The EMBO Journal vol.7 no.13 pp.4389-4395, 1988

The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus

C.J.Stirling, G.Szatmari, G.Stewart, M.C.M.Smith and D.J.Sherratt

Institute of Genetics, University of Glasgow, Church Street, Glasgow G11 5JS, UK

Communicated by D.J.Sherratt

The heritable stability in *Escherichia coli* of the multicopy plasmid ColE1 and its natural relatives requires that the plasmids be maintained in the monomeric state. Plasmid multimers, that arise through *recA*-dependent homologous recombination, are normally converted to monomers by a site-specific recombination system that acts at a specific plasmid site (*cer* in ColE1). No plasmid functions that act at this site have been identified. In contrast, two unlinked *E.coli* genes that encode functions required for *cer*-mediated site-specific recombination have been identified. Here we describe the isolation and characterization of one such gene (*xerA*) and show it to be identical to the gene encoding the repressor of the arginine biosynthetic genes (*argR*). The *argR* protein binds to *cer* DNA both *in vivo* and *in vitro* in the presence of arginine. We believe this binding is required to generate a higher order protein-DNA complex within the recombinational synapse. The *argR* gene of *Bacillus subtilis* complements an *E.coli argR* deficiency for *cer*-mediated recombination despite the two proteins having only 27% amino acid identity.

Key words: *argR*/ColE1/plasmid stability/site-specific recombination

Introduction

The heritable stability in *Escherichia coli* of the multicopy plasmid ColE1 and its natural relatives requires that the plasmid be maintained in the monomeric state. Plasmid ColE1 multimers, that result from inter-molecular homologous recombination, are resolved to monomers by a directional site-specific recombination event that requires a 250 bp plasmid region, *cer*, and chromosomally encoded products that act to mediate recombination at a specific site within *cer* (Summers and Sherratt, 1984, 1988; Summers *et al.*, 1985; Stirling *et al.*, 1988). Other natural plasmids also encode site-specific recombination systems that are necessary for stable inheritance and are used to maintain the plasmid in the monomeric state (Austin *et al.*, 1981; Dodd and Bennett, 1986; Garnier *et al.*, 1987; Hakkaart *et al.*, 1984; Lane *et al.*, 1986). Relatives of ColE1 have regions analogous to *cer*, and which can recombine with *cer*, allowing delineation of the crossover site (Summers *et al.*, 1985; Summers, personal communication).

Since *cer* acts *in cis* and its sequence reveals no significant protein-coding regions, we inferred that the proteins that act at *cer* must be chromosomally encoded. *E.coli* chromosomal mutants, isolated after Tn*5* mutagenesis, and defective in recombination at *cer*, were isolated and characterized into at least two unlinked classes (Stirling *et al.*, 1988). One of these functions is encoded by a gene (*xerB*) mapping close to *argI* and *pyrB* (96-97 min) (Stirling *et al.*, 1988). Here we show that another gene initially designated *xerA*, maps at 70.5 min on the *E.coli* chromosome, and is identical to *argR*, the gene for the arginine repressor. *argR* function is absolutely necessary for monomerizing recombination at ColE1 *cer* and we show that arginine repressor binds to a specific site in the *cer* region both *in vivo* and *in vitro*.

Results

Isolation and characterization of the E.coli chromosomal gene xerA

Transposon Tn*5*-generated *E.coli* mutants defective in stabilizing, *cer*-dependent site-specific recombination have been selected by their inability to delete a 'reporter gene' bounded by directly repeated *cer* sites in a multicopy plasmid. Mutants completely defective in *cer* recombination fall into two classes by genetic complementation (Stirling *et al.*, 1988). These two classes define two genes *xerA* and *xerB*; the latter gene maps at 97 min on the *E.coli* map, and has been physically mapped close to *pyrB*, *argI* and *valS* (Stirling *et al.*, 1988; and C.Stirling, unpublished). In order to characterize the *xerA* gene, Tn*5* mutants and wild-type *xerA* genes were isolated, physically mapped and sequenced (Figures 1 and 2; Summers *et al.*, 1986). The isolation strategy was to clone, into a multicopy plasmid vector, the Tn*5*-containing mutant gene, by virtue of the Tn*5*-mediated kanamycin resistance. This led to the isolation of a 19.4 kb plasmid pCS300 which contains, within pBR322, a 9.2 kb fragment of *E.coli* chromosomal DNA into which the 5.8 kb Tn*5* had inserted. The plasmid was restriction mapped and the map of the cloned chromosomal DNA (Figure 1) was compared to the restriction map of the *in situ* chromosomal *xerA* locus (assayed by Southern hybridization) to ensure that the cloned Tn*5* was the transposon copy that was 100% linked to the *xerA* locus. The inferred restriction map of the wild-type *xerA* gene locus predicted the presence of an 8.45 kb *Hin*dIII-*Eco*RI fragment. Fragments of this size were therefore cloned from an *Eco*RI-*Hin*dIII digest of Xer^+ chromosomal DNA into the plasmid expression vector pAT223. Xer^+ clones were screened by their ability to complement the *xerA*3::Tn*5* mutation. On such Xer^+ recombinant plasmid (pGS30) that had the restriction map of the *xerA* locus was selected for further study. Sub-cloning from the 8.45 chromosomal fragment identified a functional $XerA^+$ 0.92 kb *Sph*I-*Acc*I fragment. This fragment was sequenced (Figure 2).

Sequence of the xerA gene

The sequence shown in Figure 2 revealed three open reading frames (ORFs) longer than 50 codons: one encodes a potential 156 amino-acid residue protein; one a potential 82

residue and the other a potential 70 amino-acid residue protein. All three ORFs are disrupted by the Tn*5* insertion within *xerA*3. Minicell analysis of proteins specified by the 0.93 kb *XerA*$^+$ *Sph*I−*Acc*I fragment revealed a 17 kd polypeptide (data not shown). This could correspond to the longest 156 codon ORF. Moreover a 725 bp *Sst*I−*Dra*I fragment encoding the two smaller ORFs did not complement the mutant phenotype of *xerA*$^-$ cells. Together these results favoured the view that the *xerA* gene is the 156 ORF. This view was reinforced by computer analyses of potential promoters, potential translation starts and of codon usage; all pointed to the 156 codon ORF as having the characteristics of a functional gene.

