Archaeal genes code for GGDEF domain proteins with diguanylate cyclase activity

Roman Sidorov[1,*], Li Li[1,*], Ali Dadvar[1], Irfan Ahmad[1], Stefan Spring[2], Jörg Overmann[2,3,#], Ute Römling[1]
[1]Department of Microbiology, Tumor and Cell Biology, Biomedicum, Karolinska Institutet, Stockholm, Sweden
[2]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany
[3]Braunschweig University of Technology, Spielmannstraße 7, 38106 Braunschweig, Germany

*contributed equally

[#]current address: Bavarian State Collections of Natural History; Munich, Germany; and Chair for Molecular Diversity Research, Faculty of Biology Ludwig-Maximilians-University Munich; Munich, Germany

Corresponding author:
Ute Römling, ute.romling@ki.se

ABSTRACT

A biofilm-like tightly surface-associated mode of growth of pre-cellular life protected by an extracellular matrix has been hypothesized to be a prerequisite for the emergence of cellular life and its persistence under challenging environmental conditions. Ubiquitous in Bacteria including members of the deepest branching phyla, cyclic di-GMP, a positive regulator of biofilm formation, has not yet been detected in Archaea. Thus, whether cyclic di-GMP has been present as a primordial biofilm activator in the last universal common ancestor (LUCA) of Bacteria and Archaea remains unanswered. In this work bioinformatic analysis and structural modelling identified GGDEF domain proteins present in distinct archaeal isolates and metagenome associated genomes of confirmed archaeal origin. In particular, phenotypic and *in vivo* assays in combination with catalytic mutants in the heterologous Salmonella system indicated that the complex Rec-PAS/PAC-PocR-GGDEF-HD-GYP domain protein of *Methanocella arvoryzae* MRE50, a member of the *Stenosarchaea* order *Methanocellales* and other archaeal GGDEF domain proteins, possess diguanylate cyclase activity. While cyclic di-GMP signaling is ubiquitous in Bacteria, it can also rapidly disappear in evolution. Thus, it remains to be answered whether cyclic di-GMP signaling has been actively depleted from archaeal isolates with few synthesis genes to be sporadically maintained, whether cyclic di-GMP signaling has been secondarily introduced by horizontal gene transfer into Archaea or whether both scenarios have concomitantly occurred.

1 Introduction

Nucleotide based second messengers are found in all domains of life. For example, cyclic AMP is an ubiquitous second messenger synthesized by Bacteria, Archaea and Eukaryotes (Botsford &Harman, 1992, Leichtling, Rickenberg et al., 1986, McDonough &Rodriguez, 2011, Rall &Sutherland, 1958). Other nucleotide based second messengers have an unequal distribution. (p)ppGpp, synthesized upon monitoring the fundamental process of translation to differentially regulate gene expression is found in bacteria and plants (Potrykus &Cashel, 2008, Takahashi, Kasai et al., 2004). Cyclic di-nucleotides have only recently been discovered (Ross, Weinhouse et al., 1987), although they regulate ubiquitously fundamental processes such as the sessility-motility lifestyle switch, osmohomeostasis and phage defense (Corrigan, Campeotto et al., 2013, Duncan-Lowey, McNamara-Bordewick et al., 2021, Romling, Galperin et al., 2013, Whiteley, Eaglesham et al., 2019). Thus, nucleotide based second messengers are ubiquitous. While some of them such as 2',3' cyclic NMPs and pApppA are produced as byproducts of enzymatic reactions, dedicated enzymes are responsible for the synthesis of others. However, evidence is accumulated that such nucleotide-based signaling molecules can also be synthesized non-enzymatically in mineral-catalyzed chemical reactions indicating that molecules that serve today as central second messengers in cellular physiology have an ancient and most likely prebiotic history on earth (Romling, 2023b, Wu, Lu et al., 2025). Among the nucleotide-based signaling molecules, cyclic di-nucleotide second messengers, cyclic di-AMP and cyclic di-GMP regulate fundamental physiological processes. Cyclic di-AMP regulates osmohomeostasis, in particular potassium homeostasis throughout Bacteria and Archaea and can regulate peptidoglycan biosynthesis in Bacteria. Thereby, the DAC, diadenylate cyclase domain, a short domain of 123 amino acids, originally identified in the context of a cell cycle checkpoint regulator where DAC is covalently linked to a DNA binding domain (Witte, Hartung et al., 2008), has diversified substantially during evolution and is found connected to a broad range of different signaling domains sensing unrelated signals (Corrigan &Grundling, 2013, Romling, 2008).

On the other hand, cyclic di-GMP's predominant physiological, metabolic and ecological impact is the regulation of the fundamental bacterial life styles sessility (biofilm formation) and motility including associated physiological traits such as antimicrobial and stress tolerance (Dayton, 2020, Simm, Morr et al., 2004). The single cell level regulation by cyclic di-GMP leads to a substantial bi- and (most likely) multi-modality of the bacterial population (Grantcharova, Peters et al., 2010, Kulasekara, Kamischke et al., 2013). Thereby, the impact of cyclic di-GMP spans from the regulation of the association of microbes with abiotic surfaces and microbial cell-cell interactions to interkingdom interactions affecting organism development and ecology (Hu, Peng et al., 2024, Peng, Liang et al., 2020, Romling, 2023a, Ruiz, Castro et al., 2012). Although cyclic di-GMP can regulate the cell cycle, dysregulation of cyclic di-GMP signaling has rarely been found to be essential or to impair growth under laboratory conditions (Christen, Christen et al., 2006, Liu, Lee et al., 2020, Lori, Ozaki et al., 2015). The physiological, metabolic and ecological impact of cyclic di-GMP signaling in bacteria is reflected by enzyme domains that mediate the turnover of cyclic di-GMP to be present in all bacterial phyla. Thereby, the GGDEF domain mediates diguanylate cyclase activity, the EAL domain cyclic di-GMP specific phosphodiesterase activity with pGpG as major product output and the HD-GYP domain cyclic di-GMP specific phosphodiesterase with mainly GMP as product output. Besides their broad phylogenetic distribution, cyclic di-GMP turnover domains are frequently abundant in a bacterial genome (Romling et al., 2013, Romling, Gomelsky et al., 2005). This multitude of turnover domains which are conventionally linked to a diversity of (mostly N-terminally located) signaling domains highlights the complex integration of diverse signals spanning from the transcriptional, post-

transcriptional to the post-translational level. Local signaling by protein-protein interactions versus global signaling by highly active diguanylate cyclases further adds to the mechanistical diversity of signal transduction.

Constituting domain superfamilies in their own rights, GGDEF, EAL and HD-GYP domains, although providing a consistent enzymatic activity, have undergone substantial diversifying radiation within one genome and throughout the bacterial kingdom (Galperin &Chou, 2022, Romling, Cao et al., 2023, Schirmer &Jenal, 2009). With N-terminal signal transduction and the scaffold function for protein-protein interactions being mostly determinative for the sequence diversity, the substrate and product specificity are surprisingly conserved (Schirmer, 2016). Only few GGDEF domains with distinct scaffolds have been identified to produce the alternative signaling molecule cyclic GMP-AMP (Hallberg, Chan et al., 2019, Nelson, Sudarsan et al., 2015).

Biofilm formation, as primordial protected microbial cells closely associated with abiotic surfaces, has been hypothesized to be a prerequisite for the origin of life (Martin &Russell, 2003, Romling, 2023b, Trevors, 2011) and is predominant in Bacteria and Archaea (Flemming &Wuertz, 2019). With exclusively sessile cells not being able to survive and spread, regulation of biofilm formation might have arisen almost in parallel in a (last) universal common ancestor (LUCA) of Bacteria and Archaea. Based on bioinformatic analysis, we have previously predicted that cyclic di-GMP signaling and its turnover proteins are present in deeply branching bacteria (Romling, 2023b). In order to gain evidence for the presence of cyclic di-GMP signaling in LUCA and before, we were wondering whether cyclic di-GMP signaling is found (and regulating the motility/sessility lifestyle switch) also in the archaea. A valid argument for primordial occurrence of the cyclic di-GMP signaling molecule is the presence of protein domains in both prokaryotic life domains, we were therefore investigating whether cyclic di-GMP signaling can occur in Archaea. To this end, based on bioinformatic analysis and experimental validation, we show that diverse GGDEF domains with diguanylate cyclase activity occur in Archaea.

Materials and Methods

Strains, plasmid and growth conditions.

The archaeal isolate *Methanocella arvoryzae* MRE50 (DSM22066) (Sakai, Conrad et al., 2010) was grown in DSM medium 1318 under anaerobic conditions at 45°C for seven days. *Haloterrigena salinisoli* KLK7 (KCTC4307T; (Dong, Mao et al., 2024)) obtained from the Korean Collection for Type Cultures (KCTC) has been grown in KCTC medium No. 1762 HCM (Haloarchaeal complex medium) or respective agar plates at 37°C. *H. salinisoli* KLK7 was stored in KCTC medium No. 1762 HCM medium with 50% glycerol at -80°C.
Escherichia coli TOP10 has been used for all cloning events. *E. coli* TOP10 and derivatives were conventionally grown on LB agar plates or in LB medium at 37°C. *Salmonella typhimurium* UMR1 (ATCC14028 Nal^r $rdar_{28°C}$) and *S. typhimurium* UMR1 Δ*yhjH* were maintained on LB medium plates or Congo red agar plates (LB without NaCl medium, 40 µg/ml Congo red, 20 µg/ml Coomassie Brilliant Blue G-250. Antibiotics were added at the following concentrations; Amp (100 µg/ml) and induction of protein production was optimized using variable concentrations of IPTG from 0 to 0.75 mM and of L-arabinose from 0 to 0.1% to optimize concentrations.

Cloning of open reading frames

Synthetic gene templates of original and codon optimized open reading frames including the native or a mutated Shine-Dalgarno sequence were amplified using specific or universal primers. Two different cloning strategies were applied in order to enable protein production using the pBAD28 vector under the control of the L-arabinose inducible pBAD promoter. Initially, conventional cloning after restriction enzyme digestion of the amplified gene fragment and the vector, subsequent in vitro ligation by T4 ligase and transformation into chemocompetent *E. coli* TOP10 (alternatively DH5α and XL-1 Blue) cells was applied. In a different experimental approach, synthetic gene fragments were recombined by *in vivo* cloning using homologous vector sequences (Bubeck, Winkler et al., 1993, Jones &Howard, 1991). In detail, pBAD28 was amplified by PCR using Q5 polymerase and universal primers that included by default the XbaI and SphI restriction sites, respectively, to achieve a linear plasmid. Open reading frames were amplified using primers that contained the flanking pBAD28 nucleotide sequence at the 5' end concluding with XbaI and SphI sites at the 3' end, respectively. Both linear PCR products were transformed into chemocompetent *E. coli* TOP10 cells according to the standard protocol. Cells were subsequently plated onto LB medium containing the appropriate antibiotics to select for assembled plasmids.
The *adrA* open reading frame was recombined with plasmid pSRKGm under the control of the IPTG inducible *lac* promoter. The *adrA* open reading frame was amplified with primers that contained 15 bp vector sequencing with NdeI and NheI sites, respectively. pSRKGm was amplified by PCR using primers that included by default the NdeI and NheI sites that flank the multicloning site and the *lacZ*-α fragment at the 5' and the 3' end of the vector sequences, respectively.

Site directed Mutagenesis and construction of gene deletions

Amino acid substitutions in order to abolish the catalytic activity of GGDEF and HD-GYP domain proteins were introduced with the Q5 Quick change kit (New England Biolabs) and/or the equivalent reaction using the individual components Phusion polymerase to amplify the plasmid template with mutagenizing primers with a subsequently combined reaction of polynucleotide kinase, T4 ligase and DpnI. Alternatively, site-directed mutagenesis has been performed using mutagenized primers overlapping by at least 15 nucleotides containing the mutations to linearize the vector with the gene fragment. Circular nicked and the linear construct were transformed into chemocompetent *E. coli* K-12 cells.

A deletion of a cloned gene was constructed by amplifying the vector containing the wild type gene with primers that define the deletion and overlap by 15 bps. The linearized vector was subsequently transformed into chemocompetent *E. coli* TOP10 cells for in vivo recombination and cells with assembled vectors were selected for on LB plates containing appropriate antibiotics. All constructs were confirmed by Sanger sequencing.

Plasmid transformation

Plasmid constructs were transformed into chemocompetent E. coli TOP10 and into *S. typhimurium* UMR1 and derivatives by electroporation as described (Romling, Sierralta et al., 1998).

Biofilm formation assay

Single colony isolated isolates were streaked on Congo red agar plates or 10 µl of a 5 OD_{600} suspension was spotted on Congo red agar plates. Congo red agar plates were incubated at 28°C for up to 72 h with positive control: *S. typhimurium* UMR1 or UMR1 Δ*yhjH* and negative control: *S. typhimurium* UMR1 Δ*csgBA* Δ*bcsA* containing the respective empty vector. Colony morphology and coloration was documented in a gel documentation chamber using a conventional mobile phone,

Motility assay

Two µl of a bacterial suspension of 5 OD_{600} were inoculated into a motility agar plate (LB medium with 0.25% agar). The plate was incubated at 37°C for up to 24 h. The zone of motility was documented in a gel documentation chamber (Biorad) using a conventional mobile phone and subsequently the diameter of the motility zone was calculated. Negative control was *S. typhimurium* MAE108 (UMR1 Δ*fliC* Δ*fljB*) and positive control *S. typhimurium* UMR1 containing the respective empty vector. Motility plates were documented in a gel documentation chamber using a conventional mobile phone,

Assessment of relative cyclic di-GMP concentrations by the Vc2*lacZY* riboswitch assay

The Vc2 aptamer was used as a cyclic di-GMP sensor to qualitatively assess cyclic di-GMP synthesis and degradation in the context of the riboswitch which subsequently regulates production of the downstream β-galactosidase LacZY protein. The cyclic di-GMP sensing unit integrated in single copy into the Tn7 site of *E. coli* TOP10 inhibits production of the β-galactosidase in response to high cyclic di-GMP concentrations and activates protein production in response to cyclic di-GMP degradation by phosphodiesterases, most sensitively in the background of elevated cyclic di-GMP levels (concomitant expression of a diguanylate cyclase; (Liu, Kim et al., 2018, Liu et al., 2020)). 80 µg/ml X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was the substrate to monitor β-galactosidase activity.

Vectors containing genes and variants cloned into pBAD28 were transformed into chemocompetent *E. coli* TOP10 Tn*7*::Vc2*lacZY* to assess diguanylate cyclase activity or into chemocompetent *E. coli* TOP10 Tn*7*::Vc2*lacZY* pSRKGm-AdrA to assess phosphodiesterase activity. Three µl of a logarithmically grown suspension of OD_{600} =1 were spotted on LB agar plate containing 0 and 0.1% L-arabinose and appropriate antibiotic(s) with the degree of beta-galactosidase activity assessed after 24, 48 and 72 h.. AdrA of *S. typhimurium* UMR1 and its catalytically incompetent variant (Romling, Rohde et al., 2000, Simm et al., 2004) (Khan, Gaines et al., 2008), CdaS of *Bacillus subtilis* 3A1 and catalytically incompetent variant (Mehne, Schroder-Tittmann et al., 2014) and DncV of *E. coli* ECOR31 and catalytically incompetent variant (Li, Cimdins et al., 2019) cloned in pBAD28 served as cyclic di-nucleotide synthesis and specificity controls. YhjH and its catalytic mutants E48A and E136A cloned in pBAD28 (Simm et al., 2004) served as phosphodiesterase activity controls.

Bioinformatic analysis – identification and verification of GGDEF domain proteins

Initially, the GGDEF domain of AdrA (YaiC) of *Salmonella typhimurium* ATCC 14028 (Romling et al., 2000) was used as a query in BLAST (Altschul, Gish et al., 1990) searching the Archaea (ID: 2157) phylogenetic domain of non-redundant sequences to retrieve the proteins containing potential GGDEF domains. Of note, we previously observed that searching with the GGDEF domain of AdrA retrieved all available GGDEF domains (in contrast to single EAL domain proteins). Subsequently, results were screened for GGDEF domain proteins encoded by genomes of archaeal isolates rather than identified in megagenome-assembled genomes (MAGs). As a second step, GGDEF domain proteins encoded by MAGs were verified as derived from archaeal sequences by assessment of the phylogenetic origin of at least two proteins encoded up- and downstream of the GGDEF domain protein gene, respectively. This assessment served as a proxy for the phylogenetic origin of the respective contig. Sufficiently long contigs to reliably assess the archaeal origin of the GGDEF genes were preferred.

Bioinformatic analysis – other

GGDEF domain proteins were used as query in BLAST searching the Archaea (ID: 2157) and Bacteria (ID: 2) phylogenetic domains to obtain homologs (Altschul et al., 1990). Protein domains were identified during BLAST search which provides cumulative results from the Clusters of Orthologous Genes (COG), Conserved Domain Database, Protein Family Models (NF), InterPro/Pfam, TIGR, SMART databases. Predictions of alternative open reading frames was assessed with ORFFinder at the NCBI website. Transmembrane helices were evaluated with TMHMM (Moller, Croning et al., 2001). Identity and similarity of the proteins were calculated with the Sequence Manipulation Suite (Stothard, 2000) after sequence alignment by ClustaX 2.1 and, if required, manual curation. Domain structures were displayed with Illustrator for Biological Sequences IBS2.0 (Xie, Li et al., 2022). Structural models were retrieved from the UniProt database (AlphaFold; (UniProt, 2025)) and/or calculated using Phyre2 (Kelley, Mezulis et al., 2015) and AlphaFold 3 (Jumper, Evans et al., 2021) to identify additional domain folds. Structural homologs were identified by screening databases with DALI (Holm, Laiho et al., 2023) and FoldSeek (van Kempen, Kim et al., 2024). Structural models were displayed and compared (MatchMaker function) in Chimera 1.18 (Pettersen, Goddard et al., 2004). Homologous GGDEF open reading frames in the context of their gene neighborhood were displayed with Easyfig (Sullivan, Petty et al., 2011) using annotated gene bank files comprising +/- 10000 bps up- and downstream of the respective GGDEF gene. Most similar GGDEF domains to archaeal and all GGDEF domains were identified by BLAST search restricting the results to 500 for all non-redundant proteins. Multisequence alignment was performed with ClustalX 2.1 (Higgins &Sharp, 1988) and manually curated using GeneDoc to obtain alignments of only the homologous GGDEF domain sequences common to all proteins. If required to enhance trr visibility, proteins and/or domains with >90% or >70% similarity were removed using Jalview (Waterhouse, Procter et al., 2009). Alignments were displayed with ESPript 3.0 (Robert &Gouet, 2014). Neighborhood joining (to test feasibility) and maximum likelihood based phylogenetic trees using 1000 bootstraps were calculated and displayed in MEGA 7.0 or MEGA 11.0 (Kumar, Stecher et al., 2016) using standard parameters.

2 Results

2.1 Bioinformatic identification of GGDEF domains in archaeal isolates

BLAST search with the GGDEF domain containing diguanylate cyclase AdrA as query yielded an extensive list of potential GGDEF domain proteins encoded by archaea. However, the vast majority of GGDEF genes were found in MAGs with a true archaeal origin to be challenging to proof. Therefore the list was screened for GGDEF proteins encoded by genomes of archaeal isolates maintained as a pure culture. Indeed, we identified WP_012035199.1, WP_339106355.1 and WP_363463662.1 as encoded by archaeal genomes either on the chromosome or on a plasmid in *Methanocella arvoryzae* MRE50 (Sakai et al., 2010), *Haloterrigena salinisoli* KLK7 (Dong et al., 2024) and *Halogeometricum borinquense* E3 (Nagar, Mani et al., 2024), Of note, the three GGDEF domain proteins are encoded by isolates that are members of the Stenosarchaea, with *M. arvoryzae* MRE50 being an isolate belonging to the order *Methanocellales*, while the extreme halophilic *H. salinisoli* KLK7 and *H. borinquense* E3 belong to the Halobacteria (**Figure 1A**). In detail, *H. salinisoli* KLK7 and *H. borinquense* E3 are categorized to the order *Natrialbales* (family *Natrialbaceae*) and *Haloferacales* (family *Haloferacaceae*), respectiviely (**Supplementary Figure 1**).

2.2 Chromosomal context of GGDEF domain proteins.

We assessed whether proteins highly similar to WP_012035199.1, WP_339106355.1, WP_363463662.1 can be found encoded by MAGs on other completely sequenced genomes or contigs. Indeed, two other *Methanocella* sp. whole genome sequences (*Methanocella* sp. MCL-LM and *Methanocella* sp. SMAG_U12379) code for WP_012035199.1, WP_424359454.1 and HMK46096.1, homologous over the entire length of the protein with WP_012035199.1 with amino acid identity/similarity of 82.5/87.0% and 52.2/68.7%, respectively. The high identity of the GGDEF-HD-GYP proteins with the conservation of the domain structure points to vertical transfer from a common ancestor. While WP_424359454.1 is encoded in a similar genetic context, only the immediate genetic context of the gene environment is preserved for HMK36096.1 (**Supplementary Figure 2A**). These findings of highly homologous GGDEF domain proteins encoded by genomes of *Methanocella* spp. derived from distinct geographical regions, isolate *M. avoryzae* MRE50 from an Italian rice field soil, *Methanocella* sp. MCL-LM from the soil of a sandy desert in Israel and *Methanocella* sp. SMAG_U12379 from wetland soil in Colombia indicates that acquisition/maintenance of a chromosomally encoded WP_012035199.1 homolog, with potential diguanylate cyclase-phosphodiesterase activity, has not been an isolated event, but are potentially commonly present in the *Methanocella* genus. Of note, the genomes of these *Methanocella* isolates possess a length between 3.18 to 3.61 Mbps which is the upper size range of archaeal genomes with the largest archaeal known genome being the 5.75 Mb of *Methanosarcina acetivorans* (Kellner, Spang et al., 2018)

The GGDEF domain proteins of WP_339106355.1 of *H. salinisoli* KLK7 and WP_363463662.1 of *H. borinquense* E3 have an amino acid identity/similarity of 69.6/79.3%. A homologous, but truncated protein of 174 amino acids is found in a halite metagenome (Finstad, Probst et al., 2017). Again, the two isolates derived from geographically distant areas. While *H. salinisoli* KLK7 has been isolated from a salt lake in the Qinghai Province, China, *H. borinquense* E3 has been recovered from a solar salterns, a hypersaline sediment at Marakkaman, Tamil Nadu, India. Again, the genome size of the two strains is with 3.8 (excluding four plasmid) and 4.0 Mbps, respectively, at the upper end of archaeal genomes. Despite the high identity/similarity was the genetic context of the two GGDEF domain protein genes entirely different (**Supplementary Figure 2B).** The GGDEF protein of *H. salinisoli* KLK7 is encoded on the 212 kbp plasmid 'unknown 3' and on a 240 kbp chromosomal contig of *H. borinquense* E3.

2.3 Archaeal GGDEF proteins possess a diverse domain structure

Analyses of the domain structure showed that the 810 amino acid long WP_012035199.1 is a multidomain protein with at least a Rec, Pas/Pac, PocR, GGDEF and HD-GYP domain (**Figure 1B**). Rec receiver domains are domains present at the N-terminus of response regulators as part of bacterial two component systems. Rec receiver domains can be phosphorylated at a conserved aspartate at position equivalent 56 in order to allosterically modulate the activity of the downstream effector domain including GGDEF, EAL and HD-GYP domains (Paul, Weiser et al., 2004, Rao, Yang et al., 2008, Wigren, Liang et al., 2014). Diverse classes of PAS (Per-Ant-Sim) domains, short signaling domains that can sense a wide variety of chemical and physical signals (Dayton, 2020), are frequently associated with GGDEF and HD-GYP domains (Galperin &Chou, 2022, Schirmer, 2016). PocR is a PAS-like fold ligand binding domain that has been shown to bind hydrocarbons such as 1,2 propanediol (Anantharaman &Aravind, 2005, Teng, Gong et al., 2024). The signaling domains and the PocR domain with the GGDEF domain are connected by helix-like linkers. In particular, the PocR domain is connected with the GGDEF domain by a 57 amino acid long linker that forms a coiled-coil linker upon dimerization. Coiled-coil linkers have been shown to flexibly modulate their dimerization upon signal perception and thereby are determinative for the regulation of downstream catalytic activity such as GGDEF domain diguanylate cyclase activity which is catalytically active as a dimer (De, Navarro et al., 2009, Gourinchas, Etzl et al., 2017, Schirmer, 2016).
On the other hand, the domain structures of the 262 amino acid long WP_339106355.1 (alternative longer open reading frame used) and 272 amino acid long WP_363463662.1 (alternative longer open reading frame used) are similar with an unconventional GGDEF-(domain of unknown function) structure.

2.4 Conserved signature motifs in archaeal GGDEF domains

In order to predict by bioinformatic means whether the GGDEF domains could possess diguanylate cyclase activity, the GGDEF domains of the three proteins were aligned with previously characterized type I, II and III GGDEF domains including the GGDEF domains of the reference diguanylate cyclases PleD and WspR (De et al., 2009, Paul et al., 2004, Romling et al., 2013). Type I GGDEF domains have been experimentally characterized as diguanylate cyclases, type II GGDEF domains are evolved diguanylate cyclases with bioinformatically ambiguous assignment and type III GGDEF domain have no experimentally demonstrated catalytic activity. The alignment also contains the cyclic GMP-AMP synthesizing GGDEF domains (Hallberg et al., 2019). Detailed comparative assessment of the GGDEF domains of WP_012035199.1, WP_339106355.1 and WP_363463662.1 showed that the GGDEF domain of WP_012035199.1 possesses the GGDEF motif and other signature amino acids essential for catalytic activity indicating that this GGDEF domain is potentially catalytically active (**Figure 1C**). On the other hand, the GGDEF domains of WP_339106355.1 and WP_363463662.1 possess with GQGDEF a semi-conserved GGDEF motif with several other signature amino acids not conserved.

2.5 Alphafold models of archaeal GGDEF domain proteins are coherent with experimentally verified GGDEF domain proteins

Assessment of the 3D homology showed that the AlphaFold structural model of all three GGDEF domains from WP_012035199.1, WP_339106355.1 and WP_363463662.1 proteins showed highest structural homology to experimentally determined crystal structures of catalytically active GGDEF domains; to GcbC from *Pseudomonas fluorescens* (PDB: 5EUH), DgcB from *Caulobacter crescentus* (PDB: 6TTS) and YfiN from *Pseudomonas aeruginosa* (PDB: 4IOB), respectively (**Figure 2A**). Overlap of the structural models with the experimentally determined structures showed high overlap in the secondary structures with unstructured loops deviating most (Figure 2A). The β-hairpin containing the catalytic site motif GGDEF was, however,

coherent between the WP_012035199.1 structural model and the 5EUH GGDEF crystal structure, while the catalytic loop deviated between WP_339106355.1/6TTS and WP_363463662.1/4IOB significantly as indicated already from the sequence deviation of the GQGDEF motif compared to the conventional GGDEF motif (Figure 2B).

2.6 Archaeal GGDEF domain proteins possess homologs in Bacteria

After separation into different lineages beyond LUCA, members of the Archaea and Bacteria domains of life have continuously exchanged genetic material throughout their evolution. Indeed, novel archaeal orders had arisen through extensive horizontal gene transfer from bacteria (Gophna &Altman-Price, 2022). We were therefore wondering about the closest relatives of the GGDEF domains from the archaeal isolates. While proteins with the highest similarity over the entire length of the protein to the *M. arvoryzae* MRE50 Rec-PAS/PAC-PocR-GGDEF-HD-GYP domain protein are found in other *Methanocella* spp. (see above and Supplementary Figure 2A), we were subsequently assessing whether closely related homologs exist outside of the *Methanocella* genus in other microorganisms in the archaeal and bacterial clade. Indeed, BLAST search with the GGDEF domain of WP_012035199.1 indicated that no other proteins with the identical domain structure were identified, with the next most closely related proteins that have a high identity/similarity to the GGDEF domain of up to 46.8/56.9% are of bacterial origin (Supplementary Figure 3A). Those proteins were found to be of variable length consisting of a GGDEF-HD-GYP or $(PAS)_x$-GGDEF-HD-GYP domain structure occasionally with alternative N-terminal signaling domain, if any. The sensory domains deviated most significantly in their amino acid sequences (data not shown). On the other hand, $(PAS)_x$-GGDEF-HD-GYP domain proteins with a >80% query cover, but with the highest identity/similarity of the GGDEF domain of 39.5/54.6% are among the archaeal homologs (data not shown). In conclusion, the high similarities among the GGDEF domains of archaeal and bacterial origin which are higher than conventionally observed among GGDEF domain proteins encoded by one genome (which display conventionally approximately 35% or lower identity (El Mouali, Kim et al., 2017, Romling et al., 2023)). This high similarity of GGDEF domains can indicate significant horizontal exchange of genetic information between members the two domains of life in certain ecological niches with subsequent recombination events in order to create novel combinatorial cyclic di-GMP turnover proteins.

Equally for WP_339106355.1 and WP_363463662.1 proteins a bacterial protein from a *Methylococcales* bacterium (MCK5829069.1) is with 25.0/36.79% identity/similarity of the GGDEF domain more similar as its archaeal counterpart (MEE8112938.1 from *Nitrososphaerales* archaeon). The domain structure is in both proteins highly distinct (data not shown).

