Regulation of *E. coli* SOS Mutagenesis
by Dimeric Intrinsically Disordered *umuD* Gene Products

by

Sharotka M. Simon

B.A., Biochemistry (1999)

Rice University

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology
at the

Massachusetts Institute of Technology

June 2007

Signature of Author _____
Department of Biology
May 25, 2007

Certified by _____
Graham C. Walker
American Cancer Society Professor of Biology
Thesis Supervisor

Accepted by _____
Stephen P. Bell
Chair, Committee on Graduate Students
Department of Biology

REGULATION OF *E. coli* SOS MUTAGENESIS

BY DIMERIC INTRINSICALLY DISORDERED *umuD* GENE PRODUCTS

by

SHAROTKA SIMON

Submitted to the Department of Biology
on May 24, 2007 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

Products of the *umuD* gene in *E. coli* are involved in regulating the timing of error-free DNA repair processes and mutagenic translesion DNA synthesis (TLS) during the SOS response to DNA damage. Homodimeric $UmuD_2$ is upregulated early during the SOS response, and a slow post-translational autocleavage process removes the N-terminal 24 amino acids of each UmuD monomer. The remaining C-terminal fragment, $UmuD'_2$, activates the catalytic subunit of the Y family DNA polymerase V for mutagenic TLS. The small proteins $UmuD_2$ and $UmuD'_2$ make a large number of specific protein-protein contacts for their roles in regulation. This report chronicles experiments that indicate that *umuD* gene products share characteristics with intrinsically disordered proteins, which lack a defined secondary or tertiary structure when purified and many of which have important roles in regulation. Counterintuitively, $UmuD_2$ and $UmuD'_2$ form stable homodimers *in vitro* at concentrations where little or no α helix or β sheet is detectable. High protein concentrations and certain crowding agents can confer more typical secondary structure on $UmuD_2$ and $UmuD'_2$. The binding affinities between $UmuD_2$ and two of its interaction partners are reported; interaction with either of these interaction partners also confers secondary structure on $UmuD_2$. Intrinsic disorder in *umuD* gene products helps explain how they can make a vast number of specific protein-protein interactions despite their small size, and previous single-cysteine studies of both $UmuD_2$ and $UmuD'_2$ provide insights into the actual structures of intrinsically disordered proteins. A model is presented for how *umuD* gene products and certain hub proteins, which form highly connected nodes in protein-protein interactomes, can make sequential protein-protein interactions.

For Mike Simon

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would first and foremost like to thank my advisor, Graham Walker, for taking the risk in allowing me to work on an interesting project using a technical strategy that was not a focus of the lab at the time. I learned a great deal more than I thought possible, and he was always available for advice, refocusing, and a song. I could not imagine a better Ph.D. advisor.

Graham found a technically and politically brilliant paper [1] and contacted the first author, Ronaldo Mohana-Borges, who has acted as my collaborator throughout the Ph.D. process. The only reason I could possibly manage a biophysical Ph.D. in a genetics lab is because Ronaldo was always available via email to patiently answer my questions (no matter how trivial), to explain techniques to me before I went off to borrow another lab's equipment, and to constructively critique my data. I was very happy to finally get to meet Ronaldo face-to-face in Boston in April 2006, and one of these days I will find an excuse to visit him and his lab in Rio.

In addition to finding just the right collaborator, one of Graham's greatest strengths is his ability to recruit wonderful people into the lab. Mary Ellen Wiltrout took over my group job as I started to focus heavily on finishing, and she has the computers running much more efficiently than I ever did. Daniel Jarosz was always good for a laugh, a quote, and a good discussion about chemistry or wine. Daniel also showed me the ins and outs of SigmaPlot and wrote the vast majority of Chapter 3. Kay Jones needs to be acknowledged twice, because not only did she make sure I had the proper controls and could explain my work so that non-biochemists would understand, she was far and away the best baymate that anyone has ever had. Kay Jones: long live her Diet Coke. Rachel Woodruff kept me company and kept me sane as we both pioneered a new focus in the lab. She has single-handedly fixed my mental health multiple times. Laurie Waters showed me through the graduation process and was my biggest cheerleader when nothing was going well. Jamie Foti joined the lab as I was leaving, and I was grateful for the chance to show someone what I've learned rather than to constantly be the one asking questions. Brenda Minesinger was always available to help with anything, either scientific or otherwise. She saved some of my data through proper statistics. Susan Cohen is always good for a story, whether it is she or I who tells it. Sanjay D'Souza greeted me enthusiastically every morning, even if that was the afternoon. Bryan Davies was a great co-GRT whose Coomassie stain saved my gel at the last

minute. He also generously took my luggage to lab before I went to the airport. Lyle Simmons was always good-natured when people confused us for each other, and his feedback all along the way was incredibly helpful. Michi Taga is fearlessly bold in everything she does, whether scientific or personal, and I learned a great deal about how to be an effective researcher from her. Hajime Kobayashi is one of the most cheerful and generous people one can ever meet. Katherine Gibson taught me about politics, genetics, and the gym. Rachel Erlich may not have officially been in the lab, but she happily visited quite frequently, and I am grateful for many discussions with her about science, literature, and the rest of life. Marianne White and Judy Carlin bent over backwards to make my life easier in countless ways, administrative and personal.

Former lab members have also helped me throughout graduate school. If I had to mention one reason why I joined the lab, it would be Penny Beuning. She made the lab's research sound irresistible when I was choosing labs, constantly reminded me of the light at the end of the tunnel, and fought for me when I was too timid to do so myself. She was a fantastic collaborator on two papers and wrote the vast majority of Chapter 2. Gordon Campbell always made me laugh and taught me the value of keeping a positive attitude. Veronica Godoy knew the answer to every question I ever had, and she was always good-natured about taking my DNA sequencing to TUCF for me. Brett Pellock and Brad Smith encouraged me along the early days of my project and continued to answer questions from afar. And Mark Sutton single-handedly showed me the ropes of the Walker Lab before leaving to start up his own. Lise Barra left soon after I arrived, but she always made the world a better place when she was around.

Most of what I learned about biochemical technique I learned from neighboring labs at MIT. Amy Keating and her lab were conveniently located and full of expertise, so I often came to them for use of the FPLC and HPLC as well as for general technical advice. Amy served on my thesis committee since she began her faculty position and is the only member who came to all of my committee meetings. I always appreciated her perspective on my work. Christy Taylor offered not only long-standing friendship, but she had fixed almost every piece of equipment I used and was very generous in showing or reminding me how to use either of the two HPLCs. Kelly Elkins was an invaluable source of information and inspiration. Emiko Barre helped me with FPLC questions. Nora Zizlsperger trained me on the CD spec. Gevorg Gregarian was a

great colleague and joined me at the Chemical and Biophysics Symposium in Toronto in April 2006. And Sean Daigan fixed the FPLC computer every time it crashed.

Tania Baker and her lab have been valued friends to Walker Lab members since before anyone can remember. Tania had been on my thesis committee since I joined the Walker lab, and every time we spoke, she gave the impression that nothing was more important to her than my scientific development. Julia Flynn and Samia Siddiqui were wonderful friends who were always available for coffee or to let me use the fluorimager. Saskia Neher and I started working on *umuD* gene products together, and she and I did the first UmuD and UmuD′ growths and preps together. She also troubleshot *in vitro* ClpXP degradation of UmuD′ and let me have an aliquot of ClpXP whenever I did the assay. Igor Levchenko walked me through use of the fermentor twice and is constantly a cheerful presence. And Peter Chien showed me the fluorimeter one Saturday when the Sauer Lab was locked. Talking to Peter always made me more excited about my project; his enthusiasm is contagious! Mihaela Pruteanu knew my project inside and out, which flattered me immensely and simultaneously motivated me to finish and to keep doing experiments that would delay finishing. Anne Meyer, Aliaa Abdelhakim, and Igor were also very generous with their ClpXP.

Jon King began to chair my thesis committee only when my project began to take a protein folding turn, but he was always one of the first people I asked for help and advice. He wrote my letter of recommendation for my GRT position at Burton 4 and always gave me a straightforward scientific perspective that I could trust. I did many of my first protein preps and all diagnostic gel filtration experiments in his lab. Peter Weigle and Melissa Kosinski-Collins initially showed me how to use the Akta FPLC. Shannon Flaugh and Ishara Mills continued to remind me of the intricacies of the FPLC and discussed with me the implications of some confusing results.

Bob Sauer has not been on my thesis committee since I began in the Walker lab, but I did almost all fluorescence experiments in his lab. Shilpa Joshi first showed me how to use the fluorimeter and explained the concepts of anisotropy and polarization. Joe Yakamavich and Greg Hersch reminded me of fluorimeter procedures and provided music and entertainment while I worked. Greg only suggested once that the Walker lab might profit from its own fluorimeter. Jon Kenniston and David Wah helped me analyze results. Andreas Martin showed me how to use the refractometer. Brent Cezairliyan showed me how to analyze AUC results

using SedNTerp and WinNonLin, and he was cheerful about reminding me of what I forgot. Tony Anderson is thanked for an extremely selfless act that shall go unnamed. And Bijou Bose is all-around awesome (see Randi Rotjan's thesis at Tufts University for details).

Steve Bell and his lab were wonderful neighbors, letting me pet their dogs (thanks Steve, Hannah Blitzblau, and Laura Francis) and hold their children (thanks Laura again; QT has great parents). I borrowed FPLC equipment (thanks John Randell), chemicals we were missing (thanks John again and Andy Fan), and used the refractometer and the Typhoon scanner almost daily for a while (thanks John again and Shuyan Chen). And Milan de Vries is always great for a chat if we happen to be at lab at the same time!

I got a lot of help from people outside the Department of Biology. Debby Pheasant runs the MIT Biophysical Instrumentation Facility and showed me how to use the CD, the AUC, and dynamic light scattering. Hector Hernandez has been a good friend throughout graduate school, and if he did not know the answer to one of my questions, he certainly would know the person who did. One of these people was Carlos Bosques, who helped make sense of my CD results.

I had many helpful discussions with labs outside of MIT. Bill DeGrado told me about the HPV protein E7 and how it also appeared to be both dimeric and intrinsically disordered; my results incited much less skepticism after I started to cite that paper [2]. Kim Geise in the John Little lab suggested that I take CD spectra of cross-linked $UmuD_2$. Raquel Leiberman in the Greg Petsko/Dagmar Ringe lab suggested Ferguson plots and gave me technical advice on how to generate them. Nathan West in the Penny Beuning lab generously typed Appendix II for me so all I had to do was cut and paste. Cleo and Paul Schimmel sponsored a four-year fellowship that provided the flexibility to choose any lab I wished. Seiichi Matsuda at Rice University provided my first independent research experience; he and my immediate supervisor Jennifer Herrera inspired me to continue research as a career and were the best advisors an undergraduate could hope for.

Mitch McVey from Tufts University generously sat as the external member of my thesis committee. He asked such great questions after my Boston Area DNA Repair and Mutagenesis Meeting presentation that I could not resist asking him to be my external committee member.

And finally, for my external support. The Burton 4 Players and the rest of Burton-Conner house provided a wonderful place to live, work, and play. When big brothers were being handed out, I must have been first in line, because Bruce Godzina is certainly the best. I didn't

start out with a little sister, but finally finding one in Pam Simon was more than worth the wait. Mom and Dad Godzina always encouraged me to follow my dreams. Mom and Dad Simon don't even count as in-laws because they're too cool. And Mike Simon has more than surpassed expectations; he made thesis writing and the rest of life much more rewarding for me.

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Mechanisms of Mutagnesis

Cells of all organisms and in all environments encounter DNA damaging agents (for a comprehensive review on mechanisms and cellular responses to DNA damage, see reference [3]). In addition to environmental DNA damaging agents such as UV irradiation [4], ionizing radiation [5], and certain chemical environmental factors [6], cells produce DNA damaging agents through their own internal metabolic processes. Reactive oxygen species, alkylating agents, and other free radicals are byproducts of normal metabolism and can react with and damage cellular DNA [7]. Spontaneous DNA damage such as depurination also occurs at a significant frequency in the cell [8].

Much of this DNA damage is repaired through specialized repair pathways such as base excision repair [9], nucleotide excision repair [10, 11], or direct reversal of damage such as through photolyases [12]. However, on occasions when DNA damage is extensive or when damage is encountered during DNA replication, processes of DNA damage tolerance can occur. Tolerance allows DNA replication to continue despite the presence of a lesion. Tolerance differs from repair in that the lesion is still present in the parental DNA after tolerance mechanisms are completed [3]. DNA damage tolerance pathways include homologous recombination [13, 14] and translesion DNA synthesis (TLS). TLS involves the active use of specialized DNA polymerases that can use damaged or noncanonical DNA bases as a template for synthesis [15, 16]. When the normal replicative DNA polymerase encounters a lesion, it is more likely to stall or reprime downstream of the lesion than to success in replicating the damage DNA itself [17]. The current models of translesion DNA synthesis involve a polymerase switch from the replicative to the translesion DNA polymerase and back again [18] or recruitment of the TLS polymerase to fill in gaps left behind by the replicative DNA polymerase repriming downstream of the lesion [19, 20].

There are currently 5 known classes of DNA polymerases [21-23]. The earliest ones discovered are from the A family, represented by bacterial DNA polymerase I (Pol I) [24]. This polymerase has a proofreading subunit and fills in gaps in the DNA left by excision repair

processes. In eukaryotes, the A family is present partially as mitochondrial DNA polymerases, which are involved in the replication and repair of the mitochondrial genome [25]. Additionally, the nuclear proteins DNA polymerase θ, which appears to be involved in somatic hypermutation, and polymerase ν, whose function is unknown, are also in the A family [25]. The B family of DNA polymerases includes bacterial DNA polymerase II (Pol II) and eukaryotic polymerases α, which acts as a primase [26, 27], and polymerases δ and ε, both of which are implicated in both replication and DNA repair [28-31]. DNA Pol II is involved in early replication restart after DNA damage [32]. The C family of DNA polymerases is only represented in bacteria and includes DNA polymerase III (Pol III), the normal replicative DNA polymerase in gram negative bacteria, and the homologous replicative DNA polymerase C (Pol C) in gram positive bacteria [33]. Although there are no homologs of Pol III/Pol C in eukaryotes, these enzymes display structural homology to both B family and A family DNA polymerases, suggesting that domain rearrangements may have led to the separate replicative DNA polymerases for bacteria and eukaryotes, rather than DNA synthesis activity having evolved twice [34]. The X family of DNA polymerases, conversely, is only present in eukaryotes and is involved in base excision repair, somatic hypermutation, and non-homologous end-joining, among other functions [35]. This family of DNA polymerases is part of a larger superfamily of nucleotidyl transferases and has less homology to the A, B, or C families than those have to each other [36].

It is the relatively newly-discovered Y family of DNA polymerases that is most associated with translesion DNA synthesis [15], although other polymerases can also perform TLS i*n vitro* [37]. In fact, the eukaryotic B family DNA polymerase ζ [38, 39] is often grouped with the eukaryotic Y family DNA polymerases κ [40, 41], ι [42], and η [43, 44] due to its TLS activity, despite the lack of structural homology of polymerase ζ to Y family polymerases.

What defines Y family DNA polymerases is not so much their ability to perform TLS, as members of other polymerase families can do so both *in vivo* and *in vitro*, but rather their structures, which feature a relatively open active site [45, 46] compared to other polymerases' more closed structure [47]. Although Y family DNA polymerases share an overall "right-hand" fold with the replicative DNA polymerases, Y family DNA polymerases lack the replicative polymerases' O-helix, which is responsible for a conformational change that allows replicative DNA polymerases to discriminate against non-Watson-Crick base pairs in their active site [48], and some have extra domains, such as the Polymerase Associated Domain (PAD, or little finger

domain) [49, 50]. As a result of these structural differences, the area surrounding the DNA in Y family DNA polymerases is more solvent-exposed, which allows them to accommodate bulky adducts on the template strand [51]. Several lines of evidence suggest that even with their open active site, an open-to closed conformational change is necessary for Y family DNA polymerases to perform TLS, although this conformational change does not result in the high fidelity that replicative polymerases' conformational change effects [52-54]. Y family DNA polymerases generally also have lower processivity on undamaged DNA than replicative DNA polymerases, even in the presence of a processivity factor [55, 56].

The Y family of DNA polymerases includes the almost universally conserved DinB subfamily, the eukaryotic Rev1 and Rad30 subfamilies and the prokaryotic UmuC subfamily, which like the C family of DNA polymerases, has separate Gram negative and Gram positive counterparts [15]. Some Y family DNA polymerases recognize specific cognate lesions and can bypass them accurately [43, 53, 57]; others are more error-prone [58]. However, Y family DNA polymerases generally have lower fidelity on undamaged DNA templates than the normal replicative DNA polymerases [58-60]. In addition to the lack of O-helix, another possible reason for this is that Y family DNA polymerases do not have proofreading subunits, although it is possible that Y family DNA polymerases may use extrinsic accessory subunits to increase their fidelity on undamaged templates [61]. Because of their mutagenic potential, it is generally assumed that the activity of Y-family DNA polymerases must be regulated to prevent their action on undamaged DNA and target them only to the correct DNA lesions.

At best, translesion DNA synthesis preserves genome replication under conditions of severe DNA damage [62]. At worst, it can cause point mutations or deletions [16]. The activities of prokaryotic error-prone DNA polymerases can induce a transient mutator phenotype, increasing the genetic variation of a subset of organisms and thereby increasing the species' overall chances of survival [23, 63]. Multicellular organisms benefit less from a mutagenic bypass system because deleterious mutations have the potential to cause malignant tumors or genetic diseases that can affect the entire organism; unicellular organisms do not get tumors [64]. Eukaryotic cells have more subfamilies of Y family DNA polymerases than prokaryotic cells, presumably to provide specialists for more cognate lesions, thereby causing fewer point mutations [60].

The E. coli *SOS response*

When a bacterial cell encounters extensive DNA damage or other stressors [65] that requires a more stringent response than allowed by normal growth conditions, the cell undergoes what is termed the SOS response [66]. The SOS response is only a subset of the generalized DNA damage response, which involves more than 1,000 genes (about 25% of the bacterial genome) [67]. In *E. coli*, the SOS response involves more than 40 genes (about 1% of the bacterial genome) that are under the control of the LexA repressor [68]. LexA is a dimeric transcription factor that binds to a consensus SOS box sequence (5′-TACTGTATATATATACAGTA-3′) in the promoters of SOS response genes [69]. Among the genes upregulated during the SOS response are DNA repair genes [70, 71], certain genes involved in cell division [68], *polB,* which encodes DNA Pol II [71], and several other genes whose functions are unknown or that are less obviously involved in a response to stress [68, 71, 72]. Also upregulated are TLS genes: *dinB,* encoding the relatively accurate DNA polymerase IV (Pol IV) [71, 73]; *umuC*, encoding the catalytic subunit of DNA polymerase V (Pol V) [74-76]; and *umuD*, which encodes an accessory to both of these Y family DNA polymerases [74-77]. DNA Pol I is expressed constitutively rather than as part of the SOS response, and it has not been implicated in mutagenesis [78].

During cellular growth in the absence of significant DNA damage, LexA binding to the SOS boxes represses the expression of the SOS genes [79]. However, under conditions of extensive DNA damage, stalling of the normal replicative DNA polymerase at sites of damage produces long tracts of single stranded DNA (ssDNA) [67]. This extra ssDNA binds to cellular RecA protein, which forms the RecA nucleoprotein filament (RecA* or RecA:ssDNA) [80]. RecA* binds to LexA and facilitates a conformational change in LexA that causes autocleavage and inactivation of the repressor [81]. This process is similar to λ cI repressor cleavage at the start of lytic cycle of λ phage (extensively reviewed in [82]). A serine-lysine dyad in the C-terminus of these proteins acts as a protease to cleave an Ala-Gly bond between the N- and C-terminal domains [83, 84]. Cleavage results in separation of the LexA and λ N-terminal DNA binding domains from the C-terminal dimerization domains, destroying cooperative DNA binding and releasing repressor from the DNA [85, 86]. In the case of LexA, cleavage exposes

cryptic ClpXP proteolytic sites and results in rapid degradation of the two halves of the protein [87].

A similar cleavage reaction is thought to occur in β-lactamases [88], bacterial signal peptidases [89, 90], and UmuD [91], the subject of this work. No similar mechanism is known in eukaryotes, although the extensively-used phosphorylation reaction may represent a similar autoregulatory scheme with many different proteins undergoing a similar post-translational regulator step [92]. Although processing of LexA and similar proteins is considered "autocatalytic," it is not in fact catalytic, because the products of the reaction do not include the reactant, intact LexA [92]. In the case of UmuD (discussed below), the post-translational cleavage product UmuD′ can act as a true enzyme to cleave intact UmuD [93]. An advantage of autocatalysis is that the local concentration of substrate can be rather high, resulting in extremely fast catalysis when the correct conditions are met [92], and it is therefore unsurprising that both prokaryotes and eukaryotes use autocatalysis as a regulation scheme.

UmuD and similar proteins

Among the 40+ genes upregulated during the SOS response are the *umuDC* operon [68, 76]. This operon encodes DNA Pol V, the Y-family DNA polymerase required for most point mutagenesis in *E. coli* [75]. Members of this operon were found in a screen for cells that were nonmutable after UV irradiation [76]. The *umuDC* operon is functionally similar to the plasmid-encoded *mucAB* operon that is responsible for *Salmonella typhimurium* mutagenesis as used in the Ames test for mutagenic substances [94].

The SOS box for the *umuDC* operon (5′- CTGTATATAAAAACAG-3′) is substantially different from the consensus sequence above [95], with the result that the *umuDC* operon is among the first set of genes upregulated after DNA damage [68]. During normal cellular growth, UmuD is present in about 200 copies per cell, whereas UmuC is undetected [96]. One of the reasons why translation of UmuC is underrepresented relative to that of UmuD, even though they are on the same transcript, is that *umuC* uses rare codons [91]. Interestingly, the stop codon of UmuD and the start codon of UmuC overlap by one nucleotide [91].

UmuC is the catalytic subunit of the Y family DNA polymerase Pol V [75]. The initial product of the *umuD* gene is a 139 amino acid protein [91]. UmuD and UmuC together effect a

primitive DNA damage checkpoint-like function, in which UmuD, together with UmuC, induces a temporary decrease in the rate of DNA synthesis [97]. Accurate DNA repair processes are believed to be active during this time [97]. Overexpression of *umuDC* induces a cold-sensitive phenotype that is thought to be an exacerbation of this DNA damage checkpoint [98]. This cold-sensitive phenotype is partially relieved through mutations in either UmuC or the β processivity subunit that appear to weaken the interaction between these two proteins [99, 100].

UmuD shares considerable homology with the C-terminal domains of LexA and the λ cI repressor, including one cleavage site residue and both of the catalytic serine-lysine dyad residues [91] (Figure 1-1 p.18), although the N-terminal DNA binding domains of LexA and λ cI repressor are missing in UmuD. Based on this homology, it had been correctly predicted that UmuD may be able to undergo RecA*-facilitated autoproteolysis similar to that of LexA [101-103]. In this reaction, the N-terminal 24 amino acids are removed from UmuD, leaving the C-terminal portion, termed UmuD'. It is UmuD' that activates UmuC for its role in TLS [103]; DNA Pol V has been defined as the heterotrimeric complex $UmuD'_2C$ [75].

The mutagenic mechanism of DNA Pol V, in addition to replicating past DNA lesions, also appears to involve untargeted mutations at sites without damage [104-106]. This was experimentally demonstrated by showing that unirradiated phage injected into an irradiated host bacterium incurs greater numbers of point mutations than the same phage introduced into unirradiated bacteria [106]. Additionally, most point mutations caused by constitutive (rather than damage-induced) SOS activation can be corrected by mismatch repair systems, implying that they do not contain a non-coding DNA lesion [105]. Finally, during constitutive SOS response activation, *E. coli* strains that are defective in excision repair are not significantly more susceptible to point mutations than those that are proficient, again suggesting that these sites of mutation do not contain lesions that are excision repair substrates, but rather are the results of simple replication errors [104]. TLS by $UmuD'_2C$ is augmented in the presence of the α and β subunits of DNA Pol III in addition to single-stranded DNA binding protein (SSBP) and RecA [107-109]. The once enigmatic "third role" [103, 110] of RecA in TLS (aside from activating the SOS response and the conversion of $UmuD_2$ to $UmuD'_2$) has recently been discovered [111]. RecA acts in conjunction with *trans* ssDNA to augment TLS activity, perhaps acting as a signal that extensive DNA damage is still present [111].

Cleavage of UmuD is slower than that of LexA or even functional homologs such as MucA (discussed below). It is likely that some of the inefficiency is due to the cysteine in the cleavage site, where proteins that undergo more efficient RecA cleavage have an alanine [91]. If the Cys residue of the UmuD cleavage site is converted to Ala to make a LexA-like cleavage site, RecA* cleavage proceeds more efficiently *in vivo* than for the wild-type protein [112]. Similarly, if the Ala of LexA or MucA is converted to the Cys of UmuD, their cleavage becomes slower [113, 114]. The homologous phage 933W repressor, which also cleaves relatively slowly, has a leucine at this position [115]. Another factor that may slow the cleavage of UmuD to UmuD′ is DinI, which appears to bind to RecA and inhibit UmuD cleavage [116-118]. A similar protein is found on the plasmid-encoded *mpCAB* operon, in which *impA* encodes a homolog of *umuD*, *impB* is similar to *umuC*, and *impC* appears to function as *dinI* [119].

Protein-protein interactions involving umuD *gene products*

In addition to the dimerization as part of the UmuD′$_2$C complex, UmuD and UmuD′ can form homodimers on their own [120] and the thermodynamically most stable UmuD′D heterodimer [121]. However, it was previously unknown how stable these dimers were. Glutaraldehyde cross-linking resulted in both dimeric and monomeric species as visualized by SDS-PAGE electrophoresis [121], and there is disagreement in the literature as to the level of homodimerization of the homologous protein LexA [1, 122, 123]. Although it is now known that the dissociation constant for the homodimer LexA$_2$ is in the pM range [1], the homologous phage λ cI repressor is thought to have a K_d of 20 nM, suggesting that about 5% of the protein is monomeric at physiological concentrations [124]. Evidence is presented in Chapter 4 that indicate that all three *umuD* gene products are dimeric at a wide range of protein concentrations, from nM to mM, although higher order oligomers of both UmuD′ and UmuD have been trapped by cross-linking [100, 125].

The UmuD′D heterodimer is important at the end of the SOS response for its role in targeting the mutagenic UmuD′ for cleavage by ClpXP protease [126]. The UmuD member of the dimer acts as an adaptor protein to target the UmuD′ member for proteolysis; UmuD can be fed into ClpXP for degradation in the same manner if the dimeric complex is UmuD$_2$ [127]. UmuD$_2$ and UmuC are rapidly and continuously degraded by Lon protease [126]. Although both

$UmuD_2$ and $UmuD'_2$ can interact with UmuC *in vitro*, the interaction between the heterodimer UmuD'D and UmuC results in an insoluble complex [128]. There is no known mechanism for degradation of $UmuD'_2$.

In addition to their interactions with UmuC, both $UmuD_2$ and $UmuD'_2$ can specifically interact with a multitude of cellular proteins, many but not all of them components of DNA polymerases. This is particularly surprising given the small size of $UmuD_2$ and $UmuD'_2$ (30 and 25 kDa, respectively), as there may not be enough distinct binding sites on *umuD* gene products to account for all of the interactions. $UmuD_2$ and $UmuD'_2$ have recently been shown to interact with the other *E. coli* Y family DNA polymerase, DinB, and in contrast to the case with UmuC, the UmuD'D heterodimer can also bind to DinB [77]. Overexpression of *umuD* rescues a -1 frameshift phenotype of *dinB* overexpression, suggesting that UmuD plays a role in regulating the activity of DinB [77]. Interestingly, although there are no known homolgs of UmuD in eukaryotes, overexpression of the eukaryotic DinB homolog, Pol κ, also results in a -1 frameshift phenotype [129, 130], suggesting that eukaryotes may also have a regulator of Pol κ that performs this role of UmuD.

Both $UmuD_2$ and $UmuD'_2$ also interact with the α catalytic, β processivity, and ε proofreading subunits of the normal replicative DNA polymerase [131], and with RecA [101-103, 111]. Interestingly, despite their nearly identical primary structure, $UmuD_2$ and $UmuD'_2$ interact differently with different proteins, either to effect a different response ($UmuD_2$ in conjunction with UmuC is associated with a DNA damage checkpoint, whereas $UmuD'_2C$ is a mutageneic translesion DNA polymerase), or with different strengths of interaction. $UmuD_2$ interacts more strongly with the β processivity subunit of DNA Pol III than does $UmuD'_2$, which may cause the DNA damage checkpoint [131]. Conversely, $UmuD'_2$ interacts more strongly with the α processivity subunit of DNA Pol III [131]. It was initially thought that $UmuD'_2$ enabled mutagenesis by modifying the α subunit to become proficient for mutagenic TLS [132]; however, it is now know that $UmuD'_2C$ is a DNA polymerase in its own right [75, 108]. It is not known what the functional consequences of the interaction between $UmuD'_2$ and the Pol III α subunit are, although the "tool belt" model of DNA polymerase switching suggests that many DNA polymerases are kept in close proximity by interactions with β and can be switched as necessary for TLS [18, 133]. In the case of the DinB interaction, $UmuD_2$ interacts more strongly with DinB than does $UmuD'_2$ [77], and it is known that some of the amino acids involved in

DinB binding make up the homodimer interface for $UmuD'_2$ but not for $UmuD_2$ [134]. It seems likely that subtle differences in the conformations between $UmuD_2$ and $UmuD'_2$ may partially explain their different affinities for interaction partners. These replication proteins are not expressed at high levels, and the interactions are quite specific [77, 131]. For example, products of the *umuDC* operon interact with each other, as do products of the homologous *mucAB* operon [91]. However, although *mucB* and *umuC* are 55% identical, UmuD can interact with UmuC but not MucB, and similarly, MucA can interact with MucB but not UmuC [91].

Less well characterized interaction partners for $UmuD_2$ were found in a high-throughput pull-down assay, although it is unknown whether these interactions have relevance *in vivo* [135]. Interaction partners found in this study include: PurE (involved in purine biosynthesis), Fur (ferric uptake regulator), FldB (flavodoxin 2), RpsE (a ribosomal protein), InsM (transposase), YjgM (a putative acetlytransferase), and the hypothetical proteins YbiA and YdfU [135]. Unpublished bacterial two-hybrid results by Charles Kuang suggest that UmuD may interact with NifJ (a putative pyruvate flavodixin oxidoreductase), RseA (an anti-sigma factor), RpoE (sigma factor), TilS (involved in tRNA-Ile production), and the proteins of unknown function YhjG, YhjH, and YbfG. Using the same assay, UmuD′ has been proposed to interact with PotF (putrescine binding protein), PotG (involved in putrescine transport), Adk (has adenylate kinase activity), FruK (fructose-1-phosphate kinase), FruB (diphosphoryl transfer protein), HofF (general secretion), PyrH (uridylate kinase), BasR (transcriptional regulator), DgoT (transport of D-galactonate), GlgP (glycogen phosphorylase), NifJ as for UmuD, the small molecule transport proteins LivF, LivG, LivM, livH, and the hypothetical proteins YheG (putative export protein), YjdB (putative membrane protein), YidS (putative oxidoreductase), YjeP (a putative periplasmic binding protein), YtfE, YtfF (putatively involved in cationic amino acid transport). Experiments are underway to determine the physiological relevance of some of these interaction partners.

The roles of umuD *gene products in regulating TLS*

UmuD is present early during the SOS response, whereas UmuD′ appears later [97]. $UmuD_2$ is associated with a DNA damage checkpoint, during which error-free DNA repair processes have a chance to repair damaged DNA before tolerance mechanisms are active [131]. Conversely, $UmuD'_2$ is associated with mutagenic TLS [103]. Conversion of $UmuD_2$ to $UmuD'_2$

may represent a temporal shift from error-free DNA repair to a more error-prone DNA damage tolerance. In support of this hypothesis, it is the early $UmuD_2$ protein that is associated with the function of the more accurate Y family DNA polymerase DinB [77], and $UmuD'_2$ that is a component of the more error-prone Pol V [75]. DinB is specialized to bypass certain N^2-dG adducts that result from cellular respiration [53] and is present in relatively high copies in the cell even without SOS induction [136]; Pol V is active relatively late during the SOS response [97]. This temporal division between error-free and error-prone TLS may also occur in eukaryotes by partitioning TLS polymerases in a cell-cycle dependent manner [19].

The *umuDC* operon is well represented in bacteria, including the bacterium with the smallest genome discovered to date, *Pelagibacter ubique* [137]. Its UmuD homolog had been classified as MucA, which is functionally and similar and highly homolgous to UmuD (Figure 1-1 p.18), although the *mucAB* and *umuDC* operons have slightly different mutation spectra [138]. It appears that the *umuDC* operon may be important enough to be preserved in a minimalistic, streamlined bacterial genome.

Certain aspects of *umuDC* operon function remain enigmatic. Both $UmuD'_2C$ and DinB have some AP lyase activity *in vitro*, but the purpose of this activity in either polymerase is unknown [139]. Additionally, the SOS response seems to oscillate in a temporal fashion, and the oscillation requires the *umuDC* operon, although the functions of the *umuDC* operon in this effect and the functions of the oscillations are unknown [140].

Structures of $UmuD_2$ and $UmuD'_2$

The NMR and crystal structures of $UmuD'_2$ have been solved, and both methods of structure determination agree that $UmuD'_2$ is a dimeric protein with a C2 axis of symmetry and a mostly β-sheet structure [141, 142]. The proteins have a β sheet structured C-terminal globular domain with unstructured N-terminal arms (amino acids 25 through 39, numbering as in full-length UmuD). The NMR structure shows the correct dimerization interface [142], one that is consistent with chemical cross-linking and mutational data [143-145]. It was impossible to tell by crystallography which of the two possible dimer interfaces was the correct one, and it had been proposed that UmuD′ may form filaments of dimers [141]. Although UmuD′ has been

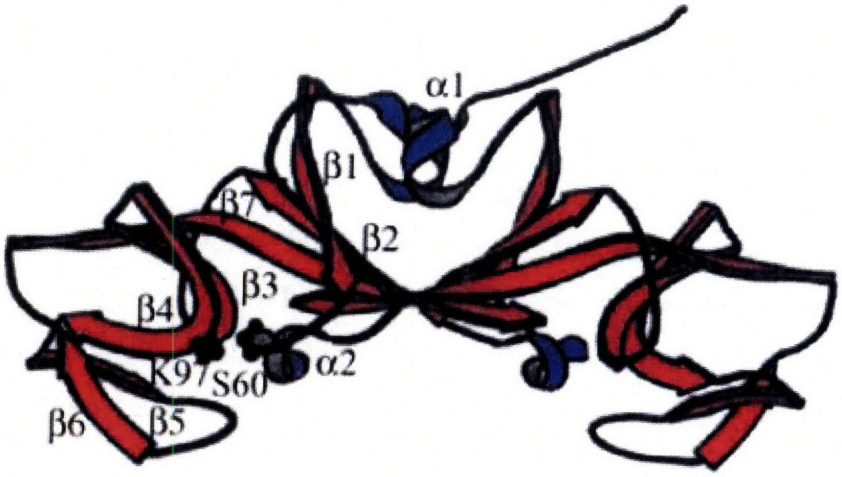
shown to cross-link as multimeric species [125], no evidence of filament formation was found by NMR [125, 142].

Despite their overall similarities, the NMR and crystal structures show several substantial differences in structure [142] (Figure 1-2 p.24). Most telling catalytically, the NMR structure of UmuD′ shows the active site serine and lysine dyad residues pointed away from each other such that their functional groups are more than 8 Å apart [142]. A conformation such as this would not be competent for catalysis. In contrast, the crystal structure shows the catalytic serine and lysine dyad residues pointed towards each other and less than 5 Å apart [141]. In this structure, the active site residues are poised for catalysis. It is possible that crystal packing forces caused $UmuD'_2$ to mimic the conformation adopted by $UmuD_2$ when it is bound to RecA*. Upon inspection, other differences become evident as well. The overall conformation of UmuD′ in the NMR structure is wider than it is tall, while that of the X-ray crystal structure shows a more square overall shape. Although the majority of the secondary structure is a β-sheet in both methods of protein structure determination, the actual conformation of many amino acid residues relative to their neighboring contacts is different between the two proposed protein structures [141, 142]. This observation may give insight into how protein-protein contacts may affect the structures of UmuD and UmuD′, and it served as a first indication that at least UmuD′ may have a flexible structure.

Despite several attempts, the structure of $UmuD_2$ has not been solved, either by NMR or by X-ray crystallography. However, the X-ray crystal structure of the homologous protein LexA has been solved [146], and the C-terminal region of this structure, together with those of the post-translationally modified $UmuD'_2$, have allowed model calculation of full-length $UmuD_2$. Four symmetric energy-minimized structures have been proposed, and they differ in the orientation of their N-terminal arms (see Figure 2-S1, p.40) [147]. It is also possible that the two N-terminal arms of $UmuD_2$ may have different conformations, resulting in asymmetry [147]. The arms may be able to adopt a *cis* conformation such as that found in the crystal structure of LexA, wherein the N-and C-termini of a single monomer remain on the same side of the protein, or a *trans* conformation, wherein the N-terminus contacts the C-terminus of the opposing monomer [147]. There is biochemical evidence that $UmuD_2$ cleavage can occur intermolecularly in the *trans* conformation [93, 148, 149], and although it has been largely accepted that $LexA_2$ cleavage occurs intramolecularly [81, 150], certain deletion mutants of $LexA_2$ have also been

Figure 1-2: **Comparison of the NMR and X-ray crystal structures of UmuD′$_2$.**

A



B

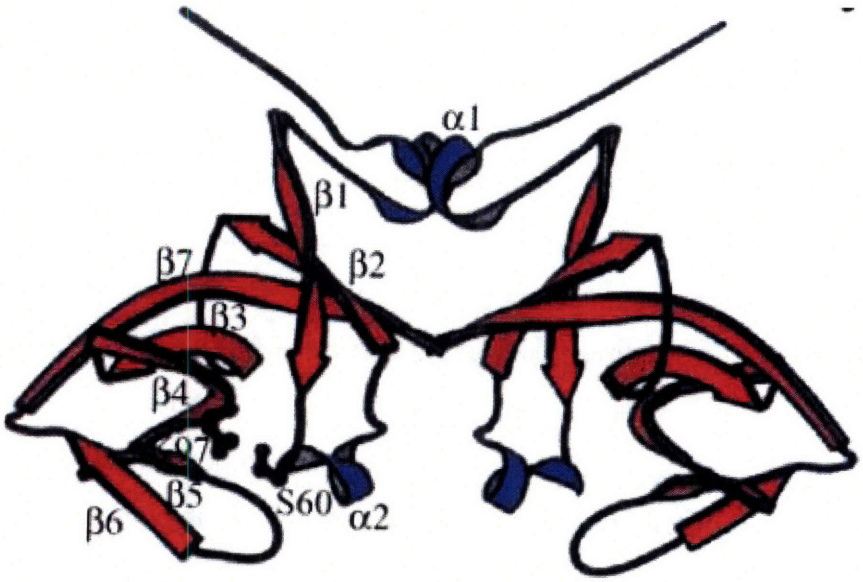


Figure 2-2: **Comparison of the NMR and X-ray crystal structures of UmuD′$_2$.**
(A) Cartoon of the NMR structure of UmuD′$_2$ as solved in [142]. (B) Cartoon of the X-ray crystal structure of UmuD′$_2$ as solved in [141]. These images were reproduced from [134].

shown to undergo intermolecular cleavage [151]. Even in the X-ray crystal structure of $LexA_2$, a minor change in a flexible loop has been proposed to enable an interchange between the *cis* and the *trans* conformations of the N-terminus [146]. In addition to the *cis* and *trans* proposed structures of $UmuD_2$, each possible conformation can adopt either a cleavable conformation, in which the protease active site and cleavage site residues are near each other and can undergo catalysis, or a noncleavable conformation, wherein the active site and cleavage site residues are far apart [147]. If these models are accurate, the actual structure of $UmuD_2$ may favor one of these possible conformations, or it may be a heterogeneous mixture of a multitude of these conformations.

Some data that suggest $UmuD_2$ may have multiple conformations are obtained from single-cysteine cross-linking and solvent accessibility studies. ^{3}H-iodoacetate has been used to modify cysteines introduced into $UmuD_2$ and to determine their extent of solvent accessibility [143]. Although certain single-cysteine derivatives had clearly higher or lower than average solvent accessibility than other residues, the majority of single cysteine derivatives were intermediately modified by iodoactate [143]. Single-cysteine cross-linking data for $UmuD_2$ are similarly unusual, as different cross-linking agents provide different results, with few clear trends and the many positions resulting in cross-linking as $UmuD_2$ homodimers with intermediate efficiency [88, 143, 144]. Most surprisingly, air oxidation by dialysis of reducing agents should be discriminatory because it is a very mild procedure and depends only on the proximity of two cysteines to each other in the native structure [152], but in the case of single cysteine derivatives of $UmuD_2$, it was not in fact very discriminatory [144]. Nevertheless, those single cysteine derivatives that cross-link robustly with iodine, which is a fast reaction that gives a snapshot of conformational states [144] generally are predicted to be in the dimer interface by our current models of $UmuD_2$ [147].

Intrinsically Disordered Proteins

These unusual features of *umuD* gene products are at least partially explained in Chapter 4, which illustrates the ways in which *umuD* gene products are similar to a relatively newly-discovered and poorly understood class of proteins currently termed intrinsically disordered proteins [153]. Intrinsically disordered proteins have been variously called rheomorphic [154],

partially folded [155], flexible [156], mobile [157], natively denatured [158], natively unfolded [159], intrinsically unstructured [160], and natively disordered [161], and the terminology is likely to change again to more fully capture the essence of these proteins as our understanding improves. Currently, they are characterized as having significantly less stable secondary or tertiary structural characteristics than other proteins when purified and at neutral pH; they are thought to undergo dynamic conformational transitions. The lack of secondary structure can be determined using a number of techniques, including NMR, CD, SAXS, fluorescence, and limited proteolysis [162, 163]. Many of these proteins obtain more typical secondary structure at an elevated or lowered pH, after a post-translational modification or cofactor assembly, upon binding to DNA, RNA, or another protein, or through other cellular conditions [164]. Some proteins, for example certain hormones [165], chaperones [153], and voltage-gated ion channels [166], maintain activity even in regions that lack secondary structure. In fact, some intrinsically disordered proteins have been shown to maintain disorder *in vivo* [167].

Examples of intrinsically disordered proteins include transcription factors [168, 169], other DNA or RNA binding proteins [163], chaperones [153], antibodies [153], proteases [163], and signaling factors [153, 170], including kinases and phosphatases. Signaling factors are particularly interesting in relation to *umuD* gene products (reviewed in [153]). Intrinsic disorder is thought to aid in the function of these proteins by enabling a large transition from disorder to order, but only under certain cellular conditions (such as the presence of a binding partner or cofactor) [171]. This transition allows specific stabilization of the ordered form of an intrinsically disordered protein and allows its functionality to be strictly regulated by the presence or absence of another cellular factor [163].

Most known intrinsically disordered proteins are from multicellular eukaryotes, although a fair number of unicellular and prokaryotic ones are also known [160, 172]. Intrinsically disordered proteins that are specific to eukaryotes include cytoskeletal proteins [173] and those that regulate cell cycle [174]. It is thought that eukaryotic cells have a greater need for precise regulation than prokaryotic ones, and intrinsically disordered proteins are often used as regulatory proteins [170, 172, 174-178].

A major advantage of intrinsically disordered proteins that enables them to have important regulatory roles is that the lack of a rigid, defined structure can allow them to respond quickly and efficiently to changes in their environment [174]. If the precise fold of a protein, or

the fact that it folds in a defined manner at all, depends on its surroundings, then the protein will only adopt a certain structure when its environment dictates the correct conditions. In the case of an interaction, the high entropic cost of the disorder-to-order transition involved in folding an intrinsically disordered protein causes many of these interactions to be low affinity and highly reversible [163]. Intrinsically disordered proteins often act as signaling molecules for precisely this reason; these proteins can detect an environmental situation and adopt a conformation, binding partner, or activity specific to the situation. The protein's change in conformation may signal downstream responses to an environmental stimulus. It has also been proposed that the relatively large volume of intrinsically disordered proteins may enable them to efficiently detect a specific binding agent, and if any of several distinct peptide units is sufficient for binding, a disordered structure with these peptides at different locations may increase binding kinetics relative to a more rigid interaction domain [175, 179]. A compact globular protein would be more limited by diffusion to find an interacting partner. Since the actual structure of an intrinsically disordered protein may depend on its specific interactions, these proteins are often able to make a large number of highly specific contacts despite a small size [180]. Intrinsically disordered proteins can often bind to multiple partners [181-183], and multiple intrinsically disordered proteins can sometimes independently conform to a single structured partner [171]. Furthermore, a lack of rigid secondary structure can sensitize proteins to degradation, and intrinsically disordered proteins are therefore often unstable *in vivo*, ensuring that their activity is limited to the correct conditions [160, 163, 174].

The actual structure of an intrinsically disordered protein can consist of a spectrum of conformations, from having large, flexible loops or unfolded domains that flank generally well-folded domains, to a non-random molten globule conformation or a more random, extended structure [153, 174]. However, a completely random and extended structure is probably rare [163, 174]. Even chemically denatured proteins may show a nonrandom structure that has some α-helix or β-sheet-like content but is more flexible than the classical secondary structures due to limited hydrogen bonding [184-186]. Proteins that have more than one distinct conformation are also thought of as intrinsically disordered if the transitions between the conformations are not cooperative [153].

Efforts to further classify and characterize the precise structures of intrinsically disordered proteins have begun. Two classes have been proposed based on extent of disorder as

measured by CD: coil-like and premoltenglobule-like intrinsically disordered proteins [187]. Coil-like proteins are more unfolded than premoltenglobules and can be truly random coils or a less random, more extended poly-proline type II coil [188, 189]. Premoltenglobules are thought of as a second unfolding intermediate, more unfolded than a molten globule, in certain protein unfolding pathways [187]. A molten globule has a hydrodynamic radius that is about 15% greater than the corresponding native globular protein, whereas the hydrodynamic radius of the premoltenglobule protein is closer to 50% greater, and there is less secondary structure in the premoltenglobule [187].

Bioinformatic approaches suggest that even regions that are currently thought of as almost completely disordered have three distinct "flavors" [178]. The flavors, called V, C, and S, were proposed based on protein disorder prediction training sets; the greatest prediction accuracy was achieved if the set of known disordered sequences was randomly divided into three subclasses and used to train separate predictors. Sequences were then tested by the disorder prediction programs and reassigned to groups based on which program predicted the disorder most accurately. The disorder predictors were retrained and the process repeated until all of the disordered proteins were sorted most accurately by their own group's prediction program. Just as ordered proteins usually have both α-helix and β-sheet components, it is likely that intrinsically disordered proteins also have more than one flavor of disorder. In fact, disordered regions, like ordered ones, can be highly conserved between species [163].

The three flavors of disorder have different sequence properties and may have different structural properties and behaviors [178]. In fact, unstructured proteins that have a particular function tend to cluster in the same flavor. Flavor C has an overrepresentation of histidine, methionine, and alanine compared to other proteins, and they often bind nucleic acids or have modification sites. Conversely, flavor S has an underrepresentation of histidine and is likely to be involved in protein-protein interactions. Interestingly, Flavor V has more cysteine, phenylalanine, isoleucine, and tyrosine than the other two flavors of protein disorder, and these amino acids are generally not considered flexible amino acids. Most unstructured ribosomal proteins are of flavor V. All of these flavors have lower sequence complexity than the general proteome [178], and although low sequence complexity is a good indication of disorder [163, 177], high sequence complexity is not a good predictor of rigid secondary structure [190, 191].

Proper *in vivo* folding and function of intrinsically disordered proteins is a topic of intense study, as intrinsically disordered proteins may be more likely than their more rigid counterparts to misfold, aggregate, and thereby cause diseases in multicellular eukaryotic organisms [192]. An example of a protein whose aggregation causes human disease is α-synuclein, a small, intrinsically disordered protein of unknown function in synaptic terminals whose misfolding is associated with Parkinson's and other neurodegenerative diseases [193-195]. In addition to having three separate alternative mRNA splicing isoforms [196], the structure of α-synuclein resembles a random coil when in an aqueous environment [197, 198] (a curiosity in the literature is that eukaryotic pre-mRNA segments that can be alternatively spliced are more likely to code for regions of disorder than for rigid structures [199]. Although α-synuclein is considered heat-stable because of its failure to form aggregates after boiling for 10 minutes [193], it can be heat-denatured when simultaneously incubated at low pH [200]. The N-terminus adopts an α-helical structure when associated with membranes [198, 201] , and a central part of the protein can become structured with protein-protein interactions [197]. The C-terminal tail remains unstructured even in a hydrophobic environment [202]. Like many intrinsically disordered proteins, α-synuclein can undergo several post-translational modifications, including ubiquitination, sumoylation, phosphorylation, nitration, and oxidation [196]. Each of these modifications influences the secondary and tertiary structure of the protein or its stability [203]. Additionally, α-synuclein can undergo a pathology-inducing transition to an α-pleated sheet, which forms aggregates and fibrils and can eventually lead to Parkinson's disease [198, 202, 204-206]. Interestingly, although α-synuclein is a neuronal protein found only in higher eukaryotes, its primary sequence shares some similarity with UmuD, including one of the protease cleavage site residues (the other maintains homology with LexA) and one of the active site residues (Figure 1-1 p.18). Although α-synuclein is thought to be post-translationally processed by proteolysis [196], the cleave site is not at the location of the semi-conserved cleavage site in UmuD and LexA, and there is no evidence for autocleavage of α-synuclein. When incubated under certain micellular conditions, α-sinuclein does form what appear to be homodimers as visualized by western blot, but these homodimers are most likely due to protein cross-linking rather protein folding forces such as hydrophobic interactions or hydrogen bonding [207-209]. Because of its lack of significant secondary structure, α-synuclein was assumed to be highly sensitive to protease digestion [210], but on the contrary, it is rather resistant to

chymotrypsin [200], cathepsin D [200], and cellular proteases [211]. The purification scheme for α-synuclein involves heat denaturation [193] or a high percentage of acetonitrile and purification by HPLC [200]. It may be interesting to see if structural features of α-synuclein differ between these harsh and other, more gentle methods of purification.

$UmuD_2$ and $UmuD'_2$ share homology with the dimerization domains of certain bacterial transcription factors, and transcription factors often have intrinsically disordered domains [168]. However, it is usually the DNA-binding domains or protein-protein interactions domains of transcription factors that are disordered rather than their homodimerization domains. Transcription factors may be divided into distinct modules that have different functions; intrinsic disorder in some of these modules facilitates interactions that promote rapid responses to cellular conditions [212].

In addition to specific interactions, nonspecific crowding effects may also induce structure in intrinsically disordered proteins [213]. Conditions in the cytosol are very different from *in vitro* conditions, partially because of macromolecular crowding [214]. Under crowded conditions, excluded volume effects artificially increase the effective concentrations of certain solutes, while steric hindrance limits the diffusion of large macromolecules [214]. Several crowding or viscosity-enhancing agents are used to mimic the effects of cellular crowding *in vitro*. Sugars have been shown to induce some CD structure in certain denatured proteins [215, 216]. High concentrations of salt also can do this, and sometimes even better than sugars [216], although in this case the effects may be due to charge-charge interactions rather than crowding. Free amino acids, methylamines, polyols, and other polymers have also been used to mimic a crowded cytosol [217]. Although the crowding mechanisms of these agents are not known in detail, multiple models have been proposed [216]. Polymeric crowding agents are thought to act through the excluded volume model [216], in which crowding is caused by steric hindrance, an increase in viscosity, and a decrease in the availability of water. Small osmolytes are thought to act through the preferential solvation model [216]. In this model, small molecules act in the opposite manner as denaturants in that they have an unfavorable interaction with the peptide backbone of proteins and provide a strong driving force for stabilization of a folded protein form [218]. Another possible model is the scaled particle theory [216].

Few intrinsically disordered proteins are known to form stable dimers in solution, although some examples are present in the literature. The homodimeric *E. coli* antitoxin MazE

has both structured and unstructured domains [219]. Like eukaryotic α-synuclein, the *Bacillus pasteurii* UreG protein, involved in urea metabolism, is thought to be nearly completely unstructured and cross-links to form dimers when oxidized [220], although earlier studies did not detect the disulfide bond [221]. The *E. coli* histone-like proteins HU-α and HU-β, like UmuD and UmuD′, form both homodimers and the heterodimer [222]. Interestingly, upon thermal denaturation, they undergo the folded to unfolded transition before the dimer to monomer transition, indicating that at certain temperatures, an unfolded dimer must be present [222]. In the case of HU-β_2, a significant amount of unfolded dimer is present even at room temperature *in vitro* [222]. Human Papilloma Virus protein E7 is homodimeric and has characteristics of intrinsically disordered proteins as shown by NMR, although the CD spectrum shows some α-helix and β-sheet structure [2].

Many intrinsically disordered proteins are also "hub" proteins, which represent local centers of an organism's interactome [223-226]. Although the exact definition of a hub protein is not yet defined, researchers in *S. cerevisiae* consider any protein that makes interactions with 5 or more other proteins a hub [227]. By this definition, both of the homodimeric *umuD* gene products would be hub proteins. Other researchers consider the definition of a hub protein to be the 10% of proteins within an organism that make the most protein-protein contacts. UmuD, though not specifically listed as a hub protein, has been found to make many interactions in a high-throughput screen [135]. Unsurprisingly, targeted deletion of hub protein genes tends to have greater and more pleiotropic phenotypic effects than deletions of other single protein-encoding genes [228].

Some evidence suggests tha hub proteins can be subdivided into party hubs, which tend to make simultaneous interactions, and date hubs, whose multiple interactions are usually separated by time or space [227]. Date hubs appear to have a greater tendency towards disorder [224], although hub proteins in general have a greater propensity for disorder than other proteins [225]. The current model is that the disorder prevalent in date hubs enables them to make multiple nonsimultaneous interactions, perhaps using overlapping surfaces [153, 224, 227]. If *umuD* gene products can be thought of as hub proteins, it is uncertain which, if any, of their interactions are made simultaneously and whether they act more as date or as party hubs. The flexibility and small size of $UmuD_2$ and $UmuD'_2$ suggest that they may act more like date hubs by making nonsimultaneous interactions that help in timing the error-free and error-prone aspects of the

SOS response. The distinction between party hubs and date hubs has recently been called into question using better-quality data sets and controlling for more variables than before [229, 230], and further research will be necessary to determine the relevance of hub protein classification.

Chapter 2 of this thesis is a study in which quantification of the interaction between β and a variant of $UmuD_2$ at the predicted β-binding motif showed that in fact, this region of $UmuD_2$ is not responsible for the strength of the interaction with β. However, the fluorescence spectrum of the complex shows that this region of UmuD does affect the conformation of the complex, explaining some of the phenotypes of the UmuD variant. Chapter 3 reports a novel interaction for *umuD* gene products—with DinB, the error-free Y family DNA polymerase in *E. coli*. Chapter 4 shows evidence that both $UmuD_2$ and $UmuD'_2$ share characteristics with intrinsically disordered proteins, whose precise 3-dimensional fold depends on the protein's environment, despite being stable homodimers. Conclusions are presented in Chapter 5.

CHAPTER 2

A NON-CLEAVABLE UmuD VARIANT THAT ACTS AS A UmuD′ MIMIC

Penny J. Beuning[1], Sharotka M. Simon[1], Adam Zemla[2], Daniel Barsky[3], and Graham C. Walker[1]

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139
[2]Computations Directorate and [3]Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550

Address correspondence to: Graham C. Walker, Massachusetts Institute of Technology, Department of Biology, 68-633, 77 Massachusetts Ave, Cambridge, MA 02139, Phone: 617-253-6716, Fax: 617-253-2643, Email: gwalker@mit.edu

ABSTRACT

$UmuD_2$ cleaves and removes its N-terminal 24 amino acids to form $UmuD'_2$, which activates UmuC for its role in UV-induced mutagenesis in *E. coli*. Cells with a non-cleavable UmuD exhibit essentially no UV-induced mutagenesis and are hypersensitive to killing by UV light. UmuD binds to the beta processivity clamp ("beta") of the replicative DNA polymerase, pol III. A possible beta-binding motif has been predicted in the same region of UmuD shown to be important for its interaction with beta. We performed alanine-scanning mutagenesis of this motif (14-TFPLF-18) in UmuD and found that it has a moderate influence on UV-induced mutagenesis but is required for the cold sensitive phenotype caused by elevated levels of wild-type UmuD and UmuC. Surprisingly, the wild-type and the beta-binding motif variant bind to beta with similar K_d values as determined by changes in tryptophan fluorescence. However, this data also implies that the single tryptophan in beta is in strikingly different environments in the presence of the wild-type versus the variant UmuD proteins, suggesting a distinct change in some aspect of the interaction with little change in its strength. Despite the fact that this novel UmuD variant is non-cleavable, we find that cells harboring it display phenotypes more consistent with the cleaved form UmuD′, such as resistance to killing by UV light and failure to exhibit the cold sensitive phenotype. Cross-linking and chemical modification experiments indicate that the N-terminal arms of the UmuD variant are less likely to be bound to the globular domain than those of the wild-type, which may be the mechanism by which this UmuD variant acts as a UmuD′ mimic.

This paper has been published as Beuning, P.J. *et al.* JBC **281**: 9633-9640 (2006).

INTRODUCTION

The *umuDC* gene products are induced as part of the SOS response and are responsible for much of the UV-induced mutagenesis in *E. coli* [231]. These gene products are subject to an elaborate set of controls that regulate their activity [231]. The LexA repressor provides transcriptional control, and there are several proteolytic controls on both the *umuD* and *umuC* gene products [231]. The homodimeric protein $UmuD_2$ is the predominant species during the first ca. 20-30 min after SOS induction [232]. $UmuD_2$, together with UmuC, plays a role in a DNA damage checkpoint, decreasing the rate of DNA synthesis and allowing time for accurate repair processes to act [232]. This correlates with the cold sensitive phenotype observed under conditions of overexpression of the *umuDC* gene products [232, 233]. As the SOS response proceeds, $UmuD_2$ binds the RecA/ssDNA nucleoprotein filament. This stimulates the latent ability of $UmuD_2$ to convert to $UmuD'_2$ by cleaving off its N-terminal 24 amino acids, resulting in $UmuD'_2$ becoming the predominant species. The RecA/ssDNA nucleoprotein filament serves to bring together the active site dyad residues Ser60 and Lys97, facilitating deprotonation of Ser60 by Lys97 [234]. The activated Ser nucleophile then cleaves the peptide bond between Cys24 and Gly25 of $UmuD_2$ [231].

