JOURNAL OF BACTERIOLOGY, June 1990, p. 3244–3249
0021-9193/90/063244-06$02.00/0

Vol. 172, No. 6

Characterization of a Gene Involved in Histidine Biosynthesis in *Halobacterium (Haloferax) volcanii*: Isolation and Rapid Mapping by Transformation of an Auxotroph with Cosmid DNA

R. KEITH CONOVER* AND W. FORD DOOLITTLE

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

Received 29 December 1989/Accepted 1 April 1990

Techniques for the transformation of halophilic archaebacteria have been developed recently and hold much promise for the characterization of these organisms at the molecular level. In order to understand genome organization and gene regulation in halobacteria, we have begun the characterization of genes involved in amino acid biosynthesis in *Halobacterium (Haloferax) volcanii*. These studies are facilitated by the many auxotrophic mutants of *H. volcanii* that have been isolated. In this project we demonstrate that cosmid DNA prepared from *Escherichia coli* can be used to transform an *H. volcanii* histidine auxotroph to prototrophy. A set of cosmid clones covering most of the genome of *H. volcanii* was used to isolate the gene which is defective in *H. volcanii* WR256. Subcloning identified a 1.6-kilobase region responsible for transformation. DNA sequence analysis of this region revealed an open reading frame encoding a putative protein 361 amino acids in length. A search of the DNA and protein data bases revealed that this open reading frame encodes histidinol-phosphate aminotransferase (EC 2.6.1.9), the sequence of which is also known for *E. coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae*.

Extreme halophiles, methanogens, and sulfur-dependent thermophiles are members of a group of organisms known as archaebacteria (32). Archaebacteria have many properties which distinguish them from eubacteria and eucaryotes (9, 31). Techniques for the genetic manipulation of extreme halophiles, *Halobacterium (Haloferax) volcanii* in particular, have been developed recently and will facilitate their study. A mating system has been demonstrated for *H. volcanii*, and it can potentially be used for strain construction and genetic mapping (22, 26). Techniques for the introduction of bacteriophage DNA and plasmid DNA have been developed for both *H. volcanii* and *Halobacterium halobium* (4–6, 15). A shuttle vector with selectable markers which can replicate in either *H. volcanii* or *Escherichia coli* has been described recently (19).

We have begun the characterization of genes involved in amino acid biosynthesis in *H. volcanii*. We wanted to see whether we could isolate genes from *H. volcanii* by transformation of auxotrophic mutants with cosmid DNA preparations from *E. coli*. Although auxotrophic mutants of *H. volcanii* can be transformed to prototrophy with *H. volcanii* genomic DNA (7), the restriction system in *H. volcanii* poses a potential problem for DNAs replicated in nonmodifying cells. Previous work indicated that the restriction system of *H. volcanii* lowers transformation efficiencies by about 4 orders of magnitude when plasmid DNA replicated in *E. coli* is used (19).

It had been shown in a previous study that the *his-1* mutation of the His^- Arg^- *H. volcanii* WR256 has a very low reversion frequency (7). For that reason, we chose this double mutant as a model. Cosmids and smaller plasmids prepared from *E. coli* were used to map the *his-1* mutation of strain WR256 to a 1.6-kilobase (kb) fragment of DNA by transformation. Sequencing of the fragment showed it to contain an open reading frame which encodes histidinol-phosphate aminotransferase, the eighth enzyme of the histidine biosynthetic pathway (2). The procedures developed here should prove to be of considerable general utility in the study of gene structure and function in *H. volcanii*.

MATERIALS AND METHODS

Strains and media. *H. volcanii* DS2 is a wild-type strain obtained from C. Woese, University of Illinois, Urbana. WFD11 is a derivative of strain DS2, which was cured of the endogenous plasmid pHV2 (4). WR256 (*his-1 arg-1*) (7) is a derivative of DS2 and was obtained from M. Mevarech, University of Tel Aviv, Tel Aviv, Israel. Cells were grown in either minimal or rich medium (22) containing 50 mM Tris hydrochloride, pH 7.2, at 37°C. Media were solidified with 2.0% (wt/vol) agar. Regeneration top agar contained 0.8% agar and 15% sucrose. Amino acids, when required, were added at 50 μg/ml. *E. coli* DH5α and DH5αF′ (Bethesda Research Laboratories, Gaithersburg, Md.) were grown by using standard procedures (20).

DNA manipulations. Standard methods for analysis of DNA, including restriction analysis, ligation, blunting of overhanging single-stranded ends, *E. coli* transformations, and plasmid DNA purifications, were performed as described by Maniatis et al. (20).

Preparation of DNA. Total DNA was isolated from *H. volcanii* by suspending cell pellets in 10 mM Tris hydrochloride (pH 8.0)–1 mM EDTA. This was followed by phenol extraction, chloroform extraction, and precipitation with ethanol. *E. coli* plasmid and cosmid DNAs were prepared by alkaline extraction (20), except that lysozyme treatment was omitted. Specific restriction fragments were isolated for subcloning or transformation of *H. volcanii* by electrophoresis in Seaplaque agarose (FMC Bioproducts, Rockland, Maine) gels, using TAE as the running buffer (20). The desired restriction fragments were excised from the gel after staining with ethidium bromide and destaining with distilled water. DNA ligations in agarose were performed by using the method described by Struhl (30).

