Volume 15 Number 18 1987 Nucleic Acids Research

Nucleotide sequence of the transcriptional repressor gene *korB* which plays a key role in regulation of the copy number of broad host range plasmid RK2

Bimal D.M.Theophilus and Christopher M.Thomas*

Department of Genetics, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK

Received August 3, 1987; Accepted September 1, 1987 Accession no.Y00448

ABSTRACT

The product of the korB gene of broad host range plasmid RK2 is one of at least two proteins which repress transcription of the essential replication gene trfA. We report here the nucleotide sequence of korB and the properties of its predicted polypeptide product KorB which has a molecular weight of 39,011 Da. KorB is likely to be a soluble protein with an overall net negative charge. However, consistent with a role in transcriptional regulation, there is a region with extensive homology to the αhelix-turn-αhelix motif of many DNA binding proteins. This region shows no significant homology to equivalent regions of the TrfB protein which is the primary transcriptional repressor of RK2 and which binds to an operator whose half sites show considerable homology to the half sites of the korB operator.

INTRODUCTION

The copy number of broad host range plasmid RK2 is apparently regulated by limiting the supply of TrfA protein (1) which activates initiation of replication from a cis-acting vegetative replication origin oriV (2,3). The trfA gene is transcribed as part of an operon which is subject to repression by the protein products of at least two genes, trfB (identical to korA and korD) and korB (4). Putative operators to which these proteins bind have been identified (5,6) and are noteworthy with respect to their similarity, differing primarily in the spacing between the two halves of the palindrome. The predicted amino acid sequence of the TrfB protein includes a region with extensive homology to the consensus αhelix-turn-αhelix motif found in many DNA binding proteins (7,8). As a further part of our study of transcriptional regulation in broad host range plasmid RK2 we have determined the nucleotide sequence of the korB gene and report here its properties together with those predicted for its protein product KorB. While comparison with the consensus αhelix-turn-αhelix motif has revealed a possible DNA binding domain this does not show significant homology to the equivalent structure in TrfB.

MATERIALS AND METHODS

Bacterial strains and plasmids

Escherichia coli K12 strain C600K⁻ (*thr-1 leu-6 thi-1 lacY1 supE44 tonA21 galK*) was used throughout. Plasmids used were: pCT413 (1); pCT413::Tn*1725*#10, containing Tn*1725* inserted at the position shown in Figure 1; pBIM15, containing the *HindIII* fragment from pCT413 containing *korB*, inserted into the *HindIII* site of pEMBL19 (8a); deletion derivatives of pBIM15 generated by linearizing plasmid DNA by cleavage at the unique *EcoRV* site, *Bal31* digestion and addition of *BamHI* linkers.

Growth media

L-broth (9) was used as liquid medium while agar was added to 1.5% (w/v) for solid medium. Penicillin resistance was selected by addition of penicillin-G to 150μg/ml (liquid medium) or 300μg/ml (solid medium). Chloramphenicol was added to 25μg/ml to select chloramphenicol resistance.

Plasmid DNA isolation, analysis and manipulation

Plasmid DNA was isolated according to Birnboim and Doly (10) with minor modifications (11). If necessary large scale preparations were further purified by CsCl/Ethidium bromide gradients (11). DNA restriction and modifying enzymes were obtained from Bethesda Research Laboratories, nbl or BCL and used essentially as recommended. Restriction fragments were analysed by standard agarose gel electrophoresis.

DNA sequence determination and analysis

DNA sequence was determined by the method of Maxam and Gilbert (12) with minor modifications as previously described (13). DNA sequence was analysed using a computer package 'DNA' courtesy of G. Ware, Department of Microbiology, University of Bristol.

Transposon insertion mutagenesis

Transposon mutagenesis using Tn*1725* (14) was as previously described (15).

