Mechanism of autonomous control of the *Escherichia coli* F plasmid: purification and characterization of the *repE* gene product

Luke Masson* and Dan S.Ray

Department of Biology and Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

Received November 9, 1987; Accepted December 2, 1987

ABSTRACT

E protein, the 29 Kd *repE* gene product, is essential for the replication of the *Escherichia coli* F plasmid. The *repE* gene has been cloned and expressed in an inducible ATG-fusion vector, and the protein product has been purified to homogeneity. The purified E protein is present as a dimer in solution and binds specifically to both the 19-bp direct repeats (*incB*) found within the mini-F origin of replication *ori2* and the *repE* operator DNA. Examination of the amino acid sequence of E protein revealed a consensus sequence for DNA binding.

INTRODUCTION

The 94.5 Kb *Escherichia coli* F plasmid is a stringently maintained, low copy number plasmid (1-2 copies per cell) belonging to the FI incompatibility group. The F plasmid *repE* gene codes for the 29 kd E protein which is required for replication (1,2,3). In addition, recent genetic and biochemical evidence has confirmed a multifunctional role for this protein during F plasmid replication (4,5,6,7). Most low copy number plasmids seem to share a common mechanism whereby replication is regulated by the stringently controlled expression of an initiator protein (4,8,9,10). However, unlike plasmids belonging to the FII incompatibility group that control expression of their replication proteins through the use of inhibitory counter transcripts, it appears that E protein is regulated by binding tightly to its own operator (5, 6,7).

In addition to the autoregulation of its own expression, E protein is required for F plasmid replication (2,3) and was found to bind to the 19-bp direct repeats present in both *incB* and *incC* and to the *ori2* region (6,7). Although these regions share a common octanucleotide consensus sequence 5'-TTGTCACA-3', gel retardation experiments indicate a much higher affinity of E protein for its operator sequence than for the 19-bp direct repeat sequences (6). Also, the pattern of protection against DNAase I cleavage differs signi-

ificantly between these binding sites. E protein gives almost complete protection to a 30-bp sequence spanning the repE operator region whereas its binding to the four 19-bp repeats in incB shows a pattern of alternating reduced and enhanced cleavages (6).

The increased replication frequency of the F plasmid when excess E protein is supplied in trans indicates that the E protein is a rate limiting factor in the replication of the entire basic replicon (5). The low intracellular concentration of E protein (11) is first set by autorepression through binding of E protein to the repE operator. The level of E protein is regulated secondarily by titration of its binding to the 19-bp repeats of the incB and incC loci (12).

An understanding of the biochemical basis for the different interactions of E protein with these distinct sites is essential for the development of a more comprehensive model of the regulatory mechanisms of the F plasmid and requires the availability of purified E protein. We describe here the purification of E protein to homogeneity and the kinetic analysis of its binding to the repE operator.

MATERIALS AND METHODS

Bacterial strains, plasmids and media.

The repE gene was placed under the control of the S. typhimurium ara BAD promoter to create the inducible ATG-fusion vector pREPE5. This plasmid was propagated in the ara-deleted host strain Escherichia coli MC1061 which was grown and induced in mineral base media as described elsewhere (6).

Enzymes and Buffers.

All restriction enzymes used were obtained from Bethesda Research Laboratory. TEMG buffer is 20 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 10 mM 2-mercaptoethanol and 10% glycerol. KPEMG buffer is identical to TEMG except that the Tris buffer was replaced by 20 mM KPO₄ buffer pH 7.0. Binding buffer is 20 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 10 mM 2-mercaptoethanol and 50 mM KCl.

Polyacrylamide gel analyses

A polyacrylamide gel retardation assay modified from Fried and Crothers (13) was used to detect and quantify the DNA-binding activity of E protein as described previously (6). The standard binding reaction used throughout the E protein purification procedure consisted of 5 ng of end-labelled BglII/AvaI fragment (fig. 1), 200 ng of non-specific competitor DNA poly(dI-dC)·Poly(dI-dC) (Pharmacia), and protein fractions in a 15 μ l final volume in gel binding buffer. Protein concentrations were determined by the Bradford method

Figure 1. Schematic representation of the E protein binding sites within the BglII/AvaI fragment of the F plasmid. E protein binding sites determined by polyacrylamide gel retardation assays and DNAase I protection studies (6) are found within each of the four 19-bp direct repeats of *incB* and within the inverted repeats overlapping the *repE* operator. Also indicated is the *repE* promoter regions and the 5' terminus of the *repE* gene.

