Cloning, Sequencing, and Expression in *Escherichia coli* of the Gene Coding for Malate Dehydrogenase of the Extremely Halophilic Archaebacterium *Haloarcula marismortui*†

Fabrice Cendrin,‡ Jadwiga Chroboczek,§ Giuseppe Zaccai,§ Henryk Eisenberg,∥ and Moshe Mevarech*,‡

Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel, CNRS, URA 1333, Institute Laue Langevin 156X, 38042 Grenoble Cedex, France, and Structural Biology Department, Weizmann Institute of Science, Rehovot 76100, Israel

Received August 27, 1992; Revised Manuscript Received December 8, 1992

ABSTRACT: The gene coding for the enzyme malate dehydrogenase (MDH) of the extremely halophilic archaebacterium *Haloarcula marismortui* was isolated and sequenced. The enzyme is composed of 303 amino acids, and its molecular mass is 32 638 Da. The deduced amino acid sequence of the enzyme was found to be more similar to the sequence of L-lactate dehydrogenase (L-LDH) from various sources than to the sequence of other MDHs. The structural gene was cloned in the *Escherichia coli* expression vector pET11a, and large amounts of a soluble but inactive form of the enzyme were produced upon its induction. Activation of the enzyme was obtained by increasing the salt concentration to 3 M NaCl. The recombinant protein was purified to homogeneity and shown to be indistinguishable from the native enzyme isolated from halobacteria. These findings present the first example of the successful expression of a halobacterial gene coding for a soluble protein in *Escherichia coli* and its recovery as a functional enzyme. Site-directed mutagenesis was employed to modify Arg^{100} on the enzyme to Gln. This modification produced an enzyme that has considerably higher specificity for pyruvate (the substrate of L-LDH) than for oxaloacetate (the substrate of MDH). The mutation also caused a modification in the relative activities of the enzyme at different salt concentrations. The greater similarity of the amino acid sequence of the halobacterial MDH to that of L-LDHs than to that of MDHs sheds light on the molecular evolution of these enzymes.

Adaptation of living organisms to extreme extracellular salinity, as can be found for instance in the Dead Sea and in the Great Salt Lake in Utah, can be manifested by two different mechanisms. Halophilic eukaryotes and eubacteria overcome the extracellular osmotic pressure by accumulating "osmoprotectants" (Yancey et al., 1982). On the other hand, halophilic archaebacteria accumulate salt (particularly K^+ ions) to concentrations that can reach and exceed saturation (Christian & Waltho, 1962). The biochemical machinery of these prokaryotes has, therefore, been adapted in the course of evolution to be able to function at salt concentrations at which most biochemical systems will cease to function. The biochemical and biophysical properties of several halophilic enzymes were studied in great detail [for a recent review, see Eisenberg et al. (1992)], and as a general rule, it was found that halobacterial enzymes are stabilized by multimolar concentration of salts and in most cases the salt also stimulates the catalytic activity. This stabilization of halophilic proteins in solvents containing high salt concentrations has been discussed in terms of apparent peculiarities in their composition, i.e., higher proportions of acidic over basic residues and borderline hydrophobic amino acid residues when compared to their nonhalophilic homologs (Lanyi, 1974). Better understanding of the nature of the interactions that maintain the native structure of the halophilic enzymes, however, awaits detailed structural information about these enzymes at atomic resolution.

The enzyme malate dehydrogenase (EC 1.1.1.37) of the extremely halophilic archaebacterium *Haloarcula marismortui* (hMDH)[1] has been studied most extensively using a wide range of biochemical and biophysical methods (Pundak et al., 1981; Pundak & Eisenberg, 1981; Zaccai et al., 1989). It was found that the enzyme is only stable in highly concentrated solutions of certain salts. In order to gain a deeper understanding of the structural features of hMDH that are responsible for its "halophilic adaptation", we have undertaken the task of isolating the gene coding for the enzyme, determining the amino acid sequence of the enzyme, and expressing the gene in *Escherichia coli* in order to enable site-directed mutagenesis and purification of mutated enzymes. Results of this molecular genetic approach are presented.

MATERIALS AND METHODS

Bacterial Strains. *E. coli* K12 71/18 (Yanisch-Peron, 1985), *E. coli* K12 HMS174, and *E. coli* B BL21 (DE3) *lys S* (Studier et al., 1990) were the bacterial strains used.

Enzymatic Assays. Standard enzymatic assays were performed in 1 mL of buffer containing 4 M NaCl, 20 mM Tris-HCl, pH 8.0, 1 mM oxaloacetate, and 0.1 mM NADH. One enzyme unit is defined as the amount of enzyme that catalyzes the oxidation of 1 μmol of NADH/min.

Determination of Protein Concentration. The concentration of hMDH was determined by measuring the light absorbance at 280 nm using an extinction coefficient of $E_{0.1\%} = 0.85$

† This work was supported by a grant from The International Scientific Cooperation France–Israel.

* To whom correspondence should be addressed.
‡ Tel Aviv University.
§ Institute Laue Langevin 156X.
∥ Weizmann Institute of Science.

[1] Abbreviations: MDH, malate dehydrogenase; hMDH, halobacterial malate dehydrogenase; L-LDH, L-lactate dehydrogenase; Tris, tris(hydroxymethyl)aminomethane; NADH, reduced nicotinamide adenine dinucleotide; SDS, sodium dodecyl sulfate; IPTG, isopropyl β-D-thiogalactopyranoside.

0006-2960/93/0432-4308$04.00/0

(Bonnete et al., unpublished results). The concentration of the R100Q mutant of hMDH was determined by the method of Bradford (1976) using hMDH as a standard.

