Gene F of plasmid RSF1010 codes for a low-molecular-weight repressor protein that autoregulates expression of the *rep*AC operon

Stefan Maeser, Peter Scholz, Sabine Otto and Eberhard Scherzinger*
Max-Planck-Institut für Molekulare Genetik, Abteilung Schuster, Ihnestrasse 73, D-1000 Berlin 33, FRG

Received August 15, 1990; Accepted September 27, 1990

ABSTRACT

The *rep*AC operon of plasmid RSF1010 consists of the genes for proteins E, F, RepA (DNA helicase), and RepC (origin-binding initiator protein) and is transcriptionally initiated by a promoter called P_4. We have studied the expression of the *rep*AC operon *in vivo* by using fusions to the *lac*Z reporter gene. The results show that the product of the second gene, F, autoregulates the operon by inhibiting transcription from P_4. To verify its properties postulated from the *in vivo* studies and to initiate its biochemical characterization, we have purified the F protein from an overproducing *E.coli* strain constructed *in vitro*. Purification was based on a gel retardation assay for detection of P_4-specific DNA binding. Subsequent DNase footprinting of the F binding sites showed clear protection around two partially symmetric P_4 sequences of 16 bp, each of which matches the symmetric consensus sequence, GCGTGAGTACTCACGC, in at least 13 positions. The native repressor, as judged from gel filtration, velocity sedimentation and crosslinking studies, exists as a dimer in dilute solution; its monomeric subunit, as predicted from DNA sequence and N-terminal protein sequence data, consists of 68 amino acids and has a calculated M_r = 7,673.

INTRODUCTION

RSF1010 is a fully sequenced Sm^r Su^r broad-host-range plasmid that is 8,684 bp in size (1). It is identical, or at least very similar, to R1162, R300B, and to several other incompatibility group Q plasmids isolated from diverse Gram-negative bacterial hosts (for a review see ref. 2). In *E.coli*, where RSF1010 is maintained at 10–12 copies per chromosome equivalent (2, 3), replication proceeds either bi- or undirectionally from a 395-bp origin region (*ori*V) (4, 5) and depends on at least three plasmid-determined proteins, the products of genes *rep*A, *rep*B' and *rep*C (5, 6). The RepC protein binds specifically to the 20-bp direct repeats of *ori*V and acts as a positive replication factor: upon induction of RepC synthesis from a cloned *rep*C-carrying fragment, the number of RSF1010 copies in the cell increases (7). RepA and RepB' (which is identical to the C-terminal half of the bifunctional *Mob*A/*Rep*B protein) provide IncQ-specific functions of a DNA helicase and a DNA primase, respectively (1, 5), and make RSF1010 independent of the bacterial *dna* gene products B, C and G (8). *E.coli* RNA polymerase and the DnaA initiator protein are also dispensable for replication of the RSF1010 plasmid (ref. 5 and unpublished data).

The strategy adopted by RSF1010 to regulate expression of its essential *rep* genes, and hence to control its replication, has not been fully elucidated. Two overlapping tandem promoters, P_1 and P_3, capable of directing transcription of all 3 *rep* genes, have been located by S1 mapping of *in vivo* RNA in the intercistronic region between the divergently transcribed genes *mob*A/*rep*B and *mob*C (9) (Fig. 1). Immediately downstream to P_1/P_3 is the origin of transfer (*ori*T) site, and from the RNA patterns observed for mob^+ and mob^- plasmids, Derbyshire et al. (9) concluded that transcription from at least P_3 is repressed by the concerted binding of proteins MobA/RepB and MobC to *ori*T. Consistent with this, Bagdasarian et al. (3) reported that deletions or insertions affecting either the *ori*T site or the 5' one-third of *mob*A/*rep*B caused an increase in the plasmid copy number. They also observed that such mutant derivatives of RSF1010 had lost the broad host range capability.

Another potential target for RSF1010 replication control is the promoter P_4 (Fig. 1). Identified originally as an RNA polymerase binding site near the *Acc*I site at nt 5470 (10), its position was confirmed later by sequence analysis. It is located just upstream of the E and F coding frames that precede *rep*A and that have been identified in the course of DNA sequencing by overproducing the respective proteins and determining part of their amino acid sequences (1). A biochemical function for these smallest known RSF1010 proteins (70 and 68 amino acids, respectively), however, has not yet been demonstrated.

Here we report our initial studies on the gene F protein. We constructed P_4/*lac*Z transcriptional fusions and found that synthesis of β-galactosidase was inhibited whenever gene F was present *in cis* or *in trans*. Then, using a fragment retention assay for detection of P_4-specific DNA binding, we purified the F protein from *E.coli* cells overproducing both E and F. Subsequent DNA cleavage protection experiments using DNaseI showed that the F-binding sequences are located within the binding region of RNA polymerase, as expected for a classical repressor of the

* To whom correspondence should be addressed

initiation of transcription. Although not examined extensively, our experiments performed *in vivo* and *in vitro* indicate that the gene E protein does not bind to the P_4 promoter/operator region, either in the presence or absence of the F repressor.

MATERIALS AND METHODS

Enzymes and biochemicals

Restriction endonucleases, T4 polynucleotide kinase, T4 DNA ligase, T4 DNA polymerase, DNA polymerase I Klenow fragment (PolIk), *E.coli* RNA polymerase holoenzyme, alkaline phosphatase (from calf intestine), DNase I (from bovine pancreas), and protein M_r standards were from Boehringer-Mannheim, New England Biolabs, or Pharmacia. The M13 universal (17-mer) sequencing primer, deoxy- and dideoxyribonucleoside triphosphates, and [γ-^{32}P] ATP were from Amersham Corp.. Isopropyl-β-D-thiogalactopyranoside (IPTG) was from Sigma and 1,3-butadiene diepoxide from Merck. All other chemicals were of the highest commercial purity available.

Bacterial strains, plasmids, and phage

The *E.coli* K-12 strains used in this study were CB454 (Δ *lac*Z, *lac*Y$^+$, *gal*K, *rec*A56) (11), used to construct and maintain *lac*Z fusion plasmids; HB101 (pVH1) = HB101 (12) harboring the ColD-based *lacI*^{q} plasmid pVH1 (7), used to construct and maintain an F repressor-overproducing plasmid; and JM101 (13), used to propagate M13mp9 and its recombinants.

