Characterization of a halobacterial gene affecting bacterio-opsin gene expression

Mary Betlach, James Friedman, Herbert W. Boyer and Felicitas Pfeifer

Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Received 2 July 1984; Revised and Accepted 24 September 1984

ABSTRACT

A substantial number of spontaneous bacterio-opsin mutants of Halobacterium halobium are the result of insertion elements up to 1400 bp upstream of the bacterio-opsin (bop) gene. The nucleotide sequence of 1800 bp upstream of the bop gene has been determined. There is a 1118 bp open reading frame (ORF) located within this region which is transcribed and which coincides with the distribution of insertion elements upstream of the bop gene in Bop mutants. Therefore, we propose that there is a gene (brp gene) 526 bp upstream of the bop gene. This putative gene is transcribed in the opposite direction as the bop gene and could encode a protein of 37,500 D (359 amino acids) with a codon usage similar to bacterio-opsin. The 5' terminus of the brp transcript has been determined. The brp transcript and the bop mRNA are complementary for 13 residues near their 5' termini and both transcripts start at or near the initiating codon of the gene. Both transcripts could form similar hairpin loop structures at their 5' termini which contain possible ribosomal binding sites. The DNA sequences immediately upstream of the bop and the brp genes have significant homologies and there is a short complementary sequence. The role of the brp gene in bacterio-opsin gene expression is unclear.

INTRODUCTION

Bacterio-opsin complexed with the chromophore retinal acts as a light driven proton pump in the purple membrane of Halobacterium halobium (1). Depending on growth conditions, the cellular concentration of bacterio-opsin varies four to five fold (2), and at low oxygen tension the purple membrane can comprise up to 50% of the total cell surface (3,4). The gene which encodes bacterio-opsin (bop gene) has been cloned, the nucleotide sequence of the gene and neighboring sequences determined and its mRNA characterized (5,6,7,8). Mutants that overproduce purple membrane and bacterio-opsin have been isolated (e.g. H. halobium S9 and ET1001) (9) but the molecular basis of the mutation(s) is not known. Bacterio-opsin deficient mutants (Bop⁻) occur spontaneously in H. halobium with high frequencies (10). We have characterized 33 such mutants, 31 of which are caused by insertion elements (6,11,12). About two-thirds of these insertions occur within the bop gene

and one-third integrate at positions up to 1400 base pairs (bp) upstream of the bop coding region. This latter finding is of interest since de novo synthesis of bop mRNA initiates only 2 nucleotides upstream of the bop coding region (8). In this paper we present evidence for a gene upstream of the bop gene which is involved in the expression of bacterio-opsin or the formation of the purple membrane. About 1800 bp of nucleotide sequence upstream of the bop gene have been determined. There are open reading frames (ORFs) in both orientations, but RNA transcripts corresponding to only one ORF can be detected. This latter ORF is oriented in a direction opposite to the bop gene and correlates with the distribution of insertion events upstream of the bop gene in Bop⁻ mutants.

MATERIALS AND METHODS

Materials: Reverse transcriptase (RNA-dependent DNA polymerase) from Avian myeloblastosis virus was obtained from J.W. Beard (Life Sciences, St. Petersburg, Fl. γ [³²P]ATP (2-3000 Ci/mM) and α [³²P] deoxynucleotide triphosphates (400 Ci/mM) were from Amersham. Deoxynucleotide triphosphates and dideoxynucleotide triphosphates were from PL Biochemicals. Synthetic M13 pentadecamer primer (5' CCCAGTCACGACGTT 3') was a gift from Genentech Inc. RNasin was from Promega Biotec (Madison, WI). Other enzymes were obtained as described previously (6,11,12).

Mutants and Strains: All halobacterial mutants and strains used and their growth conditions have been reported (6,11,12).

DNA Sequence Analysis: DNA sequences were determined by the methods of Maxam and Gilbert (13) or Sanger et al. (14). Labeling of DNA at the 3' terminus was done as described (15).