The wealth of structural data and models available for $UmuD_2$ and $UmuD'_2$ provide insight into how the two forms of the *umuD* gene products engage in multiple highly specific interactions (Figure 2-1 p.36) [141, 234-237], including with the α, β, and ε subunits of the replicative polymerase, pol III [238]. Of the two forms, $UmuD_2$ interacts more strongly with the β processivity clamp (also referred to as β or the β clamp) than does $UmuD'_2$ [238, 239]. In full-length $UmuD_2$, the 39-amino acid N-terminal arms are stably bound to the globular C-terminal domain [234, 236] and form a distinct interaction surface. In the cleaved form of the protein, $UmuD'_2$, the remaining ca. 15 amino acids at the N-terminus appear unbound from the body of the protein and solvent-exposed [141, 235], revealing the buried portion of the C-terminal globular domain [234, 236]. A series of truncations at the N-terminal arm of $UmuD_2$ indicates that the first eight amino acids of $UmuD_2$ are dispensable for the $UmuD_2$-β interaction, while deleting residues 2-18 results in a substantial decrease in, but not a complete loss of, cross-linking efficiency with the β clamp [239]. Thus, the *umuD* gene products interact with the β clamp via both the N-terminal arms of $UmuD_2$ and the globular domain of $UmuD_2$ and $UmuD'_2$

[239]. This differential interaction appears to control, at least in part, whether the *umuDC* gene products act as part of a DNA damage checkpoint or as a

Figure 2-1: **Homology model of the $UmuD_2$ dimer (trans, arms down).**

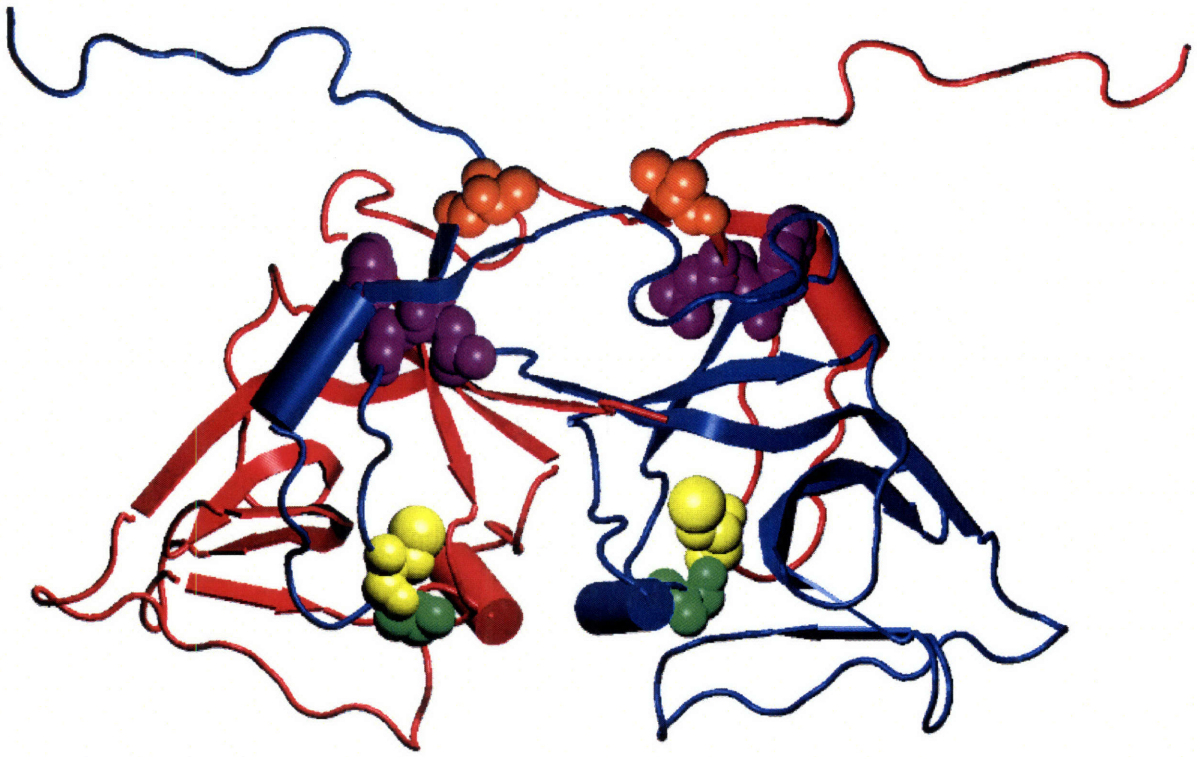


Figure 2-1: **Homology model of the $UmuD_2$ dimer (trans, arms down).** Also see Figure 2-S1 (p.40). One monomer is in blue; the other is in red. The residues mutated in UmuD-3A [T14 (orange); L17 F18 (purple)] are shown in space-filling rendering. The single Cys residue, Cys24, is shown in yellow. Ser60 is shown in green.

translesion polymerase [238, 239]. These interactions with the β clamp are of particular interest since sliding clamps play a key role in coordinating the multiple DNA polymerases present in cells [240-244]. The eukaryotic DNA sliding clamp PCNA interacts with multiple proteins. These interactions are in part regulated by covalent modification of PCNA with monoubiquitin or the small ubiquitin-like modifier (SUMO), which is distinct from the role of polyubiquitination in proteolytic degradation [245-248]. Duzen *et al* have suggested that $UmuD_2$ and $UmuD'_2$ play conceptually similar roles in modulating the various clamp interactions [249].

A version of the canonical β clamp binding motif found in eubacterial polymerases as well as other proteins involved in DNA metabolism was postulated to be present in UmuD at residues 14-18 (Figure 2-2 p.38) [250]. A yeast two-hybrid experiment with the motif of UmuD showed, however, that these five amino acids are not sufficient for the interaction with the β clamp [250]. Given the fact that this result was obtained utilizing only the five-amino acid motif in UmuD, and cross-linking experiments showed that the region of UmuD between residues 9 and 19 is important for interactions with the β clamp [239], we undertook a site-directed mutagenesis analysis of this motif. These studies led to the unexpected discovery of a new class of UmuD variant proteins that fail to undergo cleavage but whose properties resemble those of the cleaved version, UmuD'.

MATERIALS AND METHODS

Homology Model of UmuD- The models of the UmuD homodimer were created by the combined use of the program LGA [251] for protein structure comparison and superposition and the AS2TS program [252] for homology model-building. An initial model of the UmuD monomer (single chain) was constructed based on the crystal structure of UmuD' (Protein Data Bank [www.pdb.org] ID: 1ay9, chain A) [125]. The missing N-terminal arm was modeled by the LGA loop building/grafting procedure [251], using mainly the arm conformation in 1jhh_A, from the X-ray crystallographic structure of LexA [253] as well as the other LexA structures (1jhc, 1jhe, 1jhf) as template structures to guide the local and overall conformation. In the final alignment (Figure 2-S1, p.40), UmuD Asp20 had to be inserted into the LexA template (between residues 80 and 81 of LexA), and this was done by the LGA loop building procedure [251]. Finally, residues 1-14 are in an extended conformation (i.e., we make no prediction as to the placement of these residues; they are only modeled in a formal sense).

Figure 2-2: **The β-binding motif variants do not eliminate UV-induced mutagenesis.**

A

UmuD

UmuD	^{14}TLDLF^{18}
UmuD-3A	**A**LD**AA**

B

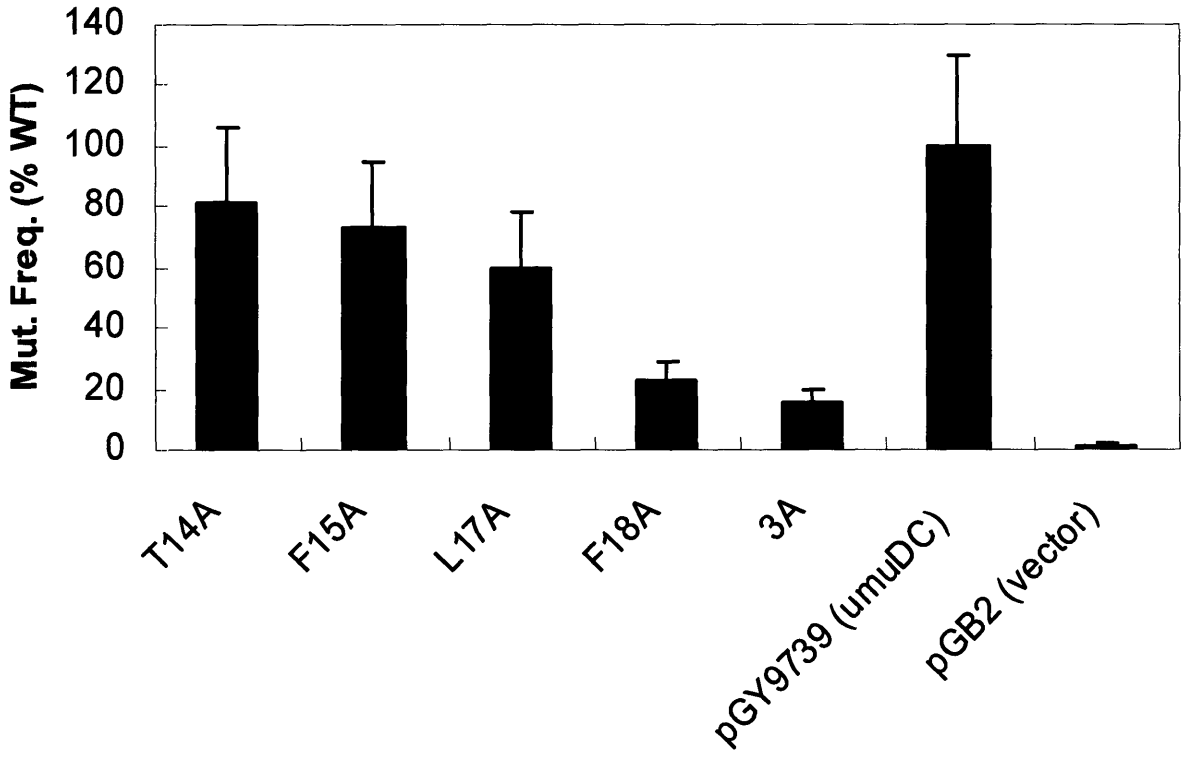


Figure 2-2: **The β-binding motif variants do not eliminate UV-induced mutagenesis.** (A) The putative β-binding motif in UmuD, with the variants indicated. (B) Induced mutation frequency of the indicated UmuD mutations in pGY9739 *umuDC* plasmids in GW8017 (Δ*umuDC*).

In creating the full UmuD homodimeric complex, we used LGA to superimpose our monomer model onto each of the template chains in the NMR structure of the UmuD' homodimer [234] (1i4v, chains A and B, model #1), but some minor clashes occurred which were alleviated by following the LexA homodimer instead [253]. This procedure creates a cis (non-domain swapped) conformation of the UmuD homodimer. Because there is a very small "shoulder" region at the top of the arms, the trans UmuD homodimer model could be constructed from the cis UmuD homodimer model by swapping the arms as follows: the first 39 residues in the chain A of our trans model were taken from the chain B of the cis model, and vice versa. This process of "arm swapping" was completed after applying the LGA loop building procedure to residues 39-41 in the shoulder regions. Finally, the LexA structures appear as both "elbows up" (N-terminal arm unbound) and "elbows down" (N-terminal arm bound to C-terminal domain), allowing us to model both conformations (Figure 2-S1 p.40). Thus, we created four models, two cis and two trans, each with an elbows up and an elbows down conformation (it is possible that heterogeneous conformations also occur with one elbow up and one elbow down as in the 1jhh LexA structure). For all the cis and trans models of the UmuD homodimer, the conformations of sidechains from residues either not present in 1ay9_A or that presented a steric clash after building the dimeric structures were modeled using the side-chain placement program SCWRL [254].

Proteins, Strains and Plasmids- A plasmid expressing UmuD-3A was constructed in pSG5 using mutagenic primers and the Quikchange kit (Stratagene) [255]. Wild-type UmuD and UmuD-3A were purified according to the published procedure [255]. The plasmid expressing His-HMK-β was a gift from Prof. M. O'Donnell (Rockefeller University), and β was purified according to the published procedure [256]. The strains and plasmids used in this study are listed in Table 2-1 (p.42). Plasmid pSJS9 was a gift from Prof. Charles McHenry (Univ. Colorado). Site-directed mutagenesis was performed using the Quikchange kit (Stratagene). Primer sequences are in Figure 2-S3 (p.43).

Mutagenesis and Survival Assays- SOS mutagenesis assays were performed according to the published method [255]. Briefly, cultures of GW8017 harboring various *umuDC*-expressing plasmids growing exponentially in LB were washed with 0.85% saline, exposed to 25 J/m^2 UV light from a germicidal lamp (General Electric), and then plated on M9 minimal plates with trace arginine (1 μg/mL). Colony-forming units were scored after 48 hr of growth at 42 °C. Survival

Figure 2-S1

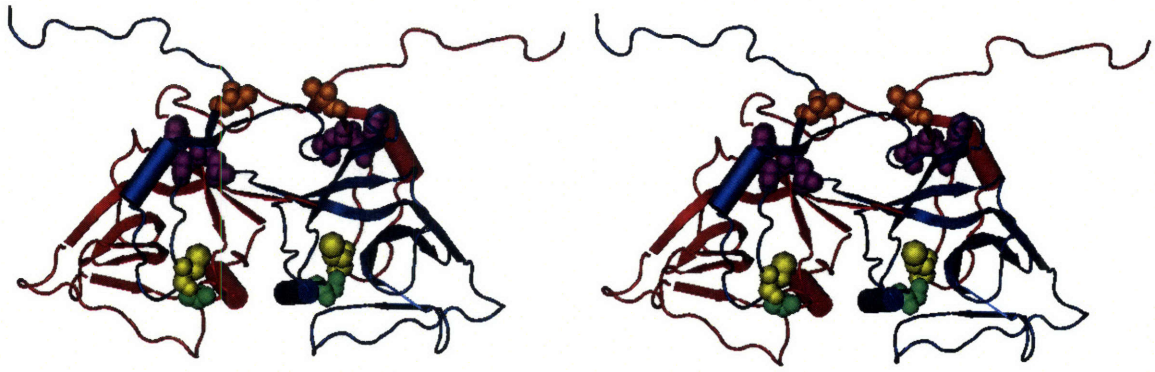
A

```
UmuD:    9   LRHIVTFPLFSDLVQCGFPSPAADYVEQRIDLNQLLIQHPSATYFVKASGDSMIDGGISD 68
               E    PL    V  G P  A  ++E    ++  L + P+A + ++ SG +M D GI D
1jhh_A: 70   QEHEEGLPLVG-RVAAGEPLLAQQHIEGHYQVDPSLFK-PNADFLLRVSGMAMKDIGIMD 127

UmuD:   69   GDLLIVDSAITASHGDIVIAAVDGEFTVKKLQLRPT-VQLIPMNSAYSPITIS-SEDTLD 126
             GDLL V       +G +V+A +D E TVK+L+ +    V+L+P NS + PI +    + +
1jhh_A: 128  GDLLAVHKTQDVRNGQVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFT 187

UmuD:   127  VFGVVIHVVKA 137
             + G+ + V++
1jhh_A: 188  IEGLAVGVIRN 198
```

B



C

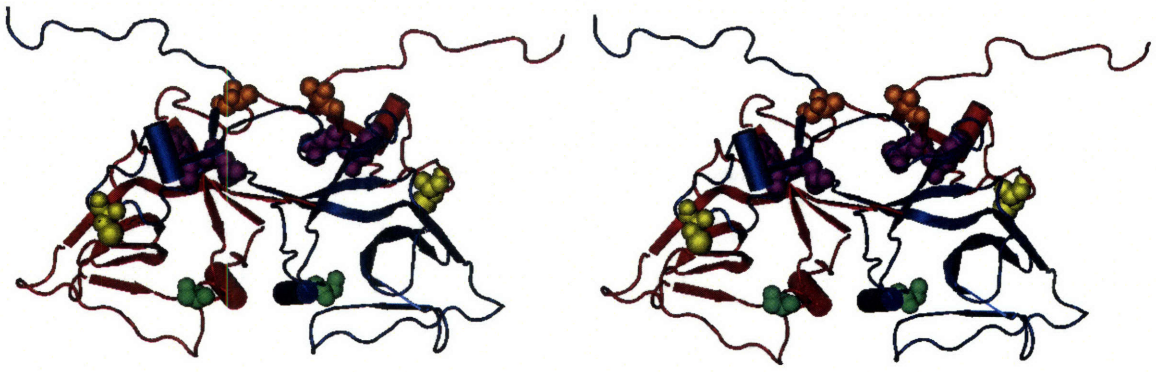
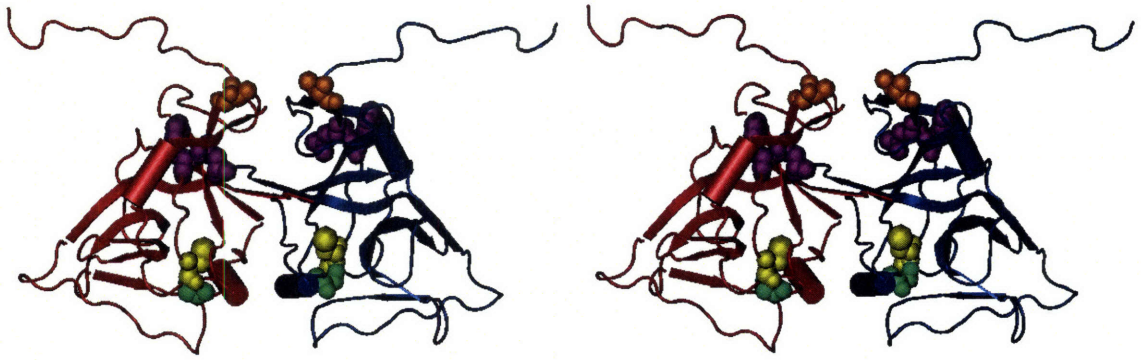


Figure 2-S1 (continued)

D



E

Figure 2-S1. (A) Alignment between sequences of UmuD and LexA proteins. The level of sequence identity is 31%. This alignment was used to construct the UmuD N-terminal residues 9 - 39 (highlighted in green). In our 3D model of UmuD the residues highlighted in blue in the alignment were taken from the PDB structure 1ay9_A. (B-E) As a result of our approach we have constructed four models: two trans versions (elbows down (B) and elbows up (C)) and two cis versions (elbows down (D) and elbows up (E)). In all cases (B-E) the coloring is the same as Figure 2-1 (p.36), as follows: one monomer is in blue; the other is in red. The residues mutated in UmuD-3A [T14 (orange); L17 F18 (purple)] are shown in space-filling rendering. The single Cys residue, Cys24, is shown in yellow. Ser60 is shown in green.

Table 2-1: **Strains and Plasmids**

Strain	**Relevant Genotype**	**Reference**
AB1157	*argE3*	Laboratory stock
GW8017	AB1157 Δ*umuDC*	[257]
Plasmid		**Reference**
pGY9738	o_1^C *umuD'C;* pSC101-derived	[258]
pGY9739	o_1^C *umuDC;* pSC101-derived	[258]
pGB2	Vector; pSC101-derived	[259]
pSJS9	Ts λ repression, Kan^R	[260]

Figure 2-S3: **Primer sequences**

DT14Acode
5'GCGAAATTGTGGCTTTTCCGCTATTTAGCG
DT14Anonc
5'CGCTAAATAGCGGAAAAGCCACAATTTCGC
DF15A
5'-CCGCGAAATTGTGACTGCTCCGCTATTTAGC
DF15A-r
5'-GCTAAATAGCGGAGCAGTCACAATTTCGCGG
DL17Acode
5'GTGACTTTTCCGGCATTTAGCGATCTTGTTC
DL17Anonc
5'GAACAAGATCGCTAAATGCCGGAAAAGTCAC
DF18Acode
5'GTGACTTTTCCGCTAGCTAGCGATCTTGTTC
DF18Anonc
5'GAACAAGATCGCTAGCTAGCGGAAAAGTCAC
D-T14AL17AF18A-C
5'GCGAAATTGTGGCTTTTCCGGCAGCTAGCGATCTTG
D-T14AL17AF18A-C-r
5'CAAGATCGCTAGCTGCCGGAAAAGCCACAATTTCGC
D-Ser60Ala
5'-GCAAGTGGTGATGCTATGATTGATGGTGG
D-Ser60Ala-r
5'-CCACCATCAATCATAGCATCACCACTTGC
Dseq
5'-GCCTGAATCAGTATTGATCTGCTGGC
UmuCseqF
5'-GGTGATCCACGTCGTTAAGGCGA
UmuC3'seq
5'-CGCTAATCCATTCGGCGCTCCTGC
pET11T-3seq
5'-GCTCAGCGGTGGCAGCAGCC

was determined by plating on M9 minimal plates with 40 μg/mL arginine. Non-UV irradiated cultures were treated identically to assess the spontaneous mutation frequency. UV survival curves were obtained after treating cells suspended in 0.85% saline in a Petri dish with the indicated doses of 254-nm light. Each sample was serially diluted and the dilutions plated on M9 minimal media plates supplemented with 1% casamino acids, 0.005% tryptophan, and 1.5% agar. Plates were incubated overnight at 42 °C.

Quantitative Transformation Assays- Transformation assays were performed essentially as described [255]. Plasmids (0.1 μg) were added to 25-50 μL competent AB1157 cells and incubated on ice for 10 min. After a 5-min heat shock at 37 °C, and a further 10-min incubation on ice, transformation mixtures were allowed to recover in 750 μL LB at 37 °C for 1.5 hr with gentle shaking. Equal volumes were plated on LB plates containing the appropriate antibiotics for incubation under different temperatures as indicated in the figure legends.

Immunoblots- To determine UmuD expression levels, cells were harvested from exponentially growing cultures in LB, lysed by boiling for 15 min, and loaded on 4-20% SDS-polyacrylamide gradient gels (Cambrex). Electrophoresed proteins were transferred to PVDF membrane (Millipore) in 10 mM CAPS, pH 8, 10% methanol. After blocking, membranes were probed with anti-UmuD/D′, and antibody interactions were detected with SuperSignal substrate (Pierce). Antibodies to UmuD/D′ were raised against purified UmuD in rabbits (Immunodynamics, Inc., La Jolla, CA). For UV-induced expression and cleavage experiments, an aliquot of ca. 2.5×10^{10} cells from an exponentially growing culture at $OD_{600} = 0.2\text{-}0.3$ was harvested, washed in 0.85% saline and UV-irradiated at 25 J/m^2. Irradiated cells were then transferred to LB and grown at 37 °C for the times indicated in the figure legend.

UmuD in vitro *Cleavage Assay*- RecA/ssDNA nucleoprotein filament-facilitated UmuD cleavage was assayed [255, 257, 261] in LG buffer for 30 min. Reactions were quenched by addition of SDS-PAGE buffer to 1x, and products were analyzed on 4-12% gradient polyacrylamide gels. Alkaline cleavage of UmuD was carried out [255, 262] in 100 mM glycine, pH 10, 10 mM $CaCl_2$, 50 mM NaCl, 10 mM DTT, and 0.25 μg/mL BSA for 48 hr at 37 °C. Reaction products were analyzed by 14% polyacrylamide gel electrophoresis. The extent of UmuD cleavage was quantified using the Line Profile feature of National Instruments Vision Assistant.

Cross-linking and Chemical Modification- Cross-linking was performed essentially as described [263] with bis-maleimidohexane (BMH, Pierce). Reactions were incubated at room temperature for the times indicated. For chemical modification with 5,5′-dithiobis(2-nitrobenzoate) (DTNB) [264], DTNB was dissolved at 2 mM final concentration in 50 mM HEPES, pH 7.5. Reactions were performed with 10-20 μM DTNB and 10-20 μM UmuD proteins in 50 mM HEPES, pH 7.5. The concentration of accessible thiols was calculated with an extinction coefficient of 13600 cm^{-1} M^{-1} at 412 nm. Several trials were performed, and representative data is shown.

Fluorescence Determination of Binding Constants- Binding constants between UmuD and β were determined essentially as described, with a PT1 QM-20000-4SE spectrofluorimeter (Lawrenceville, NJ) [255]. The β clamp has a single Trp (residue 122), while UmuD has none. The β_2 concentration was constant at 2.5 μM. Emission from UmuD or UmuD-3A without β_2 was subtracted from emission of the complex, and the center of spectral mass was calculated for each [UmuD]. Excitation was at 278 nm, and emission was monitored from 300 to 400 nm. Excitation and emission path polarizers were oriented perpendicularly. The data represent the average of at least three independent experiments ± one standard deviation.

RESULTS

Mutations in "β-binding motif" of UmuD do not result in complete loss of induced mutagenesis- We used alanine-scanning mutagenesis to make single alanine mutations in the putative β-binding motif in UmuD (Figure 2-1, p.36 and Figure 2-2, p.38) and investigated the consequences of these variants on known phenotypes of UmuD. These plasmid-borne variants were assayed for their ability to complement a Δ*umuDC* null strain for UV-induced mutagenesis. In addition to single alanine variants of UmuD, we constructed one variant with alanines at the first and last two positions in the motif (UmuD-3A, Figure 2-2, p.38), the positions most conserved among all β-binding motifs [250]. In the case of UmuC, the analogous mutation in its β-binding motif results in a complete loss of UV-induced mutagenesis [240, 241]. No single mutation or set of multiple mutations in this motif in UmuD failed completely to complement a Δ*umuDC* strain. However, plasmids expressing either the F18A UmuD variant, which is located at the top of the arm over the C-terminal globular domain, or the T14A L17A F18A ("UmuD-3A") variant resulted in substantial decreases in induced mutagenesis, down to about 15-20% of

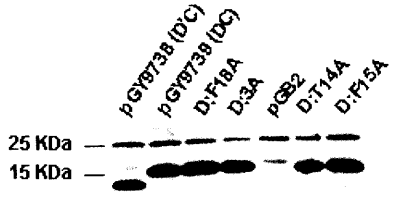
the wild-type (Figure 2-2, p.38). Curiously, this decrease in mutagenesis of cells harboring these variants was not accompanied by a corresponding decrease in survival after UV irradiation (Figure 2-5, p.49); yet, typically, increased mutagenesis due to translesion synthesis by UmuD′$_2$C is associated with increased survival after treatment with UV.

Cleavage of the N-terminal 24 amino acids from the arm of UmuD to yield UmuD′ is required to activate UmuC for its role in translesion synthesis [231]. Since the UmuD arm harboring these mutations would be removed upon cleavage, we reasoned that the defect in induced mutagenesis of strains expressing the F18A and T14A L17A F18A (UmuD-3A) variant proteins might be due to defects in cleavage. Given their positions in the N-terminal arm (Figure 2-1, p.36), it might be expected that these residues would play a role in properly positioning the arm in the active site for cleavage. We tested whether these mutations in the N-terminal arm of UmuD interfered with cleavage after UV-exposure. The F15A and L17A mutants showed a slight decrease in cleavage and an approximately 1.5-fold decrease in induced mutagenesis compared to the wild-type (Figure 2-2 p.38, Figure 2-3, p.47 and data not shown). The two UmuD variants (F18A and UmuD-3A) that showed essentially no cleavage up to 3 h after UV exposure (Figure 2-3, p.47), or even after 14 h (data not shown), resulted in the greatest reduction in induced mutagenesis (ca. 15-20% of wild-type). This is in contrast to non-cleavable UmuD active site variants that have been assayed previously, which showed essentially complete loss of induced mutagenesis (the limit of detection of this assay is ca. 1000-fold, or 0.1% of wild-type) [265, 266]. Thus, there are two groups of non-cleavable UmuD variants, one of which renders cells partially mutable and the other that renders cells essentially non-mutable.

The wild-type and UmuD-3A variant proteins were purified in order to assess their efficiency in *in vitro* cleavage facilitated by the RecA/ssDNA nucleoprotein filament. Under these conditions, UmuD$_2$ is cleaved efficiently to form UmuD′$_2$, while UmuD-3A$_2$ exhibits little detectable cleavage (Figure 2-3, p.47). Here again, the lack of cleavage is similar to that exhibited by the active site mutant of UmuD$_2$, UmuD-S60A$_2$ [265]. We note that there is a lower band present in some of the samples incubated *without* the RecA/ssDNA nucleoprotein filament. This lower band is often observed in preparations of UmuD, and even in some preparations of UmuD-S60A. However, in the case of UmuD-3A, the intensity of the lower band does not increase after incubation with the RecA/ssDNA nucleoprotein filament.

Figure 2-3: **UmuD variants are defective in cleavage.**

A



B

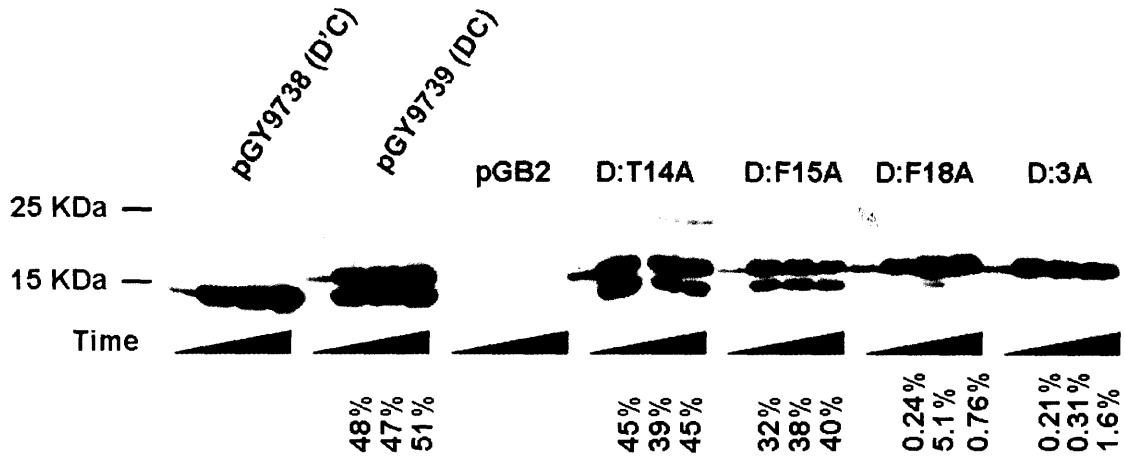


Figure 2-3 (continued): **UmuD variants are defective in cleavage.**

C

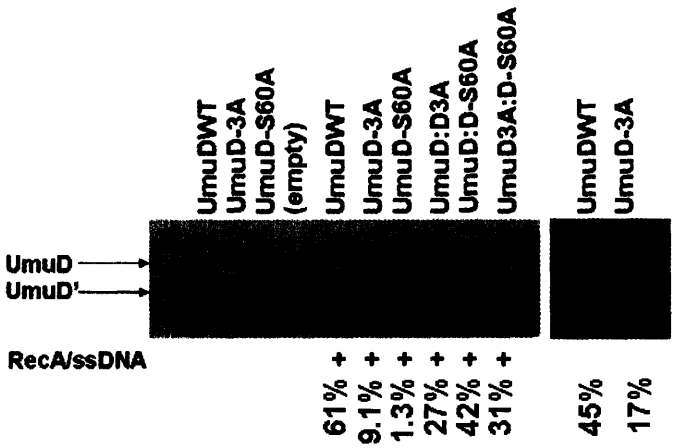


Figure 2-3. **UmuD variants are defective in cleavage.** (A) Immunoblot showing the stable production of UmuD variant proteins. (B) Immunoblot showing *in vivo* cleavage of UmuD variant proteins. Time points for each sample are 0 (before UV-irradiation) and 1, 2, and 3 h after UV irradiation. (C) UmuD cleavage *in vitro*. The (+) sign indicates the presence of the RecA/ssDNA nucleoprotein filament. The last two lanes show alkaline cleavage of UmuD. The percentage of the cleaved product was determined as the proportion of total density of each lane.

Figure 2-5: **UV survival of UmuD variants.**

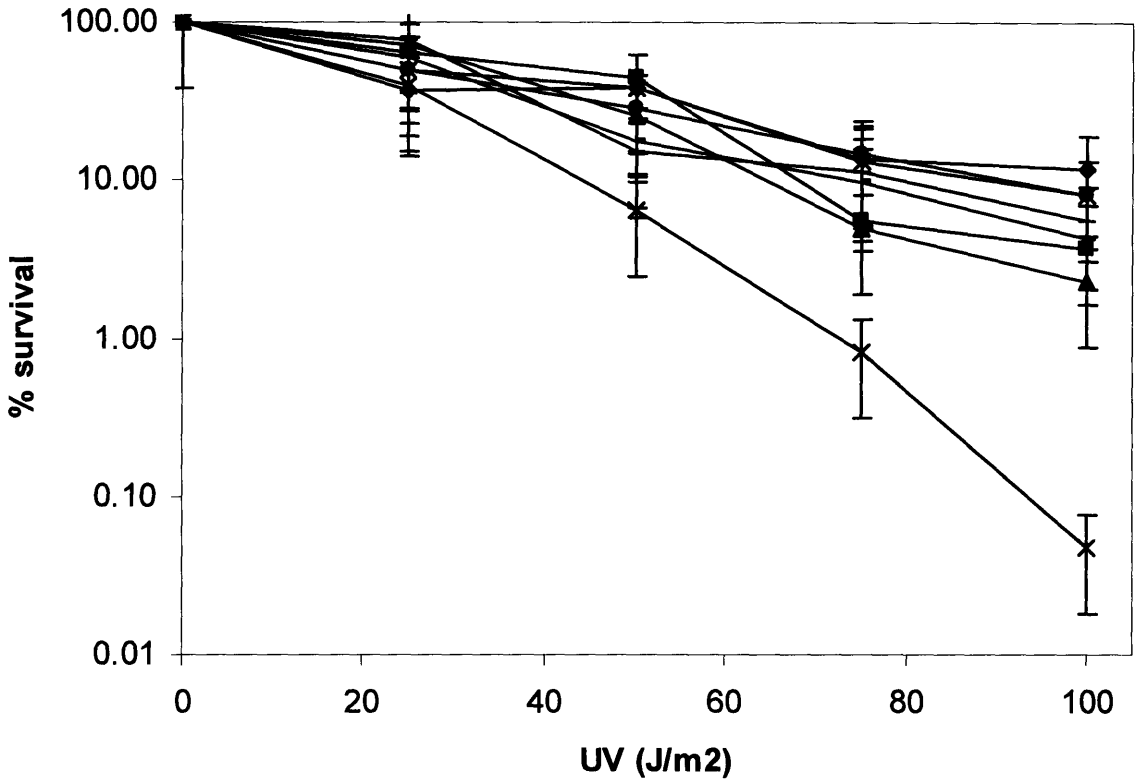


Figure 2-5: **UV survival of UmuD variants.** Assays were performed with pGY9739 plasmids and derivatives in GW8017 (Δ*umuDC*): pGY9738 (*umuD'C*, ♦); pGY9739-F18A (*umuDC* F18A, ●); pGY9739-D3A S60A (*umuDC* D3A S60A, +); pGY9739-F18A S60A (*umuDC* F18A S60A, -); pGY9739-D3A (*umuDC* D3A, ✳); pGY9739 (*umuDC*, ■); pGB2 (empty vector, ▲); pGY9739-S60A (*umuDC* S60A, ×).

UmuD forms exchangeable dimers [265], so wild-type $UmuD_2$ was combined with UmuD-$3A_2$, and cleavage was observed (Figure 2-3, p.47). Since UmuD-3A cannot cleave its own arm, the observed cleavage is likely due to the active site catalytic dyad of UmuD-3A acting on the wild-type partner's arm, although the reverse is also possible. In this experiment it is also possible that the cleavage observed is due entirely to a small population of wild-type $UmuD_2$ homodimers. To eliminate this possibility, UmuD-$3A_2$ was incubated with the active site variant UmuD-S60A_2, and some cleavage was still detected (Figure 2-3, p.47). This slight cleavage must be due to the active site residues of UmuD-3A cleaving the arm of UmuD-S60A, which suggests that the active site of UmuD-3A is proficient for cleavage and that the cleavage defect is isolated to its arm. The mutations in UmuD-3A at the top of the arm may disrupt folding of the arm over the globular domain or may interfere with specific protein-protein contacts required to facilitate cleavage (Figure 2-1, p.36).

To ensure that the cleavage defect was not due to defective interactions of UmuD-$3A_2$ with the RecA/ssDNA nucleoprotein filament, we also carried out cleavage under alkaline conditions in the absence of the RecA/ssDNA nucleoprotein filament. The RecA/ssDNA nucleoprotein filament serves to facilitate deprotonation of Ser60 by a neutral Lys97 [234, 253]. In the absence of the RecA/ssDNA nucleoprotein filament, the activation of Ser60 as a nucleophile can be accomplished under alkaline conditions. Under these conditions, $UmuD_2$ cleavage is inefficient but can be detected [262]. We found that cleavage of UmuD-$3A_2$ was substantially decreased compared to that of the wild-type (Figure 2-3, p.47). This suggests that the cleavage defect of UmuD-$3A_2$ is due to a defect intrinsic to the UmuD-$3A_2$ variant rather than deficient interactions with the RecA/ssDNA filament.

UmuD-3A fails to exhibit the cold sensitive phenotype- Strains with elevated levels of the *umuDC* gene products exhibit a cold-sensitive phenotype that correlates with a DNA damage checkpoint [232, 233]. Cells harboring plasmids overexpressing the cleavable *umuD* variants T14A and F15A (+ wild-type *umuC*) were also cold sensitive, while those overexpressing the L17A variant displayed an intermediate phenotype. The T14A and F15A variants behave similarly to wild-type in terms of their ability to exert the cold sensitive phenotype, to be cleaved to UmuD′, and to act in UV-induced mutagenesis. The cold sensitive phenotype is substantially enhanced in cells overexpressing the non-cleavable variant UmuD-S60A (Figure 2-4, p.51) [267]. Thus, we were surprised to find that strains harboring plasmids expressing the

Figure 2-4: **UmuD variants result in loss of the cold-sensitive phenotype.**

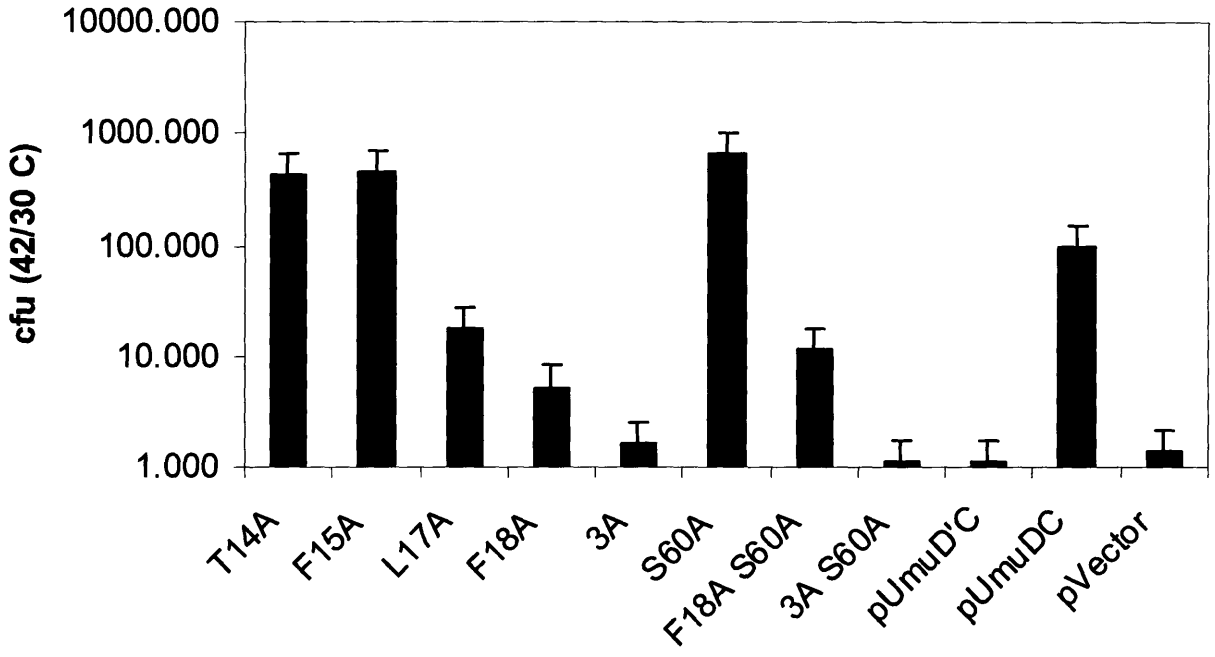


Figure 2-4: **UmuD variants result in loss of the cold-sensitive phenotype.** The ratio of colony-forming units (cfu) of AB1157 per μg of transformed plasmid DNA when grown at 42 °C versus 30 °C is plotted for each UmuD construct.

noncleavable *umuD* arm variants (UmuD-F18A and UmuD-3A) failed to display this cold-sensitive phenotype (Figure 2-1, p.36 and Figure 2-4, p.51).

We hypothesized that the loss of the cold sensitive phenotype was due to the specific arm mutations of UmuD, regardless of their cleavage defect. To test this, we combined in single constructs either the UmuD-F18A or UmuD-3A arm mutations with the S60A mutation that renders UmuD catalytically inactive. Even though they are not cleavable, the arm mutations F18A and UmuD-3A alleviated the extreme cold sensitivity exhibited by strains with elevated levels of UmuD-S60A (Figure 2-4, p.51). Although strains harboring plasmids overexpressing the UmuD F18A S60A double mutant display a cold sensitive phenotype that is intermediate between that of cells with each corresponding single mutant, the cold sensitive phenotype of cells overexpressing UmuD-S60A is suppressed by two orders of magnitude by the presence of only a single mutation in the N-terminal arm, F18A. These arm variants must disrupt a specific molecular interaction necessary to cause the cold sensitive phenotype that is independent of whether they can be cleaved.

Simultaneously elevated levels of the *umuD*, *umuC*, and *dnaN* (which codes for the β clamp) gene products cause a lethal phenotype, which has been interpreted as an exaggeration of the cold sensitive phenotype [268]. A strain harboring a plasmid expressing UmuD-3A and UmuC, when combined with high levels of the β clamp, fails to exhibit the synthetic lethal phenotype (Table 2-2, p.53). This suggests that a critical aspect of this complex formation with β is disrupted in the UmuD-3A variant.

Sensitivity to UV exposure- Given the cleavage defect of the UmuD-3A and F18A N-terminal arm variants, we decided to look more closely at the curious lack of a correlation between UV mutability and survival after exposure to UV that we had noted earlier for cells expressing the F18A and UmuD-3A variants. It is known that *E. coli* strains harboring a non-cleavable UmuD variant are hypersensitive to killing by UV light and are non-mutable [232, 265] (Figure 2-5, p.49). In order to determine whether this is also true of the non-cleavable UmuD F18A and UmuD-3A variants, strains harboring plasmid-borne *umuD* variants were assayed for their resistance to UV light. Strains with plasmids expressing either of the non-cleavable variants F18A or UmuD-3A showed a similar level of resistance to UV light as those expressing wild- type UmuD or a synthetic construct of UmuD′ (Figure 2-5, p.49). We suspected that the resistance to killing by UV of the F18A and UmuD-3A variants was due to a

Table 2-2: **Loss of Synthetic Lethality due to Mutations in UmuD β-binding motif**

AB1157 pSJS9	**cfu (37 °C, per μg DNA per mL)**
pGY9738 (WT)	0
pGB2	1040
pGY9739-UmuD-3A	1400

specific feature of the arm mutants, unrelated to their cleavage defect. Cells with plasmids expressing the non-cleavable arm variants constructed in the context of the UmuD-S60A active site variant were assayed for their resistance to UV light. Strains with plasmids expressing either the F18A or UmuD-3A arm variants combined with UmuD-S60A exhibited a striking resistance to UV light that was similar to strains with wild-type UmuD (Figure 2-5, p.49). This suggests that alterations in the N-terminal arm of UmuD are able to suppress the extreme UV-sensitive phenotype of non-cleavable UmuD-S60A, even though they are also non-cleavable. In light of the dramatic changes in the phenotypes of cells expressing the UmuD-3A variant compared to those with previously characterized non-cleavable variants of UmuD [265], we investigated the conformational consequences of the UmuD-3A variant compared to the wild-type.

Chemical cross-linking and modification of UmuD homodimers- We hypothesized that these non-cleavable UmuD variants are able to confer resistance to UV light, as well as to suppress the cold sensitive phenotype, by mimicking the conformation of UmuD′. To examine this possibility, we analyzed the conformation of the N-terminal arm of the UmuD-3A variant compared to that of the wild-type UmuD. UmuD, which possesses a C_2 axis of symmetry, has a single Cys residue, Cys24, at the cleavage site in the N-terminal arm. In order to determine whether UmuD-3A is a UmuD′ mimic with respect to the position of its arms, cross-linking was performed with the thiol-specific homobifunctional 16 Å cross-linker bis-maleimidohexane (BMH). Our model allows us to put a lower limit of 20 Å on the distance between these two Cys thiols. This lower limit represents an implausible path for the cross-linker, since it is the direct distance between the two Cys thiols (Figure 2-1, p.36). Thus, cross-linking should only be detected when the arms are "up", i.e. not bound to the C-terminal globular domain of UmuD. UmuD-3A was more readily cross-linked by BMH than either wild-type UmuD or UmuD-S60A (Figure 2-6, p.55), suggesting that the arms of UmuD-3A are less likely to be bound to the globular domain of UmuD. Therefore they are more often close enough to be cross-linked.

One of the models of UmuD that we have proposed ("trans, elbows down", see Figure 2-1, p.36 and Figure 2-S1, p.40) predicts that the thiol group of the single Cys24 residue is partially buried under the peptide backbone of the N-terminal arm. However, if the UmuD arms are in an "up", or more flexible, conformation, then the Cys should be more accessible to a thiol- specific reagent. We performed a titration of the Cys residue at the Cys24-Gly25 cleavage site

Figure 2-6: **UmuD-3A arms are more easily cross-linked with BMH and more accessible to chemical modification than wild-type.**

A

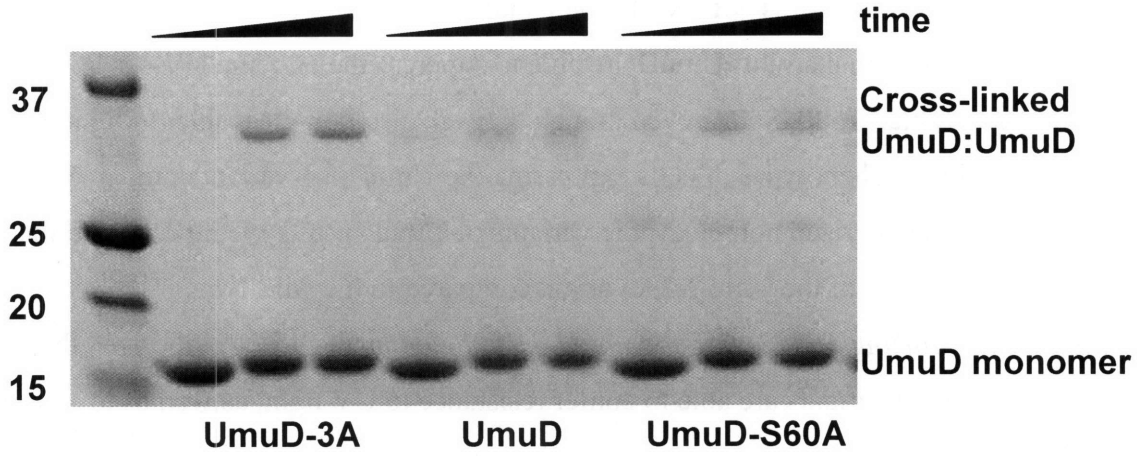


Figure 2-6 (continued): **UmuD-3A arms are more easily cross-linked with BMH and more accessible to chemical modification than wild-type.**

B

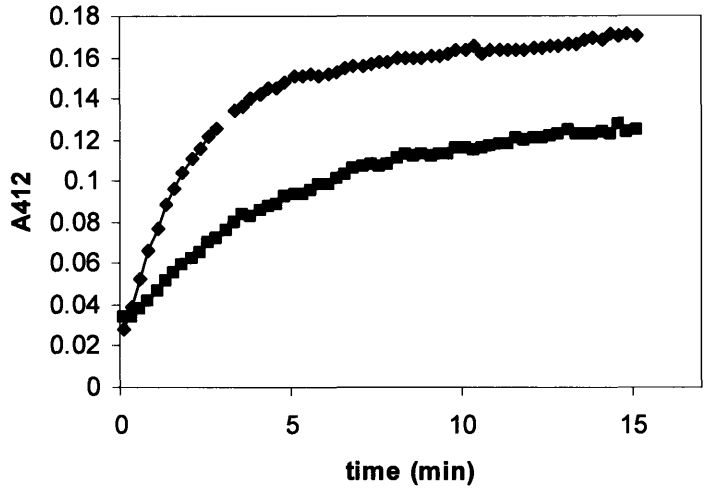


Figure 2-6: **UmuD-3A arms are more easily cross-linked with BMH and more accessible to chemical modification than wild-type.** (A) Cross-linking the UmuD arms with BMH. Each UmuD variant that was cross-linked is indicated under the lanes. The time points were 0, 15, and 30 min after addition of BMH. The first lane shows molecular weight standards, indicated in kDa. Proteins were visualized by Coomassie staining. (B) DTNB titration of free thiol in UmuD. The complete reaction without protein was used as a blank, protein was added and absorbance was recorded at 412 nm for 15 min. $UmuD_2$ (■) and $UmuD\text{-}3A_2$ (♦) were present at 10 μM with 20 μM DTNB.

with 5,5′-dithiobis(2-nitrobenzoate) (DTNB). The thiol moiety of UmuD-3A was more reactive to DTNB and therefore slightly more accessible than that of the wild-type UmuD (Figure 2-6, p.55). We determined that there is 1.0 reactive Cys residue per wild-type $UmuD_2$ and 1.2 reactive Cys residues per UmuD-$3A_2$, supporting the idea that in the UmuD-3A variant, the N-terminal arms undergo a shift in equilibrium to a less bound, arms-up state.

Determination of K_d of UmuD and the β clamp- In order to quantify the binding of UmuD and the UmuD-3A variant to the β clamp, we determined the K_d for this interaction. Surprisingly, we found that although the K_d is similar for β binding to either wild-type UmuD (5.5 ± 0.8 μM) or UmuD-3A (6.1 ± 0.5 μM), the mode of binding is different for each protein. Namely, the fluorescence emission from the tryptophan in β shifts to a longer wavelength upon binding to UmuD, while the shift is to a shorter wavelength in the presence of UmuD-3A (Figure 2-7, p.58). Tryptophan fluorescence emission peaks at a longer wavelength in a polar environment and at a shorter wavelength in a hydrophobic one, indicating that the partially-exposed tryptophan in β (Figure 2-1, p.36) becomes more solvent-exposed upon binding to wild-type UmuD and buried upon binding to UmuD-3A [269]. Accordingly, unlike canonical β-binding motifs [240, 241, 270, 271], this motif in UmuD is not responsible for the strength of the interaction with β, but rather for a qualitatively different mode of binding.

DISCUSSION

Although this work was initiated to determine the role of the putative β-binding motif (14-TFPLF-18) in UmuD function, we found that alterations in the motif do not prevent binding to the β clamp, unlike corresponding mutations in UmuC, DinB, and the pol III α subunit [240, 241, 250, 270, 271]. Instead, we show here that the UmuD-3A variant alters the N-terminal arm conformation in a way that dramatically changes UmuD activity, and seems to exhibit properties of UmuD′, particularly with respect to resistance to killing by exposure to UV light and lack of the cold sensitive phenotype when overexpressed together with UmuC. The UmuD-3A variant shows no defect in survival but decreases UV mutagenesis. This suggests that this variant may allow selective bypass of T:T cyclobutane pyrimidine dimers (CPDs) but not [6-4] photoproducts, since lethality is associated with T:T CPDs and mutagenesis with [6-4] photoproducts [231].

Figure 2-7: **UmuD binding as observed by β tryptophan fluorescence.**

A

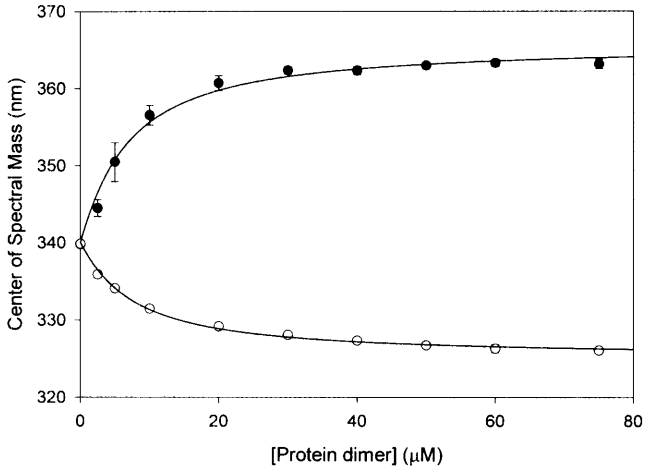


Figure 2-7 (continued): **UmuD binding as observed by β tryptophan fluorescence.**

B

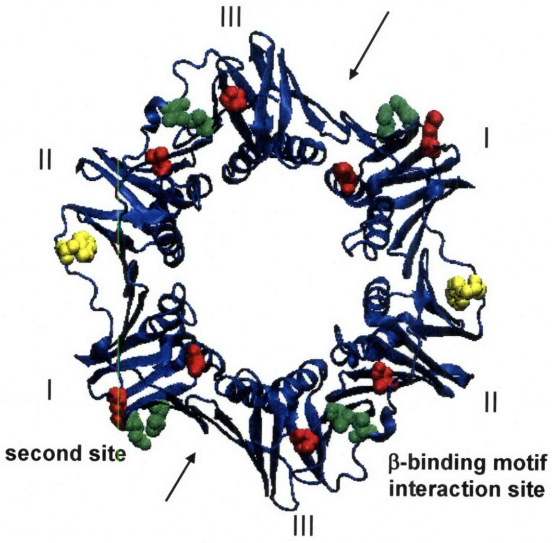


Figure 2-7: **UmuD binding as observed by β tryptophan fluorescence.** (A) UmuD (filled circles) and UmuD-3A (open circles) bind with similar affinity yet cause opposite shifts in the Trp fluorescence emisson wavelength. The K_d of $UmuD_2$ is 5.5 ± 0.8 μM, while that of UmuD-$3A_2$ is 6.1 ± 0.5 μM. Representative fluorescence emission curves, as well as the full data set of the Centers of Spectral Mass as a function of UmuD concentration, are presented as Figure 2-S2 (p.60) and Table 2-S1 (p.61). (B) Structure of the β clamp (entry 2POL from the Protein Data Bank) showing the known sites of interaction of UmuD based on cross-linking experiments (red) [249] versus the site of interaction of the canonical β-binding motif (green) and the second site of interaction observed in the co-crystal structure with DinB (green, labeled "second site") [272, 273]. The single tryptophan, W122, is also indicated in yellow. Arrows indicate the dimer interface. The domains are labeled I, II, and III. This image was prepared with VMD [274].

Figure 2-S2: **Fluorescence emission spectra of 2.5 μM β_2 excited at 278 nm.**

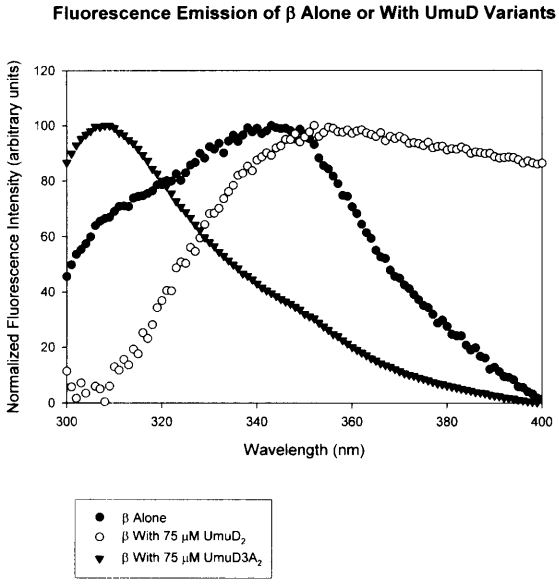


Figure 2-S2: **Fluorescence emission spectra of 2.5 μM β_2 excited at 278 nm.** The fluorescence emission of $UmuD_2$ or UmuD-$3A_2$ alone was subtracted from that of the complex as appropriate. The minimum value was set to 0 by addition or subtraction before calculation of the center of spectral mass, which was necessary because the subtracted fluorescence of β_2 in the presence of 75 μM $UmuD_2$ was negative between 300 and 335 nm.

Table 2-S1: **Fluorescence Center of Spectral Mass of 2.5 μM β_2 Bound to Wild-Type $UmuD_2$ or $UmuD3A_2$**

	$UmuD_2$		$UmuD3A_2$	
Interacting Protein Concentration (μM)	Center of Spectral Mass (nm)	Standard Deviation	Center of Spectral Mass (nm)	Standard Deviation
0	339.8	0.34	339.8	0.14
2.5	344.5	1.1	335.9	0.45
5	350.5	2.5	334.1	0.34
10	356.6	1.3	331.5	0.36
20	360.7	0.97	329.2	0.32
30	362.3	0.47	328.1	0.23
40	362.3	0.53	327.4	0.34
50	362.9	0.45	326.7	0.15
60	363.3	0.46	326.3	0.53
75	363.1	0.62	326.1	0.25

How could an uncleaved UmuD mimic the cleaved form, UmuD′? Current evidence suggests that the N-terminal arms of UmuD are usually bound to the C-terminal globular domain of the protein, i.e. it is usually in the "elbows down" conformation [236]. Even when the arms are covalently bound to the globular domain, UmuD can be cross-linked to the β clamp with almost no decrease in efficiency [239]. When UmuD is cleaved to UmuD′, the remainder of the N-terminal arm (residues 25-39) is able to move relatively freely [125, 235]. The UmuD-3A variant seems to have enough flexibility in its arms that it is at least a partial mimic of UmuD′. Although residues 14-18 of UmuD are predicted from our model to be only partially buried (Figure 2-1, p.36), the UmuD-3A variant may disrupt optimal packing of the arm against the globular domain. The mutations in UmuD-3A are at the point of the N-terminal arm that begins a downward turn over the C-terminal globular domain, and disruption of this turn is consistent with the more extended UmuD′ arm structure. In addition, the cleavage of UmuD exposes a different surface of UmuD′ for protein-protein interactions. Thus, the cleavage reaction serves the dual function of removing a portion of the N-terminal arms and presenting a dramatically different surface of the protein for interactions.

We were able to create models of the UmuD homodimer in both the elbows up and the elbows down conformation because both conformations are observed in the LexA structures [253]. The LexA structure is in a cis conformation (non-domain swapped) with respect to the positioning of the arms [253]. We have noticed, however, that in the $UmuD'_2$ structures the truncated arms point in the trans direction, suggesting that perhaps the trans conformation is actually preferred for UmuD [141, 235]. It has been shown that UmuD can undergo cleavage in trans (Fig 3) [235, 275]. The model of UmuD most consistent with the available biochemical evidence is one in which the arms are in the trans conformation [236].

In constructing the UmuD-3A variant, we have made a version of UmuD that binds to the β clamp with a similar affinity as the wild-type, but with a subtle change in the specific interaction as evidenced by the strikingly different tryptophan fluorescence emission spectra. This change would not have been detected by many of the techniques commonly used to detect protein-protein interactions, such as co-immunoprecipitation or two-hybrid analysis. Recent evidence suggests that the domains of the sliding clamps are rigid bodies joined by flexible linker regions [276]. The single tryptophan of β is on a long flexible loop between rigid Domains I and II (Figure 2-7, p.58) [277], so UmuD binding at a distal site to the tryptophan could cause a

slight conformational rearrangement of the domains that alters the environment of the tryptophan in the loop. The distinct fluorescence spectra are indicative of at least a slight change in the conformation of the β clamp when bound to either wild-type UmuD or UmuD-3A. UmuD-3A exhibits similar biological functions to those of UmuD′, yet UmuD-3A retains both the arm length and the strength of the interaction with β of wild-type UmuD. This suggests that the change in conformation of the β clamp, along with the relative conformational freedom of the UmuD-3A N-terminal arms, may be partially responsible for the UmuD′-like phenotype of UmuD-3A. UmuD and DinB/pol IV bind to β at overlapping sites (Figure 2-7, p.58) [249, 272]. One of these sites is the hydrophobic channel between β Domains II and III where all known β-binding motifs interact [272, 273, 278]. Although UmuD possesses a similar motif to a canonical β-binding motif that modulates its interaction with the β clamp, and UmuD has been shown to bind to the same site on β as other β-binding motifs, it does so in a way that is distinct from other proteins that bind β via their β-binding motifs [278].