Fig.1 shows the genetic structure of plasmid RSF1010 and its restriction sites relevant to the construction of the following recombinant plasmids and phages. General procedures for DNA purification, restriction, ligation and transformation were as described in Sambrook et al. (14). All RSF1010 fragments with 5′ or 3′ extensions were made blunt-ended prior to their insertion into *Hinc*II- or *Sma*I-cut vector DNA using the DNA polymerizing and 3′ exonuclease activities of T4 DNA polymerase (14).

pOT10, pOT11, and pOT12 were obtained by ligating the P_4 promoter-carrying fragments *Eco*RV (4477)-to-*Acc*I (5473), *Eco*RV (4477)-to-*Sfa*NI (5743), and *Eco*RV (4477)-to-*Ssp*I (5913) of RSF1010, respectively, into the unique *Sma*I site of pCB302a, a pBR322-derived promoter probe vector (11); the insert orientation in all three constructs is such that transcription from P_4 is directed towards the promoterless *lac*Z reporter gene of pCB302a.

pSO21 and pSO22 were obtained by ligating the F gene-carrying fragments ScaI (4517)-to-*Ssp*I (5913) and *Acc*I (5472)-to-*Ssp*I (5913) of RSF1010, respectively, into the unique *Hinc*II site of pACYC177 (15); the insert orientation in both constructs is such that gene F is transcribed from the β-lactamase promoter of pACYC177. The constructs were used to transform CB454(pOT10).

pSM25 was obtained by ligating the *Sca*I (5417)-to-*Ssp*I (5913) fragment of RSF1010 into the unique *Sma*I site of pKK223-3, a pBR322-derived expression vector (16); the insert orientation is such that transcription from the vector-borne *tac* promoter is directed towards genes E+F of RSF1010.

mSM1 and mSM2 were obtained by ligating fragments *Acc*I (3569)-to-*Acc*I (5473) and *Alu*I (5314)-to-*Alu*I (6485) of RSF1010, respectively, into the unique *Hinc*II site of M13mp 9 (13); the insert orientation is such that the M13 viral strand of mSM1 and mSM2 is contiguous with the RSF1010 l-strand and r-strand, respectively.

β-Galactosidase assay

Galactosidase activity was determined as described (17) in cells permeabilized with toluene.

DNA binding assay

Binding of F protein to operator DNA was assayed by a gel electrophoresis shift method (18). RSF1010 DNA was restricted to completion with *Acc*I + *Ava*I and the resulting fragments (A-G, Fig.6) were end-labeled by using T4 kinase and [γ-^{32}P] ATP, after treatment with calf intestinal alkaline phosphatase (14). The labeled DNA was extracted with phenol and precipitated with ethanol. To 0.12 μg of this DNA (10−20,000 acid-insoluble cpm) in 20 μl of binding buffer (20 mM Tris·HCl, pH7.6/50 mM NaCl/1 mM EDTA/0.1% Brij 58/50 lg BSA per ml) were added various amounts of F protein and the mixtures were incubated for 20 min at 37°C. Five μl of 20% (w/v) Ficoll 400/0.1% bromophenol blue in electrophoresis buffer (40 mM Tris acetate/5 mM Na acetate/1 mM EDTA, pH 7.9) were then added and the mixtures were immediately loaded onto a vertical 1.4% agarose gel. Following electrophoresis (3h, 21°C, 6V/cm), the gel was dried and autoradiographed. To quantitate binding, dried gel segments were cut out for liquid scintillation counting. One unit of binding activity is defined as the quantity sufficient to cause a mobility shift on 50% of the input fragment F molecules (size, 439 bp).

SDS/urea gel electrophoresis

Denaturing gel electrophoresis of proteins was carried out on a 10-cm resolving gel (1.5 mm thick) of 15% polyacrylamide (acrylamide: bis, 15:0.4) containing 0.1 M $NaPO_4$ (pH 7.2), 6 M urea, and 0.1% SDS. A 7.5 % stacking gel (2−3 cm height) was formed on top of the resolving gel using the same buffer conditions. Samples were prepared for electrophoresis by adding an equal volume of 2×sample buffer (20 mM $NaPO_4$, pH 7.2 / 10 M urea / 10% (v/v) mercaptoethanol / 2% SDS / 0,2% bromophenol blue) and heating at 95°C for 3 min. Electrophoresis was at 120 V for 3−3.5 h using a buffer containing 0.1 M $NaPO_4$ (pH 7.2) and 0.1% SDS. Proteins were visualized by staining with Commassie blue R-250, after washing the gel with 20% (v/v) methanol / 7% (v/v) acetic acid. The relative content of protein in Coomassie-stained gel bands was quantified by scanning the gel on the LKB laser densitometer Ultroscan XL (633 nm).

Protein purification

Cultures (8×1.2 liter in 5-liter flasks) of HB101 (pSM25, pVH1) were grown in a shaking water bath at 37°C in TY broth (1% Tryptone / 0.5% yeast extract / 0.5% NaCl) containing 40 mM Mops·KOH (pH 7.9), 0.2% glucose, and 20 μg/ml each of thiamine·HCl and ampicillin. At A_{600} = 0.5, IPTG was added to 0.33 mM, and shaking of the cells was continued for 4.5 h (A_{600} = 2.5). The cells (42 g) were harvested at room temperature, washed with 2 liters of 40 mM Tris·HCl, pH 8.0/0.1 M NaCl at 0°C, resuspended in 42 ml of 20 mM Tris·HCl, pH 8.0/0.1 M NaCl/1 mM DTT, and frozen in liquid N_2.

Cells (78 ml) were thawed at 10°C. Lysozyme and EDTA were added to 0.3 mg/ml and 1 mM, respectively. After 45 min at 0 °C, the cell suspension was mixed with an equal volume of

20 mM Tris·HCl, pH 8.0/0.3 M NaCl/1 mM DTT/0.5% Brij 58, and gently stirred in a 37°C-water bath until the temperature reached 20°C. The resulting lysate was centrifuged at 30.000 rpm for 90 min in a Beckman 45 Ti rotor and the supernatant was collected (fraction I, 130 ml). This and all following operations were carried out at 0–8°C.

A column of heparin-Sepharose CL-6B (5 $cm^2 \times 8$ cm) was equilibrated with buffer A (20 mM Tris·HCl, pH 8.0/1 mM DTT/0.1 mM EDTA/ 10% (v/v) glycerol) containing 0.1 M NaCl. Fraction I was diluted with 130 ml of 20 mM Tris·HCl, pH 8.0/1 mM DTT/20% (v/v) glycerol, and applied to the column at a flow rate of 20 ml/h. The column was washed with 60 ml of buffer A + 0.1 M NaCl, and bound proteins were eluted at 40 ml/h with a 400-ml linear gradient from 0.1 to 1 M NaCl in buffer A. Protein F was identified in fractions containing 0.20–0.35 M NaCl (fraction II, 60 ml).

A column of hydroxylapatite Bio-Gel HT (5 $cm^2 \times 4$ cm) was equilibrated with buffer A + 0.25 M NaCl. Fraction II was applied to the column at 10 ml/h. The column was washed with 10 ml of buffer A + 0.25 M NaCl, then with 15 ml of buffer B (10 mM KPO_4, pH 6.8/0.25 M NaCl/1 mM DTT/0.1 mM EDTA/10% glycerol), followed by a 200-ml gradient of 10 to 300 mM KPO_4, pH 6.8 in buffer B. Protein F was identified in fractions containing 50–100 mM phosphate (fraction III, 30 ml).

