Volume 10 Number 24 1982 Nucleic Acids Research

The regulatory region of the divergent *argECBH* operon in *Escherichia coli* K-12

Jacques Piette*, Raymond Cunin*, Anne Boyen*, Daniel Charlier*, Marjolaine Crabeel*, Françoise Van Vliet*, Nicolas Glansdorff*, Craig Squires+ and Catherine L.Squires+

*Microbiology, Vrije Universiteit Brussel, and Research Institute of the CERIA, 1, Ave. E. Gryson, B-1070 Brussels, Belgium, and +Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Received 24 September 1982; Revised and Accepted 24 November 1982

ABSTRACT

The nucleotide sequence of the control region of the divergent argECBH operon has been established in the wild type and in mutants affecting expression of these genes.
The argE and argCBH promoters face each other and overlap with an operator region containing two domains which may act as distinct repressor binding sites. A long leader sequence - not involved in attenuation - precedes argCBH. Overlapping of the argCBH promoter and the region involved in ribosome mobilization for argE translation explains the dual effect of some mutations. Mutations causing semi-constitutive expression of argE improve putative promoter sequences within argC. Implications of these results regarding control mechanisms in amino acid biosynthesis and their evolution are discussed.

INTRODUCTION

Divergently transcribed groups of functionally related genes are not exceptional in Escherichia coli (1-10). In only a few instances, however, are the sites responsible for the expression and the regulation of the flanking genes organized in an integrated fashion such that the gene cluster constitutes a bipolar operon, with an internal operator region flanked by promoters facing each other.

The argECBH cluster (Fig.1) is one of the earliest reported examples of such a pattern. It is transcribed from an internal region located between argE and argC (11,12) where cis-dominant regulatory mutations affecting the expression of both arms of the cluster have been localized (13,14). Certain argCB deletions, reaching into the control region, abolish the expression of argE without altering its coding part (14) ; the repressibility of argH is reduced by these deletions but its maximal level of expression is but little affected. The effects of such

0305-1048/82/1024-8031$2.00/0

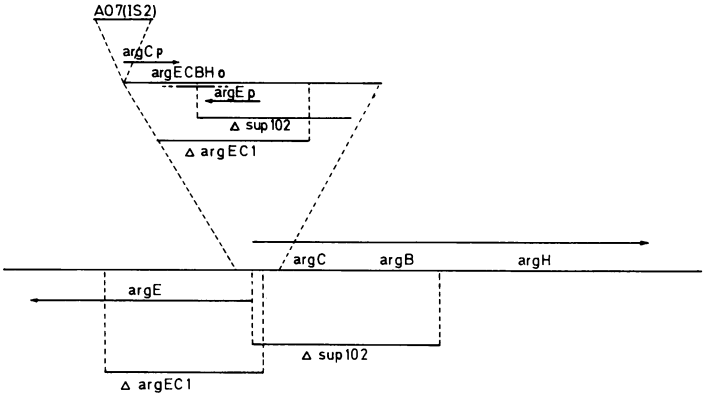

Figure 1. Genetic map of the argECBH cluster. The polarity of the argE and argCBH transcription is indicated by an arrow. Also indicated is the extent of ΔargEC1 and the Δsup102 deletions. Genetic symbols and abbreviations are as proposed by Bachmann et al. (52); Δ = deletion, bp = base pair, kb = kilobase. The control region was delimited genetically by the IS2 insertion present in mutant AO7 and by the righthant limit of ΔEC1 deletion (see text). The Δsup102 deletion abolishes the activity of argEp and affects also the repressibility of argH expression.

deletions (such as Δsup102, fig.1) were explained by assuming that the promoters lay on the far side of the control region with respect to their cognate structural genes (14). An analogous face-to-face positioning of promoters was subsequently proposed, then demonstrated in the biotin operon (bioABFCD) (5,16).

We present here the nucleotide sequence of the wild type argECBH regulatory region and of several mutations affecting gene control in the cluster. These data confirmed the face-to-face localization of the promoter sites on each side of the operator region and disclosed several unexpected features :

1. a long leader sequence was found to precede the putative translational start of argC ; it does not appear to be involved in control by attenuation.
2. mutations improving argE translation (17) were found to be located between the 5'-end of the argE message and the cognate Shine-Dalgarno (18) sequence, thus in a region not involved in base-pairing interactions with 16S RNA.
3. cis-dominant mutations located in argC can dramatically influence argE transcription.

MATERIALS AND METHODS

Strains. Derivatives of Hfr P4X (metB, str^S)

P4XB2 ; argR derivative of P4X.

P4XEC-1 : metB, str^S, ΔargEC-1.

P4Xsup102 : metB, str^S, Δ[argEp,argCB]

P4XAO-7 : metB, str^S, Δ[argEp,argCB], IS2:II.

L2, L9, L10, LL13 : Hydroxylamine-induced argECBH regulatory mutants (17).

Δ13-4 : argECBH regulatory mutant obtained from an argC-lacZ fusion strain (19).

MG545 : an argECBH regulatory mutant selected in E. coli B (13), then transduced into P4X.

