Nucleotide sequence of the *Escherichia coli* *mutH* gene

Robert H.Grafstrom and Ronald H.Hoess⁺

Department of Microbiology, Jefferson Medical College, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107 and ⁺E.I.DuPont de Nemours and Company, Central Research and Development Department, Experimental Station E328/145, Wilmington, DE 19898, USA

Received February 17, 1987; Accepted March 3, 1987

Accession no. Y00113

ABSTRACT

The complete nucleotide sequence of *mutH* gene from *E. coli* has been determined. Based on the deduced amino acid sequence, the MutH protein has a molecular weight of 25.4 kdaltons in agreement with the previous estimates based on SDS-polyacrylamide gel electrophoresis of the purified protein. Deletion analysis of the DNA sequences upstream of *mutH* has identified the promoter region for this gene. Two independently isolated temperature sensitive alleles of the *mutH* gene have also been sequenced. One mutation results in an amino acid change at position 27 (thr to leu) while the other occurs at position 156 (asp to asn).

INTRODUCTION

The spontaneous mutation rate of *Escherichia coli* is controlled by a number of genes whose products are responsible for the detection and removal of errors in the DNA. A major pathway for the removal of mismatched base pairs in the DNA is the methyl-directed mismatch repair system (for review see ref. 1). This repair pathway utilizes DNA methylation to distinguish between the parental and newly synthesized DNA strands to direct the correction of mismatched base pairs arising from errors during DNA replication. *In vivo* (2,3,4,5,6,7) and *in vitro* (8,9) experiments have implicated the products of the *dam*, *mutH*, *mutL*, *mutS*, and *uvrD* genes in the methyl-directed mismatch repair system. Functions for three of these genes have been assigned: the *dam* gene encodes the DNA adenine methylase that methylates GATC sequences (10,11,12); the *uvrD* gene encodes DNA helicase II necessary for unwinding of DNA during repair (13,14,15); and the *mutS* protein recognizes and binds to DNA containing a mismatched base pair

(9). A biochemical role for the MutL and MutH proteins has yet to be determined.

To understand the underlying biochemical mechanism of these proteins requires a detailed knowledge of their structure. We report here the complete nucleotide sequence of the mutH gene. Deletion analysis has allowed us to identify the nucleotide sequences that comprise the promoter. In addition, two temperature sensitive mutants of the mutH gene have also been isolated and their sequence as well as their biological properties have been examined.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

The mutH gene of E. coli has been cloned on a 1.5 kb HindIII fragment (16,17,18). The gene was further subcloned on a 0.9 kb XhoI-HindIII fragment following digestion with Bal31 exonuclease and the addition of a XhoI linker (19). Since mutH co-transduces with thyA (35), strain RG1051 (dam-3, mutH-34) was prepared by P1 transduction of the mutH-34 allele into GM44 (dam-3, thyA) (31). Transductants were selected for thy⁺ and screened for 2AP^r. The presence of the dam-3 allele was confirmed by isolating the DNA from this strain and demonstrating its resistance to cleavage by DpnI and sensitivity to cleavage by MboI. DpnI cleaves DNA only at GATC sites that have been methylated by the Dam methylase while MboI cleaves unmethylated DNA at GATC sites (11,12). The presence of mutH-34 was confirmed by the ability of strain RG1051 to grow in the presence of 2-aminopurine (2AP), since dam⁻mutH⁻ double mutants are resistant to this base analog (20).

Hydroxylamine Mutagenesis

Hydroxylamine mutagenesis was performed according to Kadonago and Knowles (21). Briefly, 10 ug of pRH71-17 (19) was precipitated with ethanol and resuspended in 50 ul of 1M sodium acetate, pH 5.0. The DNA was added to a 150 ul solution of 2M hydroxylamine hydrochloride (Sigma), 1M sodium acetate (pH 5.0). The sample was incubated at 55^o for 60 min. The pH was adjusted to 7.6 by the addition of 1M Tris-base (225 ul)

and Na₂EDTA was added to a final concentration of 10 mM. The DNA was precipitated with ethanol and the pellet was washed twice with 95% ethanol, dried and resuspended in 50 ul of 10 mM Tris-HCl (pH 8.0), 2mM Na₂EDTA. This DNA was then used for transformation.

