Volume 246, number 1,2, 39–43 FEB 06939 March 1989

Nucleotide sequence of the aliphatic amidase regulator gene (*amiR*) of *Pseudomonas aeruginosa*

N. Lowe, P.M. Rice* and R.E. Drew

Department of Biochemistry, University College London, Gower Street, London WC1E 6BT, England and **European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG*

Received 26 January 1989

The nucleotide sequence of a 1001 bp *ClaI/XhoI* DNA fragment encoding the amidase regulator gene (*amiR*) from *Pseudomonas aeruginosa* has been determined. The sequence derives from strain PAC433, a constitutive high expressing amidase mutant, and contains two overlapping open reading frames. Analysis of the sequence has identified one of the reading frames as *amiR*. The gene encodes a 196 amino acid polypeptide which shows a strong bias towards codons with G or C in the third position. The *amiR* gene shows no sequence homology with other bacterial regulator proteins.

Aliphatic amidase regulator; Coding sequence; (*Pseudomonas aeruginosa*)

1. INTRODUCTION

Pseudomonas aeruginosa is able to grow on a limited number of short chain length aliphatic amides by virtue of an inducible amidase activity (EC 3.5.1.4) [1,2]. Cotransduction studies have shown that the amidase structural gene (*amiE*) and regulator gene (*amiR*) are closely linked [3]. Amidase expression is regulated by at least two control systems. The product of the *amiR* gene functions as a positive controlling element [4] and the system is also subject to catabolite repression by succinate [5]. Lambda recombinant bacteriophages have been constructed which carry the amidase genes from strain PAC433 [6]. This strain expresses high levels of amidase constitutively under all growth conditions and expression is resistant to both catabolite and butyramide repression [7]. The amidase genes have been sub-cloned into plasmid pBR322 and the location of *amiE* determined [8]. The DNA sequence and amino acid sequence of amidase (*amiE*) have recently been determined [9,10]. Amidase expression from the cloned *amiE* gene regulated by trans-complementation has been investigated in *Ps. aeruginosa* and the positive control model of expression confirmed [11]. Using in vitro constructed deletions and subcloning of DNA fragments into a broad host range plasmid vector the *amiR* gene has been located to a 1 kbp *ClaI/XhoI* fragment some 2 kbp downstream of *amiE* and transcribed in the same direction [12]. Preliminary in vitro transcription/translation studies using various recombinant plasmids have identified amiRp at 22 kDa [13]. Recent investigations of the *amiE* promoter region have shown that the regulator protein (amiRp) functions as a transcription anti-termination factor [14]. In this communication we present the DNA sequence of the 1 kbp *ClaI/XhoI* fragment which encodes the amidase regulator gene.

Correspondence address: R.E. Drew, Department of Biochemistry, University College London, Gower Street, London WC1E 6BT, England

2. MATERIALS AND METHODS

2.1. *Materials*

DNA cloning and sequencing kits, restriction enzymes and T_4 ligase were from BRL (Gibco). All other reagents were of Analar quality and obtained from local suppliers.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

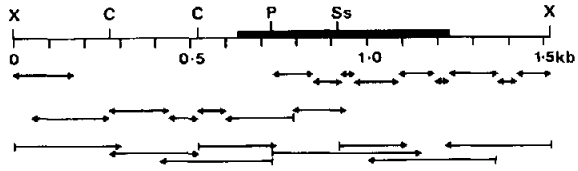

Fig.1. Strategy for sequencing the *amiR* gene. The map shows the position of some restriction enzyme targets within the 1.52 kbp *Xho*I fragment. The *amiR* coding region is shown by the thickened line. The horizontal arrows show the direction and extent of sequencing reactions. The top group are *Sau*3A fragments, the middle group *Taq*I fragments and the bottom group are unique fragments obtained using hexanucleotide recognition restriction enzymes (see section 2). The final sequence was determined by overlapping the *Sau*3A and *Taq*I sequences. Restriction sites are: *Cla*I (C), *Pvu*II (P), *Sst*II (Ss) and *Xho*I (X).

2.2. *DNA preparations*

Plasmid DNAs were prepared by detergent lysis and CsCl-ethidium bromide equilibrium-density centrifugation [15].