Derivatives of strain 6G (F^-, his, ile, ppc) (20)

PO4: argR derivative of 6G

6-8 : hydroxylamine-induced argECBH regulatory mutant (20)

Phage λ13 : λdppc argECBH bfe (22) is a derivative of λ199 (CI857, susS7, xis6, Δb515, Δb519)

Plasmids : all plasmids are derivatives of pBR322 (21) with the exception of pMC7 which is derived from pMB9 (22).

DNA fragment purification was performed either by RPC-5 chromatography (23) or by agarose gel electrophoresis followed by hydroxylapatite chromatography (24).

End labelling of fragments. Fragments were labelled at 5'-ends with γ-^{32}P-ATP and T4 polynucleotide kinase (25) after treatment with bacterial alkaline phosphatase ; 3'-ends were labelled with cordycepin triphosphate (New England Nuclear) and terminal deoxynucleotidyl transferase (26). Labelled ends were separated by a secondary restriction or a strand separation of the fragments.

Nucleotide sequencing was performed using the chemical degradation method (27). The reaction products were electrophoresed on 8 %, 10 % or 20 % denaturing polyacrylamide gels.

Mapping of the 5'-ends of transcripts. The procedure followed is that of Barry et al. (28) with a few minor modifications. End labelled DNA fragments were used as probes in hybridizations (3 hrs, 50°C, in 80 % formamide) with total RNA from the appropriate strains. After hybridization, 5000 units of S1 endo-

nuclease (Boehringer) were added and digestion allowed to proceed for 30 min at 45° C. The protected hybrid fragments were submitted to electrophoresis on 6 % or 10 % sequencing gels, in parallel with the products of G & A and C & T sequencing reactions.

RESULTS

Cloning and sequencing strategy

The genetic map of the argECBH cluster is given in Fig.1. Three sets of clones harbouring part or all of the control region of argECBH have been constructed by introducing appropriate restriction fragments of the transducing bacteriophage λdarg13 (21) into the plasmid pBR322.

In λdarg13 and derivatives thereof carrying regulatory mutations, a 18.9 kb chromosomal segment from E. coli containing the ppc argECBH and bfe genes has been substituted for the phage head and tail genes. The argECBH control region has been localized by restriction analysis and heteroduplex DNA mapping (22). It has been isolated on a 1.8 kb HindIII DNA fragment, together with the operator proximal part of argE, argC and part of argB, and cloned into the HindIII site of pBR322, yielding plasmid pMC31 and its derivatives harbouring regulatory mutations (Fig.2a) (21).

The deletion mutant P4Xsup102 has been cloned into pBR322 on a 4.7 kb HindIII-EcoRI DNA fragment including part of argE on one side and argH together with more E. coli DNA on the other side. The resulting plasmid is pSU102 (Fig.2c).

The control region of a P4Xsup102 derivative harbouring an IS2 element (P4XA07) has been cloned on a fragment defined by the HindIII site in argE and, on the other side, by the unique HindIII site present in IS2 (29). The length of that fragment is 1.12 kb (21) (Fig.2b).

The DNA sequence of 900 bp, stretching from the HindIII site in argE to well within argC, has been established by the chemical degradation method of Maxam and Gilbert (27). An equivalent length of DNA has been sequenced for some but not all of the O^c and promoter mutants. As indicated in Fig.3, a different sequencing strategy was followed for P4Xsup102 and for P4XA07.

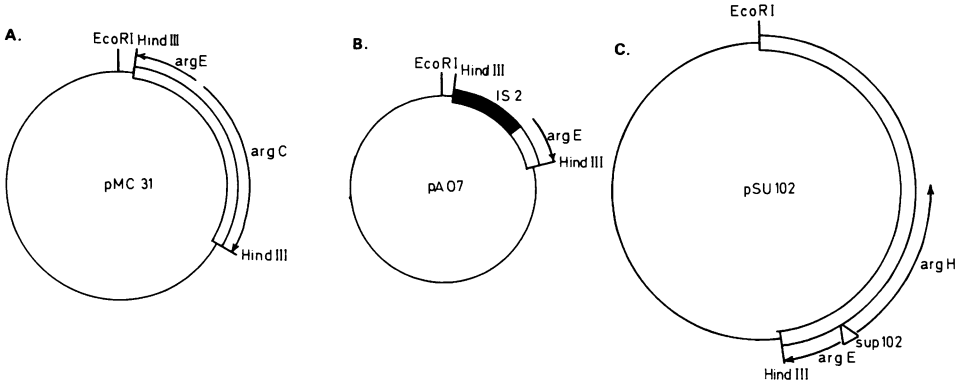

Figure 2. Plasmids used for sequencing the control region in the wild type and mutants of argECBH cluster.

a. pMC31 : a 1800 bp HindIII fragment was cloned from λdarg13 into the HindIII site of pBR322. The fragment contains the control region, argC, and part of argE. Analogous plasmids were constructed with fragments carrying the point mutations sequenced in this work.

b. pA07 : a 1200 bp HindIII fragment was cloned from λdargA07 into the HindIII site of pBR322. The fragment contains the control region, part of the argE gene and part of the IS2 element.

c. pSU102 : a 4700 bp EcoRI-HindIII fragment was cloned from λdarg13 Δsup102 in pBR322 using the EcoRI and HindIII sites of this vector. This fragment contains part of argE, argH and the adjacent parts of the E. coli chromosome.

Identification and organization of the argECBH control region.

