Minireview

Class III nucleotide cyclases in bacteria and archaebacteria: lineage-specific expansion of adenylyl cyclases and a dearth of guanylyl cyclases

Avinash R. Shenoy*, Sandhya S. Visweswariah

Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore 560012, India

Accepted 29 January 2004

First published online 12 February 2004

Edited by Robert B. Russell

Abstract The Class III nucleotide cyclases are found in bacteria, eukaryotes and archaebacteria. Our survey of the bacterial and archaebacterial genome and plasmid sequences identified 193 Class III cyclase genes in only 29 species, of which we predict the majority to be adenylyl cyclases. Interestingly, several putative cyclase genes were found to have non-conserved substrate specifying residues. Ancestors of the eukaryotic C1-C2 domain containing soluble adenylyl cyclases as well as the protist guanylyl cyclases were found in bacteria. Diverse domains were fused to the cyclase domain and phylogenetic analysis indicated that most proteins within a single cluster have similar domain compositions, emphasising the ancient evolutionary origin and versatility of the cyclase domain.

Key words: Adenylyl cyclase; Guanylyl cyclase; Bacterial genome; Archaebacterium; cAMP; cGMP

1. Introduction

Adenylyl and guanylyl cyclases are important enzymes that catalyse the formation of second messengers that regulate several signalling pathways. The ability to generate cAMP from ATP is shared by six different families of proteins that have no sequence similarity with each other [1]. Class I enzymes are involved in regulation of catabolite repression in enteric bacteria [1]. The second class comprises the toxins secreted by pathogens such as *Bacillus anthracis*, *Bordetella pertussis* and *Pseudomonas aeruginosa* [1]. Class III cyclases are found in bacteria, archaea and eukaryotes and therefore this class has also been called the 'Universal Class' [1]. Mammalian guanylyl and adenylyl cyclases share a high degree of sequence similarity and are members of this class, since the nucleotide specificity appears to be selected by a few residues that allow binding of either the adenine or the guanine purine rings [2,3]. The Classes IV, V and VI of nucleotide cyclases are exemplified by cyclases from *Aeromonas hydrophila*, *Prevotella ruminicola* and *Rhizobium etli* respectively [1,4,5].

The Class III cyclases are the best studied, both biochemically and structurally. A metal co-factor, usually Mg^{2+} or Mn^{2+}, is required for catalysis [2]. Structure elucidation of a mammalian Class III adenylyl cyclase showed that the catalytic site is formed at the interface of head-to-tail dimerisation of two Class III cyclase domains [2]. In eukaryotic adenylyl cyclases, there are two tandem cyclase domains (C1 and C2) within the polypeptide chain, which come together to form the active site. Since these two cyclases domains are similar to each other but not identical, they give rise to only a single active site [2]. As seen in the crystal structure, the second (pseudo-symmetric) site is occupied by the adenylyl cyclase activator, forskolin [2]. The two metal atoms required for catalysis are co-ordinated by two conserved aspartate residues found in the C1 domain [6], and the transition state is stabilised by a critical asparagine-arginine pair present in the C2 domain [7]. Mutations in these residues lead to inactivation of the enzyme [8]. Thus, residues can be classified as C1-like or C2-like, especially when describing corresponding residues in the homodimeric eukaryotic receptor guanylyl cyclases (Fig. 1A) [2].

Substrate specificity, as deduced from the crystal structure of the adenylyl cyclase and homology modelling of guanylyl cyclases, identified a lysine-aspartate pair of residues in adenylyl cyclases that are responsible for ATP binding, while guanylyl cyclases contained glutamate and cysteine residues at equivalent positions [9]. In addition, a glutamine residue found in mammalian adenylyl cyclases (correspondingly an arginine in guanylyl cyclases) has been shown to play an accessory role in substrate selection [9–11]. The substrate specifying pair of residues are present in the C2 domain of adenylyl cyclases and the β-subunit of the heterodimeric soluble guanylyl cyclases [9–11]. Additional residues have also been implicated in determining substrate specificity [12].

Several bacterial Class III cyclases have been characterised, and the role of cAMP in the pathogenesis of two important bacterial diseases, pertussis and anthrax, is well understood [13,14]. The aetiological agents of these diseases operate by raising cAMP levels in the host cytosol by means of a secreted adenylyl cyclase toxin. Cyclic AMP was shown to regulate virulence pathways in *Pseudomonas* [15] and sporulation in *Myxococcus* [16]. There are also reports of the role of cAMP in the survival of *Mycobacterium microti* within phagolysosomes [17,18]. To our knowledge, there is no comprehensive inventory available of the Class III cyclases found in bacterial and archaebacterial genomes, with particular refer-

*Corresponding author. Fax: (91)-80-23600999.
E-mail addresses: avirs@mrdg.iisc.ernet.in (A.R. Shenoy), sandhya@mrdg.iisc.ernet.in (S.S. Visweswariah).

0014-5793/04/$30.00
doi:10.1016/S0014-5793(04)00128-0

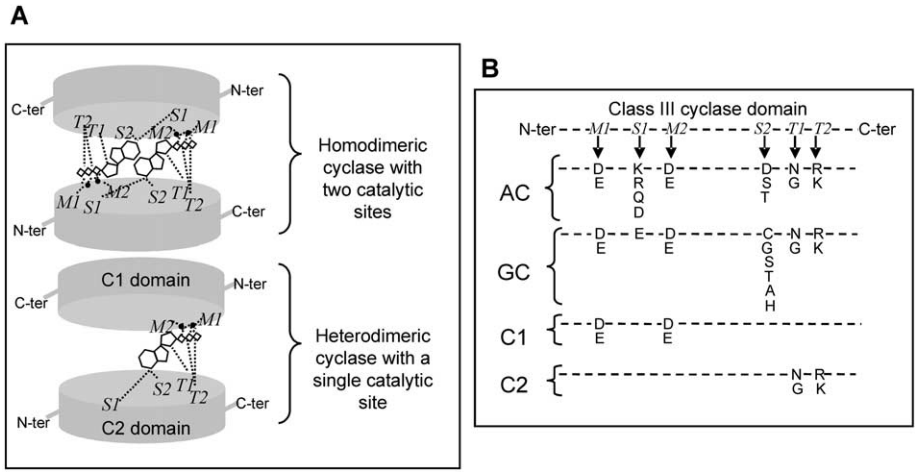


Fig. 1. Schematic representation of the critical Class III cyclase domain residues. A: Homodimeric cyclases (e.g. mammalian receptor guanylyl cyclases) have two active sites due to the presence of all residues required for catalysis in a single polypeptide chain. The heterodimeric cyclases (e.g. mammalian membrane bound adenylyl cyclases) have only a single catalytic centre. The NTP molecule is shown schematically at the active site and the metal atoms are shown as black spheres. Interactions between the nucleotide, metal and the protein are indicated by dotted lines. M1 and M2, metal binding residues; S1 and S2, substrate specifying residues; T1 and T2, transition state stabilising residues. B: A representative Class III cyclase domain indicating the positions of the two metal binding (M1 and M2), substrate specifying (S1 and S2) and transition state stabilising (T1 and T2) residues. The amino acids allowed at each position are shown. AC, adenylyl cyclases; GC, guanylyl cyclases; C1, cyclase domains containing only the C1-like residues; C2, cyclase domains containing only the C2-like residues. Cyclases with residues at positions S1 and S2 other than those found in the AC and GC groups, but having the permitted residues at the other positions, have been grouped as ambiguous cyclases. Cyclase domains that lack the canonical residues at positions M1, M2, T1 or T2 are grouped as cyclase-like domains.

ence to the presence of critical catalytic and substrate specifying residues, and associated domain fusions. In this review, we summarise our analysis of all completed bacterial and archaeal genomes and plasmid sequences for the presence of Class III cyclases. Based on the biochemical and structural information on Class III cyclases, we predict whether these putative genes are likely to be active and also comment on their substrate specificity.

