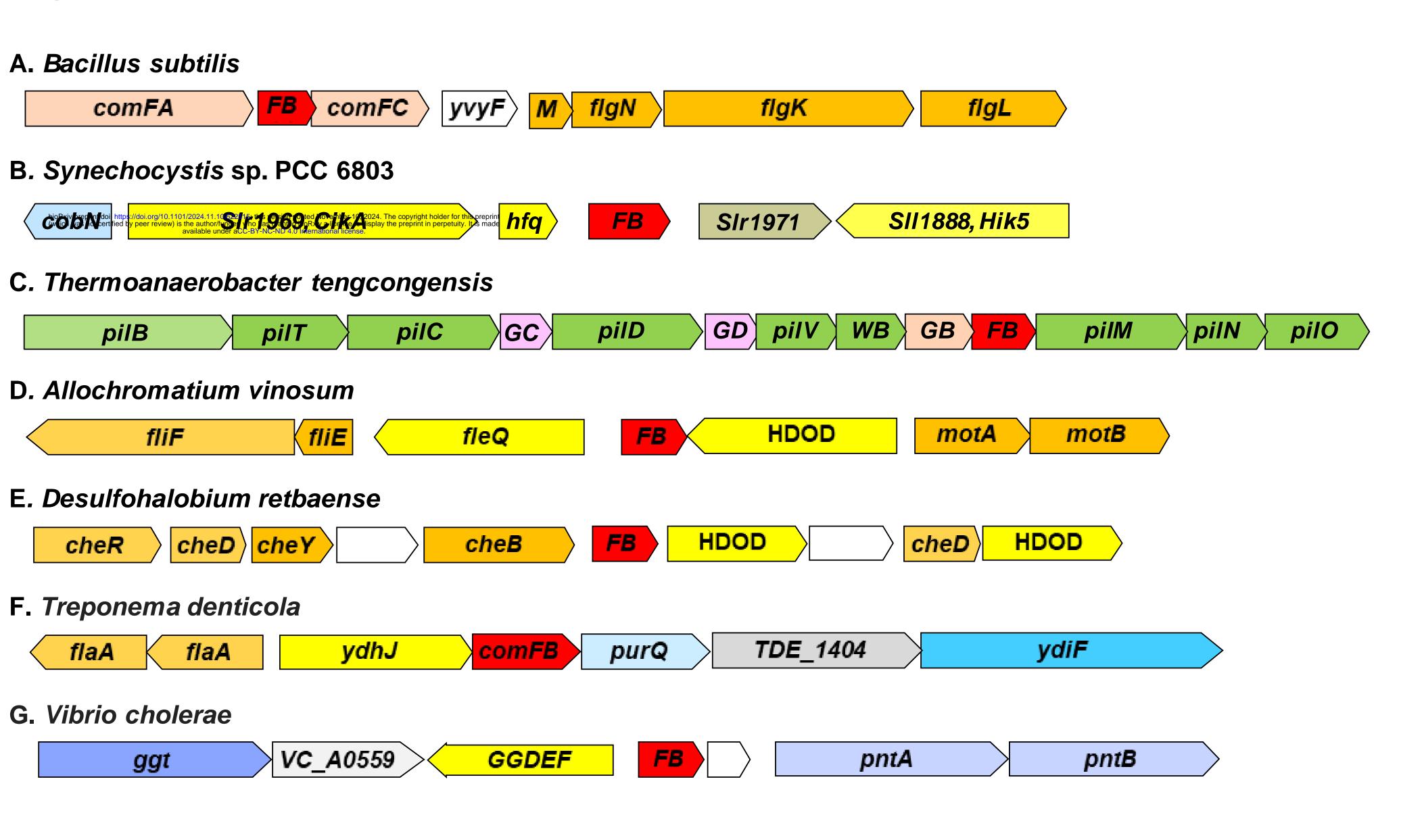
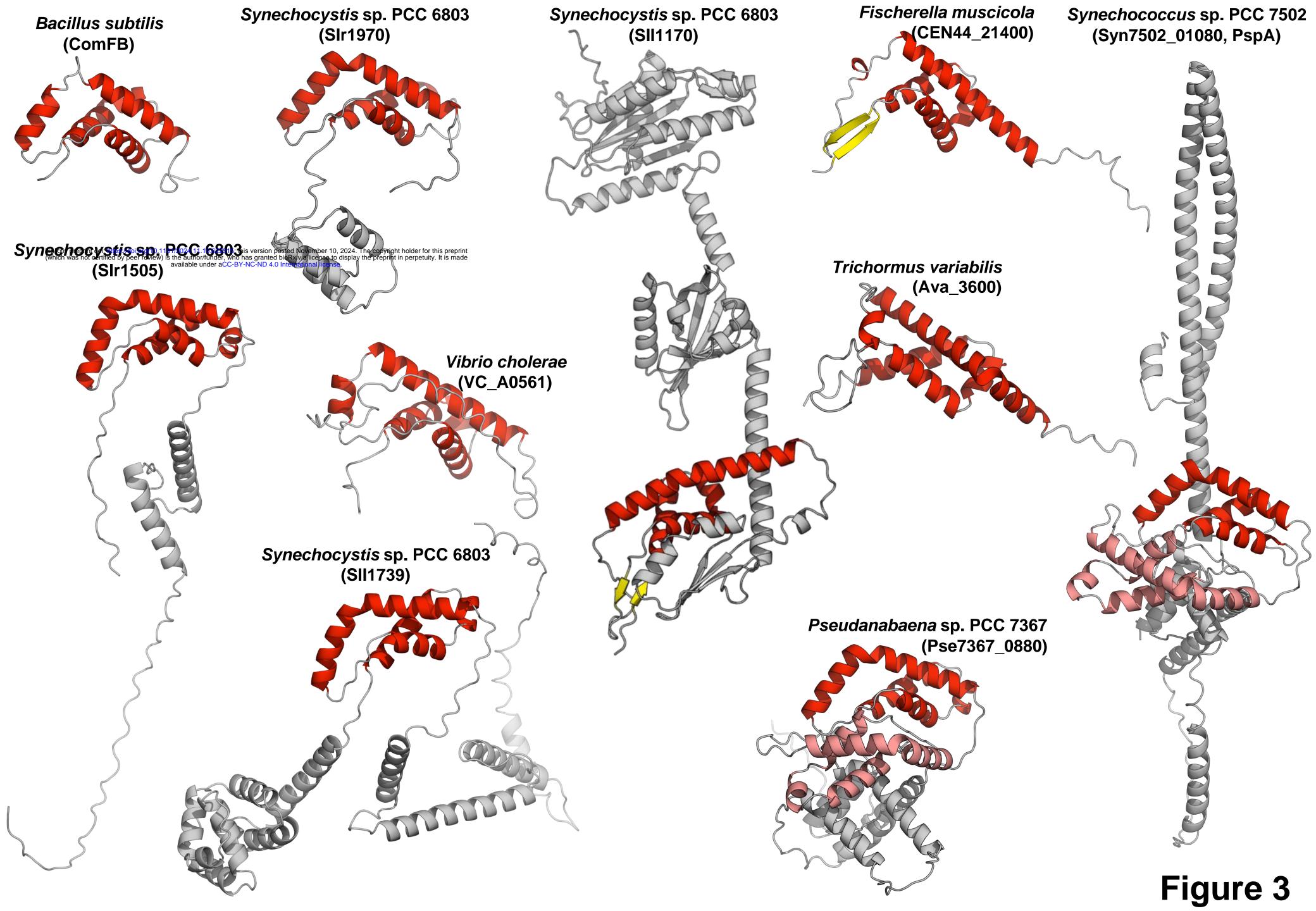


Figure 1B

```
1 --M<mark>lvnskeivm</mark>kel<mark>l</mark>dry<mark>m</mark>dqlhma<mark>c</mark>tcqvcqndvlals<mark>lnkvsp</mark>s<mark>yv</mark>tdfkkia<mark>y</mark>tkaelvdkqknta<mark>m</mark>lvi<mark>l</mark>aes<mark>a</mark>av<mark>v</mark>ses<mark>p</mark>sdl<mark>c</mark>qtk
COMFB BACSU
PDB: 4WAI
                                        59 HKK<mark>finvmeemv</mark>lte<mark>v</mark>vsq<mark>v</mark>skyqkttekqpdia<mark>diaay</mark>a<mark>lnrlpp</mark>l<mark>ya</mark>tseegae<mark>y</mark>qrqrasee-lefl<mark>i</mark>qqq<mark>v</mark>kdg<mark>l</mark>gr<mark>y</mark>fdr<mark>p</mark>qiadrkp 150
P74113 SYNY3
                                59 RKQ<mark>finvmeelv</mark>lte<mark>a</mark>iar<mark>v</mark>aeieatsessidvgdia<mark>ay</mark>a<mark>lnrlpp</mark>lyatteegas<mark>y</mark>qrirakae-lqel<mark>i</mark>aqq<mark>v</mark>sea<mark>i</mark>nlnldq<mark>p</mark>ndnrtpv 154
Q8YS15 NOSS1
PDB:8HJA
                                        1 MLE<mark>irnameeiv</mark>dew<mark>l</mark>dwv<mark>i</mark>nqld-v<mark>c</mark>gcekcradikals<mark>lnrlpp</mark>r<mark>yy</mark>vsedgki<mark>y</mark>srleelkiqfrvd<mark>v</mark>iaa<mark>l</mark>aaa<mark>a</mark>fi<mark>v</mark>rsnprhnrvse
E4S4A5 CALA7
                                 16 DMK<mark>viniveelv</mark>wec<mark>f</mark>kev<mark>i</mark>ehrpgi<mark>c</mark>rcdr<mark>c</mark>madvcala<mark>lnk</mark>mk<mark>p</mark>ryvasdkgeviskaaflekelkie<mark>l</mark>lva<mark>v</mark>ada<mark>v</mark>eivkvn<mark>p</mark>rheme--
Q0AV46 SYNWW
E8U Software print dei intro-graphic programment of the programment of
                                   9 IDK<mark>iknine</mark>kr<mark>v</mark>wel<mark>l</mark>inf<mark>y</mark>eehpef<mark>c</mark>tcrdcvldvaaitlntipphyq--vsddistaikk---isdee<mark>i</mark>lkv<mark>i</mark>lea<mark>a</mark>ir<mark>v</mark>skyphhi----
D3P9N1 DEFDS
H2J3K4 MARPK
                                    5 KYH<mark>ITNVMEHIV</mark>EEITNEM<mark>F</mark>AMPN<mark>2</mark>MCICDRCRADVIALALNHLHPKYVVTEKGRLYSELQNYTFQTRAEVLTEVLKAMEKVKEHPSHPKEES
                                   1 -MI<mark>IHNVMEDLV</mark>YTE<mark>V</mark>NKL<mark>F</mark>DEAE<mark>7</mark>TCSCMQ<mark>C</mark>RVDTM<mark>CY</mark>V<mark>LNR</mark>VK<mark>P</mark>RYIKSGRGLA<mark>H</mark>FLKFEKTE--KVQ<mark>I</mark>MAD<mark>I</mark>TSL<mark>V</mark>IEGMHRVLSTKRPH
Q73MV1 TREDE
H9UJ27 SPIAZ
                                   1 -MK<mark>v</mark>rniledevfrvikei<mark>a</mark>dedq6fattdq<mark>c</mark>rvdaa<mark>cy</mark>v<mark>lnr</mark>lppryv</mark>ssgrgqa<mark>y</mark>adtefslnpqleadils<mark>l</mark>vheglrrvttirrayytg
                                   1 -MY<mark>LVNLMEQKV</mark>SKL<mark>A</mark>DSY<mark>L</mark>KEKN-IPVNTNMRMDII<mark>AY</mark>T<mark>LNR</mark>VQ<mark>P</mark>Q<mark>YV</mark>TSARGVL<mark>H</mark>SYNRDP-IQSNTE<mark>I</mark>LSI<mark>I</mark>ADA<mark>C</mark>DI<mark>V</mark>SRRKENTLLES
D5U3J8 BRAM5
                                   5 FEQ<mark>vhnyyerlv</mark>fea<mark>v</mark>arqsaahp--gftsdmladva<mark>cv</mark>alnrlparyvrhdvdmm<mark>f</mark>ylteherhaieqs<mark>m</mark>aeg<mark>l</mark>ifa<mark>f</mark>af<mark>v</mark>adrvakrqalv
Q21X05 ALBFT
                                   5 FAG<mark>irny</mark>qeravfda<mark>v</mark>rtwrdhyp--lisddllpdva<mark>cv</mark>a<mark>lnr</mark>lpa<mark>ryi</mark>rhevdlv<mark>f</mark>yqsekerlmaerl<mark>v</mark>asa<mark>v</mark>gfa<mark>f</mark>ef<mark>v</mark>qartamrara-
A2SG25 METPP
                                   4 SED<mark>v</mark>hnymetlvgqvlgqpeyse---tydqdqladla<mark>cl</mark>a<mark>l</mark>cqlr<mark>p</mark>iyirhdidflsalpeaklvilkdh<mark>a</mark>lia<mark>v</mark>qaaesm<mark>i</mark>vndrrrnrdsd
Q9KM28 VIBCH
                                   2 QLE<mark>IRNYYEVLL</mark>MEMLSDEGLMD---ELPEDYLADLC<mark>CV</mark>T<mark>LNQLP</mark>VRYIRHLVDTY<mark>F</mark>FEDYDELQELKRE<mark>I</mark>QAA<mark>L</mark>EKSRAF<mark>L</mark>KQNLQKRMQEE
Q8EG00 SHEON
                              171 LSRTRnr<mark>ne</mark>ir<mark>v</mark>vka<mark>i</mark>rtv<mark>l</mark>geppvytpdpkdiqdiyalalnalppryaqh--gti<mark>v</mark>lrdpv----rdeq<mark>i</mark>lea<mark>v</mark>rda<mark>f</mark>irvmeh<mark>p</mark>ka-----
Q728W3 DESVH
Q313N7 OLEA2
                                 16 FYH<mark>irnrnearv</mark>lrmmekv<mark>l</mark>teppgmtpsaadlhdiyalalns<mark>lpa</mark>ryaqq--gsivlrdpv----rdedvleavrkaiatv</mark>vhn<mark>p</mark>kq-----
                                   1 -----MEELVLEELEEQLEELPDY<mark>C</mark>TCERCLADVAayaLnrlppryvvseegeayqkleserqqlraeilsavreaiervkkn<mark>p</mark>-
PF10719
                                                                                                         Function
                                        AlphaFold
                                        HHpred
                                   1 MET<mark>v</mark>vklshqs<mark>v</mark>ige<mark>i</mark>esv<mark>ldtyl6afaipd<mark>lrqel</mark>isfvinrlpgfysavpdsqisihed<mark>13</mark>plgqhlh<mark>l</mark>qnl<mark>i</mark>hqg<mark>i</mark>fs<mark>i</mark>iqqksdwisnh 109</mark>
K9ZNS1 ANACC
                                   1 MKT<mark>v</mark>vnltqqs<mark>v</mark>lge<mark>i</mark>envldt<mark>yp6</mark>afaisd<mark>lrqeliaf</mark>v<mark>lnrip</mark>cf<mark>y</mark>savseeqlllqnc<mark>13</mark>pleqqlh<mark>l</mark>qrl<mark>i</mark>hqg<mark>i</mark>ds<mark>i</mark>fqaksdwishh 109
K9QQR6 NOSS7
                                   1 MKT<mark>l</mark>vnltqqs<mark>v</mark>vge<mark>i</mark>esvldt<mark>yp6</mark>afaipd<mark>lrqelivf</mark>vltrlpsfdgamseghislaea<mark>13</mark>pleqqlh<mark>l</mark>qnl<mark>i</mark>hqg<mark>icli</mark>vqeksdwindr 109
B2J248 NOSP7
Q8YQX4 NOSS1
                                   2 IKT<mark>vvn</mark>ltqqs<mark>v</mark>lge<mark>i</mark>envldtyp6afaipd<mark>lrqeliaf</mark>vlnripclysaisdeqllipsc<mark>13</mark>pleqqqh<mark>l</mark>qrl<mark>i</mark>hqg<mark>i</mark>ys<mark>i</mark>fqeksdwishh 110
Q3M730 TRIV2
                                   2 IKT<mark>vvn</mark>ltqqs<mark>v</mark>lge<mark>i</mark>env<mark>l</mark>dt<mark>yp6</mark>afaipd<mark>lrqeliaf</mark>v<mark>lnr</mark>ipcl<mark>y</mark>saisdeqllipsr<mark>13</mark>pleqqqh<mark>l</mark>qrl<mark>i</mark>hqg<mark>i</mark>ys<mark>i</mark>fqeksdwishh 110
K9TZR9 CHRTP
                                   2 TKA<mark>v</mark>vnltlp<mark>mv</mark>veh<mark>i</mark>etaled<mark>yp6</mark>afaipd<mark>lr</mark>qkliayilsrvans<mark>y</mark>tvleegencttfgdcssyrklq<mark>i</mark>enl<mark>a</mark>kqg<mark>i</mark>qe<mark>i</mark>lqqywewivhh
                                   2 SEH<mark>iin</mark>vtai<mark>li</mark>qqe<mark>i</mark>drffvkyp7lfsrpyfqkellkkvlsqipnhh<mark>v</mark>viydkhtlinhdd<mark>3</mark>fieekrk<mark>l</mark>kal<mark>i</mark>nqa<mark>v</mark>pk<mark>v</mark>ltenyqkvgei 102
B1WU94 CROS5
                                   2 TKT<mark>ifn</mark>ltlp<mark>vv</mark>vsc<mark>i</mark>edv<mark>letyp6</mark>afsipdl<mark>r</del>qeliahvlsqvsncyvvaeegqqpclnlg<mark>4</mark>slkdkyc<mark>i</mark>edl<mark>i</mark>hqg<mark>i</mark>en<mark>i</mark>mykrekqlslq 102</mark>
G6FQQ9 9CYAN
G6FVD0 9CYAN
                                   1 MKT<mark>v</mark>lnltqqs<mark>v</mark>iae<mark>i</mark>esv<mark>l</mark>dtyp6afaipe<mark>lrqdlinyvfsrip</mark>cm<mark>y</mark>saiepkqacclnyk<mark>4</mark>pleyqih<mark>l</mark>etl<mark>i</mark>hqg<mark>i</mark>ys<mark>i</mark>mqeksdgisrh 101
ComFB-DUF1816-PAS-GGDEF
P74197 SYNY3
                                   2 SSY<mark>lvnetlavv</mark>eee<mark>i</mark>drf<mark>ltdfp7alslscfrqqlkakilstipnrhkvrtdedleftsvh<mark>4</mark>lieeklv<mark>i</mark>eer<mark>i</mark>lra<mark>m</mark>pd<mark>m</mark>idslmsrvhqe 103</mark>
                                   2 SYR<mark>v</mark>fnstls<mark>iv</mark>qte<mark>i</mark>dhfsens<mark>a7</mark>als<mark>l</mark>pyfrnklvtkilnnipnhh<mark>v</mark>lihdqtn<mark>y</mark>iddnt<mark>4</mark>vtdeill<mark>i</mark>rkk<mark>i</mark>sey<mark>i</mark>re<mark>v</mark>vteyqvlsvvd 103
AOA2K8WMZ4
                                   2 PYK<mark>lvnetlslverel</mark>adflad<mark>lp7</mark>amslpyfrqtlkakilskipnrhrvvagnemeplsdn4llqervl<mark>v</mark>eek<mark>i</mark>rsmmpnliqnmshqqvvl 103
A0A977L0T4
```