A column of CM-Sepharose Cl-6B (1.8 $cm^2 \times 7$ cm) was equilibrated with buffer C (20 mM KPO_4, pH 6.8/1 mM DTT/0.1 mM EDTA/10% glycerol) containing 50 mM KCl. Fraction III was dialyzed against the same buffer and applied to the column at 12 ml/h. The column was washed with 12 ml of buffer C + 50 mM KCl, and bound proteins were eluted with a 120-ml gradient from 50 to 600 mM KCl in buffer C. Protein F was identified in fractions containing 110–200 mM KCl (fraction IV, 18 ml).

A column of DEAE-Sephacel (0.6 $cm^2 \times 10$ cm) was equilibrated with buffer A + 25 mM NaCl. Fraction IV was dialyzed against the same buffer and applied to the column at 18 ml/h. The column was washed with 6 ml of buffer A + 25 mM NaCl, and bound proteins were eluted at 6 ml/h with a 60-ml gradient from 25 to 400 mM NaCl in buffer A. Protein F eluted between 90 and 140 mM NaCl. Fractions showing a single band of 7.2 kDa in a Coomassie-stained SDS/urea polyacrylamide gel (see Fig. 2) were pooled (fraction V, 8 ml) and stored at −70°C. Fraction V was used for all further studies described in this paper.

DNA sequencing and DNase protection

The complementary strands of the mSM1 and mSM2 viral DNAs, containing the RSF1010 *rep*B/E intergenic region and adjacent sequences of different lengths, were sequenced by the dideoxy chain-termination method (19) using PolIk and the M13 17-mer primer that had been labeled at the 5′ end with ^{32}P. The RSF1010 r-strand and l-strand sequences determined with these recombinants (nt 5473–5360 and 5314–5450, respectively) confirmed those obtained by Scholz et al. (1). To prepare a 5′-mono-labeled, duplex DNA substrate suitable for footprinting analysis, the 5′-[^{32}P] 17-mer primer, annealed to either mSM1 or mSM2 single-stranded DNA, was extended with PolIk in the absence of a chain terminator. The average extension length as estimated from the mobility of the ^{32}P-labeled products in an alkaline agarose gel was ≈ 70% of the template length. DNase protection experiments with these partially duplex DNA substrates were carried out under conditions essentially as described by Johnson et al. (20). To 25-μl reaction volumes containing approx. 0.1 pmol of mSM1 or mSM2 circular molecules various amounts of F repressor or *E.coli* RNA polymerase were added. After incubation for 15 min at 30 °C, DNase I was added to 25 μg/ml and the mixtures were incubated for an additional 5 min. Following DNase I cleavage, reaction products were resolved in a 6% polyacrylamide gel containing 8 M urea and visualized by autoradiography.

Other methods

Protein was determined by a dye binding method (21), using bovine serum albumin as a standard. All pH measurements were made at room temperature at a buffer concentration of 0.2 M.

RESULTS

Identification of the gene F protein as a repressor of transcription

In the known nucleotide sequence of RSF1010 (1), the reading frames for proteins E and F are assigned to nt 5440–5652 and 5654–5860, respectively, and the nearest upstream promoter, P_4, to the 60-bp long intergenic region between *mob*A/*rep*B and E (Fig. 1). To explore a possible role of these small proteins in the regulation of transcription initiating at P_4, RSF1010 fragments extending from a midpoint in *mob*A/*rep*B, the *Eco*RV site at nt 4477, to nt 5473 (within gene E), 5743 (within gene

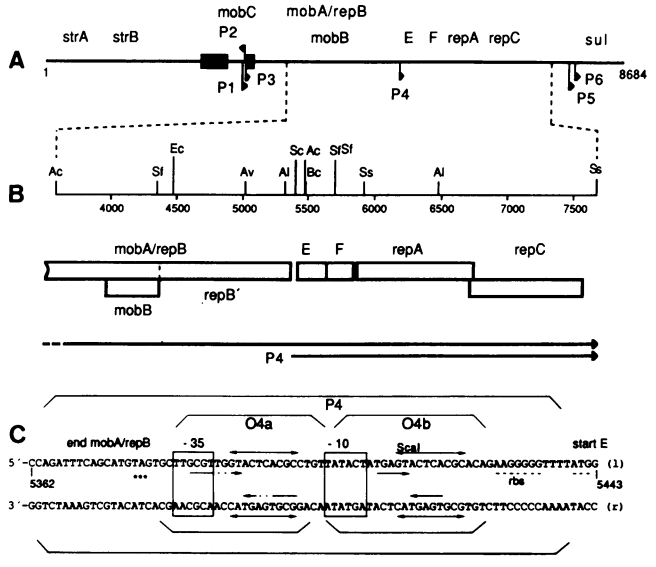

Figure 1. Map of RSF1010. (A) The gene sequence of the whole genome (8684 bp, ref. 1) is shown above the line. The positions of the six known promoters (P_{1-6}), with the direction of transcription, are indicated by arrows. The large and small black areas indicate the position of the *ori*V and *ori*T region, respectively. (B) Enlargement of the 3567–7678 bp region. Restriction sites relevant to this study are indicated above the line by the code: Ac, *Acc*I; Al, *Alu*I; Av, *Ava*I; Bc, *Bcl*I; Ec, *Eco*RV; Sf, *Sfa*NI; and Ss, *Ssp*I. Positions of the genes are indicated by boxes below the line. The horizontal arrows indicate the direction and probable extent of the transcripts initiated at $P_{1/3}$ and P_4. (C) Enlargement of the control region between nt 5362 and 5443, with the stop codon of *mob*A/*rep*B marked by asterisks and the ribosome binding site (rbs) and start codon of gene E indicated by dashed lines. The putative −10 (Pribnow box) and −35 regions of the P_4 promoter are boxed. Double arrows indicate the locations of two identical 10-bp direct repeats, and the pairs of facing arrows indicate palindromic sequences. The binding regions of *E.coli* RNA polymerase and F repressor, as determined by DNase I protection studies on both strands of the DNA, are outlined by brackets (1 base at each border) and labeled P_4 and O_{4a}/O_{4b}, respectively.

F), or 5913 (within *rep*A) were cloned upstream of *lacZ* in the promoter-probe vector pCB302a, and the level of β-galactosidase resulting from the presence of each of these P_4/*lacZ* fusion plasmids (pOT10, pOT11 and pOT12, respectively) in the *lacZ*$^-$ strain CB454 was determined. The β-gal level in cells with pOT12, which contains the full P_4 promoter region as well as E^+ and F^+, was found to be one order of magnitude lower than that of cells harboring the E^+ F^- plasmid pOT11 or the E^- F^- plasmid pOT10 (Table 1). This suggested to us that the P_4 promoter is indeed subject to autorepression, either by protein F on its own or by F in conjunction with protein E. It was also possible that a hitherto unrecognized transcription terminator, located elsewhere between genes E and *rep*A, was responsible for the decreased β-gal level obtained with the pOT12 plasmid.

