Advances in Second Messenger and Phosphoprotein Research, Vol. 27, edited by Shirish Shenolikar and Angus C. Nairn. Raven Press, Ltd., New York © 1993.

4

Phylogeny of Adenylyl Cyclases

Antoine Danchin

Régulation de l'Expression Génétique, Institut Pasteur, 75724 Paris Cedex 15, France

cAMP plays a universal role in the control of gene expression as well as in the integration of metabolic functions. It is present both in eucaryotes and in procaryotes (see 119 for early recognition of cAMP presence in bacterial genera). cAMP seems to be absent only from archaebacteria or plants but this remains controversial (160, 259 for archaebacteria, and 273 for plants). However, it has been reported in cyanobacteria (14,214,256) and in algae (220). This ubiquity explains the major interest displayed in the enzymes that produce cAMP from ATP, the adenylyl cyclases. Yet, the literature devoted to the structural aspects of the catalytic subunits of cyclases is rather scarce, mainly because they are quite elusive enzymes, synthesized in very low amount, and very unstable during purification. It has been possible, however, to clone genes encoding adenylyl cyclases of many organisms, from bacteria to higher vertebrates, and this has permitted identification of revealing features of the proteins. I shall try to classify the different cyclases by their common features, and consider the question of their origin: Are we observing a case of evolutionary divergence or convergence? The question remains open, because it appears that there are three well-defined classes of cyclases. Namely, cyclases related to enterobacterial adenylyl cyclases, toxic adenylyl cyclases isolated from bacterial pathogens, and finally, a large class which comprises cyclases from eucaryotes and procaryotes, the origin of which might predate their phylogenetic separation. This raises the question of the origin of cyclic nucleotides as regulatory molecules, as well as their universal involvement in regulatory networks. This will be discussed in later sections of this chapter.

As is usual in science—and contrary to generally accepted opinion—chance has often little to do with discovery. The discovery of cAMP is no exception. As E. Sutherland said in 1958, it is within the scope of a well-defined conceptual framework that cAMP was discovered: "When I first entered the study of hormone action, some 25 years ago, there was a wide-

spread feeling among biologists that hormone action could not be studied meaningfully in the absence of organized cell structure. However, as I reflected upon the history of biochemistry, it seemed to me there was a real possibility that hormones might act at the molecular level" (280). Sutherland therefore built up a cell-free system where well-known hormones could control glycogenolysis. Using this system he isolated a small thermostable molecule which was able to activate glycogen phosphorylase. A thorough analysis of the molecule permitted its identification as adenosine 3′-5′ cyclic monophosphate. Synthesis of cAMP was shown to be the result of the action of an enzyme, adenylyl cyclase, that generated cAMP and PPi from ATP, and which is activated by adrenaline. Since this pioneering work, the structure, function, and regulation of adenylyl cyclases, has been investigated in some detail. At that early time, it was often said that "what is true for *E. coli* is true for the elephant," so that the discovery, in 1968, of a specific effect of cAMP on catabolite repression in *E. coli* raised hopes that the study of bacterial adenylyl cyclases would help us to understand what happens in higher eucaryotes. Although this hope was not substantiated initially however, the recent discovery that several procaryotic adenylyl cyclases are related to their eucaryotic counterparts may revive this thought (13).

PURINE NUCLEOTIDE CYCLASES: GENERAL FEATURES OF THE GENES AND THE PROTEINS

The Enterobacterial Class (Class I)

After the discovery of cAMP, Mackman and Sutherland demonstrated that glucose-starved *Escherichia coli* cells accumulated cAMP (178). Ullmann and Monod further established that part of the catabolite repression phenomenon was controlled by cAMP (301). This observation gave a strong impulse to the studies involving cAMP as a mediator of gene expression (reviewed in 299). Since then numerous experiments have revealed the many features of cAMP synthesis and its mediation of transcription control in bacteria (see 40, 45, 299, 300). The first genes unambiguously recognized as adenylyl cyclase genes were cloned in 1981. Wang et al. (309) isolated from a lambda library, a DNA fragment of *Salmonella typhimurium* that carried most of the corresponding *cya* gene, while Roy and Danchin (250, 251) identified the whole of the *Escherichia coli cya* gene from a plasmid library. The transcription and translation control region of the gene was sequenced and first indicated that the protein reading frame started with an unusual TTG initiation codon (253). Subsequently Aiba and coworkers sequenced the *cya* locus *in toto*, of the *Escherichia coli* adenylyl cyclase as a protein of 862 amino acid residues (2). Study of the local organization of the gene's environment showed that it was expressed from two promoters, overlapping with a diver-

gent promoter that controlled expression of the *hemC* and *hemD* genes (3,253,289). The downstream region was initially thought to express a hydrophobic protein, encoded by a gene named *cyaX*. However, sequence comparison with several enterobacterial counterparts, in parallel with experiments measuring expression from fusions with the *lacZ* gene in the *cya* downstream region indicated that it is from the complementary strand that a gene, tentatively named *cyaY*, is expressed (P. Glaser, A. Roy, and A. Danchin, *unpublished experiments, 1988*). As a consequence, it is now established that the *cya* gene constitutes a monocistronic transcription unit, terminated by a palindromic unit typical of the *E. coli* genomic DNA (87). The gene order in the region is therefore *hemD hemC cya cyaY dapF* (194).

Cloning of the gene permitted identification of the protein produced in maxicells (251). Subsequently it was possible to demonstrate that *E. coli* adenylyl cyclase is comprised of two domains. The catalytic domain is NH_2-terminus whereas the glucose-sensitive regulatory domain is at the COOH-terminus (39,252). The organization of the gene locus as well as the protein structure is highly specific. This prompted Danchin and coworkers to analyze other bacterial types for the structure, activity, and regulation of expression of the corresponding gene and protein. Restriction minus *E. coli cya* deficient strains were constructed that could easily be transformed and complemented by DNA libraries from foreign organisms (109). It was then possible to clone *cya* genes from other bacterial types by direct complementation of such *E. coli* mutants. Work performed on *Erwinia chrysanthemi* demonstrated that both the genes' environment and the proteins' were similar in size and overall organization to that of the *E. coli* genes and proteins at the *cya* locus (41,109). Similar studies on *Salmonella typhimurium* were performed by several groups, however, complete gene sequence was not obtained, preventing extensive phylogenic comparison with the *E. coli* gene and protein counterparts (67,117,159,291). Finally, the *cya* gene from two other enterobacteria. *Yersinia intermedia* (the causative agent of intestinal diseases) and *Yersinia pestis* (the agent of plague), showed that all the major features found in the *E. coli* gene and protein organization were conserved, including the number of promoters, upstream and downstream genes (P. Glaser, O. Sismeiro, and A. Danchin, *unpublished experiments*). In particular the TTG initiation codon, together with a stretch of 18 nucleotides, CAGGCGATACGTCTTGTA, overlapping with the ribosome-binding site, is almost invariant. This nucleotide motif is probably involved in some process of translational control (236,253,254). Screening by Southern hybridization of total DNA from *E. coli* or other enterobacteria such as *Proteus mirabilis* permitted the demonstration that this nucleotide sequence is strongly conserved in this bacterial family, and is present as a single copy in the genome (Sismeiro and Danchin, *unpublished data*). It was therefore of interest to analyze the *cya* gene from other bacterial species, related to enterobacteria but distinct from them. The isolation of the gene from *Pasteurella multocida* (the causative agent of a

snout disease in the pig) permitted Mock and coworkers to demonstrate that the gene itself was very similar to the enterobacterial type, but that the contiguous genes were different, as was the 18 nucleotides segment overlapping the initiator TTG initiation codon (200). Taken together these data indicated that the regulation of the protein activity (including glucose-mediated regulation) is probably conserved, while the expression of the gene may be different in Pasteurellaceae from the expression in Enterobacteriaceae (Fig. 1).

Several general features were noticed when analyzing the polypeptide chain of the adenylyl cyclase (special features will be discussed in the sections on *Phylogenesis* and *Cyclases as a Paradigm*). In particular, no long stretch of hydrophobic amino acid residues were found as might be expected for a membrane-bound protein (122,288). Furthermore the proteins are very rich in cysteine residues, a rather uncommon feature for proteins located in the cytoplasm or at the cytoplasmic surface of the membrane. The observation that the purification of the protein is extremely difficult to achieve (in particular the enzyme becomes extremely unstable as soon as membranes are disrupted) might be related to oxidation of important cysteine residues (319). This could also explain the surprisingly low level of activity found in pure preparations obtained from overexpression vectors (238). In addition, the proteins are rich in histidine residues. This could indicate that metal ions take part in the folding of the polypeptide chain, but as yet no experimental data argue for or against this hypothesis. Finally, the protein is made of two functionally well-defined domains, as will be discussed below (*Phylogenesis* and *Cyclases as a Paradigm*). It would be interesting to know more about cyclases isolated from bacteria more or less related to enterobacteria such as the *Vibrio* species (298).

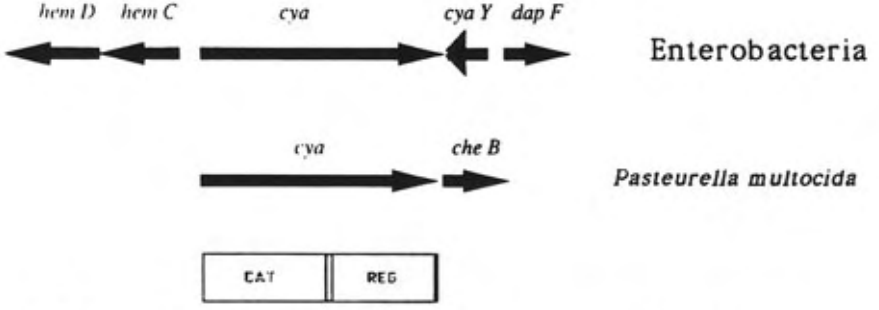


FIG. 1. The local organization of genes is conserved at the *cya* locus in enterobacteria. This is not so with *P. multocida*, where only the fine structure of the *cya* gene (CAT, catalytic region; REG, regulatory region) is similar to that of the enterobacterial counterparts.