SDS–Polyacrylamide Gel Electrophoresis. Crude extracts or purified protein was precipitated in 10% trichloroacetic acid on ice for 30 min. The pellet was suspended in "sample buffer" (5% w/v SDS, 20% w/v glycerol, and 10 mM 1,4-dithio-L-threitol), boiled, and loaded onto an SDS–polyacrylamide gel [10% w/v acrylamide and 0.05% w/v *N,N'*-methylenebis(acrylamide)].

DNA Manipulation and Sequencing. DNA was prepared from *H. marismortui* cells as described by Rosenshine et al. (1987). Partial *Mbo*I digestion of genomic DNA and sucrose gradient centrifugation for DNA size fractionation were performed according to Maniatis et al. (1982). After being subcloned into M13mp18 and M13mp19 vectors, the DNA was sequenced according to the method of Sanger et al. (1980) using the sequencing kit and protocol of International Biotechnological Inc. (New Haven, CT).

Site-Directed Mutagenesis. The gene coding for hMDH was cloned into M13mp18 phage, and site-directed mutagenesis was performed according to the oligonucleotide-directed in vitro mutagenesis method developed by Sayers and Eckstein (1989) using the "oligonucleotide-directed in vitro mutagenesis system, version 2" (Amersham, U.K.).

RESULTS

Isolation of the Gene Coding for hMDH and Sequence Determination. The enzyme was purified to homogeneity as previously described (Mevarech et al., 1977), and its amino acid sequence in the N-terminal region was determined by H. Zuber and G. Frank (ETH Zurich) (unpublished data). The sequence of the first 55 amino acids is

NH_2-TKVSVVGAAGTVGAAAGTNIALRDIADEV-
VFVDIPDKEDTVGQAADTNTTIAYVL

From the amino acid sequence of residues 24–38, an oligodeoxynucleotide probe corresponding to the complementary strand of the mRNA was synthesized. In the design of the probe, we took advantage of the fact that the codon usage of halobacteria is biased toward high G+C values. The sequence of the oligonucleotide probe is

```
5' CTC CTT GTC GGG GAT GTC GAC GAA GAC GAC CTC GTC GGC GAT GTC
          C         C   C             C
```

This probe was used to screen the *H. marismortui* genomic library obtained by cloning *Mbo*I partial digests of the chromosomal DNA into the λD69 insertional cloning vector (Mizusawa & Ward, 1982). One positive clone was discovered, and the region corresponding to the hMDH coding region was sequenced. It was found that the cloned fragment contained only the region coding for the amino-terminal part of the enzyme. Since we could not find in this library another clone that contained the entire gene, a second library was constructed, this time in the λGEM11 vector (Promega Co., Madison, WI). A fragment of the hMDH gene obtained from the λD69 library was used as a probe for the screening of the new library. Independent positive clones were isolated, and the insert of one of them was analyzed further. Figure 1 shows the restriction map of the clone containing the hMDH gene (λ-hMDH). The exact location of the hMDH gene was determined by hybridizing the partial hMDH clone to restriction fragments of the newly isolated clone.

Figure 1 also shows the direction and the extent of sequencing of the region coding for hMDH. The sequence

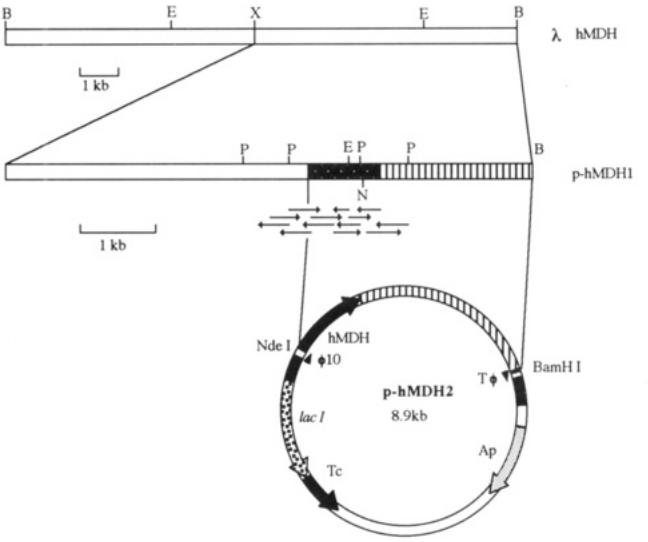

FIGURE 1: Physical map of the hMDH clones. The fragment of *H. marismortui* that was cloned in λGEM11 and contains the gene coding for hMDH (λ hMDH) was mapped by *Bam*HI (B), *Eco*RI (E), and *Xho*I (X). An *Xho*I–*Bam*HI fragment was subcloned into a pUC19 vector (p-hMDH1) and was further mapped with *Pst*I (P) and *Nde*I (N). An *Nde*I site was introduced into the gene in the locus corresponding to the N-terminus of the enzyme by PCR. The *Nde*I–*Eco*RI fragment taken from the PCR DNA and the *Eco*RI–*Bam*HI fragment taken from p-hMDH1 were ligated together with pET11a that had been digested by *Nde*I and *Bam*HI to give p-hMDH2. The arrows indicate the position, direction, and extent of the individual DNA sequencing runs.

of the gene and its flanking regions as well as the deduced amino acid sequence in hMDH is shown in Figure 2. An ATG codon at position 523 in the sequence initiates the coding region of hMDH which is extended for 303 amino acids. The molecular mass calculated following removal of the initiator methionine is 32 638 Da, significantly lower than the previously published value (Mevarech et al., 1977). The consequences of this finding will be discussed in detail elsewhere (Bonnete et al., unpublished results). Thirty-three base pairs upstream from the first codon, the sequence TTTAAC is observed. This sequence is in good agreement with the "Box A" sequence [the consensus archaebacterial promoter (Reiter et al., 1990)].