2. Analysis procedures

A combined dataset, consisting of 380 896 predicted proteins from 129 genomes (113 bacterial and 16 archaebacterial) and 456 plasmids (436 plasmids from 227 bacterial species and 20 plasmids from 11 archaebacterial species), was used for our analysis. A combination of sensitive search methods such as PSI-BLAST [19] and HMM [20] were utilised, and each of the 24 cyclase domain seed sequences in the Class III cyclase HMM (Pfam accession PF00211) of the Protein Families Database of Alignments and Models [21] was used individually in PSI-BLAST searches until convergence at a stringent inclusion value (*h*) cut-off of 10^{-4}. In addition the catalytic domain of Rv1625c, which we have earlier reported to perform better [22], was also used as a PSI-BLAST query. In HMM based searches, the Pfam accession number 00211, corresponding to the Class III cyclase alignment, was used as the model at expectation (*E*) value cut-off of 10^{-1}. Additional models were also generated for the searches.

The crystal structures of the canine type V cyclase C1 and the rat type II C2 domains have identified residues that are involved in interacting with the metal ion as well as substrate specificity [2]. Proteins that have the two critical metal binding aspartate residues (glutamate substitution was allowed) as well as the transition state stabilising asparagine/glycine-arginine (arginine to lysine substitution was allowed) pair are considered to be active nucleotide cyclases. Variations in substrate specificity residues have been observed in several enzymes and these have been used in our predictions of the catalytic activity of uncharacterised proteins. Thus, a pairing of lysine/glutamine/aspartate with either aspartate/serine/threonine has been observed in adenylyl cyclases (Fig. 1B) [2,23,24]. An arginine substitution for a lysine was allowed in classifying the proteins as an adenylyl cyclase. In guanylyl cyclases, a glutamate and cysteine/glycine/alanine/histidine/threonine/serine have been seen as being able to support guanylyl cyclase activity (Fig. 1B) [3,25–27]. However, as described below and suggested by others recently [28], the use of only these residues for substrate specificity prediction may be insufficient and additional amino acids at these positions may confer nucleotide specificity. Proteins with any other residue pairs present in cyclases, but that have the other catalytic residues described above, are labelled as ambiguous cyclases in the following analysis. Cyclase domains lacking critical active site residues have been sub-divided based on the presence of C1-like (two metal binding acidic residues) or C2-like (asparagine/glycine and arginine/lysine) residues as per the roles identified in several enzymes. In some cases, domains did not appear to contain C1- or C2-like functional residues, and are therefore referred to as cyclase-like domains (Fig. 1B).

3. Bacterial nucleotide cyclases

A total of 190 Class III nucleotide cyclase domain containing proteins are identifiable in genomes as well as in plasmids (see later) in bacteria. Table 1 shows the distribution of cyclase domains in bacteria and archaea and their predicted substrate specificity and catalytic function. *Bradyrhizobium japonicum*, *Sinorhizobium meliloti*, *Mesorhizobium loti*, *Agrobacterium tumefaciens* and *Caulobacter crescentus* are classified together (α-Proteobacteria) and show a total of 90 cyclases encoded in their genomes (Table 1). A few proteins were identified that have a cyclase domain that is predicted to be in-

Table 1
Distribution of Class III nucleotide cyclase domains in bacteria and archaebacteria

Kingdom	Clade	Organism	AC	AmbiCyc	C1	C2	Cyclike	Total
Bacteria	α-Proteobacteria	*Agrobacterium tumefaciens*	3					3
		Agrobacterium tumefaciens str. C58 (U. Washington)	3					3
		*Bradyrhizobium japonicum**	26	4	5	1	2	37
		Caulobacter crescentus CB15*	1	2	1		3	
		Mesorhizobium loti	7	3	2		1	13
		Sinorhizobium meliloti	$23^{\alpha\beta}$	2^{α}	$4^{\alpha\beta}$	1^{β}	1^{β}	31
	β-Proteobacteria	*Nitrosomonas europaea* ATCC 19718	2	1				3
	γ-Proteobacteria	*Coxiella burnetii RSA 493*			1			1
		Pseudomonas aeruginosa PA01	1					1
		Shewanella oneidensis MR-1	1					1
		Xanthomonas axonopodis pv. *citri* str. 306			1			1
		Xanthomonas campestris pv. *campestris* str. ATCC 33913			1			1
	Actinobacteria	*Arthrobacter nicotinovorans*[a]		2	1			3
		Corynebacterium efficiens YS-314	1					1
		Corynebacterium glutamicum ATCC 13032	1					1
		Mycobacterium bovis subsp. *bovis* AF2122/97	8	2	3		3	16
		Mycobacterium leprae	3				1	4
		Mycobacterium tuberculosis CDC1551	9	2	3		3	17
		Mycobacterium tuberculosis H37Rv	8	2	3		3	16
		Streptomyces avermitilis MA-4680	1					1
		Streptomyces coelicolor A3(2)	1					1
	Cyanobacteria	*Nostoc* sp. PCC 7120	6					6
		Synechocystis sp. PCC 6803[b]	1				1	3[b]
		Thermosynechococcus elongatus BP-1	1	1				2
	Spirochaetes	*Leptospira interrogans* serovar *lai* str. 56601*	16		1	1	1	18
		Treponema pallidum	1					1
	Others	*Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586			1			1
		Chlorobium tepidum TLS					1	1
		Nitrosomonas europaea	2	1				3
Archaebacteria	Euryarchaeota	*Methanosarcina acetivorans* str. C2A		1				1
		Methanopyrus kandleri			1			1
		Methanothermobacter thermoautotrophicus					1	1
		Grand total	123	21	29	4	18	193

Abbreviations: AC, adenylyl cyclase; AmbiCyc, ambiguous cyclase; C1, cyclase domain with C1-like functional residues; C2, cyclase domain with C2-like functional residues; Cyclike, cyclase domain lacking both C1-like and C2-like residues.
Proteins identified in the study were classified as described in Section 2 and numbers of each type of cyclase domain in different genomes are enlisted. The row totals indicate the number of genes identified in each genome and genomes marked by an asterisk have one gene that has two cyclase domains. Therefore the cyclase domain tally in these organisms is one more than the total number of cyclase genes identified. The numbers in the last row therefore add up to 195 cyclase domains that are present in a total of 193 gene products. One gene product is a characterised guanylyl cyclase and is not listed (see below). A detailed listing of the gene identification (gi) numbers is available from the authors on request and at www.mrdg.iisc.ernet.in/bundle/.
α and β indicate that some cyclases are present on the pSymA and pSymB mega-plasmids of *S. meliloti* respectively.
[a]The cyclases found in this organism are present on a catabolic plasmid.
[b]One cyclase in *Synechocystis* is the only known bacterial guanylyl cyclase (gi|16329561) and is not listed in the above table. The total number of cyclases in this genome is therefore three.

active (Table 1) and additionally three truncated proteins, possibly representing pseudogenes, were also found (gi|16264413, gi|16264510 and gi|27382802). Two cyclases from *S. meliloti* have been characterised through complementation (gi15964943) [23] and by heterologous expression in *Escherichia coli* (gi15964012) earlier [24]. Interestingly, it was shown that a strain of *S. meliloti* lacking both these cyclases was still capable of producing cAMP indicative of the presence of additional cyclase(s) [23].

A Class III cyclase from *P. aeruginosa* (gi|15598413) was characterised [15] and has the multi-transmembrane MASE-2 (membrane associated sensor-2) domain [29]. This domain is associated with several GGDEF domain proteins in bacteria, and this is interesting in view of the fact that the GGDEF domain is homologous to the Class III cyclase domain [29], suggesting an early fusion of the MASE-2 domain and proteins that are likely to be similar in structure to the cyclase domain. The similar MASE-1 domain is found associated with receptor histidine kinases. While the nature of the signals sensed by the MASE domains is unknown, it is interesting to note that these domains are found in proteins that are involved in signalling, and therefore could transduce an environmental signal to an intracellular response.