2.3. *Preparation of sequencing templates*

The 1.5 kbp *Xho*I fragment carrying *amiR* (fig.1) was obtained after *Xho*I digestion of plasmid pJB950 [8] and separation by electrophoresis on low-melting-point agarose gels. The fragment was then excised and purified by extraction with hot phenol and recovered by ethanol precipitation. The purified fragment was digested with additional restriction enzymes (see below) and fragments generated were ligated to appropriately cut M13mp8 or M13mp9 vectors [16]. Recombinant DNAs were recovered as phage from colourless plaques after transfection of *E. coli* JM101 or JM103 [17]. Single-stranded DNAs for use as templates were purified by standard methods [18].

2.4. *DNA sequencing*

M13mp8 and M13mp9 recombinants were obtained containing *Cla*I, *Sau*3A, *Taq*I, *Xho*I; *Cla*I/*Sau*3A, *Pvu*II/*Xho*I, *Sau*3A/*Xho*I and *Sst*II/*Xho*I fragments and sequenced using the dideoxy method [19], ^{35}S-ATP and universal primer. The sequencing strategy is shown in fig.1.

2.5. *DNA sequence analysis*

The DNA sequence analysis (fig.3) used the GCG software package (version 5.3) from the University of Wisconsin [20]. Each panel represents one of the three reading frames. The codon preference statistic [21] and third position GC bias [22] were each calculated over a window of 25 codons with a codon usage table obtained using published *Ps. aeruginosa* sequences from the EMBL library, release 14.0, plus the *amiE* sequence [9]. The reference lines at codon preference of 0.84 and GC bias of 0.65 show the values for random sequence with the same base composition. Rare codons, with a codon preference below 0.1 are indicated.

3. RESULTS AND DISCUSSION

The nucleotide sequence of the *Cla*I/*Xho*I fragment encoding the amidase regulator gene of *Ps. aeruginosa* is shown in fig.2. The sequence contains two long overlapping reading frames. RF(2) begins outside the sequence presented and terminates at residue 739, RF(3) starts at 123 and terminates at 710. We believe that *amiR* lies in RF(3) (see below). Translation of RF(3) is shown in fig.2 which comprises 588 nucleotides starting with an ATG codon and terminating with TGA. The polypeptide encoded by this sequence is 196 amino acids with a molecular mass of 21 776 Da. Previous

```
        10         20         30         40         50         60
ATCGATGCGCGCGGCGTGTTCCAGGTCCGCTGGCAGTCGCCCGAACCGATTCGCCCCGAC

        70         80         90        100        110        120
CCTTATGTCGTCGTGCATAACCTCGACGACTGGTCCGCCAGCATGGGCGGGGGACCGCTC

       130        140        150        160        170        180
CCATGAGCGCCAACTCGCTGCTCGGCAGCCTGCGCGAGTTGCAGGTGCTGGTCCTCAACC
  MetSerAlaAsnSerLeuLeuGlySerLeuArgGluLeuGlnValLeuValLeuAsnP

       190        200        210        220        230        240
CGCCGGGGGAGGTCAGCGACGCCCTGGTCTTGCAGCTGATCCGCATCGGTTGTTCGGTGC
roProGlyGluValSerAspAlaLeuValLeuGlnLeuIleArgIleGlyCysSerValA

       250        260        270        280        290        300
GCCAGTGCTGGCCGCCGCCGGAAGCCTTCGACGTGCCGGTGGACGTGGTCTTCACCAGCA
rgGlnCysTrpProProProGluAlaPheAspValProValAspValValPheThrSerI

       310        320        330        340        350        360
TTTTCCAGAATGGCCACCAGCACGAGATCGCTGCGCTGCTCGCCGCCGGGACTCCGCGCA
lePheGlnAsnArgHisHisAspGluIleAlaLeuLeuLeuAlaAlaGlyThrProArgI

       370        380        390        400        410        420
CTACCCTGGTGGCGCTGGTGGAGTACGAAAGCCCCGCGGTGCTCTCGCAGATCATCGAGC
hrThrLeuValAlaLeuValGluTyrGluSerProAlaValLeuSerGlnIleIleGluL

       430        440        450        460        470        480
TGGAGTGCCACGGCGTGATCACCCAGCCGCTCGATGCCCACGGGTGCTGCCTGTGCTGG
euGluCysHisGlyValIleThrGlnProLeuAspAlaHisArgValLeuProValLeuV

       490        500        510        520        530        540
TATCGGCGCGGCGCATCAGCGAGGAAATGGCGAAGCTGAAGCAGAAGACCGAGCAGCTCC
alSerAlaArgArgIleSerGluGluMetAlaLysLeuLysGlnLysThrGluGlnLeuG

       550        560        570        580        590        600
AGGACCGCATCGCCGGCCAGGCCGGATCAACCAGGCCAAGGTGTTGCTGATGCAGCGCC
lnAspArgIleAlaGlyGlnAlaArgIleAsnGlnAlaLysValLeuLeuMetGlnArgH

       610        620        630        640        650        660
ATGGCTGGGACGAGCGCGAGGCGCACCAGCACCTGTCGCGGGAAGCGATGAAGCGGCGCG
isGlyTrpAspGluArgGluAlaHisGlnHisLeuSerArgGluAlaMetLysArgArgG

       670        680        690        700        710        720
AGCCGATCCTGAAGATCGCTCAGGAGTTGCTGGGAAACGAGCCGTCCGCCTGAGCGATCC
luProIleLeuLysIleAlaGlnGluLeuLeuGlyAsnGluProSerAlaEnd

       730        740        750        760        770        780
GGGCCGACCAGAACAATAACAAGAGGGGTATCGTCATCATGCTGGGACTGGTTCTGCTGT

       790        800        810        820        830        840
ACGTTGGCGCGGTGCTGTTTCTCAATGCCGTCTGGTTGCTGGGCAAGATCAGCGGTCGGG

       850        860        870        880        890        900
AGGTGGCGGTGATCAACTTCCTGGTCGGCGTGCTGAGCGCCTGCGTCGCGTTCTACCTGA

       910        920        930        940        950        960
ACTTTTCCGCAGCAGCCGGGCAGGGCTGGCTGAAGGCCGGAGCGCTGACCCTGCTATTCG

       970        980        990       1000
CTTTTACCTATCTGTGGGTGGCCGCCAACCAGTTCCTCGAG
```

