UmuD and UmuD’ Proteins

**Databanks**

**MEROPS name:** UmuD protein  
**MEROPS classification:** clan SF, family S24, peptidase S24.003  
**Tertiary structure:** Available  
**Species distribution:** superkingdoms Eukaryota, Bacteria  
**Reference sequence from:** Escherichia coli (UniProt: P0AG11)

**Name and History**

The *Escherichia coli* *umuC* locus was identified in the late 1970s in genetic screens for strains that were UV-nonmutable [1]. In their manuscript, Kato and Shinoura reported the identification of three discrete loci that they called *umuA*, *umuB* and *umuC*. Mapping studies indicated that *umuA* and *umuB* mutants were likely to reside in *lexA* and *recA* respectively, while *umuC* appeared to be a novel locus involved in damage-induced mutagenesis [1]. Shortly thereafter, in an independent study, Steinborn similarly isolated mutants of *E. coli* that were also nonmutable after exposure to UV-light, which he called *uvm* (for UV-mutagenesis) [2]. The use of the *uvm* name was, however, dropped in the early 1980s when it was realized that the locus is allelic with *umuC*. Cloning and sequencing of the *umu* locus revealed that instead of encoding a single gene, it in fact consists of a two-gene operon that is regulated by the transcriptional repressor LexA [3–7]. The larger of the two genes was called *umuC*, while the smaller gene was called *umuD*. The *umuD* gene is located upstream of *umuC* with the TGA stop codon of *umuD* and the ATG start codon of *umuC* overlapping by one base-pair [6,7]. Perry *et al.* also noticed that the smaller *umuD* gene encodes a protein with similarity to the C-terminal domain of LexA, including conserved cleavage site and active site residues, suggesting that UmuD may undergo post-translational processing [7]. Indeed, in
1988, Shinagawa et al. and Burckhardt et al. demonstrated that the 15 kDa UmuD protein undergoes both RecA-mediated cleavage and autocatalysis at high pH to generate peptides of ~3 kDa and ~12 kDa [8,9]. At the same time, Nohmi et al. reported that unlike LexA, which is inactivated for its repressor functions upon proteolysis, the larger of the two UmuD cleavage products, called UmuD', is active and actually required for damage-induced mutagenesis [10]. In the years since their initial discovery, several orthologs of umuDC have been identified, cloned and characterized. Many share the same general operon organization with a umuD-like gene located immediately upstream of a umuC-like gene. Various names have been given to these orthologs including mucAB (mutagenesis, UV and chemical) [7,11]; samAB (Salmonella mutagenesis) [12]; impAB (I-group mutagenesis and protection) [13,14]; rumAB (R-plasmid umu-homolog) [15,16]; and ruLAB (resistance to UV-light) [17].

Activity and Specificity

The ability of *E. coli* UmuD [8,9] and several of its orthologs, including *Salmonella typhimurium* UmuD [18], the R-plasmid-encoded MucA [19,20], or RumA [16] proteins to undergo post-translational cleavage in vitro and in vivo has been reported. In all cases, processing is greatly stimulated by RecA protein. In vivo, moderately efficient cleavage occurs in cells that have been exposed to cellular DNA damage and in which RecA is believed to be in a so-called ‘activated’ filamentous state bound to single-stranded DNA. Constitutive cleavage occurs in *E. coli* strains expressing mutant RecAs that are in the activated state in the absence of exogenous DNA damage [8,21–23]. In vitro, the RecA-mediated reaction requires single-stranded DNA and magnesium for efficient nucleoprotein filament formation. In the absence of RecA, *E. coli* UmuD, MucA and RumA all undergo autoproteolysis at alkaline pH. The rate of autoproteolysis varies considerably between the orthologs with the fastest being MucA (t1/2 at pH 10 ~ 75 min) [20] and the slowest being *E. coli* UmuD (t1/2 at pH 10 > 10 h) [9,20]. Both of which are considerably slower than that of *E. coli* LexA (t1/2 at pH 10 ~ 8 min) [20,24].

The UmuD-like proteins usually exist as dimers in solution [25,26] and it was originally believed that cleavage occurs via an intermolecular reaction in which the substrate tail of one protomer is cleaved in the active site of a dimer mate [18,27]. However, Ollivierre et al. [28] recently reported that a umuD mutant (N41D) which is unable to dimerize, nevertheless undergoes cleavage, indicating that the UmuD cleavage reaction can occur via both intermolecular and intramolecular pathways. While both proteins form homodimers in solution, when UmuD2 and UmuD'2 are mixed together in vitro, they preferentially associate to form UmuD/UmuD' heterodimers [26].

