Volume 12 Number 5 1984 Nucleic Acids Research

Characterization of insertions affecting the expression of the bacterio-opsin gene in *Halobacterium halobium*

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

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

Received 21 December 1983; Accepted 2 February 1984

ABSTRACT

We have determined the sequence of the inverted repeats and duplicated target DNA of the halobacterial insertion elements ISH2 (520 bp), ISH23 (900 bp) and ISH24 (3000 bp) associated with bacterio-opsin (bop) mutants. ISH2 has a perfect 19 bp inverted repeat (3,5), while both ISH23 and ISH24 have imperfect inverted repeats of 29 bp and 14 bp respectively. ISH23 was shown to be highly homologous to ISH50 (6). Variable lengths of duplicated target DNA are found when ISH2 and ISH23 (ISH50) transpose into different sites. A 550 bp DNA insert ("ISH25") reverts the Bop mutation caused by ISH24. "ISH25" lacks typical structural features of a transposable element. "ISH25" and ISH24 are found adjacent to each other upstream of the bop gene. An identical arrangement of "ISH25" and ISH24 is found in the cccDNA of H. halobium NRC817. Comparative sequence analysis of both areas suggests that the translocation of "ISH25" to the bop gene region occured by a recombination event.

INTRODUCTION

Insertions are the major cause of mutations affecting the bacterio-opsin (bop) gene which encodes the apo-protein of bacteriorhodopsin in the purple membrane of Halobacterium halobium (2). Six different insertions associated with Bop mutants have been isolated (ISH1, ISH2, ISH23, ISH24, "ISH25", and ISH26; 1-4). With the exception of ISH26 these insertions have been described as to their copy number, distribution and transposability in the genome of H. halobium (5). Two insertion elements (ISH1, ISH2) have been completely sequenced (1,3). ISH1, a 1118 bp element integrates into a unique sequence at the 5' end of the bop gene. At least fifteen Bop mutants have ISH1 integrated at this same site in both possible orientations (1,2,5 and our unpublished results). ISH2, a 520 bp element integrates into at least eight different positions within or upstream of the bop gene (2,3,5 and our unpublished results). A reversion of the Bop mutation caused by ISH24 is the result of an additional insertion ("ISH25") adjacent to ISH24 (5). In this paper we present the nucleotide sequences of ISH2 integration sites in additional mutants, a characterization of the insertion elements ISH23 and ISH24, and

© IRL Press Limited, Oxford, England.

report on the nature of a revertant of the ISH24 mutant.

MATERIALS AND METHODS

Materials: $\gamma[^{32}P]$ATP (2-3000 Ci/mMol) was obtained from Amersham. Calf intestine alkaline phosphatase and T4 polynucleotide kinase were obtained from Boehringer-Mannheim. Restriction endonucleases were from New England Biolabs.
Mutants and Strains: Mutants IV-3, IV-4, IV-10, IV-12 were isolated by G. Weidinger as described (7) and are spontaneously occurring Bop$^-$ derivatives of *H. halobium* NRC817. Bop mutants M18, M89, M135, M136 and M138 were kindly provided by D. Oesterhelt (Max-Planck-Institut für Biochemie, Munich, W. Germany) and are derivatives of Vac$^-$ strain R1. The revertant (reIV-41) of mutant IV-4 was isolated by F. Pfeifer.
DNA Isolation and Southern Blot Hybridization: Growth of *H. halobium* wild type and mutant strains, purification of genomic and cccDNA, and subsequent cloning of appropriate *Pst*I fragments into positive selection vector pKG2 have been described (2). Nick-translation and Southern blot hybridization were done according to standard procedures as described (5).
DNA Sequencing: DNA sequencing was done by the Maxam and Gilbert method (8). Fragments containing integration sites of insertion elements were sequenced on one strand. Sequences upstream of the *bop* gene presented have been confirmed on both strands. Autoradiography was done at -70°C for 1-10 days using Kodak X-Omat AR5 film and Dupont Cronex Lightning-plus intensifying screens.

