CMLS, Cell. Mol. Life Sci. 62 (2005) 1234–1246
1420-682X/05/111234-13
DOI 10.1007/s00018-005-4557-x


CMLS Cellular and Molecular Life Sciences

Review

# Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae

**U. Römling**

Karolinska Institutet, Microbiology and Tumor Biology Center (MTC), Box 280, 171 77 Stockholm (Sweden),
Fax: +46 8 330477, e-mail: ute.romling@mtc.ki.se

Received 17 December 2004; received after revision 2 February 2005; accepted 14 February 2005
Available online 07 April 2005

**Abstract.** The rdar morphotype, a multicellular behaviour of *Salmonella enterica* and *Escherichia coli* is characterized by the expression of the adhesive extracellular matrix components cellulose and curli fimbriae. The response regulator CsgD, which transcriptionally activates the biosynthesis of the exopolysaccharide cellulose and curli, also transforms cell physiology to the multicellular state. However, the only role of CsgD in cellulose biosynthesis is the activation of AdrA, a GGDEF domain protein that mediates production of the allosteric activator cyclic-di-(3′-5′)guanylic acid (c-di-GMP). In *S. enterica serovar* Typhimurium a regulatory network consisting of 19 GGDEF/EAL domain-containing proteins tightly controls the concentration of c-di-GMP. c-di-GMP not only regulates the expression of cellulose, but also stimulates expression of adhesive curli and represses various modes of motility. Functions of characterized GGDEF and EAL domain proteins, as well as database searches, point to a global role for c-di-GMP as a novel secondary messenger that regulates a variety of cellular functions in response to diverse environmental stimuli already in the deepest roots of the prokaryotes.

**Key words.** Biofilm; cellulose; curli fimbriae; cyclic di-GMP; EAL domain; *Escherichia coli*; GGDEF domain; *Salmonella enterica.*

## Introduction

The tremendously broad capacity of microorganisms to communicate and co-ordinate their behavioural processes has begun to be fully appreciated in recent years. In *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) and *Escherichia coli* at least three fundamentally different modes of behaviour have been recognized: swimming, swarming and adherence of cells to surfaces or to one another (biofilm formation). The single swimming cell has imprinted the typical view of *S.* Typhimurium and *E. coli* since this originally described phenotype has been extensively studied for more than a century as a culture conventionally grown to mid-logarithmic phase in rotating laboratory flasks. Swarming behaviour was discovered as a co-ordinated behaviour, where cells move in a co-operative fashion over a surface, and whereby cells elongate and show hyperflagellation [1]. Not too long ago, it was recognized that Enterobacteriaceae also form biofilms, multicellular structures, where bacteria adhere to one another and/or to surfaces embedded into a self-produced extracellular matrix [2–5]. Through the study of biofilm formation in various experimental set-ups it has become clear that different kinds of biofilms exist, and that the character of the biofilm might depend on the environmental conditions, the surface characteristics and the particular strain under observation [6–8]. Although swarming behaviour and biofilm formation represent two apparently opposing lifestyles of bacteria characterized by motility and sessility, respectively, the two behavioural modes share several common features. Both take place on surfaces, require co-operativity and the expression of extracellular matrix components for function. Swimming, swarming and biofilm formation have been mainly studied

as separate behavioural modes. But some facets of these complex genetic programmes, which trigger the transition from one behaviour to the other have already been elucidated [9–13].

Biofilm formation has primarily been studied in long-term laboratory-passaged descendants of the *E. coli* strain K-12. Although *E. coli* K-12 has been a wonderful model organism for many years, its use in the study of biofilm formation requires some caution. First, due to its domestication in the laboratory, *E. coli* K-12 lineages frequently lost the ability to produce a full biofilm. Loss of multicellular behaviour during maintenance in the laboratory has also been observed in other bacteria [14]. Second, *E. coli* K-12 lacks a complete lipopolysaccharide (LPS) with long-chain polysaccharide (O-antigen) in the outer membrane, which is in contrast to most pathogenic and commensal strains directly isolated from patients or healthy individuals. Loss of O-antigen and additional LPS structures leads to an enhanced adherence to abiotic surfaces [15–17], probably due to exposure of binding epitopes on the cell surface or alteration of surface charge. Basically all *S.* Typhimurium and *Salmonella enterica* serovar Enteritidis (*S.* Enteritidis) strains as well as a significant number of *E. coli* strains display a multicellular behaviour that is characterized by the expression of the extracellular matrix components curli, a fimbrial component [previously called thin aggregative fibres (*agf*) in *S.enterica*] and/or cellulose [6, 18, 19, unpublished observations]. Strains of other enterobacterial species such as *Citrobacter* spp., *Enterobacter* spp. and *Klebsiella* spp. isolated from the gastrointestinal tract show a species-specific expression of curli and/or cellulose [20]. Vigorous expression of the two extracellular matrix components, proteinaceous curli and exopolysaccharide cellulose, leads to a characteristic colony morphology on agar plates: a flat colony with a rough and dry surface, netlike warpings and an undulate margin, which spreads over the agar plate within a few days. Enhanced visualization of the colony morphology can be achieved by the addition of the planar hydrophobic diazo dye Congo Red to the agar medium. Both extracellular matrix components are able to interact with Congo Red, which results in the colony having a typical dark purple colour (fig. 1). This characteristic morphotype has, therefore, been termed the rdar (red, dry and rough) morphotype. The rdar morphotype was described as early as the beginning of last century [21], but was not extensively studied until recently. A variety of biological roles have been attributed to the adhesive curli and cellulose. These roles range from the establishment of bacterial cell-cell interactions and resistance to disinfectants, to distinct roles in bacterial-host interactions ([22, 23], many of which have recently been reviewed [24]).

In this review I will focus on the emerging molecular characterization of the cell physiology of the rdar morphotype. Here, the expression 'rdar morphotype' is used as a synonym for the multicellular status that is displayed by the cells expressing the extracellular matrix components cellulose and curli. Among other characteristics cells expressing the rdar morphotype exhibit an extensive biofilm formation on abiotic surfaces [4]. Attempts to understand the physiology of the rdar morphotype led to the discovery of a novel regulatory network that utilizes cyclic-di(3′→5′)-guanylic acid (c-di-GMP) as the secondary signaling molecule in *S.* Typhimurium. The cellular turnover of c-di-GMP is mediated by proteins which contain the GGDEF and/or EAL domains that function as diguanylate cyclases and phosphodiesterases. For many years, c-di-GMP was thought to function merely as an activator of cellulose synthase in a fruit-degrading bacterium [25]; however, it has now been shown to be a widely distributed secondary messenger in prokaryotes where it affects a number of other functions and types of behaviour, including the transition from sessility to motility as shown in *S.* Typhimurium [26].

Figure 1. The rdar morphotype of *S.* Typhimurium after 7 days of growth on a Congo Red agar plate [4].

## A regulatory role for CsgD in rdar morphotype expression

Regulation of rdar morphotype expression has mainly been studied in the human pathogen *S.* Typhimurium ATCC14028 and in *E. coli* TOB1, a recently isolated fecal strain [4, 27, unpublished observations].

In *S.* Typhimurium and *E. coli,* expression of the extracellular matrix components cellulose and curli is positively regulated by CsgD (curli subunit gene D), a member of the LuxR superfamily of transcriptional regulators (fig. 2). Transcriptional regulators of the LuxR superfamily are characterized by an N-terminal activation domain and a conserved C-terminal DNA binding domain with a LuxR-like helix-turn-helix motif. In *S.* Typhimurium, CsgD has