(14). The reactions were incubated 15 min. at 37°C after which 2 μ l of agarose loading buffer was added and electrophoresis carried out as described previously (6). Any variations from this standard reaction are described in the text. To monitor the extent and purity of E protein, a discontinuous SDS-Page system was used (15).

Purification of E protein.

E. coli MC1061 (pREPE5) cells were grown overnight at 37°C in mineral base medium containing 25 μ g/ml tetracycline-HCl and subsequently used to inoculate 10 liters of the same medium. l-arabinose (Cal-Biochem) was added to the culture to a final concentration of 1% (w/v) after reaching an O.D.₅₉₅ = 0.2. Induction was allowed to proceed 6-8 hrs after which the cells were harvested and resuspended in 50 mls TEMG buffer containing 0.45 M KCl. The cells were then lysed by two passages through a French pressure cell and cellular debris and unbroken cells removed by a low speed centrifugation step (SS-34 rotor, 10,000 x g, 10 min.). The supernatant was decanted and centrifuged at 150,000 x g for 1 hr. All centrifugation steps were carried out at 4°C. The resulting supernatant was passed through a diethylaminoethyl (DEAE)-Sephacel column (10 ml bed volume) and the flow-through collected and dialyzed against TEMG buffer containing 50 mM KCl. The dialyzate (75 mls) was loaded onto another DEAE-Sephacel column (1.5 x 15 cm) equilibrated in TEMG buffer containing 50 mM KCl, washed with one bed volume of the same buffer and the bound proteins eluted with a linear KCl gradient (50 mM-500 mM). The fractions displaying binding activity to the BglII/AvaI fragment (fig. 1) in the presence of the non-specific competitor DNA poly(dI-dC)·poly(dI-dC) were pooled. The pooled fractions (approx. 2 mls) were dialyzed in TEMG buffer containing 50 mM KCl and loaded onto a Mono Q HR 5/5 FPLC anion exchange column (Pharmacia). The column was eluted with a linear KCl gradient (50

TABLE 1
E Protein Purification Scheme

Fraction	Volume (ml)	Protein ¹ (mg)	Total Units ² (x 10 ⁻⁵)	Specific Activity (x10 ⁻⁴) units/mg protein	Purification
DEAE-Sephacel (high salt)	75	380.0	1.81	0.047	-
DEAE-Sephacel	2	5.37	1.67	3.13	66.6
Mono Q	2	2.96	1.32	4.47	93.2
Mono S	1	0.17	1.13	66.7	1398.0

1. As determined by the method of Bradford *et al* (14).
2. One unit of E protein activity = the concentration of protein required to alter the mobility of 50% of the end-labelled 350 bp BglII/AvaI DNA fragment of the F ori-2 region under the conditions outlined in Materials and Methods.

Figure 2. SDS-Page analysis of pooled column fractions from table 1. 10 liters of *E. coli* MC1061 cells, containing the pREPE5 plasmid were induced as described elsewhere (6) and the E protein purified as described in the text. Pooled fractions were electrophoresed through a 15% (w/v) SDS-Page gel and stained with Coomassie Brilliant Blue. Lanes a and g contain BioRad molecular weight markers: phosphorylase b, Mr 92,000; bovine serum albumin, Mr 68,000; ovalbumin, Mr 45,000; carbonic anhydrase, Mr 31,000; soybean trypsin inhibitor, Mr 20,000; lysozyme, Mr 14,000). Lanes b and c represent 10 μ g of crude soluble cellular extract before (lane b) and after (lane c) passage through DEAE-Sephacel in the presence of 0.45 M KCl to remove nucleic acids. Lanes d, e and f represent approximately 3 μ g of pooled protein fractions (see table 1) from gradient eluted DEAE-Sephacel, mono Q and mono S columns respectively.