Six discrepancies appear between the N-terminal amino acid sequence as determined from the protein and the sequence as deduced from the gene. Four of these discrepancies are in amino acids 49, 50, 54, and 55 which are far from the amino terminus. One discrepancy is a missing aspartic acid (near aspartic acid-39) from the protein sequence. This position contains two adjacent aspartic acid residues which can lead to experimental difficulties in the amino acid sequence determination. The sixth discrepancy is at position 18 of the protein sequence analysis which shows threonine instead of tyrosine as determined from the sequence of the gene. The amino acid composition of the entire protein is in very good agreement with that determined for the purified protein (Mevarech et al., 1977) and confirms the large excess (12.8 mol %) of negatively charged amino acids typical for soluble halophilic enzymes.

Comparison of the Amino Acid Sequence of hMDH to Those of Other MDHs and L-*LDHS.* The alignment of the amino acid sequences of MDHs from pig mitochondria (PmM) and *E. coli* (EcM) and L-LDHs from dog fish (DfL) and *Bacillus stearothermophilus* (BsL) is shown in Figure 3 together with the sequence of the MDH from *H. marismortui* (HmM). The overall identity between the two nonhalophilic

```
CCCCGCTTCA GGTGGAGCGG TCGAGTCGTC GTCGTGTCCG GCGCACTCGC AGCCGCCTCG
TCACCTACGC CGAACTGGAC TCGGGCCTCG TCGCTGGGGT CGGCAACCTC GATCTGGTCA
CCGTCGTCGG TCTGGCCCTC CTGTTCGTCA GCATCCGACG GGTCCGCCCC GCTGGCCTGT
GTATCGCGAG CGACCGGCTC TTGGCAGGTG GGGCAGAACT CCTGACCGTC ATGCCGGAAG
ATGGGGTCGC  CACACTCGCT GCAGTGGGCG  TTCGTCATCG TCGCGCCCTT  CAACCAGGGG
TTCGCTCATT TGCTCGGTCG CGCCCGGTTG TCCTCCTCTT GCTCGAACTT CTCGCGGAGT
TTCTCGCGTT CGGCTTCCTT GTCGAAGTCG CTCATACGAG TCACAACGCC ACGGGGCTGA
AAACAGTTTA CGGGTGCGAG ATGTCGATAT CGCCTGAACG ATATCCGTCG GAATTGGGGT
TTCGACACGT TTAACGTTCG  CGCCCGAGCC  TTTTCGAATG GT

 M  T  K  V  S  V  V  G  A  A  G  T  V  G  A  A  A
ATG ACA AAG GTA AGC GTA GTC GGC GCA GCC GGA ACG GTC GGC GCA GCC GCA

 G  Y  N  I  A  L  R  D  I  A  D  E  V  V  F  V  D
GGG TAC AAC ATC GCG CTC CGT GAC ATC GCT GAC GAG GTC GTC TTC GTG GAC

 I  P  D  K  E  D  D  T  V  G  Q  A  A  D  T  N  H
ATC CCG GAC AAG GAA GAC GAC ACG GTC GGG CAG GCC GCT GAC ACG AAC CAC

 G  I  A  Y  D  S  N  T  R  V  R  Q  G  G  Y  E  D
GGC ATC GCC TAC GAT TCG AAC ACG CGC GTC CGC CAG GGC GGC TAC GAG GAT

 T  A  G  S  D  V  V  V  I  T  A  G  I  P  R  Q  P
ACT GCC GGC TCA GAC GTG GTA GTC ATC ACG GCC GGG ATT CCT CGC CAG CCC

 G  Q  T  R  I  D  L  A  G  D  N  A  P  I  M  E  D
GGC CAG ACA CGT ATC GAC CTC GCG GGC GAC AAC GCG CCC ATC ATG GAG GAC

 I  Q  S  S  L  D  E  H  N  D  D  Y  I  S  L  T  T
ATC CAG TCC TCA CTG GAC GAA CAC AAC GAC GAC TAC ATC TCG CTG ACC ACC

 S  N  P  V  D  L  L  N  R  H  L  Y  E  A  G  D  R
TCG AAC CCC GTC GAC CTG CTC AAC CGC CAC CTC TAC GAG GCC GGC GAC CGC

 S  R  E  Q  V  I  G  F  G  G  R  L  D  S  A  R  F
TCG CGC GAG CAG GTC ATC GGC TTC GGC GGC CGA CTG GAC TCC GCG CGG TTC

 R  Y  V  L  S  E  E  F  D  A  P  V  Q  N  V  E  G
CGT TAC GTC CTG AGC GAG GAG TTC GAC GCC CCG GTT CAG AAC GTC GAA GGA

 T  I  L  G  E  H  G  D  A  Q  V  P  V  F  S  K  V
ACG ATC CTC GGG GAA CAC GGC GAC GCA CAG GTT CCC GTG TTC TCG AAG GTC

 R  V  D  G  T  D  P  E  F  S  G  D  E  K  E  Q  L
CGC GTT GAC GGC ACC GAC CCC GAA TTC AGC GGG GAC GAG AAA GAG CAG CTG

 L  G  D  L  Q  E  S  A  M  D  V  I  E  R  K  G  A
CTC GGC GAC CTG CAG GAA TCG GCG ATG GAC GTC ATC GAG CGC AAG GGC GCG

 T  E  W  G  P  A  R  G  V  A  H  M  V  E  A  I  L
ACC GAG TGG GGG CCA GCC CGC GGT GTC GCA CAT ATG GTC GAA GCC ATC CTC

 H  D  T  G  E  V  L  P  A  S  V  K  L  E  G  E  F
CAC GAC ACC GGT GAA GTA CTG CCG GCT TCG GTC AAG CTA GAG GGT GAG TTC

 G  H  E  D  T  A  F  G  V  P  V  R  L  G  S  N  G
GGG CAC GAG GAC ACT GCC TTC GGT GTC CCG GTC CGT CTC GGG AGC AAC GGC

 V  E  E  I  V  E  W  D  L  D  D  Y  E  Q  D  L  M
GTC GAA GAG ATC GTC GAG TGG GAT CTT GAC GAC TAC GAG CAG GAC CTG ATG

 A  D  A  A  E  K  L  S  D  Q  Y  D  K  I  S
GCC GAC GCT GCC GAG AAG CTC TCG GAC CAG TAC GAC AAA ATC TCG taa

CCAGATATTA CTCGCGGTCG GCAGGGTCCG GGACGGCACC CGTCCCGAAT ATTATTTTGC
CTCGCCAGCG GTGTGAAGAC GCGGCTGTCA CTGCGAGGAG CGCTCTCTGA CAAACAGTGG
CCGTTACGCG GGCTCGTCCT GCTCGTCAAG CAAGTCGCTA TCGCGTCCGT CGTCATCGAT
ATCAGCACCG AGGAGTGATG CCCACTCTTC GAGGTGTGAG TCCGATGTCG AGAGTTCTCG
GTGGTTCATG  GCGTTATCCG AAGCATCCAC  GGCATATTAT  ATAATGGTTG TGCCGCATTG
ATACATAATG  CTGTATGCAG CGGTTCTACG  GGGTAATCCG  GGATTGGCGA TCACAAA
```