Figure 2

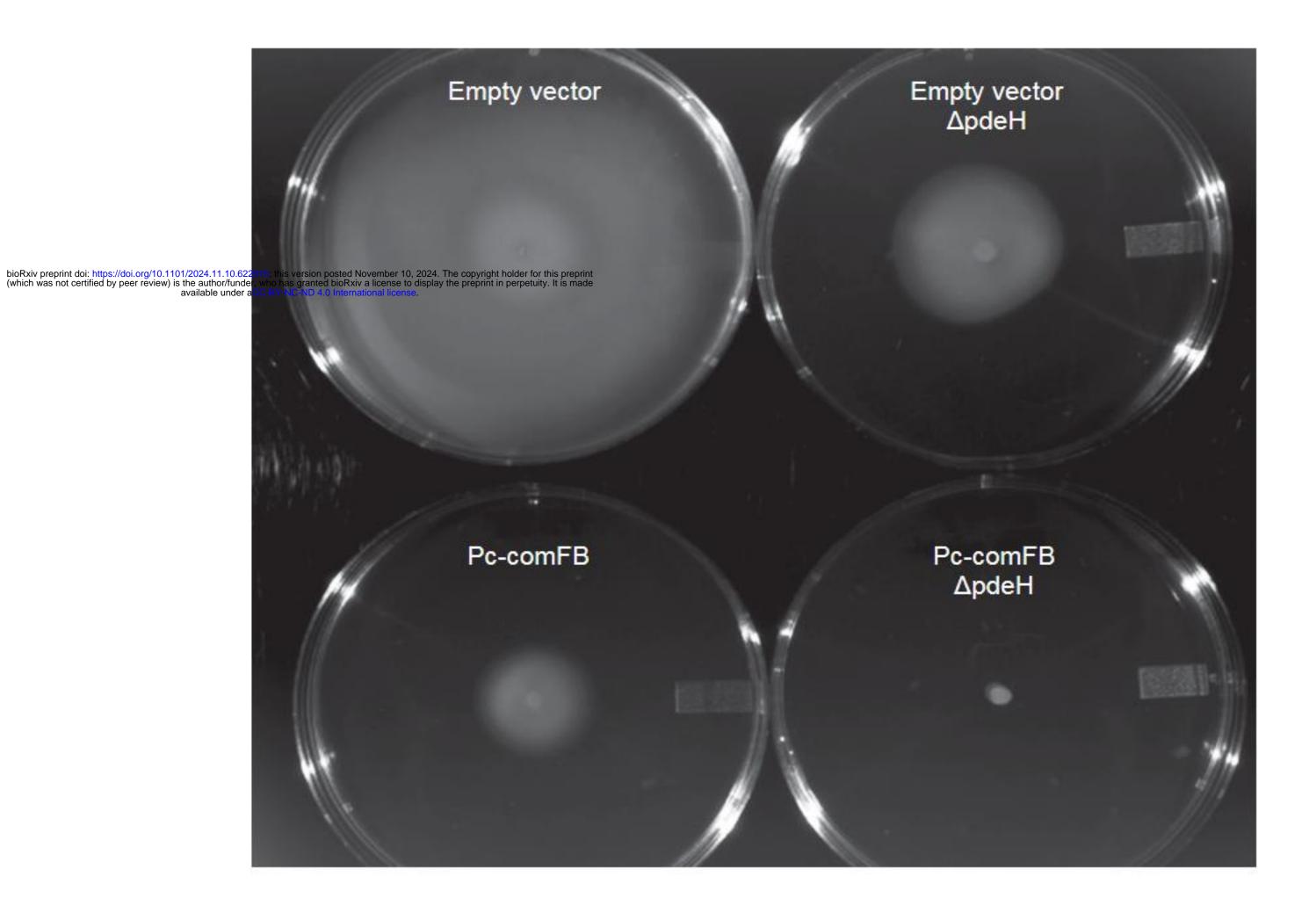




Vibrio cholerae ComFB Bacillus subtilis ComFB Figure 4 B E DP (µcal/s) DP (µcal/s) DP (µcal/s) -8.0--1.2in presence of -1.4-150 µM c-di-AMP 10 15 20 25 30 35 40 45 50 35 25 10 25 20 30 40 10 15 Time (min) Time (min) Time (min) bioRxiv preprint doi: https://doi.org/10.1101/2024.11.10.622515; this version posted November 10-12024. 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 a CC-BY-NC-ND 4.0 International license. ∆H (kcal/mol) ∆H (kcal/mol) ∆H (kcal/mol) Gram-negative Gram-positive 1.6 1.8 1.2 1.4 0.2 0.4 0.6 0.8 0.2 Molar Ratio Molar Ratio Molar Ratio Thermoanaerobacter brockii ComFB Treponema denticola ComFB **C** 0.1-D DP (µcal/s) 0.1 (hcal/s) DP (µcal/s) 0.2 0.3 in presence of 150 µM c-di-AMP ·0.5*-*0.5 5 10 15 20 25 30 35 40 45 50 5 10 15 20 25 30 35 40 45 50 15 25 30 35 10 20 Time (min) Time (min) Time (min) ΔH (kcal/mol) ΔH (kcal/mol ΔH (kcal/mol -1.5-0.2 0.3 0.4 0.5 0.6 0.2 0.3 0.4 0.5 0.6 0.1 Molar Ratio Molar Ratio Molar Ratio

Vibrio cholerae ComFB Bacillus subtilis ComFB Figure 5 E B DP (µcal/s) DP (µcal/s) 0.6 in presence of in presence of 150 µM c-di-GMP 150 μM c-di-GMP 0.2-10 15 20 25 30 35 40 45 50 15 20 25 30 35 40 45 50 Time (min) Time (min) Time (min) bioRxiv preprint doi: https://doi.org/10.1101/2024.11.10.622515; this version posted November 10, 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-NC-ND 4.0 International license ΔH (kcal/mol) ∆H (kcal/mol) ∆H (kcal/mol) negative Gram-positive 0.6 0.8 8.0 0.2 0.4 1.2 1.4 0.6 1.2 0.2 0.4 Molar Ratio Molar Ratio Molar Ratio Thermoanaerobacter brockii ComFB Treponema denticola ComFB C F D DP (µcal/s) DP (µcal/s) DP (µcal/s) 0.6-0.6-10 15 20 25 30 35 40 45 50 15 20 25 30 35 40 45 50 10 Time (min) Time (min) Time (min) in presence of in presence of 150°μM c-di-GMP 150 μM c-di-GMP ΔH (kcal/mol) ∆H (kcal/mol) ∆H (kcal/mol) 1.5-0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 Molar Ratio Molar Ratio Molar Ratio