***H. volcanii* transformations.** Transformations of *H. volca-*

* Corresponding author.

nii were performed by a modification of the method described previously (7). Cells were grown to a density of 0.5 to 1.0 A_{550} units. Portions of the cells (1.5 ml per transformation) were pelleted at 6,500 rpm in a microcentrifuge (MSE Scientific Instruments, Sussex, England) and suspended in 200 μl of spheroplasting solution. Spheroplasting was completed by the addition of 20 μl of 0.5 M EDTA, pH 8.0, gentle mixing, and incubation at room temperature for 5 min. Total genomic DNA or plasmid DNA, typically 100 ng but up to as much as 1 μg, was diluted to 10 μl with spheroplasting solution, added to the spheroplasted cells, and mixed gently. Alternatively, purified restriction fragments of DNA excised from gels after electrophoresis in Seaplaque agarose were used as the source of DNA. The piece of agarose containing the DNA of interest was melted at 75°C and then cooled to 37°C. Up to 20 μl of the melted agarose solution was added directly to the spheroplasted cells. After a 5-min incubation in the presence of DNA, an equal volume of polyethylene glycol solution (60% [vol/vol] polyethylene glycol 600, 40% spheroplasting solution) was added to the cells and mixed by inversion. After 20 min, 1 ml of rich medium containing 15% sucrose was added to the cells, and the cells were pelleted in the microcentrifuge again and then suspended in 1.5 ml of rich medium containing 15% sucrose. After incubation at 37°C overnight, the cells were pelleted and suspended in minimal medium containing 15% sucrose. Cells were then plated in regeneration top agar on minimal plates, both supplemented with arginine (50 μg/ml) if necessary. In addition to the DNA samples being tested for their ability to transform WR256, all transformation experiments included spheroplasting solution without DNA as a negative control and total DNA from DS2 or WFD11 as a positive control. The plates were sealed in plastic bags to prevent dessication and incubated at 37°C. Colonies were usually visible within 1 week, but the cells were not counted until after 2 weeks.

DNA sequencing and analysis. The DNA sequences of both strands were determined by the dideoxy-chain termination method (27) on double-stranded plasmid DNA (33), using Sequenase (United States Biochemicals, Cleveland, Ohio). Regions of the DNA sequence which were difficult to read because of compressions were subcloned into the single-stranded phage vectors M13um20 and M13um21 (International Biotechnologies, Inc., New Haven, Conn.) and sequenced with Taquence (United States Biochemicals), using the nucleotide analog 7-deaza-guanosine. Nested deletions were prepared with exonuclease III as described by Henikoff (16). Analysis of DNA sequence data was carried out with DNA Strider (21) and University of Wisconsin Genetics Computer Group software (10). Data bank searches were carried out with FASTA and TFASTA (24). Multiple sequence alignments were performed by using Multalin (8).

RESULTS

Mapping the *his-1* mutation. The cosmid library used in this study contains 35- to 45-kb fragments of *H. volcanii* DS2 DNA generated by partial digestion with *Mlu*I (3). These cosmids are being used tc construct a "bottom-up" map of restriction sites in the *H. volcanii* genome, and a physical by-product of this process is an ordered set of minimally overlapping clones. At the time this project was initiated the set contained 130 cosmids, representing perhaps 95% of the genome. In order to reduce the number of transformations required, the DNA (approximately 200 ng) from each of the 130 cosmids was pooled into 26 groups of 5 cosmids each. WR256 (*his-1 arg-1*) was transformed with each of the cosmid DNA pools and plated on minimal plates supplemented with arginine in order to select for His^+ transformants. Two of the pools (B and C) yielded hundreds of His^+ transformants each, while the other pools and the negative control resulted in fewer than 10 transformants each. The 10 cosmids in the two positive pools were then tested individually. Two cosmids, 54 and 477 (from pools B and C, respectively), were able to transform WR256 to His^+.

Restriction endonuclease analysis of the DNA inserts in cosmids 477 and 54, as described by Charlebois et al. (3), showed them to overlap by approximately 29 kb. The two cosmids have seven *Mlu*I fragments in common, with sizes of 17.5, 4.9, 2.0, 1.9, 0.9, 0.76, and 0.67 kb. The order of these fragments has not yet been determined. Each of these *Mlu*I fragments was purified by gel electrophoresis in Seaplaque agarose and subcloned into the *Mlu*I site of pIBI31 (International Biotechnologies, Inc.). Each of the resulting clones was tested in the transformation assay on WR256. Only clones containing the 17.5-kb *Mlu*I fragment (p477.1A and p477.1D have the 17.5-kb insert in opposite orientations) were able to transform the *his-1* mutation of WR256. A restriction map of this 17.5-kb fragment is shown in Fig. 1a. Also shown in Fig. 1a are four deletion derivatives of p477.1. Plasmids pAP1 and pDP1 were constructed by deletion of a *Pst*I fragment from p477.1A and p477.1D, respectively. Plasmids pAH1 and pDH1 were constructed by deletion of a *Hin*dIII fragment from p477.1A and p477.1D, respectively. pAP1 and pDH1 were unable to transform the *his-1* mutant, while pAH1 and pDP1 could (Fig. 1a). This result indicates that the affected gene lies somewhere within the 5.7-kb insert of pDP1.

