Volume 12 Number 24 1984 Nucleic Acids Research

The complete nucleotide sequence of the adenylate cyclase gene of *Escherichia coli*

Hiroji Aiba[1*], Kazuyasu Mori[1], Minoru Tanaka[1], Tatsuo Ooi[2], Anne Roy[3] and Antoine Danchin[3]

[1]Radioisotope Lab., Faculty of Medicine, Kyoto Univ., Kyoto 606, [2]Dept. Enzyme Chemistry, Inst. Chemical Research, Kyoto Univ., Uji, Kyoto 611, Japan, and [3]Dépt. Biochimie et Génétique Moléculaire, Inst. Pasteur, 28 rue du Docteur Roux, 75724, Paris Cédex 15, France

Received 19 November 1984; Accepted 23 November 1984

ABSTRACT

The complete nucleotide sequence of the cya gene from E. coli was determined. The gene encodes a polypeptide consisting of 848 amino acid residues with a calculated molecular weight of 97,542. The deduced protein structure reveals that cyclase is comprised of two domains, an amino-terminal region exhibiting catalytic activity and a carboxy-terminal region possibly carrying regulatory function. The frequent appearance of rare codons in the beginning of the gene as well as the sequence duplication in the promoter-initiator region suggest possible regulation(s) at the translational level. An unknown gene (cyaX) which seems to code for a very hydrophobic protein was found following the cya gene. Sequence analysis suggests that the cyaX is a part of the cya operon.

INTRODUCTION

While it is well recognized that cAMP is a central regulatory entity in prokaryotes as well as in eukaryotes, the mechanism by which the cellular level of cAMP is controlled is still poorly understood. However, knowledge concerning activity and expression of bacterial adenylate cyclase has been accumulated. Cyclase is sensitive to the nature of the carbon source on which the cell is feeding. The regulation involves, in a still not completely understood manner, a phosphorylation cascade known as the phosphoenolpyruvate dependent carbohydrate phosphorylation system (1,2). In this cascade a phosphorylated protein (enzyme III Glc) controls in a positive way adenylate cyclase activity (3). With respect to enzyme structure, it has been observed that adenylate cyclase is a large protein (ca 95 kb) and may comprise two domains, namely the amino-terminal domain displays cAMP synthesizing activity whereas the carboxy-terminus seems to be involved in regulation of cyclase activity by the

carbon source (4,5).

On the other hand, little is known about the regulation involved at the level of cyclase expression, although recent studies of the cloned cya gene unraveled some of the complexities of the cya regulatory region (6-8). There exist contradictory data concerning the regulation of cya expression by cAMP and its receptor protein (CRP or CAP). Some experiments suggest the involvment of cAMP-CRP as a negative effector for cya transcription (8-10) while other data refute this possibility (6,11,12).

To gain a better understanding of the structure and function of the adenylate cyclase as well as its expression, we have determined the nucleotide sequence of the entire cya gene of E. coli. The data suggest that the enzyme is made of two domains, and the gene is organized in a cya operon. Some regulatory aspects deduced from the DNA sequence are also discussed.

MATERIALS AND METHODS

Materials

Plasmid pCA2 and pCA3 carrying the cya gene (7) were used for DNA sequencing. Restriction endonucleases and T4 polynucleotide kinase were obtained from Takara Shuzo. [γ-^{32}P] ATP ($\sim$7000 Ci/m mol) was purchased from New England Nuclear.

Preparation of DNA and DNA sequence determination

Plasmid DNA was purified essentially as described by Birnboim and Doly (13). Restriction fragments were isolated from polyacrylamide gels and labeled at their 5' ends with [^{32}P-γ-ATP] by T4 polynucleotide kinase. Singly end-labeled DNA fragments were prepared either by a secondary restriction cleavage or by electrophoretic strand separation. Nucleotide sequences were determined by the method of Maxam and Gilbert (14).

RESULTS

Nucleotide sequence of the cya gene

The restriction map of the DNA region for the cya gene is given in Figure 1. The sequence was obtained on both strands for each nucleotide by several independent experiments. Each

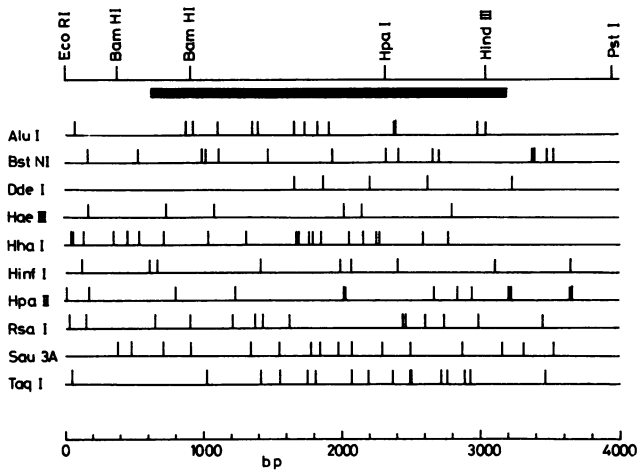

Figure 1. Restriction map of the cya gene and neighboring regions. The black bar represents the coding region for adenylate cyclase. Direction of transcription is from left to right.

site used for sequencing was also sequenced from a distinct site, at least on one strand.