2.7 Phylogenetic context of the archaeal GGDEF domains

In order to assess the broader phylogenetic context of the GGDEF proteins, specific homologs of the GGDEF domains of WP_012035199.1, and WP_339106355.1 with the highest similarity were retrieved after BLAST search from the archaeal clade and from all non-redundant proteins in the NCBI database. The phylogenetic analysis for the GGDEF domains showed that they are embedded in a broader phylogenetic context in archaea and bacteria (Supplementary Figure 3B-E). However, while closely related GGDEF domain proteins are present, the sequence diversity of the 500 most similar GGDEF domains most homologous to WP_012035199.1, and WP_339106355.1 is high (Supplementary Figure 4).

2.8 Archaeal GGDEF domains possess phenotypes in the *Salmonella* rdar biofilm and flagella-mediated motility model

In order to assess whether any of the three GGDEF domain proteins from *M. avoryzae* MRE50, *H. salinisoli* KLK7 and *H. borinquense* E3 possess catalytic activity, we first assessed the effect of production of

the wild type gene products (cloned with the assumption of the longest predicted open reading frame) expressed from the L-arabinose inducible pBAD promoter in plasmid pBAD28 (Guzman, Belin et al., 1995) on biofilm formation and motility in the *S. typhimurium* UMR1 (ATCC 14028 Nal^r $rdar_{28°C}$) model system (El Mouali et al., 2017). It has been previously shown that diguanylate cyclase activity stimulates rdar biofilm formation as assessed by enhanced dye uptake and colony morphology, while inhibiting flagella-based swimming motility (Simm et al., 2004). On the other hand, cyclic di-GMP specific phosphodiesterase activity downregulates biofilm formation while stimulated motility which is most sensitively observed in the *S. typhimurium* UMR1 Δ*yhjH* mutant deleted of the phosphodiesterase dedicated for the degradation of the cyclic di-GMP that restricts motility (Cao L.-Y., 2026, Simm et al., 2004). This assay is not absolutely selective (as e.g. cyclic di-GMP binding proteins are expected to show phenotypic alterations reminiscent of phosphodiesterase activity, while catalytically inactive EAL domains acting through protein-protein interaction can show phenotypic alterations reverse to their catalytic counterparts (Ahmad, Wigren et al., 2013, Simm, Remminghorst et al., 2009)). In combination with mutants in the catalytic site, however, this and other *in vivo* assays provide a better prediction about the catalytic activity of the archaeal GGDEF and GGDEF-HD-GYP domain proteins.

To this end, production of the wild type protein WP_012035199.1 of *M. avoryzae* MRE50 and its C-terminal HD-GYP domain deletion in the *S. typhimurium* UMR1 and UMR1 Δ*yhjH* strain did not alter biofilm formation of the rdar morphotype (Figure 3A). As a hybrid protein like WP_012035199.1 which is predicted to display diguanylate cyclase and phosphodiesterase activity can simultaneous express both activities which might nullify the phenotype, we also assessed the effect of the protein upon abolishment of the phosphodiesterase and diguanylate cyclase activity by introducing the D551A/E552A in the GGDEF motif of the GGDEF domain and D680A in the HD_{680} motif of the HD-GYP domain. Production of these mutant proteins did, however, not alter rdar biofilm formation compared to the *S. typhimurium* UMR1 vector control. In contrast, the motility assay showed upregulated motility upon production of WP_012035199.1 in *S. typhimurium* UMR1 Δ*yhjH* (Figure 3B). A similar degree of stimulation of motility was observed in the GGDEF catalytic mutant. However, the HD-GYP catalytic mutant and the HD-GYP deletion mutant showed vector control motility indicating that the HD-GYP domain is an apparent cyclic di-GMP phosphodiesterase.

2.9 Linker mutations trigger diguanylate activity of the WP_012035199.1 GGDEF domain

As the phenotypic assessment upon expression of the complex GGDEF-HD-GYP domain protein of *M. avoryzae* MRE50 was only informative with respect to the phosphodiesterase activity of the HD-GYP domain in the motility assay, we aimed to design a protein that displays substantial diguanylate activity by the GGDEF domain. While the observed cyclic di-GMP dependent phosphodiesterase activity can be accidental as a minor catalytic activity, diguanylate cyclase activity of the GGDEF domain will ultimately indicate the synthesis of cyclic di-GMP by this archaeal protein. First, we inspected the three signaling domains (Figure 1B). Most obvious, phosphorylation of the N-terminal receiver domain on the conserved aspartate at position 56 could affect the activity of downstream catalytic domains. To this end, the D56A mutant did not affect the behaviour of WP_012035199.1 (data not shown).

The ligands that bind to the PAS/PAS and PocR domain are currently unknown. We noted however that the 57 aa long linker which connects the PocR with the GGDEF domain includes the sequence of an incomplete heptad repeat (Figure 4). This type of linker can dimerize by allosteric transmission of the signal via coiled-coil interactions by switching between a repressive and active register dependent on signal transduction mediated allosteric conformation changes (Baral, Ho et al., 2025, De et al., 2009, Gourinchas et al., 2017). When we created a structural model by AlphaFold 3, however, WP_012035199.1 formed a non-productive dimer with the GGDEF domains located at opposite sites

(Supplementary Figure 4). When the N-terminal REC, PAS and PocR domains were deleted, a protein dimer with GGDEF domains facing each other with their catalytic sites was formed (Figure 4C). Naturally or engineered, this type of linker can exist as a constitutive dimerization module (Baral et al., 2025, De et al., 2009, Gourinchas et al., 2017, O'Shea, Klemm et al., 1991, Rasmussen, Benvegnu et al., 1991). The linker sequence of WP_012035199.1 deviates from the sequence that can provide effective constitutive dimerization leading to constitutively active downstream domains (Figure 4B). Consequently, we therefore first introduced single amino acid substitutions designed to provide tight interactions between linker dimers and assessed their effects on the performance of the enzyme. To this end, we constructed the R454L, the E465V and the S472V substitution. Surprisingly, while the R454L mutant had the most pronounced effect, any of the single amino acid substitutions significantly triggered rdar biofilm formation in *S. typhimurium* UMR1 and substantially downregulated motility in *S. typhimurium* UMR1 Δ*yhjH* indicating apparent diguanylate cyclase activity effectively hydrolized by the YhjH phosphodiesterase (Figure 3A,B). Of note, motility was not diminished in the *S. typhimurium* UMR1 strain by any of the linker mutants. Consistent with the hypothesis that with optimized linker configuration diguanylate cyclase activity of the downstream GGDEF domain is triggered, combination of the S472V linker mutant with the D551A/E552A catalytic site mutant of the GGDEF domain displayed a rdar morphotype colony and motility equally as the wild type and vector control (Figure 3).

2.10 Archaeal GGDEF domain protein show diguanylate cyclase activity in an *in vivo* riboswitch assay

The rdar biofilm formation and motility assays of wild type proteins in combination with (potential) catalytic mutants indicated that the observed phenotypic alterations are due to the catalytic activity of the domains associated with cyclic di-GMP turnover. In order to verify whether indeed alterations in cyclic di-GMP are responsive for the phenotypic alterations we used a previously developed assay based on the Vc2 riboswitch located in single copy at the Tn*7* site in *E. coli* TOP10 (El Mouali et al., 2017, Liu et al., 2018, Liu et al., 2020). This assay is based on the alteration of beta-galactosidase activity in response to alternating concentrations of cyclic di-GMP which inhibit the translation of the gene product as it binds with nanomolecular affinity to the Vc2 aptamer to cause a conformational change in the riboswitch. Diguanylate cyclase activity can be readily sensitively monitored resulting in significant downregulation of the beta-galactosidase activity (El Mouali et al., 2017, Liu et al., 2018, Liu et al., 2020). To this end, this riboswitch assay indicated, consistent with the phenotypic biofilm assay, no cyclic di-GMP turnover in the wild type, the HD-GYP deletion mutant and the catalytic mutants (Figure 3C). The three linker mutants, however, indicated significant apparent diguanylate cyclase activity as beta-galactosidase activity was abolished throughout the experimental time span up to the 72 h endpoint. Recovery of beta-galactosidase activity by the introduction of the D551A/E552A mutation in the GGDEF motif in the S472V linker mutant recover the activity indicating loss of the catalytic activity.

Assessment of phosphodiesterase activity is most sensitively and consistently monitored in the background of elevated cyclic di-GMP concentrations produced by a diguanylate cyclase (El Mouali et al., 2017). As previously engineered constructs expressing the diguanylate cyclase AdrA were incompatible with expression of the WP_012035199.1 ORF under L-arabinose inducible pBAD promoter, we cloned AdrA in the pSRKGm vector under the control of the leaky plac promoter. To this end, this riboswitch assay indicated, consistent with the phenotypic biofilm assay, no cyclic di-GMP degradation in the wild type, the HD-GYP deletion mutant and the catalytic mutants (Figure 3D). The linker mutants, however, downregulated the beta-galactosidase activity even more indicating

diguanylate cyclase activity which was relieved by the introduction of the D551A/E552A mutation in the GGDEF motif.

The catalytic activities of WP_339106355.1 and WP_363463662.1 were monitored in biofilm, motility and Vc2 riboswitch assay (Figure 3A-D). To this end, production of the two proteins and its GQGAAF mutants showed upregulation of rdar biofilm formation in the *S. typhimurium* UMR1 Δ*yhjH* mutants without arabinose induction.

2.11 Selected GGDEF domain proteins from MAGs encode functional diguanylate cyclases

Having demonstrated that the archaeal isolate *M. avoryzae* MRE50 encodes an enzymatically functional GGDEF domain protein with diguanylate cyclase activity, we were wondering whether GGDEF domains with diguanylate cyclase activity are more widespread in Archaea. To this end, we searched the NCBI database of non-redundant Archaeal protein sequences with the diguanylate cyclase AdrA from *S. typhimurium* ATCC 14028 as a query. The phylogenetic origin of the contigs of the retrieved archaeal candidate GGDEF protein was assessed by investigating at least two additional gene products flanking the GGDEF domain protein gene up- and down-stream, respectively.

We subsequently selected twenty GGDEF domain proteins to be encoded mainly by archaeal contigs of the DPANN and class II archaeal clade (Supplementary Figure 5). These proteins were either only GGDEF proteins or proteins with a short N-terminal or C-terminal amino acid sequences too short to comprise a domain or with undefined short domain. Short only GGDEF domain proteins are rarely found in bacteria. Other proteins have diverse N-terminal signaling domains containing a GAF, REC or PAS/PAC domain, a PocR-GAF domain combination and a 7TM domain. All those signaling domains have been identified in bacterial GGDEF domain proteins with PAS/PAC, GAF and REC domains being most frequent (Anantharaman &Aravind, 2005, Xing, Gumerov et al., 2023). Alignment of the GGDEF domains of those proteins with group I, II and III GGDEF domains indicated that the majority of these proteins had the GG(D/E)EF motif and additional signature amino acid required for catalytic activity conserved while three of the proteins showed unconventional GGDEF motifs or insertions close to the central catalytic GGDEF motif (Supplementary Figure 6). These were MBI5036943,1 with a RIFHPSGDEF motif, UCD02967.1 with a RYGGDKGDEF and MBN1216112.1 with a 12 amino acid long insertion upstream of the RxxGGDEF motif. None of the 20 additional GGDEF domain proteins contained the signature serine that is indicative for cyclic AMP-GMP activity of a GGDEF domain (Hallberg et al., 2019). Subsequent construction of a phylogenetic tree of the 20 GGDEF domains with experimentally confirmed GGDEF domain proteins showed a cluster of archaeal GGDEF domains closest to functional WspR from *Pseudomonas aeruginosa* and AdrA of *S. typhimurium*. Other archaeal GGDEF domains are, however, also found outside of this cluster, dispersed throughout this phylogenetic tree with diverse GGDEF domains (Figure 5).

2.12 Phenotypic assays for archaeal GGDEF proteins from MAGs indicate catalytic activity

To this end, we cloned eight selected GGDEF domain proteins and constructed their GGAEF catalytic mutants in the GG(D/E)EF motif for assessment of the catalytic functionality to synthesize cyclic di-GMP by in vivo assays, the *S. typhimurium* rdar biofilm formation and motility and the Vc2 riboswitch assay. We found indications for catalytic activity of GGDEF proteins in all three assays, however, the proteins which showed indications for catalytic functionality were different in the three assays (Figure 6; Supplementary Figure 7). For example, KOH53289.1 and MBN1216112.1 indicated diguanylate cyclase activity in the Vc2*lacZY* riboswitch assay as relieve of the downregulation of the beta-galactosidase activity occurred in the GGAEF mutant. Of note, the protein KOH53915.1 homologous to

KOH53289.1 showed an opposite behaviour in the riboswitch assay and the motility assay. On the other hand, KY476720.1, KHO53915.1, MBN1216112.1 and MBT3408960.1 showed upregulated biofilm behavior either in *S. typhimurium* UMR1 upon 0.1% L-arabinose induction or in *S. typhimurium* UMR1 Δ*yhjH* without L-arabinose, respectively.

2.13 The PAS-PAS-GGDEF domain protein KYK2270.1 possesses pronounced diguanylate cyclase activity in the *S. typhimurium* rdar biofilm and motility model

As none of the previously MAG-derived GGDEF domain proteins consistently indicated (apparent) catalytic activity in all three *in vivo* assays, we investigated additional GGDEF domain proteins. To this end consistent (apparent) diguanylate cyclase activity was observed for the PAS-PAS-GGDEF protein KYK22704.1 compared to its GGAAF catalytic variant in the biofilm and motility assays equally as in the *E. coli* TOP10 Tn*7*::Vc2*lacZY* riboswitch assay which monitors intracellular levels of cyclic di-GMP (Figure 7). As pronounced diguanylate cyclase activity was observed, we assessed the activity of the enzyme also at elevated temperature of 37°C. Indeed catalytic activity as indicated by elevated rdar biofilm formation upon expression of the wild type protein, but not its GGAAF variant, was also observed. As KYK22704.1 was highly active, we assessed the role of the RxxD motif located six amino acids N-terminal of the GGDEF motif which is conventionally part of the allosteric inhibitory (I)-site (Christen et al., 2006). While the R330A mutant showed extended enhanced rdar biofilm formation compared to the wild type protein as expected from the impairment of the I-site, the D333A mutant displayed enhanced rdar morphotype development only at 28°C, but in contrast, a reduced apparent catalytic activity displayed as a downregulated rdar morphotype at 37°C.

We further assessed whether any of the three additional MAG derived GGDEF protein, UCD02967.1, HJJ27798.1 and RJQ22058.1 demonstrated apparent diguanylate cyclase activity (Supplementary Figure 8). Indeed, production of RJQ22058.1 upregulated biofilm formation, repressed motility and β-galactosidase activity in the riboswitch assay while the D256A catalytic mutant acted as the vector control indicating diguanylate cyclase activity of RJQ22058.1 in the *Salmonella* and *E. coli* model. In contrast, Production of UCD02967. and HJJ27798.1 and its catalytic mutants did not significantly alter the phenotypes compared to the vector control.

Discussion

Cyclic di-GMP is a multifunctional nearly ubiquitous second messenger in Gram-negative and Gram-positive Bacteria regulating fundamental single cell life style switch between motility and sessility (Purcell &Tamayo, 2016, Romling et al., 2013). In this work we provided evidence for cyclic di-GMP as a primordial molecule dedicated to the regulation of biofilm formation at the origin of life already in LUCA and before as we show that GGDEF domain proteins and their predominantly produced product, the second messenger cyclic di-GMP, are encoded by Archaea. These findings indicate cyclic di-GMP and presumably an entire signaling cascade including cyclic di-GMP specific phosphodiesterases and receptors to be indeed present in multiple branches of Archaea supporting the hypothesis of an ancient origin of this ubiquitous second messenger molecule.

Importantly, we identified GGDEF domain proteins that synthesize cyclic di-GMP in archaeal isolates from different branches of the archaeal phylogenetic tree, isolates *M. arvoryzae* MRE50, *H. salinisoli* KLK7 and *H. borinquense* E3 and *Methanococcoides* sp. SA1 (NCBI accession number JABKBV010000000, not experimentally investigated). Indeed, the Methanococcoides sp. SA1 genome codes for two GGDEF domain proteins (Figure 5; Supplementary Figure 6). This work also shows that GGDEF domain proteins from Archaea show activity in the *S. typhimurium* and *E. coli* model systems indicating a universal functionality of the cyclic di-GMP signaling system. In case of phenotypes not mediated by the catalytic activity of the enzymes, the model systems can serve to assess potentially novel regulatory pathways and protein-protein interactions that will support the universal functionality and exceptional modularity of this signaling system.

While we show in this work that cyclic di-GMP signaling is indeed present in Archaea, open questions remain. First, we have not systematically investigated the presence of GGDEF domain proteins throughout the archaeal domain of life. Thus, at the moment, the fraction of isolates that possess a GGDEF domain protein and potentially a cyclic di-GMP signaling network is unknown. However, our preliminary overall assessment of the occurrence of GGDEF domains in Archaea indicates that GGDEF domain mediated cyclic di-GMP signaling might only rarely be found in this domain of life. Also in contrast to Bacteria preliminary evidence indicates that turnover proteins are not found in multiple copies encoded by one chromosome. However, a scarce occurrence might not necessarily be an indication for horizontal gene transfer from bacteria at some point in evolution. Although cyclic di-GMP signaling is (nearly) ubiquitous in Bacteria, it occurs rarely in distinct phyla such as *Bacteroidota* (Morehouse, Govande et al., 2020). In addition, cyclic di-GMP signaling and its synthesizing and hydrolizing proteins can rapidly disappear in evolution (Liu et al., 2020). Although cyclic di-GMP signaling is predominantly present in *Gammaproteobacteria* with turnover proteins present in multiple (>60 copies of GGDEF domain proteins encoded by one chromosome), cyclic di-GMP signaling is (nearly) absent in the family *Morganellaceae* despite free living bacteria thriving in multiple ecological niches with a large genome size to exist in this genus (Liu et al., 2020). Verified reasons for the rapid disappearance of cyclic di-GMP are niche restriction such as in *Shigella* spp., genome reduction and the use of cyclic di-GMP as a molecule mediating phage defense (Lai, Kumagai et al., 2009, Morehouse et al., 2020, Ojha, Dittmar et al., 2021). On the other hand, there are also examples of regain of a cyclic di-GMP signaling system such as the horizontal gene transfer of entire signaling modules comprising the diguanylate cyclase and a corresponding effector cellulose synthase like enzyme such as in *Streptococcus* spp. (Liu et al., 2020). Thus, although the cyclic di-GMP signaling system might have been present in Archaea originally, the possibility exists that it could have been (nearly) depleted early in evolution only being maintained occasionally. In this context GGDEF domain based di-GMP signaling might have interfered with frequently occurring antiphage defense systems such as the cyclic-oligonucleotide-based anti-phage signaling system (CBASS) (Morehouse et al., 2020). The diverse

cGAS/DncV-like nucleotidyltransferases (CD-NTases)/SMODS nucleotide synthases of the CBASS system (which are structurally unrelated to the GGDEF domain) synthesize a number of different cyclic di-nucleotides among them cyclic di-GMP. However, whether CBASS CD-NTases which are present in Archaea synthesize cyclic di-GMP as part of anti-phage defense systems and whether this occurs mutually exclusive with GGDEF synthesized cyclic di-GMP is currently unknown.

On the other hand, extensive horizontal gene transfer has occurred and is occurring between Bacteria and Archaea that even determinatively contributed to the emerging archaeal order Halobacterales (Gophna &Altman-Price, 2022). Thus it cannot be excluded that horizontal transfer of cyclic di-GMP turnover protein genes might have occurred that also contributed to the occurrence of GGDEF domain proteins in the archaeal *Halobacteria*. If so, GGDEF domain proteins will not have been maintained, if they would not provide a competitive advantage for the organism. It can therefore be hypothesized that the transferred GGDEF domain proteins, perhaps together with their receptors, have opened new possibilities for forming biofilms, regulating motility or affecting other physiological features and metabolic pathways (Liu et al., 2020, Mikkelsen, Ball et al., 2009).

In conclusion this work opens the way to analyze cyclic di-GMP signaling in archaea. Immediate follow-up questions at the level of individual genomes include the (co)occurrence of EAL and/or HD-GYP domain phosphodiesterases with GGDEF domains, the identity of archaeal cyclic di-GMP receptors and the physiological and metabolic role of cyclic di-GMP signaling. On the phylogenetic level, to analyze domain composition between archaeal and bacterial proteins will provide indications of functional dissection to different signals between Archaea and Bacteria. As initial observation indicated archaeal GGDEF domains deviating from the consensus catalytic GGDEF motif, the function of GGDEF domains which deviate substantially in the signature amino acids needs to be defined. In addition, the phylogenetic context of cyclic di-GMP turnover proteins is open questions. Last but not least, in conclusion, the discovery of cyclic di-GMP synthesizing GGDEF domain proteins present in Archaea opens question of their occurrence, function and role.

Acknowledgements

We cordially thank Mahfuzur Rahman for the help with the cloning of three archaeal GGDEF protein genes, and Niels Theys and Sebahat Hale Ersahin for support with motility assays during their intershipps. This work has been supported by a grant from the Swedish Research Council for Natural Sciences and Engineering (diary number 2022-04865) to UR and the Karolinska Institutet.

References

Ahmad, I., Wigren, E., Le Guyon, S., Vekkeli, S., Blanka, A., El Mouali, Y. et al. (2013) The eal-like protein stm1697 regulates virulence phenotypes, motility and biofilm formation in salmonella typhimurium. *Mol Microbiol* **90**: 1216-1232.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Anantharaman, V. & Aravind, L. (2005) Meds and pocr are novel domains with a predicted role in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. *Bioinformatics* **21**: 2805-2811.

Aouad, M., Flandrois, J.P., Jauffrit, F., Gouy, M., Gribaldo, S. & Brochier-Armanet, C. (2022) A divide-and-conquer phylogenomic approach based on character supermatrices resolves early steps in the evolution of the archaea. *BMC Ecol Evol* **22**: 1.

Baral, R., Ho, K., Kumar, R.P., Hopkins, J.B., Watkins, M.B., LaRussa, S. et al. (2025) A general mechanism for initiating the bacterial general stress response. *Elife* **13**:

Botsford, J.L. & Harman, J.G. (1992) Cyclic amp in prokaryotes. *Microbiol Rev* **56**: 100-122.

Bubeck, P., Winkler, M. & Bautsch, W. (1993) Rapid cloning by homologous recombination in vivo. *Nucleic Acids Res* **21**: 3601-3602.

Cao L.-Y., Z.X., Bai F.-W., Römling U. (2026) Previously uncharacterized aliphatic amino acid positions modulate the apparent catalytic activity of the eal domain of zmo_1055 and other cyclic di-gmp specific eal phosphodiesterases. *Microb Biotechnol* **accepted**:

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P. et al. (2006) Allosteric control of cyclic di-gmp signaling. *J Biol Chem* **281**: 32015-32024.

Corrigan, R.M., Campeotto, I., Jeganathan, T., Roelofs, K.G., Lee, V.T. & Grundling, A. (2013) Systematic identification of conserved bacterial c-di-amp receptor proteins. *Proc Natl Acad Sci U S A* **110**: 9084-9089.

Corrigan, R.M. & Grundling, A. (2013) Cyclic di-amp: Another second messenger enters the fray. *Nat Rev Microbiol* **11**: 513-524.

Dahlstrom, K.M., Giglio, K.M., Collins, A.J., Sondermann, H. & O'Toole, G.A. (2015) Contribution of physical interactions to signaling specificity between a diguanylate cyclase and its effector. *mBio* **6**: e01978-01915.

Dayton, H., Smiley, M.K., Forouhar, F., Harrison, J.J., Price-Whelan, A., Dietrich, L.E.P. (2020) Sensory domains that control cyclic di-gmp-modulating proteins: A critical frontier in bacterial signal transduction. In *Microbial cyclic di-nucleotide signaling*, Chou, S.-H., Guiliani, N., Lee, V.T., Römling, U. (ed) pp 137-158. Cham: Springer Cham

De, N., Navarro, M.V., Raghavan, R.V. & Sondermann, H. (2009) Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator wspr. *J Mol Biol* **393**: 619-633.

Dong, X.Y., Mao, Y.L., Zhang, Q.K., Zhu, L.R., Hou, J. & Cui, H.L. (2024) Genome-based classification of the family natrialbaceae and description of four novel halophilic archaea from three saline lakes and a saline-alkaline land. *Extremophiles* **28**: 47.

Duncan-Lowey, B., McNamara-Bordewick, N.K., Tal, N., Sorek, R. & Kranzusch, P.J. (2021) Effector-mediated membrane disruption controls cell death in cbass antiphage defense. *Mol Cell* **81**: 5039-5051 e5035.

El Mouali, Y., Kim, H., Ahmad, I., Brauner, A., Liu, Y., Skurnik, M. et al. (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**:

Finstad, K.M., Probst, A.J., Thomas, B.C., Andersen, G.L., Demergasso, C., Echeverria, A. et al. (2017) Microbial community structure and the persistence of cyanobacterial populations in salt crusts of the hyperarid atacama desert from genome-resolved metagenomics. *Front Microbiol* **8**: 1435.

Flemming, H.C. & Wuertz, S. (2019) Bacteria and archaea on earth and their abundance in biofilms. *Nat Rev Microbiol* **17**: 247-260.

Galperin, M.Y. & Chou, S.H. (2022) Sequence conservation, domain architectures, and phylogenetic distribution of the hd-gyp type c-di-gmp phosphodiesterases. *J Bacteriol* **204**: e0056121.
Giardina, G., Paiardini, A., Fernicola, S., Franceschini, S., Rinaldo, S., Stelitano, V. et al. (2013) Investigating the allosteric regulation of yfin from pseudomonas aeruginosa: Clues from the structure of the catalytic domain. *PLoS One* **8**: e81324.
Gophna, U. & Altman-Price, N. (2022) Horizontal gene transfer in archaea-from mechanisms to genome evolution. *Annu Rev Microbiol* **76**: 481-502.
Gourinchas, G., Etzl, S., Gobl, C., Vide, U., Madl, T. & Winkler, A. (2017) Long-range allosteric signaling in red light-regulated diguanylyl cyclases. *Sci Adv* **3**: e1602498.
Grantcharova, N., Peters, V., Monteiro, C., Zakikhany, K. & Romling, U. (2010) Bistable expression of csgd in biofilm development of salmonella enterica serovar typhimurium. *J Bacteriol* **192**: 456-466.
Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose pbad promoter. *J Bacteriol* **177**: 4121-4130.
Hallberg, Z.F., Chan, C.H., Wright, T.A., Kranzusch, P.J., Doxzen, K.W., Park, J.J. et al. (2019) Structure and mechanism of a hypr ggdef enzyme that activates cgamp signaling to control extracellular metal respiration. *Elife* **8**:
Higgins, D.G. & Sharp, P.M. (1988) Clustal: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244.
Holm, L., Laiho, A., Toronen, P. & Salgado, M. (2023) Dali shines a light on remote homologs: One hundred discoveries. *Protein Sci* **32**: e4519.
Hu, X.M., Peng, L., Wang, Y., Ma, F., Tao, Y., Liang, X. et al. (2024) Bacterial c-di-gmp triggers metamorphosis of mussel larvae through a sting receptor. *NPJ Biofilms Microbiomes* **10**: 51.
Jones, D.H. & Howard, B.H. (1991) A rapid method for recombination and site-specific mutagenesis by placing homologous ends on DNA using polymerase chain reaction. *Biotechniques* **10**: 62-66.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with alphafold. *Nature* **596**: 583-589.
Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. (2015) The phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845-858.
Kellner, S., Spang, A., Offre, P., Szollosi, G.J., Petitjean, C. & Williams, T.A. (2018) Genome size evolution in the archaea. *Emerg Top Life Sci* **2**: 595-605.
Khan, S.R., Gaines, J., Roop, R.M., 2nd & Farrand, S.K. (2008) Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of trar and tram expression on ti plasmid quorum sensing. *Appl Environ Microbiol* **74**: 5053-5062.
Kulasekara, B.R., Kamischke, C., Kulasekara, H.D., Christen, M., Wiggins, P.A. & Miller, S.I. (2013) C-di-gmp heterogeneity is generated by the chemotaxis machinery to regulate flagellar motility. *Elife* **2**: e01402.
Kumar, S., Stecher, G. & Tamura, K. (2016) Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**: 1870-1874.
Lai, T.H., Kumagai, Y., Hyodo, M., Hayakawa, Y. & Rikihisa, Y. (2009) The anaplasma phagocytophilum plec histidine kinase and pled diguanylate cyclase two-component system and role of cyclic di-gmp in host cell infection. *J Bacteriol* **191**: 693-700.
Leichtling, B.H., Rickenberg, H.V., Seely, R.J., Fahrney, D.E. & Pace, N.R. (1986) The occurrence of cyclic amp in archaebacteria. *Biochem Biophys Res Commun* **136**: 1078-1082.
Li, F., Cimdins, A., Rohde, M., Jansch, L., Kaever, V., Nimtz, M. et al. (2019) Dncv synthesizes cyclic gmp-amp and regulates biofilm formation and motility in escherichia coli ecor31. *mBio* **10**:
Liu, Y., Kim, H. & Romling, U. (2018) In vivo analysis of cyclic di-gmp cyclase and phosphodiesterase activity in escherichia coli using a vc2 riboswitch-based assay. *Bio Protoc* **8**: e2753.
Liu, Y., Lee, C., Li, F., Trcek, J., Bahre, H., Guo, R.T. et al. (2020) A cyclic di-gmp network is present in gram-positive streptococcus and gram-negative proteus species. *ACS Infect Dis* **6**: 2672-2687.
Lori, C., Ozaki, S., Steiner, S., Bohm, R., Abel, S., Dubey, B.N. et al. (2015) Cyclic di-gmp acts as a cell cycle oscillator to drive chromosome replication. *Nature* **523**: 236-239.