Transformation

Plasmid DNA preparation and transformation of strain RG1051 (dam⁻mutH⁻) by various plasmids were performed by standard techniques (22). Transformants were selected by ampicillin resistance and screened for sensitivity to 2AP.

Killing by 2-aminopurine

Bacteria to be tested for killing by 2AP were grown to mid-log phase in L-Broth plus ampicillin (100 ug/ml). The bacteria were diluted and plated on L-plates containing ampicillin (100 ug/ml) and 2AP (nitrate salt, Sigma) at the concentrations listed in the Figures.

DNA Sequencing

The initial DNA sequence of the wild type mutH gene was performed by the chemical method of Maxam and Gilbert (23). The two temperature sensitive mutants, mutH^{ts7} and mutH^{ts28} were cloned into the M13-plasmid hybrid vectors pZ150 and pZ152 (24). Single stranded DNA derived from the hybrid vectors were then sequenced using the dideoxy chain termination method (25). A set of six sequencing primers were synthesized, three hybridizing to each strand of the mutH gene (Fig. 1).

RESULTS

Sensitivity of dam⁻mutH⁻ strains to 2-aminopurine

The mutH gene of E. coli is contained on a 1.5 kb HindIII fragment that was cloned into pBR322 (16). Based on in vitro transcription/translation analysis the gene was shown to be translated upstream from the BglII site and proceed past the SmaI site (Fig. 1) (19). The gene was further subcloned from the original 1.5 kb HindIII fragment, during which a series of Bal31 deletions of sequences upstream of the mutH gene were generated (19). Of the three isolates that were characterized further, two of them, pRH71-17 and pRH71-6, were still capable of restoring a wild-type spontaneous mutation frequency to a

Figure 1. Restriction map of the cloned *mutH* gene from *E. coli* and the DNA sequencing strategy. The restriction map of subclones of pRH71 containing the entire *mutH* gene on a 0.9 kb HindIII/XhoI fragment with the relevant restriction enzyme sites is shown. The end points of the three Bal31 deletion mutants are shown as 71-17, 71-6, and 71-155. The orientation of transcription is also indicated (⇒). DNA sequences determined by Maxam-Gilbert technique are shown as —→, while that determined by the dideoxy procedure are shown by ▷ .

mutH⁻ strain (19). The third isolate, pRH71-155, was unable to restore the wild-type spontaneous mutation rate in a *mutH*⁻ strain (19). When the *mutH* gene of pRH71-155 was placed under the control of the λ_{PL} promoter, the wild-type spontaneous mutation frequency was restored (19). This suggests that the deletion extends into the promoter region but not into the coding sequences for *mutH*.

Each of the above plasmids were tested for their ability to complement a *dam*⁻*mutH*⁻ strain in regards to the sensitivity of these strains to the base analog, 2AP. *dam*⁻*mutH*⁺ strains are extremely sensitive to killing by this analog (26) while *dam*⁻*mutH*⁻ strains are resistant (20). As shown in Figure 2, the *dam*⁻*mutH*⁻ strain containing pBR322 is completely resistant to killing by 2AP, while only 0.1% of a *dam*⁻*mutH*⁺ strain survive treatment with 2AP at 20 ug/ml. Plasmid pRH71-17 completely restores the *mutH*⁺ phenotype to the *dam*⁻*mutH*⁻ strain. pRH71-155, which was unable to restore *mutH*⁻ strains to the wild-type spontaneous mutation frequency, is also *mutH*⁻ with respect to killing by 2AP. Interestingly, pRH71-6 which can restore *mutH*⁻ strains to the wild-type mutation frequency (19) was only partially sensitive

Figure 2. Effect of 2AP on the survival of $dam^{-}mutH^{-}$ strain. Bacteria were grown to mid-log phase and plated on L-plates containing ampicillin (100 ug/ml) for plasmid containing strains, and increasing concentrations of 2AP. $dam^{-}mutH^{+}$ (▲); $dam^{-}mutH^{-}$ containing either pBR322 (●), pRH71-17 (△), pRH71-6 (X), pRH71-155 (○).

to killing by 2AP, suggesting that there is enough Muth protein made from this plasmid to repair spontaneous DNA damage but insufficient protein synthesized to repair the large amount of DNA damage resulting from exposure to 2AP.