Figure 2 shows the complete nucleotide sequence of the cya gene along with its 5' and 3' flanking regions. Comments on the cya promoter region are described elsewhere (6-8). It has been previously shown that the cya translation starts with TTG codon at position 652 (5). Starting with the TTG, a long open reading frame followed by two TGA termination codons at position 3196 is identified. This reading frame consists of 2544 nucleotides which is sufficient to specify adenylate cyclase protein. A second open reading frame preceded by a conserved Shine and Dalgarno sequence and fitting the Stormo rule (15) is also visible in the region after the cya gene.

Following the two TGA stop codons, a palindromic unit, TGCCGGAAA---TATCCGGCA, is identified between two open reading frames. This type of sequence, which was originally found in the intercistronic region of some operons (16,17), appears to be widely present outside structural genes on the E. coli chromosome and is believed to play roles in regulating expression of the operon as well as in genome evolution (18).

Adenylate cyclase

The primary structure of adenylate cyclase protein deduced from the nucleotide sequence is also shown in Figure 2. The protein consists of 848 amino acids and its molecular weight is calculated to be 97,542, which is in good agreement with those of the purified cyclase (19) or the protein specified by the cloned cya gene (20). The amino acid composition of adenylate cyclase given in Table 1 indicates that the protein is slightly acidic (net charge of -18). This is consistent with the result of the isoelectric focusing gel (pI=6.1) of the purified adenylate cyclase (19). The secondary structure of the protein predicted by the method of Robson and Suzuki (21) is shown in Figure 3. The contents of α-helix and β-sheet structures were predicted to be 30.0 % and 19.3 %, respectively. The polypeptide can be roughly divided into two distinct regions based on the primary and secondary structures. The first domain (residues 1-395) is more acidic (net charge of -17) and shows a high content of α-helix (42.5 %). Especially the region from 240 to 395 reveals an extremely high content of α-helix (65 %). This region is preceded by the alternative stretches of α-helix and β-sheet that are believed to participate in the binding of adenine nucleotide (22). The second region (residues 396-848) is electrically neutral (net charge of -1) and is less organized, showing low content (19.0 %) of α-helix. We believe that the first domain carries adenylate cyclase catalytic activity, since carboxy-terminal half can be deleted without a significant loss of cyclase activity (4). The active domain of the protein has been precisely defined by several in vitro deletions of the gene as well as Bal 31 nuclease digestion (Danchin, unpublished results) and it has been found that deletions extending upstream from the Bcl I site located at position 1781 (amino acid 377) rapidly destroy adenylate cyclase activity. With respect to the predicted secondary structure, this corresponds to destruction of the last α-helix present in the first domain (see Figure 3).