The Calmodulin-Activated Toxic Class (Class II)

Whooping cough is caused by *Bordetella pertussis*, a Gram negative procaryote that secretes into the medium many proteins having toxic activities. Among these proteins Wolff and Cook, as early as 1973, recognized the presence of an adenylyl cyclase activity (315). Hewlett and Wolff purified a polypeptide from the culture medium of the bacteria and partially characterized its activity by demonstrating that it was extremely efficient in converting ATP to cAMP (112). Only in 1980 the most unexpected and fascinating observation about this enzyme was made: it is activated by a *host* protein, calmodulin, which is not known to occur in bacteria (95,316). Two years later Leppla demonstrated that another toxic adenylyl cyclase, secreted by a Gram positive organism, *Bacillus anthracis*, the etiological agent of anthrax, was also dependent on the presence of host calmodulin (163). These remarkable observations stimulated intense efforts for the study of these unusual proteins, but in spite of numerous attempts, it took several years before the *cya* genes from either organism could be cloned (23,89,199,293). This can be readily understood if one notes the total lack of activity of the adenylyl cyclase toxins when expressed in bacteria used to screen libraries by complementation of a *cya* defect, and may reflect the absence of significant calmodulin-like activity in the host strains. The introduction of a plasmid directing synthesis of calmodulin into an *E. coli* strain deficient in adenylyl cyclase activity, permitted the cloning of adenylyl cyclase genes coding for the calmodulin-dependent cyclases [similar multiple component cloning procedure can be used for instance, for cloning calmodulin cDNAs (42)]. In 1988 the *Bordetella pertussis* gene was cloned and sequenced by Glaser et al., who used a *cya* strain of *E. coli* expressing a plant-mammal hybrid calmodulin gene (89). The same year, the *Bacillus anthracis* gene was cloned by Tippetts and Robertson using hybridization with an oligonucleotide designed after the amino-terminal peptide of the protein (293) and by Mock et al. who used the calmodulin-dependent complementation technique (199). It was subsequently sequenced by Escuyer et al. (65), followed several months later by Robertson et al. (239), who published a slightly different sequence. Among the differences is the presence of a glutamate residue in place of a valine residue in the conserved catalytic region (suggesting a sequencing error), and a stretch of amino acid differences, corresponding to frameshifts and deletions in a region that is extremely different between *Bordetella pertussis* and *Bacillus anthracis*. This could correspond to true polymorphism, although this seems unlikely.

The *Bordetella pertussis* adenylyl cyclase was found to be a large bifunctional polypeptide of 1,706 amino acid residues (89). Previously, different groups had published various values for the molecular weight (from 43 to 70 kDa) of the purified protein. The biochemical data was explained by the fact that an N-terminal segment of the protein (400 residues) displayed calmodu-

lin-activated ATP-cyclizing activity, while the rest of the molecule contained the domain responsible for the hemolytic activity of the pathogen (58,88,89, 91,246) similar to *E. coli* hemolysin toxin (94). Indeed, fragments of the large precursor purified from extracts of *B. pertussis* as a 200 kDa protein (88,90, 92,107,114,166,190,245), were often found in culture media as low-molecular-mass forms of 50, 45, or 43 kDa (112,113,148,166,190,244). The name cyclolysin was therefore coined for the toxic adenylyl cyclase from *B. pertussis* (90).

The unusually long sequence of the protein prompted Glaser and coworkers to analyze the adjacent regions for functionally related genes. This led them to discover that the protein is secreted using a mechanism specific to Gram negative bacteria and similar to that permitting release in the external medium of *E. coli* hemolysin or cognate proteins in related organisms (90, and discussed in the section on *Cyclases as a Paradigm*). Three genes located downstream from the cyclolysin gene were identified as required for secretion, *cyaB*, *cyaD* and *cyaE*. No other genes were identified by Glaser et al. presumably because of a one base sequencing error in the upstream region of the *cyaA* gene. In this region, Weiss and coworkers discovered the counterpart of the *E. coli hlyC* gene, which they termed *cyaC* (8). The gene order in the region is therefore *cyaC*, *cyaA*, *cyaB*, *cyaD*, and *cyaE* (Fig. 2). This contrasted with the case of *E. coli* hemolysin where hemolysin secretion was supposed to require only two genes, *hlyB* and *hlyD* (314), a third gene *hlyC*, being only required for activation of the protein (211).

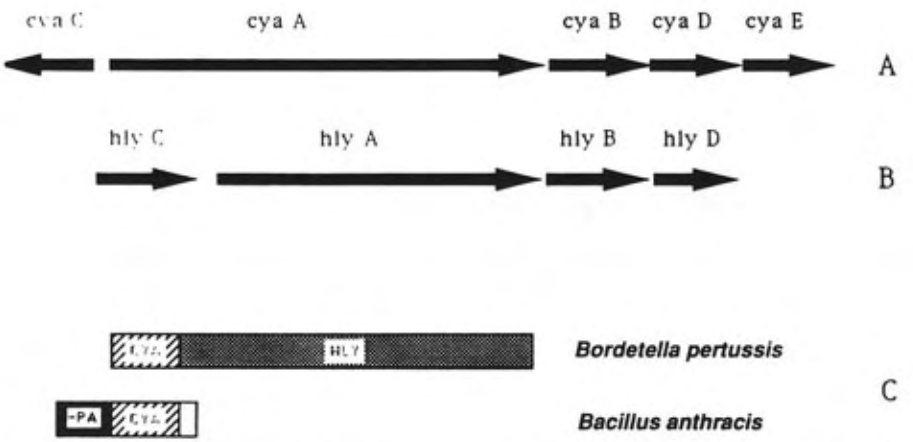


FIG. 2. The local gene organization of the *cya* locus in *B. pertussis* **(A)** is similar to that of the *hly* locus in pathogenic strains of *E. coli* **(B)**. Note however the opposite transcription orientation of *cya*C and *hly*C, and the lack of the *cya*E counterpart on the *hly* locus. When *B. pertussis* and *B. anthracis* proteins are compared **(C)**, their modular organization is prominent. The catalytic center is amino-terminal in the former and central in the latter, other domains are related to secretion or pathogenic activity (see text).

The adenylyl cyclase of *B. anthracis* has been termed the edema factor (EF), named after the symptom it triggers in the infected host. It is encoded in a plasmid, together with another toxin, named the lethal factor (LF) and a carrier protein, the protective antigen (PA), which is necessary for internalization of both EF and LF in host target cells (9,66,96,164,290). The adenylyl cyclase gene product shows corresponding structural features, comprised of three segments (Fig. 2). The first segment is a signal peptide, permitting secretion of the protein. Its carboxy-terminal end is not precisely known, but the NH_2-terminus of the active protein, MNEHT, is known (293), which may indicate that the signal peptide sequence is MFIPNKFSIISFSVLLFAISSSQAIEVNA, similar to known peptide sequences of bacteria (82). The second segment corresponds to the domain permitting binding with PA. The third domain encodes the adenylyl cyclase function. It is followed by a region of unknown function (Fig. 2).

These toxic adenylyl cyclases have been subjected to a most thorough biochemical analysis and they are at present the best known cyclases from the point of view of biochemical studies (105,148,149,165). It is possible that some mammalian adenylyl cyclases also belong to this class (see 59,201,248). For example, the calmodulin activable adenylyl cyclase activity present in cyanobacteria may belong to this class (14), although preliminary sequencing data do not support this (Lemeignan and Tandeau de Marsac, *unpublished data*). Finally work by Mishankin et al. suggested that *Yersinia pestis* possesses a secreted form of cyclase (197), which may belong to class II cyclases.

The "Universal" Class (Class III)

An ancestral class of purine nucleotide cyclases, comprising both adenylyl and guanylyl cyclase, has been discovered recently, following the sequencing of the genes or cDNAs from a variety of organisms (eucaryotic as well as procaryotic). In this section, the main features of the corresponding genes and proteins (where known) will be described.

Adenylyl Cyclase Types

Adenylyl cyclases from multicellular eucaryotes have remained elusive for a very long time, because the purification of the corresponding catalytic subunit is extremely difficult. In addition, the enzyme is subject to a complex regulation, so that one can never be sure that the catalytic activity will be expressed. Techniques such as the one used for cloning of bacterial adenylyl cyclase toxins are accordingly very difficult to implement. The use of affinity chromatography using forskolin (226) permitted purification of small amounts of the protein (35,227–230,271) and resulted in the cloning of adenylyl cyclase cDNA from bovine brain (144).

In *S. cerevisiae*, there existed mutants of the cell cycle that were also

affected in cAMP synthesis [*CDC35*, *CYR1*, and *TSM085* (186,191,192,302)]. As most genes in this organism are not interrupted by introns, this permitted cloning of chromosomal DNA by direct complementation of suitable mutants. Masson and coworkers cloned the wild type allele complementing *TSM085* (186), and Wigler and coworkers determined the complete gene sequence by direct complementation of a *S. cerevisiae CYR1* mutant (125). In parallel, and despite the observation of Uno et al. indicating that the *RAS* gene product was necessary for activity of *S. cerevisiae* cyclase expressed in *E. coli* (303), Masson and coworkers attempted to clone the gene directly in a *cya* deficient *E. coli* strain. This procedure yielded a gene that was expressed and able to synthesize cAMP in *E. coli* (187). The same authors then demonstrated that the corresponding gene was identical with the 3′ part of the gene described by Wigler's group, and coded for the catalytic domain of the adenylyl cyclase protein. They further demonstrated that this corresponded to a short major adenylyl cyclase mRNA transcript (187). This noteworthy feature (76) revealed that *two* different mRNAs of 6kb and 3kb could be generated from a single yeast adenylyl cyclase gene depending on the growth conditions (187). This is consistent with the fact that the amino-terminal part of the protein, which contains repetitions of a leucine rich motif (125,187) plays a regulatory role (249,260). It should be noted that the gene isolated by complementing the *TSM085* mutant harbors a TY1 sequence upstream of the promoter region for the long transcript, a feature which is absent from the gene isolated in Wigler's laboratory (162). Whether this is relevant to the expression mode resulting in a short and a long transcript remains open to question. Finally although there are slight sequence discrepancies in the very distal part of the genes cloned by both groups (in a region that does not seem to be involved in catalysis) their sequences are similar.