Fig.2. The nucleotide sequence of the *Cla*I/*Xho*I *amiR* gene fragment. The DNA strand shown has the same orientation as the mRNA. Nucleotides are numbered from the intact *Cla*I target. Translation of RF(3) is shown.

studies have indicated that the promoter for *amiR* expression lies some distance from the gene itself and is not expected to be present in this sequence [12]. Starting at nucleotide −14 from the translation start the sequence GGGA is partly complementary to the 3′-end of the 16 S rRNA from *Ps. aeruginosa* and presumably acts as the ribosome-binding site [23]. This region is less complementary to the 3′-end of the *E. coli* 16 S rRNA [24]. Unlike the *amiE* gene [9] there are no translation termination codons upstream of *amiR* and in fact from the analysis (fig.3) it appears that there is an additional open reading frame, RF(1), in this region.

For *amiR*, of the 196 codons used (table 1), codon usage is remarkably similar to that for *amiE* [9]. Only 41 of the possible 61 sense codons are used, and of these, only 17 out of 196 show A or T in the third position (8.7%). For some amino acids a single codon is used exclusively; UUC for phenylalanine (3), CAG for glutamine (14) and AAG for lysine (6). Additionally all arginines (14) are encoded by CGC/G. This highly biased codon usage shown by some *Pseudomonas* genes is reported to occur in genes capable of being expressed at high levels [9]. At present we are unsure about the regulation of *amiR* but previous studies [12] have shown that poor amidase expression in *E. coli* is due to low *amiR* expression. Thus in *Ps. aeruginosa* it remains a possibility that the positive regulator can be expressed at high levels.

