Mol Gen Genet (1989) 217:488–498

MGG

Characterization of the ColE1 mobilization region and its protein products

A. Christopher Boyd[1], John A.K. Archer[2], and David J. Sherratt[3]

[1] Department of Molecular Biology, University of Edinburgh, King's Buildings, Mayfield Rd., Edinburgh EH9 3JR, UK
[2] Department of Applied Biological Sciences, Room 56-121, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3] Institute of Genetics, University of Glasgow, Church St., Glasgow G11 5JS, UK

Summary. A third of the 6.6 kb genome of ColE1 is devoted to mobilization (*mob*) genes necessary to promote its specific transfer in the presence of conjugative plasmids. The *mob* region is genetically complex: two *mob* genes are entirely overlapped by a third. Oligonucleotide-directed mutagenesis was used to insert an amber codon into one of the overlapped genes and make possible a full complementation analysis of *mob*. Four *mob* genes essential for mobilization by R64*drd11* were thus identified. Fragments of *mob* were subcloned under control of the P_{tac} promoter in a suitable vector, overexpressed in minicells and the mobilization proteins visualized. A comprehensive alignment of the *mob* region of ColE1 with those of its close relatives ColK and ColA demonstrating that the four essential *mob* genes are conserved is also presented.

Key words: ColE1 – Plasmid mobilization – Overlapping genes – Site-directed mutagenesis – Relaxation complex

Introduction

The conjugative transfer of plasmid DNA is a process comparable in complexity to DNA replication and requires the concerted action of many gene products (Willetts and Wilkins 1984; Ippen-Ihler and Minkley 1986). While large, conjugative plasmids like F can transfer autonomously, much smaller plasmids, belonging to a class typified by ColE1, are only transferred – mobilized – in the presence of a conjugative plasmid. The role of the conjugative plasmid in mobilization is almost certainly to provide the initial, pilus-mediated cell contact, the subsequent formation of a conjugation pore and other morphological functions. ColE1 exists in vivo largely as "relaxation complex" (Clewell and Helinski 1969), a specific association of proteins with *oriT*. The existence of the complex correlates with the presence of a full complement of mobilization (*mob*) genes, so we believe the proteins are *mob*-encoded and constitute the machine that nicks ColE1 prior to mediating the transfer of the nicked strand. It seems likely therefore that the *mob* genes of ColE1 and similar plasmids represent a minimal subset required only for the DNA metabolism at the core of the process, i.e. the specific nicking at *oriT* (=*bom*; Finnegan and Sherratt 1982), the "piloting" of the 5′ end of the nicked strand into the recipient (Boyd and Sherratt 1986), recircularization, and perhaps the priming of complementary strand synthesis in the recipient (Willetts and Wilkins 1984). The fact that ColE1 can be mobilized by any of a highly diverse set of conjugative plasmids suggests that it is able to carry out these processes independently. Thus, scrutiny of the mode of action of *mob* genes and their products should lead to a clearer understanding of the central steps of plasmid transfer.

In this study, we present evidence for four genes involved in mobilization of ColE1. We show that less than 2.1 kb of ColE1 is required for *trans*-acting mobilization functions and that the overlapping gene organization suggested by the DNA sequence (Chan et al. 1985) does indeed exist. Moreover, the same organization is observed in two related plasmids ColK (Archer 1985) and ColA (Morlon et al. 1988a). Implications for the expression of *mbe* genes are discussed.

Note. Following the example of Summers et al. (1985) with the resolution sites of ColE1-like plasmids (*cer* for ColE1, *ckr* for ColK etc.) we propose that the specific *mob* genes of ColE1, ColK and ColA be designated *mbe, mbk* and *mba*, respectively; the name *mob* being used therefore only as a generic term to describe genes (or proteins) common to all three plasmids. Thus, "*mobA*" refers to *mbeA, mbkA* and *mbaA* as a class of conserved *mob* genes rather than to any one in particular.

Materials and methods

Bacterial strains and plasmid vectors. (Only relevant characteristics shown.) The hosts DS631, CB133 and CB257 used in mobilization experiments were R64*drd11*-containing derivatives of AB2463 (Boyd and Sherratt 1986), CSH26 (suppressor-free; Miller 1972) and CR63 (cured of F by acridine orange treatment; *supD*; Bachmann 1972), respectively. The recipient used was CB56, a spontaneous Rif^r mutant of CSH26. The minicell-producing strain was CB175, a derivative of DS410 (Warren and Sherratt 1977) carrying an F′*lacI*^{q}*ZΔ*M15 to repress the P_{tac} promoter of pACB62. Transformations after cloning were into strains JM83, JM101 or BMH71-18, all of which carry the *lacZΔ*M15 allele for alpha-complementation. JM101 and BMH71-18 also carry the *lacI*^{q} mutation (Norrander et al. 1983).

For complementation experiments, the vectors used were: pMB1 replicons; pUC9 (Ap^r, Bom^-, 2.7 kb; Vieira

Offprint requests to: A.C. Boyd

and Messing 1982), pHG171 (Ap^r, Bom$^+$, 3.4 kb; Stewart et al. 1986); and λdv replicons; pACB101 (Cm^r, Bom$^-$, 5.0 kb; Boyd and Sherratt 1986), pACB104 (similar to pACB101 but with the pUC18 polylinker of Norrander et al. 1983; Cm^r, Bom$^-$, 4.3 kb; manuscript in preparation), pACB107 (a pACB104 derivative with the pUC19 polylinker and containing ColE1 *bom* on a *TaqI*–*PvuII* fragment; Cm^r, Bom$^+$, 4.6 kb). (Morlon et al. (1988b) were mistaken in stating that pUC plasmids are mobilizable; all have had the *Acc*I site and flanking regions originally sited adjacent to *oriT* deleted and are demonstrably Bom$^-$.) For overexpression of *mbe* proteins, pACB62 (Ap^r, 3.9 kb), a *bom*$^-$ *rom*$^-$ derivative of the P_{tac} promoter vector pKK223-3 (Pharmacia; Analects (1984) 12(3): pp 1–2; Amann et al. 1983), was constructed essentially by deleting the 705 bp of DNA corresponding to the *Hae*II B and H fragments of pBR322 (Twigg and Sherratt 1980). By deleting *bom* we expected to eliminate any sequestering or other effects of the vector on the cloned *mob* gene products. The removal of *rom* had the useful property of increasing the copy number by two- to threefold (Twigg and Sherratt 1980).

Media and antibiotics. These were as described previously (Boyd and Sherratt 1986). Ap, ampicillin; Cm, chloramphenicol.

Mobilization assays. Mid-log donor cells (0.4 ml) were mixed with 0.4 ml of stationary phase recipient cells, spread on an L plate and incubated at 37° C for 2–3 h. Viable counts of donors and recipients at the beginning of mating were performed: there was always at least a tenfold excess of recipients. Under these conditions, R64*drd11* transfer approaches 100%. Cells were harvested into 2 ml L broth using a glass spreader. If the mobilizable plasmid was Ap^r, the harvested cells were then immediately washed three times with L broth to remove β-lactamase from the medium. Dilutions were made and spread on plates selective for transfer of R64*drd11* only and for transfer of both R64*drd11* and the Bom$^+$ plasmid, so that co-transfer frequencies could readily be calculated.