Figure 6



Cyanobacteriota

P74113 SYNY3/60-143 KK<mark>finvmeemv</mark>lte<mark>v</mark>vsqvskyq-----kttekqpdia<mark>di</mark>aaya<mark>lnrlpp</mark>l<mark>y</mark>atsee<mark>s</mark>ae<mark>y</mark>qrqrasee---lefl<mark>i</mark>qqq<mark>v</mark>kdg<mark>l</mark>gr<mark>y</mark>fdr**p** PDB:8HJA **Q8YS15** NOSS1/60-143 GVHVN<mark>vmetlv</mark>yee<mark>i</mark>dkolrfyp-----knlrnylnlt<mark>ev</mark>atyalnrlpplyassvk<mark>t</mark>veeorrvaaro---yrseltsa<mark>v</mark>rralaa<mark>v</mark>erd**p** P73943 SYNY3/15-98 P73385_SYNY3/125-207 KT<mark>C</mark>LN<mark>LLEVVV</mark>KDE<mark>I</mark>YNE<mark>L</mark>QKLP-----VNVAKFVSIAE<mark>V</mark>ASYT**L**NRL**PP**M<mark>Y</mark>VASEE<mark>C</mark>KA<mark>E</mark>QMQRVEQ----MRGE<mark>I</mark>RTA<mark>V</mark>LRG**I**GA<mark>I</mark>MRD**P** kQ<mark>f</mark>in<mark>vmeelv</mark>ise<mark>v</mark>vkr<mark>v</mark>sevd------gndsaldvg<mark>di</mark>a<mark>aya<mark>l</mark>nrl**pp**ly</mark>attee<mark>g</mark>ad<mark>y</mark>qrqrakee---lrdl<mark>i</mark>aeQ<mark>v</mark>qea<mark>m</mark>grylnr**p** K9T1S5 9CYAN/60-142 E0U819_GLOV7/60-142 K9YPR0_CYASC/60-143 kq<mark>f</mark>in<mark>vmeelv</mark>isq<mark>a</mark>inr<mark>v</mark>aelk-----ssndsldvg<mark>di</mark>a<mark>ayalnrlpp</mark>l<mark>y</mark>attee<mark>s</mark>an<mark>y</mark>qrqrakee---lqel<mark>i</mark>kqq<mark>v</mark>eea<mark>i</mark>yr<mark>y</mark>lar**p** kq<mark>f</mark>in<mark>vmeelv</mark>vse<mark>v</mark>iar<mark>v</mark>aeie-----ttsnrlldvg<mark>di</mark>a<mark>ayalnrlpp</mark>l<mark>y</mark>attee<mark>e</mark>as<mark>e</mark>qrerakee---lsdi<mark>i</mark>nsq<mark>v</mark>eqa<mark>i</mark>sl<mark>y</mark>ler**p** K9WJV2 9CYAN/60-143 KOFINVMEELVLTEAVAQIAEIE----AKSDRILDVGDIAAYALNRLPPLYATTEEGARYORORAKEE---LHDLITQOVSAALTRNLDRP K9UBH3_CHAP6/60-144 ĸQ<mark>f</mark>in<mark>vmeelv</mark>ite<mark>t</mark>isQ<mark>i</mark>taie----vntddvqldvg<mark>di</mark>a<mark>ay</mark>a<mark>lnrlpp</mark>l<mark>y</mark>attee<mark>g</mark>as<mark>y</mark>qreraqte---lkdl<mark>i</mark>rqq<mark>v</mark>esa<mark>i</mark>aqnldr**p** Q116P4_TRIEI/60-143 K9YGV0 HALP7/60-143 kQ<mark>f</mark>in<mark>vmeelv</mark>lse<mark>a</mark>iar<mark>v</mark>aene-----gssdQtldvg<mark>di</mark>a<mark>ayalnrlpp</mark>l<mark>y</mark>attee<mark>a</mark>qQ**f**Qrskakde---lQhl<mark>i</mark>skQ<mark>v</mark>sea<mark>i</mark>dQnitrk KK<mark>CINVMEELV</mark>LTEAITI<mark>V</mark>AEIE-----VTSDQSIDLSD<mark>I</mark>TAYA**LNRLPP**L<mark>Y</mark>ATTEE<mark>ANY</mark>QKQRAQNE---LQEL<mark>I</mark>NQR<mark>V</mark>QEA<mark>I</mark>THHLKSD B2J7T3 NOSP7/65-148 kq<mark>f</mark>in<mark>vmeelv</mark>lte<mark>a</mark>iar<mark>v</mark>aeie-----atsessldvg<mark>di</mark>a<mark>ayalnrlpp</mark>l<mark>y</mark>attee<mark>a</mark>as<mark>e</mark>qrqtaqae---lqel<mark>i</mark>sqq<mark>v</mark>sea<mark>i</mark>nr<mark>y</mark>ldr**p** K7WD54_9NOST/60-143 K9U537_CHRTP/60-143 rq<mark>fin<mark>vmeelv</mark>lsn<mark>v</mark>iak<mark>v</mark>aeie-----vtseisldvg<mark>di</mark>t<mark>ay</mark>a<mark>lnrlpp</mark>l<mark>y</mark>attee<mark>g</mark>ak<mark>y</mark>qrerante---ldal<mark>i</mark>sqq<mark>v</mark>sea<mark>i</mark>ck<mark>y</mark>ldt**p**</mark> KQ<mark>f</mark>in<mark>vm</mark>ee<mark>lv</mark>lte<mark>v</mark>iar<mark>v</mark>aeik-----vtsghsldvg<mark>di</mark>aaya<mark>lnrlpp</mark>l<mark>x</mark>attee<mark>s</mark>an<mark>y</mark>qrqrakee---laal<mark>i</mark>vqq<mark>i</mark>dea<mark>i</mark>arsldr**p** K9SQX5 9SYNE/60-146 KQFIN<mark>vmeelv</mark>laesvar<mark>v</mark>aeie-----satntvldlg<mark>di</mark>aaya**l**nrlpply</mark>attee<mark>s</mark>aa<mark>y</mark>qrkralae---lqtl<mark>i</mark>tqq<mark>v</mark>kea<mark>i</mark>drsldr**p** KQFINVMEELVLTEAMTRIAAIE-----AEQKCALDIG<mark>DI</mark>AAYALNRLPPLYATTEEBAQYQRQRAQEE---IMDIIRSQVQEAINHNMARP Q8DKL0_THEEB/60-143 B8HM73_CYAP4/60-143 B0C3D2_ACAM1/64-147 ko<mark>f</mark>in<mark>vmeelv</mark>lte<mark>a</mark>itr<mark>v</mark>aaie-----angeotldlg<mark>di</mark>aaya<mark>lnrlpp</mark>l<mark>y</mark>atted<mark>a</mark>an<mark>f</mark>orkhaoee---lmdl<mark>i</mark>ioo<mark>v</mark>eea<mark>i</mark>hrnlar**p** KQFIN<mark>vmeelv</mark>lteaitr<mark>v</mark>skie-----atsnitldlgd<mark>i</mark>aaya**l**nrl**pp**l<mark>y</mark>attee<mark>a</mark>an<mark>y</mark>qrqrarqe---lldl<mark>i</mark>iqq<mark>v</mark>eta<mark>i</mark>srnldq**p** KQ<mark>F</mark>IN<mark>VMEELV</mark>LSE<mark>A</mark>MSQ<mark>I</mark>AQQE-----SKESCQLDIG<mark>DI</mark>A<mark>AY</mark>A<mark>LNRLPP</mark>L<mark>Y</mark>ATTHE<mark>A</mark>AD<mark>Y</mark>QRQRAKEE---LKTL<mark>I</mark>AQQ<mark>V</mark>KEG<mark>I</mark>QR<mark>Y</mark>FER**P** B1WRQ0 CROS5/67-150 K9SSR2_9SYNE/196-279 K9SGU8_9CYAN/2-85 lr<mark>i</mark>sn<mark>vmekiv</mark>eea<mark>v</mark>lni<mark>m</mark>sews-----temkaqvntea<mark>i</mark>vafa<mark>lnrlpp</mark>m<mark>y</mark>atsel<mark>g</mark>cr**f**lrsrainq---lhqe<mark>i</mark>nli<mark>v</mark>sda<mark>l</mark>vh<mark>v</mark>rrm**p** VS<mark>lkntiedlv</mark>vee<mark>a</mark>khq<mark>l</mark>qrlg-----nsisqsiels<mark>dv</mark>aaya<mark>l</mark>nll<mark>pp</mark>m<mark>y</mark>astdr<mark>s</mark>wlqqrkrahse---mrpk<mark>v</mark>isa<mark>v</mark>qka<mark>m</mark>mn<mark>v</mark>srds K9SGU8 9CYAN/211-294 YG<mark>i</mark>tn<mark>vleklv</mark>lte<mark>s</mark>eeq<mark>lkrls-----slvsrqvkge**dv**a<mark>ayalnrlpp</mark>m<mark>y</mark>atsaq<mark>e</mark>ykrqqqrahle---lade<mark>i</mark>esa<mark>v</mark>vqa<mark>i</mark>lt<mark>l</mark>skt**p**</mark> as<mark>l</mark>kn<mark>vlediv</mark>lse<mark>a</mark>iaq<mark>l</mark>kaid------ksdhsnls<mark>ei</mark>vaya**lnrlpp**l<mark>y</mark>astdr<mark>e</mark>wlqqrkrahse---lisk<mark>i</mark>qst<mark>v</mark>rqa<mark>i</mark>lg<mark>a</mark>kfd**p** K9SXG6_9SYNE/2-83 YA<mark>FTNALEKLV</mark>VTN<mark>V</mark>EQQMTQLG-----SVLARAVKIE**DA**AAYA**LNRLP**TM<mark>Y</mark>ATSAGGLQ<mark>Y</mark>QQQKAKTD---LADE<mark>I</mark>SSTVIQALLT<mark>L</mark>GKA**P** ENCRNAIEEIVLEEINTQLNRLG-----QDARRNINMGEVVAYA<mark>LNRLPP</mark>MYATTQKGWIQQRKLARDR---LQHQVNSAVRQALMGVRRDP K9SXG6_9SYNE/194-277 K9SGA4_9CYAN/2-85 YN<mark>ysn<mark>ilenlv</mark>tsv<mark>a</mark>lrq<mark>i</mark>akln-----pevvdqlnmd<mark>dv</mark>v<mark>aya<mark>l</mark>nrl**pp**my</mark>atsdr<mark>g</mark>lkklrqkikde---mthe<mark>i</mark>aav<mark>v</mark>rra<mark>v</mark>sk<mark>i</mark>aqe**p**</mark> K9SGA4 9CYAN/196-279 K9SNQ3_9SYNE/4-87 K9SNQ3_9SYNE/200-283 QG<mark>C</mark>RN<mark>AIEELV</mark>VEE<mark>I</mark>KAQ<mark>I</mark>SRLS-----SAVKAKPSLD<mark>EV</mark>A<mark>AY</mark>A<mark>LNRLPP</mark>M<mark>Y</mark>ATTRR<mark>G</mark>YIQQQKRAHTD---MKQE<mark>I</mark>AQT<mark>I</mark>SKG<mark>L</mark>IG<mark>V</mark>KKDS wn<mark>fvntlenlv</mark>asv<mark>a</mark>qrq<mark>i</mark>skla-----dnlsqrvtle<mark>ev</mark>sayalnrlppmx</mark>atses<mark>tlkx</mark>wrerarte---lssd<mark>i</mark>lvt<mark>v</mark>rqg<mark>v</mark>it<mark>i</mark>lksp K9SH83 9CYAN/200-283 fd<mark>fvnvlenpv</mark>lav<mark>v</mark>hrq<mark>l</mark>arlh-----palidqikpe**ev**a<mark>aytlnrlpp</mark>m<mark>y</mark>atsqr<mark>e</mark>lkeqryrvkae---lakd<mark>i</mark>ilm<mark>v</mark>reg<mark>i</mark>si<mark>v</mark>vns**p** SS<mark>LRNALEDIV</mark>IRE<mark>A</mark>QKQLAPLK-----SAIGTSIRLAD<mark>V</mark>VAYTLNRLPPLYATSHHGWQLQRRRLVKE---MRSQ<mark>I</mark>IEA<mark>V</mark>SVA<mark>I</mark>AA<mark>F</mark>STKN K9SP86_9CYAN/2-85 K9SP86_9CYAN/263-346 K9SNM1_9CYAN/2-85 LE<mark>fsnvleklv</mark>llv<mark>a</mark>ghl<mark>v</mark>rnfd-----peiraqinlte<mark>vmayalnrlppmy</mark>aasdr<mark>d</mark>yrrorqyaqse---lanr<mark>v</mark>tdl<mark>v</mark>krg<mark>i</mark>el<mark>v</mark>las**p** AS<mark>FRNVLEDLV</mark>VRE<mark>A</mark>KLR<mark>V</mark>SHLR-----HEAIPKLNVG<mark>EV</mark>I<mark>AYSLNRLPP</mark>M<mark>Y</mark>ATTQQ<mark>=</mark>CLRLRARIRNE---IGYQ<mark>I</mark>SET<mark>V</mark>RRA<mark>I</mark>SA<mark>V</mark>QVGD LG<mark>y</mark>cn<mark>vlerlv</mark>ltl<mark>a</mark>elq<mark>i</mark>srla-----peirdridaa<mark>ev</mark>t<mark>aftinrlpp</mark>m<mark>y</mark>atssr<mark>i</mark>lqquhqrvkae---fass<mark>i</mark>nrtiyha<mark>i</mark>tvivdsp K9SNM1 9CYAN/206-289 K9SXA1_9SYNE/201-284 WN<mark>YTNILENLV</mark>MAVAELQ<mark>L</mark>KRME-----PAAKERITLNEVAAYT<mark>L</mark>NRLPPMY</mark>ATSNK<mark>E</mark>LQQLRIRAKTE---MAHQ<mark>I</mark>VAL<mark>V</mark>REAVVRVREYP SFLVNVIEYPVIQEAISQHSQLA-----VDIRIKFKPIEVAAYALNRLPSMY</mark>VTTLDDYNQQIRFVSND---MKPK<mark>I</mark>SDAVKRAIHSLRMGD K9SSQ8_9SYNE/152-235 K9SZS6_9CYAN/10-93 QT<mark>Y</mark>QN<mark>IMELLV</mark>REE<mark>L</mark>EKQ<mark>L</mark>KQCP-----ETLAQYINTV<mark>EV</mark>AT<mark>YA<mark>LNRLPA</mark>L<mark>Y</mark>ASCEK<mark>G</mark>KN<mark>M</mark>QKLLAQKQ---YREE<mark>I</mark>KKA<mark>V</mark>RQG<mark>L</mark>AA<mark>I</mark>QRD**P**</mark> QSNKN<mark>amellv</mark>aqe<mark>i</mark>kko<mark>l</mark>epcs-----aeikelvnqv<mark>ev</mark>atya<mark>lnrlps</mark>l<mark>y</mark>assee<mark>claw</mark>okorgoke---mool<mark>i</mark>tat<mark>v</mark>rea<mark>l</mark>ev<mark>v</mark>rkd**p** K9T3B6 9CYAN/18-101 KSSRN<mark>VMEVLV</mark>VEEVEKQFQSLP-----AKTAKYIKAS<mark>EV</mark>I<mark>AYA<mark>LNRLPA</mark>LYATSKRGWQRQWHRGKTE---MYQKI</mark>TTAVRQGIIAVQQDP QI<mark>TRNVMELLV</mark>EEEIERQISRLP-----VPISQYINRVEVATYA<mark>LNRLPP</mark>LYASSHEGFNKQKLKGKAE---FSAD<mark>V</mark>TKAVRQGFAAVQKDL K9WUV3_9NOST/22-105 K9XSF2_STAC7/8-91 GVKVN<mark>VMETLV</mark>AQE<mark>V</mark>EQQ<mark>L</mark>KALP-----RASAECITKL<mark>DV</mark>I**TYALNRLPP**L<mark>Y</mark>AASKE<mark>C</mark>IAHQKEEAKQK---HQQA<mark>I</mark>KSA<mark>V</mark>KRA<mark>I</mark>TA<mark>V</mark>QRD**P** K9YBB1 HALP7/326-409 QVHLN<mark>vmeilv</mark>qqe<mark>i</mark>ekq<mark>l</mark>kfyp-----snikaylnri<mark>ei</mark>a<mark>tya<mark>lnrlpa</mark>l<mark>y</mark>asset<mark>c</mark>keqqlkvgqqk---ykse<mark>i</mark>vaa<mark>v</mark>rra<mark>l</mark>aa<mark>v</mark>erd**p**</mark> B7K1H4_RIPO1/10-93 B1WUF5_CROS5/10-93 L8MYS8 9CYAN/2-85 KIHV<mark>NIMELLV</mark>QEE<mark>I</mark>EKQ<mark>L</mark>KLYP-----KNLKNYINKVE<mark>VATYA<mark>LNRLPP</mark>L<mark>Y</mark>ASSAM<mark>G</mark>KE<mark>Y</mark>QKRTGKKQ---YQSQ<mark>I</mark>NLA<mark>V</mark>RRA<mark>L</mark>AA<mark>I</mark>ERD**P**</mark> TS<mark>l</mark>kn<mark>vleeivvvea</mark>qaq<mark>l</mark>kqis-----qpireainls<mark>ev</mark>a**a**fa**lnrle**vl<mark>y</mark>astsrewlqqrkrahse---lknq<mark>v</mark>tsa<mark>v</mark>qqa<mark>l</mark>lg<mark>i</mark>krd**p** YF<mark>l</mark>vn<mark>vledla</mark>ire<mark>v</mark>knq<mark>l</mark>tyma-----nvlprkvgvd<mark>dv</mark>c<mark>ayv<mark>l</mark>nrl**pa**my</mark>atseq<mark>g</mark>vi<mark>w</mark>qtqkakeq---lssq<mark>i</mark>est<mark>v</mark>iqs<mark>l</mark>mtlgkt**p** L8MYS8_9CYAN/193-276 E0UCY2_GLOV7/13-96 E0UCY3_GLOV7/17-99 QI<mark>Y</mark>QN<mark>IMEVLV</mark>KEE<mark>I</mark>QRQ<mark>L</mark>KIYP-----AQIRSSFNIV<mark>EI</mark>EAYA<mark>LNRLP</mark>VL<mark>Y</mark>ASSEK<mark>G</mark>RDMQKEIGKQK---YRKE<mark>I</mark>GIA<mark>V</mark>RRA<mark>L</mark>AT<mark>V</mark>EKN**P** CIHIN<mark>vmevlv</mark>qqe<mark>i</mark>ekq<mark>l</mark>rdys-----crlkpylnrv<mark>ev</mark>atfa**lnrl<mark>pa</mark>ly**astid<mark>e</mark>kn<mark>y</mark>qialgkk----yqqq<mark>i</mark>tla<mark>v</mark>rra<mark>i</mark>aa<mark>v</mark>erd**p** HI<mark>YENTMECLVKEE IEKOL</mark>KNYP-----KTVIKFINKT**E IEAYALNRL PA**LYASSEQGREKQKE IGQQK---YQKQ**I**SLAVSQA**L**VTVREN**P** GIRQNIMELLVKQE IEKQ<mark>L</mark>KHYS-----PQLKPYINKIEVATFALNRL PALYASTIQ KNHQIELAKK----YQEQITLAVRRG IAAVERD P QV<mark>YKNVMEE</mark>LVEEE IDRQTRNFK-----PETAKALNRIDVVSYALNRL PPLYASSQE SVYRQKQRGQQQ---FGQRL RAAVHQS LKVVAQQ P B7K9K2 GLOC7/12-95 B7K9K3_GLOC7/17-99 B0BZM3 ACAM1/9-92 B0CEZ4_ACAM1/9-90 ns<mark>y</mark>kn<mark>imeilv</mark>dee<mark>i</mark>dqqtcaws-----leeaqrinri<mark>ev</mark>a<mark>Aha<mark>lnhlpp</mark>l<mark>y</mark>assqe<mark>s</mark>va<mark>l</mark>qyeraqre---hqgd<mark>i</mark>tta<mark>v</mark>tda<mark>l</mark>ta<mark>v</mark>krip</mark> PL<mark>yrnaleply</mark>leesoro<mark>l</mark>oolp--pkmlgslkpervlao<mark>v</mark>vayalnrl<mark>pg</mark>l<mark>y</mark>atsor<mark>s</mark>wofoohoaok----lrpo<mark>l</mark>vma<mark>v</mark>rogfaa<mark>v</mark>ord**p** K9TDR7 9CYAN/10-96 QT<mark>yknvmeafv</mark>qeeTefqLqnnk----alsrsadyinllevatfa<mark>lnrlpq</mark>yyassiesierqrrrikek-relkqkIsfvvsqafaaverdp ni<mark>yknvmeil</mark>vdeeTeyqLihnr---tvnrnlkkyinpv<mark>ev</mark>atfa<mark>lnrlps</mark>lyasstesinkqrkraliq---ykkeTrqavtqgfaaverdp K9ZA58_CYAAP/6-93 K9YPH5_CYASC/6-92 hk<mark>y</mark>in<mark>vmeelv</mark>eie<mark>v</mark>nko<mark>l</mark>knlp---pmflkyrdclnpv<mark>ei</mark>qt<mark>yalnhlpo</mark>l<mark>y</mark>assta<mark>e</mark>kf**y**qlekgktq---lseq<mark>i</mark>qtt<mark>v</mark>tra<mark>iasv</mark>lrd**p** E0U686 GLOV7/6-92 onsrn<mark>vmemlv</mark>aee<mark>v</mark>ekq<mark>i</mark>eslp-----vktasfikss<mark>ev</mark>a<mark>ay</mark>a<mark>lnrlps</mark>l<mark>y</mark>atsqk<mark>g</mark>wqkqwhhgkte---lcqk<mark>i</mark>tma<mark>v</mark>gqg<mark>i</mark>va<mark>v</mark>qrd**p** D7DY23_NOSA0/22-105 YT<mark>Y</mark>RN<mark>VMESLV</mark>VEE<mark>V</mark>ERQ<mark>M</mark>MSLP-----PKVLQYFNKT<mark>EA</mark>I<mark>AYALNRLPP</mark>L<mark>Y</mark>ATSER<mark>B</mark>WEQQRAKASRE---MQLQ<mark>I</mark>TAA<mark>V</mark>RQA<mark>L</mark>VA<mark>V</mark>QRD**P** K8GHT0 9CYAN/8-91 KS<mark>LENQMERAVLEMV</mark>NEILLMES----QQRY<mark>CFC</mark>EKFCNDAAALALNNLQPRYATSFHGSLRTLEAIQAD-QELQRLTRLEVVKAMDKVAANP -----MERAVLEMVNELLLLES----QQRYCSCERFCHDAAALALNNLQPRYTTSFEGSIYTLEAIQAD-QELQSLTRREVGKAMEIVAANP Q2JNL8 SYNJB/9-96 Q2JUI8 SYNJA/1-82