FIGURE 2: Nucleotide sequence of the region coding for the *H. marismortui* MDH and the derived amino acid sequence of the coded protein (GenBank Accession Number M97218).

MDHs is 60% and between EcM and HmM is only 21%. On the other hand, the sequence identity between HmM and DfL is 33%, which is comparable to the value obtained when the two L-LDHs are related. Closer inspection of the sequence comparison shows that the two regions that are involved in the binding of the dinucleotide coenzyme (Branden & Tooze, 1991) (amino acids 22–53)[4] are highly conserved in all five molecules. In addition, several amino acids that are involved in catalysis (His^{193}, Asp^{166}, and Arg^{106}) and in substrate binding (Arg^{170}) are also conserved in all molecules. Recent work (Wilks et al., 1988) demonstrated that BsL which is usually highly specific to pyruvate can be modified to assume the activity of MDH and reduce oxaloacetate, by mutating Asp^{195} to Asn, or Thr^{244} to Gly, or Gln^{100} to Arg. Interestingly, in HmM, positions 195 and 244 are occupied by Asp and Thr, respectively, as in the L-LDHs and not as in MDHs. The greatest shift in the BsL specificity was achieved when Gln^{100} was mutated to Arg. This mutation changed the specificity of the enzyme to practically that of MDH. It seems significant to note that in HmM the corresponding residue is Arg. In order to establish the structural relationships between hMDH and L-LDHs, the corresponding Arg^{100} residue of hMDH was mutated to Gln (R100Q) as described under Materials and Methods, and the mutated gene was expressed in *E. coli* as will be described in the following section.

[4] All numbering of amino acid residues in this text is referred to the numbering of dogfish L-LDH as indicated in Figure 3.

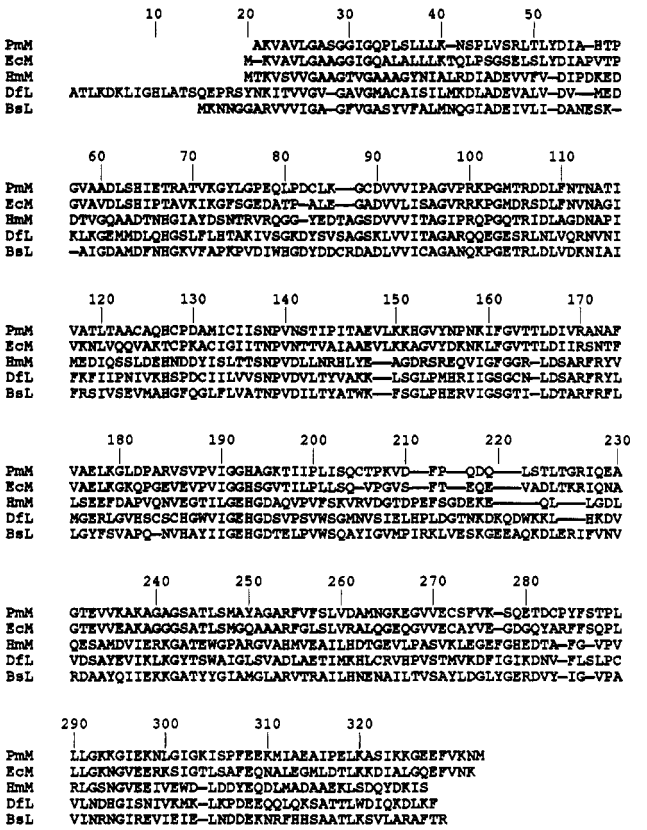

FIGURE 3: Combined alignment of MDH and L-LDH. MDH: PmM, pig mitochondria; EcM, *E. coli*; HmM, *H. marismortui*. L-LDH: DfL, dogfish; BsL, *B. stearothermophilus*. The numbering of the amino acid residues is that of dogfish L-LDH.