RESULTS AND DISCUSSION

Figure 1 shows a physical and genetic map of the region of RK2 containing *korB*. The nucleotide sequence for the N-terminal section of this gene has previously been reported by us as part of the *trfB* region of mini plasmids derived from RK2 (7). The strategy for obtaining the nucleotide sequence of both DNA strands for the rest of *korB* is also shown in Figure 1. Figure 2 shows the nucleotide sequence of *korB* and the predicted amino acid sequence of the KorB protein. Position 1181 was

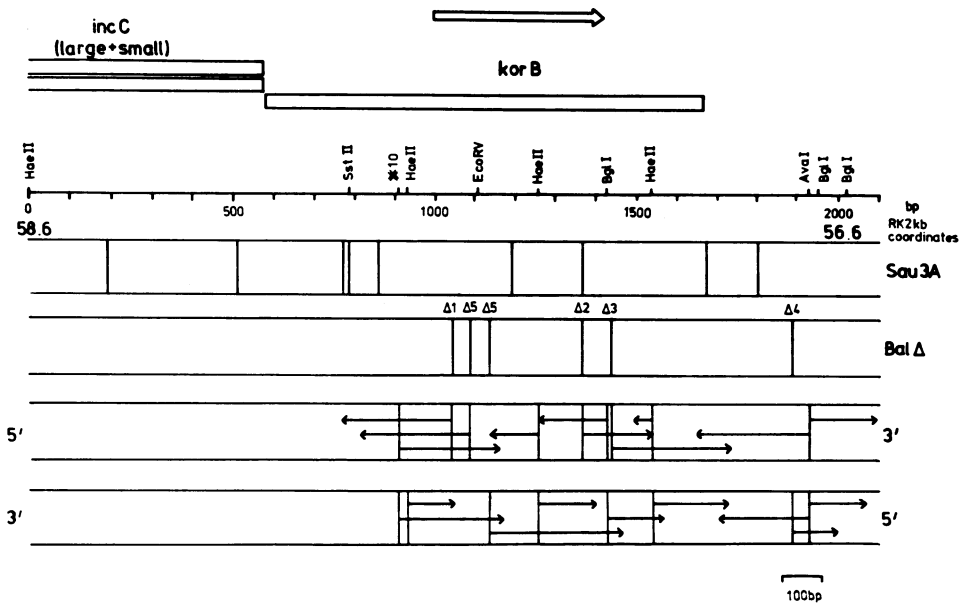
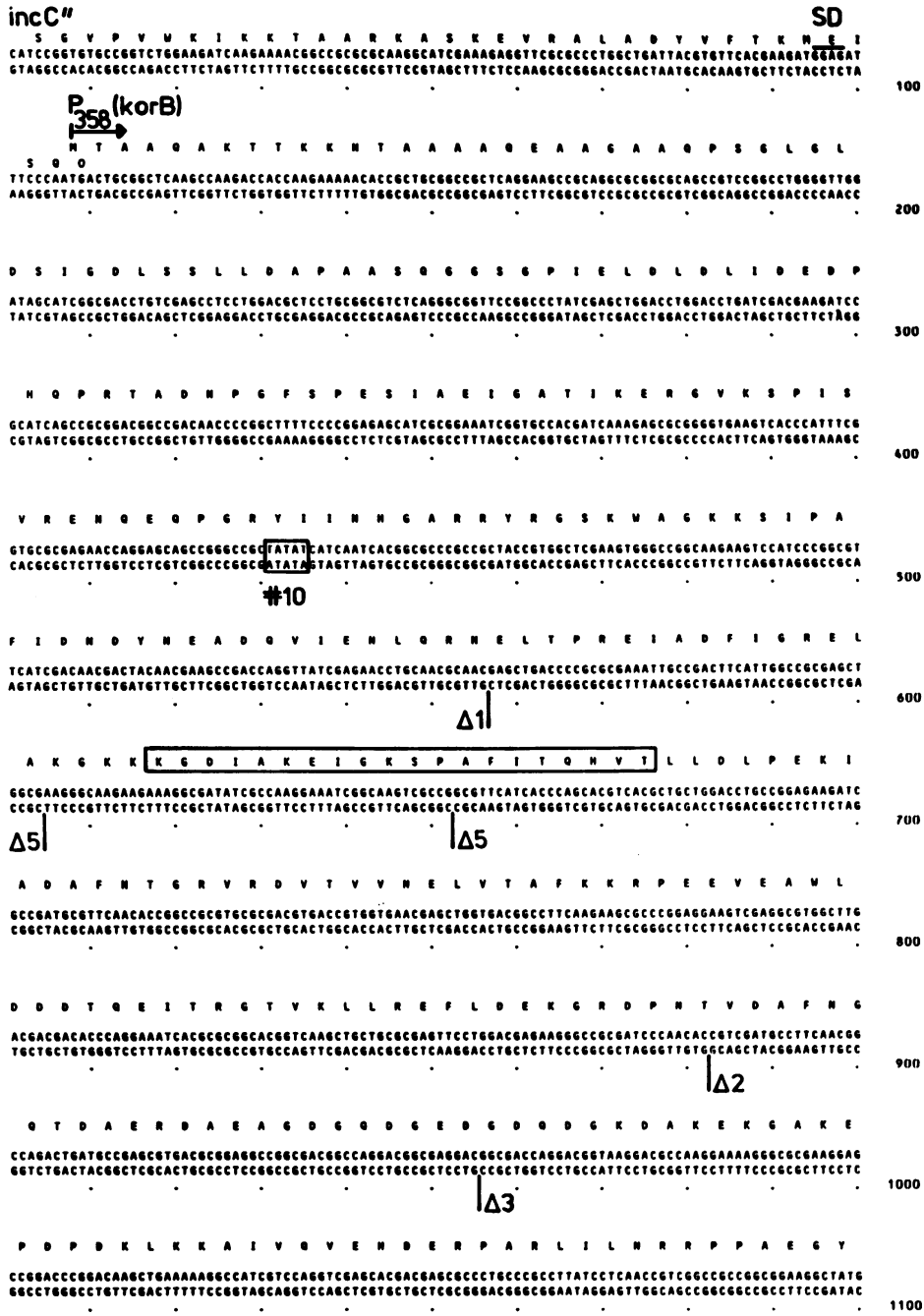