Recombinant DNA and single clone plasmid screening techniques. These were as described previously (Boyd and Sherratt 1986). Restriction enzymes, ligase and Klenow fragment were purchased from Boehringer and used as recommended.

Construction of F9. A plasmid containing the F1 fragment of ColE1 (Fig. 1), with unique sites for *Cla*I and *Stu*I in the *mbeA* reading frame, was cleaved with both enzymes. After overnight self-ligation, the mixture was heat treated to kill ligase activity and re-incubated with *Cla*I and *Stu*I to remove reconstituted parent plasmid. BMH71-18 was then transformed by the mixture and spread on Ap plates. One colony was recovered from an original input of 2 μg of plasmid DNA and proved to contain a plasmid with a restriction pattern profile consistent with the deletion of ca. 250 bp between the restriction sites. We assume that it arose by some illegitimate event such as ligation between the blunt *Stu*I end and the 2 base 5′ single-strand protrusion of the *Cla*I end: its exact provenance is of no consequence since the F9 fragment it provided had the expected Mbe$^-$ phenotype. F9 was subcloned into pHG171 to generate pACB97 (Fig. 2).

Oligonucleotide mutagenesis. A 17-mer oligonucleotide, 5′-CCTCTGTCTAGGCTGTG-3′ complementary to the *mbe* sense strand between coordinates 3032 and 3048 (Chan et al. 1985), except that the underlined T residue replaces the original G residue, was synthesized by John Keyte (Department of Biochemistry, University of Leicester, UK) as described (Boyd et al. 1986). The single base change to be effected by the oligonucleotide was chosen such that the protein sequence of *mbeA* was not affected (the leucine codon at 3038–3040 is changed from CTC to CTA) while an amber stop codon TAG is inserted in place of the serine codon TCG (3039–3041) within the overlapping *mbeB* reading frame. The F4 fragment of ColE1 (Fig. 1) was subcloned into the *Hin*cII site of M13mp9 in the appropriate orientation. The single-stranded DNA preparation, oligonucleotide annealing, extension by Klenow polymerase, and ligation steps were essentially as described by Grundstrom et al. (1985). The duplex circles were then cleaved immediately with *Hin*dIII and *Eco*RI to excise the heteroduplex F4 fragment and ligated to *Hin*dIII/*Eco*RI cleaved pACB62 which had been purified from an agarose gel slice (Tautz and Renz 1983). JM101 was transformed with the ligation mix and Ap^r clones selected. Plasmid minipreparations (Birnboim and Doly 1979) were screened for the presence of the extra *Mae*I restriction site (CTAG) created by the oligonucleotide. DNA from one such positive clone was purified and used as the source of mutant *mbeB* gene in the construction of pACB97 (Table 2).

Minicells. These were prepared from CB175 carrying the appropriate pACB62 derivatives as described (Hallewell and Sherratt 1976), except for the inclusion of an isopropyl β-D-thiogalactoside (IPTG) induction step. Minicells were incubated in M9 minimal medium at 37° C for 30 min and then split into two aliquots. IPTG to 2 mM was added to one aliquot 15 min prior to the addition of [^{35}S]methionine to both. After incubation for 45 min, L broth containing 40 μg/ml methionine was added as chase and incubation continued for a further 45 min. Minicells were then lysed and loaded onto a 12.5% SDS/polyacrylamide gel (Laemmli 1970).

Sequencing of ColK mob. Restriction fragments of ColK were shotgun cloned into M13 vectors and sequenced by the dideoxy method (Archer 1985; Messing et al. 1981).

Results

Identification of reading frames in mbe

Open reading frames (ORFs) in the ColE1 *mob* region were identified by plotting start and stop codons on one-dimensional graphs of the sequence for each frame (Fig. 1). A program based on work described by Stormo et al. (1982) was used to locate potential ribosome binding sites. While the Perceptron algorithm used by the program is not ideal, and although many genes have poor or absent ribosome binding sites, sequences yielding high scores usually correlate with real ribosome binding sites. A combination of these computer methods allowed us provisionally to assign ORFs to genes (as highlighted in Fig. 1 and tabulated in Table 1), two of which are entirely overlapped by a third. We include the *mbe*8 reading frame which, although not