The gene was mapped more precisely in a slightly different way. In order to avoid some of the subcloning which would have otherwise been required, we investigated the possibility of transforming *H. volcanii* with small restriction fragments. We have already shown that auxotrophic mutants of *H. volcanii* can be transformed with linear *H. volcanii* genomic DNA (7). In that study, we showed that transformation efficiency is dependent on the length of input DNA and that DNA less than 5 kb in length transforms *H. volcanii* poorly. We first transformed WR256 with p477.1, pAH1, and pDP1 that had each been linearized by digestion with a restriction enzyme. In all cases, transformation frequencies were nearly as high as with the circular plasmids (data not shown). We next tried transformation with the 5.7-kb gel-purified (Seaplaque agarose) insert of pDP1. The presence of agarose in the DNA preparation had little effect on the transformation. A more detailed restriction map of pDP1 and the results of transformations of WR256 with different portions of this clone are shown in Fig. 1b. The sequences flanking the 3.3-kb internal *Xho*I-*Nru*I fragment and the 3.0-kb *Pst*I-*Sac*I fragment did not transform the *his-1* mutation in WR256, while the 3.3-kb *Xho*I-*Nru*I fragment and the 2.7-kb *Sac*I-*Mlu*I fragment did. These results indicate that the *his-1* defect is located in the region of overlap between the latter two fragments, which is a 1.6-kb fragment between the *Sac*I and *Nru*I sites. The 2.7-kb *Sac*I-*Mlu*I fragment of pDP1 was subcloned by making the ends of the fragment blunt by treatment with T4 DNA polymerase (20) and ligating the resulting DNA into the *Sma*I site of pBGS18, a kanamycin-resistant derivative of pUC18 (29). Both orientations of the desired clone were obtained and were called pT4C and pT4G, respectively.

The insert in pT4C was further characterized by using it as a probe in a Southern hybridization experiment with DNA