Martin, W. & Russell, M.J. (2003) On the origins of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* **358**: 59-83; discussion 83-55.
McDonough, K.A. & Rodriguez, A. (2011) The myriad roles of cyclic amp in microbial pathogens: From signal to sword. *Nat Rev Microbiol* **10**: 27-38.
Mehne, F.M., Schroder-Tittmann, K., Eijlander, R.T., Herzberg, C., Hewitt, L., Kaever, V. et al. (2014) Control of the diadenylate cyclase cdas in bacillus subtilis: An autoinhibitory domain limits cyclic di-amp production. *J Biol Chem* **289**: 21098-21107.
Mikkelsen, H., Ball, G., Giraud, C. & Filloux, A. (2009) Expression of pseudomonas aeruginosa cupd fimbrial genes is antagonistically controlled by rcsb and the eal-containing pvrr response regulators. *PLoS One* **4**: e6018.
Moller, S., Croning, M.D. & Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**: 646-653.
Morehouse, B.R., Govande, A.A., Millman, A., Keszei, A.F.A., Lowey, B., Ofir, G. et al. (2020) Sting cyclic dinucleotide sensing originated in bacteria. *Nature* **586**: 429-433.
Nagar, D.N., Mani, K. & Braganca, J.M. (2024) Genomic insights on carotenoid synthesis by extremely halophilic archaea haloarcula rubripromontorii bs2, haloferax lucentense bbk2 and halogeometricum borinquense e3 isolated from the solar salterns of india. *Sci Rep* **14**: 20214.
Nelson, J.W., Sudarsan, N., Phillips, G.E., Stav, S., Lunse, C.E., McCown, P.J. et al. (2015) Control of bacterial exoelectrogenesis by c-amp-gmp. *Proc Natl Acad Sci U S A* **112**: 5389-5394.
O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. (1991) X-ray structure of the gcn4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**: 539-544.
Ojha, R., Dittmar, A.A., Severin, G.B. & Koestler, B.J. (2021) Shigella flexneri diguanylate cyclases regulate virulence. *J Bacteriol* **203**: e0024221.
Paul, R., Weiser, S., Amiot, N.C., Chan, C., Schirmer, T., Giese, B. et al. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* **18**: 715-727.
Peng, L.H., Liang, X., Chang, R.H., Mu, J.Y., Chen, H.E., Yoshida, A. et al. (2020) A bacterial polysaccharide biosynthesis-related gene inversely regulates larval settlement and metamorphosis of mytilus coruscus. *Biofouling* **36**: 753-765.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. et al. (2004) Ucsf chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612.
Potrykus, K. & Cashel, M. (2008) (p)ppgpp: Still magical? *Annu Rev Microbiol* **62**: 35-51.
Purcell, E.B. & Tamayo, R. (2016) Cyclic diguanylate signaling in gram-positive bacteria. *FEMS Microbiol Rev* **40**: 753-773.
Rall, T.W. & Sutherland, E.W. (1958) Formation of a cyclic adenine ribonucleotide by tissue particles. *J Biol Chem* **232**: 1065-1076.
Rao, F., Yang, Y., Qi, Y. & Liang, Z.X. (2008) Catalytic mechanism of cyclic di-gmp-specific phosphodiesterase: A study of the eal domain-containing rocr from pseudomonas aeruginosa. *J Bacteriol* **190**: 3622-3631.
Rasmussen, R., Benvegnu, D., O'Shea, E.K., Kim, P.S. & Alber, T. (1991) X-ray scattering indicates that the leucine zipper is a coiled coil. *Proc Natl Acad Sci U S A* **88**: 561-564.
Robert, X. & Gouet, P. (2014) Deciphering key features in protein structures with the new endscript server. *Nucleic Acids Res* **42**: W320-324.
Romling, U. (2008) Great times for small molecules: C-di-amp, a second messenger candidate in bacteria and archaea. *Sci Signal* **1**: pe39.
Romling, U. (2023a) Cyclic di-gmp signaling-where did you come from and where will you go? *Mol Microbiol* **120**: 564-574.
Romling, U. (2023b) Is biofilm formation intrinsic to the origin of life? *Environ Microbiol* **25**: 26-39.
Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. & Normark, S. (1998) Curli fibers are highly conserved between salmonella typhimurium and escherichia coli with respect to operon structure and regulation. *J Bacteriol* **180**: 722-731.

Romling, U., Cao, L.Y. & Bai, F.W. (2023) Evolution of cyclic di-gmp signalling on a short and long term time scale. *Microbiology (Reading)* **169**:
Romling, U., Galperin, M.Y. & Gomelsky, M. (2013) Cyclic di-gmp: The first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77**: 1-52.
Romling, U., Gomelsky, M. & Galperin, M.Y. (2005) C-di-gmp: The dawning of a novel bacterial signalling system. *Mol Microbiol* **57**: 629-639.
Romling, U., Rohde, M., Olsen, A., Normark, S. & Reinkoster, J. (2000) Agfd, the checkpoint of multicellular and aggregative behaviour in salmonella typhimurium regulates at least two independent pathways. *Mol Microbiol* **36**: 10-23.
Romling, U., Sierralta, W.D., Eriksson, K. & Normark, S. (1998) Multicellular and aggregative behaviour of salmonella typhimurium strains is controlled by mutations in the agfd promoter. *Mol Microbiol* **28**: 249-264.
Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R. et al. (1987) Regulation of cellulose synthesis in acetobacter xylinum by cyclic diguanylic acid. *Nature* **325**: 279-281.
Ruiz, L.M., Castro, M., Barriga, A., Jerez, C.A. & Guiliani, N. (2012) The extremophile acidithiobacillus ferrooxidans possesses a c-di-gmp signalling pathway that could play a significant role during bioleaching of minerals. *Lett Appl Microbiol* **54**: 133-139.
Sakai, S., Conrad, R., Liesack, W. & Imachi, H. (2010) Methanocella arvoryzae sp. Nov., a hydrogenotrophic methanogen isolated from rice field soil. *Int J Syst Evol Microbiol* **60**: 2918-2923.
Schirmer, T. (2016) C-di-gmp synthesis: Structural aspects of evolution, catalysis and regulation. *J Mol Biol* **428**: 3683-3701.
Schirmer, T. & Jenal, U. (2009) Structural and mechanistic determinants of c-di-gmp signalling. *Nat Rev Microbiol* **7**: 724-735.
Simm, R., Lusch, A., Kader, A., Andersson, M. & Romling, U. (2007) Role of eal-containing proteins in multicellular behavior of salmonella enterica serovar typhimurium. *J Bacteriol* **189**: 3613-3623.
Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U. (2004) Ggdef and eal domains inversely regulate cyclic di-gmp levels and transition from sessility to motility. *Mol Microbiol* **53**: 1123-1134.
Simm, R., Remminghorst, U., Ahmad, I., Zakikhany, K. & Romling, U. (2009) A role for the eal-like protein stm1344 in regulation of csgd expression and motility in salmonella enterica serovar typhimurium. *J Bacteriol* **191**: 3928-3937.
Stothard, P. (2000) The sequence manipulation suite: Javascript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**: 1102, 1104.
Sullivan, M.J., Petty, N.K. & Beatson, S.A. (2011) Easyfig: A genome comparison visualizer. *Bioinformatics* **27**: 1009-1010.
Takahashi, K., Kasai, K. & Ochi, K. (2004) Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppgpp) in plants. *Proc Natl Acad Sci U S A* **101**: 4320-4324.
Teng, Y., Gong, X., Zhang, J., Obideen, Z. & Yan, Y. (2024) Investigating and engineering an 1,2-propanediol-responsive transcription factor-based biosensor. *ACS Synth Biol* **13**: 2177-2187.
Trevors, J.T. (2011) Hypothesized origin of microbial life in a prebiotic gel and the transition to a living biofilm and microbial mats. *Comptes Rendus Biologies* **334**: 269–272.
UniProt, C. (2025) Uniprot: The universal protein knowledgebase in 2025. *Nucleic Acids Res* **53**: D609-D617.
van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M. et al. (2024) Fast and accurate protein structure search with foldseek. *Nat Biotechnol* **42**: 243-246.
Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. (2009) Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.
Whiteley, A.T., Eaglesham, J.B., de Oliveira Mann, C.C., Morehouse, B.R., Lowey, B., Nieminen, E.A. et al. (2019) Bacterial cgas-like enzymes synthesize diverse nucleotide signals. *Nature* **567**: 194-199.
Wigren, E., Liang, Z.X. & Romling, U. (2014) Finally! The structural secrets of a hd-gyp phosphodiesterase revealed. *Mol Microbiol* **91**: 1-5.

Witte, G., Hartung, S., Buttner, K. & Hopfner, K.P. (2008) Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell* **30**: 167-178.
Wu, Q., Lu, S., Zhang, C., Zhong, W., Zhao, H., Zhao, Y. et al. (2025) Significant impacts of metal ions from ancient oceans on nucleoside phosphorylation. *BMC Chem* **19**: 252.
Xie, Y., Li, H., Luo, X., Li, H., Gao, Q., Zhang, L. et al. (2022) Ibs 2.0: An upgraded illustrator for the visualization of biological sequences. *Nucleic Acids Res* **50**: W420-W426.
Xing, J., Gumerov, V.M. & Zhulin, I.B. (2023) Origin and functional diversification of pas domain, a ubiquitous intracellular sensor. *Sci Adv* **9**: eadi4517.

Figures and figure legends

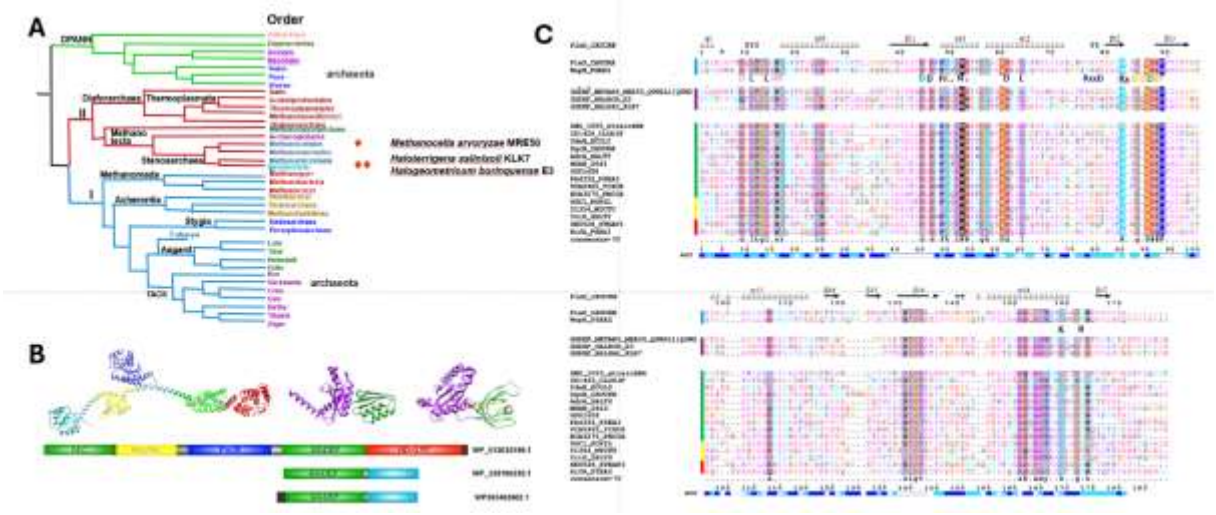


Figure 1
Initial characterisation of GGDEF domain protein from archaeal isolates *M. arvoryzae* MRE50, *H. salinisoli* KLK7 and *H. borinquense* E3.
Figure 1A Phylogenetic tree of Archaea with phylogenetic position of isolates encoding GGDEF domain proteins. Phylogenetic tree has been redrawn from (Aouad, Flandrois et al., 2022).
Figure 1B Structural models and domain structure of GGDEF domain proteins WP_012035199.1 from *M. arvoryzae* MRE50, WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3. Structural models have been calculated with AlphaFold 3 (Jumper et al., 2021) with colors according to domain structure.
Figure 1C Alignment of GGDEF domain of proteins WP_012035199.1 from *M. arvoryzae* MRE50, WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3 with GGDEF domains from well investigated catalytically active GGDEF domains from the PleD from *Caulobacter vibrioides* CB15 (CC_2462, PleD_CAUCRE) and WspR from *Pseudomonas aeruginosa* PAO1 (PA3702, WspR_PSEAI). Green bar, catalytically active GGDEF domains; yellow bar, domains with unconventional signature motifs; red, catalytically inactive GGDEF domains. Protein designation in supplementary material. Alignment was created with ClustalX2 (Higgins &Sharp, 1988), manually curated in GeneDoc and visualized with ESPript 3.0 (Robert &Gouet, 2014). Secondary structure of PleD from 2V0N (RCSB Protein Data Bank, RCSB PDB).

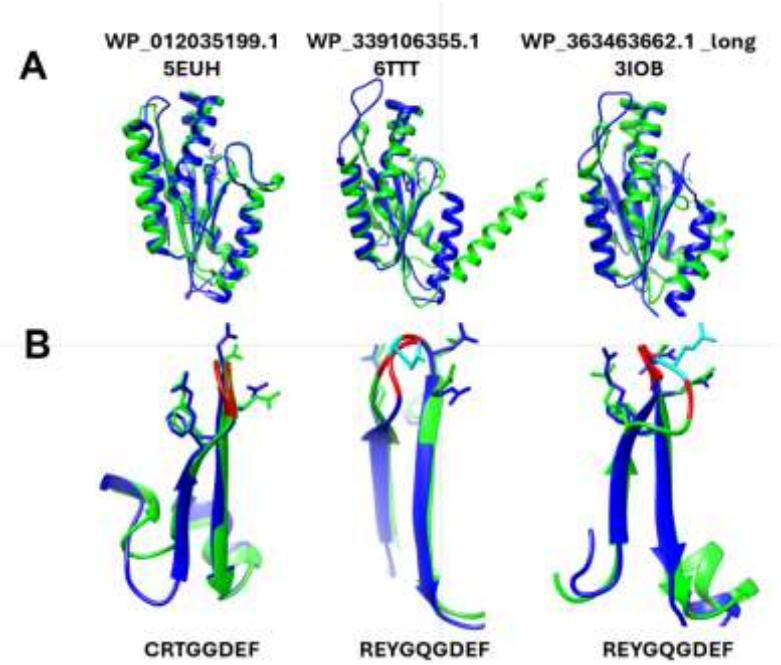


Figure 2

Structural models of GGDEF domains of proteins WP_012035199.1 from *M. arvoryzae* MRE50, WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3 overlayed with most homologous experimentally determined structures of catalytically active GGDEF domains.

Figure 2A. Structural models (in green) of GGDEF domains of proteins WP_012035199.1 from *M. arvoryzae* MRE50, WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3 were determined by AlphaFold 3 (Jumper et al., 2021). Overlay with 5EUH (GcbC of *Pseudomonas fluorescens* (Dahlstrom, Giglio et al., 2015) , 6TTS (DgcB from C. vibrioides CB15) and 4IOB (PA1120/YfiN/TpbB from *P. aeruginosa* PAO1 (Giardina, Paiardini et al., 2013)) in blue, respectively. Structural homologs to the structural models were searched for with DALI (Holm et al., 2023) and FoldSeek (van Kempen et al., 2024) and subsequently overlayed with MatchMaker in Chimera 1.18 (Pettersen et al., 2004).

Figure 2B. Close-up on the overlay between the structural models (green) of GGDEF domains of proteins WP_012035199.1 from *M. arvoryzae* MRE50, WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3 and experimentally determined structures (blue) 5EUH, 6TTS and 4IOB, respectively focusing on the of the β-hairpin with the GGDEF motif. Amino acids of the GGDEF motifs are displayed with their side chains with glycine in red and inserted glutamine Q in light blue.

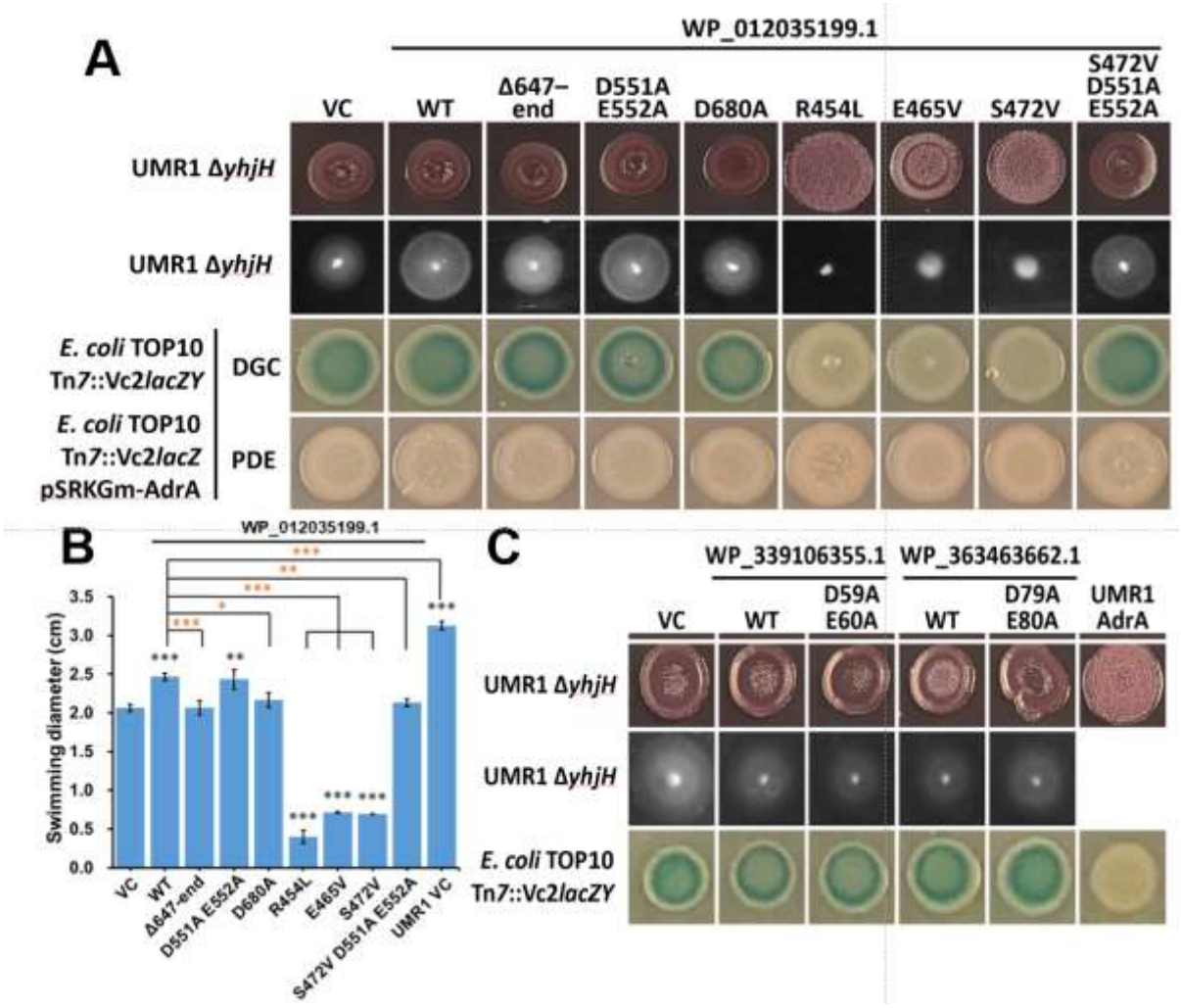


Figure 3

In vivo assays rdar biofilm formation and flagella based swimming motility in the S. typhimurium UMR1 model and Vc2 riboswitch assay in *E. coli* TOP10 Tn*7*::Vc2*lacZY* to assess catalytic activity of the GGDEF-HD-GYP domain protein WP_012035199.1 from *M. arvoryzae* MRE50 and its variants under different growth conditions.

Figure 3A Rdar colony morphology assay to assess the catalytic activity of WP_012035199.1 from *M. arvoryzae* MRE50 and its variants in *S. typhimurium* UMR1 Δ*yhjH* on CR agar plates using LB without salt medium. Ten µl of a bacterial suspension of OD_{600}=5 were placed on the agar, dried and subsequently incubated at 28 °C. Proteins were produced under the control of the pBAD promoter induced with 0.1% L-arabinose and development of colony morpohology and color assessed up to 72 h. displayed is rdar biofilm development after 48 h.

Flagella based swimming motility assay to assess the catalytic activity of WP_012035199.1 from *M. arvoryzae* MRE50 and its variants in *S. typhimurium* UMR1 Δ*yhjH*. Two µl of a bacterial suspension of OD_{600}=5 were inoculated into the 0.25% agar with LB as medium with the plate subsequently incubated at 37 °C. Motiity was documented after 6 h. Proteins were produced under the control of the pBAD promoter induced with 0.1% L-arabinose. MAE108 (*S. typhimurium* UMR1 Δ*fliC* Δ*fljB*) served as a negative control.

Vc2-based riboswitch assay to assess diguanylate cyclase catalytic activity GGDEF and GGDEF-HD-GYP domain proteins. Catalytic activity of WP_012035199.1 from *M. arvoryzae* MRE50 and its variants was assessed in *E. coli* TOP10 Tn*7*::Vc2*lacZY* via downregulation of beta-galactosidase activity. Three µl of a bacterial suspension of OD_{600}=5 were placed onto an LB agar plate containing 80 µg/ml X-gal and appropriate antibiotics. Plates were incubated up to 72 h to assess diminished development of the blue precipitate 5,5'-

dibromo-4,4'-dchloroindigo after dimerisation and oxidation. Cyclic di-nucleotide controls for cyclic di-GMP were *E. coli* TOP10 Tn*7*::Vc2*lacZY* pAdrA and the pAdrA G288A, G289A mutant; pDncV and its catalytic pDncV D129A, D131A mutant and pCdaS and its catalytic pCdaS D135A mutant (in Supplementary Figure 5).

Vc2-based riboswitch assay to assess phosphodiesterase catalytic activity of EAL and HD-GYP domain proteins. Catalytic activity of WP_012035199.1 from *M. arvoryzae* MRE50 and its variants was assessed in *E. coli* TOP10 Tn*7*::Vc2*lacZY* pAdrAGm by upregulation of beta-galactosidase activity. Three µl of a bacterial suspension of OD_{600}=5 were placed onto an LB agar plate containing 80 µg/ml X-gal and appropriate antibiotics. Plates were incubated up to 72 h to assess development of the blue precipitate 5,5'-dibromo-4,4'-dichloroindigo after dimerisation and oxidation. Cyclic di-nucleotide controls for cyclic di-GMP hydrolysis were pYhjH and its pYhjH E48A and pYhjH E136A catalytic mutants (in Supplementary Figure 5).

Figure 3B Quantitative estimation of flagella based swimming motility to assess the catalytic activity of WP_012035199.1 from *M. arvoryzae* MRE50 and its variants in *S. typhimurium* UMR1 Δ*yhjH*. p*<0.05, p**<0.01, p***<0.001 by student's t-test compared to *S. typhimurium* UMR1 Δ*yhH* vector control pBAD28 (black stars) and *S. typhimurium* UMR1 Δ*yhH* with WP_012035199.1 cloned in pBAD28 (red stars).

Figure 3C *In vivo* assays rdar biofilm formation and flagella based swimming motility in the S. typhimurium UMR1 model and Vc2 riboswitch assay in *E. coli* TOP10 Tn*7*::Vc2*lacZY* to assess catalytic activity of the GQGDEF domain proteins WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3 and its catalytic GQGAAF motif mutants under different growth conditions. Assays were conducted as described for Figure 3A besides that L-arabinose has been omitted in the rdar biofilm assay and flagella-based motility was documented after 5 h.

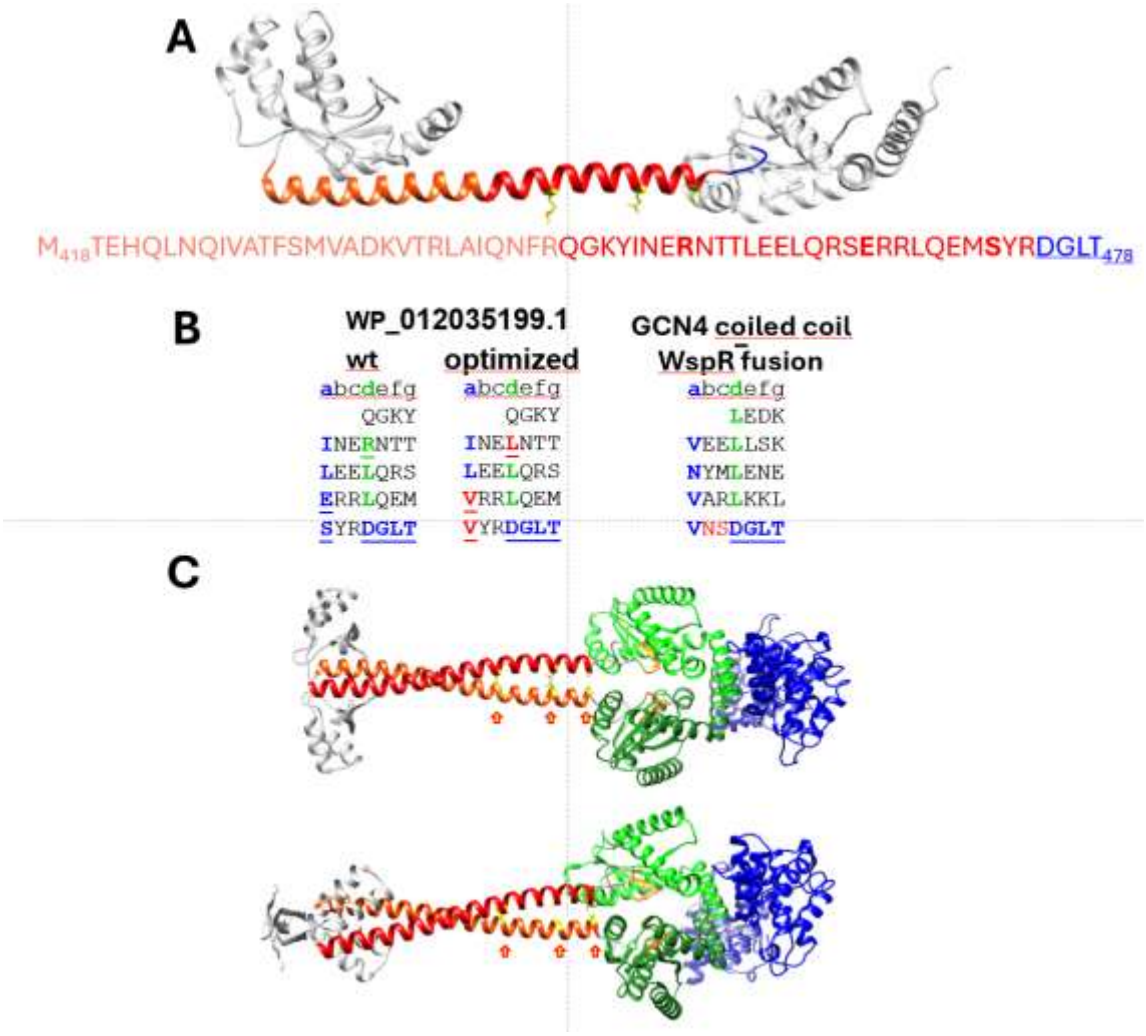


Figure 4
Characterisation of the linker region between the PocR and the GGDEF domain of the WP_012035199.1 from *M. arvoryzae* MRE50 as a hydrophobic heptad repeat sequence potentially capable to form coiled-coil interactions.
Figure 4 A. Left, wild type heptad repeat sequence of the linker region ending C-terminal at the DGLT motif which marks the beginning of a wid turn compared to a middle, optimized dimerizing as a heptad repeat linker sequence and right the GCN4-WspRGGDEF hybrid linker sequence. Positions a and d comprise amino acids with an aliphatic side chain such as valine and leucine.
Amino acid sequence of the linker between the PocR and the GGDEF domain.
AlphaFold 3 structural model of the linker (in red) with the wide-turn at the C-terminal end starting with the DGLT motif (in blue). The linker is flanked by the PocR and the GGDEF domain (in grey), respectively. The amino acids to be substituted for an optimal heptad repeat sequence are indicated with their side chains in yellow.
Figure 4 B. Dimer formation of the wild type linker-GGDEF-HD-GYP domain protein WP_012035199.1 from *M. arvoryzae* MRE50 and the three amino acid substitution linker variant. The amino acids to be substituted for an optimal heptad repeat sequence and substituted are indicated with their side chains in yellow. The GGDEF domains of the monomer face each other in an apparently active constellation.

