Volume 15 Number 20 1987 Nucleic Acids Research

Mutation spectrum in *Escherichia coli* DNA mismatch repair deficient (*mutH*) strain

Caroline Rewinski and M.G.Marinus

Department of Pharmacology, University of Massachusetts Medical School, Worcester, MA 01655-2937, USA

Received July 31, 1987; Revised and Accepted September 14, 1987

ABSTRACT

The Dam-directed post-replicatve mismatch repair system of Escherichia coli removes base pair mismatches from DNA. The products of the mutH, mutL and mutS genes,among others,are required for efficient mismatch repair. Absence of any of these gene products leads to persistence of mismatches in DNA with a resultant increase in spontaneous mutation rate. To determine the specificity of the mismatch repair system in vivo we have isolated and characterized 47 independent mutations from a mutH strain in the plasmid borne mnt repressor gene. The major class of mutations comprises AT to GC transitions that occur within six base pairs of the only two 5'-GATC-3' sequences in the mnt gene. In the wild type control strain, insertion of the IS1 element was the major spontaneous mutational event. A prediction of the Dam-directed mismatch repair model, that the mutation spectra of dam and mutH strains should be the same, was confirmed.

INTRODUCTION

Spontaneous base substitution mutations in Escherichia coli probably arise mainly as the consequence of errors introduced into DNA during replication. At least two mechanisms are known which remove base pair mismatches from DNA: proofreading and post-replicational Dam-directed mismatch repair. Proofreading is intimately associated with replication and errors are removed as soon as they arise (1). The proofreading function is determined by the alpha and epsilon subunits of DNA polymerase III holoenzyme which are specified by the polC and mutD(dnaQ) genes respectively (2). Genetic data support this assignment since mutD mutant alleles confer a strong mutator phenotype and certain polC alleles impart enhanced spontaneous mutation frequencies.

The Dam-directed mismatch repair system removes errors in the newly synthesized daughter DNA strand (for reviews,see 3-7). The discrimination between daughter and parental DNA strands

© IRL Press Limited, Oxford, England.

resides in the differential state of methylation of 5'-GATC-3' (Dam recognition) sequences. The daughter strand is thought to be partially undermethylated, whereas the parental strand is fully methylated at these sequences (7). The differential methylation results in selective repair of the undermethylated strand in a heteroduplex containing a methylated and an unmethylated strand, both in vitro and in vivo. No Dam-directed mismatch repair occurs on fully methylated heteroduplexes containing a susceptible base pair mismatch.

The dam locus of E. coli specifies a DNA adenine methylase which modifies the adenine residue in 5'-GATC-3' sequences (7). In dam mutant strains, which have unmethylated DNA, strand discrimination for repair should be lost, and half the time the repair mechanism should fix the replication error into the DNA sequence. For example, repair of an A/C (adenine/cytosine) mismatch (where A is in the parental strand) should produce A/T (adenine/thymine) and G/C (guanine/cytosine) base pairs with equal probability and lead to an increased mutation frequency of AT to GC transition mutations. Hypermutability is a characteristic feature of dam mutants, and we have recently determined that the principal class of mutations from a dam-3 strain is base pair transition mutations (8).

Strains of E. coli which are deficient in Dam-directed mismatch repair should also show a mutator phenotype. The products of the mutH, mutL and mutS genes, among others, are required for efficient repair of mismatched heteroduplexes(3-7). The MutS protein specifically recognizes mismatches in DNA heteroduplexes and the MutH product has weak endonuclease activity in vitro (5). No function has yet been assigned to the MutL gene product.

In such mut strains the mutation spectrum should be identical to that found in the dam-3 strain if the model for Dam-directed repair is correct. That is, an A/C mismatch would not be subject to repair and would yield AT and GC base pairs following chromosome replication. It follows that the increased spontaneous mutation frequency in mut strains should then be principally AT to GC transitions.

In this paper we show that, indeed, the mutation spectra for mutH34 and dam-3 strains are very similar. This result supports the Dam-directed mismatch repair model (4).

MATERIALS AND METHODS

Bacterial strains,phages and plasmids.