Figure 1. Restriction map of the korB region showing sequencing strategy. RK2 coordinates shown are defined clockwise relative to the unique EcoRI site on the 60 Kb genome. Restriction sites with 5' protruding ends were 5'-labelled, after calf-intestinal phosphatase treatment, with T4 polynucleotide kinase and $[\gamma\text{-}^{32}P]$-ATP, or 3'-labelled using DNA polymerase large fragment and $[\alpha\text{-}^{32}P]$ dNTP; restriction sites with 3' protruding or blunt ends were 3'-labelled with terminal transferase and $[\alpha\text{-}^{32}P]$ ddATP. Sequence determined is shown separately for each strand. #10 indicates an EcoRI site generated by Tn1725 insertion into pCT413. BalΔ 1-5 represent BamHI sites from linkers ligated into Bal31 deletion derivatives generated from the EcoRV site of pBIM15. The C-terminal of the incC ORFs and the extent of the korB ORF are shown (▭) as is the direction of transcription through the region (⟹).

changed from C to T on examining our gels in light of a conference communication by J. Kornacki and D. Figurski in which they also presented the korB sequence. Previous estimates of the size of the KorB polypeptide based on SDS/PAGE analysis of radioactively labelled proteins synthesized in maxicells are 49 kDa (16) and 52 kDa (17). There is only one ORF in this region capable of coding for a polypeptide approaching this size. The predicted size of this polypeptide is 39,011 Da. This discrepancy between apparent mobility and predicted size from DNA sequence analysis may be explained by the high overall negative charge of KorB, interfering with its complete denaturation by SDS. The korB ORF runs from an ATG start codon overlapping the stop codon of incC through to an opal stop codon (TGA) at


incC''
SD
P358 (korB)
#10
Δ1
Δ2
Δ3
Δ5
Δ5
100
200
300
400
500
600
700
800
900
1000
1100