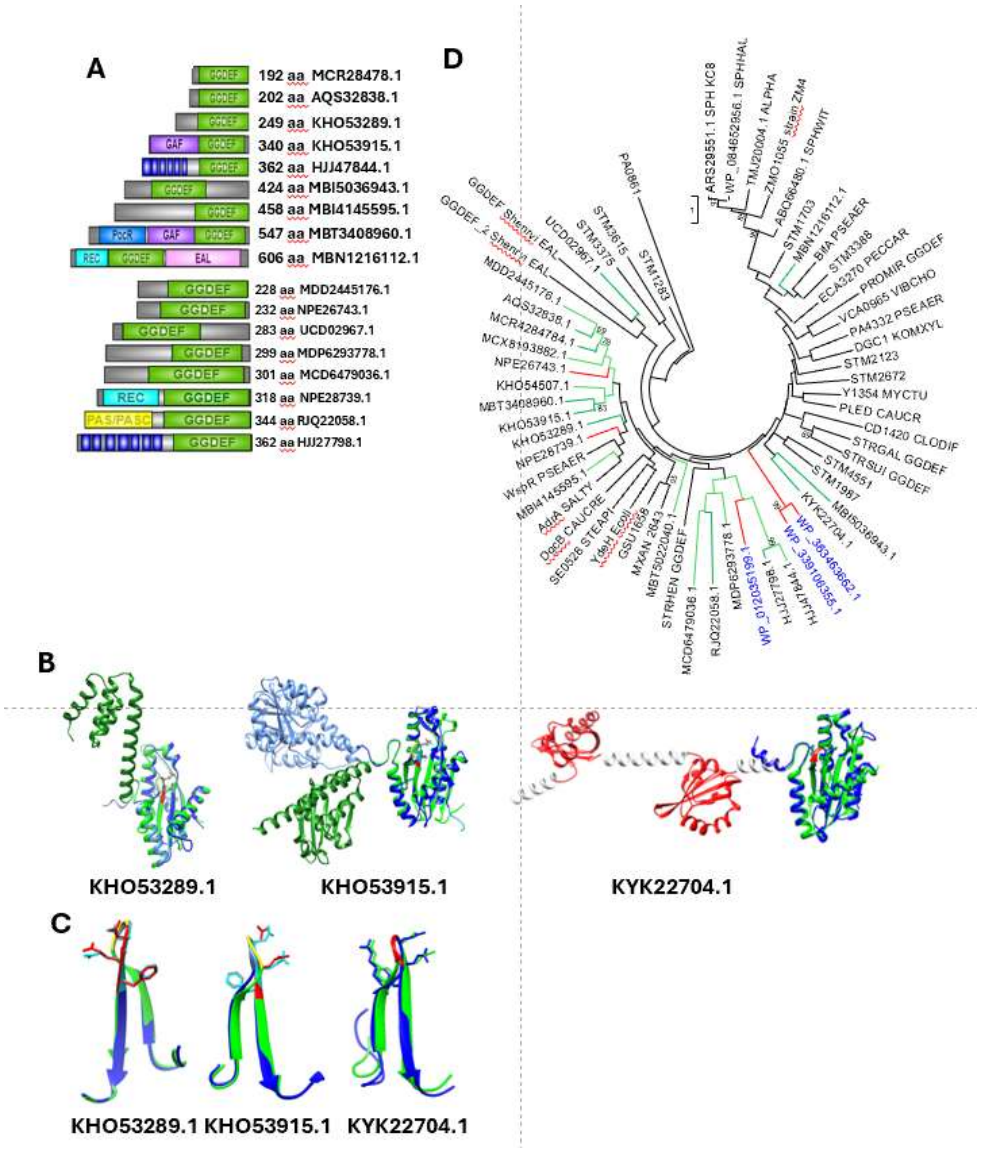


Figure 5

Characterisation of the 20 archaeal GGDEF domain proteins considered for analysis.

Figure 5A. Domain structure of the 20 selected GGDEF domain proteins from archaeal MAGs. Domain structures were grown with IBS 2.0 (Xie et al., 2022). The eight proteins eventually cloned were indicated by an asterix.

Figure 5B. Top, overlays of AlphaFold 3 structural models of GGDEF domain proteins KHO53289.1 and KHO53915.1 with the most similar experimentally determined GGDEF domain structures. Bottom, close up of the beta-hairpin with the GGEEF motif. The beta-hairpin overlays are highly congruent with the experimentally determined structures of *Marinobacter aquaeolei* VT8 (PDB: 3IGN) and the activated diguanylate cyclase PleD of *C. vibrioides* (PDB: 2WB4), respectively.

Figure 5C The GGDEF domains of the 20 archaeal GGDEF domain proteins from MAGs and the three GGDEF domain proteins from archaeal isolates phylogenetically compared to GGDEF domains from experimentally analyzed GGDEF domain proteins. The identity of the GGDEF doamain proteins is indicated in the Supplementary Material. After alignment with ClustalX2 and manual curation in GeneDoc, phylogenetic assessment was performed in MEGA 7.0 using the Maximum Likelihood approach and statistical significance of the phylogenetic tree branches was assessed with 100 bootstraps.

Figure 5D Alignment of the 20 GGDEF domain proteins from archaeal MAGs selected for further analysis in order to assess their catalytic functionality. After alignment with ClustalX2 and manual curation in GeneDoc, the alignment was visualized with ESPript 3.0 (Robert &Gouet, 2014). References were GGDEF domains from PleD and WspR, two well investigated diguanylate cyclases. Other GGDEF domains are from catalytically competent GGDEF domains (blue line), bioinformatically uncertain GGDEF domains (green) and catalytically incompetent GGDEF domains (red).

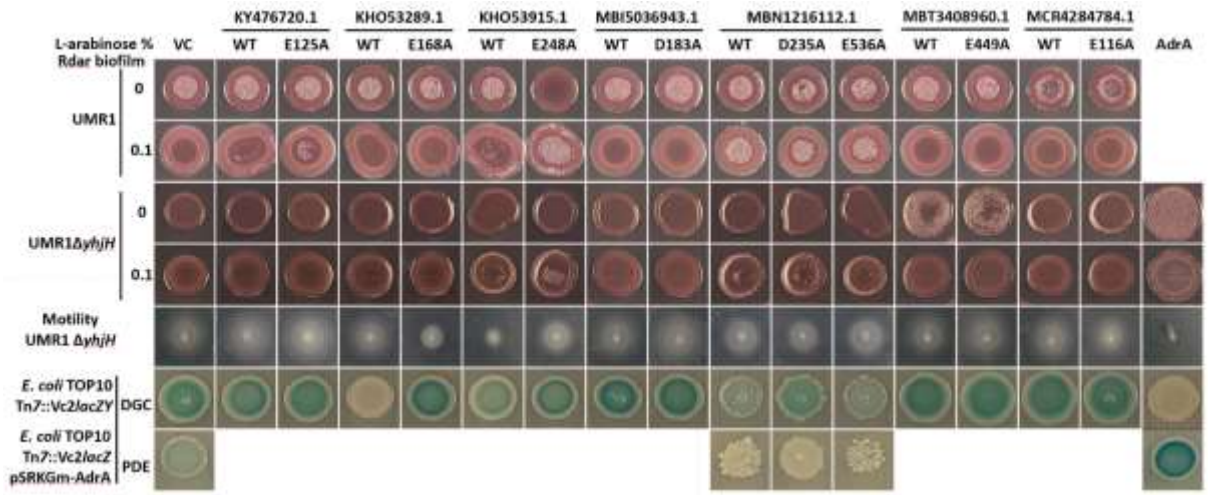


Figure 6

In vivo assays rdar biofilm formation and flagella based swimming motility in the S. typhimurium UMR1 and UMR1 Δ*yhjH* model and Vc2 riboswitch assay in *E. coli* TOP10 Tn*7*::Vc2*lacZY* to assess the catalytic activity of GGDEF domain proteins from metagenome-assembled genomes and its catalytic variants. The genes and variants had been cloned into pBAD28 vector. Assay conditions as described in Figure 3C. Biofilm formation has been assessed without and with 0.1% L-arabinose induction.

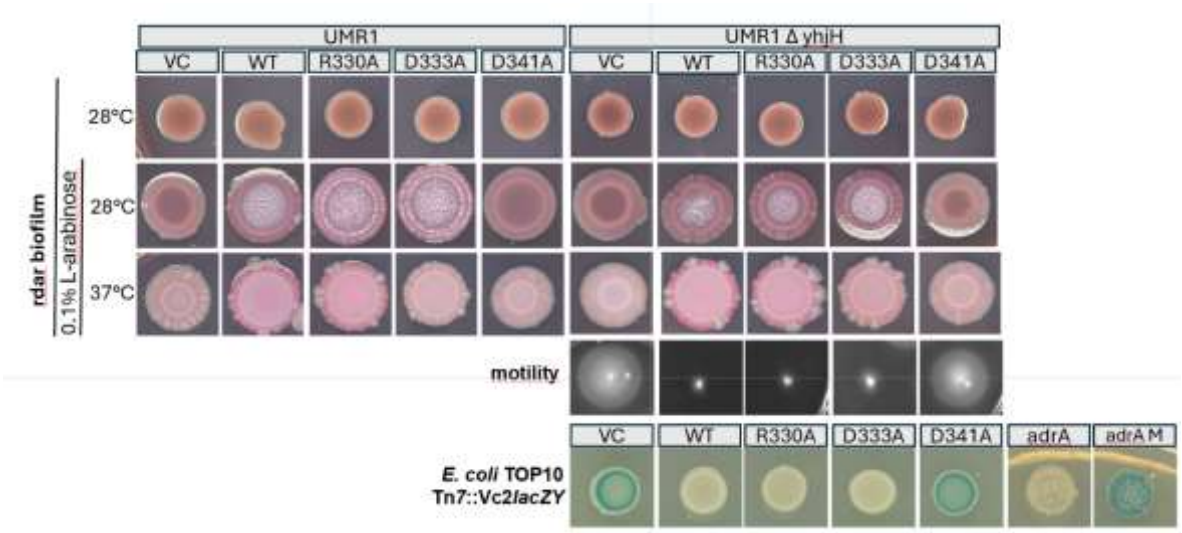


Figure 7

In vivo assays rdar biofilm formation and flagella based swimming motility in the S. typhimurium UMR1 and UMR1 Δ*yhjH* model and Vc2 riboswitch assay in *E. coli* TOP10 Tn*7*::Vc2*lacZY* to assess the catalytic activity of the PAS-PAS-GGDEF domain protein KYK22704.1 from a metagenome-assembled genome, its catalytic variant with the GGAAF motif and alanine substitutions in the RxxD motif which as part of the allosteric inhibitory I-site. Assay conditions as described in Figure 3C. The gene and variants had been cloned into pBAD28 vector. Biofilm formation has been assessed without and with 0.1% L-arabinose induction.

.

Table S1. Strains used in this study

Strain name	Genotype	Reference
***Escherichia coli* strains**		
TOP10	F$^-$ *mcrA* (*mrr*-hsdRMS-*mcrBC*) 80*lacZ* M15 *lacX74 recA1 ara139* (*ara-leu*)*7697 galU galK rpsL* (StrR) *endA1 nupG*	Invitrogen
DH5α	F– ϕ80 *lacZ*ΔM15 Δ(*lacZYA–argF*)U169 *recA1 endA1 hsdR17*(rK–, mK+) *phoA supE44* λ– *thi-1 gyrA96 relA1*	Laboratory collection
XL-1 Blue MR	Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F′ *proAB lacI*^q^ Δ(*lacZ*)M15 Tn10 (*tet*^R^)]	Laboratory collection
TOP10 Vc2Rib	attTn*7*::[Vc2 riboswitch–*lacZY*]	(Liu et al., 2018)

Other bacterial strains		
Salmonella typhimurium UMR1	*S. typhimurium* ATCC 14028s rdar$_{28°C}$, Nal^r	(Romling, Bian et al., 1998)
Salmonella typhimurium UMR1 Δ*yhjH*	Δ*yhjH*	(Simm, Lusch et al., 2007)
Bacillus subtilis 3A1	Wild type	
Archaeal strains		
Haloterrigena salinisoli KCTC4307	Wild type	(Dong et al., 2024)
Methanocella arvoryzae MRE50	Wild type	(Sakai et al., 2010)

Table S2. Plasmids used in this study

Plasmid name	Features	Reference
Empty plasmids		
pBAD28	pACYC184 origin, M13, Amp^r, Cm^r, P_{BAD} promotor, L-arabinose-inducible	(Guzman et al., 1995)
pBAD30	pACYC184 origin, M13, Am^r, P_{BAD} promotor, L-arabinose-inducible	(Guzman et al., 1995)
pSRKGm	pBBR1 origin, Gm^r, P_{lac} promotor, IPTG-inducible	(Khan et al., 2008)
Plasmid constructs containing		
Gene products from archaeal isolates		
pBAD28-WP_339106355.1	Codon optimized, includes alternative start-codon, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag, from *H. salinisoli* KCTC4307	This study
pBAD28-WP_339106355.1 (D59A/E60A)	Substitution in GQGDEF motif	This study
pBAD28-WP_363463662.1	Codon optimized, includes alternative start-codon, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag, from *H. borinquense* E3	This study
pBAD28-WP_363463662.1 (D79A/E80A)	Substitution in GQGDEF motif	This study
pBAD28-WP_012035199.1	Codon optimized, includes alternative start-codon, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag, from *M. arvoryzae* MRE50	This study
pBAD28-WP_012135199.1 (Δ647–end)	Deletion of C-terminal HD-GYP domain	This study
pBAD28-WP_012035199.1 (D551A/E552A)	Substitution in GGDEF motif	This study
pBAD28-WP_012035199.1 (D680A)	Substitution in HD motif	This study
pBAD28-WP_012035199.1 (R454L)	Linker mutation	This study
pBAD28-WP_012035199.1 (E465V)	Linker mutation	This study
pBAD28-WP_012035199.1 (S472V)	Linker mutation	This study
pBAD28-WP_012035199.1 (S472V/D551A/E552A)	Linker mutation and substitution in GGDEF motif	This study
Gene products from metagenome-assembled genomes		
pBAD28-KY476720.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study

pBAD28-KY476720.1 (E125A)	Substitution in GGEEF motif	This study
pBAD28-KHO53289.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28-KHO53289.1 (E168A)	Substitution in GGEEF motif	This study
pBAD28-KHO53915.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28-KHO53915.1 (E248A)	Substitution in GGEEF motif	This study
pBAD28-MBI5036943.1	Original sequence, includes alternative start-codon, native RBS, inserted in XbaI and SphI sites	This study
pBAD28-MBI5036943.1 (D183A)	Substitution in SGDEF motif	This study
pBAD28-MBN1216112.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28-MBN1216112.1 (D235A)	Substitution in GGDEF motif	This study
pBAD28-MBN1216112.1 (E536A)	Substitution in EGI motif	This study
pBAD28-MBT3408960.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28-MBT3408060.1 (E449A)	Substitution in GGEEI motif	This study
pBAD28-MCR4284784.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28-MCR4284784.1 (E116A)	Substitution in GGEEF motif	This study
pBAD28- KYK22704.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28- KYK22704.1 (D341A)	Alanine substitution in the GGDEF motif	This study
KYK22704.1 (D330A)	Alanine substitution in the DxxR I-site motif	This study
KYK22704.1-D341A (D333A)	Alanine substitution in the DxxR I-site motif	This study
pBAD28- HJJ27798.1	Codon optimized, native RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28- HJJ27798.1 (D291A)	Alanine substitution in the GGDEF motif	This study
pBAD28- RJQ22058.1	Altered RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study

pBAD28- RJQ22058.1 (D256A)	Alanine substitution in the GGDEF motif	This study
pBAD28- UCD02967.1	Altered RBS, inserted in XbaI and SphI sites, C-terminal 6xHisTag	This study
pBAD28- UCD02967.1 (D107A)	Alanine substitution in the GGDEF motif	This study
Control plasmid constructs		
pBAD30-*yhjH*	Phosphodiesterase from *S. typhimurium*, EAL domain protein control	(Simm et al., 2004)
pBAD30-*yhjH* (E48A)	Substitution in ELL motif	2009?
pBAD30-*yhjH* (E136A)	Substitution in catalytic residue	(Simm et al., 2004)
pBAD30-*adrA*	Diguanylate cyclase from *S. typhimurium* UMR1, GGDEF domain protein control	(Simm et al., 2004)
pBAD30-*adrA* (G288A/G289A)	Substitution in GGEEF motif	(Simm et al., 2004)
pBAD28-*adrA*	Diguanylate cyclase from *S. typhimurium* UMR1, GGDEF domain protein control	(Li et al., 2019)
pBAD28-*adrA* (E290A/E291A)	Substitutions in GGEEF motif	This study
pBAD28-*dncV*	Dinucleotide cyclase from *E. coli* ECOR31	(Li et al., 2019)
pBAD28-*dncV* (D129A, D131A)	Substitution in catalytic residues	(Li et al., 2019)
pBAD28-*cdaS Bsub*	Diadenylate cyclase from *B. subtilis* 3A1	This study
pBAD28-*cdaS Bsub* (D135A)	Substitution in catalytic residue	This study
pSRKGm-*adrA*	Diguanylate cyclase from *S. typhimurium* UMR1, for phosphodiesterase assay in the *E. coli* TOP10 Tn*7::*Vc2*lacZY*	This study

Table S3. Primers used in this stud

Name	Sequence	Gene
Cloning: vector linearization		
pBAD homology Fd	GACCTGCAGGCATGCAAGCT	pBAD28 vector
pBAD homology Re	GACTCTAGAGGATCCCCGGG	
pBAD_Lin.FOR	AGAGTCGACCTGCAGGC	pBAD28 vector
pBAD_Lin.REV	AGAGGATCCCCGGGTAC	
F-pRSK	GCTAGCAATTCGAAAGCAAATTCG	pSRKGm vector
R-pRSK-ATG	CATATGCTGTTTCCTGTGTGAAATTG	
Cloning: amplification of insert		
Arch_WP_Upstream	GTACCCGGGGATCCTCTAGAGATA	WP_012035199.1 upstream part
Arch_WP_re-Upstream	GCGTACGTCCGCCGGG	
Arch_WP_re-Downstream	TTTCCCCGGCGGACGTA	WP_012035199.1 downstream part
Arch_WP_Downstream	AAACAGCCAAGCTTGCATGCTTAAT	
SFA_XbaI_Fd	GTACCCGGGGATCCTCTAGA	WP_339106355.1
SFA_SphI_Re	AAACAGCCAAGCTTGCATGC	
KY476720_1_Xba_f	ttttttgggctagcgaattcgagctcggtacccggGGATCCTCTAGACCTTGTTTCAGG	KY476720.1
KY476720_1_Sph_r	caggctgaaaatcttctctcatccgccaaaacagccAAGCTTGCATGCTCAGTGGTGGT	
KHO53289_1_Xba_f	ttttttgggctagcgaattcgagctcggtacccggGGATCCTCTAGACTTTAAAGGTAGT	KHO53289.1
KHO53289_1_Sph_r	caggctgaaaatcttctctcatccgccaaaacagccAAGCTTGCATGCTTAGTGGTGATG	
KHO53915_1_XbaI_f	ttttttgggctagcgaattcgagctcggtacccggGGATCCTCTAGATATTATAGGACGT	KHO53915.1
KHO53915_1_Sph_r	caggctgaaaatcttctctcatccgccaaaacagccAAGCTTGCATGCTCAATGATGGTG	
Woese_Fragment.FOR	acccggggatcctctAGACTACAAGCACCACCTGAAG	MBI4145595.1
Woese_Fragment.REV	ctgcaggtcgactctTCAGGCGGTTGCGC	
Microarch_Fragment.FOR	acccggggatcctctTAGCGCAAGAGCTTAAATACTTCTGAA	MBI5036943.1
Microarch_Fragment.Rev	ctgcaggtcgactcTTTAGCAAACTTCGGATTTTATGAGCAAG	
MBN1216112_Xba_up	GCGCTCTAGAAACAGTTTTGGGAAG	MBN1216112.1
down_universal_His_TAA_Hind	GGCAAGCTTAGTGATGGTGATGG	
MBT3408960_1_Xba_f	cccgtttttttgggctagcgaattcgagctcggtacccggGGATCCTCTAGATCAAATAGGAAGT	MBT3408960.1
MBT3408960_1_Sph_r	aggctgaaaatcttctctcatccgccaaaacagccAAGCTTGCATGCTTAATGATGATGG	
MCR4284784_1_Xba_f	ttttttgggctagcgaattcgagctcggtacccggGGATCCTCTAGAGCAGAAAGGATG	MCR4284784.1

MCR4284784_1_Sph_r	ggctgaaaatcttctctcatccgccaaaacagccAAGCTTGCATGCTTAGTGGTGATGA	
IVC-XbaI-ATG-Forward_universal	GTACCCGGGGATCCTCTAGA	HJJ27798.1 RJQ22058.1 UCD02967.1 KYK22704.1
IVC-SphI-Stop-Reverse_universal	CAAAACAGCCAAGCTTGCATGC	
cdaS_Bsub_IVC_Fd_originalRBS_FLgene	cccggggatcctctagagtcAAATTTAAACAGGAGATGAAGGCT	*cdaS* from *Bacillus subtilis* 3A1
cdaS_Bsub_IVC_Re	tgcatgcctgcaggtcttaatggtgatggtgatggtgCGTTCTTGGTGAAATTAACGGAT	
adrA_pSRKGm_IVC_Fd	cacacaggaaacagcatATGTTCCCAAAAATAATGAATGATGAAAATTTTTACCGA	*adrA* from *Salmonella typhimurium* UMR1
adrA_pSRKGm_IVC_Re	ctttcgaattgctagctcagtgatggtgatggtgatgTGCCGCCACTTCGGTGC	
Site-directed mutagenesis		
WP_012035199_1_DE551AA_R	GGAGGTGcgGcgTTTTGCATCCTTCTTCCC	WP_012035199.1
WP_012035199_1_DE551AA_F	GCAAAAcgCcgCACCTCCGGTACGGCAGAC	
WP_012035199_1_D680A_F	TTGCACGcgCTTGGGAAGATCGGG	WP_012035199.1
WP_012035199_1_D680A_R	CCCAAGcgCGTGCAAACGAGCAAG	
WP_012035199_1_R454L_F	AACGAGCtgAACACGACGCTTGAAGAGTTGCA	WP_012035199.1
WP_012035199_1_R454L_R	CGTGTTcaGCTCGTTGATATACTTCCCTTGG	
WP_012035199_1_E465V_F	CGCTCGGtgCGCCGTTTGCAGGAAATGTCCT	WP_012035199.1
WP_012035199_1_E465V_R	ACGGCGcaCCGAGCGCTGCAACTCTTCAAG	
WP_012035199_1_S472V_F	GAAATGgtgTACCGCGATGGCTTGACAGGA	WP_012035199.1
WP_012035199_1_S472V_R	GCGGTacaCCATTTCCTGCAAACGGCGTTCC	
HS_KLK7_QGAAF_f	AGGGTGcgGcGTTTTTGATGGTCTTACCTGATG	WP_339106355.1
HS_KLK7_QGAAF_r	AAAACgCcgCACCCTGTCCGTATTCCCGAT	
HB_DE59AA_Fd	AGGGCGcCGcGTTCCTGCTGATTCTGCCGAATGAAG	WP_363463662.1
HB_DE59AA_Re	AGGAACgCGgCGCCCTGGCCGTATTCACGG	
KY476720_D125A_F	CGCTTTGGTGGCGcGGAATTCTTCATC	KY476720.1
KY476720_D125A_F	CGCGATAATGTCACTCTTACGAA	
KHO53289_G168A_F	GTATCGGCGGTGcGGAATTCTTCATT	KHO53289.1
KHO53289_G168A_R	GGCAAAGAATGTCCGTTCCACGAAA	
KHO53915_D248A_Fd	AGAGTTTATCATTATCTTACCCGAGACCGATAATGAGA	KHO53915.1
KHO53915_D248A_Re	gCGCCACCATAGCGACCAATAATATCATT	
MA_GAAF_F	CTCAGGGGcCGcGTTTATCGTGAT	MBI5036943.1
MA_GGAEF_R	GGATGAAATATCCTGACATGCTC	
MBN1216112_D235A_F	CCTGGGGGGCgcgGAATTCGTGATTT	MBN1216112.1
MBN1216112_D235A_Rev	CGAGCCACGACGTCG	
MBN1216112_E536A_F	GTTATCTGCGcAGGAATTGAGAAC	MBN1216112.1
MBN1216112_E536A_Rev	ACTTAAATTCAGATTATGGCCCA	

MBT_3408960_D449A_Forw	ACGGCGGTGcAGAGATTATTGTCTT	MBT3408960.1
MBT_3408960_D449A_rev	AGCGTCCAATGATATCGATCTCGC	
MCR4284784_D116A_Forw	CTTTGGAGGGGcAGAGTTCCTTGTG	MCR84784.1
MCR4284784_D116A_Rev	CGGGCCAAAACGTCATACTTGCGG	
KYK22704.1-D341A-fw	GGCGGT**GCG**GAGTTTATTATTGTGTGGAACAATAC	KYK22704.1 (GGDEF-GGAEF)
KYK22704.1-D341A-rev	AAACTC**CGC**ACCGCCCAGACGGAAGAAAAT	
KYK22704.1-R330A-fw	AACCTG**GCG**GAAAGCGATATTTTCTTCCGTCTG	KYK22704.1 (RxxD-AxxD)
KYK22704.1-R330A-rev	GCTTTC**CGC**CAGGTTATCTTTGATCAGGCTTGC	
KYK22704.1-D333A-fw	GAAAGC**GCG**ATTTTCTTCCGTCTGGGCGGTGA	KYK22704.1 (RxxD<RxxA)
KYK22704.1-D333A-rev	GAAAAT**CGC**GCTTTCACGCAGGTTATCTTTG	
HJJ27798_1_D291A_fw	gtggcg**CG**gaatttgttgtactgctgcacgatt	HJJ27798.1
HJJ27798_1_D291A_rev	caaattcCGcgccaccgatacggaataccg	
RJQ22058_1_D256A_fw	gcggg**g**CGgagtttcgcgtggtgactgatgc	RJQ22058.1
RJQ22058_1_D256A_rv	gaaactcCGccccgccaatgcggtacacct	
UCD02967_1_D107 A_fw	aagga**g**CGgagttccttgccggaattctcgg	UCD02967.1
UCD02967_1_D107 A_rv	ggaactcCGctcctttgtctccaccataccttc	
adrA-GGAAF-fw	CGGC**GCGGCG**TTTGCGGTGATTATGTGCGGAAC	AdrA
adrA-GGAAF-rev	GCAAA**CGCCGC**GCCGCCAAAGCGCCCGATAA	
cdaS_Bsub_D135A_IVC_Fd	AGCGCCGgCATGAAGAGGGTTTCCGGGAAA	*cdaS* from *Bacillus subtilis* 3A1
cdaS_Bsub_D135A_IVC_Re	CTTCATGcCGGCGCTTTATTAGTGAGGGAAAACAAACTC	
Deletion of C-terminal domain		
WP_IVC_del_Fd	gcgattatatggatcacggtCATCACCACCACCATCATTAAGCATGC	WP_012035199.1
WP_IVC_del_Re	ATCGACAATCACCGTCAGCACA	
Screening and sequencing		
F-pBAD-screen	GTCTATAATCACGGCAGAAAAGTCCAC	pBAD28
R-pBAD-screen	CTGTTTTATCAGACCGCTTCTGC	
WP_012035199_1_InternalSeq	TGAATCAACGTTCTAACGTGACG	WP_012035199.1
F-screen-pRSK	CTATTTAACGACCCTGCCCT	pSRKGm
R-screen-pRSK	GCGGATAACAATTTCACACAGG	

Supplementary Information

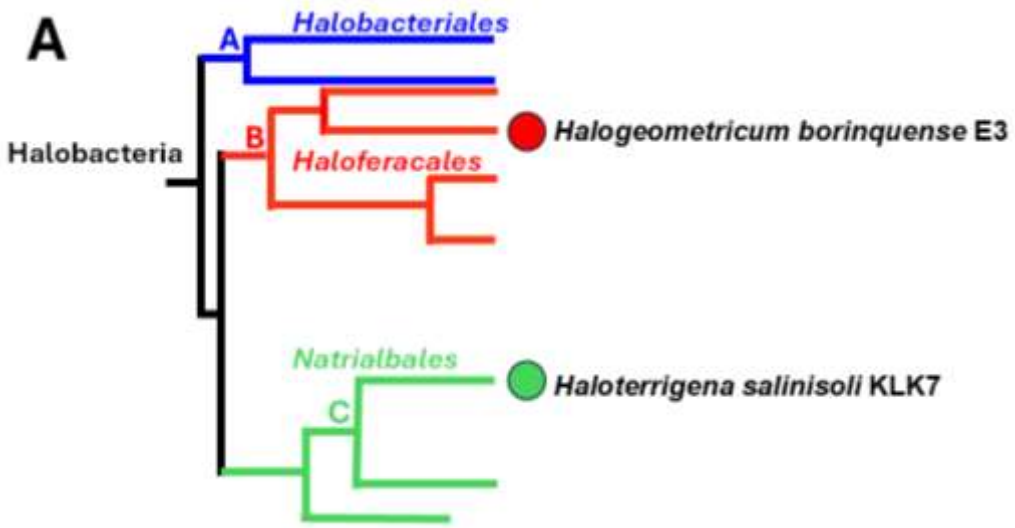
Archaeal genes code for GGDEF domain proteins with diguanylate cyclase activity

Roman Sidorov[1,*], Li Li[1,*], Ali Dadvar[1], Irfan Ahmad[1], Stefan Spring[2], Jörg Overmann[2,3], Ute Römling[1]
[1]Department of Microbiology, Tumor and Cell Biology, Biomedicum, Karolinska Institutet, Stockholm, Sweden
[2]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany
[3]Braunschweig University of Technology, Spielmannstraße 7, 38106 Braunschweig, Germany

*contributed equally

Corresponding author:
Ute Römling, ute.romling@ki.se

Supplementary Figures



Supplementary Figure 1 Characterisation of archaeal isolates

Figure S1A. Phylogenetic tree of Halobacteria with the position of GGDEF domain proteins WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3. Phylogenetic tree of Halobacteria has been redrawn from (Blanquart, Groussin et al., 2021).