The dissimilatory metal reducing bacteria [30] are capable of accepting various metal atoms as terminal electron acceptors and are important in bioremediation. Cyclic AMP in one such organism, *Shewanella oneidensis*, is required for anaerobic respiration [31] and a Class III cyclase could play a role in regulating cAMP levels (Table 1). The cyclase domain containing protein in the two plant pathogenic *Xanthomonas* species has C1-like residues (gi|21241033 and gi|21229718; Table 1) and the transition state stabilising asparagine residue is replaced by proline, while the arginine residue is conserved. In order for this protein to be active, it would need to hetero-

dimerise to provide C2-like residues for nucleotide binding at the catalytic site. However, given the lack of any other putative cyclase-like gene in the genome, it might be possible that the homodimer is active as a consequence of additional mutations that could provide a functional catalytic centre.

Coxiella burnetii, which resides in phagolysosomal vacuoles and is the agent of human Q fever [32], is predicted to have a cyclase (gi|29654377) that is C1-like due to the substitution of the transition state stabilising arginine residue by glycine. The Institute for Genome Research (TIGR) annotation reveals the presence of a frameshift truncated Class III cyclase gene (gi|29541643) in this organism as well [33]. *C. burnetii* is an obligate intracellular pathogen with a reduced genome size and several pseudogenes and it is possible that it has lost cyclase genes in this process of gene degradation [32]. Interestingly, *Pseudomonas*, *Shewanella*, *Xanthomonas* and *Coxiella* are classified as γ-Proteobacteria and have only a single Class III cyclase in their genomes, unlike the α-Proteobacteria which have multiple Class III cyclase domains in their genomes (Table 1).

The Actinobacteria such as *Mycobacterium*, *Corynebacterium*, *Streptomyces* and *Arthrobacter* have significant medical and industrial importance and constitute a group with a large number (60 in all) of cyclases (Table 1). Many of these organisms were found to have Class III cyclases and cyclase-like putative regulatory domains (Table 1). Several cyclase-like domains were found fused to an ATPase and a helix-turn-helix type DNA binding domain (Fig. 2) and such domain organisation is unique to the Actinobacteria. These transcription factor-like and other actinobacterial cyclases have been described in greater detail elsewhere and will not be discussed here [22].

Cyclic AMP in Cyanobacteria is known to regulate heterocyst patterns [34], respiration and gliding [35], and its levels are known to change in response to changes in light and pH [36,37]. Several cyanobacterial cyclases have been studied biochemically [38–45]. Extensive biochemical characterisation by cloning, expression and purification has been carried out on a cyclase, CyaC, from *Spirulina* which shows intricate regulation of the cyclase activity by a two-component phosphorylation event [45]. The adenylyl cyclase, CyaG, from *Spirulina* has guanylyl cyclase-like dimerisation regions N-terminal to the cyclase domain and greater sequence similarity to guanylyl cyclases [40]. Site-directed mutagenesis of substrate specifying residues in this enzyme could convert it to a guanylyl cyclase, indicating that this enzyme appeared to possess the structural features required of a guanylyl cyclase. In this context, it is interesting to note that the Rv1625c gene product from *Mycobacterium tuberculosis* could not be converted to a guanylyl cyclase by mutation, despite its overall higher sequence similarity to a guanylyl cyclase, again emphasising the subtle changes in amino acids and, as a consequence, the structures of closely related bacterial Class III cyclases.

Six Class III cyclases from *Anabaena* spp. (now *Nostoc*) were cloned through a genetic screen [42] and have been identified in the genome analysis described here (Table 1), and one of them has similar domain fusions described for the *Spirulina platensis* CyaC enzyme [45]. A GAF (cGMP specific phosphodiesterase, adenylyl cyclase and FhlA) domain containing cyclase from *Anabaena* [38] is interesting in that it is regulated by cAMP binding to one of its two N-terminal GAF domains, suggesting a positive feed-back for regulation of its enzymatic activity. To date, no such domain fusion has been observed in any cyclase from higher organisms.

Spirochaetes are a group of bacteria with a unique cellular structure and are agents of diseases such as Lyme disease, syphilis and leptospirosis [46]. The three Spirochaete genome sequences available [47–49] show a stark contrast as far as the presence of Class III cyclase genes. Thus, *Leptospira interrogans* contains as many as 18 putative cyclase domains in its genome, while the genome of *Borrelia burgdorferi* does not appear to harbour any. This is obviously indicative of the diverse nature of this small group of bacteria (Table 1) and their adaptation to their mammalian host.

A single soluble cyclase was detected in the obligate anaerobe *Fusobacterium nucleatum* (Table 1). *Chlorobium tepidum*, a photosynthetic green-sulphur bacterium that utilises sulphide as the electron source, also encodes a Class III cyclase (Table 1). Both these cyclases are highly divergent members of the cyclase domain and therefore have variations in the active site residues. The *F. nucleatum* cyclase has threonine and asparagine residues in the place of the transition state stabilising asparagine and arginine respectively. *Nitrosospira europaea* (a β-Proteobacterium) shows the presence of three cyclase domain containing genes (Table 1).

4. Archaebacterial cyclases

The sequences of the cyclase domains in the three archaeal methanogens, namely *Methanosarcina*, *Methanopyrus* and *Methanothermobacterium*, are very divergent and do not allow for the easy identification of the typical Class III cyclase catalytic residues. The residues required for C2 functions are not the conserved asparagine and arginine residues and therefore these proteins might have interesting and unique biochemical properties. These proteins are >300 amino acids in length; however, no other known domain was identifiable in these proteins. Cloning and biochemical characterisation of these cyclases would be of interest.

5. Cyclases on bacterial plasmids

Plasmids are well known vehicles for horizontal gene transfer and the presence of cyclases on plasmids could indicate that this mechanism is operative in horizontal transmission of Class III cyclase genes [50]. Our analysis shows that three cyclase genes are present on the nicotine utilisation catabolic plasmid of *Arthrobacter nicotinovorans*, a soil Actinobacterium [51]. These enzymes could be possibly involved in regulation of expression of genes on the large plasmid itself. The substrate specificity of two active cyclases (gi|25169176 and gi|25169179) is not predictable from their sequence and the third cyclase (gi|25169178) lacks critical catalytic residues and therefore has C1-like functional residues (Table 1). The lack of a large number of cyclases on plasmids suggests that these may not be the main source of horizontal transfer of this gene.

Several cyclases are present on the large plasmids (pSymA and pSymB) found in *Sinorhizobium* (Table 1). However, the size and the additional presence of several important genes on these symbiotic plasmids fuels the debate as to whether these plasmids should be classified as such, or are actually chromosomes [52].