Figure 3. Elution profile of purified E protein on Superose-12 FPLC sizing column. The Superose-12 column was pre-equilibrated with KPEMG buffer containing 100 mM KCl. Protein markers (20-30 μ g each) were combined and loaded in a 250 μ l volume onto the Superose-12 column (panel A.). The protein markers used were bovine serum albumin dimer (peak A, Mr 136,000), bovine serum albumin monomer (peak B, Mr 68,000), ovalbumin (peak C, Mr 45,000), chymotrypsinogen A (peak D, Mr 25,700) and lysozyme (peak E, Mr 14,000). B. E protein (25 μ g in 250 μ l KPEMG buffer plus 100 mM KCl) was chromatographed under the same conditions as in A.

mM-500 mM) and the fractions containing DNA binding activity (eluting between 180 and 260 mM KCl) were pooled. The appropriate fractions, after being assayed for DNA binding activity, were dialyzed in KPEMG buffer containing 50 mM KCl. The dialyze (approx. 2 mls) was loaded onto the Mono S HR 5/5 FPLC cation exchange column and eluted with a linear KCl gradient (50 mM-450 mM). The peak of binding activity (eluting at 350 mM KCl) was pooled and dialyzed against KPEMG containing 50 mM KCl with the glycerol level increased to 50% (v/v) and the pure protein stored at -20°C .

RESULTS

Purification and physical characterization of E protein.

The purification of E protein presented here is rapid (2-3 days) and results in approximately 0.17 mg of E protein from 10 liters of culture with

Figure 4. Binding of the Superose-12 peak fractions to the 19-bp direct repeat fragment or *repE* operator fragment. 12 μ l from each fraction surrounding the purified E protein peak shown in fig. 3 were added to reaction mixtures containing either 5×10^{-10} M end-labelled *Bgl*II/*Alu*I 19-bp direct repeat DNA (A.) or *Alu*I/*Ava*I *repE* operator DNA (B.) in the absence of poly(dI-dC)•(dI-dC) under standard binding conditions. A smaller volume (2 μ l) of E protein from fractions 26 and 27 was used when binding the operator fragment due to the higher affinity of E protein for this inverted repeat binding site. After incubation for 15 min at 37°C, the reaction mixtures were electrophoresed on a 6% polyacrylamide gel for 60 min. The gel was then dried and autoradiographed.

an overall yield of 62% (Table 1). Gradient elution of E protein from DEAE-Sephacel gives a 66-fold purification. Subsequent purification on a Mono Q anion exchange column gave only a small increase in the purification factor but was found to be necessary to remove a single high molecular weight protein contaminant (data not shown). In the further purification on a Mono S cation exchange column, most of the protein passed through the column in the flow-through fraction. Analysis of the single peak of A₂₈₀ absorbing material

Figure 5. Titration of purified E protein using end-labelled AluI/AvaI fragment containing the *repE* operator. A. Reaction mixtures consisted of 5×10^{-10} M of specific end-labeled fragment, binding buffer, and variable levels of E protein in a final volume of 15 μ l. The mixtures were incubated for 15 min at 37°C and subsequently electrophoresed through a 6% (w/v) polyacrylamide gel for 60 min. The gel was then dried onto Whatmann 3MM paper and autoradiographed. The unbound *repE* operator fragment is indicated by an arrow. B. The autoradiogram of the gel was quantitated by densitometry and the percentage of DNA bound was plotted as a function of the amount of E protein monomer.

revealed a single species of approximately 29 kDa on a SDS-polyacrylamide gel (fig. 2).

Specific binding of proteins to sites having two-fold rotational symmetry generally involves some multimeric form of the protein (16). Previous results demonstrated that E protein forms distinctly different complexes at the *incB* and *repE* operator sites (6). Binding of E protein at these two sites might possibly reflect binding by different forms of the E protein. In order to determine whether the purified E protein might exist as a higher oligomeric form, an aliquot of the mono S fraction of the E protein was analyzed by gel

Figure 6. Dissociation rate of purified E protein complexed with the AluI/AvaI repE operator fragment. A. The reaction mixtures consisted of 5×10^{-10} M end-labelled fragment, binding buffer, and 1.5×10^{-8} M of purified E protein dimer in a final volume of 180 μ l. After 15 min. incubation at 37°C, a 100-fold excess of unlabelled AluI/AvaI fragment was added. At the times indicated (including the time required for entry of the complex into the gel matrix), 15 μ l samples were removed and electrophoresed. Lane a represents a typical mixture but lacks added E protein. Lanes b-j represent samples taken at 1.5, 3.5, 5.5, 7.5, 13.5, 16.5, 19.5, 22.5 and 25.5 min, respectively. The unbound repE operator fragment is indicated by an arrow. B. The slower migrating band representing E protein complexed with DNA was quantitated and plotted as a function of time.