At the time the sequence was determined, no similar sequences were within the data bases. However the availability of a restriction map of the *E.coli* chromosome (Kohara *et al.*, 1987) allowed us provisionally to place the *xerA* locus at 70.5 min, close to the mapped *argR* and *mdh* genes. The *argR* gene encodes a protein product of 17 kd and W.Maas kindly supplied his then unpublished sequence of *argR* (now published as Lim *et al.*, 1987). The 156 codon ORF tentatively identified as *xerA* is identical to *argR*. Henceforth, for consistency we continue to describe our mutant alleles of *xerA*/*argR* as *xerA* mutants.

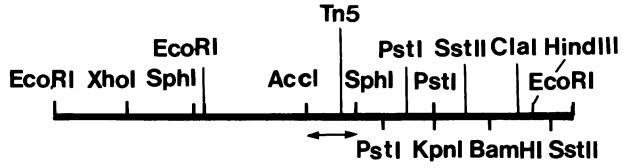

Fig. 1. Restriction map of 9.2 kb *Eco*R1 fragment of *E.coli* chromosomal DNA containing the *xerA* (*argR*) gene. The position of the Tn*5* insertion in *xerA*3::Tn*5* is indicated as is the position of the *xerA* (*argR*) ORF (←—).

The 920 bp annotated sequence (Figure 2) extends the 818 bp sequence of Lim (1987). Our own sequence contains five additional base pairs outside of the *xerA* coding sequence (at coordinates 766, 781, 803, 805 and 808). In addition, we note that Lim *et al.* (1987) reported the cloning of *argR* as an 880 bp *Sph*I−*Acc*I fragment as compared to the 920 bp *Sph*I−*Acc*I fragment isolated from our strain. The ARG boxes, that are bound by repressor (Lim *et al.*, 1987; this manuscript) are indicated, as are two functional promoters (p1 and p2, Lim *et al.*, 1987). The sequence also contains a third promoter-like sequence (p3), detected by computer analyses of the sequence. We have also noted a 61 bp sequence within *argR* that is similar to a region within *cer*. The optimal alignment includes 61 bp of *argR* and a 61 bp portion of *cer*, which with the insertion of two single base-pair gaps into each sequence, results in a total of 40 out of 61 matches. This portion of the *cer* sequence encompasses the 34 bp region known to contain the crossover site (Summers *et al.*, 1985), and which is highly conserved in ColK, pMB1, CloDF13 and ColA (Summers and Sherratt, 1988; Figure 3) This *cer*-like site in *argR* is not detectably a substrate for Xer-mediated site-specific recombination (deletion or inversion), either with itself or with *cer* (data not shown). The significance of this similarity therefore remains unclear.

Also within the *argR* gene (Figure 2) we have noticed a 13 bp sequence showing very strong similarity to the consensus integration host-factor (IHF) binding site (Leong *et al.*, 1985).

argR (261−273 bp) CAAGGCTTTGACA

IHF consensus $^{C}_{T}$AANNNNTTGAT$^{A}_{T}$

The fact that a consensus IHF binding sequence is close both to the *cer*-like site within *argR* and to the crossover region