To distinguish between these possibilities, we constructed plasmids pSO21 (pACYC177 with an RSF1010 *ScaI* -to- *SspI* insert carrying both E^+ and F^+ but lacking the P_4 promoter, Fig. 1) and pSO22 (pACYC177 with an RSF1010 *AccI* -to- *SspI* insert carrying only F^+) and introduced them singly into cells harboring the P_4/*lacZ* construct pOT10. Estimation of plasmid DNA levels in these strains revealed no significant variation in the pOT10 copy number. Galactosidase assays (Table 1) showed that while pACYC177, the vector plasmid, had no discernible effect on the β-gal level, the presence of either pSO21 or pSO22 resulted in a reduction in β-gal expression to the background level. Hence, the gene F product by itself acts as a repressor of the promoter P_4.

Overproduction and purification of protein F

Plasmid pSM25 is a pKK223−3-derived recombinant in which the RSF1010 E-F region without the P_4 promoter has been put under the inducible control of the *tac* promoter. When cultures of *E.coli* HB101 harboring both pSM25 and a *lacI*-expressing compatible plasmid (pVH1) are induced with IPTG, two low-molecular-weight proteins migrating in a 15% polyacrylamide gel containing 6 M urea and 0.1% SDS at approximately M_r 5,000 and 7,200 accumulate over a 4−5 hr period (data not shown). Both overproduced proteins are soluble in extracts, representing ≈ 14 and 4% of the cellular protein, respectively (Fig. 2, lane I). The M_r 7,200 polypeptide was identified as the gene F product by further purification (see below) and determination of the sequence of 18 amino acids at its N-terminus (1). Similarly, the M_r 5,000 polypeptide was identified as the gene E product by further purification (unpubl. procedure) and determination of the 5 last amino acid residues released by carboxypeptidase P digestion (1). In both cases, the amino acid sequence determined experimentally precisely matched that predicted from the DNA sequence. However, while for F the predicted M_r of the encoded polypeptide is in reasonable agreement with that determined by SDS/urea-PAGE (see below), the predicted and measured M_r's for E differ considerably (7,563 *versus* 5,000); the reason for this discrepancy is not known.

The procedure used for purification of protein F is summarized in Table 2. Purification was monitored by the gel retardation assay described under Materials and Methods (see also Fig. 6) as well as by tracing the co-overproduced M_r 5,000 and 7,200 polypeptides electrophoretically. Fig. 2 shows that the procedure will only select the M_r 7,200 species which after the final DEAE-Sephacel step was estimated to be > 95% pure. Approximately 1 mg of pure repressor was obtained per gram (wet weight) of induced cells. The major purification was achieved during chromatography on heparin-Sepharose, where F, typical for a DNA binding protein, binds quite tightly, while the bulk of the cellular protein including the RSF1010 E protein elutes in the flowthrough fraction. Unlike the case of F, no interaction of protein E with RSF1010 DNA could be detected under a variety of conditions.

Physical properties of F protein

The subunit M_r of protein F calculated from the nucleotide sequence and N-terminal amino acid sequence data is 7,673 (1), close to the 7,200 value estimated from its migration in a

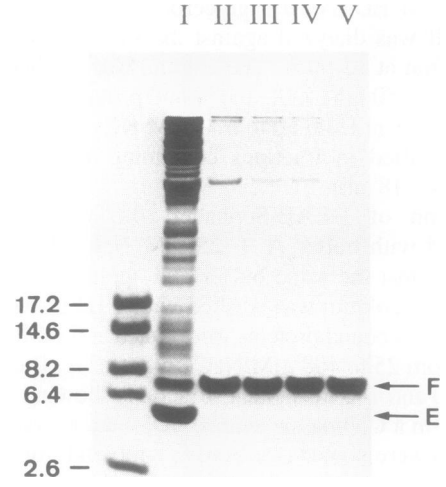

Figure 2. Purification of protein F. F protein fractions I through V, generated during its purification (Table 2), were analyzed by 15% SDS/urea-PAGE. The amount of total protein applied to each slot was: I (57 lg), II (8 lg), III (7 lg), IV (6.5 lg), and V (6.1 μg). The marker lane at the left contains a mixture of cyanogen bromide peptides from sperm-whale myoglobin whose M_r's are given $\times 10^{-3}$.

Table 1. Effect of gene F product on expression of *lacZ* fused to the P_4 promoter.

lacZ plasmid	Coresident plasmid	Average β-gal units[a]
pCB302a	–	< 1
pOT 10	–	505
pOT 11	–	473
pOT 12	–	26
pOT 10	pACYC177	489
pOT 10	pSO21	< 1
pOT 10	pSO22	< 1

[a] β-Galactosidase activity was measured three times at three different bacterial growth stages, and the average value is given in Miller units.

Table 2. Purification of F protein

Fraction	Step	Total protein	Total activity	% activity recovered
		mg	units $\times 10^{-6}$	
I	Crude extract[a]	2.940	694	
II	Heparin-Sepharose	85	444	64
III	Hydroxyapatite	57	367	53
IV	CM-Sepharose	47	307	44
V	DEAE-Sephacel	39	260	38

[a] From 42 g of IPTG-induced HB101 (pSM25, pVH1) cells

denaturing SDS/urea gel relative to marker polypeptides of known size (Fig.2).

In an analytical gel filtration on Sephadex G-100 (Fig. 3), F protein eluted as a symmetric peak at about the same position as chymotrypsinogen A (M_r 25,000), suggesting that the native protein is an oligomer of at least two subunits. The Stokes radius was estimated to be 21Å (Fig. 3, *insert*).

