A highly conserved sequence motif defining the family of MutT-related proteins from eubacteria, eukaryotes and viruses

Eugene V.Koonin
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received July 9, 1993; Revised and Accepted September 2, 1993

Escherichia coli protein MutT is a component of the GO system that prevents the deleterious effect of 7,8-dihydro-8-oxoGTP (8-oxo-GTP), the spontaneously formed mutagenic substrate of DNA replication (1). Inactivation of the *mutT* gene results in a drastic specific increase in the frequency of A-T/C-G transversions (2). MutT protein has a weak dGTPase activity (3), resulting in the formation of dGMP and pyrophosphate, and shows a much higher activity with (8-oxo-GTP) as the substrate (4). A similar enzymatic activity has been isolated from mammalian cells (5).

Comparison of the amino acid sequence of MutT with the sequence databases using the BLAST program (6) revealed limited similarities with putative products of uncharacterized open reading frames from *E.coli, Xenopus laevis* and mammals, and with gene products of poxviruses and African swine fever virus whose function(s) is not known either. While not highly significant statistically, all these similarities consistently highlighted the same segment of the MutT protein consisting of about 30 amino acid residues, suggesting its possible functional importance. Further database comparisons with the sequences similar to MutT allowed the identification of additional members of the putative MutT protein family. Multiple alignment of these proteins revealed one highly conserved block (probability of finding it by chance alone was computed to be below 10^{-15}; closely related sequences were omitted from this calculation), with six positions containing strictly conserved amino acid residues (Figure 1). A unique sequence signature for this family was derived (legend to Figure 1). Secondary structure prediction (7) strongly suggested that the best conserved portion of the common block in the proteins of the MutT family forms an α-helix flanked by too loops (Figure 1), a structure very unusual for catalytic cores of NTPases (8).

The proteins of the MutT family may possess pyrophosphate-releasing NTPase activity, with the conserved block comprising part of the active center of a novel type. It remains to be determined whether they are involved in inactivation of mutagenic substrates of DNA replication.

ACKNOWLEDGEMENT

Thanks are due to Kenn Rudd for helpful discussions.

REFERENCES

1. Michaels,M.L. and Miller,J.H. (1992) *J. Bacteriol.* **174**, 6321–6325.
2. Yanofsky,C. *et al.* (1966) *Proc. Natl. Acad. Sci. USA* **55**, 274–281.
3. Bhatnagar,S.K. *et al.* (1991) *J. Biol. Chem.* **266**, 9050–9054.
4. Maki,H. and Sekiguchi,M. (1992) *Nature* **355**, 273–275.
5. Mo,J.-Y. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11021–11025.
6. Altschul,S.F. *et al.* (1990) *J. Mol. Biol.* **215**, 403–410.
7. Rost,B. and Sander,C. (1992) *Nature* **360**, 540.
8. Schulz,G.E. (1992) *Curr. Opin. Struct. Biol.* **2**, 61–67.
9. Schuler,G.D. *et al.* (1991) *Proteins Struct. Funct. Genet.* **9**, 180–190.
10. Binns,M.M. *et al.* (1990) *J. Gen. Virol.* **71**, 2873–2881.

```
consensus              .....G.....E.....U.RE&.EE.........&
                                        A
MutT     E.coli     32 LEFPGGKIEMGETPEQAVVRELQEEVGITPQHFSL  62  P08337
MutT     P.vulgaris 35 WEFPGGKLEDNETPEQALLRELQEEIGIDVTQCTL  42  L07579(G)
YjaD     E.coli     42 HTVLAGFVEVGETLEQAVAREVMEESGIKVKNLRY  69  D12624(G)
YebD     E.coli     35 WQSVTGSVEEGETAPQAAMREVKEEVTIDVVAEQL  80  P24236
ORF154   S.ambo.    42 WELPGGVLELDETPETGVAREVWEETGIRVEVDEL  77  Z19590(G)
D10      VV        120 AIYPGGIPKRGENVPECLSREIKEEVNIDNSFVFI  93  P21012
D10      SFV       122 IIFPGGLPKNEEDPIMCLSREIKEEINIDSKDIYI 103  M74532(G)
D10      FPV       110 IIFPGGKIKDLESITNCLVREIKEELNIDSSYLAI  80  ref. 10
D9       VV        105 LILLGGKLDKKESIKDCLRRELKEESDERITVKEF  73  P21011
D9       SFV       110 LILLGGKLNKSETIDDCIRREIKEETDSKLTIKSI  73  M74532(G)
D250     AFSV      126 WEIPKGKPKEDESDLTCAIREFEEETGITREYYQI  89  L07263(G)
antiFGF  X.l.       71 WKFPGGLSDQGEDIGATAVREVLEETGIHSEFKSL 111  P13420
antiFGF  human       ? ------------VGDTAVREVFEETGIKSEFRSV 111  M27968(G)
antiFGF  bovine      ? ------------VGDTAVREVFEETGIKSEFRSL 111  M13440(G)
                       HHlllllllllllHHHHHHHHHHHHHlllHHHHHH
```

Figure 1. Conserved sequence block in the proteins of the MutT family. The alignment was generated using the MACAW program (9). Distances to the protein termini are indicated for each sequence. Consensus includes amino acid residues conserved in all of the aligned sequences (U — a bulky aliphatic residue, i.e. I, L, V or M; and &— a bulky hydrophobic residue, i.e. I, L, V, M, F, Y or W). The signature $\mathbf{Gx_5Ex_5}$[UA]xRE&xEE$\mathbf{x_9}$& was found to be a unique identifier of the MutT family when the complete non-redundant amino acid sequence database (National Center for Biotechnology Information) was screened. Secondary structure prediction is shown: H, α-helix, l-loop. The sequences were from current databases except for D10 of FPV. Each sequence is accompanied by a SWISSPROT or GenBank (G) accession number. Abbreviations: P.vulgaris, *Proteus vulgaris*; S.ambo., *Streptomyces ambofaciens* (a plasmid gene product); VV, vaccinia virus (the closely related sequences of variola virus are not shown); FPV, fowlpox virus; SFV, Shope fibroma virus; AFSV, African swine fever virus; X. l., *Xenopus laevis*, antiFGF, putative protein encoded by the antisense RNA of the basic fibroblast growth factor gene (the human and bovine antiFGF sequences are incomplete at their N-termini).