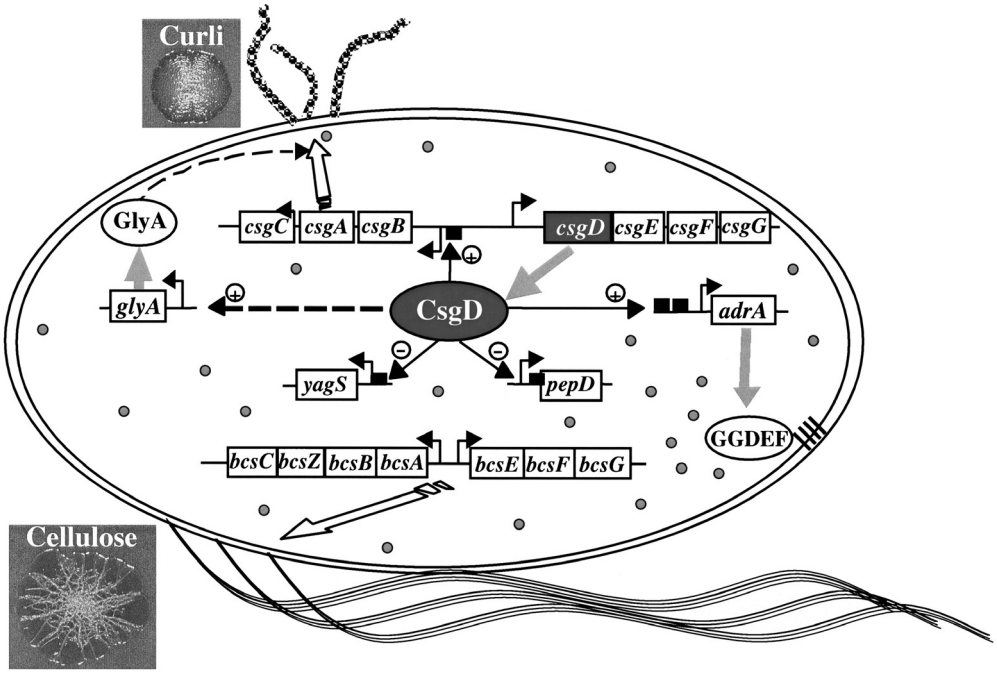



Figure 2. Cell physiology of the rdar morphotype. In the stationary phase of growth expression of CsgD triggers the biosynthesis of the extracellular matrix components curli and cellulose [31]. CsgD activates the transcription from the *csgBA*(*C*) operon coding for the structural components of curli. After transcriptional activation by CsgD, AdrA stimulates production of c-di-GMP by its GGDEF domain [26, 31]. C-di-GMP binds to the BcsB subunit of the cellulose synthase, thus stimulating cellulose biosynthesis [51]. On the other hand, CsgD represses the transcription of the biofilm inhibitors *yagS*, coding for a putative purine dehydrogenase and *pepD* encoding an unspecific dipeptidase [34]. *GlyA* transcription is activated by CsgD through an unknown mechanism. Figure modified after [32]. Reprinted with permission.

previously been referred to as AgfD (thin aggregative fimbriae gene D).

Based on the homology of the N-terminal activation domain of CsgD with the activation domains of BvgA in *Bordetella pertussis* and NarP in *E. coli*, CsgD is classified as a member of the UhpA (FixJ) family of response regulators. FixJ family members are part of prokaryotic signalling systems known as two-component systems, which use phosphor transfer for signal transduction. The N-terminal activation domain of response regulators, also called the receiver domain, is usually phosphorylated at a conserved aspartate at position 58 by a cognate sensor kinase located in the cytoplasmic membrane. CsgD contains an unconventional receiver domain; although Asp58 is present, only two of the five amino acids involved in phosphorylation are conserved. Thus, the mode of CsgD activation remains unknown and a cognate sensor kinase has not been identified.

There are few CsgD homologues encoded by sequenced genomes of microorganisms. Environmental bacteria such as *Microbulbifer degradans* 2-40, *Shewanella oneidensis* MR-1 and *Magnetococcus* sp. MC-1 harbour one homologous gene product. More than one homologue per genome is found in different *Vibrio* species that are human pathogens, and which have their reservoir in the aquatic environment. *Vibrio vulnificus* and *Vibrio parahaemolyticus* have three copies of a CsgD homologue and *Vibrio cholerae* has two copies. *V. cholerae vpsT* encoding a CsgD homologue on chromosome II exhibits enhanced biofilm formation and transcriptional activation of genes required for the production of the $VPS^{EL\ TOR}$ exopolysaccharide [28]. In *V. parahaemolyticus* the CsgD homologue CpsS acts as a repressor that downregulates transcription of the operon encoding capsular polysaccharide genes [29].

In *S.* Typhimurium the *csgD* gene is an integral part of the curli biosynthesis operon *csgDEFG-csgBAC* (fig. 2). CsgD is expressed in the stationary phase of growth, when phosphorous, nitrogen and other (unknown) trace elements are scarce, but sufficient carbon is provided [4, 30]. It has been implicated, but never shown, that CsgD binds to the promoter and activates transcription of the divergently transcribed *csgBAC* operon, which encodes the structural genes required for the formation of curli [27, 31, 32]. On the other hand, CsgD stimulates cellulose biosynthesis indirectly via transcriptional activation of AdrA [31, 33]. Putative conserved CsgD binding sequences of 11 bps have been identified upstream of the *csgBAC* and *adrA* transcriptional start sites, which supports a direct role of CsgD in the transcriptional activation of these two coding regions. Yet CsgD might activate *csgBAC* and *adrA* by distinct mechanisms, since the

putative CsgD binding sites are differentially arranged and the distance of the CsgD binding site(s) from the transcriptional start site is different for these two coding regions [34].

Activation of the biogenesis of the extracellular matrix components curli and cellulose is not the only function of CsgD. CsgD also transforms the cellular physiology, thus favoring biofilm formation and expression of curli.

DNA array analysis in an *E. coli* K-12 strain showed that CsgD strongly downregulated two genes: *pepD,* which encodes a broad-spectrum dipeptidase [35], and *yagS*, which codes for the molybdopterin-binding subunit of an oxidoreductase protein complex encoded by *yagTSRQ* [34]. In accordance with the repressive function of CsgD, putative CsgD binding sites were found downstream of the transcriptional start sites of the two genes. Once expressed, the two gene products prevented biofilm formation. How biofilm formation is inhibited by *pepD* and *yagS* is not yet clear, although it has been speculated that both genes are involved in the degradation of different kinds of signalling molecules that are required for the formation of biofilms [34]. In *S.* Typhimurium, CsgD does not target these two genes since *pepD* does not contain a putative CsgD binding site [unpublished observations], while the *yagTSRQ* genes are not at all present on the chromosome. Whether *S.* Typhimurium harbours genes besides *csgBAC* and *adrA* that are directly regulated by CsgD remains to be shown. It is important to note that the genome-wide search for CsgD regulated genes in *E. coli* had its drawbacks. The array analysis did not identify *yaiC*, the *E. coli* homologue of *adrA* as a target regulated by CsgD, although subsequent analysis of transcriptional fusions to *yaiC* confirmed the findings previously carried out in *S.* Typhimurium [31, 34]. Thus it cannot be excluded that there are more genes that may be directly regulated by CsgD.

Apart from transcriptional activation, CsgD also exerts other functions. Chirwa and Herrington [36] showed that when overexpressed, the first 70 N-terminal amino acids of the 216-amino acid long CsgD in combination with chromosomally encoded CsgD enabled an *E. coli* K-12 laboratory strain not capable of producing curli to restore curli biosynthesis. In this strain the transcription of *glyA* encoding serine hydroxymethyltransferase (SHMT), which is the major source of glycine in the cell, was enhanced. Thus, limited glycine production restricted the biogenesis of curli, which are mainly composed of CsgA subunits containing a glycine-rich N-terminus, but enhancement of the activity of SHMT could relieve repression in this laboratory strain. The authors suggested that a small molecular weight compound binds to the N-terminus of CsgD, which is titrated out by overexpression of the N-terminus, thus enabling enhanced transcription of *glyA* [36].

DNA arrays revealed that CsgD caused a genome-wide up- and downregulation of the transcription of several genes in *E. coli* [34], although no putative CsgD binding site could be found up- or downstream of the respective promoters, suggesting an indirect role for CsgD in the regulation of those genes. The indirect role of CsgD could be exerted through the putative binding function of its N-terminal domain [36] or the activity of the gene products regulated by CsgD [26, 31]. Alternative explanations include an altered cell physiology, which would channel resources into the production of extracellular matrix components or an altered access to oxygen or nutrients caused by the production of extracellular matrix components, which would cause expression of the biofilm phenotype, such as cell clumping and surface adherence.

In conclusion, CsgD regulates the expression of the rdar morphotype by stimulating expression of the extracellular matrix components, curli and cellulose, and transforms cellular physiology in favour of biofilm formation and expression of extracellular matrix components. In accordance with its role as a major switch for the multicellular rdar morphotype behaviour, expression of CsgD itself is controlled by a complex network of globally regulatory proteins on a transcriptional and post-transcriptional level [32, and unpublished observations].