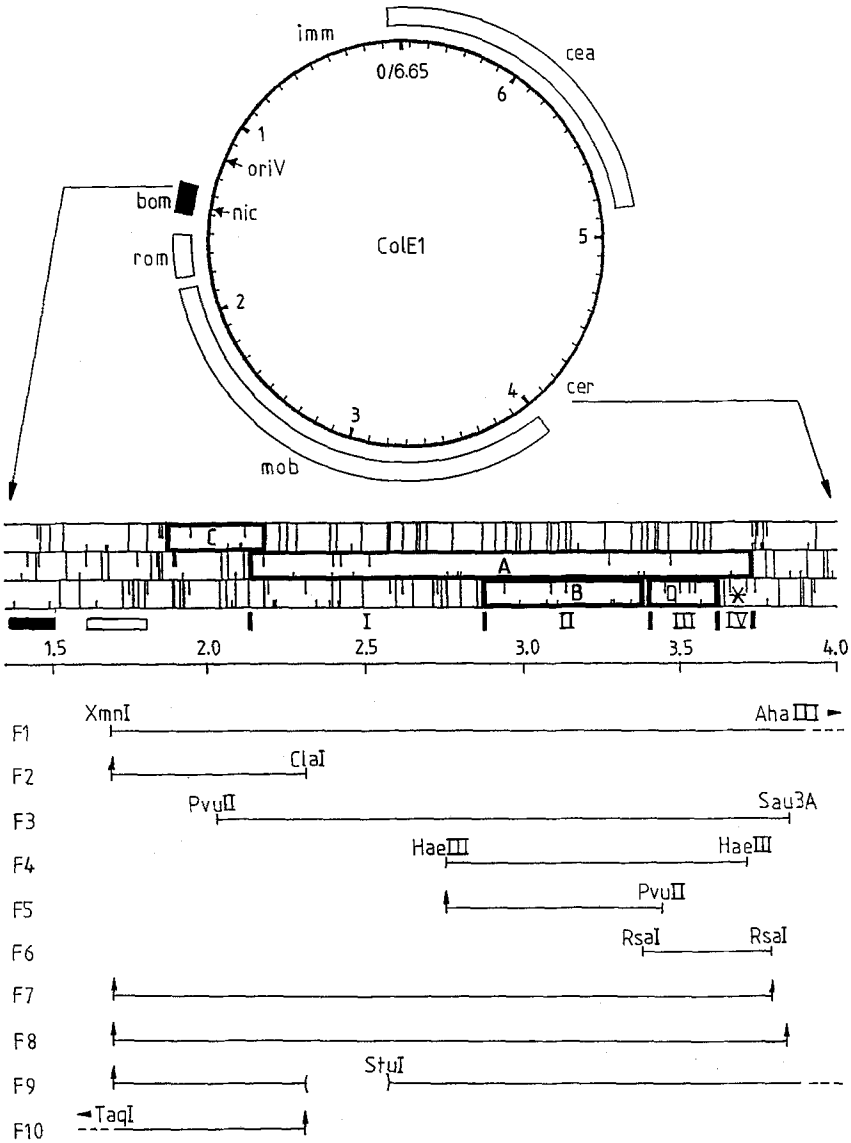

Fig. 1. Reading frames in *mbe*. The three anticlockwise frames of the *mbe* region extending from 1350–4000 are shown in linear form below the circular ColE1 map (cea, colicin gene; imm, colicin immunity region). *Vertical lines* extending right across the bands denote stop codons; those extending half way down, ATG codons; and those extending a quarter of the way up, GTG codons. The essential *mbe* genes are *boxed:* * is the *mbe*8 reading frame. The four regions of *mbeA* discussed in the text are indicated in Roman numerals beneath the three frames. Labels (F1–F10) and extents of *mbe* DNA fragments subcloned for complementation and expression experiments (Table 2) are shown beneath the scale line, together with the restriction enzymes defining their boundaries. Shared boundaries are indicated by arrows (↑). Scale in kilobases

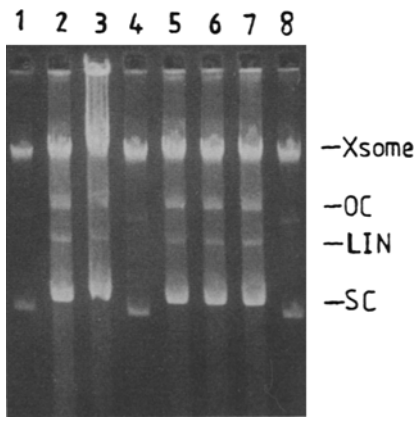

Fig. 2. The effect of overexpression of *mbeB*. Tracks 2, 3, 5, 6 and 7 of a 1% agarose gel were loaded with single-clone extracts (Boyd and Sherratt 1986) of cells containing pACB4: the others were from cells containing a control plasmid with an insert lacking *mbeB*. Xsome, chromosome; OC, open circular DNA; LIN, linear DNA; SC, supercoiled DNA

an essential *mbe* gene (see below), appears to be expressed in the minicell system described here.

In Table 1 the *mbe* ORFs are renamed in accordance with the proposals of Demerec et al. (1966), the letters being assigned in order of decreasing ORF length. Our interpretation of the reading frame assignments to genes is in broad agreement with that described by Chan et al. (1985).

Similar analyses were carried out for the complementary strand of the *mbe* region. Although there are several ORFs, in particular one very long one in-phase with *mbeA*, none have likely ribosome binding sites or conserved start codons. Significantly, no complementary strand ORFs are conserved in the *mob* regions of ColK (Archer 1985) and ColA (Morlon et al. 1988a). We conclude, in agreement with Chan et al. (1985), that the complementary strand does not encode protein.

Identification of the four essential mbe *genes*

A series of subclones of the *mbe* region was made, initially in pUC9 and later in other vectors (see Fig. 1 for the rele-

Table 1. ColE1 *mob* genes and proteins: coordinates and molecular weights

Old[a] name	New name	ColE1[a] coordinates	Figure 4 coordinates	Number of amino acids	Calculated size (daltons)	Estimated size[b]
*mob*2	*mbeC*	1867–2190	1– 324	107	12883	12000
*mob*3	*mbeA*	2180–3733	314–1924	517	57744	60000
NA[c]	*mbeA*-I	2180–2866	314–1000	229	NA	NA
NA	*mbeA*-II	2867–3394	1001–1585	176	NA	NA
NA	*mbeA*-III	3395–3604	1586–1795[d]	70	NA	NA
NA	*mbeA*-IV	3605–3733	1796–1922	42	NA	NA
*mob*6	*mbeB*	2868–3386	1002–1577	172	19525	17000
*mob*7	*mbeD*	3393–3626	1584–1817	77	9299	9000
*mob*8	NA	3657–3818	1847–2008	53	6116	6000

[a] Chan et al. (1985)
[b] Estimated from gel migration in Fig. 3
[c] Not applicable
[d] Corresponds to coordinate of *mbaD* stop codon (Fig. 4)

Table 2. Directory of pACB plasmids containing ColE1 fragments

Fragment	Fragment[a] coordinates	Vectors				*mbe* genotype				
		pUC9	λdv[b]	P_{tac}[c]	pHG171	A	B	C	D	8
F1	1685–4076	7	52*	66		+	+	+	+	+
F2	1685–2309			63		−	−	+	−	−
F3	2023–3836	9	48	65		+	+	−	+	+
F4	2746–3715		42			−	+	−	+	−
F5	2746–3435		47			−	+	−	−	−
F6	3371–3786		50			−	−	−	+	−
F7	1685–3786		53*			+	+	+	+	−
F8	1685–3836				95[d], 96	+	±	+	+	+
F9	1685–4076				97	−	+	+	+	+
F10	1324–2309		41*			−	−	+	−	−

[a] From Chan et al. (1985)
[b] Starred plasmids contain *bom* and are derivatives of pACB107, except pACB41 which is a derivative of pACB101 and contains *bom* on the F10 fragment inserted in its polylinker. The other λdv plasmids are derived from pACB101 or pACB104
[c] pACB62
[d] pACB95 has the amber mutation in *mbeB*: pACB96 is its wild-type equivalent

vant fragments subcloned and Table 2 for a cross reference of inserts and vectors). The parent plasmid for most of the subclonings was pACB4 (Boyd and Sherratt 1986), a pUC9 derivative containing the 2.9 kb *Xmn*I B fragment of ColE1 cloned into the polylinker *Hin*cII site. In the construction of pACB4, we obtained only one orientation of the insert, that is, with transcription from P_{lac} proceeding anti-sense to the *mbe* genes. This suggests that expression of one or more *mbe* genes from P_{lac} on a high copy number vector like pUC9 is deleterious to the cell. Even without transcription from P_{lac} the presence of pACB4 in cells reduces their viability: colonies are translucent and flattened and, in the absence of selection, plasmid-free segregants rapidly dominate. Gel analysis of crude extracts of strains containing pACB4 display marked plasmid DNA degradation (Fig. 2), including the appearance of linear forms. We later identified *mbeB* as the ORF responsible for this effect (see below). All the pUC9, λdv and pHG171 derivatives used in this work have inserts in the same, anti-P_{lac} orientation.

Table 3 shows the results of typical experiments demonstrating that all four ORFs highlighted in Fig. 1 are essential *mbe* genes. In each case, cells harbouring R64*drd11* in which all but one *mbe* ORF is intact were shown not to mobilize the *oriT*-containing plasmid. When the wild-type *mbe* ORF was supplied, however, normal mobilization proficiency was restored, thus verifying that the ORF encoded essential protein. We can interpret these results to confirm that *mbeC, mbeA, mbeB* and *mbeD* are coding regions as follows:

1. mbeC. The plasmid pACB9 has all of *mbe* except *mbeC*, and does not mobilize a λdv plasmid containing only *bom* (pACB107). If a λdv plasmid containing both *mbeC* and *bom* (pACB41) is co-resident however, complementation occurs and pACB41 is efficiently mobilized.