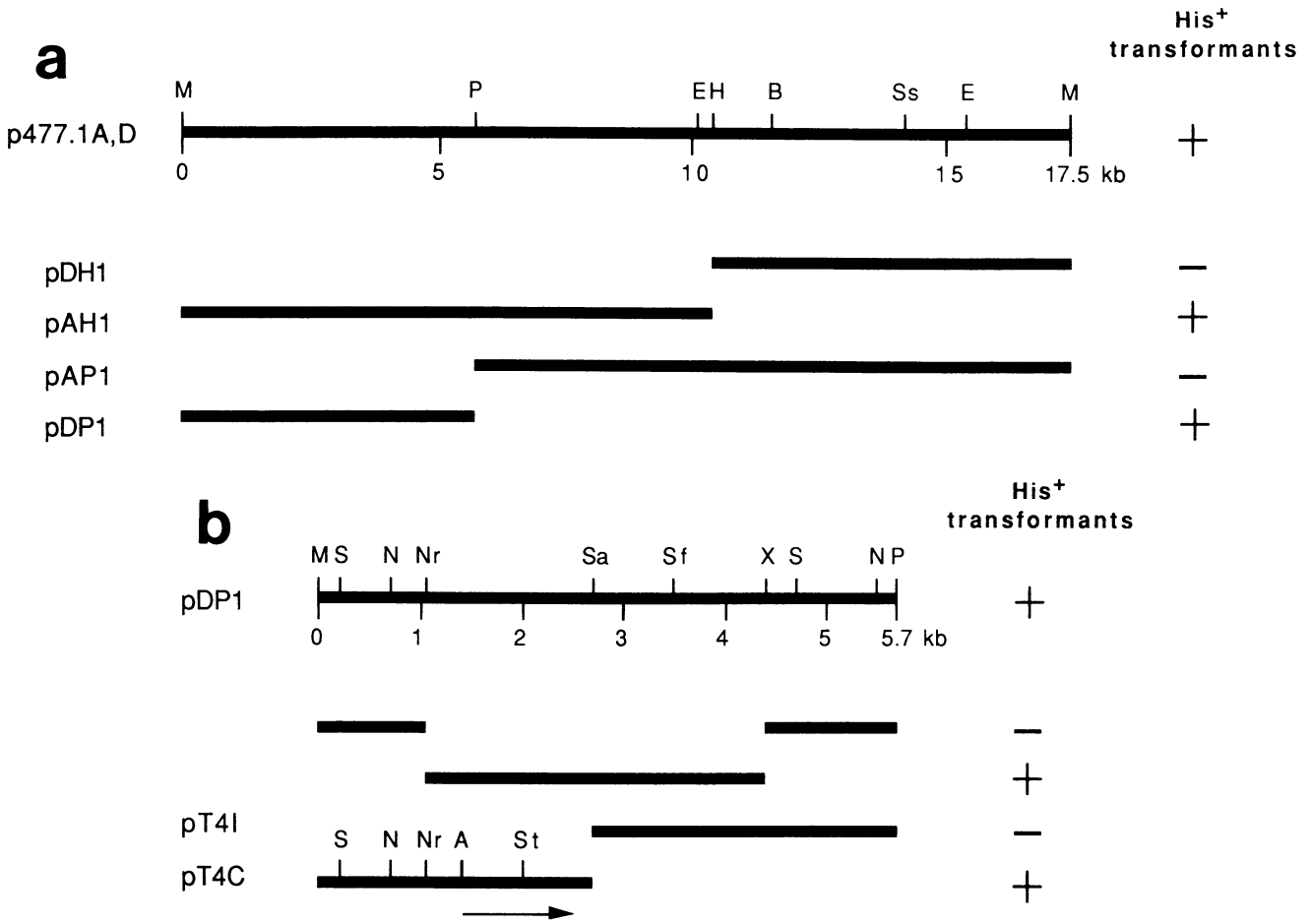

FIG. 1. (a) Deletion analysis of p477.1A and -D. The 17.5-kb *Mlu*I fragment from cosmid 477 was subcloned in both orientations into the *Mlu*I site of pIBI31, resulting in p477.1A and p477.1D. Plasmids pDH1 and pAH1 were constructed by deletion of a *Hin*dIII fragment from p477.1D and p477.1A, respectively. Plasmids pAP1 and pDP1 were constructed by deletion of a *Pst*I fragment from p477.1A and p477.1D, respectively. Only the inserts are shown here. Each of these constructs was tested for its ability to transform the *his-1* mutation in WR256 to His^+. Restriction sites are abbreviated as follows: M, *Mlu*I; P, *Pst*I; E, *Eco*RI; H, *Hin*dIII; B, *Bgl*II; Ss, *Ssp*I. (b) Deletion analysis of pDP1. Different regions of the pDP1 insert were tested for their ability to transform the *his-1* mutation of WR256 to His^+. Restriction fragments were purified by gel electrophoresis, excised from the gel, and used in the transformation assay. The *Mlu*I-*Sac*I and *Sac*I-*Pst*I fragments were subcloned into the cloning vector pBGS18, resulting in pT4C and pT4I, respectively. The position and direction of the open reading frame are indicated by the arrow. The abbreviations used for restriction sites are the same as those in panel a, with the following additions: S, *Sma*I; N, *Not*I; Nr, *Nru*I; Sa, *Sac*I; Sf, *Sfi*I; X, *Xho*I; A, *Afl*II; St, *Stu*I.

from *H. volcanii* DS2, WFD11, and WR256 digested separately with *Mlu*I and *Sma*I. The hybridization patterns were identical for all three strains, with a single band of the expected size in each lane. This indicates that this fragment of DNA is unique in the genome of *H. volcanii* and that the mutagenesis procedures which were used to isolate WFD11 (4) and WR256 (22) probably did not produce genomic rearrangements in the vicinity of this fragment. The hybridization pattern of pT4C to 12 isolates of WR256 which were transformed to His^+ with pT4C DNA (circular) was also studied. In all cases, the hybridization pattern was identical to that obtained with DS2 (data not shown). This indicates that vector sequences in pT4C are not present in these cells and that the *his-1* mutation in WR256 was probably corrected by two homologous recombination events between the input DNA and the chromosomal *his-1* locus, although in other situations we have observed incorporation of vector DNA (19).

Nucleotide and derived amino acid sequences. The DNA sequence of the 1.6-kb *Sac*I-*Nru*I restriction fragment was determined by the dideoxy-chain termination method and was found to include only a single large open reading frame (ORF). The DNA sequence of the ORF and the amino acid sequence derived from it are shown in Fig. 2. The ORF encodes 361 amino acids, corresponding to a protein with a molecular mass of 39 kilodaltons. The predicted protein is acidic (isoelectric point = 4.23), which is expected for halophilic proteins (11). The DNA sequences immediately upstream and downstream to the ORF do not appear to encode proteins, so this ORF is probably not part of a multigene transcriptional unit. The sequence CTTAAG is found about 30 bases upstream from the putative start codon. This sequence may be an element of halophilic promoters (25).