## Regulation of cellulose biosynthesis in *S.* Typhimurium by AdrA and c-di-GMP

Cellulose is an exopolysaccharide consisting of glucose monomers connected by a 1-4-$\beta$-glycosidic bond. The individual linear glucan chains self-assemble to crystalline fibers, macromolecular structures that display the particular properties of cellulose [37]. Two divergently transcribed operons, *yhjRQbcsABZC-bcsEFG*, encode the structural genes required for cellulose biosynthesis. In *S.* Typhimurium (and *E. coli*) the operon *bcsABZ* encodes the structural genes required for cellulose biosynthesis. Sequence homologues to the *bcsABZ* region are common to all bacteria which harbour a homologue of the bacterial-like cellulose synthase [38]. BcsA and BcsB form the cellulose synthase complex, while BcsZ is a cellulase [39]. BcsA, the catalytic subunit of the cellulose synthase, is highly conserved, especially in the cytoplasmically located catalytic domain of 350 amino acids, which contains six conserved amino acid motifs clustered around highly conserved positively or negatively charged amino acids [39]. The other genes of the cellulose biosynthesis operon might be required for accessory features of the cellulose synthase protein complex or the cellulose macromolecule.

Regulation of cellulose biosynthesis by CsgD is indirect. Once CsgD is expressed in the stationary phase of growth, it activates transcription of *adrA* (AgfD regulated gene) and subsequently leads to cellulose biosynthesis [4, 31]. This regulatory network has been validated under various

environmental conditions such as different oxygen tensions and temperatures. Besides activation of *adrA* transcription, CsgD lacks an additional role in cellulose biosynthesis, since subsequent analysis showed that expression of AdrA in a *csgD* knockout background is sufficient to activate cellulose biosynthesis [33]. These results demonstrate that AdrA is the protein crucial for cellulose biosynthesis.

To evaluate the mechanism by which AdrA activates cellulose biosynthesis, transcriptional analysis of the structural genes for cellulose biosynthesis, *bcsA* and *bcsC*, was performed. However, transcription of the *bcs* operon was constitutive and not dependent on CsgD, and consequently AdrA [33], indicating that AdrA acts on a post-transcriptional level to activate cellulose biosynthesis.

AdrA is a modular protein of 371 amino acids with two domains: an N-terminal MASE2 domain [40] and a C-terminal GGDEF domain, also called DUF1 (domain of unknown function 1) domain in the database. The N-terminal MASE2 domain, which has been suggested to be a sensor domain for an unknown stimulus, consists of six trans-membrane helices. The MASE2 domain is only occasionally found in other bacteria and is restricted to the branch of γ-proteobacteria. The GGDEF domain is approximately 175 amino acids long. It received its name from the characteristic GGD(E)EF motif, but highly conserved amino acids are found throughout the domain [41].

Subsequent in vivo studies demonstrated that the expression of AdrA enhanced the intracellular concentration of the novel nucleotide c-di-GMP [26, 42]. c-di-GMP is made up of two molecules of GMP linked 3′→5′ by a phosphodiester bond. Diguanylate cyclase activity has previously been shown to be mediated by three GGDEF/EAL domain proteins (see below) in vitro and in vivo [43]. Lateron, sequence and structural homology between the GGDEF domain and eukaryotic adenylate cyclase has indicated that the GGDEF domain could function as a nucleotide cyclase [44]. On the basis of these previous investigations diguanylate cyclase activity could be directly assigned to AdrA. Replacement of the two glycine residues in the GGDEF motif by alanine resulted in an inactive enzyme. As a consequence, the mutant protein did not enhance the intracellular level of c-di-GMP and did not stimulate cellulose biosynthesis. These in vivo studies showed that c-di-GMP is the effector molecule through which the GGDEF domain-containing protein AdrA activates cellulose biosynthesis in *S.* Typhimurium.

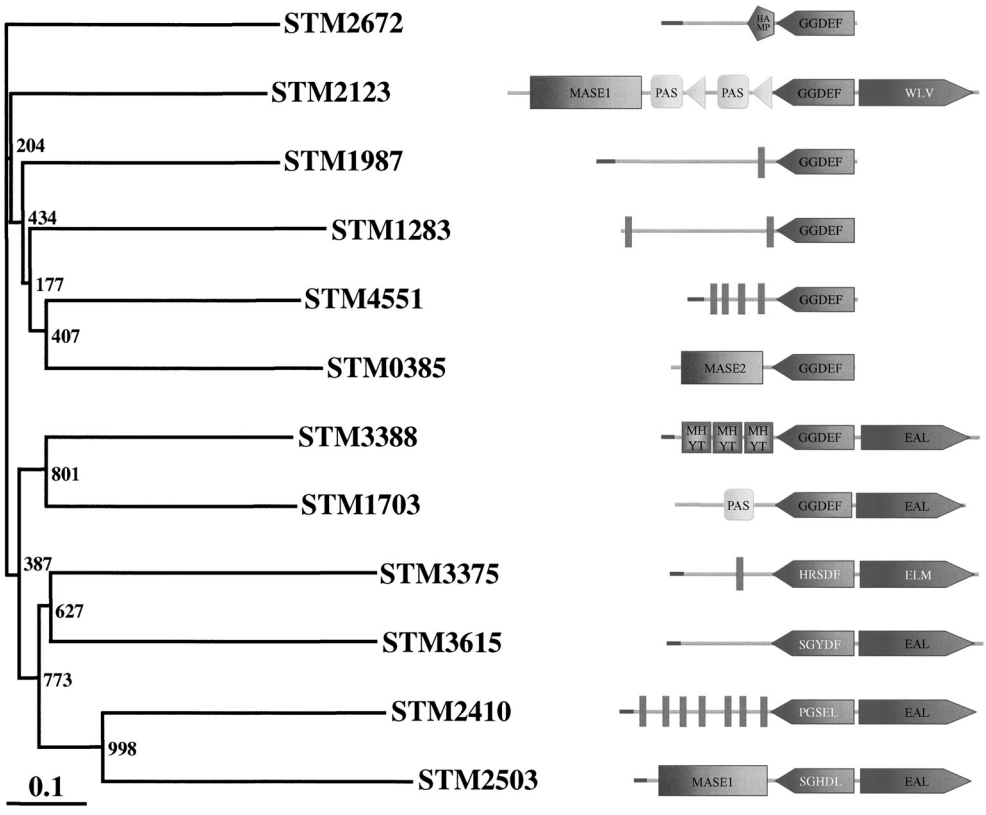
More than 15 years ago, the group of Benziman identified c-di-GMP as an allosteric activator of cellulose synthase in the fruit-utilizing bacterium *Gluconacetobacter xylinus,* using a combination of chemical and biochemical methods [25]. Our recent findings in *S.* Typhimurium have shown that c-di-GMP is a general activator of cellulose biosynthesis in bacteria.