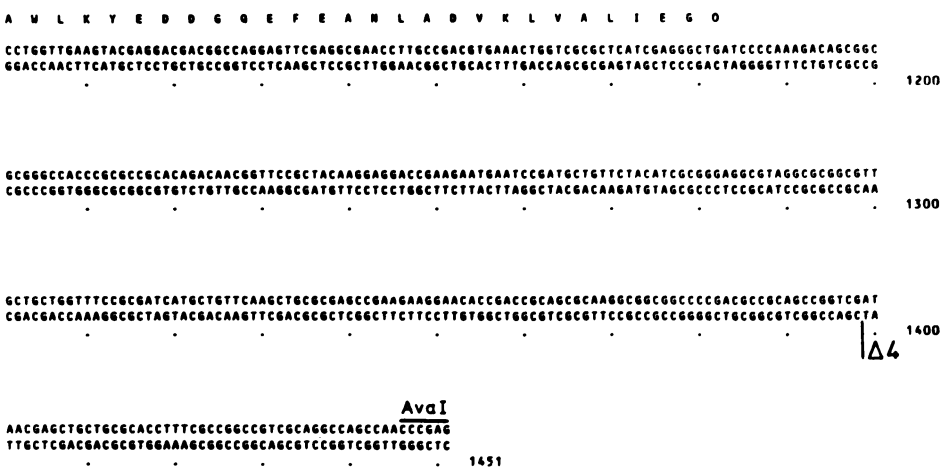


Figure 2. Nucleotide sequence of the korB region and amino acid sequence of the predicted polypeptide. The target site duplicated by Tn1725 insertion #10 into pCT413 is shown (boxed 5bp sequence), as are the end points of Bal31 deletions (Δ 1-5). The putative Shine-Dalgarno sequence for korB is marked. The standard one letter amino acid code is used to show the predicted KorB polypeptide, and the C-terminal section of IncC (incC"). the putative αhelix-turn-αhelix DNA binding region is boxed.

position 1181-1183. While this may lead to some translational coupling between incC and korB the korB ORF is preceded by a good Shine-Dalgarno sequence and korB expression does not apparently depend on incC translation (16). No obvious candidate for a transcriptional promoter is found within a few hundred base pairs upstream of korB, consistent with our previous observation that korB is transcribed from the trfB promoter some 1100 bp upstream. Previously the functional C-terminal end point of korB has been predicted by the study of Sau3A deletion derivatives (16) and Bal31-derived deletion derivatives (17). In the latter case (17) a KorB$^-$ phenotype was only obtained when deletions had gone approximately 200 bp past the AvaI site downstream of korB. The end point of the proposed korB ORF is consistent with this functional mapping of the korB gene.

It has been found that functional ORFs can be predicted by analysis of the ratio of RNY to YNR triplets (where R is purine; N is any base: Y is pyrimidine) in all three phases (18). The functional ORF is that which has the largest excess of RNY over YNR. Sequential analysis of all subsections of the korB sequence in all three phases reveals a large excess of RNY over YNR only in the proposed korB ORF, providing further support for its