Processivity clamps play a critical role in controlling traffic at the replication fork and in cell cycle checkpoints. Polymerase binding to the β clamp regulates access to the primer terminus by replicative or translesion synthesis DNA polymerases [242, 243]. Moreover, it has been shown that UmuD interacts more strongly with β than UmuD′ does [238]. The interaction of UmuD with the β clamp seems to be important for facilitating a DNA damage checkpoint in *E. coli* [232, 233], and the cleavage of UmuD to UmuD′ may attenuate this checkpoint function [239]. The UmuD-3A variant seems to bypass this switch, yet still binds the β clamp. In eukaryotes, covalent modification of PCNA with monoubiquitin or SUMO determines whether the cell utilizes DNA repair or potentially mutagenic translesion synthesis [246]. Thus, access to sliding clamps is universally important for control and regulation of proteins acting at the replication fork.

FOOTNOTES

We thank Prof. Robert Sauer (MIT) and his laboratory for use of the fluorimeter and for technical advice. We also thank Prof. Charles McHenry (Univ. of Colorado) for pSJS9 and helpful discussions. We thank Prof. Mike O'Donnell (Rockefeller Univ.) for the plasmid expressing His-HMK-β, Daniel Jarosz for assistance in preparing Figure 2-7 (p.58), and Michael Simon for careful reading of the manuscript. This work was supported by grant number CA21615 from the National Cancer Institute, NIEHS Center Grant Number P30ES02109 from the MIT Center for Environmental Health Sciences, and an American Cancer Society Professorship to G.C.W. and a postdoctoral fellowship from the Damon Runyon Cancer Research Foundation to P.J.B. S.M.S. was supported by a Cleo and Paul Schimmel Fellowship. Some of this work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

CHAPTER 3

The Mutagenic Activity of DinB is Modulated in *Escherichia coli*

Veronica G. Godoy[1,3,*], Daniel F. Jarosz[2,*], Sharotka M. Simon[1], Alexev Abyzov[3], Valentin Ilyin[3], and Graham C. Walker[1]

Departments of Biology[1] and Chemistry[2], Massachusetts Institute of Technology, Cambridge, MA 02139; [3]Department of Biology, Northeastern University, Boston, MA 02115

Corresponding author: Graham C. Walker (gwalker@mit.edu), Department of Biology, 68-633, 77 Massachusetts Ave., Cambridge, MA 02139. Phone: 617-253-6716, Fax: 617-253-2643. E-mail:

*These authors contributed equally to this work

Running Title: $UmuD_2$ Modulates the Mutagenic Function of DinB

This document has been submitted to *Molecular Cell.*

Summary

DinB is the only mutagenic Y-family DNA polymerase conserved among bacteria, archaea, and eukaryotes. DinB and its orthologs possess a specialized lesion-bypass function but also display potentially deleterious -1 frameshift mutagenic phenotypes when overproduced. We show that the mutagenic capability of *E. coli* DinB is genetically separable from its translesion synthesis function and that the DNA damage inducible proteins RecA and $UmuD_2$ act in concert to modulate its potentially mutagenic activity. Structural modeling suggests that the relatively open active site of DinB is enclosed by interaction with these proteins. Intriguingly, residues that define the $UmuD_2$ interacting surface on DinB statistically co-vary throughout evolution, suggesting a driving force for the maintenance of a regulatory protein-protein interaction at this site. Together, these observations indicate that proteins like RecA and $UmuD_2$ may be responsible for managing the mutagenic function of DinB orthologs throughout evolution.

Introduction

Decades after their discovery [71, 76], DinB (DNA pol IV) [73] and UmuD′$_2$C (DNA pol V) [75, 108] were shown to be specialized Y-family DNA polymerases that catalyze the insertion of deoxyribonucleoside triphosphates (dNTPs) opposite to potentially lethal replication blocking lesions in a process termed translesion synthesis (TLS) [15, 58]. TLS can proceed with a range of fidelities [58], but in all cases Y-family polymerases replicate undamaged DNA with a reduced fidelity relative to the enzymes that replicate the majority of the genome [279]. Therefore, Y-family polymerases must be excluded from improper access to the replication fork to maintain genomic integrity [23].

DinB is the only Y-family DNA polymerases conserved among all domains of life [15] and under conditions of DNA damage it is the most abundant DNA polymerase in *E. coli* [136]. We recently elucidated the function of DinB and its orthologs in proficient and accurate TLS past DNA damage at the N^2 position of deoxyguanosine [53]. On the surface, these observations seemed incompatible with known, albeit enigmatic, DinB associated -1 frameshift mutagenic phenomena [280-282]. Although the molecular details of these phenomena have been the subject of intense debate [283-285], the role of elevated DinB function is clearly established [281].

Both *dinB* and *umuDC* are regulated transcriptionally by the SOS regulatory network [286], and biochemical regulation of the UmuD′$_2$C is remarkably complex and relatively well characterized. The RecA nucleoprotein filament promotes *umuDC* expression, the conversion of UmuD$_2$ to UmuD′$_2$ required for activation of UmuC polymerase function [75, 108] and activation of UmuD′$_2$C *in trans* [111, 287] via multiple protein-protein interactions. RecF, RecO, and RecR also cooperate to alleviate

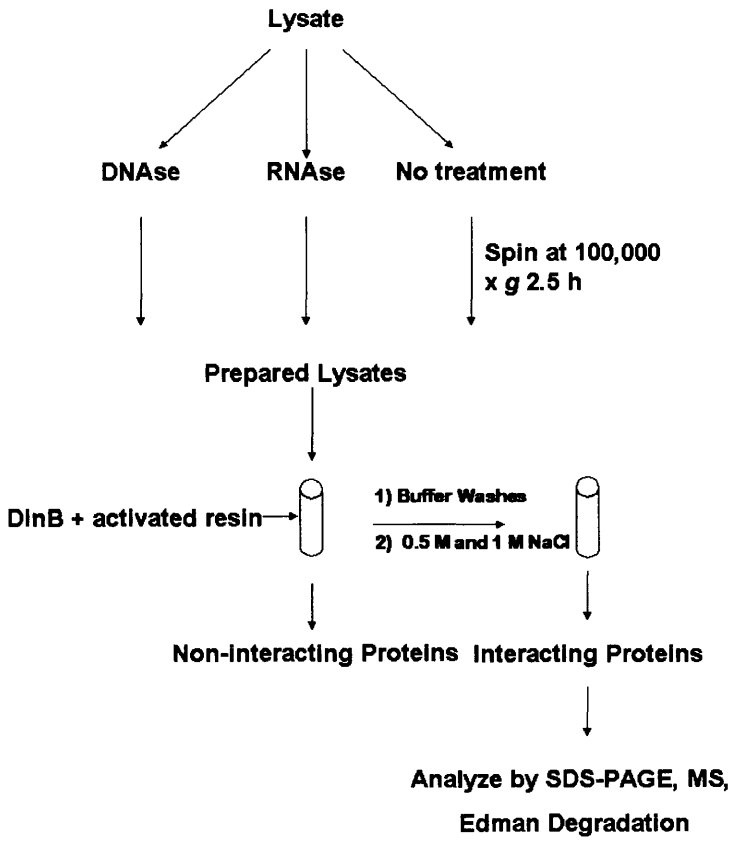
the inhibition of UmuD′$_2$C mediated TLS that is brought about by DNA pol III [288]. Despite the remarkable evolutionary conservation of DinB, the details of its biochemical regulation are comparatively unknown [289]. Previous studies have focused on the pivotal role of processivity clamps in regulation of TLS and mutagenesis promoted by DinB, UmuD′$_2$C, and other Y-family DNA polymerases in both prokaryotic and eukaryotic systems [18, 290-296]. However, regulation of DinB function is crucial given its deleterious -1 frameshift mutagenic signature relative abundance [136, 297]. We therefore sought to identify and characterize additional regulatory factors that might manage this potentially problematic function. Remarkably, we found that UmuD, UmuD′, and RecA regulate both the activity and mutagenic properties of DinB via protein-protein interactions that enclose its active site. Intriguingly, features of these mechanisms may be maintained in eukaryotes, suggesting a common pattern of regulation for these DNA polymerases through evolution.

Results

DinB Interacts with Numerous Cellular Factors

To identify proteins that regulate DinB function we employed an affinity chromatography scheme in which purified recombinant DinB was covalently coupled to an affinity resin (Figure 3-S1, p.68). Lysates from constitutively SOS-induced *E. coli* were passed over the DinB column and interacting proteins were eluted and separated by SDS-PAGE. Those that bound in a DinB dependent fashion were excised and identified. Included in this set of proteins were the chaperones GroEL and DnaK and also the ribosomal protein L3. The presence of the chaperones was anticipated since GroEL has been shown to regulate the function of both DinB [298] and UmuD′$_2$C (Pol V) [299] and DnaK regulates

Figure 3-S1: **Experimental scheme for identifying DinB interaction partners**



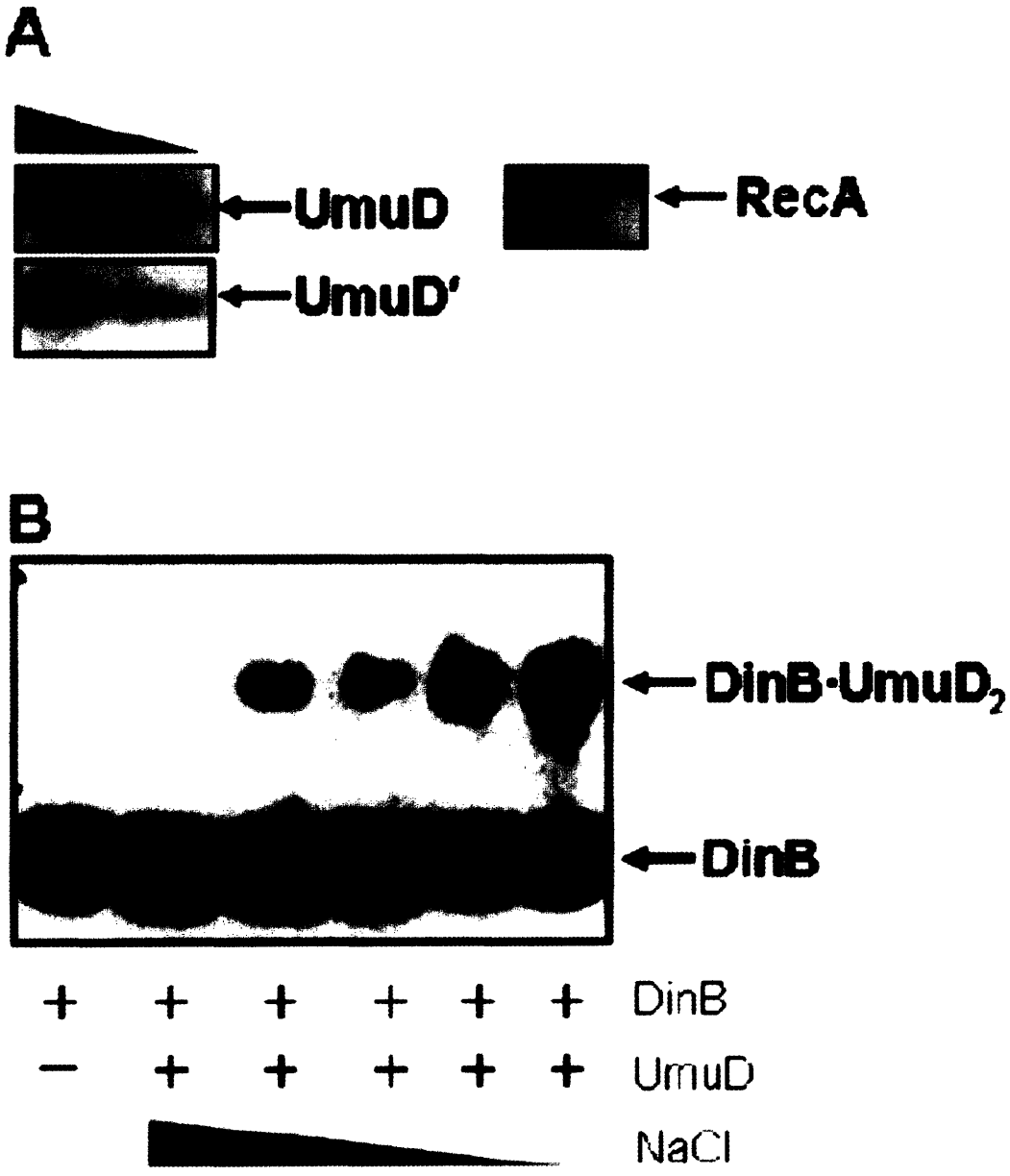
UmuC levels *in vivo* [300]. Ribosomal protein L3 seems less likely to be a *bona fide* regulator of DinB function although this remains a formal possibility. However, we also identified UmuD, UmuD′, and RecA as DinB interacting proteins. This was a complete surprise because, despite their intensively studied highly nuanced roles in regulating UmuC function [286, 297, 301], none of these factors have hitherto been implicated in regulating DinB function aside from the indirect role of RecA in mediating DinB induction via the SOS regulatory network. We therefore investigated the ability of these proteins to affect DinB function *in vivo* and *in vitro*.

DinB Forms a Stable Interaction with $UmuD_2$ and RecA

To ascertain whether the interactions we observed between DinB, RecA, UmuD, and UmuD′ were direct, we performed a farwestern blot in which we probed membranes containing UmuD, UmuD′, and RecA with DinB. Each of the interactions appears to be direct in nature (Figure 3-1A, p.70). To analyze the stoichiometry of the DinB·UmuD interaction, we crosslinked DinB and UmuD with formaldehyde and analyzed the products by immunoblot using an antibody against DinB. The crosslinked species corresponds to the molecular weight of a DinB·$UmuD_2$ complex and the reaction appears to be inhibited by high concentrations of NaCl (Figure 3-1B, p.70), suggesting that the interface may partly involve ionic or polar interactions. The propensity of RecA to multimerize [302] made it difficult to establish the stoichiometry of the DinB·RecA interaction.

To test whether DinB, $UmuD_2$ and/or $UmuD'_2$, and RecA form a stable ternary complex in solution, a DinB variant with a hexahistidine affinity tag at its C-terminus

Figure 3-1: **DinB interacts directly with $UmuD_2$ and RecA.**



Chapter 3-1 (continued): **DinB interacts directly with $UmuD_2$ and RecA**.

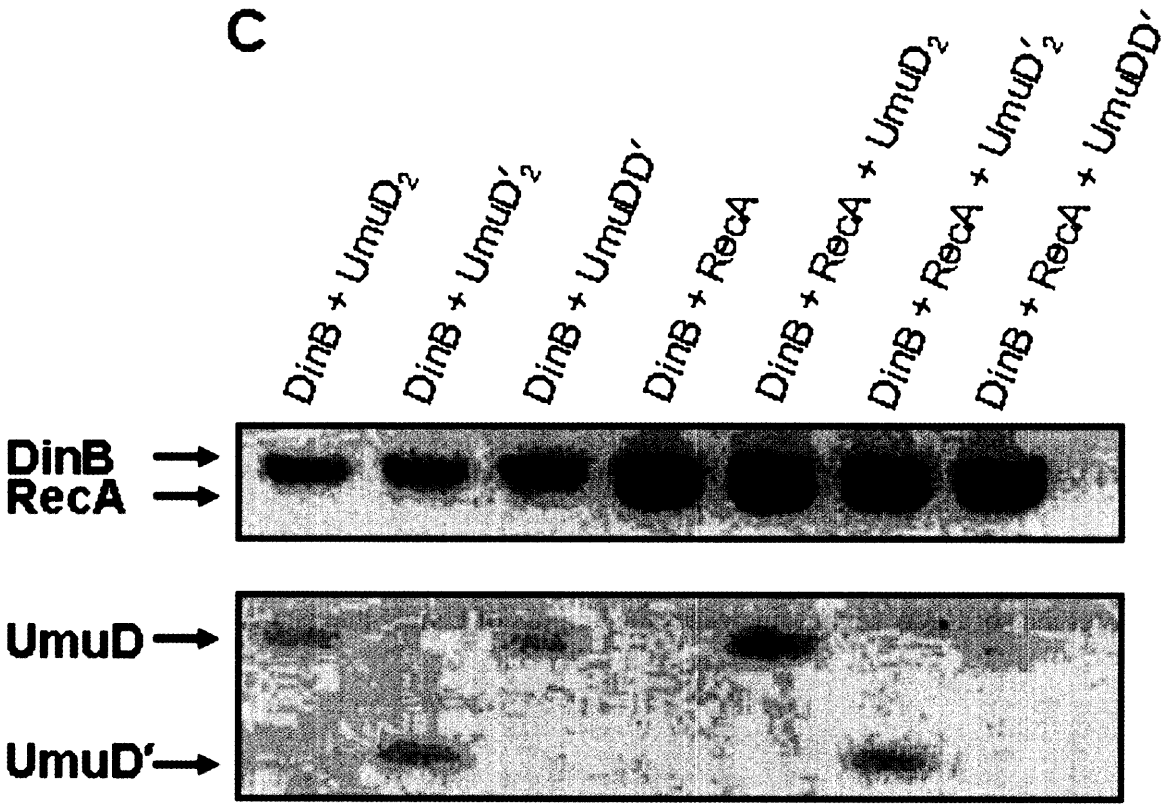


Figure 3-1 (continued): **DinB interacts directly with $UmuD_2$ and RecA**. (A) Farwestern blot demonstrates that UmuD directly interacts with ^{32}P-labeled $(His)_6$HMK-DinB. Either 50 or 100 pmols of UmuD or UmuD′ were separated by 12% SDS-PAGE and transferred to a PDVF membrane. The HMK-DinB protein probe was radioactively labeled and incubated with the membrane for 15 min. after which the membrane was exposed to film. (B) Crosslinking experiment suggests that DinB interacts with the $UmuD_2$ homodimer. One hundred pmol of DinB and $UmuD_2$ were mixed in a 10 μL volume in 50 mM Hepes pH 7.5, 25-500 mM NaCl, and 1 mM DTT and incubated for 10 min at 25°C. **C.** DinB forms a stable binary and ternary complexes with RecA and $UmuD_2$. DinB$(His)_6$ pulls down UmuD, UmuD′, and RecA on a Ni^{2+} affinity resin. The presence of RecA slightly increases the amount of UmuD that is recovered.

(DinB(His)$_6$) was incubated with RecA – both alone and in combination with UmuD$_2$, UmuD′$_2$, and the heterodimeric species UmuDD′ [121]. Complexes that formed with DinB were isolated using Ni^{2+} affinity resin. Using approximately physiological concentrations (2.5 μM) of each protein, we determined that DinB interacts stably with each of the UmuD species used in the experiment. We also observed the formation of a stable stoichiometric complex between DinB and RecA (Figure 3-1C, p.70), despite the previous report that fluorescence anisotropy does not reveal such an association [287]. Moreover, it appears that RecA stimulates DinB association with UmuD, but not UmuD′ or UmuDD′, by approximately two-fold (Figure 3-1C, p.70). Taken together, these data indicate that DinB, RecA, and UmuD$_2$ and to a lesser extent UmuD′$_2$, can form ternary complexes under physiological conditions.

UmuD Suppresses DinB-dependent Mutagenic Phenomena *in vivo*

We investigated the influence of the interactions we observed on the TLS function of DinB. Surprisingly, deletion of *umuD* has no effect on NFZ resistance. Although *ΔrecA* strains do exhibit sensitivity to this agent (data not shown), the multitudinous roles of RecA [302] complicate the interpretation of this result. Because the cellular levels of UmuD are much higher than those of UmuC and instead mirror those of DinB [297], we also investigated the consequences of overproducing the *umuD* gene products. Strikingly, we observed that overproduction of UmuD and to a lesser extend UmuD′ profoundly inhibits the ability of DinB to promote -1 frameshift mutagenesis (Figure 3-2A, p.74). Indeed, overproduction of UmuD reduces the frequency of CC108 -1 frameshift mutagenesis nearly to the extent that it is DinB dependent, while UmuD′

Figure 3-2: **$UmuD_2$ regulates the -1 frameshift activity of DinB *in vivo*.**

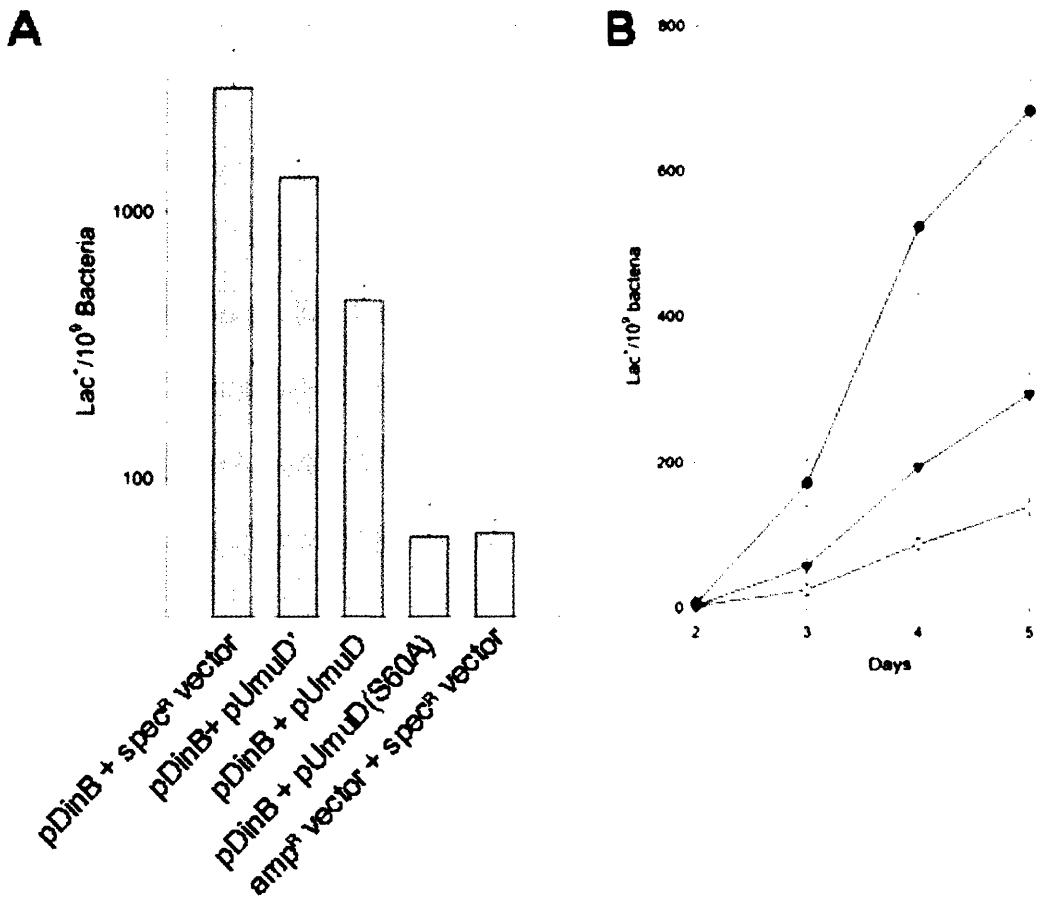


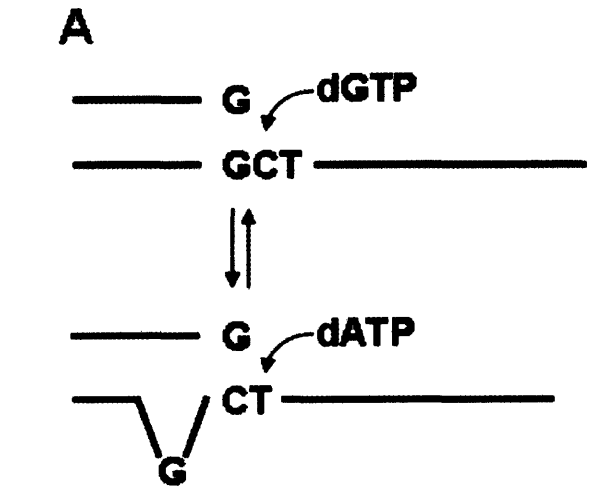
Figure 3-2: **$UmuD_2$ regulates the -1 frameshift activity of DinB *in vivo*. (A)** Lac^+ reversion assay of the frameshift allele in CC108 demonstrates that $UmuD_2$ modulates DinB's -1 frameshift function (B) $UmuD_2$ also affects the number of Lac^+ revertants in an adaptive mutagenesis experiment.

causes a more modest reduction. A noncleavable $UmuD_2$ variant, UmuD(S60A), was able to fully suppress DinB dependent frameshift mutagenesis (Figure 3-2A, p.74), suggesting that full-length UmuD is sufficient for maximal inhibition.

We then considered that the *umuD* gene products might also modulate the phenomenon of adaptive mutagenesis [303], which is dependent on DinB promoted -1 frameshifts during stationary phase [281, 304]. Under the conditions required to observe such mutagenesis DinB levels are elevated by approximately four-fold [136]. Remarkably, overproduction of UmuD or UmuD′ profoundly reduced adaptive mutagenesis (Figure 3-2B, p.74). The UmuD-dependent 5-fold reduction is comparable to the degree of adaptive mutagenesis that is DinB dependent [281, 304]. These observations suggest that the *umuD* gene products are able to modulate the -1 frameshift activity of DinB in both exponential and stationary phase.

To investigate whether DinB reciprocally affects UmuD function, we examined the effect of DinB overproduction on UV-induced mutagenesis, a *umuDC*-dependent phenomenon [76]. Coexpression of *dinB* from a low copy number plasmid suppressed UV-induced mutagenesis by a factor of 8.9 ($\pm$ 2.2). This effect is likely related to DinB dependent inhibition of $UmuD_2$ autocleavage, which we observed *in vitro* (Figure 3-5E, p.76). These observations are consistent with the notion that DinB interacts with $UmuD_2$ and in doing so sequesters it from an interaction, perhaps with the RecA-nucleoprotein filament, that facilitates its autoproteolysis.

Figure 3-5: **$UmuD_2$ and RecA directly modulate DinB -1 frameshift function.**



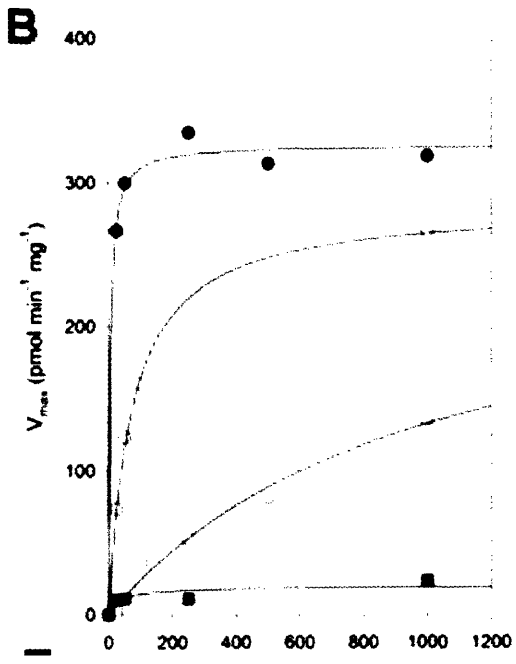


Figure 3-5 (continued): **$UmuD_2$ and RecA directly modulate DinB -1 frameshift function.**

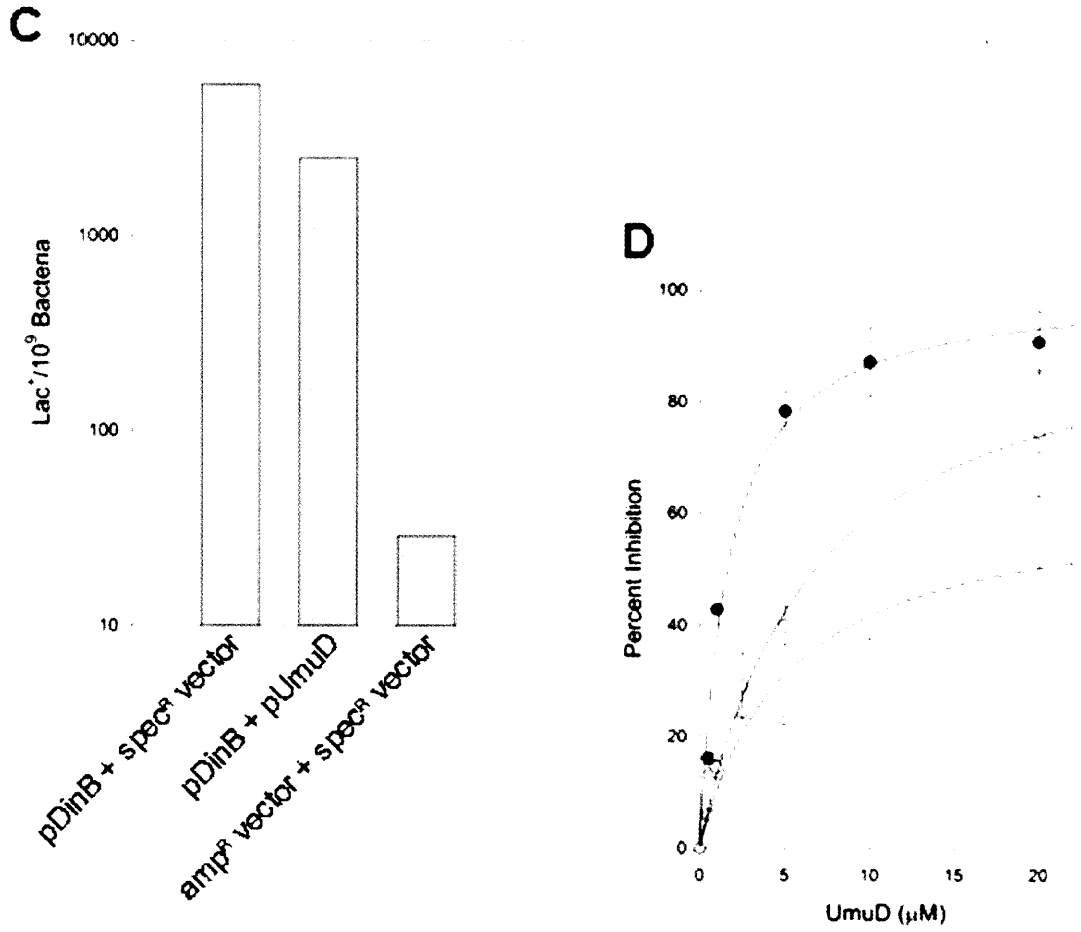


Figure 3-5 (continued): **$UmuD_2$ and RecA directly modulate DinB -1 frameshift function.**

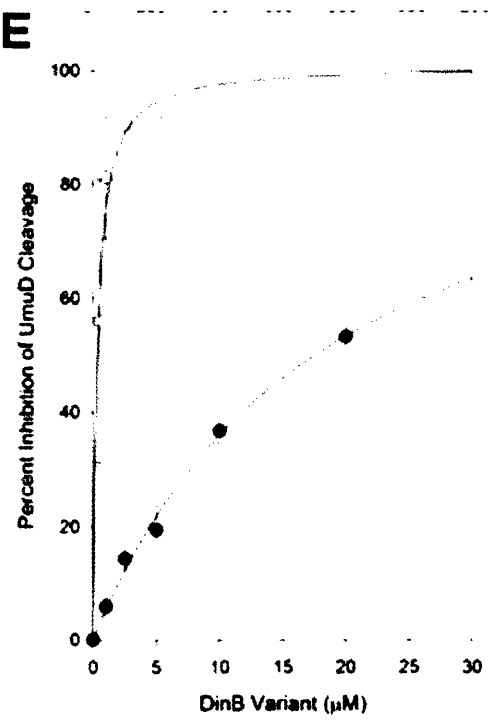


Figure 3-5 (continued): **$UmuD_2$ and RecA directly modulate DinB -1 frameshift function.** (A) Schematic of a mismatched DNA substrate that can be extended either by dGTP to generate a full length product or by dATP, thereby forming a dNTP stabilized misalignment and generating a -1 frameshift product. (B) Plot of reaction velocity vs. dNTP substrate concentration for DinB alone and in combination with RecA and $UmuD_2$. Extension of the GG mismatch by DinB alone with dATP (open squares) is detectable but weaker than extension of a GC basepair in the same sequence context (open circles). The addition of RecA in stoichiometric ratios with DinB and saturating (10 μM) $UmuD_2$ profoundly inhibits DinB activity on a GG mismatch (closed squares) but stimulates DinB activity on a GC (closed circles) by more than 20-fold. (C) *recA*$^+$ is required for UmuD dependent inhibition of DinB promoted -1 frameshift mutagenesis *in vivo*. Overeproduction of DinB promotes -1 frameshift mutagenesis in a *ΔrecA* background but the co-overproduction of $UmuD_2$ has little effect on mutation frequency. (D) Plot of percent frameshift inhibition vs. UmuD variant concentration. The frameshift activity of DinB is efficiently inhibited by $UmuD_2$ (closed circles) but the frameshift activity of DinB(F172A) is more inert to $UmuD_2$ supression (open circles). The UmuD(D91A) variant is also very inefficient at inhibiting the -1 frameshift activity of wild-type DinB (open triangles). (E) DinB efficiently inhibits $UmuD_2$ autocleavage *in vitro*. Wild-type DinB (open circles) inhibits $UmuD_2$ autocleavage far better than DinB(F172A) (closed circles).

Identification of the Molecular Interface Between DinB and $UmuD_2$

Our observation that UmuD suppresses the -1 frameshift activity of DinB *in vivo* was especially intriguing in light of structural studies of archaeal DinB homologs, which have been shown to possess remarkably open active sites. We therefore analyzed the interaction between DinB and $UmuD_2$ using cellulose filter peptide arrays. The membranes were probed with either DinB or $UmuD_2$, and interacting peptides were identified and mapped onto structural models of DinB [53] or $UmuD_2$ [134, 147]. Interestingly, the $UmuD_2$ interacting peptides on DinB localize to a single face of the protein (Figure 3-3A, p.81). Further, the presence of an extended interacting surface on DinB suggest that its interaction with $UmuD_2$ is qualitatively different from its interaction with the β processivity clamps, which depends on a highly conserved peptide motif [294, 305]. Most intriguingly, the interaction interface suggests that $UmuD_2$ may enclose the relatively open active site of DinB, which has been proposed to play a crucial role in its mutagenic function [45, 306]. The DinB interacting interface forms a somewhat less contiguous surface when mapped onto a $UmuD_2$ model. However, the DinB binding interace forms a contiguous surface when mapped onto certain isoenergetic models of $UmuD_2$ [147], in which its N-terminal arm is raised to reveal an interacting surface across the side of the protein (Figure 3-3B, p.81). These observations may hint at biological function for alternative $UmuD_2$ conformers.

In an effort to design a DinB variant that is unable to interact with $UmuD_2$, we identified a strongly interacting group of peptides from the DinB peptide array and examined conservation of this region in numerous *umuD*-containing organisms. Three residues, P166, F172, and L176 were strikingly conserved and we determined the effect

Figure 3-3: **Molecular characterization of the interaction between DinB and $UmuD_2$.**

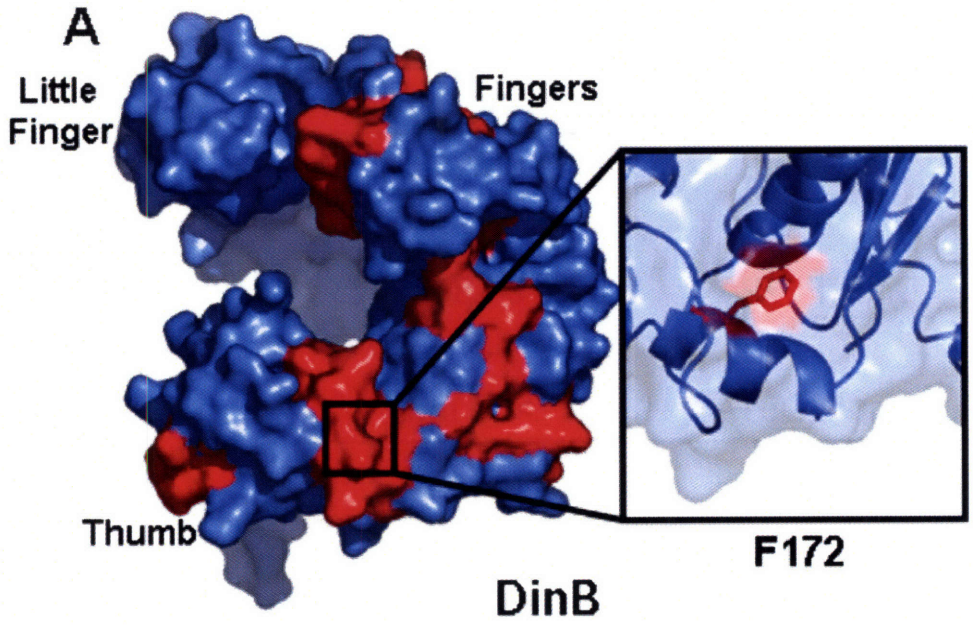


Figure 3-3 (continued): **Molecular characterization of the interaction between DinB and $UmuD_2$.**

C

DinB/UmuD Variant	K_d for $UmuD_2$
WT DinB	0.62 μM
DinB(F172A)	34.8 μM
UmuD(D91A)	>15 μM?

D

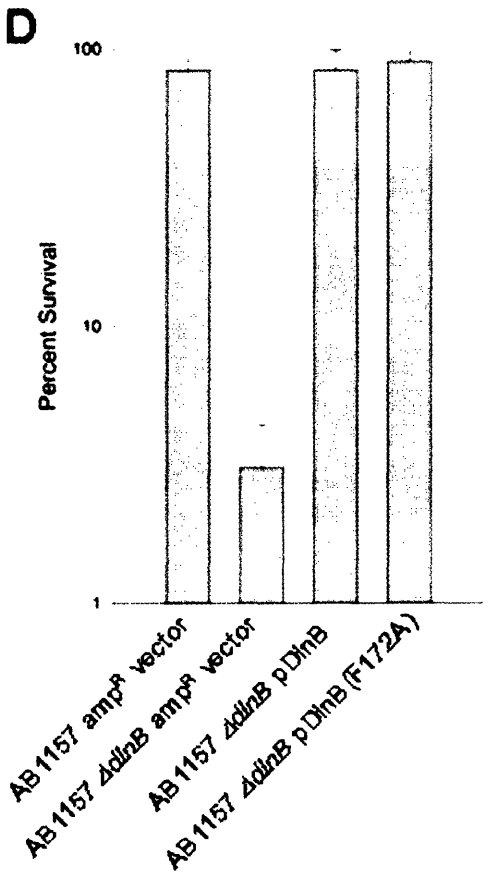


Figure 3-3 (continued): **Molecular characterization of the interaction between DinB and $UmuD_2$.** (A) Peptide array mapping of the UmuD binding interface on DinB reveals a surface composed of the thumb and finger domains of the polymerase. Several hydrophobic residues in the most strongly interacting peptide are conserved among DinB orthologs from organisms containing *umuD*. (B) Peptide array mapping of the DinB binding interface on $UmuD_2$ reveals a discontinuous interface on a structural model of *trans*-$UmuD_2$ that is rendered contiguous in an alternative isoenergetic *trans*-$UmuD_2$ conformer. (C) Alanine mutants of DinB Phe172 or UmuD Asp 91 result in a weakened interaction determined by fluorescence spectroscopy. (D) A low-copy number plasmid encoding DinB(F172A) is fully able to rescue the NFZ sensitivity of a *ΔdinB E. coli* strain.

of changing each residue to an alanine (Figure 3-3A, p.81). Although the mutant proteins DinB(P166A) and DinB(L176A) were insoluble (data not shown), we were able to express and purify DinB(F172A) in soluble form (Figure 3-S3, p.85). Moreover, we found that the *dinB* allele encoding DinB(F172A) complements the NFZ sensitivity of a *ΔdinB* strain (Figure 3-3D, p.81), indicating that this mutant is TLS proficient *in vivo*. A reciprocal approach was used to generate a $UmuD_2$ mutant that would be impaired with respect to its ability to interact with DinB. The variant UmuD(D91A) is soluble and proficient for facilitated autoproteolysis (Figure 3-S4, p.86).

We determined that the dissociation constant between DinB and $UmuD_2$ is approximately 0.620 μM using fluorescence spectroscopy (Figure 3-3C, p.81; Figure 3-S5, p.87). Provocatively, the levels of $UmuD_2$ rise from *ca.* 0.35 μM under non-SOS-induced conditions to *ca.* 4.5 μM under conditions of SOS-induction [136], indicating that DinB and $UmuD_2$ are capable of interaction within the range of physiologically relevant concentrations. Futhermore, the DinB(F172A) and $UmuD(D91A)_2$ proteins were each greatly impaired with respect to their ability to bind $UmuD_2$ and DinB, respectively (Figure 3-3C, p.81).

Mutation of the Interface Between DinB and $UmuD_2$ Impairs Function *in vivo*

To determine whether the physical interaction between DinB and $UmuD_2$ that we observed *in vitro* is important for modulation of DinB dependent frameshift mutagenesis *in vivo*, we examined whether the -1 frameshifts produced by DinB(F172A) could be inhibited by UmuD. Overproduction of DinB(F172A) results in an increase in -1 frameshift mutagenesis by approximately 6-fold (Figure 3-4A, p.88). Unlike the

Figure 3-S3: **Purified DinB(F172A) is soluble and migrates identically to wild-type DinB on SDS-PAGE.**

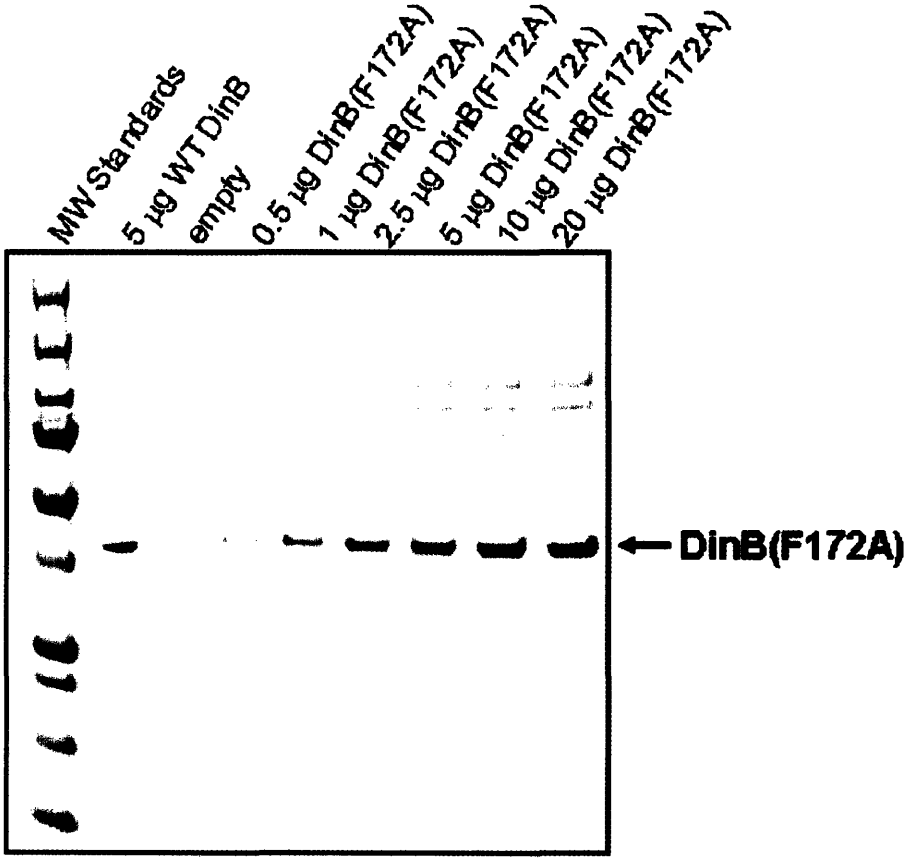
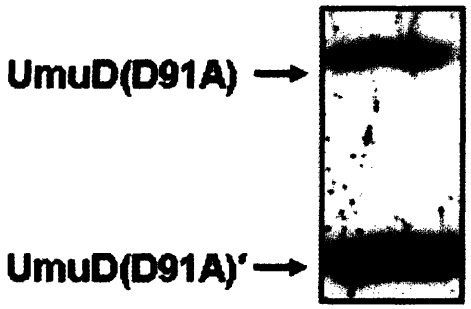


Figure 3-S4: **Purification of UmuD(D91A).**



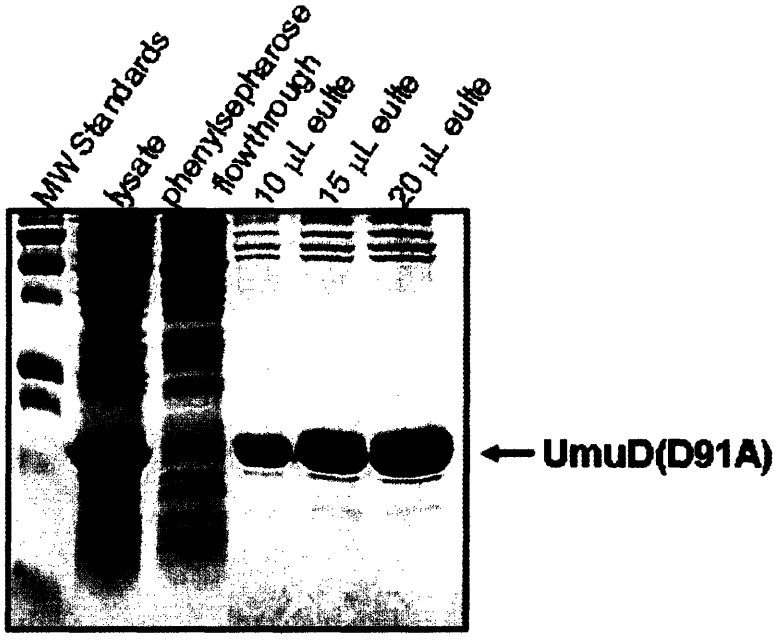


Figure 3-S5: **Interaction between $UmuD_2$ and DinB or DinB(F172A).**

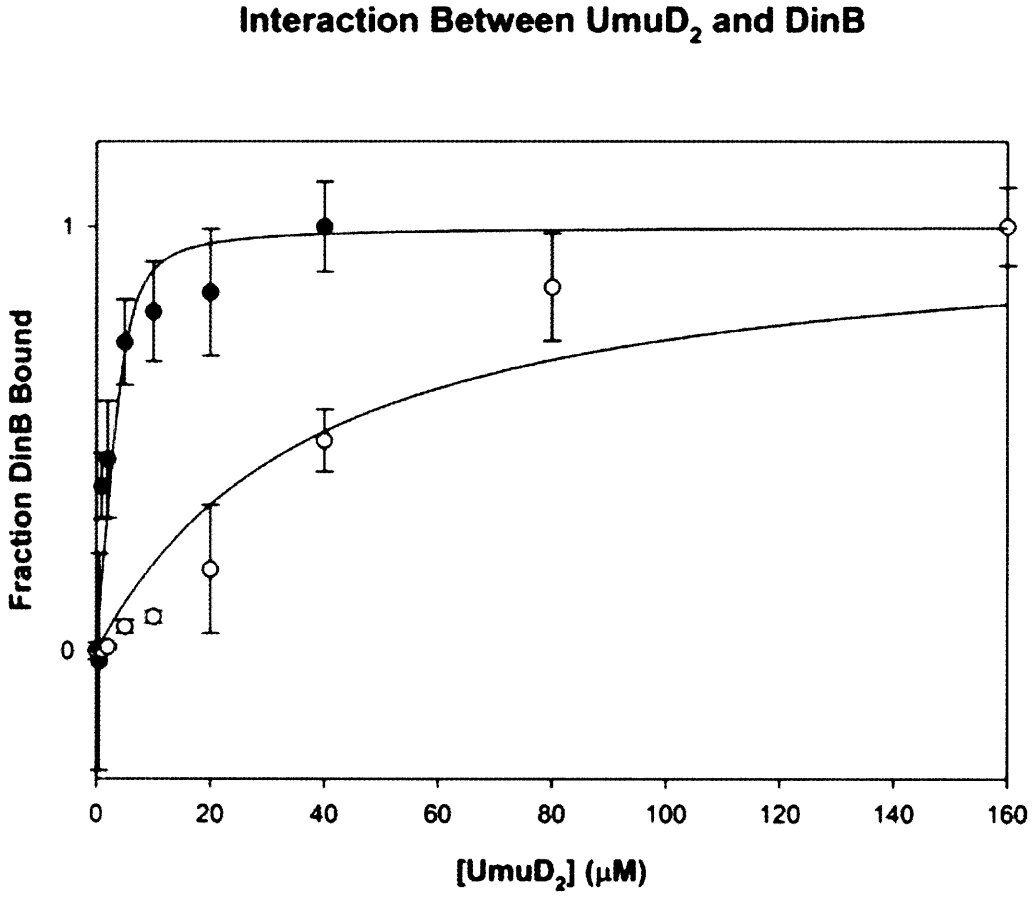


Figure 3-4: **Single amino acid changes on the interface between DinB and UmuD perturb regulation of -1 frameshift activity.**

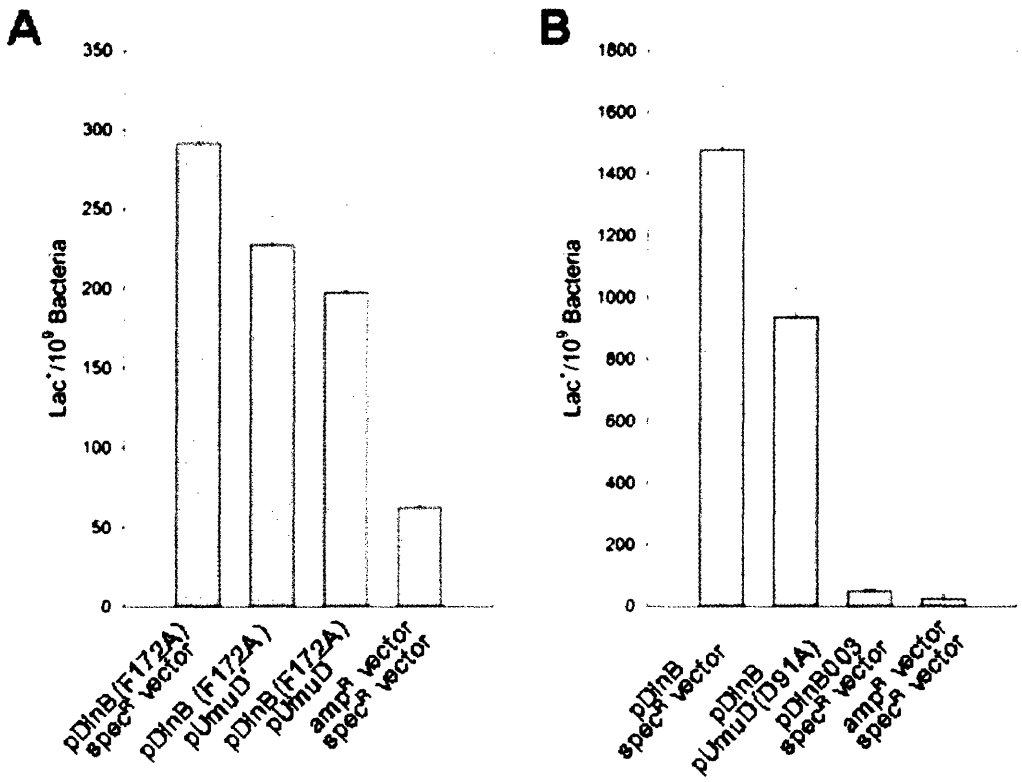


Figure 3-4: **Single amino acid changes on the interface between DinB and UmuD perturb regulation of -1 frameshift activity.** (A) The DinB(F172A) variant has a lower affinity for $UmuD_2$ and is not as responsive as wild type DinB to regulation by $UmuD_2$. (B) The UmuD(D91A) variant has a lower affinity for DinB and does not regulate -1 frameshift activity as well as wild type UmuD.

situation with wild-type DinB (Fig 2A), co-overproduction of UmuD or UmuD′ does not substantially reduce the -1 frameshift mutation frequency (Figure 3-4A, p.88). These data suggest that a direct interaction between DinB and $UmuD_2$ or $UmuD'_2$ at the interface we have identified is important for modulation of -1 frameshift mutagenesis.

Reciprocally we examined whether UmuD(D91A) could suppress the -1 frameshift mutagenesis promoted by overproduction of wild-type DinB. Overproduction of UmuD(D91A) only modestly suppresses -1 frameshift mutagenesis (*ca.* 1.5-fold vs. >25-fold for WT UmuD) (Figure 3-4B, p.88). Moreover, even significant overproduction of DinB(F172A) was insufficient to impair UV-induced mutagenesis, in contrast to wild-type DinB (data not shown). These observations suggest that the direct interaction between DinB and $UmuD_2$ is crucial for the ability of each protein to modulate the function of the other *in vivo*.

$UmuD_2$ Inhibits DinB-dependent -1 Frameshift Activity *in vitro*

To understand the mechanisms governing modulation of DinB function by UmuD *in vivo*, we reconstituted DinB dependent -1 frameshift activity *in vitro* with a substrate containing a G:G mispair that can be extended with either with dGTP to generate a full-length product or with dATP to generate a -1 frameshift product that is one nucleotide shorter than the template (Figure 3-5A, p.76) [73, 307]. DinB is unable to extend from this mispair using dGTP under our experimental conditions (data not shown). In contrast, DinB can readily act on this substrate using dATP albeit at a rate that is 10-fold lower than its ability to extend from a G:C basepair in the same sequence context (Figure 3-5B, p.76). We were initially surprised when we found that the addition of $UmuD_2$ alone did

not alter the -1 frameshift activity of DinB (data not shown). Unexpectedly, the addition of $UmuD_2$ nearly completely inhibits the reaction when RecA is added in a stoichiometric ratio with DinB (Figure 3-5B, p.76). Strikingly, the addition of $UmuD_2$ and RecA enhances the ability of DinB to extend from a G:C basepair in the same sequence context (Figure 3-5B, p.76). These observations indicate that $UmuD_2$ and RecA act in concert to sculpt DinB function in a highly nuanced manner. The requirement of RecA to reconstitute $UmuD_2$ modulation of DinB mutagenic function *in vitro* led us to examine whether *recA*$^+$ is required for suppression of -1 frameshift mutagenesis *in vivo*. Indeed, overproduction of UmuD was virtually unable to alter the frequency of frameshift mutagenesis in a *ΔrecA* strain (Figure 3-5C, p.76). Taken together, these findings indicate that RecA is required for $UmuD_2$ dependent modulation of DinB function.

In vitro, UmuD reduced the maximal -1 frameshift activity of wild-type DinB by one half at a concentration of 1.7 μM (840 nM $UmuD_2$), while a concentration of 7.1 μM (3.6 μM $UmuD_2$) is needed to cause an equivalent effect on DinB(F172A) (Figure 3-5D, p.76). UmuD(D91A) also shows a marked 10-fold decrease in its ability to inhibit DinB dependent -1 frameshift activity relative to wild-type $UmuD_2$. Additionally, we observed that DinB was able to inhibit the facilitated autoproteolysis of $UmuD_2$ *in vitro*, and moreover that the DinB(F172A) variant was unable to efficiently do so (Figure 3-5E, p.76). All of these data underscore the notion that a physical interaction between DinB and $UmuD_2$ exists under physiological conditions and is required for $UmuD_2$ dependent modulation of DinB function. Furthermore, perturbation of that interaction compromises the ability of $UmuD_2$ to control the -1 frameshift mutagenic function of DinB.

A TLS-deficient DinB Variant is Proficient for -1 Frameshift Function

It has been suggested that the -1 frameshift mutagenic signature of DinB arises from its open active site because of its specialized TLS function [45]. To ascertain whether DinB dependent -1 frameshift activity is separable from its function in TLS, we employed a DinB(F13V) variant, which is able to catalyze DNA synthesis on undamaged DNA, but is virtually unable to perform TLS on certain adducted templates [53]. We found that this mutant is able to promote -1 frameshift mutagenesis *in vivo* (Figure 3-6A, p.92), indicating that DinB function in TLS and frameshift mutagenesis can be separated. Curiously, the mutation frequency induced by overexpression of DinB(F13V) is about 5-fold greater than that produced by overexpression of wild-type DinB even though the levels of each protein are comparable *in vivo* (Figure 3-S6, p.92). This observation suggests either that DinB(F13V) has an increased -1 frameshift activity or that some other mechanism is responsible for this phenomenon *in vivo*. Although the first explanation is formally possible, the -1 frameshift activity of DinB(F13V) *in vitro* is slightly reduced (*ca.* 3-fold) relative to wild-type DinB. This observation suggests that DinB(F13V) promotes increased frameshifts by virtue of either its recruitment to, or association with, frameshift intermediates. Curiously, we were unable to observe $UmuD_2$ dependent inhibition of DinB(F13V) -1 frameshift activity either *in vivo* or *in vitro* (Figure 3-6 A-B, p.92). These observations may be at least in part due to the fact that DinB(F13V) has a reduced affinity for $UmuD_2$ relative to the wild-type enzyme (Figure 3-S7, p.95).

Figure 3-6: **A TLS deficient variant of DinB is proficient for -1 frameshift function.**

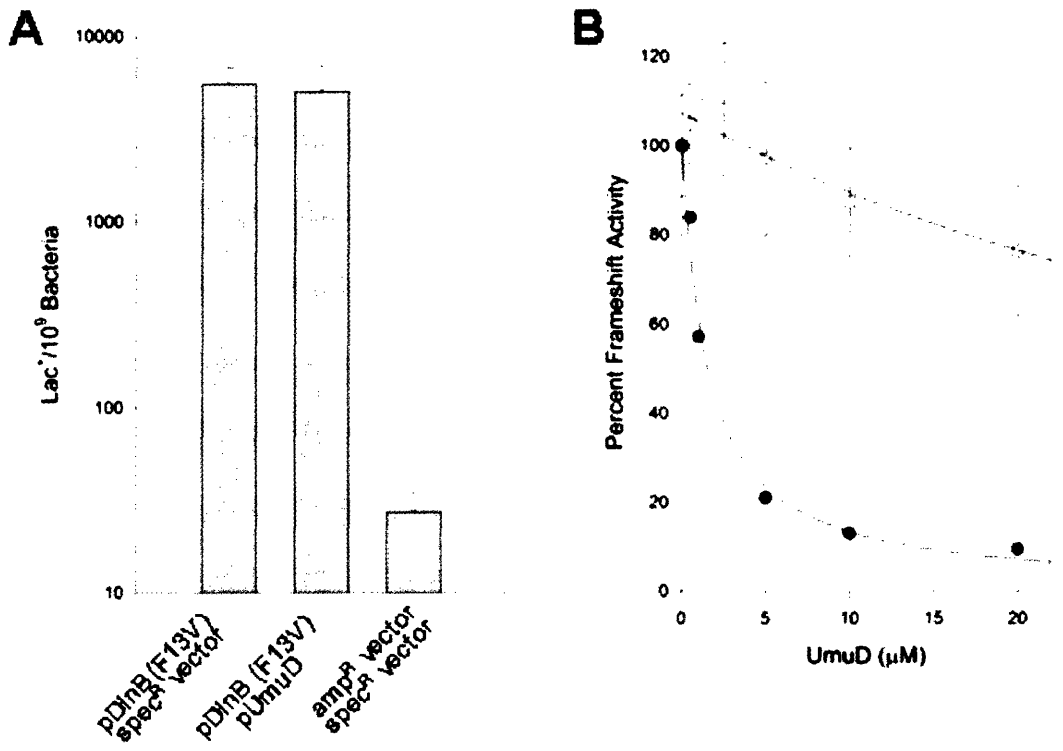
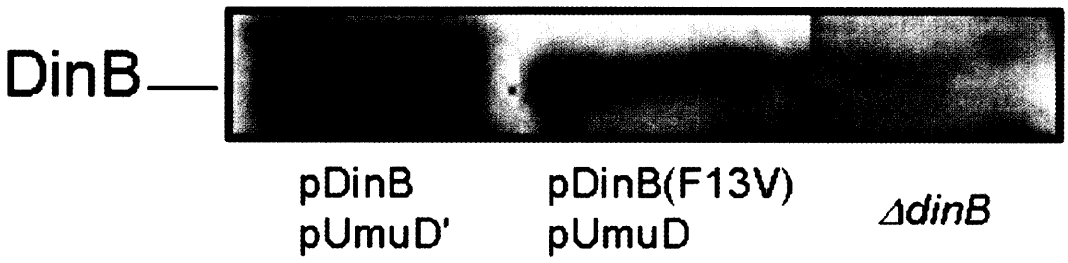


Figure 3-6: **A TLS deficient variant of DinB is proficient for -1 frameshift function.** (A) The DinB(F13V) variant can promote -1 frameshift mutagenesis but is not controlled by co-overproduction of UmuD. (B) The -1 frameshift activity of DinB(F13V) is poorly inhibited by $UmuD_2$ *in vitro*. Plot of frameshift activity vs. UmuD concentration indicates that DinB(F13V) (open squares) retains much of its frameshift activity at concentrations of UmuD that inhibit virtually all DinB frameshift activity (closed circles). All reactions contain RecA in stoichiometric ratios with DinB.