It was clear then that this eucaryotic adenylyl cyclase was completely different from the enterobacterial class, not only because of the sequence differences in the catalytic domain, but also because the organization of the gene was different: the catalytic domain is located at the COOH-terminus in *S. cerevisiae* cyclase, whereas it can be found at the NH_2-terminus in *E. coli*. This raised the question of a possible case of phylogenetic convergence for adenylyl cyclase evolution (37) and suggested that this new protein constituted the first member of a *third* adenylyl cyclase class. The yeast enzyme remained the only instance of its class until Garbers, Goeddel, and coworkers recognized *guanylyl cyclase* genes cloned from several metazoans, were clearly derived from the same ancestor as the yeast adenylyl cyclase (31,81, 176,261,270). Another member of this putative new class was then discovered by Young et al. (1989) who cloned the adenylyl cyclase gene from *Schizosaccharomyces pombe* by hybridization using the catalytic domain gene sequence from *S. cerevisiae* as a probe (320). Finally, the first higher eucaryote adenylyl cyclase gene isolated in Gilman's laboratory from bovine brain, displayed features clearly reminiscent of this class (144). Later, Beuve et al.

(1990) described an adenylyl cyclase gene, isolated from the Gram negative *procaryote Rhizobium meliloti* (133), that was also a member of the same class (13,13a). This was a remarkable observation as procaryotes are supposed to have separated from eucaryotes at least 2.5 billion years ago. Since then many other genes or cDNAs for adenylyl cyclase or guanylyl cyclase belonging to this universal class have been isolated and sequenced, not only from eucaryotes [*Saccharomyces kluyveri, S. pombe, Trypanosoma brucei* and *Trypanosoma equiperdum, Plasmodium falciparum, Dictyostelium discoideum, Drosophila melanogaster, Neurospora crassa*, human olfactive cells (4,7,81,140a,166a,219,230a,235,261,292,318,320,321)], but also from procaryotes such as the Gram positive organism, *Brevibacterium* (now *Corynebacterium*) *liquefaciens* (224,225,283). It was also recently discovered that the Gram negative sliding bacteria *Stigmatella aurantiaca* harbored *two* cyclase genes, both corresponding to the same adenylyl cyclase class (176a). These results substantiate the hypothesis that cyclases must have been present in cells *before* the separation between eucaryotes and procaryotes. An alternative explanation would be that horizontal gene exchange would have permitted capture of the corresponding eucaryotes' genes by many different procaryotes, but this seems doubtful. It appears therefore that synthesis of cAMP and cGMP (found for example in *Myxococcus xanthus* (220a) and *S. aurantiaca* (B. Lubochinsky, *personal communication, 1991*) predated not only the origin of eucaryotic cell differentiation but also the separation between eubacteria and eucaryotes. The question of the origin of adenylyl cyclases thus becomes an important phylogenic issue (see section *Are Cyclases the Result of Convergent Evolution?*).

Mammalian adenylyl cyclases have been informally grouped into several types, according to their tissue location and activity regulation (7,144,287). Type I enzymes were described as calmodulin-activated enzymes from brain (35,59,102,201,203,208,248), type II proteins are found in brain, lung, and other tissues (7,69,71a,128,201,268), type III are abundant in olfactory tissue (7,156,158,229), and the presumably smaller type IV enzymes are present in testicular tissue (19,78a,103,215). Several cDNAs corresponding to these types have been cloned, and some have been sequenced (7,71a,78a,144,219). It should be noted however that the classification into types is somewhat arbitrary (287); note for instance that type IV can also be calmodulin-activated (103,295). For this reason it cannot be concluded that other, completely different types are also present in mammalian tissue (see also section on *Cyclases as a Paradigm*).

The widespread origin of class III adenylyl cyclases is reflected by the wide variation in their structural organization and molecular weight. The smallest adenylyl cyclase is the *R. meliloti* enzyme, which contains only the catalytic domain. Yet other data suggests that upstream sequences may yield a longer protein showing a complex regulatory pattern (157). The next smallest cyclase protein is expressed by *B. liquifaciens* (225). *S. aurantica* con-

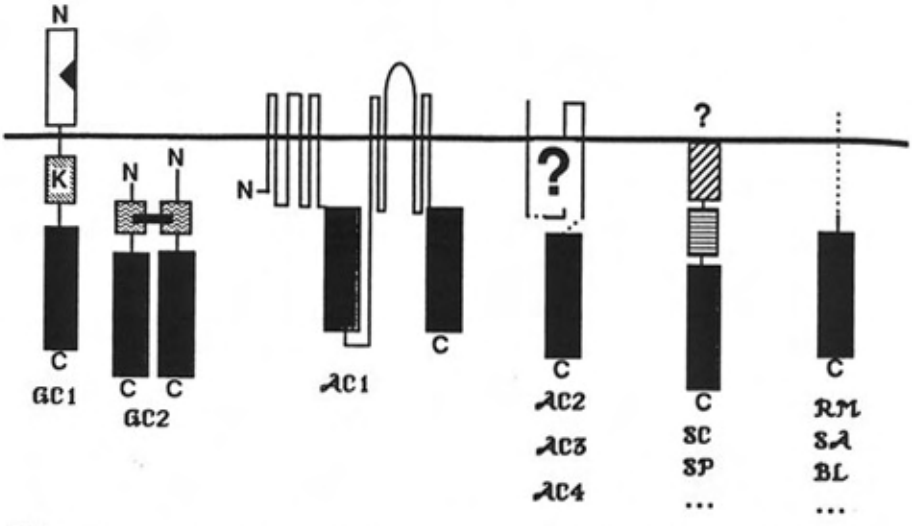


FIG. 3. Guanylyl and adenylyl cyclases, class III types, vary according to their modular organization. Two main types have been described for guanylyl cyclases, a membrane-bound type comprising a polypeptide with a single transmembrane segment (GC1), and a cytoplasmic type (GC2), containing a heme (for a more detailed description see text and 264). The situation of adenylyl cyclases is more complex, and less well understood. At least four types have been described, classified as a function of cell localization, pattern of activation, and subunit composition. Two main types exist: a polypeptide made of two catalytic domains, coupled to several transmembrane segments, as in the case of the bovine brain enzyme (AC1), and a polypeptide made of a single COOH-terminal catalytic center, coupled to a variety of different domains, as in all other cases (see text). Initials underscoring the schematic representation of enzymes stand for the name of the source of enzyme (e.g., Sc, *Saccharomyces cerevisiae*; Rm, *Rhizobium meliloti*, etc.).

tains two enzymes, of differing size (*unpublished data*). In all cases, the catalytic domain is located at the COOH-terminus. The type I enzyme, isolated from brain tissue is remarkable because it reveals a duplication of the domain thought to be responsible for catalysis within the same polypeptide chain (144). The question of the subunit composition of active form of the enzymes remains however open, especially when one recognizes that the related form of guanylyl cyclase is sometimes active as a (hetero)dimer (Fig. 3).

Guanylyl Cyclases

cGMP was discovered shortly after cAMP. Its concentration was found to vary according to tissues and environmental conditions, in a way reminiscent of cAMP. The role of cyclic GMP as a second messenger was perceived for some time as antagonistic to the role of cyclic AMP, the concentration of the former rising when the concentration of the latter went down and conversely.

Some data could be interpreted in this framework, but this hypothesis proved not to be universally true. Indeed cGMP is active in many highly specific processes, most of which differ significantly from those involving cAMP (79,195,296). This is the case in particular of phototransduction, in which cGMP plays a major role (see section on *Space and Time*). Although this process is formally similar to phenomena involving cAMP (274), it is controlled in a completely different manner. It became generally accepted that guanylyl cyclases may be enzymes quite distinct from adenylyl cyclases (for a recent review on cGMP see 296). In particular, two well-defined forms of the enzyme, cytoplasmic and particulate, displaying different activity and regulation, were found (81,264). As a consequence, they were described as significantly different proteins (27,80,81,138,139,261,262,263,292). The cytoplasmic enzyme was found to be a heterodimer containing two subunits, of 70 and 82 kDa, each having a carboxy-terminal catalytic domain, and unable to function as a monomer (206,207,264). They were shown to contain a heme as a prosthetic group, which modulates enzyme activity after binding of nitric oxide or other nitrogen containing gaseous molecules derived from arginine metabolism (e.g., 120,183,264). In contrast, the plasma membrane form of the enzyme was a multidomain protein made from the association of a peptide receptor domain (e.g., atrial natriuretic peptide receptor) a protein kinase-like domain and a guanylyl cyclase domain (31) (Fig. 3). Finally, Chinkers and coworkers demonstrated that guanylyl cyclases were structurally similar to each other as well as to adenylyl cyclase from yeast. This revealed that in spite of their differences in activity and overall organization, their catalytic domains were derived from a common ancestor (31,140) (Fig. 3).

PHYLOGENY OF THE CATALYTIC DOMAIN

A vast amount of experimental data has documented that eucaryotic adenylyl cyclases are enzymes regulated by GTP-binding proteins (G-proteins). Because many review articles (and books) have been written on the subject, I shall only quote a few of the recent ones, which permit the reader to browse through the immense literature on the subject (see for instance 16,86,100, 116,151,152,167,168,174,188,210,218,221,242,243,269,276,278,312). I shall not discuss the many fascinating features of the regulatory subunits nor their phylogenic relationships (17), in order to concentrate only on the polypeptide chain containing the catalytic center of the proteins. In the present section I shall restrict the discussion to the phylogenetic relationships between the catalytic domains, and consider the modular organization of the corresponding polypeptide chain in a later section (*Cyclases as a Paradigm*).

Enterobacteria and Related Bacteria

Generally speaking the nucleotide sequences of the enterobacterial genes are very similar, and variation occurs mainly at the third codon position

(41, and P. Glaser, O. Sismeiro, and A. Danchin, *unpublished data*), where transitions dominate over transversions. This is indicative of a recent divergence of the species, and could be used for analysis of constraints operating at the DNA level as compared to constraints operating at the protein level (48,196).