Identification of Two Temperature Sensitive Mutants of Muth

In order to obtain temperature sensitive mutants of muth, in vitro mutagenesis of the purified pRH71-17 DNA was performed using either UV irradiation or hydroxylamine. Mutants were selected for resistance to ampicillin and 2AP at 42° and screened for sensitivity to 2AP at 32°. No temperature

Figure 3. Effect of 2AP on the survival of *dam*⁻*mutH*⁻ strain containing temperature sensitive mutants of the cloned *mutH* gene. The experimental procedure was the same as that described in Figure 2 except that the bacteria were plated at either 30° or 42° in the presence of 2AP. The *dam*⁻*mutH*⁻ strain contained either pBR322 (●), pRH71-17 (○), pMuth^{ts7} (□) or pMuth^{ts28} (Δ).

sensitive mutants were found in 120 mutants obtained with UV irradiation while two temperature sensitive mutants were found in 36 mutants obtained with hydroxylamine. The mutants were tested for temperature sensitivity by plating at 30° or 42° on plates containing increasing concentrations of 2AP. As shown in Figure 3 each of these mutants differed in their sensitivity to base analog at 30°. At 42° pMuth^{ts7} was nearly completely inactivated while pMuth^{ts28} still had some residual *mutH*⁺ activity.

Nucleotide Sequence and Coding Region

The nucleotide sequence of the complete *mutH* 0.8 kb

GCAATATAAA TCAGAGAATT GAACAACGCA TGTGGAAAAA GTTACACTGC
 Δ71-6

GAATATTCGG CACATAATTG CTGTTTGT TTAAATCAAG GTATCATGAC
 Δ153

1
 ser gln pro arg pro leu leu ser pro pro glu thr glu
 ATG TCC CAA CCT CGC CCA CTG CTC TCT CCT CCC GAA ACT GAA

20
 glu gln leu leu ala gln ala gln gln leu ser gly tyr thr
 GAA CAG TTG TTA GCG CAA GCA CAG CAA CTT TCT GGT TAT ACA
 ile

30
 leu gly glu leu ala ala leu val gly leu val thr pro glu
 TTG GGA GAA CTG GCG GCA CTT GTC GGG CTG GTT ACG CCA GAG

40
 asn leu lys arg asp lys gly trp ile gly val leu leu glu
 AAT TTA AAA CGC GAT AAA GGC TGG ATT GGC GTG TTA CTG GAG

50
 ile trp leu gly ala ser ala gly ser lys pro glu gln asp
 ATC TGG CTA GGT GCC AGC GCA GGG AGT AAA CCT GAG CAA GAT

60
 phe ala ala leu gly val glu leu lys thr ile pro val asp
 TTT GCT GCT CTG GGC GTG GAA CTT AAA ACT ATC CCT CTG GAT

70
 ser leu gly arg pro leu glu thr thr phe val cys val ala
 AGT CTT GGT CGT CCG CTG GAA ACA ACA TTC GTT TGT GTT GCC

80
 pro leu thr gly asn ser gly val thr trp glu thr ser his
 CCG TTA ACG GGC AAT AGC GGG GTG ACC TGG GAA ACC AGC CAC

90
 val arg his lys leu lys arg val leu trp ile pro val glu
 GTG CGC CAC AAA CTC AAA CGC GTA CTG TGG ATA CCG GTT GAA

100
 gly glu arg arg leu pro leu ala gln ala arg val gly ser
 GGC GAG CGC AGA CTC CCG CTG GCG CAG GCT CGC GTT GGA TCA