## General role of GGDEF domain proteins in the activation of cellulose biosynthesis

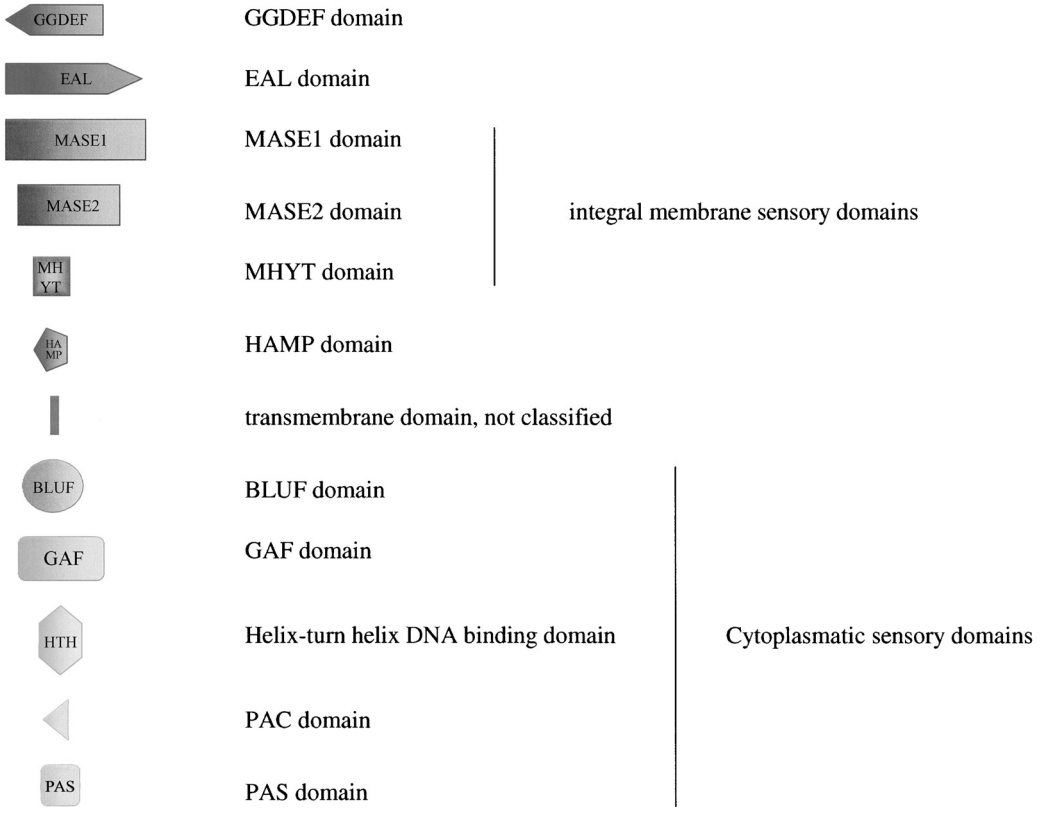
The activation of cellulose biosynthesis in *S.* Typhimurium is not restricted to AdrA. When overexpressed, a panel of other GGDEF domain-containing proteins encoded in the *S.* Typhimurium chromosome (fig. 3) activated cellulose biosynthesis. Activation occurred under growth conditions optimal for rdar morphotype expression [rich laboratory medium (Luria broth agar plates without the salt component)] as well as alternative growth conditions where the rdar morphotype is usually not expressed [26, 45]. The GGDEF domain protein STM1987 was required for the activation of cellulose biosynthesis at 37 °C in a starvation medium. This medium was devoid of trace elements, and therefore did not support active growth of the bacteria, but was rich in glucose as an energy source. Surprisingly AdrA was not able to complement a STM1987 knockout under these conditions, although other GGDEF domain proteins were functional [45]. Since more detailed studies on AdrA expression and function were not reported, any speculation about the significance of this finding would be premature.

Furthermore, activation of cellulose biosynthesis by GGDEF domain proteins is not unique to *S.* Typhimurium. In fact, three highly homologous GGDEF domain-containing proteins gradually activated cellulose biosynthesis in *G. xylinus* [43]. Several other bacterial species also require a GGDEF domain protein for the production of cellulose or a cellulose-like polymer [31, 46, 47]. It has also been shown that proteins from different species, with only the GGDEF domain in common, can activate cellulose biosynthesis in bacteria such as *Rhizobium leguminosarum* and *S.* Typhimurium, thus demonstrating a universal role of the GGDEF domain in the activation of cellulose biosynthesis in different bacteria [26, 48]. When YcdT (in some databases called YhcK) from *E. coli* was expressed in *S.* Typhimurium, c-di-GMP production was shown in accordance with activation of cellulose biosynthesis [26]. On the other hand AdrA is a promiscuous protein that is capable of producing c-di-GMP and subsequently activating exopolysaccharide biosynthesis in bacteria such as *Pseudomonas aeruginosa* and *Yersinia pestis* which do not harbour a cellulose biosynthesis operon [26, 49], but have other operons that may be involved in the biosynthesis of exopolysaccharides. Although production of the secondary messenger molecule c-di-GMP by GGDEF domain proteins has been experimentally proven in only a few cases [26, 42, 50], the available data can be extrapolated to suggest that several of the more than 1300 GGDEF domain proteins in the database have the ability to produce c-di-GMP or related compounds.

The mechanism by which c-di-GMP activates the biosynthesis of cellulose and other polysaccharides needs to be




## Domain symbols



studied in detail. The only study undertaken to address this issue showed binding of c-di-GMP to a 67-kDa fragment derived by proteolytic cleavage of BcsB [51]. The c-di-GMP binding protein BcsB together with the catalytic subunit BcsA forms the core of the cellulose synthase complex in all prokaryotes that harbour bacterial-like cellulose synthases [38, 39].

## The role of elevated c-di-GMP concentrations on bacterial behaviour

As mentioned earlier, the function of c-di-GMP is not restricted to the activation of cellulose biosynthesis. In *S.* Typhimurium and *E. coli* the expression of AdrA (and other GGDEF domain proteins) and the resulting increase in the levels of c-di-GMP also stimulate the expression of curli [unpublished observations]. Thus, the enhancement of biofilm formation, and the rdar morphotype that is observed when AdrA is expressed, is due to increased cellulose as well as curli biosynthesis [26]. Whether c-di-GMP acts post-transcriptionally on the curli biosynthesis genes or stimulates curli expression via CsgD, thus providing a positive feedback loop remains to be established.

On the other hand, motility was inhibited by the expression of AdrA and high levels of c-di-GMP [26]. This holds true for swimming as well as swarming motility. Thus, in conclusion, high c-di-GMP levels favoured bacterial sessility and production of adhesive matrix molecules, and repressed all known modes of motility in *S.* Typhimurium.

## Role of EAL domain proteins in c-di-GMP metabolism and bacterial behaviour

Since high levels of c-di-GMP promoted sessility and repressed motility, low c-di-GMP levels were expected to lead to the opposite behaviour. A putative phosphodiesterase for the breakdown of c-di-GMP was thought to indicate this correlation between c-di-GMP levels and bacterial behaviour. An obvious candidate for such a phosphodiesterase was found in the EAL domain, also called domain of unknown function 2 (DUF2). EAL domains are functionally coupled to GGDEF domains since they are frequently found C-terminal of a GGDEF domain in the same protein (fig. 4). In fact, the six highly homologous diguanylate cyclases and phosphodiesterases required for c-di-GMP synthesis and degradation in *G. xylinus* each contain both effector domains. This finding has made the functional assignment of distinct enzymatic functions to the GGDEF and EAL domains inconclusive [43].

The EAL domain is approximately 250 amino acids in total, with the characteristic EAL motif at the N-terminal end. However, other motifs, such as the DDFGTG motif, which are more conserved and may therefore be more indicative of the function of the domain, are found throughout the domain.

Expression of YhjH (STM3611), a 255-amino acid protein which consists solely of an EAL domain (fig. 5) mediated degradation of c-di-GMP and repression of cellulose biosynthesis in vivo. To achieve high c-di-GMP levels within the cell, the GGDEF domain protein AdrA was overexpressed. Subsequent expression of YhjH resulted in a significantly decreased c-di-GMP concentration [26]. In this experimental setup, it is possible that the EAL domain requires cooperation with the GGDEF domain to degrade c-di-GMP. The replacement of the highly conserved glutamate at position 136 by alanine resulted in abolished phosphodiesterase activity of YhjH in vivo.

The reduction of the c-di-GMP level following expression of YhjH not only affected cellulose biosynthesis, but also the whole rdar morphotype behaviour including biofilm formation was abolished. On the other hand, both modes of motility, swimming as well as swarming, were enhanced by YhjH overexpression. Thus expression of the EAL domain has precisely the opposite effect on sessility, expression of adhesive matrix components and motility compared with the GGDEF domain. The role of the EAL domain in c-di-GMP degradation was shown to be a general mechanism by the expression of another EAL domain protein from *S.* Typhimurium, STM1827, which triggered the same phenotypes [26]. Comparison of the phenotypes mediated by YhjH and STM1827 indicated that STM1827 was more effective in the degradation of c-di-GMP than YhjH under these growth conditions, although a precise analysis of the enzymatic activity of the two proteins is required to show this formally.

In conclusion, recent genetic and analytical data have shown that c-di-GMP has a general role in the regulation

Figure 3. Phylogram of the relationships between the 12 GGDEF domains of proteins encoded by the *S.* Typhimurium genome. On the right, the domain structure of the GGDEF domain proteins are shown. Knockout of AdrA (STM0385) and STM1987 abolished cellulose biosynthesis in LB without salt and ATM medium, respectively [31, 45]. Four GGDEF domain proteins, STM1283, STM2123, STM3388 and STM4551, activated cellulose biosynthesis in *S.* Typhimurium ATCC14028, when overexpressed. However, no phenotypes were reported for the knockout mutants [45]. GGDEF domains of proteins STM2410, STM2503, STM3375 and STM3615 do not show a consensus GGDEF motif, as indicated in the figure by pale letters. Sequences were aligned in Clustalx using the default mode and afterwards processed manually. The tree was constructed using the neighbourhood joining (NJ) method [65] and subjected to 1000 bootstrap trials. Numbers on nodes indicate percentages of bootstrap values. The tree was drawn with TreeView [66]. Domain symbols were taken from the SMART database [67] and modified.