Comparison of the catalytic domain polypeptide sequence of the *E. coli* enzyme with known sequences present in the protein data libraries did not reveal significant identities with other known proteins. There was weak homology to *E. chrysanthemi* dehydrogenases, perhaps indicating the presence of a nucleotide-binding fold. A very weak similarity with the alpha subunit of membrane ATPases was also apparent, and the first 112 residues of the protein bear some similarity with pyrophospate-binding proteins (for example tRNA synthetases), perhaps indicating that the very amino terminus of the protein is involved in pyrophosphate binding. These similarities are however not sufficient to indicate that this class of adenylyl cyclases is derived from an ancestral ATP-binding protein.

Knowledge of the catalytic domain sequence (420 residues) from five class I genes permits us to construct a phylogenic tree (Fig. 4). Adenylyl cyclases from *Salmonella typhimurium* and *Escherichia coli* are clustered together, as are the two *Yersinia* species. This is in line with the known ecological properties of the strains (21), and indicates that the taxonomic criteria that have been used to define the corresponding genera are not entirely satisfactory: from the point of view of adenylyl cyclase evolution, if *Yersinia inter-*

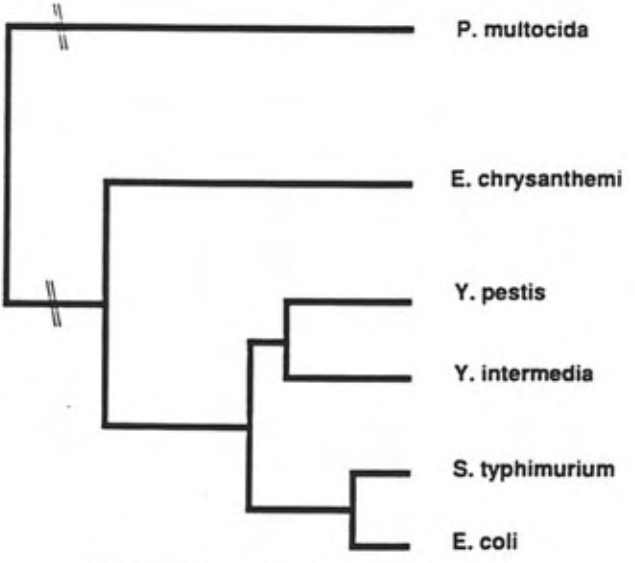


FIG. 4. A phylogenic tree for class I cyclases.

media and *Yersinia pestis* belong to the same genus, *Salmonella typhimurium* and *Escherichia coli* should belong to a common genus. The *Erwinia chrysanthemi* enzyme is clearly distinct from both groups, which is at variance with the apparent proximity derived from the analysis of other DNA sequences (20). However the position within a taxonomical tree cannot necessarily be compared when the various organisms live in very different ecological niches. Indeed differences in evolution of a specific protein might be irrelevant from the point of view of taxonomical phylogeny, but rather be related to a selection pressure that results in special physicochemical properties of the protein. In the present situation the original ecological niche of *Erwinia* species [which are known to grow on media very different from the media generally used for typical enterobacteria (21)] certainly differs from the environment of bacteria whose hosts are warm-blooded animals. Indeed, *E. chrysanthemi* cannot grow at temperatures higher than 30°C. It may therefore be that the structure of its adenylyl cyclase is better adapted to low temperatures than its counterparts in bacteria meant to grow at 37°C. The *Pasteurella multocida* enzyme is clearly distant from all enterobacterial adenylyl cyclases. This is in agreement with the accepted distance between the corresponding families, and substantiates that the previous taxonomic reorganization was instituted some time ago (182).

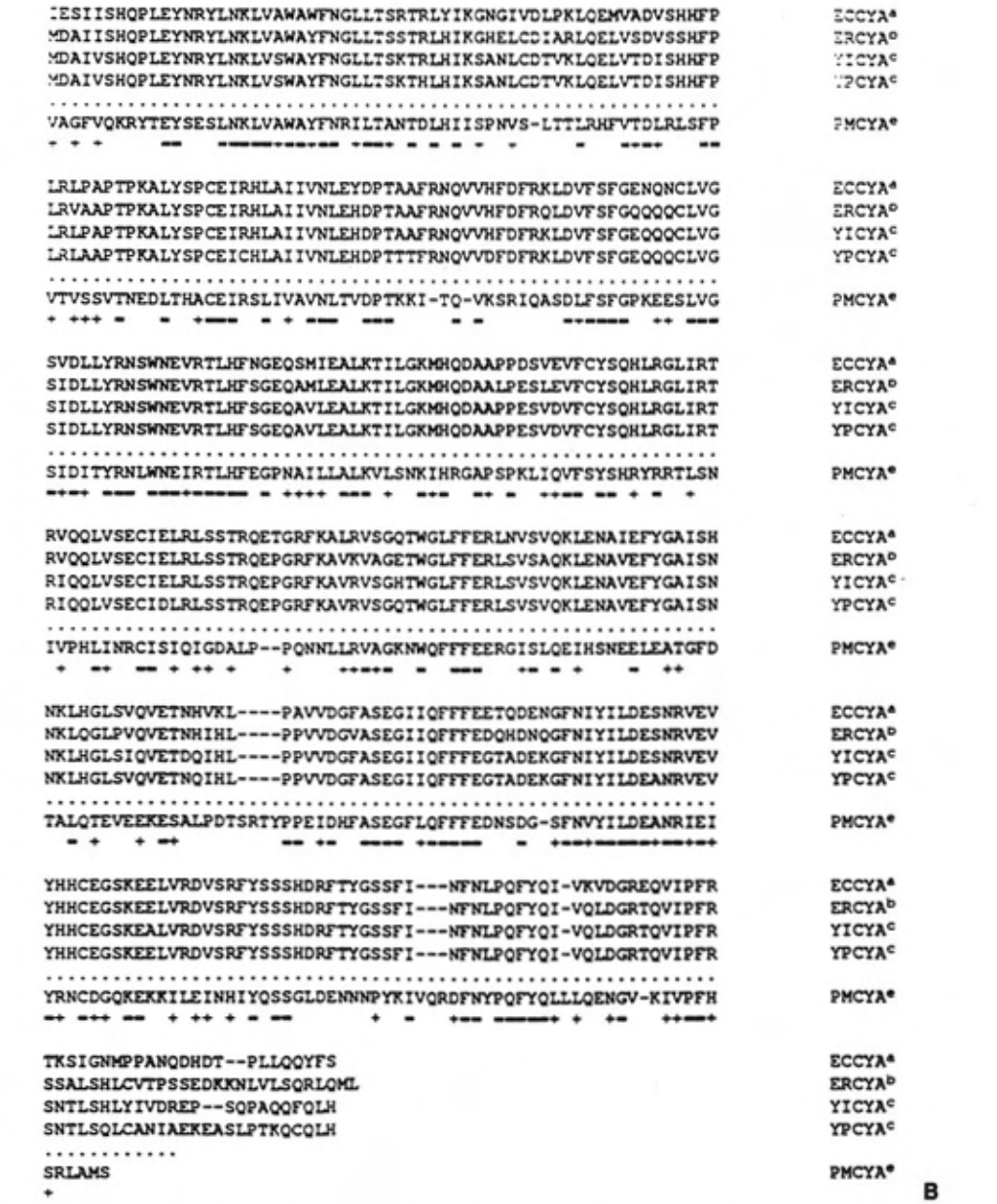
Four insertions must be included in the catalytic domain of the *Pasteurella* enzyme, when compared to the enterobacterial counterparts, and the overall identity score between residues remains significant (38%). This comparison permits identification of regions of major conservation, that must be important for catalysis and/or proper folding of the protein (Fig. 5). Several clustered histidine or cysteine residues are conserved, suggesting that a metal ion may be involved in the protein structure or activity.

Toxic Cyclases

The bacterial cyclases, which are activated by a eucaryotic host protein, calmodulin, is an intriguing puzzle. In spite of several attempts to isolate other members of this class, only two examples of such proteins have been isolated from extremely distant organisms, a Gram positive one and a Gram negative one. Comparison between the catalytic region of the *B. pertussis* and *B. anthracis* enzymes permitted us to identify four conserved regions that should be involved in catalysis, calmodulin binding and activation. The first region (see Fig. 6) comprises a peptide, G----G(A)KS, having a sequence similar to the nucleotide-binding region found in many ATP- or GTP-binding proteins (106). It was therefore proposed to be part of the catalytic site (85,91) and *in vitro* mutagenesis of the enzyme substantiated this interpretation (93,257). A second region, PLTADID, displaying some similarity with 6-phosphofructokinase, was also demonstrated to be involved in several as-