Isolation of Single-Stranded DNA Probes: Single-stranded DNA probes were prepared from a 1103 bp BamHI fragment upstream of the bop gene as follows. The fragment was cloned into M13mp8 and recombinant single-stranded phage DNA of both orientations was prepared (16). The single-stranded DNA of each orientation (1 pmole) and M13 primer (6 pmoles) were heated in sealed capillary tubes at 100°C for 3 min and annealed at 65°C for >15 min in 50 mM NaCl, 6 mM Tris-HCl, pH 7.5, 6 mM MgCl₂, 1 mM DTT. 25 μ Ci of each α [³²P] deoxynucleotide triphosphate and 5 units of DNA polymerase I (large fragment) were added, followed by incubation at 37°C for 30 min. The reactions were chased by addition of unlabeled deoxynucleotide triphosphates to a concentration of 100 μ M for 15 min. The reactions were heated to 65°C and purified through Sephadex G25 to remove unincorporated label. Both DNA samples, each labeled

in only one strand, were digested with EcoRI endonuclease which cuts at one site within M13mp8 on the opposite side of the insert relative to the M13 pentadecamer priming site. Samples were phenol extracted, ethanol precipitated, denatured in 90% formamide, 2 mM EDTA at 100°C for 3 min and electrophoresed on 5% acrylamide, 7 M urea gels. The gels were autoradiographed and bands corresponding to full size single-stranded inserts were excised, electroeluted (13), and used for Northern hybridizations.

RNA Gels and Northern Hybridization: Total RNA from H. halobium purified according to Chirgwin (17) was denatured 3 min at 100°C in 50% formamide, 2.2 M formaldehyde, 50 mM HEPES, 1 mM EDTA. The RNA samples were electrophoresed on 1.5% agarose slab gels in 2.2 M formaldehyde, 50 mM HEPES, 1 mM EDTA, 5 mM NaOAc pH 7, with no formaldehyde in the reservoir buffer (18) as modified by Barry Drees (UCSF). The fractionated RNA was transferred to Sartorius nitrocellulose filters in 20X SSC for 20 h and the filters were subsequently baked at 80°C for 2 h in a vacuum oven. Filters were prehybridized, hybridized with 10^6 - 10^7 cpm α [32 P] single stranded DNA, washed, and dried according to the procedure of Thomas (19). Kodak X-Omat AR5 films were used for autoradiography (-70°C, 3-18 h with Dupont Cronex lightning plus screens).

cDNA Synthesis: Two pmoles of an 86 bp AvaII/NaeI fragment (see Fig. 3) and 25 μ g of total ET1001 RNA in cDNA buffer (50 mM KCl, 50 mM Tris-HCl, pH 8.3, 10 mM MgCl₂, 10 mM DTT) were sealed in a capillary tube, denatured at 100°C for 2 min, and reannealed at 65°C for 2 hr (20,21). 60 μ Ci of α [32 P]-dCTP, 100 μ M dGTP, dATP, dTTP, and 20 units RNAsin followed by 16 units of reverse transcriptase were added to the annealed mixture in the above buffer. The reaction (10 μ l volume) was incubated at 42°C for 60 min (20,21,22) and chased by addition of 100 μ M dCTP at 42°C for 15 min followed by the addition of 25 mM EDTA. The reaction was phenol extracted, ethanol precipitated, denatured and electrophoresed on a 5% acrylamide, 7 M urea gel alongside a sequence ladder. Autoradiography was for 7 h with an intensifying screen at -70°C.

cDNA Sequencing: Two pmoles of the 86 bp AvaII/NaeI primer and 200 μ g of total ET1001 RNA were denatured and reannealed as above. Pulse-labeling followed by extension with reverse transcriptase using chain terminating inhibitors was done according to a modification of Bina-Stein et al. (23) as follows. The annealed mixture was added to 130 μ Ci each of α [32 P]-dCTP and α [32 P]-dGTP, 200 pmoles dTTP, 10 units of RNAsin and 16 units of reverse transcriptase in cDNA buffer. The reaction (20 μ l) incubated at 42°C for 2 min was divided between four tubes for sequencing. The concentration of

deoxynucleotide triphosphates was adjusted to 75 μ M in each reaction and dideoxynucleotide triphosphates were added to the appropriate tubes at a concentration of 50 μ M for ddGTP, ddATP and ddCTP, and 150 μ M for ddTTP. An