Expression of the Gene Coding for hMDH and the R100Q Mutant in E. coli. In order to express the gene coding for hMDH, it was cloned into the *E. coli* expression vector pET11a (Studier et al., 1990). An *Nde*I site was introduced by PCR into the gene in the position corresponding to the first ATG codon. The *Nde*I–*Eco*RI fragment synthesized by PCR and the fragment *Eco*RI–*Bam*HI taken from p-hMDH1 (see Figure 1) were ligated together with *Nde*I–*Bam*HI-digested pET11a vector to give p-hMDH2. Genes that are cloned in this expression vector are transcribed from the T7 promoter by T7 polymerase whose gene resides in the chromosome of the *E. coli* strain BL21 (DE3) and can be expressed from the *lac* promoter by induction with IPTG. The production of recombinant protein after incubation with IPTG is demonstrated in Figure 4. The mutation R100Q was introduced into the gene as described under Materials and Methods by using the oligonucleotide 5′GGCCGGGCTGCTGAGGAATCCCGG3′ which corresponds to the untranslated strand. The mutated gene was cloned into pET11a and expressed in the same way as the wild-type enzyme. Both hMDH wild-type and the R100Q mutant polypeptides produced upon induction are soluble but inactive. Activation of the enzymes can be achieved by increasing the salt concentration to 3 M NaCl or higher. The enzymes were purified to homogeneity as described below. The physical (SDS gel electrophoresis,

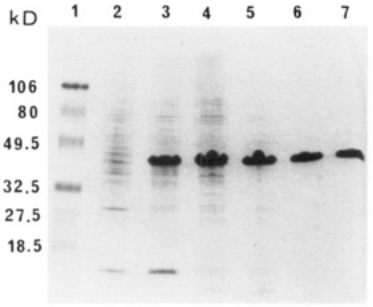

FIGURE 4: SDS–polyacrylamide gel electrophoresis of the halobacterial malate dehydrogenase (hMDH) at various stages of purification. Molar mass marker (lane 1); crude extract from cells not carrying the expression vector (lane 2); crude extract from cells carrying the expression vector and grown for 2 h with 0.4 mM IPTG (lane 3). For an explanation of lanes 4–7, see the text.

sedimentation velocity, light and neutron scattering) and biochemical (salt-dependent enzymatic activity and stability) properties of the recombinant hMDH were compared to those of the native hMDH isolated from *H. marismortui* and found to be identical (Bonnete et al., unpublished results).

Purification of Recombinant Enzymes. E. coli BL21 (DE3) carrying the plasmid p-hMDH2 was grown at 37 °C with vigorous shaking to A_{600} = 0.8. IPTG was added to a final concentration of 0.4 mM, and the culture was grown for an additional 2 h. The cells were centrifuged, and the pellet was resuspended in 0.2 M NaCl, 0.1 M Tris-HCl, pH 8, and 0.2 mM NADH, cooled on ice, and sonicated (Figure 4, lane 4). A solution of 5 M NaCl/50 mM Tris-HCl, pH 8, was added to the crude extract to a final concentration of 4 M NaCl, and the solution was centrifuged. The supernatant (Figure 4, lane 5) was dialyzed against 2 M ammonium sulfate/50 mM Tris-HCl, pH 8, then diluted with 3.5 M ammonium sulfate/50 mM Tris-HCl, pH 8, to a final ammonium sulfate concentration of 2.5 M, and centrifuged. The supernatant (Figure 4, lane 6) was loaded on a Sepharose 4B (Pharmacia) column equilibrated with 2.5 M ammonium sulfate/50 mM Tris-HCl, pH 8. The column was washed with the same buffer and then developed by a descending concentration gradient from 2.5 to 1 M ammonium sulfate. Fractions were collected and assayed for activity. Active fractions were pooled (Figure 4, lane 7). In order to remove the tightly bound NADH, the pooled fractions were dialyzed against 4 M NaCl/10 mM sodium phosphate, pH 7, loaded on a hydroxylapatite-Bio-Gel HT (Bio-Rad) column, washed with 500 mL of the same buffer, and eluted with 4 M NaCl/0.3 M sodium phosphate. The R100Q mutant was expressed, activated, and purified in similar ways. The specific activity of the purified hMDH at 4 M NaCl is 70 units/mg, and the specific activity of the purified R100Q mutant at 0.15 M NaCl and 1 mM pyruvate is 4.6 units/mg.

Effect of Salt Concentration on Enzymatic Activities of hMDH and the R100Q Mutant. The enzymatic activities of the two purified enzymes were measured as a function of oxaloacetate and pyruvate concentrations at various NaCl concentrations. A summary of the k_{cat}, K_m, and k_{cat}/K_m values at 0.15, 2, and 4 M NaCl is given in Table I. The k_{cat} of the wild-type hMDH at 2 M NaCl is 2.8-fold higher when compared to that at 4 M when oxaloacetate is used as a substrate. On the other hand, the K_m values are only slightly altered by the salt concentration. The evaluation of the above steady-state parameters at 0.15 M NaCl was impossible since at this salt concentration substrate inhibition is observed at oxaloacetate concentrations as low as 0.3 mM. Substrate inhibition is observed also at the higher salt concentrations

Table I: Steady-State Kinetic Parameters of hMDH and the R100Q Mutant[a]

enzyme	salt concn (M)	pyruvate k_{cat} (s^{-1})	pyruvate K_m (mM)	pyruvate k_{cat}/K_m (M^{-1} s^{-1})	oxaloacetate k_{cat} (s^{-1})	oxaloacetate K_m (mM)	oxaloacetate k_{cat}/K_m (M^{-1} s^{-1})
wt hMDH	0.15	*b*			*c*	*c*	
	2	*b*			197	0.9	2.0×10^5
	4	*b*			69	0.7	1.0×10^5
R100Q mutant	0.15	12.3	2.2	5.6×10^3	2	29.4	6.7
	2	2.1	2.2	1.0×10^3	0.3	21.8	1.4
	4	3.5	2.9	1.2×10^3			

[a] Enzymatic activities were measured with 0.1 mM NADH at the indicated salt concentrations in 20 mM phosphate buffer, pH 7. [b] The steady-state parameters could not be determined since only barely detectable activity was observed with hMDH amounts as high as 8.5 nmol and pyruvate concentrations as high as 10 mM. [c] The steady-state parameters for wt hMDh at 0.15 M NaCl could not be determined since at this salt concentration the oxaloacetate is inhibitory to the reaction already at a concentration of 0.3 mM.

but at much higher oxaloacetate concentrations. Pyruvate is a very bad substrate for the wild-type enzyme (specific activity of less than 5 milliunits/mg at 10 mM pyruvate and 2 M NaCl), and therefore it was impossible to measure the kinetic constants for pyruvate.