MYLYIETLKQRLDAINQLRVDRALAAMGPAFQQVYSLLPTLLHYHHPLMPGYLDGNVPKG ECCYA*
MYFYIETLKQRLDAINQLRVDRALEAMKPAFQQVYSLLPVLLHHHHPLMPGYLEGKVPHG ERCYA*
MYLYIETLKQRLDAINQLRVDRALAAMGPTFQKVYSLLPTLLHCHHPLMPGYLDGNVPHG ITCYA*
MYLYIETLKQRLDAINQLRVDRALAAMGPAFQKVYSLLPILLHCHHPQMPGYLDGNVPHG YPCYA*
MYLYIETLKQRLDAINQLRVDRALAAMGPAFQQVYSLLPTLLHYHHPLMPGYLDGNVPSG STCYA*
MNYDLFSAQKKVEYLDKLRIERALSGSSGEFQHVFQLLTLLLHINHPNLPGYV-ADAPVG PMCYA*
ICLYTPDETQRHYLNELELYRGMSVQDPPKGELPITGVYTMGSTSSVGQSCS----SDLD ECCYA*
ICLFSPDEKQQHYLDSVELRWGELSAPDRKGELPITGVYSMGSTSSIGQSCS----SDLD ERCYA*
VCLFTPNETQQDYLSEVEAKWGEPLQQSVGGELPITGVYSMGSTSSIGQCHT----SDLD YICYA*
VCLFTPNEIQQDYLADVEARWGEPLAPSAGGELPITGVYSMGSTSSIGQCHT----SDLD YPCYA*
ICFYTPDETQRHYLNELELYRGMTPQDPPKGELPITGVYTMGSTSSVGQSCS----SDLD STCYA*
IADFVISPYQKQYLLTTVPSLEANQSLLPSFSYRSTNAILGVYVMGSIASISQTPKSDLD PMCYA*
IWVCHQSWLDSEERQLLQRKCSLLENWAASLGVEVSFFLIDENRFRHNESG-SLGGEDCG ECCYA*
IWVCHQSWLDNEERQ-LQQKCSLLEKWAAGQGVDVSFFLMDENRFRHNESG-SLGGEDCG ERCYA*
VWVCHQAWLDSEERNRLQEKCSLLEKWAASMGVEVSFFLIDENRFRHNASG-SLGGEDCG YICYA*
IWVCHQAWLDTEERNQLQQKCSLLEKWAASMGVEVSFFLVDENRFRHNASG-SLGGEDCG YPCYA*
IWVCHQSWLDGEERQLLQRKCSLLESWAASLGVEVSFFLIDENRFRHNESG-SLGGEDCG STCYA*
TWVCHRDDLSTKEKEALQRKTHLLKNWAKQFNIEINFYLMDQKRFRCFRYAEPLTAENCG PMCYA*
STQHILLLDEFYRTAVRLAGKRILWNMVPCDEEEHYDDYVMTLYAQGGLTPNEWLDLGGL ECCYA*
STQHILLLDEFYRTAVRMAGKRILWNMVPVEEEAHYDEFVLSLYARGALAPNEWLDLGGL ERCYA*
STQHILLLDEFYRSAVRLAGKRILWNMVPVEEENNYDDYVLSLYAQGVLTPNEWLDLGGL YICYA*
STQHILLLDEFYRSAVRLAGKRILWNMVPVKEEEHYDDYVLSLYREGVLTPNEWLDLGGL YPCYA*
STQHILLLDEFYRTAVRLAGKRILWSMVPCDEEEHYDDYVMTLYAQGVLTPNEWLDLGGL STCYA*
SAQYMLLLDEFYRSAIRLAGKPLLWLHLLIEQEENYESEVERLVRTQQICLDDWVDFGGL PMCYA*
SSLSAEEYFGASLWQLYKSIDSPYKAVLKTLLLEAYSWEYPNPRLLAKDIKQRLHDGEIV ECCYA*
SALSAEEYFGASLWQLYKSIDSPYKAVLKTLLLEAYSWEYPNTRLLSSEIKARLHKGEIV ERCYA*
STLSAEEYFGASLWQLYKSIDSPYKAVLKTLLLEAYSWEYPNSQLLAMEIKQRLHAGEIV YICYA*
STLSAEEYFGASLWQLYKSIDSPYKAVLKTLLLEAYSWEYPKSQLLAMEIKQHLHAGEIV YPCYA*
SSLSAEEYFGATLWQLYKSIDSPYKAVLKTLLLEAYSWEYPNPRLLAKDIKQRLHDGEIV STCYA*
GQLSANEYFGASLWQLYKGIDAPYKSVIKILLLETYSSEYPNTYLIARQFKEELLTGKLN PMCYA*
-SFGLDPYCMMLERVTEYLTAIEDFTRLDLVRRCFYLKVCEKLSR--ERACVGWRRAVLS ECCYA*
-SFGLDPYCMMLERVTQYLDAINDQTRLDLVRRCFYLKVCEKLSR--ERACTAWRRQILT ERCYA*
-AFGLDAYCMMLDRVTRYLTQINDTTRLNLVRRCFYLKVCEKLSR--SPASVGWRREILS YICYA*
-SFGLDAYCMMLDRVTRYLIQINDTTRLNLVRRCFYLKVCEKLSR--TPASTGWRREVLS YPCYA*
-SFGLDPYCMMLERVTEYLTAIEDPTGLDLVRRCFYLKVCEKLSR--ERACVGWRREVLS STCYA*
PSHHFDPYLAMLQRATRYLTKHNELKRLGFVRRSVYLKATEGMCWQDPNATNNWRLQHLQ PMCYA*
QLVSEWGWDEARLAMLDNRANWKIDQVREAHNELLDAMMQSYRNLIRFARRNNLSVSASP ECCYA*
QMVQAWGWSDERLVMLDNRANWKIGQVREAHNELLDAMMQSYRNLIRFARRNNLSVSASP ERCYA*
QLVSEWGWSDESLAVLDNRANWKIERVREAHNELLDAMMQSYRNLIRFARRNNLSVSASP YICYA*
QLVSEWGWSNEKLAVLDNRANWKIERVREAHNELLDAMMQSYRNLIRFARRNNLSVSASP YPCYA*
QLVSEWGWDDARLTMLDNRANWKIDQVREAHNELLDAMMQSYRNLNRFARRNNLSVSASP STCYA*
KLIQEWDWSDALIEELNQRANWKIKQVKKAHNSLIKFLMLSYRNLVAFARKHKVNSSIMP PMCYA*
QDIGVLTRKLYAAFEALPGKVTLVNPQISPDLSEPNLTFIYVPPGRANRSGWYLYNRAPN ECCYA*
QDIGVLTRKLYAAFEALPGKVTLVNPQISPDLSETNLTFIYVPAGRANRSGWYLYNQAPS ERCYA*
QDIGVLTRKLYAAFEALPGKVTLVNPQISPDLSEEHLTFIHVPAGRANRAGWYLYNQAPS YICYA*
QDIGVLTRKLYAAFEALPGKVTLVNPQISPDLSEEHLTFIHVPAGRANRPGWYLYNQAPS YPCYA*
QDIGVLT..... STCYA*
QDISVLTRKLYTAFEELPGKITLLNPQISLNLSEKNLLFFEVKGSKTFKAGWYVVNQTPS PMCYA*

A



B

FIG. 5A,B. A direct comparison of the known sequences of class I proteins reveals blocks of identity (=) or conservative replacements (+). [a], ref. 2; [b], ref. 41; [c], *unpublished results*; [d], ref. 291; [e], ref. 200.

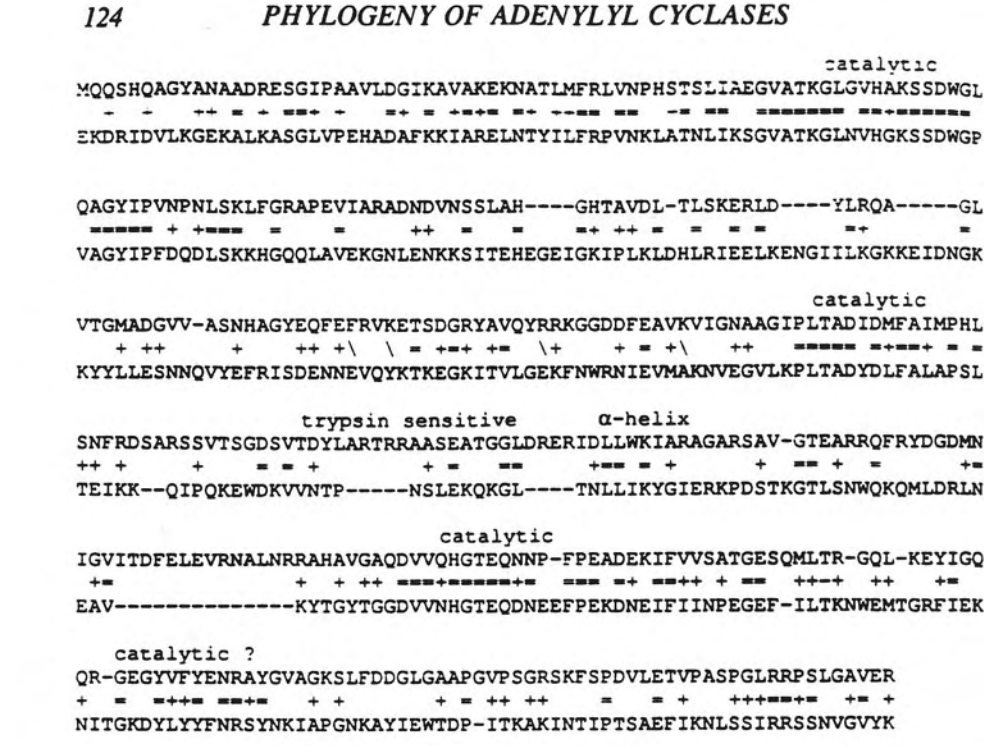


FIG. 6. Comparison of the catalytic domain of class II enzymes reveals three zones of extensive similarity (= for identity, and + for conservative replacements). A fourth zone at the carboxy-terminal end is also probably of major importance.

pects of catalysis, and it was proposed that the aspartate residues present in this region are involved in ribose and magnesium-phosphate binding (93). In spite of this strong conservation, the first proline residue does not seem very important since it can be replaced by a leucine residue without any measurable influence on the activity or calmodulin activation of the native enzyme (93).

