Nucleic Acids Research, 1998, Vol. 26, No. 18 **4153–4159**

Domain organization and functional analysis of *Thermus thermophilus* MutS protein

Hidehisa Tachiki, Ryuichi Kato, Ryoji Masui, Koichi Hasegawa, Hiroyuki Itakura, Keiichi Fukuyama and Seiki Kuramitsu*

Department of Biology, Graduate School of Science, Osaka University, 1-1 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan

Received June 11, 1998; Revised and Accepted July 31, 1998

ABSTRACT

MutS protein binds to DNA and specifically recognizes mismatched or small looped out heteroduplex DNA. In order to elucidate its structure–function relationships, the domain structure of *Thermus thermophilus* MutS protein was studied by performing denaturation experiments and limited proteolysis. The former suggested that *T.thermophilus* MutS consists of at least three domains with estimated stabilities of 12.3, 22.9 and 30.7 kcal/mol and the latter revealed that it consists of four domains: A1 (N-terminus to residue 130), A2 (131–274), B (275–570) and C (571 to C-terminus). A gel retardation assay indicated that *T.thermophilus* MutS interacts non-specifically with double-stranded (ds), but not single-stranded DNA. Among the proteolytic fragments, the B domain bound to dsDNA. On the basis of these results we have proposed the domain organization of *T.thermophilus* MutS and putative roles of these domains.

INTRODUCTION

Various chemical compounds, such as mutagens and radicals, some physical agents, such as ultraviolet (UV) light and ionizing radiation, and inaccurate DNA replication lead to damage or errors in DNA (1). The DNA lesions can be a source of induction of mutagenesis and carcinogenesis. All living cells have DNA repair systems, which involve photoreactivation and base excision, nucleotide excision, mismatch and recombinational repair, to counteract DNA damage (1). The mismatch repair system acts on mismatched DNA produced by DNA replication errors, genetic recombination and modified bases. In *Escherichia coli*, this repair system consists of MutS, L and H proteins (2) and MutS and MutL homologues have been found in other bacteria and in eukaryotes (3,4), which suggests that the mismatch repair system is ubiquitous. Kindred analysis of homologues of MutS and MutL in humans has indicated that they play roles in hereditary nonpolyposis colorectal cancer (HNPCC) (3,4). Molecular analysis of the mismatch repair system is required in order to elucidate the mechanisms responsible for mutagenesis and tumor development.

In the mismatch repair system, MutS and its homologues bind to mismatched or base looped out heteroduplex DNA. The three-dimensional structure of MutS is unknown and there are few clues to its structure–function relationships. It is important to identify which regions of MutS interact with DNA, and some studies have been carried out in attempts to do so. The finding that the *E.coli* MutS dimer bound to mismatched DNA and formed an α-shaped loop structure (5) suggests that this protein has two DNA binding sites: one binds mismatched DNA and the other binds homoduplex DNA. Malkov *et al*. (6) reported that phenylalanine (Phe39) at the N-terminal region was crucial for heteroduplex DNA binding by *Thermus aquaticus* MutS, whereas the C-terminal region of hMSH2, a human homologue of MutS, was reported to be sufficient to bind mismatched DNA (7). The importance of the C-terminal region was also implied by the reduced affinity of a C-terminal mutant of *Salmonella typhimurium* MutS for heteroduplex DNA (8).

In order to understand the structure–function relationships of proteins, enzymatic and physicochemical analysis must be carried out, and proteins from thermophilic bacteria are particularly useful for such studies, because they are stable and easily crystallized. *Thermus thermophilus* HB8 is an aerobic, rod-shaped, non-sporulating Gram-negative eubacterium, which can grow at temperatures >80°C (9). In a previous study, we cloned and overexpressed the *T.thermophilus mutS* gene and purified the gene product (10). We found that the purified protein was stable between pH 1.5 and 12 at 25°C and at temperatures up to 80°C at a neutral pH. We also found that it had ATPase activity and recognized G-T mismatches.

Generally, large proteins contain multiple domains (11) and the functions of multifunctional proteins are often associated with individual domains. We have studied several repair systems, including mismatch repair, using *T.thermophilus* HB8 (10,12–18) and the *T.thermophilus* UvrB protein, which plays an essential role in the nucleotide excision repair system, was found to consist of five domains (15). As MutS is a relatively large (~90 kDa) multifunctional protein, it is probably a multidomain protein like UvrB. The domain structure of a protein can be elucidated if the structural stabilities of the domains differ and protein domains can

*To whom correspondence should be addressed. Tel: +81 6 850 5433; Fax: +81 6 850 5442; Email: kuramitu@bio.sci.osaka-u.ac.jp

be dissected by limited proteolysis if they are loosely assembled and the connecting regions are exposed to the environment.

In this study, we subjected *T.thermophilus* MutS to denaturation and limited proteolysis and the results suggest that this protein possesses at least four domains. We also studied the DNA-binding activities of *T.thermophilus* MutS and its proteolytic fragments and found that the MutS protein interacted with double-stranded (ds) DNA but not single-stranded (ss) DNA and that the central domain of the protein interacted directly with dsDNA. On the basis of these results, we have proposed a model of the domain organization of *T.thermophilus* MutS and discussed the relationships between its function and structure.