filtration on a Superose-12 HR 10/30 FPLC column (Pharmacia) in KPEMG buffer containing 100 mM KCl. Its elution pattern was compared with that of proteins of known molecular weights. Protein markers were run through the column either separately or combined as illustrated in figure 3. Purified E protein eluted from the column between the bovine serum albumin monomer (68 kDa) and ovalbumin (45 kDa) at a position consistent with a globular protein of

approximately 60 kDa. This result suggests that E protein is a dimer under these conditions. Examination of the eluted E protein fractions revealed that the dimeric form has the ability to bind to both the 19-bp direct repeats of incB and to the inverted repeats of the repE operator (fig. 4). Because of the greater affinity of E protein for the repE operator fragment than for the incB fragment (6) only 2 μ l of fractions 26 and 27 was used in the binding assays with the labeled operator fragment. The multiple bands observed when E protein binds to the fragment containing the 19-bp direct repeats are presumably due to the presence of multiple binding sites in the fragment.

Kinetic analysis of the binding of purified E protein to the repE operator.

Figure 5 shows a typical titration of the purified E protein against end-labelled AluI/AvaI repE operator DNA showing the altered mobility of the DNA-protein complex. The relative amount of the labelled operator fragment that was altered in mobility was determined by densitometry and plotted as a function of protein concentration (fig 5). This data was then used to calculate an apparent equilibrium binding constant of 1.9×10^{-10} M under the conditions used here (17). In our purified E protein preparation, the active protein concentration calculated by titration of E protein against known amounts of AluI/AvaI operator DNA fragment indicated that approximately 50% of the protein retained binding activity. The equilibrium constant was calculated assuming that the protein is a dimer and that only 50% of it is active in binding to the operator.

By using a slight modification of the gel binding assay, we were able to measure the dissociation rate constant for complexes of E protein bound to the repE operator DNA (AluI/AvaI fragment). Purified E protein was added to end-labelled repE operator fragment to the level where >95% of the input DNA was bound after reaching equilibrium. After a 15 min incubation, a 100-fold excess of unlabelled repE operator fragment was added to bind any free E protein. The addition of unlabelled specific DNA also allows for the preferential binding of E protein released from the labelled DNA, thus ensuring accurate dissociation measurements. Samples (15 μ l) were removed with time and electrophoresed. Figure 6 shows a typical dissociation experiment. The bands of unbound and complexed DNA appear skewed due to ongoing electrophoresis between sampling times. Dissociation of E protein from the repE operator appears to follow first-order kinetics with a half-time of approximately 22 min at 50 mM KCl. The dissociation rate constant for the repE operator calculated from the data in figure 6 is 5×10^{-4} s⁻¹ (18).

Figure 7. Comparison of the consensus amino acid sequence of the E protein with other known DNA binding proteins. The sequences have been aligned according to the conserved residues present in the α -2 and α -3 helices of cro. The boxes represent homologous amino acids between E protein and either cro or lexR proteins.

DISCUSSION

It has been shown for F plasmid (19,20), as well as for a large number of bacterial plasmids examined to date, that replication is modulated in a negative manner. This is usually accomplished through the regulation of an essential, positively acting factor required for plasmid replication (9). In the case of F plasmid, the E protein has been shown to be required for replication both at the primary (*ori1*) and at the secondary (*ori2*) F plasmid origins of replication. E protein appears to be present at a low intracellular concentration and is rate-limiting for F plasmid replication (4,5,11). Direct biochemical evidence of the binding of the *repE* gene product to its autogenous operator as well as to the 19-bp direct repeats comprising the incompatibility regions *incB* and *incC* has been presented recently (6,7). These results directly support the results of Tsutsui *et al.* (21) who have postulated a role for these *inc* regions in titrating E protein and thus controlling, along with the *repE* operator, the availability of this essential replication protein for initiating replication. The ability of the E protein to bind to multiple sites reflects its functional complexity and together with previous evidence that different protein-DNA complexes exist at these different sites, underscores the importance of the availability of purified E protein.

By cloning the *repE* gene under the control of the inducible *araB* promoter, E protein can be expressed upon the addition of arabinose to the culture. The assay of this protein throughout the purification scheme exploited the specific DNA binding properties of E protein, thus resulting in its rapid

purification. Gel filtration of the purified protein was consistent with it being a dimer in 0.1 M KCl. Neither monomers nor multimeric forms larger than the dimer were observed under these conditions. The dimer was found to bind to both the 19-bp repeats and to the repE operator. This result is of interest in view of the model proposed by Trawick and Kline (22) in which different forms of E protein are postulated to exist, one of which binds to the direct repeats and a different form of which may bind to the repE operator. Our results indicate that E protein is largely in a dimeric form which appears capable of binding to both the repE operator and to the direct repeats. Previous results (6) showing the absence of intermediate bands during gel retardation analysis of the AluI/AvaI repE operator fragment, is consistent with the direct binding of a dimeric form of the protein to the rotationally symmetric operator sequence.