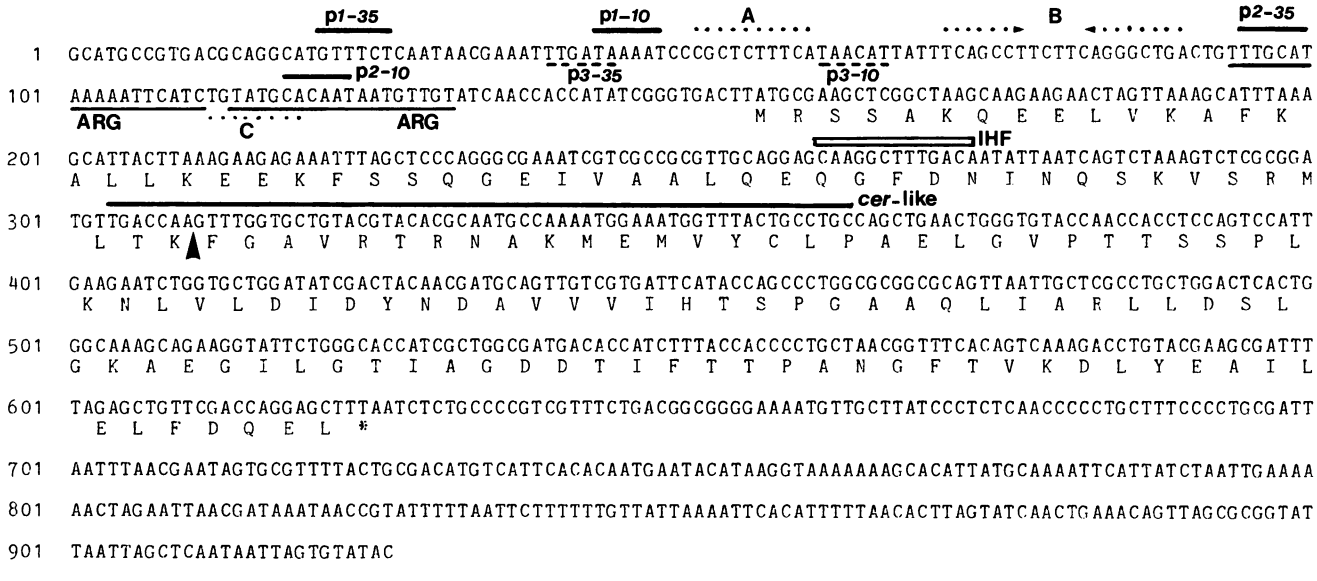

Fig. 2. DNA sequence of the 920 bp *Sph*I−*Acc*I fragment containing the *xerA* (*argR*) locus. The sequence was determined on both strands using the Sanger dideoxy chain-termination technique. Ambiguities were rigorously resolved. The 156 amino-acid residue coding seqeunce for *argR* is shown. We identified presumptive promoters p1, p2 and p3 by computer analysis. p1 and p2 correspond to functional promoters identified by Lim *et al.* (1987). The presumptive IHF binding site is shown ( ), as is the *cer*-like sequence ( ); nucleotides 301−364) and the two 18 bp arginine repressor binding sites (ARG boxes). the *nut*-like site containing 'box A', 'box B' and 'box C' is also shown ($\overset{A}{\ldots}\ldots\overset{B}{\ldots}\ldots\overset{C}{\ldots}$). The site of Tn*5* insertion in *xerA* is shown by a vertical arrow (between nucleotides 310 and 311).

```
          -35                    -10
argR;     ttTGcATaaaaATtCAtctgtaTGcAcaataATgttgt---172nt---TGACCAAGTTTGGTGCTGTACGTACACGCAATGCCAAAATGG-AAATGG-TTTACTGCCTGCC
cer       gtTGcATaggtATtCAta--cggttAaaatttatCAgg---65nt----TGAACGCGTTT-TAGCGGTGCGTACAATTAAGGG-ATTATGGTAAATCCACTTACTGTCTGCC
          a      ta         a      ta
CONSENSUS aTGAATa  nATnCAntnn aTGAATa  nATnCAnt
          t      at         t      at
```

Fig. 3. Comparative alignment of the ARG boxes of ColE1 *cer* and the *argR* gene and of the crossover region of *cer* and the similar sequence in *argR*. The consensus ARG box sequence shown is derived from the sequence of the various arginine operators (Glansdorff, 1987; Lim *et al.*, 1987). In the characterized *arg* operators, two directly repeated ARG boxes occur, separated by two (as shown in the consensus sequence) or three nucleotides. *argR* has a poor second box and *cer* an almost non-existent second box. The promoters in the ARG boxes of *argR* and *cer* are indicated. The crossover region in *cer* is indicated (———) as are the inverted repeats that contain putative IHF binding sites. Bold face type in the *argR* and *cer* sequences indicates identity of bases in the two sequences. Within the ARG boxes of *argR* and *cer*, upper case type indicates identity to conserved nucleotides in the consensus.

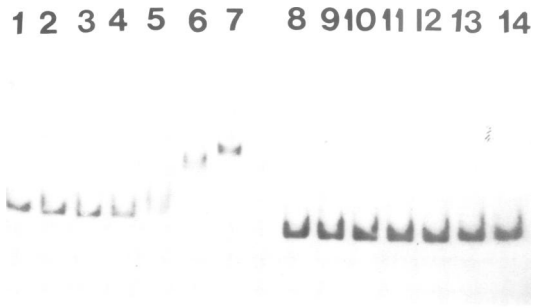

Fig. 4. Detection of *cer*-specific DNA binding activity of purified *argR* protein using the gel retardation assay. **Lanes 1–7**: ^{32}P-labelled 481 bp *Eco*R1–*Pvu*II restriction fragment of pKS492 (consisting of the 280 bp *cer* sequence with some flanking pUC18 DNA) incubated with 0, 0.45, 0.9, 1.8, 2.7, 3.6 and 4.5 pmol of *argR* protein, respectively. **Lanes 8–14** were treated in the same manner as lanes 1–7 except that the ^{32}P-labelled DNA used was the 409 bp pUC19 *Dde*I restriction fragment.

within *cer* itself (Figures 2 and 3), is particularly intriguing in view of IHF involvement in other site-specific recombination systems (Leong *et al.*, 1985; Sadowski, 1986).

We also note that just 5′ of the *argR* translational start there is bacteriophage λ *nut*-like site (Friedman and Gottesman, 1983; Figure 2). Such sites are where λ N antitermination protein and host factors become associated with transcribing RNA polymerase, rendering it insensitive to downstream termination sites (Friedman and Gottesman, 1983). Candidate *nut*A, B and C boxes are all present adjacent to the *argR* gene. The B box is an almost perfect inverse of the λ *nut*L-B box, and is also a potential transcription termination site. Perhaps significantly, several *E.coli* terminators and attenuation sites are preceded by a *nut*A (Friedman and Gottesman, 1983). Again, we have no insight into the biological significance, if any, of these sequences 5′ of the ARG box that are present in a wild-type transcription initiating upstream.

The argR gene product binds to ColE1 cer DNA in vitro and in vivo

In vitro binding. The *argR* protein was overexpressed from plasmid pGS38-containing cells after IPTG treatment, and arginine repressor was isolated from DNase-treated sonicates of these cells based on the ability of arginine repressor to be precipitated by arginine (5 mM) in low salt buffer as a purification regime (Lim *et al.*, 1987). Arginine repressor purified in this way (>50% pure) was assayed for its specific DNA binding by the gel retardation assay of Fried and Crothers (1983).

As shown in Figure 4, the arginine repressor specifically retards DNA fragments containing the *cer* site. This retardation is absolutely dependent on the presence of L-arginine. Failure to include L-arginine in the binding reactions, or running buffer results in the failure to detect specific DNA binding (not shown).

The specific binding of the arginine repressor to *cer* was further localized by splitting the *cer* site in two at the *Pvu*I site, giving a promoter-containing half and a crossover-containing half. When a gel retardation assay was performed on these two fragments, the *argR* protein was found to bind to the promoter-containing half of the *cer* site (not shown).