MATERIALS AND METHODS

Materials

The enzymes and reagents used were purchased as follows: recrystallized nagarse (subtilisin BPN′) from Nagase Industrial, thermolysin from Daiwa Chemical; phenylmethylsulfonyl fluoride (PMSF) from Wako Pure Chemical, poly(dC) and poly(dI)·(dC) from Pharmacia Biotech and Immobilon-P^{SQ} polyvinylidene difluoride (PVDF) membranes from Millipore. *Thermus thermophilus* MutS was prepared as described previously (10) and all the other chemicals and reagents used were purchased from commercial sources.

Spectroscopic measurements

Circular dichroism (CD) measurements were carried out using a Jasco spectropolarimeter, model 720W. The samples comprised 2 µM *T.thermophilus* MutS, 100 mM KCl, 1 mM EDTA, 1 mM dithioerythritol (DTE), 50 mM Tris–HCl (pH 7.5) and the required concentrations of guanidine hydrochloride (GdnHCl). After incubation at 25°C for 1 h, the CD ellipticity values at 222 nm were measured using a 1 mm light-path cell. No further changes in the ellipticity values were observed after incubation for 24 h.

Reaction mixtures comprising 1 µM *T.thermophilus* MutS in 100 mM KCl, 1 mM EDTA, 1 mM DTE, 50 mM Tris–HCl (pH 7.5) and the required concentrations of GdnHCl were incubated at 25°C for 1 h, then placed in 5 × 5 mm cells and their fluorescence spectra at an excitation wavelength of 280 or 295 nm were recorded using a Hitachi spectrofluorometer, model F-4500.

Thermodynamic analysis of ellipticity (at 222 nm) denaturation curves

The three denaturation states of the protein would be expected to obey equation **1**:

$$N \underset{\leftrightarrow}{\overset{K_1}{}} I_1 \underset{\leftrightarrow}{\overset{K_2}{}} I_2 \underset{\leftrightarrow}{\overset{K_3}{}} U \qquad \mathbf{1}$$

where N, I_1, I_2 and U represent the native, intermediate I_1, intermediate I_2 and unfolded molecules, respectively, and K_1, K_2 and K_3 are the equilibrium constants for the N to I_1, I_1 to I_2 and I_2 to U, respectively. The ellipticity measured at 222 nm was expressed by equation **2**:

$$\theta_{222} = f_N\theta_N + f_{I_1}\theta_{I_1} + f_{I_2}\theta_{I_2} + f_U\theta_U \qquad \mathbf{2}$$

where θ_N, θ_{I_1}, θ_{I_2} and θ_U are the molar ellipticities of the native, intermediate I_1, intermediate I_2 and unfolded proteins, respectively, and f_N, f_{I_1}, f_{I_2} and f_U are the fractions of the molecules in the native, intermediate I_1, intermediate I_2 and unfolded states, respectively ($f_N + f_{I_1} + f_{I_2} + f_U = 1$). The fractions are expressed by the following equations:

$$f_N = \frac{1}{1 + K_1 + K_1K_2 + K_1K_2K_3} \qquad \mathbf{3}$$

$$f_{I_1} = \frac{K_1}{1 + K_1 + K_1K_2 + K_1K_2K_3} \qquad \mathbf{4}$$

$$f_{I_2} = \frac{K_1K_2}{1 + K_1 + K_1K_2 + K_1K_2K_3} \qquad \mathbf{5}$$

$$f_U = \frac{K_1K_2K_3}{1 + K_1 + K_1K_2 + K_1K_2K_3} \qquad \mathbf{6}$$

The equilibrium constants are related to the changes in Gibbs energy as follows:

$$\Delta G_1 = -RT \ln K_1 \qquad \mathbf{7}$$

$$\Delta G_2 = -RT \ln K_2 \qquad \mathbf{8}$$

$$\Delta G_3 = -RT \ln K_3 \qquad \mathbf{9}$$

The effects of GdnHCl on protein stability can be accounted for by the following equations:

$$\Delta G_1 = \Delta G_{01} - m_1[\text{GdnHCl}] \qquad \mathbf{10}$$

$$\Delta G_2 = \Delta G_{02} - m_2[\text{GdnHCl}] \qquad \mathbf{11}$$

$$\Delta G_3 = \Delta G_{03} - m_3[\text{GdnHCl}] \qquad \mathbf{12}$$

where ΔG_{01}, ΔG_{02} and ΔG_{03} are the changes in free energy upon unfolding of the native to intermediate I_1, intermediate I_1 to intermediate I_2 and intermediate I_2 to the unfolded states (all extrapolated to a denaturant concentration of zero), respectively, and m_1, m_2 and m_3 represent the dependency of their respective Gibbs free energies on the GdnHCl concentration.

Limited proteolysis

Thermus thermophilus MutS was treated with nagarse or thermolysin, at a protein to protease weight ratio of 100:1, for various times at 37 or 60°C, respectively. The reaction mixtures comprised 1 µg/ml protein, 10 ng/ml protease and 50 mM Tris–HCl (pH 7.5). $CaCl_2$ was added to the thermolysin-containing reaction mixture to produce a final concentration of 10 mM. Each reaction was stopped by adding the polyacrylamide gel electrophoresis (PAGE) loading dye, to produce final concentrations of 1% (w/v) SDS, 2.5% β-mercaptoethanol, 5% glycerol, 0.05% bromophenol blue and 30 mM Tris–HCl (pH 7.2), and boiling the mixture for 3 min. The digests were separated by SDS–PAGE using 12.5% (w/v) acrylamide gel (19) and then stained with Coomassie brilliant blue (CBB).