Since adenylate cyclase has been reported to be membrane associated (10,23,24), we have determined the hydropathy profile of cyclase protein by the method of Kyte and Doolittle (25) to

```
                                                                                                                ATTCAACCGGCACATCTTTCATTGAGTGTACGGCGATATCGGCGCGATTTT    51

CGAGGAGCGCGACTTCCAGCTCTTTTACAAATAAGCCTTTTCCGCCTACTTTCGCCAGCGGCGTATCAAGAATCACATCGCCGCGCGTCACCATCGGTACCAGTTCAACGACCAGGCCCG   171

GATGGCTCGCCATCAACTTGTCTTTGACATAGTGTGCCTGCCAGAGTGCAAGTGGGCTTTGGCGTGTGGCAATTCTTAAAACATTGTCTAACATGCTTGTTACCGTCATTATCATCCGTG   291

GTCCATCCTAACATCCTTGCCAGAGTGATGTCAGTGTTGTGGTGAAACGTAGACGCCTGCGCAAACCGTAAAATGAGGTCTGGCAGTGGATCCTGACAGGCGTTTCACGCCGTTGTAATA   411

AGGAATTTACAGAGAATAAACGGTGCTACACTTGTATGTAGCGCATCTTTCTTTACGGTCAATCAGCAAGGTGTTAAATTGATCACGTTTTAGACCATTTTTTCGTCGTGAAACTAAAAA   531

AACCAGGCGCGAAAAGTGGTAACGGTTACCTTTGACATACGAAATATCCCGAATGCCGCGTGTTACCGTTGATGTTGGCGGAATCACAGTCATGACGGGTAGCAAATCAGGCGATACGTC   651

TTG TAC CTC TAT ATT GAG ACT CTG AAA CAC AGA CTG GAT GCC ATA AAT CAA TTG CGT GTG GAT CGC GCG CTT GCT GCT ATG GGG CCT GCA   741
Met Tyr Leu Tyr Ile Glu Thr Leu Lys Gln Arg Leu Asp Ala Ile Asn Gln Leu Arg Val Asp Arg Ala Leu Ala Ala Met Gly Pro Ala    30

TTC CAA CAG GTC TAC AGT CTA CTG CCG ACA TTG TTG CAC TAT CAC CAT CCG CTA ATG CCG GGT TAC CTT GAT GGT AAC GTT CCC AAA GGC   831
Phe Gln Gln Val Tyr Ser Leu Leu Pro Thr Leu Leu His Tyr His His Pro Leu Met Pro Gly Tyr Leu Asp Gly Asn Val Pro Lys Gly    60

ATT TGC CTT TAC ACG CCT GAT GAA ACT CAA CGC CAC TAC CTG AAC GAG CTT GAA CTG TAT CGT GGA ATG TCA GTA CAG GAT CCC CCG AAA   921
Ile Cys Leu Tyr Thr Pro Asp Glu Thr Gln Arg His Tyr Leu Asn Glu Leu Glu Leu Tyr Arg Gly Met Ser Val Gln Asp Pro Pro Lys    90

GGT GAG CTT CCA ATT ACT GGT GTA TAC ACC ATG GGC AGC ACC TCG TCC GTA GGG CAA AGT TGT TCC TCT GAC CTG GAT ATC TGG GTC TGT  1011
Gly Glu Leu Pro Ile Thr Gly Val Tyr Thr Met Gly Ser Thr Ser Ser Val Gly Gln Ser Cys Ser Ser Asp Leu Asp Ile Trp Val Cys   120

CAT CAA TCC TGG CTC GAT AGC GAA GAG CGC CAA TTG CTA CAA CGT AAA TGT AGC CTG CTG GAA AAC TGG GCC GCC TCG CTG GGT GTG GAA  1101
His Gln Ser Trp Leu Asp Ser Glu Glu Arg Gln Leu Leu Gln Arg Lys Cys Ser Leu Leu Glu Asn Trp Ala Ala Ser Leu Gly Val Glu   150

GTC AGC TTC TTC CTG ATT GAT GAA AAC CGC TTC CGT CAT AAT GAA AGC GGC AGC CTG GGG GGC GAA GAT TGT GGC TCC ACC CAG CAT ATA  1191
Val Ser Phe Phe Leu Ile Asp Glu Asn Arg Phe Arg His Asn Glu Ser Gly Ser Leu Gly Gly Glu Asp Cys Gly Ser Thr Gln His Ile   180

CTG CTG CTT GAC GAA TTT TAT CGT ACC GCC GTG CGT CTC GCC GGT AAG CGT ATT CTG TGG AAT ATG GTG CCG TGC GAC GAA GAA GAG CAT  1281
Leu Leu Leu Asp Glu Phe Tyr Arg Thr Ala Val Arg Leu Ala Gly Lys Arg Ile Leu Trp Asn Met Val Pro Cys Asp Glu Glu Glu His   210

TAC GAC GAC TAT GTG ATG ACG CTT TAC GCG CAG GGC GGT CTG ACG CCA AAT GAA TGG CTG GAT CTC GGT GGC TTA AGC TCG CTT TCT GCT  1371
Tyr Asp Asp Tyr Val Met Thr Leu Tyr Ala Gln Gly Gly Leu Thr Pro Asn Glu Trp Leu Asp Leu Gly Gly Leu Ser Ser Leu Ser Ala   240

GAA GAG TAC TTT GGT GCC AGC CTT TGG CAG CTC TAC AAG AGT ATC GAT TCC CCA TAC AAA GCG GTA CTG AAA ACA CTG CTG CTG GAA GCC  1461
Glu Glu Tyr Phe Gly Ala Ser Leu Trp Gln Leu Tyr Lys Ser Ile Asp Ser Pro Tyr Lys Ala Val Leu Lys Thr Leu Leu Leu Glu Ala   270