Figure 2. Hybridization of single stranded DNA probes with total RNA of *H. halobium* fractionated on agarose gels, and transferred to nitrocellulose. (1) Total RNA (10 μ g) probed with one strand of the *Bam*HI 1103 bp fragment (see Materials and Methods). This strand is contiguous to the *bop* gene coding strand (3 hr exposure). (2) Total RNA probed with the strand complementary to that used in (1) (16 hr exposure). (3) and (4) are filters (1) and (2) overlaid with the *bop* gene specific nick-translated probe (Ref. 6) (18 hr exposure). The sizes of the two major *bop* mRNA species (8) are indicated. The larger species is not readily discernible on this autoradiogram.

additional 4 units of reverse transcriptase were added per tube followed by incubation at 42°C for 30 min. The sequencing reactions were terminated as described (23) and electrophoresed on a 5% acrylamide, 7M urea gel. The gel was subsequently dried on Whatman 3 MM paper and autoradiographed for 48 h with an intensifying screen.

RESULTS AND DISCUSSION

DNA sequence upstream of the *bop* gene contains a large open reading frame which is transcribed

We have determined 1800 bp of nucleotide sequence upstream of the *bop* gene in *H. halobium* strain R1, a Vac^- derivative of wild type (Fig. 1). Comparison of this sequence to 360 nucleotides upstream of the *bop* gene in the purple membrane overproducing strain S9 (5) reveals several differences (see Fig.1). This sequence has been revised to include the clustered G residues at positions 143 to 149 (see Fig. 1) (G. Khorana, personal communication). Limited independent sequence data from strain R1 agrees with our sequence in the region of the transition changes (see Fig. 1) (D. Oesterhelt, personal communication). It is not clear at this time whether the differences are relevant to the overproducer phenotype of strain S9.

Three substantial open reading frames (ORFs) are found in this segment of DNA. Two overlapping ORFs are on the same strand of DNA as the coding strand of the *bop* gene and the third is on the opposite strand. Total wild type RNA fractionated on agarose gels and transferred to nitrocellulose was hybridized with single-stranded DNA probes corresponding to each strand of the *Bam*HI 1103 bp fragment upstream of the *bop* gene (see Materials and Methods) (Fig. 2, lanes 1 and 2). Hybridization was observed only with the

Figure 3. Restriction map illustrating the distribution of insertion elements upstream of the bacterio-opsin gene in various Bop mutants, and the nucleotide sequence strategy. The coding regions of the *brp* gene and part of the *bop* gene are indicated by heavy black bars. The 5' terminus of both the *brp* transcript and the *bop* mRNA are denoted by heavy arrows. The bracket below the *brp* gene sequence delineates the 86 bp *AvaII*/*NaeI* primer used for characterization of the 5' terminus of the *brp* transcript. The integration sites of insertions in various Bop mutants (IV-4, IV-3, M135, M138, M89, M18, IV-12, IV-10 and IV-8) and a revertant (reIV-41) (11) are marked by vertical bars. Restriction sites are: (□) *Bam*HI, (■) *Ava*I, (o) *Acc*I, (●) *Nae*I, (∇) *Ava*II, (Δ) *Mlu*I, (◇) *Hinf*I and (◆) *Nde*I. Nucleotide sequence strategy is shown by the array of arrows beneath the map which indicate the direction and extent of sequence. Nucleotide sequence determined by the chemical degradation method is marked by (—) for those fragments 5'-end labeled and (o—) for those fragments 3'-end labeled. (■—) indicates nucleotide sequence determined by the dideoxynucleotide chain termination procedure. Approximate positions of differences in the nucleotide sequence upstream of the *bop* gene from that published by Dunn et al. (5) are indicated by asterisks.

probe which would detect an mRNA transcript initiating in the opposite direction as the *bop* gene mRNA. The translation of the ORF corresponding to this transcript is presented in Fig. 1. The entire ORF extends for 1118 nucleotides and encompasses all of the insertion events upstream of the *bop* gene which result in the Bop⁻ phenotype (Fig. 1, 3) (6,11,12). Based on these observations, we propose that this ORF is a gene whose product is somehow associated with the expression of bacterio-opsin or purple membrane formation. We designate this putative gene *brp* (*b*acterio-*o*psin-*r*elated protein gene).