To define precisely the region of *cer* being specifically bound by the arginine repressor *in vitro*, a DNase I footprint analysis (Galas and Schmitz, 1978) was performed on both strands of the *cer* region. As shown in Figure 5, (lanes a–f) the *argR* protein appears to protect ~19 bp of the top strand of *cer* corresponding to the region of the consensus ARG box. The bottom strand (data not shown) displays a protected region spanning ~21 nucleotides. In addition, two symmetrically-located enhanced DNase I cleavages occur. These enhanced cleavages both occur between the C and A of the imperfect ATTCA inverted repeat found in the ARG box. Two other enhanced cleavages are found on the top strand, occurring 6 and 8 bp to the 5′ side of the ARG box (Figure 5). The region 3′ (top strand) of the ARG box is intrinsically resistant to DNase I, (Figure 5, lanes c–f) presumbaly because of compression of the DNA minor groove in this region. This prevented us from determining by DNase 'footprinting' whether arginine repressor binds to the second pseudo-ARG box in this region (Figure 4).

In vivo binding. Binding the arginine repressor to ARG boxes *in vivo* was assessed by *in vivo* footprinting using protection to dimethylsulphate (DMS) (Martin *et al.*, 1986) and by the ability of repressor to inhibit transcription of the *argR* and *cer* promoters linked to a *galK* 'reporter' gene on a multicopy plasmid (McKenney *et al.*, 1981). In the presence of an expressed *argR* gene product, a guanine of *cer* is protected against DMS methylation within the ARG box *in vivo* (Figure 5) demonstrating that binding to this region of *cer* occurs *in vivo*.

In vivo binding of the arginine repressor to ColE1 *cer* was also assayed by exploiting the fact that the ARG-box within *cer* contains a functional promoter (whose biological function is unknown) that can be used to 'drive' a promoterless *galK* gene on a pBR322 related plasmid. Binding of repressor *in vivo* should repress *galK* expression. As a positive control we used a fragment containing the chromosomal *argR* promoter from which expression is known to be repressed by repressor binding. The results are shown in Table I. In

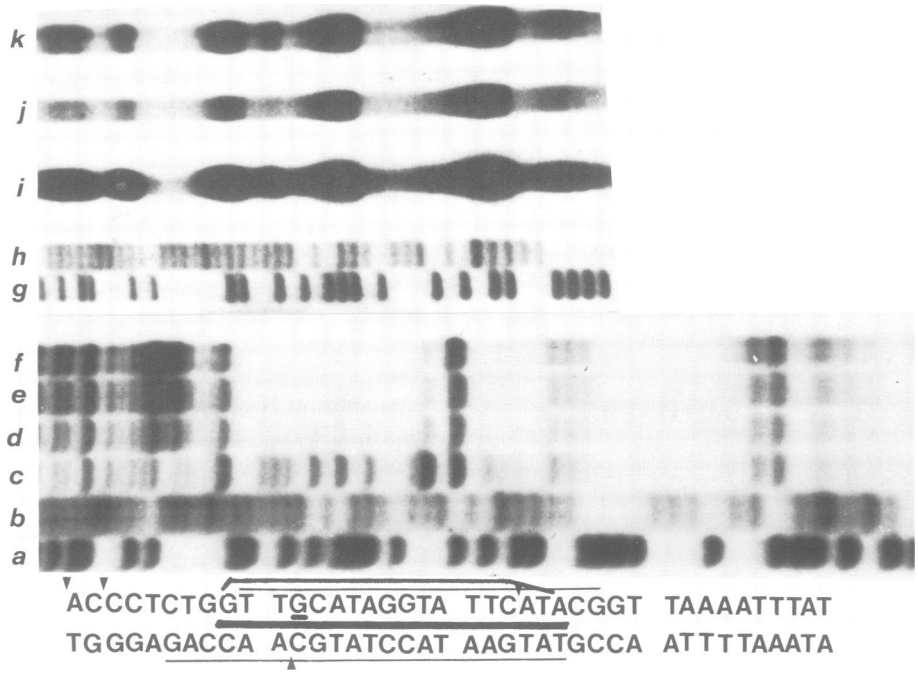

Fig. 5. *In vitro* and *in vivo* footprints of the *argR* protein on *cer* DNA. An *in vitro* DNase footprint experiment (Galas and Schmitz, 1978) is shown in **lanes a–f**. **Lanes a, b**, purine and pyrimidine sequencing reactions respectively; **lane c**, 0 pmol *argR* protein; **lane d**, 15 pmol *argR* protein; **lane e**, 30 pmol *argR* protein; **lane f**, 60 pmol *argR* protein. An *in vivo* DMS footprint (Martin *et al.*, 1986) is shown in **lanes g–k**. **Lanes g, h**, purine and pyrimidine sequencing reactions respectively. **Lanes i, j, k** show *cer* DNA modified by DMS *in vivo* and isolated from cells containing plasmid vector (**lane i**) *argR*$^+$ plasmid (**lane j**); and *argR*$^-$ plasmid (**lane k**). Underneath the footprints is the sequence of the ARG box-containing *cer* region. The consensus ARG box is indicated by a thick horizontal line and the region footprinted by DNase in our experiments is demonstrated by thin horizontal lines. Only the gel of the top strand reactions is shown. Arrows indicate enhanced cleavages. The G residue protected from DMS methylation *in vivo* is underlined.

Table I. Binding of the arginine repressor to its own promoter and ColE1 *cer in vivo*. Binding was assayed by its effect on *galK* expression in plasmids in which *galK* expression was either from the *argR* promoter (pKO-p*argR*) or from the *cer* promoter-containing region (pKO-p*cer*). Galactokinase levels were normalized by comparing them to β-lactamase levels (the pKO plasmids encode *bla*), thus correcting for deviations in copy number etc. Results for two separate experiments are shown.