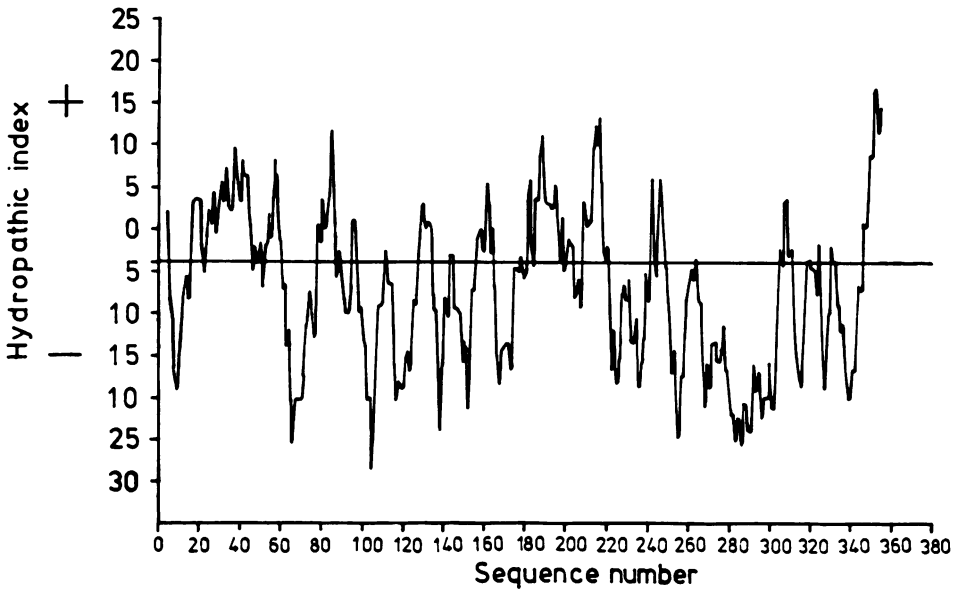


Figure 3. Hydropathy plot of the predicted KorB polypeptide. This was calculated and plotted according to ref. 19, using a 9 residue span for successive sums of overlapping segments. Increasing positive values represent increasing hydrophobicity.

validity. Codon usage by the korB ORF is similar to that found previously for trfA (13) and incC (7) and agrees with the bias expected due to the higher G+C content of RK2 plasmid DNA (an average of 65% for korB, 59% for trfA and 63% for incC) compared to the value of 50% G+C for E.coli.

Figure 3 shows a hydropathy plot for KorB which reveals KorB as being remarkably hydrophilic and provides no evidence for it being membrane associated. It is therefore likely to exist in a soluble form in the bacterial cytoplasm.

The KorB polypeptide includes a region (amino acid residues 171-190) which should be able to adopt the αhelix-turn-αhelix motif found in many regulatory proteins which bind to DNA, Figure 4 (8). While this region was initially identified by simple homology searches followed by visual inspection, the recent systematic method of Dodd and Egan (20) confirmed our proposal and revealed no other possible candidate. The score achieved by this method for this region of KorB was 1161, compared to 1141 for the equivalent structure in TrfB (7). The consensus for this structure comprises a distribution of hydrophobic and hydrophilic residues consistent

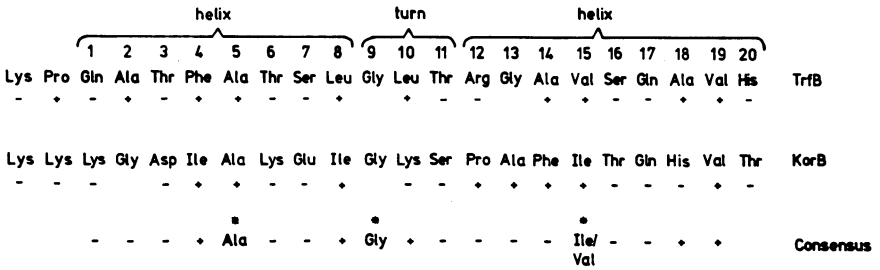


Figure 4. Potential α-helix-turn-αhelix regions of KorB and TrfB. Regions of KorB and TrfB with potential to form an α-helix-turn-αhelix are aligned with a concensus for this motif (8). Asterisks indicate highly conserved residues. Hydrophobic residues are indicated by +, while hydrophilic residues are indicated by -.