Figure 3-S6: **DinB(F13V) is expressed at levels comparable to wild-type DinB *in vivo*.**



RecA and $UmuD_2$ Modulate DinB Function by Restricting its Open Active Site

In an effort to understand how $UmuD_2$ suppresses the intrinsic -1 frameshift mutator activity of DinB in a RecA dependent manner, we generated a model of a ternary complex among DinB, RecA, and $UmuD_2$. The structure of RecA [308] and models of DinB [53] and $UmuD_2$ [147] were docked using several constraints. First, $UmuD_2$ was positioned on DinB using the peptide array data (Figure 3-3A-B, p.81), ensuring that there were no steric collisions between the two proteins. We then used distance constraints between RecA and $UmuD_2$ from published monocysteine crosslinking studies [88] to orient RecA relative to $UmuD_2$. Finally, we analyzed the RecA binding interface on DinB with a peptide array experiment (Figure 3-S8). Together, these data were used to generate a working model shown in Figure 3-7A-B (p.97), which indicates that RecA and $UmuD_2$ act in concert to enclose the relatively open active site of DinB, perhaps thereby restricting its mutagenic potential. It is also curious that the molecule of RecA is positioned appropriately to interact with the end of a RecA-nucleoprotein filament, suggesting that this interaction may also play a pivotal role in targeting DinB to RecA-coated substrates, a concept that has been previously proposed for $UmuD'_2C$ [110, 309].

Although the UmuD protein is only conserved among certain bacteria, we wondered whether the UmuD-binding interface on DinB might be maintained throughout evolution to interact with a highly diverged $UmuD_2$ or a different but functionally equivalent partner protein. The eukaryotic ortholog of DinB, pol κ, also promotes frameshift mutagenesis in overproduction [310], and it is possible that a putative eukaryotic protein might interact in a similar location on pol κ. Although residues on

Figure 3-S7: **Interaction between $UmuD_2$ and DinB(F13V)**

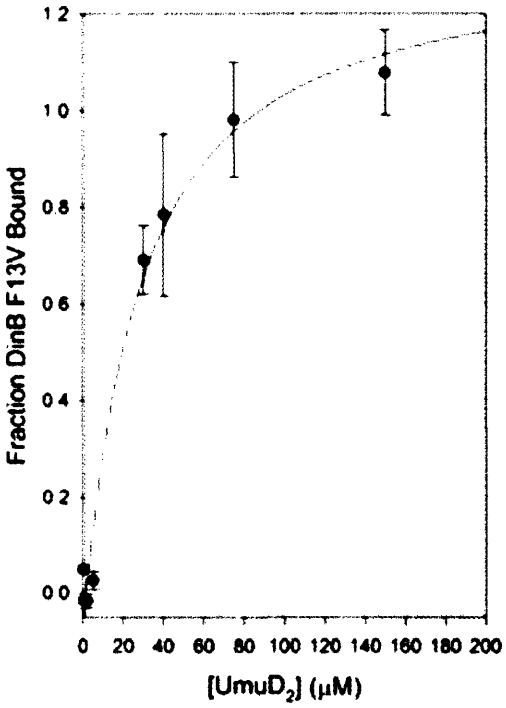


Figure 3-S8: **Predicted binding interface with RecA is shown in red on a model of DinB.**

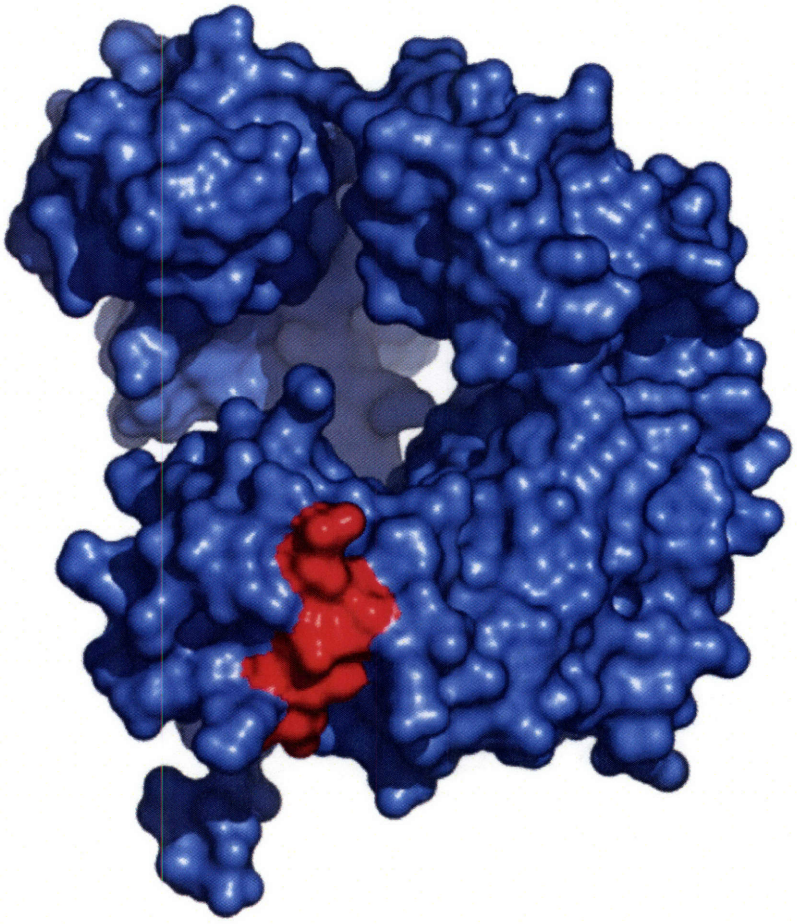


Figure 3-7: **RecA and $UmuD_2$ enclose the open active site of DinB.**

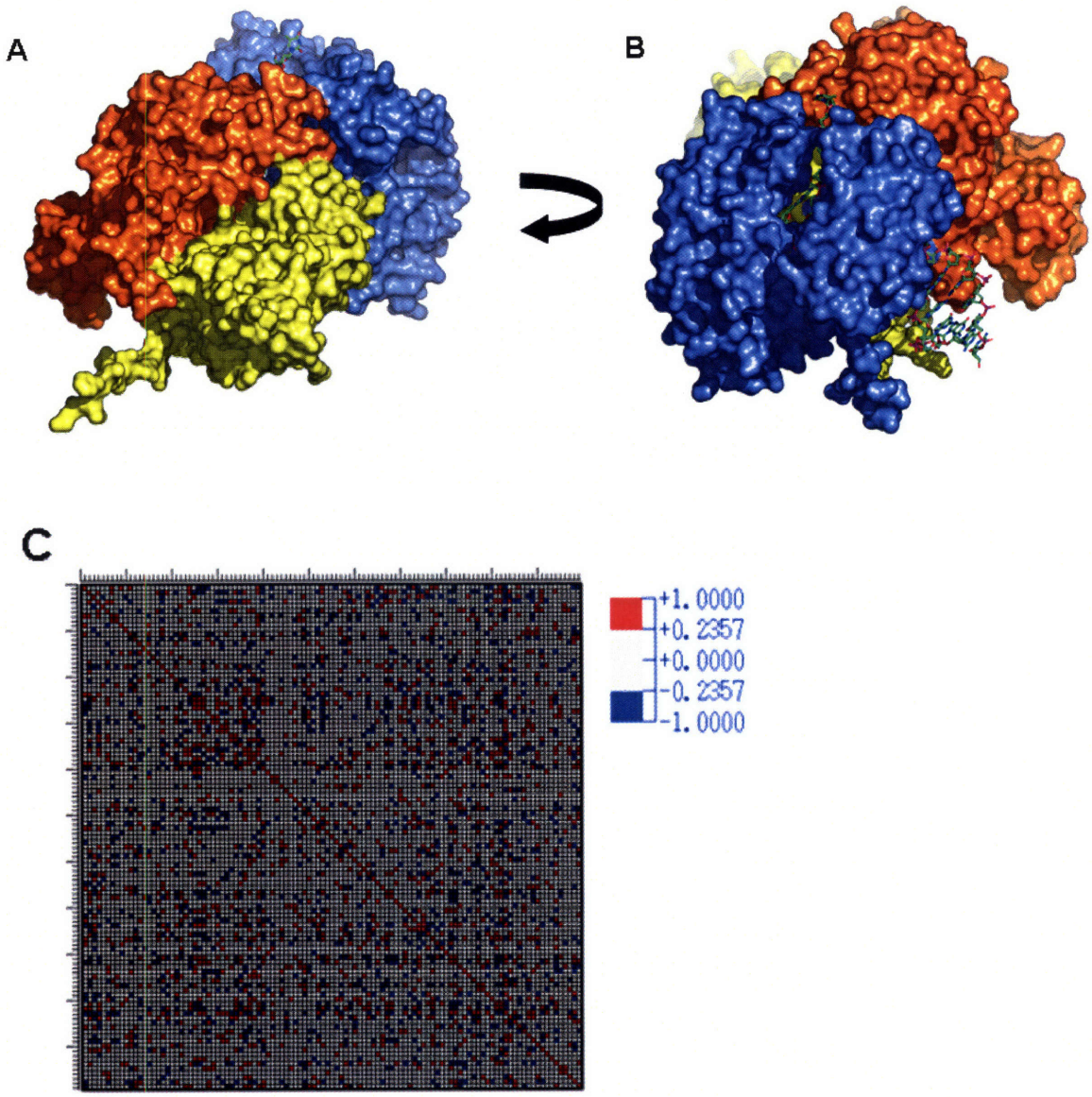


Figure 3-7 (continued):

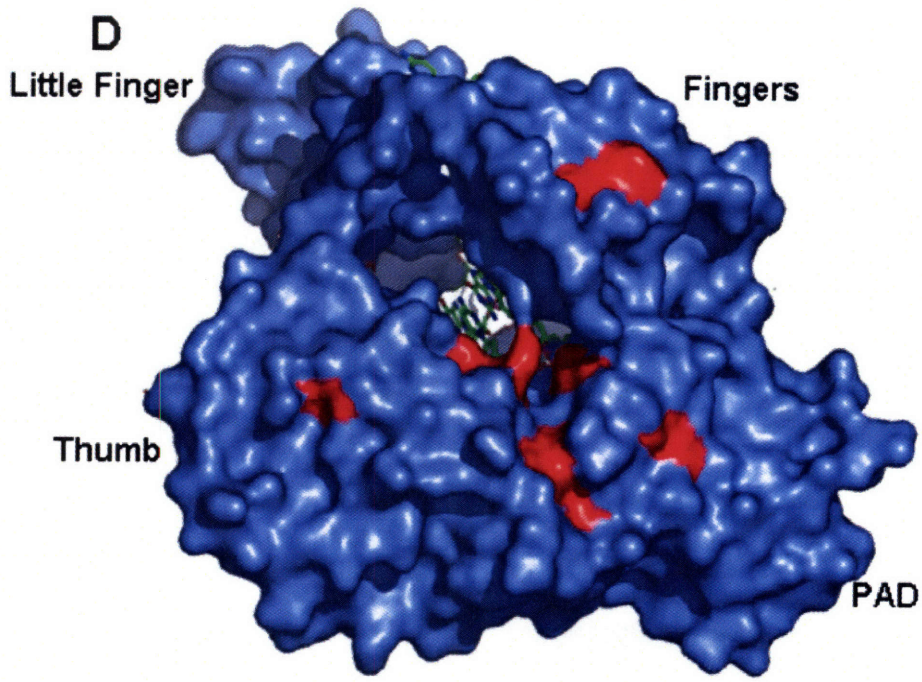


Figure 3-7: **RecA and $UmuD_2$ enclose the open active site of DinB. A-B.** *In silico* modeling of a ternary complex of the proteins. The surface representation of DinB is shown in blue, $UmuD_2$ in yellow, and RecA in orange. The DNA is relatively enclosed in the complex. **C-D.** Statistical covariance of DinB/pol κ residues across evolution. Residues that display statistical covariance with the $UmuD_2$ binding interface on *E. coli* DinB define an interface in a similar position on pol κ, suggesting a possible rationale for the maintenance of this interface as a site of regulatory protein-protein interactions.

DinB that participate in the interface are only cryptically conserved, we considered whether they would exhibit statistical covariance through evolution. We therefore assembled an alignment of 84 DinB and pol κ sequences from all domains of life and examined which positions showed statistically significant covariance with at least two of three residues that define the $UmuD_2$ interacting interface of DinB using the CRASP algorithm (Figure 3-7C, p.97) [311]. Strikingly, these residues define a similar interface on pol κ as that identified experimentally on DinB (Figure 3-7D, p.97). These observations suggest that there may be an evolutionary driving force for the maintenance of this interface, perhaps as a site for regulatory protein-protein interactions.

Discussion

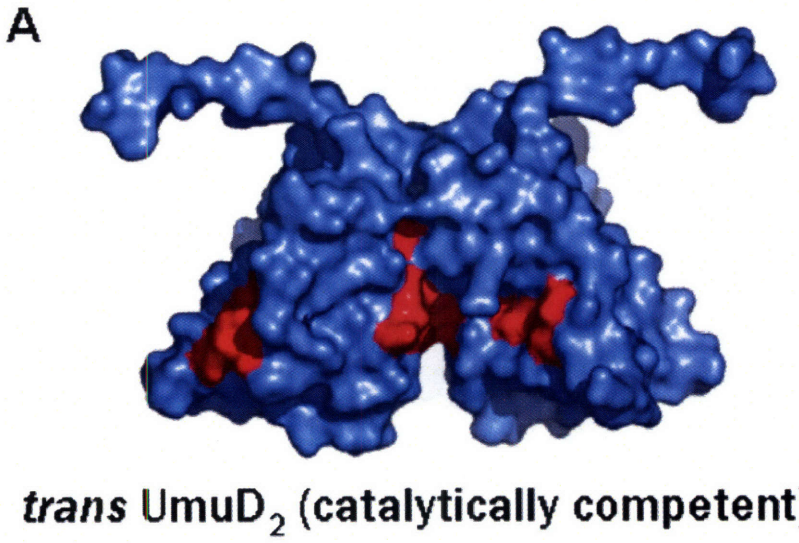
The SOS network regulates the transcription of numerous gene products, including Y-family DNA polymerases, in response to DNA damage and environmental stress. In addition to a demonstrated biochemical preference for adducted DNA substrates [43, 53] exquisite protein-protein interactions extending well beyond the replicative processivity subunit [23, 312] govern the activity of Y-family DNA polymerases. $UmuD'_2C$ is regulated by diverse cellular factors including RecA, $UmuD_2$, and $UmuD'_2$ [75, 98, 108, 111, 131, 313]. Strikingly, we report that DinB is subject to regulation by these same proteins *in vivo* and *in vitro*.

We observe direct interactions between DinB, $UmuD_2$, $UmuD'_2$, and RecA using several methods (Figure 3-1A-C, p.70) and importantly these interactions are physiologically relevant. The dissociation constant between DinB and $UmuD_2$ is 620 nM while the uninduced cellular concentrations of DinB and UmuD are approximately 400 nM and 350 nM respectively. Thus, there are sufficient cellular quantities of UmuD to

interact with nearly every molecule of DinB. The estimated K_d between DinB and RecA is *ca.* 1 μM and RecA concentration under normal conditions is 2.5 μM. The cellular levels for these proteins rise at least 10-fold upon SOS induction [96, 136, 314]. Thus, DinB is likely to exist as at least a binary and perhaps ternary complex under many physiological conditions. Indeed, we have been able to isolate a ternary complex of these three proteins that is stable on the minute timescale (Figure 3-1C, p.70).

We find that the -1 frameshift mutagenesis that is induced by overproduction of DinB can be suppressed by co-overproduction of UmuD (Figure 3-2A-B, p.74). Moreover, a noncleavable UmuD variant, UmuD(S60A), fully suppresses -1 frameshift mutagenesis, indicating that only the function of full-length UmuD is required to control it (Figure 3-2A, p.74). The DinB-binding deficient UmuD(D91A) variant (Figure 3-3C, p.81), although perfectly proficient at facilitated autoproteolysis (Figure 3-S4, p.86), does not efficiently reduce DinB dependent-1 frameshift mutagenesis (Figure 3-4B, p.88). Curiously, the residue Asp91 in UmuD does not form a contiguous interface with the other residues identified as part of the DinB-binding interface as it is occluded by the curled N-terminal arm of UmuD (Figure 3-3B, p.81). However, computational modeling has suggested that an isoenergetic conformation of $UmuD_2$ can be formed in which the N-terminal arm of $UmuD_2$ is raised [147], thereby potentially favoring interaction with DinB but also rendering it incompetent for autoproteolysis (Figure 3-S2, p.101). Our observation that DinB inhibits $UmuD_2$ autoproteolysis *in vitro* and UV-induced mutagenesis *in vivo*, is consistent with this notion. Overproduction of UmuD′ can also suppress -1 frameshift mutagenesis, albeit to a lesser degree, indicating that the

Figure 3-S2: **Models of $UmuD_2$ with DinB-interacting residues highlighted in red.**



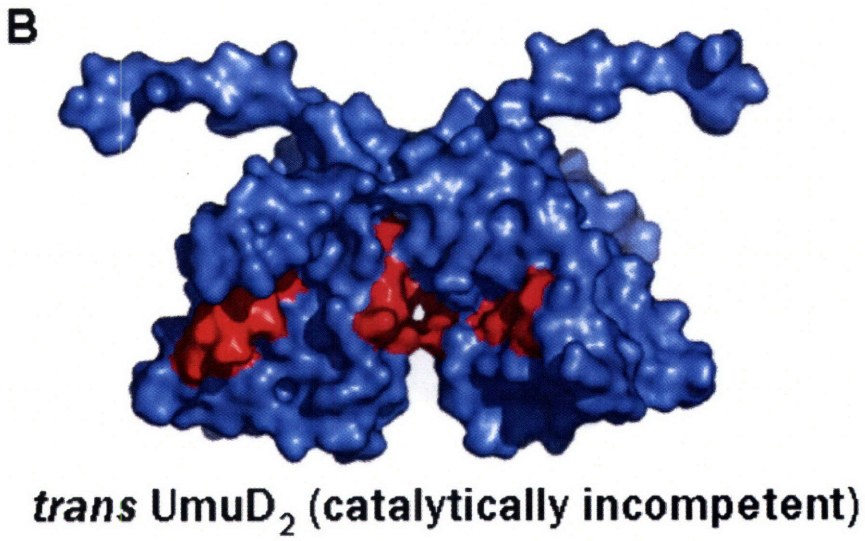
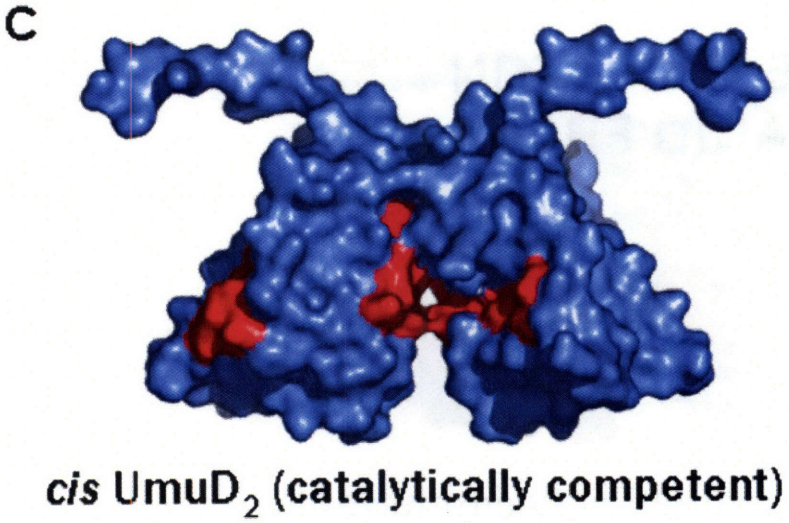
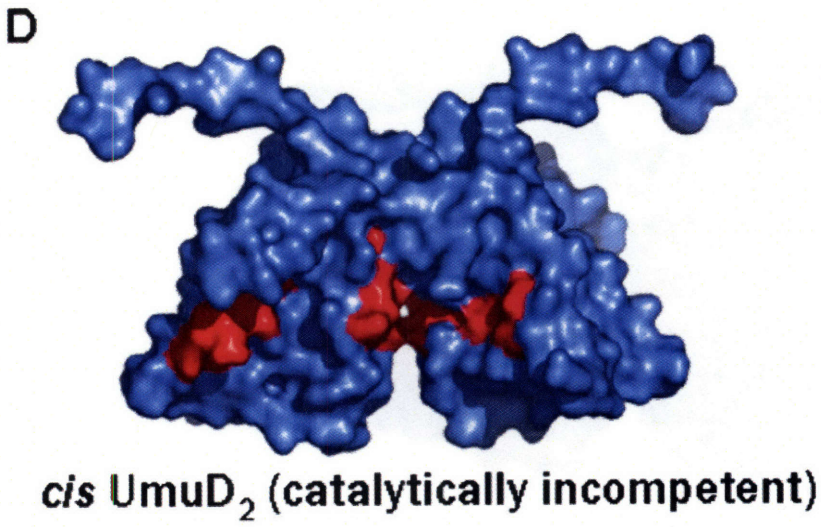


Figure 3-S2 (continued): **Models of $UmuD_2$ with DinB-interacting residues highlighted in red.**





interface between DinB and $UmuD_2$ or $UmuD'_2$ does not substantially involve the UmuD N-terminus, in agreement with our observations using cellulose filter peptide arrays (Figure 3-3B, p.81). Overproduction of UmuD, and to a lesser extent UmuD′, also inhibits adaptive mutagenesis (Figure 3-2B, p.74), which occurs via a -1 frameshift event [303]. Irrespective of the precise molecular mechanism of this type of mutagenesis, it is clear that the *umuD* gene products play a role in modulating it.

Perhaps the *umuD* gene products also regulate the fidelity of the Y-family DNA polymerases in response to DNA damage or environmental stress. Under normal conditions the levels of $UmuD_2$ are relatively low, increasing over approximately 30 minutes upon SOS induction. During this phase of the SOS response in which high fidelity repair mechanisms predominate [100], the mutagenic activity of DinB may be restricted. The subsequent, and highly regulated, conversion of $UmuD_2$ to $UmuD'_2$ marks the beginning of a more mutagenic phase of the SOS response (Figure 8).

Remarkably, examination of the molecular interface between $UmuD_2$ and DinB by peptide array mapping indicates that $UmuD_2$ may enclose the relatively open active site of DinB (Figures 3-3A, p.81 and 3-7A-B, p.97). Using this information, we constructed a mutant, DinB(F172A), which has a >50-fold reduced affinity for $UmuD_2$ (Figure 3-3C, p.81). Overproduction of this DinB variant induces -1 frameshift mutagenesis, albeit to a slightly lesser extent than the wild-type enzyme, and is relatively inert to the action of UmuD both *in vivo* and *in vitro* (Figure 3-4A, p.88; Figure 3-5D, p.76). Overproduction of mammalian pol κ has similarly deleterious mutagenic consequences as DinB overproduction [315], despite the fact that its active site appears to be comparatively closed [316]. The *umuD* gene is only present in certain prokaryotic

species, but it is tempting to speculate that there may be a functional ortholog in eukaryotes that might similarly modulate Pol κ function. Moreover, even though the UmuD-binding interface of DinB is only cryptically conserved between prokaryotes and eukaryotes, the residues that define this interface display statistical covariance through evolution (Figure 3-7C-D, p.97), suggesting the possible presence of a driving force for the maintenance of this surface.

To our surprise, *in vitro* reconstitution the $UmuD_2$ modulation of DinB frameshift activity using a mismatched substrate [73, 307] is only efficient when RecA is present in stoichiometric quantities with DinB. Moreover, the addition of RecA and $UmuD_2$ to an assay in which DinB replicates a normal template of the same sequence context results in a profound (*ca.* 20-fold) increase in DinB catalytic proficiency (Figure 3-5B, closed circles, p.76). We also found that $recA^+$ is required for the suppression of DinB dependent -1 frameshift mutagenesis *in vivo*, as the co-overproduction of $UmuD_2$ does not suppress this mutagenesis in a *recA* strain (Figure 3-5C, p.76). Based on the observation that a noncleavable UmuD(S60A) variant is fully proficient for suppression of -1 frameshift mutagenesis (Figure 3-2A, p.74), we infer that the function of RecA in this inhibition is not to promote $UmuD_2$ autoproteolysis. These observations suggest that $UmuD_2$ is able to modulate the activity of DinB in a highly nuanced fashion that depends on the presence of RecA.

We anticipated that deletion of either *umuD* or *recA* would affect DinB TLS function. Surprisingly, deletion of *umuD* had no effect on *dinB* dependent resistance to NFZ in a wild-type *E. coli* strain. Deletion of *recA* caused sensitivity to NFZ, likely due to the central role of RecA in coordinating numerous aspects of the DNA damage

response [286], but also perhaps because of its interaction with DinB. We wondered whether the frameshift activity of DinB is genetically separable from its function in TLS. We have previously characterized a separation of function mutant of DinB's steric gate residue, DinB(F13V), which is entirely active as a conventional DNA polymerase but is virtually unable to catalyze TLS on N^2-dG adducted templates [53]. *In vivo*, DinB(F13V) is exceptionally proficient at promoting -1 frameshift mutagenesis, although its levels are comparable to wild type by immunoblot (Figure 3-S6, p.93), indicating that TLS function is not required for this property of DinB (Figure 3-6A, p.92). *In vitro*, DinB(F13V) can catalyze frameshift formation, albeit at a 3-fold reduced rate relative to wild-type DinB, suggesting that this variant may either obtain or retain increased access to frameshift intermediates *in vivo*. These observations indicate that TLS and -1 frameshift function are distinct phenomena.

We propose, as have others [45, 306], that the modest -1 frameshift activity of DinB and its orthologs may arise as a result of its unique active site that is specialized for TLS function. Under most cellular conditions, $UmuD_2$ and RecA are present at sufficient levels not only to prevent rampant -1 frameshift mutagenesis but also to control DinB polymerase function, suggesting a mechanism through which the mutagenic potential of DinB is regulated, and perhaps exploited, by the cell. Elevation of DinB levels above those of its $UmuD_2$ manager protein, as occurs in both adaptive and spontaneous -1 frameshift mutagenesis, renders the cell vulnerable to the full mutagenic potential of DinB. Under conditions of DNA damage, the $UmuD_2$ predominant phase of the SOS response may serve to delay the lower fidelity facets of Y-family polymerase function. Indeed, the precise tuning of DinB levels relative to those of $UmuD_2$ under both basal and

SOS induced conditions may facilitate the modulation of its function in response to environmental stress.

Experimental Procedures

Protein Expression and Purification

DinB, UmuD, UmuD′, and RecA were purified as described previously [53, 317, 318], but 50 mM Hepes pH 7.2 was used exclusively. Expression plasmids for DinB(F172A) and UmuD(D91A) were constructed from pDFJ1 and pSG5 [317], respectively, using a Quikchange kit (Stratagene). Both DinB(F172A) and UmuD(D91A) behaved as wild-type DinB and UmuD during purification. A plasmid encoding a $(His)_6$HMK-DinB derivative [319] was constructed in pET16B using standard cloning procedures. The HMK-DinB protein derivative was purified using Ni^{2+}-NTA affinity resin (Qiagen) following the manufacturer's instructions.

Affinity Chromatography

The details of the procedure are described in supplemental material accompanying this manuscript.

Crosslinking and Binding Measurements

UmuD and DinB were mixed and crosslinking reactions were initiated as described [131] and allowed to proceed for 10 min before quenching with SDS-PAGE loading buffer containing 5% β-mercaptoethanol. Fluorescence center of spectral mass were performed and analyzed as described previously [317].

Farwestern assays and Peptide Array Experiments

Farwestern blots were performed previously described [131]. Abimed cellulose filter peptide arrays were synthesized with overlapping 12-mer peptides offset by two residues scanning the primary sequences of DinB and UmuD (MIT Center for Cancer Research Core Facility). The arrays were probed with 1 μM $UmuD_2$ or DinB, and washed and developed as described [320]. Control arrays were performed using DinB or $UmuD_2$ antibodies alone.

Mutagenesis Assays

DinB dependent -1 frameshift mutagenesis experiments were performed as described [282] except that IPTG was not added to the media. Ampicillin (100 μg/mL) and spectinomycin (60 μg/mL) were used as necessary for plasmid maintenance. Adaptive mutagenesis was performed as previously described [321]. A table of the strains and plasmids used in this study is available as supplemental material (Table 3-S1).

DinB and $UmuD_2$ Activity Assays

DinB was assayed as described previously [53] except 50 nM enzyme and 10 nM primer/template was used. The oligonucleotides 5′-ATCCTAGTCCAGGCTGCTGACAACTCGGGAACGTGCTACATGAAT-3′, 5′-ATTCATGTAGCAGCGTTCCC-3′, and 5′ATTCATGTAGCAGCGTTCCG-3′ were designed based on those used previously [73]. Reactions were initiated with the appropriate dNTP, quenched after 20 min with 95% formamide, 25 mM EDTA, 0.5%

bromophenol blue, and 0.5% xylene cyanol, and separated on a 16% denaturing polyacrylamide gel, which was quantified using a Typhoon phosphorimager (GE Healthcare). $UmuD_2$ autocleavage reactions were performed as described previously [317].

Molecular Modeling and Statistical Covariance

A model of the DinB-$UmuD_2$ complex was constructed using the application 3D-dock based on the following constraints: E168 < 6 Å from either UmuD chain, and L176, P177, K180, and F172 < 8 Å from D91 of either UmuD chain. The $UmuD_2$-RecA model was made using the following constraints: RecA S117 < 7 Å from either UmuD chain [88], UmuD residues L101, R102, V34, S81 are 6-12 Å from RecA, and UmuD residue E11 is 6-25 Å from RecA. Resulting complexes were filtered based the RecA residues T243 and R244 < 10 Å from DinB to generate a model of the DinB·RecA·$UmuD_2$ ternary complex. We performed the same procedure by docking DinB to the $UmuD_2$·RecA models using the constraints described above and obtained similar results. Statistical covariance was performed by aligning 84 DinB and pol κ sequences from diverse organisms using ClustalW and analyzing significant pairwise correlation of alignment positions with the CRASP algorithm [311].

Supplemental Data

Supplemental Data includes 8 figures, supplemental References, and supplemental experimental procedures.

Acknowledgements

This work was supported by NIH grant CA021615 to G.C.W and NIEHS grant P30 ES002109 to the MIT Center for Environmental Health Sciences. G.C.W. is an American Cancer Society Professor. SMS was a Cleo and Paul Schimmel fellow.

CHAPTER 4

Regulation of *E. coli* SOS Mutagenesis by Dimeric Intrinsically Disordered *umuD* Gene Products

S.M. Simon[1], F. J. R. Sousa[2], R. Mohana-Borges[2], G.C. Walker[1]

[1]Massachusetts Institute of Technology, Department of Biology

[2]Federal University of Rio de Janeiro, Department of Structural Biophysics

This paper will be submitted to PNAS.

Abstract

Products of the *umuD* gene in *E. coli* play key roles in coordinating the switch from DNA repair to mutagenic translesion DNA synthesis (TLS) during the SOS response to DNA damage. Homodimeric $UmuD_2$ is upregulated approximately 10-fold during the beginning of the SOS response, after which a slow post-translational autocleavage step removes the N-terminal 24 amino acids of each UmuD monomer. The remaining C-terminal fragment, $UmuD'_2$, is required for mutagenic TLS. The small proteins $UmuD_2$ and $UmuD'_2$ make a large number of specific protein-protein contacts for their roles in regulation, including interactions with three of the five known *E. coli* DNA polymerases, parts of the replication machinery, and RecA recombinase. We show that, despite forming stable homodimers at a wide range of protein concentrations, $UmuD_2$ and $UmuD'_2$ have a CD spectrum with almost no α helix or β sheet signal at physiological concentrations *in vitro*. High protein concentrations, osmolytic crowding agents, and specific interactions with a partner protein can induce CD spectra that more closely resemble the expected β sheet signature. A lack of secondary structure *in vitro* is characteristic of a class of proteins termed intrinsically disordered proteins, many of which act as regulators. A stable homodimer that lacks significant secondary structure is unusual but not unprecedented. Furthermore, previous single-cysteine cross-linking studies of $UmuD_2$ and $UmuD'_2$ show that they have a nonrandom structure at physiologically relevant concentrations *in vitro*. Our results offer insights into structural characteristics of the relatively poorly understood class of proteins designated as intrinsically disordered and provide a model for how the *umuD* gene products can regulate diverse aspects of the bacterial SOS response.

Introduction

The bacterial SOS response is a tightly regulated cellular reaction to stress-induced DNA damage (reviewed in [3]). It is temporally divided into at least two phases; an early, relatively accurate DNA repair phase and a later, more mutagenic damage tolerance phase for persistent damage [97]. This timing is regulated in part by products of the *umuD* gene. The initial product of *umuD* is a homodimer of 139-amino acid subunits that appears early after SOS induction [97]. Expression of $UmuD_2$ is associated with a reduction in the rate of DNA replication and an increase in survival after UV-induced DNA damage [97]. $UmuD_2$ has also been shown to interact with DinB (the Y family DNA Pol IV) and reduce the frequency of DinB-induced -1 frameshift mutations [77]. Damage-induced RecA:ssDNA nucleoprotein filaments mediate a slow autocleavage reaction of $UmuD_2$ that is mechanistically similar to inactivation of the LexA and λ CI repressors [101-103, 322]. The N-terminal 24 amino acids of each subunit of $UmuD_2$ are removed, producing a homodimer of the C-terminal 115 amino acid subunits, $UmuD'_2$ [101-103]. $UmuD'_2$ interacts with UmuC, the catalytic subunit of the Y family DNA Pol V, thereby activating UmuC for TLS [75, 108, 120]. $UmuD'_2C$ is responsible for most damage-induced point mutations in *E. coli* [91, 120]. Full-length $UmuD_2$ is rapidly and continuously degraded by Lon protease, whereas $UmuD'_2$ first exchanges with $UmuD_2$ to form the thermodynamically most stable UmuD'D heterodimer [121, 126]. It is in the context of the heterodimer that the UmuD' subunit is degraded by ClpXP protease to end the mutagenic phase of the SOS response [323].

For such small proteins, $UmuD_2$ and $UmuD'_2$ make a remarkable number of highly specific protein-protein contacts, many but not all of them to DNA polymerases

(Figure 4-5A, p.114). Both proteins have been shown to interact with UmuC [120], DinB [77], and the catalytic, processivity, and proofreading subunits of the *E. coli* replicative DNA polymerase, Pol III [131]. Additionally, both interact with RecA:ssDNA nucleoprotein filaments [101-103, 111].

However, despite the nearly identical primary sequence between $UmuD_2$ and $UmuD'_2$, their interactions with the same partner can differ in affinity and functional significance. As mentioned above, only $UmuD_2$ prevents DinB-induced -1 frameshifts [77], whereas only $UmuD'_2$ activates UmuC for TLS [75, 108, 120]. $UmuD_2$ interacts more strongly with the β processivity subunit of Pol III, presumably to slow DNA replication, whereas $UmuD'_2$ interacts preferentially with the α catalytic subunit, which may facilitate TLS by $UmuD'_2C$ [131]. RecA:ssDNA interacts with $UmuD_2$ to promote cleavage to $UmuD'_2$ [103], whereas $UmuD'_2C$ requires *trans* RecA:ssDNA for efficient TLS [111]. The fact that such small proteins (no more than 30 kDa in their dimeric forms) can make so many highly specific protein-protein interactions is intriguing, and so high resolution structural studies were undertaken in an effort to find an explanation.

The X-ray crystal structure [141] and an NMR solution structure [142] of the cleaved form $UmuD'_2$ offer some insight. Although both methods of structure determination indicated that $UmuD'_2$ has an overall β sheet fold with a C2 axis of symmetry, a detailed comparison between the two structures reveals substantial differences [142]. The overall shape of the protein is less globular in the NMR structure [142], and the protease active site residues are only poised for catalysis in the X-ray crystal structure [141]. The differences in structural features between the two methods of structure determination suggested that $UmuD'_2$ may have considerable plasticity.

Figure 4-5: **A model for sequential protein-protein interactions by intrinsically disordered proteins.**

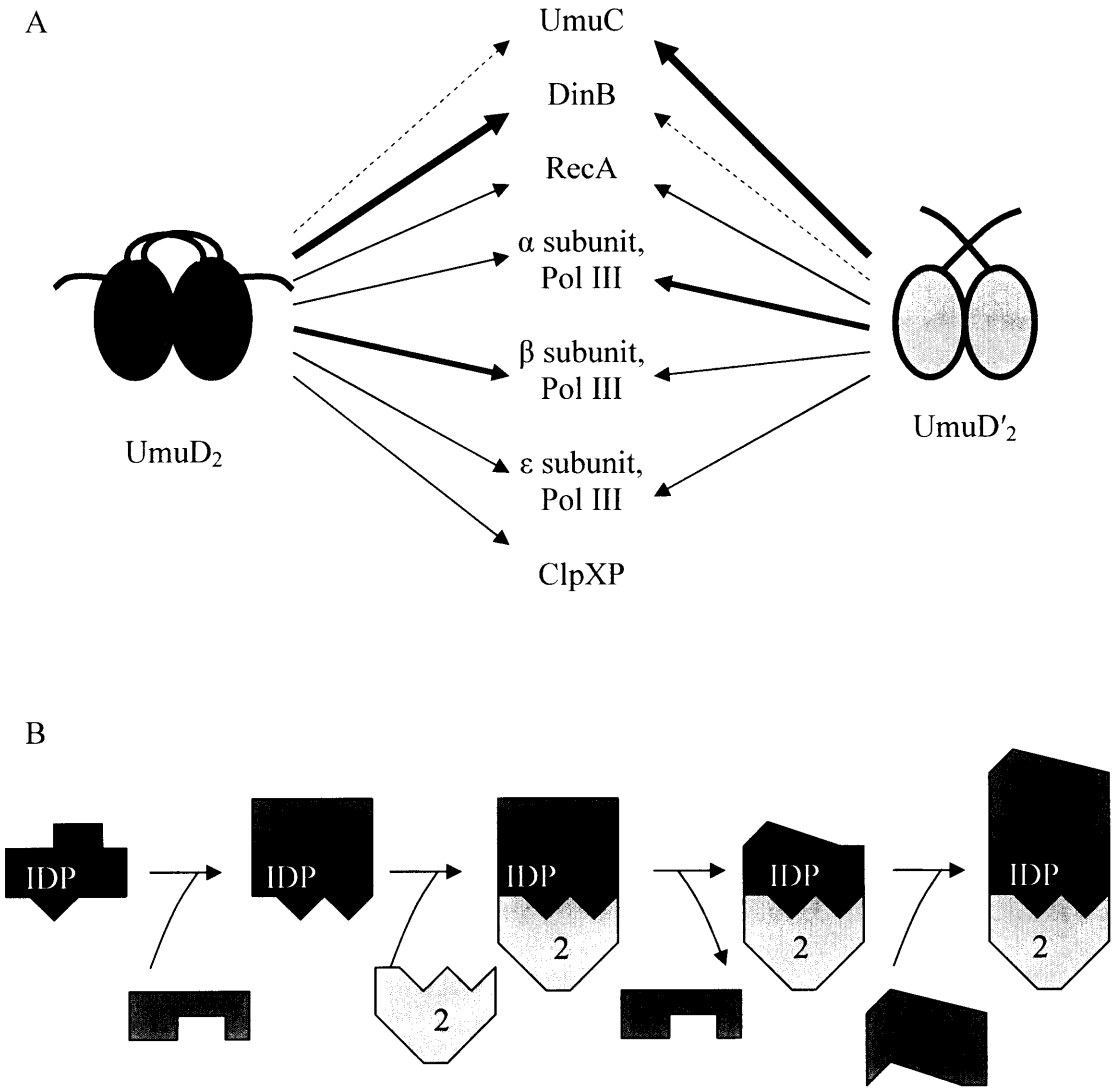


Figure 4-5 (continued): **A model for sequential protein-protein interactions by intrinsically disordered proteins.**

C

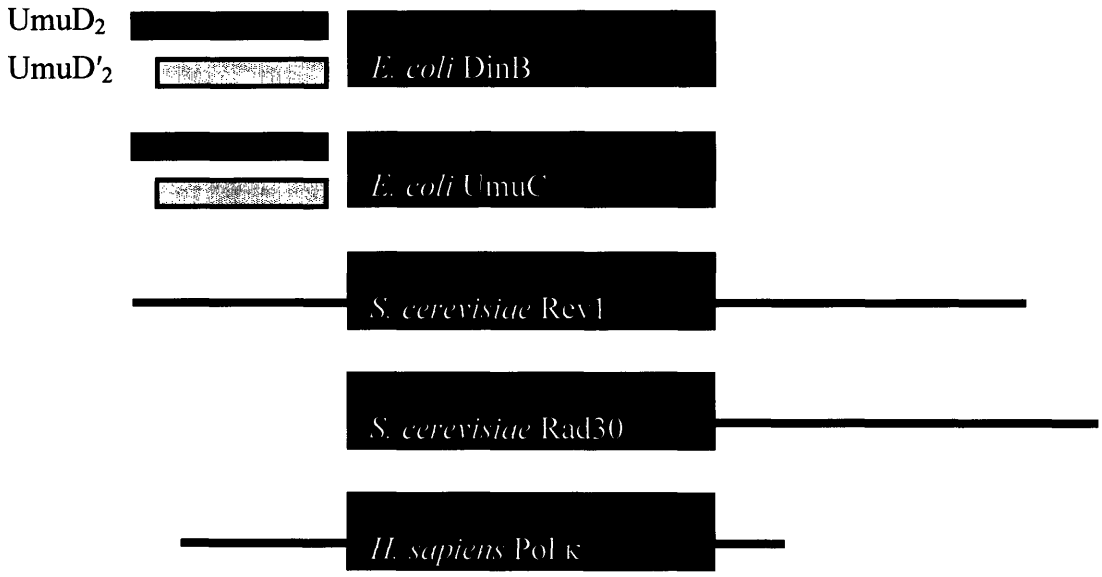


Figure 4-5 (continued): **A model for sequential protein-protein interactions by intrinsically disordered proteins.** (A) $UmuD_2$ and $UmuD'_2$ make a variety of distinct protein-protein interactions. Where such information is available, the relative binding affinities between $UmuD_2$ or $UmuD'_2$ and an interaction partner are represented by thick arrows (for strong interactions) or thin ones (for weak interactions). (B) Model for sequential protein-protein interactions with an intrinsically disordered protein. An intrinsically disordered protein (IDP) may first bind to one interaction partner (1), which stabilizes a particular conformation. If a second binding interface becomes exposed in this conformation, another protein may now bind (2). Subtle conformational changes in this context may destabilize the first protein-protein interaction, causing the original protein to exit the complex and possibly exposing a different interface. If so, a different partner (3) can bind at this site. (C) $UmuD_2$ and $UmuD'_2$ may act as interchangeable protein-protein interaction domains for *E. coli* Y family DNA polymerases. Y family DNA polymerases have conserved catalytic domains (large boxes), and many eukaryotic Y family polymerases have extended N-terminal or C-terminal protein-protein interaction domains (lines on bottom three representations). Although these interaction domains are missing in the two *E. coli* Y family DNA polymerases, both of them interact with *umuD* gene products, which may serve as interchangeable protein-protein interaction domains in a streamlined genome (top two representations).

Although no high-resolution structural data are available for $UmuD_2$, we have recently proposed four energy-minimized symmetrical models [147] based on the $UmuD'_2$ structures [141, 142], the homologous protein LexA [146], and single-cysteine studies [134]. The models all suggest that the C-terminal domain of $UmuD_2$ has substantial β sheet content, but they differ in the locations of the N-terminal amino acids relative to the C-terminal globular domains. Previous single cysteine studies of $UmuD_2$, which probed the structure of $UmuD_2$ in solution at physiologically relevant concentrations, have generally been consistent with our current models of $UmuD_2$. Single cysteine derivatives of many amino acids that are predicted to be close to the C2 axis of symmetry robustly cross-link to covalent dimers [143, 144]. Interestingly, some positions that are predicted to be far away from the C2 axis of symmetry also cross-link to form covalent dimers [88, 143]. However, these residues can form a dimer interface if the N-terminal arms are in an intermediate conformation between the *cis* and *trans* proposed models [147]. These results are consistent with the hypothesis that $UmuD_2$ can interchange between multiple conformations that resemble those presented in our four models.

We used CD spectroscopy to compare the secondary structure of $UmuD_2$ with that of the known β-sheet protein $UmuD'_2$, but we were surprised to find that, at physiological concentrations, both proteins show a spectrum that more resembles a random coil than the expected β-sheet structure. These results are typical of a previously defined class of proteins, called intrinsically disordered proteins, whose precise structures depend on their environment and which often have regulatory functions [160, 163, 164, 324]. However, $UmuD_2$ and $UmuD'_2$ form stable homodimers at physiological concentrations *in vitro*, and previous single-cysteine studies show that both proteins have a nonrandom structure

at these concentrations [88, 134, 143, 144]. Our results provide a rare opportunity to probe the actual structure of proteins that appear unfolded by CD and limited proteolysis.

Results

$UmuD_2$ and $UmuD'_2$ have extremely different CD spectra at μM and mM concentrations.

As part of our effort to compare the secondary structure of $UmuD_2$, whose three-dimensional structure is unknown, to the known structure of its derivative $UmuD'_2$ [141, 142], we measured the CD spectrum of $UmuD'_2$ at 5 μM, the concentration found in SOS-induced cells [77]. We were startled to discover that the CD spectrum of $UmuD'_2$ at the physiologically relevant concentration more closely resembles a random coil than the expected β-sheet structure (Figure 4-1A, p.119). In an attempt to reconcile these results with the two previous high resolution analyses of $UmuD'_2$, which had revealed it to be a β-sheet rich protein [141, 142], we took the CD spectrum of $UmuD'_2$ at the high, nonphysiological protein concentration used to solve the solution structure by NMR [142]. Consistent with the previous structural analysis [141, 142], the CD spectrum of $UmuD'_2$ at 2 mM displays more typical β-sheet character (Figure 4-1A, p.119). Examination of $UmuD_2$ reveals exactly the same striking anomaly, a CD spectrum resembling a random coil at μM concentrations and one consistent with a β-sheet rich protein at mM concentrations (Figure 4-1B, p.119).

These unexpected findings suggested to us that $UmuD_2$ and $UmuD'_2$ must be undergoing a transition from a folded state at mM concentrations to a more disordered state upon dilution to μM concentrations. We therefore examined the effect of dilution on the susceptibility of $UmuD'_2$ and $UmuD_2$ to limited proteolysis over a 5-minute time window. $UmuD'_2$ or $UmuD_2$ that has been pre-equilibrated at 10 μM results in more

Figure 4-1: **$UmuD_2$ and $UmuD'_2$ have extremely different CD spectra at μM and mM concentrations**

A

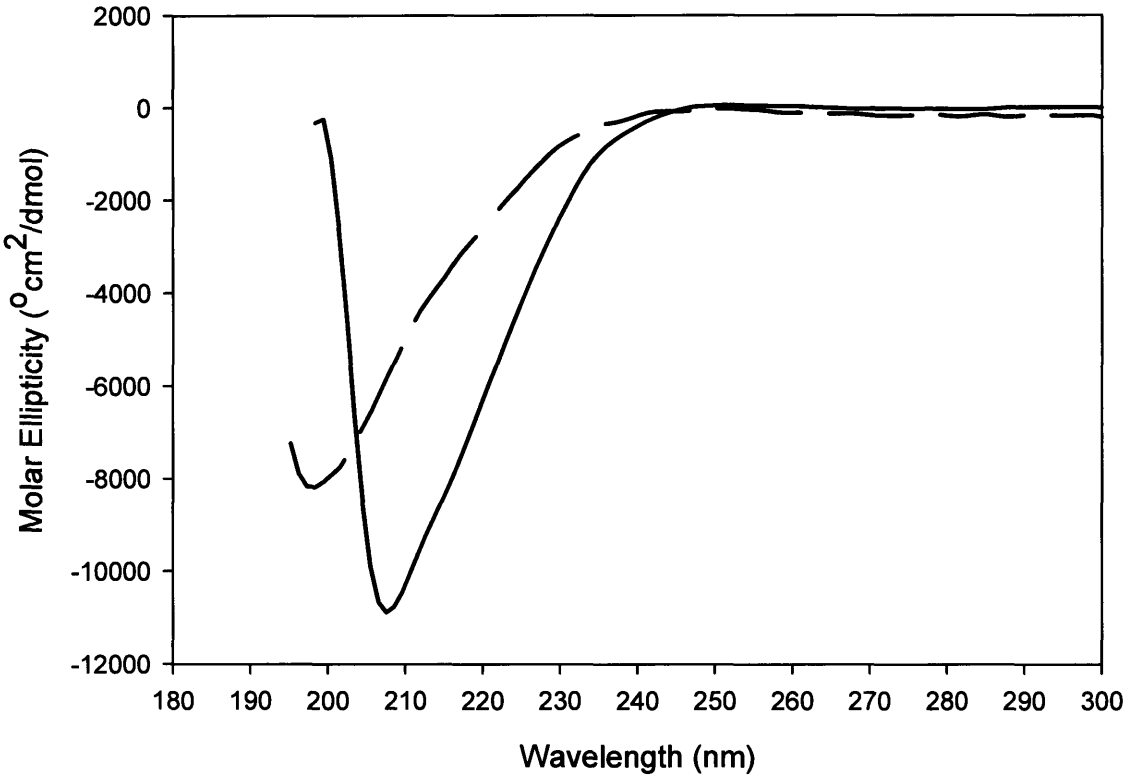
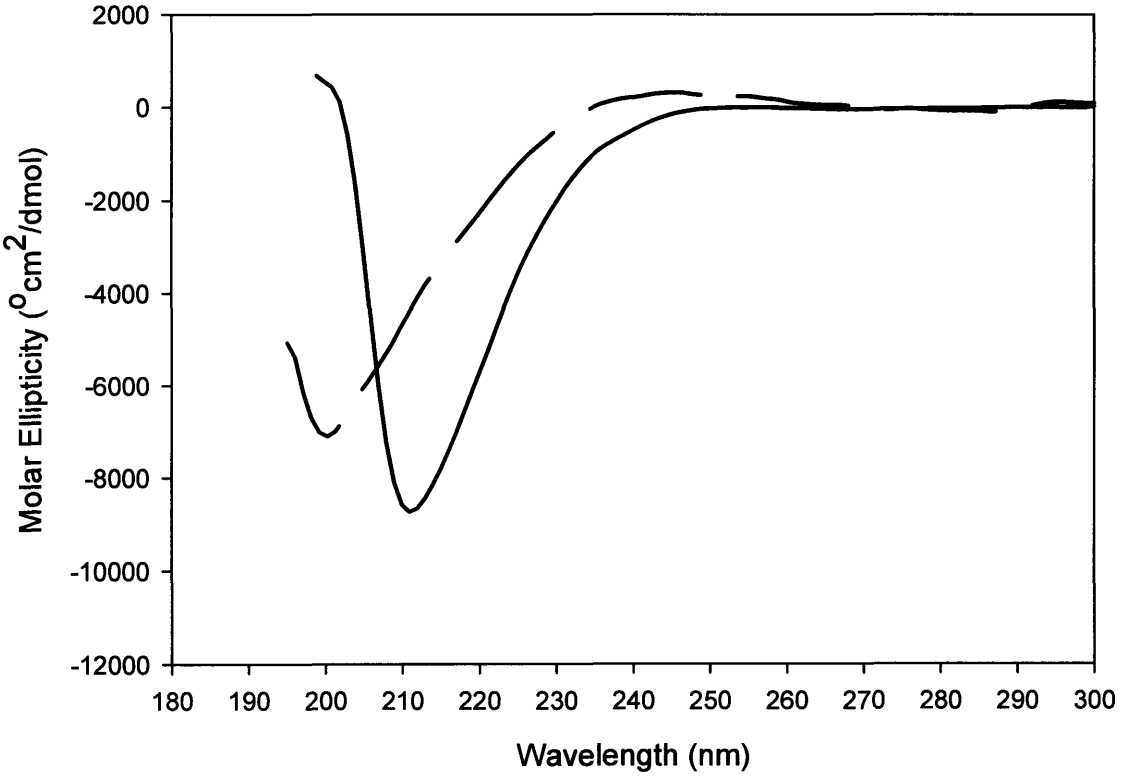


Figure 4-1 (continued): **$UmuD_2$ and $UmuD'_2$ have extremely different CD spectra at μM and mM concentrations**

B



C Limited Proteolysis

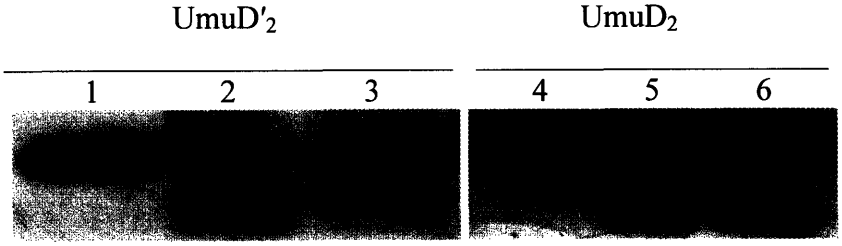


Figure 4-1 (continued): **$UmuD_2$ and $UmuD'_2$ have extremely different CD spectra at μM and mM concentrations.** CD spectra of 5 μM (dashed line) and 2 mM (solid line) $UmuD'_2$ (A) or $UmuD_2$ (B) at 25 °C. (C) Limited chymotrypsin proteolysis of 5 μM $UmuD'_2$ (lanes 1-3) and 5 μM $UmuD_2$ (lanes 4-6) at 37 °C for 5 minutes. Lanes are: (1) 5 μM $UmuD'_2$ with no protease (2) $UmuD'_2$ pre-equilibrated at 10 μM and diluted 1:1 with 5 mg/mL chymotrypsin (3) $UmuD'_2$ freshly diluted to 10 μM 1 minute before 1:1 dilution with 5 mg/mL chymotrypsin (4) 5 μM $UmuD_2$ with no protease (5) $UmuD_2$ pre-equilibrated at 10 μM and diluted 1:1 with 5 mg/mL chymotrypsin (6) $UmuD_2$ freshly diluted to 10 μM 1 minute before 1:1 dilution with 5 mg/mL chymotrypsin.

complete proteolysis by chymotrypsin than UmuD′$_2$ or UmuD$_2$ that has been freshly diluted from a 2 mM stock (Figure 4-1C, p.119). The extent of degradation of freshly diluted UmuD′$_2$ is about 60% of that of pre-equilibrated UmuD′$_2$, whereas freshly diluted UmuD$_2$ is degraded at about 85% the level of pre-equilibrated UmuD$_2$. These results would be consistent with a model in which UmuD′$_2$ and UmuD$_2$ fold only at higher than physiologically relevant concentrations.

The CD and proteolysis results at physiological concentrations are typical of a class of proteins called intrinsically disordered proteins, which lack significant α helix and β sheet structure *in vitro* (reviewed in [160, 163, 187, 324]). We therefore used PONDR protein disorder prediction programs [325, 326] to test the similarity of UmuD′$_2$ and UmuD$_2$ to known disordered sequences and found that the extreme N-terminus of UmuD′$_2$ and much of the C-terminal regions of both UmuD′$_2$ and UmuD are predicted to be disordered (Figure 4-S1, p.123). Nevertheless, both proteins are active *in vitro* at physiologically relevant concentrations (Figure 4-S2, p.125).

Crowding agents and specific protein-protein interactions induce secondary structure in UmuD$_2$ and UmuD′$_2$.

To test whether the β-sheet rich CD spectra of UmuD$_2$ and UmuD′$_2$ at mM concentrations result from specific self-self interactions or from more general crowding effects, we took the CD spectrum of *umuD* gene products in the presence of the osmolytic crowding agent proline [217]. Proline at 200 mM increases the secondary structure content of both UmuD′$_2$ (Figure 4-2A, p.126) and UmuD$_2$ (Figure 4-2B, p.126). Less profound but consistent results are obtained with 2.5 M glucose (Figure 4-S4, p.131).