E.coli strains	*galK* expression in the presence of the undernoted additional plasmids					
	None		pC348 (*argR*$^+$)		pCS119 (*argR*$^-$)	
	galK sp. act.	*galK/bla*	*galK* sp. act.	*galK/bla*	*galK* sp. act.	*galK/bla*
DS941 (Xer$^+$) pKO-p*argR*	216, 302	2.5, 2.8	49, 44	0.43, 0.44	119, 115	1.8, 1.9
DS941*xerA*3 pKO-p*argR*	323, 249	2.9, 3.2	48, 42	0.54, 0.48	160, 160	2.7, 2.8
DS941 (Xer$^+$) pKO-p*cer*	85, 80	0.87, 0.92	15, 9	0.22, 0.18	42, 38	0.42, 0.40
DS941*xerA*3 pKO-p*cer*	95, 90	1.0, 0.95	15, 11	0.23, 0.16	47, 49	0.36, 0.48

either a Xer$^+$ or *xerA*3 *E.coli* strain, high levels of *argR* expression *in trans* from a compatible λ *dv* based multicopy plasmid pCS348 resulted in the repression of transcription from both the *cer* promoter (4- to 6-fold reduction in expression) and the *argR* promoter (5- to 7-fold reduction) as measured by comparison of the normalized *galK/bla* ratios. The presence of a similar *argR*$^-$ plasmid (pCS119) reduced *galK* expression by up to 30% from the *argR* promoter and by about a factor of 2 for the *cer* promoter. We do not understand the reason for these effects, but they are substantially less than the repression from pCS348. Also note that the chromosomal status of the *argR* gene has little effect on *galK* expression in the absence of additional plasmid. This was unexpected and the reason is still puzzling, though the inability of chromosomally synthesized arginine repressor to act efficiently on ARG boxes on plasmids has been reported previously (e.g. see Glansdorff, 1987). This failure of a wild-type chromosomal *argR*$^+$ gene to repress a plasmid-borne *argR* promoter is apparently not due to titration of all of the repressor synthesized by the multicopy promoter, since in such strains the two chromosomal arginine repressor-controlled ornithine carbamoyl transferase (OCTase) genes are fully repressed (data not shown). Functional compartmentalization (for example, by differences in DNA topology) would appear to be the obvious explanation of this apparent preferred *cis* action of arginine repressor.

The B.subtilis argR gene product can act at cer in E.coli to mediate ColE1 site-specific recombination

The *B.subtilis* arginine repressor represses the arginine biosynthetic genes in that organism. It has 27% identical

Table II. Action of *B.subtilis* arginine repressor at ColE1 *cer*

(a) Plasmids	Proportion of transformants retaining Tc^R 'reporter' marker of pCS202 after transforming indicated plasmids into strains shown below	
	DS903*xerA*3	DS903(*XerA*$^+$)
pCS202 + pUL2033	50/50	0/50
pCS202 + pGLW11	50/50	0/50
(b) strain origin of independent transformants from (a) used for plasmid reisolation	Numbers of retransformants retaining Tc^R 'reporter' marker of pCS202	
	DS903*xerA*3	DS903(*XerA*$^+$)
DS903*xerA*3 pCS202+ pUL2033	(i) 3/50	0/6
	(ii) 1/50	0/6
	(iii) 3/50	0/6
	(iv) 1/50	0/6
	(v) 5/50	0/6
DS903*xerA*3 pCS202+ pGLW11	(i) 49/50	not done

DS903*xerA*3 and DS903 were separately co-transformed either with the 2-*cer* 'reporter' plasmid pCS202 and pUL203 which expresses the *B.subtilis argR* gene, or with pCS202 and pGLW11, the parental vector of pUL2033. Presence of the 'reporter' gene was assayed for (Tc^R) in 50 individual transformants of each strain and then small-scale plasmid preparations were made from several transformants and used to transform both DS903*xerA*3 and DS903. Transformants were then tested for the presence/loss of the reporter marker (TC^R) as indicated in panel (b). The presence/loss of Tc^R after retransformation was correlated with *cer*-mediated resolution of pCS202 analysed by gel electrophoresis (not shown). Indeed, electrophoretic analysis of primary transformants showed that resolution was occurring despite the fact the Tc^R had not been lost from all cells in the colony at this stage.

amino acid residues to the *E.coli* enzyme (A.North, M.Smith and S.Baumberg, unpublished). The *E.coli* strain DS903*xerA*3 was co-transformed with the 2-cer 'reporter' plasmid pCS202 and pUL2033, a compatible plasmid in which the *B.subtilis argR* gene is expressed from the *tac* promoter. Analysis of transformants by using small-scale plasmid preparation from them to transform both *xerA*3 and *xer*$^+$ strains, and testing for the fraction that had just the 'reporter' marker is shown in Table II. It is clear that *B.subtilis* arginine repressor can complement the deficiency in the *E.coli xerA*3 strain. These results of phenotypic analysis were confirmed by electrophoretic analysis of resolution of pCS202 in DS903*xerA*3 pUL2033$^+$ strain (not shown).

We believe the complementation is the consequence of *B.subtilis* repressor binding to *cer*. Supporting this view is the demonstration that *B.subtilis* repressor can repress transcription from both *E.coli* OCTase genes and can bind to the *argR* operator *in vitro* (Smith *et al.*, 1988). This is despite the differences in the *E.coli* and *B.subtilis* enzymes and their rather different natural operators.

Discussion

As our studies on ColE1 *cer* site-specific recombination have progressed we have been surprised to find an absence of a plasmid encoded *trans*-acting protein function for recombination. This contrasts to other characterized site-specific recombination systems which normally have a recombinase gene adjacent to the recombination site (for a review see Sadowski, 1986). Moreover, our studies revealed no obvious similarity of the *cer* region with the recombination sites used by either the λ integrase-like class of enzymes or the resolvase/invertase enzymes. The use by ColE1 and related plasmids of a monomerizing site-specific recombination system to ensure stable inheritance is not novel. Indeed most, if not all, natural plasmids encode such a system. However, plasmids other than the ColE1 group that have been investigated utilize systems belonging to either the integrase or resolvase class. For example, plasmids P1 and F use different integrase-like systems (Austin *et al.*, 1981; Lane *et al.*, 1986), while R46, found in enteric bacteria and pIP04 of *Clostridium perfringens* use resolvase-like systems (Dodd and Bennett, 1986; Garnier *et al.*, 1987).