110
 pro leu leu trp ser pro asn glu glu glu asp arg gln leu
 CCG TTG CTG TGG AGC CCG AAT GAA GAG GAA GAC CGC CAG CTA

120
 arg glu asp trp glu glu leu met asp met ile val leu gly
 CGC GAA GAC TGG GAA GAA TTA ATG GAT ATG ATT GTT CTC GGT

130
 gln val glu arg ile thr ala arg his gly glu tyr leu gln
 CAG GTT GAG CGG ATT ACC GCT CGT CAC GGG GAG TAT TTA CAG

140
 ile arg pro lys ala ala asn ala lys ala leu thr glu ala
 ATA CGA CCG AAA GCA GCG AAT GCG AAA GCG CTT ACC GAA GCC

150
 ile gly ala arg gly glu arg ile leu thr leu pro arg gly
 ATT GGT GCC CGG GGC GAA CCG ATT CTG ACG CTG CCA CCG GGC

160
 phe tyr leu lys lys asn phe thr ser ala leu leu ala arg
 TTT TAT TTG AAG AAG AAT TTC ACC AGT GCA CTA CTG GCC CGT

170
 his phe leu ile gln
 CAT TTT CTG ATC CAG TAG

180
 210
 220
 228

Figure 4. Nucleotide sequence of the mutH gene. The nucleotide sequence was determined as described in the text. The amino acid sequence deduced from the DNA sequence is also shown. Homologies with the -35 and -10 promoter sequences and the putative Shine-Delgarno sequences are underlined. The end points of the various Bal31 deletions and the point mutations of the two temperature sensitive mutants are also shown.

XhoI-HindIII fragment from pRH71-17 is shown in Figure 4. There is only one open reading frame of 229 amino acids that will accommodate a protein of 25 kdaltons. The size of the protein based on the DNA sequence has been calculated to be 25.4 kdaltons agreeing with the size of radioactively labeled protein from muth containing maxi-cells, in vitro transcription-translation of muth, and overproduction of Muth using the λ_{PL} promoter (16,19). The deduced sequence of the first 24 amino acids has been confirmed by sequencing the amino terminus of the purified Muth protein. In addition, the protein sequencing data also reveals that the first methionine is cleaved from the mature protein.

The DNA sequence of both hydroxylamine induced temperature sensitive mutations was determined and is shown in Figure 4. Each results from a single transition mutation, which is expected since hydroxylamine was the mutagen. pMuth^{ts7} results in the change of threonine to isoleucine at amino acid number 27 and pMuth^{ts28} results in a change of aspartate asparagine at position 156. The entire gene was sequenced for each mutant. Although it is impossible to determine from this data if either of these mutations occur at the active site it is interesting to note that they occur in very different areas of the gene and that pMuth^{ts28} occurs in a very acidic region of the protein.

The Muth protein contains 31 basic amino acids, 4 of which are histidine, and 28 acidic amino acids. The carboxy half of the molecule contains 77% of the basic amino acids. Although the 28 acidic amino acids are evenly distributed over the entire protein, there is a region between #147 to #171 that contains a cluster of 10 acidic amino acids. This is the same region where one of the temperature sensitive mutants (ts28) occurs.

The muth gene contains 19 rarely used codons (27) that are not randomly distributed throughout the genome. Ten of these rare codons are clustered in the first third of the gene. The proportion of rarely used codons to the total codons in the reading frame is 8.3% for muth versus a 4.24% for other non-regulatory proteins of E. coli (27). Genes with a high

percentage of rarely used codons, such as dnaG and the regulatory proteins of the lacI, araC, and trpR usually encode products that are present in only a few copies per cell because of the limiting availability of tRNA's for these rare codons (27). Thus far, we have not determined the amount of Muth protein found in wild-type E. coli or in any of our plasmid strains.