Bacillota (Firmicutes)

Daciiii	
COMFB	BACSU /1-84
PDB: 4V	VA I

C8WQ62 ALIAD/1-85 Q65EA8 BACLD/1-84 A8FHY2_BACP2/1-84 U5L9B8_9BACI/1-81

F2F6C1 SOLSS/4-90

E8USF0_THEBF/1-84 U5MS12 CLOSA/2-85

Clostridia

M1N0J0_9CLOT/2-85 G8LWC2 HUNCD/12-99 I5ARU2 EUBCE/50-134 D7GSS2 9FIRM/151-235 D9R9Z4 CLOSW/206-289

MP<mark>I</mark>VNMTEV<mark>LARQIL</mark>KDPELQKL-----LP<mark>CHC</mark>DQCLD<mark>DV</mark>LAMALNHLPPRY</mark>ATTDE QLYVQVEYLRP--QLRSD<mark>I</mark>LRELTQAALRVSSHP -MLVNAKEAVLEELFDQY<mark>I</mark>DQLH-----MS**CTC**SR**C**RE**D**VL<mark>A</mark>LALNAVKPQ<mark>Y</mark>VTDQSKLA<mark>Y</mark>IKAELVDK--QKNTSMLVTLAEAARKVNENP -MLINAKETLLKEILYQYLNQLN-----MI<mark>CHC</mark>EK<mark>C</mark>IED<mark>V</mark>LAIS<mark>L</mark>NQVKPQYTDDINKISYSKSEMVDK--QKNTAM</mark>LVILTEAASKVSAFP -----<mark>meeiv</mark>atl<mark>v</mark>kvl<mark>m</mark>lspd----yqtf**cqc**mk**c**ra<mark>di</mark>i<mark>ais<mark>lnnlp</mark>nh<mark>y</mark>vttee<mark>g</mark>rk<mark>m</mark>vfeqlnte--enrkw<mark>i</mark>nkr<mark>i</mark>ina<mark>i</mark>hv<mark>v</mark>sky**p**</mark> IKLIN<mark>v</mark>teeivkglvsflhgve----yqtf**chc**eq**c**emdvnaivlnvlpar<mark>y</mark>vtsdetrdavfkqmstp--eylee<mark>i</mark>nkq<mark>i</mark>iralhivkqy**p**

MQ<mark>L</mark>KN<mark>YM</mark>ED<mark>AV</mark>EQM<mark>M</mark>DRV<mark>L</mark>KDLD------V<mark>CKC</mark>DR<mark>C</mark>RM**DI**K<mark>AL</mark>ALNNL<mark>PP</mark>K<mark>YV</mark>VSEEGEL<mark>Y</mark>VKTNELVR--QFEVD<mark>I</mark>IKA<mark>I</mark>TMA<mark>A</mark>IKVNNNK ng<mark>vinymevlv</mark>eehmeil<mark>l</mark>kesn-----sctcorcksolfalalnnlkeyyvvvtorerviktlomteo--ofotolttievtraleovknnp YN<mark>VINYMEIWU</mark>NEYMDVL<mark>L</mark>ERSG-----GCKCDNCKRDIFTLSLNNLKPCY</mark>VATPIGNIMARLESTKQ--QFETDIIVEITKAINKINSNP AG<mark>LENLTENLV</mark>FQELNDF<mark>I</mark>EKIK-NTPDSQFCICNICLAD<mark>V</mark>AAIV<mark>L</mark>NDLKPQYCSNFIDKHKSKDYYIKH----RLEVREKI</mark>VKAFDMVKKNP YH<mark>v</mark>vn<mark>vmsevm</mark>krenvle<mark>i</mark>mqql-----gv<mark>ctc</mark>qr<mark>c</mark>qa**dv**laltetqtkekycvmydnektllisdyar--ihqke<mark>i</mark>hak<mark>i</mark>mraciv<mark>v</mark>qqhe YR<mark>vvnimeeil</mark>tpe<mark>l</mark>ild<mark>a</mark>lksn-----dt**ctc**sr**c**qadvkalm<mark>ltrlpp</mark>k<mark>y</mark>ivadnttv<mark>p</mark>mlltfyrn--kfrva<mark>v</mark>lsq<mark>s</mark>mracme<mark>v</mark>keh**p** YE<mark>YLNVMEYVV</mark>KNR<mark>V</mark>KDY<mark>M</mark>EKFD------V<mark>C</mark>LCGRCIAD<mark>V</mark>TALALTNLPPKY</mark>IVVEPPSASPLLNFYSN--RYSQH<mark>I</mark>IVE<mark>L</mark>TKACSA<mark>V</mark>KENP