Characterization of the *brp* transcript

Northern analysis suggested that the *brp* transcript is heterogeneous (Fig. 2, lane 1). However, S1 nuclease protection mapping with RNA from wild type and purple membrane overproducer strain ET1001, indicated that the 5' terminus of this transcript is unique. Both wild type and ET1001 RNAs yielded S1 resistant hybrids of identical size although ET1001 RNA yielded more hybrid than wild type (data not shown). In order to determine the 5'

Figure 4. Mapping of the 5' terminus of the *brp* transcript by cDNA analyses. (A) Full length primed cDNA synthesis. In the center lane is the cDNA product made by extension from an *Ava*II/*Nae*I fragment primer (see Fig. 3) and ET1001 total RNA. Flanking the cDNA is a nucleotide sequence ladder obtained by the chain termination method with the same DNA primer and single-stranded DNA template from the area of interest. Pertinent sequence from this area is bracketed on the right illustrating that the cDNA terminates at the T residue marked by an asterisk. Thus the 5' terminus of the transcript begins AUGAGCAAUA... (B) cDNA sequence. An extension reaction similar to the cDNA synthesis was performed in the presence of chain terminators (see Materials and Methods). Sequence obtained corresponds to that of (A) and terminates in all four lanes at the position labeled "stop."

terminus of the transcript, cDNA was synthesized using an 86 bp *Ava*II-*Nae*I DNA fragment (see Fig. 3) as a primer for reverse transcriptase with ET1001 total RNA. The cDNA products were fractionated on an acrylamide urea gel (Fig. 4A). The largest significant cDNA product displayed an extension to the T nucleotide complementary to the A residue of the second in phase methionine codon of the ORF (Fig. 1). This result was confirmed by cDNA nucleotide sequence analysis using chain terminators and reverse transcriptase with the 86 bp primer and total RNA as template (Fig. 4B).

The apparent heterogeneity of the *brp* transcript in formaldehyde agarose gels (Fig. 2, lane 1) is more extensive than is the case for the *bop* mRNA which consists of at least two mRNA subspecies with different 3' termini (Fig. 2, lane 4) (8). Most of the *brp* RNA population migrated at approximately the same position or more slowly than would be expected of an mRNA

Table 1. Codon usage of bop and brp genes. Codon usage of the bop gene is according to Dunn et al. (5).

bop brp			bop brp			bop brp			bop brp		
UUU	Phe	0 1	UCU		0 0	UAU	Tyr	3 1	UGU	Cys	0 0
UUC		13 18	UCC	Ser	3 2	UAC		8 12	UGC	Op	0 0
UUA	Leu	2 0	UCA		0 1	UAA	OC	0 0	UGA	OP	1 1
UUG		6 6	UCG		5 12	UAG	AM	0 0	UGG	Trp	8 9
CUU		2 4	CCU		0 0	CAU		0 2	CGU		3 0
CUC	Leu	12 17	CCC	Pro	3 5	CAC	Nis	0 3	CGC	Arg	3 6
CUA		2 0	CCA		3 3	CAA		0 0	CGA		0 2
CUG		15 19	CCG		6 8	CAG	Gln	4 8	CGG		1 9
AUU		0 1	ACU		0 2	AAU	Asn	0 1	AGU	Ser	1 0
AUC	Ile	15 12	ACC		7 4	AAC		3 2	AGC		5 6
AUA		0 0	ACA	Thr	2 0	AAA	Lys	3 1	AGA	Arg	0 0
AUG	Met	10 9	ACG		10 8	AAG		4 1	AGG		0 1
GUU		3 2	GCU		1 1	GAU	Asp	2 0	GGU		4 7
GUC	Val	8 26	GCC	Ala	8 27	GAC		8 15	GGC	Gln	10 22
GUA		3 0	GCA		5 5	GAA		4 0	GGA		4 3
GUG		9 16	GGG		16 24	GAG	Glu	7 5	GGG		8 10

predicted from the ORF. The diversity in electrophoretic mobility could be due to marginal RNase degradation, secondary structure or a heterogeneous 3' terminus.