The mutH Regulatory Region

The DNA sequence upstream of the ATG start codon for mutH has been examined for sequences that could act as promoters for the gene (28). Figure 4 shows one potential promoter with its -10 and -35 regions underlined. The underlined -10 region has 5 out of 7 matches with the consensus sequence heptamer TATAATG including the invariant T in the sixth position and the strongly conserved initial TA. In addition there are 4 out of 6 matches with the consensus -35 region. The plasmid pRH71-17 contains 100 bp prior to the ATG start codon and, thus, possesses all of the regulatory regions listed above. In pRH71-6, which was shown to have a wild-type spontaneous mutation frequency (19) but an intermediate mutH phenotype when exposed to 2AP, part of the proposed -35 region has been deleted. This could account for the presumed ability of this plasmid to make sufficient amount of enzyme to repair spontaneous damage but insufficient amount of enzyme to repair increased damage caused by the exposure to 2AP. In pRH71-155, the complete upstream sequence containing the proposed -10 and -35 regions was deleted. Hence, the completely negative mutH phenotype of this plasmid in both the spontaneous mutation frequency as well as 2AP sensitivity is observed. The substitution of the λ_{PL} promoter completely restores the wild-type spontaneous mutation frequency to this plasmid indicating that the structural gene for mutH was not affected.

DISCUSSION

The complete nucleotide sequence of the mutH gene from E. coli has been determined. Based on deletion analysis, the promoter and coding regions for this gene have been identified. The size of the Muth protein predicted from the

DNA sequence agrees with the size determined by SDS-polyacrylamide gel electrophoresis of the purified protein (19). Two different temperature sensitive mutants of mutH have been isolated and have been shown to result in amino acid changes in two distinct regions of the protein.

The deletion analysis in combination with the sequence data suggests that the promoter for mutH is immediately upstream of the coding sequences. At this time it is not known whether this gene is under some additional regulatory control aside from the intrinsic strength of its own promoter. It is interesting to note that contained within the promoter region is the sequence (₅,TgCTGTtTgTtttTtaAtca₃,) which shares considerable homology with half of the consensus sequence for the lexA binding site (₅,taCTGTaTat-cat-CAGga₃,) (29). The CAG present in the lexA binding site which forms part of a dyad symmetry, is absent in the mutH sequence making it an unlikely target for LexA binding. Furthermore, the *E. coli* mutH has not been identified as a damage unducible gene (din) (32). Nevertheless, it is interesting to speculate whether this sequence may represent what once was a LexA binding site, and through evolution has been altered so that mutH is no longer under lexA control.

Others have reported the cloning of the mutH gene and the purification of the protein (16,17,18) but, to date, the biochemical function of MutH in methyl directed mismatch repair has not been determined. Recent genetic evidence suggests that MutH may be involved in the recognition of unmethylated GATC sites (33). The expression of the bacteriophage Mu mom gene had previously been shown to be regulated by Dam methylation: in dam⁻ cells the mom gene is turned off (34). However, expression of the mom gene is restored if the cells are now made mutH⁻ (this is not the case if either mutL⁻ or MutS⁻ are introduced into these cells) (33). A simple model to explain these results is that the mutH protein recognizes the unmethylated GATC sequences in the mom promoter, binds to them thereby acting as a repressor. Interestingly, not all mutH⁻ alleles can restore full expression of mom suggesting that involvement in mismatch repair (mutator phenotype) and mom

expression may represent two separate functions of the protein (33). Clearly, further analysis of the temperature sensitive mutH mutants, as well as isolation of other mutants in the gene should prove useful in testing the validity of such models and a basis for further biochemical characterization.

ACKNOWLEDGMENTS

The authors wish to thank A. Amsterdam and K. Zachariasewycz (Thomas Jefferson University) for their expert technical assistance, and Dr. R. Kutny and B. Krieger (E.I. DuPont deNemours and Co.) for determining the amino acid composition of purified MutH, and E. Doran (E.I. DuPont deNemours and Co.) for oligonucleotide synthesis. They also thank Drs. T. Block, R. Zagursky and R. LaRossa for critically reviewing the manuscript, and K. Givens for the preparation of the manuscript. This work was supported by USPHS Grant No. RR-05414 and NSF Grant No. DMB-8517662.