RESULTS

Insertion elements (ISH2, ISH23 and ISH24) cloned from various Bop mutants have been characterized by restriction mapping and sequence analyses of their integration sites. These elements have terminal inverted repeats which are flanked by direct repeats of duplicated target DNA (see Fig. 1 and Table 1).

a) ISH2

ISH2 is the most frequently occurring insertion element associated with Bop mutants. Sixteen out of a total of 33 Bop mutants have ISH2 integrated in or upstream of the *bop* gene. Nine of these ISH2 elements are integrated in the bacterio-opsin gene within the *Ava*I 429 bp fragment, but their integration sites have not been sequenced. Seven mutants have ISH2 integrated in various positions in a region 525 bp to 1367 bp upstream of the gene (in mutants IV-3, IV-12, M18, M89, M135, M136, and M138) (5; and our unpublished results). All but one of the ISH2 insertions (mutant M89) have the same orientation relative

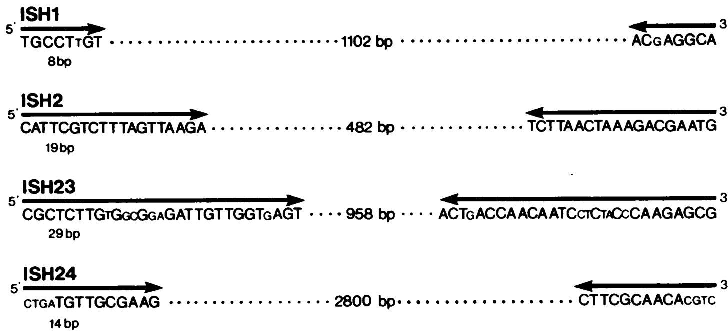

Figure 1. Nucleotide Sequences of Inverted Repeats. Mismatched bases are indicated by smaller capital letters. The sequence of the inverted repeat of ISH1 is according to Simsek et al. (1). Analysis of one terminus of ISH1 in our mutant IV-8 revealed the same sequence as found for ISH1 in mutant SD17. ISH1, ISH2 and ISH23 (ISH50) have been completely sequenced (1,3,6). The number of base pairs between the inverted repeats was obtained by subtracting the size of the inverted repeats from the total size of the insertion element. The number of base pairs in ISH24 was estimated by restriction mapping.

to the bop gene (data not shown).

The complete nucleotide sequence of ISH2 has been determined (3). There is a perfect 19 bp inverted repeat (Fig. 1) and in three insertion events the target DNA duplication is either 10 bp or 20 bp (3; see Table 1). We have

TABLE 1. Duplicated Target DNA Sequences of Known Halobacterial Insertion Elements

Insertion Element	Bop Mutant	Sequence of Duplicated Target DNA	Length
ISH1	SD17*	AGTTATTG	8 bp
ISH2	SD19*	CGGCTCCGTG	10 bp
ISH2	L33*	TGGCCTCACA	10 bp
ISH2	R1mR*	ACCCCATCTACTGGGCGCGG	20 bp
ISH2	IV-3	TCCAGGGCGT	10 bp
ISH2	IV-12	GAATACACACG	11 bp
ISH2	M18	ACCCAACAGGT	11 bp
ISH23	IV-10	TACACACAT	9 bp
ISH50	R1*	TTGTGGAT	8 bp
ISH24	IV-4	CCCACGA	7 bp

*Sequence of target DNA of ISH1 in mutant SD17 is according to Simsek et al. (1), the target DNA of ISH2 in SD19, L33 and R1mR is according to DasSarma (3), and the ISH50 target DNA sequence in strain R1 was determined by Xu and Doolittle (6).

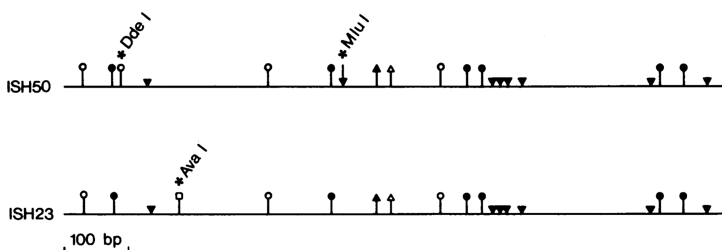

Figure 2. Comparison of ISH23 and ISH50 by Restriction Site Mapping. Restriction sites are shown for (♠) EcoRI; (↑) BglI; (♀) AvaI; (↓) MluI; (♦) HpaII; (♀) DdeI; and (▼) HinfI. Differences in restriction sites are marked by (*). Restriction sites in ISH50 were determined from nucleotide sequence (6) and in ISH23 by restriction mapping. Analysis of ISH50 in the 50 kb cccDNA of our R1 strain revealed an AvaI site as shown by hybridization of ISH23 to AvaI digested cccDNA of R1 (data not shown).

sequenced both terminal regions of ISH2 in mutants IV-3, IV-12 and M18, and the length of duplicated target DNA in these mutants is either 10 bp or 11 bp (see Table 1). The inverted repeats and additional ISH2 sequence determinations are in complete agreement with that already published. The duplicated target DNAs are not homologous nor do they share homology with sequences within ISH2 as found in the case of ISH1 (1,5). These results emphasize the nonspecific integration of ISH2. Limited insertional hot spots may occur; for example, in mutants M89 and M138, ISH2 is integrated in exactly the same position 700 bp upstream of the bop gene but in different orientations. In mutant M18, ISH2 integrated 1 bp further upstream than in mutants M89 and M138.

b) ISH23

ISH23 integrated 30 bp upstream of the bop gene initiation codon in mutant IV-10. It has an imperfect 29 bp inverted repeat, and 9 bp of duplicated target DNA (Fig. 1 and Table 1).