2. mbeA. The *bom*$^+$ plasmid pACB97 contains intact all *mbe* ORFs except *mbeA*, which has had an internal deletion of ca. 250 bp (see Materials and methods for construction of F9). In the presence of pACB48, a λdv plasmid containing *mbeA* (and therefore necessarily *mbeB* and *mbeD*), pACB97 is mobilized, demonstrating complementation of the *mbeA* lesion. The control shows that providing *mbeB* and *mbeD* in *trans* alone is not sufficient.

3. mbeB and *mbeD.* In the case of the overlapped genes, a more complex interpretation is required. We observe that mobilization of the plasmid carrying the mutated *mbeB* ORF (pACB95) is reduced 40-fold compared to its wild-type analogue (pACB96). (That the lesion in *mbeB* is an oligonucleotide-directed amber mutation (see Materials and methods) is confirmed by its suppression in the *supD* host.) This is a much less significant reduction in mobilization proficiency than for *mbeC* and *mbeA*, and is probably an effect of the small amount of spurious suppression in the host strain. More surprisingly, the mutation is not complemented at all when *mbeB* alone is provided in *trans* (on pACB47), but is complemented in the presence of a plasmid containing both *mbeB* and *mbeD* (pACB42). This pattern of complementation confirms that both *mbeB* and *mbeD* are essential genes. It could be that the *mbeB* mutation has a strong polar effect on the expression of the downstream *mbeD* gene, or that wild-type *mbeB* and *mbeD* genes need to be present in *cis* for mobilization to occur. (There are precedents for this: at low efficiency, UGA codons can be mistranslated as tryptophan and cause translational readthrough (Fox 1987). The *mbeB* gene terminates in

Table 3. Results of mobilization experiments to identify *mbe* genes

Host strain	Test[a] plasmid	Co-resident[a] plasmid	Percentage co-transfer[b] with R64*drd11*	Effect shown
DS631	9[c] (F3)	107	<0.005	*mbeC*$^-$ implies Mbe$^-$
DS631	9 (F3)	41 (F10)	17	Complementation by *mbeC*$^+$ restores Mbe$^+$
DS631	7 (F1)	107	30	Control: pACB107 is Bom$^+$
CB133	97 (F9)	–	<0.005	*mbeA*$^-$ implies Mbe$^-$
CB133	97 (F9)	48 (F3)	47	Complementation by *mbeA*$^+$ restores Mbe$^+$
CB133	97 (F9)	42 (F4)	<0.005	Control: no complementation by *mbeB*$^+$/*mbeD*$^+$ alone
CB133	95 (F8)	–	2.2	*mbeB*^{am} implies Mbe$^-$
CB257	95 (F8)	–	61	Suppression in *supD* background
CB133	96 (F8)	–	88	Control: plasmid without amber mutation
CB257	96 (F8)	–	81	Control: plasmid without amber mutation
CB133	95 (F8)	47 (F5)	0.44	Non-complementation by *mbeB*$^+$ alone
CB133	95 (F8)	50 (F6)	0.18	Effect of *mbeD*$^+$ alone
CB133	95 (F8)	42 (F4)	69	Complementation by *mbeB*$^+$/*mbeD*$^+$

[a] Number of cloned ColE1 fragment (if present) shown in brackets
[b] Calculated as 100× (transfer frequency of the mobilized plasmid/transfer frequency of R64*drd11*). Host strains with the test and/or co-resident plasmids were mated with CB56 and the transfer of the mobilized plasmid was selected for using ampicillin or chloramphenicol as appropriate
[c] pACB prefixes omitted for clarity

UGA.) Since the complementing plasmid pACB42 expresses *mbeB* and *mbeD* in the wild-type *cis* configuration, we are investigating the latter possibility by performing complementation experiments in which two plasmids express *mbeB* and *mbeD* separately. We are also constructing an amber mutation of *mbeD* in order to obtain direct confirmatory evidence that it is an essential *mbe* gene. (It is interesting also to note that the mobilization proficiency of the mutant plasmid pACB95 is reduced somewhat in the presence of λdv plasmids pACB47 and pACB50, carrying ORFs *mbeB* and *mbeD*, respectively. We do not yet know whether this is due to specific inhibition by *mbe* gene products, or whether some competitive effect of the λdv replicon itself is responsible.)

The *mbe* region was initially crudely delimited by transposon insertions (Dougan and Sherratt 1977). We used the restriction map deduced from the sequence to help construct a minimal Mbe$^+$ subclone of ColE1. A λdv derivative, pACB53, containing 2.1 kb of ColE1 (fragment F8, Fig. 1) from the *Xmn*I site at 1685 (within *rom*) to the *Rsa*I site at 3786 (within *mbe*8) was constructed in a series of steps (not shown) and was found to be Mbe$^+$ with an efficiency indistinguishable from that of pACB52 which includes *mbe*8. While this construct removes only the last ten amino acids of the putative *mbe*8 protein, this still constitutes a deletion of approaching 20%. It is likely that such a deletion in an essential *mbe* gene would have a measurable effect on mobilization proficiency, so we conclude that the intact *mbe*8 ORF is not essential for mobilization.

The complementation results described above, together with the fact that the plasmid pACB53 (containing intact only *mbeA, mbeB, mbeC* and *mbeD*) is Mbe$^+$, suggest strongly that these four genes are necessary and sufficient for the mobilization of ColE1 by R64*drd11*.

Overexpression of mbe *genes in minicells*

Previous attempts to identify proteins encoded by the ColE1 *mob* region have yielded disappointing results (Ebina et al. 1979; Inselburg and Applebaum 1978; Collins 1979; Chen and Zubay 1983). We believe this to be a consequence of the low level of *mbe* gene expression. To identify which *mbe* ORFs described above were protein coding, therefore, we subcloned suitable fragments (including F1, F2 and F3; Fig. 1) into a specially constructed expression vector, pACB62 (see Materials and methods).

Overexpression of genes under the control of P_{tac} is often deleterious to the cell (Brosius 1984). We have found that any *mbeB*-containing fragments (e.g. F1 and F3) cloned into pACB62 and oriented so that the *mbeB* gene is expressed by P_{tac} can only be transformed into strains like JM101 that overproduce the Lac repressor. Further-