Interestingly a cyclase from the pSymB plasmid of *S. meli-*

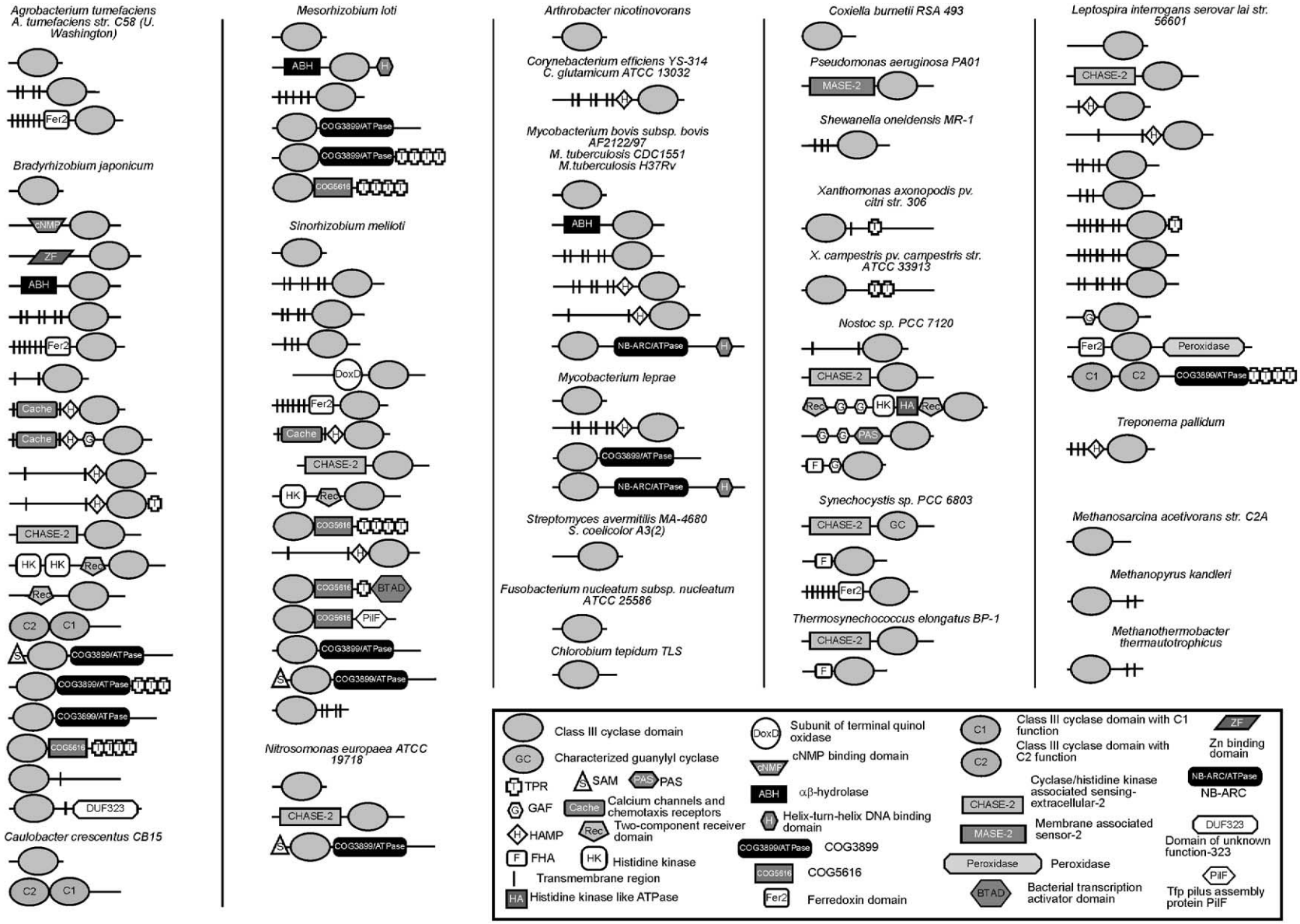


Fig. 2. Domain compositions of Class III cyclases in bacteria and archaea. Schematic domain compositions (not to scale) of cyclases are shown based on the SMART, CD search and Pfam domain entries. Transmembrane regions were predicted using TMHMM2 (http://www.cbs.dtu.dk/services/TMHMM/). Proteins shown are representative of each domain composition in a genome, and several proteins of similar domain compositions were identified as is seen in Fig. 3. The number of tetratricopeptide repeats (TPR) varies within proteins that have other domains in common (e.g. when found along with the COG5616 and COG3899/ATPase domains) and only one representative has been shown (see Fig. 3).

loti (gi|16264509) appears to be a truncated cyclase pseudogene. Full-length cyclase genes with 65–67% sequence identity are found in *M. loti* (gi|13472842) and *Rhizobium etli* (gi|21616128) respectively, possessing all residues required for catalytic activity.

6. Eukaryotic-like double cyclase domain containing proteins

Several eukaryotic adenylyl cyclases have two cyclase domains within a single polypeptide chain that associate to form the catalytic centre [2]. In *B. japonicum*, a cyclase gene (gi|27376782) was found to have two tandem cyclase domains (Fig. 2). Closer analysis of the *B. japonicum* protein indicates that the N-terminal cyclase domain in the sequence (called C1 by convention) has residues that classify it functionally as a C2 domain, and the second domain has C1-like residues. This organisation is reminiscent of the 22-transmembrane guanylyl cyclases in eukaryotic microbes such as *Plasmodium* where such switching of the C1 and C2 functional domains in the polypeptide chain was first observed [25]. A closely related cyclase with switched C1 and C2 domains was found in another α-Proteobacterium, *Caulobacter crescentus* (gi|16126933). This raises the interesting possibility that the protist guanylyl cyclases could have evolved from these double cyclase domain cyclases found in bacteria. Further biochemical characterisation of these forerunners of mammalian-like cyclases and a study of differences in their biochemical properties would provide insight into the evolution of the enzymes found in higher organisms.

Another interesting case is that of a gene in *L. interrogans* (gi|24216707) that encodes a soluble adenylyl cyclase with C1-like and C2-like domains arranged in tandem, similar to the mammalian adenylyl cyclases. The cyclase domains in this protein are followed by the COG3899 ATPase domain that is fused C-terminal to the cyclase domains. This therefore makes it similar in domain composition to the mammalian testicular adenylyl cyclase (Fig. 2). It has been suggested earlier that the progenitor of the mammalian testicular cyclase must have arisen in bacteria, where the fusion of the cyclase and an ATPase domain is found [26]. We suggest that the bacterial double cyclase domain and ATPase containing proteins are the most likely ancestors of the mammalian soluble enzyme. These genes contain additional tetratricopeptide repeats (TPR) at the C-terminal region, allowing for protein–protein interactions and formation of higher oligomers, which are absent in the mammalian enzyme. It would be interesting to test whether the stimulatory effect of bicarbonate on the mammalian enzyme [53] is also found in the *L. interrogans* protein. The substrate specifying residues in the predicted C2 domain of this protein are a serine-threonine pair of residues, and recently the presence of a threonine residue at the second position was suggested to be essential for bicarbonate stimulation [54].

7. Guanylyl cyclases in bacteria

There are only very early reports of the occurrence of cGMP in bacteria and this is also reflected in the relatively few candidate guanylyl cyclases in bacteria and archaea. In our analysis, only two cyclases with substrate specifying residue pairs as seen in eukaryotic guanylyl cyclases (glutamate-glycine and glutamate-threonine) are seen in the Cyanobacteria *Synechocystis* and *Nostoc*. A guanylyl cyclase identified through genetic analysis in *Synechocystis* (gi|16329561) [27] has four transmembrane regions, a CHASE-2 domain and the single Class III cyclase domain, and deletion of this gene resulted in reduced intracellular cGMP levels in the cell. However, a thorough biochemical characterisation of this protein would be a fruitful area of study, since no direct measure of the catalytic activity of the purified gene product has been described so far.

The similar CHASE-2 domain containing cyclase from *Nostoc* (gi|17228613) that has a glutamate-threonine pair at the substrate specifying positions has been earlier characterised through genetic screening via complementation of an adenylyl cyclase deficient strain of *E. coli* [42]. It is possible, however, that at least in a subset of bacterial cyclases there may be contributions of as yet unknown additional amino acids towards substrate specificity. This is also apparent from the failure of conversion of an adenylyl cyclase from *M. tuberculosis* to a guanylyl cyclase, despite high overall sequence similarity of that protein to the guanylyl cyclases [55]. Indeed, sequence analysis has already identified more residues that differ between the mammalian adenylyl and guanylyl cyclases that could contribute to substrate specificity [12].