Two mutants were used to delimit the regulatory region of the cluster : P4Xsup102 which harbours a deletion covering the promoter for argE (argEp), argC and argB, but leaves argCp unaffected, and a derivative of P4Xsup102, P4XA07, which carries an IS2 element abutting on argCp on its argE side (Table 1, ref. 17 and 21). Between the IS2 element of P4XA07 and the sup102 deletion lies a region of about 60 bp that exhibits two overlapping imperfect palindromes (Pal1 and Pal2, Fig.4) and promoter-like sequences. As discussed below, this region was indeed shown to contain repressor binding sites and the promoter for argCBH.

The translation startpoints of argE and argC could be ascribed tentatively to two AUG codons, 150 bp apart, one on each DNA strand, which set open reading frames of 97 and 145 codons,

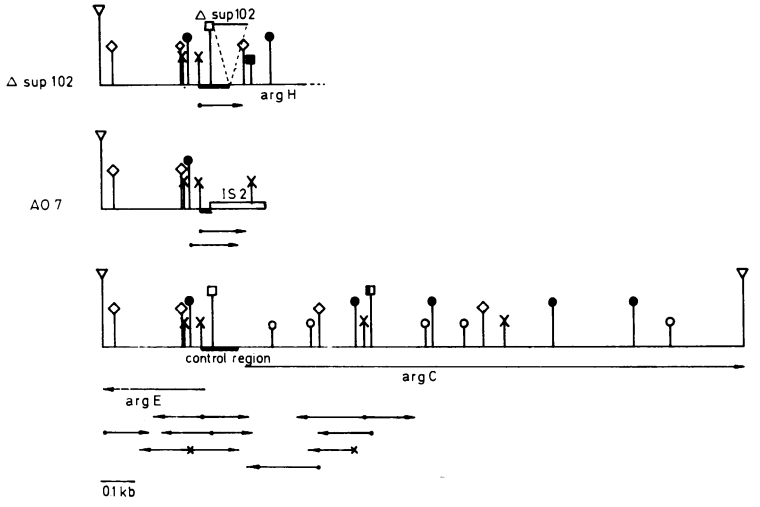

Figure 3. Restriction map of the cloned arg fragment and sequencing strategy. Symbols for restriction sites are as follows : ∇ = HindIII, □ = HincII, ■ = HinfI, ◧ = HpaI, ● = HeaII, x = Taq1, O = AluI, ◇ = SauA3. Sequenced regions are represented by arrows. 5'-labelled ends are indicated by dots, 3'-labelled ends by crosses.

respectively, as far as present sequencing data extend. They are preceded by Shine-Dalgarno sequences displaying extensive matches to the 3'OH extremity of 16S RNA (18). In the case of argC, the putative Shine-Dalgarno-AUG interval contains a TAG codon ; this would constitute one of the few exceptions to the striking absence of TAG codons generally observed at this location (30). Confirmation of the role of these sequences awaits the purification and sequencing of the argE and argC gene products, presently in progress. The location of promoter sites for argCBH and argE is established in the next sections.

Characterization of the argCBH promoter site

A HincII restriction site coïncides with the -35 RNA polymerase recognition region (TTGACA, 31, 32) of a putative argCBH promoter site (Fig.4). Preincubation with RNA polymerase of the purified 1800 bp HindIII fragment carrying the regulatory region prevented HincII cleavage at that site, though not at another HincII site within argC. This indicated the participation of the former site to a sequence involved in RNA polymerase binding.

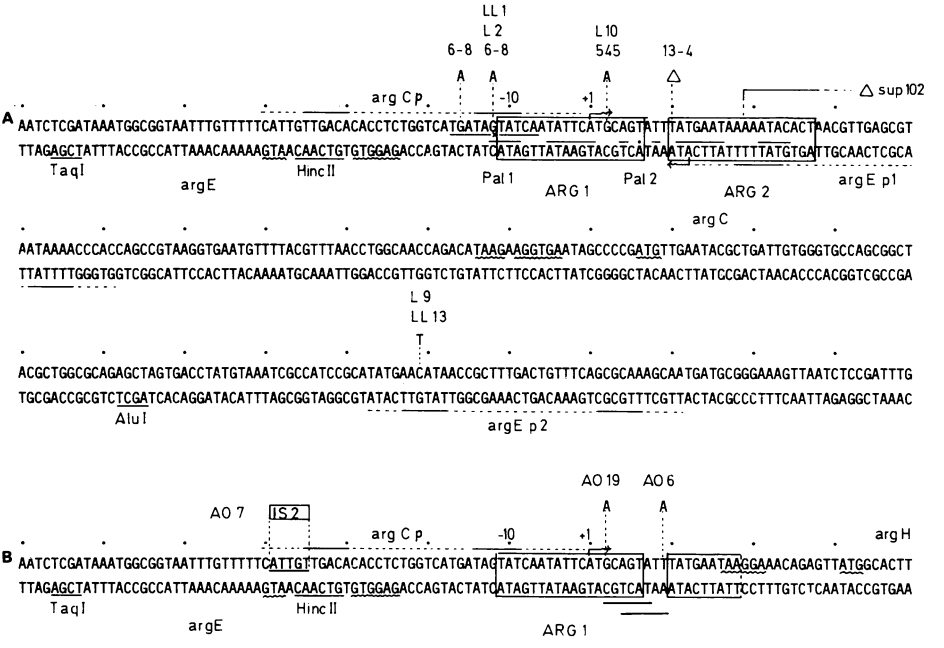

Figure 4. Nucleotide sequence of the argECBH control region.