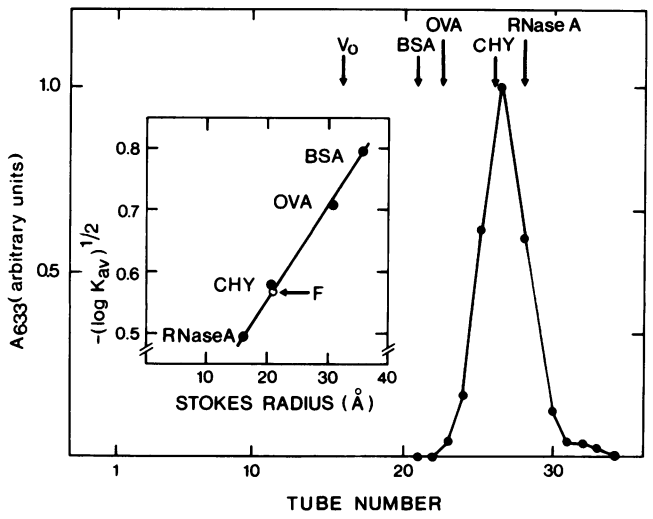

Figure 3. Sephadex G-100 filtration of protein F and estimation of the Stokes radius. Purified F protein (0.5 mg) in 0.75 ml of a buffer containing 20 mM Tris·HCl (pH 7.6), 0.1 M NaCl, 1 mM DTT, and 5% (v/v) ethylene glycol was filtered at 14 ml/h through a 134-ml G-100 (superfine) column (1.5×76 cm) which had been equilibrated in the same buffer at 4°C. Fractions of 3.6 ml were collected, and an aliquot (0.15 ml) of each included fraction was subjected to SDS/urea-PAGE. The gel was stained with Coomassie blue, and the relative content of F protein in the peak fractions was quantified by scanning laser densitometry. *Insert*, the column was calibrated with protein markers to determine the Stokes radius according to Siegel and Monty (22). Abbreviations used are: BSA, bovine serum albumin, OVA, ovalbumin; CHY, chymotrypsinogen A; V_0, void volume (measured with blue dextran).

Velocity sedimentation in a 10–30% glycerol gradient yielded a sedimentation coefficient ($s_{20,\omega}$) of ≈ 2.1 S (Fig. 4). The partial specific volume calculated from the predicted amino acid composition (1) was 0.72 ml/g. The M_r of the native F protein calculated from the Stokes radius, $s_{20,\omega}$, and apparent partial specific volume was 17,000, close to the 15,300 value expected for a dimer. The frictional coefficient (f/f_0) was calculated to be 1.22 from the equation: $f/f_0 = a/(3\ \upsilon\ M_r/4\ \pi\ N)^{1/3}$, in which a = Stokes radius, υ = partial specific volume, and N = Avogadro's number (22).

The dimeric structure of the native repressor as predicted from the above data was essentially confirmed by cross-linking studies. In the experiment of Fig. 5, F protein at concentrations ranging from 0.5 to 4 mg/ml was reacted with 1,3-butadiene diepoxide and the resulting products were analyzed by SDS/urea-PAGE. Both at low and high protein concentrations, the predominant reaction product was a cross-linked dimer. At F protein concentrations of ≥ 1 mg/ml (lanes b–e), two minor bands corresponding to molecular weights of the tri- and tetramer were also observed. In a control experiment using aprotinin (lanes f–j), the cross-linking reaction resulted only in the generation of modified monomers. Thus, a major contribution of unspecific intermolecular reactions to the results obtained with protein F can be ruled out.

When purified F protein (Fract. V) was dialyzed to remove the dithiothreitol, virtually all of the protein migrated in an SDS/urea gel at the position of the dimer unless it was reduced prior to electrophoresis by boiling in the presence of 5% mercaptoethanol and 1% SDS (data not shown). Apparently, in the nonreduced F protein the monomeric subunits are reversibly cross-linked via a disulfide bridge. The DNA sequence predicts

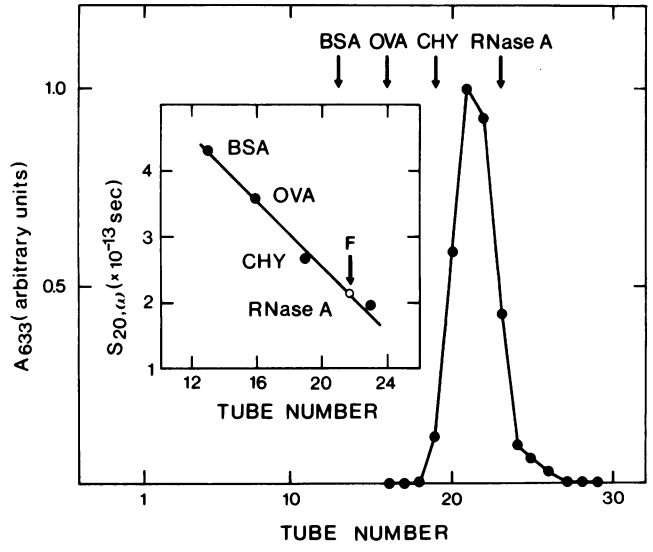

Figure 4. Glycerol gradient sedimentation of protein F and estimation of the sedimentation coefficient. A solution (0.2 ml) containing 120 μg of F protein and 60 μg each of the marker proteins BSA and RNase A was layered onto a 3.8-ml, 10 to 30% (v/v) glycerol gradient in 20 mM Tris·HCl (pH 7.6), 0.1 M NaCl, and 1 mM DTT. In a separate tube a mixture of ovalbumin (OVA), chymotrypsinogen A (CHY), and RNase A was applied to the same gradient. Centrifugation was for 24 h at 53.000 rpm at 4°C in a Spinco SW60 rotor. Fractions of 0.13 ml were collected from the bottom of the tube. Aliquots (15 μl) were taken and assayed for protein content as described in the legend to Fig.3.

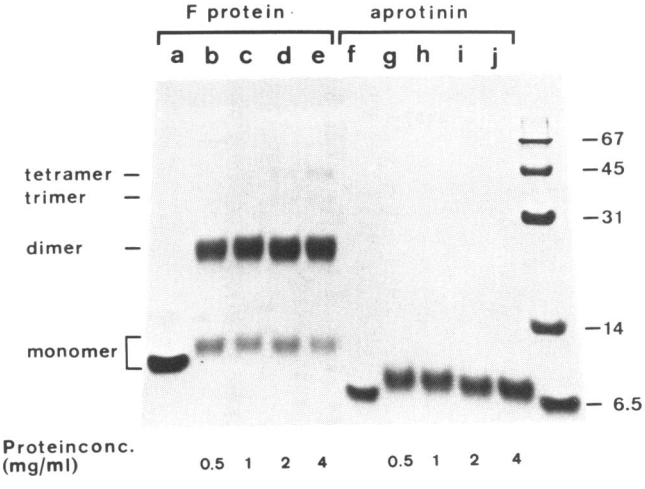

Figure 5. Cross-linking of protein F subunits. Protein was cross-linked according to an established method (23). Reaction mixtures (100 μl) contained 20 mM KPO_4 (pH 6.8), 200 mM KCl, 1 mM DTT, 5% (v/v) ethylene glycol, 2% (v/v) 1,3-butadiene diepoxide, and F protein or aprotinin (control for intermolecular reactions) at the indicated concentrations. After incubation for 2h at 37°C, reactions were stopped by adding 11 μl of 2 M methylamine HCl, pH 7.6. The cross-linked products were denatured and reduced, and a portion of each sample (10 μg protein) was analyzed by 15% SDS/urea-PAGE. Lanes a and f contain 10 μg of nonreacted F protein and aprotinin, respectively. The marker lane at the right contains BSA, ovalbumin, carbonic anhydrase, lysozyme, and aprotinin whose M_r's are given $\times 10^{-3}$.