8. Ambiguous cyclases

Several cyclases that have all the residues required for catalysis but do not have either adenylyl or guanylyl cyclase-like substrate binding residues were found and classified as a group of ambiguous cyclases (Tables 1 and 2). Biochemical characterisation would be required for resolution of their status as adenylyl or guanylyl cyclases. Cloning, purification and initial characterisation of the arginine-leucine pair containing cyclase (gi|15608031) from *M. tuberculosis* (Rv0891c) has indicated it to be an adenylyl cyclase (unpublished observations) and therefore represents a new substrate specificity pair that allows for the conversion of ATP. This combination however

Table 2
Substrate specifying residues found in bacterial cyclases that are not found in characterised adenylyl or guanylyl cyclases

Organism	GI number	S1	S2
Arthrobacter nicotinovorans	gi|25169179	S	V
	gi|25169176	D	V
Bradyrhizobium japonicum	gi|27375549	N	D
	gi|27380748	G	T
	gi|27375470	K	A
	gi|27381818	R	A
Caulobacter crescentus CB15	gi|16126936	S	V
Methanosarcina acetivorans str. C2A	gi|20090420	N	V
Mycobacterium bovis subsp. *bovis* AF2122/97	gi|31791563	Q	N
	gi|31792079	R	L
Mesorhizobium loti	gi|13471136	G	T
	gi|13472842	S	S
	gi|13473854	N	A
Mycobacterium tuberculosis CDC1551	gi|15839770	Q	N
	gi|15840309	R	L
Mycobacterium tuberculosis H37Rv	gi|15607527	Q	N
	gi|15608031	R	L
Nitrosomonas europaea ATCC 19718	gi|30249278	D	R
Sinorhizobium meliloti	gi|15965819	N	D
	gi|16263440	K	G
Thermosynechococcus elongatus BP-1	gi|22299823	K	G

The amino acid at the S1 position corresponds to the lysine and that at S2 to the aspartate found in most adenylyl cyclases.

is present in only three orthologous mycobacterial proteins (gi|31792079, gi|15840309, gi|15608031; Table 2). A few cyclases have lysine paired with glycine (gi|22299823, gi|16263440) or alanine (gi|27375470) and thus have at least one conserved substrate specificity residue that is present in adenylyl cyclases (Table 2).

9. Domain compositions of Class III cyclases

Class III cyclases in bacteria have been shown to be regulated by pyruvate [56], aspartate and histidine phosphorylation [45] and cAMP [38], suggesting that domains fused to the cyclase domain have an important regulatory role. We therefore have identified additional domains fused to the cyclase domain using HMMPfam (Pfam version 9) and the IMPALA program [57], and version 1.62 of the Conserved Domain Database consisting of 11 088 protein family profiles (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Representative domain compositions of the nucleotide cyclases identified in this study are shown in Fig. 2. A larger variety of domain compositions is observed within genomes that have multiple cyclases, such as the α-Proteobacteria and the Actinobacteria (Fig. 2). A total of 119 proteins are soluble and a majority of the membrane bound cyclases have six transmembrane domains (30).

Within the α-Proteobacteria, there are several multidomain proteins (Fig. 2). One such protein (gi|13472512) has several WD40 repeats (composite E value 2.3×10^{-98}) along with a region that picks up comparatively weak stretches of similarity to the flavivirus non-structural protein NS2B and a cytochrome cd_1 haem binding domain (E value ~0.06). It is interesting to note that in eukaryotic systems, the β subunit of the heterotrimeric G proteins that along with the γ subunit shows isoenzyme specific regulation of mammalian cyclase activity, is a prototypical protein with WD40 domain repeats [21]. Here, in bacteria, we find a part of the cyclase domain and the WD40 repeats within the same polypeptide chain, thus dating back the functional relation between these two domains to the bacteria.

A large number of cyclases, especially within the α-Proteobacteria, have TPRs (Fig. 2). This domain is found in several proteins, and is possibly involved in protein–protein interactions that could regulate the oligomeric status of the proteins [58]. Additionally, the SAM domain, involved in protein–protein interaction and formations of homo- and heteromeric complexes [59], was also found fused to the cyclase domain. The requirement for dimerisation for the formation of the active site could be the reason behind the presence of domains that allow for protein–protein interaction in cyclases. The 12-transmembrane eukaryotic adenylyl cyclases are activated by forskolin and $G_{s\alpha}$ protein by facilitating tighter association of the protomers [2]. Unlike these double cyclase domain enzymes however, the regulation of bacterial homodimeric cyclases might require additional mechanisms to bring the two protomers together, and domains fused to the cyclase domain may assist in this. Alternatively, a particularly interesting implication is the possibility of generating a very large repertoire of cyclases in organisms with multiple cyclases, by the heterodimerisation of different cyclases. Moreover, the dimerisation of an active cyclase with a cyclase domain that lacks catalytic residues could lead to inactivation of one or both of the active sites. Earlier, the co-expression of two receptor cyclases from *Leishmania*, only one of which was active, led to a complete loss of cyclase activity [60]. We have shown recently that the substrate specifying residues in the Rv1625c protein of *M. tuberculosis* were found to contribute to dimerisation, thus indicating the intricate mechanisms that have evolved in regulating oligomerisation in bacterial cyclases [61].

A striking domain composition is found in the gi|27376781 cyclase in *B. japonicum* which has a cNMP binding domain in addition to the cyclase domain (Fig. 2). This domain probably binds cAMP thereby allowing for allosteric modulation of the cyclase activity, in a manner similar to the GAF domain cyclase from *Anabaena* that is activated by cAMP [38]. A novel domain composition is found with domains involved in electron transport, such as the quinol oxidase subunit (DoxD) and the ferredoxin (Fer2) domain containing cyclases (seven proteins). The cyclase domains found within these proteins (gi|15889837, gi|15966204, gi|16263053, gi|16330418, gi|17936455, gi|24213884, gi|27382367) are predicted to be active as adenylyl cyclases (Table 1 and Fig. 2). The Fer2 domain has been implicated to bind small organic molecules or metal ions [62]. This domain might therefore act as an allosteric regulator of the cyclase domain. Fusions of inactive cyclase-like domains are also found with the αβ-hydrolase domain (Fig. 2). In this scenario it is likely that these divergent cyclase domains have evolved novel regulatory functions. They may retain the ability to bind nucleotides that could in turn lead to the regulation of the associated effector domain.

A predicted adenylyl cyclase in *Nostoc*, *Synechocystis* and *Thermosynechococcus* has a forkhead associated (FHA) domain (gi|17228238, gi|16330472, gi|22299953 respectively), which is a domain involved in binding to phospho-serine and phospho-threonine residues [63]. Interestingly, eukaryotic-like kinases have been identified in Cyanobacteria [64,65] and probably this could be a link between the cyclic nucleotide and kinase signalling pathways in these organisms. The versatile GAF, HAMP (histidine kinases, adenylyl cyclases, methyl accepting chemotactic receptors and phosphatases) as well as the histidine kinase and receiver domains are found associated with the cyclase domain from *Nostoc*, and may therefore allow the regulation of diverse functions by cAMP in these organisms [42].

L. interrogans has four membrane bound HAMP domain cyclases and two soluble cyclases with GAF domain fusions (Figs. 2 and 3). Deletion and mutagenesis of HAMP domains in proteins leads to deregulation of the downstream effector domains [66]. A large number of cyclases in *M. tuberculosis* and Actinobacteria also have HAMP domains (Figs. 2 and 3). The HAMP domain is the most commonly occurring domain (present in 30 proteins) found fused to the catalytic domain, followed by the ATPase domain (22 proteins) and A variable number of TPRs (found in 21 proteins). The COG5616 and the CHASE-2 domain were found in 14 and 11 proteins respectively.