Analysis of the DNA sequence is shown in fig.3. Previous studies using in vitro constructed deletions [12] had shown that *amiR* lay entirely within the *ClaI/XhoI* fragment and SDS-polyacrylamide gel electrophoresis of in vitro transcription/translation reaction products showed amiRp to be 22 kDa [13]. Measurement of the codon preference statistic, third position GC bias and rare codon usage identifies *amiR* in RF(3), the only RF carrying a complete gene of the correct size and showing the expected analysis patterns. The alter-

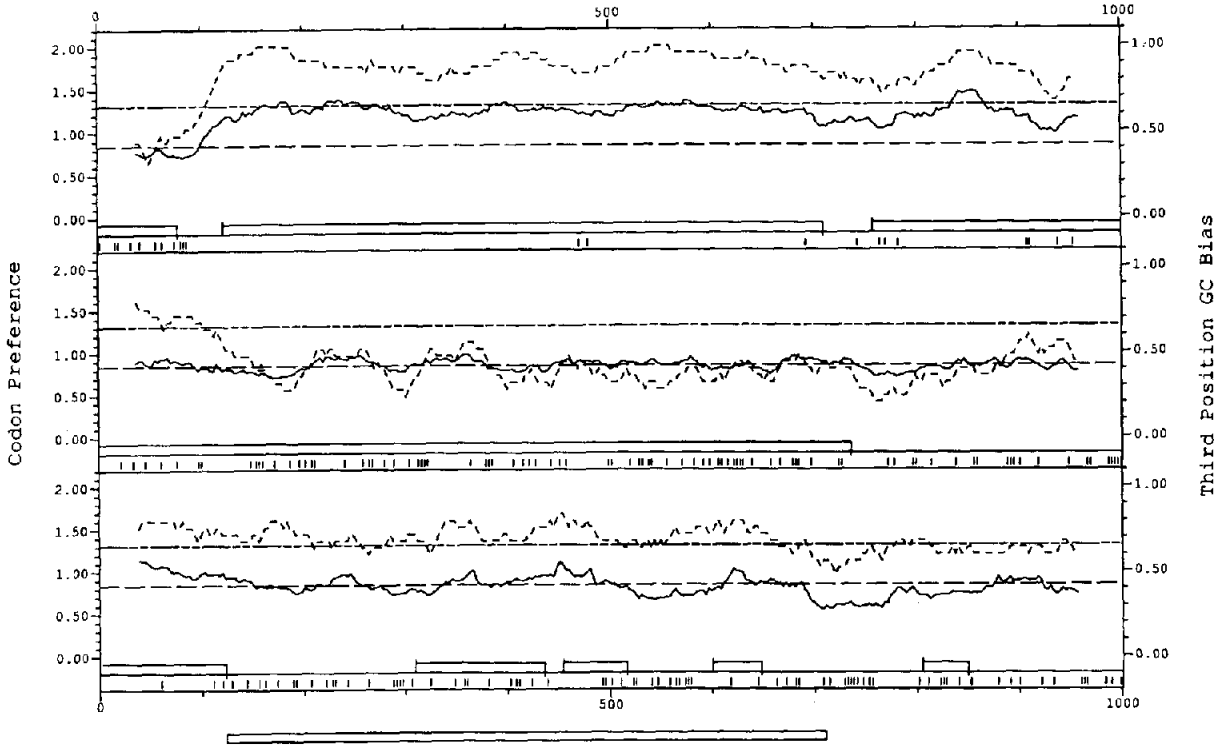

Fig.3. Codon preference analysis of *ClaI/XhoI* DNA fragment. The continuous line (——) shows the codon preference statistic, the dashed line (– – –) shows the third position GC bias and vertical bars (|) show rare codons. The reference lines at codon preference of 0.84 and GC bias of 0.65 show the values for random sequence with the same base composition. Boxes (⌐¬) show open reading frames from an ATG start and the box underneath delineates *amiR*. Each panel represents one reading frame, RF(3) top, RF(2) middle and RF(1) bottom.

Table 1

Codon-usage within the *amiR* gene

Phe	UUU	0	Ser	UCU	0	Tyr	UAU	0	Cys	UGU	1
	UUC	3		UCC	1		UAC	1		UGC	2
Leu	UUA	0		UCA	0	Ter	UAA	0	Ter	UGA	1
	UUG	4		UCG	5	Ter	UAG	0	Trp	UGG	2
Leu	CUU	0	Pro	CCU	1	His	CAU	1	Arg	CGU	0
	CUC	6		CCC	1		CAC	6		CGC	9
	CUA	0		CCA	0	Gln	CAA	0		CGA	0
	CUG	16		CCG	10		CAG	14		CGG	5
Ile	AUU	1	Thr	ACU	2	Asn	AAU	1	Ser	AGU	0
	AUC	11		ACC	4		AAC	4		AGC	6
	AUA	0		ACA	0	Lys	AAA	0	Arg	AGA	0
Met	AUG	4		ACG	0		AAG	6		AGG	0
Val	GUU	0	Ala	GCU	2	Asp	GAU	1	Gly	GGU	1
	GUC	4		GCC	10		GAC	6		GGC	5
	GUA	1		GCA	0	Glu	GGA	4		GGA	1
	GUG	12		GCG	7		GAG	13		GGG	2