When Arg^{100} of the wild-type hMDH is mutated to Gln (the R100Q mutant), the substrate specificity is changed, and the mutated enzyme reduces pyruvate more efficiently than oxaloacetate. This mutation affects the k_{cat} values of the enzyme for the reduction of pyruvate and oxaloacetate by factors of 16 and 99, respectively. However, whereas the K_m value for pyruvate is increased about 2-fold, that for oxaloacetate increased 30-fold. The R100Q mutant is not subjected to substrate inhibition, at least not at substrate concentrations up to 30 mM, and the highest k_{cat} value is obtained at the lowest salt concentration used (0.15 M NaCl).

These results show that the structural resemblance between hMDH and the various L-LDHs demonstrated above can be manifested by the ability to modify its specificity to that of L-LDH by a single mutation.

DISCUSSION

One way to gain a better understanding of the molecular basis of the adaptation of enzymes to function at high salt concentrations is to modify the halophilic enzymes in a way that (a) will stabilize the enzyme at low salt concentrations and (b) will abolish the requirement of high salt concentration for maximum activity. The development of molecular genetic methodologies in recent years enabled the modification of the primary structure of enzymes of known structures by site-directed mutagenesis of their genes. Although several halophilic enzymes have been crystallized (Sussman et al., 1986; Harel et al., 1987), none of the structures have yet been determined. We therefore used an approximate approach based on the observation that many enzymes show a high degree of three-dimensional structural resemblance in spite of limited amino acid sequence homology (Wagner & Benkovic, 1990). One of the systems that had been studied extensively is that of related L-LDH and MDH (Birktoft et al., 1982). The three-dimensional structures of MDH and L-LDH from various sources have been determined at high resolution (Abad-Zapatero et al., 1987; Birktoft et al., 1989; Wigley et al., 1992), and it was shown that many structural motifs are shared among all the enzymes. It was thought that if the amino acid sequence of the halobacterial MDH demonstrates homologies to the nonhalophilic MDH it may enable the construction of an approximate structure of the

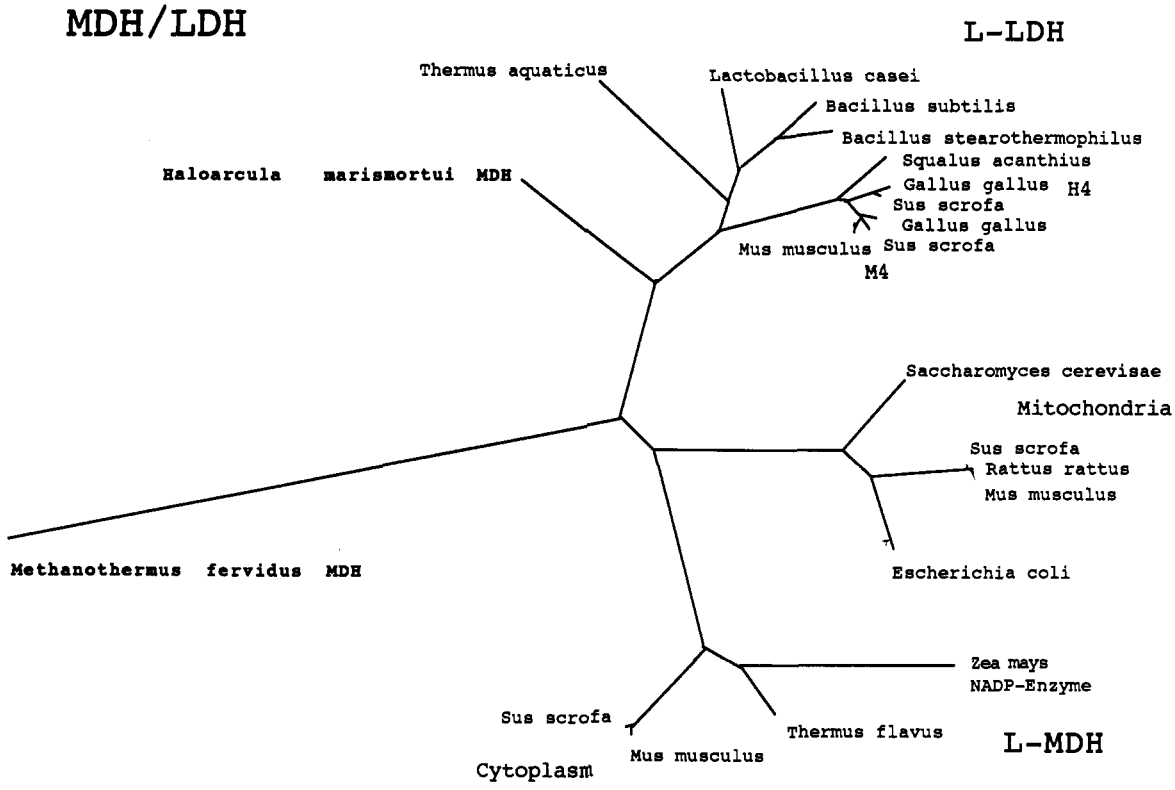

FIGURE 5: Combined phylogenetic tree of MDH and L-LDH constructed according to the method of Fitch and Margoliash (1967). The phylogenetic distances were calculated from similarity scores for protein sequences corrected for multiple mutations as proposed by Feng et al. (1985). The similarity scores were deduced from the Dayhoff mutational matrix (Dayhoff et al., 1978).

halophilic enzyme from the known information about the nonhalophilic enzyme.