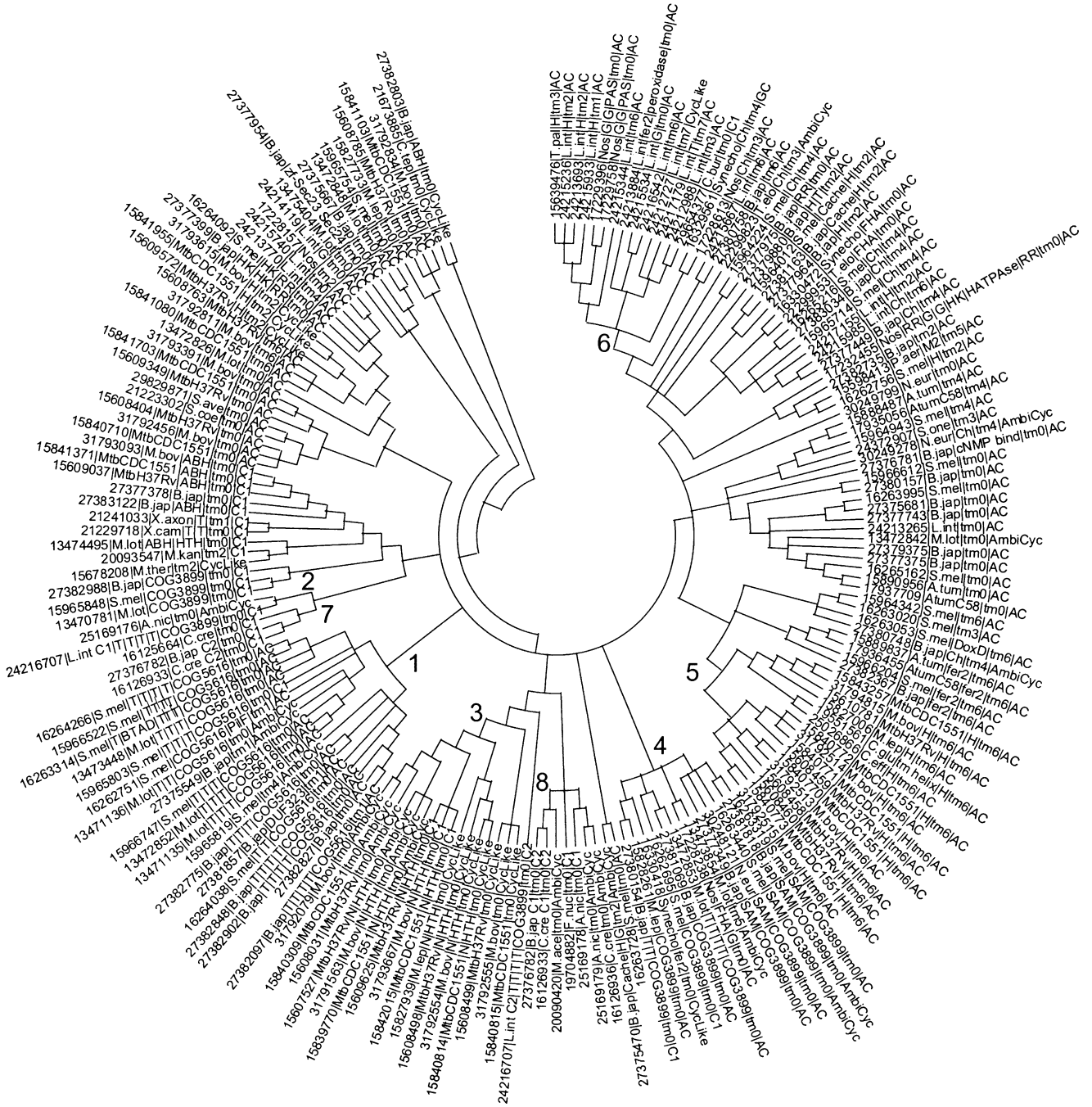
10. Phylogenetic analysis of the cyclases

A phylogenetic analysis of the cyclases was carried out using an alignment of the cyclase domains from the hits described above excluding eight (gi|13472512, gi|15608260, gi|15840558, gi|16264413, gi|16264509, gi|16264510, gi|27382802, gi|31792314) truncated cyclases that could represent pseudogenes. Multiple sequence alignments were gener-

ated using hmmalign [20] and ClustalX [67], and edited with Jalview [http://www.jalview.org]. The phylogenetic tree representation shown in Fig. 3 is based on the neighbour-joining method using Molecular Evolution Genetics Analysis software [68]. Branches with $>50\%$ support in a 1000 replicate interior branch test are included in the tree. The tree is drawn and labelled to include the domains, number of predicted transmembrane helices and the predicted nature of the domain classified as explained earlier.

As can be seen in Fig. 3, several proteins with similar domain compositions are found on neighbouring clusters, indicating that these domains have co-evolved from a common ancestor and then diverged and/or expanded within specific lineages. Cases in point are the cyclases in the branch labelled 1, which contains several cyclases with the TPR and the COG5616 integral membrane domains. Interestingly, the ATPase domain represented by COG3899 and the NB-ARC (nucleotide binding domain common to Apaf1, Resistance proteins and CED4) domain are found in cyclases in clusters 2–4. This indicates that different ATPase domains may have fused to the cyclase domain multiple times during evolution. The actinobacterial transcription factor-like proteins with a cyclase domain are found in cluster 3. Cluster 5 has the six-transmembrane HAMP domain cyclases from Actinobacteria, close to which are found cyclases with a Fer2 or DoxD domain. Several cyclases from *L. interrogans* are found within cluster 6, and this indicates that the cyclases in this genome are probably related by paralogy.

The C1-like and C2-like functional domains of the double cyclase domain containing proteins are found in separate clus-



ters labelled 7 and 8. Note that the domains cluster based on the functionality of these domains and not their order of appearance in the polypeptide chain. Interestingly, cyclases from *C. crescentus* and the *Arthrobacter* plasmid are found in each of these two clusters. It would be interesting to test whether the *Arthrobacter* enzymes act as functional heterodimers. The existence of bacterial cyclases with two cyclase domains also predates the evolution of such cyclases to bacteria. Moreover, the switching of C1 and C2 domain is also found within bacteria. The distribution of such cyclases within bacteria is limited at present, probably due to the bias in the type of bacteria whose genomes have been sequenced. However, the identification of these eukaryotic-like proteins has opened up new ancestral candidates of the protist guanylyl cyclases and the mammalian soluble adenylyl cyclases.

11. Conclusions

Class III cyclases were found in relatively few (29 species) of the 129 genomes that were surveyed. The major bacterial groups that lack the Class III cyclases are the *Bacillus*/*Clostridium* group, the *Deinococcus* group, the Chlamydiae, and the ε- and δ-Proteobacteria. Phylogenetic trees based on multiple genes and multiple methods have yielded reference trees for inferring phylogeny [69–71]. Based on such trees, the Class III cyclase domains are present in most branches and appear to be absent mainly from the ancestors of the *Bacillus*/*Clostridium* group and that of the β/γ-Proteobacteria. Although the *Deinococci* have been integrated with the Actinobacteria-Cyanobacteria clade, *D. radiodurans* lacks the Class III cyclase. This organism however has 16 GGDEF domain containing proteins (TIGR, http://www.tigr.org). Interestingly, several Proteobacteria, which also lack Class III cyclases, have multiple GGDEF domain containing proteins. The GGDEF domain is thought to have diguanylate cyclase activity [72] but very few proteins have been biochemically characterised. Indeed, the diversification and expansion of the GGDEF domains in several bacteria that lack Class III cyclases stresses the need for further characterisation of these proteins.

Class III cyclases are restricted only to a few γ-Proteobacteria, whose genomes encode nucleotide cyclases of the other classes as well. This indicates the presence of the cyclase domain in a common ancestor and its loss or diversification in some lineages. The presence of archaeal cyclases in several clusters of the phylogenetic tree (Fig. 3) indicates the possible horizontal transfer of these enzymes from bacteria. Moreover, their presence in archaebacteria also indicates the structural modification of the cyclase domain into that which can retain its function even under extreme conditions.

Fusions to various domains increase the versatility of a protein family and allow its recruitment into various cellular regulatory pathways. Since this is seen often in cyclases within bacteria and eukaryotes, the Class III cyclase domain itself appears to be structurally amenable to undergoing such fusions with a variety of diverse domains.

The presence of an overwhelming majority of adenylyl cyclases in bacteria is reflected in only a few reported cases of cGMP based signalling in prokaryotes. Indeed, even in eukaryotes such as yeast, there is no conclusive evidence to demonstrate the use of cGMP as a second messenger. We can only hypothesise as to the reasons for the more widespread occurrence of adenylyl cyclases in all phyla. It has been suggested that cAMP may have been the precursor for the generation of ATP in the reverse cyclase reaction [1]. Therefore, the retention and subsequent evolution of the utilisation of cAMP for signalling may have led to the ubiquitous presence of cAMP rather than cGMP in most life forms.