What can be said about the calmodulin-binding sites of the enzymes? Calmodulin-dependent enzymes represent a heterogeneous class of catalysts. Although many such enzymes have been identified, the precise mechanism of activation by calmodulin is still not clearly understood. In several cases, limited proteolysis released active calmodulin-independent forms of enzymes. Accordingly, it was proposed that the calmodulin-binding domain of these enzymes blocks the access of substrates to the active site and that activation results because following binding of calmodulin an inhibitory domain is removed. This phenomenon could be mimicked by removal of this domain with proteases. Nucleotide-binding experiments on *Bacillus anthracis* adenylyl cyclase, showing that the enzyme binds the ATP analog 3′dATP only (or preferentially) in the presence of calmodulin, may support

a mechanism involving a catalytic site blocked by the calmodulin-binding site (257). Tryptic cleavage of the low-molecular-mass forms of *B. pertussis* adenylyl cyclase yields two complementary fragments of 25 kDa (T28) and 18 kDa (T19) (149). By combining genetic and biochemical information it has been shown that the 25-kDa fragment harbors most of the active site, whereas the 18-kDa fragment corresponds mainly to the calmodulin-binding domain of the enzyme (91,93,149). Highly active forms of the enzyme were obtained when the calmodulin-binding domain (T19) was reassociated with the catalytic domain in the presence of calmodulin. This suggests that both T28 and T19 possess well-defined three-dimensional structures enabling them to recognize each other and their respective ligands. The higher sensitivity of isolated fragments to denaturation by urea compared to that of uncleaved adenylyl cyclase suggests that interactive contacts between ordered portions of T28 and T19 in intact protein are essential not only for full catalytic activity, but also participate both in their own stabilization and in the stabilization of the whole tertiary structure. Although biochemical and genetic experiments have established unambiguously that the α-helical sequence situated around Trp242 in *B. pertussis* adenylyl cyclase is involved in binding of calmodulin, the exact extent of the calmodulin-binding sequence in the intact protein has not yet been firmly established. Experiments using random insertional mutagenesis of DNA fragments into the *B. pertussis* adenylyl cyclase gene have indicated, as already suggested by the lack of identity between the *B. pertussis* and *B. anthracis* enzymes in this region (Fig. 6), that the region situated immediately upstream of Arg235 is not of major importance for calmodulin binding. However, the carboxy-terminal end of the binding domain could not be precisely mapped using this technique (149a). It is possible that this domain spans a region located somewhat upstream of the C-terminal end of T28 (see the regular distribution of charged residues upstream of the PLTADID region in the *B. pertussis* and *B. anthracis* enzymes), and the first 30 residues of T19, which display similar secondary structure features in the *B. pertussis* and *B. anthracis* enzymes (Fig. 7). One might speculate that interactions with T28-calmodulin in the intact protein, though weak, are able to place the calmodulin-binding site in a proper orientation. Indeed, upon reassociation of T28 and T19 into an active species, the affinity of reconstituted protein for calmodulin increased, as shown by titration of dansyl-calmodulin with T19 in the absence and in the presence of T28 (93). Thus some of the catalytic center must be located in the latter peptide, whereas some interaction with calmodulin exists in the former (93). This suggests that the active protein is somehow folded around calmodulin, a central helical structure of adenylyl cyclase being involved as a scaffolding device for this interaction (91,105,147,205) (Fig. 7).

Proper adjustment of the protein conformation for efficient catalytic activity is certainly mediated by interaction with water. If this is true, addition of appropriate solvents should permit recovery of some activity, even in the

absence of calmodulin, as has been observed in the case of 5'AMP-mediated activation of glycogen phosphorylase *b* (54,55). This may be the reason for the sometimes only partial requirement of *B. pertussis* adenylyl cyclase for calmodulin.

Thus, the analysis of mutants in the region conserved between the *B. anthracis* and *B. pertussis* enzymes indicates that the protein may require the cooperation of two halves of the molecule, the function of calmodulin (or lipophilic solvents, in the case of the *B. pertussis* enzyme) being to trigger the appropriate conformational change to form the active catalytic center. This is reminiscent of the situation for cAMP-dependent protein kinases (135,136), and it is worth considering whether there is any evolutionary relationship between this class of ATP-binding proteins and toxic cyclases.

The Class III Catalytic Center

Analysis of the class III cyclases indicates several amino acid motifs specific for these proteins, such as F-DI--FT-L, QVV-LLND, VK(E)T-GD--M, R-GIH-G, GDTVN-ASR (Fig. 7). At first sight, they are likely to be involved in the cyclization reaction rather than in the heterocyclic base recognition, since they are common both to adenylyl and guanylyl cyclases. Analysis of the upstream part of the domain suggested, however, that the third motif could be involved in base discrimination, because it is different in adenylyl cyclases (VKT) and guanylyl cyclases (VET) (225). However the set of individual adenylyl cyclase examples encompasses a wider series of organisms, from the point of view of phylogeny, than the guanylyl cyclase set, which was obtained only from metazoans.

As a step toward understanding evolution and function of class III cyclases, a genetic screening procedure was designed to allow direct analysis of the evolution from adenylyl cyclase to guanylyl cyclase and vice versa. The system utilized an adenylyl cyclase deficient *E. coli* strain, mutated in its cAMP receptor so that it could recognize cGMP. Mutated adenylyl cyclase genes were then introduced into the strain under conditions in which only bacteria synthesizing cGMP could grow. Enzymes displaying significant guanylyl cyclase activity having evolved from an adenylyl cyclase ancestor were thus isolated. A similar experiment was performed, driving evolution from guanylyl cyclase to adenylyl cyclase. It seems noteworthy that, as in the case of cAMP-dependent protein kinases made to recognize cGMP (267,311), a single amino acid residue change, in the region located between R-GIH and GDTVN changed the specificity of the enzyme. Thus, it is likely that a pocket situated in this region of the protein accommodates the heterocyclic base and that amino acid side chains discriminate between alanine and glycine by appropriate occupancy of a small portion of the pocket. This is substantiated by the observation that a further mutation (from GDTVN to

```
GIESGHFGGRAAWRQEVTAMFTDIYDFTTI----SEGRSPEEVVAMLSEYFDLFSEVVA---AHDGTIIQFHGDSVFAMWNAP-VADTRHA    A RHIME
TTLRRLQPEISPPTGNLAMVFTDIKSSTFL-----WELFPNAMRTAIKTHNDIMRRQLR---IYGGYEVKTEGDAFMVAFPTPTSGLTWCL    A SACCE
SSLARMNREVSPPKGCIAMVFTDIKNSTLL-----WERHPIAMRSAIKTHNTIMRRQLR---ATGGYEVKTEGDAFMVCFQTVPAALLWCF    A SCHPO
MSNPRNMDLYYQSYSQVGVMFASIPNFNDFYIELDGNNMGVECLRLLNEIIADFDELMDKDFYKDLEKIKTIGSTYMAAVGLAPTAGTKA-    A BOVINb
PPERIFHKIYIQRHDNVSILFADIVGFTGL----ASQCTAQELVKLLNELFGKFDELAT---ENHCRRIKILGDCYYCVSGLTQ-PKTDHA    A BOVINa
NQDRGEDDGGTPLPLARAVGFADLVSYTSL----SRRMNERTLAQLVQRFEAKCAEIIS---VGGGRLVKTIGDEVLYVAETPQAGAQIAL    A BRELI
RLQDLDRSETSPELREVTLLFADIRDFTSL----SERLRPEQVVTLLNEYYGRMVEVVF---RHGGTLDKFIGDALMVYFGAPIADP-AHA    A STIAU1
ILKSPDTVVLTGEKREVTVLFADIRNFTGL----AESLPPEQVVGVLNQVLGRLSDAVL---TCGGTLDKFLGDGLMAVWGAPVHRTDDAL    A STIAU2
RNKRDNDNAPKELADPVTLIFTDIESSTAQ-----WATQPELMPDAVATHHSMVRSLIE---NYDCYEVKTVGDSFMIACKSPFAAVQLA-    A TRYBR2
HNTRDNNLAPKELTDPVTLIFTDIESSTAL-----WAAHPELMPDAVATHHRLIRSLIG---RYGCYEVKTVGDSFMIASKSPFAAVQLA-    A TRYEQ
GSKKRDEELYSQSYDEIGVMFASLPNFADFYTEESINNGGIECLRFLNEIISDFDSLLDNPKFRVITKIKTIGSTYMAASGVTPDVNTNG-    A RAT
EQLKRGETVQAEAFDSVTIYFSDIVGFTAL----SAESTPMQVVTLLNDLYTCFDAIID---NFDVYKVETIGDAYMVVSGLPGRNGQRHA    G ANPBRAT
EQLKRGETVQAEAFDSVTIYFSDIVGFTAL----SAESTPMQVVTLLNDLYTCFDAIID---NFDVYKVETIGDAYMVVSGLPGRNGQRHA    G ANPBHUMAN
EQLKRGETVQAEAFDSVTIYFSDIVGFTAL----SAESTPMQVVTLLNDLYTCFDAVID---NFDVYKVETIGDAYMVVSGLPVRNGRLHA    G ANPAHUMAN
EQLKRGETVQAEAFDSVTIYFSDIVGFTAL----SAESTPMQVVTLLNDLYTCFDAVID---NFDVYKVETIGDAYMVVSGLPVRNGQLHA    G ANPAMOUSE
SQLIKGIAVLPETFEMVSIFFSDIVGFTAL----SAASTPIQVVNLLNDLYTLFDAIIS---NYDVYKVETIGDAYMLVSGLPLRNGDRHA    G STRPU
EQLKRGETVQAEAFDSVTIYFSDIVGFTAL----SAESTPMQVVTLLNDLYTCFDAVID---NFDVYKVETIGDAYMVVSGLPVRNGQLHA    G ANPARAT
NELRHKRPVPAKRYDNVTILFSGIVGFNAFCSKHASGEGAMKIVNLLNDLYTRFDTLTDSRKNPFVYKVETVGDKYMTVSGLPEPCI-HHA    G CYTRAT
NELRHKRPVPAKRYDNVTILFSGIVGFNAFCSKHASGEGAMKIVNLLNDLYTRFDTLTDSRKNPFVYKVETVGDKYMTVSGLPEPCI-HHA    G CYTBOVIN
KSLKEKGIVEPELYEEVTIYFSDIVGFTTIC-K---YSTPMEVVDMLNDIYKSFDQIVD---HHDVYKVETIGDAYVVASGLPMRNGNRHA    G ARBPU
RQLWQGHAVQAKRFGNVTMLFSDIVGFTAIC----SQCSPLQVITMLNALYTRFDRQCG---ELDVYKVETIGDAYCVAGGL-HKESDTHA    G CYTBOVIN2

A   eqLkrgetVqaeafdsVTIyFSdIVGFtAl    saestpmqvVtlLNDLYTcFDa*id   nfdVYKVETiGDaYmvvSGLP rng* HA    G
        r           v***FaDI  fT l     *    pe *   *ne      el*      gg vvKt*GD**m   g p       a    A
```

FIG. 7A–C. A comparison of class III cyclases catalytic domains. The guanylyl cyclase set (*G*) seems less variable, but this may be due to the fact that it is derived from metazoan sequences only. The enzymes are identified by the name of the organism or tissue (references in text). Two sequences, corresponding to two putative catalytic domains, are indicated for the bovine brain enzyme (*BOVINa* and *b*). The most frequent residues are noted underneath the sequences, for adenylyl cyclases (*A*) and guanylyl cyclases (*G*). Asterisks indicate conservative replacements (from the classes ASTPG, MLIV, DNQE, KHR, FYW).