Codon usage in the putative protein is very biased and is shown in Table 1. There is a strong preference for C or G in position 3 of the codon, with C almost twice as frequent as G (59 and 32%, respectively). The preference for C or G

```
                                                -80                                              -40          Afl II
AGTCGTTCGGGCGGCGCTCGGCTGACGGCCGTCGGTCGTCGCGTCCCCAACCCGACCCCCTACCGCCACGTCCGACCCGGAGTACGCACCCTTAAGAACCGCGACCCGCATTTTCCGACC
                                                                                                                ------

                                         30                                      60
ATG CAA CCA CGG GAC CTC TCC GCG CAC GCT CCC TAC GTA CCC GGC CGC GGG ACA GAG GAG GTC GCC CGC GAA CTC GGA ATG GAC CCC GAG
 M   Q   P   R   D   L   S   A   H   A   P   Y   V   P   G   R   G   T   E   E   V   A   R   E   L   G   M   D   P   E
 1
                                        120                                     150
GAC CTG ACG AAA CTC TCC TCG AAC GAG AAC CCC CAC GGC CCG AGT CCG AAG GCG GTC GCC GCC ATC GAA GAC GCC GCG CCG ACC GTG AGC
 D   L   T   K   L   S   S   N   E   N   P   H   G   P   S   P   K   A   V   A   A   I   E   D   A   A   P   T   V   S
31
                                        210                                     240
GTC TAC CCG AAG ACC GCC CAC ACG GAC CTG ACC GAA CGC CTC GCC GAC AAG TGG GGC CTC GCA CCC GAA CAG GTG TGG GTG TCT CCC GGC
 V   Y   P   K   T   A   H   T   D   L   T   E   R   L   A   D   K   W   G   L   A   P   E   Q   V   W   V   S   P   G
61
                                        300                                     330
GCG GAC GGC TCT ATC GAC TAC CTG ACC CGC GCG GTG CTC GAA CCG GAC GAC CGG ATT CTC GAA CCC GCG CCC GGC TTT TCG TAC TAC TCG
 A   D   G   S   I   D   Y   L   T   R   A   V   L   E   P   D   D   R   I   L   E   P   A   P   G   F   S   Y   Y   S
91
                                        390                                     420
ATG AGC GCC CGC TAC CAC CAC GGC GAC GCC GTC CAG TAC GAG GTG TCG AAG GAC GAC GAC TTC GAA CAG ACC GCC GAC CTC GTC CTC GAC
 M   S   A   R   Y   H   H   G   D   A   V   Q   Y   E   V   S   K   D   D   D   F   E   Q   T   A   D   L   V   L   D
121
                                        480                                     510
GCC TAC GAC GGC GAG CGC ATG GTC TAC CTC ACA ACG CCG CAC AAC CCC ACC GGT TCC GTG CTC CCG CGG GAG GAA CTC GTC GAA CTG GCC
 A   Y   D   G   E   R   M   V   Y   L   T   T   P   H   N   P   T   G   S   V   L   P   R   E   E   L   V   E   L   A
151
                                        570             Stu 1                   600
GAG TCG GTC GAA GAG CAC ACG CTC CTC GTC GTC GAC GAG GCC TAC GGC GAG TTC GCC GAG GAG CCG TCG GCC ATC GAC CTC TTG TCG GAG
 E   S   V   E   E   H   T   L   L   V   V   D   E   A   Y   G   E   F   A   E   E   P   S   A   I   D   L   L   S   E
181
                                        660                                     690
TAC GAC AAC GTC GCG GCC CTG CGG ACG TTC TCG AAG GCG TAC GGG CTG GCC GGC CTC CGC ATC GGC TAC GCC TGC GTG CCC GAG GCG TGG
 Y   D   N   V   A   A   L   R   T   F   S   K   A   Y   G   L   A   G   L   R   I   G   Y   A   C   V   P   E   A   W
211
                                        750                                     780
GCC GAC GCC TAC GCC CGC GTG AAC ACG CCG TTC GCC GCC AGC GAG GTC GCC TGC CGC GCC GCG CTC GCC GCG CTC GAC GAC GAG GAA CAC
 A   D   A   Y   A   R   V   N   T   P   F   A   A   S   E   V   A   C   R   A   A   L   A   A   L   D   D   E   E   H
241
                                        840                                     870
GTC GAG AAA TCC GTC GAG TCG GCC CGG TGG TCC CGC GAC TAT CTC CGC GAA CAC CTC GAC GCG CCG ACG TGG GAA AGC GAG GGC AAC TTC
 V   E   K   S   V   E   S   A   R   W   S   R   D   Y   L   R   E   H   L   D   A   P   T   W   E   S   E   G   N   F
271
                                        930                                     960
GTC CTC GTC GAG GTC GGC GAC GCC ACG GCC GTC ACC GAG GCC GCC CAG CGC GAG GGC GTC ATC GTC CGC GAC TGC GGG AGC TTC GGC CTG
 V   L   V   E   V   G   D   A   T   A   V   T   E   A   A   Q   R   E   G   V   I   V   R   D   C   G   S   F   G   L
301
                                       1020                                    1050
CCG GAG TGC ATC CGC GTC TCC TGC GGC ACG GAA ACC CAG ACC AAG CGC GCC GTG GAC GTG CTC AAC CGC ATC GTC TCG GAG GTG CCG ACG
 P   E   C   I   R   V   S   C   G   T   E   T   Q   T   K   R   A   V   D   V   L   N   R   I   V   S   E   V   P   T
331

GCG  TGA  GAGACGACGACACCGGCACGCCCGGCACCGGAAAGACCACGGCGACCGAGCCGGTCGCCGCCGACCTCGACCTCGACGTGGTCCACCTCAACCGACTCGTGAAAGACGAGG
 A   End
361
```

FIG. 2. DNA and derived amino acid sequences of the *H. volcanii hisC* gene. The DNA sequence is numbered above, and the amino acid sequence is numbered below. The restriction sites for *Afl*II and *Stu*I are indicated. The sequence data reported here have been submitted to GenBank and have been assigned accession number M33161.

TABLE 1. Codon usage in the *hisC* gene of *H. volcanii*[a]

Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used
Phe	UUU	1	Ser	UCU	2	Tyr	UAU	1	Cys	UGU	0
Phe	UUC	6	Ser	UCC	6	Tyr	UAC	14	Cys	UGC	5
Leu	UUA	0	Ser	UCA	0	OCH	UAA	0	OPA	UGA	1
Leu	UUG	1	Ser	UCG	10	AMB	UAG	0	Trp	UGG	5
Leu	CUU	0	Pro	CCU	0	His	CAU	0	Arg	CGU	0
Leu	CUC	22	Pro	CCC	10	His	CAC	9	Arg	CGC	16
Leu	CUA	0	Pro	CCA	1	Gln	CAA	1	Arg	CGA	0
Leu	CUG	7	Pro	CCG	12	Gln	CAG	5	Arg	CGG	5
Ile	AUU	1	Thr	ACU	0	Asn	AAU	0	Ser	AGU	1
Ile	AUC	7	Thr	ACC	9	Asn	AAC	7	Ser	AGC	5
Ile	AUA	0	Thr	ACA	2	Lys	AAA	2	Arg	AGA	0
Met	AUG	4	Thr	ACG	10	Lys	AAG	6	Arg	AGG	0
Val	GUU	0	Ala	GCU	1	Asp	GAU	0	Gly	GGU	1
Val	GUC	22	Ala	GCC	31	Asp	GAC	28	Gly	GGC	16
Val	GUA	1	Ala	GCA	1	Glu	GAA	14	Gly	GGA	1
Val	GUG	11	Ala	GCG	13	Glu	GAG	25	Gly	GGG	3