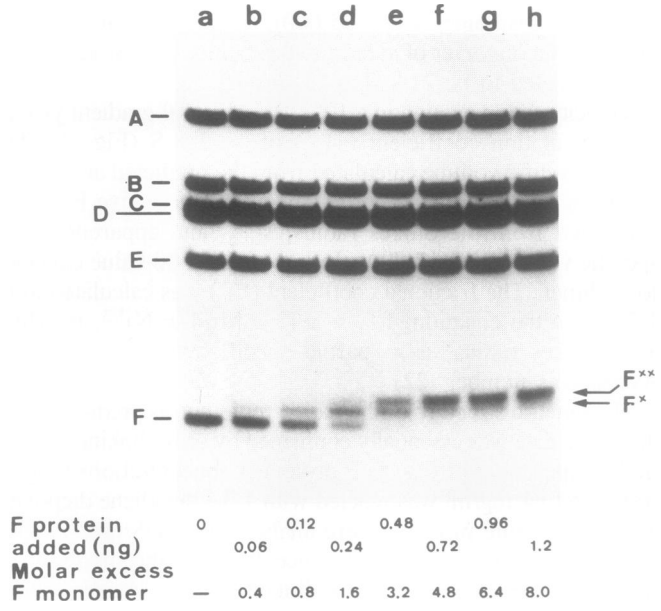

Figure 6. Gel electrophoresis of F protein-DNA complexes. Binding reactions (25 μl) were as described under MATERIALS AND METHODS, with 0.12 μg (20 fmol) of *AvaI* + *AccI*-digested, 5′ end-labeled RSF1010 DNA and various amounts of F protein as indicated. The protein-DNA complexes were electrophoresed on a 1.4% agarose gel and were visualized by autoradiography. Arrows labeled F* and F** point to two major complexes formed by F binding to the P_4 promoter-bearing fragment F (see the text). The percentages of fragment F to F* + F** conversion were determined to be 0, 18, 42, 66 and 90 for lanes a−e and > 90 for lanes f−h, respectively. The smallest RSF1010 fragment produced by *AvaI* + *AccI* digestion, G (nt 1924−2142, Fig.1) has run off this particular gel; no specific F binding to this fragment was detected in other experiments.

a single cysteine (residue 51) for the 68-residue F polypeptide (1). Dialysis of the native protein against destilled water caused it to precipitate, but is was readily soluble in dilute acetic acid (pH 4.0−4.2) or low salt buffer of pH 7.5−8.0 (e. g., binding buffer).

Gel retardation analysis of operator DNA binding

F protein binds specifically to the P_4 promoter region in RSF1010, as shown in Fig. 6. In this experiment, *AccI*/AvaI doubly digested RSF1010 DNA, labeled at the 5′ ends with ^{32}P, was incubated at 37°C with varying amounts of F protein and then subjected to electrophoresis in a neutral agarose gel. The P_4 promoter is contained on a 439-bp *AvaI*-to-*AccI* fragment (nt 5032−5471, Fig. 1). In the presence of sufficient F protein, this fragment (labeled F in Fig. 6) was specifically converted to protein-DNA complexes as revealed by its reduced electrophoretic mobility. At subsaturating concentrations of protein F, two distinct bands of retardation are observed (arrows labeled F* and F**), which suggested to us that two separate F binding sites (or operators) may exist. This was confirmed in subsequent DNA cleavage protection experiments (see below).

Under standard binding conditions with 20 fmol of fragmented RSF1010 DNA (10^{-9}M), 50% fragment F→F* + F** conversion was observed upon the addition of an approximately equimolar amount of F protein (as monomer); the change in band position of the fragment was essentially complete (greater than 80% F→F** conversion) when the F monomer: DNA molar ratio was ≥5. Essentially the same result was obtained when a 10-fold higher concentration of DNA (10^{-8}M) was used. Thus, at saturation two F protein molecules, presumably as a preformed dimer (see DISCUSSION), are apparently bound to

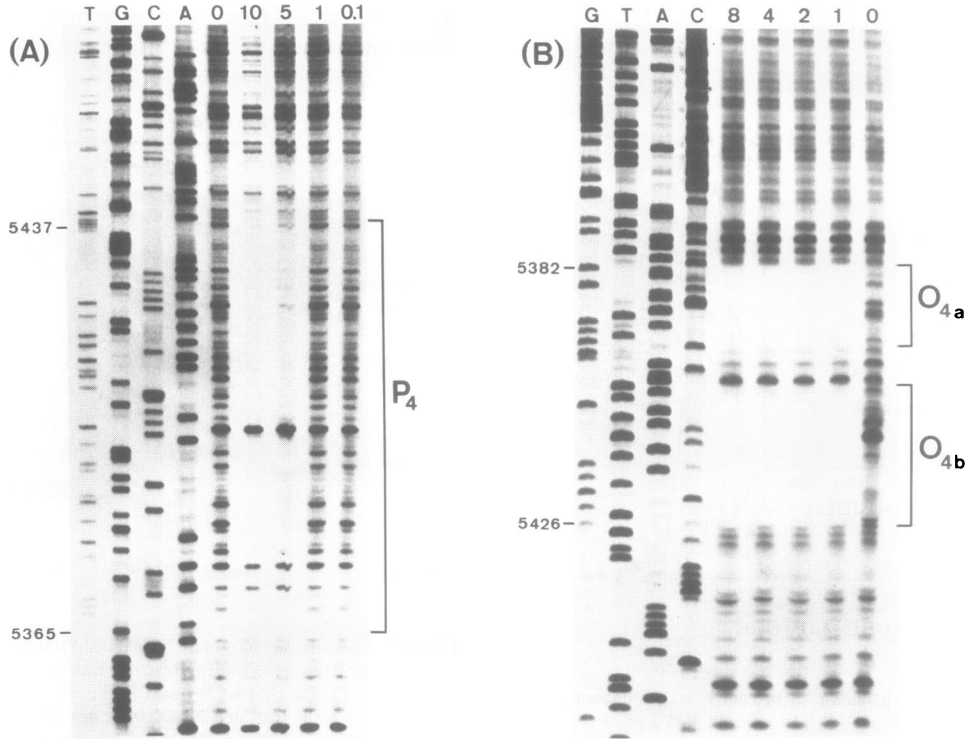

Figure 7. Visualization of RNA polymerase (A)- and F repressor (B)- binding in the P_4 promoter/operator region. DNaseI footprinting was performed on 5′-^{32}P end-labeled, partially duplex mSM2 (A) and mSM1 (B) DNA substrates as described under MATERIALS AND METHODS. Products were fractionated on a 6% sequencing gel and were visualized by autoradiography. Numbers above the lanes indicate picomoles of protein used in each reaction. Positions of the RNA polymerase—and F repressor—protected regions (indicated by the brackets P_4 and O_{4a}/O_{4b}, respectively) were determined by comparison with dideoxy sequencing products carrying identical 5′-[^{32}P] ends (lanes G, T, A, C). For their location in the nucleotide sequence of the *rep* B/E intergenic region, see Fig.1C.

each operator site present on the fragment F. The presence of purified RSF1010 E protein during incubation of F protein with the DNA had no discernible effect on its binding to the operator fragment (data not shown).