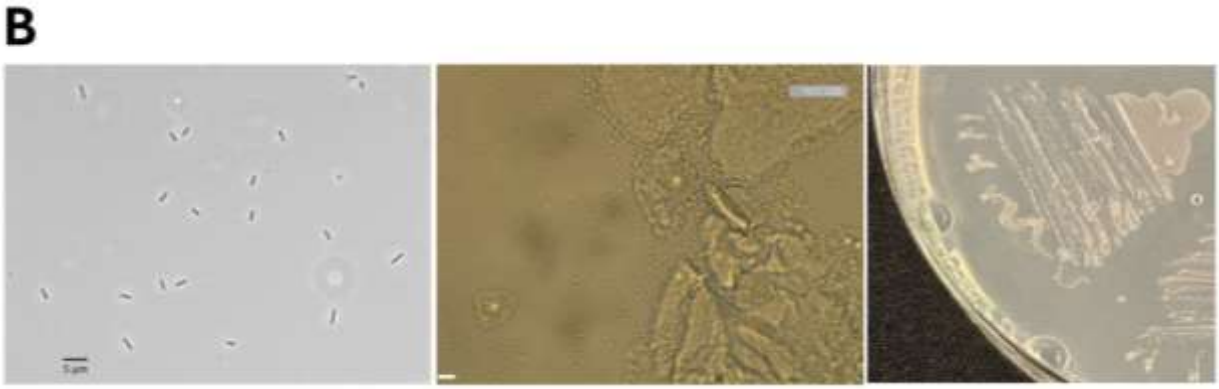
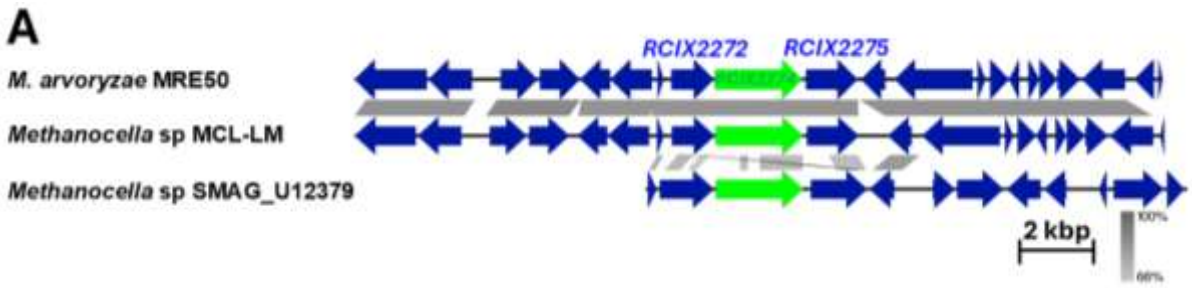


Figure S1B Micrographs and plate-grown archaeal isolates. Left, *M. arvoryzae* MRE50 grown in DSM medium 1318 under anaerobic conditions at 45°C for seven days. Middle, H. *salinisoli* KLK7 grown in KCTC medium No. 1762 HCM. Right, *H. salinisoli* KLK7 grown on KCTC medium No. 1762 HCM medium for 14 days as 37 °C.



Supplementary Figure 2 Genomic localization of the GGDEF domain proteins from isolates.

Figure S2A. Chromosomal location of the Pac-PAS/PAC-PocR-GGDEF-HD-GYP domain protein genes from *Methanocella* spp. Three genomes of isolates belonging to the *Methanocella* genus encode the Rec-PAS/PAC-PocR-GGDEF-HD-GYP domain protein gene. In all three isolates the Rec-PAS/PAC-PocR-GGDEF-HD-GYP domain protein gene *RCIX2274* is present at the same chromosomal location flanked by HEAT repeat domain-containing protein genes *RCIX2272* and *RCIX2275*. Of note, the three *Methanocella* isolates are of distinct geographical origin. While *M. arvoryzae* MRE50 has been isolated at the Italian Rice Research Institute, Vercelli, Italy, *Methanocella* sp. MCL-LM from a biocrust at Midreshet Ben-Gurion, Israel and *Methanocella* sp. SMAG_U12379 from a soil metagenome, Colombia.

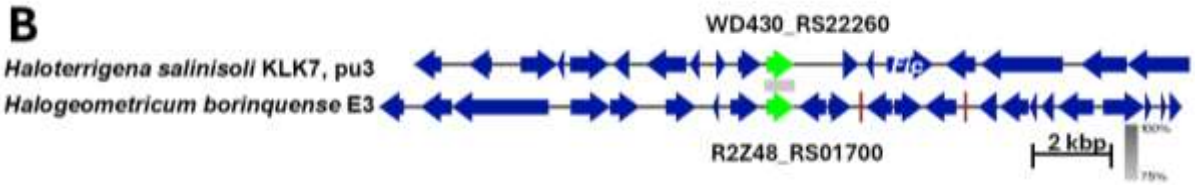
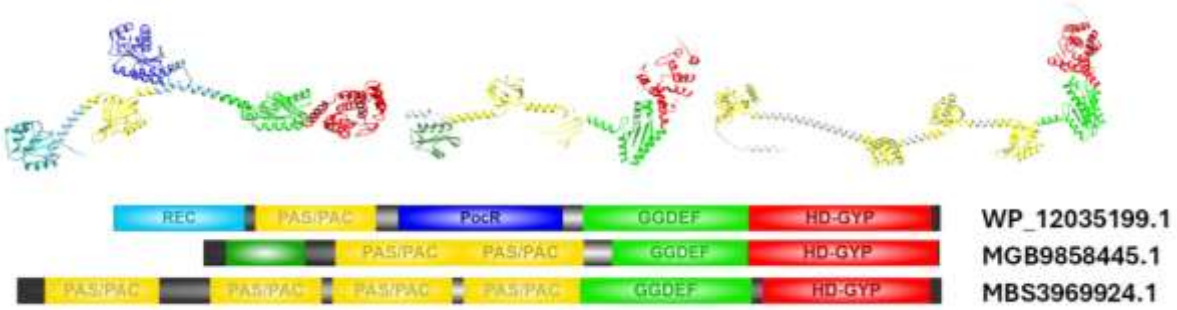


Figure S2B. Genome location of the GGDEF-unknown domain protein genes from *H. salinisoli* KLK7 and *H. borinquense* E3. The genome location of the two proteins is distinct. WD430_RS22260 encoding WP_339106355.1 is located on the plasmid pu3 of *H. salinisoli* KLK7, while R2Z48_RS01700 encoding WP_363463662.1 of *H. borinquense* E3 is located on the chromosome.



Supplementary Figure 3 Structural and phylogenetic analysis of GGDEF domains of GGDEF domains of proteins WP_012035199.1 from *M. arvoryzae* MRE50, WP_339106355.1 from *H. salinisoli* KLK7 and WP_363463662.1 from *H. borinquense* E3.

Figure S3A. Structural models and domain structures of bacterial proteins MGB9858445.1 and MBS3969924.1 which possess a GGDEF domain most similar to WP_012035199.1 from *M. arvoryzae* MRE50. Structural models were created with AlphaFold 3 (Jumper et al., 2021) and processed in Chimera 1.18 (Pettersen et al., 2004). Domain structure of proteins were constructed in IBS 2.0 (Xie et al., 2022).

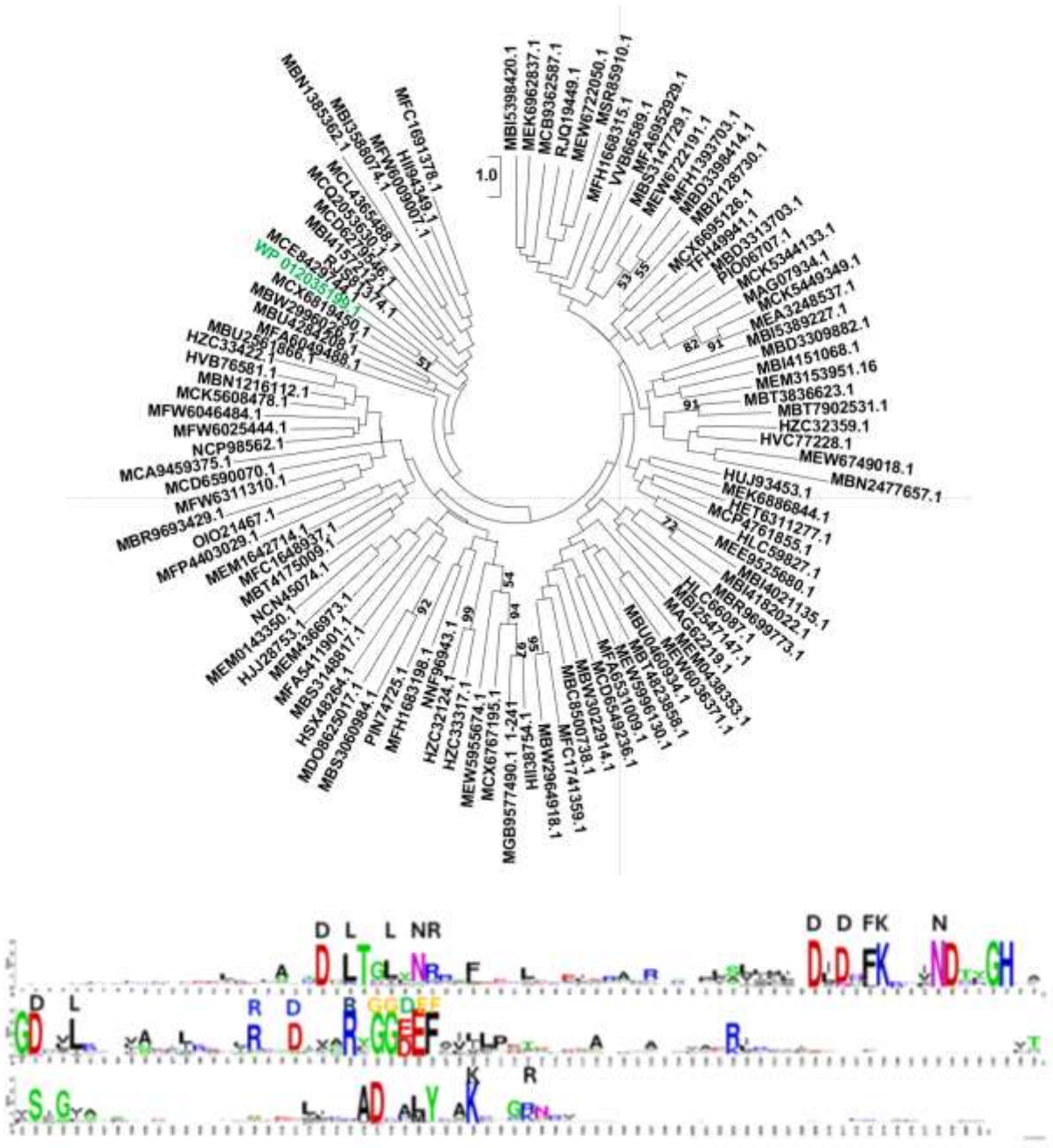


Figure S3B. Phylogenetic tree of archaeal GGDEF domain proteins with the GGDEF domains most similar to the GGDEF domain of WP_012035199.1 from *M. arvoryzae* MRE50. Proteins were retrieved after BLAST search using the GGDEF domain of WP_012035199.1 as a query. After alignment with ClustalX2 and manual curation in GeneDoc, phylogenetic assessment was performed in MEGA 7.0 using the Maximum Likelihood approach and statistical significance of the tree was assessed with 100 bootstraps. Of 500 domains, domains with <70% identity are displayed.
WebLogo (Crooks, Hon et al., 2004) of 500 most similar archaeal GGDEF domains with indicatings of signautr amino acids required for catalysis.

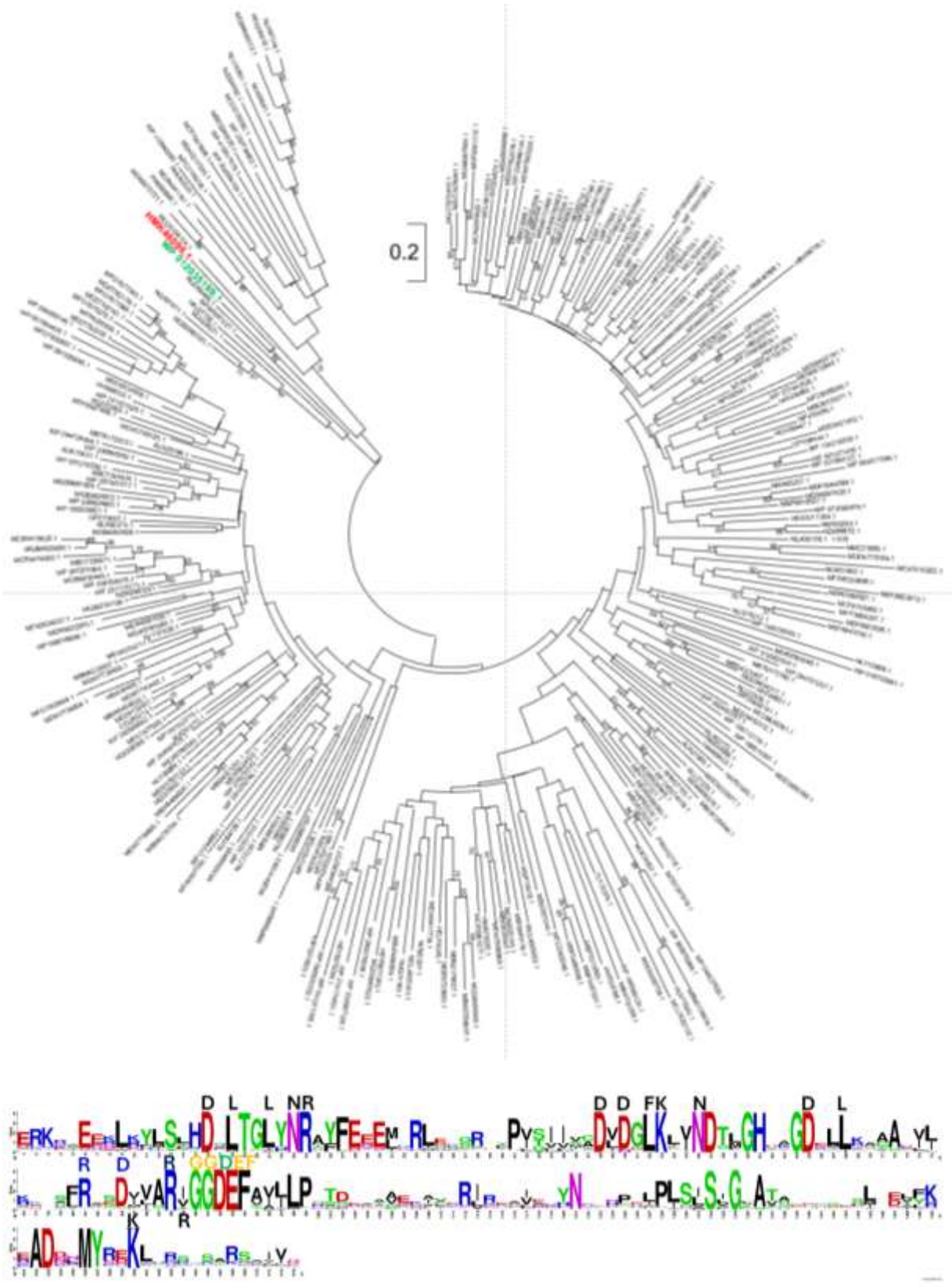


Figure S3C. Phylogenetic tree of all GGDEF domain proteins with the GGDEF domains most similar to the GGDEF domain of WP_012035199.1 from *M. arvoryzae* MRE50. Proteins were retrieved after BLAST search using the GGDEF domain of WP_012035199.1 as a query. Of 500 domains, domains with <90% identity are displayed.

WebLogo (Crooks et al., 2004) of 500 most similar GGDEF domains from the bacteriala dn archaealdomain with indicating signature amino acids required for catalysis.

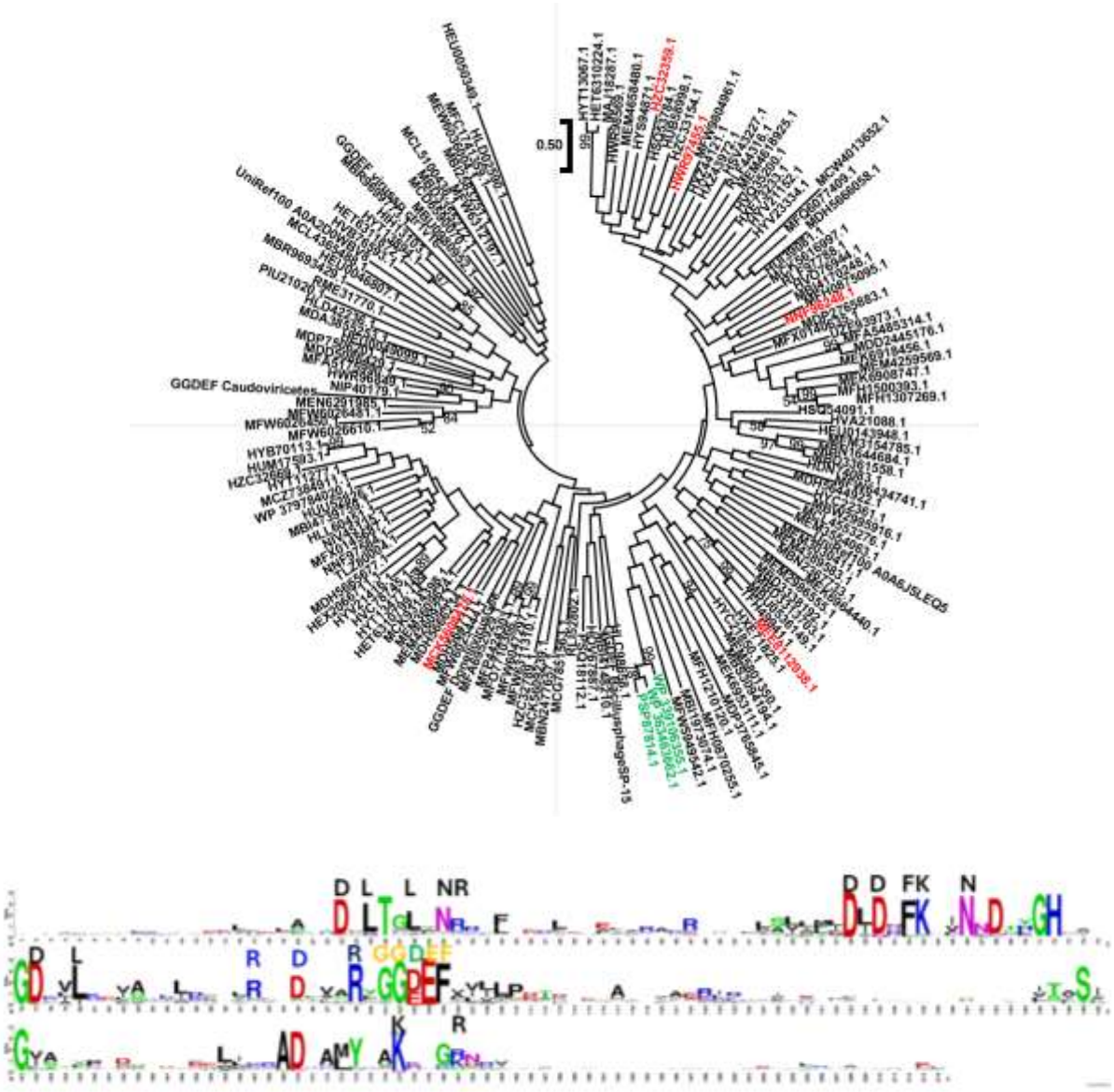


Figure S3D. Phylogenetic tree of archaeal GGDEF domain proteins with the GGDEF domains most similar to the GGDEF domain of WP_339106355.1 from *H. salinisoli* KLK7 . Proteins were retrieved after BLAST search using the GGDEF domain of WP_339106355.1 as a query. Of 500 domains, domains with <90% identity are displayed.
WebLogo (Crooks et al., 2004) of 500 most similar archaeal GGDEF domains indicating signature amino acids required for catalysis.

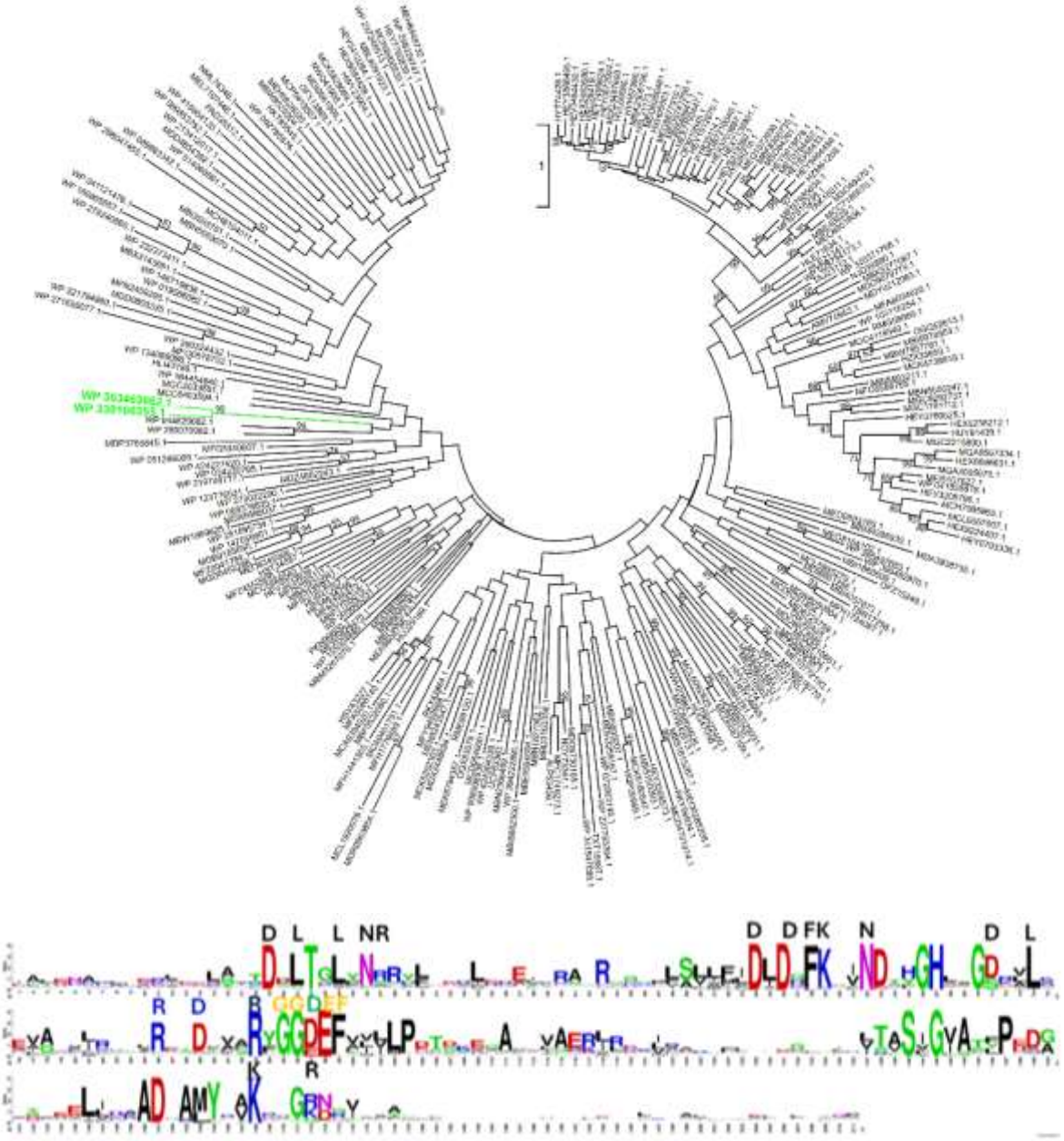
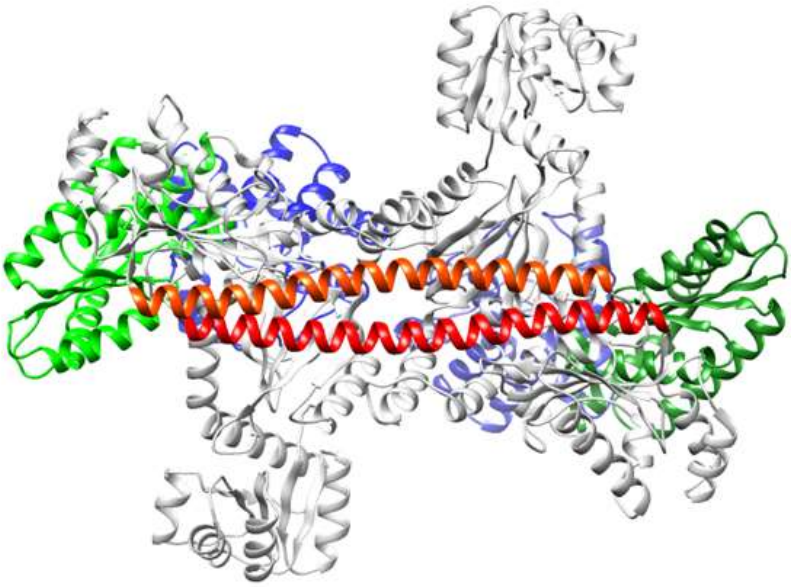
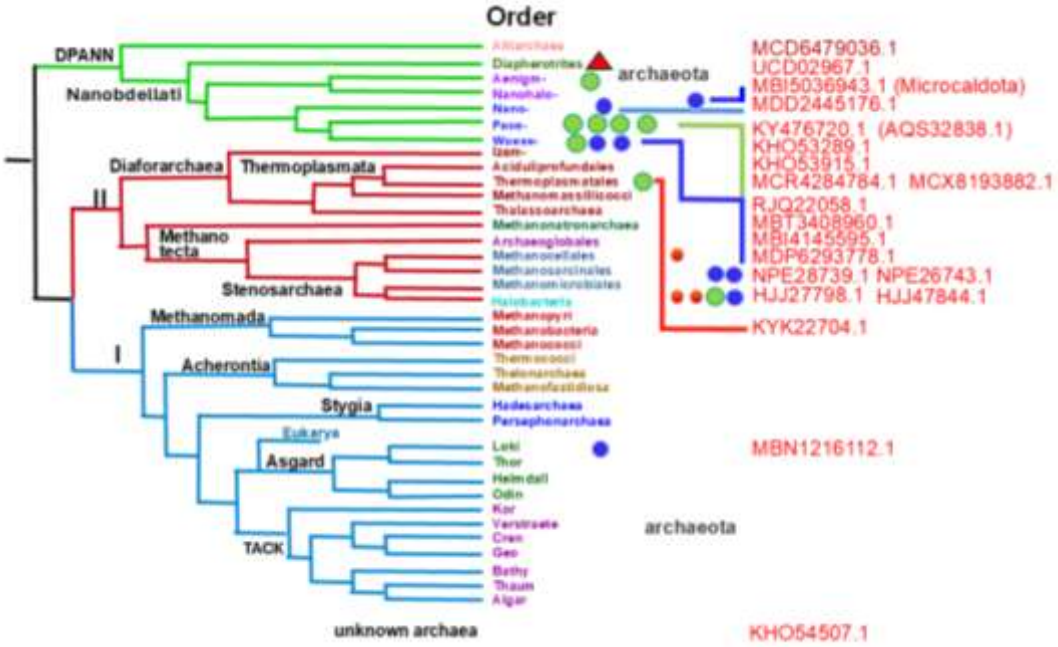


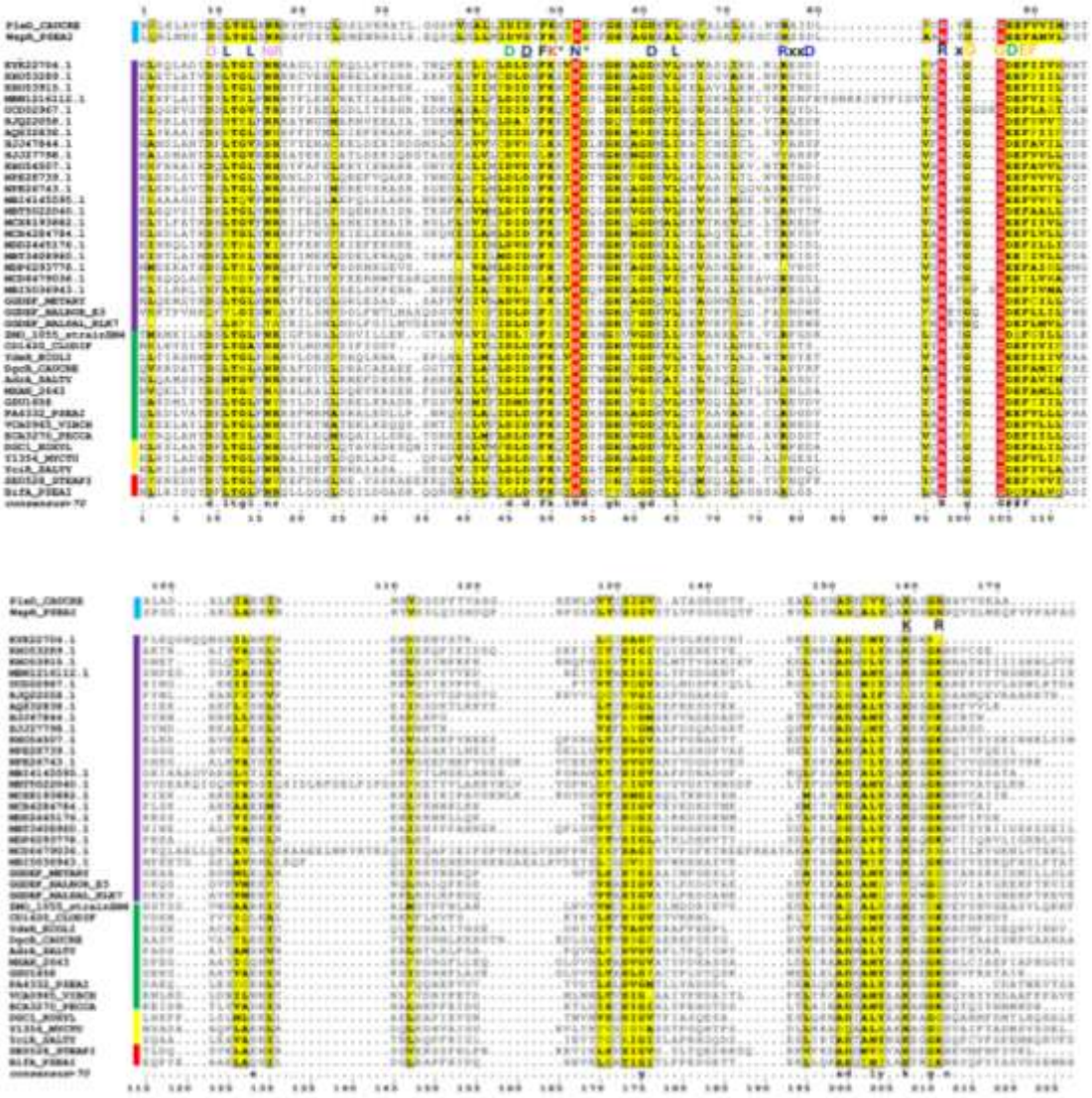
Figure S3E. Phylogenetic tree of all GGDEF domain proteins with the GGDEF domains most similar to the GGDEF domain of WP_339106355.1 from *H. salinisoli* KLK7 with domains <90% identity displayed. Proteins were retrieved after BLAST search using the GGDEF domain of WP_339106355.1 as a query. WebLogo (Crooks et al., 2004) of 500 most similar GGDEF domains from the bacterial and archaeal domain indicating signature amino acids required for catalysis.