The proteolytic fragments separated by SDS–PAGE were electroblotted onto a PVDF membrane (20). The bands visualized by Ponceau S staining were excised from the membrane and their N-terminal amino acid sequences were determined using a protein sequencer (Applied Biosystems, model 473A).

DNA binding assay

DNA binding was assayed by native PAGE (21) using poly(dC) and poly(dI)·(dC) as ssDNA and dsDNA, respectively. The reaction mixtures, which comprised 5 µM *T.thermophilus* MutS, 200 µM DNA, 10 mM $MgCl_2$ and 25 mM Tris–HCl (pH 7.5), were incubated on ice for 30 min, loaded onto a 5% acrylamide gel containing 10 mM $MgCl_2$, electrophoresed at 5°C in a buffer containing 10 mM $MgCl_2$ under non-denaturing conditions (21) and the gel was stained with CBB to visualize the bands.

DNA binding by the proteolytic fragments was assayed as follows. *Thermus thermophilus* MutS (1 µg/ml) was digested with nagarse (20 ng/ml) in 50 mM Tris–HCl (pH 7.5) at 37°C for 30 min. The reaction was stopped by adding PMSF to produce a final concentration of 100 µM, and a portion of the reaction mixture was subjected to SDS–PAGE to confirm the protein had fragmented. Then, mixtures comprising 5 µM each digest, 200 µM substrate DNA, 10 mM $MgCl_2$ and 25 mM Tris–HCl (pH 7.5) were incubated on ice for 30 min and electrophoresed on a 5% acrylamide gel at 5°C in the presence of 10 mM $MgCl_2$. After electrophoresis, the gel was stained with CBB to visualize the bands or the fragments were electroblotted onto a PVDF membrane and their N-terminal amino acid sequences were determined, as described above.

RESULTS

Denaturation of *T.thermophilus* MutS

In order to study the organization of the structural domains, the conformations of *T.thermophilus* MutS in the presence of various concentrations of the denaturant GdnHCl were examined and the stability of each domain in the absence of denaturant was determined. The effects of increasing the GdnHCl concentration on the protein were studied using CD and fluorescence spectroscopy. As shown in Figure 1A, the intensity of the mean residue ellipticity at 222 nm ($[\theta]_{222}$) decreased as the denaturant concentration increased. The denaturation curve showed that two highly cooperative transitions and a gradual increase in ellipticity had occurred and the midpoints of the transitions occurred with 1.5, 3.0 and 4.5 M GdnHCl. These three transitions suggest that *T.thermophilus* MutS consists of at least three structural domains with different stabilities.

The denaturation curve was analyzed thermodynamically according to a four-state model as described in Materials and Methods. The measured values of $[\theta]_{222}$ fitted the theoretical curve well (Fig. 1A) and the stabilities of the three domains in the absence of denaturant were calculated to be 12.3, 22.9 and 30.7 kcal/mol. The fractional distributions of each intermediate of *T.thermophilus* MutS are shown in Figure 1B and those of intermediate I_1 and the I_2 were maximal at 2.0 and 3.5 M GdnHCl, respectively. At ~2.0 M GdnHCl, 99% of the molecules had one domain unfolded, whereas at ~3.5 M GdnHCl, ~90% of the molecules were in the unfolded state. These observations suggest that the denaturation pathway involves at least four-conformational states: native, intermediate I_1 (at ~2.0 M GdnHCl), intermediate I_2 (at ~3.5 M GdnHCl) and unfolded (Fig. 1B).

In order to obtain further information on the aromatic residues, fluorescence spectroscopy was carried out. First, the fluorescence intensities of the protein excited by 280 nm light in the presence of various concentrations of GdnHCl were measured and they decreased as the denaturant concentration increased (Fig. 2A).

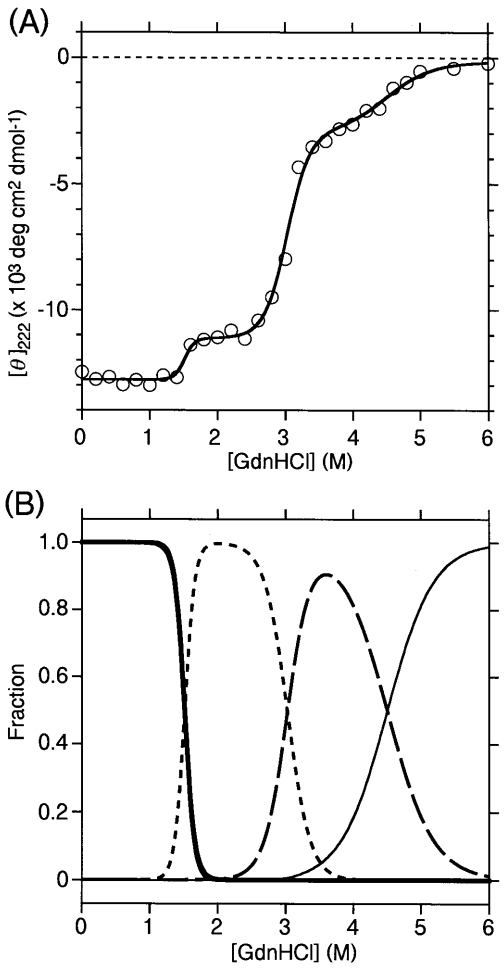

Figure 1. Effect of GdnHCl concentration on the CD spectra of *T.thermophilus* MutS. (**A**) Changes in the mean residue ellipticity at 222 nm are shown. *Thermus thermophilus* MutS (2 µM) was denatured with the indicated concentrations of GdnHCl. The solid line is a theoretical curve calculated according to a four-state model (see text for details). (**B**) Fractions of the protein in its native (thick line), intermediate I_1 (dotted line), intermediate I_2 (broken line) and unfolded (thin line) states.