The wild type strain, GM3136, is AB1157 (9; a gift of E.A. Adelberg,Yale University) containing F42 (F-lac) and plasmid pPY97. Strain GM3135 is identical to GM3136, except for the mutH34 mutation, and is derived from ES1590,a gift of E. Siegel, Tufts University. Strain MM294 (10) was obtained from the Coli Genetic Stock Center, Department of Biology, Yale University, Box 6666, New Haven CT 06520 USA. The plasmid pPY97 is identical to the previously described plasmid pMQ151 (8) except that it contains an M13 origin of replication (11) and was obtained from P. Youderian, University of Southern California. Phages lambda immP22 hy1 and lambda immP22 dis were gifts of N. Yamamoto, Temple University.

Isolation of tetracycline-resistant strains

Brain Heart infusion broth (Difco; 20 g per liter) containing 40 µg ampicillin per ml was inoculated with about 100 cells per ml and incubated overnight at 37 C. Portions of 0.1-0.3 ml were inoculated onto BH plates containing 40 µg ampicillin per ml and 3.5 µg tetracycline per ml and the plates incubated at 37 C. Single mutant colonies were purified three times on media containing 10 µg tetracycline per ml.

Identification of mnt mutants.

Tetracycline-resistant bacteria were lysogenized with lambda immP22 hy1. Lysogens were identified as immune to lambda immP22 hy1 but sensitive to lambda vir. Lysogens immune to challenge with lambda immP22 dis were considered to harbor plasmid mutations in the Mnt operator while lysogens sensitive to this phage contained mutations in the mnt gene (8,18).

Isolation of plasmid DNA

Plasmid DNA from tetracycline-resistant mnt derivatives of strains GM3135 and GM3136 was isolated (12) and used to transform strain MM294 to tetracycline-resistance. We used this step to ensure that plasmid DNA for sequencing was derived from a single

molecule and because DNA isolated from an endA mutant strain (MM294) gave more consistent DNA sequencing results (13). Plasmid DNA for sequence analysis was isolated as described by Birnboim and Doly (14) except that two phenol extractions were included.

DNA sequence analysis

Supercoiled plasmid DNA was sequenced by the enzymatic method of Sanger et al (15) as modified by Chen and Seeberg (13) and Zagursky et al (16). EcoRl site primer was obtained from New England Biolabs Inc. Primers MM-1 (5'-CTCACAATACAGGTC-3'), MM-4 (5'-GATAGAGCATCTTGG-3') and MM-5 (5'-CCAAGATGCTCTATC-3') were obtained from Dr K.L. Taneja, DNA Synthesis Facility,Univ. of Massachusetts Medical School. Primers were purified by electrophoresis through and elution from a 20% acrylamide-7 M urea gel (17) followed by chromatography over a Sep-Pak C_{18} cartridge (Waters Associates,Inc.). These primers allow the sequence of the entire mnt gene and its promoter to be determined on both strands.

RESULTS

Experimental System

The target DNA sequence we have used to monitor spontaneous mutations is the mnt (maintenance of lysogeny) repressor gene of bacteriophage P22 (18). This gene and part of the arc (antirepressor control) region are present on a pBR322 derivative, pPY97, as a 500 base pair EcoRl-HindIII fragment replacing the 29 base pair sequence of pBR322. This location places the tetA gene under the control of the P22 Mnt repressor which prevents expression from the ant promoter and therefore cells containing the wild type plasmid are tetracycline sensitive. Mutation in mnt or its operator allow transcription from the ant promoter in the operon fusion and enable the plasmid to confer a tetracycline-resistant phenotype on the cell carrying it (Figure 1).

After isolation of spontaneous tetracycline-resistant variants from mut^+ or mutH34 strains, they were lysogenized with a hybrid lambda phage (lambda immP22 hyl). The lysogens were challenged with another hybrid lambda phage, lambda immP22 dis, and those permissive for phage growth were assumed to contain mnt

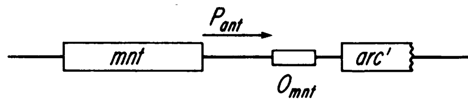

Figure 1. Genetic map of the mnt-arc region. The mnt and arc genes are separated by the ant promoter and the Mnt operator. Mutation in either mnt or its operator in pPY97, promotes transcription from the ant promoter through the tetA gene to yield a tetracycline-resistant phenotype. The arc gene is fused to the tetA gene at the HindIII recognition site of pBR322.

mutated plasmids (8). Plasmid DNA from the presumptive mnt isolates was introduced into strain MM294 for DNA sequencing.