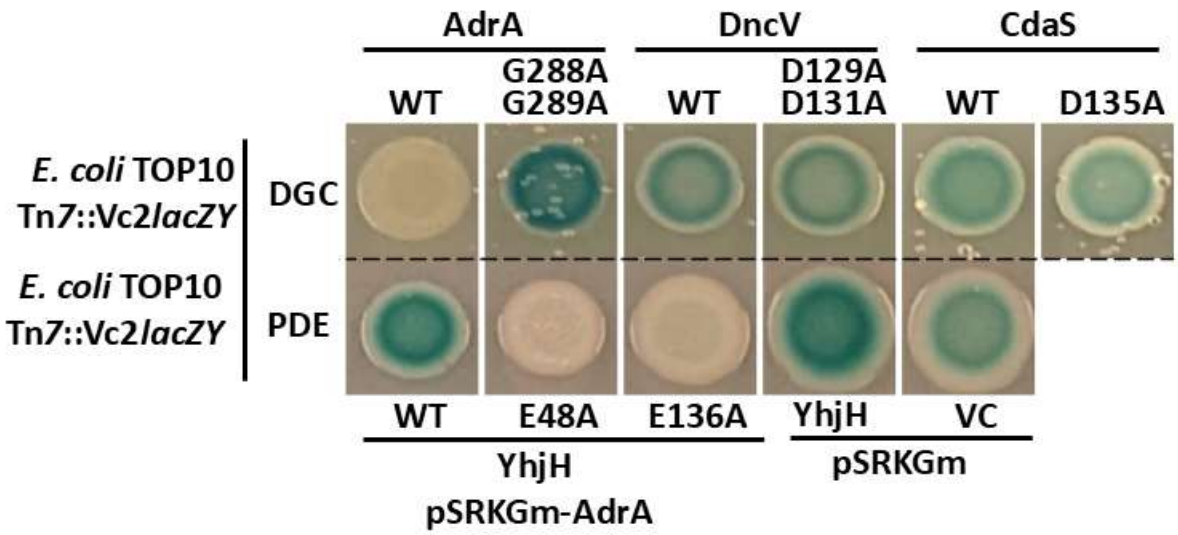
Supplementary Figure 4 AlphaFold 3 structural model of dimer formation of the wild type full length protein WP_012035199.1 from *M. arvoryzae* MRE50. The monomers and their GGDEF domains (in green) are arranged oppositely preventing a catalytic competent arrangement. Blue, HD-GYP domain; red, linker.



Supplementary Figure 5 Phylogenetic tree of Archaea with phylogenetic position of 16 additional GGDEF domain proteins identified in archaeal MAGs. Phylogenetic tree has been redrawn from (Aouad et al., 2022).



Supplementary Figure 6. Alignment of GGDEF domain of archaeal GGDEF domain proteins from MAGs with GGDEF domains from well investigated catalytically active GGDEF domains from the PleD from *Caulobacter vibrioides* CB15 (CC_2462, PleD_CAUCRE) and WspR from *Pseudomonas aeruginosa* PAO1 (PA3702, WspR_PSEAI). Green bar, catalytically active GGDEF domains; yellow bar, domains with unconventional signature motifs; red, catalytically inactive GGDEF domains. Protein designation in supplementary material. Alignment was created with ClustalX2 (Higgins &Sharp, 1988), manually curated in GeneDoc and visualized with ESPript 3.0 (Robert &Gouet, 2014). Secondary structure of PleD from 2V0N (RCSB Protein Data Bank, RCSB PDB).



Supplementary Figure 7 Assessment of catalytic activity of archaeal GGDEF domain proteins from MAGs.

Figure S7A *In vivo* Vc2 riboswitch assay in *E. coli* TOP10 Tn*7*::Vc2*lacZY* to assess catalytic activity of the controls for the diguanylate cyclase (DGC) and cyclic di-GMP specific phosphodiesterase (PDE) assay. Cyclic di-nucleotide controls for cyclic di-GMP were *E. coli* TOP10 Tn*7*::Vc2*lacZY* pAdrA and the pAdrA G288A, G289A mutant; pDncV and its catalytic pDncV D129A, D131A mutant and pCdaS and its catalytic pCdaS D135A mutant. The PDE assay expressed pSRKGm-AdrA per default without IPTG induction with PDE activity to be observed by relieve of repression of translation. VC is pBAD28.

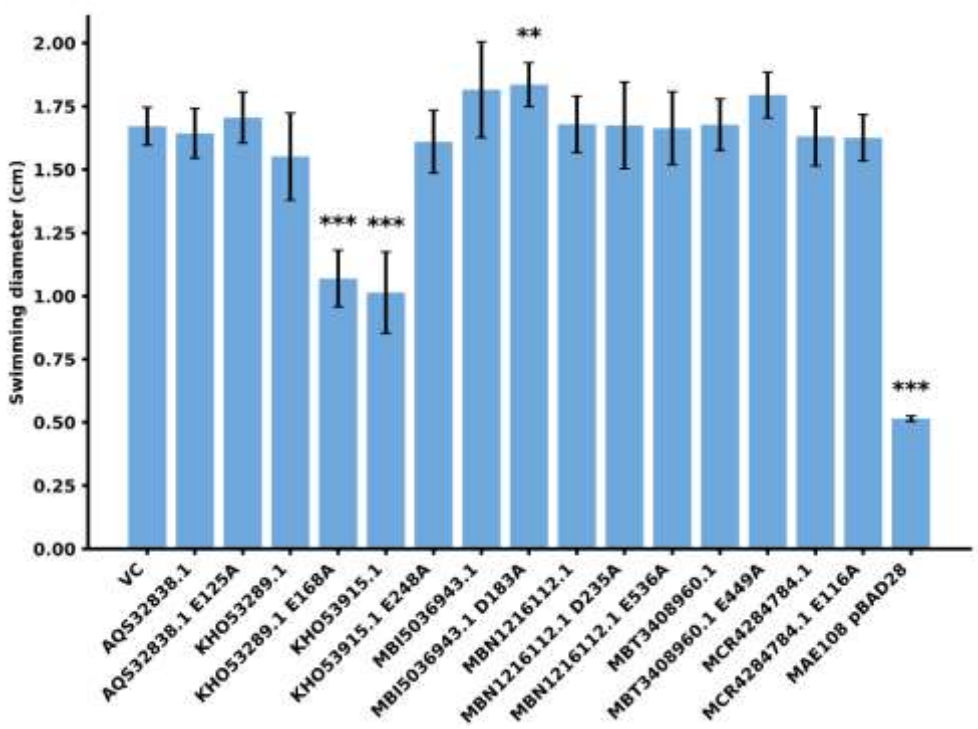
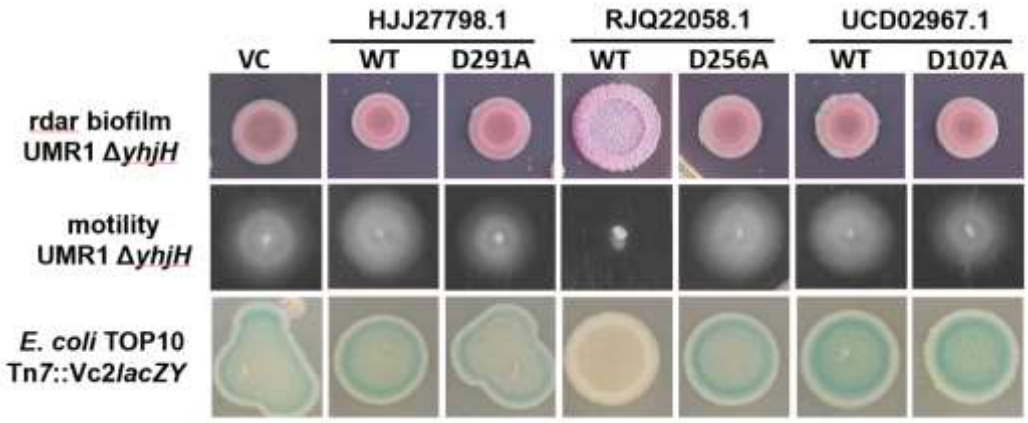


Figure S7B Quantitative estimation of flagella based swimming motility to assess the catalytic activity of MAGs derived GGDEF domain proteins and its variants in *S. typhimurium* UMR1 Δ*yhjH*. p*<0.05, p**<0.01, p***<0.001 by student's t-test compared to *S. typhimurium* UMR1 Δ*yhH* vector control pBAD28 and negative control *S. typhimurium* MAE108 (Δ*fliC* Δ*fljB*) pBAD28.



Supplementary Figure 8 *In vivo* assays rdar biofilm formation and flagella based swimming motility in the *S. typhimurium* UMR1 and UMR1 Δ*yhjH* model and Vc2 riboswitch assay in *E. coli* TOP10 Tn*7*::Vc2*lacZY* to assess the catalytic activity of GGDEF domain proteins UCD02967.1, HJJ27798.1 and RJQ22058.1 from metagenome-assembled genomes and its catalytic variants. Assay conditions as described in Figure 3A. Genes and variants have been cloned into the pBAD28 vector. Induction of protein production with 0.1% L-arabinose.

References

Ahmad, I., Wigren, E., Le Guyon, S., Vekkeli, S., Blanka, A., El Mouali, Y. et al. (2013) The eal-like protein stm1697 regulates virulence phenotypes, motility and biofilm formation in salmonella typhimurium. *Mol Microbiol* **90**: 1216-1232.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Anantharaman, V. & Aravind, L. (2005) Meds and pocr are novel domains with a predicted role in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. *Bioinformatics* **21**: 2805-2811.

Aouad, M., Flandrois, J.P., Jauffrit, F., Gouy, M., Gribaldo, S. & Brochier-Armanet, C. (2022) A divide-and-conquer phylogenomic approach based on character supermatrices resolves early steps in the evolution of the archaea. *BMC Ecol Evol* **22**: 1.

Baral, R., Ho, K., Kumar, R.P., Hopkins, J.B., Watkins, M.B., LaRussa, S. et al. (2025) A general mechanism for initiating the bacterial general stress response. *Elife* **13**:

Blanquart, S., Groussin, M., Le Roy, A., Szollosi, G.J., Girard, E., Franzetti, B. et al. (2021) Resurrection of ancestral malate dehydrogenases reveals the evolutionary history of halobacterial proteins: Deciphering gene trajectories and changes in biochemical properties. *Mol Biol Evol* **38**: 3754-3774.

Botsford, J.L. & Harman, J.G. (1992) Cyclic amp in prokaryotes. *Microbiol Rev* **56**: 100-122.

Bubeck, P., Winkler, M. & Bautsch, W. (1993) Rapid cloning by homologous recombination in vivo. *Nucleic Acids Res* **21**: 3601-3602.

Cao L.-Y., Z.X., Bai F.-W., Römling U. (2026) Previously uncharacterized aliphatic amino acid positions modulate the apparent catalytic activity of the eal domain of zmo_1055 and other cyclic di-gmp specific eal phosphodiesterases. *Microb Biotechnol* **accepted**:

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P. et al. (2006) Allosteric control of cyclic di-gmp signaling. *J Biol Chem* **281**: 32015-32024.

Corrigan, R.M., Campeotto, I., Jeganathan, T., Roelofs, K.G., Lee, V.T. & Grundling, A. (2013) Systematic identification of conserved bacterial c-di-amp receptor proteins. *Proc Natl Acad Sci U S A* **110**: 9084-9089.

Corrigan, R.M. & Grundling, A. (2013) Cyclic di-amp: Another second messenger enters the fray. *Nat Rev Microbiol* **11**: 513-524.

Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004) Weblogo: A sequence logo generator. *Genome Res* **14**: 1188-1190.

Dahlstrom, K.M., Giglio, K.M., Collins, A.J., Sondermann, H. & O'Toole, G.A. (2015) Contribution of physical interactions to signaling specificity between a diguanylate cyclase and its effector. *mBio* **6**: e01978-01915.

Dayton, H., Smiley, M.K., Forouhar, F., Harrison, J.J., Price-Whelan, A., Dietrich, L.E.P. (2020) Sensory domains that control cyclic di-gmp-modulating proteins: A critical frontier in bacterial signal transduction. In *Microbial cyclic di-nucleotide signaling*, Chou, S.-H., Guiliani, N., Lee, V.T., Römling, U. (ed) pp 137-158. Cham: Springer Cham

De, N., Navarro, M.V., Raghavan, R.V. & Sondermann, H. (2009) Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator wspr. *J Mol Biol* **393**: 619-633.

Dong, X.Y., Mao, Y.L., Zhang, Q.K., Zhu, L.R., Hou, J. & Cui, H.L. (2024) Genome-based classification of the family natrialbaceae and description of four novel halophilic archaea from three saline lakes and a saline-alkaline land. *Extremophiles* **28**: 47.

Duncan-Lowey, B., McNamara-Bordewick, N.K., Tal, N., Sorek, R. & Kranzusch, P.J. (2021) Effector-mediated membrane disruption controls cell death in cbass antiphage defense. *Mol Cell* **81**: 5039-5051 e5035.

El Mouali, Y., Kim, H., Ahmad, I., Brauner, A., Liu, Y., Skurnik, M. et al. (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**:

Finstad, K.M., Probst, A.J., Thomas, B.C., Andersen, G.L., Demergasso, C., Echeverria, A. et al. (2017) Microbial community structure and the persistence of cyanobacterial populations in salt crusts of the hyperarid atacama desert from genome-resolved metagenomics. *Front Microbiol* **8**: 1435.
Flemming, H.C. & Wuertz, S. (2019) Bacteria and archaea on earth and their abundance in biofilms. *Nat Rev Microbiol* **17**: 247-260.
Galperin, M.Y. & Chou, S.H. (2022) Sequence conservation, domain architectures, and phylogenetic distribution of the hd-gyp type c-di-gmp phosphodiesterases. *J Bacteriol* **204**: e0056121.
Giardina, G., Paiardini, A., Fernicola, S., Franceschini, S., Rinaldo, S., Stelitano, V. et al. (2013) Investigating the allosteric regulation of yfin from pseudomonas aeruginosa: Clues from the structure of the catalytic domain. *PLoS One* **8**: e81324.
Gophna, U. & Altman-Price, N. (2022) Horizontal gene transfer in archaea-from mechanisms to genome evolution. *Annu Rev Microbiol* **76**: 481-502.
Gourinchas, G., Etzl, S., Gobl, C., Vide, U., Madl, T. & Winkler, A. (2017) Long-range allosteric signaling in red light-regulated diguanylyl cyclases. *Sci Adv* **3**: e1602498.
Grantcharova, N., Peters, V., Monteiro, C., Zakikhany, K. & Romling, U. (2010) Bistable expression of csgd in biofilm development of salmonella enterica serovar typhimurium. *J Bacteriol* **192**: 456-466.
Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose pbad promoter. *J Bacteriol* **177**: 4121-4130.
Hallberg, Z.F., Chan, C.H., Wright, T.A., Kranzusch, P.J., Doxzen, K.W., Park, J.J. et al. (2019) Structure and mechanism of a hypr ggdef enzyme that activates cgamp signaling to control extracellular metal respiration. *Elife* **8**:
Higgins, D.G. & Sharp, P.M. (1988) Clustal: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244.
Holm, L., Laiho, A., Toronen, P. & Salgado, M. (2023) Dali shines a light on remote homologs: One hundred discoveries. *Protein Sci* **32**: e4519.
Hu, X.M., Peng, L., Wang, Y., Ma, F., Tao, Y., Liang, X. et al. (2024) Bacterial c-di-gmp triggers metamorphosis of mussel larvae through a sting receptor. *NPJ Biofilms Microbiomes* **10**: 51.
Jones, D.H. & Howard, B.H. (1991) A rapid method for recombination and site-specific mutagenesis by placing homologous ends on DNA using polymerase chain reaction. *Biotechniques* **10**: 62-66.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with alphafold. *Nature* **596**: 583-589.
Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. (2015) The phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845-858.
Kellner, S., Spang, A., Offre, P., Szollosi, G.J., Petitjean, C. & Williams, T.A. (2018) Genome size evolution in the archaea. *Emerg Top Life Sci* **2**: 595-605.
Khan, S.R., Gaines, J., Roop, R.M., 2nd & Farrand, S.K. (2008) Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of trar and tram expression on ti plasmid quorum sensing. *Appl Environ Microbiol* **74**: 5053-5062.
Kulasekara, B.R., Kamischke, C., Kulasekara, H.D., Christen, M., Wiggins, P.A. & Miller, S.I. (2013) C-di-gmp heterogeneity is generated by the chemotaxis machinery to regulate flagellar motility. *Elife* **2**: e01402.
Kumar, S., Stecher, G. & Tamura, K. (2016) Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**: 1870-1874.
Lai, T.H., Kumagai, Y., Hyodo, M., Hayakawa, Y. & Rikihisa, Y. (2009) The anaplasma phagocytophilum plec histidine kinase and pled diguanylate cyclase two-component system and role of cyclic di-gmp in host cell infection. *J Bacteriol* **191**: 693-700.
Leichtling, B.H., Rickenberg, H.V., Seely, R.J., Fahrney, D.E. & Pace, N.R. (1986) The occurrence of cyclic amp in archaebacteria. *Biochem Biophys Res Commun* **136**: 1078-1082.
Li, F., Cimdins, A., Rohde, M., Jansch, L., Kaever, V., Nimtz, M. et al. (2019) Dncv synthesizes cyclic gmp-amp and regulates biofilm formation and motility in escherichia coli ecor31. *mBio* **10**:
Liu, Y., Kim, H. & Romling, U. (2018) In vivo analysis of cyclic di-gmp cyclase and phosphodiesterase activity in escherichia coli using a vc2 riboswitch-based assay. *Bio Protoc* **8**: e2753.

Liu, Y., Lee, C., Li, F., Trcek, J., Bahre, H., Guo, R.T. et al. (2020) A cyclic di-gmp network is present in gram-positive streptococcus and gram-negative proteus species. *ACS Infect Dis* **6**: 2672-2687.
Lori, C., Ozaki, S., Steiner, S., Bohm, R., Abel, S., Dubey, B.N. et al. (2015) Cyclic di-gmp acts as a cell cycle oscillator to drive chromosome replication. *Nature* **523**: 236-239.
Martin, W. & Russell, M.J. (2003) On the origins of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* **358**: 59-83; discussion 83-55.
McDonough, K.A. & Rodriguez, A. (2011) The myriad roles of cyclic amp in microbial pathogens: From signal to sword. *Nat Rev Microbiol* **10**: 27-38.
Mehne, F.M., Schroder-Tittmann, K., Eijlander, R.T., Herzberg, C., Hewitt, L., Kaever, V. et al. (2014) Control of the diadenylate cyclase cdas in bacillus subtilis: An autoinhibitory domain limits cyclic di-amp production. *J Biol Chem* **289**: 21098-21107.
Mikkelsen, H., Ball, G., Giraud, C. & Filloux, A. (2009) Expression of pseudomonas aeruginosa cupd fimbrial genes is antagonistically controlled by rcsb and the eal-containing pvrr response regulators. *PLoS One* **4**: e6018.
Moller, S., Croning, M.D. & Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**: 646-653.
Morehouse, B.R., Govande, A.A., Millman, A., Keszei, A.F.A., Lowey, B., Ofir, G. et al. (2020) Sting cyclic dinucleotide sensing originated in bacteria. *Nature* **586**: 429-433.
Nagar, D.N., Mani, K. & Braganca, J.M. (2024) Genomic insights on carotenoid synthesis by extremely halophilic archaea haloarcula rubripromontorii bs2, haloferax lucentense bbk2 and halogeometricum borinquense e3 isolated from the solar salterns of india. *Sci Rep* **14**: 20214.
Nelson, J.W., Sudarsan, N., Phillips, G.E., Stav, S., Lunse, C.E., McCown, P.J. et al. (2015) Control of bacterial exoelectrogenesis by c-amp-gmp. *Proc Natl Acad Sci U S A* **112**: 5389-5394.
O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. (1991) X-ray structure of the gcn4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**: 539-544.
Ojha, R., Dittmar, A.A., Severin, G.B. & Koestler, B.J. (2021) Shigella flexneri diguanylate cyclases regulate virulence. *J Bacteriol* **203**: e0024221.
Paul, R., Weiser, S., Amiot, N.C., Chan, C., Schirmer, T., Giese, B. et al. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* **18**: 715-727.
Peng, L.H., Liang, X., Chang, R.H., Mu, J.Y., Chen, H.E., Yoshida, A. et al. (2020) A bacterial polysaccharide biosynthesis-related gene inversely regulates larval settlement and metamorphosis of mytilus coruscus. *Biofouling* **36**: 753-765.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. et al. (2004) Ucsf chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612.
Potrykus, K. & Cashel, M. (2008) (p)ppgpp: Still magical? *Annu Rev Microbiol* **62**: 35-51.
Purcell, E.B. & Tamayo, R. (2016) Cyclic diguanylate signaling in gram-positive bacteria. *FEMS Microbiol Rev* **40**: 753-773.
Rall, T.W. & Sutherland, E.W. (1958) Formation of a cyclic adenine ribonucleotide by tissue particles. *J Biol Chem* **232**: 1065-1076.
Rao, F., Yang, Y., Qi, Y. & Liang, Z.X. (2008) Catalytic mechanism of cyclic di-gmp-specific phosphodiesterase: A study of the eal domain-containing rocr from pseudomonas aeruginosa. *J Bacteriol* **190**: 3622-3631.
Rasmussen, R., Benvegnu, D., O'Shea, E.K., Kim, P.S. & Alber, T. (1991) X-ray scattering indicates that the leucine zipper is a coiled coil. *Proc Natl Acad Sci U S A* **88**: 561-564.
Robert, X. & Gouet, P. (2014) Deciphering key features in protein structures with the new endscript server. *Nucleic Acids Res* **42**: W320-324.
Romling, U. (2008) Great times for small molecules: C-di-amp, a second messenger candidate in bacteria and archaea. *Sci Signal* **1**: pe39.
Romling, U. (2023a) Cyclic di-gmp signaling-where did you come from and where will you go? *Mol Microbiol* **120**: 564-574.

Romling, U. (2023b) Is biofilm formation intrinsic to the origin of life? *Environ Microbiol* **25**: 26-39.
Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. & Normark, S. (1998) Curli fibers are highly conserved between salmonella typhimurium and escherichia coli with respect to operon structure and regulation. *J Bacteriol* **180**: 722-731.
Romling, U., Cao, L.Y. & Bai, F.W. (2023) Evolution of cyclic di-gmp signalling on a short and long term time scale. *Microbiology (Reading)* **169**:
Romling, U., Galperin, M.Y. & Gomelsky, M. (2013) Cyclic di-gmp: The first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77**: 1-52.
Romling, U., Gomelsky, M. & Galperin, M.Y. (2005) C-di-gmp: The dawning of a novel bacterial signalling system. *Mol Microbiol* **57**: 629-639.
Romling, U., Rohde, M., Olsen, A., Normark, S. & Reinkoster, J. (2000) Agfd, the checkpoint of multicellular and aggregative behaviour in salmonella typhimurium regulates at least two independent pathways. *Mol Microbiol* **36**: 10-23.
Romling, U., Sierralta, W.D., Eriksson, K. & Normark, S. (1998) Multicellular and aggregative behaviour of salmonella typhimurium strains is controlled by mutations in the agfd promoter. *Mol Microbiol* **28**: 249-264.
Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R. et al. (1987) Regulation of cellulose synthesis in acetobacter xylinum by cyclic diguanylic acid. *Nature* **325**: 279-281.
Ruiz, L.M., Castro, M., Barriga, A., Jerez, C.A. & Guiliani, N. (2012) The extremophile acidithiobacillus ferrooxidans possesses a c-di-gmp signalling pathway that could play a significant role during bioleaching of minerals. *Lett Appl Microbiol* **54**: 133-139.
Sakai, S., Conrad, R., Liesack, W. & Imachi, H. (2010) Methanocella arvoryzae sp. Nov., a hydrogenotrophic methanogen isolated from rice field soil. *Int J Syst Evol Microbiol* **60**: 2918-2923.
Schirmer, T. (2016) C-di-gmp synthesis: Structural aspects of evolution, catalysis and regulation. *J Mol Biol* **428**: 3683-3701.
Schirmer, T. & Jenal, U. (2009) Structural and mechanistic determinants of c-di-gmp signalling. *Nat Rev Microbiol* **7**: 724-735.
Simm, R., Lusch, A., Kader, A., Andersson, M. & Romling, U. (2007) Role of eal-containing proteins in multicellular behavior of salmonella enterica serovar typhimurium. *J Bacteriol* **189**: 3613-3623.
Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U. (2004) Ggdef and eal domains inversely regulate cyclic di-gmp levels and transition from sessility to motility. *Mol Microbiol* **53**: 1123-1134.
Simm, R., Remminghorst, U., Ahmad, I., Zakikhany, K. & Romling, U. (2009) A role for the eal-like protein stm1344 in regulation of csgd expression and motility in salmonella enterica serovar typhimurium. *J Bacteriol* **191**: 3928-3937.
Stothard, P. (2000) The sequence manipulation suite: Javascript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**: 1102, 1104.
Sullivan, M.J., Petty, N.K. & Beatson, S.A. (2011) Easyfig: A genome comparison visualizer. *Bioinformatics* **27**: 1009-1010.
Takahashi, K., Kasai, K. & Ochi, K. (2004) Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppgpp) in plants. *Proc Natl Acad Sci U S A* **101**: 4320-4324.
Teng, Y., Gong, X., Zhang, J., Obideen, Z. & Yan, Y. (2024) Investigating and engineering an 1,2-propanediol-responsive transcription factor-based biosensor. *ACS Synth Biol* **13**: 2177-2187.
Trevors, J.T. (2011) Hypothesized origin of microbial life in a prebiotic gel and the transition to a living biofilm and microbial mats. *Comptes Rendus Biologies* **334**: 269–272.
UniProt, C. (2025) Uniprot: The universal protein knowledgebase in 2025. *Nucleic Acids Res* **53**: D609-D617.
van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M. et al. (2024) Fast and accurate protein structure search with foldseek. *Nat Biotechnol* **42**: 243-246.
Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. (2009) Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

Whiteley, A.T., Eaglesham, J.B., de Oliveira Mann, C.C., Morehouse, B.R., Lowey, B., Nieminen, E.A. et al. (2019) Bacterial cgas-like enzymes synthesize diverse nucleotide signals. *Nature* **567**: 194-199.
Wigren, E., Liang, Z.X. & Romling, U. (2014) Finally! The structural secrets of a hd-gyp phosphodiesterase revealed. *Mol Microbiol* **91**: 1-5.
Witte, G., Hartung, S., Buttner, K. & Hopfner, K.P. (2008) Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell* **30**: 167-178.
Wu, Q., Lu, S., Zhang, C., Zhong, W., Zhao, H., Zhao, Y. et al. (2025) Significant impacts of metal ions from ancient oceans on nucleoside phosphorylation. *BMC Chem* **19**: 252.
Xie, Y., Li, H., Luo, X., Li, H., Gao, Q., Zhang, L. et al. (2022) Ibs 2.0: An upgraded illustrator for the visualization of biological sequences. *Nucleic Acids Res* **50**: W420-W426.
Xing, J., Gumerov, V.M. & Zhulin, I.B. (2023) Origin and functional diversification of pas domain, a ubiquitous intracellular sensor. *Sci Adv* **9**: eadi4517.