The midpoints of the transitions occurred with 1.5 and 3.0 M GdnHCl. The emission maximum shift also showed two transitions. In the presence of 1.5 M GdnHCl, the emission maximum shifted toward higher wavelengths and the shift increased as the denaturant concentration increased. This was due to the exposure of tryptophan residues to the solvent as a result of denaturation. Then, it shifted toward a lower wavelength of ~305 nm, which is characteristic of exposure of tyrosine residues, with the second transition midpoint occurring at 3.0 M GdnHCl (Fig. 2B). These transition midpoints agreed well with those determined from the CD measurements.

The fluorescence intensity denaturation curve (Fig. 2A) was analyzed thermodynamically in the same way as the CD denaturation curve, but as only two transitions were observed, a three-state model, which corresponds to $\theta_{I_2} = \theta_U$ in equation **2**, was applied. A theoretical curve (Fig. 2A, continuous line) was constructed using the corresponding free energy values obtained

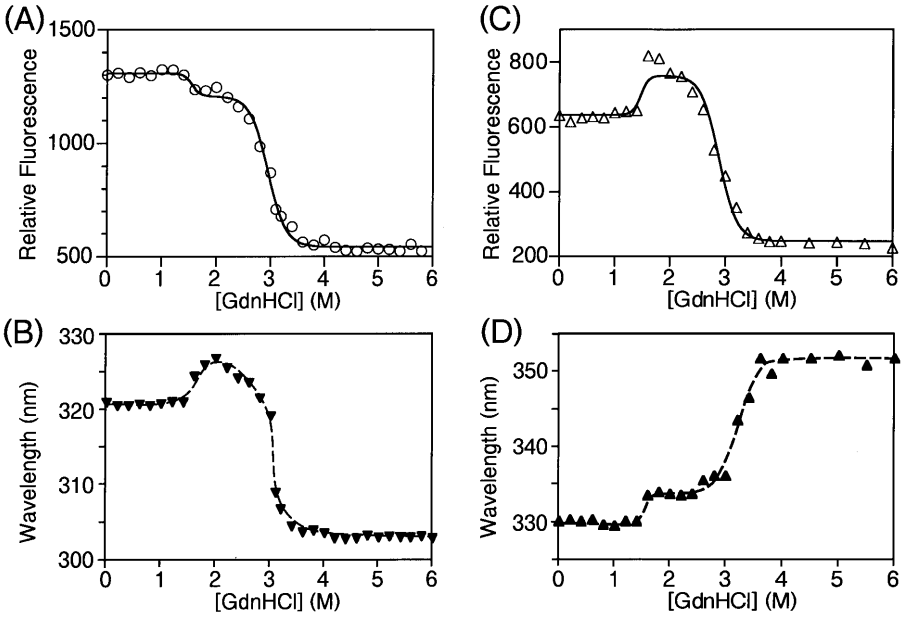

Figure 2. Effect of GdnHCl concentration on the fluorescence of *T.thermophilus* MutS. *Thermus thermophilus* MutS (1 µM) was denatured with the indicated concentrations of GdnHCl and the spectrum of each reaction mixture in a 5 × 5 mm cell at 25°C was measured. (**A**) Changes of fluorescence intensity at 320 nm with an excitation wavelength of 280 nm. (**B**) Changes of wavelength at the maximum fluorescence intensity at an excitation wavelength of 280 nm. (**C**) Changes of fluorescence intensity at 335 nm at an excitation wavelength of 295 nm. (**D**) Changes of wavelength at the maximum fluorescence intensity at an excitation wavelength of 295 nm. The continuous lines in panels (A) and (C) are theoretical curves calculated according to a three-state model (see text for details).

from the CD measurements (Fig. 1) and the measured values fit the theoretical curve well.

In order to gain more detailed information about the domains which possessed a tryptophan residue(s), the fluorescence intensities of the protein excited by 295 nm light in the presence of various concentrations of denaturant were measured. As shown in Figure 2C, two characteristic fluorescence intensity changes occurred. As the denaturant concentration increased, the fluorescence intensity increased initially and then decreased. As with 280 nm light excitation, the observed values fit the theoretical curve calculated according to the three-state model well (Fig. 2C, continuous line) and the transition midpoints occurred with 1.5 and 3.0 M GdnHCl. The emission maximum shifted toward higher wavelengths as the denaturant concentration increased (Fig. 2D). The two transitions, with 1.5 and 3.0 M GdnHCl, were also observed in the denaturation curve of the emission maximum. As *T.thermophilus* MutS contains only two tryptophan residues (Trp151 and Trp299, Fig. 4), we concluded that these transitions corresponded to the environmental change of each tryptophan residue. These findings strongly suggest that *T.thermophilus* MutS consists of at least three domains with different stabilities.