Figure 4-S1: **PONDR analyses predict that $UmuD_2$ and $UmuD'_2$ have regions of intrinsic disorder.**

A

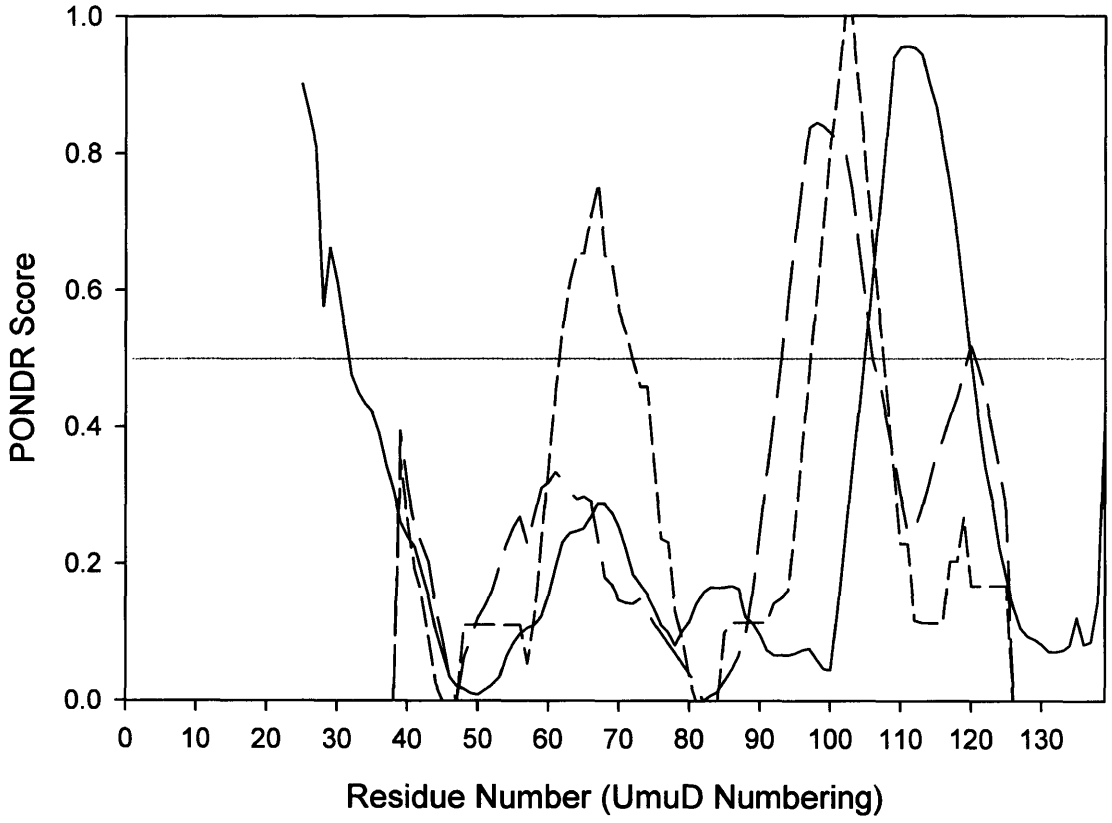


Figure 4-S1 (continued): **PONDR analyses predict that $UmuD_2$ and $UmuD'_2$ have regions of intrinsic disorder.**

B

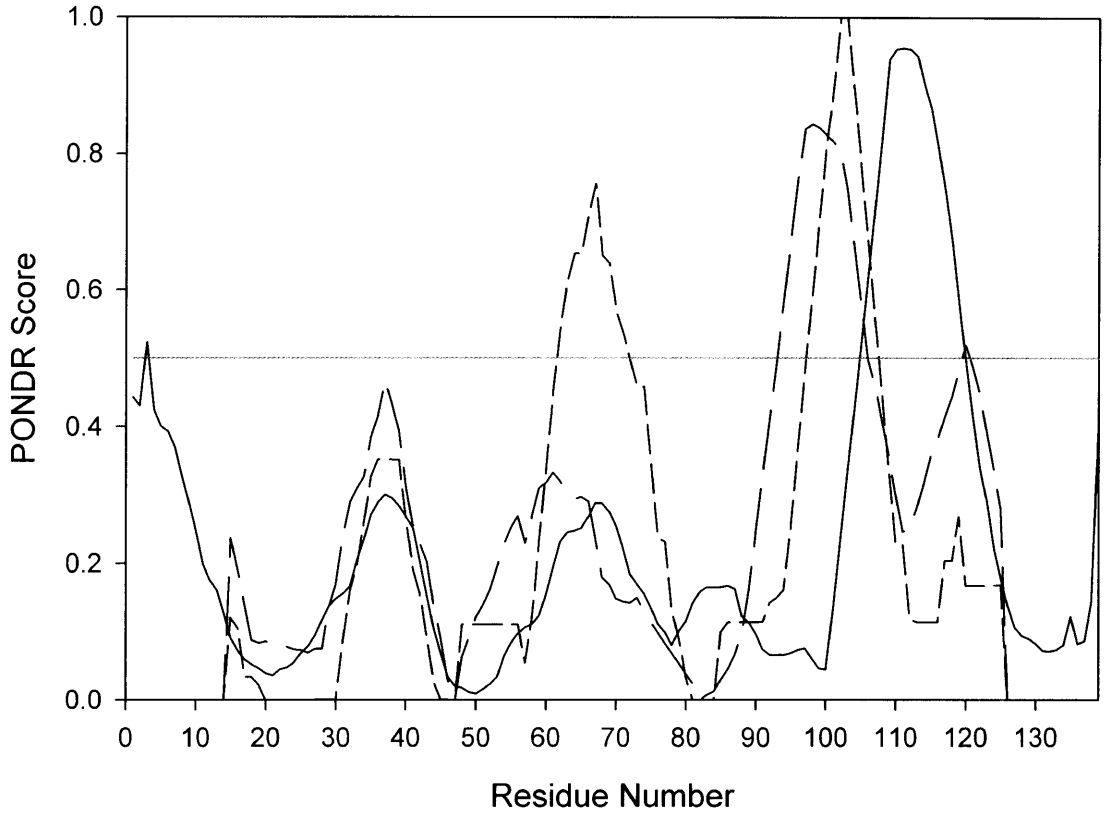


Figure 4-S1: **PONDR analyses predict that $UmuD_2$ and $UmuD'_2$ have regions of intrinsic disorder.** Prediction of disordered regions in the sequence of $UmuD'_2$ (A) and $UmuD_2$ (B) using PONDR protein disorder prediction programs VLXT (solid line), XL1-XT (long dash), and CAN-XT (short dash). Residues with a PONDR score of > 0.5 using any of these programs are predicted to be disordered.

Figure 4-S2: **$UmuD_2$ and $UmuD'_2$ are active *in vitro* at physiologically relevant concentrations.**

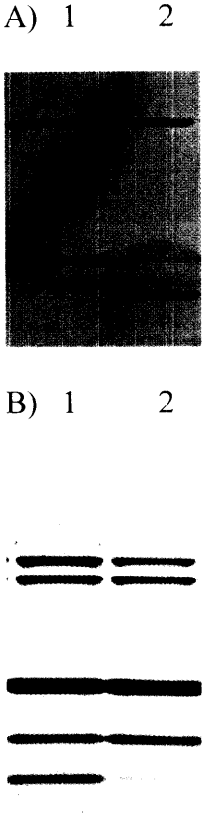


Figure 4-S2: **$UmuD_2$ and $UmuD'_2$ are active *in vitro* at physiologically relevant concentrations.** (A) $UmuD_2$ is active for *in vitro* RecA*-mediated cleavage to yield $UmuD'_2$. RecA* was formed from RecA, ssDNA, and ATP-γ-S and added to 5 μM $UmuD_2$. Aliquots were removed at t=0 (lane 1) and t=1 hr (lane 2), after which most $UmuD_2$ had been converted to $UmuD'_2$. (B) Both $UmuD_2$ and $UmuD'_2$ are active for heterodimer formation and ClpXP proteolysis. $UmuD_2$ and $UmuD'_2$ were coincubated for 30 min to allow formation of the UmuD'D heterodimer and added to ClpXP protease and ATP. Aliquots were removed at t=0 (lane 1) and t=2 hr (lane 2), after which most $UmuD'_2$ had been degraded by ClpXP. Assay procedures are in [317].

Figure 4-2: **Crowding agents and specific protein-protein interactions induce secondary structure in $UmuD_2$ and $UmuD'_2$.**

A

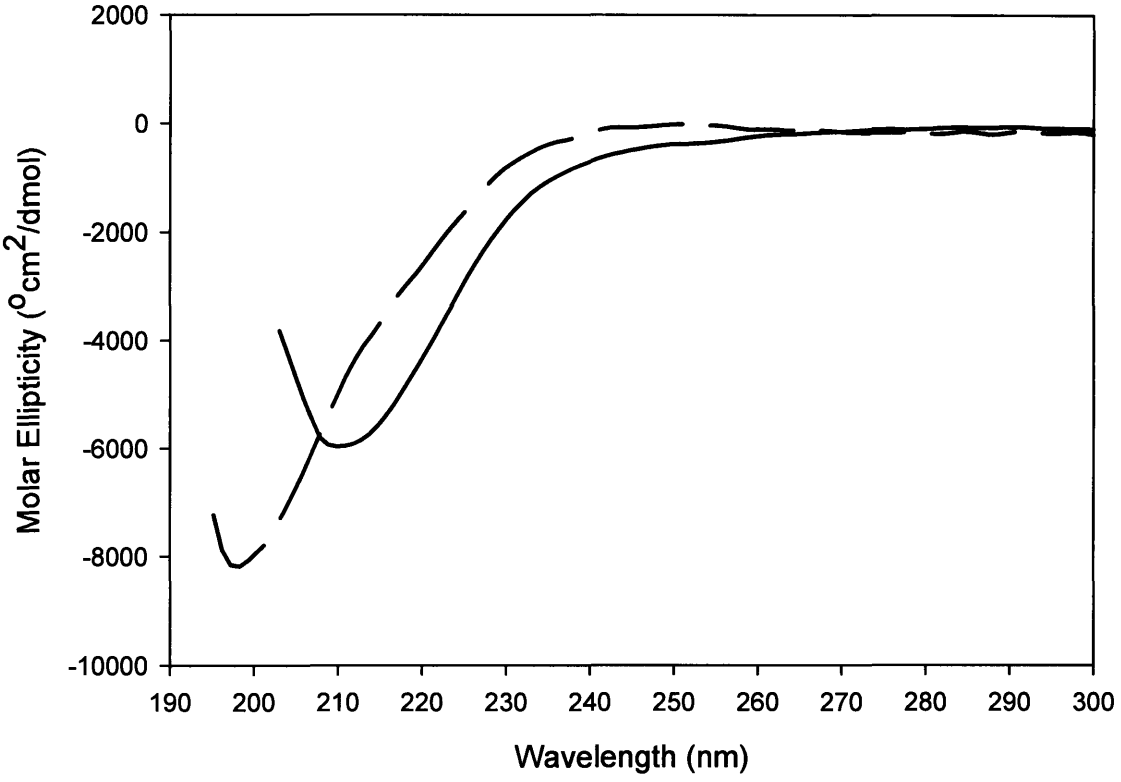


Figure 4-2 (continued): **Crowding agents and specific protein-protein interactions induce secondary structure in $UmuD_2$ and $UmuD'_2$.**

B

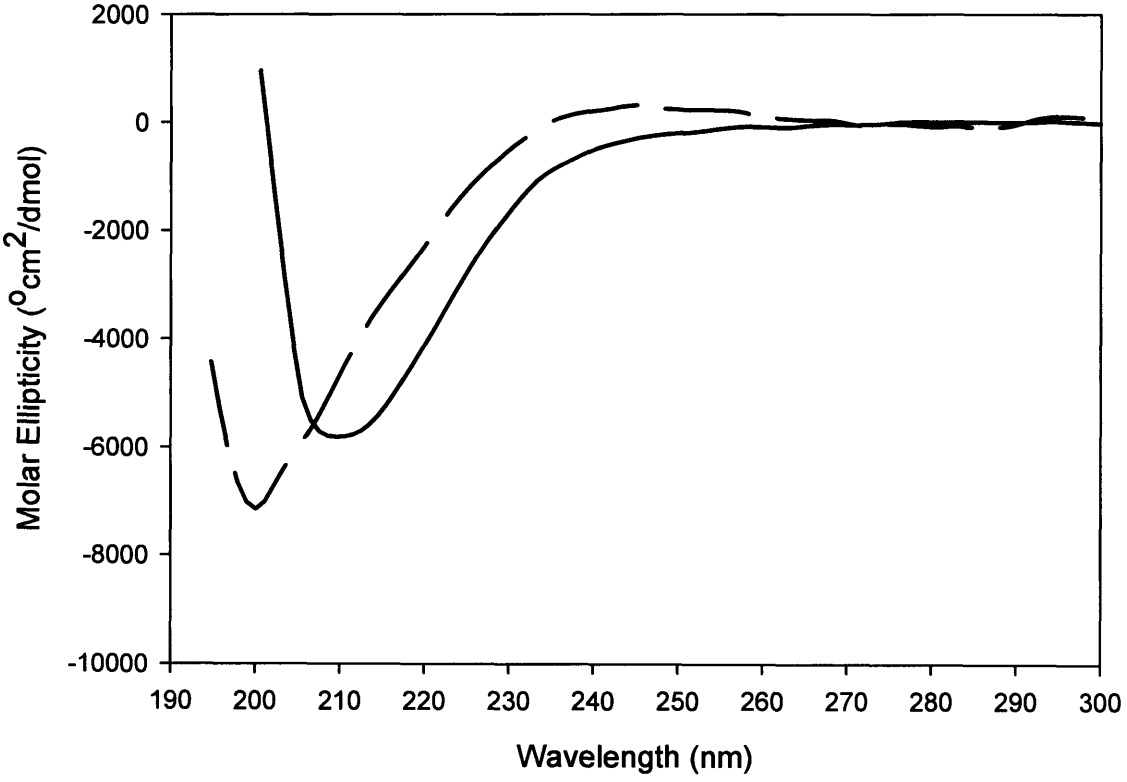


Figure 4-2 (continued): **Crowding agents and specific protein-protein interactions induce secondary structure in $UmuD_2$ and $UmuD'_2$.**

C

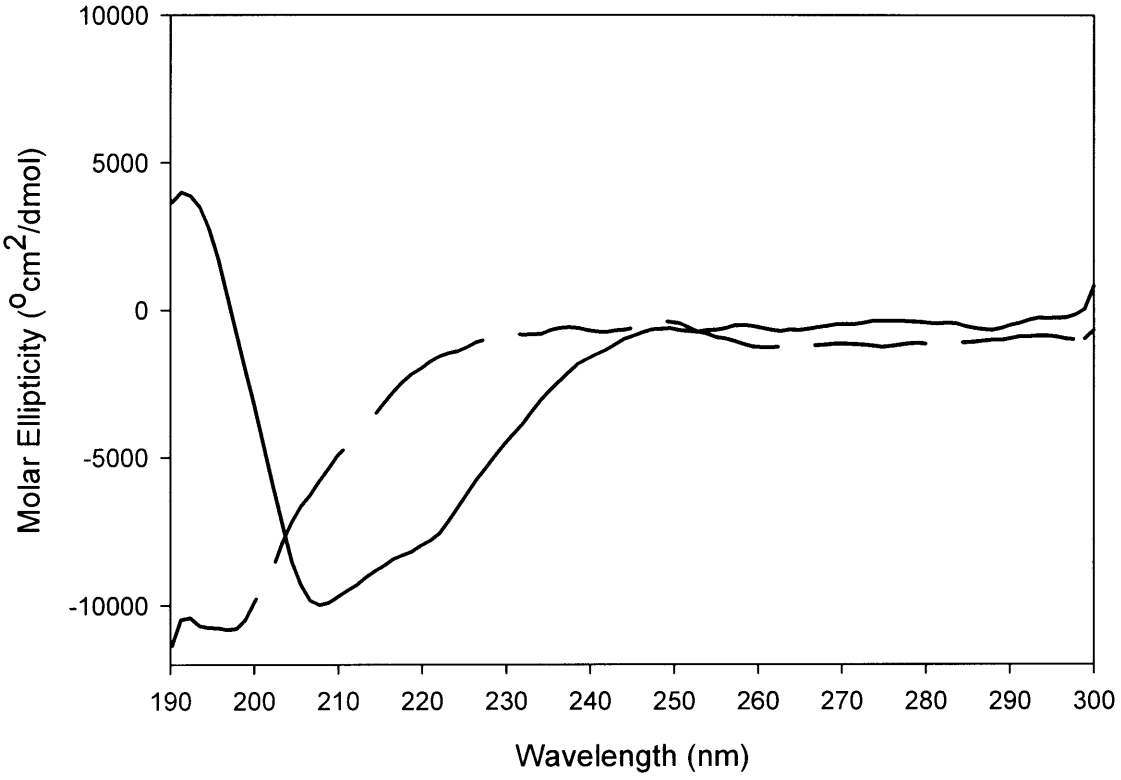


Figure 4-2 (continued): **Crowding agents and specific protein-protein interactions induce secondary structure in $UmuD_2$ and $UmuD'_2$.**

D

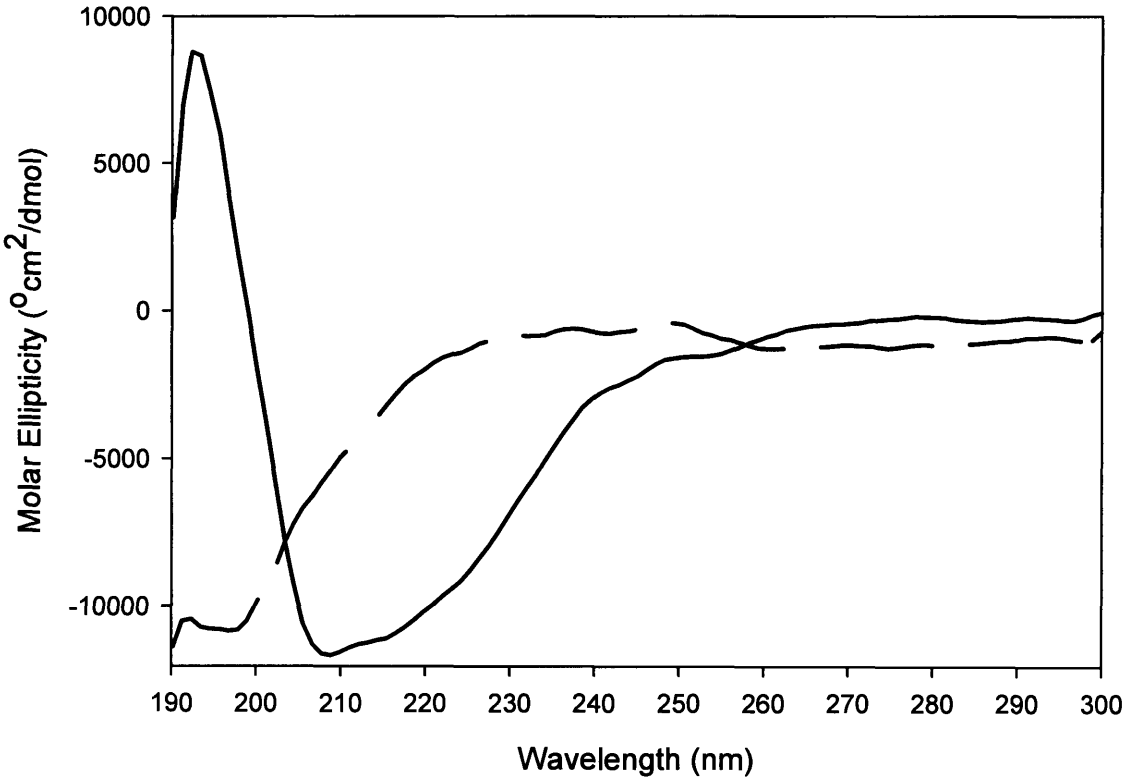


Figure 4-2 (continued): **Crowding agents and specific protein-protein interactions induce secondary structure in $UmuD_2$ and $UmuD'_2$.** CD spectra of $UmuD'_2$ (A) or $UmuD_2$ (B) in the absence (dashed line) or presence (solid line) of 200 mM proline. (C) CD spectrum of $UmuD_2$ alone (dashed line) or in the presence of DinB (solid line). The CD signal of DinB alone was subtracted from that of the complex to obtain the latter spectrum. (D) CD spectrum of $UmuD_2$ alone (dashed line) or in the presence of the β processivity subunit of Pol III (solid line). The CD signal of β was subtracted from that of the complex.

Figure 4-S4: **The osmolytic crowding agent glucose causes some secondary structure formation in both $UmuD_2$ and $UmuD'_2$.**

A

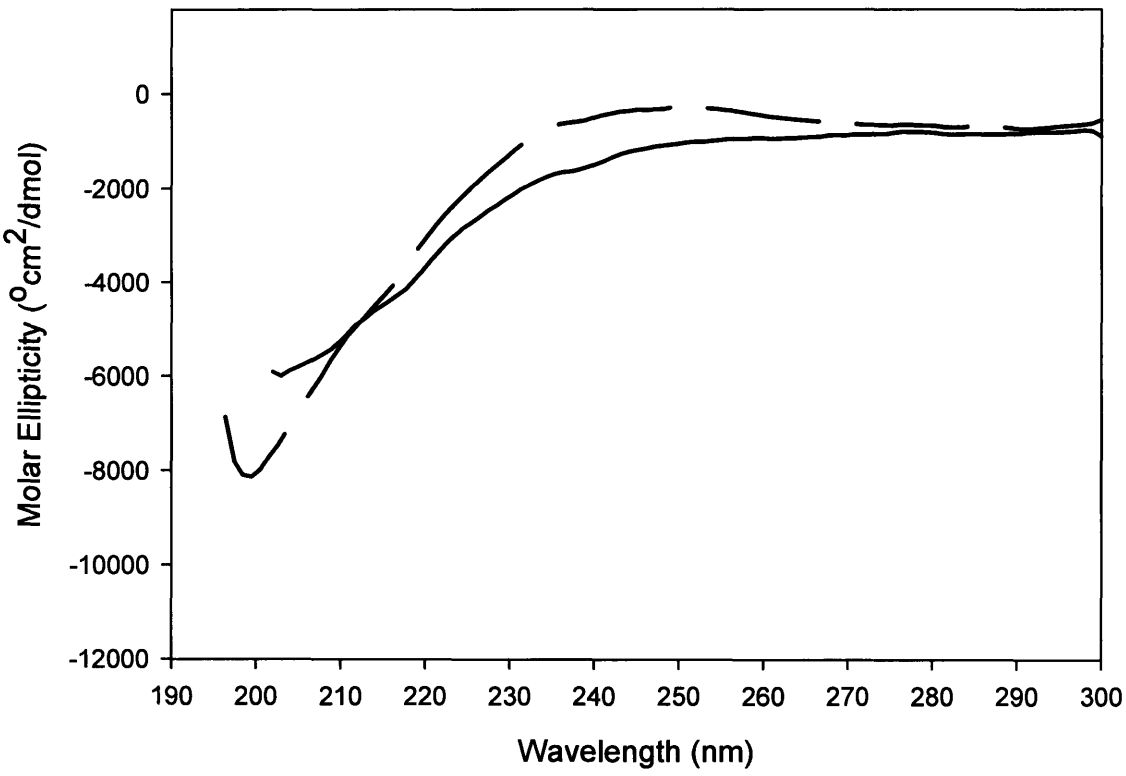


Figure 4-S4 (continued): **The osmolytic crowding agent glucose causes some secondary structure formation in both $UmuD_2$ and $UmuD'_2$.**

B

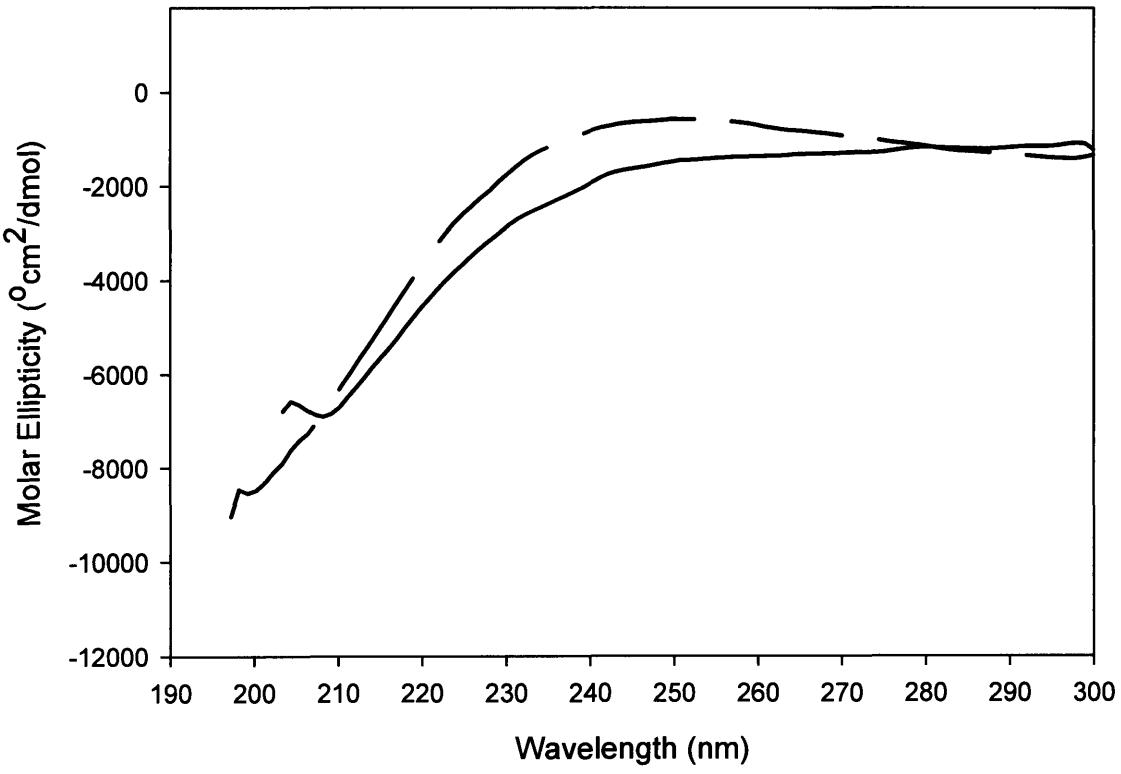


Figure 4-S4: **The osmolytic crowding agent glucose causes some secondary structure formation in both $UmuD_2$ and $UmuD'_2$.** The CD spectrum of 5 μM $UmuD'_2$ (A) or 5 μM $UmuD_2$ (B) in the absence (dashed lines) or presence of 2.5 M glucose (solid lines).

Interestingly, certain other crowding agents such as PEG 8000, 1M NaCl, and glycerol did not increase the secondary structure of either $UmuD_2$ or $UmuD'_2$ (data not shown).

We have previously shown that $UmuD_2$ interacts with DinB (K_D = 0.64 μM) [77] and with the β processivity subunit of DNA Pol III (K_D = 5.5 μM) [147]. To test whether these specific protein-protein interactions induce secondary structure in $UmuD_2$ at μM concentrations, we took the CD spectrum of 50 μM $UmuD_2$ in the presence of 50 μM interacting protein. After subtracting out signal from the interacting protein alone, the resulting spectra of $UmuD_2$ in the presence of DinB (Figure 4-2C,p.126) or of the β subunit (Figure 4-2D,p.126) reveal an increase in the β sheet content. Both spectra show nearly identical secondary structure content. Since both the β subunit and DinB have more typical secondary structure than $UmuD_2$ (data not shown), it is likely that the increase in secondary structure in the complex is mostly due to an increased β-sheet content of $UmuD_2$. However, we cannot rule out the possibility that binding of $UmuD_2$ may cause a conformational change in the interacting proteins as well [77, 147].

$UmuD_2$ and $UmuD'_2$ are dimeric at physiologically relevant concentrations.

The above data would be consistent with a model wherein *umuD* gene products are dimeric and folded at mM concentrations but dissociate into monomers and unfold at physiological concentrations. However, several lines of evidence are inconsistent with monomeric *umuD* gene products at physiological concentrations. Gel filtration of $UmuD'_2$ or $UmuD_2$ shows that their elution volume is between the expected size of a dimer and that of a trimer (See [120] and Figure 4-S3A,p.134). Independent biochemical and biophysical evidence shows a dimeric form of $UmuD_2$, $UmuD'_2$, and UmuD'D [120, 121, 327, 328], whereas no evidence for a trimer has been found. Although the proteins

Figure 4-S3: **$UmuD_2$ and $UmuD'_2$ are dimeric at physiological concentrations *in vitro*.**

A

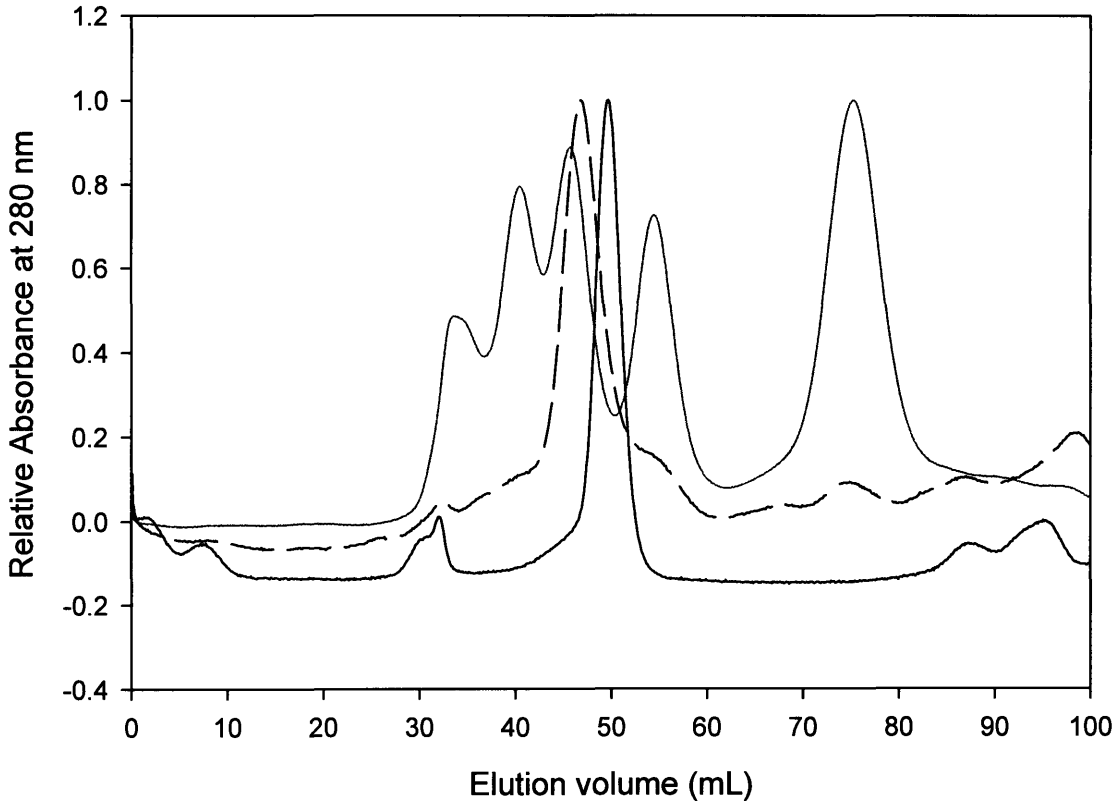


Figure 4-S3 (continued): **$UmuD_2$ and $UmuD'_2$ are dimeric at physiological concentrations *in vitro*.**

B

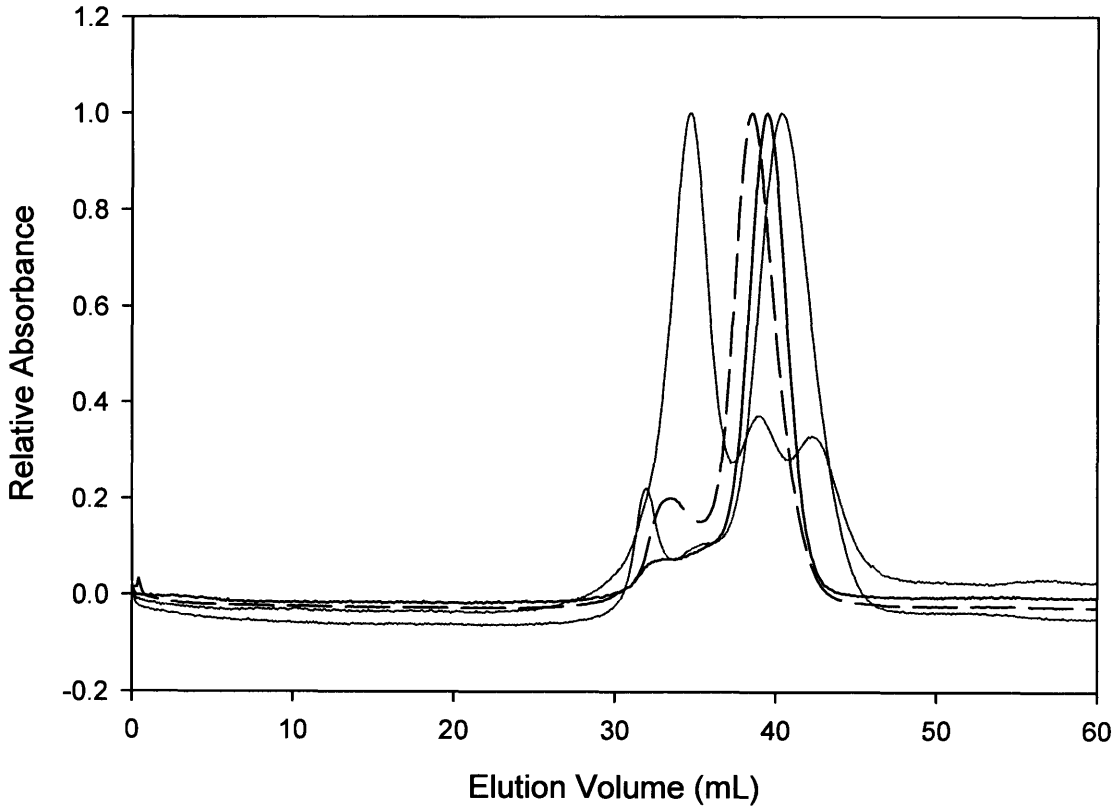


Figure 4-S3 (continued): **$UmuD_2$ and $UmuD'_2$ are dimeric at physiological concentrations *in vitro*.**

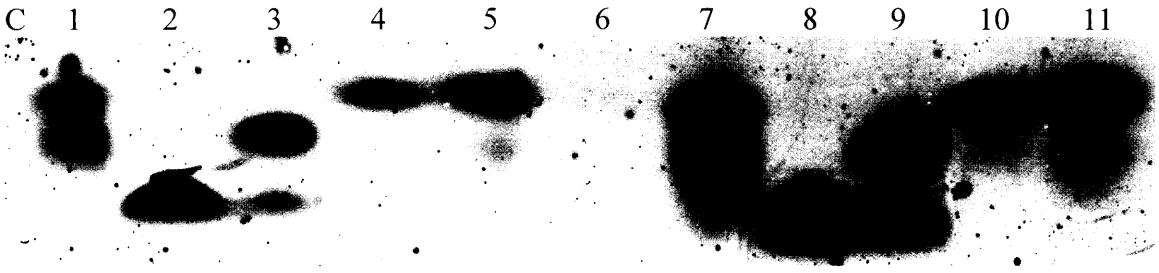


Figure 4-S3: **$UmuD_2$ and $UmuD'_2$ are dimeric at physiological concentrations *in vitro*.** (A) Superdex 75 gel filtration of 5 μM $UmuD'_2$ (black solid line) or 5 μM $UmuD_2$ (black dashed line) as compared to size standards (gray solid line). Size standards are as follows: BSA (dimer of 132 kDa eluting at 34 mL and monomer of 66 kDa eluting at 41 mL), ovalbumin (43 kDa, eluting at 46 mL), chymotrypsin (26 kDa, eluting at 55 mL), and lysozyme (14 kDa, eluting at 75 mL). (B) Lysozyme, chymotrypsin, and 5 μM $UmuD'_2$ or $UmuD_2$ were denatured in 6 M guanidinium for 2 hours before loading and elution from a Superdex 75 gel filtration column in buffer containing 6M guanidinium hydrochloride. Note that all proteins eluted at a lower volume than their native counterparts in (A). (C) Native gel electrophoresis of 500 nM (lanes 1-5) or 5 μM (lanes 7-11) $UmuD_2$ (lanes 1 and 7), $UmuD'_2$ (lanes 2 and 8), UmuD'D (lanes 3 and 9), Cross-linked $UmuD(F94C)_2$ (lanes 4 and 10), and mock-treated $UmuD(F94C)_2$ (lanes 5 and 11).

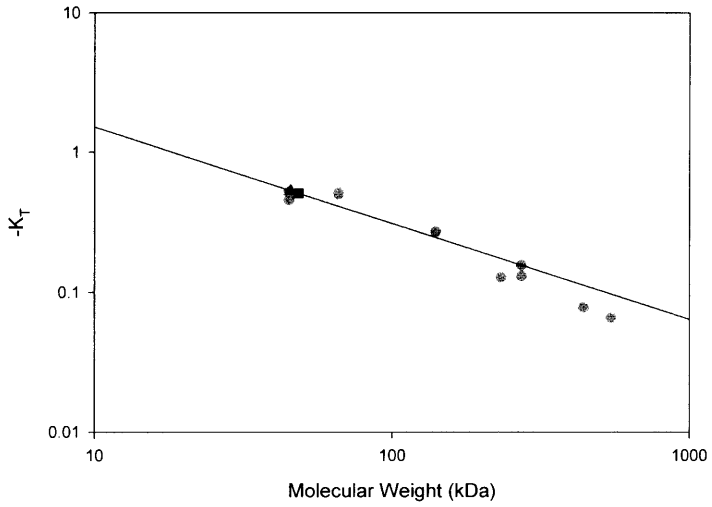
elute slightly earlier than expected for a globular dimer, these data are best explained by a nonglobular dimer of both $UmuD_2$ and $UmuD'_2$, consistent with other proteins that lack significant secondary structure [329]. Denatured UmuD′ and UmuD behave as monomers, eluting earlier than their native counterparts and just before denatured chymotrypsin (14 kDa) (Figure 4-S3B, p.134).

A native PAGE gel of $UmuD_2$ and $UmuD'_2$ at 500 nM (20 μL) and 5 μM (2 μL) shows that the proteins are dimeric at both uninduced and SOS-induced physiological concentrations (Figure 4-S3C, p.134), consistent with earlier studies at 6 μM [328]. The major $UmuD_2$ band runs nearly identically to a UmuD derivative, $UmuD(F94C)_2$, that has been covalently cross-linked in the dimeric form by disulfide bonds [134, 330]. An equimolar mixture of UmuD and UmuD′ at these concentrations shows a predominant intermediate band, corresponding to the UmuD′D heterodimer, rather than two distinct monomeric bands. The theoretical pI of all of these proteins is 4.5, making charge effects on migration negligible. A Ferguson plot of $UmuD_2$ and $UmuD'_2$ compared to native PAGE standards shows that both $UmuD_2$ and $UmuD'_2$ are most similar to the 45 kDa size standard (Figure 4-3A, p.138), consistent with gel filtration and inconsistent with a monomeric form of UmuD or UmuD′ at physiological concentrations.

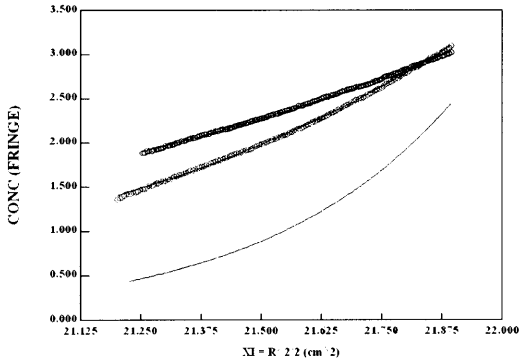
In an effort to determine the K_D of $UmuD_2$ and $UmuD'_2$ homodimers, equilibrium analytical ultracentrifugation was performed at three rotor speeds. The best fit of the data is to a single species model (Figures 4-3 B and C, p.138). The predicted molecular weight of UmuD′ at 20 μM is 25.4 kDa, compared to the monomer molecular weight of 12.5 kDa (Figure 4-3B, p.138). The same model for UmuD at 40 μM (Figure 4-3C, p.138) results in a fitted molecular weight of 31.0 kDa, in comparison with the predicted

Figure 4-3: **$UmuD_2$ and $UmuD'_2$ are dimeric at physiologically relevant concentrations.**

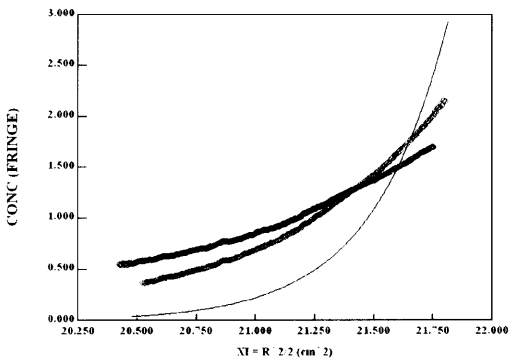
A



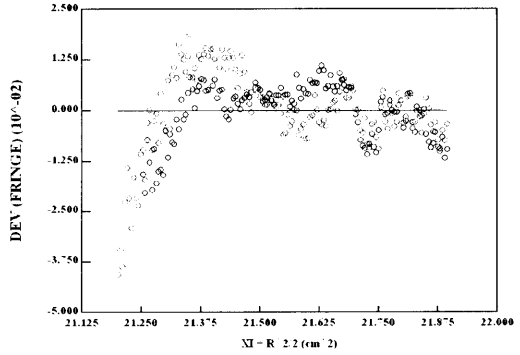
B



C



D



E

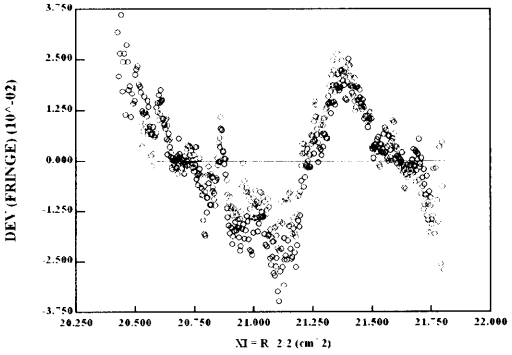


Figure 4-3 (continued): **$UmuD_2$ and $UmuD'_2$ are dimeric at physiologically relevant concentrations.** (A) Ferguson plot of native PAGE size standards (gray circles), $UmuD'_2$ (black square), and $UmuD_2$ (black triangle). The K_T for each protein was determined by plotting the retention factor (log scale) vs. acrylamide concentration (linear scale) for each protein and finding the slope of the best fit regression line. The best fit of the plot of $-K_T$ vs. molecular weight is to $y=7.3408x^{-0.6868}$. R=0.958. Solving for the molecular weight of UmuD gives an estimate of 46 kDa and for UmuD′ an estimate of 49 kDa; the difference is not statistically significant. Native gel standards are: Jack bean urease hexamer (545 kDa) equine spleen ferritin (440 kDa), jack bean urease trimer (272 kDa), bovine liver catalase (232 kDa), bovine heart lactate dehydrogenase (140 kDa), bovine serum albumin (66 kDa), and chicken egg white ovalbumin (45 kDa). Where more than one data point is present, multiple protein isoforms were analyzed. Equilibrium analytical ultracentrifugation of 20 μM $UmuD'_2$ (B) and 40 μM $UmuD_2$ (C). Data for 3 different rotor speeds (16,000 rpm, black circles; 20,000 rpm, dark gray circles; and 30,000 rpm, light gray circles) are plotted with the best fit theoretical curve (single species of dimeric molecular weight) overlaid. (D) Residuals from data fitting to (B). (E) Residuals from data fitting to (B).

monomeric molecular weight of 15.1 kDa. If data are fit to a monomer-dimer equilibrium model, the K_D generated is infinitely low. Residuals, though somewhat nonrandom (Figures 4-3 D and E, p.138), are small and do not improve with fits to other theoretical models. The lower limit of K_D determination for a simple monomer-dimer equilibrium using analytical ultracentrifugation is about 10^{-11} M [331]. Thus, despite the CD spectra at low concentrations, both $UmuD_2$ and $UmuD'_2$ are dimers with K_Ds of <10 pM. Interestingly, the K_D of the related protein LexA has also been shown to be in the pM range [1].

A covalently linked variant of $UmuD_2$ has as CD spectrum resembling a random coil.

To confirm that the random coil CD signal of $UmuD_2$ does not require a monomeric species, we took the spectrum of disulfide cross-linked $UmuD(F94C)_2$, which has been shown to bind the β processivity subunit of DNA Pol III in a similar manner to wild-type [330]. Surprisingly, even though this variant cross-links nearly quantitatively (Figure 4-4A, p.141), it shows slightly less propensity for secondary structure than an otherwise equivalent mock-treated sample of UmuD(F94C) (Figure 4-4B, p.141).

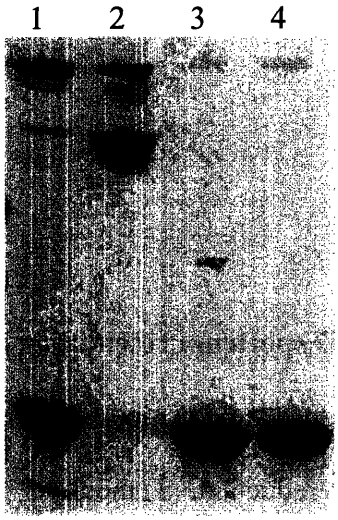
We have no evidence of stable higher-order oligomers of $UmuD_2$ or $UmuD'_2$ at 2 mM, wherein the CD spectrum shows considerable secondary structure, and $UmuD'_2$ at these concentrations has been shown to be dimeric [142]. Native PAGE of 5 μM (20 μL) and 2 mM (0.5 μL) $UmuD_2$, $UmuD'_2$, and UmuD'D shows that all proteins have a consistent retention factor regardless of starting concentration (Figure 4-4C, p.141).

Discussion

These studies have led us to conclude that, at physiologically relevant concentrations, $UmuD'_2$ and $UmuD_2$ have structural characteristics of intrinsically

Figure 4-4: **A covalently linked variant of $UmuD_2$ has a CD spectrum resembling a random coil.**

A



B

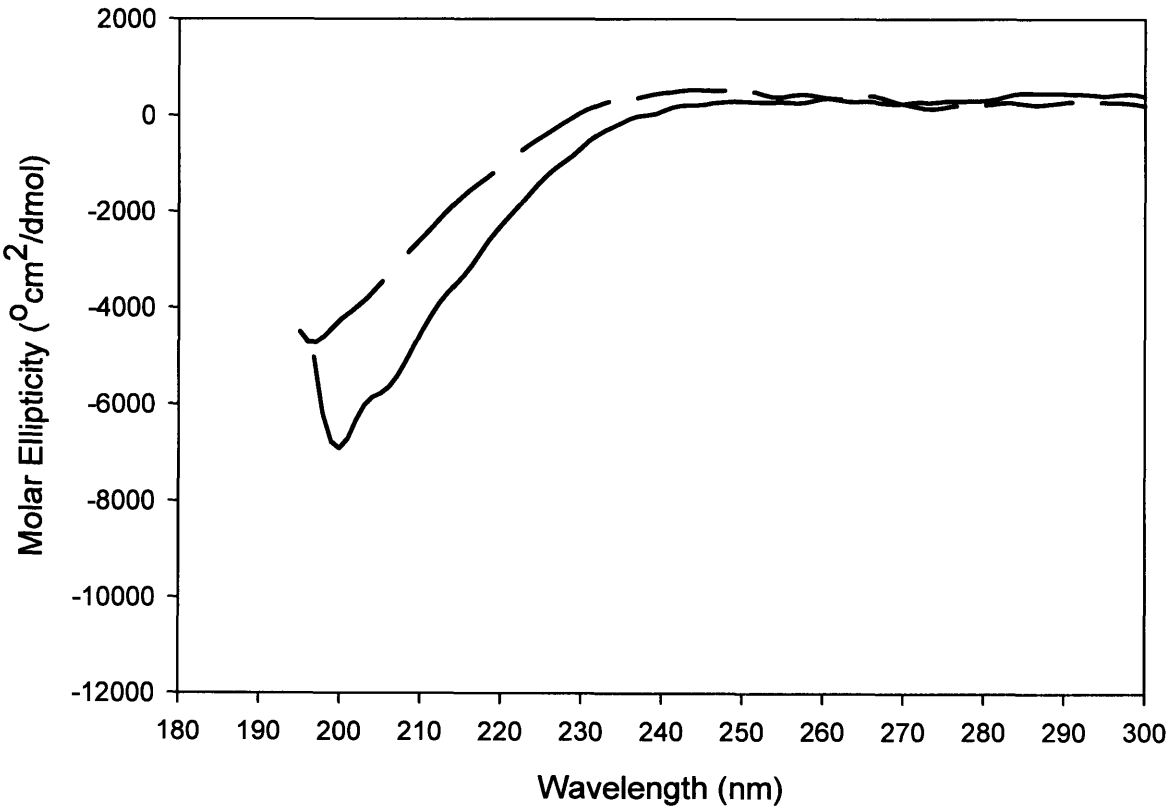


Figure 4-4 (continued): **A covalently linked variant of $UmuD_2$ has a CD spectrum resembling a random coil.**

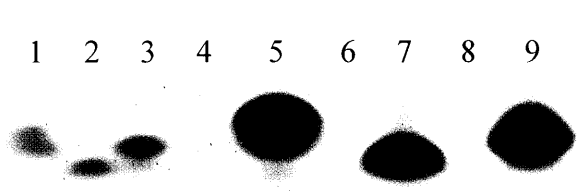


Figure 4-4: **A covalently linked variant of $UmuD_2$ has as CD spectrum resembling a random coil.** (A) Extent of cross-linking of UmuD(F94C). Lanes are: (1) mock-treated UmuD(F94C), no reductant (2) cross-linked UmuD(F94C), no reductant, (3) mock-treated UmuD(F94C) with 1 mM DTT (4) cross-linked UmuD(F94C) with 1 mM DTT. (B) CD spectra of cross-linked (dashed line) or mock-treated (solid line) UmuD(F94C). (C) Native gel electrophoresis of physiological and high concentrations of *umuD* gene products. Lanes are: 5 μM (lanes 1-3) or 2 mM (lanes 5-9) $UmuD_2$ (lanes 1 and 5), $UmuD'_2$ (lanes 2 and 7), or UmuD'D (lanes 3 and 9).

disordered proteins. Little is known about the precise structures of intrinsically disordered proteins, although several efforts to further characterize them have begun [178, 187]. However, in the case of $UmuD_2$ and $UmuD'_2$, a considerable amount of structural information is already available from studies of single cysteine cross-linking and solvent accessibility at physiologically relevant concentrations [88, 134, 143, 144]. Consistent with an unfolded or flexible structure, cross-linking of single cysteine derivatives of $UmuD_2$ by slow, gentle methods such as dialyzing out the reducing agent shows that most derivatives will cross-link to form covalent $UmuD_2$, with only a few positions that cross-link much more or less than average [144]. However, certain amino acid positions are consistently more solvent-exposed than others, and faster methods of cross-linking that give a snapshot of protein-protein contacts distinguish residues that are near the homodimer interface [88, 143, 144]. The homodimer interface shown by these methods is consistent with the structures of $UmuD'_2$ [141, 142] and our four models of $UmuD_2$ [147], or with intermediate conformations between these models, suggesting that $UmuD_2$ is likely to have a flexible but nonrandom structure in solution.

The insights obtained here suggest that at least one type of disorder may have a structure that is not very dissimilar to a β sheet. This is consistent with theories of protein unfolding that suggest some secondary-like structure persists in unfolded proteins, but it involves fewer hydrogen bonds and allows more possible conformations than native secondary structure [184, 185]. Therefore, the high resolution structures of $UmuD'_2$ [141, 142] both may have relevance to its structure *in vivo*, although inside the cell, *umuD* gene products are likely to be surrounded by interaction partners that may influence their actual structure (Figure 4-5A, p.114).

It is not known whether multiple interactions with *umuD* gene products occur simultaneously or in a stepwise fashion. Hub proteins, which are found in the interactomes of many organisms and make a multitude of protein-protein contacts, can be distinguished by which mechanism they use [227]. We suggest that the properties of intrinsically disordered proteins such as $UmuD_2$ and $UmuD'_2$ might provide a simple mechanism for temporally ordering multiple protein interactions. One can easily imagine that an initial protein-protein interaction may constrain the conformations of an intrinsically disordered protein in such a manner as to expose a preferential binding interface for a second protein. After the second protein binds, the structure may change even more to expose or occlude other binding interfaces (Figure 4-5B, p.114). The interactions between $UmuD_2$ and the β subunit of Pol III or DinB result in a similar secondary structure for $UmuD_2$ (Figure 4-4, p.141), but the amino acids in $UmuD_2$ that are important for each interaction are different. $UmuD_2$ seems to interact with β mainly through a patch of amino acids near the N-terminus [330], whereas the interaction with DinB requires a distinct patch of amino acids in the C-terminal domain [77]. It is possible that DinB may bind first by virtue of its lower K_D and induce a conformation of $UmuD_2$ that exposes a binding site to the β processivity subunit, increasing the affinity of the $UmuD_2$-β interaction.

Although many proteins bind to the β processivity subunit of Pol III, the interaction between β and $UmuD_2$ is unusual in that a specific three-dimensional fold exposing particular amino acids of $UmuD_2$ appears to be necessary [330]. Other β-interacting proteins often have a variant of a conserved pentapeptide β-binding motif that is proposed to bind to a single position on β [305]. Since β is a dimer [332], a maximum

of two of these proteins may bind to β at any given time. However, $UmuD_2$ does not have this β binding motif [147, 333]; its interactions with β therefore might not interfere with proteins that bind through the pentapeptide motif. For intrinsically disordered proteins, a requirement for a particular fold ensures that protein-protein interactions only occur when the necessary conditions are met [160]. Perhaps for the $UmuD_2$-β interaction, one necessary condition is a prior contact with DinB.

The crystal structures of the catalytic domains of several DNA polymerases have been solved, but N-terminal or C-terminal protein-protein interaction domains are sometimes removed to enable crystallization [334]. It is possible that a tendency towards disorder in these regions in the absence of their interaction partners precludes their representation in the crystal structures. Although UmuC and DinB do not have regions that align with these protein-protein interactions domains, they both interact with the quasi-disordered *umuD* gene products [77, 120]. We suggest that, instead of being fused to a particular DNA polymerase, $UmuD_2$ and $UmuD'_2$ may act as interchangeable protein-protein interaction domains for the two Y family DNA polymerases in *E. coli*, thus allowing for a streamlined genome while maintaining the regulatory sensitivity of an intrinsically disordered interaction domain (Figure 4-5C, p.114). A flexible structure that can adapt to multiple distinct protein-protein interactions helps explain how the small *umuD* gene products can make many specific protein-protein interactions, and a post-translational modification further differentiates these interactions (Figure 4-5A, p.114).

Although intrinsically disordered proteins are often involved in protein-protein contacts, few are known homodimers in solution. Stable quaternary structure in the absence of rigid secondary structure is counterintuitive to the current protein folding

paradigm [162]. However, limited examples are present in the literature. *E. coli* MazE antitoxin has both a disordered domain and a structured dimerization domain [219]. The homodimeric bacterial histone-like proteins HU-α_2 and HU-β_2 first undergo an order-to-disorder transition upon thermal denaturation and only subsequently dissociate [222]. Most similar to $UmuD_2$ and $UmuD'_2$ is human papillomavirus (HPV) protein E7, which is both dimeric *in vitro* and has characteristics of intrinsically disordered proteins [2]. Interestingly, $UmuD'_2$ and $UmuD_2$ show less secondary structure by CD than HPV E7 [2].

How much of a protein must be rigidly structured in order to allow homodimerization? Type II restriction endonucleases such as EcoRV undergo massive domain rearrangements upon binding to DNA while maintaining a constant dimer interface representing about 10% of the amino acid sequence [335]. The dimerization interface of the 115 amino acid protein UmuD′ comprises about 16 amino acids, including the C-terminal β sheet and the α helix at the base of the N-terminal arms [142]. The dimer interface of the 139 amino acid UmuD is predicted to involve about 21 residues, again including the C-terminal β sheet in addition to extensive contacts of the N-terminal arms along the C-terminal globular domain [147]. These dimer interfaces represent between 10% and 15% of amino acid composition, and a rigid β-sheet like structure stabilizing the dimer through these residues would not make a significant contribution to an overall random coil CD spectrum. Interestingly, two of three regions of the *umuD* gene products that disorder prediction programs calculate to have the greatest propensity for secondary structure (Figure 4-S1, p.123) are known to be at the

dimer interface of $UmuD'_2$ (amino acids 40-45 and 131-139), as would be expected for proteins that are largely disordered but still form stable dimers [142].

$UmuD_2$ and $UmuD'_2$ share homology with the dimerization domains of certain bacterial transcription factors [91]. Transcription factors often have large regions of intrinsic disorder, either in their DNA binding domains or in protein interaction domains [168]. The seemingly unrelated tendencies for transcription to be homodimeric and intrinsically disordered suggests that more homodimers with a large degree of structural plasticity may be found soon.

Materials and Methods

Materials. Copper phenanthroline, jack bean urease, chymotrypsin, chicken egg white ovalbumin, BSA, and lysozyme were purchased from Sigma-Aldrich. RecA protein was purchased from New England Biolabs. High molecular weight native PAGE standards were obtained from GE Healthcare. 7.5%, 10%, 12%, and 15% acrylamide PAGE gels were obtained from BioRad. 4-20% PAGEr tris-glycine gradient gels were obtained from Cambrex Pharmaceuticals.

Protein Purification. Purification of $UmuD'_2$, $UmuD_2$, and $UmuD(F94C)_2$ was performed as previously described [317]. A plasmid encoding UmuD(F94C) was produced from pSG5 using the Stratagene QuikChange site-directed mutagenesis kit [317]. Copper phenanthroline cross-linking was performed as previously published [112]. Protein concentration was determined using the BioRad Protein Assay. DinB was a kind gift from Daniel Jarosz [317]; the β processivity subunit of Pol III was generously provided by the Beuning Lab at Northeastern University [147] and ClpXP protease from the Baker Lab at MIT [127].

CD Spectroscopy. Circular dichroism was performed on an Aviv Model 202 CD spectrometer. Spectra were recorded at 25 °C; each data point represents the average of three seconds of data collection. Proteins at physiological concentrations were monitored using a 350 μL 0.1-cm cuvette (Hellma Glassware), whereas proteins at 50 μM or greater concentrations were recorded using a 4 μL 0.01 cm cuvette (Wilmad Glass, Buena, NJ). Spectra of *umuD* gene products alone were recorded in 10 mM $Na_3(PO_4)$ pH 6.8, 100 mM NaCl, 0.1 mM EDTA, 1mM DTT. For protein-protein interaction studies, the buffer was 50 mM Hepes pH 7.4, 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol. The spectrum of a buffer blank was subtracted from the spectrum of protein. Spectra in the presence of crowding agents were taken in CD buffer with addition of 0.2 M proline or 2.5 M glucose as indicated.

Limited proteolysis. $UmuD'_2$ or $UmuD_2$ was diluted to 10 μM in CD buffer and either incubated on ice for 2 hours or used within 1 minute. Proteolysis reactions were begun by adding 10 μL 5 mg/mL chymotrypsin to 10 μL $UmuD'_2$ or $UmuD_2$ and incubating at 37 °C for 5 minutes. Reactions were stopped by addition of 4 μL 6x SDS-PAGE loading buffer (1x is 25 mM Tris–HCl, pH 6.8, 5% glycerol, 0.1% bromphenol blue, 2% SDS, 1 mM DTT) and freezing in liquid nitrogen. Proteins were run on 4-20% Tris-glycine gels, stained with 1x SYPRO Orange (Molecular Probes) in 7.5% acetic acid, and quantified using ImageQuant software.

Protein Disorder Prediction. Access to PONDR® was provided by Molecular Kinetics (6201 La Pas Trail - Ste 160, Indianapolis, IN 46268; 317-280-8737; E-mail: main@molecularkinetics.com). VL-XT is copyright©1999 by the WSU Research

Foundation, all rights reserved. PONDR® is copyright©2004 by Molecular Kinetics, all rights reserved.

Gel filtration chromatography. Gel filtration was performed using a 100 mL Superdex 75 column on an Akta FPLC system (GE Healthcare). One mL of protein solution was injected; $UmuD_2$ and $UmuD'_2$ were 5 μM at injection. CD buffer described above was used as running buffer. For denatured gel filtration, $UmuD'_2$, $UmuD_2$, and each size standard was denatured separately in CD buffer + 6 M guanidinium hydrochloride for 2 hours, centrifuged for 1 minute in a microcentrifuge at 14,000 rpm to pellet aggregates, and injected as above. CD buffer + 6 M guanidinium hydrochloride was used for elution.

Native PAGE. Proteins were diluted into 1x PAGE Loading Buffer lacking SDS (25 mM Tris–HCl, pH 6.8, 5% glycerol, 0.1% bromphenol blue, 1 mM DTT), incubated for 30 min at 25 °C to promote heterodimer formation, and run at 20 V at 4 °C overnight. Cross-linked $UmuD(F94C)_2$ was diluted into 1x PAGE Loading Buffer lacking both SDS and DTT. Gels were soaked in 0.05% SDS for 30 min and stained with 1x SYPRO Orange (Molecular Probes) in 7.5% acetic acid after running. Ferguson plots were calculated as described using 5 μM $UmuD_2$, $UmuD'_2$, and UmuD'D [336].