Structural Chemistry

The *E. coli* UmuD protein is 139 amino acids in length and has a calculated molecular mass of 15 063 Da and a pI of 4.5. The 115 residue UmuD' protein has a molecular mass of 12 285 Da and also has a pI of 4.5 [6,7].

The crystal structure of the UmuD2 protein was solved at 2.5 Å resolution revealing an extended N-terminal tail and a globular C-terminal catalytic and dimerization domain containing a mostly beta protein fold [29,30] (Figure 773.1A). NMR spectroscopy has been used to map the UmuD' dimerization interface [31]; to solve the UmuD' solution structure [32] (Figure 773.1B); and to propose a structure for the UmuD/UmuD' heterodimer [32,33].

**FIGURE 773.1**  (A) The crystallographic structure of the UmuD'2 dimer. The protein fold is shown as a ribbon with the side chains of the nucleophilic Ser 60 and general base Lys 97 shown in ball-and-stick. The atomic coordinates 1UMU (pdb code) were used to produce this figure after generating the symmetry-related molecules; (B) The NMR-determined structure of the UmuD'2 dimer. Molecules A and B from the atomic coordinates 1I4V (pdb code) were used to produce this figure.
UmuD utilizes a serine-lysine dyad mechanism. Site-directed mutagenesis [10] and structural studies [30] are consistent with Ser60 serving as the nucleophile and Lys97 the general base. A structural alignment of UmuD with the acyl-enzyme of signal peptidase (another member of the clan SF) suggested an orientation for the cleavage site in the UmuD binding site. These studies also revealed that the nucleophilic Ser60 hydroxyl of UmuD attacks the scissile bond (located between Cys24 Gly25 of UmuD) from the si-face rather than the re-face as seen in most serine proteases [34] (Figure 773.2).

NMR analysis of the UmuD/UmuD′ dimer [32] and crystallographic analysis of the analogous protein LexA [35] are consistent with the proposed substrate orientation. The crystal structure of LexA with its bound cleavage site suggests that the main chain amide hydrogens from Ser60 and Asp59 in UmuD could serve as the oxyanion hole [35] (Figure 773.2B). Crystal structures are now available for members of the clan SF: UmuD′, λ CI repressor, LexA repressor, and signal peptidase [30,35–38]. A superposition of their active sites reveals that the Nζ of the lysine general base is coordinated by three hydrogen bonds. In the case of UmuD′, the neutral ε-amino group of Lys97 (the deprotonated state is a requirement for it serving as the general base) would have two hydrogen bond acceptors (Val96 O and Thr95 Oγ) and one hydrogen bond donor (Ser60 Oγ-H). It is proposed that the pKa of the ε-amino group of Lys97 is depressed by its burial upon binding of the cleavage site in an energetically unfavorable position [35] (Figure 773.2).

The NMR solution structure suggests that the UmuD′2 dimer is structurally dynamic and that Ser60 and Lys97 are not within hydrogen bonding distance in solution [32] (Figure 773.1B). In contrast, all crystal structures of the clan SF proteases so far have shown the nucleophilic serine and general-base lysine to be within hydrogen bonding distance. Ferentz and colleagues [32] suggest that the crystal packing forces may, therefore, result in the stabilization of the catalytically competent conformation and that an interaction with a RecA nucleoprotein filament in vivo facilitates the cleavable conformation.

Other proteases utilizing the serine-lysine catalytic dyad mechanism whose structure have recently been solved include the clan SJ proteases: Lon-A peptidase (family S16, PDB: 1RRE) [39] and birnavirus VP4 protease from blotched snakehead virus (family S50, PDB: 2GEF) [40], infectious pancreatic necrosis virus (family S50, PDB: 2PNL, 2PNM) [41] and tellina virus 1 (family S69, PDB: 3P06) [42]. Clan SK proteases that utilize the serine-lysine dyad mechanism include: C-terminal processing peptidase-1 (family S41, PDB: 1FC6) [43], and bacterial signal peptide peptidase A (family S49, PDB: 3BF0, 3BEZ) [44].