Footprint analysis of operator DNA binding

A 1904-bp *AccI* fragment and a 1171-bp *AluI* fragment of RSF1010, each containing the *rep*B/E intergenic region with the putative binding sites for RNA polymerase and F protein, were cloned into M13 mp9 to create mSM1 and mSM2, respectively. The viral DNAs of these recombinant phages, carrying opposite strands of RSF1010 DNA, were isolated and converted *in vitro* to a duplex open circular form using PolIk and the M13 17-mer primer that had been labeled at the 5′ end with ^{32}P. To footprint, these 5′-mono-labeled DNAs were mixed with different amounts of purified F protein or *E.coli* RNA polymerase, then treated with a low amount of DNaseI, and the resulting fragments were analyzed in sequencing gels. The results of some of these experiments are shown in Fig. 7 and are summerized in Fig. 1C.

In the presence of RNA polymerase, we observed clear protection over about 73 bp, covering the region from the 3′ end of *mob*A/*rep*B (nt 5365 ± 1) to the beginning of gene E (nt 5437 ± 1). The protected segment contains considerable homology to the *E.coli* promoter consensus sequence (1). Within the RNA polymerase binding region, the F repressor was found to protect two distinct 19−21 bp regions in both strands of the DNA. They are separated by 3−4 bp and overlap the putative +1, −10 and −35 regions of P_4. At the center of each F-protected region lies an inverted repeat sequence of different length (facing arrows in Fig. 1C). Furthermore, each protected region includes an identical 10-bp direct repeat sequence, suggesting that this sequence element (5′ -GTACTCACGC) is the key recognition determinant for the repressor protein. We propose that the two F-binding sites be called O_{4a} and O_{4b} (as depicted in Fig. 1C).

Independence of F binding to O_{4a} and O_{4b}

Because the O_{4a} and O_{4b} operator sites are adjacent, it seemed possible that F binding to one site might positively affect binding of the protein to the other. However, in gel shift experiments using a mixture of RSF1010 *AvaI*/*ScaI* fragments, half-maximal binding of the *AvaI*(5032)-to-*ScaI*(5416) fragment, lacking the right half-site of O_{4b} (see Fig. 1), occurred at about the same F protein: DNA molar ratio as that observed for 50% binding of the O_{4a} + O_{4b}- carrying AvaI-to-AccI fragment F (data not shown). As expected, only one band of retardation was produced with this incomplete operator fragment. Thus, F binding to at least the O_{4a} site appears to be independent of the presence of the second operator site.

DISCUSSION

The results presented above show that the RSF1010 gene F protein acts as a repressor of the *rep*AC promotor P_4. This was demonstrated in three ways: (i) elimination of the sequences 3′ to the gene E end from a P_4-E-F-*rep*A′/*lac*Z transcriptional fusion leads to a derepression of the P_4 promoter, (ii) expression of gene F *in trans* leads to an inhibition of transcription from P_4, and (iii) purified F protein binds to two adjacent operator sites, which overlap the putative +1, −10 and −35 regions of P_4. When RNA polymerase binds to a promoter, it is also thought to contact the DNA at or near these regions (24). So, the F protein apparently prevents transcription from P_4 by blocking the binding of RNA polymerase.

Gene F is the second gene to be transcribed from the P_4 promoter and precedes the essential RSF1010 replication genes *rep*A and *rep*C. There is no recognizable transcription terminator in the nucleotide sequence between the 3′ end of gene F and the 3′ end of *rep*C and, therefore, protein F most probably autoregulates not only its own formation but also that of RepA and RepC. The RepA protein has been characterized as a DnaB-like helicase, and RepC is an origin-binding initiator protein whose concentration in the cell is limiting for RSF1010 replication (5, 7). Hence, in general terms the action of the F repressor resembles that of the Cro protein from bacteriophage lambda, which also negatively controls a transcript that, downstream from the *cro* message, contains two essential, positive-acting replication genes, O and P (25). However, as mentioned in the Introduction, the RSF1010 replication control system contains at least one other autoregulatory loop, that at the *mob* promoters P_{1-3}. Finally, for R1162 (probably identical with RSF1010) it has been reported that a 75-base antisense RNA molecule inhibits translation of *rep*A mRNA (26), and since in RSF1010 *rep*C translation is coupled to that of *rep*A (5), both may be inhibited.

To overproduce the F repressor, we placed the RSF1010 E-F region without the P_4 promoter downstream to the *tac* promoter present on the multicopy plasmid pKK223−3. When induced, cells bearing this plasmid (pSM25) produce the E and F proteins at a level equivalent to about 14 and 4 % by weight of soluble cell protein, respectively. We have attempted to increase the F expression by cloning the F gene alone (on a 442-bp AccI-to-SspI fragment, Fig. 1) onto pKK223−3. Strains bearing this plasmid did produce F repressor, as detected by an *in vivo* complementation assay, but not in quantities sufficient to display a prominent band in the gel pattern of Coomassie-stained cellular proteins (data not shown). We presume that *in situ* F expression is translationally coupled to that of gene E.

The purification protocol that we describe provides homogeneous F protein in quantities of about 1 mg per gram cell paste. In a subsequent purification, the CM-Sepharose chromatography step was omitted and the final preparation was still greater than 95% electrophoretically pure. It is interesting that in low salt buffer of pH 7.4−7.6 the F repressor will adsorb to both CM-Sepharose and DEAE-Sephacel, an indication for an uneven distribution of charges over the protein surface (27). In fact, the predicted amino acid sequence of the F polypeptide (1) shows a total of 13 basic residues (Arg and Lys, no His) and 13 acidic residues (Asp and Gln), which are distributed such that the N-terminal half (residues 1−37) contains an excess of 6 positive charges and the C-terminal half (residues 38−68) an excess of 6 negative charges (overall net charge, ±0).

The molecular weight of the native F protein is estimated to be 17,000 with a frictional coefficient of 1.22, suggesting near symmetry. Based on the calculated subunit molecular weight of 7,673 and cross-linking studies, the repressor most likely exists as a dimer in dilute solution; at concentrations above about 1 mg/ml, tetrameric, and even larger aggregates may also exist, however (Fig. 5).