Sedimentation Equilibrium. Experiments were performed on a Beckman Model XL-I analytical ultracentrifuge at 20 °C. Proteins were dialyzed against 3 changes of 500 mL CD buffer at 4 °C over 12 hours. The reference solution was the final dialysis buffer. Protein gradients were monitored by interference. Each rotor speed was centrifuged for 12 hours to attain equilibrium, scanning every hour, and WinMatch software was used to confirm that equilibrium had been reached. Rotor speeds were 16000, 20000, and 30000 rpm; the same protein samples experienced all three rotor speeds. Only the last scan for

each speed was used in data analysis. Protein concentration was determined by direct analysis of each sample after the last scan. Data analysis was performed using the software WinNonlin.

Acknowledgements

We would like to thank members of the G.C. Walker, T.A. Baker, R.T. Sauer, A.E. Keating, J.A. King, P.J. Beuning, W. DeGrado, J.W. Little (especially Kim Giese), and G.A. Petsko/D. Ringe (especially Raquel Lieberman) Labs for helpful discussions and use of equipment; Debbie Pheasant for help with biophysical instrumentation; and Brent Cezairliyan for help in AUC analysis. CD and AUC were performed in the MIT Biophysical Instrumentation Facility. This work was supported by grant number CA21615 from the National Cancer Institute. S.M.S. was a Cleo and Paul Schimmel Fellow.

CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

I have shown that *umuD* gene products are a new member of the family of intrinsically disordered proteins, which lack a precise three-dimensional fold *in vitro* at physiological concentrations but instead are made up of an ensemble of freely exchanging conformations [153, 162, 163, 174, 325, 337, 338]. In one respect, this characterization is completely consistent with previous literature. The mystery of how such small proteins can be involved in such a diverse array of protein-protein contacts can finally be explained by *umuD* gene products' high degree of flexibility. Many intrinsically disordered proteins are thought to be able to use their flexible nature to make a number of highly specific but low affinity interactions to effect diverse regulatory processes [96, 120, 131, 135, 339]. UmuD shares qualities with other intrinsically disordered proteins via its protease and regulatory activities and through its homology to several dimeric transcription factors [168], although *umuD* gene products do not bind DNA themselves [91, 340]. The primary structures of UmuD and UmuD′ also share similarities to other known intrinsically disordered proteins by virtue that they are small proteins with repetitive sequences and have a pI value (pI = 4.5) is far from neutral [341].

However, one might not have predicted this intrinsically disordered characterization outright because a wealth of structural information is already available for *umuD* gene products. UmuD′$_2$ has been crystallized [141, 342] and its structure confirmed by NMR [142, 327], and although several attempts at high resolution structure determination of UmuD$_2$ have not yielded results, it is thought that UmuD$_2$ is less flexible in solution than UmuD′$_2$ [100, 134]. This characterization is consistent with preliminary data that suggest that polymeric crowding agents can induce a β-sheet conformation in UmuD$_2$ but not in UmuD′$_2$ (Chapter 4). Finally, a number of single and double-cysteine studies suggest that, although they may be flexible in solution, both UmuD$_2$ and UmuD′$_2$ have defined homodimer interfaces and nonrandom structures in general [88, 134, 143, 144]. These studies, in conjunction with those presented in Chapter 4, may shed light on the nonrandom characteristics of intrinsically disordered proteins. Preliminary results show that the UmuD′D heterodimer, which is thermodynamically more stable than either

$UmuD_2$ or $UmuD'_2$, is also disordered at 5 μM, and its spectrum is very close to the averaged spectra of UmuD and UmuD′ (Figure 5-1, p.153).

Because both $UmuD_2$ and $UmuD'_2$ share the common feature of intrinsic disorder at physiological concentrations and a more ordered structure at higher concentrations, the differences between $UmuD_2$ and $UmuD'_2$ have not been highlighted in this report. It would be interesting in the future to study the differences between all three dimeric *umuD* gene products and the underlying causes of these differences. Intrinsic disorder helps explain how $UmuD_2$ and $UmuD'_2$ can make as many highly specific protein-protein contacts as they do, but it remains a mystery how the relatively small change of removing 24 amino acids causes a large difference in preferences for different binding partners. One possibility is that when the N-terminal arms of $UmuD_2$ fold over the C-terminal gobular domain, a binding interface on $UmuD'_2$ is obscured, and a new one is formed by the arms. However, the fact that residues on the N-terminal arms of $UmuD_2$ can form a dimer interface [143] suggests that the N-terminal arms may not always obscure the C-terminal globular domain (Figure 5-2, p.154). In favor of this hypothesis, cross-linked single-cysteine derivatives on the N-terminal arm of $UmuD_2$ cannot undergo RecA*-mediated cleavage, whereas cross-linked derivatives on the C-terminal globular domain can [112], suggesting that the arms are far away from the active site in these cross-linked conformations. Computational studies of the dynamics of $UmuD_2$ conformational exchange may be able to estimate the relevance of an "arms-up" conformation of $UmuD_2$ and test the hypothesis that differences in binding preference result from different binding interfaces with and without the N-terminal arms.

Several biochemical differences between $UmuD_2$ and $UmuD'_2$ become evident when working with these proteins. $UmuD'_2$ purification generally results in a better yield and cleaner protein than $UmuD_2$ purification [317]. However, addition of a single protein-protein interactor in stoichiometric quantities tends to cause aggregation of $UmuD'_2$ instead of the stable complex usually seen with $UmuD_2$. Furthermore, equilibrium guanidinium denaturation curves of $UmuD_2$ and $UmuD'_2$ at two different protein concentrations differ (Figure 5-3, p.156). Whereas the curves for $UmuD'_2$ are offset relative to each other, confirming that denaturation is not a simple unimolecular process

Figure 5-1: **CD spectrum of 10 μM UmuD'D.**

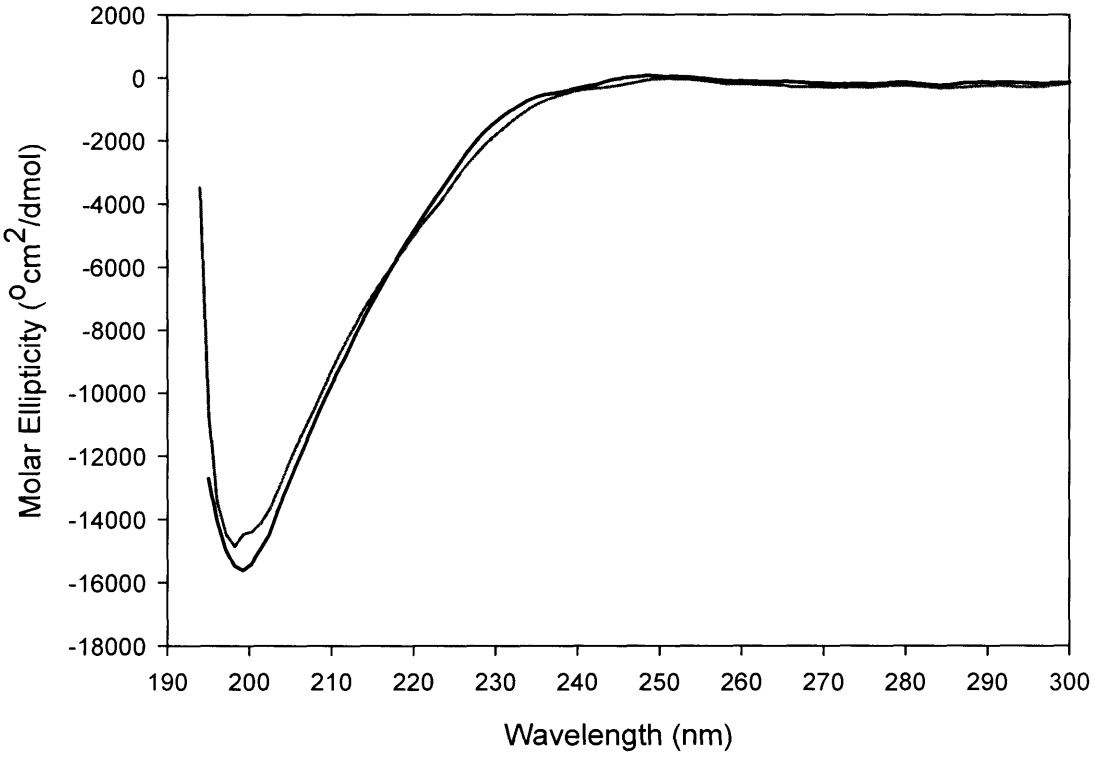


Figure 5-1: **CD spectrum of 10 μM UmuD'D.** The CD spectra of 5 μM $UmuD_2$ and 5 μM $UmuD'_2$ were added mathematically and are plotted as the gray line. The CD spectrum of 10 μM UmuD'D (5 μM $UmuD_2$ and 5 μM $UmuD'_2$ combined) is plotted as the black line. CD signal was measured at 1 nm intervals.

Figure 5-2: **A model for an alternate $UmuD_2$ dimer interface**

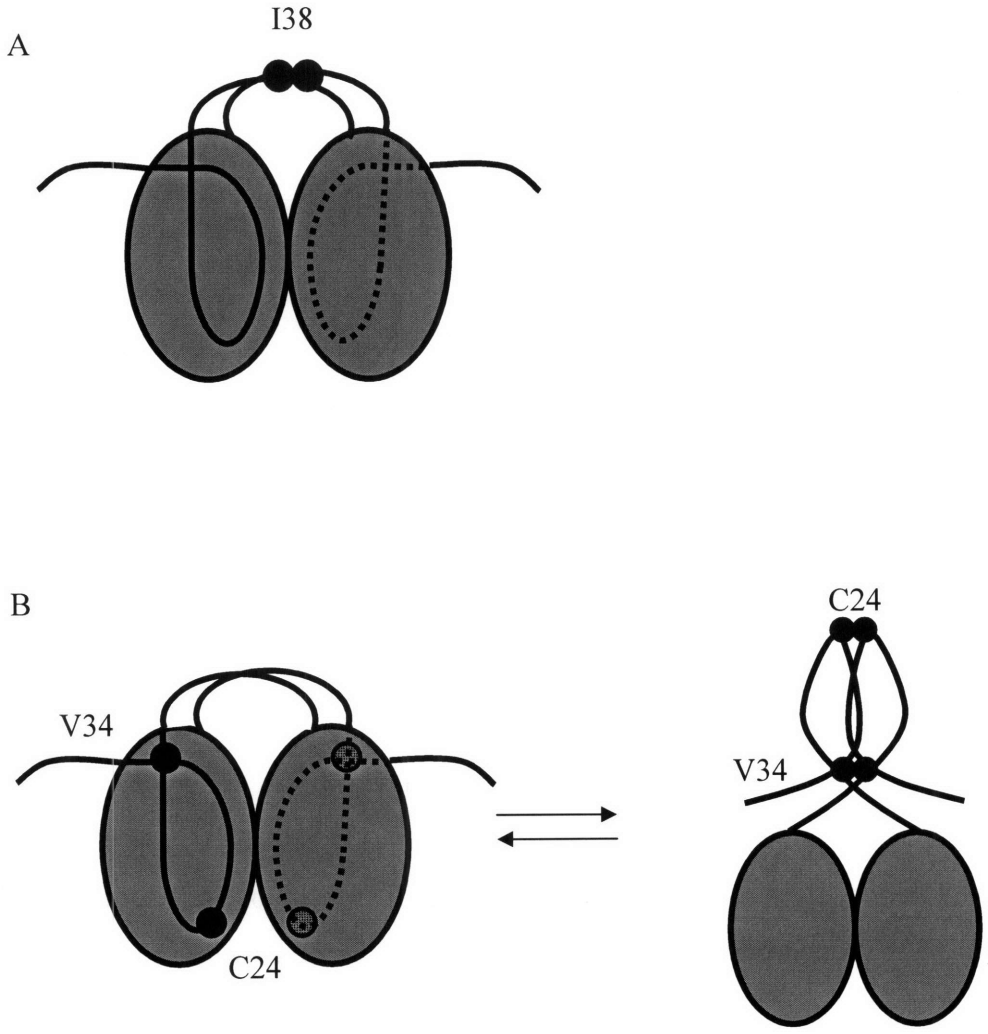


Figure 5-2 (continued): **A model for an alternate $UmuD_2$ dimer interface.** (A) Single-cysteine cross-linking studies have generally found that residues that are near the C2 axis of symmetry in $UmuD'_2$ [142] cross-link robustly as homodimers in $UmuD_2$ (A representative residue from [143] is shown in cartoon form). (B) Some residues that are on the N-terminal arms also cross-link robustly even though they are predicted to be far from the dimer interface (data from [143]). Based on our four proposed models of $UmuD_2$ [147], it is possible that the N-terminal arms may spend considerable time between the *cis* and *trans* forms, forming a temporary dimer interface that would explain the unexpected cross-linking results.

Figure 5-3: **Differences in equilibrium guanidinium denaturation activity of $UmuD_2$ and $UmuD'_2$**

A

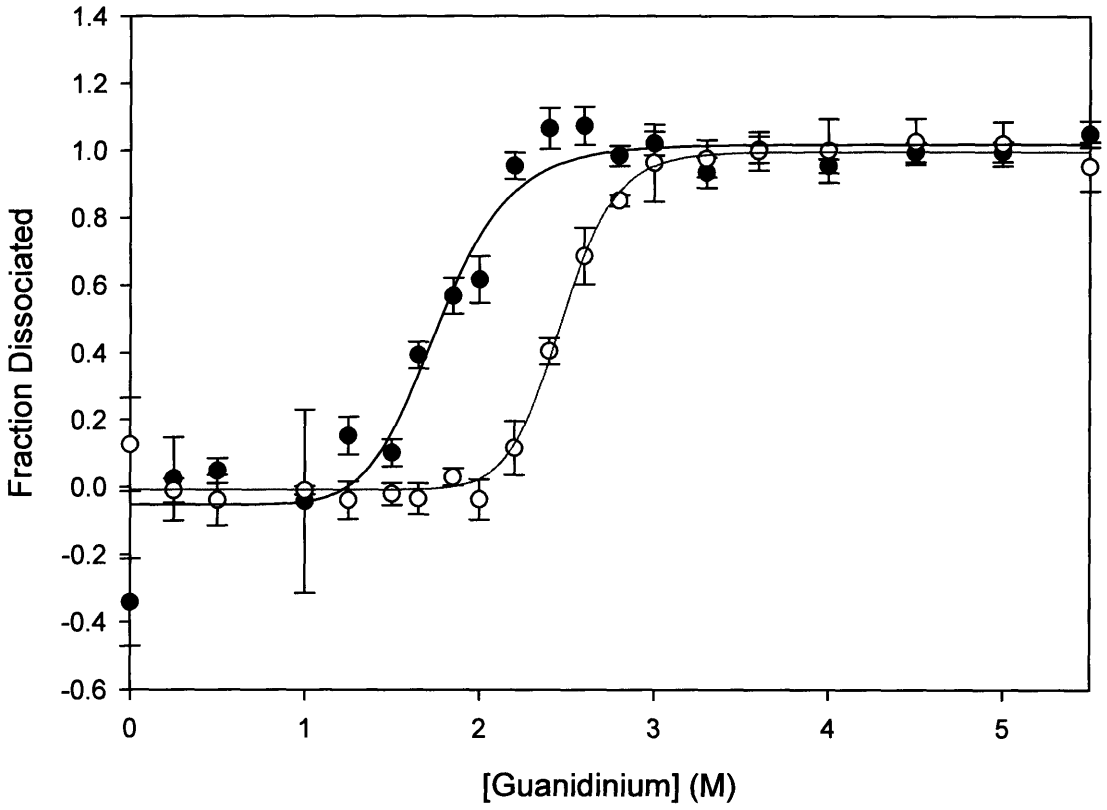


Figure 5-3 (continued): **Differences in equilibrium guanidinium denaturation activity of $UmuD_2$ and $UmuD'_2$**

B

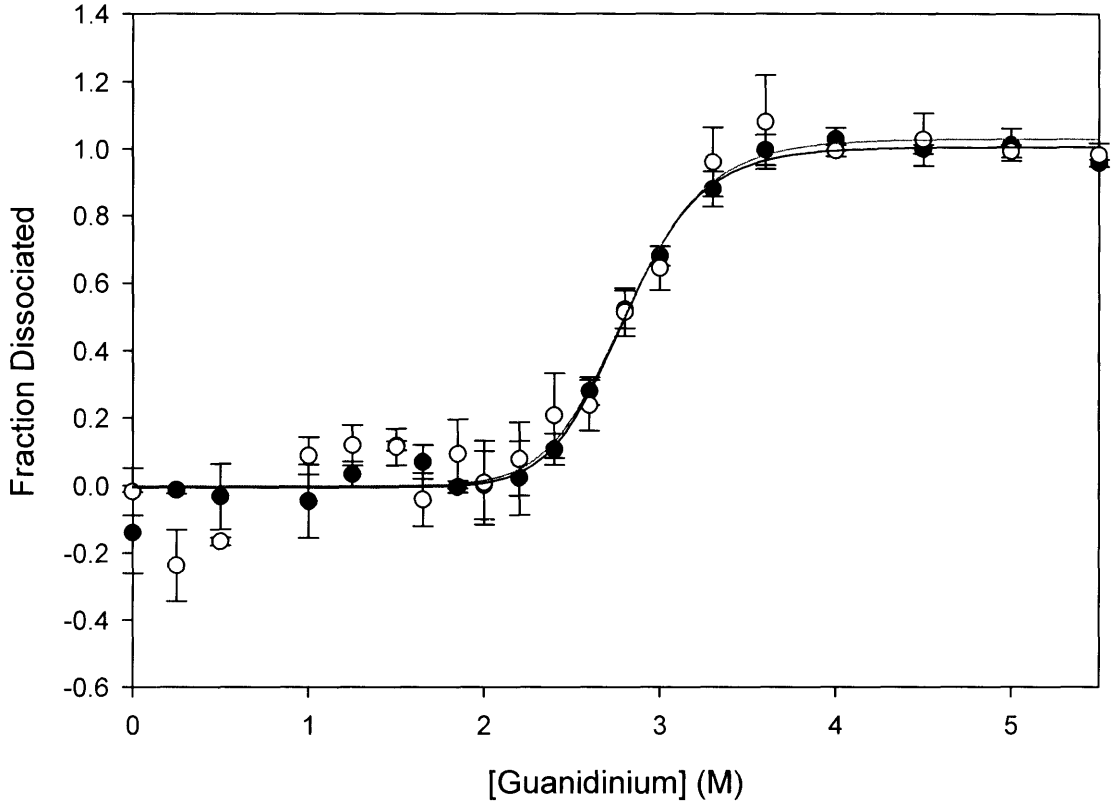


Figure 5-3: **Differences in equilibrium guanidinium denaturation activity of $UmuD_2$ and $UmuD'_2$.** (A) Guanidinium denaturation of $UmuD'_2$ at 5 μM (black symbols) and 50 μM (white symbols). (B) Guanidinium denaturation of $UmuD_2$ at 5 μM (black symbols) and 50 μM (white symbols).

[343], those for full-length $UmuD_2$ overlap almost completely. Nevertheless, denaturation of $UmuD_2$, like $UmuD'_2$, involves a dimer-to-monomer transition (Chapter 4). These differences in guanidinium denaturation behavior are not well understood but may result from multiple distinct conformations of $UmuD_2$ in solution [344], which have been proposed to occur [147]. Experiments are currently underway in P.J. Beuning's lab to further characterize this phenomenon.

The structure of the dimer interface of UmuD'D appears to be more similar to that of $UmuD_2$ than $UmuD'_2$ [134]. It is likely that the heterodimer has many of the same structural characteristics of the two homodimers (CD in Figure 5-1, p.153, and native PAGE in Chapter 4), but aside from studies of ClpXP [126, 127, 323, 328], functions and characteristics of the thermodynamically most-stable [121] heterodimer are poorly understood. Given its similarity to $UmuD_2$ and $UmuD'_2$, the heterodimer may make many more protein-protein contacts than is currently known, and it may have other important roles in regulating the timing of the SOS response than are currently appreciated. In fact, UmuD'D is known to bind DinB (V. Godoy and D. Jarosz, personal communication), although its interaction with UmuC results in aggregation [128].

The concept of an intrinsically disordered protein that has a stable quaternary structure is counterintuitive to the current protein folding paradigm, where primary structure dictates secondary structure, which is required for tertiary structure and finally quaternary structure [162]. However, there are known examples of dimeric proteins that are either characterized as intrinsically disordered (such as HPV protein E7 [2]) or otherwise exhibit substantial flexibility (*E. coli* MazE [219] and the histone-like proteins HU-α and HU-β [222]), showing that exceptions to the ordered protein folding paradigm are already known.

It is likely that the paradoxical of a dimeric intrinsically disordered protein stems from a general lack of understanding about the actual structures of intrinsically disorderd proteins. In fact, many of these proteins have nonrandom structures [174], and just as there are several types of rigid secondary structure, three types of disorder have been proposed [178]. Disordered regions can also be highly conserved [163], although it has also been proposed that intrinsically disordered proteins might have evolved through their ability to perform a function while maintaining a flexible structure that is resistant to

mutation [191]. Point mutations in *umuD* gene products, however, may not favor the latter hypothesis. Almost every point mutation that has been made for *umuD* gene products, including very conservative ones, disrupts at least one of its functions (Appendix II). Furthermore, *umuD* gene products do appear to be highly conserved, including in *Pelagibacter ubique*, the bacterium with the smallest known genome to date [137].

Products of *umuD* share homology to the dimerization domains of certain transcription factors, proteins that are more likely than many others to be intrinsically disordered [168]. However, it is usually the DNA-binding or protein-protein interaction domains of transcription factors that are intrinsically disordered, not their dimerization domains. $LexA_2$, the transcription factor most homologous to *umuD* [91], is nevertheless relatively well folded *in vitro* [345]. However, once its N-terminal domain is removed, the C-terminus, which has the most homology to UmuD′, tends to aggregate, consistent with the notion that it, like UmuD′ may not have its own stable fold *in vitro* (R. Mohana-Borges, personal communication). The C-terminal domain of LexA is rapidly degraded *in vivo*, preventing aggregation inside the cell [87]. Additionally, X-ray crystallography of $LexA_2$ shows substantial disorder in one of the two monomers of the dimeric unit [146]. Furthermore, the C-terminal domain of the λ CI repressor protein, which has homology to LexA, unfolds at a higher temperature than the N-terminal domain [82], consistent with disordered protein segments [159].

Examples of dimeric transcription factors with intrinsically disordered domains include the viral Tax protein [188] and the bacterial ParG required for plasmid separation [346, 347]. In both of these cases, individual domains have been studied and classified as intrinsically disordered. The well-structured dimerization domain is removed for these studies, similarly to how poorly folded domains are often removed for crystallization.

Our current model for the functions of *umuD* gene products is that they act as interchangeable protein-protein interaction domains for the two *E. coli* Y-family DNA polymerases (Chapter 4), whereas Y family DNA polymerases in eukaryotes tend to have protein-protein interaction domains that are contiguous with the catalytic domain [334]. In *E. coli*, the constitutively-expressed DNA Pol I, involved in error-free DNA repair processes, does not interact with *umuD* gene products and has a contiguous intrinsically

disordered domain [348]. *E. coli* DNA Pol III, the replicative polymerase, also has an extended C-terminal region that was removed to allow crystallography [34].

Intrinsically disordered proteins have been further classified into more-unstructured "coil-like" proteins and intermediate "premoltenglobule-like" proteins based on the ratio of their CD signals at 200 nm (indicative of disorder) vs 222 nm (indicative of rigid secondary structure) [187]. $UmuD_2$ and $UmuD'_2$, have CD spectra more consistent with the intermediate premoltenglobule-like intrinsically disordered proteins, as might be predicted due to their stable quaternary structures. This is also true of HPV protein E7, which is another dimeric intrinsically disordered protein [2], and human α-synuclein, which has some primary structural similarity to UmuD (Figure 1-1) [159].

It is not certain how much rigid secondary structure, if any, must be present for a protein to form a stable dimer. However, studies of the type II restriction endonuclease EcoRV may provide a clue [335]. Several crystal structures of this protein have been solved under many conditions, showing that despite massive domain rearrangements that depend on crystallization conditions, the dimer interface remains constant. The dimer interface represents about 10% of the protein's size, suggesting that as little as 10% of a protein must be rigid in order to maintain dimerization; the rest of the protein may be free to move about independently. The dimer interfaces of $UmuD_2$ and $UmuD'_2$ make up more than 10% of these proteins, but even if they are rigid under all circumstances, the dimer interfaces may not significantly contribute to an otherwise random coil CD spectrum. Alternatively, dimerization may not always require recognizable α-helix or β-sheet structures for stabilization. As an example, the T7 gene 2.5 protein is a well-folded dimer, but the dimer interface is composed mostly of loops [349].

One aspect of *umuD* gene products that has not been studied in detail yet is the wealth of UmuD variants that appear to be deficient in dimerization, particularly mutants at the dimer interface [143, 145, 350]. Some of these proteins, though not all, are less stable than wild-type UmuD *in vivo* [145, 350]. If the monomer of UmuD or UmuD' is even structured than the dimer, it would be an ideal substrate for rapid proteolysis *in vivo* [160, 163, 174]. It would be interesting to characterize the proteins that are stable and yet apparently deficient in homodimerization. They may be stabilized by other protein-protein contacts even in the monomeric form, either by enabling an otherwise inefficient

dimerization or simply through binding as a monomer. Protein-protein contacts involving RecA are already known to stabilize $UmuD'_2$ *in vivo* [351]. Alternatively, even a monomeric UmuD, though rare, may have enough nonrandom structure to be protected from proteolysis.

It is not known if *umuD* gene products are unfolded under cellular conditions as they are *in vitro*. Increased folding *in vivo* seems likely, since the cellular environment both is more crowded and contains a number of proteins with which *umuD* gene products interact. UmuD and UmuD′ activity assays, including primer extension, RecA*-mediated autocleavage, and ClpXP proteolysis, almost invariably involve interactions with one or more additional proteins (the exception is base-catalyzed autocleavage of $UmuD_2$, which is an excessively slow *in vitro* reaction) [317]. Additionally, the cellular environment is crowded with macromolecules, and certain crowding agents have an effect on both $UmuD_2$ and $UmuD'_2$ (Chapter 4). Therefore, even in the absence of protein-protein interactions, $UmuD_2$ and $UmuD'_2$ may adopt a better defined structure *in vivo* than at physiological concentrations *in vitro*. A small group of scientists has proposed more emphasis on biochemical studies in the presence of crowding agents rather than in simple aqueous solution, as cellular crowding may have important effects on proteins that biochemical researchers are missing [213].

Since *umuD* gene products are likely to be highly flexible, how relevant are the X-ray and NMR structures of $UmuD'_2$ [141, 142] and the models of $UmuD_2$ [134, 147]? It is likely that they have some relevance *in vivo*, since the two high resolution structures both agree on several aspects of the protein fold (it is dimeric, it is largely a β-sheet, and the regions of α-helicity are in the same locations), and the X-ray crystal structure shows the known active site residues in a conformation that is poised for catalysis [141]. Under conditions used to solve the NMR structure of $UmuD'_2$ [142], even the full-length $UmuD_2$ protein, whose structure has not been solved, shows a signal that is consistent with the expected β-sheet [134, 147]. Additionally, both the high-resolution structures [141, 142] and the models [134, 147] are consistent with the single-cysteine data [88, 134, 143, 144]. Although the crystallography and NMR studies were performed at protein concentrations that are much higher than are physiologically relevant [77, 141, 142], macromolecular crowding by high concentrations of protein may mimic *in vivo*

conditions in a way that *in vitro* experiments at physiologically relevant concentrations do not [213, 214].

An interesting characteristic of $UmuD_2$ is that, although it is not strictly an enzyme [92], it does have proteolyic activity [101-103]. However, in contrast to most metabolic enzymes, and even relative to other similar autocleavage proteins, this activity is not very fast [113]. Many proteases are intrinsically disordered, and some of them rely on this disorder to accept a wide range of substrates [163]. Disorder may help $UmuD_2$ regulate the timing of its transition to $UmuD'_2$ by undergoing efficient cleavage only in the presence of RecA* [142]. However, disorder may also play a role in allowing efficient cleavage. At low concentrations wherein *umuD* gene products are more disordered than at high concentrations, more spontaneous UmuD′ reproducibly forms during native gel electrophoresis (Chapter 4). It is possible that some of the more ordered forms present at higher protein concentrations are less likely to undergo spontaneous cleavage.

One explanation for why UmuD conversion to UmuD′ may appear inefficient *in vivo* is that low levels of UmuD′ may be present mostly in the UmuD′D heterodimer, which would rapidly be degraded by ClpXP [323]. Heterodimer formation at the beginning of the SOS response may occur through subunit exchange [121], or it is possible that the initial product of $UmuD_2$ cleavage is the UmuD′D heterodimer rather than $UmuD'_2$. An intriguing result is that a UmuD variant, UmuD(C24A), that has been shown to undergo RecA*-mediated cleavage more quickly than wild-type [112] has a defect in *in vivo* mutagenesis (Appendix II), suggesting a role of intact $UmuD_2$ in mutagenesis. This result is dependent on the wild-type promoter sequence for reasons that are not yet understood (Appendix II). Additionally, a Δ*clpXP* strain is nonmutable, as would not be expected given the role ClpXP has in ending the mutagenic phase of the SOS response [126], and Lon protease that degrades $UmuD_2$ does not appear to efficiently degrade UmuD′D heterodimers [323]. It is possible that, if UmuD′D is the immediate product of $UmuD_2$ cleavage, $UmuD_2$ may also serve as a competitive substrate for ClpXP [127], protecting a subset of UmuD′D until it can form the mutagenically active $UmuD'_2$ [93].

It is uncertain whether multiple protein-protein interactions with *umuD* gene products do, or even can, happen simultaneously. Even though they can participate in a relatively large number of protein-protein interactions, *umuD* gene products are small proteins. Therefore, unless each of the interactions took up only a very small amount of surface space and did not occlude nearby binding sites by steric hindrance, there cannot be more than a few distinct simultaneous protein-protein interactions with *umuD* gene products.

However, $UmuD'_2$ is symmetrical except for the highly unstructured N-terminal 15 amino acids [142], and it is quite possible that $UmuD_2$ may also be symmetrical [134, 147]. Most of the known interaction partners are asymmetrical. Presumably an identical surface interface is available for a protein-protein interaction on each side of the dimer. In the case of the interactions between $UmuD_2$ and DinB or the β processivity subunit of Pol III, all three proteins interact with each other, and the predicted binding sites do not overlap to a great extent (Chapter 4). It is possible that all three proteins are in contact with each other at the same time, at least transiently. The different affinities between $UmuD_2$ and DinB or β suggest that there may be an order of addition effect such that $UmuD_2$ binds DinB first, exposing a binding site for β on $UmuD_2$. Alternatively, DinB and β may already be in a complex before binding to $UmuD_2$, facilitating $UmuD_2$'s interaction with β.

Modern biophysical techniques allow determination of detailed, high-resolution protein structures, but the prevalence of intrinsically disordered proteins (an estimated 60% of eukaryotic proteins and 10-30% of prokaryotic ones) may illustrate the limits of this type of focus. Currently, high-resolution structural data are only available for a subset of proteins, mostly enzymes that make up much but certainly not all of the proteome [162]. Although certain membrane proteins have recently been crystallized, they, like intrinsically disordered proteins, remain technically difficult [352, 353]. Even as structure-determination technology improves on a daily basis, concerns as to the relevance of structures gathered at such high homogenous protein concentrations and the current technical limitations ensure that other methods of studying and computing protein structure, dynamics, function, and mechanistic details will be important to capture the full essence of proteins and their genetic and biochemical interactions. In the case of the

structures of UmuD′$_2$, two possible dimer interfaces were seen in the X-ray crystal structure [141], and NMR and biochemical studies were necessary to determine the correct dimer interface [88, 134, 142-144, 327]. As is becoming increasingly obvious, interdisciplinary approaches to problem-solving are necessary to give the most complete answers to biological questions.

APPENDIX I

Characterization of *Escherichia coli* Translesion Synthesis Polymerases and their Accessory Factors

Published as part of: Beuning, P.J., Simon, S.M., Godoy, V.G., Jarosz, D.F., and Walker, G.C. "Characterization of Escherichia coli Translesion Synthesis Polymerases and their Accessory Factors." *Methods in Enzymology* Volume 408 (2006) p. 318-340.

Only sections to which S.M. Simon contributed significantly are included.

ABSTRACT

Members of the Y family of DNA polymerases are specialized to replicate lesion-containing DNA. However, they lack 3′-5′ exonuclease activity and have reduced fidelity compared to replicative polymerases when copying undamaged templates, and thus are potentially mutagenic. Y family polymerases must be tightly regulated to prevent aberrant mutations on undamaged DNA while permitting replication only under conditions of DNA damage. These polymerases provide a mechanism of DNA damage tolerance, confer cellular resistance to a variety of DNA-damaging agents, and have been implicated in bacterial persistence. The Y family polymerases are represented in all domains of life. *Escherichia coli* possesses two members of the Y family, DNA pol IV (DinB) and DNA pol V ($UmuD'_2C$), and several regulatory factors, including those encoded by the *umuD* gene that influence the activity of UmuC. This chapter outlines procedures for in vivo and in vitro analysis of these proteins. Study of the *E. coli* Y family polymerases and their accessory factors is important for understanding the broad principles of DNA damage tolerance and mechanisms of mutagenesis throughout evolution. Furthermore, study of these enzymes and their role in stress-induced mutagenesis may also give insight into a variety of phenomena, including the growing problem of bacterial antibiotic resistance.

Purification and Characterization of *umuD* Gene Products

Expression and Purification of UmuD and UmuD'

Expression and purification of the *umuD* gene products have been typically accomplished using *umuD*$^+$ *E. coli* strain BL21(DE3) harboring a *umuD* expression plasmid. Unfortunately, a drawback of the use of this strain for the preparation of UmuD variants is that some contaminating wild-type protein will be present due to expression from the chromosomal copy of the gene. Therefore, a chloramphenicol-resistant Δ*umuDC* derivative of *E. coli* BL21 was constructed by P1 transduction from GW8017 [144, 354, 355]. UmuD and UmuD′ expression plasmids, pSG5 and pSG4, respectively, are based on pET11t [356] and were constructed from pMAD (UmuD) or pMADp (UmuD′), respectively [357]. An *Nde*I site was added 5′ of the *umuD* gene in pMAD or the *umuD′* gene in pMADp, by site-directed mutagenesis, and the *Nde*I–*Bam*HI fragment was ligated into the *Nde*I–*Bam*HI linearized pET11t. The integrity of the construct was confirmed by sequencing. Colony-dependent variation in expression has been observed in our laboratory when using the pET11t expression plasmid so several colonies are routinely screened for efficient overexpression.

The major advantage of this purification protocol over those published previously [102, 143, 327] is the use of fast flow FPLC columns, reducing purification time greatly. Using a single FPLC, two UmuD variants can be purified simultaneously, with one preparation on the instrument while the other is assayed. One should allow two consecutive days after growth, induction, and harvesting for the preparation of UmuD proteins.

Transformed BL21(DE3) Δ*umuDC* are grown at 37°C in LB medium or M9 medium [358] supplemented with 100 μg/ml ampicillin in baffled flasks until an OD_{600} of 0.6 to 0.8 is reached. Isopropyl-β-D-thiogalactoside (IPTG) is added to 1 mM, at which time the culture is transferred to 30°C for 3–4 h. Cells are harvested by centrifugation at 5000g for 30 min and resuspended in 15 ml lysis buffer (50 mM Tris–HCl, pH 8.0, 2 mM EDTA) per liter of culture. After resuspension, cells can be frozen dropwise in liquid nitrogen and stored at -80°C indefinitely or can be lysed immediately. Typically, protein from 2 to 4 liters of culture is purified at once.

Frozen cells are thawed on ice overnight at 4° C. All subsequent steps are performed at 4° C. Cell lysis is accomplished by two passages at 10 kpsi through a French press (Thermo Spectronic) or via sonication at 50% output with alternating 15-s bursts followed by 15-s rest periods for 3–4 min. After lysis, cell debris is pelleted by centrifugation at 14,000g for 30 min.

The first ammonium sulfate precipitation of the supernatant is to 20% saturation (0.121 g/ml). The precipitate is removed by centrifugation at 14,000g for 30 min, and the supernatant is subjected to a second ammonium sulfate precipitation, this time adding ammonium sulfate to 40% saturation (0.151 g/ml). Centrifugation is the same as for the first ammonium sulfate precipitation, but the pellet is retained. The pellet itself may be stored on ice overnight. However, once chromatography is started, it must be completed within the same day, as either extensive storage at 4°C or freezing and thawing results in a loss of yield.

The pellet is resuspended in 50% PSA (10 mM sodium phosphate, pH 6.8, 0.1 mM EDTA, 1 mM dithiothreitol [DTT], 1 M $[NH_4]_2SO_4$) and 50% PSB (10 mM sodium phosphate, pH 6.8, 0.1 mM EDTA, 1 mM DTT) plus one Complete, EDTA-free protease inhibitor cocktail tablet (Roche Molecular Biochemicals), until there is no visible debris (about 10 ml per liter original culture). The sample is centrifuged at 14,000g for 30 min to pellet remaining debris and then filtered through a 0.25-μm Millex GV syringe filter. The filtered sample is loaded onto a HiTrap fast flow, low substitution phenyl-Sepharose column (GE Healthcare) preequilibrated with PSA. The protein is eluted with a gradient of 0–100% PSB over 5 column volumes at a flow rate of 5 ml/min. Fractions are collected at the end of the gradient, after 80% PSB is reached, and during a 2 column volume wash with PSB. UmuD proteins will elute as the last peak, starting at 100% PSB. Aliquots (10 μl) of each fraction are assayed by 14% SDS–PAGE and stained with Coomassie blue.

Fractions that contain UmuD are pooled and applied to a HiTrap fast flow Q-Sepharose column (GE Healthcare) equilibrated with QA (10 mM sodium phosphate, pH 6.8, 0.1 mM EDTA, 1 mM DTT, 100 mM NaCl). The protein is eluted with a gradient of 0–100% QB (10 mM sodium phosphate, pH 6.8, 0.1 mM EDTA, 1 mM DTT, 1 M NaCl) over 10 column volumes at a flow rate of 5 ml/min. UmuD elutes at approximately 400 mM NaCl. Fractions are analyzed by SDS–PAGE.

Fractions chosen for further purification are concentrated using spin concentrators with a 5-kDa molecular weight cutoff membrane (VivaSpin) to 1–3% of the volume of a Superdex 75 gel filtration column (GE Healthcare). After concentration, the sample is filtered through a Millex GV syringe filter before injection onto the column. The column is run isocratically at 1.5 ml/min with buffer QA. UmuD elutes as the last peak. Fractions are analyzed for purity by SDS–PAGE. Fractions with pure UmuD are pooled, concentrated to 0.1–4 mM, and flash frozen in liquid N_2 in conveniently sized aliquots. The purified proteins are stored at -80° C, and once an aliquot is thawed, it should be used the same day. The protein is more stable when concentrated than when left dilute after purification. Each liter of induced BL21(DE3) will yield 10–20 mg protein. Both

$UmuD_2$ and $UmuD'_2$ can be purified with this protocol, although purifying $UmuD'_2$ will tend to result in higher yields than $UmuD_2$.

ClpXP Degradation Assay

UmuD and UmuD′ exist in three dimeric species: $UmuD_2$ or $UmuD'_2$ homodimers or the most thermodynamically stable UmuD′D heterodimer [121, 143, 339]. When $UmuD_2$ or $UmuD'_2$ homodimers are coincubated *in vitro* at 25°C in equal concentration, the only dimeric species detectable after 30 min is the UmuD′D heterodimer [121]. In this context, the purified ClpXP protease will specifically target the UmuD′ partner of the heterodimer for degradation in the presence of ATP [328].

This assay allows for evaluation of the ability of a protein to act as either a substrate of ClpXP degradation, as do UmuD′ and UmuD, or an adaptor protein delivering substrates to ClpXP, as does UmuD [127]. It is important when making these comparisons that a wild-type UmuD′D control is included in each assay, as the extent of UmuD′ degradation can vary with ClpXP preparation and length of storage.

ClpX and ClpP were a gift from Professor Tania Baker at MIT [127, 359]. UmuD′D is formed by coincubating 4.5 μM $UmuD_2$ and 4.5 μM $UmuD'_2$ with 0.8 μM $ClpP_{14}$ in 50 mM Tris–HCl, pH 8.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT at 30°C in 38-μl reactions for 30 min. A 2-μl aliquot of 20× ATP regeneration mix [50 mM creatine phosphate, 1 mg/ml rabbit muscle creatine kinase (both from Sigma-Aldrich), and 80 mM ATP] is added. The reaction is initiated with addition of $ClpX_6$ to 0.3 μM. At this point, a 20-μl aliquot of the reaction is removed and quenched with 5 μl 4× SDS–PAGE loading buffer as a t = 0 control. The quenched reaction is flash frozen in liquid N_2 and stored at -20°C until analysis. The remaining reaction is incubated at 30°C for a given time, at which point it is quenched and frozen as described earlier. Typically, the reaction is complete after 2 h.

Samples are boiled 2 to 3 min and analyzed by 14% SDS–PAGE. Due to the small molecular weight difference between UmuD and UmuD′, the gel must be run to completion to resolve both species. The gel is stained with SYPRO Orange (Molecular Probes), and UmuD/UmuD′ are visualized and quantitated by phosphorimager and ImageQuant software (GE Healthcare). It is possible that UmuD bands will become fainter over the time course of the reaction as well, as ClpXP will degrade one UmuD in the context of the homodimer in the absence of the UmuD′D heterodimer (Figure AI-2, p.170) [127].

Figure AI-2

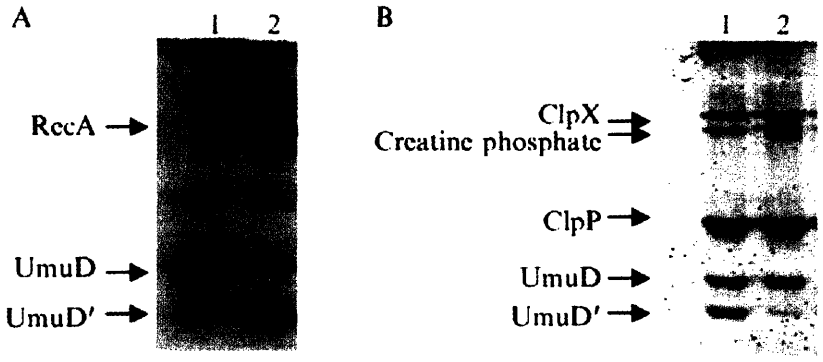


FIG. 2. (A) *In vitro* RecA/ssDNA filament-facilitated cleavage of UmuD. Lane 1, −RecA; lane 2, +RecA $t = 1$ h. A small amount of autocleavage is observed in lane 1. Unlabeled bands are impurities in the UmuD preparation. (B) *In vitro* ClpXP proteolysis of wild-type UmuD' with wild-type UmuD as adaptor protein. Lane 1, $t = 0$; lane 2, $t = 2$ h.

Using Intrinsic Tryptophan Fluorescence to Determine Binding Constants between UmuD Proteins and Their Interaction Partners

Dimeric forms of the UmuD proteins interact with a variety of other *E. coli* proteins [75, 109, 120, 131, 287, 328, 339, 340, 360, 361]. Although the interactions made with $UmuD_2$ or with $UmuD'_2$ largely overlap, some differ in magnitude [131, 330] or result in different effects on the interacting protein. Namely, $UmuD'_2$ activates UmuC as a translesion polymerase [75, 108, 109, 357], while $UmuD_2C$ is involved in a DNA damage checkpoint [97, 100]. In an effort to gain insight into the myriad roles that the different forms of the UmuD protein play in regulation of the *E. coli* SOS response, it is important to quantitate the affinity of the dimeric UmuD proteins for their interactors.

Because UmuD has no tryptophan, binding of UmuD proteins to most other proteins will not contribute significantly to the overall emission spectrum of the complex. Samples are prepared in 50 mM HEPES, pH 7.5, 0.1 mM EDTA, 100 mM NaCl. Unless there is danger of aberrant disulfide bond formation, no reductant is added due to the intrinsic fluorescence of DTT and the high volatility of β-ME. To measure the effect of UmuD binding on the intrinsic tryptophan fluorescence of other proteins, the tryptophan-containing protein should be kept at a convenient concentration (on the order of 5 μM), while the concentration of the interacting UmuD protein can be varied from 0 to 200 μM or more, if necessary. A separate sample of UmuD at the same concentration is made in the absence of an interactor. The fluorescence of UmuD alone is subtracted from the spectrum of the combined proteins to eliminate noise from the aromatic residues in UmuD. For each binding curve, the fluorescence from a sample of the interacting partner in the absence of UmuD is made as a starting point.

Samples with more than one protein present are preincubated at room temperature for 2 h prior to data collection in order to allow the protein complex to reach equilibrium. The samples are excited at either 278 nm for measurement of all aromatic residues or at 295 nm to selectively excite tryptophan. Emission is measured from 300 to 400 nm. To ensure that the complex has reached equilibrium, the emission spectrum is acquired at additional 1- to 2-h intervals until no change in the spectrum is observed.

The center of spectral mass $\Sigma(\lambda \times I)_i/\Sigma I_i$ is calculated for each sample, where λ is the wavelength in nanometers and I is the intensity in arbitrary units for each step i. The center of spectral mass is plotted vs. [UmuD], and the plot is fit to:

$$\lambda_{obs}=\lambda_0\pm[\lambda_1[(C_0+D+K_D)-((C_0+D+K_D)^2-4C_0\times D)^{0.5}]]/(2C_0)$$

[362], where λ_{obs} is the variable center of spectral mass at UmuD protein concentration D, with λ_0 as λ_{obs} at $D = 0$, and C_0 is the fixed concentration of interacting protein.

If UmuD′D is tested, an additional step is required to generate UmuD′D from purified UmuD and UmuD′. The two proteins are mixed in equimolar concentrations and are allowed to equilibrate for 30 min at room temperature before addition of the interacting protein.

Conclusions

We have illustrated common techniques for *in vivo* and *in vitro* analysis of Y family polymerases from *E. coli*. The experiments described here lend insight into the functions of these specialized polymerases and their accessory factors in cellular survival and mutagenesis. Major unanswered questions about these polymerases include detailed methods of regulation, especially temporal control of activity and polymerase switching according to specific lesions. Further characterization will undoubtedly continue to address these questions and reveal the properties of the Y family polymerases and their accessory factors.

APPENDIX II

Mutations in *umuD* Gene Products and Their Resultant Phenotypes

Residue #	Side Chain	Observations
1	M	
2	L	
3	F	
4	I	
5	K	
6	P	
7	A	
8	D	UmuDΔ2-8 cannot target UmuD′ for ClpXP proteolysis; it can be targeted for proteolysis but not as well as UmuD′ (Figure AII-1, p.175). The defect may be in heterodimerization or in recognition.
9	L	Implicated in targeting UmuD′ for proteolysis by ClpXP but not by Lon [127, 328, 363]. UmuD L9A R10A E11A I12A is proficient for RecA cleavage and has greater stability *in vivo* than wild-type UmuD [328]; it has greater than wild-type *in vivo* mutagenesis activity [363].
10	R	Implicated in targeting UmuD′ for proteolysis by ClpXP but not by Lon [127, 328, 363]. UmuD L9A R10A E11A I12A is proficient for RecA cleavage and has greater stability *in vivo* than wild-type UmuD [328]; it has greater than wild-type *in vivo* mutagenesis activity [363].
11	E	Implicated in targeting UmuD′ for proteolysis by ClpXP but not by Lon [127, 328, 363]. UmuD L9A R10A E11A I12A is proficient for RecA cleavage and has greater stability *in vivo* than wild-type UmuD [328]; it has greater than wild-type *in vivo* mutagenesis activity [363]. UmuD E11V I12V V13K is more efficient than wild-type UmuD at targeting UmuD′ for ClpXP-mediated proteolysis [127].

Figure AII-1

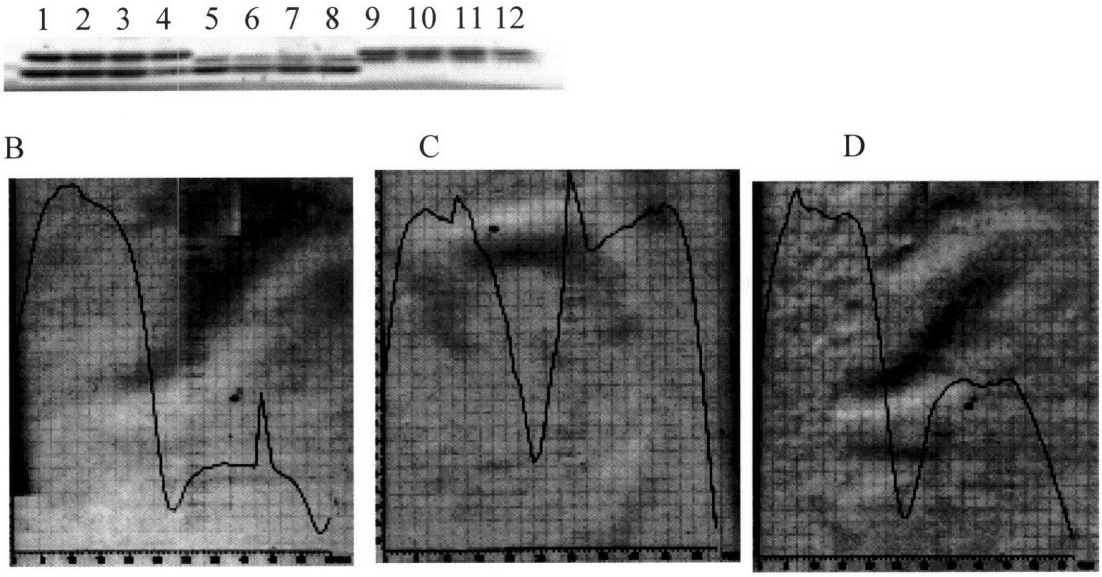


Figure legend: ClpXP proteolysis of UmuDΔ2-8. (A) Gel of proteolysis reactions. Lanes are: (1) wild-type UmuD and UmuD′, no ClpXP, t = 0 (2) wild-type UmuD and UmuD′, no ClpXP, t = 2 hr (3) wild-type UmuD and UmuD′, + ClpXP, t = 0 (4) wild-type UmuD and UmuD′, + ClpXP, t = 2 hr (5) UmuDΔ2-8 and UmuD′, no ClpXP, t = 0 (6) UmuDΔ2-8 and UmuD′, no ClpXP, t = 2 hr (7) UmuDΔ2-8 and UmuD′, + ClpXP, t = 0 (8) UmuDΔ2-8 and UmuD′, + ClpXP, t = 2 hr (9) wild-type UmuD and UmuDΔ2-8, no ClpXP, t = 0 (10) wild-type UmuD and UmuDΔ2-8, no ClpXP, t = 2 hr (11) wild-type UmuD and UmuDΔ2-8, + ClpXP, t = 0 (12) wild-type UmuD and UmuDΔ2-8, + ClpXP, t = 2 hr. (B-D) Density scan of the lower bands of +ClpXP lanes in the gel in (A). In each case, t = 0 is on the left; t = 2 hr is on the right. (B) Wild-type UmuD and UmuD′. (C) UmuDΔ2-8 and UmuD′. (D) Wild-type UmuD and UmuDΔ2-8. Procedure is in [317].

12	I	Implicated in targeting UmuD′ for proteolysis by ClpXP but not by Lon [127, 328, 363]. UmuD L9A R10A E11A I12A is proficient for RecA cleavage and has greater stability *in vivo* than wild-type UmuD [328]. it has greater than wild-type *in vivo* mutagenesis activity [363]. UmuD E11V I12V V13K is more efficient than wild-type UmuD at targeting UmuD′ for ClpXP-mediated proteolysis [127].
13	V	UmuD E11V I12V V13K is more efficient than wild-type UmuD at targeting UmuD′ for ClpXP-mediated proteolysis [127].
14	T	
15	F	Implicated in targeting UmuD for proteolysis by Lon protease [363]. UmuD F15A P16A L17A F18A is noncleavable and nonmutable *in vivo* [363].
16	P	Implicated in targeting UmuD for proteolysis by Lon protease [363]. UmuD F15A P16A L17A F18A is noncleavable and nonmutable *in vivo* [363].
17	L	Implicated in targeting UmuD for proteolysis by Lon protease [363]. UmuD F15A P16A L17A F18A is noncleavable and nonmutable *in vivo* [363].
18	F	Implicated in targeting UmuD for proteolysis by Lon protease [363]. UmuD F15A P16A L17A F18A is noncleavable and nonmutable *in vivo* [363].
19	S	Mutation to Cysteine in UmuD has a defect in mutagenesis and RecA* cleavage [143]. The corresponding residue in LexA also has a RecA*-cleavage defect [364, 365]. Single cysteine at this position is moderately solvent accessible; cross-links as $UmuD_2$ poorly with iodine, robustly with BMH, and only moderately with Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ does not depend on protein concentration [143]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144]. A single-cysteine variant of $UmuD_2$ at this position cross-links poorly to RecA* [88, 112]. Single cysteine derivative of UmuD at this position cross-links as $UmuD_2$ moderately with AIA [88]. UmuD S19C C24A does not cross-link efficiently to β [330].

20	D	UmuD D20C C24A F94C forms poor covalent iodine-induced cross-links [134]. Interestingly, two dimers form, presumably one with two disulfide bonds and one with only one [134]. UmuD D20C has wild-type or greater UV-mutagenesis activity (Figure AII-2, p.178).
21	L	UmuD L21C has wild-type or greater UV-mutagenesis activity (Figure AII-2, p.178).
22	V	UmuD V22C C24A F94C robustly forms iodine-induced cross-links to homodimer [134]. Interestingly, two dimers form, presumably one with two disulfide bonds and one with only one [134].
23	Q	ClpXP degradation of $UmuD_2$ results in cleavage after this residue [127].
24	C	Mutation to Alanine in UmuD gives a slight hypermutable phenotype when under control of T7 promoter [143]. C24A has normal extent of glutaraldehyde cross-linking to $UmuD_2$ [143]. Position is relatively solvent accessible; cross-links as $UmuD_2$ well with iodine and moderately with BMH or Cu phenanthroline [143]. Extent of Cu phenanthroline cross-linking as $UmuD_2$ does not depend on protein concentration [143]. This cleavage site position is usually an alanine in other RecA*-dependent autocleavage enzymes, but the bacteriophage Φ80 repressor also has a cysteine in the wild-type sequence [143]. Single-cysteine at this site (wild-type UmuD) cross-links as $UmuD_2$ moderately well by dialyzing out DTT [144]. UmuD C24A undergoes RecA* cleavage to form UmuD′ more efficiently than wild-type UmuD [112]. Although in "data not shown" it is suggested that UmuD C24A does not undergo Cu phenanthroline-catalyzed formation of $UmuD_2$, cleavage of UmuD C24A is inhibited by treatment with Cu phenanthroline [112]. Wild-type $UmuD_2$ disulfide cross-linked at C24 is unable to undergo *in vitro* RecA* cleavage in the absence of reductant [112]. Wild-type $UmuD_2$ is unable to cross-link to RecA*, suggesting that RecA* is not the enzyme in the cleavage reaction but merely a facilitator; also cross-links robustly as $UmuD_2$ by AIA [88]. When expressed from a low-copy plasmid under control of the native promoter, UmuD C24A causes UV-induced mutation frequency intermediate between wild-type and non-cleavable UmuD (Figure AII-3, p.179). However, when expressed from a constitutive promoter,

Figure AII-2

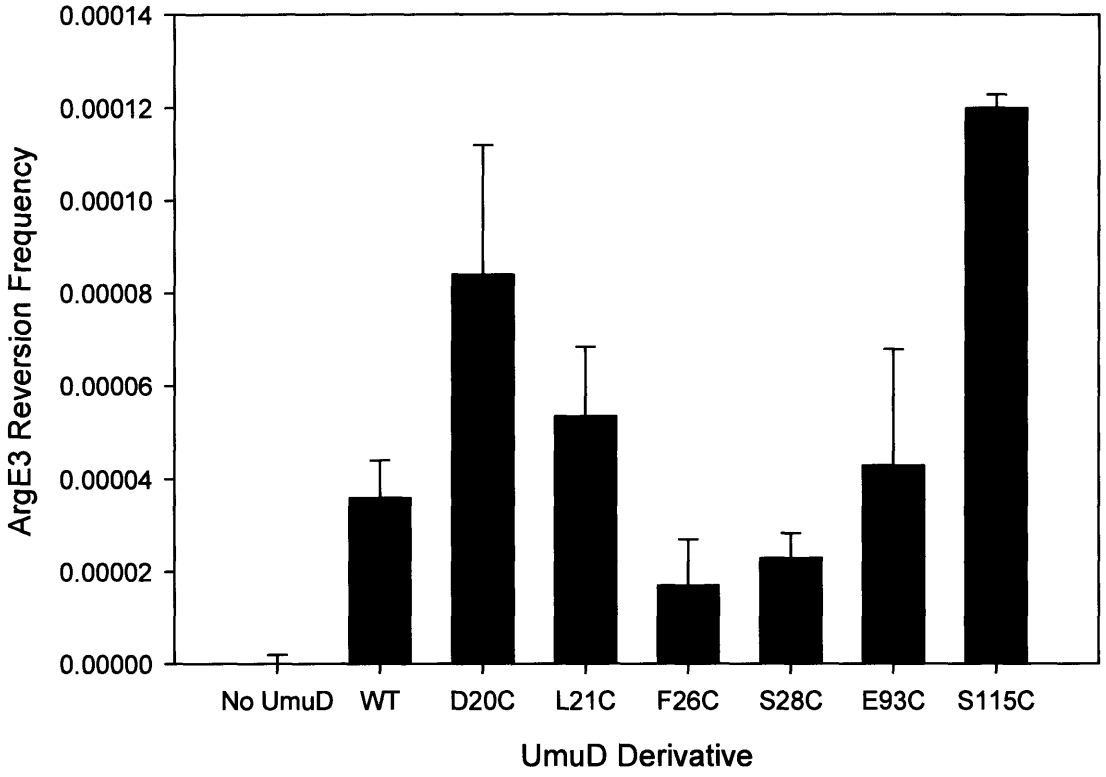


Figure Legend: ArgE3 reversion assay was performed as described at 25 J/m^2 [317], using bacterial strain GW8017 and plasmids pSE101 (empty vector), pSE115 (positive control), pSE115 D20C, pSE115 L21C, pSE115 F26C, pSE115 S28C, pSE115 E93C, and pSE115 S115C.