a. Wild type : the -35 and -10 regions of the promoter for argCBH (argCp) and two putative promoter sites for argE (argEp1 and argEp2) are indicated by solid lines ; the direction of transcription is shown by an arrow.
The Pal1 and Pal2 palindromes are shown by lines centered respectively on a cross for Pal1 and a dot for Pal2.
The ARG boxes, ARG1 and ARG2 are indicated as such.
The putative ribosome binding sites and the translation initiation codons for argE and argC are indicated by wavy lines. The new base pairs introduced by the point mutations and the lefthand extremity of the sup102 deletion are shown above the sequence. Restriction sites for Taq1, HaeII and AluI are indicated below. See text for furter details.

b. The sup102 deletion and mutations reactivating argE in this context. Symbols are as in figure 4a. Putative ribosome binding sites and startpoints codons, followed by a free reading frame for argH are also indicated.
A postulated 5 bp duplication accompanying the IS2 insertion is underlined ; IS2 is in orientation II with respect to the direction of the transcription of argE.
Sequences modified into putative Pribnow boxes by mutations AO19 and AO6 are underlined.

This was supported by the results of footprinting experiments (33) showing that a stretch of DNA, ranging from positions -48 to +25 on Figure 4, was protected by RNA polymerase against

Table 1

Strain	Acetylornithine lyase Enzyme E (EC 3.5.1.16)		Argininosuccinate lyase Enzyme H (EC 4.3.2.1)	
Supplement to minimal medium	---	L-arginine (100 µg/ml)	---	L-arginine (100 µg/ml)
P4X	19.0	5.0	1.5	0.15
P4XB2 (P4XargR)	84.0	-	9.0	-
P4XΔsup102	-	<0.1	-	0.65
P4XΔsup102 argR	0.2[a]	-	5.1[a]	-
P4XA07	-	1.9	-	0.55
P4XA07 argR	2.8[a]	-	1.6[a]	-
P4XA06	-	1.2	-	0.85
P4XA06 argR	6.5[a]	-	4.1[a]	-
P4XA019	-	3.5	-	14.0
P4XA019 argR	5.8[a]	-	18.0[a]	-
Δ13-4	97.0	45.0	6.7	3.9
L10	104.0	14.5	-	0.75
L10 argR	116.0	-	4.4	-
MG545	-	23.0	-	0.80
MG545 argR	87.0	-	5.3	-
LL13	79.0	52.0	1.4	0.15
LL13 argR	130.0	-	6.2	-
L9	49.0	40.0	1.2	0.18
L9 argR	121.0	-	8.1	-
LL1	63.0	17.0	0.5	0.08
LL1 argR	141.0	-	3.2	-
L2	38.0	12.0	0.7	0.08
L2 argR	147.0	-	2.1	-
6G[b]	25.3	5.8	2.2	-
PO4[b](6G argR)	46.0	-	7.7	-
6-8[b,c]	149.0	12.0	0.7	0.10
PO4-8[b](6-8 argR)	143.0	-	1.7	-

Relevant enzyme specific activities, expressed in µM/hour x mg protein. Enzyme assays have been described previously (17,51). All mutants are derivatives of Hfr P4X with two exceptions : MG545, which was selected in Escherichia coli B (13), then transduced into P4X, and 6-8 which was selected in strain 6-G (20). argR strains are genetically derepressed. The data are from Boyen et al. (17) and Beny et al. (19)

a Minimal medium supplemented with L-ornithine 100 µg/ml.
b Enzyme determinations were performed in the genetic context of strain 6-G.
c Low argH expression in this mutant limits endogenous arginine production, leading to derepression of the whole arginine biosynthesis when 6-8 is grown on unsupplemented minimal medium .

digestion by DNAse I.

Conclusive evidence about the localization of argCp came from the DNA sequence of four mutants : three independently isolated argCp-down mutants (L2,LL1 and 6-8), in which the rate of transcription of argCBH is 2- to 4-fold less than in the wild type (Table 1, and ref.17), and an IS2 insertion mutant (P4XA07). All three argCp-down mutants show a GC → AT transition at position -12 in the Pribnow box (31,34) of the promoter locus indicated by the previous experiments. Mutant 6-8, which exhibits the strongest promoter-down phenotype of all three mutants, harbours a second GC → AT transition 4 bp upstream of the first one (position -16). There is little constraint on the third nucleotide in the Pribnow box and it is difficult to predict whether the substitution of a G by an A would have a negative effect. The only instance of a single base pair mutation at this position is the galp2 promoter, where a T → G change has a weak promoter-down effect (35).

The second G → A transition in mutant 6-8 (position -16 on the sense strand for argCBH) could be expected to slightly impair promoter function since it affects a region which is weakly conserved in various promoters (32).

The IS2 element present in strain P4XA07 is inserted in orientation I (36) with respect to argH in such a way that the first nucleotide of the IS element adjoins nucleotide -40 of argCp. In this strain the expression from argCp amounts to only 30 % of the normal rate. This result provides direct evidence that RNA polymerase makes specific contacts with promoter DNA at least as far upstream as that position. The lac mutant L305, which has suffered a single base deletion at position -37 and therefore shifts the -45 region out of register with the -35 domain, has also been interpreted in this way (32,37).