Limited proteolysis of *T.thermophilus* MutS

In order to investigate the domain structure further, *T.thermophilus* MutS was subjected to limited proteolysis with two endoproteases with different substrate specificities, nagarse and thermolysin. In the presence of low endoprotease concentrations, a substrate protein would be expected to be cleaved preferentially at the sites that are exposed to the solvent. Such sites often lie within inter-domain peptide linkers. As shown in Figure 3A, electrophoresis of the peptide fragments produced by nagarse digestion resulted in discrete bands on the gel. Bands of ~43, 32 and 15 kDa appeared 20 min after the addition of nagarse, and as more time elapsed, the first two bands disappeared and smaller bands appeared. The major nagarse-digested fragments with molecular masses of 32, 27, 19, 17 and 15 kDa were designated N-1, N-2, N-3, N-4 and N-5, respectively. This result suggests that *T.thermophilus* MutS has some protease-accessible sites, which may be inter-domain linkers.

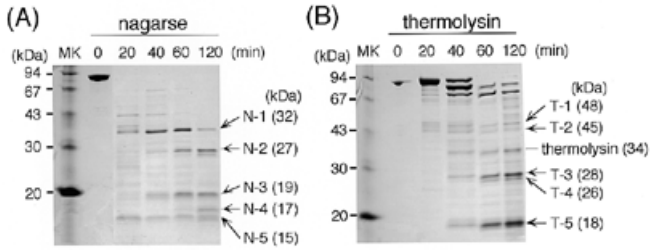

Figure 3. Limited proteolysis of *T.thermophilus* MutS. (**A**) *Thermus thermophilus* MutS (1 µg/ml) was treated with nagarse (protein to protease weight ratio of 100:1) at 37°C. (**B**) *Thermus thermophilus* MutS (1 µg/ml) was treated with thermolysin (protein to protease weight ratio of 100:1) at 60°C. The resolving gel contained 12.5% (w/v) acrylamide and about 2–20 µg protein was loaded into each lane of the gel. Molecular mass markers (MK) are indicated to the left of each panel (rabbit muscle phosphorylase A, 94 kDa; bovine serum albumin, 67 kDa; hen egg white ovalbumin, 43 kDa; bovine erythrocyte carbonic anhydrase, 30 kDa; soybean trypsin inhibitor, 20 kDa).

If inter-domain linkers are digested with nagarse, another protease should yield similar results. In order to verify this, thermolysin digestion was carried out, which yielded similar discrete bands (Fig. 3B). The molecular masses of the major thermolysin-digested fragments were 48, 45, 28, 26 and 18 kDa and they were designated T-1, T-2, T-3, T-4 and T-5, respectively.

In order to identify the cleavage sites, the N-terminal amino acid sequences of each band were determined. The results of limited proteolysis are summarized in Table 1 and Figure 4. The N-1 and N-3 fragments yielded by nagarse digestion started at Arg275. Their C-terminal amino acid residues were estimated from their apparent molecular masses, determined from their relative mobilities on the gel, and found to be 566 and 448, respectively. The N-2, N-4 and N-5 fragments started at Val321, Leu134 and Met9, respectively, and were estimated to end at 566, 288 and 140, respectively. The T-1 and T-2 fragments yielded by thermolysin digestion both started at Leu13 and their molecular masses indicated their C-terminals ended at positions of 449 and 422, respectively. The T-3 fragment contained two peptide fragments, one starting at Phe320 and the other at Val321, both of which were estimated to end at 575. The T-4 fragment, which also contained two peptide fragments, starting at Leu333 and Pro354. As its molecular mass was slightly smaller than that of the T-3 fragment, the T-4 fragment may have been a partially degraded product of the T-3 fragment. The T-5 fragment started at Leu134 and was estimated to end at 257.

Table 1. Peptides obtained by partial digestion with proteases

Fragment	Apparent molecular mass (kDa)	Determined N-terminal amino acid sequence
N-1	32	275-RGQDTLFGVL
N-2	27	321-VREGALREGV
N-3	19	275-RGQDTLFGVL
N-4	17	134-LLPREANYLA
N-5	15	9-MEGMLKGEGP
T-1	48	13-LKGEGPGPL
T-2	45	13-LKGEGPGPL
T-3	28	320-FVREGAL
		321-VREGALE
T-4	26	333-LLFRLADL
		354-PRDLAALR
T-5	18	134-LLPREANYL

Peptides N-1 to N-5 were obtained by nagarse digestion, and peptides T-1 to T-5 by thermolysin.

As summarized in Figure 4, cleavage by both nagarse and thermolysin occurred around some specific sites of *T.thermophilus* MutS. The N-terminal fragment, N-5, ended at about residue 140 and both the N-4 and T-5 fragments started from residue 134, suggesting the presence of an inter-domain linker near residue 130. The region around residue 275 is the second candidate for an inter-domain linker, because the N-1, N-2, N-3, T-3 and T-4 fragments started from or near residue 275 and the C-termini of the N-4 and T-5 fragments were estimated to end around residue 275. The third candidate for an inter-domain linker site lies near residue 570, as the C-termini of the N-1, N-2, T-3 and T-4 fragments were estimated to end near this residue.