TAT TCC TGG GAA TAC CCG AAC CCA CGT CTG CTG GCG AAA GAT ATC AAA CAG CGT TTG CAC GAC GGC GAG ATT GTA TCG TTT GGT CTC GAT  1551
Tyr Ser Trp Glu Tyr Pro Asn Pro Arg Leu Leu Ala Lys Asp Ile Lys Gln Arg Leu His Asp Gly Glu Ile Val Ser Phe Gly Leu Asp   300

CCA TAC TGC ATG ATG CTG GAG CGT GTT ACT GAA TAC CTG ACG GCG ATT GAA GAT TTT ACC CGT CTG GAT TTA GTA CGT CGC TGC TTC TAT  1641
Pro Tyr Cys Met Met Leu Glu Arg Val Thr Glu Tyr Leu Thr Ala Ile Glu Asp Phe Thr Arg Leu Asp Leu Val Arg Arg Cys Phe Tyr   330
```

```
TTA AAA GTG TGC GAA AAG CTC AGC CGT GAA CGC GCC TGC GTA GGC TGG CGT CGC GCA GTG TTG AGC CAG TTA GTG AGC GAG TGG GGT TGG   1731
Leu Lys Val Cys Glu Lys Leu Ser Arg Glu Arg Ala Cys Val Gly Trp Arg Arg Ala Val Leu Ser Gln Leu Val Ser Glu Trp Gly Trp    360

GAC GAA GCT CGT CTG GCA ATG CTC GAT AAC CGC GCT AAC TGG AAG ATT GAT CAG GTG CGT GAG GCG CAC AAC GAG TTG CTC GAC GCG ATG   1821
Asp Glu Ala Arg Leu Ala Met Leu Asp Asn Arg Ala Asn Trp Lys Ile Asp Gln Val Arg Glu Ala His Asn Glu Leu Leu Asp Ala Met    390

ATG CAG AGC TAC CGT AAT CTG ATC CGC TTT GCG CGT CGC AAT AAC CTT AGC GTC TCC GCC AGT CCG CAG GAT ATC GGC GTG CTG ACG CGT   1911
Met Gln Ser Tyr Arg Asn Leu Ile Arg Phe Ala Arg Arg Asn Asn Leu Ser Val Ser Ala Ser Pro Gln Asp Ile Gly Val Leu Thr Arg    420

AAG CTG TAT GCC GCG TTT GAA GCA TTA CCA GGT AAA GTG ACG CTG GTA AAC CCG CAG ATT TCA CCC GAT CTC TCG GAA CCG AAT CTG ACC   2001
Lys Leu Tyr Ala Ala Phe Glu Ala Leu Pro Gly Lys Val Thr Leu Val Asn Pro Gln Ile Ser Pro Asp Leu Ser Glu Pro Asn Leu Thr    450

TTT ATT TAT GTG CCG CCG GGC CGG GCT AAC CGT TCA GGT TGG TAT CTG TAT AAC CGC GCG CCA AAT ATT GAG TCG ATC ATC AGC CAT CAG   2091
Phe Ile Tyr Val Pro Pro Gly Arg Ala Asn Arg Ser Gly Trp Tyr Leu Tyr Asn Arg Ala Pro Asn Ile Glu Ser Ile Ile Ser His Gln    480

CCG CTG GAA TAT AAC CGT TAC CTG AAT AAA CTG GTG GCG TGG GCA TGG TTT AAC GGC CTG CTG ACC TCG CGC ACC CGT TTG TAT ATT AAA   2181
Pro Leu Glu Tyr Asn Arg Tyr Leu Asn Lys Leu Val Ala Trp Ala Trp Phe Asn Gly Leu Leu Thr Ser Arg Thr Arg Ley Tyr Ile Lys    510

GGT AAC GGC ATT GTC GAT TTG CCT AAG TTG CAG GAG ATG GTC GCC GAC GTG TCG CAC CAT TTC CCG CTG CGC TTA CCT GCA CCG ACA CCG   2271
Gly Asn Gly Ile Val Asp Leu Pro Lys Leu Gln Glu Met Val Ala Asp Val Ser His His Phe Pro Leu Arg Leu Pro Ala Pro Thr Pro    540

AAG GCG CTC TAC AGC CCG TGT GAG ATC CGC CAT CTC GCG ATT ATC GTT AAC CTG GAA TAT GAC CCG ACA GCG GCG TTC CGC AAT CAG GTC   2361
Lys Ala Leu Tyr Ser Pro Cys Glu Ile Arg His Leu Ala Ile Ile Val Asn Leu Glu Tyr Asp Pro Thr Ala Ala Phe Arg Asn Gln Val    570

GTG CAT TTC GAT TTC CGT AAG CTG GAT GTC TTC AGC TTT GGC GAG AAT CAA AAT TGC CTG GTA GGT AGC GTT GAC CTG CTG TAC CGC AAC   2451
Val His Phe Asp Phe Arg Lys Leu Asp Val Phe Ser Phe Gly Glu Asn Gln Asn Cys Leu Val Gly Ser Val Asp Leu Leu Tyr Arg Asn    600

TCG TGG AAC GAA GTG CGT ACG CTG CAC TTC AAC GGC GAG CAA TCG ATG ATC GAA GCC CTG AAA ACT ATT CTC GGC AAA ATG CAT CAG GAC   2541
Ser Trp Asn Glu Val Arg Thr Leu His Phe Asn Gly Glu Gln Ser Met Ile Glu Ala Leu Lys Thr Ile Leu Gly Lys Met His Gln Asp    630

GCC GCA CCG CCA GAT AGC GTG GAA GTC TTC TGT TAT AGC CAG CAT CTG CGC GGC TTA ATT CGT ACT CGC GTG CAG CAA CTG GTT TCT GAG   2631