	al	/allable under	acc-br-inc-ind 4.0	miemational	ilcerise.			
F6DRU1 DESRL/5-88	FI <mark>S</mark> N NVMEQIV DCK <mark>L</mark> F							
F3ZVH4 MAHA5/3-86	HA <mark>M</mark> R NYMEDVVEHT<mark>L</mark>F							
Q8RAF2_CALS4/1-84	MQ <mark>L</mark> K NYMEDAVEQM<mark>M</mark>I							
A3DDQ6_HUNT2/2-85	SQ <mark>I</mark> K nymeevvfsl<mark>m</mark>f							
L7VNC9_THES1/1-84	ME <mark>L</mark> K NYMEEIV FNF <mark>I</mark> I							
B8I7C0_RUMCH/2-85	VT <mark>L</mark> K n<mark>ymeevv</mark>fnli							
A6TR22_ALKMQ/1-84	-M <mark>v</mark> k <mark>n</mark> ymeevvnqe <mark>i</mark> n							
H6LJ92_ACEWD/1-84	-M <mark>lknyme</mark> dt <mark>v</mark> daf <mark>l</mark> i							
A0PY62_CLONN/2-87	CQ <mark>l</mark> kn <mark>y</mark> seeavdkl <mark>l</mark> a							
E3DQL1_HALPG/12-98	SE <mark>lnn</mark> iteeivlem <mark>l</mark> e							
E4RNZO_HALHG/12-98	SR <mark>L</mark> H n hted <mark>iv</mark> mqt <mark>l</mark> f							
B8CZE1_HALOH/12-98	KD <mark>l</mark> k n nt e k <mark>iv</mark> let <mark>m</mark> e							
L0K911_HALHC/11-97	ER <mark>l</mark> h n ht e dvvlek <mark>i</mark> e							
F4LVH4_TEPAE/6-89	LK <mark>l</mark> k <mark>nyme</mark> dvvall <mark>l</mark> i							
D9RXL1_THEOJ/5-88	FK <mark>l</mark> k nymedvvfsf<mark>l</mark>i							
BOTEG7_HELMI/1-84	-M <mark>IH<mark>NLME</mark>GVVANF<mark>L</mark>I</mark>							
K4L1G7_9FIRM/2-86	YE <mark>L</mark> K <mark>NF</mark> S E VIVKKT <mark>L</mark> E							
F0T2R8_SYNGF/2-86	CE <mark>ikn</mark> lt e llvrka <mark>l</mark> i							
J7IRW7_DESMD/2-86	LE <mark>L</mark> K N HT E IVVQQA <mark>L</mark> N							
W0EB03_9FIRM/2-86	YE <mark>L</mark> K N HM E S <mark>VA</mark> QQA <mark>L</mark> F							
Q24W94_DESHY/2-86	YE <mark>L</mark> K N HT E D <mark>VV</mark> QHV <mark>L</mark> F							
L0F6Z2_DESDL/2-86	YE <mark>L</mark> K N HT E SVVQHVL							
I4DC42_DESAJ/7-91	HQ <mark>V</mark> K N YSEILVNQAL							
Q3ACZ4_CARHZ/2-85	PV <mark>LRNYTEEVV</mark> DEI <mark>L</mark> (-MLKNYVEVAVAEVFI							
B1I3R7_DESAP/1-88 F6B5A0 DESCC/1-88	-MIRNYIELTVEETVI							
A4J5H1 DESRM/1-88	-MIKNYIELAVEDML							
F6DKP9 DESRL/1-88	-MIRNYIELAVEETLI							
D5XEK7 THEPJ/20-104	KKLN NYTET VVEQVL							
B2A554 NATTJ/4-88	LK <mark>LQNYMEKVVEKKL</mark>							
Q2RMC9 MOOTA/5-89	LFLKNYMEDCVWELLI							
R7INC8 9FIRM/129-212	PDYKNVMEELVIEQII							
Q0AV46 SYNWW/17-101	MK <mark>VINIVEELV</mark> WECF							
D7CL50 SYNLT/1-85	ME <mark>V</mark> K NVVESLVWNNL							
K4LEW8 THEPS/2-86	LR <mark>L</mark> V N<mark>VMEILV</mark>EET<mark>I</mark>E	EDI <mark>M</mark> KSHP	DF <mark>C</mark> TCERCRYD <mark>y</mark>	VAAIA <mark>LNK</mark> L <mark>PI</mark>	S <mark>Y</mark> VVTME <mark>G</mark> EA	<mark>I</mark> LRANSLKQ-	-QVRVD <mark>L</mark> IRA <mark>v</mark>	<mark>/</mark> TEA <mark>M</mark> II <mark>V</mark> GKN P
A5D5P8 PELTS/19-103	YR <mark>L</mark> V N<mark>VME</mark>ILV KET <mark>I</mark> E	EEI <mark>L</mark> RSAK	NV <mark>C</mark> GCERCLLD <mark>:</mark>	IAAIA <mark>LNK</mark> L P I	L <mark>y</mark> vvtee <mark>g</mark> ev	<mark>L</mark> LRTTSLRQ-	-QAKVD <mark>I</mark> IRA <mark>v</mark>	<mark>/</mark> TEA <mark>I</mark> DI <mark>V</mark> GQK P
R5T5Q3_9CLOT/2-85	QT <mark>L</mark> K nlmedtv lek <mark>i</mark> i							
R6D9T8_9CLOT/1-78	<mark>Me</mark> etvlhk <mark>i</mark> i							
A9KP56_LACP7/3-86	KI <mark>L</mark> H <mark>N</mark> LMEDLVIKQ <mark>I</mark> I	DHI <mark>V</mark> DSFN	C <mark>C</mark> K <mark>C</mark> EQ <mark>C</mark> RLD	IASYA <mark>LNR</mark> L P A	AK <mark>y</mark> vvttq <mark>g</mark> el	ITRLDSLDL-	-QFETN <mark>V</mark> TTA <mark>:</mark>	TQG <mark>I</mark> IL <mark>V</mark> GNN P
Negativicutes								
C9LS53_SELS3/1-85	MP <mark>IKnTMEEFV</mark> QQN <mark>I</mark> I							
F7NPR6_9FIRM/1-85	MQ <mark>LRNYMEDLVWQRL</mark>							
A0A517DSN4/1-85	MQ <mark>IRNFMEDLV</mark> WQRLI							
A0A075KC76/1-85	ME <mark>L</mark> KN <mark>YME</mark> KLVMEK <mark>L</mark> I MEIKNYMEILVMEKLI	SIV <mark>I</mark> QANP	TMCTCQRCQYD	LAALALNALP.	PR <mark>Y</mark> VATSTGET	XIKISSLDQ-	-QFHVD <mark>V</mark> VSA.	TTQAIKI <mark>V</mark> KKQ P
I8TZQ0_9FIRM/1-58	WE <mark>IKNAWE</mark> IPAWEK <mark>P</mark> I	OIV <mark>I</mark> KANR	TTCNCKRCRYD	LAALA <mark>L</mark> NSL P '.	rr <mark>y</mark> vatss <mark>g</mark> qh	LLK*DSLDQ-	-QFHVD <mark>I</mark> VSV <mark>.</mark>	ITQA <mark>i</mark> ki <mark>v</mark> kks p
Tissierellia KOB3T4 GOTA9/1-84	-MLKNYMEDFVDNTL	אעשים <mark>ד</mark> וזמה	DICKCERCIMD	TENTO T DE	T A EXWED COM	VCKI MOI EA	OEKED <mark>T</mark> /WE	TOULT CECE
M1YRU4 9FIRM/1-83	-MVKNYMEVLVDEIYE							
M1ZKTO 9FIRM/1-86	MELRNLMEDEVIYAIN	IDI I KUKK	DICACDRCKID	TAATATAMITKI	ALMOINGEI	ACKADAI DA-	-OEDVDITKE.	TANKA WILLIAM TO THE
R1AY00 9FIRM/1-86	MELHNFMEKEVLNTI	ORI. <mark>I.</mark> KDRP	DICHCEKCRMD	TAATA T.NN T.K	KYWTEKCEA	YTKVSELEV-	-OFENDITRE	TKATEKVSKE P
R1CDA7 9FIRM/1-84	-MTKNYMEILVKNAL							
							~	
Deferribacterota								
D3PEJ3 DEFDS/9-89	NY <mark>FVnlneor<mark>v</mark>val<mark>v</mark>f</mark>	ZDZ <mark>I</mark> GAND	OVOCO A COLUMN	7 T T 7 M CO T TO T	TINT TURNOR	7 E377 VOI	TODORTIDU	TAINI <mark>T</mark> E T <mark>V</mark> AZZE S
E4THP4 CALNY/10-90	EOLKNVNEKRVWDIL							
D4H3C0 DENA2/10-90	DMLKNINEKRAWEIIS							
D3PEZ0 DEFDS/10-89	EL <mark>LQNV</mark> NEKR <mark>V</mark> WNL <mark>I</mark>]							
D3P9N1 DEFDS/10-89	DKIKNINEKRVWELL	NF <mark>Y</mark> EEHP	EFCTCRDCVLD	VAAIT L NTI P I	H <mark>Y</mark> OVSDDI	STAIKK	-ISDEEILKV	LEA <mark>A</mark> IR <mark>V</mark> SKY P
	- <u>-</u> <u>-</u> -	- <mark>-</mark>		<u></u>	_		-	
Pseudomonadota (Pr	otoobactoria\							
•	oteonacteria)							
Betaproteobacteria Q21X05 RHOFT/6-88	EQ <mark>V</mark> H NYYE RLVFEA <mark>V</mark>	/D/G// IID	СЕПОЛИТ Л	77 C777 T NTD T -	D V (/DUD//DIA)	יימשטשים דע <mark>ס</mark>	_ 7 TEAC <mark>M</mark> 7 EAT	. T & Y & Y & Y & Y & Y & Y & Y & Y
IOHU96 RUBGI/6-90	TSIHNHNENAVFEAVI							
B1Y1I4 LEPCP/6-88	SSVHNHYERPVFDAVA							
A2SG25 METPP/6-88	AGIRNYQERAVFDAV							
Gammaproteobacteria			TIODDIHI D		V	*^	- I I I I I I I I I I I I I I I I I I I	<mark>-</mark> <u>*</u> 2.11/1
G0A0I1 METMM/2-85	FD <mark>ICNYYEQLV</mark> TDQ <mark>L</mark> V	VOLKEOSE	ETLPPSFIM <mark>D</mark>	Va <mark>claln</mark> st. <mark>d</mark> n	NC <mark>Y</mark> VRNLVDKG	AGMTEKEHO-	-DMRDA <mark>a</mark> mka <mark>-</mark>	EOA <mark>I</mark> LV <mark>V</mark> NRH P
D3PPD5 ATTVD/3-86	A STONVYEDI WIEST		EEODADEVAD	TACTATNOT DA	M TOWN HANDIN		_ 7 MD \ E 7 E T 7	TEGATATMODDD