**Preparation**

The UmuD protein was initially overexpressed in *E. coli* from a temperature-inducible λP_L promoter [9]. Homodimeric UmuD′ can be purified from the same UmuD overproducing strain if the cells are also exposed to the DNA-damaging agent, mitomycin C, so as to promote in vivo conversion of UmuD to UmuD′ [25]. Both UmuD and recombinant UmuD′ are now routinely expressed from an IPTG-inducible T7 promoter [31,45]. Purification is relatively simple and involves ammonium
sulfate precipitation, ion-exchange and gel-filtration chromatography. Under these conditions, up to 10 mg of highly purified UmuD or UmuD’ protein can be isolated from 1 liter of an induced *E. coli* culture.

**Biological Aspects**

Since their discovery, the Umu proteins have been hypothesized to participate in damage-induced mutagenesis. For many years it was believed that the Umu proteins somehow modified the cell’s main replicase, so that it would traverse otherwise replication-blocking lesions. However, in the late 1990s, the *E. coli* UmuC protein was shown to possess intrinsic DNA polymerase activity [46–48] and is now considered one of the founding members of the ‘Y-family of DNA polymerases’ [49]. *In vitro* studies suggest that homodimeric UmuD binds to the UmuC protein [48], to form UmuD₂C, or *E. coli* DNA polymerase V (polV) [25,47,50]. The Y-family polymerases are found in all three kingdoms of life, yet interestingly, UmuD-like orthologs have only been identified in Gram-negative bacteria, their self-transmissible R-plasmids, or bacteriophages. Even more intriguing, is the fact that the PI and N15 bacteriophage orthologs actually encode for a preprocessed UmuD-like protein and do not undergo post-translational cleavage to become biologically active, nor are they associated with a cognate *umuC*-like gene [51]. Together, these observations have led to the suggestion that the UmuD and UmuD’-like proteins may participate in other biochemical pathways unique to Gram-negative bacteria. One such role might be in a ‘cell-cycle’ DNA damage-checkpoint pathway [32,52,53]. Whatever their role(s) in addition to translesion replication, it is clear that *E. coli* has gone to great lengths to minimize the cellular concentrations of both the UmuD and UmuD’ proteins in *vivo* [21]. In addition to being tightly regulated at the transcriptional level by LexA, UmuD protein is rapidly degraded by the Lon protease [54,55]. Some molecules of UmuD that escape Lon-mediated proteolysis are nevertheless converted to UmuD’ upon cellular DNA damage. But instead of forming homodimers, which are resistant to proteolysis [54,56], the UmuD’ protomers preferentially associate with intact UmuD to form a UmuD/UmuD’ heterodimer, where the UmuD’ protomer becomes a substrate for another serine protease, ClpXP [54,57]. Significant levels of UmuD₂ are only formed when the cell is exposed to high levels of DNA damage and as a consequence, error-prone polV-dependent translesion DNA synthesis is only utilized as a last resort to enable cell survival.

**Distinguishing Features**

Polyclonal rabbit antibodies have been produced against both UmuD and UmuD’ [25,54] that can detect endogenous levels of the chromosomally encoded *E. coli* proteins [21,22].

**Related Peptidases**

The C-terminal proteolytic/dimerization domain of the UmuD protein (residues 50–136) has sequence and structural similarity to the proteolytic/dimerization domain of the large family (family S24) of λ CI [37] and LexA-like repressors [35]. It is also structurally related to the central catalytic domain of bacterial signal peptidase (family S26; [34,36,58]). The families S24 and S26 both belong to the clan SF of serine proteases.

**Acknowledgments**

Funding for this review was provided, in part, by the NICHD/NIH Intramural Research Program to R.W., the Canadian Institute of Health Research, the Michael Smith Foundation for Health Research and the National Science and Engineering Research Council of Canada to M.P.

**Further Reading**

A structural analysis and comparison of the Ser/Lys protease has been previously reported [58]. For a general discussion on serine-lysine proteases see Paetzel & Dalbey [59]. For a recent review of serine proteases utilizing unconventional catalytic mechanisms see Ekici et al. [60]. A recent review on UmuD and its role in the SOS response can be found in Ollivierre et al. [61]. There have also been several recent reviews on the cellular functions of Y-family DNA polymerases; see Jarosz et al. [62], Fuchs et al. [63], Yang & Woodgate [64], and Pata [65] (and additional references therein).

**References**


Mark Paetzel

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada. Email: mpaetzel@sfu.ca

Roger Woodgate

Laboratory of Genomic Integrity, National Institute of Child Health and Human Development, National Institutes of Health, 9800 Medical Center Drive, Bethesda, MD 20892-3371 USA. Email: woodgate@nih.gov