Ala Ala Pro Pro Asp Ser Val Glu Val Phe Cys Tyr Ser Gln His Leu Arg Gly Leu Ile Arg Thr Arg Val Gln Gln Leu Val Ser Glu    660

TGT ATT GAA TTG CGT CTT TCC AGC ACC CGC CAG GAA ACC GGG CGT TTC AAG GCG CTG CGC GTT TCT GGT CAA ACC TGG GGG TTG TTC TTC   2721
Cys Ile Glu Leu Arg Leu Ser Ser Thr Arg Gln Glu Thr Gly Arg Phe Lys Ala Leu Arg Val Ser Gly Gln Thr Trp Gly Leu Phe Phe    690

GAA CGC CTG AAT GTA TCG GTA CAG AAA CTG GAA AAC GCC ATC GAG TTT TAT GGC GCG ATT TCG CAT AAC AAA CTG CAC GGC CTG TCA GTG   2811
Glu Arg Leu Asn Val Ser Val Gln Lys Leu Glu Asn Ala Ile Glu Phe Tyr Gly Ala Ile Ser His Asn Lys Leu His Gly Leu Ser Val    720

CAG GTT GAA ACC AAT CAC GTC AAA TTA CCG GCG GTG GTG GAC GGC TTT GCC AGC GAA GGG ATC ATC CAG TTC TTT TTC GAA GAA ACG CAA   2901
Gln Val Glu Thr Asn His Val Lys Leu Pro Ala Val Val Asp Gly Phe Ala Ser Glu Gly Ile Ile Gln Phe Phe Phe Glu Glu Thr Gln    750

GAC GAG AAT GGC TTT AAT ATC TAC ATT CTC GAC GAA AGC AAC CGG GTT GAG GTA TAT CAC CAC TGC GAA GGC AGC AAA GAG GAG CTG GTA   2991
Asp Glu Asn Gly Phe Asn Ile Tyr Ile Leu Asp Glu Ser Asn Arg Val Glu Val Tyr His His Cys Glu Gly Ser Lys Glu Glu Leu Val    780
```

```
CGT GAC GTC AGT CGC TTC TAC TCG TCA TCG CAT GAC CGC TTT ACC TAC GGC TCA AGC TTC ATC AAC TTC AAC CTG CCG CAG TTC TAT CAG     3081
Arg Asp Val Ser Arg Phe Tyr Ser Ser Ser His Asp Arg Phe Thr Tyr Gly Ser Ser Phe Ile Asn Phe Asn Leu Pro Gln Phe Tyr Gln      810

ATT GTG AAG GTT GAT GGT CGT GAA CAG GTG ATT CCG TTC CGC ACA AAA TCT ATC GGT AAC ATG CCG CCT GCC AAT CAG GAT CAC GAT ACG     3171
Ile Val Lys Val Asp Gly Arg Glu Gln Val Ile Pro Phe Arg Thr Lys Ser Ile Gly Asn Met Pro Pro Ala Asn Gln Asp His Asp Thr      840

CCG CTA TTA CAG CAA TAT TTT TCG TGA TGA ACGTGCCGGAAAGCGAGGCTTATCCGGCATGCAATCTTAGCGGAAACTGACTGTTTCACCCGCCTGCTGCGTCGCCGCCT     3281
Pro Leu Leu Gln Gln Tyr Phe Ser                                                                                            848

GTTCCAGCAAATCCCAGAAGGTTTCGCCGCTGCCATCACAAATCCACTCATCGCCCTTTCAGGTCAAA ATG GTA GCC GCC CTG TTT GGT TGC CAG CCA TAC CTG GTG     3387
                                                                      Met Val Ala Ala Leu Phe Gly Cys Gln Pro Tyr Leu Val       13

CAG CGG CTC CTG GCG GTT GAT AAT GAT TTT GCT GCC ATT CTC AAA GGT AAT GGT CAG TAC GCC GCC GTT GAT TTC GCA GTC GAT ATC GCT     3477
Gln Arg Leu Leu Ala Val Asp Asn Asp Phe Ala Ala Ile Leu Lys Gly Asn Gly Gln Tyr Ala Ala Val Asp Phe Ala Val Asp Ile Ala       43

GTC GCC ATC CCA GTC GTC CAG GCG TTC TTC AAT GGT CAG CCA GAG TTG ATC AGC CAG GCG ATG AAA TTC ACT GTC GTT CAT TGT TGT ATC     3567
Val Ala Ile Pro Val Val Gln Ala Phe Phe Asn Gly Gln Pro Glu Leu Ile Ser Gln Ala Met Lys Phe Thr Val Val His Cys Cys Ile       73

CTG TTT TTA AGT GAT GGC GGC AGT ATA GCG GCA TGG GGT CAG GGC TTC AAA GTT TGC ACC TCT GCG GCT GCG TTC CGG CAC GAT TCA TCC     3657
Leu Phe Leu Ser Asp Gly Gly Ser Ile Ala Ala Trp Gly Gln Gly Phe Lys Val Cys Thr Ser Ala Ala Ala Phe Arg His Asp Ser Ser      103

GTC ACC GGA ATA ATG ATG TCT CTG TGT AGC GAA AGA TTT G                                                                       3697
Val Thr Gly Ile Met Met Ser Leu Cys Ser Glu Arg Phe                                                                         116
```