Analysis of spontaneous mutations from the wild type (mut^+) strain.

Tetracycline-resistant colonies arose at a frequency of 0.8 per 10^8 cells plated. From 518 tetracycline-resistant isolates, four (0.8%) were shown to have mutations in the mnt gene (Table 1). One of these is a deletion removing bases +7 through +15 (5'-ACTTGGAGT-3') which eliminates the ribosome binding site (The start site of transcription is +1). The other three isolates contained IS1 insertions at bases +56/57, +121/122 and +138/139, all in the region coding for the amino-terminal end of Mnt, which is critical for DNA binding (19).

The remaining 514 tetracycline-resistant variants were assumed to contain mutations in the Mnt operator and were not examined further.

Distribution and frequency of spontaneous mutations from the mutH34 strain.

In contrast to the results from the wild type, both the frequency of tetracycline-resistant colonies and their genetic constitution are altered in the mutH34 strain. The frequency to tetracycline-resistance was 52-fold higher, and of these 63% were mapped to the mnt gene. Therefore, isolates containing mutations within the mnt target gene have a high probability of being generated by the presence of the mutH34 allele (63% versus 0.8%).

Position of mutations in the mnt gene from the mutH34 strain

Table 2 lists 47 independent mutations identified by DNA sequencing in the mnt gene from the mutH34 strain. Transition

Table 1. Location of mutations in the mnt gene arising spontaneously in the wild type.

Number	Position	Type
1	+7 to +15	Deletion
1	+55/+56	IS1L
1	+121/+122	IS1L
1	+138/+139	IS1R

Base pairs are numbered with respect to the startpoint of transcription (+1) of the mnt gene. The translation initiation codon is at 21-23. IS1 insertions were identified by their DNA sequences and R and L indicate if the right or left end respectively is closer to the mnt promoter (23).

mutations comprise the major class (45/47) and of these the AT to GC class predominates (32/45). Only 13/42 (31%) are GC to AT transitions. The two remaining mutations are a deletion and an insertion. The deletion includes bases -32 to -44 which eliminates the -35 region of the mnt promoter. An IS1 insertion was identified at position +130/131, and this probably represents background "noise" because its frequency (1/47) is approximately that of the mutH34/mut^+ spontaneous mutation frequency (1/52).

Comparison of mutation spectra between dam-3 and mutH34 bacteria

Figure 2 displays the AT to GC mutation spectra in the mnt gene from dam-3 and mutH34 strains. The three predominant hotspots at positions +41, +106 and +116, seen previously in the dam-3 mutant, are also present in the spectrum of the mutH cells. These three hotspots are within six base pairs of the only two 5'-GATC-3' sequences present in the mnt gene. These Dam-recognition sites are located at positions +34 to +37 and +107 to +110. The AT to GC mutations in these three hotspots comprise 87% of all AT to GC mutations identified in the mnt structural gene isolated from the mutH34 strain.

There are some differences in the mutation spectra between the two strains in mutational classes other than AT to GC transitions. For example, no frameshift or transition mutations were identified from the mutH34 strain. This is probably due to the small sample size of mutations analyzed and the relatively low frequency of frameshifts (5/91) and transversions (4/91) in the dam-3 sample.

Table 2. Location of mutations in the mnt gene arising in a mutH34 strain.

Number	Position	Codon change	Amino acid change
1	-32/-44	Deletion	-
1	+28	AGA to GGA	Arg2 to Gly
8	+41	CAC to CGC	His6 to Arg
1	+46	AAC to GAC	Asn8 to Asp
1	+47	AAC to AGC	Asn8 to Ser
3	+52	CGT to TGT	Arg10 to Cys
1	+53	CGT to CAT	Arg10 to His
2	+70	AGG to GGG	Arg16 to Gly
12	+106	AGA to GGA	Arg28 to Gly
3	+110	TCA to TTA	Ser29 to Leu
8	+116	AAC to AGC	Asn31 to Ser
2	+121	GAG to AAG	Glu33 to Lys
1	+129/+130	IS1R	-
3	+273	TGA to TGG	Stop to Trp

Legend as in Table 1.