From the data presented in the previous paragraphs, it appears that argC transcription should be initiated a few nucleotides downstream from the Pribnow box affected by mutants L2, LL1 and 6-8, most probably within the consensus sequence CAT which is present in this region. The S1 nuclease mapping experiment reported on Fig.5a (lane D) indeed showed that the derepressed mutant P4XB2 synthesized a major RNA species whose 5'end maps

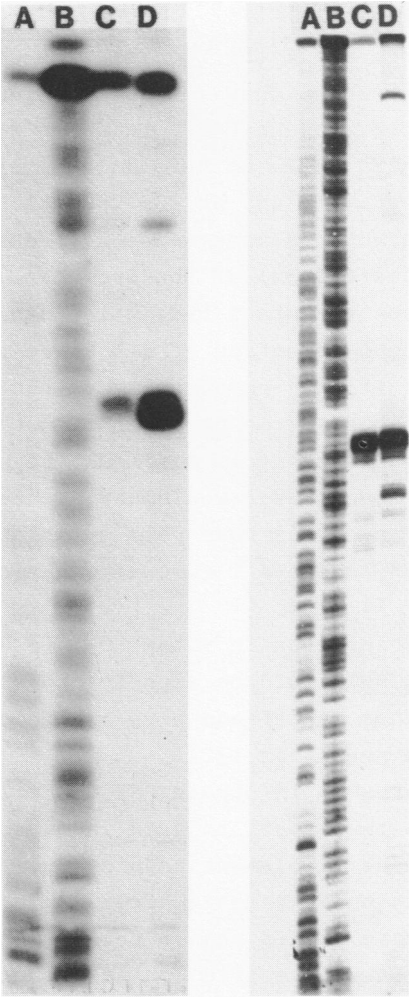

Figure 5. S1 endonuclease mapping of the argECBH transcripts.
a. S1 endonuclease mapping of the argCBH transcript.
A 390 bp SauA3-HaeII fragment containing the control region was 5' labelled at its SauA3 end and was hybridized with RNA from a strain containing the argECBH plasmid pMC7 (22) grown on minimal medium without arginine (lane C) and from P4XB2 (argR) (17) grown on rich medium (lane D).
The protected fragments were loaded in parallel with the products of the G and A (lane A) and T and C (lane B) reactions. A major band of approximately 285 bases would correspond with a start at position +1 (Fig.4a). A minor band, a few nucleotides shorter (but more important with RNA transcribed from the argECBH plasmid pMC7 - ref.22) would correspond with a start at positon -4.

b. S1 endonuclease mapping of the argE transcript. A 410 bp SauA3 fragment was 5'labelled, strand-separated and the strand complementary to the argE transcript was used as probe. Lane A : G and A specific reaction products, lane B : T and C specific reaction products , lane C : protected fragment after hybridization with P4XB2 RNA, and lane D : protected fragment after hybridization with pMC7 RNA. The major transcript corresponds with a start at position +13 on Fig.4a, weaker transcripts are also seen with pMC7, possibly an effect of the supercoiling of the plasmid. The upper band could correspond to transcription started at argEp2.

at or near this site (position +1 on Fig.4a). The fact that the minor band in lane D becomes the major one in lane C, loaded with RNA from a strain containing the argECBH plasmid pMC7 (22), suggests that the preferred site for transcription initiation on the plasmid is different from that on the chromosome.

Superimposed genetic sites at the argE end of the regulatory region

Another remarkable effect of the argCp-down mutations just described is that they simultaneously increase the rate of argE translation by 2- to 3-fold (17). Here too, the effect is more pronounced in 6-8 than in the other mutants. The most likely explanation of this phenomenon is that the nucleotide changes create a more effective ribosome loading site on the argE messenger RNA. The two mutations in 6-8 fall respectively 26 and 22 bp upstream of the putative argE translation initiation codon ; they provide direct evidence that the stretch of DNA between the sequence involved in 16S RNA recognition (Shine and Dalgarno, ref.18) and the 5'-end of the argE mRNA is involved in ribosome mobilization ; these mutants are discussed in detail elsewhere (38).

Initiation of argE transcription

No point mutants were available to characterize argEp but the study of phage μ insertions (19) had located it within 150 bp on the right-hand side of the sup102 deletion endpoint (Fig.4). The exact site could be deduced from S1 nuclease mapping experiments which located the 5'-end of the argE mRNA at position +13 (Fig.4a and 5). The relative positions of argCp and argEp are such that the transcripts they initiate must overlap over a short (12 bp) sequence. Transcription initiation at the two face-to-

face promoters should therefore be mutually exclusive. Another putative promoter site -argEp2- (Fig.4) deserves attention. Mutations in this sequence, which lies within argC, result in high, partially constitutive expression of argE (mutants L9 and LL13, Table 1). Those mutations have identical CG → AT transitions at position +198 (Fig.4). They appear to create a new promoter for argE or to activate a weak promoter already present in the wild type (see upper band in lane D of Fig.5b, and further when the mechanism of repression is discussed). Although L9 and LL13 fall within the coding part of argC, the Arg^+ phenotype of these mutations is readily explained since they modify one asparagine codon (AAT) into a synonymous one (AAC).