[a] OCH, AMB, and OPA, ochre, amber, and opal termination codons, respectively.

reflects the high G+C content of the *H. volcanii* genome (65%). The preference for C over G may be a result of the lower abundance of tRNAs capable of recognizing G in the third position of codons in *H. volcanii* (13, 14, 18). The same preference in position 3 of the codon is seen in *Halobacterium cutirubrum* (28), while *Halobacterium halobium* shows the strong preference for C or G but no preference for C over G (1).

The putative protein sequence was used to search the protein and DNA data banks for similar sequences with FASTA and TFASTA, respectively (24). Only three significant matches were found. The matches were *hisC* from *E. coli* (12), *hisH* from *Bacillus subtilis* (17), and *his5* from *Saccharomyces cerevisiae* (23). All three genes encode histidinol-phosphate aminotransferase (EC 2.6.1.9), which catalyzes the conversion of imidazolylacetolphosphate to L-histidinol phosphate, the eighth step in histidine biosynthesis (2). An alignment of all four protein sequences is shown in Fig. 3. This alignment was performed with the hierarchical clustering program of Corpet by using the Dayhoff matrix and a gap penalty of 7 (8). On the basis of these sequence similarities, we propose that the gene in which the *his-1* mutation occurs be called *hisC*, consistent with the nomenclature for *E. coli*. A summary of the pairwise sequence identities obtained from this alignment is shown in Table 2. The *H. volcanii hisC* sequence is closer to the eubacterial sequences than it is to the eucaryotic (*S. cerevisiae*) sequence. Introduction of the *H. volcanii* gene into an *E. coli hisC* auxotroph did not restore this strain to prototrophy. This is not surprising, given the difference between eubacterial and archaebacterial promoters (25) and the unusual stability requirements of proteins from halophiles (11).

DISCUSSION

We have shown that cosmid DNA may be used to transform auxotrophic mutations in *H. volcanii* to prototrophy. *H. volcanii* does appear to have a restriction-modification system, which lowers the transformation efficiency of *E. coli* isolated DNA by approximately 4 orders of magnitude (19). Yet, given the high frequency of *H. volcanii* transformation (approximately 10^7 transformants per μg of plasmid DNA) and the low reversion frequency of the mutant used in this study (approximately 10^{-9}), we can very efficiently map genes in *H. volcanii* by transforming auxotrophs with cosmid DNA isolated from *E. coli*. Transformation is probably the result of replacement of the defective gene by means of two homologous recombination events. DNA fragments as small as 2 kb in size work efficiently. We have now achieved transformation with less-than-gene-sized fragments of DNA, so this technique can also be used to map the location of a mutation within a gene (data not shown). Fine mapping of genes is relatively fast because one can bypass subcloning steps and transform directly with DNA isolated from low-melting-temperature agarose.