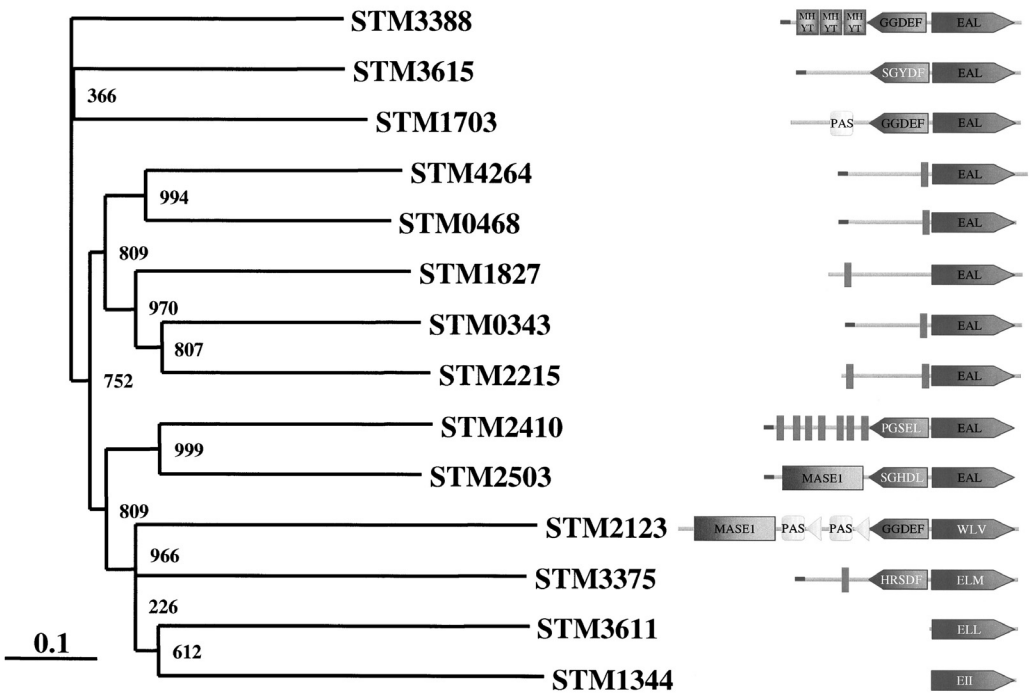



Figure 4. Phylogram of relationships between the 14 EAL domains of proteins encoded by the *S.* Typhimurium genome. On the right, the domain structure of the EAL domain proteins are shown. Three subgroups of EAL domains can be discriminated. STM3611 and STM1827 EAL domain proteins representing two different subgroups were shown to stimulate degradation of c-di-GMP and suppression of cellulose biosynthesis [26]. The EAL domain proteins STM1344, STM2123 and STM3375 do not show a complete signature of conserved amino acids, as indicated in the figure by pale letters. Sequences were aligned in Clustalx using the default mode and afterwards processed manually. The tree was constructed using the neighbourhood joining (NJ) method [65] and subjected to 1000 bootstrap trials. Numbers on nodes indicate percentages of bootstrap values. The tree was drawn with TreeView [66].

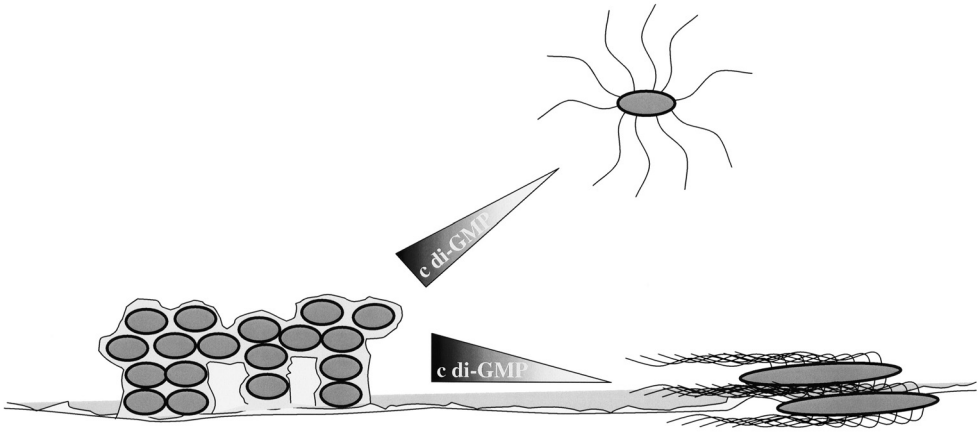



Figure 5. c-di-GMP concentrations in sessile and motile cells. In sessile cells c-di-GMP concentrations are high, thus triggering the expression of adherent matrix components such as cellulose and curli. In motile cells, c-di-GMP concentrations are low. Those cells exist either as single free-swimming (planktonic) cells or as highly differentiated swarmer cells. The signals which trigger the development into either a swimmer or a swarmer cell have not yet been resolved.

of exopolysaccharide synthesis (for example cellulose), biosynthesis of adhesive components (curli) as well as swimming and swarming motility (fig. 5). In addition, gene products with GGDEF and EAL domains play a role in the synthesis and degradation of c-di-GMP, respectively. Whether other gene products are directly involved in c-di-GMP metabolism remains to be shown. In *G. xylinus* only 10% of the c-di-GMP is in the cytoplasm accessible to the activity of phosphodiesterase, while 90% is inaccessible through its binding to a membrane-located c-di-GMP binding protein, which is distinct from the cellulose synthase subunit BcsB [52].

## Complexity of c-di-GMP metabolism in *S.* Typhimurium

In *S.* Typhimurium, 12 proteins with a GGDEF domain and 14 proteins with an EAL domain were identified by database searches: 7 of these proteins belong to both groups, since they contain both domains (figs. 3, 4, 6). The genome of *E. coli* encodes even more proteins with these domains (19 GGDEF and 18 EAL domain proteins, whereby 7 have both domains; fig. 7). One gene encoding an EAL domain protein resides on the fertility plasmid. A significantly higher diversity of temperature and expression variants of the rdar morphotype is observed in *E. coli* in contrast to *S.* Typhimurium strains [6, unpub-

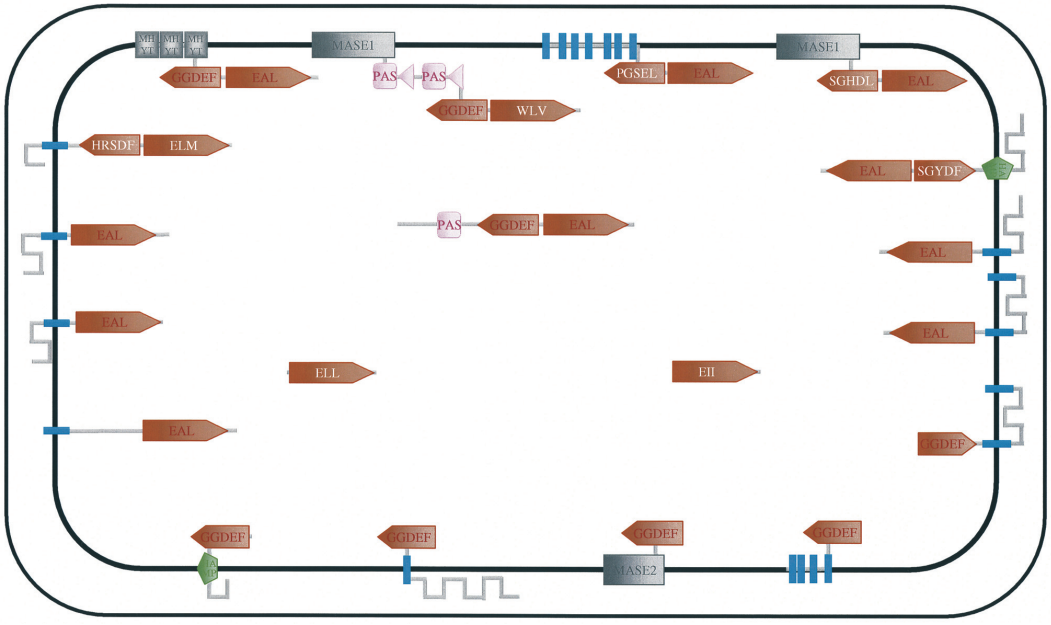



Figure 6. Predicted sub-cellular localization of GGDEF and EAL domain proteins in *S.* Typhimurium.