The brp gene sequence

The brp gene could encode a protein of 359 amino acids (MW 37,500D; pK 6.74). The codon usage of the brp gene is very similar to that of the bop gene (Table 1). The predicted secondary structure for the brp protein has six hydrophobic helices of sufficient length to span the cell membrane and the net hydrophobicity of the brp protein is equivalent to that of bacterio-opsin (0.95 Kcal/mole/residue; Robert Stroud, personal communications) (24). There is no amino acid sequence homology to bacterio-opsin nor to the limited amino acid sequence of another retinal binding membrane protein, halorhodopsin (Peter Scherrer, personal communication).

Comparison of brp and bop mRNAs

The bop mRNA and the putative brp mRNA have some interesting similarities. Both start at or near the initiating codon of the gene. The 5' terminus of the brp transcript has an inverted repeat of 6 bp which could form a hairpin loop secondary structure similar to that proposed for the bacterio-opsin mRNA (5) except that the loop size for the brp transcript is slightly larger (Fig. 5A). Possible functional roles for these putative secondary structures include ribosomal interaction as structural recognition elements, and/or stabilization of the mRNA 5' terminus. It is noteworthy that 13 ribonucleotides at the 5' termini of both transcripts (including the

Figure 5. Comparison of the 5' termini of the bop mRNA and putative brp mRNA. (A) Possible hairpin loop secondary structures. Potential ribosomal binding sites are underlined. (B) Complementarity between the 5' termini.

loops of both these hairpins) are complementary to each other (Fig. 5B). There are potential ribosomal binding sites for each mRNA near the 5' termini; the sequence GGAG (position 6-9 from the bop AUG (5)) and the sequence AGGU (position 10-13 from the brp AUG) both occur within the loops of their respective mRNAs (Fig. 5A). These sequences are complementary to the sequence ACCUCC near the 3' terminus of H. halobium 16s rRNA (25). Other potential ribosomal binding sites occurring further downstream are the sequences GUGG (position 25-28 from the bop AUG, and position 30-33 from the brp AUG) (Fig. 5A).

Sequence homologies upstream of brp and bop genes

Little is known about gene regulation in H. halobium. The lack of a genetic exchange system and problems inherent with in vitro studies using H. halobium RNA polymerase at physiologically high ionic strength (26 and W. Zillig, personal communication) have impeded progress in this area. However, the availability of the brp gene sequence in addition to that of the

Figure 6. Comparison of 120 nucleotides upstream of the *brp* and the *bop* genes. Both DNA sequences are aligned relative to their respective mRNA 5' termini (nucleotide +1). Homologous areas are boxed. Complementary areas are bracketed.

bop gene enables an initial comparison of sequences of two potential promoter regions of halobacteria. Indeed, there are significant homologies in the 120 nucleotides of sequence immediately upstream of the brp and bop genes (Fig. 6). Two homologous sequences (the -10 and -45 regions) are reminiscent of the -10 and -35 sequences of procaryotic promoters but other homologies exist beyond -70 as well. The homologies could represent DNA-dependent RNA polymerase binding sites and/or regulatory sequences common to both genes. The 120 nucleotides of upstream sequence of each gene have a high A+T content (49% G+C) relative to their flanking genes (bop gene 62% G+C; brp gene 68% G+C) and to the organism as a whole (>58% G+C (27)). The homologies upstream of these two genes should provide a foundation for in vitro transcription experiments which with an analysis of additional halobacterial genes will help to define a consensus halobacterial promoter.

Also located within the 120 nucleotides of sequence upstream of both these genes is a complementary sequence (see Fig. 6) which is a target site for insertion elements ISH23 (Bop mutant IV-10) and ISH2 (Bop mutant IV-12) (see Fig. 3) (12).

The brp gene affects bop gene expression

We previously proposed that the sequence upstream of the bop gene contains a gene encoding a regulatory protein or an unknown structural gene which may be cotranscribed with the bop gene (6). In this report we provide evidence for the existence of a gene (brp gene) located upstream of the bop gene and transcribed in the opposite direction as the bop gene. Insertion elements transposing in the brp gene result in the Bop⁻ phenotype (6,11,12) suggesting that the brp gene in some way affects expression of the bop gene or the formation of the purple membrane. It is not yet clear at what level the brp gene is involved in bop gene expression. Characterization of the putative brp protein should aid in the determination of its function and relationship to bacterio-opsin. In addition, an analysis of the levels of bop and brp mRNA in wild type, overproducer and mutant strains should help to clarify the brp gene's effect on bop gene expression.