Supplementary data

Materials and Methods

Gene templates used in this study

>**WP_012035199.1** Q0W2L1_METAR Signal transduction protein OS=Methanocella arvoryzae (strain DSM 22066 / NBRC 105507 / MRE50) OX=351160 GN=RCIX2274 PE=4 SV=1

LRTGEDSVLAGINGDVLLLTGDGAEVAAIGEALALAGLDYRHAVDVASACFLAADTAFDAVLLGSDQS
VSAAIAFLKNSGTQVPVIVMSGTERCDMPVEAIKYVLSGALVQVAGLQCPGSGAIAAILETISRNRTR
NLGARLQDEVNKLSYIADTFPECLLMIDRAGTIVYANRACYETYEYSNGALKSCNISGLLAGEADFTK
LKELLRVTRSGGSNTELATISRRGQKRLSLAVITPRSDDAGAIFAYVILLQDITDSRYAETRMERLIS
ALTSPEAGLSSLDFRDLITDPEIQVMQDTLAAAMGVSAVIVTPGGTPLNQRSNVTSLCSSLARPGTVS
AQICGKCITDLAGRADAGQQARCTCPCTGMTEVAIPLVVNGKPAAFWIVGQVSLGEPAREKLQDLLTD
QGISPADVRNLIAGVHRMTEHQLNQIVATFSMVADKVTRLAIQNFRQGKYINERNTTLEELQRSERRL
QEMSYRDGLTGLHNRAYFEQELGRLESASSAFPVSIVAADVDGLKLINDTLGHTAGDQHLKVAGNMIS
AVFSASGDVCRTGGDEFCILLPGCGEAAAGKMLELLREEIRRYNERQPGPPLSISTGIAAINPGESLY
DAYKRADRLMYEHKLKRDDYARSKSIDMILLSLSERDYMDHGHTRVLTVIVDAMASELGLGEDARANL
KLLARLHDLGKIGVPDEIVMKPAKLDEHEWVLMRSHVEIGANIASRSLHMNHIAPLIRHHHERYDGTG
YPSGLSGSQIPLECRIMAIADSYDAMIADRPYRKGLSHECALEEILRNAGTQFDPEIVKVFAGIVRSG
RL

Synthax for all gene fragments

Underlined are start and stop codons; in italics codons for 6xHis-tag, in red, regions complementary to the pBAD30/28 vector and in bold mutated nucleotides. Bold and underlined is the sequence present in each of two synthetic gene fragments. For in vivo cloning mimicking conventional restriction enzyme cloning, the gene fragment is flanked by the XbaI and SphI sites (red and underlined) located within the multicloning site of pBAD30/28 vector.

Codon optimized gene template for WP_012035199.1 from *Methanocella arvoryzae* MRE50

GTACCCGGGGATCCTCTAGAGATAGT**T**G**A**CGAAA**T**GCGGGCTGAGAACAGGCGAGGATTCAGTTTTGG
CCGGAATCAATGGTGACGTCTTACTTCTTACGGGAGACGGTGCTGAGGTTGCTGCGATTGGAGAGGCC
CTTGCGTTAGCCGGGCTGGATTACCGTCACGCCGTCGATGTGGCTTCAGCCTGCTTTTTAGCCGCGGA
CACGGCCTTTGACGCTGTACTGCTGGGATCAGATCAATCTGTGAGTGCAGCGATCGCCTTTCTTAAAA
ATTCCGGGACGCAGGTACCGGTTATCGTGATGTCAGGGACCGAGCGTTGCGACATGCCAGTAGAAGCA
ATTAAATATGTATTGAGCGGAGCTCTTGTACAGGTCGCCGGATTACAGTGCCCCGGTTCTGGTGCGAT
TGCGGCCATCTTGGAAACGATCTCACGCAACCGCACTCGCAATCTGGGTGCGCGTCTGCAGGACGAGG
TTAACAAGCTGAGTTATATCGCAGATACGTTTCCAGAATGCTTGTTAATGATCGATCGCGCAGGTACC
ATCGTGTATGCCAATCGCGCTTGCTACGAGACGTATGAGTACTCGAATGGCGCACTTAAATCTTGTAA
TATCTCTGGACTTTTGGCCGGGGAAGCGGATTTTACAAAGCTGAAAGAATTATTGCGCGTGACACGCT
CGGGCGGATCCAATACCGAGTTGGCCACGATTAGTCGTCGTGGGCAGAAACGTCTGTCTCTGGCGGTT
ATCACGCCACGCAGTGACGATGCTGGTGCAATCTTTGCATATGTAATCTTGTTACAGGACATCACGGA
TTCACGCTATGCGGAAACCCGCATGGAGCGCCTTATCAGTGCTCTGACCTCGCCCGAGGCTGGGCTGT
CGTCGCTTGATTTTCGCGATCTTATTACGGACCCGGAAATCCAGGTTATGCAAGATACTCTTGCCGCA
GCAATGGGGGTCTCGGCGGTAATTGTCACTCCTGGCGGCACGCCGCTGAATCAACGTTCTAACGTGAC
GTCTTTATGTTCCTCTCTTGCCCGCCCGGGGACTGTTAGCGCCCAAATTTGTGGTAAATGTATCACAG
ATTTAGCGGGGCGTGCAGACGCGGGACAGCAAGCTCGCTGCACGTGCCCGTGCACCGGTATGACGGAA
GTGGCTATCCCCTTAGTGGTTAACGGCAAGCCCGCGGCATTCTGGATTGTAGGCCAGGTTTCGTTGGG
TGAACCAGCTCGTGAAAAGCTTCAGGACCTGCTTACCGATCAAGGAA**TTTCCCCGGCGGACGTACGC**A
ACCTTATTGCCGGGGTGCATCGCATGACCGAACACCAATTAAATCAAATTGTTGCCACTTTTTCCATG
GTAGCTGACAAAGTCACCCGCCTGGCGATTCAAAATTTTCGCCAAGGGAAGTATATCAACGAGCGTAA
CACGACGCTTGAAGAGTTGCAGCGCTCGGAACGCCGTTTGCAGGAAATGTCCTACCGCGATGGCTTGA
CAGGATTGCACAACCGTGCCTACTTCGAACAGGAATTAGGGCGTCTTGAGAGTGCTTCGTCTGCCTTC
CCTGTGTCTATCGTAGCAGCGGACGTTGATGGATTGAAACTTATCAACGACACATTGGGTCATACTGC

GGGGGATCAGCACTTAAAAGTTGCTGGGAACATGATTTCTGCAGTGTTCTCTGCATCAGGGGACGTCT
GCCGTACCGGAGGTGATGAATTTTGCATCCTTCTTCCCGGTTGCGGAGAGGCGGCTGCAGGGAAAATG
TTGGAATTGTTACGCGAGGAGATCCGTCGCTATAATGAGCGTCAACCGGGCCCTCCATTATCTATTTC
CACTGGAATTGCGGCAATTAATCCTGGGGAAAGCTTATACGACGCTTATAAACGCGCCGATCGTTTGA
TGTACGAGCATAAACTTAAACGCGACGACTACGCGCGTTCTAAGTCTATTGATATGATCTTATTAAGT
TTGTCCGAACGCGATTATATGGATCACGGTCACACACGTGTGCTGACGGTGATTGTCGATGCGATGGC
CTCCGAATTGGGATTAGGCGAAGACGCCCGCGCTAACTTAAAGTTGCTTGCTCGTTTGCACGACCTTG
GGAAGATCGGGGTACCAGACGAGATCGTTATGAAGCCTGCTAAGCTTGACGAACACGAGTGGGTTCTT
ATGCGTAGTCATGTTGAGATCGGAGCGAATATTGCCAGTCGTTCATTACACATGAACCACATTGCACC
CTTAATCCGCCACCATCACGAGCGTTATGATGGGACTGGGTATCCGTCAGGCCTGTCTGGAAGCCAGA
TTCCCCTTGAATGTCGTATCATGGCTATTGCCGATTCGTACGACGCTATGATTGCGGACCGTCCGTAC
CGCAAAGGCTTGTCTCATGAATGTGCGTTGGAAGAAATTCTTCGTAATGCGGGCACACAGTTCGATCC
CGAAATTGTTAAAGTATTCGCCGGTATTGTTCGCTCTGGACGTCTG*CATCACCACCACCATCAT*TAAG
CATGCAAGCTTGGCTGTTT

.

>**WP_339106355.1** *Haloterrigena salinisoli* strain KLK7 lcl|ORF19

LLSFSLYATKILSNLNDLFGILMVDEHNVVSYGYIDLDGFSEINNEHGHTFGDKIIEEVVKFGRRFSD
SKWEFNREYGQGDEFLMVLPDEDKESAVEVMEEFLKKLRSLRPEGQVVTASIGVATLPEDGANEGEIT
EKADLAMLNVEKWGGDGVIAYGNEKPTKEVEVWFEEAMLVEQDDLVKIQKWIDDSPSIRALKIYNESK
GIPNESRSSKTMMSTTKYEGEIKGIVKEIKEWDSSKVRFEMRGDKIQLEKAGF

Codon optimized gene template for WP_339106355.1 from *Haloterrigena salinisoli* KLK7

GTACCCGGGGATCCTCTAGACAAAGGATGAAATCAGCCTACTGTTGAGTTTTTCTCTATATGCAACCA
AAATATTAAGTAATCTCAATGACTTGTTTGGCATACTCATGGTAGATGAACATAACGTTGTTTCTTAC
GGTTACATTGACTTAGACGGCTTCTCTGAGATCAATAACGAACACGGCCACACCTTCGGCGATAAAAT
AATAGAAGAAGTCGTTAAATTCGGTAGAAGGTTTTCGGATTCTAAATGGGAATTCAATCGGGAATACG
GACAGGGTGACGAGTTTTTGATGGTCTTACCTGATGAAGATAAGGAATCAGCGGTAGAGGTGATGGAA
GAGTTTCTTAAAAAGTTACGTTCGCTCCGACCGGAAGGTCAGGTTGTCACCGCCAGTATTGGTGTAGC
TACGTTACCCGAAGATGGAGCAAACGAAGGTGAGATCACTGAAAAAGCAGACCTTGCGATGCTGAATG
TCGAGAAATGGGGCGGAGACGGTGTAATTGCATATGGTAACGAAAAGCCAACGAAAGAAGTAGAGGTT
TGGTTTGAAGAAGCCATGTTGGTAGAGCAGGATGATCTTGTTAAAATACAGAAATGGATTGATGATAG
CCCCTCAATTAGAGCACTGAAAATATATAACGAATCCAAAGGCATCCCTAACGAAAGTAGATCGTCAA
AGACGATGATGTCCACGACAAAATACGAGGGGGAAATAAAAGGGATAGTTAAAGAGATCAAGGAGTGG
GATTCCTCAAAAGTAAGGTTTGAGATGAGAGGTGATAAGATACAATTAGAAAAAGCAGGGTTC*CATCA*
*CCATCACCATCAC*TAAGCATGCAAGCTTGGCTGTTT

.

>WP_363463662.1 GGDEF domain-containing protein [Halogeometricum borinquense]

LCEGVRKTPVHKQFYLGISNLAKILSNPDGLFNTLMASQSGVVSYGFIDLDGFSEINNEHGHAFGDKV
LDRIVKFGQDFAGNDWEFSREYGQGDEFLLILPNEDKQSGVEVMEEFLNQLHALQPEGESVSASIGVA
TLPDDGTAEDEVIEKADLAMLNVEQWGGDGVIAYGEEKPTKVIEVWFEEAMLIEQDDLITIQKWMDDG
RSIRALKIHNESKGIPNESKSSNTMMSSTKYEGEIKGIVKEIKEWDSSRVTFEMRGHKEQLEEAGF

Codon optimized gene template for WP_363463662.1 from *Halogeometricum borinquense* E3

GTACCCGGGGATCCTCTAGATTTATCT**A**GGAATATCGAACTTGGCTAAAATCTTAAGTAACCCGGATG
GCCTTTTTAACACCTTGATGGCATCGCAAAGCGGTGTTGTTAGCTATGGCTTTATTGATCTTGACGGA
TTTTCGGAAATTAACAACGAACACGGCCACGCATTTGGAGATAAAGTGTTAGATCGTATTGTTAAGTT
CGGTCAGGACTTTGCGGGCAACGACTGGGAATTTAGCCGTGAATACGGCCAGGGCGACGAGTTCCTGC
TGATTCTGCCGAATGAAGATAAGCAGTCAGGTGTCGAAGTTATGGAGGAATTTCTTAACCAACTTCAC
GCGCTGCAGCCTGAGGGAGAGAGCGTCAGCGCGTCAATTGGTGTTGCCACTTTACCCGATGATGGTAC
GGCGGAAGATGAGGTAATTGAAAAAGCGGATCTGGCTATGTTAAATGTAGAACAGTGGGGTGGAGACG
GCGTGATTGCCTATGGTGAGGAAAAACCTACCAAGGTGATTGAGGTATGGTTTGAAGAGGCCATGCTG
ATCGAACAGGACGACCTGATCACGATTCAGAAATGGATGGATGATGGCCGCTCGATCCGTGCCTTAAA
GATTCATAATGAATCTAAAGGCATCCCGAACGAATCTAAATCTAGTAATACCATGATGTCGTCGACGA
AATATGAAGGCGAGATCAAAGGAATTGTTAAAGAAATCAAAGAGTGGGATTCCTCACGTGTGACTTTT
GAAATGCGTGGTCATAAAGAGCAGCTGGAAGAGGCGGGTTTC*CATCACCATCACCATCAC*TGAGCATG
CAAGCTTGGCTGTTT

>**MBI5036943.1** MAG: GGDEF domain-containing protein [Candidatus Micrarchaeota archaeon]

MSFWERMLNKHMENAALLRSGVFAEAQKTRGKLAGFIRKKKRKERAKLPLMLARIGVREPLERGRREMEL
LAAERKGLLADYFKKSTAAKEGSKLILRRLYNPLTGLHSKEFYKEVESELSKPEHANSEIAIASCDLKGF
KNVNDQHGHEAGDKMLKRVAELLEEARRKEGLEHVRIFHPSGDEFIVMAVRTMPEETGAELAVRNLLEQF
QLIKERENERDKKRGAEALHVDETELIGGVSSNVWKKNASVRGITASADTMSYVAKLLGHWGYVSEKQFN
SLPTATQKKILQTLPKVFRREGDRVPREGKAKKGYTRLLCTFSCGRSRVPCHLQNSFYSPIFAKADPRWG
KEKQFLSTTRIDASPKLDLPKGRRKKLNEKQLSPIDYCNNRAKQEYIPLHLLRFAHKSLLAYKTQNLLIK
SEVC

MBI5036943_1 template

TAGCGCAAGAGCTTAAATACTTCTGAAACCAAATCAGATTATGTCGTTTTGGGAAAGAATGTTAAACAAGCACATGGAAAAC
GCGGCATTGCTACGCAGCGGAGTGTTTGCTGAAGCACAGAAAACCAGGGGCAAGCTAGCAGGCTTTATAAGAAAAAAGAAGA
GGAAAGAAAGAGCGAAGCTTCCTCTGATGCTAGCGAGAATTGGCGTGCGGGAGCCCCTGGAGAGGGGGCGGCGGGAAATGGA
ACTACTGGCGGCTGAAAGGAAAGGGCTGCTGGCAGATTATTTCAAGAAAAGTACGGCTGCTAAGGAAGGAAGCAAACTTATT
CTGAGGAGGCTCTACAACCCTCTCACAGGCTTGCACAGCAAGGAATTCTATAAGGAGGTGGAGAGCGAGCTTAGTAAACCAG
AGCATGCAAACTCTGAAATCGCAATCGCTTCATGTGATTTGAAGGGTTTCAAGAACGTAAACGACCAACACGGACATGAAGC
CGGCGACAAGATGTTGAAGAGAGTCGCGGAGTTGCTTGAGGAAGCCCGGAGAAAGGAAGGACTTGAGCATGTCAGGATATTT
CATCCCTCAGGGGACGAGTTTATCGTGATGGCGGTCAGAACCATGCCTGAGGAGACGGGAGCAGAGCTGGCGGTGAGGAACC
TGCTTGAACAGTTCCAGCTCATAAAGGAAAGGGAGAACGAGAGGGACAAAAAAAGGGGCGCAGAAGCGCTTCACGTGGATGA
AACCGAGCTCATCGGCGGGGTTTCCTCGAACGTCTGGAAGAAAAACGCTTCGGTAAGAGGGATAACGGCTTCAGCTGACACG
ATGAGTTATGTAGCTAAATTGCTGGGCCACTGGGGGTACGTTTCGGAAAAACAGTTCAACTCCCTCCCTACTGCAACCCAGA
AAAAGATTCTGCAAACGCTACCCAAAGTATTCCGACGAGAGGGGGATAGGGTACCGAGGGAGGGAAAGGCGAAAAAGGGTTA
TACGCGGTTGCTTTGTACATTTTCCTGTGGCCGAAGCCGCGTCCCCTGTCATCTCCAAAACTCATTCTATTCACCTATTTTT
GCTAAGGCTGACCCGCGCTGGGGGAAAGAAAAACAATTCCTGTCCACAACCCGAATCGACGCCTCACCTAAACTCGATCTTC
CGAAAGGCCGTCGCAAAAAACTAAACGAAAAGCAGCTTTCTCCAATCGACTACTGTAACAATAGGGCCAAGCAAGAATATAT
ACCTTTGCATCTCTTGCGCTTTGCTCATAAAAGCTTGCTTGCTTACAAAACCCAAAACTTGCTCATAAAATCCGAAGTTTGC
TAAAAGCCTGTTTGTCCTAACCCTTTGG

>**MBI4145595.1** MAG: GGDEF domain-containing protein [Candidatus Woesearchaeota archaeon]

MKEEDARRIIQLLASLTSDEQLQLHSLEAELGTIEPHRQLVRGLIFALRQNDTATVRQVIVRIMADIFPS
FMVTLRPLDQEFESVRQQILQQSKIQFEDTEALQQELEQKGMAAAFAERLALAAKRKQITDISAFITARL
GDWRARQSDKKRLLEQEQATLRSIEQTPQPGPNQAVAWADDLDAVERALELVLRKQKDALDALNLALAKQ
RTVERIILDLQQANKMTADALLKVIAGFSKPSEFQELQRIILRHKVALGPEAERVLNLAGTFAVHRTEHV
ETEAAAGGIDFLTQVPNRRTFLQLAEPQLSLARRNRWPAALLVVDIDHFKGINDTYGHDAGDRALALVAR
IMHENVRTSDIVGRWGGEEFVVFLPNTDKIHAAGVAEHLRTLIADKTVTLMSELRKGEPGRANLTVSIGV
AAFPDNADGFNQLFSSADTALYQAKNSGRNRVVSSATA

MBI4145595.1 template

AGACTACAAGCACCACCTGAAGCTTTAAATATAGTTCTTGTCGTATCCCGCCTGTGAAAGAGGAGGACGCGCGACGGATCAT
CCAGCTTCTCGCCAGCCTCACTTCTGACGAACAGCTGCAATTGCACAGCCTCGAGGCAGAGCTAGGCACGATTGAGCCGCAC
CGGCAGCTTGTCAGGGGGTTGATTTTCGCGTTGCGCCAGAACGACACCGCGACCGTCCGGCAGGTCATCGTTCGCATCATGG
CTGATATCTTCCCTTCGTTCATGGTGACGTTGCGACCGCTCGATCAGGAGTTCGAGTCAGTGCGCCAGCAGATCCTGCAGCA
GTCGAAGATCCAGTTCGAGGATACGGAAGCTCTGCAGCAGGAACTCGAACAGAAGGGCATGGCAGCCGCGTTCGCTGAAAGG
CTGGCGCTCGCGGCCAAGCGCAAACAGATCACGGACATCTCCGCGTTCATCACCGCGAGATTGGGCGATTGGCGGGCCAGGC
AGTCGGACAAGAAACGCCTCCTGGAGCAGGAACAGGCAACGCTTCGGAGCATCGAGCAGACGCCGCAGCCGGGCCCGAACCA
GGCCGTCGCGTGGGCCGACGACCTCGATGCCGTCGAACGCGCACTTGAGCTCGTGTTGCGGAAGCAGAAGGACGCTCTCGAC
GCGTTGAACCTCGCGCTGGCCAAACAGCGTACGGTCGAGCGGATCATCCTCGATCTTCAGCAGGCCAACAAGATGACCGCTG
ACGCGTTGCTGAAGGTGATCGCGGGGTTCTCCAAGCCTTCAGAATTCCAGGAGCTCCAACGCATAATCCTCAGGCACAAGGT
TGCGCTCGGGCCGGAGGCAGAACGCGTGTTGAACCTCGCCGGGACGTTCGCCGTGCACCGGACAGAGCACGTCGAGACCGAG
GCTGCCGCGGGAGGGATCGACTTCCTCACGCAAGTCCCGAACAGGCGCACGTTCCTGCAATTGGCAGAACCCCAACTGTCGC
TGGCGCGGCGCAACCGTTGGCCCGCCGCGTTGCTGGTGGTCGACATCGACCATTTCAAGGGCATCAACGACACCTACGGCCA
TGACGCGGGCGACCGCGCGTTGGCGTTGGTCGCGAGGATCATGCACGAGAACGTACGCACCTCGGATATAGTCGGACGCTGG
GGCGGCGAGGAGTTCGTCGTGTTCCTGCCGAACACCGACAAGATTCATGCTGCGGGTGTCGCTGAGCATCTGAGGACGTTGA
TTGCGGACAAGACCGTGACTTTGATGAGCGAACTGCGGAAAGGCGAGCCAGGCCGCGCCAATCTCACCGTCAGCATAGGCGT
GGCAGCGTTCCCCGACAACGCGGATGGGTTCAACCAGCTTTTCTCCTCGGCTGACACAGCGCTGTACCAGGCGAAAAACTCG
GGCAGGAACCGCGTCGTTTCTAGCGCAACCGCC *CATCACCATCACCATCAC* TGA

>**MBN1216112.1** MAG: EAL domain-containing protein [Candidatus Lokiarchaeota archaeon]

MENTSLDKINILIIEDNPGDVKIIEKILSEIETTIYNLVCKVNLSDGIKYALIKKPDIILLDLMLSDSHG
LETLKIIREKVSRIPIVVLTGIDDEEIGVQAVKFGAQDYLVKNYIQSRIFSRVILYAIERSKAEEEIKFL
AYYDYLTKLPNRRYFLERFNATIASAGRYNRISAILFLDIDDFKLINDSLGHKIGDLLLIDISKRLKKCI
RKSNFNYSNKKIETFIDVVARLGGDEFVISLIEISNPEDSSKIAKRIVHELSHPYYVEDHEIYITASIGI
ALYPSDGENTETLLKHADIAMYQAKKSGKNNFKYFTNSMNKSIIKLVSLEKDLRKALDNEEFFLCYQPKM
DLNTGNIIGFEALIRWQQQGKKILLQEDFMSIAEEINLIVPIGYWVIRNIALFLSELKKENLEIPISINI
SLKQLKEVELIKIINNLLIEYNISKELIEFEISESIFLGNLEEIKDNINEINKLGIKIALDNFGEGCSSF
EILKKLKASTIKIGRYFTENILKSNEDASIVSTLISLGHNLNLSVICEGIENFSQLNYLKSKNCDIIQGF
LLSRPVPAEKTRLIMIKEKEGNGIGKKLLRTRRNKNYEQ

MBN1216112.1 template

GCGCTCTAGAAACAGTTTTGGGAAGAATATATGGAGAACACTTCTTTGGATAAAATCAACATCCTGAT
TATCGAAGATAACCCGGGGGATGTTAAGATTATCGAAAAGATTCTGTCAGAAATCGAGACAACGATCT
ACAATTTGGTATGTAAGGTGAACTTGAGCGACGGTATTAAATATGCCTTGATTAAGAAACCGGATATC
ATTCTTTTAGATCTTATGTTAAGTGATAGCCACGGTTTAGAGACGCTTAAGATTATTCGCGAGAAAGT
TAGCCGCATCCCAATTGTTGTTCTGACAGGTATTGATGACGAAGAAATTGGAGTGCAGGCTGTTAAGT
TCGGGGCACAGGATTATCTGGTAAAAAATTACATCCAGTCGCGCATTTTTAGCCGTGTAATCCTTTAT
GCAATTGAACGCTCCAAGGCCGAAGAGGAAATTAAATTTCTGGCTTACTATGATTACTTAACGAAGTT
ACCGAACCGCCGCTACTTTCTGGAACGTTTCAATGCGACCATTGCAAGCGCGGGACGTTATAACCGCA
TCTCCGCGATCCTGTTCCTTGACATTGACGACTTCAAATTAATCAACGATAGCTTGGGGCATAAAATT
GGGGACCTTTTGCTTATCGACATCTCAAAGCGCCTGAAAAAATGCATTCGCAAAAGCAACTTCAACTA
TTCTAACAAAAAGATTGAGACTTTCATCGACGTCGTGGCTCGCCTGGGGGGCGACGAATTCGTGATTT
CGTTAATCGAAATCAGCAACCCTGAGGATTCCTCTAAAATTGCAAAGCGCATTGTACATGAATTATCA
CATCCTTACTATGTCGAGGACCACGAAATCTACATTACCGCGAGCATCGGTATTGCCTTGTATCCGAG
CGATGGGGAAAACACGGAAACACTTTTGAAGCACGCCGACATCGCTATGTACCAGGCAAAAAAGTCAG
GGAAAAATAACTTTAAATATTTTACAAACAGCATGAATAAATCAATTATTAAGTTGGTGAGCTTGGAA
AAGGATCTTCGTAAGGCTTTAGACAATGAAGAGTTTTTCCTTTGTTACCAGCCAAAAATGGACTTAAA
CACAGGTAACATTATTGGCTTCGAGGCCTTAATTCGTTGGCAGCAACAAGGTAAGAAAATTCTGTTAC
AGGAGGATTTTATGTCGATCGCAGAGGAAATTAACCTTATTGTGCCAATTGGATACTGGGTGATCCGC
AACATCGCCCTTTTTCTGTCGGAACTTAAAAAGGAGAATTTGGAGATTCCAATTTCTATCAACATTAG
CTTGAAACAGTTGAAAGAGGTCGAATTAATCAAAATTATCAACAACTTATTGATTGAATACAATATTT
CAAAAGAGTTGATTGAATTCGAAATCTCAGAGTCTATCTTCTTGGGTAATTTAGAAGAGATTAAAGAC
AACATCAATGAAATCAACAAACTTGGGATTAAAATCGCGCTTGACAATTTTGGAGAAGGGTGTTCCAG
CTTCGAGATCCTGAAAAAGCTGAAAGCGTCAACTATCAAGATCGGCCGCTATTTTACGGAGAACATCC
TTAAGTCAAACGAGGATGCGAGCATTGTGTCAACTCTGATCTCCCTGGGCCATAATCTGAATTTAAGT
GTTATCTGCGAAGGAATTGAGAACTTTAGTCAGCTTAACTACCTGAAGAGCAAGAACTGCGATATTAT
TCAAGGTTTTCTGTTAAGCCGTCCTGTCCCTGCTGAAAAGACTCGTTTGATTATGATCAAAGAAAAGG
AGGGGAACGGGATTGGAAAAAAGCTGCTGCGCACGCGTCGTAACAAAAACTATGAACAG*CATCACCAT*
*CACCATCAC*TAAGCTTGCCG

>**AQS32838.1 = KY476720.1** hypothetical protein [uncultured archaeon]

MPKKDYLTEKLNKKVSALDPKFREKITELFSTMHESISLLYEAAIHDEKTGLYNNKFFDTMLDIEFEKAK
RGRQRLCLFVIDIDFFKKINDAYGHLMADKLLKKLAEILQRSLRKSDIIARFGGEEFFIIFPETSIEKAK
FLTSRLREIIKSDKTLKKYSLTISGGLTEFKKSDTKKTLMKRADKALYKAKNSGRDRFVVLK

Codon optimized gene template for **AQS32838.1 = KY476720.1**

GGATCCTCTAGACAATAAGGTTAAAAGTCCTTGTTTCTTAC**G**AT**C**AT**C**ATGCCCAAGAAAGATTACTTGAC
GGAGAAGCTTAACAAGAAGGTGTCGGCGTTAGACCCGAAATTTCGCGAAAAGATCACAGAGTTGTTCAGCA
CGATGCATGAGTCCATTTCTTTATTATATGAAGCCGCAATCCATGATGAGAAGACAGGACTTTACAATAAC
AAATTCTTCGATACGATGCTTGACATCGAATTCGAGAAAGCTAAACGCGGTCGTCAACGTCTGTGTTTATT
CGTCATCGATATCGACTTCTTTAAGAAGATCAATGATGCATACGGCCACTTGATGGCCGATAAATTACTTA
AAAAGCTTGCCGAGATTCTGCAGCGCTCGCTTCGTAAGAGTGACATTATCGCGCGCTTTGGTGGCGAGGAA
TTCTTCATCATTTTCCCTGAAACGAGCATCGAAAAGGCTAAGTTTCTTACGAGTCGCCTTCGTGAGATTAT
TAAGAGCGACAAAACATTAAAGAAGTATTCATTAACAATCAGCGGTGGGCTGACTGAGTTCAAAAAGTCGG
ACACTAAAAAGACATTGATGAAGCGCGCAGACAAGGCTCTTTATAAAGCAAAGAACTCCGGTCGTGACCGC
TTTGTGGTCCTTAAA*CATCACCACCACCACCAC*TGAGCATGCAAGCTT