Figure AII-3

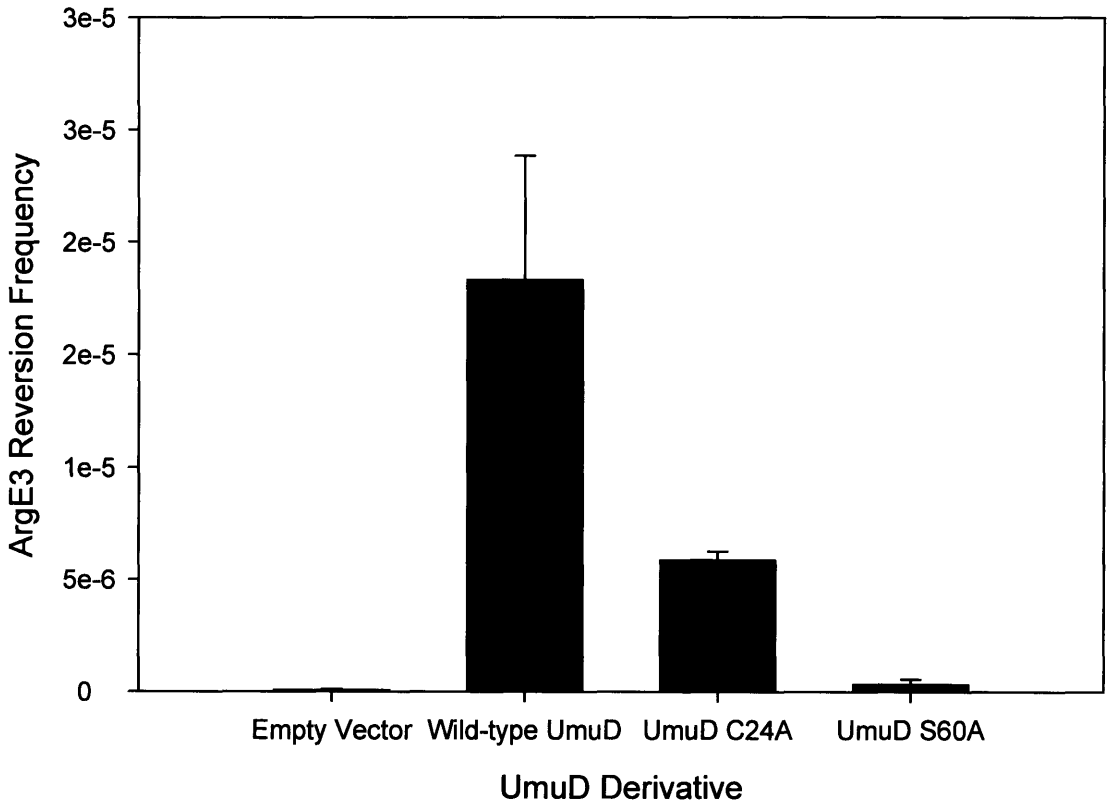


Figure Legend: ArgE3 reversion assay was performed as described at 25 J/m^2 [317], using bacterial strain GW8017 and plasmids pSE101 (empty vector), pSE115 (positive control), pSE115 C24A, and pSE115 S60A.

		UmuD C24A has wild-type activity (Jamie Foti, personal communication). UmuD cross-links robustly to β through amino acid C24 [330].
25	G	
26	F	UmuD C24A F26C F94C form iodine-induced cross-links very poorly [134]. Interestingly, two dimers form, presumably one with two disulfide bonds and one with only one [134]. UmuD F26C has somewhat reduced UV-mutagenesis activity compared to wild-type (Figure AII-2, p.178). Part of a secondary site for Lon proteolysis [363]. UmuD F26A P27A S28A P29A is deficient in cleavage by *recA730* [328] and *in vivo* mutagenesis activity [363].
27	P	P at this position is conserved among UmuD-like proteins, but not among structural homologs [125, 366]. S at this position blocks cleavage when expressed as full-length UmuD [121, 367, 368]. S mutation slightly diminishes steady-state levels when expressed as UmuD′ and reduces MMS and spontaneous mutagenesis to a similar extent (about 50%) [350]. Part of a secondary site for Lon proteolysis [363]. UmuD F26A P27A S28A P29A is deficient in cleavage by *recA730* [328] and *in vivo* mutagenesis activity [363]. UmuD P27S is not degraded by ClpXP, and UmuD(P27S) in the context of UmuD(P27S)UmuD′ is not degraded by Lon (the relevance of the Lon data to the wild-type situation is unknown) [126, 323].
28	S	UmuD C24A S28C F94C cross-links as a homodimer moderately efficiently [134]. Interestingly, two dimers form, presumably one with two disulfide bonds and one with only one [134]. UmuD S28C has somewhat reduced UV-mutagenesis activity compared to wild-type (Figure AII-2, p.178). Part of a secondary site for Lon proteolysis [363]. UmuD F26A P27A S28A P29A is deficient in cleavage by *recA730* [328] and *in vivo* mutagenesis activity [363].
29	P	Part of a secondary site for Lon proteolysis [363]. UmuD F26A P27A S28A P29A is deficient in cleavage by *recA730* [328] and *in vivo* mutagenesis activity [363].
30	A	May be involved in RecA* interaction [143]. Mutations in the corresponding region in λCI protein were found to lower RecA*-mediated cleavage but not base-induced self-cleavage [369]. Single cysteine at this position in $UmuD_2$ is hardly

		impaired at all for UV mutagenesis or RecA* cleavage; cross-links as $UmuD_2$ moderately with iodine and well with Cu phenanthroline [144]. Position is solvent-accessibile [144]. Single-cysteine mutant cross-links as $UmuD_2$ very well by dialyzing out DTT [144]. UmuD′ E30V is defective in UV and MMS-induced mutagenesis, but it is not dominant negative to chromosomal *umuD*$^+$ [145]. Protein levels of UmuD′ E30V are at least as high as wild-type UmuD′ when expressed constitutively, but there may be a slight defect in *in vivo* formaldehyde cross-linking to $UmuD'_2$ [145].
31	A	Single cysteine at this position in $UmuD_2$ is impaired for UV mutagenesis but not RecA* cleavage; cross-links as $UmuD_2$ moderately with both iodine and Cu phenanthroline [144]. Position is solvent-accessibile [144]. Single-cysteine mutant cross-links as $UmuD_2$ moderately well by dialyzing out DTT [144].
32	D	Single cysteine at this position in $UmuD_2$ is very impaired in both UV mutagenesis and RecA* cleavage; cross-links as $UmuD_2$ poorly with iodine and moderately with Cu phenanthroline [144]. Position is moderately solvent-accessible [144]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144].
33	Y	Single cysteine at this position in $UmuD_2$ is impaired in both UV mutagenesis and RecA* cleavage, but the RecA* cleavage defect is much more pronounced *in vitro* than *in vivo* [144]. Position is solvent-accessible [144]. Single cysteine at this position cross-links as $UmuD_2$ moderately poorly with iodine and moderately well with Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144]. This position may be involved in binding ClpXP [127].
34	V	Mutation to Cysteine in UmuD has a defect in mutagenesis and RecA* cleavage [143]. Single cysteine has essentially normal extent of glutaraldehyde cross-linking to $UmuD_2$; forms heterodimers with UmuD′ as efficiently as wild-type UmuD [143]. Single cysteine at this position has low to moderate solvent accessibility; cross-links robustly with iodine or Cu phenanthroline but moderately with BMH [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ increases slightly with protein concentration [143]. Position is not particularly solvent-accessible, and single cysteine at this position cross-links as $UmuD_2$ moderately

		with both iodine and Cu phenanthroline (results are not entirely consistent with those of reference [143]; the discrepancy may be due to higher DTT content in reaction buffer) [144]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144]. A disulfide cross-linked single cysteine variant at this position is unable to undergo *in vitro* RecA* cleavage without reductant, and cleavage is less efficient than wild-type even with reductant [112]. A single cysteine variant of UmuD at this position cross-links efficiently to RecA* (and there are many resulting cross-linked species, as opposed to other single cysteine variants that cross-link to RecA*); also cross-links as $UmuD_2$ moderately well with AIA [88]. UmuD′ V34C does not cross-link robustly to UmuD′ $V34C_2$, but UmuD V34C does cross-link to UmuD $V34C_2$ [134]. V34 is more solvent accessible in UmuD′ than in UmuD [134]. UmuD C24A V34C cross-links robustly to β [330]. This position may be involved in binding ClpXP [127].
35	E	May be involved in RecA* interaction [143]. Mutations in the corresponding region in λCI protein were found to lower RecA*-mediated cleavage but not base-induced self-cleavage [369]. Single cysteine at this position in $UmuD_2$ is defective in both UV mutagenesis and RecA* cleavage [144]. Position is solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ moderately poorly with iodine but better with Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144]. UmuD′ E35K is defective in UV and MMS-induced mutagenesis, but it is not dominant negative to chromosomal $umuD^+$ [145]. Protein levels of UmuD′ E35K are at least as high as wild-type UmuD′ when expressed constitutively, and there does not appear to be a defect in *in vivo* formaldehyde cross-linking to $UmuD'_2$ [145]. This position may be involved in binding ClpXP [127].
36	Q	Single cysteine at this position in $UmuD_2$ is not defective in either UV mutagenesis or RecA* cleavage [144]. Position is moderately solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ poorly with iodine but moderately with Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144]. This position may be involved in binding ClpXP [127].
37	R	May be involved in RecA* interaction [143]. Mutations in the corresponding region in λCI protein were found to lower

		RecA*-mediated cleavage but not base-induced self-cleavage [369]. Single cysteine at this position in $UmuD_2$ is slightly defective in both UV mutagenesis and *in vitro* RecA* cleavage, but it does not appear affected in *in vivo* RecA* cleavage [144]. Position is not solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ robustly with both iodine and Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ very well by dialyzing out DTT [144]. UmuD′ R37C does not cross-link robustly to UmuD′ $R37C_2$, but UmuD R37C cross-links very robustly to UmuD $R37C_2$ [134]. R37 is more solvent accessible in UmuD′ than in UmuD [134]. UmuD R37C and UmuD′ R37C cross-linked together form a predominant heterodimer band and weak bands for the UmuD $R37C_2$ and UmuD′ $R37C_2$ homodimers [134]. Both spin-labeled UmuD R37C and UmuD′ R37C show spin-spin interactions as homodimer despite the different cross-linking activities of UmuD $R37C_2$ and UmuD′ $R37C_2$, suggesting that there are only modest structural differences between $UmuD_2$ and $UmuD'_2$ in this region [134]. UmuD′ R37A is deficient for proteolysis by ClpXP, but UmuD R37A can target wild-type UmuD′ for proteolysis [127]. Both UmuD R37A and UmuD′ R37A can form homodimers [127].
38	I	May be involved in RecA* interaction [143]. Mutations in the corresponding region in λCI protein were found to lower RecA*-mediated cleavage but not base-induced self-cleavage [369]. Single cysteine at this position in $UmuD_2$ is slightly defective in both UV mutagenesis and *in vitro* RecA* cleavage, but it does not appear affected in *in vivo* RecA* cleavage [144]. Position is not solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ robustly with both iodine and Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ very well by dialyzing out DTT [144]. A single-cysteine variant of UmuD at this position undergoes some RecA*-induced cleavage to UmuD′ *in vitro* even without reductant, and addition of reductant makes RecA*-induced cleavage slightly more efficient than wild-type UmuD (in contrast to results from [144]) [112]. UmuD′ I38C does not cross-link robustly to UmuD′ $I38C_2$, but UmuD I38C cross-links very robustly to UmuD $I38C_2$ [134]. I38 is more solvent accessible in UmuD′ than in UmuD [134]. UmuD I38C and UmuD′ I38C cross-linked together form a predominant heterodimer band and weak bands for the UmuD $I38C_2$ and UmuD′ $I38C_2$ homodimers [134].

39	D	Single cysteine at this position in $UmuD_2$ is slightly defective in both UV mutagenesis and RecA* cleavage (more defective in *in vitro* cleavage than *in vivo*) [144]. Position is highly solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ moderately with both iodine and Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ very well by dialyzing out DTT [144]. UmuD′ D39G is defective in UV and MMS-induced mutagenesis, but it does augment the activity of chromosomal *umuD*$^+$ [145]. Protein levels of UmuD′ E35K are very low compared to wild-type UmuD′ when expressed constitutively, but only in *lexA*$^+$ background, and there is a severe defect in *in vivo* formaldehyde cross-linking to $UmuD'_2$ [145]. Side chain may be important for dimerization to $UmuD'_2$ [145].
40	L	May be involved in RecA* interaction [143]. R variant in cleavage site mutant background (C24D/G25D) greatly reduces ability to cleave UmuD S60A over otherwise wild-type cleavage site mutant [149, 370]. R variant in cleavage site mutant background cannot cleave UmuD S60A L40R [370]. Mutations in the corresponding region in λCI protein were found to lower RecA*-mediated cleavage but not base-induced self-cleavage [369]. Single cysteine at this position in $UmuD_2$ is impaired in both UV mutagenesis and RecA* cleavage, but the RecA* cleavage defect is much more pronounced *in vitro* than *in vivo* [144]. Position is solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ poorly with iodine but moderately well with Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ very well by dialyzing out DTT [144]; also cross-links robustly as $UmuD_2$ with AIA [88]. UmuD′ L40R is severely defective in UV and MMS-induced mutagenesis, and it is dominant negative to chromosomal *umuD*$^+$ [145]. Protein levels of UmuD′ L40R as high as wild-type UmuD′ when expressed constitutively, but there does is a severe defect in *in vivo* formaldehyde cross-linking to $UmuD'_2$ [145]. Side chain may be important for dimerization to $UmuD'_2$ [145]. ClpXP degradation of $UmuD_2$ sometimes results in cleavage after this residue [127].
41	N	Single cysteine at this position in $UmuD_2$ is defective in both UV mutagenesis and RecA* cleavage (more defective in *in vitro* cleavage than *in vivo*) [144]. Position has moderately poor solvent accessibility; single cysteine at this position cross-links as $UmuD_2$ poorly with iodine but moderately with Cu phenanthroline [144]. Single-cysteine mutant cross-links

		as $UmuD_2$ moderately well by dialyzing out DTT [144]. Side chain may be important for dimerization to $UmuD'_2$ [145]. ClpXP degradation of $UmuD_2$ sometimes results in cleavage after this residue [127]. This position may be involved in binding ClpXP [127].
42	Q	Single cysteine at this position in $UmuD_2$ is slightly defective in UV mutagenesis and *in vitro* RecA* cleavage but not *in vivo* RecA* cleavage [144]. Position is solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ poorly with both iodine and Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ poorly by dialyzing out DTT [144]. ClpXP degradation of $UmuD_2$ sometimes results in cleavage after this residue [127]. This position may be involved in binding ClpXP [127].
43	L	This position may be involved in binding ClpXP [127].
44	L	Single cysteine at this position has defect in glutaraldehyde cross-linking to $UmuD_2$ but forms heterodimers with UmuD′ as efficiently as wild-type UmuD [143]. Single cysteine at this position has moderate solvent accessibility and cross-linking to $UmuD_2$ with iodine, BMH, or Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ does not depend on protein concentration [143]. Mutations in the corresponding region in λCI protein were found to lower RecA*-mediated cleavage but not base-induced self-cleavage [369]. Position is moderately solvent-accessible; single cysteine at this position cross-links as $UmuD_2$ poorly with iodine (again, discrepancy with reference [143] may be due to greater DTT concentration in experiments in [144]) but moderately with Cu phenanthroline [144]. Single-cysteine mutant cross-links as $UmuD_2$ moderately well by dialyzing out DTT [144]. A disulfide cross-linked single cysteine variant of $UmuD_2$ at this position undergoes some *in vitro* RecA*-mediated cleavage in the absence of reductant, but cleavage in the presence of reductant is as efficient as wild-type (but produces an extra intermediate band) [112]. A single-cysteine variant of $UmuD_2$ at this position cross-links poorly to RecA* [88, 112]. Side chain may be important for dimerization to $UmuD'_2$ [145]. UmuD′ L44C cross-links robustly to UmuD′ $L44C_2$, but UmuD L44C does not cross-link robustly to UmuD $L44C_2$ [134]. L44C is more solvent accessible in UmuD than in UmuD′ [134]. This position may be involved in binding ClpXP [127].

45	I	This position may be involved in binding ClpXP [127].
46	Q	This position may be involved in binding ClpXP [127].
47	H	Side chain may be important for dimerization to UmuD′$_2$ (by providing a hydrogen bond to stabilize a loop) [145]. This position may be involved in binding ClpXP [127].
48	P	This position may be involved in binding ClpXP [127].
49	S	UmuD′ S49N-A50T is defective in UV and MMSinduced mutagenesis, but it slightly augments mutagnesis activity of chromosomal *umuD*$^+$ [145]. Protein levels of UmuD′ S49N-A50T are at least as high as wild-type UmuD′ when expressed constitutively, and there does not appear to be a defect in *in vivo* formaldehyde cross-linking to UmuD′$_2$ [145]. This position may be involved in binding ClpXP [127].
50	A	Highly conserved among UmuD-like proteins (except *S. typhimurium* [371, 372] and RulA [373]. Mutations to either T or V disrupt homodimerization of UmuD′ by yeast two-hybrid (V variant disrupted slightly more) [350]. Steady-state levels are approximately equal to wild-type when expressed as UmuD′, although V mutation has slightly lowered levels [350]. T variant slightly and V variant greatly reduces both spontaneous and MMS-induced mutagenesis as UmuD′ [350]. S variant in S60A construct can be recognized for cleavage by *S. typhimurium* UmuD enzyme, whereas otherwise wild-type S60A construct cannot [149]. UmuD′ S49N-A50T is defective in UV and MMS-induced mutagenesis, but it slightly augments mutagnesis activity of chromosomal *umuD*$^+$ [145]. Protein levels of UmuD′ S49N-A50T are at least as high as wild-type UmuD′ when expressed constitutively, and there does not appear to be a defect in *in vivo* formaldehyde cross-linking to UmuD′$_2$ [145]. This position may be involved in binding ClpXP [127].
51	T	Conserved in 9 out of 10 UmuD-like proteins [350]. UmuD′ T51I has reduced steady-state levels, a pronounced decrease in MMS-induced and spontaneous mutagenesis, and a defect in self-association of UmuD′ by yeast two-hybrid [350]. UmuD T51I in cleavage site mutant construct (C24D/G25D) diminishes its ability to cleave otherwise wild-type cleavage

		site mutant UmuD; UmuD T51I cleavage site mutant cannot cleave UmuD T51I S60A [370]. UmuD′ T51I is defective in UV and MMS-induced mutagenesis, but it slightly augments mutagnesis activity of chromosomal *umuD*$^+$ [145]. Protein levels of UmuD′ T51I are at least as high as wild-type UmuD′ when expressed constitutively, but there is a defect in *in vivo* formaldehyde cross-linking to UmuD′$_2$ [145]. Side chain may be important for dimerization to UmuD′$_2$ (by providing a hydrogen bond to stabilize a loop) [145]. This position may be involved in binding ClpXP [127].
52	Y	UmuD′ Y52Q is probably deficient in UV-induced mutagenesis [145]. UmuD Y52W is deficient for UV-induced mutagenesis (Figure AII-4, p.188). UmuD C24A Y52C cross-links robustly to β [330].
53	F	
54	V	
55	K	
56	A	
57	S	Mutation to Cysteine in UmuD has a defect in mutagenesis and RecA* cleavage [143]. Single cysteine at this position has high solvent accessibility, low cross-linking as UmuD$_2$ with iodine or BMH, and moderate cross-linking with Cu pehnanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as UmuD$_2$ increases slightly with protein concentration [143]. Single-cysteine mutant cross-links as UmuD$_2$ poorly by dialyzing out DTT [144]. A single cysteine variant of UmuD at this position cross-links moderately to RecA*; also cross-links as UmuD$_2$ moderately with AIA [88]. UmuD′ S57C cross-links very robustly to UmuD′ S57C$_2$, but UmuD S57C cross-links very poorly to UmuD S57C$_2$ (this would not have been predicted by the structures of UmuD′) [134]. S57C is solvent accessible in both UmuD and UmuD′ [134]. UmuD S57C seems to have more contaminating UmuD′ than other UmuD derivatives, and when it is cross-linked as a homodimer, the only band visible corresponds to the size of a UmuD′D heterodimer (so it either runs fast on a gel or only cross-links to contaminating UmuD′ S57C) [134]. UmuD S57C and UmuD′ S57C cross-linked together form both a band that corresponds to the size of the heterodimer and one that

Figure AII-4

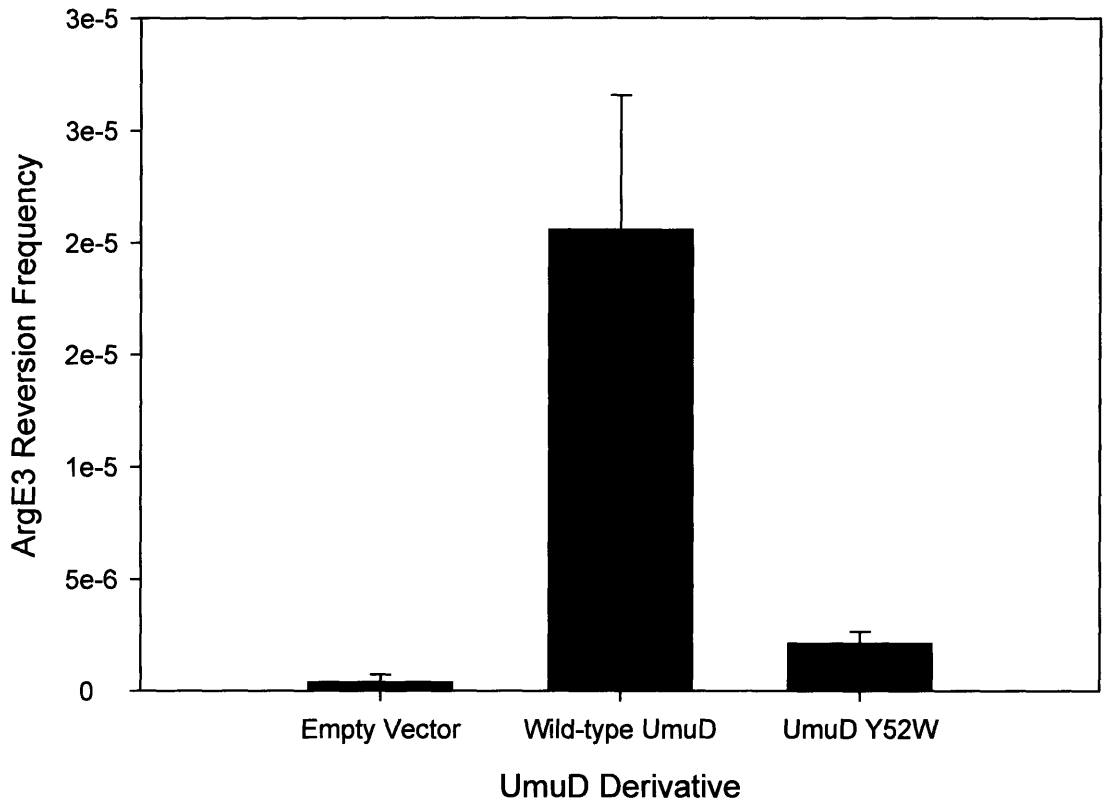


Figure Legend: ArgE3 reversion assay was performed as described at 25 J/m^2 [317], using bacterial strain GW8017 and plasmids pSE101 (empty vector), pSE115 (positive control), and pSE115 Y52W.

		appears to be UmuD′ $S57C_2$ [134]. Spin-labeled UmuD S57C and UmuD′ S57C both showed a lack of spin-spin interactions by EPR [134]. UmuD C24A S57C does not cross-link robustly to β [330].
58	G	
59	D	
60	S	Mutation to Cysteine in UmuD has a defect in mutagenesis and RecA* cleavage [143]. Single cysteine at this position has essentially normal extent of glutaraldehyde cross-linking to $UmuD_2$; forms heterodimers with UmuD′ as efficiently as wild-type UmuD; reacts well to iodoacetate (but since this is the active site nucleophile, that does not necessarily mean it is solvent accessible but merely reactive); cross-links as $UmuD_2$ poorly with iodine or BMH but cross-links moderately with Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ depends on protein concentration [143]. UmuD S60C and S60A are both impaired for UV mutagenesis, but UmuD′ S60A is much less impaired [103]. Single-cysteine mutant cross-links as $UmuD_2$ moderately poorly by dialyzing out DTT [144]. A single-cysteine variant of $UmuD_2$ at this position cross-links poorly to RecA* [88, 112]. Single cysteine derivative of UmuD at this position cross-links as $UmuD_2$ poorly with AIA [88]. UmuD C24A S60C does not cross-link robustly to β [330].
61	M	Conserved among most UmuD-like proteins, except P1 HumB protein [366]. Despite a reduced steady-state level when expressed as UmuD′, failure of this variant to self-associate by yeast two-hybrid, and a lowered spontaneous mutagenesis frequency under the same conditions, the I mutation does not have a lowered MMS-induced mutation frequency [350].
62	I	
63	D	
64	G	G64D mutation in UmuD′ has about half of wild-type UmuD′'s spontaneous mutagenesis frequency, but has at least wild-type MMS-induced mutagenesis (interestingly, the authors stated in the discussion that this mutation resulted in about half of wild-type MMS-induced mutagenesis, but this

		is contrary to the table of results. I am not sure which is the type-o) [350]. G64D in UmuD′ runs lower on SDS-PAGE than wild-type [350]. Mutations in this residue in UmuD′ cause less than a 2-fold reduction in UV mutagenesis [145].
65	G	R mutation nonmutable in UmuD [76, 121, 367]. R mutation has only slight defect in spontaneous and MMS-inducted mutagenesis when expressed as UmuD′; therefore defect in UmuD is probably due to lack of autocleavage [350]. Position is expected to be solvent-exposed [141]. Interestingly, yeast-two hybrid experiment shows a 4-fold increase in homodimerization signal with G65R in UmuD′ than with wild-type UmuD′ [350]. G65R mutation runs lower than wild-type on SDS-PAGE gel [350].
66	I	
67	S	Mutation to Cysteine in UmuD has a defect in mutagenesis, but much less of a defect in RecA* cleavage [143]. Single cysteine at this position has moderate solvent accessibility, low cross-linking as $UmuD_2$ with iodine, moderate cross-linking with BMH, and robust cross-linking with Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ does not depend on protein concentration, but it depends on pH (cross-linking is more efficient at pH 8.1 than pH 7.3) [143]. Single-cysteine mutant cross-links as $UmuD_2$ poorly by dialyzing out DTT [144]. A single cysteine variant of UmuD at this position cross-links moderately to RecA*; but it cross-links as $UmuD_2$ poorly with AIA [88]. UmuD′ S67C cross-links robustly to UmuD′ $S67C_2$ (this would not have been predicted by the structures of UmuD′), but UmuD S67C cross-links very poorly to UmuD $S67C_2$ [134]. S67C is solvent accessible in both UmuD and UmuD′ [134]. Spin-labeled UmuD S67C and UmuD′ S67C both showed spin-spin interactions by EPR [134].
68	D	
69	G	
70	D	
71	L	
72	L	

73	I	
74	V	
75	D	Highly conserved residue; UmuD′ D75A appears to significantly alter the structure of UmuD′ [145]. UmuD′ D75A is defective in UV and MMS-induced mutagenesis, but it slightly augments mutagnesis activity of chromosomal *umuD*$^+$ [145]. Protein levels of UmuD′ D75A are much less than wild-type UmuD′ when expressed constitutively, but only in a *lexA*$^+$ background, and there is a severe defect in both the extent and the protein size of *in vivo* formaldehyde cross-linking to UmuD′$_2$ [145].
76	S	
77	A	
78	I	
79	T	Mutations in this residue in UmuD′ cause less than a 2-fold reduction in UV mutagenesis [145].
80	A	
81	S	This and surrounding positions are not well conserved in UmuD-LexA family of proteins [143]. Mutation to Cysteine in UmuD results in moderate deficiency in both mutagenesis and RecA* cleavage [143]. Single cysteine at this position has essentially normal extent of glutaraldehyde cross-linking to $UmuD_2$ and forms heterodimers with UmuD′ as efficiently as wild-type UmuD [143]. Single cysteine at this position has moderate solvent accessibility, low cross-linking as $UmuD_2$ with iodine or BMH, but high cross-linking propensity with Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ does not depend on protein concentration [143]. Single-cysteine mutant cross-links as $UmuD_2$ poorly by dialyzing out DTT [144]. Single-cysteine variant of UmuD at this position cross-links robustly to RecA*; but cross-links as $UmuD_2$ poorly with AIA [88]. UmuD′ S81C cross-links rather robustly to UmuD′ S81C$_2$ (this would not have been predicted by the structures of UmuD′), but UmuD S81C cross-links poorly to UmuD S81C$_2$ [134]. S81C is more solvent accessible in UmuD than in UmuD′ [134]. Spin-

		labeled UmuD S81C and UmuD′ S81C both showed a lack of spin-spin interactions by EPR [134]. UmuD C24A S81C does not cross-link robustly to β [330].
82	H	Conserved in 7 out of 10 UmuD-like proteins (no reference listed in [350], but try 34, 36, and 38). H82Y mutation in UmuD′ results in mild defect in spontaneous and MMS-induced mutagenesis, but yeast two-hybrid suggests a loss of ability to self-associate [350]. UmuD′ H82Y is much more defective in UV mutagenesis than MMS mutagenesis [125].
83	G	
84	D	
85	I	This position may be involved in binding ClpXP [127].
86	V	This position may be involved in binding ClpXP [127].
87	I	This position may be involved in binding ClpXP [127].
88	A	This position may be involved in binding ClpXP [127].
89	A	Not particularly solvent accessible in UmuD; may play important role in protein architecture [143]. Mutation to Cysteine in UmuD results in moderate deficiency in both mutagenesis and RecA* cleavage [143]. Single cysteine at this position has low solvent accessibility and low cross-linking as $UmuD_2$ with iodine, BMH, or Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ depends on protein concentration [143]. UmuD A89C defective in cross-linking as heterodimer with UmuD′ (but data not shown) [143]. Single-cysteine mutant cross-links as $UmuD_2$ poorly by dialyzing out DTT [144]. This position may be involved in binding ClpXP [127].
90	V	This position may be involved in binding ClpXP [127].
91	D	UmuD D91A is soluble, slightly hypercleavable, proficient for RecA*-mediated autoproteolysis, and deficient in binding to DinB [77]. This position may be involved in binding ClpXP [127].
92	G	G92D is nonmutable in UmuD [76, 121, 374]. This is a particularly interesting result because both LexA and the 434 repressor have a D at this position in their wild-type forms

		[125, 367], and there is essential no defect in mutagenesis or dimerization for G92D in UmuD′ [350]. However, UmuD′ G92D runs higher than wild-type UmuD′ on SDS-PAGE [350]. This position may be involved in binding ClpXP [127].
93	E	Mutations in this residue in UmuD′ cause less than a 2-fold reduction in UV mutagenesis [145]. UmuD E93C has wild-type UV mutagenesis activity (Figure AII-2, p.178). This position may be involved in binding ClpXP [127].
94	F	Spin-labeled UmuD F94C shows spin-spin interactions by EPR with position C24 [134]. Spin-labeled UmuD C24A F94C shows no spin-spin interactions [134]. Position 94 appears buried and rigid in UmuD but solvent exposed and flexible in UmuD′ by EPR [134]. UmuD F94C cross-links as UmuD $F94C_2$ almost quantitatively by copper phenanthroline (First mentioned in [134]; data shown in Chapter 4). UmuD C24A F94C dicysteine derivatives with an additional cysteine on the N-terminal arms cross-link as dimers with iodine in the following order: robust cross-linking if the addition cysteine is V22C, less if S28C, even less if D20C, and hardly at all if F26C [134]. Interestingly, two dimers form, presumably one with two disulfide bonds and one with only one [134]. UmuD F94C is completely inactive for UV-induced mutagenesis *in vivo* (Figure AII-5, p.194). The proximity of position F94 is to the intermolecular C24, not the intramolecular one [134]. UmuD C24A F94C does not cross-link robustly to β [330]. This position may be involved in binding ClpXP [127].
95	T	Mutations in this residue in UmuD′ cause less than a 2-fold reduction in UV mutagenesis [145]. This position may be involved in binding ClpXP [127].
96	V	This position may be involved in binding ClpXP [127].
97	K	This position may be involved in binding ClpXP [127].
98	K	This position may be involved in binding ClpXP [127].
99	L	This position may be involved in binding ClpXP [127].
100	Q	Not particularly solvent accessible in UmuD; may play important role in protein architecture [143]. Mutation to Cysteine in UmuD results in moderate deficiency in both

Figure AII-5

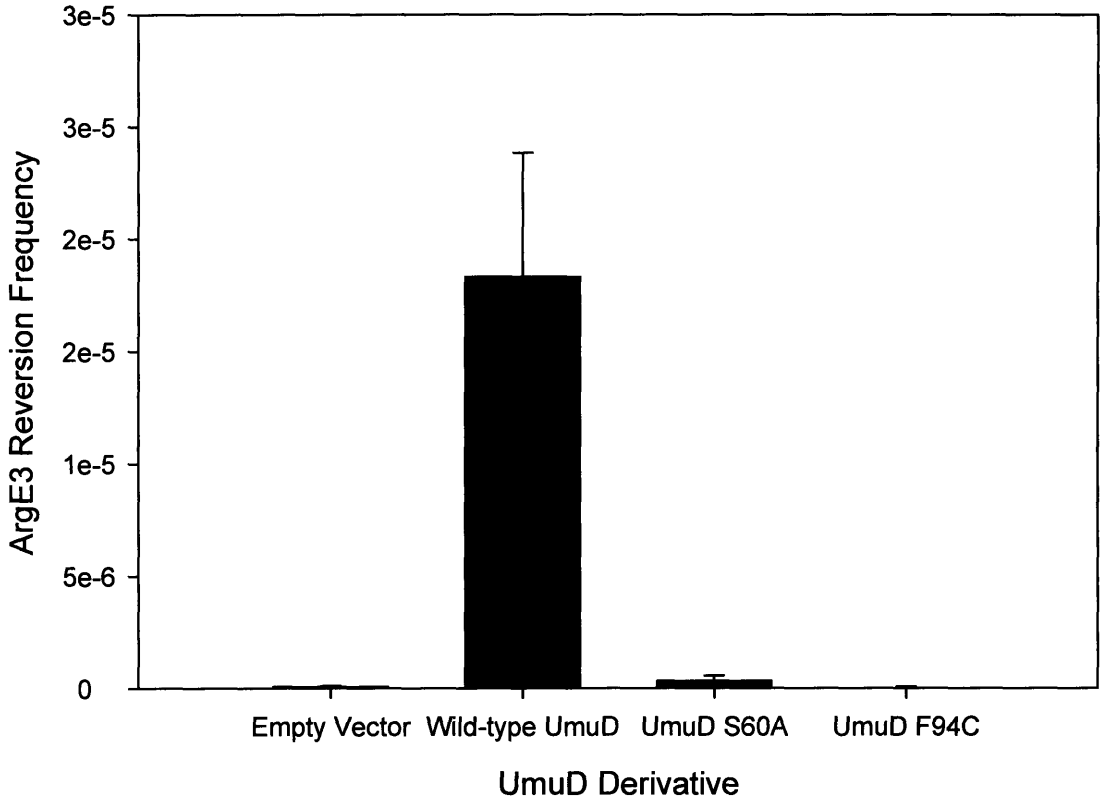


Figure Legend: ArgE3 reversion assay was performed as described at 25 J/m^2 [317], using bacterial strain GW8017 and plasmids pSE101 (empty vector), pSE115 (positive control), pSE115 S60A, and pSE115 F94C.

		mutagenesis and RecA* cleavage [143]. Single cysteine at this position has low solvent accessibility and low cross-linking as $UmuD_2$ with iodine, BMH, or Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ depends on protein concentration [143]. Single-cysteine mutant cross-links as $UmuD_2$ poorly by dialyzing out DTT [144]. This position may be involved in binding ClpXP [127].
101	L	Conserved in UmuD-like proteins but not in LexA-like ones [340]. UmuD L101G R102G is deficient in UV-induced mutagenesis, but UmuD′ L101G R102 G is not particularly deficient [340]. UmuD L101G R102G is deficient in *in vivo* and *in vitro* RecA*-mediated cleavage, although the effect is more dramatic *in vivo*, and it exacerbates cold sensitivity relative to wild-type UmuD [340]. However, UmuD L101G R102G interacts with RecA* just as efficiently as wild-type UmuD [340]. UmuD′ L101G R102G is defective in inhibiting recombination [340]. This position may be involved in binding ClpXP [127].
102	R	Relatively conserved in UmuD-like proteins but not in LexA-like ones [340]. UmuD L101G R102G is deficient in UV-induced mutagenesis, but UmuD′ L101G R102 G is not particularly deficient, suggesting a role in RecA* cleavage [340]. UmuD L101G R102G is deficient in *in vivo* and *in vitro* RecA*-mediated cleavage, although the effect is more dramatic *in vivo*, and it exacerbates cold sensitivity relative to wild-type UmuD [340]. However, UmuD L101G R102G interacts with RecA* just as efficiently as wild-type UmuD [340]. UmuD′ L101G R102G is defective in inhibiting recombination [340]. This position may be involved in binding ClpXP [127].
103	P	This position may be involved in binding ClpXP [127].
104	T	This position may be involved in binding ClpXP [127].
105	V	This position may be involved in binding ClpXP [127].
106	Q	This position may be involved in binding ClpXP [127].
107	L	This position may be involved in binding ClpXP [127].
108	I	This position may be involved in binding ClpXP [127].

109	P	Mutations in this residue in UmuD′ cause less than a 2-fold reduction in UV mutagenesis [145]. This position may be involved in binding ClpXP [127].
110	M	
111	N	
112	S	Mutation to Cysteine in UmuD results in slight defect in both mutagenesis and RecA* cleavage [143]. Single cysteine at this position has moderate solvent accessibility (but large error bars), low cross-linking as $UmuD_2$ with iodine, and moderate cross-linking with BMH or Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ depends slightly on protein concentration [143]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144]. A single cysteine variant of UmuD at this position cross-links moderately to RecA*; but cross-links as $UmuD_2$ poorly with AIA [88]. Both UmuD S112C and UmuD′ S112C cross-link poorly as homodimers [134]. S112C is solvent accessible in UmuD′; the error bars are very high for the same experiment UmuD [134]. Spin-labeled UmuD S112C and UmuD′ S112C both showed a lack of spin-spin interactions by EPR [134]. UmuD C24A S112C does not cross-link robustly to β [330].
113	A	
114	Y	
115	S	UmuD S115C has augmented UV mutagenesis activity (Figure AII-2, p.178).
116	P	
117	I	
118	T	
119	I	
120	S	
121	S	
122	E	

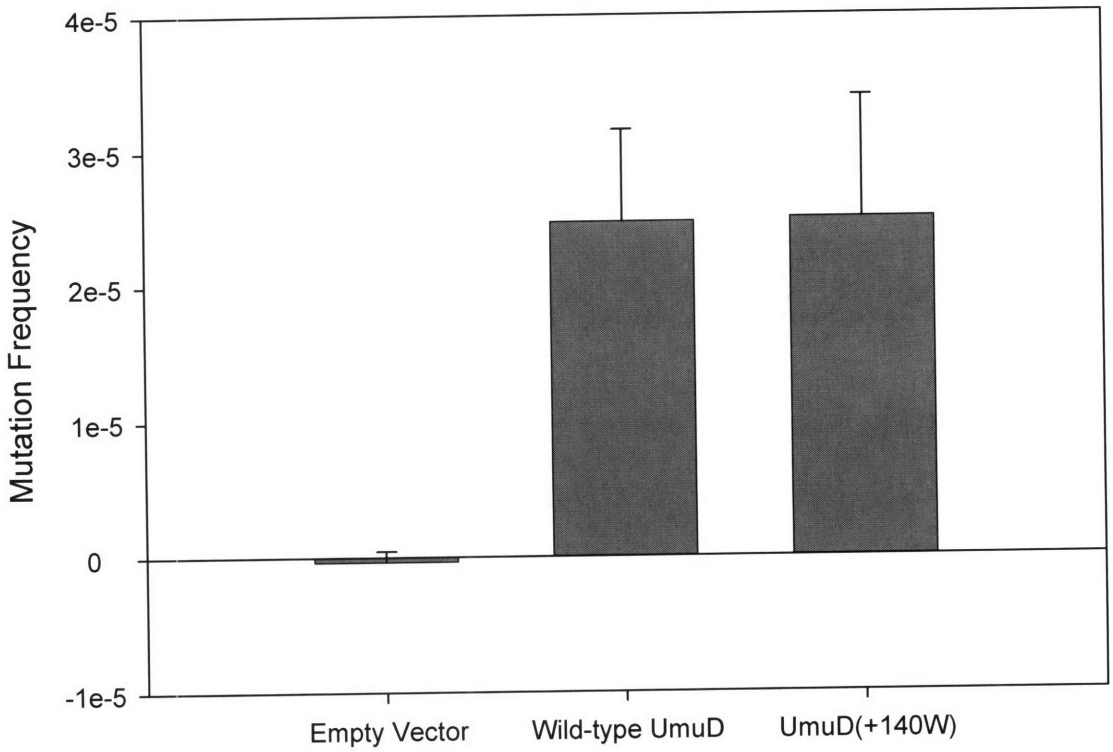
123	D	
124	T	
125	L	
126	D	Not particularly solvent accessible in UmuD; may play important role in protein architecture [143]. Mutation to Cysteine in UmuD results in very slight deficiency in mutagenesis but a large defect in RecA* cleavage [143]. D126C has slight defect in glutaraldehyde cross-linking to $UmuD_2$ and a great defect in heterodimer formation with UmuD′ (this inability to form the UmuD′-depleting heterodimer might account for its discrepancy in mutagenesis/cleavage phenotype) [143]. Single cysteine at this position has low solvent accessibility and low cross-linking as $UmuD_2$ with iodine, BMH, or Cu phenanthroline [143]. Extent of single-cysteine Cu phenanthroline cross-linking as $UmuD_2$ depends on protein concentration [143]. Single-cysteine mutant cross-links as $UmuD_2$ moderately by dialyzing out DTT [144]. UmuD C24A D126 cross-links somewhat robustly to β [330].
127	V	
128	F	
129	G	Position is conserved among homologous proteins [125, 145, 366]. G129D is non-cleavable in UmuD [121]. UmuD′ G129D is severely reduced for both spontaneous and MMS-induced mutagenesis, and it produced the strongest signal in a yeast two-hybrid assay (over 5-fold the signal of wild-type UmuD′) [350]. UmuD′ G129D is defective in UV and MMS-induced mutagenesis, and it is strongly dominant negative to chromosomal *umuD*$^+$ [145]. UmuD′ G129S is also severely reduced in spontaneous and MMS-mutagenesis, but less so than G129D [350]. G129S, in contrast to G129D, shows a defect in homodimer formation to $UmuD'_2$ [350]. UmuD′ G129D is not dominant negative in a *recA730* background [350]; possible reasons for this discrepancy are discussed in [145]. Protein levels of UmuD′ G129D are at least as high as wild-type UmuD′ when expressed constitutively, and there is a slight defect in formaldehyde cross-linking to $UmuD'_2$ [145].

130	V	Position is conserved in UmuD-like adaptor proteins but not in homologous repressors (no reference listed in [350], but try 28, 34, 36, 38 and 40). V130M in UmuD′ greatly reduces spontaneous and mildly reduces MMS-induced mutagenesis, but the *in vivo* degradation of this protein is much more efficient than wild-type, suggesting that these defects might be due more to the number of proteins in the cell than to any specific functional defect [350]. Similarly, a yeast two-hybrid of UmuD′ V130M shows a defect in self-association, but that might still be due to hyperdegradation of the protein [350].
131	V	
132	I	
133	H	UmuD′ H133C cross-links efficiently to UmuD′ V135C but not to V134C [134]. UmuD′ H133C cross-links somewhat to UmuD′ A137C, UmuD′ M138C, and UmuD′ R139C [134]. UmuD′ H133C cross-links inefficiently as a homodimer [134]. UmuD H133C cross-links efficiently to UmuD V135C and slightly less efficiently to UmuD M138C [134]. UmuD H133C cross-links inefficiently to UmuD V134C, UmuD A137C, and UmuD R139C [134]. UmuD H133C is expressed at wild-type levels but has only 17.4% of wild-type TLS activity; UmuD′ H133C is expressed almost as well as wild-type levels but has 49.4% TLS activity *in vivo* [134]. Neither UmuD H133C nor UmuD′ H133C cross-links robustly to homodimers with iodine [134]. UmuD C24A H133C does not cross-link robustly to β [330].
134	V	UmuD′ V134C cross-links inefficiently to UmuD′ H133C [134]. UmuD V134C cross-links inefficiently to UmuD H133C [134]. UmuD V134C is expressed at 70% of wild-type levels and has 94.6% of wild-type TLS activity; UmuD′ V134C is expressed almost as well as wild-type levels but has 169.6% TLS activity *in vivo* [134]. Neither UmuD V134C nor UmuD′ V134C cross-links robustly to homodimers with iodine [134]
135	V	UmuD V135D in cleavage site mutant construct (C24D/G25D) diminishes its ability to cleave otherwise wild-type cleavage site mutant UmuD; UmuD V135D cleavage site mutant cannot cleave UmuD S60A V135D [370]. UmuD′ V135C cross-links efficiently to UmuD′ H133C [134]. UmuD′ V135C cross-links less efficiently to UmuD′

		V134C, UmuD′ A137C, and UmuD′ M138C [134]. UmuD′ V135C cross-links inefficiently as a homodimer or to UmuD′ R139C [134]. UmuD V135C cross-links efficiently to UmuD H133C [134]. UmuD V135C cross-links much less efficiently as a homodimer or to UmuD 134C, UmuD A137C, UmuD M138C, or UmuD R139C [134]. UmuD V135C is expressed slightly more than wild-type levels but has only 12.1% of wild-type TLS activity; UmuD′ V135C is expressed 60% as well as wild-type levels but has 38.5% TLS activity *in vivo* [134]. Neither UmuD V135C nor UmuD′ V135C cross-links robustly to homodimers with iodine [134].
136	K	UmuD K136C is expressed at 80% of wild-type levels and has 78.5% of wild-type TLS activity; UmuD′ K136C is expressed at 40% of wild-type levels but has 195.3% TLS activity *in vivo* [134]. UmuD K136C easily precipitated *in vitro* (Graham Walker, personal communication). UmuD′Δ136-139 had only a slight destabilization in a *clp*+ strain but was significantly less stable than wild-type UmuD′ in a *clpX::kan* strain [323].
137	A	UmuD′ A137C cross-links somewhat to UmuD′ H133C [134]. UmuD A137C cross-links inefficiently to UmuD H133C [134]. UmuD A137C is expressed almost as well as wild-type but has only 47.3% of wild-type TLS activity; UmuD′ A137C is expressed 40% as well as wild-type levels but has 189.5% TLS activity *in vivo* [134]. Neither UmuD A137C nor UmuD′ A137C cross-links robustly to homodimers with iodine [134]. Spin-labeled UmuD A137C and UmuD′ A137C both show spin-spin interactions by EPR, suggesting that A137 is close to the dimer interface [134]. UmuD′Δ137-139 is less stable than wild-type UmuD′ in both a *clp*$^+$ and a *clpX::kan* strain [323].
138	M	UmuD′ M138C cross-links somewhat to UmuD′ H133C [134]. UmuD M138C cross-links relatively efficiently to UmuD H133C [134]. UmuD M138C is expressed at slightly above wild-type levels but has only 22.6% of wild-type TLS activity; UmuD′ M138C is expressed at 40% of wild-type levels but has 80.5% TLS activity *in vivo* [134]. Neither UmuD M138C nor UmuD′ M138C cross-links robustly to homodimers with iodine [134]. Spin-labeled UmuD M138C and UmuD′ M138C both show spin-spin interactions by EPR, suggesting that M138 is close to the dimer interface [134]. UmuD′Δ138-139 behaves like wild-type in a *clp*$^+$ strain but is

		more unstable than wild-type in a *clpX::kan* strain [323]. This residue appears to interfere with selenium-soaking into UmuD′ crystals; mutation to UmuD′ M138V or UmuD′ M138T allows soaking [342]. UmuD′ M138V and UmuD′ M138T have increased UV mutagenesis activity [125].
139	R	UmuD′ R139C cross-links somewhat to UmuD′ H133C [134]. UmuD R139C cross-links inefficiently to UmuD H133C [134]. UmuD R139C is expressed at 30% of wild-type levels and has 46.7% of wild-type TLS activity; UmuD′ R139C is expressed at 25% of wild-type levels but has 122% TLS activity *in vivo* [134]. Neither UmuD R139C nor UmuD′ R139C cross-links robustly to homodimers with iodine [134]. UmuD′Δ139 has no phenotype related to *in vivo* ClpXP proteolysis [323]. UmuD C24A R139C does not cross-link robustly to β [330].
140		W in this position of $UmuD_2$ has wild-type activity *in vivo* but a slight decrease in ClpXP-mediated proteolysis for UmuD′(+140W) (figure AII-6, p.201).

Figure AII-6
A



B

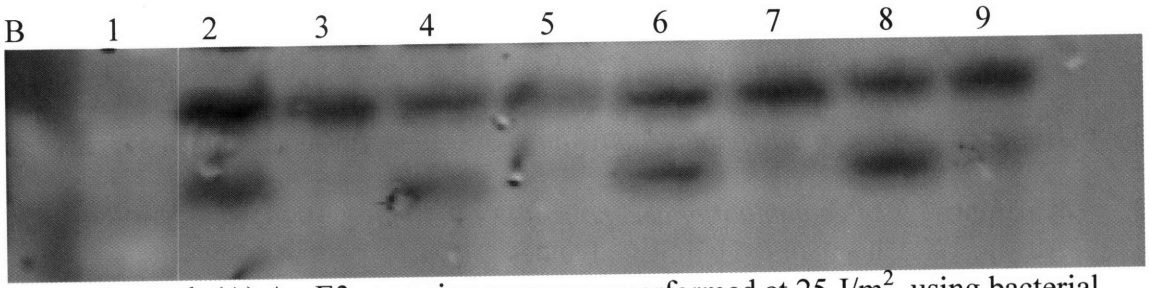


Figure Legend: (A) ArgE3 reversion assay was performed at 25 J/m^2, using bacterial strain GW8017 and plasmids pSE101 (empty vector), pSE115 (positive control), and pSE115(+140W). (B) ClpXP degradation assay of UmuD(+140W) and UmuD′(+140W). Upper bands are UmuD; lower bands are UmuD′. Lanes are: (1) molecular weight markers (2) wild-type UmuD and UmuD′, t= 0 (3) wild-type UmuD and UmuD′, t = 2 hr (4) UmuD(+140W) and wild-type UmuD′, t = 0 (5) UmuD(+140W) and wild-type UmuD′, t = 2 hr (6) wild-type UmuD and UmuD′(+140W), t = 0 (7) wild-type UmuD and UmuD′(+140W), t = 2 hr (8) UmuD(+140W) and UmuD′(+140W), t = 0 (9) UmuD(+140W) and UmuD′(+140W), t = 0. All procedures are described in [317].

REFERENCES

1. Mohana-Borges, R., et al., *LexA repressor forms stable dimers in solution. The role of specific dna in tightening protein-protein interactions.* J Biol Chem, 2000. **275**(7): p. 4708-12.
2. Alonso, L.G., et al., *High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners.* Biochemistry, 2002. **41**(33): p. 10510-8.
3. Friedberg, E.C., et al., *DNA repair and mutagenesis*. 2nd ed. 2006, Washington, D.C.: ASM Press. xxix, 1118 p.
4. Setlow, R.B., *The wavelengths in sunlight effective in producing skin cancer: a theoretical analysis.* Proc Natl Acad Sci U S A, 1974. **71**(9): p. 3363-6.
5. Sonntag, C.v., *The chemical basis of radiation biology*. 1987, London; New York: Taylor & Francis. 515 p.
6. Lawley, P.D., *Mutagens as carcinogens: development of current concepts.* Mutat Res, 1989. **213**(1): p. 3-25.
7. Adelman, R., R.L. Saul, and B.N. Ames, *Oxidative damage to DNA: relation to species metabolic rate and life span.* Proc Natl Acad Sci U S A, 1988. **85**(8): p. 2706-8.
8. Lindahl, T., *Instability and decay of the primary structure of DNA.* Nature, 1993. **362**(6422): p. 709-15.
9. Lindahl, T., *New class of enzymes acting on damaged DNA.* Nature, 1976. **259**(5538): p. 64-6.
10. Setlow, R.B. and W.L. Carrier, *The Disappearance Of Thymine Dimers From Dna: An Error-Correcting Mechanism.* Proc Natl Acad Sci U S A, 1964. **51**: p. 226-31.
11. Boyce, R.P. and P. Howard-Flanders, *Release Of Ultraviolet Light-Induced Thymine Dimers From Dna In E. Coli K-12.* Proc Natl Acad Sci U S A, 1964. **51**: p. 293-300.
12. Rupert, C.S., *Enzymatic photoreactivation: overview.* Basic Life Sci, 1975. **5A**: p. 73-87.
13. Kuzminov, A., *Recombinational repair of DNA damage*. 1996, New York Austin, TX: Chapman & Hall; R.G. Landes. 210 p.
14. Michel, B., et al., *Rescue of arrested replication forks by homologous recombination.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8181-8.
15. Ohmori, H., et al., *The Y-family of DNA polymerases.* Mol Cell, 2001. **8**(1): p. 7-8.
16. Friedberg, E.C. and V.L. Gerlach, *Novel DNA polymerases offer clues to the molecular basis of mutagenesis.* Cell, 1999. **98**(4): p. 413-6.
17. Rupp, W.D. and P. Howard-Flanders, *Discontinuities in the DNA synthesized in an excision-defective strain of Escherichia coli following ultraviolet irradiation.* J Mol Biol, 1968. **31**(2): p. 291-304.
18. Becherel, O.J., R.P. Fuchs, and J. Wagner, *Pivotal role of the beta-clamp in translesion DNA synthesis and mutagenesis in E. coli cells.* DNA Repair (Amst), 2002. **1**(9): p. 703-8.

19. Waters, L.S. and G.C. Walker, *The critical mutagenic translesion DNA polymerase Rev1 is highly expressed during G(2)/M phase rather than S phase.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 8971-6.
20. Langston, L.D. and M. O'Donnell, *DNA replication: keep moving and don't mind the gap.* Mol Cell, 2006. **23**(2): p. 155-60.
21. Barry, E.R. and S.D. Bell, *DNA replication in the archaea.* Microbiol Mol Biol Rev, 2006. **70**(4): p. 876-87.
22. Pavlov, Y.I., P.V. Shcherbakova, and I.B. Rogozin, *Roles of DNA polymerases in replication, repair, and recombination in eukaryotes.* Int Rev Cytol, 2006. **255**: p. 41-132.
23. Sutton, M.D. and G.C. Walker, *Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8342-9.
24. Kornberg, A. and T.A. Baker, *DNA replication.* 2nd ed. 1992, New York: W.H. Freeman. xiv, 931 p.
25. Sweasy, J.B., J.M. Lauper, and K.A. Eckert, *DNA polymerases and human diseases.* Radiat Res, 2006. **166**(5): p. 693-714.
26. Wang, T.S., S.Z. Hu, and D. Korn, *DNA primase from KB cells. Characterization of a primase activity tightly associated with immunoaffinity-purified DNA polymerase-alpha.* J Biol Chem, 1984. **259**(3): p. 1854-65.
27. Frick, D.N. and C.C. Richardson, *DNA primases.* Annu Rev Biochem, 2001. **70**: p. 39-80.
28. Waga, S. and B. Stillman, *Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro.* Nature, 1994. **369**(6477): p. 207-12.
29. Waga, S., G. Bauer, and B. Stillman, *Reconstitution of complete SV40 DNA replication with purified replication factors.* J Biol Chem, 1994. **269**(14): p. 10923-34.
30. Johnson, A. and M. O'Donnell, *Cellular DNA replicases: components and dynamics at the replication fork.* Annu Rev Biochem, 2005. **74**: p. 283-315.
31. Fuss, J. and S. Linn, *Human DNA polymerase epsilon colocalizes with proliferating cell nuclear antigen and DNA replication late, but not early, in S phase.* J Biol Chem, 2002. **277**(10): p. 8658-66.
32. Rangarajan, S., R. Woodgate, and M.F. Goodman, *A phenotype for enigmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated Escherichia coli.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 9224-9.
33. Alba, M., *Replicative DNA polymerases.* Genome Biol, 2001. **2**(1): p. REVIEWS3002.
34. Lamers, M.H., et al., *Crystal structure of the catalytic alpha subunit of E. coli replicative DNA polymerase III.* Cell, 2006. **126**(5): p. 881-92.
35. Ramadan, K., I. Shevelev, and U. Hubscher, *The DNA-polymerase-X family: controllers of DNA quality?* Nat Rev Mol Cell Biol, 2004. **5**(12): p. 1038-43.
36. Aravind, L. and E.V. Koonin, *DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history.* Nucleic Acids Res, 1999. **27**(7): p. 1609-18.

37. Napolitano, R., et al., *All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis.* Embo J, 2000. **19**(22): p. 6259-65.
38. Gibbs, P.E., et al., *A human homolog of the Saccharomyces cerevisiae REV3 gene, which encodes the catalytic subunit of DNA polymerase zeta.* Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6876-80.
39. Morrison, A., et al., *REV3, a Saccharomyces cerevisiae gene whose function is required for induced mutagenesis, is predicted to encode a nonessential DNA polymerase.* J Bacteriol, 1989. **171**(10): p. 5659-67.
40. Suzuki, N., et al., *Translesion synthesis by human DNA polymerase kappa on a DNA template containing a single stereoisomer of dG-(+)- or dG-(-)-anti-N(2)-BPDE (7,8-dihydroxy-anti-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene).* Biochemistry, 2002. **41**(19): p. 6100-6.
41. Avkin, S., et al., *Quantitative analysis of translesion DNA synthesis across a benzo[a]pyrene-guanine adduct in mammalian cells: the role of DNA polymerase kappa.* J Biol Chem, 2004. **279**(51): p. 53298-305.
42. McDonald, J.P., et al., *DNA polymerase iota and related rad30-like enzymes.* Philos Trans R Soc Lond B Biol Sci, 2001. **356**(1405): p. 53-60.
43. Johnson, R.E., S. Prakash, and L. Prakash, *Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta.* Science, 1999. **283**(5404): p. 1001-4.
44. Washington, M.T., et al., *Accuracy of thymine-thymine dimer bypass by Saccharomyces cerevisiae DNA polymerase eta.* Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3094-9.
45. Ling, H., et al., *Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication.* Cell, 2001. **107**(1): p. 91-102.
46. Trincao, J., et al., *Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis.* Mol Cell, 2001. **8**(2): p. 417-26.
47. Johnson, S.J., J.S. Taylor, and L.S. Beese, *Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3895-900.
48. Patel, P.H., et al., *Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection.* J Mol Biol, 2001. **308**(5): p. 823-37.
49. Yang, W., *Damage repair DNA polymerases Y.* Curr Opin Struct Biol, 2003. **13**(1): p. 23-30.
50. Yang, W., *Portraits of a Y-family DNA polymerase.* FEBS Lett, 2005. **579**(4): p. 868-72.
51. Ling, H., et al., *Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase.* Proc Natl Acad Sci U S A, 2004. **101**(8): p. 2265-9.
52. Fiala, K.A. and Z. Suo, *Mechanism of DNA polymerization catalyzed by Sulfolobus solfataricus P2 DNA polymerase IV.* Biochemistry, 2004. **43**(7): p. 2116-25.
53. Jarosz, D.F., et al., *A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates.* Nature, 2006. **439**(7073): p. 225-8.