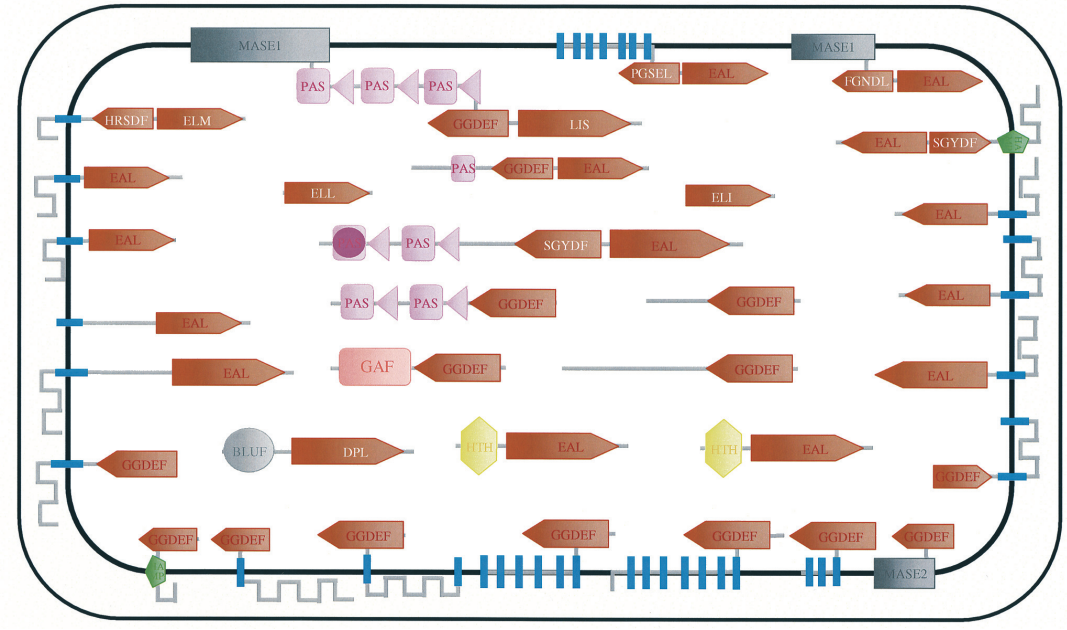



Figure 7. Predicted sub-cellular localization of GGDEF and EAL domain proteins in *E. coli* K12. GGDEF and EAL domain proteins present in *E. coli,* but not *S.* Typhimurium are drawn on a larger scale. Oxygen binding to the first PAS domain of DOS [58] is indicated by a red dot.

lished observations]. This may be a consequence of the more complex c-di-GMP signaling network as suggested by the increased numbers of GGDEF and EAL domain proteins in *E. coli*.

Although it cannot be excluded that some of the GGDEF/EAL domains have a slightly different function, my group has shown that unrelated GGDEF and EAL domain proteins in these two species are involved in c-di-GMP metabolism [26]. Additional proof for a general role of GGDEF and EAL domain proteins in c-di-GMP metabolism comes from the in vitro and in vivo activities of those proteins in *Caulobacter crescentus, P. aeruginosa, V. cholerae* and *Y. pestis* [26, 42, 49, 50, 53, unpublished observations]. Thus, the evidence available at present suggests that metabolism of c-di-GMP is surprisingly complex.

How could such a complex regulatory network have been overlooked? EAL domain proteins have popped up occasionally in the literature, some of these reports being somewhat early in the era of molecular biology. They were suggested to be involved in several phenotypes of *E. coli* such as resistance to bacteriophages $\lambda$ and N14, enhanced transposition of Tn21 and mercury resistance [54–56]. However, from their sequence a function for these proteins could not be assigned, and the abundance of EAL proteins could not be recognized before the era of whole-genome sequencing.

Besides their number and redundancy of function, there is further complexity on the level of the individual GGDEF/EAL protein sequences (figs. 3, 4). In *S.* Typhimurium, the proteins YhjH and STM1344 are made up solely of an EAL domain. The other GGDEF and EAL proteins have a modular structure with at least one up to multiple N-terminal domains (figs. 4, 6, 7). Common are integral membrane domains of which only the MASE1, MASE2 and MHYT domains have been defined [40, 57]. A sensing function has been suggested for all these domains. Membrane topology of the MHYT domain indicated that it may be involved in the sensing of oxygen, NO or CO [57]. Other N-terminal domains are predicted to consist of large periplasmic loops with uncharacterized sensing function. The PAS domain (first recognized in the eukaryotic Per, ARNT and Sim proteins) that binds small molecules is the only cytoplasmic sensing domain coupled to the GGDEF/EAL domains in *S.* Typhimurium (fig. 6). Indeed, the PAS domain of DOS, a GGDEF/EAL domain containing *E. coli* protein is a heme-binding oxygen sensor ([58], fig. 7). Extension of this line of study had shown that oxygen binding to the respective PAS domain of the homologous protein from *G. xylinus* stimulated phosphodiesterase activity [59]. In *E. coli* the spectrum of proteins containing cytoplasmic N-terminal sensing domains is broader than in *S.* Typhimurium (fig. 7). Besides PAS, GGDEF/EAL proteins have a GAF domain [first recognized in mammalian cyclic GMP (cGMP) binding, in *Anabaena* adenylyl cyclase, and *E. coli* FhlA], which binds small molecules such as cGMP or cyclic AMP (cAMP). YcgF has even a BLUF domain, which senses blue light and the redox status [60, 61]. The functional significance of a combined EAL domain and N-terminal DNA binding domain with a LuxR type helix-turn-helix motif remains to be explored. Thus, from the domain structure it can be concluded that the activity of GGDEF/EAL domain-containing proteins can be regulated by the integration of environmental as well as physiological signals, which are sensed by periplasmic, integral membrane or cytoplasmic domains (figs. 6, 7).

Since a set of proteins contains a GGDEF as well as an EAL domain, and the GGDEF domain is commonly found N-terminal to the EAL domain (figs. 3, 4), then domains of opposite function in the same protein might ensure elevated concentrations of c-di-GMP to stay spatially restricted. One might imagine a scenario in which as soon as c-diGMP is produced, it will be degraded. Another mechanism to create a spatial concentration gradient for c-di-GMP is the localization or redirection of GGDEF domain proteins to certain compartments within the cell [42, 62].

In *S.* Typhimurium the most complex protein with GGDEF/EAL domains is STM2123, which has a MASE1 domain, an integral membrane domain with eight transmembrane helices, at the N-terminus, two PAS-PAC domains as cytoplasmic sensory domains, and a GGDEF and an EAL domain as C-terminal output domains (figs. 3, 4). Although such a complex protein is the exception in *S.* Typhimurium and *E. coli,* more complex GGDEF/EAL domain proteins are found in free-living aquatic Gram-negative bacteria such as *P. aeruginosa* and *V. vulnificus*, which contain relative to their genome size even more GGDEF and EAL domain proteins.