As shown above, the halobacterial hMDH shares homologies with both the L-LDHs and the MDHs, being more homologous to the former. The structural resemblance between hMDH and L-LDH could be further manifested by the change of the specificity of hMDH from oxaloacetate to pyruvate in the R100Q mutant. Although it was previously shown that the specificity of *B. stearothermophilus* L-LDH can be modified to assume MDH activity by the inverse mutation (Gln to Arg) (Wilks et al., 1988), the modification of MDH to assume L-LDH activity was never reported before.

The mutagenesis of R100Q of hMDH converted the specificity of the enzyme for oxaloacetate to that of pyruvate $\{(k_{cat}/K_m[\text{pyruvate}])/(k_{cat}/K_m[\text{oxaloacetate}]) = 820\}$. This mutation caused also a change in the effect of salt concentration on the enzymatic activity. The highest activity of R100Q hMDH is at the lowest salt concentration and is reduced as the salt concentration increases. The wild-type hMDH, on the other hand, is subjected to substrate inhibition that is more pronounced at the low salt concentrations, and as a result, the enzymatic activity has a maximum between 1 and 2 M NaCl. The molecular basis for these changes is unclear at the moment since unlike the conversion of *B. stearothermophilus* L-LDH to MDH described by Wilks et al. (1988) by the inverse mutation of Q100R, in which the K_{cat} and K_m of the former L-LDH activity and the resulted MDH activity remained the same, in our case the mutation was associated with considerable changes in these parameters and therefore cannot be attributed to changes in the affinity of the substrate to the enzyme alone. Further characterization of the catalytic properties of the wild-type and mutated hMDH is underway.

The results presented above demonstrate that a halobacterial gene can be expressed with high efficiency in *E. coli*. Previous attempts to express the gene coding for the membrane protein bacteriorhodopsin of *Halobacterium halobium* in *E. coli* encountered great difficulties (Braiman et al., 1987; Dunn et al., 1987; Karnik et al., 1987; Nassal et al., 1987). These difficulties may have been due to the special nature of the membrane protein rather than to the halobacterial origin of the gene. It was shown previously (Mevarech & Neumann, 1977) that hMDH is stable only at salt concentrations above 2 M. At lower concentrations, the enzyme is dissociated into subunits, and subsequently, the polypeptide unfolds. Whereas the dissociation is readily reversible, the unfolding process seemed to be irreversible. Refolding of the unfolded chain could be achieved only after complete denaturation of the polypeptide chain by guanidine hydrochloride (Hecht & Jaenicke, 1989). It is therefore surprising that the hMDH produced in an inactive form in *E. coli*, presumably at low salt concentration, could be readily activated by raising the salt concentration above 2 M. It seems, therefore, that the in vivo folding of the nascent polypeptide chain has a different folding pathway than the in vitro folding of the already complete chain or that additional factors are involved in the folding. The kinetics of the reactivation of the *E. coli* produced enzyme are more complex (data not shown) than the second-order reactivation observed previously when dissociated subunits are diluted into concentrated salt solutions (Mevarech & Neumann, 1977). Detailed analysis of the state of the primary product of the *E. coli* produced hMDH is currently underway.

The structural resemblance of hMDH to the dog fish and *B. stearothermophilus* L-LDHs for which high-resolution crystal structures were determined (Abad-Zapatero et al., 1987; Wigley et al., 1992) and the ability to express the halobacterial gene in *E. coli* to obtain large quantities of protein that can be easily activated and purified will facilitate detailed studies of those structural features of the enzyme that are relevant to its functional adaptation to high salinity.

The sequence information of the halobacterial malate dehydrogenase has an additional significant aspect. It can shed light on the evolution of the MDH–L-LDH gene family. The proposal of Woese and his colleagues [summarized in Woese (1987)] that all living organisms can be classified into three primary kingdoms is now widely accepted. Assignment of organisms into one kingdom or another has usually been based on comparison of sequences of RNA or protein molecules for construction of phylogenetic trees. However, the exact evolutionary relationships among the primary kingdoms cannot be determined because the roots of the phylogenetic trees cannot be determined uniquely. Recently, Iwabe et al. (1989) suggested the establishment of composite phylogenetic trees from comparisons of sequences of pairs of duplicated genes that exist in all organisms. On the basis of the composite trees for the translation elongation factors EF-Tu and EF-G and for the F_1-α and F_1-β components of ATPase, they confirmed that archaebacteria are phylogenetically more closely related to eukaryotes than to eubacteria. A similar analysis using the amino acid sequences of the enzymes L-LDH and MDH, which are also believed to be the result of gene duplication, was incomplete due to the fact that no sequence information for these two enzymes from archaebacteria was available at that time. Meanwhile, the sequence of the gene coding for MDH of the archaebacterium *Methanothermus fervidus* was determined (Honka et al., 1990) and found to diverge significantly from both the L-LDH and MDH genes. This divergence was also observed when the sequence of the enzyme glyceraldehyde-3-phosphate dehydrogenase of the same organism was compared to the sequences of equivalent enzymes in both eukaryotes and eubacteria (Fabry & Hensel, 1988). An intriguing question that arises is whether the *M. fervidus* MDH sequence is an exception or whether other archaebacterial MDH sequences are as divergent. A composite phylogenetic tree constructed for all MDH and L-LDH sequences available is shown in Figure 5. It shows clearly that all eubacterial and eukaryotic L-LDHs are grouped in one branch and all eubacterial and eukaryotic MDHs are grouped in another branch. The two archaebacterial MDH lineages emerge from the connection between the two groups. However, whereas the ramification of the *M. fervidus* sequence from the connecting line is located near the MDHs, the *H. marismortui* sequence emerges at the basis of the L-LDH branch. Discussion of evolutionary implications (Zillig et al., 1991) is outside the scope of this contribution.