ACKNOWLEDGMENTS

We thank Patricia Clausen for preparation of the manuscript. This work was supported by a grant from the National Institute of General Medical Sciences (GM31785).

REFERENCES

1. Oesterhelt, D. and Stoekenius, W. (1971) *Nature* **262**, 416-418.

2. Sumper, M., and Herrmann, G. (1976) FEBS Lett. 69, 149-152.
3. Oesterhelt, D. and Stoebenius, W. (1971) Nature New Biol. 233, 149-152.
4. Henderson, R.E. (1977) Annu. Rev. Biophys. Bioeng. 6, 87-109.
5. Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S.H., RajBhandary, U.L. and Khorana, H.G. (1981) Proc. Nat. Acad. Sci. (USA) 78, 6744-6748.
6. Betlach, M., Pfeifer, F., Friedman, J. and Boyer, H.W. (1983) Proc. Nat. Acad. Sci. (USA) 80, 1416-1420.
7. Vogelsang, H., Oertel, W. and Oesterhelt, D. (1983) In: Methods in Enzymology. (Fleischer, S. and Fleischer, B., eds.) In preparation.
8. DasSarma, S., RajBhandary, U.L., and Khorana, H.G. (1984) Proc. Nat. Acad. Sci. (USA) 81, 125-129.
9. Stoebenius, W., Lozier, R.H. and Bogomolni, R.A. (1979) Biochem. Biophys. Acta 505, 215-278.
10. Pfeifer, F., Weidinger, G. and Goebel, W. (1981) J. Bacteriol. 145, 375-381.
11. Pfeifer, F., Betlach, M., Martienssen, R., Friedman, J. and Boyer, H.W. (1983) Mol. Gen. Genet. 191, 182-188.
12. Pfeifer, F., Friedman, J., Boyer, H.W. and Betlach, M. (1984) Nucleic Acids Res. 12, 2489-2497.
13. Maxam, A.M. and Gilbert, W. (1980) In: Methods in Enzymol. 65 (L. Grossman, K. Moldave, eds.), 499-560.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. (USA) 74, 5463-5467.
15. Rütger, U., Koenen, M., Otto, K. and Müller-Hill, B. (1981) Nucleic Acids Res. 9, 4087-4098.
16. Messing, J. (1983) In: Methods in Enzymol. 101 (Wu, R., Grossman, L., Moldave, K., eds.) 20-78.
17. Chirgwin, J., Przybyla, A., MacDonald, R. and Rutter, W. (1979) Biochemistry 18, 5294-5299.
18. Lehrach, H., Diamond, D., Wozney, J. and Boedtker, H. (1977) Biochem. 16, 4743-4751.
19. Thomas, P. (1980) Proc. Nat. Acad. Sci. (USA) 77, 5201-5205.
20. Lee, D.C., Roeder, R.G. (1981) Mol. Cell Biol. 1, 635-651.
21. Luse, D.S., Haynes, J.R., Van Leeuwen, D., Schon, E.A., Cleary, M.L., Shapiro, S.G., Lingrel, J.B. and Roeder, R.G. (1981) Nuc. Acids Res. 9, 4339-4354.
22. Maniatis, T., Fritsch, D.R., and Sambrook, J. (eds.) (1982) In: Molecular Cloning - A Laboratory Manual (Cold Spring Harbor Laboratory), pp. 229-234.
23. Bina-Stein, M., Thoren, M., Salzman, N. and Thompson, J.A. (1979) Proc. Nat. Acad. Sci. (USA) 76, 731-735.
24. Finer-Moore, J. and Stroud, R. (1984) Proc. Nat. Acad. Sci. USA 81, 151-159.
25. Kagramanova, V.K., Mankin, A.S., Baratova, L.A., and Bogdanov, A.A. (1982) FEBS Lett. 144, 177-180.
26. Madon, J. and Zillig, W. (1983) Eur. J. Biochem. 133, 471-474.
27. Moore, R.L. and McCarthy, B.J. (1969) J. Bacteriol. 99, 248-254.