A 1.6-kb region of DNA found to be essential for transformation of the *his-1* mutation in strain WR256 was sequenced and found to contain a single large ORF. The putative protein contains 361 amino acids and is acidic. Most halophilic proteins are acidic, probably because being highly charged enables them to be soluble in the highly saline intracellular environment (11). The putative protein has significant sequence identity with the histidinol-phosphate aminotransferases from *E. coli*, *B. subtilis*, and *S. cerevisiae*. Given this sequence identity and the fact that the *his-1* mutation of WR256 results in histidine auxotrophy, we propose that the *his-1* mutation affects expression of the *hisC* gene in *H. volcanii*. The sequence of the *hisC* gene product from *H. volcanii* shares greater identity with the two eubac-

```
     1
Hv   MQPRD-LSAH APYVPGRGTE EVARELGMDP EDLTKLSSNE NPHGPSP--K
Bs   MRIKEHLKQL KPYQPGKPIE AVKSEYGLD- -KVVKLASNE NPYGCSE--A
Ec   MSTVTITDLA RENVRNLTPY QSARRLGGN- -GDVWLNANE YP-TAVEFQL
Sc   M-VFDLKRIV RPKIYNLEPY RCARDDFTE- -G-ILLDANE NAHGPTPVEL
     M           p          ar  g         L  NE np g

     51
Hv   AVAAIEDAAP TVSVYPKTAH TDLT---ERL ADKWGLAP-- -EQVWVSPGA
Bs   AKEALHHEIQ QLALYPDGYS AALR---TRL SKHLNVSE-- -TSLIFGNGS
Ec   TQQTLNRYPE CQP---KAVI ENY----AQY AG---VKP-- -EQVLVSRGA
Sc   SKTNLHRYPD PHQLEFKTAM TKYRNKTSSY ANDPEVKPLT ADNLCLGVGS
       l         lypk       r        a    v p           G

     101
Hv   DGSIDYLTRA V-LEPDDRIL EPAPGFSYYS MSARYHHGDA VQYE-VSKDD
Bs   DEIIQIICRA F-LNDKTNTV TAAPTFPQYK HNAVIEGAEV REIA-LRPDG
Ec   DEGIELLIRA FCEPGKDAIL YCPPTYGMYS VSAETIGVE- --CRTVPTPD
Sc   DESIDAIIRA CCVRE-EKIL VLPPTYSMSS VCANINDIEV VQCPLIVSDG
     De I    RA          il    Pt   ys   A     ev vq      d

     151
Hv   DFEQTADLVL DAYDGE---R MVYLTTPHNP TGSVLPREEL VELAESVEEH
Bs   S--HDLDAML EAIDEQ--TQ VVWICSPNNP TGTYTSEGEL LAFLERVPSR
Ec   NWQLDLQGIS DKLDG---VK AVYVCSPNNP TGQLINPQDF RTLLELTRGK
Sc   SFQMDTEAVL TILKNDSLIK LMFVTSPGNP TGAKIKTSLI EKVLQ-NWDN
      f  d    l    d        v   sP NP TG            le

     201
Hv   TLLVVDEAYG EF--AEE-PS AIDLLSEYDN VAALRTFSKA YGLAGLRIGY
Bs   VLVVLDEAYY EYVTAEDYPE TVPLLSKYSN LMILRTFSKA YGLAALRVGY
Ec   AIVVADEAYI EF--CPQ-AS LAGWLAEYPH LAILRTLSKA FALAGLRCGF
Sc   GLVVVDEAYV DF--CGG-ST -APLVTKYPN LVTLQTLSKS FGLAGIRLGM
      lvV DEAY  ef            ll  Y n l  LrT SKa  gLAglR G

     251
Hv   ACVPEAWADA YARVNTPFAA SEVACRAALA ALDDEEHVEK SVESARWS-R
Bs   GIADENLIRQ IEPAREPFNT SRLGQAAAIA ALYDQAFIAS CVEQNNAGLQ
Ec   TLANEEVINL LMKVIAPYPL STPVADIAAQ ALSPQGIVAM RERVAQIIAE
Sc   TYATAELARI LNAMKAPYNI SSLASEYALK AVQDSNLKKM EATSKIINEE
       a e             P    S      A   Al d

     301
Hv   DY--LREHLD APT----WES --EGNFVLVE VGDATA-VTE AAQ-----RE
Bs   QYYDFAKTHG LKC----YPS --QTNFVLID FKRPADELFQ ALL-----EK
Ec   REYLIAALKE IPCVEQVFDS --ETNYILAR FKASSAVF-K SLW-----DQ
Sc   KMLLLKELTA LDYVDDQYVG GLDANFLLIR INGGDNVLAK KLYYQLATQS
       yl                s     Nf L                 l

     351
Hv   GVIVR-DCGS FGLPECIRVS CGTETQTKRA VDVLNRIVSE VPTA
Bs   GYIVR-SGNA LGFPTSLRIT IGTKEQNEEI LAILAEIL-- ----
Ec   GIILRDQNKQ PSLSGCLRIT VGTREESQRV IDALRAEQV- ----
Sc   GVVVRFRGNE LGCSGCLRIT VGTHEENTHL IKYFKETLYK LANE
     G ivR       g   clRit  GT e         l
```

FIG. 3. Alignment of four histidinol-phosphate aminotransferase protein sequences. Hv, *H. volcanii*; Bs, *B. subtilis*; Ec, *E. coli*; and Sc, *S. cerevisiae*. A consensus sequence is indicated below the other sequences. Positions where all four amino acids are identical are indicated with a capital letter. Positions where three of four amino acids are identical (or two of three in cases where one of the sequences has a gap) are indicated with a lowercase letter. Gaps are indicated by dashes.

TABLE 2. Percent sequence identity among the four histidinol-phosphate aminotransferases

Organism	% Histidinol-phosphate aminotransferase amino acid sequence identity with[a]:			
	H. volcanii	*B. subtilis*	*E. coli*	*S. cerevisiae*
H. volcanii		31	25	21
B. subtilis	33		23	25
E. coli	32	30		30
S. cerevisiae	27	31	34	

[a] Values in the upper right half were derived from the multiple alignment shown in Fig. 3. Values in the lower left half were derived from pairwise alignments.

terial sequences (approximately 30%) than with the eucaryotic sequence (approximately 25%). The most surprising thing about the multiple alignment of all four sequences (Fig. 3 and Table 2) is that the *E. coli* and *B. subtilis* sequences are so different from one another, given that they share a common ancestor much more recently than do *H. volcanii*, *S. cerevisiae*, and the eubacteria.

As the sequences immediately upstream and downstream from the *H. volcanii hisC* gene do not appear to encode proteins, the gene may not be part of an operon. Beginning 30 bases upstream from the putative start codon, there is a sequence which matches part of the halobacterial consensus promoter. Initiation of transcription usually occurs approximately 30 nucleotides downstream from this sequence (25), so transcription of the *H. volcanii hisC* gene may initiate very close to the start codon.

The availability of auxotrophic mutants and a method for the introduction of DNA into *H. volcanii* enables one to isolate genes from *H. volcanii* either by the method described in this paper or by using the shuttle vector described by Lam and Doolittle (19). Although shuttle vector cloning is simpler, cosmid transformation has several advantages. Recipients need not be cured of endogenous plasmids. Adjacent genes are identified automatically, and the gene of interest can be characterized fully. Most importantly, our cosmids will soon be ordered into a set covering the entire *H. volcanii* chromosome (R. L. Charlebois, L. C. Schalkwyk, J. D. Hofman, and W. F. Doolittle, unpublished results). Cosmid transformation will immediately map genes to known regions corresponding to less than 1% of the genetic map.