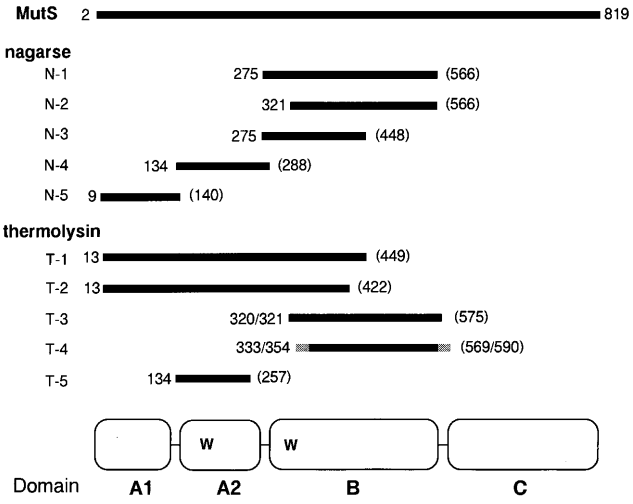

Figure 4. Schematic representation of the fragmentation of *T.thermophilus* MutS. The N- and C-terminal amino acid residues of the fragments were determined using a protein sequencer and estimated from their SDS–PAGE mobilities, respectively. The structural domains suggested by the results of limited proteolysis are shown at the bottom and W indicates the positions of the tryptophan residues at positions 151 and 299 (10).

The substrate specificity of nagarse is broad (22,23), whereas thermolysin acts on amino groups of hydrophobic amino acid residues (24). Our results indicate that cleavage by both nagarse and thermolysin occurred at some specific linker regions of *T.thermophilus* MutS. The presence of these preferential cleavage sites suggests that *T.thermophilus* MutS is organized into at least four domains: the N-terminus to residue 130, residues 131–274, residues 275–570 and residue 571 to the C-terminus. Hereafter, these domains are referred to as the A1, A2, B and C, respectively. It should be emphasized that nagarse and thermolysin, which are relatively non-specific proteases, produced only a limited number of fragments.

DNA binding

The binding of *T.thermophilus* MutS to DNA was analyzed by carrying out native PAGE at 5°C. As shown in Figure 5, the addition of dsDNA, poly(dI)·(dC), resulted in upward smearing of the *T.thermophilus* MutS band, reflecting reduced mobility, indicating that this protein can bind to dsDNA. However, ssDNA, poly(dC), did not result in such a mobility reduction. These observations suggest that *T.thermophilus* MutS can interact with dsDNA, but not with ssDNA.

In order to establish which domain interacts with DNA, the nagarse-digested peptide fragments N-1 to N-5 were mixed with DNA and then subjected to native PAGE. In the absence of DNA, these peptide fragments separated to form discrete bands on the gel (Fig. 5, lane 4). Mixing the fragments with dsDNA, but not with ssDNA, resulted in reduced densities of two of the bands (Fig. 5, lane 5). The mobilities of intact MutS and these two fragments were reduced by dsDNA but not by ssDNA. Therefore, like intact MutS, the peptide fragments contained in the two bands bound specifically to dsDNA, but not to ssDNA. In order to

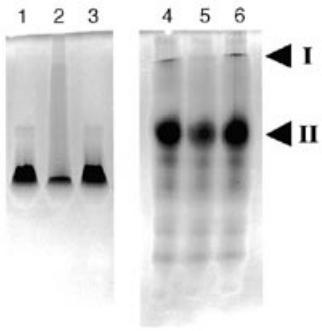

Figure 5. Binding of *T.thermophilus* MutS and its fragments to DNA. Aliquots (5 µM) of *T.thermophilus* MutS (lanes 1–3) or its proteolytic fragments (lanes 4–6) were incubated with 200 µM substrate DNA on ice for 30 min, electrophoresed on a 5% polyacrylamide gel at 5°C under non-denaturing conditions and then stained with CBB. Lanes 1 and 4 contain no DNA, lanes 2 and 5 contain dsDNA [poly(dI)·(dC)] and lanes 3 and 6 contain ssDNA [poly(dC)]. The arrows I and II indicate the bands that showed reduced densities in the presence of dsDNA.

identify which peptide fragment(s) interacted with dsDNA, the N-terminal amino acid sequences of the two bands were determined and their sequences were compared with those of the nagarse-digested fragments. The band with the lowest mobility (Fig. 5, band A) contained the fragment that started at Arg275 and the band showing intermediate mobility (Fig. 5, band B) contained two peptide fragments that started at Ala133 and Val321. As summarized in Table 2, we concluded that band I contained the N-1 or N-3 fragment and band II contained N-2 and N-4. The first alanine of the N-4 fragment was missing when its N-terminus was determined by SDS–PAGE analysis (Table 1), because boiling the sample with SDS before loading the gel removed it. The N-1, N-2 and N-3 fragments corresponded to the B domain and the N-4 fragment corresponded to the A2 domain (Fig. 4). These results suggest that the B domain of *T.thermophilus* MutS interacts with dsDNA and the A2 domain may also do so.

Table 2. The peptide fragments which interact with dsDNA

Band	Determined N-terminal amino acid sequence	Identified fragment
I	275-RGQDTLF	N-1 or N-3
II	321-VREGAL	N-2
	133-ALLPREAN	N-4

DISCUSSION

Denaturation with GdnHCl and limited proteolysis were found to be useful for studying the domain structure of *T.thermophilus* MutS. The results of limited proteolysis suggested that *T.thermophilus* MutS is organized into at least four domains (Figs 3 and 4). The discrete fragmentation of this protein is characteristic of thermophilic proteins, such as UvrB protein (15), whereas mesophilic proteins often yield bands tailing toward a lower molecular mass. Up to now, it was not clear whether MutS has a domain structure. The denaturation curves of *T.thermophilus* MutS we obtained suggested that this protein has at least three domains (Figs 1 and 2). Domain organization can also be verified from the amino acid sequence conservation pattern. Comparison of *T.thermophilus* MutS with known MutS homologues revealed that the region around the putative linker between the A1 and A2 domains (at around residue 130) has a large gap (10). The second putative linker region between the A2 and B domains (at around residue 275) was also observed to have a large gap, whereas that of the third putative linker region between the B and C domains (at around residue 570) was small. As the C-terminal region following the third linker (C domain) is a highly conserved region between prokaryotic and eukaryotic MutS homologues and the remaining N-terminal region (A1, A2 and B domains) was less highly conserved (10), it would appear that MutS can be roughly divided into two domains.