Recently, the complete sequence of a 996 bp insertion element (ISH50) has been determined (6). This insertion element was isolated from the 50 kb cccDNA of the Vac$^-$ strain R1. Comparison of the ISH50 sequence with a partial DNA sequence of ISH23 as well as restriction maps of these elements indicates a high degree of homology (Fig. 2). The inverted repeats of these elements are identical, but restriction map analysis of ISH23 in comparison with the computer generated restriction map of ISH50 reveals some possible nucleotide polymorphisms (see Fig. 2). The target sites of ISH23 (9 bp) and ISH50 (8 bp) have no apparent homology (see Table 1).

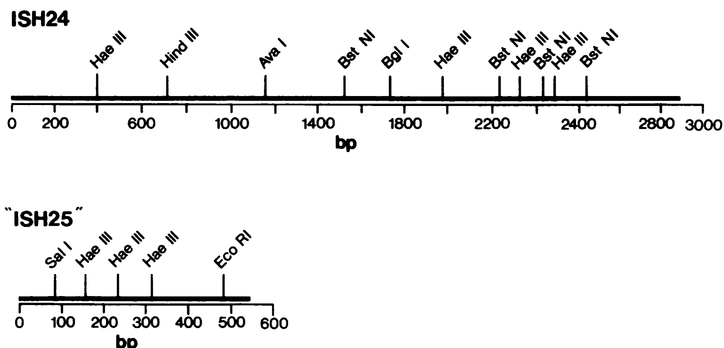

Figure 3. Internal Restriction Maps of ISH24 and "ISH25". ISH24 has no restriction sites for BamHI, EcoRI, MluI, NdeI, PstI and SalI. "ISH25" has no sites for BamHI, HindIII and PstI. Restriction maps have been determined by incomplete digests with end-labeled fragments as well as by mapping of isolated fragments from mutant IV-4 and the revertant.

c) ISH24 and "ISH25"

The 3000 bp insertion element ISH24, integrated 1405 bp upstream of the bop gene in mutant IV-4 (2,5), has an imperfect inverted repeat of 14 bp (Fig. 1). This insertion event has a duplication of 7 bp of target DNA (Table 1). The ISH24 mutant is particularly interesting because it is the only Bop mutant from which revertants could be isolated. The two revertants obtained from screening 10,000 colonies are both the result of a translocation of a 550 bp sequence ("ISH25") rather than a deletion of the initial ISH24 insertion, and they appear to be identical (2). ISH24 and "ISH25" are contiguous in the revertants. Restriction maps of ISH24 and "ISH25" are given in Fig. 3. Nucleotide sequencing at the integration site of "ISH25" revealed that "ISH25" is upstream of the inverted repeat (IR_L) of ISH24. No terminal inverted repeats nor duplicated target DNA were found with "ISH25" (Fig. 4b) confirming a preliminary observation that "ISH25" might not be a transposable element (5). Furthermore, when the sequences from mutant and revertant were compared, there was a difference in the DR_L associated with ISH24, i.e. a deletion of five base pairs in the DR_L of ISH24 (5'CACGA 3'; see Fig. 4 a,b).

One of the two copies of ISH24 located on pHH1, the main cccDNA species of H. halobium NRC817 (9), is closely associated with "ISH25" (5; see Fig. 5). The nucleotide sequence determined for the junction of "ISH25" and ISH24 in pHH1 fragment E7 is identical to that found in the revertant (Fig. 4b,c). A sequence of 8 bp (1 bp mismatch) upstream of "ISH25" in the revertant is homologous to the distal "ISH25" terminus in pHH1 (Fig. 4c).