DNA binding by Cro, lambda repressor and CAP proteins involves DNA contacts by two α -helices named α -2 and α -3 (or E and F in CAP) (16). Since these three proteins are structurally homologous within these two adjacent helices, it was predicted that other DNA-binding proteins would reflect this homology. Indeed a list of proteins sharing structural homology to these two helices have been compiled which include both the lac repressor and araC protein (16). Examination of the amino acid sequence of E protein shows a similar pattern of conserved residues and residue types suggesting that E protein also forms such a helix-turn-helix structure. At the amino-terminus of the E protein, between amino acids 64 to 84 (12), the highly conserved alanine residue is found at position 5 and glycine at position 9 (fig. 7). Of the DNA-binding proteins examined (16), E protein shares the most homology to the α -2 and α -3 helices of Cro and the LexR protein. These structural similarities as well as the observed binding to the repE operator are consistent with the proposed role of E protein in the autogenous regulation of the repE gene. A direct demonstration of the specific repression of the repE gene in an in vitro transcription system should now be possible.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Institutes of Health (AI10752). Equipment for these studies was provided in part by a grant from the California Biotechnology Research and Education Program. L.M. was the recipient of a Postdoctoral fellowship from the Medical Research Council of Canada.

*Present address: Genetic Engineering Section, Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada

REFERENCES

1. Watson, L.A., Phua, S.H., Berquist, P.L. and Lane H.E.D. (1982) *Gene* **19**, 173-178.
2. Maki, S., Kuribayashi, M., Miki, T. and Horiuchi, T. (1983) *Mol. Gen. Genet.* **191**, 231-237.
3. Tolun, A. and Helinski, D.R. (1982) *Mol. Gen. Genet.* **186**, 372-377.
4. Kline, B.C. (1985) *Plasmid* **14**, 1-16.
5. Rokeach, L.A., Sogaard-Andersen, L. and Molin, S. (1985) *J. Bacteriol.* **164**, 1262-1270.
6. Masson, L. and Ray, D.S. (1986) *Nucl. Acids Res.* **14**, 5693-5711.
7. Tokino, T., Murotsu, T. and Matsubara, K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4109-4113.
8. Manch-Citron, J.N., Gennaro, M.L., Majumder, S. and Novick, R.P. (1986) *Plasmid* **16**, 108-115.
9. Scott, J.R. (1984) *Microbiol. Rev.* **48**, 1-23.
10. Riise, E. and Molin, S. (1986) *Plasmid* **15**, 163-171.
11. Komai, N., Nishizawa, T., Hayakawa, Y., Murotsu, T. and Matsubara, K. (1982) *Mol. Gen. Genet.* **186**, 193-203.
12. Murotsu, T., Matsubara, K., Sugisaki, H. and Takanami, M. (1981) *Gene* **15**, 257-271.
13. Fried, M. and Crothers, D.M. (1981) *Nucl. Acids. Res.* **9**, 6505-6525.
14. Bradford, M. (1976) *Anal. Biochem.* **72**, 248-254.
15. Laemmli, U.K. (1970) *Nature (London)* **227**, 680-685.
16. Pabo, C.O. and Sauer, R.T. (1984) *Ann. Rev. Biochem.* **53**, 293-322.
17. Riggs, A.D., Suzuki, H. and Bourgeois, S. (1970) *J. Mol. Biol.* **48**, 67-83.
18. Riggs, A.D., Bourgeois, S. and Cohn, M. (1970) *J. Mol. Biol.* **53**, 401-417.
19. Tsutsui, H. and Matsubara, K. (1981) *J. Bacteriol.* **155**, 337-344.
20. Kline, B.C. and Seelke, R. (1982) *Mol. Gen. Genet.* **187**, 218-224.
21. Tsutsui, H., Fujiyama, A., Murotsu, T. and Matsubara, K. (1983) *J. Bacteriol.* **155**, 334-337.
22. Trawick, J.D. and Kline, B.C. (1985) *Plasmid* **13**, 59-69.