Using Tn*5* mutagenesis, we have isolated and partly characterized two classes of *E.coli* chromosomal mutant that are completely defective in monomerizing recombination between *cer* sites. The discovery that one of these classes defines a gene, *xerA*, that is identical to *argR*, the gene for the arginine repressor, initially astounded us. The demonstration that *argR* protein binds specifically to *cer* DNA within a region required for *cer* activity suggests that it has a direct role in *cer* recombination, though almost certainly not as the recombinase. The *argR* protein sequence bears none of the hallmarks of a recombinase and it binds *cer* some 200 bp away from the crossover region. We think it likely that its bindings to the *cer* ARG box is necessary for the organization of DNA and protein into a highly organized synaptic complex which is necessary for the observed directional intramolecular recombination. This role would therefore be similar to that of the other site-specific recombination accessory factors IHF, FIS, (factor for inversion stimulation; Kahmann *et al.*, 1985) and Tn*3* resolvase, in organizing and providing specificity to other synaptic complexes (Echols, 1986; Boocock *et al.*, 1987; Thompson *et al.*, 1987; Benjamin and Cozzarelli, 1988). In support of this is the observation by D.Summers (personal communication) that certain *cer* derivatives altered in the crossover region can recombine in the absence of *argR* and *xerB* function. Such mutant sites can also dispense with the region to which *argR* protein binds. However, directionality is lost, with intermolecular recombination and intramolecular recombination occurring at comparable frequencies. Binding of *argR* protein to *cer* clearly results in DNA bending (Figure 4) and the region of DNA immediately 5′ of the ARG box is known to have a sequence motif that is likely to promote binding after specific protein binding (Satchwell *et al.*, 1986; Summers and Sherratt, 1988).

The significance of the promoter within the ARG box of *cer* remains unclear. Transcription from the promoter occurs *in vivo*, yet no significant coding sequence within *cer* can be expressed from the RNA. Site-directed mutagenesis of this promoter that led to an ~20-fold reduction in transcription *in vivo* (to an almost undetectable level) had no discernible effect on the recombination activity of the mutant *cer* site (Summers and Sherratt, 1988). Nevertheless, it seems likely that the promoter has some biological function, and that arginine repressor binding to it may modulate that function. Again the significance of the failure of a chromosomally encoded wild-type *argR* gene to repress

efficiently transcription *in vivo* from the promoter within the ARG box of *cer* remains unclear. The multicopy plasmid containing *cer* in these assays contains all of the *cer* sequences 5′ of the ARG box that are present in a wild-type *cer*. Similarly, chromosomally encoded repressor acts inefficiently on a plasmid-borne *argR* promoter. The fact that expression of arginine repressor is autoregulated, and that in cells containing either a multicopy cer^+ plasmid, or a multicopy $argR^+$ plasmid, the chromosomal *argF* and *argI* genes are regulated normally suggests some physical or topological compartmentalization between the multicopy plasmids and the chromosome. Notwithstanding this, we have direct evidence that arginine repressor can bind to *cer in vivo*, and we believe that this binding is central to the requirement for arginine repressor in *cer* monomerizing recombination.

What features of the arginine repressor and its interactions make it a suitable accessory factor for ColE1 *cer* recombination? Is it a particularly good protein at inducing DNA bending and the formation of higher order structures? Does its use allow plasmid multimeric state and heritable stability to be regulated by the status of arginine and polyamine biosynthesis in a cell? Could the arginine repressor be particularly good at synapsing distant DNA segments? If so, could this explain why the arginine biosynthetic genes are dispersed on the *E.coli* chromosome rather than organized as a single operon? Binding of arginine repressor to the different *arg* operators, followed by repressor−repressor interactions could organize the dispersed arginine genes into physical proximity within a cell, so that on derepression, the different arginine biosynthetic enzymes are synthesized in physical proximity to allow enzyme coupling. As yet we do not know the answers to these questions. The requirement of arginine for binding of arginine repressor to *cer* DNA is unlikely to be a problem for its role in *cer* recombination, since it has been estimated that under most physiological conditions, there is sufficient arginine present in a cell for the arginine regulon to be substantially repressed (e.g. see Glansdorff, 1987).

We also believe that the arginine repressor may have roles in other cellular processes since we have observed (our unpublished work) that arginine repressor can bind specifically to other DNA fragments that are not involved in either arginine metabolism or plasmid stability. This may compare with the diverse role of IHF, which appears to be involved in site-specific recombination, transposition, plasmid replication, conjugal transfer and transcriptional control of biosynthetic and catabolic operons (Friedman *et al.*, 1984; Leong *et al.*, 1985; Dempsey, 1987; Morisato and Kleckner, 1987; Prentki *et al.*, 1987; Stenzel *et al.*, 1987; Thompson *et al.*, 1987). Despite there being consensus IHF binding sequences close to the *cer*-like region of *argR* and within the crossover region of *cer*, we have not observed reduced recombination in IHF-deficient strains (our unpublished work). There are other precedents for such behaviour in systems known to involve IHF (Morisato and Kleckner, 1987; Prentki *et al.*, 1987) and it is possible that HU and similar proteins can at least partly substitute for IHF *in vivo*.

Since *xerA*/*argR* protein is unlikely to be the recombinase acting at *cer*, is the other chromosomal gene that we have isolated and characterized (*xerB*) the recombinase gene? Our experiments to date suggest not (unpublished); indeed it is likely that the *E.coli xerB* gene is equivalent to the *pepA* gene of *Salmonella typhimurium* which encodes aminopeptidase A (Miller and MacKinnon, 1974). Recently we have identified a third chromosomal locus (*xerC*) by a Tn*5* mutation which reduces but does not abolish monomerizing recombination at *cer*. This locus, which is physically distinct from *xerA* and *xerB* is the subject of current study.