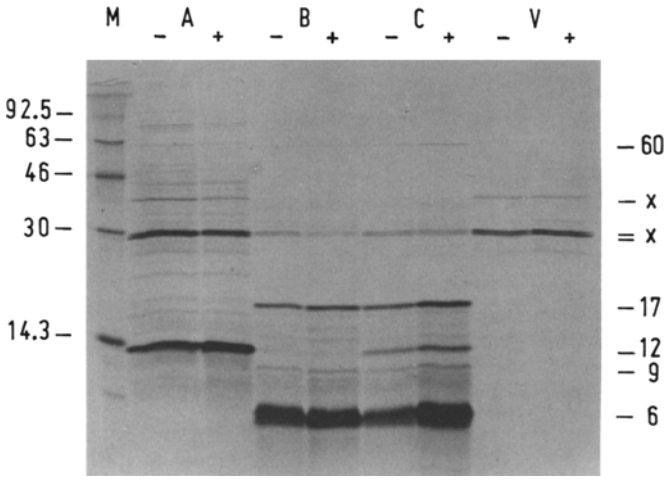

Fig. 3. SDS/polyacrylamide gel analysis of ^{35}S-labelled minicell extracts. Track M contains the size markers; the A, B and C tracks show expression of cloned *mbe* genes in constructs pACB63, pACB65 and pACB66, respectively; the V tracks show expression of vector-specific (pACB62) polypeptides. For each of the paired tracks, + and – refer to incubation with or without isopropyl β-D-thiogalactoside prior to minicell lysis. Sizes are shown in kDa: bands marked x are forms of β-lactamase

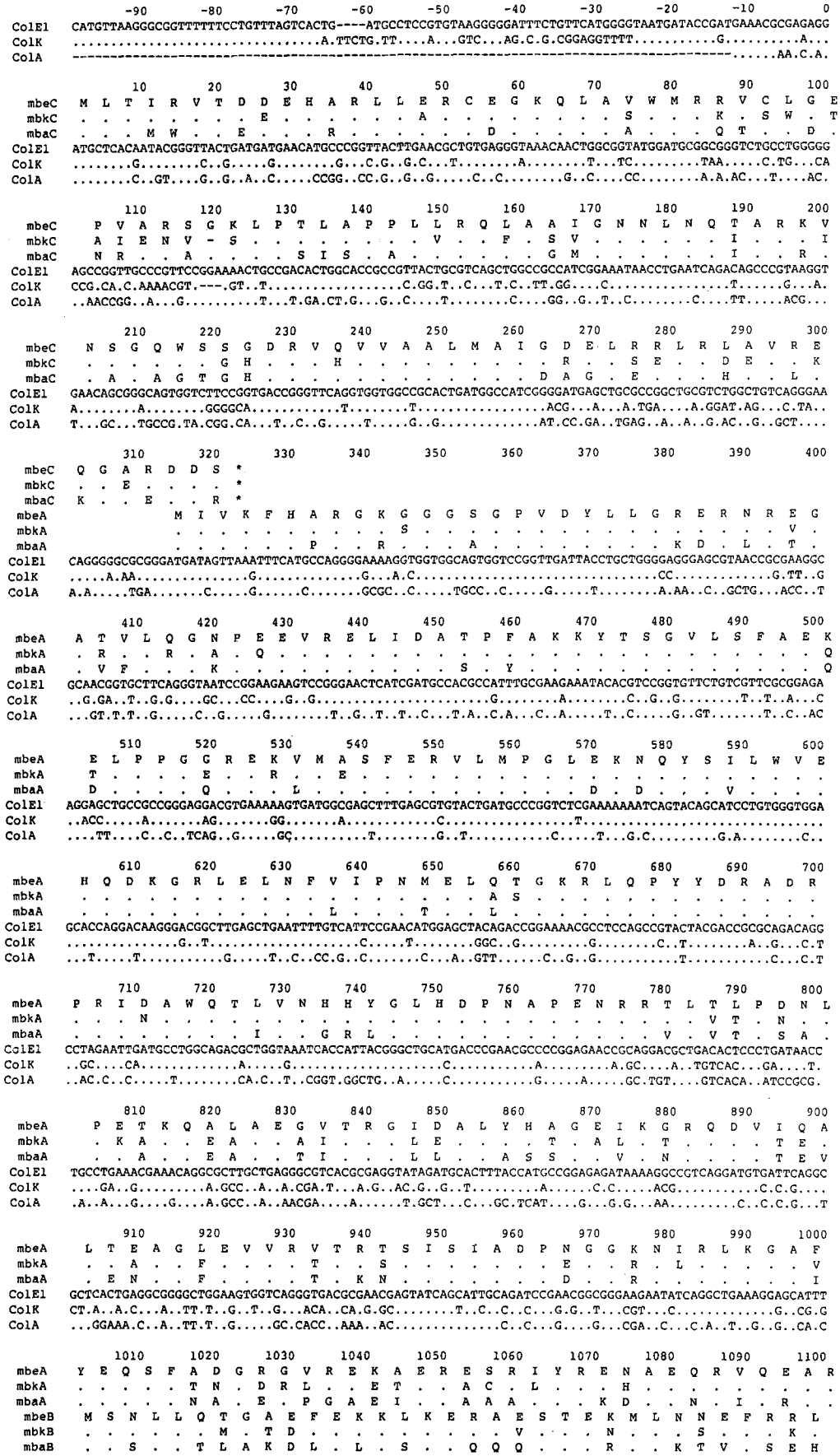

Fig. 4 (continuation see page 494)