I0HU96_RUBGI/6-90	TS <mark>IHNHNENAV</mark> FEA <mark>V</mark> LAA <mark>A</mark> PAHP	glagdgella <mark>dv</mark> a <mark>c</mark> fa <mark>lnrl</mark> q e r <mark>y</mark> irhavdfs <mark>f</mark> ylserereesarq <mark>l</mark> aea <mark>v</mark> dfa <mark>f</mark> gf <mark>v</mark> qar
B1Y1I4 LEPCP/6-88	SS <mark>V</mark> HNH <mark>YE</mark> RPVFDA <mark>V</mark> ARL <mark>A</mark> ATYP	yldedalp <mark>dv</mark> a <mark>cv</mark> a <mark>lnrlga</mark> r <mark>y</mark> irhtvdla <mark>f</mark> yltekerlsmdqt <mark>i</mark> nea <mark>v</mark> sfa <mark>f</mark> ef <mark>v</mark> qar:
A2SG25 METPP/6-88	AG <mark>I</mark> RN <mark>Y</mark> QER <mark>AV</mark> FDA <mark>V</mark> RTWRDHYP	lisddllp <mark>dv</mark> a <mark>cv</mark> a <mark>lnrlpa</mark> r <mark>y</mark> irhevdlv <mark>f</mark> yqsekerlmaerl <mark>v</mark> asa <mark>v</mark> gfa <mark>f</mark> ef <mark>v</mark> qar:
Gammaproteobacteria		
G0A0I1 METMM/2-85	FD <mark>I</mark> C n<mark>yye</mark>q<mark>lv</mark>tdq<mark>l</mark>wqlkeqse	etlppsflm <mark>dv</mark> a <mark>c</mark> la <mark>l</mark> ns <mark>lp</mark> nc <mark>y</mark> vrnlvdkg <mark>a</mark> gmtekehqdmrda <mark>a</mark> mka <mark>i</mark> eqa <mark>i</mark> lv <mark>v</mark> nrhi
D3RRD5 ALLVD/3-86	AS <mark>IQNYYE</mark> R <mark>LV</mark> LES <mark>I</mark> QDK <mark>L</mark> RDRD	eeqdadfva <mark>dl</mark> a <mark>c</mark> la <mark>l</mark> ng <mark>lpa</mark> r <mark>y</mark> vrhavdlwshlgdadrtamrqe <mark>v</mark> eta <mark>v</mark> esa <mark>l</mark> at <mark>m</mark> qrri
W8KJ90 9GAMM/3-86	GN <mark>IQNYYEHLV</mark> YDR <mark>I</mark> RHV <mark>L</mark> LELG	DALDTTHMD <mark>DL</mark> A <mark>CV</mark> A <mark>LNRLPP</mark> R <mark>Y</mark> VRFSVDLTTRLTDEDWELISSQ <mark>V</mark> NEA <mark>V</mark> DYA <mark>M</mark> RTTGRR
I3Y605 THIV6/3-86	SS <mark>IGNYFE</mark> R <mark>LV</mark> MER <mark>I</mark> HGV <mark>L</mark> GENG	sefdttyie <mark>dl</mark> a <mark>cv</mark> a <mark>l</mark> ny <mark>lpp</mark> r <mark>y</mark> vrhavdlashlsdsdhrnmree <mark>v</mark> sda <mark>v</mark> nfa <mark>i</mark> attqrr(
W0E3F1 MARPU/3-86	SS <mark>IENHYE</mark> R <mark>LV</mark> AAR <mark>I</mark> SEL <mark>L</mark> AERG	etfdagyfd <mark>dl</mark> a <mark>c</mark> va <mark>l</mark> ny <mark>lpp</mark> r <mark>y</mark> vrhsidfashqsdeersrmaee <mark>v</mark> ada <mark>v</mark> afa <mark>f</mark> ettrrri
Q2SK09_HAHCH/5-86	DN <mark>V</mark> Y N<mark>YYE</mark>K<mark>LV</mark>LEE<mark>L</mark>AKQ<mark>M</mark>EGQE	pdedflt <mark>da</mark> v <mark>cv</mark> a <mark>lnhlpa</mark> r <mark>y</mark> frhgvdma <mark>f</mark> ylssdeflkmqee <mark>v</mark> vga <mark>v</mark> aka <mark>v</mark> ty <mark>v</mark> kegi
B3PKD8 CELJU/14-96	DF <mark>I</mark> H <mark>NFYE</mark> R <mark>LV</mark> VQE <mark>A</mark> FDQSPRIQ	QGDRDFLA <mark>DV</mark> A <mark>CV</mark> A <mark>LNRLPP</mark> R <mark>Y</mark> IRHDVDMTFFMSPQDMMEIENK <mark>V</mark> ATA <mark>V</mark> GDA <mark>L</mark> HY <mark>V</mark> EARI
C5BLT2 TERTT/17-98	DS <mark>V</mark> H NYYES<mark>LV</mark>IAQ<mark>L</mark>LRASDRAN	edaefma <mark>dv</mark> t <mark>cv</mark> a <mark>lnrlpp</mark> r <mark>y</mark> vrhdvdmtfflspaemeemetk <mark>v</mark> aha <mark>v</mark> nha <mark>m</mark> ay <mark>v</mark> etri
Q21KP7 SACD2/50-131	ES <mark>I</mark> H N<mark>YYE</mark>A<mark>MV</mark>FEQ<mark>L</mark>VRSSDRAA	adpefma <mark>da</mark> a <mark>cv</mark> a <mark>lnrlpp</mark> r <mark>y</mark> vrhdvdmtfflsptemeemlik <mark>v</mark> aha <mark>v</mark> nda <mark>v</mark> sy <mark>v</mark> eeri
W0DQ54 9GAMM/4-83	DA <mark>I</mark> QN <mark>YYE</mark> K <mark>LV</mark> FDE <mark>I</mark> VQH <mark>Y</mark> WTHE	LSEDELE <mark>DIACIA<mark>L</mark>NT<mark>MSP</mark>R<mark>Y</mark>IRHHVDLC<mark>F</mark>FMSQEERNRMVDE<mark>V</mark>KIA<mark>V</mark>AFA<mark>Y</mark>KK<mark>V</mark>FS-</mark>
W0DV99_9GAMM/4-85	DT <mark>I</mark> H <mark>NFYE</mark> RH <mark>V</mark> INE <mark>I</mark> TDN <mark>Y</mark> LKSG	lneeqla <mark>dm</mark> g <mark>ci</mark> a <mark>l</mark> nli <mark>pp</mark> r <mark>y</mark> irhdidms <mark>f</mark> ymtpdeydeiqmk <mark>v</mark> kla <mark>v</mark> qka <mark>f</mark> kr <mark>v</mark> qesi
031IJ0 HYDCU/5-86	OSVHNYYERPVFNHIOENYLNVG	LTENOLA <mark>DM</mark> ACIA <mark>LNRIAPRY</mark> IRHDIDMSFYMSSEEYOEIOTR <mark>V</mark> EKA <mark>V</mark> KKA <mark>F</mark> KK <mark>V</mark> KKLI

```
050WT6 IDILO/5-86
                                             DD<mark>Ihnyyerlv</mark>adr<mark>i</mark>eelelekq------hSqefladlsclvlnqlppryirhevdmafflppskrl--dmemq<mark>v</mark>hka<mark>i</mark>aea<mark>l</mark>eflkgrk
G4QKV4 GLANF/5-86
                                              DD<mark>I</mark>HN<mark>YYEHLV</mark>LEK<mark>I</mark>EALGLNQN------KTADYLA<mark>DL</mark>T<mark>C</mark>LV<mark>L</mark>NQLPPRY</mark>IRFEVDMA<mark>F</mark>YLPQSERR--QMEMN<mark>V</mark>SHS<mark>V</mark>NNACKY<mark>L</mark>DAHP
E1SW73_FERBD/3-84
A1S4Z5_SHEAM/3-84
                                              LD<mark>irnynetlv</mark>rqv<mark>l</mark>eqvglhle------lnedqltdltclalnk<mark>lpa</mark>r<mark>y</mark>irypvdlagytsstele--smitd<mark>a</mark>rna<mark>i</mark>dva<mark>l</mark>th<mark>l</mark>klhp
                                             LE<mark>IRNYYEV</mark>LLMEILSDEGLMDE------LPEDYLADLSCITLNQLEVR<mark>Y</mark>IRHLVDTYEFEDYQELK--QMRNE<mark>I</mark>QTA<mark>L</mark>EKSRAF<mark>L</mark>KKRN
Q07ZT9 SHEFN/3-84
                                             LE<mark>IRNYYEVLL</mark>MEM<mark>L</mark>RDEGLMEE------LPEEYLADLCCVTLNQLPVR<mark>Y</mark>IRHLVDTY<mark>F</mark>FENYQELK--QMKTE<mark>I</mark>SEA<mark>L</mark>ERSRQF<mark>L</mark>KANL
Q5E0J4_ALIF1/3-84
                                              SN<mark>I</mark>HN<mark>YMEVLV</mark>NKR<mark>F</mark>FALEFHQI------YTLDQIA<mark>DM</mark>K<mark>CIA<mark>L</mark>NQL<mark>P</mark>TL<mark>Y</mark>IRYSLDML<mark>A</mark>ATSQKKLM--QYDEM<mark>V</mark>AAA<mark>V</mark>ENAEKM<mark>I</mark>VNDR</mark>
Q87HN9_VIBPA/5-86
U4KIM0_9VIBR/5-86
                                              SD<mark>VHNYMETLV</mark>GQV<mark>L</mark>SSDQYTSQ------FDNEQLA<mark>DI</mark>A<mark>CL</mark>ALIQLK<mark>E</mark>VYIRHDIDFLSALPERKLT--QFKQS<mark>V</mark>DVA<mark>V</mark>HNA<mark>A</mark>AM<mark>I</mark>LEDR
                                             VD<mark>V</mark>HN<mark>YMETLV</mark>GNR<mark>L</mark>GDPDYADA------YDSEQLADLA<mark>C</mark>IALNQLRFIYIRHDIDFLSALPEERLV--VLRKQ<mark>V</mark>NDA<mark>L</mark>QASESM<mark>I</mark>KDDR
                                             ED<mark>vhnymetlv</mark>gqv<mark>l</mark>gqpeyset------ydqdqla<mark>dl</mark>acla<mark>l</mark>cq<mark>lre</mark>i<mark>y</mark>irhdidflsalpeaklv--ilkdh<mark>a</mark>lia<mark>v</mark>qaaesm<mark>i</mark>vndr
Q9KM28 VIBCH/5-86
Q3IIU2_PSET1/5-86
K6Z2N5_9ALTE/5-86
                                             ED<mark>IHNYYEKFV</mark>LEE<u>I</u>INRKLNEQ------YTDDVMA<mark>DFCCT</mark>V<mark>LNQL<mark>PA</mark>R<mark>Y</mark>IRSDVDMA<mark>Y</mark>YLSQTERM--HMQQR<mark>V</mark>QTA<mark>I</mark>DVA<mark>I</mark>SQ<mark>I</mark>AKKK</mark>
                                             DD<mark>I</mark>HN<mark>YYEHLV</mark>LDR<mark>I</mark>ETLGLNKS------KGEDYLADLCCLALNQIPPR<mark>Y</mark>IRFEVDMS<mark>F</mark>YLPQSERQ--QMEMN<mark>A</mark>VNA<mark>V</mark>SQA<mark>A</mark>SFLDQNV
A0KPQ2 AERHH/32-113
                                              TD<mark>ihnfyehlv</mark>lke<mark>i</mark>erqglnqk------lsanqma<mark>dlcc</mark>ltlnq<mark>lpshy</mark>irhdvdmiyfltdakrt--emekn<mark>v</mark>ies<mark>l</mark>klaqek<mark>v</mark>qsap
A0A8A7FCX2/5-86
                                              DD<mark>L</mark>HN<mark>YYEKIV</mark>LEN<mark>I</mark>EERGLDHK-------YDEDVMA<mark>DLCC</mark>TA<mark>L</mark>NRLPARY<mark>I</mark>RYDVDME<mark>F</mark>YLPTDERA--EMEGR<mark>V</mark>KEA<mark>V</mark>DYA<mark>L</mark>AQ<mark>I</mark>AKKK
BB2AT2 0549/5-86
                                              DD<mark>ihnfyekiv</mark>vee<mark>i</mark>ikrkldkv------ysedvmad<mark>fcc</mark>tv<mark>l</mark>nql<mark>pp</mark>r<mark>yi</mark>rydvdma<mark>f</mark>ylpqseri—hmeer<mark>v</mark>qta<mark>i</mark>dva<mark>i</mark>sq<mark>i</mark>skkk
Spirochaetota
D5U3J8 BRAM5/1-80
                                             MY<mark>LVNLMEQKV</mark>SKL<mark>A</mark>DSY<mark>L</mark>KEKN-----IPVNTNMRMD<mark>I</mark>IA<mark>YTLNRVQEQY</mark>VTSAR<mark>G</mark>VLHSYNRDPI-----QSNTEI<mark>L</mark>SII<mark>A</mark>DA<mark>C</mark>DIVS
V5WHS4 9SPIO/10-93
                                             ed<mark>lyneaeelv</mark>fqelerqlelag----edvpkdqdsvv<mark>dm</mark>aayalnlvkemyranllervyapalsqe----hkde<mark>i</mark>asa<mark>v</mark>nqa<mark>v</mark>akisenp
H9UJ27 SPIAZ/1-91
                                             MK<mark>vrnIledev</mark>frv<mark>i</mark>kei<mark>a</mark>dedq-csssphfattdq<mark>c</mark>rv<mark>daacy</mark>vlnrlppr<mark>y</mark>vssgr<mark>c</mark>qa<mark>y</mark>adtefsl-npqlead<mark>i</mark>lsl<mark>v</mark>heglrr<mark>v</mark>ttir
H9UKKO_SPIAZ/10-92
V5WIR9_9SPIO/1-91
                                             ed<mark>lvneserfv</mark>lee<u>leqql</u>etaa-----da<mark>c</mark>rtte<mark>c</mark>vldmaayalnkvppr<mark>y</mark>rvnllerlessvpdad----yqta<mark>v</mark>qqa<mark>v</mark>sealri<mark>v</mark>sknp
                                             ME<mark>IHNLMEDKV</mark>LNILNEI<mark>C</mark>DDEE-QSSEYSYCTTPQCRMDAACFVLNR<mark>IP</mark>QR<mark>Y</mark>VSSGR<mark>G</mark>FAHIESDIQD-DPQIQID<mark>I</mark>VTL<mark>A</mark>HEGLRR<mark>V</mark>TTIQ
GOGAJ9 SPITZ/1-91
                                             MK<mark>vhnlmediv</mark>trv<mark>v</mark>ddl<mark>c</mark>keka-htepdky<mark>c</mark>ltee<mark>c</mark>rq<mark>divcyvlnrvppr<mark>y</mark>vssar<mark>c</mark>va<mark>y</mark>tsvhyee-dpqllid<mark>i</mark>ltl<mark>a</mark>teg<mark>l</mark>kr<mark>i</mark>tsir</mark>
                                             MK<mark>vhnlmediv</mark>trv<mark>v</mark>ddlckeka-htepgky<mark>c</mark>lteecrqd<mark>ivcyvlnrvpp</mark>ryvssar<mark>g</mark>va<u>y</u>tsvhyee-----dpq<mark>l</mark>lid<mark>i</mark>ltl<mark>a</mark>teglkri
EORR51 SPITD/1-87
F5YB03_TREAZ/10-95
F5YR57 TREPZ/10-95
                                              DL<mark>LINEAEKLV</mark>LAD<mark>L</mark>GRQ<mark>L</mark>EAYP----EAI<mark>CKO</mark>ND<mark>C</mark>VVDMAA<mark>M</mark>ALNS<mark>VKE</mark>L<mark>Y</mark>RVSLL<mark>C</mark>TI<u>Y</u>ASRAMDE--KAYATS<mark>I</mark>REA<mark>V</mark>FKA<mark>IEKV</mark>RKNP
                                             EN<mark>l</mark>ane<mark>aehlv</mark>hde<mark>l</mark>grq<mark>l</mark>esfq----gei<mark>clc</mark>nd<mark>c</mark>vv<mark>dm</mark>aa<mark>mal</mark>ntvkpl<mark>y</mark>rysll<mark>g</mark>tlwassamsd--gayaes<mark>v</mark>rea<mark>v</mark>sna<mark>i</mark>ek<mark>v</mark>rknp
                                             EY<mark>L</mark>ANE<mark>AERLV</mark>LEK<mark>L</mark>EYY<mark>L</mark>SSEE----FSSV<mark>C</mark>KCQDCVLDMAAYALNNVRPLYRVSLMCTLYAHNVDDT---DYSRD<mark>V</mark>EKS<mark>V</mark>LEAIRRISANP
E1R291 SEDSS/10-95
                                              DQ<mark>l</mark>vne<mark>aerlv</mark>ide<mark>l</mark>ekq<mark>v</mark>lsyp----eplc</mark>lcedcvld<mark>m</mark>atfa<mark>l</mark>nsvkply</mark>rvsllcsm<mark>y</mark>aahamde--asyaes<mark>v</mark>kna<mark>v</mark>asa<mark>i</mark>ak<mark>v</mark>kah<b>p
F8F3J2_TRECH/10-95
F5YJZ6_TREPZ/1-92
                                             ME<mark>IHNTTEDIV</mark>FTT<mark>I</mark>DEI<mark>C</mark>ASIESKGNPDKL<mark>CLC</mark>ERCRTD<mark>V</mark>ACFVLNRVPPY<mark>Y</mark>IISNRSAARIEQETIK-KQQSAAD<mark>T</mark>ATL<mark>V</mark>FEA<mark>F</mark>KR<mark>V</mark>SHNQ
                                             MQ<mark>vfnvmeelv</mark>lem<mark>v</mark>hei<mark>f</mark>eekr--gkgfplvd<mark>c</mark>eq<mark>c</mark>rl<mark>dv</mark>acyv<mark>lnriepey</mark>lisgr<mark>c</mark>vvhvqnelrq-nfqkrad<mark>i</mark>vsl<mark>v</mark>neg<mark>i</mark>ri<mark>i</mark>tqnq
E1R1U3 SEDSS/1-90
                                             MT<mark>V</mark>HN<mark>IMEDRV</mark>AER<mark>V</mark>DAL<mark>Y</mark>EAIQKEHPGFLS<mark>C</mark>DCESCRADTLCYVLNRIPAK<mark>Y</mark>IVSDRCVTYSSLAGST---QLKAD<mark>I</mark>DAL<mark>V</mark>IEGMKI<mark>V</mark>NTAK
F4LKY9_TREBD/1-90
S6A514_9SPIO/-6-87
Q73MV1_TREDE/1-93
                                              MI<mark>I</mark>HNVMED<mark>IV</mark>FNE<mark>V</mark>NKMFDEAEKNNEKWLTCSCMQCRLDTICYVLNRVKPRYIKSGR<mark>G</mark>LAHFLKFGGSEKNQIMAD<mark>I</mark>TAL<mark>V</mark>IEG<mark>M</mark>HR<mark>V</mark>LSTR
                                             MI<mark>IHNVMEDLV</mark>YTEVNKLFDEAEEKKESWLT<mark>CSC</mark>MQCRVDTMCYVLNRVKERYIKSGRCLAHFLKFEKTEKVQIMAD<mark>I</mark>TSL<mark>V</mark>IEGMHR<mark>V</mark>LSTK
F8EZ03 TRECH/1-92
                                             MD<mark>IHNTIEDRV</mark>LNV<mark>V</mark>HEIFDSIEKAGRPDKP<mark>CTC</mark>YQCRLDTACYVLNRLPPRYVISSRCVARSESETLE-KQQEDAD<mark>I</mark>VSL<mark>V</mark>YEGLNK<mark>I</mark>AKSM
F5YAE5 TREAZ/1-92
                                             MELHNTMEDKILAKVEDIFNTISKDGNPDNFCTCSQCRMDTACYVLNRTVERYIVSNRGAARVQQETID-QQQKDADIAALIYEGLKRVNHNQ
Thermodesulfobacteriota
                                              YH<mark>IRNRNEARV</mark>LRMMEKV<mark>L</mark>TEPP-----GMTPSAADLH<mark>DI</mark>YA<mark>L</mark>ALNSLPAR<mark>Y</mark>A--QQ<mark>E</mark>SI<mark>V</mark>LRDPVRD-----ED<mark>V</mark>LEA<mark>V</mark>RKA<mark>L</mark>AT<mark>V</mark>VHNP
Q313N7_DESAG/17-95
Q728W3_DESVH/172-250
                                              SR<mark>TRNRNEIRV</mark>VKA<mark>I</mark>RTV<mark>L</mark>GEPP-----VYTPDPKDIQ<mark>DI</mark>YA<mark>LAL</mark>NAL<mark>PP</mark>R<mark>Y</mark>A--QH<mark>C</mark>TIVLRDPVRD------Q<mark>I</mark>LEA<mark>V</mark>RDAFIR<mark>V</mark>MEHP
F0JBV6 9DELT/14-93
                                              SK<mark>I</mark>VNRNERR<mark>V</mark>AEL<mark>V</mark>PQI<mark>I</mark>AEYY----EDYIFEDLDIQ<mark>DI</mark>YALALNLL<mark>PA</mark>G<mark>Y</mark>A--QA<mark>S</mark>SI<mark>V</mark>LSDRISD-----YE<mark>I</mark>RTQ<mark>I</mark>RNA<mark>V</mark>ER<mark>V</mark>LDNP
M1WPD9 PSEP2/14-93
                                              GM<mark>I</mark>KNRNESR<mark>V</mark>AELLPQLMEEHY-----VDFIFDPLDIQD<mark>I</mark>YALSLNLIPAHYA--QK<mark>G</mark>SI<mark>V</mark>LSDRLSD-----FE<mark>I</mark>RSR<mark>I</mark>RDA<mark>I</mark>ER<mark>V</mark>LDNP
E6VUX7_PSEA9/14-93
C8X0L1_DESRD/18-96
                                              NK<mark>irnrnekrv</mark>akf<mark>m</mark>vei<mark>l</mark>dqyy-----edylfeqldle<mark>di</mark>ya<mark>lti</mark>nll<mark>pa</mark>r<mark>y</mark>v--qr<mark>e</mark>si<mark>i</mark>lsdrlsd-----fv<mark>i</mark>ksk<mark>i</mark>rea<mark>t</mark>erw</mark>lenp
                                              TG<mark>I</mark>VNRNERR<mark>V</mark>REQ<mark>L</mark>PRI<mark>L</mark>SEYE-----EFPLDRLAVQDVYA<mark>L</mark>TLNS<mark>LSP</mark>R<mark>Y</mark>T--QAFSI<mark>V</mark>LKEPIQD-----AE<mark>I</mark>EEA<mark>L</mark>RAA<mark>I</mark>DK<mark>V</mark>IASP
C6BRV9 DESAD/11-98
                                              DE<mark>I</mark>VNLTEEIVYNE<mark>L</mark>QAL<mark>I</mark>GRAE----IEF<mark>CQC</mark>DK<mark>C</mark>LF<mark>DI</mark>VCVV<mark>L</mark>NKIPSLYSSSIADRTYPSAEFKAEYDRLKLL<mark>A</mark>AEE<mark>I</mark>PRA<mark>I</mark>EQ<mark>I</mark>KDRL
Thermotogota
H2J3K4 MARPK/6-91
                                             YH<mark>I</mark>TN<mark>VMEHIV</mark>EEITNEMFAMPN----IDMCICDRCRADVIALALNHLHPKY</mark>VVTEKCRLYSELQNYT--FQTRAE<mark>V</mark>LTE<mark>V</mark>LKAMEK<mark>V</mark>KEHP
A0A1M4SN03/6-91
                                             YH<mark>ifnvlehlv</mark>editnem<mark>f</mark>smpn-----vdm<mark>cvc</mark>drcrad<mark>v</mark>iaLalnhlnek<mark>y</mark>vvtekcrifseletyt--fqmrae<mark>v</mark>lte<mark>v</mark>lka<mark>m</mark>ek<mark>v</mark>krkp
                                             LK<mark>vfnlmealv</mark>sev<mark>v</mark>eemflmpn-----ldmctcnrckldtlimvlnkfppk<mark>y</mark>svtlkckvfteletlk--tqyrad<mark>i</mark>lretlnamei<mark>v</mark>knnp
A0A7G1G726/6-91
                                             IQ<mark>lknimediv</mark>edv<mark>y</mark>ndl<mark>l</mark>ktnl-----iki<mark>cnc</mark>pkchad<mark>v</mark>la<mark>ivlnnikeky</mark>vvtek<mark>s</mark>eaiertnelr--dqirvd<mark>v</mark>leq<mark>i</mark>lka<mark>v</mark>ei<mark>v</mark>sknp
A0A7G1G3P1/7-92
A0A4Z0VZR5/14-99
                                             EMLVNIMEKIVDEIYEDILETHE----INFCDCPKCREDVKAIVLNEITEKYVSTDKGLAISKAEMME--VQLRIDILDKIVKAIIKVGNNP
                                              -MIKNIMEDIVGECLTEIIKDTN----IKV<mark>CKC</mark>QKCISDIMALTLNHVKPKYVSTKKEEIYTRVELQN--SQLKVD<mark>L</mark>MDT<mark>I</mark>IKSILI<mark>V</mark>ARNP
UYO99622.1/1-85
PF10719/1-79
                                              -----<mark>meelv</mark>lee<mark>l</mark>eeq<mark>l</mark>eelp-----dyctcercla<mark>dv</mark>aa<mark>yalnrlpp</mark>r<mark>y</mark>vvsee<mark>s</mark>ea<mark>y</mark>qkleser--qqlrae<mark>i</mark>lsa<mark>v</mark>rea<mark>i</mark>er<mark>v</mark>kknp
Non-bacterial sequences
Methanocella sp.
                                              VQ<mark>M</mark>VNRMEP<mark>MV</mark>AQA<mark>V</mark>DETLVREK-----kVCSCERCRHDITA<mark>L</mark>ALNMLPPR<mark>Y</mark>VVTPLGEV<mark>V</mark>TNVDLQS--SQWKAD<mark>I</mark>MMA<mark>V</mark>YRA<mark>I</mark>EI<mark>V</mark>RGRP
A0A1V4Z792/4-88
Cand. Methanomethylicus sp.
                                              YVTKNLMEE<mark>LV</mark>LKF<mark>Y</mark>DKVISDLN-----VC<mark>C</mark>KCER<mark>C</mark>KA<mark>DI</mark>IA<mark>LARLPP</mark>R<mark>Y</mark>YVTKECEM<mark>Y</mark>LKLKELE--IQFEVD<mark>I</mark>VAA<mark>L</mark>AAA<mark>A</mark>FI<mark>V</mark>NNKP
A0A7C4KWL5/2-85
Aduncisulcus paluster
```