The proteolytic peptide fragment corresponding to the C domain was not found after nagarse or thermolysin digestion (Fig. 4), possibly because the C domain becomes unstable when the B–C linker is cut. The C domain may interact with other domains and these interactions may be required for folding of the C domain. Another possible explanation is that the N- and/or C-terminus of the C domain is/are digested by proteases leading to rapid digestion of the domain.

Three and two denaturation midpoints were revealed by far-UV CD (Fig. 1A) and fluorescence excitation with 295 nm light (Fig. 2C and D), respectively. Protein fluorescence in response to 295 nm light is derived from tryptophan residues and as *T.thermophilus* MutS contains only two tryptophan residues (Trp151 and Trp299, Fig. 4), we consider the two transitions shown in Figure 2C corresponded to denaturation of the A2 and B domains. The shift of the emission maximum toward higher wavelengths indicates that the tryptophan residues moved from a hydrophobic to a hydrophilic environment as denaturation progressed (Fig. 2D). Generally, tryptophan fluorescence is quenched in a hydrophilic environment, but the fluorescence intensity increased in the presence of 1.5 M GdnHCl (Fig. 2C), which suggests that one tryptophan residue in native MutS is in a hydrophobic environment but its fluorescence is quenched by an environmental quencher, e.g. a peptide.

From the changes in the ellipticity at 222 nm evoked by the denaturant, the stabilities of the *T.thermophilus* MutS domains in the absence of denaturant were calculated to be 12.3, 22.9 and 30.7 kcal/mol (Figs 1 and 2). The free energy for the most unstable domain, 12.3 kcal/mol, is the same as that for usual proteins, the values of which range from 5 to 15 kcal/mol (25). The most stable domain, however, was more stable than a stable four-helix bundle protein, the stability of which was found to be 22 kcal/mol (26). Therefore, *T.thermophilus* MutS has a very stable domain.

As shown in Figure 5, *T.thermophilus* MutS bound to dsDNA. In a previous study, we showed that this protein bound to dsDNA (10). Recently, the human MutS homologue was found to bind to homoduplex DNA, as well as to mismatched DNA (27). In the light of these observations, we propose the following: *T.thermophilus* MutS has two DNA binding modes, one is non-specific binding to dsDNA and the other mode is specific binding to mispaired bases in dsDNA. The results of native PAGE in Figure 5 suggested that the affinity between *T.thermophilus* MutS and ssDNA is very weak or none. This protein may not be able to recognize directly the ssDNA region formed by base mispairing or loopout.

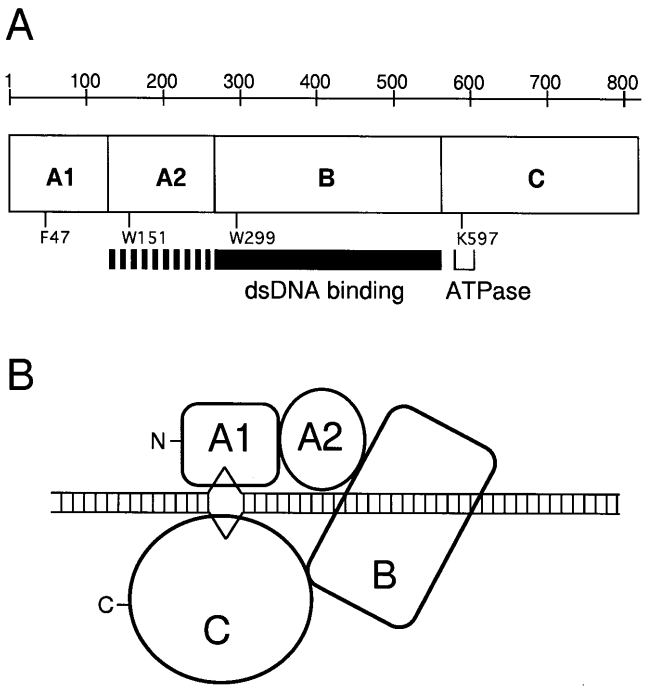

Figure 6. Domain organization of *T.thermophilus* MutS. (**A**) Relationship between its putative domain structure and function. F47 represents a phenylalanine residue that corresponds to Phe39 in *T.aquaticus* MutS and is a candidate for a mismatched DNA binding site (6), K597 represents an essential lysine residue of Walker's A-type nucleotide binding motif (10,28) and W151 and W299 represent the positions of tryptophan residues. (**B**) Schematic model of the domain organization and DNA binding. A1, A2, B and C represent the domains suggested by the results of this study.