The F species active in operator DNA binding is presumably the dimer, since a repressor preparation consisting mainly (> 90%) of reversibly cross-linked dimers (obtained by dialysis of the purified F protein against a buffer lacking a reducing agent) behaved in the DNA binding studies reported here just as the nondialyzed, reduced material. With both forms of F protein, two equal-sized DNA regions of 19−21 bp were protected against

DNaseI attack, and in titration experiments using the gel retardation assay, near saturation of the two operator sites was achieved when there were 2.5 F dimers per *Acc*I/*Ava*I-restricted RSF1010 molecule in the reaction mixture (Figs. 6 and 7, and data not shown).

The two F-protected operator sequences (O_{4a} and O_{4b}, Fig. 1) show partial 2-fold symmetry and include an identical direct repeat sequence of 10 bp. By alignment of the two sequences, the following 16-bp, completely symmetric F consensus operator sequence can be derived:

$$\begin{array}{l} \text{G C G T } \overline{\text{G A}} \text{ G T A C T C A C G C} \\ \underline{\text{C G C}} \text{ A C T C A T G A G T G C G} \end{array}$$

where underlines and overlines denote the nonconserved base pairs that occur only in either O_{4a} or O_{4b}. By analogy to other oligomeric repressors that bind to symmetric operators (e.g., λ Cro) (for a review see ref. 28), it seems likely that each subunit of a bound F dimer will contact one half-site of the operator and that the bases conserved in each operator site are important for mediating the contacts.

The secondary structure of the F polypeptide, as predicted by the method of Garnier et al. (29), is approximately 47% α-helical, yet lacks obvious homology with the conserved helix-turn-helix domains that form the DNA binding surfaces of many phage and bacterial duplex DNA binding proteins (30). That the repressor uses a 'zinc finger' for DNA binding is also unlikely as it contains only one cysteine residue per subunit. Thus, the F protein appears to be a member of a different class of site-specific DNA binding proteins, perhaps falling into that represented by the phage P22 Arc and Mnt proteins. These small, structurally related repressors (53 and 82 residues/subunit, respectively) use short, probably nonhelical regions of N-terminal residues for operator recognition and binding (31). Although the F repressor does not share direct sequence homology with either Arc or Mnt, 8 out of its first 9 residues (Met-Lys-Asp-Gln-Lys-Asp-Lys-Gln-Thr...) are polar and hence could readily make hydrogen bonds to bases in the DNA helix or interact with the sugar-phosphate backbone. Further structural studies as well as the isolation and analysis of F protein and F operator mutants will be needed to determine the mode by which this small repressor contacts its two partially symmetric operator sites. The crystallization of protein F for x-ray diffraction is being attempted.

ACKNOWLEDGEMENT

We thank H. Schuster for generous support and helpful discussions during the course of this work.

REFERENCES

1. Scholz, P., Haring, V., Wittmann-Liebold, B., Ashman, K., Bagdasarian, M. and Scherzinger, E. (1989) Gene 75, 271–288.
2. Frey, J. and Bagdasarian, M. (1989) In Thomas, C. and Franklin, F.C.H. (eds.), Molecular Biology of Broad Host Range Plasmids, Academic Press, London, pp. 79–94.
3. Bagdasarian, M.M., Scholz, P., Frey, J. and Bagdasarian, M. (1987) in Novick, R. and Levy, S. (eds.), Evolution and Environmental Spread of Antibiotic Resistance Genes, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 209–223.
4. De Graaf, J., Crosa, J. H., Heffron, F. and Falkow, S. (1978) J. Bacteriol. 134, 1117–1122.
5. Haring, V. and Scherzinger, E. (1989) In Thomas, C. and Franklin, F.C.H. (eds.), Molecular Biology of Broad Host Range Plasmids, Academic Press, London, pp. 95–124.
6. Scherzinger, E., Bagadsarian, M. M. Scholz, P., Lurz, R., Rückert, B. and Bagdasarian, M. (1984), Proc. Natl. Acad. Sci. USA 81, 654–658.
7. Haring, V., Scholz, P., Scherzinger, E., Frey, J. Derbyshire, K., Hatfull, G., Willetts, N.S. and Bagdasarian, M. (1985) Proc. Natl. Acad. Sci. USA 82, 6090–6094.
8. Scholz, P., Haring, V., Scherzinger, E., Lurz, R., Bagdasarian, M. M., Schuster, H. and Bagadasarian, M. (1984) In Helinski, D. R., Cohen, S. N., Clewell, D. B. and Jackson, D. A. (eds.), Plenum Press, New York, pp. 243–259.
9. Derbyshire, K. M., Hatful, G. and Willetts, N. S. (1987) Mol. Gen. Genet. 206, 161–168.
10. Bagdasarian, M., Lurz, R., Rückert, B., Franklin, F. C. H., Bagdasarian, M. M., Frey, J. and Timmis, K. N. (1981) Gene 16, 237–247.
11. Schneider, K. and Beck, C. F. (1987) Methods Enzymol. 153, 452–461.
12. Boyer, H. W. and Roulland-Dussoix, D. (1969) J. Mol. Biol. 41, 459–472.
13. Messing, J. (1983) Methods Enzymol. 101, 20–78.
14. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning–A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
15. Chang, A. C. Y. and Cohn, S. N. (1978) J. Bacteriol. 134, 1141–1156.
16. Brosius, J. and Holy, A. (1984) Proc. Natl. Acad. Sci. USA 81, 6929–6933.
17. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
18. Fried, M. G. and Crothers, D. M. (1981) Nucl. Acids Res. 9, 6505–6525.
19. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
20. Johnson, A. D., Meyer, B. J. and Ptashne, M. (1979) Proc. Natl. Acad. Sci. USA 76, 5061–5065.
21. McKnight, G. S. (1977) Anal. Biochem. 78, 86–92.
22. Siegel, L. M. and Monty, K. J. (1966) Biochim. Biophys. Acata 112, 346–362.
23. Bäumert, H. G., Sköld, S. E. and Kurland, G. G. (1978) Eur. J. Biochem. 89, 353–359.
24. Siebenlist, U., Simpson, R. B. and Gilbert, W. (1980) Cell 20, 269–281.
25. Matsubara, K. (1976) J. Mol. Biol. 102, 427–439.
26. Kim, K. and Meyer, R. J. (1986) Nucl. Acids Res. 14, 8027–8046.
27. Scopes, R. K. (1982) Protein Purification–Principles and Practice, Springer-Verlag, New York·Heidelberg·Berlin.
28. Pabo, C. O. and Sauer, R. T. (1984) Annu. Rev. Biochem. 53, 293–321.
29. Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
30. Dodd, J. B. and Egan, J. B. (1987) J. Mol. Biol. 194, 557–564.
31. Knight, K. L., Bowie, J. U., Vershon, A. K., Kelley, R. D. and Sauer,R. T. (1989) J. Biol. Chem. 264, 3639–3642.