Figure S1. Sequence alignment of representative members of the ComFB/CdgR family.

GKT31366/8-91

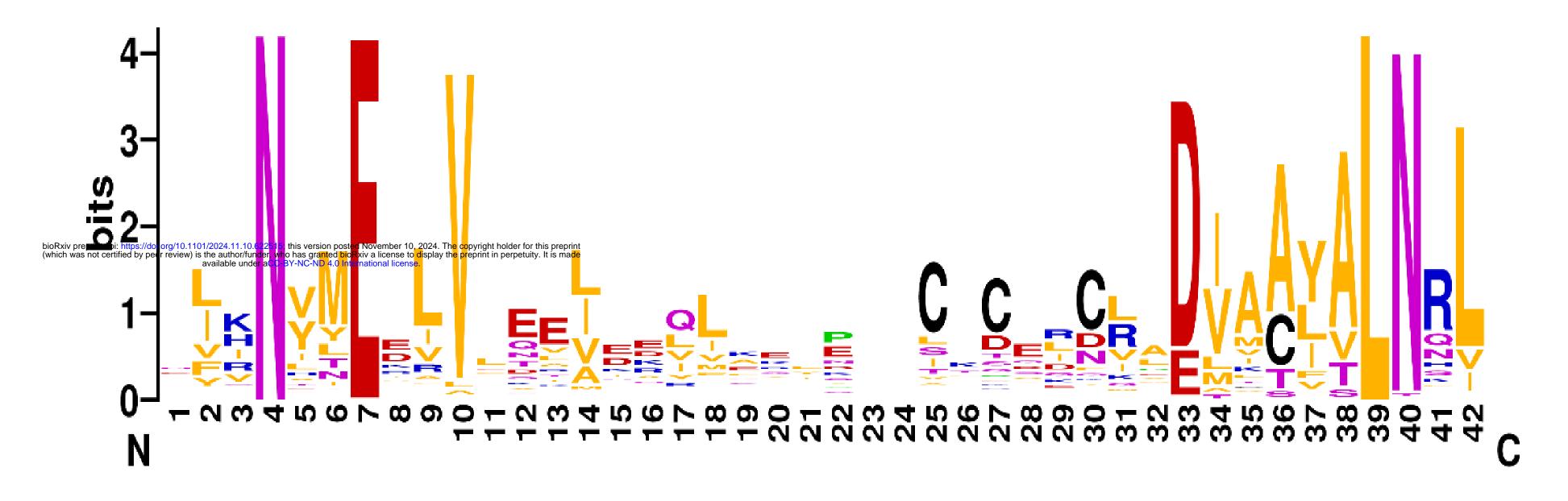
Arabidopsis thaliana A0A178U577/213-296

Proteins are shown under their UniProt identifiers and sorted by taxonomy. Conserved negatively (D, E) and positively (K, R) charged residues are shown in red and blue, respectively; amide residues (N, Q) are in purple. Conserved hydrophobic residues are indicated with yellow shading, conserved turn residues (G, P, S, A) are shaded green. Zinc-binding Cys residues of ComFB and the correspondding Cys residues in other proteins are shown on light blue background. The last sequence in Thermotogota phylum is followed by a consensus sequence from Pfam entry PF10719. The last four sequences represent non-bacterial entries, a potential contamination.

DE<mark>I</mark>VNLS**EE<mark>VV</mark>YDE<mark>L</mark>QALIARAE----IEF<mark>CQC</mark>DK<mark>C</mark>LFD<mark>I</mark>AC<mark>VV<mark>L</mark>NKI<mark>PS</mark>L<mark>Y</mark>SSSIADRT<mark>Y</mark>PSAEFKA----EYDK<mark>L</mark>KVL<mark>A</mark>AKE<mark>I</mark>PR<mark>A</mark>IEQI**</mark>