ACKNOWLEDGMENT

We are grateful to Dr. H. Zuber and Dr. G. Frank for the determination of the N-terminal amino acid sequence of hMDH, to Ms. Yehudith Navon for excellent technical assistance, to Dr. R. F. Doolittle and Dr. P. Palm for the phylogenetic tree construction, and to Dr. W. Zillig for fruitful discussions.

REFERENCES

Abad-Zapatero, C., Griffith, J. P., Sussman, J. L., & Rossmann, M. G. (1987) *J. Mol. Biol. 198*, 445–467.

Birktoft, J. J., Fernley, R. T., Bradshaw, R. A., & Banaszak, L. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6166–6170.

Birktoft, J. J., Rhodes, G., Banaszak, L. J. (1989) *Biochemistry 28*, 6065–6081.

Bradford, M. (1976) *Anal. Biochem. 72*, 248–254.

Braiman, M. S., Stern, L. J., Chao, B. H., & Khorana, H. G. (1987) *J. Biol. Chem. 262*, 9271–9276.

Branden, C., & Tooze, J. (1991) *Introduction to Protein Structure*, pp 148–151, Gerland Publishing Inc., New York and London.

Cristian, J. H. B., & Waltho, J. A. (1962) *Biochim. Biophys. Acta 65*, 506–508.

Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978) *Atlas of protein sequence and structure* (Dayhoff, M. O., Ed.) Vol. V (Suppl. 3), pp 345–352, National Biomedical Research Foundation, Washington, D.C.

Dunn, R. J., Hackett, N. R., McCoy, J. M., Chao, B. H., Kimura, K., & Khorana, H. G. J. (1987) *J. Biol. Chem. 262*, 9246–9254.

Eisenberg, H., Mevarech, M., & Zaccai, G. (1992) *Adv. Protein Chem. 43*, 1–61.

Fabry, F., & Hensel, R. (1988) *Gene 64*, 189–197.

Feng, D. F., Johnson, M. S., & Doolittle, R. F. (1985) *J. Mol. Evol. 21*, 112–125.

Fitch, W. M., & Margoliash, E. (1967) *Science 155*, 279–284.

Harel, M., Shoham, M., Frolow, F., Eisenberg, H., Mevarech, M., Yonath, A., & Sussman, J. L. (1987) *J. Mol. Biol. 200*, 609–610.

Hecht, K., & Jaenicke, R. (1989) *Biochemistry 28*, 4979–4985.

Honka, E., Fabry, S., Niermann, T., Palm, P., & Hensel, R. (1990) *Eur. J. Biochem. 188*, 623–632.

Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S., & Miyata, T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9355–9359.

Karnik, S. S., Nassal, M., Doi, T., Jay, E., Sagramella, V., & Khorana, H. G. J. (1987) *J. Biol. Chem. 262*, 9255–9263.

Lanyi, J. (1974) *Bacteriol. Rev. 38*, 272–290.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mevarech, M., & Neumann, E. (1977) *Biochemistry 16*, 3786–2792.

Mevarech, M., Eisenberg, H., & Neumann, E. (1977) *Biochemistry 16*, 3781–3785.

Mizusawa, S., & Ward, D. F. (1982) *Gene 20*, 317–322.

Nassal, M., Mogi, T., Karnik, S. S., & Khorana, H. G. (1987) *J. Biol. Chem. 262*, 9264–9270.

Pundak, S., & Eisenberg, H. (1981) *Eur. J. Biochem. 118*, 463–470.

Pundak, S., Aloni, H., & Eisenberg, H. (1981) *Eur. J. Biochem. 118*, 471–477.

Reiter, W.-D., Huedepohl, U., & Zillig, W. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9509–9513.

Rosenshine, I., Zusmann, T., Werczberger, R., & Mevarech, M. (1987) *Mol. Gen. Genet. 208*, 518–522.

Sayers, J. R., & Eckstein, F. (1989) in *Protein Function (a practical approach)* (Creighton, T. E., Ed.) pp 279–295, IRL Press, Oxford.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–89.

Sussman, J. L., Brown, D. H., & Shoham, M. (1986) *Frontiers of Iron–Sulfur Protein Research* (Matsubara, H., Katsube, Y., & Wada, K., Eds.) pp 69–82, Japan Scientific Societies, Tokyo, and Springer-Verlag, Berlin.

Wagner, C. R., & Benkovic, S. J. (1990) *Trends Biotechnol. 8*, 263–270.

Wigley, D. B., Gamblin, S. J., Turkenburg, J. P., Dodson, E. J., Piontek, K., Muirhead, H., & Holbrook, J. J. (1992) *J. Mol. Biol. 223*, 317–335.

Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Bartow, D. A., Atkinson, T., Clarke, A. R., & Holbrook, J. J. (1988) *Science 242*, 1541–1544.

Woese, C. R. (1987) *Microbiol. Rev. 51*, 221–271.

Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D., & Somero, G. N. (1982) *Science 217*, 1214.

Yanisch-Perron, C., Viera, J., & Messing, J. (1985) *Gene 33*, 103–119.

Zaccai, G., Cendrin, F., Haik, Y., Borochov, N., & Eisenberg, H. (1989) *J. Mol. Biol. 208*, 491–500.

Zillig, W., Palm, P., Klenk, H.-P., Puehler, G., Gropp, F., & Schleper, C. (1991) *General and Applied Aspects of Halophilic Microorganisms* (Rodriguez-Valera, F., Ed.) pp 321–332, NATO ASI Series, Plenum Press, New York.