Figure 2. DNA sequence of the region for the cya gene and the predicted polypeptide sequences. The DNA sequence of the noncoding strand is shown. Nucleotide numbering starts with the Eco RI site. The vertical arrow at position 498 indicates the initiation site of the major cya RNA. The first open reading frame from positions 652 to 3195 represents adenylate cyclase. The second one, which corresponds to an unknown gene (cyaX), start at position 3349. The putative Shine-Dalgarno sequences for both genes are double-underlined. A palindromic sequence in the region between two genes is shown by horizontal arrows. Clusters of the penta nucleotide, GCCGC, are underlined.

Table I Amino acid compositions of *cya* and *cyaX* proteins

Amino acid	Adenylate Cyclase						CyaX	
	residues 1-848		residues 1-395		residues 396-848		residues 1-116	
Nonpolar	379	(44.7%)	172	(32.4%)	207	(45.7%)	65	(56.4%)
Ala	47		23		24		17	
Val	56		21		35		12	
Leu	104		58		46		9	
Ile	39		13		26		7	
Pro	39		15		24		3	
Met	16		12		4		4	
Phe	37		9		28		10	
Trp	16		11		5		1	
His	25		10		15		2	
Polar	281	(33.1%)	128	(32.4%)	153	(33.8%)	37	(31.9%)
Gly	48		24		24		9	
Ser	63		28		35		8	
Thr	33		15		18		3	
Cys	15		10		5		5	
Tyr	40		22		18		2	
Asn	42		12		30		3	
Gln	40		17		23		7	
Negatively charged	103	(12.1%)	56	(14.2%)	47	(10.4%)	8	(6.9%)
Asp	43		25		18		6	
Glu	60		31		29		2	
Positively charged	85	(10.0%)	39	(9.9%)	46	(10.2%)	6	(5.2%)
Lys	29		13		16		3	
Arg	56		26		30		3	
Total	848		395		453		116	

see whether the membrane-spanning segments are present or not. Although the profile also suggests the existence of the two separable domain, it does not reveal significant features of the membrane-bound protein (data not shown).

Codon usage

Table II summarizes the overall codon usage of the entire *cya* gene. It should be noted that rare codons are used frequently throughout the coding region; the frequency of the use of optimal codons (Fop), which was defined by Ikemura (26), is calculated to be 0.66 in the *cya* gene. This value is

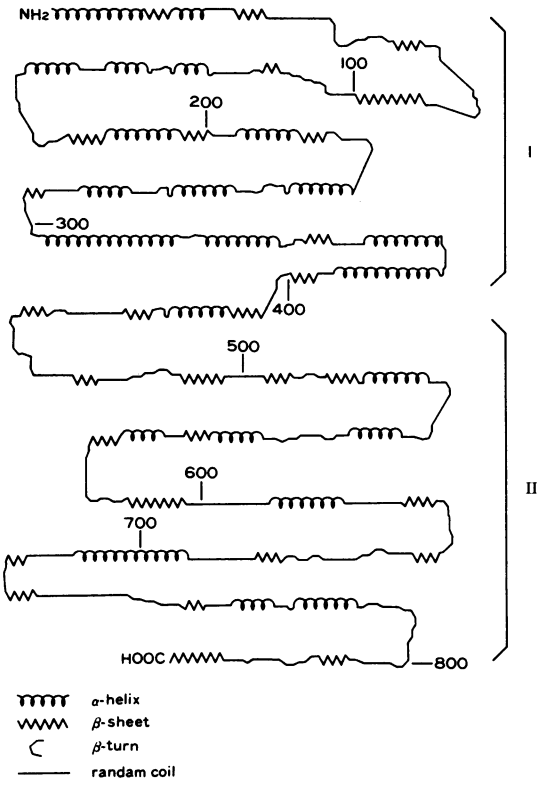

Figure 3. Predicted secondary structure of adenylate cyclase. Two domains are shown by brackets.

comparable to those of trp genes, suggesting that the cya gene is expressed at a relatively low level. It is also noteworthy that rare codons are clustered in the very beginning of the protein, namely 11 out of the first 20 codons are nonoptimal codons.