RPYRNRPIVTDYFREHILGKYQL-----QCDCAKCRLDTIVLTLNQLPPQYTSTQACAAYIKALYMN-PQMQSDVLQALARSVQTVNRLL

Figure S2



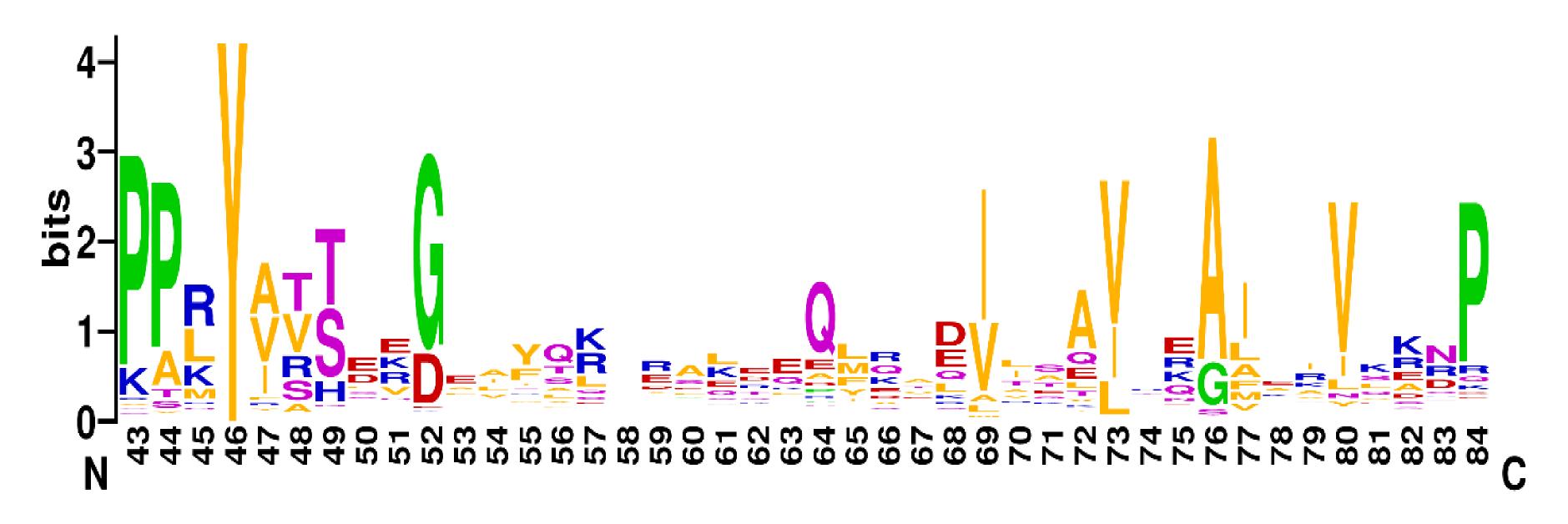


Figure S3

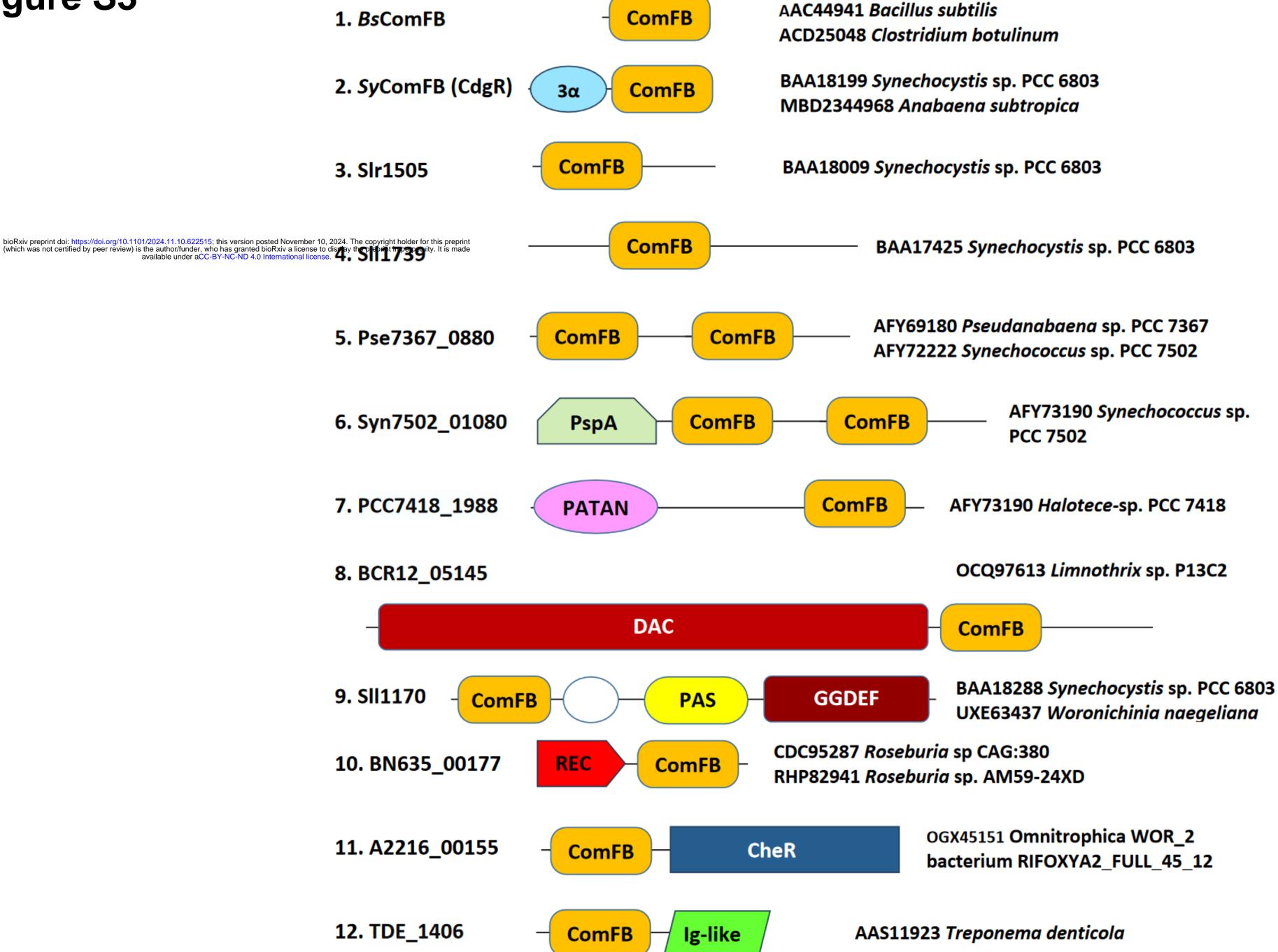
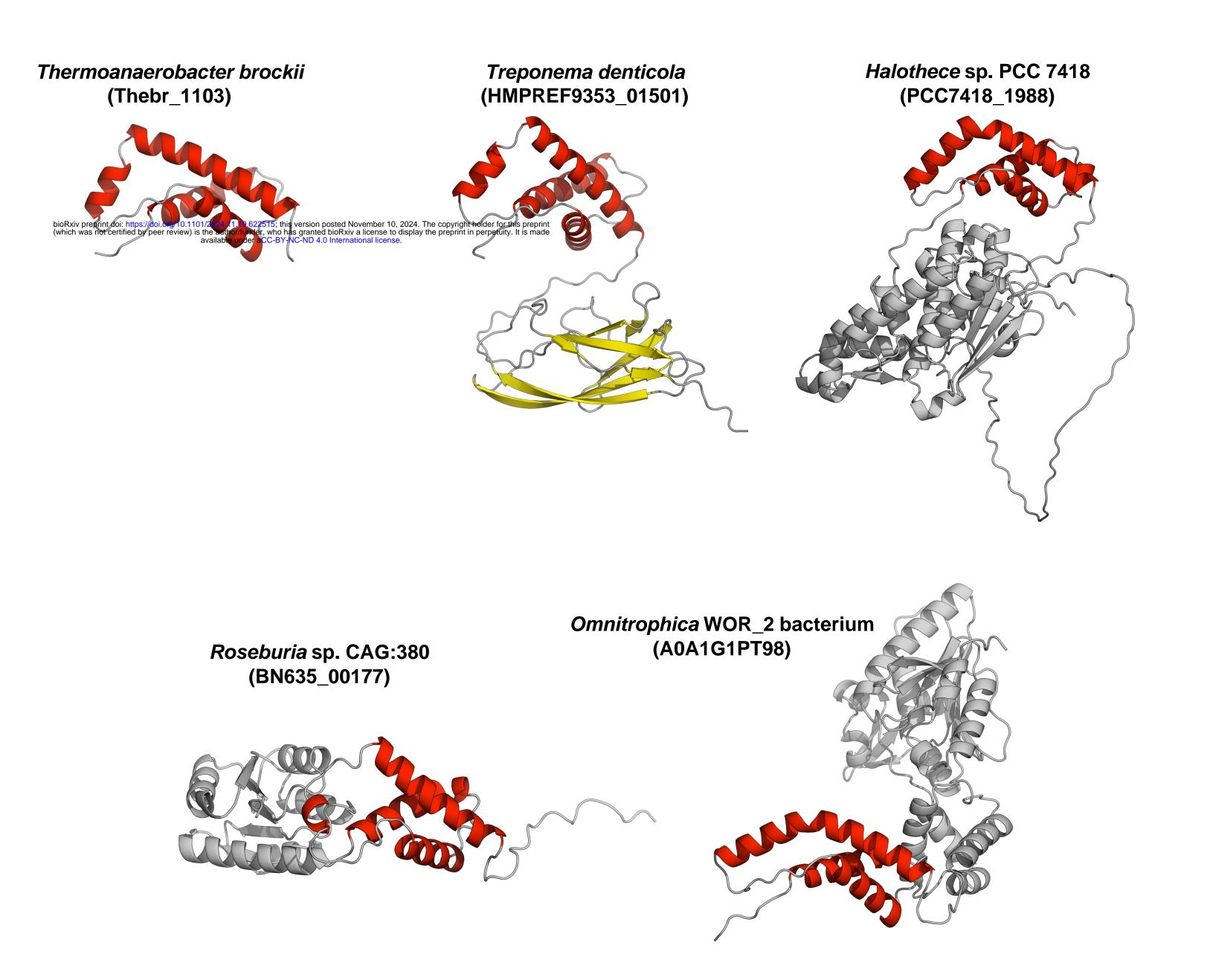
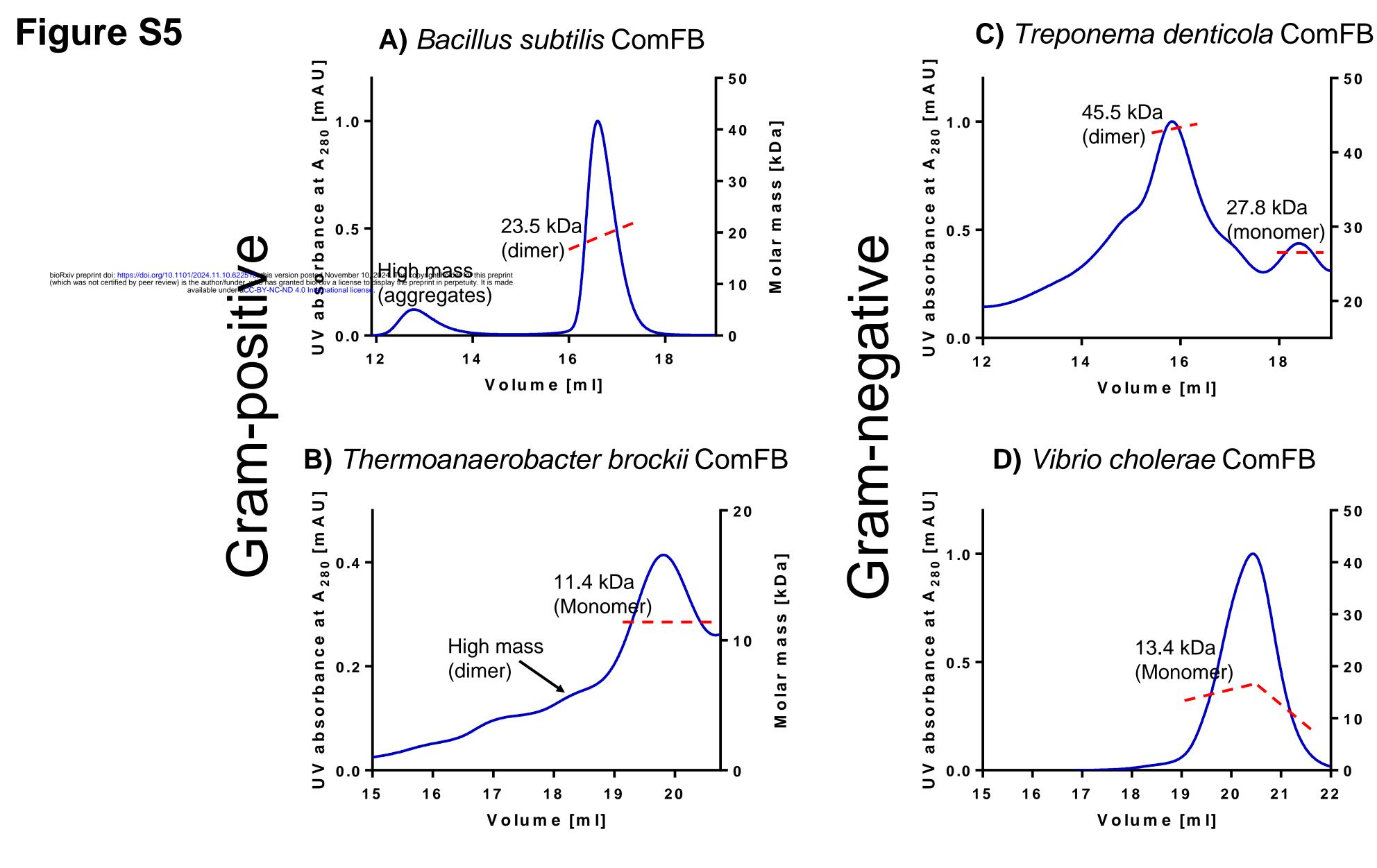


Figure S4





F 50

40

- 20

F 50

- 20

10

22

18

21

Figure S6

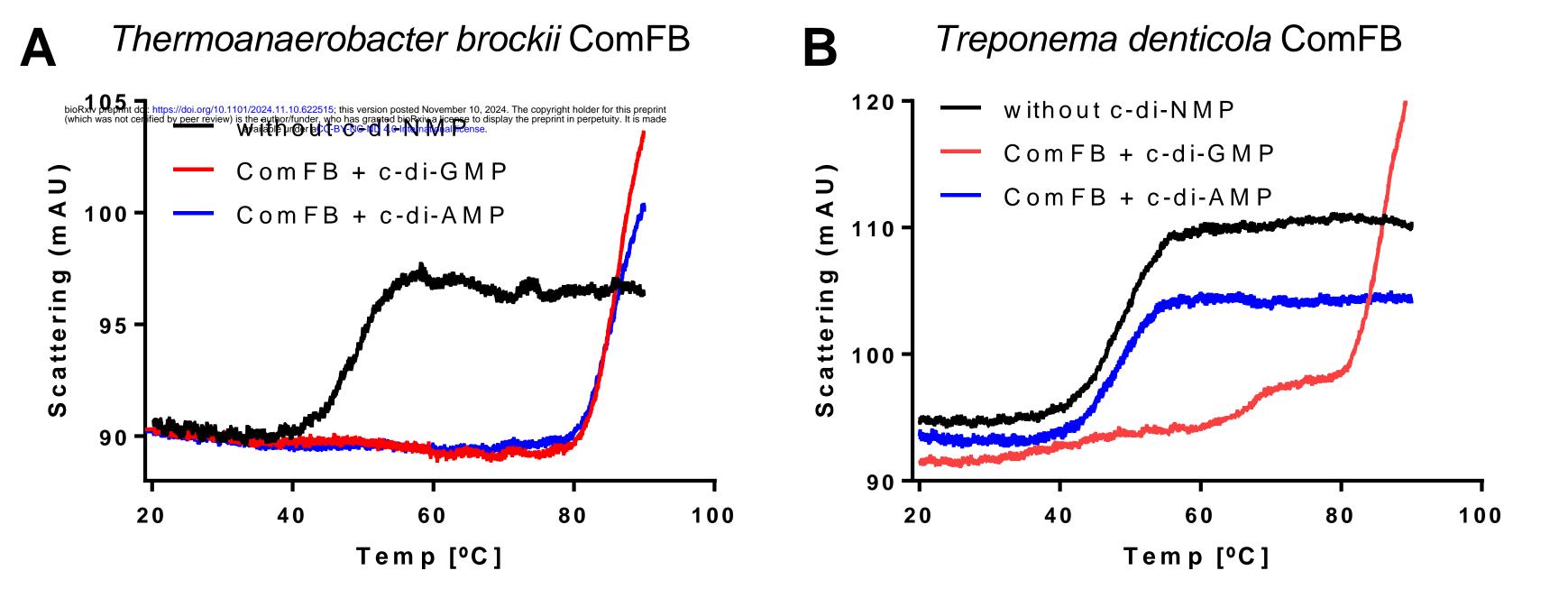


Figure S7