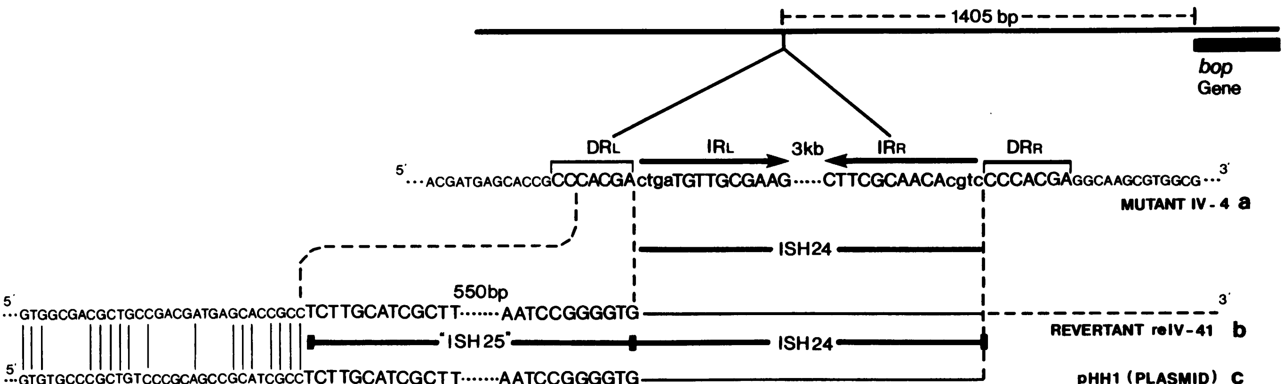

Figure 4. Sequence Comparison of the Junctions Between "ISH25" and ISH24 Upstream of the bop Gene in Revertant reIV-41, and on pHH1.

(a) Sequence of ISH24 integration site 1405 bp upstream of the bop gene in mutant IV-4. The 14 bp inverted repeat is marked by arrows. The 7 bp direct repeat of duplicated target DNA is indicated by large capital letters and a bracket. The remainder of the genomic sequence is given in small capital letters.

(b) Sequence of "ISH25" termini in revertant reIV-41. "ISH25" specific sequence is given in large capital letters, and genomic sequence upstream of the integration site is in small capital letters.

(c) Sequence of "ISH25" termini on pHH1. Homologies between the integration site in pHH1 and that upstream of the bop gene are indicated by lines.

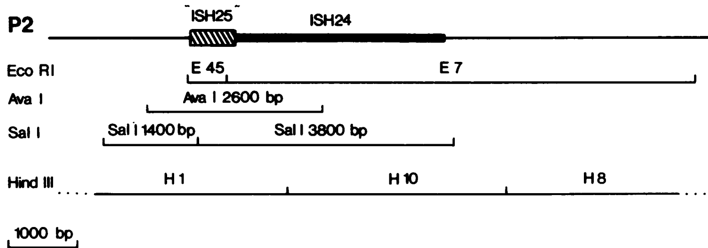

Figure 5. Restriction Map of the "ISH25" and ISH24 on pHH1. PstI fragment P2 (22 kb) was cloned into the positive selection vector pKG2. Relevant EcoRI, AvaI, SalI and HindIII restriction sites found within the 10 kb of P2 shown in this figure are indicated. The EcoRI and HindIII restriction fragments are designated E45 (560 bp), E7 (7,000 bp), H1 (30,000 bp), H10 (3,300 bp) and H8 (7,800 bp). Hybridization experiments using a nick-translated ^{32}P-labeled internal fragment from "ISH25" (430 bp SalI-EcoRI fragment) and a fragment of ISH24 (800 bp BamHI-HindIII fragment from mutant IV-4) as probes against SalI, EcoRI and AvaI digested P2 revealed hybridization to the same 3800 bp SalI and 2600 bp AvaI fragments indicating that "ISH25" and ISH24 are contiguous (data not shown). The two EcoRI fragments E7 and E45 were used for DNA sequence analysis. The arrangement of "ISH25"and ISH24 in the cccDNA of both wild type and revertant is identical.

DISCUSSION

Structural characteristics of halobacterial insertion elements

ISH2, the smallest known insertion element in halobacteria, has the highest copy number (8-10 copies in *H. halobium*), and the highest transposition frequency (5). ISH2 occurs in *H. halobium* and closely related species (*H. salinarium*, *H. cutirubrum*) but not in a number of other natural isolates producing purple membrane (5) nor in "square" halobacteria (10) (unpublished results). The integration of ISH2 appears to be more nonspecific than is the case for ISH1, which integrated at a specific target site at the 5' end of the *bop* gene in all fifteen Bop mutants caused by ISH1 (1,5 and unpublished results).

ISH23 derived from the Bop mutant IV-10 and ISH50 isolated from the cccDNA of strain R1 are highly homologous. ISH50 could be a derivative of the ISH23 element found on pHH1 in *H. halobium* NRC817, since the 50 kb cccDNA of strain R1 from which ISH50 was isolated although smaller is homologous to pHH1 (5).