The use of backbone peptide bond interactions in binding ATP in the Class III adenylyl cyclases [7] has led to the generation of enzymes that are less sensitive to mutations that abolish adenylyl cyclase activity and completely alter substrate specificity. For example, mutations of the substrate specifying residues in adenylyl cyclases resulted in a general purine cyclase, which could synthesise both cAMP and cGMP, while corresponding mutations in guanylyl cyclases completely abolished the ability to use GTP as a substrate, and converted the enzyme to an adenylyl cyclase [10]. Thus, guanylyl cyclases appear to be more specialised enzymes, with stringent structural requirements, and could have evolved later by refinement of the adenylyl cyclase core. Finally, the relative stability of cGMP as compared to that of cAMP (free energy of hydrolysis of the 3′-phosphate bond being −10.5 kcal/mol for cGMP and −14.1 kcal/mol for cAMP [73]) may also have contributed to the absence of cGMP in bacteria, where a rapid turn-over of signalling molecules is required to generate

←

Fig. 3. Phylogenetic analysis of Class III cyclases in bacteria and archaea. The topology of a neighbour-joining consensus tree with branches having >50% support in an interior branch test is shown. Names of genes begin with the gene identifier number, followed by the abbreviation for the name of organism along with the mention of the C1 or the C2 domain (by order of appearance in the polypeptide chain) when more than one cyclase domain was present. This is followed by an abbreviation of the other domains fused to the cyclase domain and the number of transmembrane (tm) regions. Finally, the predicted substrate specificity when active, or the C1-like or C2-like nature of the domain is mentioned (see Section 2). Branches with numbers are described in greater detail in the text. Abbreviations used for domains are: ABH, αβ-hydrolase; Cache, Ca channels and chemotaxis receptors; Ch, CHASE-2; cNMP bind, cNMP binding; COG3899, an ATPase; COG5616, an integral membrane domain; DoxD, quinol oxidase subunit; DUF323, domain of unknown function-323; fer2, ferredoxin; FHA, forkhead associated; G, GAF; H, HAMP; HATPase, histidine kinase-like ATPase; HK, histidine kinase; HTH, helix-turn-helix (HTH) type DNA binding domains; M2, MASE-2; N, NB-ARC; PilF, Tfp pilus assembly protein PifF; RR, response regulator/receiver; SAM, sterile α motif; zf-Sec23 Sec24, Sec23/Sec24 zinc finger; T, TPR. Names of organisms are A.nic, *Arthrobacter nicotinovorans*; A.tum, *Agrobacterium tumefaciens*; A.tumC58, *Agrobacterium tumefaciens* str. C58 (U. Washington); B.jap, *Bradyrhizobium japonicum*; C.cre, *Caulobacter crescentus* CB15; C.tep, *Chlorobium tepidum* TLS; C.eff, *Corynebacterium efficiens* YS-314; C.glu, *Corynebacterium glutamicum* ATCC 13032; C.bur, *Coxiella burnetii* RSA 493; F.nuc, *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586; L.int, *Leptospira interrogans* serovar *lai* str. 56601; M.lot, *Mesorhizobium loti*; M.kan, *Methanopyrus kandleri* AV19; M.ace, *Methanosarcina acetivorans* str. C2A; M.ther, *Methanothermobacter thermoautotrophicus*; M.bov, *Mycobacterium bovis* subsp. *bovis* AF2122/97; M.lep, *Mycobacterium leprae*; MtbCDC1551, *Mycobacterium tuberculosis* CDC1551; MtbH37Rv, *Mycobacterium tuberculosis* H37Rv; N.eur, *Nitrosomonas europaea* ATCC 19718; Nos, *Nostoc* sp. PCC 7120; P.aer, *Pseudomonas aeruginosa* PA01; S.one, *Shewanella oneidensis* MR-1; S.mel, *Sinorhizobium meliloti*; S.ave, *Streptomyces avermitilis* MA-4680; S.coe, *Streptomyces coelicolor* A3(2); Synecho, *Synechocystis* sp. PCC 6803; T.elo, *Thermosynechococcus elongatus* BP-1; T.pal, *Treponema pallidum*; X.axon, *Xanthomonas axonopodis* pv. *citri* str. 306; X.cam, *Xanthomonas campestris* pv. *campestris* str. ATCC 33913.

quick responses. It is important to note that bacteria utilise ppGpp and cyclic diguanylate as intracellular messengers and further studies on enzymes that generate these molecules are likely to prove illuminating.

The Class III cyclase domains are found in bacteria that occupy varied ecological niches and have different physiological prowess. These range from oxygenic and non-oxygenic photosynthetic bacteria such as Cyanobacteria and *Chlorobium*, pathogens such as *Mycobacterium leprae* and *M. tuberculosis*, biotechnologically important nitrogen fixing α-Proteobacteria, *Shewanella* as well as free living bacteria. However, the presence of members of the other classes of cyclases and the absence of the Class III domain in many bacteria [33], indicates that perhaps evolutionary tinkering has selected for the function of generating cAMP, rather than the Class III cyclase fold itself. Nevertheless, the regulation by cAMP of important secondary metabolite pathways and virulence in several bacteria of interest to humans make the Class III cyclase family a prominent target for structural and biochemical studies. It is hoped that many of the putative cyclase genes that we have identified in this study will provide an impetus for biochemical characterisation and evaluation of the roles of these enzymes in diverse bacteria.

Acknowledgements: We would like to thank A. Krupa and N. Srinivasan for useful discussions and Urvashi Bahadur for critical reading of the manuscript. Support from the Department of Biotechnology, Government of India and the Wellcome Trust, UK, is gratefully acknowledged.