Sequence comparison

The cya sequence reveals several more or less extended duplications. The most interesting one corresponds to the sequence immediately upstream from the initiation TTG and a region upstream of the major promoter (Figure 4). This duplication, along with an unusual initiation codon TTG, might carry a specific information for regulating the level of translation.

Table II Codon Usage in the *cya* Gene

Leu	UUA	9	Ser	UCU	5
	UUG	12		UCC	8
	CUU	11		UCA	6
	CUC	13		UCG	15
	CUA	4		AGU	5
	CUG*	55		AGC	24
Arg	CGU*	29	Thr	ACU*	6
	CGC*	24		ACC*	13
	CGA	0		ACA	5
	CGG	2		ACG	9
	AGA	1			
	AGG	0	Ile	AUU	21
				AUC*	16
Pro	CCU	5		AUA	2
	CCC	2			
	CCA	8	Asn	AAU	17
	CCG*	24		AAC*	25
Gln	CAA	13	Phe	UUU	15
	CAG*	27		UUC*	22
Lys	AAA*	19	Tyr	UAU	19
	AAG	10		UAC*	21
Ala	GCU*	6	Glu	GAA*	38
	GCC	16		GAG	22
	GCA*	7			
	GCG*	18	Cys	UGU	7
				UGC	8
Val	GUU*	9			
	GUC	10	His	CAU	13
	GUA*	13		CAC	12
	GUG*	24			
			Asp	GAU	26
Gly	GGU*	18		GAC	17
	GGC*	23			
	GGA	1	Trp	UGG	16
	GGG	6			
			Met	AUG	15
				UUG	1

Optimal codons (26) are shown by asterisks(*).

Another feature in the *cya* sequence is the existence of clusters of the penta nucleotide GCCGC, namely 7 such sequence can be found at the end of the gene as well as a cluster of three sequences in the middle of the gene. This prompted us to look for similar clusters in other genes. The most surprising data came from comparison with pBR322 (27). We found extensive homology between the interdomain sequence of *cya* and

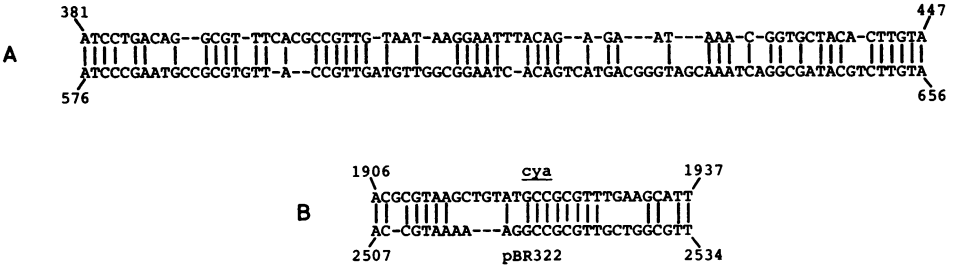

Figure 4. DNA sequence comparison. A, a sequence duplication in the region upstream of the cya structural gene. B, a sequence homology between the interdomain region of cya and the replication origin of pBR322. The nucleotides in the cya gene are numbered as in Figure 2. The DNA sequence of pBR322 is taken from the reference (27).

the replication origin of pBR322 (Figure 4). Whether this is of functional significance remains to be seen.

The second open reading frame: a cya operon?

Downstream from the cya reading frame and after the Clément and Hofnung's palindromic unit one can see an open reading frame of more than 116 amino acids. This amino acid sequence appears to be extremely hydrophobic, in a way which is characteristic of membrane-bound proteins. No visible promoter sequence can be identified in the intercistronic region, suggesting that the second reading frame is part of an operon. We therefore tentatively term this gene cyaX.

DISCUSSION

The DNA sequence presented here suggests some noteworthing features about the organization and expression of cya gene as well as the structure of adenylate cyclase. First, the cyclase protein appears to be composed of two distinct domains. This substantiates the physiological and biochemical data which have shown that cyclase catalytic activity is located in the amino-terminal region whereas the glucose sensitive regulatory site may be located in the carboxy-terminal region of the protein (4,5). The existence of a domain structure in cyclase suggests that the cya gene results from the fusion of two ancestral genes as, for instance, does the aspartokinase homoserine dehydrogenase gene (28).