REFERENCES

1. Claverys, J.-P. and Lacks, S.A. (1986) Microbiol. Rev. 50,133-165.
2. Rydberg, B. (1978) Mutat. Res. 52,11-24.
3. Wagner, R. and Meselson, M. (1976) Proc. Natl. Acad. Sci. U.S.A. 73,4135-4139.
4. Pukkila, P., Peterson, J., Herman, G., Modrich, P. and Meselson, M. (1983) Genetics 104,571-582.
5. Kramer, B., Kramer, W. and Fritz, H.-J. (1984) Cell 38,879-887.
6. Dohet, C., Wagner, R., and Radman, M. (1985) Proc. Natl. Acad. Sci. U.S.A. 82,503-505.
7. Fishel, R.A., Siegel, E.C. and Kolodner, R. (1986) J. Mol. Biol. 188,147-157.
8. Lu, A.-L, Clark, S. and Modrich, P. (1983) Proc. Natl. Acad. Sci. U.S.A. 80,4639-4643.
9. Su, S.-S. and Modrich, P. (1986) Proc. Natl. Acad. Sci. U.S.A. 83,5057-5061.
10. Marinus, M.G. and Morris, N.R. (1973) J. Bacteriol. 114,1143-1150.
11. Lacks, S. and Greenberg, B. (1977) J. Mol. Biol. 114,153-168.
12. Geier, G.E. and Modrich, P. (1979) J. Biol. Chem. 254,1408-1413.
13. Arthur, H.M., Bramhill, D., Eastlake, P.B. and Emmerson, P.T. (1982) Gene 19,285-295.
14. Maples, V.F. and Kushner, S.R. (1982) Proc. Natl. Acad. Sci. U.S.A. 79,5616-5620.

15. Kumura, K. and Sekiguchi, M. (1984) *J. Biol. Chem.* 259,1560-1565.
16. Grafstrom, R.H. and Hoess, R.H. (1983) *Gene* 22,245-253.
17. Lu, A.-L., Welsh, K., Clark S., Su,S.-S. and Modrich, P. (1984) Cold Spring Harbor Symp. Quant. Biol. 49,589-596.
18. Pang, P.P., Tsen, S.-D., Lundberg, A.S. and Walker, G.C. (1984) Cold Spring Harbor Symp. Quant. Biol. 49,597-602.
19. Grafstrom, R.H., Bear, J. and Hoess, R.H. (1983) In: Cellular Responses to DNA Damage (E.C. Friedberg and B.A. Bridges eds.) Alan R. Liss Inc., New York, pp 299-307.
20. Glickman, B.W. and Radman, M. (1980) *Proc. Natl. Acad. Sci. U.S.A.* 77,1063-1067.
21. Kadonaga, J.T. and Knowles, J.R. (1985) *Nucleic Acids Res.* 13,1733-1745.
22. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
23. Maxam, A.M. and Gilbert W. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74,560-564.
24. Zagursky, R.J. and Berman, M.L. (1984) *Gene* 27, 183-191.
25. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74,5463-5467.
26. Glickman, B., van den Elsen, P. and Radman, M. (1978) *Molec. Gen. Genet.* 163,307-312.
27. Konigsberg, W. and Godson, G.N. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80,687-691.
28. Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genet.* 13,319-353.
29. Little, J.W. and Mount D.W. (1982) *Cell* 29,11-22.
30. Wertman, K.F. and Mount, D.W. (1985) *J. Bacteriol.* 163,376-384.
31. Marinus, M.G. and Morris, N.R. (1974) *J. Mol. Biol.* 85,309-322.
32. Pang, P.P., Lundberg, A.S. and Walker, G.C. (1985) *J. Bacteriol.* 163,1007-1015.
33. Seiler, A., Blocker, H., Frank, R. and Kahmann, R. (1986) *EMBO J.* 5,2719-2728.
34. Hattman, S. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79,5518-5521.
35. Hoess, R.H. and Herman, R.K. (1975) *J. Bacteriol.* 122,474-484.