The operator region. Analysis of operator-constitutive mutants

Several groups of dyad symmetries can be recognized in the promoter region (Fig.4) in a domain which is especially A-T rich (about 80 %), a characteristic shared by many operator sites. A first palindrome (Pal1), from nucleotide -20 to -4, is almost entirely comprised within argCp. A second one, Pal2, extends from within argCp (nucleotide -12) to nucleotide + 25 in argEp. We can also distinguish two similar sequences of 18 bp (boxed in fig.4a) which themselves exhibit dyad symmetry and are entirely contained with Pal2 : TATCAATAT.TCATGCAGT and TATGAATAA.AAATACACT. These regions are noteworthy because a comparison between the regulatory regions of several arg genes (argF, argI, argECBH, argR) has revealed the existence in all of them of partly conserved 18 bp sequences which we call the "ARG boxes" (39,40,41,Cunin et al., in preparation). We have sequenced 6 mutants where this region in argECBH is affected, namely sup102, P4XAO19, P4XAO6, L10, 545 and 13-4.

The sup102 deletion, which destroys argEp and lowers the repressibility of argH from 60- to 8-fold (Table 1) removes half of the righmost ARG box (ARG2). Mutants P4XAO6 and P4XAO19 are spontaneous derivatives of P4Xsup102, selected for the reactivation of argE expression (14,17). Both exhibit a single additional base pair change (Fig.4b) which may be interpreted in each case as creating a new, albeit weak promoter for argE (the corresponding enzyme specific activities are about the fully repressed level - Table 1). The mutations result in the emergence of

Pribnow boxes that could function through their juxtaposition with putative -35 recognition sequences on the other side of Δsup102. The two mutations differ,however, when their effect on the repressibility of argE and argH are compared (Table 1). P4XA019 displays a GC → AT transition, 16 bp on the argE side from Δsup102 in the leftmost ARG box (ARG1) (Fig.4b). Both argE and argH expression are fully constitutive, which suggests that the mutation affects one of the remaining targets for repressor binding in P4Xsup102.

P4XA06 harbours a TA → AT transversion, 10 bp from Δsup102, with no further effect on the repressibility of argH. argE is repressible to the same extent as argH in this strain. It is worth noting that A06 which does not perturb regulation, falls just between ARG1 and ARG2.

Mutant L10 was isolated after mutagenesis with hydroxylamine and selection for high expression of argE. It harbours a GC → AT transition at the same position as in P4XA019 but, this time, in the wild type context (Fig.4a). It is partially constitutive for both argE and argCBH which are now about 7-fold repressible.

Mutant 545, isolated by G. Jacoby (13) after selection for enhanced argC expression, harbours a GC → AT transition at the same position as in L10 and P4XA019.

Mutant 13-4 was obtained from an argC-lac operon fusion strain (19). It harbours a single base pair deletion at the first position in the ARG2 box. It exhibits a 2-fold residual repression of both argE and argCBH (Table 1).

The operator mutations define two regions which appear essential to repressor-operator interactions ; (i) the argC proximal part of Pal2, which largely overlaps ARG2 (mutations 13-4 and Δsup102) and (ii) the left hand portion of Pal2, which largely overlaps ARG1 (mutations L10, 545, A019). The respective roles of these regions as determinants in the mechanism of repression are discussed in the next section.

DISCUSSION

Mechanism of repression

Since the quaternary structure of the arginine repressor is unknown, it is not possible yet to define with absolute certain -

ty which of the two following symmetries is the most important for the DNA-repressor interactions : the Pal2 palindrome which extends over 37 bp, or the shorter ARG boxes, which exhibit internal symmetry and could constitute two adjacent repressor binding sites.

However, the following considerations lead us to favour the view that the two ARG boxes are the relevant features of the dyad symmetry region for repressor control. First of all, similar sequences were found in the control region of argF, argR (39,40,41,42 ; Cunin et al., in preparation), and argI (41).

Moreover, were the Pal2 palindrome the target of repressor binding, we would expect repressibility to be more severely affected by the sup102 deletion than by point mutations. In fact, residual repressibility of argH in P4Xsup102 is about 8-fold as compared to 7-fold in mutants L10 and 545, and 2-fold in mutant 13-4. Rather, these data suggest that repressor molecules could bind at the two adjacent ARG boxes considered as distinct repressor targets and in a cooperative fashion ; ARG2 alone is not sufficient for strong repression as shown by the already low residual repressibility displayed by L10 and 545, both located in ARG1. However, ARG2 appears necessary for complete repression since in the P4Xsup102 mutant, where 50 % of ARG2 has been deleted, the repressibility of argH is no more than 8-fold, compared to the 60-fold ratio observed in the wild type.

The behaviour of mutant 13-4 is also compatible with this hypothesis ; indeed the strong constitutivity exhibited by this single base pair deletion could stem from its putting the two binding regions out of register by shortening the distances between them. The hypothesis of two adjacent operator sites is further supported by the observation that complete constitutivity is achieved only when mutations in each box are combined (as in P4XA019) ; it is also in keeping with the fact that a single base pair substitution between the two ARG boxes (in P4XA06) does not alter regulation.

The operator -whether Pal2 or the two ARG boxes- overlaps with both argCp and argEp in such a way that bound repressor must sterically prevent RNA polymerase from binding at these promoters.