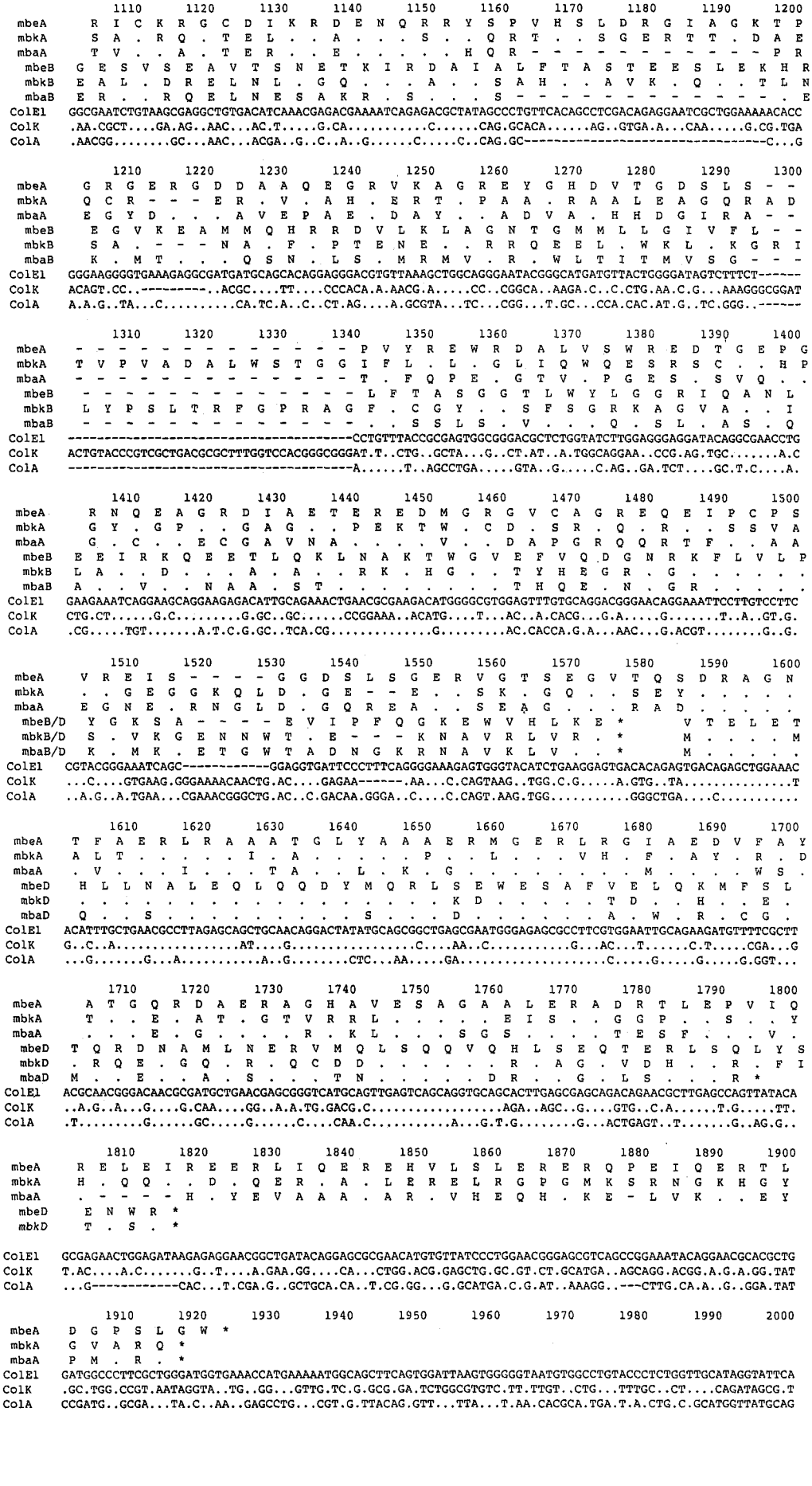

Fig. 4. Alignment of *mob* sequences from ColE1, ColK and ColA. The coordinate system is arbitrary and is defined such that coordinate 1 = first base of *mobC* gene. For both the protein and DNA sequences, a *dot* represents identity with the ColE1 amino acid or base at that position. *Asterisks* represent stop codons: hyphens are inserted where strictly necessary to maximize similarity

Table 4. Percentage similarity scores[a] for *mob* genes of ColE1 (E1), ColK (K) and ColA (A)

Comparison	*mobC*	*mobA* (all)	*mobA*-I	*mobA*-II	*mobA*-III	*mobA*-IV	*mobB*	*mobD*
E1/K	72	56	83	32	56	18	38	66
E1/A	67	57	76	38	69	25	41	70
K/A	54	49	77	26	47	10	30	61
Mean	64	54	78	32	57	18	36	66

[a] Calculated as 100 × (no. of matching amino acid residues/total number of residues)

more, if JM101 derivatives carrying such constructs are plated on IPTG-containing media, there is little or no growth. Gel analysis of IPTG-induced cultures of this kind show the same pattern of plasmid degradation as that seen in analyses of pACB4-containing strains (Fig. 2). The likeliest explanation is that the *mbeB* gene product is a nuclease which cleaves non-specifically when overexpressed.

In the experiment shown (Fig. 3), minicells were prepared from CB175, CB175(pACB63), CB175(pACB65) and CB175(pACB66) (see Table 2). Half of each batch was treated with IPTG to derepress the P_{tac} promoter driving expression of insert genes (see Materials and methods). The effect of IPTG addition was surprisingly limited: the intensity of the bands resulting from IPTG-treated minicells is less than twice that of corresponding bands from untreated minicells. This indicates that repression of P_{tac} is incomplete in this system, and could be the effect of poor segregation of the F′*lacI*^{q} into minicells. Alternatively, Lac repressor binding to the P_{tac} operator may be adversely affected in the physiological conditions of minicells.

The V tracks show vector-specific β-lactamase bands common to all the minicell extracts. The A tracks show a strong insert band migrating at 12 kDa. This is clearly the product of *mbeC* (Table 1), the only ORF in pACB63 capable of encoding a protein of this size. Similarly, the 17 kDa, 9 kDa and 6 kDa bands in the B tracks are assigned to ORFs *mbeB, mbeD* and *mbe*8, respectively. Also faintly visible in the B tracks are 60 kDa bands which must be the product of the *mbeA* ORF encoded by pACB65. In the C tracks can be seen the products of the complete *mbe* region contained in pACB66.

Genetic experiments outlined above demonstrated that *mbe*8 was not essential for mobilization; also, no ORF equivalent to *mbe*8 exists in ColK or ColA (see below). Nevertheless, we see a band corresponding to its product in minicell extracts. It may be that the *mbe*8 protein has a non-essential accessory role in mobilization under conditions yet to be discovered.

ColE1, ColK and ColA mob *regions compared*

Comparison of ColE1 *mob* with the closely related *mob* sequences of ColK and ColA reveals extensive similarity. We find that the four ORFs found to encode essential proteins in ColE1 are conserved in the other two plasmids. Moreover, no other ORF of any significant size, including the *mbe*8 ORF, is conserved in all three plasmids. Figure 4 shows a comprehensive alignment of the sequences both at the DNA and amino acid sequence level. (Morlon et al. (1988a; their Fig. 7) also aligned the *mob* polypeptide sequences of ColA with those of ColE1, but arrived at a less parsimonious alignment than that presented here. In addition, their comparison of *mbaA* with *mbeA* displays few matches to the first 30 amino acids of the ColE1 gene: we strongly suspect that this is the result of a DNA sequencing mistake which inserted an additional C residue at coordinate 462. When this spurious base is removed, the deduced proteins match as shown in Fig. 4.) Initial alignment was of the amino acid sequences and was achieved by use of the WORDSEARCH program (Devereux et al. 1984), which inserts gaps only where necessary to maximise matches. It was found that gaps introduced into the *mobB* genes were introduced by the program into similar regions of the *mobA* genes: this reinforces the validity of the alignment here. Incorporation of the DNA sequence alignments into the figure was achieved manually and required only trivial changes in the automatically aligned polypeptide sequences. While the figure does not represent the best possible alignment of the three sequences, we believe that it is sufficiently accurate to permit a valid assessment of evolutionary relatedness.

Inspection of the aligned DNA sequences reveals that many of the base differences are at codon third base (wobble) positions and therefore do not affect the amino acid sequence. This type of variation is peculiar to protein coding sequences and gives further credence to the gene designations arrived at for ColE1. The conservation of gene length is considerable and only the *mobB* genes and the corresponding regions of *mobA* have a large number of deletions and insertions. We now describe pertinent features of each gene in turn. (Refer throughout to Fig. 4 for sequence and coordinate data and to Table 4 for similarity scores.)

1. mobC. The *mbkC* gene is shorter than the other two by a single codon (coordinates 119–121); otherwise, *mobC* is highly conserved. The *mbeC* and *mbkC* reading frames extend upstream from their designated start points, and contain extra in-phase ATG codons: however, the *mbaC* sequence is divergent upstream, so we propose that the ATG at coordinate 1 is the true start codon for all three genes. Here we are at variance with Chan et al. (1985) who place the *mbeC* start at coordinate −23 and suggest that the GGGGT sequence just preceding it is a ribosome binding site. As this sequence is completely unconserved in ColK, we reject their placement. (The long deletion in the ColA sequence designated by the hyphens up to coordinate −13 is indicative of the fact that ColA lacks an analogue of the *rom* gene (Cesareni et al. 1982) which is encoded by the strand complementary to *mob* and starts with a GTG codon at coordinate −67 in both ColE1 and ColK. This clearly has implications for the control of ColA copy number.) All three genes terminate in a TAA codon, five

bases downstream of the ATG start codon of the *mobA* genes.

