Nonconserved segment of the MutL protein from *Escherichia coli* K-12 and *Salmonella typhimurium*

Ho-Ching Tiffany Tsui, Bela S.Mandavilli and Malcolm E.Winkler*
Department of Microbiology and Molecular Genetics, University of Texas Medical School, Houston, TX 77030, USA

Submitted March 19, 1992

EMBL accession no. Z11831

The *mutL* gene product is required for methyl-directed mismatch repair *in vivo* (1–3) and *in vitro* (4,5), although its exact biochemical function is unclear (5). According to one recent model, MutL protein may function as an interface between the MutS protein, which binds to base-pair mismatches, and the MutH d(GATC) endonuclease (5). The *mutL* gene of *Escherichia coli* K-12 and *Salmonella typhimurium* seems to be in a complex operon with at least two other genes (6–8). The upstream gene encodes a 47,000 dalton protein of unknown function. The downstream gene is *miaA*, which encodes a tRNA modification enzyme, and the translation stop and start of *mutL* and *miaA* probably overlap (7). To understand the regulation of this complex locus, it was necessary to determine the DNA sequence of *E. coli mutL* and other potential upstream genes. Because *E. coli* MutL has been the subject of several biochemical studies (5), we wanted to make its deduced amino acid sequence available as soon as possible.

Comparison of *E. coli* MutL with the previously published sequence of the *S. typhimurium* enzyme showed one noteworthy feature. The first 330 and last 189 amino acids of MutL from *E. coli* and *S. typhimurium* are generally highly conserved with no gaps and few nonidentical or nonsimilar amino acids. In contrast, the central region of the MutL protein (amino acids 331–425 relative to the *E. coli* sequence) consists of a mixture of conserved and nonconserved segments, including three distinct gaps (Fig. 1). In this regard, it is interesting that a recent comparison of the *E. coli* and *S. typhimurium* MutS proteins also revealed nonconserved segments, including gaps, at the middle of the protein, but highly conserved, nearly identical sequences at both ends (9). Moreover, when we compared other analogous protein pairs from *E. coli* and *S. typhimurium*, we found that such distinct internal gaps are relatively rare. Of 133 pairs available in the SwissProt database, approximately 94% contained no gaps or only one gap of one or two amino acids. Besides MutL and MutS, only five other protein pairs (CheA, Dyr3, FliC, OmpA, and PhoQ) showed sizeable internal gaps. One speculation is that the nonconserved segments in MutL and MutS may represent hinge regions between different protein domains which interact with MutS and MutH or with DNA and MutL, respectively. As expected from previous studies, the *E. coli* MutL amino acid sequence reported here showed significant homology to *Streptococcus pneumoniae* HexB and *Saccharomyces cerevisiae* PMS1, which are thought to represent a family of repair proteins with a common evolutionary origin (8, 10, 11).

ACKNOWLEDGEMENTS

This work was supported by Public Health Service grant GM43070 from the National Institute of General Medical Sciences to M.W.

REFERENCES

1. Cox,E.C. (1976) *Ann. Rev. Genet.* **10**, 135–156.
2. Choy,H.E. and Fowler,R.G. (1985) *Mutat. Res.* **142**, 93–97.
3. Schaaper,R.M. and Dunn,R.L. (1987) *Proc. Natl. Acad. Sci.* USA **84**, 6220–6224.
4. Grilley,M., Welsh,K.M., Su,S-S. and Modrich,P. (1989) *J. Biol. Chem.* **264**, 1000–1004.
5. Modrich,P. (1991) *Ann. Rev. Genet.* **25**, 229–253.
6. Connolly,D.M. and Winkler,M.E. (1989) *J. Bacteriol.* **171**, 3233–3246.
7. Connolly,D.M. and Winkler,M.E. (1991) *J. Bacteriol.* **173**, 1711–1721.
8. Mankovich,J.A., McIntyre,C.A. and Walker,G.C., (1989) *J. Bacteriol.* **171**, 5325–5331.
9. Schlensog,V. and Böck,A., (1991) *J. Bacteriol.* **173**, 7414–7415.
10. Prudhomme,M., Martin,B., Mejean,V. and Claverys,J.-P. (1989) *J. Bacteriol.* **171**, 5332–5338.
11. Kramer,W., Kramer,B., Williamson,M.S. and Fogel,S. (1989) *J. Bacteriol.* **171**, 5339–5346.

```
EcMutL   300 DVNVHPAKHEVRFHQSRLVHDFIYQGVLSVLQQQLETPLPLDDEPQPAPR 349
             ||||||||||||||||||||||||||||||||||||||||||| ||:||| ||  ||||
StMutL   300 DVNVHPAKHEVRFHQSRLVHDFIYQGVLSVLQQQTETALPL.EEIAPAPR 348

EcMutL   350 SIPENRVAAGRNHFAEP....AAREPVAPRYTPAPASGS...RPAAPWPN 392
             |.||||||||||||| |    |||||  |||  :: :|     |: ||:
StMutL   349 HVQENRIAAGRNHFAVPAEPTAAREPATPRYSGGASGGNGGRQSAGGWPH 398

EcMutL   393 AQPGYQKQQGEVYRQLLQTPAPMQKLKAPEPQEPALAANSQSFGRVLTIV 442
             |||||||||||||||| |||||:    |||   ||| ::|||||||||||
StMutL   399 AQPGYQKQQGEVYRTLLQTPTASP...APESVTPALDGHSQSFGRVLTIV 445

EcMutL   443 HSDCALLERDGNISLLSLPVAERWLRQAQLTPGEAPVCAQPLLIPLRLKV 492
             :|||||: | | ||||||||||||||||||: ||||||||||||||||
StMutL   446 GGDCALLEHAGTIQLLSLPVAERWLRQAQLTPGQSPVCAQPLLIPLRLKV 495
```

Figure 1. Alignment of the central region of the *E. coli* (Ec) and *S. typhimurium* (St) MutL proteins that contains several gaps and nonconserved segments. The alignment was generated by using Bestfit from the University of Wisconsin Computer Group programs with gap and length weights of 3.0 and 0.1, respectively.

* To whom correspondence should be addressed