The band shift assay results indicate that at least the B domain interacts with dsDNA (Fig. 6A), as may the A2 domain adjacent to it. Haber and Walker (8) reported that a C-terminal mutant of *S.typhimurium* MutS, in which Walker's A-type motif (28) had been destroyed, was able to bind to heteroduplex DNA with reduced affinity in comparison with the native protein. Whitehouse *et al*. (7) reported that the C-terminal region of the human MutS homologue, which corresponds to the C domain of *T.thermophilus* MutS, was required for both ATPase activity and binding to mismatched heteroduplex DNA. These findings suggest that the C-terminal region is important for mismatched DNA recognition and involved in protein–DNA interactions. However, Malkov *et al*. (6) reported that Phe39 at the N-terminal region of *T.aquaticus* MutS was affinity-labeled by one-base looped out duplex DNA, which suggests that the N-terminal region is also involved in mismatched DNA interactions. In the light of these observations, we propose the following model for the interaction between MutS and DNA (Fig. 6B). First, dsDNA is bound non-specifically by the B domain of MutS and then, the mismatched region of dsDNA is recognized by the A1 and C domains. This model is rather simple and we must consider the effects of magnesium ions and adenine nucleotides, because interactions with these cofactors change the DNA binding mode of MutS (27). However, even though the stages of the process of mismatched DNA recognition may be affected by these cofactors, the basic mechanism responsible for the interaction between MutS and DNA would be similar as described above.

MutS homologues form protein complexes with themselves and with MutL homologues. The three-dimensional structure of MutS is unknown and the sites of interaction of MutS with itself and with MutL have not been identified. In this study, we showed that the B domain of MutS is a dsDNA binding domain. The sites of interaction with MutS and MutL may be localized in the other domains, A1, A2 and C. In order to verify this hypothesis, we are trying to isolate each individual domain and studies on these fragments should yield important information about the structure–function relationships of MutS.

ACKNOWLEDGEMENT

This work was supported by a Grant-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan.

REFERENCES

1 Friedberg,E.C., Walker,G.C. and Siede,W. (1995) *DNA repair and mutagenesis.* ASM Press, Washington, DC.
2 Modrich,P. (1991) *Annu. Rev. Genet.*, **25**, 229–253.
3 Kolodoner,R. (1996) *Genes Dev.*, **10**, 1433–1442.
4 Fishel,R. and Wilson,T. (1997) *Curr. Opin. Genet. Dev.*, **7**, 105–113.
5 Allen,D.J., Makhov,A., Grilley,M., Taylor,J., Thresher,R., Modrich,P. and Griffith,J.D. (1997) *EMBO J.*, **16**, 4467–4476.
6 Malkov,V.A., Biswas,I., Camerini-Otero,R.D. and Hsieh,P. (1997) *J. Biol. Chem.*, **272**, 23811–23817.
7 Whitehouse,A., Deeble,J., Taylor,G.R., Guillou,P.J., Phillips,S.E.V., Meredith,D.M. and Markham,A.F. (1997) *Biochem. Biophys. Res. Comm.*, **232**, 10–13.
8 Haber,L.T. and Walker,G.C. (1991) *EMBO J.*, **10**, 2707–2715.
9 Oshima,T. and Imahori,K. (1974) *Int. J. Syst. Bacteriol.*, **24**, 102–112.
10 Takamatsu,S., Kato,R. and Kuramitsu,S. (1996) *Nucleic Acids Res.*, **24**, 640–647.
11 Doolittle,R.F. (1995) *Annu. Rev. Biochem.*, **64**, 287–314.
12 Kato,R. and Kuramitsu,S. (1993) *J. Biochem.*, **114**, 926–929.
13 Yamamoto,N., Kato,R. and Kuramitsu,S. (1996) *Gene*, **171**, 103–106.
14 Kato,R., Yamamoto,N., Kito,K. and Kuramitsu,S. (1996) *J. Biol. Chem.*, **271**, 9612–9618.
15 Nakagawa,N., Masui,R., Kato,R. and Kuramitsu,S. (1997) *J. Biol. Chem.*, **272**, 22703–22713.
16 Kato,R., Hasegawa,K., Hidaka,Y., Kuramitsu,S. and Hoshino,T. (1997) *J. Bacteriol.*, **179**, 6499–6503.
17 Hiramatsu,Y., Kato,R., Kawaguchi,S. and Kuramitsu,S. (1997) *Gene*, **199**, 77–82.
18 Mikawa,T., Kato,R., Sugahara,M. and Kuramitsu,S. (1998) *Nucleic Acids Res.*, **26**, 903–910.
19 Laemmli,U.K. and Favre,M. (1973) *J. Mol. Biol.*, **80**, 575–599.
20 Matsudaira,P. (1987) *J. Biol. Chem.*, **262**, 10035–10038.
21 Davis,B.J. (1964) *Ann. N. Y. Acad. Sci.*, **121**, 404–427.
22 Canfield,R.E. (1963) *J. Biol. Chem.*, **238**, 2698–2707.
23 Steinman,H.M., Naik,V.R., Abernethy,J.L. and Hill,R.L. (1974) *J. Biol. Chem.*, **249**, 7326–7338.
24 Matsubara,H., Sasaki,R.M., Singer,A. and Jukes,T.H. (1966) *Arch. Biochem. Biophys.*, **115**, 324–331.
25 Privalov,O.L. (1979) *Stability of Proteins.* In Anfinsen,C.B., Edsall,J.T. and Richards,F.M. (eds), *Advances in Protein Chemistry.* Academic Press, NY, Vol. 33, pp. 167–241.
26 Regan,L. and DeGrado,W.F. (1988) *Science*, **241**, 476–478.
27 Gradia,S., Acharya,S. and Fishel,R. (1997) *Cell*, **91**, 995–1005.
28 Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945–951.