Materials and methods

Bacterial strains

All strains for plasmid experiments were derivatives of *E.coli* K12 AB1157 (Bachman, 1972) DS903 is AB1157 *rec*F and DS941 is AB1157 *rec*F $lacI^q$ *lacZ*delM15.

Plasmids

pUC8 and pUC9 are described in Vieira and Messing (1982). pKO-p*cer* is a derivative of pK01 (McKenney *et al.*, 1981) containing the *cer* region from coordinates 3737−3831 (Summers and Sherratt, 1988). pKO-p*argR* contains 192 bp of the *argR* region from the *Sph*I site downstream to a *Dra*I site at coordinate 195 (Figure 2) cloned into the *galK* vector pK0500 which is a polylinker derivative of pK01. pCS348 is a Cm^R λ *dv*-based plasmid vector (pCB104) in which the *argR* gene is cloned into a polylinker expressed from p*lac* (Stirling *et al.*, 1988). pCS119 is the same vector containing a non-*argR* fragment of DNA in the polylinker. The fragment in this vector is a functional *xer*B gene whose product does not interact with ARG boxes (unpublished). pCS202 is a Cm^RTc^R λ *dv* based 2-*cer* reporter plasmid that deletes Tc^R on *cer*-mediated resolution (Stirling *et al.*, 1988) pAT223 is derived from pKK223-3 (Pharmacia), is pBR322 related, and contains a polylinker downstream of the p*tac* promoter. pGLW11 and pUL2033 are described in Smith *et al.* (1988) and are derived from pKK223-3. pKS492 and pKS493 are pUC18 derivatives containing the 280 bp *Hpa*II−*Taq*I *cer* fragment in either orientation in the polylinker *Acc*I site.

Bacterial growth media and conditions

L-broth (Kennedy, 1971) was used for routine growth, supplemented with agar and antibiotics where appropriate. Cultures for galactokinase assays were grown in minimal medium with appropriate supplements and 0.3% fructose as carbon source. Ornithine was used to supplement the *argE* deficiency of the AB1157 strains. Galactokinase assays and specific activities determinations were as described in McKenney *et al.* (1981). β-lactamase assays were determined on the same samples using the method of Perret (1954).

In vitro DNA manipulations

Conditions for enzymatic reactions were those described by the suppliers. Protocols were essentially as described in Maniatis *et al.* (1982). DNA sequencing was by the chain termination technique of Sanger *et al.* (1977) using M13mp vectors (Norrander *et al.* (1983). Exonuclease III (Henikoff, 1984) was used to generate deletions for sequencing.

Overexpression and partial purification of the xerA gene product

The *argR* gene product was overexpressed from plasmid pGS38, a pUC18-derived plasmid containing the 0.92 kb *Sph*I−*Acc*I fragment encoding a functional *argR* gene (Figure 2). This plasmid was transformed into strain DS941, and grown in 1 litre of L-broth containing 50 μg/ml ampicillin and 1 mM IPTG to an A_{550} of 1.5. Arginine repressor was then partially purified using the procedure of Lim *et al.* (1987) with the following modifications: after sonication, the crude lysate was treated with DNase for 30 min on ice, fractionated with ammonium sulphate and precipitated with 5 mM L-arginine overnight at 4°C. The arginine-precipitated pellet was resuspended in 10 mM Tris pH 7.5, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 100 mM NaCl, 50% glycerol. This fraction was used for subsequent DNA binding assays, had at least 50% of its total protein as arginine repressor. The amounts of *argR* protein indicated in Figures 4 and 5 assume 75% purity of this preparation and a hexameric native protein (Lim *et al.*, 1987).

Gel retardation assays

The binding of the *argR* gene product to DNA was assayed using a modification of the band competition assay (Fried and Crothers, 1983; Strauss and Varshavsky, 1984; Tolias and DuBow, 1985). 10^{-2}−5×10^{-2} pmol of ^{32}P-labelled DNA restriction fragment was incubated with 2.5 μg of sonicated calf thymus DNA and up to 4 pmol of arginine repressor in 36

μl of binding buffer (10 mM Tris−HCl, pH 7.5, 1 mM EDTA, 100 mM KCl, 1 mM L-arginine 25 mM NaCl, 10 mM 2-mercaptoethanol, 5% glycerol) for 15 min at 37°C. Bromophenol blue loading dye was then added to the reactions, which were then loaded immediately onto 5% polyacrylamide gels in TBE buffer containing 1 mM L-arginine. The gel was pre-electrophoresed at 7 V/cm for 30 min, after which the samples were loaded and electrophoresis was performed at room temperature for 2 h. The gel was dried and autoradiographed on Kodak XS-1 X-ray film with Dupont Cronex intensifying screens at −70°C.

DNaseI footprinting in vitro

The procedure used was essentially that of Lim *et al.* (1987) with the following modifications: protein was allowed to bind to end-labelled DNA ($\sim 10^{-2}$ pmol) in 40 μl of binding buffer containing 10 mM $MgCl_2$ for 15 min at room temperature, after which 4 U (16 ng) of FPLC-pure DNase I (Pharmacia) was added and the reaction was stopped after 2 min by the addition of 20 μl of stop solution (3 M ammonium acetate, 250 mM EDTA, 10 μg sonicated calf thymus DNA). The reaction was phenol extracted, ethanol precipitated and electrophoresed on 8% polyacrylamide—7 M urea sequencing gels (Maxam and Gilbert, 1980). As a reference, the G+A and T+C sequencing reactions were performed on the same radiolabelled DNA fragment and electrophoresed with the footprinting reactions. In the footprint shown in lanes a−f of Figure 5, the *cer*$^+$ DNA was a 216 bp *Hpa*II−*Mlu*I fragment end-labelled with ^{32}P at the *Mlu*I site.