With respect to the expression of the cya, the DNA duplication found around the cya regulatory region including the translational start site suggests a regulation at the level of cya translation. This is corroborated by the frequent appearance of rare codons in the very beggining of the gene, which may disturb the translation especially under exponential growth conditions. The previous finding that cyclase translation start with an unusual UUG codon also suggests a regulation at the level of translation. In addition, the finding that the interdomain sequence has a strong homology with pBR322 replication origin may indicate a coupling between transcription of the cya gene (at least its distal part) and a component of the replication machinery.

Finally, the sequence determination revealed that the cya gene is followed by an unknown gene (cyaX) which may specify a very hydrophobic protein, thus suggesting the existence of a cya operon. This is particularly interesting in view of the fact that the cyclase protein, as determined from the DNA sequence, is not very hydrophobic, whereas most studies suggest that the adenylate cyclase is membrane bound. On the other hand, an isolation of the protein as a cytoplasmic one, albeit with a low yield, indicates that adenylate cyclase is a soluble enzyme (19). These data may be reconciled under the hypothesis that the adenylate cyclase catalytic subunit is in fact bound loosely to an integral membrane protein. The cyaX gene located downstream from cya seems to specify a hydrophobic protein, and this might correspond to the cyclase membrane "adapter". The phenotype of a cyaX$^-$ strain is difficult to predict (mainly because cytoplasmic adenylate cyclase still synthesizes cAMP), which might explain why no corresponding mutations have yet been identified.

The nucleotide sequence of the cya region thus reveals a complex organization of the gene(s) and their products. Physiological and biochemical experiments, as well as comparison with cya genes from other organisms, are required to unravel the complexity of this system.

ACKNOWLEDGEMENTS

We are grateful to Drs. K. Izui and N. Fujita for useful discussions, and M. Tabata for her typing assistance. This work was supported by a Grant-in-Aid for Special Project Research from the Ministry of Education, Science and Culture of Japan.

*To whom reprint requests should be sent

REFERENCES

1. Botsford, J. L. (1981) Microbiol. Rev. 45, 620-642.
2. Ullmann, A. and Danchin, A. (1983) Advan. Cyclic Nucl. Res. 15, 1-53.
3. Postma, P. W. (1982) Ann. Microbiol. (Inst. Pasteur), 133A, 261-267.
4. Roy, A., Danchin, A., Joseph, E. and Ullmann, A. (1983) J. Mol. Biol. 165, 197-202.
5. Danchin, A., Guiso, N., Roy, A. and Ullmann, A. (1984) J. Mol. Biol. 175, 403-408.
6. Roy, A., Haziza, C. and Danchin, A. (1983) EMBO J. 2, 791-797.
7. Aiba, H., Kawamukai, M. and Ishihama, A. (1983) Nucl. Acids Res. 11, 3451-3465.
8. Aiba, H. J. Biol. Chem. in press.
9. Botsford, J. L. and Drexler, M. (1978) Mol. Gen. Genet. 165, 47-56.
10. Majerfeld, I. H., Miller, D., Spitz, E. and Rickenberg, H. V. (1981) Mol. Gen. Genet. 181, 470-475.
11. Bankaitis, V. A. and Bassford, P. J. Jr. (1982) J. Bacteriol. 151, 1346-1357.
12. Joseph, E., Bernsley, C., Guiso, N. and Ullmann, A. (1982) Mol. Gen. Genet. 185, 262-268.
13. Birnboim, H. C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.
14. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
15. Stormo, G. D., Schneider, T. D., Gold, L. and Ehrenfeucht, A. (1982) Nucl. Acids Res. 10, 2997-3011.
16. Clément, J. M. and Hofnung, M. (1981) Cell 27, 507-514.
17. Higgins, C. F., Ames, G. F., Barnes, W. M., Clément, J. M. and Hofnung, M. (1982) Nature 298, 760-762.
18. Gilson, E., Clément, J. M., Brutlag, D., and Hofnung, M. (1984) EMBO J. 3, 1417-1421.
19. Yang, J. K. and Epstein, W. (1983) J. Biol. Chem. 258, 3750-3758.
20. Roy, A. and Danchin, A. (1982) Mol. Gen. Genet. 188, 465-471.
21. Robson, B. and Suzuki, E. (1977) J. Mol. Biol. 107, 899-904.
22. Rossmann, M. G., Moras, D. and Olsen, K. W. (1974) Nature 250, 194-199.
23. Ide, M. (1971) Arch. Biochem. Biophys. 144, 262-268.

24. Janeček, J., Náprstek, J., Dobrová, Z., Jirešová, M. and Spižek, J. (1980) Folia Microbiol. 25, 361-368.
25. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.
26. Ikemura, T. (1981) J. Mol. Biol. 151, 389-409.
27. Sutcliffe, J. G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.
28. Zakin, M. M., Duchange, N., Ferrara, P. and Cohen, G. N. (1983) J. Biol. Chem. 258, 3028-3031.