with one side of the α-helices exposed for binding DNA and with the other side facing towards the centre of the protein. This possible DNA-binding domain of KorB also shares with the consensus three highly conserved residues (ala, gly and ile/val at relative positions 5, 9 and 15, Figure 4) which are important in forming and maintaining the correct angle between the helices. Whereas the predicted KorB polypeptide has an overall net excess charge of -21 (70 aspartate and glutamic acid residues compared to 49 arginine and lysine residues) this region has a net excess charge of +1, and moreover three of the four residues preceding it are positively charged (the fourth, gly, is uncharged). This is consistent with a role for this region in DNA binding. Despite the close similarity of their half-operator sites, the proposed DNA-binding domains of TrfB and KorB share only 4/20 amino acids in common (Figure 4). Two of these occur at positions which are highly conserved in all such structures and KorB shows in this region as much homology to other proteins with such DNA-binding domains as to TrfB. In addition, no significant homology between other regions of TrfB and KorB has been detected. This suggests that TrfB and KorB have not evolved by duplication of an ancestral plasmid repressor gene.

It is of interest that since the combination of TrfB and KorB repress 100-fold transcription from the trfA and trfB promoters while on their own they repress transcription only 10 and 2-4 fold respectively (4,6) it is possible that they interact cooperatively. In addition, it is unclear why KorB is so much larger than TrfB, given that so far we only know of its role in transcriptional repression. Further analysis of the action of TrfB and KorB will depend on purification of these repressor proteins which is

currently underway and has been facilitated by the DNA sequence reported here.

ACKNOWLEDGEMENTS

We thank Dr. C.A. Smith for helpful discussions and Mrs. E. Badger for typing the manuscript. BDMT was supported by an SERC Quota Studentship while the research was financed, in part, by MRC Project Grant No. G8511172CB.

*To whom correspondence should be addressed

REFERENCES

1. Thomas, C.M. and Hussain, A.A.K. (1984) EMBO J. 3, 1513-1519.
2. Meyer, R. and Helinski, D.R. (1977) Biochim. Biophys.Acta 478, 109-113.
3. Thomas, C.M., Meyer, R. and Helinski, D.R. (1980) J. Bacteriol. 141, 213-222.
4. Shingler V. and Thomas, C.M. (1984) Mol. Gen. Genet. 195, 523-529.
5. Smith, C.A., Shingler, V. and Thomas, C.M. (1984) Nucleic Acids Res. 12, 3619-3630.
6. Theophilus, B.D.M., Cross, M.A., Smith, C.A. and Thomas, C.M. (1985) Nucleic Acids Res. 13, 8129-8142.
7. Thomas, C.M. and Smith, C.A. (1986) Nucleic Acids Res. 14, 4453-4469.
8. Pabo, C.O. and Sauer, R.T. (1984) Ann. Rev. Genet. 17, 499-525.
8a. Dente, L., Solazzo, M. Baldari, C., Cesareni, G. and Cortese, R. (1985) in DNA Cloning. A practical approach Vol. I. Glover, D.M. ed., pp. 101-107. IRL Press, Oxford.
9. Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut, E. and Helinski, D.R. (1979) in 'Methods in Enzymology Vol. 68', Wu, R. ed., pp. 268-280. Academic Press, New York.
10. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
11. Smith, C.A. and Thomas, C.M. (1983) Mol. Gen. Genet. 190, 245-254.
12. Maxam, A.M. and Gilbert, W. (1980) in 'Methods in Enzymology Vol. 65', Grossman, L. and Moldave, K. Eds., pp. 499-560. Academic Press, New York.
13. Smith, C.A. and Thomas, C.M. (1984) J. Mol. Biol. 175, 229-250.
14. Ubben, D. and Schmitt, R. (1986) Gene 41, 145-152.
15. Shingler, V. and Thomas, C.M. (1984) J. Mol. Biol. 175, 229-250.
16. Smith, C.A. and Thomas, C.M. (1984) J. Gen. Microbiol. 130, 1651-1663.
17. Bechhofer, D.H., Kornacki, J.A., Firshein, W. and Figurski, D.H. (1986) Proc. Natn. Acad. Sci. USA 83, 394-398.
18. Shepherd, J.C.W. (1981) Proc. Natn. Acad. Sci. USA 78, 1596-1600.
19. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
20. Dodd, I.B. and Egan, J.B. (1987) J. Mol. Biol. 194, 557-564.