Protection against DMS methylation in vivo

The procedure followed was a modification of that described by Martin *et al.* (1986) and Thompson *et al.* (1987). DS941 cells containing pKS493 and either pCB104 (vector), pCS119 (*argR*$^-$) or pCS348 (*argR*$^+$) were grown in L-broth with the appropriate antibiotic to mid-log phase. One ml of DMS was added to each 200 ml culture, and the culture was shaken vigorously for 1 min. The reaction was stopped by pouring the culture onto ice. The cells were harvested and plasmid DNA was extracted using procedure of Birnboim and Doly (1979). Plasmid DNA was then cleaved with *Pst*1 and *Mlu*1, and the 280 bp *cer* fragment from pKS493 was purified from 5% acrylamide gels (Maxam and Gilbert, 1980). This DNA was 3′ end-labelled at the *Mlu*1 site using ^{32}P and then cleaved at the methylated sites using the piperidine cleavage reaction (Maxam and Gilbert, 1980) and electrophoresed on 8% DNA sequencing gels.

Acknowledgements

We give our special thanks to Werner and Renata Maas, and Dongbin Lim for communicating unpublished reports on *argR* and for their continuing interest and encouragement. We also wish to thank David Summers for his collaboration and interest in the project, and our many colleagues in the Department of Genetics for providing ideas, encouragement, and technical support. Sue Halley helped prepare the manuscript. The work was supported by the MRC and SERC.

References

Austin,S., Ziese,M. and Sternberg,N. (1981) *Cell*, **25**, 729−736.
Bachman,B.J. (1972) *Bact. Rev.*, **36**, 525−557.
Benjamin,H.W. and Cozzarelli,N.R. (1988) *EMBO J.*, **1**, 1897−1905.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1515−1523.
Boocock,M.R., Brown,J.L. and Sherratt,D.J. (1987) In Kelly,T.J. and McMacken,R. (eds) *DNA Replication and Recombination*. Alan R.Liss, New York, pp. 703−718.
Dempsey,W.B. (1987) *J.Bacteriol.*, **169**, 4391−4392.
Dodd,H.M. and Bennett,P.M. (1986) *J. Gen. Microbiol.*, **132**, 1009−1020.
Echols,H. (1986) *Science*, **233**, 1050−1056.
Fried,M.G. and Crothers,D.M. (1983) *Nucleic Acids Res.*, **11**, 141−158.
Friedman,D.I. and Gottesman,M. (1983) In Hendrix,R.W., Roberts, J.W., Stahl,F.W. and Weisberg,R.A. (eds) *Lambda II*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp.21−52.
Kahmann,R., Rudt,F., Koch,C. and Mertens,G. (1985) *Cell*, **41**, 771−780.
Kennedy,C.K. (1971) *J. Bacteriol.*, **108**, 10−19.
Kohara,Y., Akiyama,K. and Isono,K. (1987) *Cell*, **50**, 495−508.
Lamond,A.I. and Travers,A.A. (1983) *Nature*, **305**, 248−250.
Glansdorff,N. (1987) In Ingraham,J.L., Low,B., Magasanik,B., Schaechter, M. and Umbarger,H.E. (eds) *Escherichia coli and Salmonella typhimurium*. ASM, Washington, D.C., pp. 321−344.
Lane,D., de Feyter,R., Kennedy,M., Phua-S.-H. and Semon,D. (1986) *Nucleic Acids Res.*, **14**, 9713−9728.
Leong,J.M., Nunes-Duby,S., Lesser,C.F., Youderian,P., Susskind,M.M. and Landy,A. (1985) *J. Biol. Chem.*, **260**, 4468−4477.
Lim,D., Oppenheim,J.D., Eckhardt,T. and Maas,W.K. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6697−6701.
McKenney,K., Shimatake,H., Court,D., Schmeissner,U., Brady,C. and Rosenberg,M. (1981) In Chirikjian,J.G. and Papa,T.S. (eds) *Gene Amplification and Analysis*. Elsevier, Amsterdam, Vol. II, pp. 383−415.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York.
Martin,K., Huo,L. and Schleif,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3654−3658.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499−560.
Miller,C.G. and MacKinnon,K. (1974) *J. Bacteriol.*, **120**, 355−363.
Morisato,D. and Kleckner,N. (1987) *Cell*, **51**, 101−111.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101−106.
Perret,C.J. (1954) *Nature*, **174**, 1012.
Prentki,P., Chandler,M. and Galas,D.J. (1987) *EMBO J.*, **6**, 2479−2487.
Sadowski,P. (1986) *J. Bacteriol.*, 341−347.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.
Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) *J. Mol. Biol.*, **191**, 659−675.
Smith,M.C.M., Czaplewski,L., North,A.K., Baumberg,S. and Stockley,P.G. (1988) *Mol. Microbiol.*, in press.
Stenzel,T.T., Patel,P. and Bastia,D. (1987) *Cell*, **50**, 709−717.
Stirling,C., Stewart,G. and Sherratt,D.J. (1988) *Mol. Gen. Genet.*, **214**, 80−84.
Strauss,F. and Varshavsky,A. (1984) *Cell*, **37**, 889−901.
Summers,D.K. and Sherratt,D.J. (1984) *Cell*, **36**, 1097−1103.
Summers,D.K. and Sherratt,D.J. (1988) *EMBO J.*, **7**, 851−858.
Summers,D., Yaish,S., Archer,J. and Sherratt,D. (1985) *Mol. Gen. Genet.*, **201**, 334−338.
Thompson,J.F., de Vargas,L.M., Koch,C., Kahmann,R. and Landy,A. (1987) *Cell*, **50**, 901−908.
Tolias,P.P. and DuBow,M.S. (1985) *EMBO J.*, **4**, 3031−3037.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259−268.

Received on August 25, 1988