FIG. 7. (*Continued*)

```
GDTVNVASRLEG-MNKDY-GTSV-LAS-GAVVAQ-CKDMVKFRP---L-GTAKAKGRSTALDIYEVVGVVRAVNTTEAGTAA*    A RHIME
GPMVNKAARVQGVADGGQIAMSSDFYSEFNKI---MKYHER-----------VVKGKESLKEVYGEEIIGEVLEREIAMLESI    A SACCE
GPVVNRTSRVVSVADGGQIAVSAEVVSVLNQLDSETMSSEKTNVNEMEVRALKQIGYIIHNLGEFKLKGLDTTEMISLVYPVQ    A SCHPO
GNTVNVASRMDS--T-GVQ-GRIQVTEEVHRLLRRGSYRF-------VCGKVSVKGKGEMLTYFLEGRTDGNGSQTRSLNSER    A BOVINb
SNDVTLANVMEA--A-GLP-GKVHITKTTLACLN-GDYEVEP-----------GHGHERNSFLKTHNIETFFIVPSHRRKIFP    A BOVINa
GPTVNMAARLTSLAEPGTVLTDAITANT-LRNDARFVLTAQEITAVRGFGDIQPYELSAGEGAGLVID*                  A BRELI
GDTVNLASRIESLTKTRDVPILASRATRE-QAGDTFLWNEMAP------ASVPGKSQPVAIFTPRNRTPAQQAGAPAAA*      A STIAU1
GDAVNVAARLCALAGPGE--ILAGERTRELVSHREMPFEDLPP------VRLKGKQQPVPLYRVL*                     A STIAU2
GQTANTAARTESVGNGGQV-LMTCETYHSLSTAERSQFD-VTP----L-GGVPLRGVSEPVEVYQLNAVPGRSFAELRLDRVL    A TRYBR2
GRTSNMAARTESVGNGGQV-LMTTAAYMSLSAEEREQID-VTA----L-GDVPLRGVAKPVEMYQLNAVPGRTFAGLRLEHEL    A TRYEQ
GNTVNVASRMESTGVMGNI----QVVEETQVILREYGFRFVR------RGPIFVKGKGELLTFFLKGRDRPAAFPNGSSVTLP    A RAT
GDTVNTASRMES--N-GQ-ALKIHVSSTTKDALDELGCFQLE-----LRGDVEMKGKGKMRTYWLLGERKGPPGLL*         G ANPBRAT
GDTVNTASRMES--N-GQ-ALKIHVSSTTKDALDELGCFQLE-----LRGDVEMKGKGKMRTYWLLGERKGPPGLL*         G ANPBHUMAN
GDTVNTASRMES--N-GE-ALKIHLSSETKAVLEEFGGFELE-----LRGDVEMKGKGKVRTYWLLGER-GSSTRG*         G ANPAHUMAN
GDTVNTASRMES--N-GE-ALRIHLSSETKAVLEEFDGFELE-----LRGDVEMKGKGKVRSYWLLGDR-GCSSRA*         G ANPAMOUSE
GDTVNTASRMES--N-GL-ALRIHVSPWCKQVLDKLGGYELE-----DRGLVPMNGKGEIHTFWLLGQDPSYKITKVKPPPQK    G STRPU
GDTVNTASRMES--N-GE-ALKIHLSSETKAVLEEFDGFELE-----LRGDVEMKGKGKVRTYWLLGERGCSTRG*         G ANPARAT
GNTVNLTSRTET--T-GEKG-KINVSEYTYRCLMSPENSDPQ-FHLEHRGPVSMKGKKEPMQVWFLSRKN--TGTEETNQDEN    G CYTRAT
GNTVNLTSRTET--T-GEKG-KINVSEYTYRCLMTPENSDPQ-FHLEHRGPVSMKGKKEPMQVWFLSRKN--TGTEETEQDEN    G CYTBOVIN
GDTVNTASRMES--T-G-LPLRIHMSSSTIAILRRTDCQFL----YEVRGETYLKGRGTETTYWLTGMKDQEYNLPTPPTVEN    G ARBPU
GNNVTLANKFESCSVPRKINVSPTTYRLLKDCPGFVFTPRSREELPPNFPSDIPGICHFLEAYQQGTTSKPWFQKKDVEEANA    G CYTBOVIN2

  GdTVNtASRMES  n Ge alkIhvSs Tk  L e   f le     lRGdVeMkGKgk rtyWlLg                  G
C G tVN AaR*es***gGq    *      l       *         l      *G       *                     A
```

FIG. 7. (*Continued*)

GDTIN) shifts the enzyme activity toward a general purine nucleoside triphosphate cyclizing activity (13a).

The apparent facility with which it was possible to build up a purine nucleoside triphosphate cyclase of broad specificity might be relevant to phylogeny of the catalytic center of class III enzymes. As stated above, Garbers, Goeddel, and coworkers found that the catalytic centers of guanylyl cyclases were strongly related to eucaryotic class III adenylyl cyclases. In general, gene organization revealed conservation in both types of enzymes, of the localization of at least one catalytic domain in the carboxy-terminal part of the protein, coupled to a large variety (in length and in sequence) of amino-terminal parts. In the case of the bovine brain enzyme, a duplication of the catalytic domain was separated by a hydrophobic stretch of amino acid residues. The common origin between adenylyl cyclases and guanylyl cyclases suggested the existence of an ancestral purine nucleotide triphosphate cyclase as a precursor. As a consequence, one may wonder whether evolution has not permitted some association between both nucleotide syntheses, such that a given cyclase, under appropriate conditions, could synthesize either cAMP or cGMP.

Are Cyclases the Result of Convergent Evolution?

The discovery that a class of cyclases can derive from a common ancestor predating the eubacterial and eucaryotic separation suggests that originally, the function of cAMP differed from that of a secondary messenger. It is clear that the adenylyl cyclase reaction is easily reversed, because of the equilibrium constant for ATP synthesis (173, and O. Barzu, *personal communication, 1991*):

$$K_{eq} = \frac{[cAMP][PP_i][H^+]}{[ATP]} = 2.1 \pm 0.2\ 10^{-9}\ M^2$$

and that under appropriate conditions one can generate ATP (or GTP) from cAMP (or cGMP) and pyrophosphate. The hypothesis that cAMP was originally involved in energy scavenging or production, rather than regulation seems therefore worth considering. There is a weak similarity between this general class of cyclases and ATP synthases (Fig. 8). Indeed phosphates (and polyphosphates) certainly play a very significant role in biological processes (12,25,38,306). Generation of a chemiosmotic function allowing synthesis of ATP (from ADP) requires numerous steps. It may therefore be considered that originally cAMP could have been a building block for ATP. This requires a source for cAMP synthesis. This could have been included in the many catalytic activities of self-splicing introns involving processes of transesterification. One may wonder whether cAMP or cGMP could not be by-products of such reactions, providing an early source for these molecules. A first class

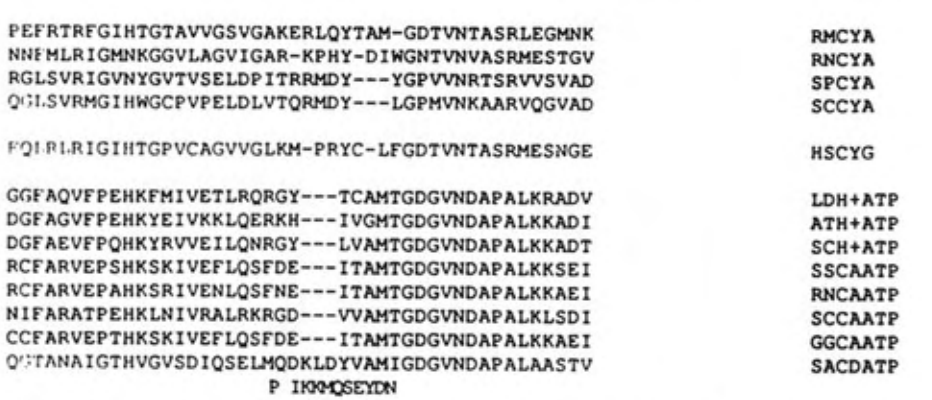


FIG. 8. Comparison of class III proteins segments with ATP synthases (α-subunits) and ATP-dependent ion pumps. The four first sequences are from adenylyl cyclases, the fifth is from human guanylyl cyclase. The other sequences correspond to proton, calcium, and cadmium pumps, respectively (24, 57, 108, 124, 193, 212, 255, 266).

of enzymes predating class III adenylyl cyclases would therefore be involved in ATP synthesis; a second, very important enzyme, adenylyl kinase, would then scavenge AMP, generating ADP. A general source of ADP thus available might have triggered evolution toward generation of ATP synthase. Some similarity can be discovered with such enzymes (284), as well as with enzymes using ATP for positive ion translocation (265). Finally it should be noticed that the class of toxic adenylyl cyclases is phylogenetically related to adenylyl kinase, at least when considering the first block of conserved amino-acid residues in *B. pertussis* and *B. anthracis* enzymes. This may provide a link between both evolution routes (from 5'AMP to ATP and from 3'5'cAMP to ATP).