In spite of the relatively symmetrical position of argEp and

argCp with respect to the operator region, argE and argCBH exhibit different repression-derepression ratios (17- and 60-fold respectively). This differential repressibility could arise from a different efficiency of the operator-bound repressor in blocking transcription initiation on both sides of the operator region. However, another factor should be taken into account. In mutant L9 and LL13, a GC→AT transition at position +198 (Fig.4a) results in the emergence of a strong promoter site (argEp2) from a sequence which may already function as a weak initiation site in the wild type. Transcription from this weak promoter could be relatively insensitive to repression because of the total absence of overlap between the promoter and the operator region, and therefore be the cause of the lower repression-derepression ratio for argE.

The situation of the ARG boxes in the arginine regulon presents similarities to that of the operator sites present in the control region of the lexA and recA genes (43) ; argR is preceded by one putative repressor binding site like recA while argECBH, argF and argI exhibit two such sites in their control region, like lexA (39,40,41,42).

Lack of attenuation control

Most amino acid biosynthetic operons are controlled by an attenuation mechanism which correlates early termination of transcription to the intracellular concentration of the end product of the corresponding biosynthetic chain (44). Previous investigations on the pattern of messenger synthesis in the cluster, both in vitro and in vivo, failed to bring to light a control by attenuation for the argCBH arm of the cluster (45). Although the data presented here suggest the existence of a 115bp leader RNA sequence for argCBH, this region exhibits none of the features associated with an attenuation mechanism (successive arg codons and potential stem-loop secondary structures that could act as terminators) ; the only stem-loop structure that could be formed would not be stable at 25°C. The argF and argI regulatory regions are also devoid of attenuation features (41, 42,46). The function, if any, of the argCBH leader is unknown.

It is tempting to draw a parallel between the absence of attenuation in the control region of the arg genes investigated

so far and the scattered genetic organization of the regulon. If the arg regulon arose by chromosomal rearrangements scattering genes originally grouped into a primeval, attenuated arg operon, one would not expect these genes to be translocated systematically with the attenuator site of the operon. Besides, depending on the sequences that would become fused to each other in such a process, the translocated genes could find themselves preceded by a stretch of DNA having no relation whatsoever with an attenuation mechanism. The presence of a leader of unknown function in front of the aroH (47) could be another example of such a situation. If the primeval arg cluster had already been a divergent operon expressed from a site between argE and argC, the structure of the argECBH control region would require another explanation ; the leader sequence, whether it presently serves a function or not, might be the consequence of the rearrangement which created the divergent operon. We have indeed observed the de novo formation of divergent operons consisting in inverted repeats of argE separated by a control region of variable length (48 ; Charlier et al., manuscript submitted).

Another possibility would be that the arg genes were originally attenuated but that this type of control has been lost, leaving the present genes regulated only by repression and preceded by remnants of formerly attenuated leaders. The teleological significance, if any, of the loss of an attenuation control or of its replacement by repression is not obvious.

Finally, we would like to emphasize the overall similarity between the genetic organization of the control region of argECBH and that of another bipolar operon : bioABFCD (16,49) as well as between their patterns of regulation. It seems quite probable that both operons have originated from similar sequences of events. Both clusters seem to be exclusively under repression control (this work and ref.50) and harbour face-to-face promoter sites, the bio promoters lying closer to each other, however, so that bio transcripts do not overlap at all. A point of difference between the two systems resides in the greater complexity of the operator region in argECBH and the asymmetrical situation of the translation starting points of argE and argC.

Acknowledgements

The authors are indebted to Mrs Else Saederup and Mrs N.Huysveld for their assistance with the experiments and to Dr. Eva Van den Bergh (University of Leyden, The Netherlands) for computer analysis of the sequence.

This work was supported by grant nr.S2/5 AME21 of the Belgian NFWO to R.C. and by grant nr.2.9005.79 of the Belgian FKFO to N.G. J.P. is a fellow of the Belgian IWONL.