native RF(2) carries an open reading frame of 951 nucleotides starting outside the sequence presented (−215) coding for a protein of approx. 35 kDa. This RF has the characteristics of random sequence with a high GC ratio (65%), many rarely used codons and codons of the type GGX(Gly), CCX(Pro), GCX(Ala) and CGX(Arg). This together with the predicted molecular mass make it most unlikely that RF(2) represents *amiR*. From the analysis it appears that *amiR* lies within a polycistronic transcript. There is the upstream reading frame, RF(1), whose termination codon overlaps the *amiR* initiation codon and downstream of *amiR*, in the same frame, is a further open reading frame. As yet we are unsure of the role of these genes, if any, in the regulation of amidase synthesis.

Homology searches of the EMBL, GenBank and SwissProt Libraries for DNA and amino acid sequences related to *amiR*, a transcription antitermination factor, have been unsuccessful. This DNA sequence will appear in the EMBL/GenBank/DDBJ nucleotide sequence databases under the accession number X13776 *amiR*.

Acknowledgements: We are grateful, as always, to Pat Clarke and Bill Brammar for advice and encouragement; to Alison Sparrow for her drawing expertise and to Stuart Wilson for his help with the mail link between UCL and Heidelberg. This project was funded by the SERC.

REFERENCES

[1] Kelly, M. and Clarke, P.H. (1962) J. Gen. Microbiol. 27, 305–316.
[2] Brammar, W.J. and Clarke, P.H. (1964) J. Gen. Microbiol. 37, 307–319.
[3] Brammar, W.J., Clarke, P.H. and Skinner, A.J. (1967) J. Gen. Microbiol. 47, 87–102.
[4] Farin, F. and Clarke, P.H. (1978) J. Bacteriol. 135, 379–392.
[5] Smyth, P.F. and Clarke, P.H. (1975) J. Gen. Microbiol. 90, 81–90.
[6] Drew, R.E., Clarke, P.H. and Brammar, W.J. (1980) Mol. Gen. Genet. 177, 311–320.
[7] Smyth, F.P. and Clarke, P.H. (1975) J. Gen. Microbiol. 90, 91–99.
[8] Clarke, P.H., Drew, R.E., Turberville, C., Brammar, W.J., Ambler, R.P. and Auffet, A.D. (1981) Biosci. Rep. 1, 299–307.
[9] Brammar, W.J., Charles, I.G., Matfield, M., Cheng-Pin, L., Drew, R.E. and Clarke, P.H. (1987) FEBS Lett. 215, 291–294.
[10] Ambler, R.P., Auffret, A.D. and Clarke, P.H. (1987) FEBS Lett. 215, 285–290.
[11] Drew, R.E. (1984) J. Gen. Microbiol. 130, 3101–3111.
[12] Cousens, D.J., Clarke, P.H. and Drew, R.E. (1987) J. Gen. Microbiol. 133, 2041–2052.
[13] Cousens, D.J. (1985) PhD Thesis, University of London.
[14] Drew, R.E. and Lowe, N. (1989) J. Gen. Microbiol., in press.
[15] Clewell, D. and Helinski, D. (1969) Proc. Natl. Acad. Sci. USA 62, 1159–1166.
[16] Norrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101–106.
[17] Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309–321.

[18] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[19] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[20] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[21] Gribskov, M., Devereux, J. and Burgess, R.R. (1984) Nucleic Acids Res. 12, 539–549.
[22] Bibb, M.J., Findlay, P.R. and Johnson, M.W. (1984) Gene 30, 157–166.
[23] Shine, J. and Dalgarno, L. (1975) Nature 254, 34–38.
[24] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.