Sample size ,however, does not explain a difference in the GC to AT class between the two strains. The GC to AT class represents 15% and 29% of the total transitions in the dam and mutH strains respectively. Furthermore, in mutH34 GC to AT transitions are present at positions +53, +110 and +121 (46% of the total) but do not occur at all in the dam-3 spectrum.

DISCUSSION

Our results indicate that the principal base pair mismatches which are substrates for the Dam-directed mismatch repair system in vivo, are those leading to transitions, i.e., A/C or G/T. Furthermore, we have shown that the mutation spectra in dam-3 and mutH34 strains are very similar, as predicted by the model (4).

Our results on the specificity of repair complement those

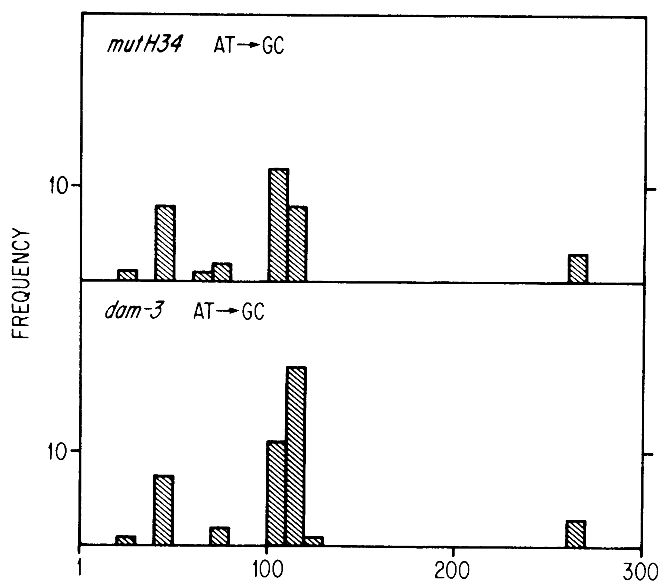

Figure 2. Distribution of AT to GC mutations in the mnt gene obtained from dam-3 and mutH34 strains. Base numbering begins with the start site of transcription. Note the three hotspots at positions 41,106 and 116. The mutations at base 273 are in the transcription termination codon.

obtained using artificially constructed heteroduplexes containing defined base pair mismatches (3-6). Those studies indicated that the repair system preferentially corrected transition mismatches either after transfection of phage heteroduplexes into repair competent cells or by measuring excision in vitro. The in vivo and in vitro results are thus in good agreement.

During the course of this work, Leong et al (20) have determined the mutation spectrum in mutH34, mutS3 and mutU4 (uvrD) bacteria using the lacI gene as a target. The spectra for all three strains were essentially the same and consisted principally of GC to AT transitions. Although this result seems in contradiction to our own, the lacI system does not monitor AT to GC transitions. Taken together, however, the two studies are in agreement that transition mutations predominate in the absence of Dam-directed mismatch repair.

The AT to GC mutations in the mnt gene isolated from mutH34 cells are clustered in three hotspots close to the only two 5'-

GATC-3'sequences in the gene. At present, we do not know the mechanism by which these hotspots arise. We can, however, rule out that methylation of these sequences is involved because of the similarity of the mutation spectra in dam-3 cells, in which DNA is not methylated, and mutH34 in which the DNA is methylated.

The MutH protein has been implicated in genetic experiments to be responsible for strand discrimination during mismatch repair (21). This is probably because it has weak nuclease activity and cuts DNA at the 5' side of the G in GATC sequences on the unmethylated strand in a hemi-methylated duplex, or on one of the two strands in an unmethylated duplex (5). Symmetrically methylated duplexes are resistant to the action of the endonuclease. This suggests that there may be increased MutH-dependent cleavage in a dam mutant and that subsequent excision and/or resynthesis might introduce errors in DNA near Dam sites. This explanation for the generation of the hotspots must also be discarded since these hotspots occur in the mutH34 cells which we presume to have little, if any, nuclease activity although this has not been measured directly.