ACKNOWLEDGMENTS

We thank Robert Charlebois and Jason Hofman for providing the cosmid clones used in this study, Moshe Mevarech for providing *H. volcanii* WR256, and David Spencer for help with the protein sequence alignments.

W.F.D. is a Fellow of the Canadian Institute for Advanced Research. This work was supported by grants from the U.S. Office of Naval Research (N00014-88-J-1030) and the Canadian Medical Research Council (MT4467).

LITERATURE CITED

1. **Blanck, A., and D. Oesterhelt.** 1987. The halo-opsin gene. II. Sequence, primary structure of halorhodopsin and comparison with bacteriorhodopsin. EMBO J. **6**:265–273.
2. **Brenner, M., and B. N. Ames.** 1971. The histidine operon and its regulation, p. 349–387. *In* H. J. Vogel (ed.), Metabolic pathways, vol. 5. Metabolic regulation. Academic Press, Inc., New York.
3. **Charlebois, R. L., J. D. Hofman, L. C. Schalkwyk, W. L. Lam, and W. F. Doolittle.** 1989. Genome mapping in halobacteria. Can. J. Microbiol. **35**:21–29.
4. **Charlebois, R. L., W. L. Lam, S. W. Cline, and W. F. Doolittle.** 1987. Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. Proc. Natl. Acad. Sci. USA **84**:8530–8534.
5. **Cline, S. W., and W. F. Doolittle.** 1987. Efficient transfection of the archaebacterium *Halobacterium halobium*. J. Bacteriol. **169**:1341–1344.
6. **Cline, S. W., W. L. Lam, R. L. Charlebois, L. C. Schalkwyk, and W. F. Doolittle.** 1989. Transformation methods for halophilic archaebacteria. Can. J. Microbiol. **35**:148–152.
7. **Cline, S. W., L. C. Schalkwyk, and W. F. Doolittle.** 1989. Transformation of the archaebacterium *Halobacterium volcanii* with genomic DNA. J. Bacteriol. **171**:4987–4991.
8. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16**:10881–10890.
9. **Dennis, P. P.** 1986. Molecular biology of archaebacteria. J. Bacteriol. **168**:471–478.
10. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
11. **Eisenberg, H., and E. J. Wachtel.** 1987. Structural studies of halophilic proteins, ribosomes, and organelles of bacteria adapted to extreme salt concentrations. Annu. Rev. Biophys. Biophys. Chem. **16**:69–92.
12. **Grisolia, V., M. S. Carlomagno, A. G. Nappo, and C. B. Bruni.** 1985. Cloning, structure, and expression of the *Escherichia coli* K-12 *hisC* gene. J. Bacteriol. **164**:1317–1323.
13. **Gupta, R.** 1984. *Halobacterium volcanii* tRNAs. J. Biol. Chem. **259**:9461–9471.
14. **Gupta, R.** 1986. Transfer RNAs of *Halobacterium volcanii*: sequences of five leucine and three serine tRNAs. Syst. Appl. Microbiol. **7**:102–105.
15. **Hackett, N. R., and S. DasSarma.** 1989. Characterization of the small endogenous plasmid of *Halobacterium* SB3 and its use in transformation of *Halobacterium halobium*. Can. J. Microbiol. **35**:86–91.
16. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. Gene **28**: 351–359.
17. **Henner, D. J., L. Band, G. Flaggs, and E. Chen.** 1986. The organization and nucleotide sequence of the *Bacillus subtilis hisH*, *tyrA* and *aroE* genes. Gene **49**:147–152.
18. **Ikemura, T.** 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**:13–34.
19. **Lam, W. L., and W. F. Doolittle.** 1989. Shuttle vectors for the archaebacterium *Halobacterium volcanii*. Proc. Natl. Acad. Sci. USA **86**:5478–5482.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Marck, C.** 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
22. **Mevarech, M., and R. Werczberger.** 1985. Genetic transfer in *Halobacterium volcanii*. J. Bacteriol. **162**:461–462.
23. **Nishiwaki, N., N. Hayashi, S. Irie, D.-H. Chung, S. Harashima, and Y. Oshima.** 1987. Structure of the yeast his5 gene responsive to general control of amino acid biosynthesis. Mol. Gen. Genet. **208**:159–167.
24. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
25. **Reiter, W. D., P. Palm, and W. Zillig.** 1988. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res. **16**:1–19.
26. **Rosenshine, I., R. Tchelet, and M. Mevarech.** 1989. The mechanism of DNA transfer in the mating system of an archaebacterium. Science **245**:1387–1389.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Shimmin, L. C., and P. Dennis.** 1989. Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. EMBO J. **8**:1225–1235.
29. **Spratt, B. G., P. J. Hedge, S. Heesen, A. Edelman, and J. K. Broome-Smith.** 1986. Kanamycin-resistant vectors that are analogues of plasmids pUC8, pUC9, pEMBL8, and pEMBL9. Gene **41**:337–342.
30. **Struhl, K.** 1985. A rapid method for creating recombinant DNA molecules. Biotechniques **3**:452–453.
31. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.
32. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74**:5088–5090.
33. **Zhang, H., R. Scholl, J. Browne, and C. Somerville.** 1988. Double stranded DNA sequencing as a choice for DNA sequencing. Nucleic Acids Res. **16**:1220.