2. mobA. For convenience, this gene is split into four regions (Fig. 1).

mobA-I comprises 43% of the gene and is the most conserved, with no gaps in the alignment. Although the sequences upstream of the start codons show little similarity to classical SD sequences (Shine and Dalgarno 1974), they are strikingly purine rich: the 16 bp upstream of *mbeA*, *mbkA* and *mbaA* have only three, two and three pyrimidines respectively. This is a common feature of genes with slight terminal overlaps: many such purine-rich tracts are found, for example, preceding the ORFs in the *nin* region of bacteriophage λ (Sanger et al. 1982). It has been suggested that they function to promote translational coupling between the overlapping genes (Oppenheim and Yanofsky 1980).

mobA-II and *mobA*-III are those regions of *mobA* which are overlapped by *mobB* and *mobD*, and constitute 36% and 13% of the gene, respectively. *mobA*-II is the second least conserved region and requires the insertion of most gaps to align the sequences. It should be noted, however, that the 5′ end of the region (coordinates 1001–1156) is noticeably better conserved than the rest of the region. Surprisingly, *mobA*-III needs no gaps and is much more conserved than *mobA*-II.

mobA-IV is the extreme 3′ end of the gene, comprising only 8%, and is not overlapped. Nevertheless, it is highly diverged – the most diverged region in *mob*, much more so than *mobA*-II. The introduction of more gaps would increase the similarity between the sequences in this region, but we do not think this is justified in such a short region. Note that the *mbeA* gene ends two codons downstream of the other two genes, and with a TGA codon instead of a TAG codon.

3. mobB. Unlike the *mobA* gene starts, the *mobB* genes are preceded by sequences resembling normal ribosome binding sites (coordinates 989–995). This poses the problem of accounting for how translation of *mobB* is initiated (see Discussion). The alignment obviously requires gaps corresponding to those inserted in *mobA*-II, and the similarity pattern varies concordantly with that region of *mobA*. The *mobB* gene is overall only slightly better conserved than *mobA*-II. All three genes terminate at equivalent positions (coordinate 1575): *mbeB* and *mbaB* with TGA; *mbkB* with TAA.

4. mobD. The *mobD* genes start two codons downstream of the *mobB* stop codons, the intergenic gap being completely unconserved: the putative ribosome binding site is at coordinates 1568–1574, immediately preceding the *mobB* stop codons. The *mbeD* gene is the only *mob* gene which has a GTG start. The most noteworthy feature of *mobD* is its high degree of conservation: the alignment requires no gaps, and the similarity is slightly greater than that of the unoverlapped *mobC* genes. Again, the *mobD* gene is somewhat better conserved than the *mobA*-III region it overlaps. *mbeD* terminates with a TAG codon at coordinate 1794, while the other two genes terminate with TAAs seven codons further downstream, at coordinate 1815.

It is clear from the above analysis and the similarity figures in Table 4 that the *mob* sequences of ColE1, ColK and ColA have evolved in a strikingly non-uniform manner. The *mobA*-II/*mobB* and *mobA*-IV regions have diverged to a significantly greater extent than the rest of *mob*. The high variability of the *mobA*-IV segment suggests that this small carboxy-terminal sequence is not important for *mobA* protein function. More noteworthy is the other highly variable region, since it occurs in a region of overlap where it might be expected that evolutionary change is constrained: indeed, the adjacent *mobA*-III/*mobD* overlap region is significantly better conserved. Such high divergence may indicate that the *mobB* protein, and the corresponding part of the *mobA* protein can drift significantly in sequence without affecting function. The *oriT* sites of all three plasmids are completely interchangeable (Archer 1985; Morlon et al. 1988b): there are some differences, however, in their mobilization properties – for example, ColK is not mobilized by F (Archer 1985). It is tempting to speculate that such differences are a consequence of the drift in the *mobB* overlap region.

An intriguing feature of the *mob* region sequences is the distribution of GATC sequences. In the sequences displayed in Fig. 4, ColE1 has a single GATC (coordinate 959); ColK has none; yet ColA has six (coordinates 128, 226, 316, 637, 832, 1758). We believe that GATC sequences in *mob* have been selected against in ColE1 and ColK, whose natural host *Escherichia coli* possesses the *dam* enzyme which methylates GATC sequences and is responsible for permitting strand discrimination in mismatch repair (Glickman et al. 1978). Why this selection should have occurred is not clear, but the near absence of GATC sequences in a 2 kb stretch of DNA is highly significant. Since there are the expected number of GATC sequences in *mba*, we would predict that *Citrobacter freundii* CA31, the natural host of ColA, lacks a *dam*-like methylation system.

Discussion

The minicell and genetic experiments described above provide strong evidence that ColE1 *mob* encodes four genes essential for plasmid transfer. Furthermore, these genes are conserved in two related plasmids ColK and ColA, suggesting that all three have diverged from a common ancestral plasmid capable of being mobilized. While CloDF13 appears to have functionally similar mobilization properties (Van Putten et al. 1987), it is notable that its *mob* protein sequences have no detectable similarity to those of any of the above plasmids. It also seems to possess a much simpler structure, with only two *mob* genes in tandem array (Van Putten et al. 1987). This dissimilarity is reflected in the fact that CloDF13 *oriT* is not mobilized by ColE1, ColK or ColA *mob* (and vice-versa). We must conclude that CloDF13 belongs to a distinct lineage. The greater complexity of the ColE1 group might enable its members to be mobilized by a larger range of conjugative plasmids than CloDF13, but this remains to be tested.

The basic genetic characterization of *mob* described in this paper allows us to proceed with the study of how ColE1 genes are expressed: the overlapping gene structure suggests that such expression may be complex. We assume that the cAMP-dependent promoter (Queen and Rosenberg 1981) just upstream of *oriT* and conserved in ColE1, ColK and ColA is responsible for the transcription of some or all of the *mob* genes, since there are no other genes under its control. It is possible, though, that it serves merely to provide an RNA molecule involved in mobilization. The overlapping nature of two *mob* genes necessarily complicates their expression: if there is a single *mob* transcript, for example, it is easy to see how *mobC* and *mobA* might

be translated, but it is more difficult to conceive of how ribosomes could initiate translation of *mobB* and *mobD* when the mRNA is loaded with ribosomes already actively translating *mobA*. However, in minicells (Fig. 3), the relative yield of *mbeA* product is very low even though transcription is driven by the strong P_{tac} promoter: this suggests that *mbeA* is poorly translated. If this holds true for *mbk* and *mba*, it may be that the proposed translational coupling (see Results) between *mobC* and *mobA* is inefficient, or that there is a translational pause site downstream of *mobC* causing ribosomes to stall: in either case ribosomes could attach to the naked mRNA upstream of *mobB* and initiate translation of the overlapping genes, at least on a significant fraction of the transcripts. Alternatively, there may be a separate transcript for *mobB/mobD* which could be produced either by processing of a long, primary message or by de novo synthesis from a promoter within the *mobA* reading frame. Although we have no experimental data, we have searched for promoter-like sequences (Harr et al. 1983) within *mobA*-I and find several in all three plasmids (data not shown). None of these is significantly conserved however, so transcript mapping and experiments with promoter-probe vectors need to be performed.