Usually, target DNA of a specific length is duplicated for each insertion element found in *E. coli* and other eubacteria. Only a few exceptions are known including a variant of IS1 (11) and a copy of IS4 in the galactose operon in *E. coli* (12). In the case of halobacterial insertion elements, both

ISH2 and ISH23 duplicate target DNA of different lengths (10 bp, 11 bp and 20 bp in the case of ISH2; and 8 bp and 9 bp for ISH50 and ISH23).

The largest insertion element, ISH24 (3,000 bp) has an imperfect 14 bp terminal inverted repeat and duplicates 7 bp of target DNA upon integration. ISH24 is confined to H. halobium and closely related species (5). Comparison of the inverted repeat of ISH24 with those of other halobacterial insertion elements reveals a common pentanucleotide (5'TGTTG 3') in the inverted repeat of ISH23 (ISH50) and ISH24. ISH1 has part of this sequence (5'TGTT 3'), whereas it is absent in the ISH2 inverted repeat. This pentanucleotide is also present in inverted repeats of E. coli transposable elements IS102, IS903 (13) and IS26 (14). The relevance of this common pentanucleotide is unknown.

Translocation mechanism of "ISH25"

Bop mutant IV-4 is the only mutant from which revertants could be recovered. The "ISH25" insertion in the revertant is not a transposable element since it lacks the typical structural features of a transposable element and no additional copy of "ISH25" is found in any other mutant tested (5). The "ISH25" insertion appears to be the result of a recombination event between two ISH24 elements and short sequence homologies at the end of "ISH25" and the target site upstream of the bop gene.

Intramolecular recombinations between two insertion elements (ISH1.8 in halobacterial phage ϕH) leading to inversion or circularization of the enclosed phage ϕH DNA have been described by Schnabel (15,16). Thus insertion elements serve not only as templates for transpositions but also as hot spots for recombinations which result in rearrangements, deletions and translocations of adjacent DNA sequences.

It is not yet known how "ISH25" overcomes the initial insertional inactivation of the bop gene by ISH24 and restores the expression of the bop gene. Further analyses of "ISH25" and the upstream region of the bop gene are in progress and should provide a better understanding of the functional relationship between this region and the expression of the bacterio-opsin gene in H. halobium.

ACKNOWLEDGEMENTS

We thank Dieter Oesterhelt for providing Bop mutants and Patricia Clausen for preparation of the manuscript. This work was supported by a grant number GM31785-01 awarded by the National Institute of General Medical Sciences. F.P. is supported by a fellowship (PF165/1-1) from the Deutsche Forschungsgemeinschaft (DFG).

REFERENCES

1 Simsek, M., DasSarma, S., RajBhandary, U.L., and Khorana, H.G. (1982) Proc. Natl. Acad. Sci. USA 79, 7268-7272
2 Betlach, M., Pfeifer, F., Friedman, J. and Boyer, H.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1416-1420
3 DasSarma, S., RajBhandary, U.L., and Khorana, H.G. (1983) Proc. Natl. Acad. Sci. USA 80, 2201-2205
4 Ebert, K., Goebel, W., and Pfeifer, F. (1984) Molec. Gen. Genet, in press.
5 Pfeifer, F., Betlach, M., Martienssen, R., Friedman, J. and Boyer, H.W. Molec. Gen. Genet. 191, 182-188
6 Xu, W.L. and Doolittle, W.F. (1983) Nucleic Acids Res. 11, 4195-4199
7 Pfeifer, F., Weidinger, G., and Goebel, W. (1981) J. Bacteriol. 145, 375-381
8 Maxam, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
9 Weidinger, G., Klotz, G., and Goebel, W. (1979) Plasmid 2, 377-386
10 Javor, B., Requardt, C., and Stoeckenius, W. (1982) J. Bacteriol. 151, 1532-1542
11 Iida, S., Marcoli, R., and Bickle, T.A. (1981) Nature (London) 294, 374-376
12 Klaer, R., Kuhn, S., Fritz, H.J., Tillmann, E., Saint Girons, I., Habermann, P., Pfeifer, D., and Starlinger, P. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 215-224.
13 Johnsrud, L. (1979) Mol. Gen. Genet. 169, 213-218
14 Mollet, B., Iida, S., Shepherd, J., and Arber, W. (1983) Nucl. Acids Res. 11, 6319-6330
15 Schnabel, H., Schramm, E., Schnabel, R., and Zillig, W. (1982) Mol. Gen. Genet. 188, 370-377.
16 Schnabel, H. (1984) Proc. Nat. Acad. Sci. USA, submitted.