>**MCR4284784.1** MAG: GGDEF domain-containing protein [archaeon]
MINKGNSVLNLEFKKEVSKVLKKFKKSLYELHDLATRDDKTGLYNHRFFTNVFEIELEKARRGKQNISLA
MLDIDCFKKYNDKYGHVMGDEILYDLAQTLLKTTRKYDVLARFGGEEFLVLLPETPLSKAKKAAERMRKG
LYKNNKLKKYGITISIGVTEYKDKDTMKKMITRTDKALYQSKQDGRNRVTAI

Codon optimized gene template for **MCR4284784.1**
GGATCCTCTAGAGCAGAAA**GG**AT**G**TATAGTCTATGATCAATAAGGGTAACAGCGTCCTGAACTTGGAATTT
AAAAAGGAAGTGTCGAAGGTTTTGAAGAAGTTCAAGAAGTCTTTATACGAGTTGCACGACTTGGCTACACG
TGACGACAAGACCGGCTTATACAATCATCGTTTCTTCACAAATGTATTTGAGATTGAGCTGGAAAAAGCAC
GTCGTGGAAAGCAAAATATTTCTTTAGCGATGCTGGATATTGATTGCTTTAAGAAGTATAATGATAAATAC
GGACACGTCATGGGTGACGAGATTCTTTATGATTTGGCCCAAACATTGTTAAAGACAACCCGCAAGTATGA
CGTTTTGGCCCGCTTTGGAGGGGAAGAGTTCCTTGTGCTGTTGCCTGAGACTCCGTTATCGAAGGCCAAAA
AGGCTGCGGAACGTATGCGCAAGGGGCTGTACAAGAATAACAAACTGAAGAAGTATGGCATCACAATCAGT
ATCGGAGTTACGGAATATAAAGACAAGGACACAATGAAGAAGATGATCACGCGCACAGACAAAGCACTTTA
TCAGTCCAAACAAGACGGTCGCAATCGCGTCACGGCGATC*CACCACCATCATCACCAC*TAAGCATGCAAGC
TT

>**KHO53915.1** MAG: ggdef protein [archaeon GW2011_AR18]
MEEGINKKIEEYEYQIKLMSIFQNFLIDITKKLEIKELAENLLKFLNESFKIDKCSVIINRQRYFLGN
LKDEKLIESENRIMKEVYKVSIPYIVKNLKNDTILFDLKNDSEESLLVIPIINDNKIITYVNIYDQPE
NIHKKDVKLINYFLSKINYAIINSLEYTQVKDKSITDSLTGLYNRNYFIEKLRYESRNFKKYLSIIMT
DIDYFKNYNDKNGHQAGDYLLKELAVLLKNNFRTNDIIGRYGGEEFIIILPETDNETGLQVCERLRKS
VESYNFKFKENQPNNAVTISIGLMTTVSKEIEVEDLIKEADNNLYKSKTNGRNRATNSIIISKNLSVK

Codon optimized gene template for KHO53915.1

GGATCCTCTAGATATTAT**AGGACGT**ATGAACCATGGAGGAAGGTATTAACAAGAAGATTGAAGAATATGAG
TACCAAATCAAGTTAATGTCCATTTTCCAGAACTTTTTGATTGACATCACAAAGAAATTGGAGATCAAGGA
ATTGGCCGAAAATTTGTTAAAATTTCTTAATGAGAGCTTTAAGATCGACAAGTGTAGCGTGATTATTAATC
GCCAACGTTATTTTCTGGGAAATCTGAAAGACGAAAAACTTATCGAGAGTGAAAATCGCATCATGAAGGAG
GTGTATAAGGTGTCGATCCCATACATTGTTAAGAACTTGAAGAATGATACCATTTTATTTGACTTGAAGAA
TGACAGCGAGGAGAGTTTGCTGGTCATTCCTATTATTAACGACAACAAGATTATCACATACGTGAATATCT
ACGATCAGCCTGAAAACATCCATAAGAAGGATGTTAAACTGATCAACTACTTTTTATCTAAAATCAATTAT
GCCATCATCAATTCCTTGGAGTACACACAGGTGAAGGATAAAAGTATTACTGATTCCTTGACCGGTCTGTA
TAATCGCAACTATTTCATTGAGAAGTTGCGCTATGAGTCGCGTAACTTCAAGAAATACCTGTCCATCATTA
TGACAGACATTGATTACTTTAAGAATTACAACGACAAGAATGGGCACCAAGCAGGCGATTATCTGCTTAAA
GAGCTTGCTGTGCTTCTTAAGAATAATTTTCGTACGAATGATATTATTGGTCGCTATGGTGGCGAAGAGTT
TATCATTATCTTACCCGAGACCGATAATGAGACCGGCTTGCAGGTGTGCGAGCGTCTTCGCAAGTCGGTGG
AATCGTACAATTTCAAGTTTAAAGAGAACCAACCGAACAATGCTGTCACGATTTCGATCGGCCTTATGACT
ACCGTGTCGAAAGAGATTGAAGTTGAAGATTTAATTAAAGAGGCGGATAATAACTTGTATAAATCCAAGAC
CAATGGCCGTAATCGTGCAACAAATAGCATTATCATCTCTAAGAACTTATCTGTTAAG*CATCACCACCACC*
*ATCAT*TGAGCATGCAAGCTT

>**KHO53289.1** MAG: Diguanylate cyclase with PAS/PAC sensor [archaeon GW2011-AR13]

MHNEECLDKIVKNVIDSNHFVSPITIKRWLDTMKNSNDKIYVTNRILTGLLEKNNLYMQGTIPYIQEV
INNEIKNLKNEKSELETLATVDSLTGLFNRRRCVENLNRELKRSAREKKPLSVIMCDLDHFKSINDTY
GHDAGDSVLKTISGVAKNNFRGTDILCRIGGEEFFIILPNTAKTNAIYVADRLRRNIEKQPIKIDSQD
KPIYITTSIGIYQIGENETVETIMKNADLRLYNAKKSGRNKVCSE

Codon optimized gene template for **KHO53289.1**

GGATCCTCTAGACTTTAA**AGGTAGT**ATTTTTTATGCATAACGAGGAATGTCTTGATAAAATCGTCAAA
AACGTTATCGACTCAAACCATTTCGTCTCGCCTATTACAATTAAACGTTGGCTGGACACTATGAAAAA
CTCCAACGATAAGATCTATGTAACTAACCGCATCCTTACGGGGTTACTGGAAAAGAACAACCTTTACA
TGCAGGGCACCATTCCCTATATCCAGGAGGTAATTAATAACGAAATCAAGAATCTTAAGAACGAGAAA
TCGGAGCTTGAGACCTTAGCTACAGTCGATAGCCTGACCGGACTTTTCAATCGCCGCCGCTGCGTGGA
AAACCTGAATCGTGAATTGAAGCGCAGTGCGCGCGAGAAGAAACCGCTGTCAGTGATCATGTGTGACT
TAGACCATTTCAAGTCCATTAACGATACATATGGACACGACGCGGGAGACTCGGTGCTTAAGACAATT
AGTGGAGTAGCTAAGAATAACTTTCGTGGAACGGACATTCTTTGCCGTATCGGCGGTGAGGAATTCTT
CATTATTCTTCCCAACACGGCTAAGACGAATGCAATTTATGTAGCGGACCGTCTTCGCCGCAATATTG
AGAAACAACCTATTAAAATCGATTCCCAGGACAAACCTATTTACATCACCACCTCCATCGGCATTTAT
CAGATTGGCGAGAACGAGACAGTCGAAACTATCATGAAGAATGCTGATTTACGCCTGTACAATGCCAA
GAAATCTGGTCGTAATAAAGTTTGCTCAGAG*CACCATCACCATCACCAC*TAAGCATGCAAGCTT

>**MBT3408960.1** MAG: diguanylate cyclase [Candidatus Woesearchaeota archaeon]
VSMNKKEVSFTSIENIKKEIDKATSKQGISDEINDISKIIDIDTWVKIQDNFATAINLPVFLIEESGT
EIAVSNDFPYFCKLINSKEKGRKACINCRKKNYKELKEEDTNIKYFECHAGLFNIMTPIKINNKMVGA
INCSSILKSSRNIAKVIRVAEELNIESVELIDAVNELPIKQREDLVFNGTLLYVLSQTIPGLVNEHKT
NIEKINELEIINKFSNEIQSTLQLNETLKIILNKIVKLTNAIESSITIYKEDKDNKILERYTNSKNDL
QLFQNIEKQIINKITNEKNLKKIYEFEKIEDEFNIKTTFIKKIKTIPLLMRQKIIGFIFIYLDNIEKD
KNLELLQIISSQSAFGIINAQQFKEINTLAIMDKLTNLYNRRYIMEKLDDELKRAQRTERFLSIILMD
IDHFSNYNNNNGHPEGDKLLTKISALIKENTREIDIIGRYGGEEIIVLLPDANINEALPVAERIRKAI
ENYPFANREKQPLGKVTTCIGLVSHRGKNKTKDDLIKLSDDLLYKSKKAGRNKTSYRIIDEKSSEILQ
SQTHQ

Codon optimized template for MBT3408960.1

GGATCCTCTAGATCAAAT**AGGAAGT**GGTGTCAATGAATAAGAAGGAGGTCTCCTTCACAAGTATTGAG
AATATTAAGAAGGAAATTGATAAGGCCACCAGTAAGCAAGGCATTAGTGATGAGATTAACGACATTTC
CAAGATTATTGATATTGACACATGGGTTAAAATTCAGGATAACTTTGCCACCGCAATCAATCTTCCGG
TATTCCTTATTGAAGAGTCCGGTACAGAGATTGCGGTAAGTAATGACTTTCCCTATTTCTGCAAGTTG
ATCAACTCAAAAGAAAAAGGACGCAAAGCGTGTATTAACTGTCGTAAGAAGAACTACAAAGAGCTGAA
AGAAGAAGATACCAACATCAAATATTTCGAATGCCATGCCGGGCTTTTCAACATTATGACCCCGATTA
AAATCAACAATAAGATGGTTGGTGCCATTAATTGTAGCAGTATTCTTAAATCGAGCCGCAACATTGCT
AAAGTTATTCGCGTAGCAGAAGAACTGAACATTGAAAGTGTCGAGCTGATTGACGCCGTAAACGAATT
ACCCATTAAACAGCGCGAAGACCTTGTGTTTAATGGCACCTTACTTTACGTCTTATCCCAAACAATCC
CAGGTCTGGTGAATGAGCATAAGACGAATATCGAGAAAATTAACGAACTTGAAATTATTAACAAGTTC
TCTAATGAGATTCAGTCCACTCTGCAATTGAACGAGACCTTAAAGATTATCTTAAACAAGATTGTGAA
ACTTACAAACGCCATTGAGAGTAGTATTACTATCTATAAAGAAGACAAAGACAATAAGATTCTGGAAC
GTTATACAAATTCCAAAAATGACCTTCAATTGTTTCAAAACATCGAAAAACAGATTATTAATAAGATT
ACAAATGAAAAGAACCTTAAAAAGATCTATGAGTTTGAGAAAATCGAAGACGAGTTCAATATTAAGAC
CACCTTCATCAAAAAGATTAAGACCATTCCACTTCTGATGCGTCAAAAGATCATCGGCTTCATTTTTA
TCTATCTGGACAACATCGAAAAAGATAAAAACTTAGAGTTATTACAAATCATTTCCTCCCAGAGTGCT
TTTGGCATTATCAACGCACAGCAATTCAAGGAAATCAATACCCTTGCGATTATGGATAAGTTGACCAA
TCTTTACAACCGCCGTTACATCATGGAGAAACTGGACGATGAGCTGAAGCGCGCTCAGCGCACTGAGC
GTTTCTTATCCATTATCCTTATGGACATTGATCATTTTAGCAACTATAACAACAACAACGGTCATCCT
GAAGGTGATAAGTTACTGACAAAAATCTCGGCGCTTATTAAAGAGAATACCCGCGAGATCGATATCAT
TGGACGCTACGGCGGTGAAGAGATTATTGTCTTGCTTCCAGATGCTAATATCAACGAGGCGCTGCCGG
TTGCTGAACGCATCCGCAAGGCAATCGAAAATTATCCTTTCGCTAACCGCGAAAAACAACCCCTGGGA
AAAGTTACGACCTGTATCGGACTGGTTTCGCATCGTGGAAAGAACAAGACAAAAGACGACTTAATTAA
ACTTAGCGACGACTTGTTGTATAAATCAAAGAAAGCTGGCCGCAATAAAACAAGCTATCGCATCATTG
ACGAAAAATCGAGTGAAATCCTGCAGAGCCAGACACACCAG*CATCATCACCATCATCAT*TAAGCATGC
AAGCTT

>**HJJ27798.1** MAG TPA: GGDEF domain-containing protein [Methanocorpusculum sp.]
MATLAQWFYGGKDKEFFIRHSEEIRLANTKTAGTFGIVITALVLLFTIFVCASRLETDLTALFIFTAV
LLVLNIYYWVWGREHPKLTVYYIQVILLILIAFLILRDIMEPEEYAVFIPLFLIVMPLLFINPMHKTL
ISMLVVLAGFACITILVKGPTVYGIYDIVDVCITASFGFFMGQNILRLRIGTIEAYEKLKQSSESEVS
KALDMANTDALTGVRSRSAYEKICATLDEKIQEGTASEFALVLCDVNGLKKTNDTMGHEMGDKLIKSC
CHEICVVYAHSPVFRIGGDEFVVLLHDSDYMDRKALFEKLHEANMKTNVSFAYGMAEFDSQRDARTQH
VFARADAQMYLCKKKMKAARSD

Codon optimized template for HJJ27798.1

GTACCCGGGGATCCTCTAGATATACATA**GG**CAATAATAAAATGGCCACCCTGGCGCAGTGGTTTTACG
GGGGAAAAGATAAGGAATTCTTTATTCGTCACTCCGAAGAAATTCGCCTGGCCAATACCAAGACCGCG
GGCACGTTCGGCATTGTAATTACCGCGCTGGTGCTGCTGTTTACGATTTTTGTGTGTGCCTCGCGCCT
GGAAACTGATCTGACCGCCTTATTCATTTTTACCGCGGTCCTGCTGGTGCTGAATATCTACTATTGGG
TATGGGGTCGCGAACATCCGAAACTTACTGTTTACTACATCCAAGTGATTTTGCTGATTCTTATTGCG
TTTTTAATTCTTCGCGATATTATGGAACCCGAGGAATATGCAGTTTTTATCCCTCTTTTCCTGATCGT
AATGCCTTTGCTGTTTATTAACCCAATGCATAAAACCTTGATTAGCATGCTGGTGGTGCTGGCAGGAT
TTGCGTGCATCACGATCCTTGTAAAGGGACCAACGGTGTATGGTATTTATGATATTGTGGATGTGTGC
ATCACAGCATCGTTTGGTTTTTTTATGGGCCAAAACATTTTGCGTTTACGTATTGGTACGATTGAAGC
TTATGAAAAGCTGAAACAGAGTAGCGAAAGCGAGGTAAGTAAAGCCCTGGATATGGCGAACACCGACG
CGCTTACCGGCGTTCGCTCGCGTAGTGCGTACGAGAAAATCTGTGCGACGCTGGATGAGAAAATTCAA
GAAGGAACCGCCTCTGAATTCGCGTTAGTTCTGTGCGATGTCAATGGTTTAAAGAAAACCAACGATAC
TATGGGTCACGAGATGGGAGATAAACTTATCAAGAGTTGCTGCCACGAAATTTGTGTGGTGTATGCTC
ATTCTCCGGTATTCCGTATCGGTGGCGATGAATTTGTTGTACTGCTGCACGATTCGGACTATATGGAT
CGCAAAGCCCTGTTTGAAAAACTGCACGAAGCTAATATGAAAACCAATGTTTCTTTTGCCTATGGCAT
GGCCGAATTTGATAGCCAACGCGATGCGCGTACCCAGCATGTCTTTGCCCGTGCGGACGCCCAGATGT
ACCTTTGCAAAAAGAAAATGAAAGCAGCACGCTCTGAT*CATCACCATCATCACCAT*TAAGCATGCAAG
CTTGGCTGTTTTG

>**RJQ22058.1** MAG: GGDEF domain-containing protein [Candidatus Woesearchaeota archaeon]

MIEQETSAVNRLERIVESAPQVIQVVYALPPSRDAPYTQHCARDVPKSFLGRGPDGVFIDDRIAPNAA
LYLPRAEGDVFGYTTSMSDAGGRRRLLILSDTQIPLDPEQDSLPPEFRRFYERFIDADQGLSPLPVIE
GRQPEYWEVMEEAMREKSRQLKQSRRQLQATTRRFRRLAYMDDNTELPNRRAYWGDMARNVEEAIAYG
KKMHVLALDAIGFKKINDEYGQVEGDGVIRQLGEILKRVTRAEDKVYRIGGDEFRVVTDADPYMLKAR
FEKVVPVATMSVTGPGGTGEEVYLQGVYVGIASFDGARSGVLTEEERRAIFVGLEKKADAAMQEVKAA
RKTR

Template for RJQ22058.1

GTACCCGGGGATCCTCTAGAAAAATCAAA**G**AAGTTGGTGTCTGATGATTGAACAGGAAACAAGCGCGG
TTAACCGGTTGGAGCGAATTGTTGAGTCAGCGCCGCAAGTGATTCAGGTTGTCTATGCTCTTCCGCCG
TCAAGGGACGCTCCGTACACGCAGCATTGCGCGCGGGATGTGCCCAAATCGTTTCTGGGGCGCGGGCC
GGATGGTGTTTTTATTGATGACCGCATCGCGCCGAATGCCGCATTGTATTTGCCAAGGGCAGAGGGCG
ATGTTTTTGGGTATACGACGTCAATGAGTGATGCGGGCGGGCGTCGCCGTCTTTTGATTTTATCGGAC
ACGCAGATTCCTCTTGACCCCGAGCAGGATTCATTGCCGCCGGAGTTCAGGCGGTTCTATGAGCGGTT
CATTGATGCGGACCAGGGGTTGTCTCCGCTTCCGGTAATTGAGGGTAGGCAGCCGGAGTATTGGGAAG
TGATGGAAGAGGCGATGCGGGAAAAGTCGCGTCAGCTCAAGCAGTCACGGCGTCAGTTGCAGGCGACA
ACGCGCCGGTTCCGAAGGCTTGCGTATATGGATGATAATACGGAGTTGCCGAACCGCCGCGCGTACTG
GGGAGATATGGCAAGGAATGTGGAAGAGGCGATTGCGTACGGGAAAAAAATGCACGTGCTTGCGCTTG
ACGCGATAGGGTTCAAGAAGATTAATGATGAGTACGGGCAGGTGGAAGGCGATGGTGTTATCAGGCAG
CTTGGTGAGATTTTGAAGCGCGTCACGCGTGCCGAGGATAAGGTGTACCGCATTGGCGGGGATGAGTT
TCGCGTGGTGACTGATGCTGACCCGTATATGCTCAAGGCGCGGTTTGAGAAGGTGGTGCCGGTTGCGA
CGATGAGTGTAACGGGCCCGGGCGGAACAGGTGAGGAAGTTTATCTGCAGGGCGTGTATGTTGGCATT
GCAAGTTTTGATGGGGCTCGCTCGGGAGTGTTGACGGAAGAGGAGCGTCGGGCAATTTTTGTCGGTCT
TGAGAAGAAAGCGGATGCGGCAATGCAGGAAGTGAAGGCCGCGCGCAAAACACGG*CATCACCATCACC*
*ATCAC*TAAGCATGCAAGCTTGGCTGTTTTG

>**UCD02967.1** MAG: GGDEF domain-containing protein [Candidatus Aenigmatarchaeota archaeon]

MDSNSIESMGKSELREYARELQEEIERLQGEVSTDDLTGVLTRKYFIRELGDLIYRSHREDKMASAGV
IDIDYFKQYNRDGHELGDRVLKGVAEAGRRVLRQYDLIGRYGGDKGDEFLAGILGWEINGKNIIDRWR
REVGTIEVPPGFTTVTITGGVADLNDSPEIQLLRGKRGTDAIMSYKEGEFDRAREYVGVLADWLKTDA
KADIDPIELYGALKRLDRSARDFMESMGAEGDLVTYLEDPDFRYPLAATILTNYADRALMAGKRKKKN
RIWTPMDISSL

template for UCD02967.1

GTACCCGGGGATCCTCTAGAATTAAT**A**GGTAGTGATATTATGGATAGTAACTCTATAGAATCAATGGG
AAAAAGCGAACTAAGAGAATACGCCCGTGAATTACAAGAAGAGATTGAAAGACTCCAGGGCGAAGTAT
CCACCGATGATCTTACAGGCGTATTGACAAGAAAATATTTTATCAGAGAACTTGGGGATCTAATTTAC
AGATCTCACCGTGAGGATAAAATGGCCTCAGCTGGTGTTATCGATATTGACTATTTCAAGCAATATAA
CAGGGACGGACACGAACTCGGTGACAGAGTATTGAAAGGGGTGGCTGAAGCGGGCCGAAGAGTCTTAA
GGCAATACGACCTTATCGGAAGGTATGGTGGAGACAAAGGAGACGAGTTCCTTGCCGGAATTCTCGGA
TGGGAGATAAATGGAAAGAACATAATAGATAGATGGAGACGTGAGGTCGGGACAATTGAGGTGCCGCC
GGGGTTCACAACAGTCACAATCACTGGCGGTGTTGCTGATCTAAACGACAGCCCTGAGATCCAACTGT
TAAGAGGAAAAAGGGGAACGGATGCGATAATGAGCTATAAAGAAGGAGAGTTTGATAGAGCCAGAGAA
TACGTGGGTGTATTGGCCGACTGGCTCAAAACTGATGCAAAGGCAGATATTGACCCGATTGAATTGTA
TGGCGCTTTAAAACGGCTTGATAGAAGCGCCCGCGATTTCATGGAGTCCATGGGGGCCGAAGGGGACT
TGGTTACATACCTGGAGGATCCTGATTTCAGGTACCCGCTAGCTGCAACCATTCTGACTAATTATGCA
GACAGGGCCCTGATGGCCGGAAAGAGAAAGAAGAAGAACAGAATCTGGACGCCTATGGACATATCATC
ATTG*CATCACCATCACCATCAC*TGAGCATGCAAGCTTGGCTGTTTTG

>KYK22704.1 gi|1008839333|gb|KYK22704.1| MAG: hypothetical protein AYK24_00320 [Thermoplasmatales archaeon SG8-52-4]

MALKEYILNNFQYLIENNPDGIVIHSDGIIRYANKSILTIFGASSADEIIGRNIFDFIQPDYHEAAKI
RIAKALNNMEPAPLAEYEAIKLDGAKIYIEILSIPFMRGEAKALQVLVKDISERKRQYKALQYSEEKF
RALFQTAPDAILLSNKETNIISLNSALEDMFGYTYDELIGKSWMLLVAKRLRNDRMERIEYIREKLKA
GEFLKEEVYGTTKFGKDIPVEIKVGRWQINGEPFYTTIIRDISERREFEAKLKQLADIDHLTGIYNRR
AGLILTEQLLKTSRRTNQPITLCYLDLDNFKYVNDSHGHVAGDKVLNAVASLIKDNLRESDIFFRLGG
DEFIIVWNNTFLEQGHQQWSRILNKFNEWNKENYSTRLGISAGFVCPDLKKDINIDRIIDIADKEMYK
SKGNK

>Codon optimized template for KYK22704.1

GTACCCGGGGATCCTCTAGAGCTGACA**A**T**G**TAGTACCACTATGGCGTTAAAAGAATATATTCTGAATAATTTTCAATATCTGA
TTGAAAATAACCCGGATGGCATTGTGATTCACTCGGACGGAATTATCCGCTACGCAAATAAATCGATTCTGACCATTTTTGGT
GCGTCCAGCGCAGATGAAATTATTGGTCGCAATATTTTTGATTTCATTCAGCCTGACTACCATGAAGCTGCCAAGATCCGTAT
CGCTAAAGCACTGAACAATATGGAACCTGCACCTCTGGCCGAGTACGAGGCAATTAAGCTGGACGGGGCTAAAATTTATATTG
AAATTCTGTCAATCCCATTCATGCGCGGTGAAGCCAAAGCCTTGCAAGTATTGGTGAAAGATATTAGTGAACGTAAACGCCAG
TACAAAGCGTTACAATATAGCGAGGAAAAATTCCGCGCCCTGTTTCAAACCGCTCCTGATGCGATTTTACTGAGCAATAAAGA
GACGAATATCATTAGTCTTAATTCAGCGCTGGAAGATATGTTTGGATATACCTATGATGAGCTGATCGGAAAGAGCTGGATGC
TTTTAGTGGCCAAACGCCTGCGCAACGATCGCATGGAGCGTATTGAATACATTCGTGAAAAATTAAAGGCGGGTGAATTTCTG
AAGGAAGAGGTTTATGGGACCACCAAATTCGGCAAAGATATTCCGGTTGAGATTAAAGTTGGCCGTTGGCAGATTAACGGTGA
GCCGTTTTACACTACGATTATTCGTGACATCTCGGAGCGTCGCGAATTCGAAGCGAAACTTAAACAGCTGGCGGATATTGACC
ATCTGACTGGCATCTATAACCGTCGCGCGGGTTTGATCTTAACCGAGCAGCTGCTGAAAACCAGCCGCCGCACCAATCAGCCG
ATTACCCTGTGTTATTTAGATCTTGATAATTTCAAATATGTAAACGACAGCCATGGCCATGTAGCAGGCGATAAGGTTCTGAA
CGCCGTGGCAAGCCTGATCAAAGATAACCTGCGTGAAAGCGATATTTTCTTCCGTCTGGGCGGTGATGAGTTTATTATTGTGT
GGAACAATACGTTCCTGGAACAGGGGCATCAACAGTGGTCGCGTATCCTTAACAAGTTCAACGAGTGGAACAAAGAAAATTAT
AGCACGCGTTTAGGCATCAGCGCCGGTTTCGTTTGTCCTGATCTGAAGAAAGACATCAATATCGATCGCATTATCGATATTGC
CGATAAAGAAATGTACAAAAGTAAAGGCAATAAA*CATCACCATCACCATCAC*TAAGCATGCAAGCTTGGCTGTTTTG

Proteins used for construction of the phylogenetic tree in Figure 5 D

WSPR_PSEAER, Q9HXT9 (*Pseudomonas aeruginosa*); PleD_CAUCR, Q9A5I5 (Caulobacter crescentus); ARS29551_1_SPH_KC8, ARS29551.1 (*Sphingomonas* sp. KC8); STRGAL_GGDEF, WP_012961431.1 (*Streptococcus gallolyticus* UCN34); STRSUI_GGDEF, WP_079269016.1 (*Streptococcus suis*); DGCB_CAUCRE, A0A0H3CAN8 (*Caulobacter crescentus*); DGC1_KOMXYL, O87374 (*Komagataeibacter xylinus*); Y1354_MYCTU Rv1354c/P9WM13 S4 (*Mycobacterium tuberculosis*); ECA3270_PECCAS, Q6D226 (*Pectobacterium atrosepticum*); PROMIR_GGDEF, B3F0R5 (*Proteus mirabilis*); VCA0965_VIBCHO, Q9KKY5 (*Vibrio cholerae*); STRHEN_GGDEF, WP_018163948.1 (*Streptococcus henryi*); ZMO_1055_strainZM4, ZMO1055 (*Zymomonas mobilis* ZM4); WP_084652956_1_SPHHAL, WP_084652956.1 (*Edaphosphingomonas haloaromaticamans* strain P3); TMJ20004_1_ALPHA, TMJ20004.1 (Alphaproteobacteria bacterium AP_24); ABQ66580_1_SPHWIT, ABQ66580.1 (*Sphingomonas wittichii* RW1); BifA_PSEAER, Q9HW35 (*P. aeruginosa*); YdeH_Ecoli, P31129 (*Escherichia coli* MG1655); SE0528_STEAS, Q8CTF5 (*Staphylococcus epidermidis*); CD1420_CLODIF, Q18BU4 (*Clostridioides difficile*); PA0861_PSEAE, Q9I580 (*P. aeruginosa*);); PA4332_PSEAE, Q9HW69 (*P. aeruginosa*); STM1703, STM3388, STM2123, STM2672, STM4551, STM1987, STM1283, STM3375, STM3615 (*Salmonella typhimurium* ATCC14028); AdrA_SALTY, Q9L401 (*S. typhimurium*); EAL_Shenryi_DSM19005, EAL_Shenryi_DSM19005_endPart, WP_018165242.1 (*S. henryi* DSM19005); GSU1658_GEOSL, Q74CL4 (Geobacter sulfurreducens); MXAN_2643, Q1D911 (*Myxococcus xanthus*); In VCA0965, the sequence 'RATNQHDY' was deleted to optimize the alignment.