The observation of Lovett and Helinski (1975) that the relaxation complex of ColE1 contains three proteins of 60 kDa, 16 kDa and 11 kDa correlates well with the results of our minicell experiments: it seems likely that these proteins are encoded by the *mbeA, mbeB* and *mbeC* genes, respectively (Fig. 3). However, we need to confirm this by reconstitution experiments. Our study is currently concentrated, therefore, on purifying *mbe* proteins from overproducing strains in order to analyse their interactions with each other and with *oriT*. The initial aim is to reproduce the nicking reaction at *oriT* with purified components.

Acknowledgements. This work was supported by the MRC and SERC.

References

Amann E, Brosius J, Ptashne M (1983) Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. Gene 25:167–178

Archer JAK (1985) Sequence analysis of plasmid ColK. PhD Thesis, University of Glasgow, Scotland

Bachmann BJ (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol Rev 36:525–557

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Boyd AC, Sherratt DJ (1986) Polar mobilization of the *E. coli* chromosome by the ColE1 transfer origin. Mol Gen Genet 203:496–504

Boyd AC, Charles IG, Keyte JW, Brammar WJ (1986) Isolation and computer-aided characterization of *Mme*I, a Type II restriction endonuclease from *Methylophilus methylotrophus*. Nucleic Acids Res 14:5255–5274

Brosius J (1984) Toxicity of an overproduced foreign gene product in *Escherichia coli* and its use in plasmid vectors for the selection of transcription terminators. Gene 27:161–172

Cesareni G, Muesing MA, Polisky B (1982) Control of ColE1 DNA replication: the *rop* gene product negatively affects transcription from the replication primer promoter. Proc Natl Acad Sci USA 79:6313–6317

Chan PT, Ohmori H, Tomizawa J-I, Lebowitz J (1985) Nucleotide sequence and gene organization of ColE1 DNA. J Biol Chem 260:8925–8935

Chen H-Z, Zubay G (1983) Analysis of ColE1 expression *in vitro* after chromosome fragmentation. J Bacteriol 154:650–655

Clewell DB, Helinski DR (1969) Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. Proc Natl Acad Sci USA 62:1159–1166

Collins J (1979) Cell-free synthesis of protein coding for mobilization functions of ColE1 and transposition functions of Tn*3*. Gene 6:29–42

Demerec M, Adelberg EA, Clark AJ, Hartman PE (1966) A proposal for a uniform nomenclature in bacterial genetics. Genetics 54:61–76

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dougan G, Sherratt DJ (1977) The transposon Tn*1* as a probe for studying ColE1 structure and function. Mol Gen Genet 151:151–160

Ebina Y, Kishi F, Nakazawa T, Nakazawa A (1979) Gene expression *in vitro* of colicin E1 plasmid. Nucleic Acids Res 7:636–649

Finnegan J, Sherratt D (1982) Plasmid ColE1 conjugal mobility: the nature of *bom*: a region required in *cis* for transfer. Mol Gen Genet 185:344–351

Fox TD (1987) Natural variation in the genetic code. Annu Rev Genet 27:67–91

Glickman B, Van Den Elsen P, Radman M (1978) Induced mutagenesis in *dam*$^-$ mutants of *E. coli:* a role for 6-methyladenine residues in mutation avoidance. Mol Gen Genet 163:307–312

Grundstrom T, Zenke WM, Wintzerith M, Matthes HWD, Staub A, Chambon P (1985) Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis. Nucleic Acids Res 13:3305–3316

Hallewell RA, Sherratt DJ (1976) Isolation and characterization of ColE2 plasmid mutants unable to kill colicin-sensitive cells. Mol Gen Genet 146:239–245

Harr R, Haggstrom M, Gustaffson P (1983) Search algorithm for pattern matching analysis of nucleic-acid sequences. Nucleic Acids Res 11:2943–2957

Inselburg J, Applebaum B (1978) Proteins synthesized in minicells containing plasmid ColE1 and its mutants. J Bacteriol 133:1444–1451

Ippen-Ihler KA, Minkley EG (1986) The conjugation system of F, the fertility factor of *Escherichia coli*. Annu Rev Genet 20:593–624

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lovett MA, Helinski DR (1975) Relaxation complexes of plasmid DNA and protein. II. Characterization of the proteins associated with the unrelaxed and relaxed complexes of plasmid ColE1. J Biol Chem 250:8790–8795

Messing J, Crea R, Seeburg PH (1981) A system for shotgun DNA sequencing. Nucleic Acids Res 9:309–321

Miller JH (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, New York

Morlon J, Chartier M, Bidaud M, Lazdunski C (1988a) The complete nucleotide sequence of the colicinogenic plasmid ColA. High extent of homology with ColE1. Mol Gen Genet 211:231–243

Morlon J, Sherratt D, Lazdunski C (1988b) Identification of functional regions of the colicinogenic plasmid ColA. Mol Gen Genet 211:223–230

Norrander J, Kempe T, Messing J (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101–106

Oppenheim DS, Yanofsky C (1980) Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics 95:785–795

Queen C, Rosenberg M (1981) A promoter of pBR322 activated by cAMP receptor protein. Nucleic Acids Res 9:3365–3377

Sanger F, Coulson AR, Hong GF, Hill DF, Peterson GB (1982)

Nucleotide sequence of bacteriophage λ DNA. J Mol Biol 162:729–773

Shine J, Dalgarno L (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342–1346

Stewart GSAB, Lubinsky-Mink S, Jackson CG, Cassel A, Kuhn J (1986) pHG165: a pBR322 copy number derivative of pUC8 for cloning and expression. Plasmid 15:172–181

Stormo GD, Schneider TD, Gold L, Ehrenfeucht A (1982) Use of the 'Perceptron' algorithm to distinguish translational initiation sites in *E. coli*. Nucleic Acids Res 10:2997–3011

Summers D, Yaish S, Archer J, Sherratt D (1985) Multimer resolution systems of ColE1 and ColK: localisation of the crossover site. Mol Gen Genet 201:334–338

Tautz D, Renz M (1983) An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. Anal Biochem 132:14–19

Twigg AJ, Sherratt DJ (1980) Trans-complementable copy-number mutants of plasmid ColE1. Nature 283:216–218

Van Putten AJ, Jochems GJ, de Lang R, Nijkamp HJJ (1987) Structure and nucleotide sequence of the region encoding the mobilization proteins of plasmid CloDF13. Gene 51:171–178

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268

Warren G, Sherratt D (1977) Complementation of transfer deficient ColE1 mutants. Mol Gen Genet 151:197–201

Willetts NS, Wilkins B (1984) Processing of plasmid DNA during bacterial conjugation. Microbiol Rev 48:24–41

Communicated by W. Goebel

Received December 20, 1988