## Regulatory network for c-di-GMP in other prokaryotes

The c-di-GMP signalling network seems to be present in bacteria alone, since archaea lack homologues of GGDEF and EAL domains. GGDEF and EAL domains are already present in the genomes of bacterial species located in the deepest branches of prokaryotes, suggesting that c-di-GMP is an ancient secondary messenger. However, c-di-GMP signalling is not essential and has been subsequently reduced or lost in individual species. Although there is the tendency that host-adapted bacteria have fewer GGDEF/EAL domain proteins than free-living bacteria, the factors that determine the abundance of GGDEF/EAL domain proteins are not entirely clear. The $\delta$-proteobacteria *Wolinella succinogenes*, a commensal in the gastrointestinal tract of cattle, codes for 19 GGDEF and 13 EAL domain proteins, yet its close relatives *Helicobacter pylori*, a colonizer of the human stomach, and *Campy-*

*lobacter jejunii*, a commensal of avians and human pathogen, either have none, or only one protein that does not seem to be functional. Interestingly, *W. succinogenes* has a genome size of 2.1 Mbp, which is less than half the size of the genomes of *S.* Typhimurium and *E. coli*. GGDEF/EAL domain proteins are less abundant in Gram-positive bacteria compared with Gram-negative bacteria. Few prokaryotes, such as the obligate intracellular or highly host-adapted bacteria such as *Chlamydia, Haemophilus influenzae* and *Streptococci*, and the free-living cyanobacteria *Gloeobacter violoceus* and *Prochlorococcus marinus*, the dominant photosynthetic organism in the ocean, entirely lack GGDEF/EAL domain proteins.
Homologous proteins have rarely been found in prokaryotes: as such the discovery of the c-di-GMP network came as a surprise. Still, one can ask, what the precise function of such a sophisticated regulatory network for the regulation of c-di-GMP concentration is. Is the system highly redundant, as some complementation data suggest, or are there distinct functions for each of these proteins? What is already evident is that different networks for the expression of GGDEF domain proteins exist since CsgD is only able to stimulate the transcription of AdrA [45], at least under the standard conditions of rdar morphotype expression. Thus there is a complex temporal and spatial expression of individual proteins, the activity of which can be further regulated by environmental cues. As a consequence, cellulose biosynthesis can be uncoupled from the multicellular rdar morphotype behaviour [45]. In any case, this highly abundant signalling pathway has been overlooked in bacteriological research for more than a century.

## Conclusions and future perspectives

Studies of the regulation of multicellular behaviour in *S.* Typhimurium, the rdar morphotype, led to the discovery of the highly abundant c-di-GMP signalling and sensing network that is present already in the oldest forms of prokaryotes. This ancient signalling pathway, which regulates the transition between sessility and motility and other morphological changes [61, 63, 64], was subsequently lost in highly host-adapted, but also some environmental bacteria. There are striking similarities between the prokaryotic regulatory network of c-di-GMP signalling and the regulatory networks, which make use of the secondary messengers cAMP and cGMP in eukaryotic cells. A variety of phenotypes such as cell morphology and development are regulated by cyclic nucleotides in these two branches of the tree of life. The proteins, which balance synthesis and breakdown of the cyclic nucleotides, are representatives of large superfamilies of enzymes, which show a modular structure, a distinct subcellular location, tightly regulated temporal expression and activation of enzymatic function by a variety of stimuli. Intriguingly, these enzymes are associated with the same sensing and activation domains such as PAS, GAF and BLUF in both prokaryotes and eukaryotes. Although the fundamental principles of the complex c-di-GMP signalling network in prokaryotes have been recognized, the challenge remains to precisely define the diverse roles of the c-di-GMP signalling network in the context of prokaryotic cellular physiology and behaviour.

*Acknowledgements.* I thank Inder Saxena for critically reading this review. This work in the laboratory of U.R. is supported by the Karolinska Institutet (‘elitforskartjänst’) and the Mukoviszidose e.V.

1 Harshey R. M. (1994) Bees aren’t the only ones: swarming in gram-negative bacteria. Mol. Microbiol. **13:** 389–394
2 Pratt L. A. and Kolter R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. Mol. Microbiol. **30:** 285–293
3 Genevaux P., Müller S. and Bauda P. (1996) A rapid screening procedure to identify mini-Tn10 insertion mutants of *Escherichia coli* K-12 with altered adhesion properties. FEMS Microbiol. Lett. **142:** 27–30
4 Römling U., Sierralta W. D., Eriksson K. and Normark S. (1998) Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the *agfD* promoter. Mol. Microbiol. **28:** 249–264
5 Vidal O., Longin R., Prigent-Combaret C., Dorel C., Hooreman M. and Lejeune P. (1998) Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new *ompR* allele that increases curli expression. J. Bacteriol. **180:** 2442–2449
6 Römling U., Bokranz W., Rabsch W., Zogaj X., Nimtz M. and Tschäpe H. (2003) Occurrence and regulation of the multicellular morphotype in *Salmonella* serovars important in human disease. Int. J. Med. Microbiol. **293:** 273–285
7 Prouty A. M. and Gunn J. S. (2003) Comparative analysis of *Salmonella enterica* serovar Typhimurium biofilm formation on gallstones and on glass. Infect. Immun. **71:** 7154–7158
8 Boddicker J. D., Ledeboer N. A., Jagnow J., Jones B. D. and Clegg, S. (2002) Differential binding to and biofilm formation on, HEp-2 cells by *Salmonella enterica* serovar Typhimurium is dependent upon allelic variation in the *fimH* gene of the *fim* gene cluster. Mol. Microbiol. **45:** 1255–1265
9 Wolfe A. J., Chang D. E., Walker J. D., Seitz-Partridge J. E., Vidaurr, M. D., Lange C. F. et al. (2003) Evidence that acetyl phosphate functions as a global signal during biofilm development. Mol. Microbiol. **48:** 977–988
10 Cano D. A., Dominguez-Bernal G., Tierrez A., Garcia-Del Portillo F. and Casadesus J. (2002) Regulation of capsule synthesis and cell motility in *Salmonella enterica* by the essential gene *igaA*. Genetics **162:** 1513–1523
11 Jackson D. W., Suzuki K., Oakford L., Simecka J. W., Hart M. E. and Romeo T. (2002) Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. J. Bacteriol. **184:** 290–301
12 Li X., Rasko D. A., Lockatell C. V., Johnson D. E. and Mobley H. L. (2001) Repression of bacterial motility by a novel fimbrial gene product. EMBO J. **20:** 4854–4862
13 Clegg S. and Hughes K. T. (2002) FimZ is a molecular link between sticking and swimming in *Salmonella enterica* serovar Typhimurium. J. Bacteriol. **184:** 1209–1213
14 Branda S. S., Gonzalez-Pastor J. E., Ben-Yehuda S., Losick R. and Kolter R. (2001) Fruiting body formation by *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **98:** 11621–11626

15 Landini P. and Zehnder A. J. (2002) The global regulatory *hns* gene negatively affects adhesion to solid surfaces by anaerobically grown *Escherichia coli* by modulating expression of flagellar genes and lipopolysaccharide production. J. Bacteriol. **184:** 1522–1529

16 Thomsen L. E., Chadfield M. S., Bispham J., Wallis T. S., Olsen J. E. and Ingmer H. (2003) Reduced amounts of LPS affect both stress tolerance and virulence of *Salmonella enterica* serovar Dublin. FEMS Microbiol. Lett. **228:** 225–231

17 Mireles J. R. 2nd, Toguchi A. and Harshey R. M. (2001) *Salmonella enterica* serovar typhimurium swarming mutants with altered biofilm-forming abilities: surfactin inhibits biofilm formation. J. Bacteriol. **183:** 5848–5854

18 Ben Nasr A., Olsen A., Sjöbring U., Müller-Esterl W. and Björck L. (1996) Assembly of human contact phase proteins and release of bradykinin at the surface of curli-expressing *Escherichia coli*. Mol. Microbiol. **20:** 927–935

19 Bian Z., Brauner A., Li Y. and Normark S. (2000) Expression of and cytokine activation by *Escherichia coli* curli fibers in human sepsis. J. Infect. Dis. **181:** 602–612

20 Zogaj X., Bokranz W., Nimtz M. and Römling U. (2003) Production of cellulose and curli fimbriae by members of the family Enterobacteriaceae isolated from the human gastrointestinal tract. Infect. Immun. **71:** 4151–4158

21 von Lingelsheim A. (1913) Frage der Variation der Typhusbacillen und verwandter Gruppen. Centralbl. Bakterio. (Orig) **68:** 577–582

22 Cogan T. A., Jorgensen F., Lappin-Scott H. M., Benson C. E., Woodward M. J. and Humphrey T. J. (2004) Flagella and curli fimbriae are important for the growth of *Salmonella enterica* serovars in hen eggs. Microbiology **150:** 1063–1071

23 Persson K., Russell W., Mörgelin M. and Herwald H. (2003) The conversion of fibrinogen to fibrin at the surface of curliated *Escherichia coli* bacteria leads to the generation of proinflammatory fibrinopeptides. J. Biol. Chem. **278:** 31884–31890

24 Römling U., Bokranz W., Gerstel U., Lünsdorf H., Nimtz M., Rabsch W. et al. (2003) Dissection of the genetic pathway leading to multicellular behaviour in *Salmonella enterica* serotype Typhimurium and other Enterobacteriaceae. In: Medical Implication of Biofilms, pp. 231–261, Wilson, M. and Devin, D. (eds), Cambridge University Press, Cambridge