REFERENCES

1. Hofnung, M. (1974) Genetics 76, 169-184.
2. Bedouelle, M. and Hofnung, M. (1982) Molec. gen. Genet. 185, 82-87.
3. Debarbouillé, M., Cossart, P. and Raibaud, O. (1982) Molec.gen. Genet. 185, 88-92.
4. Guha, A., Saturen, Y. and Szybalski, W. (1971) J. Mol.Biol. 56, 53-62.
5. Cleary, P.P., Campbell, A. and Chang, R. (1972) Proc. Nat. Acad. Sci. USA 69, 2219-2223.
6. Reichardt, L. and Kaiser, A.D. (1971) Proc. Nat. Acad. Sci. USA 68, 2185-2189.
7. Chou, J., Lemaux, P.G., Casadaban, M.J. and Cohen, S.N. (1979) Nature 282, 801-806.
8. Smith, B.R. and Schleif, R. (1078) J. Biol. Chem. 253, 6931-6933.
9. Greenfield, L., Boone, T. and Wilcox, G. (1978) Proc. Nat. Acad. Sic. USA 75, 4724-4728.
10. Wray, L.V., Jorgensen, R.A. and Reznikoff, W.S. (1981) J. Bacteriol. 147, 297-304.
11. Pouwels, P., Cunin, R. and Glansdorff, N. (1974) J. Mol. Biol. 83, 421-424.
12. Panchal, C.J., Bagchee, S.N. and Guha, A. (1974) J. Bacteriol. 117, 675-680.
13. Jacoby, G.A. (1972) Molec. gen. Genet. 117, 337-348.
14. Elseviers, D., Cunin, R., Glansdorff, N., Baumberg, S. and Ashcroft, E. (1972) Molec. gen.Genet. 117, 349-366.
15. Ketner, G. and Campbell, A. (1975) J. Mol. Biol. 96,13-27.
16. Otsuka, A. and Abelson, J. (1978) Nature 276, 689-694.
17. Boyen, A., Charlier, D., Crabeel, M., Cunin, R., Palchaudhuri, S. and Glansdorff, N. (1978) Molec. gen. Genet. 161, 185-196.
18. Shine, J. and Dalgarno, L. (1974) Proc. Nat. Acad. Sci. USA 71, 1342-1346.
19. Beny, G., Boyen, A., Charlier, D., Lissens, V., Feller, A., and Glansdorff, N. (1982) J. Bacteriol. 152, 62-67.
20. Bretscher, A.P. and Baumberg, S. (1976) J. Mol. Biol. 102, 205-220.
21. Crabeel, M., Charlier, D., Cunin, R. and Glansdorff, N. (1979) Gene 5, 207-231.
22. Crabeel, M., Charlier, D., Glansdorff, N., Palchaudhuri, S. and Maas, W.K. (1977) Molec.gen. Genet. 151, 161-168.
23. Wells, R.D., Hardies,S.C., Horn, G.T., Klein, B., Larson, J.E., Neuendorf, S.K., Panayotatos, N., Patient, R.K. and

Selsing, E. (1980) Methods in Enzymol. 65, 327-347.
24. Tabak, H.F. and Flavell, R.A. (1978) Nucl. Ac. Res. 5, 2321-2332.
25. Van de Sande, J.H., Kleppe, K. and Khorana, H.G. (1973) Biochemistry 12, 5050-5055.
26. Tu, C.P. and Cohen, S.N. (1980) Gene 10, 177-183.
27. Maxam, A. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.
28. Barry, G., Squires, C.L. and Squires C. (1979) Proc. Nat. Acad. Sci. USA 77, 3331-3335.
29. Ghosal, D., Sommer, H. and Saedler, H.(1979) Nucl. Ac. Res. 6, 1111-1121.
30. Atkins, J.F. (1979) Nucl. Ac. Res. 7, 1035-1041.
31. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genetics 13, 319-353.
32. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
33. Galas, D.J. and Schmitz, A. (1978) Nucl. Ac. Res. 5, 3157-3170.
34. Pribnow, D.(1975) Proc. Nat. Acad. Sci. USA 72, 784-788.
35. Busby, S., Aiba, H. and de Crombrugghe, B. (1982) J. Mol. Biol. 154, 211-227.
36. Saedler, H., Reif, H.J., Hsu, S. and Davidson, N. (1974) Molec. gen. Genet. 132, 265-289.
37. Dickson, R.C., Abelson, J., Barnes, W.M. and Reznikoff, W.S. (1975) Science 187, 27-35.
38. Boyen, A., Piette, J., Cunin, R. and Glansdorff, N. (1982) J. Mol. Biol. in the press.
39. Cunin, R., in "Biotechnology, vol.1 : Amino Acid Biosynthesis and Regulation" (eds. Hermann, K. and Sommerville, R.) Addison Wesley, New York, in the press.
40. Cunin, R., Eckhardt, T., Piette, J., Boyen, A. and Glansdorff, N. (1981) Arch. Intern. Physiol. Biochim.89, B-94.
41. Piette, J., Cunin, R., Van Vliet, F., Charlier, D., Crabeel, M., Ota, Y. and Glansdorff, N. (1982) EMBO Journal 1,853-857.
42. Piette, J., Cunin, R., Crabeel, M. and Glansdorff, N. (1981) Arch. Intern. Physiol. Biochim. 89, B126-B127.
43. Brent, R. and Patshne, M. (1981) Proc. Nat. Acad. Sci.USA 78, 4204-4208.
44. Yanofsky, C. (1981) Nature 289, 751-758.
45. Beny, G., Cunin, R., Glansdorff, N., Boyen, A., Charlier, D. and Kelker, N. (1982) J. Bacteriol. 151, 58-61.
46. Moore, S.K., Garvin, R.T. and James, E. (1981) Gene 16, 119-132.
47. Zurawski, G., Gunsalus, R.P., Brown, K.D. and Yanofsky, C. (1981) J. Mol. Biol. 145, 47-73.
48. Charlier, D., Crabeel, M., Cunin, R., and Glansdorff, N. (1979) Molec. gen. Genet. 174, 75-88.
49. Barker, D.F., Kuhn, J. and Campbell, A.M. (1981) Gene 13, 89-102.
50. Barker, D.F. and Campbell, A.M. (1981) J. Mol. Biol. 146, 469-492.
51. Cunin, R., Elseviers, D., Sand, G., Freundlich, G. and Glansdorff, N. (1969) Molec. gen. Genet. 106, 32-47.
52. Bachmann, B.J., Low, B.K. and Taylor, A.L. (1976) Bact. Rev. 40, 116-167.