Another observation may add substance to the validity of this hypothetical relationship. Antibodies raised against *B. pertussis* adenylyl cyclase, or against a peptide motif common to the *B. pertussis* and *B. anthracis* enzymes recognized a human brain adenylyl cyclase (98,202,216). This suggested that higher eucaryotic enzymes might display similarities with the calmodulin-activated toxic proteins. Yet, an adenylyl cyclase gene from bovine brain, identified by the Gilman group, was clearly more similar to the yeast enzyme (class III) than to adenylyl cyclase toxins (class II) (144). It is clear however that immune recognition does not demonstrate true phylogenetic relationships. The case of the sweet proteins thaumatin and mellitin is revealing in this respect, because, although recognized by the same receptors and the same antibodies, they do not share any common feature in amino acid sequence primary structure (134). Furthermore one should still remain open to the possibility that several classes of brain adenylyl cyclases existed, among which one class might be similar to the class II toxic type. Therefore it is

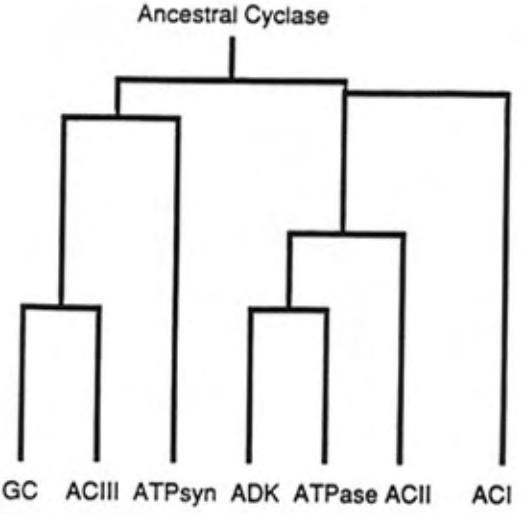


FIG. 9. A hypothetical tree relating the three classes of cyclases. Some relationships are highly speculative and should only be taken as a tentative hypothesis.

difficult at the moment to know whether there exists a true similarity between class II and class III enzymes. A tentative phylogenetic tree could be proposed (Fig. 9). This corresponds to an integrated view, where all cyclases derive from an ancestral nucleotide-binding protein, made of small independent modules as seen in protein kinases (135,136). Alternatively, we could be faced with a case of convergent evolution (146). For instance, the different adenylyl cyclases classes could have evolved independently, with the oldest activity corresponding to class III enzymes, and being derived from an ancestral NTP synthesizing enzyme (13b). This clearly shows that we are much in need of crystallographic data for all three classes; a challenge one can hope will be taken up in the next few years. In parallel, determination of the sequence of adenylyl cyclase genes from the many organisms that are sensitive to cAMP, should permit more convincing phylogenetic comparisons (119,121,127,132,221,305).

CYCLASES AS A PARADIGM FOR THE EVOLUTION OF MODULAR PROTEIN CONSTRUCTION

The appearance of multicellular organisms has created a new regulatory constraint, which requires cell-to-cell communication. Generally speaking, hormones are the mediators of such integrated pattern. How then are hormone-mediated interactions integrated into a behavioral pattern that can accommodate the coordinated functioning of every single cell into appropriate

grouping remains unclear. It is generally accepted that such integration proceeds through a cascade of events within each cell. Hormones are recognized by specific receptors that trigger, through appropriate relay proteins, enzymes that synthesize so-called "second" messengers, which, in turn, activate a cascade of events generating a specialized behavior of the cell (metabolic regulation and/or specific gene expression). cAMP or cGMP are such second messengers, and adenylyl cyclases or guanylyl cyclases are enzymes whose activity is regulated by such relay proteins (the GTP-binding proteins, G-proteins, or transducins (reviewed in 81). Cyclases must, accordingly, have regions that permit regulation by specific G-proteins, and/or by other modulator proteins. This explains the wide variety of non-catalytic domains found in the proteins whose genes have thus far been characterized.

In all cases it has been found that intracellular levels of cAMP and cGMP are determined by extracellular signaling molecules. The coupling between signal and activity is generally transmitted through quaternary interactions, which are sometimes relayed by the single polypeptide chain containing the catalytic center. It has been shown that in many cases GTP-binding proteins act as relays for such information. The role for such proteins and other organisms remains unclear, however. For instance, Engleberg and coworkers (1990) have shown that the *S. pombe* adenylyl cyclase is not regulated by guanyl nucleotides (61).

Class I Cyclases (Enterobacteria and Related Bacteria)

The class of enterobacterial adenylyl cyclases is made of proteins comprising two functional domains. The amino-terminal moiety is responsible for catalytic activity, whereas the carboxy-terminal end is responsible for its control. We have already discussed the evolution of the catalytic domain. Most general conclusions that have been drawn also hold for the regulatory domains. The separation between the catalytic and regulatory domain is not well-defined, but one could suggest that it lies at a proline/alanine-rich region similar to the PAPAPKA consensus described by Erni (63). Analysis of the regulation of adenylyl cyclase activity suggests that *in vivo*, the activity can vary over a wide range. Indeed, it is known that cells deficient in the receptor for cAMP, CAP, overproduce the nucleotide by a factor of 30- to 100-fold, and that this cannot be accounted for by an increase in adenylyl cyclase *synthesis* (see 299 for a review, and 53 for an early tentative interpretation). This indicates that the carboxy-terminal end acts, together with other factors, as a *tonic* inhibitor of activity, which under appropriate conditions, displays increased synthetic capacity. As we shall see this property is not linked only to class I enzymes.

It is well established that the phosphoenolpyruvate-dependent phosphotransferase system enzymeIII^{Glc} is able to regulate cyclase (72). Only the

phosphorylated form of the enzyme can act as a stimulating factor, the unphosphorylated form being inactive (231). The actual process of activation is unknown. However in complementation experiments it has been shown that adenylyl cyclase from heterologous organisms such as *P. multocida*, is still activated by the *E. coli* enzymeIII^{Glc} (200). Because the divergence between *P. multocida* and *E. coli* is large (see Fig. 5), this suggests that the interaction involves a recognition mechanism independent of any complex quaternary interactions. Another likely explanation would be that phosphorylated enzymeIII^{Glc} phosphorylates a residue of the regulatory domain of adenylyl cyclase, thereby relieving inhibition. The enzymeIII^{Glc} is a member of the phosphorylation cascade which involves phosphohistidine residues. It is therefore tempting to assume that a histidine residue is involved: comparison of the different adenylyl cyclases of this class identifies two conserved histidine residues, and one of them appears to be in an environment that bears some resemblance with the residues that are phosphorylated by enzymeIII^{Glc} (141) (Fig. 10). Other residues such as cysteine [as in enzymeII^{Mtl} (240)] or aspartate [such as in the chemotactic response cascade (161)] could also be involved, but these cannot be clearly identified by a sequence comparison. However, analysis of phosphorylated proteins from *E. coli* failed to reveal a protein of 95 kDa as expected for adenylyl cyclase (*unpublished data, 1990*). The question of the enzymeIII^{Glc}-mediated phosphorylation of adenylyl cyclase activity remains therefore an open question. However, the overall cell content in adenylyl cyclase is very low (~400 molecules per cell), and the presumed phosphorylated form of the enzyme might be unstable. Experi-

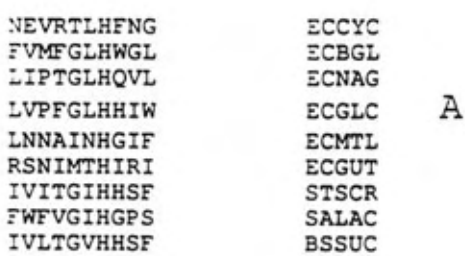


FIG. 10. The putative phosphorylated histidine residue in class I adenylyl cyclases. The sequences that are presented for comparison carry a residue that has been identified as phosphorylated by biochemical analysis **(A)**. A further comparison between the system responsible for lactose transport in *Staphylococcus aureus*, and a portion of the regulatory region of the *E. coli* cyclase **(B)** shows further similarities including conservation of the two histidine residues.

ments with class III adenylyl cyclases indicate that they are phosphorylated, suggesting that such mode of control indeed exists, at least in another situation (307).

Several other factors have been proposed to interact with adenylyl cyclase and to stimulate its activity. A protein somewhat similar to RAS (activating *S. cerevisiae* cyclase, see below) exists in *E. coli*, but its relationship to adenylyl cyclase has not been established (1). Trying to demonstrate that as in the case of most eucaryotes, enterobacterial cyclase could be regulated by G-proteins, Reddy and coworkers have shown that elongation factor Tu, a GTP-binding protein essential for protein synthesis, could serve as an activator of *E. coli* adenylyl cyclase (237). These experiments utilized a partially purified (10% pure) preparation of adenylyl cyclase, and should therefore be considered with some caution. The same group showed that all three cytoplasmic proteins of the phosphoenolpyruvate-dependent phosphotransferase system were required for full activation, and that they participated in a phosphate-mediated control of cyclase activity (222). This is consistent with the genetic observation of Lévy et al. (1990) who have shown by generating progressive deletions of the cytoplasmic phosphoenolpyruvate-dependent phosphotransferase system genes, that enzymeIII^{Glc} alone could not account for all of the phosphoenolpyruvate-dependent phosphotransferase system-mediated regulation of adenylyl cyclase activity (169). Finally, the stimulatory effect revealed by the absence of the cAMP receptor protein CAP is mediated by enzymeIII^{Glc} (36), leading one to think that another factor exists, which is controlled by the CAP-cAMP complex, and that this regulates adenylyl cyclase activity. Several observations substantiate the idea that, one or more other factors are involved, perhaps acting on the carboxy-terminal domain of the protein. Firstly, Gershanovitch and coworkers have shown that in strains deficient in some of the fructose operon genes, adenylyl cyclase activity was altered (83); secondly it was shown that catabolite repression by 2-ketoglutarate, a carbon substrate that does not enter the cell through a phosphoenolpyruvate-dependent phosphotransferase system, involves at least in part cAMP synthesis (43). The expression of other enzymeIII-like domains did not stimulate cAMP synthesis in the absence of enzymeIII (50). This suggests that FPr can replace enzymeIII^{Glc} under certain circumstances (and this would still be compatible with phosphorylation of adenylyl cyclase) and/or acts by interacting with another (phosphorylating ?) factor that binds to adenylyl cyclase.

The meaning of such a complex regulatory pattern will be addressed in the last part of this chapter. However, it should be pointed out that the phosphoenolpyruvate-dependent phosphotransferase system-mediated control of cAMP synthesis faces paradoxical constraints, since enterobacteria are facultative anaerobes. Indeed, under anaerobic growth conditions, the amount of carbon supply required for sustained growth of the cells is five to ten times higher than under aerobic growth conditions (78, see also 76). In