References

[1] Barzu, O. and Danchin, A. (1994) Prog. Nucleic Acid Res. Mol. Biol. 49, 241–283.
[2] Tang, W.J. and Hurley, J.H. (1998) Mol. Pharmacol. 54, 231–240.
[3] Wedel, B. and Garbers, D. (2001) Annu. Rev. Physiol. 63, 215–233.
[4] Tellez-Sosa, J., Soberon, N., Vega-Segura, A., Torres-Marquez, M.E. and Cevallos, M.A. (2002) J. Bacteriol. 184, 3560–3568.
[5] Shenoy, A.R., Srinivasan, N. and Visweswariah, S.S. (2002) J. Biosci. 27, 85–91.
[6] Tesmer, J.J., Sunahara, R.K., Johnson, R.A., Gosselin, G., Gilman, A.G. and Sprang, S.R. (1999) Science 285, 756–760.
[7] Tesmer, J.J., Sunahara, R.K., Gilman, A.G. and Sprang, S.R. (1997) Science 278, 1907–1916.
[8] Yan, S.Z., Huang, Z.H., Shaw, R.S. and Tang, W.J. (1997) J. Biol. Chem. 272, 12342–12349.
[9] Liu, Y., Ruoho, A.E., Rao, V.D. and Hurley, J.H. (1997) Proc. Natl. Acad. Sci. USA 94, 13414–13419.
[10] Sunahara, R.K., Beuve, A., Tesmer, J.J., Sprang, S.R., Garbers, D.L. and Gilman, A.G. (1998) J. Biol. Chem. 273, 16332–16338.
[11] Tucker, C.L., Hurley, J.H., Miller, T.R. and Hurley, J.B. (1998) Proc. Natl. Acad. Sci. USA 95, 5993–5997.
[12] Hannenhalli, S.S. and Russell, R.B. (2000) J. Mol. Biol. 303, 61–76.
[13] Confer, D.L. and Eaton, J.W. (1982) Science 217, 948–950.
[14] Hoover, D.L., Friedlander, A.M., Rogers, L.C., Yoon, I.K., Warren, R.L. and Cross, A.S. (1994) Infect. Immun. 62, 4432–4439.
[15] Wolfgang, M.C., Lee, V.T., Gilmore, M.E. and Lory, S. (2003) Dev. Cell 4, 253–263.
[16] Kimura, Y., Mishima, Y., Nakano, H. and Takegawa, K. (2002) J. Bacteriol. 184, 3578–3585.
[17] Lowrie, D.B., Jackett, P.S. and Ratcliffe, N.A. (1975) Nature 254, 600–602.
[18] Lowrie, D.B., Aber, V.R. and Jackett, P.S. (1979) J. Gen. Microbiol. 110, 431–441.
[19] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[20] Eddy, S.R. (2001) HMMER: Profile hidden Markov models for biological sequence analysis (http://hmmer.wustl.edu/).
[21] Bateman, A. et al. (2002) Nucleic Acids Res. 30, 276–280.
[22] Shenoy, A.R., Sivakumar, K., Krupa, A., Srinivasan, N. and Visweswariah, S.S. (2003) Comp. Funct. Genomics 5, 17–38.
[23] Archdeacon, J., Talty, J., Boesten, B., Danchin, A. and O'Gara, F. (1995) FEMS Microbiol. Lett. 128, 177–184.
[24] Beuve, A., Boesten, B., Crasnier, M., Danchin, A. and O'Gara, F. (1990) J. Bacteriol. 172, 2614–2621.
[25] Linder, J.U. and Schultz, J.E. (2002) Mol. Cell. Biochem. 230, 149–158.
[26] Roelofs, J. and Van Haastert, P.J. (2002) Mol. Biol. Evol. 19, 2239–2246.
[27] Ochoa De Alda, J.A., Ajlani, G. and Houmard, J. (2000) J. Bacteriol. 182, 3839–3842.
[28] Linder, J.U. and Schultz, J.E. (2003) Cell Signal. 15, 1081–1089.
[29] Nikolskaya, A.N., Mulkidjanian, A.Y., Beech, I.B. and Galperin, M.Y. (2003) J. Mol. Microbiol. Biotechnol. 5, 11–16.
[30] Nealson, K.H. and Cox, B.L. (2002) Curr. Opin. Microbiol. 5, 296–300.
[31] Saffarini, D.A., Schultz, R. and Beliaev, A. (2003) J. Bacteriol. 185, 3668–3671.
[32] Seshadri, R. et al. (2003) Proc. Natl. Acad. Sci. USA 100, 5455–5460.
[33] TIGR. The Insitute for Genomic Research (http://www.tigr.org).
[34] Smith, G. and Owsby, J.D. (1981) FEMS Microbiol. Lett. 11, 175–180.
[35] Ohmori, K., Hirose, M. and Ohmori, M. (1993) Plant Cell Physiol. 34, 169–171.
[36] Ohmori, M., Ohmori, K. and Hasunuma, K. (1988) Arch. Microbiol. 150, 203–204.
[37] Ohmori, M. (1989) Plant Cell Physiol. 30, 911–914.
[38] Kanacher, T., Schultz, A., Linder, J.U. and Schultz, J.E. (2002) EMBO J. 21, 3672–3680.
[39] Kasahara, M., Yashiro, K., Sakamoto, T. and Ohmori, M. (1997) Plant Cell Physiol. 38, 828–836.
[40] Kasahara, M., Unno, T., Yashiro, K. and Ohmori, M. (2001) J. Biol. Chem. 276, 10564–10569.
[41] Katayama, M., Wada, Y. and Ohmori, M. (1995) J. Bacteriol. 177, 5197.
[42] Katayama, M. and Ohmori, M. (1997) J. Bacteriol. 179, 3588–3593.
[43] Bianchini, G.M., Pastini, A.C., Muschietti, J.P., Tellez-Inon, M.T., Martinetto, H.E., Torres, H.N. and Flawia, M.M. (1990) Biochim. Biophys. Acta 1055, 75–81.
[44] Yashiro, K., Sakamoto, T. and Ohmori, M. (1996) Plant Mol. Biol. 31, 175–181.
[45] Kasahara, M. and Ohmori, M. (1999) J. Biol. Chem. 274, 15167–15172.
[46] Stanier, R.Y., Ingraham, J.L., Wheelis, M.L. and Painter, P.R. (1995) in: General Microbiology, pp. 464–469, Macmillan, London.
[47] Fraser, C.M. et al. (1998) Science 281, 375–388.
[48] Ren, S.X. et al. (2003) Nature 422, 888–893.
[49] Fraser, C.M. et al. (1997) Nature 390, 580–586.
[50] Davison, J. (1999) Plasmid 42, 73–91.
[51] Igloi, G.L. and Brandsch, R. (2003) J. Bacteriol. 185, 1976–1986.
[52] Galibert, F. et al. (2001) Science 293, 668–672.
[53] Chen, Y., Cann, M.J., Litvin, T.N., Iourgenko, V., Sinclair, M.L., Levin, L.R. and Buck, J. (2000) Science 289, 625–628.
[54] Cann, M.J., Hammer, A., Zhou, J. and Kanacher, T. (2003) J. Biol. Chem. 278, 35033–35038.
[55] Shenoy, A.R., Srinivasan, N., Subramaniam, M. and Visweswariah, S.S. (2003) FEBS Lett. 545, 253–259.
[56] Peters, E.P., Wilderspin, A.F., Wood, S.P., Zvelebil, M.J., Sezer, O. and Danchin, A. (1991) Mol. Microbiol. 5, 1175–1181.
[57] Schaffer, A.A., Wolf, Y.I., Ponting, C.P., Koonin, E.V., Aravind, L. and Altschul, S.F. (1999) Bioinformatics 15, 1000–1011.
[58] Lamb, J.R., Tugendreich, S. and Hieter, P. (1995) Trends Biochem. Sci. 20, 257–259.
[59] Peterson, A.J., Kyba, M., Bornemann, D., Morgan, K., Brock, H.W. and Simon, J. (1997) Mol. Cell. Biol. 17, 6683–6692.

[60] Sanchez, M.A., Zeoli, D., Klamo, E.M., Kavanaugh, M.P. and Landfear, S.M. (1995) J. Biol. Chem. 270, 17551–17558.
[61] Shenoy, A.R., Srinivasan, N., Subramaniam, M. and Visweswariah, S.S. (2003) FEBS Lett. 545, 253–259.
[62] Anantharaman, V., Koonin, E.V. and Aravind, L. (2001) J. Mol. Biol. 307, 1271–1292.
[63] Hofmann, K. and Bucher, P. (1995) Trends Biochem. Sci. 20, 347–349.
[64] Wang, L., Sun, Y.P., Chen, W.L., Li, J.H. and Zhang, C.C. (2002) FEMS Microbiol. Lett. 217, 155–165.
[65] Kamei, A., Yoshihara, S., Yuasa, T., Geng, X. and Ikeuchi, M. (2003) Curr. Microbiol. 46, 296–301.
[66] Appleman, J.A. and Stewart, V. (2003) J. Bacteriol. 185, 89–97.
[67] Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and Gibson, T.J. (1998) Trends Biochem. Sci. 23, 403–405.
[68] Kumar, S., Tamura, K., Jakobsen, I.B. and Masatoshi, N. (2001) Bioinformatics 17, 1244–1245.
[69] Daubin, V., Gouy, M. and Perriere, G. (2002) Genome Res. 12, 1080–1090.
[70] Wolf, Y.I., Rogozin, I.B., Grishin, N.V., Tatusov, R.L. and Koonin, E.V. (2001) BMC Evol. Biol. 1, 8.
[71] Wolf, Y.I., Rogozin, I.B., Grishin, N.V. and Koonin, E.V. (2002) Trends Genet. 18, 472–489.
[72] Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M. and Lindberg, M. (2001) FEMS Microbiol. Lett. 204, 163–167.
[73] Simon, L.N., Shuman, D.A. and Robins, R.K. (1973) Adv. Cyclic Nucleotide Res. 3, 225–353.