Another model for the generation of the hotspots may involve proteins that bind specifically to 5'-GATC-3' sequences in DNA. This may expain why two of the hotspots are ten base pairs apart, which represents one turn of the DNA helix, because both hotspot sites would be on the same face of the DNA. Protein binding may also explain why only one hotspot is present at base pair +41. The sequence at +29,30 31 and 32 is GAGA. The A at position 30 is at a wobble position and thus if mutations do occur at this site they would remain undetected. No mutations at position 32 have been isolated and this may mean that a change from asparagine to glycine may be tolerated at this position. Neither G could be a hotspot position.

Two proteins are known which can bind to Dam sites: Dam and MutH (7,22). In a dam mutant strain MutH may bind to 5'-GATC-3'sequences and in the mutH34 cells Dam protein binds to the same sequence. The effect of protein binding can be tested by determining if hotspots are present in a strain deleted for both the dam and mutH genes.

An alternative model for the presence of hotspots is that codons for amino acids of Mnt critical to DNA binding are by chance located close to the 5'-GATC-3'sequences. We are testing this model by constructing derivatives of our mnt plasmid in which either or both 5'-GATC-3' sequences are altered but in which the amino acid sequence is the same as wild type. Once obtained we will assess if the hotspots are present or not in such derivatives.

ACKNOWLEDGMENTS

We thank Drs. E.A. Adelberg,E.C. Siegel,N. Yamamoto and P. Youderian for gifts of plasmids,bacterial strains and phages. This work was supported by grant GM33233 from the US Public Health Service.

REFERENCES

1. Kornberg,A. (1980) DNA Replication,Freeman,San Francisco.
2. Maki,H. and Kornberg,A. (1987) Proc. Natl. Acad. Sci.,USA. 84, 4389-4392.
3. Claverys,J-P. and Lacks,S.A. (1986) Microbiol. Rev. 50,133-165.
4. Meselson,M. (1987) In Low,K.B. (ed.), The Recombination of Genetic Material,Academic Press,New York,pp
5. Modrich,P. (1987) Ann. Rev. Biochem. 56,
6. Radman,M. and Wagner,R. (1986) Ann. Rev. Genet. 20,523-538.
7. Marinus,M.G. (1987) In Neidhardt,F.C.,Ingraham,J.L.,Low,K.B., Magasanik,B.,Scheachter,M. and Umbarger,H.E. (eds.),Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology,pp.697-702. Am. Soc. Microbiol.,Washington,D.C.
8. Carraway,M.,Youderian,P. and Marinus,M.G. (1987) Genetics 116, 343-347.
9. Howard-Flanders,P.,Simson,E. and Theriot,L. (1964) Genetics 49,237-246.
10. Meselson,M. and Yuan,R. (1968) Nature 217,1111-1114.
11. Zagursky,R.J. and Berman,M.L. (1984) Gene 27,183-191.
12. Davis,R.W.,Botstein,D. and Roth,J.R. (1980) Advanced Bacterial Genetics.pp124-125. Cold Spring Harbor Laboratories, New York.
13. Chen,E.Y. and Seeberg,P.H. (1985) DNA 4,165-170.
14. Birnboim,M.C. and Doly,J. (1979) Nucl. Acids Res. 7,1513-1523.
15. Sanger,F.,Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci.,USA. 74,5463-5467.
16. Zagursky,R.J.,Baumeister,K.,Lomax,N. and Berman,M.L. (1985) Gene Anal. Techn. 2,89-94.
17. Maxam.A.M. and Gilbert,W. (1980) Methods Enzymol. 65,499-560.

18. Susskind,M.M. and Youderian,P. (1983) In Hendrix,R.,Roberts,J. W.,Stahl,F. and Weisberg,P. (eds.), Lambda II,Cold Spring Harbor Laboratory,pp. 347-363.
19. Youderian,P.,Vershon,D.,Bouvier,S., Sauer,R.T. and Susskind, M.M. (1983) Cell 35,777-783.
20. Leong,P-M.,Hsia,H.C. and Miller,J.H. (1986) J. Bacteriol. 168,412-416.
21. Kramer,B.,Kramer,W. and Fritz,H-J. (1984) Cell 38,879-887.
22. Seiler,A.,Blocker,H.,Frank,R. and Kahmann,R. (1986) EMBO J. 5,2719-2728.
23. Ohtsubo,M. and Ohtsubo,E. (1978) Proc. Natl. Acad. Sci.,USA, 75,615-619.