25 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

26 Simm R., Morr M., Kader A., Nimtz M. and Römling U. (2004) GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. Mol. Microbiol. **53:** 1123–1134

27 Hammar M., Arnqvist A., Bian Z., Olsen A. and Normark S. (1995) Expression of two *csg* operons is required for production of fibronectin- and congo red-binding curli polymers in *Escherichia coli* K-12. Mol. Microbiol. **18:** 661–670

28 Casper-Lindley C. and Yildiz F. H. (2004) VpsT is a transcriptional regulator required for expression of *vps* biosynthesis genes and the development of rugose colonial morphology in *Vibrio cholerae* O1 El Tor. J. Bacteriol. **186:** 1574–1578

29 Guvener Z. T. and McCarter L. L. (2003) Multiple regulators control capsular polysaccharide production in *Vibrio parahaemolyticus*. J. Bacteriol. **185:** 5431–5441

30 Gerstel U. and Römling U. (2001) Oxygen tension and nutrient starvation are major signals that regulate *agfD* promoter activity and expression of the multicellular morphotype in *Salmonella typhimurium*. Environ. Microbiol. **3:** 638–648

31 Römling U., Rohde M., Olsen A., Normark S. and Reinköster J. (2000) AgfD, the checkpoint of multicellular and aggregative behaviour in *Salmonella typhimurium* regulates at least two independent pathways. Mol. Microbiol. **36:** 10–23

32 Gerstel U. and Römling U. (2003) The *csgD* promoter, -control unit for biofilm formation in *Salmonella typhimurium*-. Res. Microbiol. **154:** 659–667

33 Zogaj X., Nimtz M., Rohde M., Bokranz W. and Römling U. (2001) The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. Mol. Microbiol. **39:** 1452–1463

34 Brombacher E., Dorel C., Zehnder A. J. and Landini P. (2003) The curli biosynthesis regulator CsgD co-ordinates the expression of both positive and negative determinants for biofilm formation in *Escherichia coli*. Microbiology **149:** 2847–2857

35 Miller C. G. (1996) Protein degradation and proteolytic modification. In: *Escherichia coli* and *Salmonella*, pp. 938–954, Neidhardt F. C. (ed.), ASM Press, Washington, DC

36 Chirwa N. T. and Herrington M. B. (2003) CsgD, a regulator of curli and cellulose synthesis, also regulates serine hydroxymethyltransferase synthesis in *Escherichia coli* K-12. Microbiology **149:** 525–535

37 Ross P., Mayer R. and Benziman M. (1991) Cellulose biosynthesis and function in bacteria. Microbiol. Rev. **55:** 35–58

38 Nobles D. R., Romanovicz D. K. and Brown R. M. Jr (2001) Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? Plant Physiol. **127**: 529–542

39 Römling U. (2002) Molecular biology of cellulose production in bacteria. Res. Microbiol. **153:** 205–212

40 Nikolskaya A. N., Mulkidjanian A. Y., Beech I. B. and Galperin M. Y. (2003) MASE1 and MASE2: two novel integral membrane sensory domains. J. Mol. Microbiol. Biotechnol. **5:** 11–16

41 Galperin M. Y., Nikolskaya A. N. and Koonin E. V. (2001) Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol. Lett. **203:** 11–21

42 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

43 Tal R.. Wong H. C., Calhoon R., Gelfand D., Fear A. L., Volman G. et al. (1998) Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. J. Bacteriol. **180:** 4416–4425

44 Pei J. and Grishin N. V. (2001) GGDEF domain is homologous to adenylyl cyclase. Proteins **42:** 210–216

45 Garcia B., Latasa C., Solano C., Portillo F. G., Gamazo C. and Lasa I. (2004) Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. Mol. Microbiol. **54:** 264–277

46 Ausmees N., Jonsson H., Höglund S., Ljunggren H. and Lindberg M. (1999) Structural and putative regulatory genes involved in cellulose synthesis in *Rhizobium leguminosarum* bv. *trifolii*. Microbiology **145:** 1253–1262

47 Spiers A. J., Kahn S. G., Bohannon J., Travisano M. and Rainey P. B. (2002) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. Genetics **161:** 33–46

48 Ausmees N., Mayer R., Weinhouse H., Volman G., Amikam D., Benziman M. et al. (2001) Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. FEMS Microbiol. Lett. **204:** 163–167

49 Kirillina O., Fetherston J. D., Bobrov A. G., Abney J. and Perry R. D. (2004) HmsP, a putative phosphodiesterase, and HmsT, a putative diguanylate cyclase, control Hms-dependent biofilm formation in *Yersinia pestis*. Mol. Microbiol. **54:** 75–88

50 Tischler A. D. and Camilli A. (2004) Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. Mol. Microbiol. **53:** 857–869

51 Mayer R., Ross P., Weinhouse H., Amikam D., Volman G., Ohana P. et al. (1991) Polypeptide composition of bacterial

cyclic diguanylic acid-dependent cellulose synthase and the occurrence of immunologically crossreacting proteins in higher plants. Proc. Natl. Acad. Sci. USA **88:** 5472–5476
52 Weinhouse H., Sapir S., Amikam D., Shilo Y., Volman G., Ohana P. et al. (1997) c-di-GMP-binding protein, a new factor regulating cellulose synthesis in *Acetobacter xylinum*. FEBS Lett, **416:** 207–211
53 Chan C., Paul R., Samoray D., Amiot N. C., Giese B., Jenal U. et al. (2004) Structural basis of activity and allosteric control of diguanylate cyclase. Proc. Natl. Acad. Sci. USA **101:** 17084–17089
54 Chae K. S. and Yoo O. J. (1986) Cloning of the lambda resistant genes from *Brevibacterium albidum* and *Proteus vulgaris* into *Escherichia coli*. Biochem. Biophys. Res. Commun. **140:** 1101–1105
55 Hyde D. R. and Tu C. P. (1985) *tnpM*: a novel regulatory gene that enhances Tn21 transposition and suppresses cointegrate resolution. Cell **42:** 629–638
56 Brown N. L., Misra T. K., Winnie J. N., Schmidt A., Seiff M. and Silver S. (1986) The nucleotide sequence of the mercuric resistance operons of plasmid R100 and transposon Tn501: further evidence for *mer* genes which enhance the activity of the mercuric ion detoxification system. Mol. Gen. Genet. **202:** 143–151
57 Galperin M. Y., Gaidenko T. A., Mulkidjanian A. Y., Nakano M. and Price C. W. (2001) MHYT, a new integral membrane sensor domain. FEMS Microbiol. Lett. **205:** 17–23
58 Delgado-Nixon V. M., Gonzalez G. and Gilles-Gonzalez M. A. (2000) Dos, a heme-binding PAS protein from *Escherichia coli*, is a direct oxygen sensor. Biochemistry **39:** 2685–2691
59 Chang A. L., Tuckerman J. R., Gonzalez G., Mayer R., Weinhouse H., Volman G. et al. (2001) Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. Biochemistry **40:** 3420–3426
60 Gomelsky M. and Klug G. (2002) BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. Trends Biochem. Sci. **27:** 497–500
61 Han Y., Braatsch S., Osterloh L. and Klug G. (2004) A eukaryotic BLUF domain mediates light-dependent gene expression in the purple bacterium *Rhodobacter sphaeroides* 2.4.1. Proc. Natl. Acad. Sci. USA **101:** 12306–12311
62 Huang B., Whitchurch C. B. and Mattick J. S. (2003) FimX, a multidomain protein connecting environmental signals to twitching motility in *Pseudomonas aeruginosa*. J. Bacteriol. **185:** 7068–7076
63 Thomas C., Andersson C. R., Canales S. R. and Golden S. S. (2004) PsfR, a factor that stimulates *psbA1* expression in the cyanobacterium *Synechococcus elongatus* PCC 7942. Microbiology **150:** 1031–1040
64 Gronewold T. M. and Kaiser D. (2001) The act operon controls the level and time of C-signal production for *Myxococcus xanthus* development. Mol. Microbiol. **40:** 744–756
65 Saitou N. and Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425
66 Page R. D. (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:** 357–358
67 Letunic I., Copley R. R., Schmidt S., Ciccarelli F. D., Doerks T., Schultz J. et al. (2004) SMART 4.0: towards genomic data integration. Nucleic Acids Res.: **32 Database issue:** D142–144