54. Washington, M.T., L. Prakash, and S. Prakash, *Yeast DNA polymerase eta utilizes an induced-fit mechanism of nucleotide incorporation.* Cell, 2001. **107**(7): p. 917-27.
55. Tissier, A., et al., *poliota, a remarkably error-prone human DNA polymerase.* Genes Dev, 2000. **14**(13): p. 1642-50.
56. Boudsocq, F., et al., *Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta.* Nucleic Acids Res, 2001. **29**(22): p. 4607-16.
57. Washington, M.T., et al., *Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase eta.* J Biol Chem, 1999. **274**(52): p. 36835-8.
58. Goodman, M.F., *Error-prone repair DNA polymerases in prokaryotes and eukaryotes.* Annu Rev Biochem, 2002. **71**: p. 17-50.
59. Kunkel, T.A., *DNA replication fidelity.* J Biol Chem, 2004. **279**(17): p. 16895-8.
60. Prakash, S., R.E. Johnson, and L. Prakash, *Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function.* Annu Rev Biochem, 2005. **74**: p. 317-53.
61. McElhinny, S.A., Y.I. Pavlov, and T.A. Kunkel, *Evidence for extrinsic exonucleolytic proofreading.* Cell Cycle, 2006. **5**(9): p. 958-62.
62. Rattray, A.J. and J.N. Strathern, *Error-prone DNA polymerases: when making a mistake is the only way to get ahead.* Annu Rev Genet, 2003. **37**: p. 31-66.
63. Sutton, M.D., et al., *The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance.* Annu Rev Genet, 2000. **34**: p. 479-497.
64. Friedberg, E.C., P.L. Fischhaber, and C. Kisker, *Error-prone DNA polymerases: novel structures and the benefits of infidelity.* Cell, 2001. **107**(1): p. 9-12.
65. Miller, C., et al., *SOS response induction by beta-lactams and bacterial defense against antibiotic lethality.* Science, 2004. **305**(5690): p. 1629-31.
66. Radman, M., *SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis.* Basic Life Sci, 1975. **5A**: p. 355-67.
67. Khil, P.P. and R.D. Camerini-Otero, *Over 1000 genes are involved in the DNA damage response of Escherichia coli.* Mol Microbiol, 2002. **44**(1): p. 89-105.
68. Courcelle, J., et al., *Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli.* Genetics, 2001. **158**(1): p. 41-64.
69. Walker, G.C., *Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli.* Microbiol Rev, 1984. **48**(1): p. 60-93.
70. Kenyon, C.J. and G.C. Walker, *Expression of the E. coli uvrA gene is inducible.* Nature, 1981. **289**(5800): p. 808-10.
71. Kenyon, C.J. and G.C. Walker, *DNA-damaging agents stimulate gene expression at specific loci in Escherichia coli.* Proc Natl Acad Sci U S A, 1980. **77**(5): p. 2819-23.
72. Fernandez De Henestrosa, A.R., et al., *Identification of additional genes belonging to the LexA regulon in Escherichia coli.* Mol Microbiol, 2000. **35**(6): p. 1560-72.
73. Wagner, J., et al., *The dinB gene encodes a novel E. coli DNA polymerase, DNA pol IV, involved in mutagenesis.* Mol Cell, 1999. **4**(2): p. 281-6.
74. Ciesla, Z., *Plasmid pKM101-mediated mutagenesis in Escherichia coli is inducible.* Mol Gen Genet, 1982. **186**(2): p. 298-300.

75. Tang, M., et al., *UmuD'(2)C is an error-prone DNA polymerase, Escherichia coli pol V.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 8919-24.
76. Kato, T. and Y. Shinoura, *Isolation and characterization of mutants of Escherichia coli deficient in induction of mutations by ultraviolet light.* Mol Gen Genet, 1977. **156**(2): p. 121-31.
77. Jarosz, D.F., Godoy, V.G., Simon, S.M., Delaney, J.C., Essigman, J.E., Walker, G.C., *Substrate Specificity and Novel Regulation of DinB Activity in DNA Damage Tolerance.* Submitted.
78. Bates, H., et al., *Spontaneous and UV-induced mutations in Escherichia coli K-12 strains with altered or absent DNA polymerase I.* J Bacteriol, 1989. **171**(5): p. 2480-4.
79. Gudas, L.J. and A.B. Pardee, *Model for regulation of Escherichia coli DNA repair functions.* Proc Natl Acad Sci U S A, 1975. **72**(6): p. 2330-4.
80. Rehrauer, W.M., et al., *Interaction of Escherichia coli RecA protein with LexA repressor. I. LexA repressor cleavage is competitive with binding of a secondary DNA molecule.* J Biol Chem, 1996. **271**(39): p. 23865-73.
81. Little, J.W., *Autodigestion of lexA and phage lambda repressors.* Proc Natl Acad Sci U S A, 1984. **81**(5): p. 1375-9.
82. Ptashne, M., *A genetic switch: phage lambda revisited.* 3rd ed. 2004, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. xiv, 154 p.
83. Little, J.W., *Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease.* Biochimie, 1991. **73**(4): p. 411-21.
84. Slilaty, S.N. and J.W. Little, *Lysine-156 and serine-119 are required for LexA repressor cleavage: a possible mechanism.* Proc Natl Acad Sci U S A, 1987. **84**(12): p. 3987-91.
85. Little, J.W., et al., *Cleavage of the Escherichia coli lexA protein by the recA protease.* Proc Natl Acad Sci U S A, 1980. **77**(6): p. 3225-9.
86. Little, J.W. and D.W. Mount, *The SOS regulatory system of Escherichia coli.* Cell, 1982. **29**(1): p. 11-22.
87. Neher, S.B., et al., *Latent ClpX-recognition signals ensure LexA destruction after DNA damage.* Genes Dev, 2003. **17**(9): p. 1084-9.
88. Lee, M.H. and G.C. Walker, *Interactions of Escherichia coli UmuD with activated RecA analyzed by cross-linking UmuD monocysteine derivatives.* J Bacteriol, 1996. **178**(24): p. 7285-94.
89. van Dijl, J.M., et al., *Identification of the potential active site of the signal peptidase SipS of Bacillus subtilis. Structural and functional similarities with LexA-like proteases.* J Biol Chem, 1995. **270**(8): p. 3611-8.
90. Paetzel, M. and N.C. Strynadka, *Common protein architecture and binding sites in proteases utilizing a Ser/Lys dyad mechanism.* Protein Sci, 1999. **8**(11): p. 2533-6.
91. Perry, K.L., et al., *umuDC and mucAB operons whose products are required for UV light- and chemical-induced mutagenesis: UmuD, MucA, and LexA proteins share homology.* Proc Natl Acad Sci U S A, 1985. **82**(13): p. 4331-5.
92. Little, J.W., *LexA cleavage and other self-processing reactions.* J Bacteriol, 1993. **175**(16): p. 4943-50.

93. McDonald, J.P., et al., *Intermolecular cleavage by UmuD-like mutagenesis proteins.* Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1478-83.
94. McCann, J., et al., *Detection of carcinogens as mutagens: bacterial tester strains with R factor plasmids.* Proc Natl Acad Sci U S A, 1975. **72**(3): p. 979-83.
95. Kitagawa, Y., et al., *Structural analysis of the umu operon required for inducible mutagenesis in Escherichia coli.* Proc Natl Acad Sci U S A, 1985. **82**(13): p. 4336-40.
96. Woodgate, R. and D.G. Ennis, *Levels of chromosomally encoded Umu proteins and requirements for in vivo UmuD cleavage.* Mol Gen Genet, 1991. **229**(1): p. 10-6.
97. Opperman, T., et al., *A model for a umuDC-dependent prokaryotic DNA damage checkpoint.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 9218-23.
98. Marsh, L. and G.C. Walker, *Cold sensitivity induced by overproduction of UmuDC in Escherichia coli.* J Bacteriol, 1985. **162**(1): p. 155-61.
99. Sutton, M.D., et al., *Genetic interactions between the Escherichia coli umuDC gene products and the beta processivity clamp of the replicative DNA polymerase.* J Bacteriol, 2001. **183**(9): p. 2897-909.
100. Sutton, M.D. and G.C. Walker, *umuDC-mediated cold sensitivity is a manifestation of functions of the UmuD(2)C complex involved in a DNA damage checkpoint control.* J Bacteriol, 2001. **183**(4): p. 1215-24.
101. Shinagawa, H., et al., *RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis.* Proc Natl Acad Sci U S A, 1988. **85**(6): p. 1806-10.
102. Burckhardt, S.E., et al., *UmuD mutagenesis protein of Escherichia coli: overproduction, purification, and cleavage by RecA.* Proc Natl Acad Sci U S A, 1988. **85**(6): p. 1811-5.
103. Nohmi, T., et al., *RecA-mediated cleavage activates UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation.* Proc Natl Acad Sci U S A, 1988. **85**(6): p. 1816-20.
104. Fijalkowska, I.J., R.L. Dunn, and R.M. Schaaper, *Genetic requirements and mutational specificity of the Escherichia coli SOS mutator activity.* J Bacteriol, 1997. **179**(23): p. 7435-45.
105. Caillet-Fauquet, P. and G. Maenhaut-Michel, *Nature of the SOS mutator activity: genetic characterization of untargeted mutagenesis in Escherichia coli.* Mol Gen Genet, 1988. **213**(2-3): p. 491-8.
106. Ichikawa-Ryo, H. and S. Kondo, *Indirect mutagenesis in phage lambda by ultraviolet preirradiation of host bacteria.* J Mol Biol, 1975. **97**(1): p. 77-92.
107. Rajagopalan, M., et al., *Activity of the purified mutagenesis proteins UmuC, UmuD', and RecA in replicative bypass of an abasic DNA lesion by DNA polymerase III.* Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10777-81.
108. Reuven, N.B., et al., *The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication.* J Biol Chem, 1999. **274**(45): p. 31763-6.
109. Tang, M., et al., *Biochemical basis of SOS-induced mutagenesis in Escherichia coli: reconstitution of in vitro lesion bypass dependent on the UmuD'2C mutagenic complex and RecA protein.* Proc Natl Acad Sci U S A, 1998. **95**(17): p. 9755-60.

110. Frank, E.G., et al., *Targeting of the UmuD, UmuD', and MucA' mutagenesis proteins to DNA by RecA protein.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 8169-73.
111. Schlacher, K., et al., *RecA acts in trans to allow replication of damaged DNA by DNA polymerase V.* Nature, 2006. **442**(7105): p. 883-7.
112. Lee, M.H., A. Guzzo, and G.C. Walker, *Inhibition of RecA-mediated cleavage in covalent dimers of UmuD.* J Bacteriol, 1996. **178**(24): p. 7304-7.
113. Granger-Schnarr, M., P. Oertel, and M. Schnarr, *A mutant LexA repressor harboring a cleavage motif cysteine-glycine remains inducible.* FEBS Lett, 1988. **231**(2): p. 437-9.
114. Shiba, T., et al., *Proteolytic processing of MucA protein in SOS mutagenesis: both processed and unprocessed MucA may be active in the mutagenesis.* Mol Gen Genet, 1990. **224**(2): p. 169-76.
115. Koudelka, A.P., L.A. Hufnagel, and G.B. Koudelka, *Purification and characterization of the repressor of the shiga toxin-encoding bacteriophage 933W: DNA binding, gene regulation, and autocleavage.* J Bacteriol, 2004. **186**(22): p. 7659-69.
116. Yasuda, T., et al., *Inhibition of Escherichia coli RecA coprotease activities by DinI.* Embo J, 1998. **17**(11): p. 3207-16.
117. Yasuda, T., et al., *Physical interactions between DinI and RecA nucleoprotein filament for the regulation of SOS mutagenesis.* Embo J, 2001. **20**(5): p. 1192-202.
118. Yasuda, T., T. Nagata, and H. Ohmori, *Multicopy suppressors of the cold-sensitive phenotype of the pcsA68 (dinD68) mutation in Escherichia coli.* J Bacteriol, 1996. **178**(13): p. 3854-9.
119. Lodwick, D., D. Owen, and P. Strike, *DNA sequence analysis of the imp UV protection and mutation operon of the plasmid TP110: identification of a third gene.* Nucleic Acids Res, 1990. **18**(17): p. 5045-50.
120. Woodgate, R., et al., *UmuC mutagenesis protein of Escherichia coli: purification and interaction with UmuD and UmuD'.* Proc Natl Acad Sci U S A, 1989. **86**(19): p. 7301-5.
121. Battista, J.R., et al., *Dominant negative umuD mutations decreasing RecA-mediated cleavage suggest roles for intact UmuD in modulation of SOS mutagenesis.* Proc Natl Acad Sci U S A, 1990. **87**(18): p. 7190-4.
122. Kim, B. and J.W. Little, *Dimerization of a specific DNA-binding protein on the DNA.* Science, 1992. **255**(5041): p. 203-6.
123. Schnarr, M., et al., *Large-scale purification, oligomerization equilibria, and specific interaction of the LexA repressor of Escherichia coli.* Biochemistry, 1985. **24**(11): p. 2812-8.
124. Pabo, C.O., et al., *The lambda repressor contains two domains.* Proc Natl Acad Sci U S A, 1979. **76**(4): p. 1608-12.
125. Peat, T.S., et al., *The UmuD' protein filament and its potential role in damage induced mutagenesis.* Structure, 1996. **4**(12): p. 1401-12.
126. Frank, E.G., et al., *Regulation of SOS mutagenesis by proteolysis.* Proc Natl Acad Sci U S A, 1996. **93**(19): p. 10291-6.

127. Neher, S.B., R.T. Sauer, and T.A. Baker, *Distinct peptide signals in the UmuD and UmuD' subunits of UmuD/D' mediate tethering and substrate processing by the ClpXP protease.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13219-24.
128. Shen, X., R. Woodgate, and M.F. Goodman, *Escherichia coli DNA polymerase V subunit exchange: a post-SOS mechanism to curtail error-prone DNA synthesis.* J Biol Chem, 2003. **278**(52): p. 52546-50.
129. Kai, M. and T.S. Wang, *Checkpoint activation regulates mutagenic translesion synthesis.* Genes Dev, 2003. **17**(1): p. 64-76.
130. Ogi, T., et al., *Mutation enhancement by DINB1, a mammalian homologue of the Escherichia coli mutagenesis protein dinB.* Genes Cells, 1999. **4**(11): p. 607-18.
131. Sutton, M.D., T. Opperman, and G.C. Walker, *The Escherichia coli SOS mutagenesis proteins UmuD and UmuD' interact physically with the replicative DNA polymerase.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12373-8.
132. Bridges, B.A. and R.P. Mottershead, *Mutagenic DNA repair in Escherichia coli. III. Requirement for a function of DNA polymerase III in ultraviolet-light mutagenesis.* Mol Gen Genet, 1976. **144**(1): p. 53-8.
133. Indiani, C., et al., *A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously.* Mol Cell, 2005. **19**(6): p. 805-15.
134. Sutton, M.D., et al., *A model for the structure of the Escherichia coli SOS-regulated UmuD2 protein.* DNA Repair (Amst), 2002. **1**(1): p. 77-93.
135. Arifuzzaman, M., et al., *Large-scale identification of protein-protein interaction of Escherichia coli K-12.* Genome Res, 2006. **16**(5): p. 686-91.
136. Kim, S.R., et al., *Roles of chromosomal and episomal dinB genes encoding DNA pol IV in targeted and untargeted mutagenesis in Escherichia coli.* Mol Genet Genomics, 2001. **266**(2): p. 207-15.
137. Giovannoni, S.J., et al., *Genome streamlining in a cosmopolitan oceanic bacterium.* Science, 2005. **309**(5738): p. 1242-5.
138. O'Grady, P.I., et al., *Intrinsic polymerase activities of UmuD'(2)C and MucA'(2)B are responsible for their different mutagenic properties during bypass of a T-T cis-syn cyclobutane dimer.* J Bacteriol, 2000. **182**(8): p. 2285-91.
139. Shen, X., R. Woodgate, and M.F. Goodman, *Lyase activities intrinsic to Escherichia coli polymerases IV and V.* DNA Repair (Amst), 2005. **4**(12): p. 1368-73.
140. Friedman, N., et al., *Precise temporal modulation in the response of the SOS DNA repair network in individual bacteria.* PLoS Biol, 2005. **3**(7): p. e238.
141. Peat, T.S., et al., *Structure of the UmuD' protein and its regulation in response to DNA damage.* Nature, 1996. **380**(6576): p. 727-30.
142. Ferentz, A.E., G.C. Walker, and G. Wagner, *Converting a DNA damage checkpoint effector (UmuD2C) into a lesion bypass polymerase (UmuD'2C).* Embo J, 2001. **20**(15): p. 4287-98.
143. Lee, M.H., T. Ohta, and G.C. Walker, *A monocysteine approach for probing the structure and interactions of the UmuD protein.* J Bacteriol, 1994. **176**(16): p. 4825-37.
144. Guzzo, A., et al., *Analysis of the region between amino acids 30 and 42 of intact UmuD by a monocysteine approach.* J Bacteriol, 1996. **178**(24): p. 7295-303.

145. Ohta, T., et al., *Mutations affecting the ability of the Escherichia coli UmuD' protein to participate in SOS mutagenesis.* J Bacteriol, 1999. **181**(1): p. 177-85.
146. Luo, Y., et al., *Crystal structure of LexA: a conformational switch for regulation of self-cleavage.* Cell, 2001. **106**(5): p. 585-94.
147. Beuning, P.J., et al., *A non-cleavable UmuD variant that acts as a UmuD' mimic.* J Biol Chem, 2006. **281**(14): p. 9633-40.
148. McDonald, J.P., et al., *Regulation of UmuD cleavage: role of the amino-terminal tail.* J Mol Biol, 1998. **282**(4): p. 721-30.
149. McDonald, J.P., et al., *Intermolecular cleavage by UmuD-like enzymes: identification of residues required for cleavage and substrate specificity.* J Mol Biol, 1999. **285**(5): p. 2199-209.
150. Slilaty, S.N., J.A. Rupley, and J.W. Little, *Intramolecular cleavage of LexA and phage lambda repressors: dependence of kinetics on repressor concentration, pH, temperature, and solvent.* Biochemistry, 1986. **25**(22): p. 6866-75.
151. Kim, B. and J.W. Little, *LexA and lambda Cl repressors as enzymes: specific cleavage in an intermolecular reaction.* Cell, 1993. **73**(6): p. 1165-73.
152. Hughes, R.E., et al., *Protein-protein interactions directing resolvase site-specific recombination: a structure-function analysis.* Embo J, 1993. **12**(4): p. 1447-58.
153. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling.* J Mol Recognit, 2005. **18**(5): p. 343-84.
154. Holt, C.a.S., L, *Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the αS1-, β- and κ-caseins.* J Chem Soc Faraday Trans, 1993. **29**: p. 2683-2692.
155. Boyer, P.D., *The Enzymes*. 2d ed. 1959, New York: Academic Press. 8 v.
156. Pullen, R.A., et al., *The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulins, glucagon and human placental lactogen.* Mol Cell Biochem, 1975. **8**(1): p. 5-20.
157. Cary, P.D., T. Moss, and E.M. Bradbury, *High-resolution proton-magnetic-resonance studies of chromatin core particles.* Eur J Biochem, 1978. **89**(2): p. 475-82.
158. Schweers, O., et al., *Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure.* J Biol Chem, 1994. **269**(39): p. 24290-7.
159. Weinreb, P.H., et al., *NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded.* Biochemistry, 1996. **35**(43): p. 13709-15.
160. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* J Mol Biol, 1999. **293**(2): p. 321-31.
161. Romero, P., Z. Obradovic, and A.K. Dunker, *Natively disordered proteins: functions and predictions.* Appl Bioinformatics, 2004. **3**(2-3): p. 105-13.
162. Sickmeier, M., et al., *DisProt: the Database of Disordered Proteins.* Nucleic Acids Res, 2007. **35**(Database issue): p. D786-93.
163. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
164. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins.* Curr Opin Struct Biol, 2002. **12**(1): p. 54-60.

165. Boesch, C., et al., *1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution.* Eur J Biochem, 1978. **91**(1): p. 209-14.
166. Armstrong, C.M. and F. Bezanilla, *Inactivation of the sodium channel. II. Gating current experiments.* J Gen Physiol, 1977. **70**(5): p. 567-90.
167. Bryant, J.E., et al., *Protein dynamics in living cells.* Biochemistry, 2005. **44**(26): p. 9275-9.
168. Liu, J., et al., *Intrinsic disorder in transcription factors.* Biochemistry, 2006. **45**(22): p. 6873-88.
169. Mitchell, P.J. and R. Tjian, *Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.* Science, 1989. **245**(4916): p. 371-8.
170. Iakoucheva, L.M., et al., *Intrinsic disorder in cell-signaling and cancer-associated proteins.* J Mol Biol, 2002. **323**(3): p. 573-84.
171. Bustos, D.M. and A.A. Iglesias, *Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins.* Proteins, 2006. **63**(1): p. 35-42.
172. Dunker, A.K., et al., *Intrinsic protein disorder in complete genomes.* Genome Inform Ser Workshop Genome Inform, 2000. **11**: p. 161-71.
173. Ward, J.J., et al., *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.* J Mol Biol, 2004. **337**(3): p. 635-45.
174. Fink, A.L., *Natively unfolded proteins.* Curr Opin Struct Biol, 2005. **15**(1): p. 35-41.
175. Dunker, A.K. and Z. Obradovic, *The protein trinity--linking function and disorder.* Nat Biotechnol, 2001. **19**(9): p. 805-6.
176. Liu, J. and B. Rost, *Comparing function and structure between entire proteomes.* Protein Sci, 2001. **10**(10): p. 1970-9.
177. Tompa, P., *Intrinsically unstructured proteins evolve by repeat expansion.* Bioessays, 2003. **25**(9): p. 847-55.
178. Vucetic, S., et al., *Flavors of protein disorder.* Proteins, 2003. **52**(4): p. 573-84.
179. Pontius, B.W., *Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association.* Trends Biochem Sci, 1993. **18**(5): p. 181-6.
180. Gunasekaran, K., et al., *Extended disordered proteins: targeting function with less scaffold.* Trends Biochem Sci, 2003. **28**(2): p. 81-5.
181. Landsteiner, K., *The specificity of serological reactions*. 1936, Springfield, Ill., Baltimore, Md.: C.C. Thomas. vii, 178 p.
182. Pauling, L., *A Theory of the Structure and Process of Formation of Antibodies.* Journal of the American Chemical Society, 1940. **62**: p. 2643-2657.
183. Kriwacki, R.W., et al., *Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11504-9.
184. Ptitsyn, O.B., *Molten globule and protein folding.* Adv Protein Chem, 1995. **47**: p. 83-229.
185. Shakhnovich, E.I. and A.V. Finkelstein, *Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition.* Biopolymers, 1989. **28**(10): p. 1667-80.

186. Dyson, H.J. and P.E. Wright, *Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance.* Adv Protein Chem, 2002. **62**: p. 311-40.
187. Uversky, V.N., *Natively unfolded proteins: a point where biology waits for physics.* Protein Sci, 2002. **11**(4): p. 739-56.
188. Vendel, A.C., S.J. McBryant, and K.J. Lumb, *KIX-mediated assembly of the CBP-CREB-HTLV-1 tax coactivator-activator complex.* Biochemistry, 2003. **42**(43): p. 12481-7.
189. Shi, Z., et al., *Polyproline II structure in a sequence of seven alanine residues.* Proc Natl Acad Sci U S A, 2002. **99**(14): p. 9190-5.
190. Romero, P., Z. Obradovic, and A.K. Dunker, *Folding minimal sequences: the lower bound for sequence complexity of globular proteins.* FEBS Lett, 1999. **462**(3): p. 363-7.
191. Romero, P., et al., *Sequence complexity of disordered protein.* Proteins, 2001. **42**(1): p. 38-48.
192. Karlin, S., et al., *Amino acid runs in eukaryotic proteomes and disease associations.* Proc Natl Acad Sci U S A, 2002. **99**(1): p. 333-8.
193. Jakes, R., M.G. Spillantini, and M. Goedert, *Identification of two distinct synucleins from human brain.* FEBS Lett, 1994. **345**(1): p. 27-32.
194. Dev, K.K., et al., *Part II: alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease.* Neuropharmacology, 2003. **45**(1): p. 14-44.
195. Nakajo, S., et al., *A new brain-specific 14-kDa protein is a phosphoprotein. Its complete amino acid sequence and evidence for phosphorylation.* Eur J Biochem, 1993. **217**(3): p. 1057-63.
196. Beyer, K., *Alpha-synuclein structure, posttranslational modification and alternative splicing as aggregation enhancers.* Acta Neuropathol (Berl), 2006. **112**(3): p. 237-51.
197. Davidson, W.S., et al., *Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes.* J Biol Chem, 1998. **273**(16): p. 9443-9.
198. Perrin, R.J., et al., *Interaction of human alpha-Synuclein and Parkinson's disease variants with phospholipids. Structural analysis using site-directed mutagenesis.* J Biol Chem, 2000. **275**(44): p. 34393-8.
199. Romero, P.R., et al., *Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8390-5.
200. Hossain, S., et al., *Limited proteolysis of NACP/alpha-synuclein.* J Alzheimers Dis, 2001. **3**(6): p. 577-584.
201. Bisaglia, M., et al., *The 11-mer repeats of human alpha-synuclein in vesicle interactions and lipid composition discrimination: a cooperative role.* Biopolymers, 2006. **84**(3): p. 310-6.
202. Jao, C.C., et al., *Structure of membrane-bound alpha-synuclein studied by site-directed spin labeling.* Proc Natl Acad Sci U S A, 2004. **101**(22): p. 8331-6.
203. Clark, R.S., H. Bayir, and L.W. Jenkins, *Posttranslational protein modifications.* Crit Care Med, 2005. **33**(12 Suppl): p. S407-9.

204. Bussell, R., Jr. and D. Eliezer, *Effects of Parkinson's disease-linked mutations on the structure of lipid-associated alpha-synuclein.* Biochemistry, 2004. **43**(16): p. 4810-8.
205. Chandra, S., et al., *A broken alpha -helix in folded alpha -Synuclein.* J Biol Chem, 2003. **278**(17): p. 15313-8.
206. Trojanowski, J.Q. and V.M. Lee, *Parkinson's disease and related alpha-synucleinopathies are brain amyloidoses.* Ann N Y Acad Sci, 2003. **991**: p. 107-10.
207. Souza, J.M., et al., *Dityrosine cross-linking promotes formation of stable alpha -synuclein polymers. Implication of nitrative and oxidative stress in the pathogenesis of neurodegenerative synucleinopathies.* J Biol Chem, 2000. **275**(24): p. 18344-9.
208. Dickson, D.W., et al., *Widespread alterations of alpha-synuclein in multiple system atrophy.* Am J Pathol, 1999. **155**(4): p. 1241-51.
209. Giasson, B.I., et al., *Oxidative damage linked to neurodegeneration by selective alpha-synuclein nitration in synucleinopathy lesions.* Science, 2000. **290**(5493): p. 985-9.
210. Kim, J., *Evidence that the precursor protein of non-A beta component of Alzheimer's disease amyloid (NACP) has an extended structure primarily composed of random-coil.* Mol Cells, 1997. **7**(1): p. 78-83.
211. Paik, S.R., et al., *Aluminum-induced structural alterations of the precursor of the non-A beta component of Alzheimer's disease amyloid.* Arch Biochem Biophys, 1997. **344**(2): p. 325-34.
212. Frankel, A.D. and P.S. Kim, *Modular structure of transcription factors: implications for gene regulation.* Cell, 1991. **65**(5): p. 717-9.
213. Ellis, R.J., *Macromolecular crowding: obvious but underappreciated.* Trends Biochem Sci, 2001. **26**(10): p. 597-604.
214. Luby-Phelps, K., *Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area.* Int Rev Cytol, 2000. **192**: p. 189-221.
215. Davis-Searles, P.R., et al., *Sugar-induced molten-globule model.* Biochemistry, 1998. **37**(48): p. 17048-53.
216. Davis-Searles, P.R., et al., *Interpreting the effects of small uncharged solutes on protein-folding equilibria.* Annu Rev Biophys Biomol Struct, 2001. **30**: p. 271-306.
217. Liu, Y. and D.W. Bolen, *The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes.* Biochemistry, 1995. **34**(39): p. 12884-91.
218. Qu, Y., C.L. Bolen, and D.W. Bolen, *Osmolyte-driven contraction of a random coil protein.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9268-73.
219. Lah, J., et al., *Energetics of structural transitions of the addiction antitoxin MazE: is a programmed bacterial cell death dependent on the intrinsically flexible nature of the antitoxins?* J Biol Chem, 2005. **280**(17): p. 17397-407.
220. Neyroz, P., B. Zambelli, and S. Ciurli, *Intrinsically disordered structure of Bacillus pasteurii UreG as revealed by steady-state and time-resolved fluorescence spectroscopy.* Biochemistry, 2006. **45**(29): p. 8918-30.

221. Zambelli, B., et al., *UreG, a chaperone in the urease assembly process, is an intrinsically unstructured GTPase that specifically binds Zn2+.* J Biol Chem, 2005. **280**(6): p. 4684-95.
222. Ruiz-Sanz, J., et al., *Thermodynamic analysis of the unfolding and stability of the dimeric DNA-binding protein HU from the hyperthermophilic eubacterium Thermotoga maritima and its E34D mutant.* Eur J Biochem, 2004. **271**(8): p. 1497-507.
223. Dunker, A.K., et al., *Flexible nets. The roles of intrinsic disorder in protein interaction networks.* Febs J, 2005. **272**(20): p. 5129-48.
224. Singh, G.P., M. Ganapathi, and D. Dash, *Role of intrinsic disorder in transient interactions of hub proteins.* Proteins, 2006. **66**(4): p. 761-765.
225. Dosztanyi, Z., et al., *Disorder and sequence repeats in hub proteins and their implications for network evolution.* J Proteome Res, 2006. **5**(11): p. 2985-95.
226. Haynes, C., et al., *Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.* PLoS Comput Biol, 2006. **2**(8): p. e100.
227. Han, J.D., et al., *Evidence for dynamically organized modularity in the yeast protein-protein interaction network.* Nature, 2004. **430**(6995): p. 88-93.
228. Cooper, T.F., et al., *Effect of random and hub gene disruptions on environmental and mutational robustness in Escherichia coli.* BMC Genomics, 2006. **7**: p. 237.
229. Batada, N.N., L.D. Hurst, and M. Tyers, *Evolutionary and physiological importance of hub proteins.* PLoS Comput Biol, 2006. **2**(7): p. e88.
230. Batada, N.N., et al., *Stratus not altocumulus: a new view of the yeast protein interaction network.* PLoS Biol, 2006. **4**(10): p. e317.
231. Friedberg, E.C., et al., *DNA Repair and Mutagenesis*. Second ed. 2005, Washington, D.C.: ASM Press.
232. Opperman, T., et al., *A model for a umuDC-dependent prokaryotic DNA damage checkpoint.* Proceedings of the National Academy of Sciences, USA, 1999. **96**(16): p. 9218-23.
233. Sutton, M.D. and G.C. Walker, *umuDC-Mediated Cold Sensitivity is a Manifestation of Functions of the $UmuD_2C$ Complex Involved in a DNA Damage Checkpoint Control.* Journal of Bacteriology, 2001. **183**: p. 1215-1224.
234. Ferentz, A.E., G.C. Walker, and G. Wagner, *Converting a DNA damage checkpoint effector ($UmuD_2C$) into a lesion bypass polymerase ($UmuD'_2C$).* EMBO Journal, 2001. **20**(15): p. 4287-4298.
235. Ferentz, A.E., et al., *Dimerization of the UmuD' protein in solution and its implications for regulation of SOS mutagenesis.* Nature Structural Biology, 1997. **4**(12): p. 979-83.
236. Sutton, M.D., et al., *A model for the structure of the Escherichia coli SOS-regulated $UmuD_2$ protein.* DNA Repair, 2002. **1**: p. 77-93.
237. Shen, X., R. Woodgate, and M.F. Goodman, *Escherichia coli DNA polymerase V subunit exchange: a post-SOS mechanism to curtail error-prone DNA synthesis.* Journal of Biological Chemistry, 2003. **278**(52): p. 52546-50.
238. Sutton, M.D., T. Opperman, and G.C. Walker, *The Escherichia coli SOS mutagenesis proteins UmuD and UmuD' interact physically with the replicative DNA polymerase.* Proceedings of the National Academy of Sciences, USA, 1999. **96**(22): p. 12373-12378.

239. Sutton, M.D., I. Narumi, and G.C. Walker, *Posttranslational modification of the umuD-encoded subunit of Escherichia coli DNA polymerase V regulates its interactions with the beta processivity clamp.* Proceedings of the National Academy of Sciences, USA, 2002. **99**(8): p. 5307-12.
240. Becherel, O.J., R.P.P. Fuchs, and J. Wagner, *Pivotal role of the β-clamp in translesion DNA synthesis and mutagenesis in E. coli cells.* DNA Repair, 2002. **1**: p. 703-708.
241. Beuning, P.J., et al., *Two Processivity Clamp Interactions Differentially Alter the Dual Activities of UmuC.* Molecular Microbiology, 2006. **59**: p. 460-474.
242. Fujii, S. and R.P. Fuchs, *Defining the position of the switches between replicative and bypass DNA polymerases.* EMBO Journal, 2004. **23**: p. 4342-4352.
243. Indiani, C., et al., *A Sliding-Clamp Toolbelt Binds High- and Low-Fidelity DNA Polymerases Simultaneously.* Molecular Cell, 2005. **19**: p. 805-815.
244. Yang, J., et al., *The dynamic processivity of the T4 DNA polymerase during replication.* Proceedings of the National Academy of Sciences, USA, 2004. **101**: p. 8289-8294.
245. Kannouche, P.L., J. Wing, and A.R. Lehmann, *Interaction of Human DNA Polymerase eta with Monoubiquinated PCNA: A Possible Mechanism for the Polymerase Switch in Response to DNA Damage.* Molecular Cell, 2004. **14**: p. 491-500.
246. Gill, G., *SUMO and ubiquitin in the nucleuc: different functions, similar mechanisms?* Genes and Development, 2004. **18**: p. 2046-2059.
247. Vivona, J.B. and Z. Kelman, *The diverse spectrum of sliding clamp interacting proteins.* FEBS Letters, 2003. **546**(2-3): p. 167-72.
248. Warbrick, E., *The puzzle of PCNA's many partners.* Bioessays, 2000. **22**(11): p. 997-1006.
249. Duzen, J.M., G.C. Walker, and M.D. Sutton, *Identification of specific amino acid residues in the E. coli processivity clamp involved in interactions with DNA polymerase III, UmuD and UmuD'.* DNA Repair, 2004. **3**: p. 301-312.
250. Dalrymple, B.P., et al., *A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems.* Proceedings of the National Academy of Sciences, USA, 2001. **98**: p. 11627-11632.
251. Zemla, A., *LGA – A Method for Finding 3D Similarities in Protein Structures.* Nucleic Acids Research, 2003. **31**: p. 3370-3374.
252. Zemla, A., et al., *AS2TS System for Protein Structure Modeling and Analysis.* Nucleic Acids Research, 2005. **33 (Web Server issue)**: p. W111-W115.
253. Luo, Y., et al., *Crystal Structure of LexA: A Conformational Switch for Regulation of Self-Cleavage.* Cell, 2001. **106**: p. 585-594.
254. Bower, M., F.E. Cohen, and J.R.L. Dunbrack, *Sidechain prediction from a backbone-dependent rotamer library: A new tool for homology modeling.* Journal of Molecular Biology, 1997. **267**: p. 1268-1282.
255. Beuning, P.J., et al., *Characterization of E. coli translesion synthesis polymerases and their accessory factors.* Methods in Enzymology, 2006. **408**: p. 318-340.
256. Turner, J., et al., *The internal workings of a DNA polymerase clamp-loading machine.* EMBO Journal, 1999. **18**(3): p. 771-83.

257. Guzzo, A., et al., *Analysis of the region between amino acids 30 and 42 of intact UmuD by a monocysteine approach.* Journal of Bacteriology, 1996. **178**(24): p. 7295-303.
258. Sommer, S., et al., *Specific RecA amino acid changes affect RecA-UmuD'C interaction.* Molecular Microbiology, 1998. **28**(2): p. 281-291.
259. Churchward, G., D. Belin, and Y. Nagamine, *A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors.* Gene, 1984. **31**(1-3): p. 165-71.
260. Oh Johanson, K., T.E. Haynes, and C.S. McHenry, *Chemical Characterization and Purification of the β Subunit of the DNA Polymerase III Holoenyme from an Overproducing Strain.* Journal of Biological Chemistry, 1986. **261**: p. 11460-11465.
261. Lee, M.H., A. Guzzo, and G.C. Walker, *Inhibition of RecA-mediated cleavage in covalent dimers of UmuD.* Journal of Bacteriology, 1996. **178**(24): p. 7304-7.
262. Kulaeva, O.I., et al., *Characterization of the umu-complementing operon from R391.* Journal of Bacteriology, 1995. **177**(10): p. 2737-43.
263. Lee, M.H., T. Ohta, and G.C. Walker, *A monocysteine approach for probing the structure and interactions of the UmuD protein.* Journal of Bacteriology, 1994. **176**(16): p. 4825-37.
264. Wright, S.K. and R.E. Viola, *Evaluation of methods for the quantitation of cysteines in proteins.* Analytical Biochemistry, 1998. **265**: p. 8-14.
265. Battista, J.R., et al., *Dominant negative umuD mutations decreasing RecA-mediated cleavage suggest roles for intact UmuD in modulation of SOS mutagenesis.* Proceedings of the National Academy of Sciences, USA, 1990. **87**(18): p. 7190-4.
266. Nohmi, T., et al., *RecA-mediated cleavage activates UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation.* Proceedings of the National Academy of Sciences, USA, 1988. **85**(6): p. 1816-20.
267. Opperman, T., S. Murli, and G.C. Walker, *The genetic requirements for UmuDC-mediated cold sensitivity are distinct from those for SOS mutagenesis.* Journal of Bacteriology, 1996. **178**(15): p. 4400-11.
268. Sutton, M.D., et al., *Genetic Interactions between the Escherichia coli umuDC Gene Products and the β Processivity Clamp of the Replicative DNA Polymerase.* Journal of Bacteriology, 2001. **183**: p. 2897-2909.
269. Harris, D.A., *Light Spectroscopy.* 1996, Oxford, UK: BIOS Scientific Publishers, Ltd.
270. Dohrmann, P.R. and C.S. McHenry, *A bipartite polymerase-processivity factor interaction: only the internal beta binding site of the alpha subunit is required for processive replication by the DNA polymerase III holoenzyme.* Journal of Molecular Biology, 2005. **350**(2): p. 228-39.
271. Wagner, J., et al., *The β clamp targets DNA polymerase IV to DNA and strongly increases its processivity.* EMBO Reports, 2000. **1**: p. 484-488.
272. Bunting, K.A., S.M. Roe, and L.H. Pearl, *Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the β-clamp.* EMBO Journal, 2003. **22**: p. 5883-5892.

273. Jeruzalmi, D., et al., *Mechanism of Processivity Clamp Opening by the Delta Subunit Wrench of the Clamp Loader Complex of E. coli DNA Polymerase III.* Cell, 2001. **106**: p. 417-428.
274. Humphrey, W., A. Dalke, and K. Schulten, *VMD - Visual Molecular Dynamics.* Journal of Molecular Graphics, 1996. **14**: p. 33-38.
275. McDonald, J.P., et al., *Intermolecular cleavage by UmuD-like mutagenesis proteins.* Proceedings of the National Academy of Sciences, USA, 1998. **95**(4): p. 1478-83.
276. Kazmirski, S.L., et al., *Out-of-plane motions in open sliding clamps: Molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen.* Proceedings of the National Academy of Sciences, USA, 2005. **102**: p. 13801-13806.
277. Kong, X.-P., et al., *Three-dimensional structure of the β subunit of E. coli DNA Polymerase III Holoenzyme: A Sliding DNA Clamp.* Cell, 1992. **69**: p. 425-437.
278. Wijffels, G., et al., *Inhibition of Protein Interactions with the β_2 Sliding Clamp of Escherichia coli DNA Polymerase III by Peptides from β_2-Binding Proteins.* Biochemistry, 2004. **43**: p. 5661-5671.
279. Friedberg, E.C., R. Wagner, and M. Radman, *Specialized DNA polymerases, cellular survival, and the genesis of mutations.* Science, 2002. **296**(5573): p. 1627-30.
280. Brotcorne-Lannoye, A. and G. Maenhaut-Michel, *Role of RecA protein in untargeted UV mutagenesis of bacteriophage lambda: evidence for the requirement for the dinB gene.* Proc Natl Acad Sci U S A, 1986. **83**(11): p. 3904-8.
281. McKenzie, G.J., et al., *SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification.* Mol Cell, 2001. **7**(3): p. 571-9.
282. Kim, S.R., et al., *Multiple pathways for SOS-induced mutagenesis in Escherichia coli: an overexpression of dinB/dinP results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA.* Proc Natl Acad Sci U S A, 1997. **94**(25): p. 13792-7.
283. Roth, J.R., et al., *Origin of mutations under selection: the adaptive mutation controversy.* Annu Rev Microbiol, 2006. **60**: p. 477-501.
284. Slack, A., et al., *On the mechanism of gene amplification induced under stress in Escherichia coli.* PLoS Genet, 2006. **2**(4): p. e48.
285. Stumpf, J.D., A.R. Poteete, and P.L. Foster, *Amplification of lac Cannot Account for Adaptive Mutation to Lac+ in Escherichia coli.* J. Bacteriol., 2007. **189**(6): p. 2291-2299.
286. Friedberg, E.C., et al., *DNA Repair and Mutagenesis*. 2 ed. 2006, Washington, DC: ASM Press.
287. Schlacher, K., et al., *DNA polymerase V and RecA protein, a minimal mutasome.* Mol Cell, 2005. **17**(4): p. 561-72.
288. Fujii, S., A. Isogawa, and R.P. Fuchs, *RecFOR proteins are essential for Pol V-mediated translesion synthesis and mutagenesis.* Embo J, 2006. **25**(24): p. 5754-63.
289. Fuchs, R.P., S. Fujii, and J. Wagner, *Properties and functions of Escherichia coli: Pol IV and Pol V.* Adv Protein Chem, 2004. **69**: p. 229-64.

290. Lenne-Samuel, N., et al., *The processivity factor beta controls DNA polymerase IV traffic during spontaneous mutagenesis and translesion synthesis in vivo.* EMBO Rep, 2002. **3**(1): p. 45-9.
291. Frampton, J., et al., *Postreplication repair and PCNA modification in Schizosaccharomyces pombe.* Mol Biol Cell, 2006. **17**(7): p. 2976-85.
292. Ogi, T., P. Kannouche, and A.R. Lehmann, *Localisation of human Y-family DNA polymerase kappa: relationship to PCNA foci.* J Cell Sci, 2005. **118**(Pt 1): p. 129-36.
293. Kannouche, P.L., J. Wing, and A.R. Lehmann, *Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage.* Mol Cell, 2004. **14**(4): p. 491-500.
294. Bunting, K.A., S.M. Roe, and L.H. Pearl, *Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp.* EMBO J, 2003. **22**(21): p. 5883-92.
295. Bergoglio, V., et al., *Localisation of human DNA polymerase kappa to replication foci.* J Cell Sci, 2002. **115**(Pt 23): p. 4413-8.
296. Beuning, P.J., et al., *Two processivity clamp interactions differentially alter the dual activities of UmuC.* Mol Microbiol, 2006. **59**(2): p. 460-74.
297. Nohmi, T., *Environmental stress and lesion-bypass DNA polymerases.* Annu Rev Microbiol, 2006. **60**: p. 231-53.
298. Layton, J.C. and P.L. Foster, *Error-prone DNA polymerase IV is regulated by the heat shock chaperone GroE in Escherichia coli.* J Bacteriol, 2005. **187**(2): p. 449-57.
299. Donnelly, C.E. and G.C. Walker, *groE mutants of Escherichia coli are defective in umuDC-dependent UV mutagenesis.* J Bacteriol, 1989. **171**(11): p. 6117-25.
300. Grudniak, A.M., M. Kuc, and K.I. Wolska, *Role of Escherichia coli DnaK and DnaJ chaperones in spontaneous and induced mutagenesis and their effect on UmuC stability.* FEMS Microbiol Lett, 2005. **242**(2): p. 361-6.
301. Jarosz, D.F., et al., *Y-family DNA polymerases in Escherichia coli.* Trends Microbiol, 2007.
302. McGrew, D.A. and K.L. Knight, *Molecular design and functional organization of the RecA protein.* Crit Rev Biochem Mol Biol, 2003. **38**(5): p. 385-432.
303. Cairns, J. and P.L. Foster, *Adaptive reversion of a frameshift mutation in Escherichia coli.* Genetics, 1991. **128**(4): p. 695-701.
304. Tompkins, J.D., et al., *Error-prone polymerase, DNA polymerase IV, is responsible for transient hypermutation during adaptive mutation in Escherichia coli.* J Bacteriol, 2003. **185**(11): p. 3469-72.
305. Dalrymple, B.P., et al., *A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11627-32.
306. Potapova, O., N.D. Grindley, and C.M. Joyce, *The mutational specificity of the Dbh lesion bypass polymerase and its implications.* J Biol Chem, 2002. **277**(31): p. 28157-66.

307. Kobayashi, S., et al., *Fidelity of Escherichia coli DNA polymerase IV. Preferential generation of small deletion mutations by dNTP-stabilized misalignment.* J Biol Chem, 2002. **277**(37): p. 34198-207.
308. Story, R.M., I.T. Weber, and T.A. Steitz, *The structure of the E. coli recA protein monomer and polymer.* Nature, 1992. **355**(6358): p. 318-25.
309. Bailone, A., et al., *A RecA protein mutant deficient in its interaction with the UmuDC complex.* Biochimie, 1991. **73**(4): p. 479-84.
310. Bavoux, C., J.S. Hoffmann, and C. Cazaux, *Adaptation to DNA damage and stimulation of genetic instability: the double-edged sword mammalian DNA polymerase kappa.* Biochimie, 2005. **87**(7): p. 637-46.
311. Afonnikov, D.A. and N.A. Kolchanov, *CRASP: a program for analysis of coordinated substitutions in multiple alignments of protein sequences.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W64-8.
312. Bienko, M., et al., *Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis.* Science, 2005. **310**(5755): p. 1821-4.
313. Opperman, T., S. Murli, and G.C. Walker, *The genetic requirements for UmuDC-mediated cold sensitivity are distinct from those for SOS mutagenesis.* J Bacteriol, 1996. **178**(15): p. 4400-11.
314. Salles, B. and C. Paoletti, *Control of UV induction of recA protein.* Proc Natl Acad Sci U S A, 1983. **80**(1): p. 65-9.
315. Bavoux, C., et al., *Up-regulation of the error-prone DNA polymerase kappa promotes pleiotropic genetic alterations and tumorigenesis.* Cancer Res, 2005. **65**(1): p. 325-30.
316. Lone, S., et al., *Human DNA Polymerase [kappa] Encircles DNA: Implications for Mismatch Extension and Lesion Bypass.* Molecular Cell, 2007. **25**(4): p. 601-614.
317. Beuning, P.J., et al., *Characterization of Escherichia coli translesion synthesis polymerases and their accessory factors.* Methods Enzymol, 2006. **408**: p. 318-40.
318. Konola, J.T., et al., *Mutations at Pro67 in the RecA protein P-loop motif differentially modify coprotease function and separate coprotease from recombination activities.* J Biol Chem, 1995. **270**(15): p. 8411-9.
319. Chen, B.P. and T. Hai, *Expression vectors for affinity purification and radiolabeling of proteins using Escherichia coli as host.* Gene, 1994. **139**(1): p. 73-5.
320. Niehbuhr, K. and J. Wehland, *Screening of Antibody Epitopes and Regions of Protein-Protein Interaction Sites Using SPOT Peptides*, in *Immunology Methods Manual: The Comprehensive Sourcebook of Techniques*, I. Lefkovits, Editor. 1997, Academic Press: San Diego, CA.
321. Godoy, V.G., F.S. Gizatullin, and M.S. Fox, *Some features of the mutability of bacteria during nonlethal selection.* Genetics, 2000. **154**(1): p. 49-59.
322. Roberts, J.W., C.W. Roberts, and N.L. Craig, *Escherichia coli recA gene product inactivates phage lambda repressor.* Proc Natl Acad Sci U S A, 1978. **75**(10): p. 4714-8.
323. Gonzalez, M., et al., *Structural insights into the regulation of SOS mutagenesis.* Acta Biochim Pol, 1998. **45**(1): p. 163-72.

324. Tompa, P., *The interplay between structure and function in intrinsically unstructured proteins.* FEBS Lett, 2005. **579**(15): p. 3346-54.
325. Le Gall, T., et al., *Intrinsic disorder in the protein data bank.* J Biomol Struct Dyn, 2007. **24**(4): p. 325-42.
326. Iakoucheva, L.M., et al., *Identification of intrinsic order and disorder in the DNA repair protein XPA.* Protein Sci, 2001. **10**(3): p. 560-71.
327. Ferentz, A.E., et al., *Dimerization of the UmuD' protein in solution and its implications for regulation of SOS mutagenesis.* Nat Struct Biol, 1997. **4**(12): p. 979-83.
328. Gonzalez, M., et al., *Subunit-specific degradation of the UmuD/D' heterodimer by the ClpXP protease: the role of trans recognition in UmuD' stability.* Embo J, 2000. **19**(19): p. 5251-8.
329. Csizmok, V., et al., *A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins.* Mol Cell Proteomics, 2006. **5**(2): p. 265-73.
330. Sutton, M.D., I. Narumi, and G.C. Walker, *Posttranslational modification of the umuD-encoded subunit of Escherichia coli DNA polymerase V regulates its interactions with the beta processivity clamp.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5307-12.
331. Tennyson, R.B. and J.E. Lindsley, *Type II DNA topoisomerase from Saccharomyces cerevisiae is a stable dimer.* Biochemistry, 1997. **36**(20): p. 6107-14.
332. Johanson, K.O. and C.S. McHenry, *Purification and characterization of the beta subunit of the DNA polymerase III holoenzyme of Escherichia coli.* J Biol Chem, 1980. **255**(22): p. 10984-90.
333. Duzen, J.M., G.C. Walker, and M.D. Sutton, *Identification of specific amino acid residues in the E. coli beta processivity clamp involved in interactions with DNA polymerase III, UmuD and UmuD'.* DNA Repair (Amst), 2004. **3**(3): p. 301-12.
334. Gerlach, V.L., et al., *Human and mouse homologs of Escherichia coli DinB (DNA polymerase IV), members of the UmuC/DinB superfamily.* Proc Natl Acad Sci U S A, 1999. **96**(21): p. 11922-7.
335. Perona, J.J. and A.M. Martin, *Conformational transitions and structural deformability of EcoRV endonuclease revealed by crystallographic analysis.* J Mol Biol, 1997. **273**(1): p. 207-25.
336. Gallagher, S.R., *Native Discontinuous Electrophoresis and Generation of Molecular Weight Standard Curves (Ferguson Plots).* Current Protocols in Protein Science, 1995: p. 10.3.5.
337. Williams, R.M., et al., *The protein non-folding problem: amino acid determinants of intrinsic order and disorder.* Pac Symp Biocomput, 2001: p. 89-100.
338. Dunker, A.K., et al., *Intrinsic disorder and protein function.* Biochemistry, 2002. **41**(21): p. 6573-82.
339. Jonczyk, P. and A. Nowicka, *Specific in vivo protein-protein interactions between Escherichia coli SOS mutagenesis proteins.* J Bacteriol, 1996. **178**(9): p. 2580-5.
340. Sutton, M.D., M. Kim, and G.C. Walker, *Genetic and biochemical characterization of a novel umuD mutation: insights into a mechanism for UmuD self-cleavage.* J Bacteriol, 2001. **183**(1): p. 347-57.

341. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. **41**(3): p. 415-27.
342. Peat, T.S., et al., *Production and crystallization of a selenomethionyl variant of UmuD', an Escherichia coli SOS response protein.* Proteins, 1996. **25**(4): p. 506-9.
343. Freire, E., *Statistical thermodynamic linkage between conformational and binding equilibria.* Adv Protein Chem, 1998. **51**: p. 255-79.
344. Erijman, L. and G. Weber, *Oligomeric protein associations: transition from stochastic to deterministic equilibrium.* Biochemistry, 1991. **30**(6): p. 1595-9.
345. Sousa, F.J., et al., *Tetramerization of the LexA repressor in solution: implications for gene regulation of the E.coli SOS system at acidic pH.* J Mol Biol, 2006. **359**(4): p. 1059-74.
346. Golovanov, A.P., et al., *ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure.* Mol Microbiol, 2003. **50**(4): p. 1141-53.
347. Carmelo, E., et al., *The unstructured N-terminal tail of ParG modulates assembly of a quaternary nucleoprotein complex in transcription repression.* J Biol Chem, 2005. **280**(31): p. 28683-91.
348. Mullen, G.P., J.B. Vaughn, Jr., and A.S. Mildvan, *Sequential proton NMR resonance assignments, circular dichroism, and structural properties of a 50-residue substrate-binding peptide from DNA polymerase I.* Arch Biochem Biophys, 1993. **301**(1): p. 174-83.
349. Hollis, T., et al., *Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7.* Proc Natl Acad Sci U S A, 2001. **98**(17): p. 9557-62.
350. McLenigan, M., et al., *Novel Escherichia coli umuD' mutants: structure-function insights into SOS mutagenesis.* J Bacteriol, 1998. **180**(17): p. 4658-66.
351. Frank, E.G., et al., *In vivo stability of the Umu mutagenesis proteins: a major role for RecA.* J Bacteriol, 1996. **178**(12): p. 3550-6.
352. Raman, P., V. Cherezov, and M. Caffrey, *The Membrane Protein Data Bank.* Cell Mol Life Sci, 2006. **63**(1): p. 36-51.
353. Fethiere, J., *Three-dimensional crystallization of membrane proteins.* Methods Mol Biol, 2007. **363**: p. 191-223.
354. Silhavy, T.J., et al., *Experiments with gene fusions*. 1984, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory. xv, 303 p., [1] p. of plates.
355. Woodgate, R., *Construction of a umuDC operon substitution mutation in Escherichia coli.* Mutat Res, 1992. **281**(3): p. 221-5.
356. Nguyen, L.H., D.B. Jensen, and R.R. Burgess, *Overproduction and purification of sigma 32, the Escherichia coli heat shock transcription factor.* Protein Expr Purif, 1993. **4**(5): p. 425-33.
357. Reuven, N.B., G. Tomer, and Z. Livneh, *The mutagenesis proteins UmuD' and UmuC prevent lethal frameshifts while increasing base substitution mutations.* Mol Cell, 1998. **2**(2): p. 191-9.
358. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning: a laboratory manual*. 2nd ed. 1989, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Figure 1-1: **Sequence alignment of UmuD and similar proteins.**

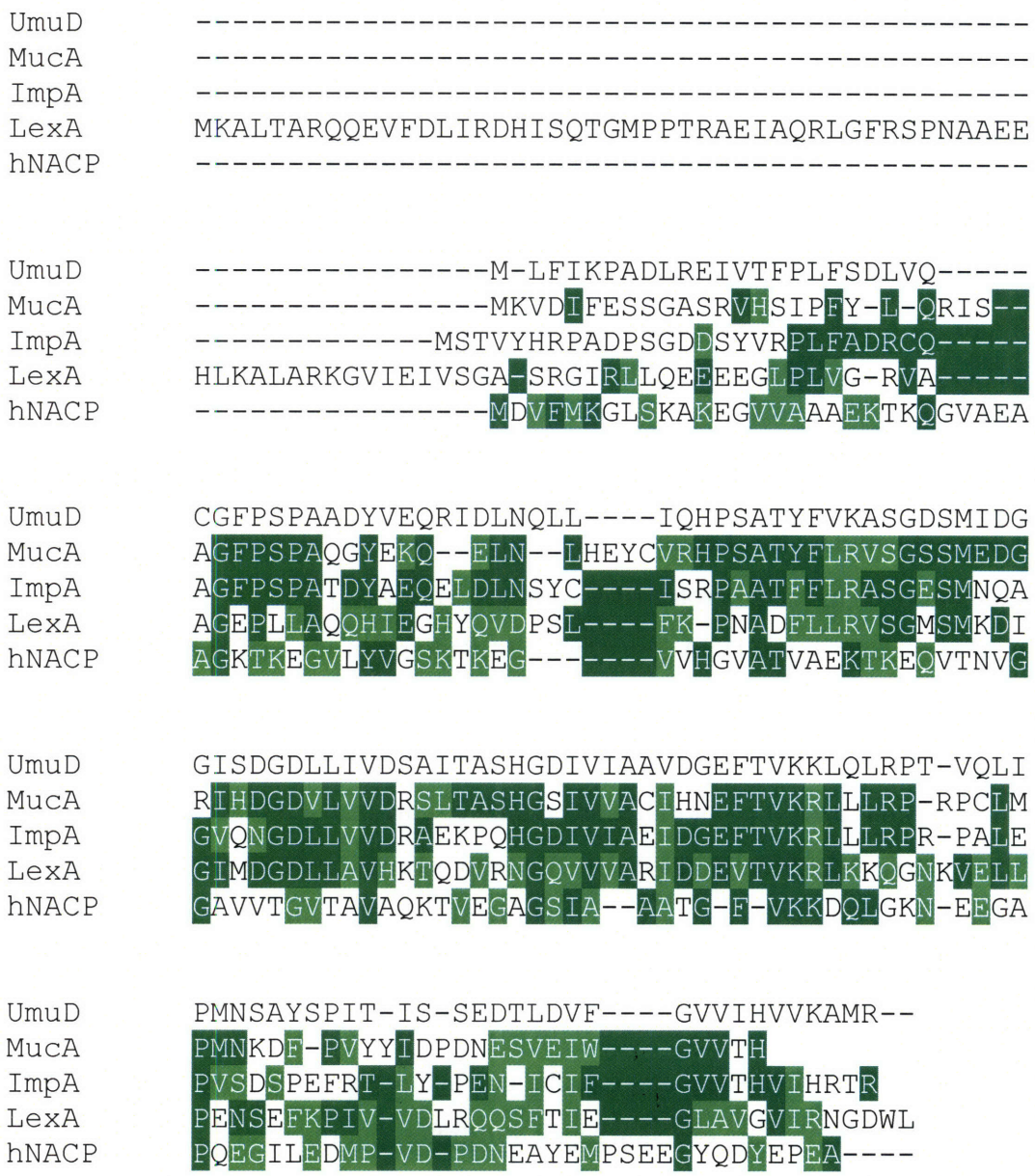


Figure 1-1: **Sequence alignment of UmuD and similar proteins.** Sequence alignment of *E. coli* UmuD and plasmid-borne homologs (MucA and ImpA), the similar protein LexA, and the full-length intrinsically disordered human α-synuclein, which has some primary structural similarity. Residues that are identical in UmuD are in dark green; residues that are similar are in light green.

359. Kim, Y.I., et al., *Dynamics of substrate denaturation and translocation by the ClpXP degradation machine.* Mol Cell, 2000. **5**(4): p. 639-48.
360. Mustard, J.A. and J.W. Little, *Analysis of Escherichia coli RecA interactions with LexA, lambda CI, and UmuD by site-directed mutagenesis of recA.* J Bacteriol, 2000. **182**(6): p. 1659-70.
361. Sutton, M.D., et al., *umuDC-dnaQ Interaction and its implications for cell cycle regulation and SOS mutagenesis in Escherichia coli.* J Bacteriol, 2001. **183**(3): p. 1085-9.
362. Segel, I.H., *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems*. 1975, New York: Wiley. xxii, 957 p.
363. Gonzalez, M., et al., *Lon-mediated proteolysis of the Escherichia coli UmuD mutagenesis protein: in vitro degradation and identification of residues required for proteolysis.* Genes Dev, 1998. **12**(24): p. 3889-99.
364. Lin, L.L. and J.W. Little, *Autodigestion and RecA-dependent cleavage of Ind- mutant LexA proteins.* J Mol Biol, 1989. **210**(3): p. 439-52.
365. Lin, L.L. and J.W. Little, *Isolation and characterization of noncleavable (Ind-) mutants of the LexA repressor of Escherichia coli K-12.* J Bacteriol, 1988. **170**(5): p. 2163-73.
366. Woodgate, R. and A.S. Levine, *Damage inducible mutagenesis: recent insights into the activities of the Umu family of mutagenesis proteins.* Cancer Surv, 1996. **28**: p. 117-40.
367. Koch, W.H., et al., *Escherichia coli umuDC mutants: DNA sequence alterations and UmuD cleavage.* Mol Gen Genet, 1992. **233**(3): p. 443-8.
368. Steinborn, G., *Uvm mutants of Escherichia coli K12 deficient in UV mutagenesis. I. Isolation of uvm mutants and their phenotypical characterization in DNA repair and mutagenesis.* Mol Gen Genet, 1978. **165**(1): p. 87-93.
369. Gimble, F.S. and R.T. Sauer, *Lambda repressor inactivation: properties of purified ind- proteins in the autodigestion and RecA-mediated cleavage reactions.* J Mol Biol, 1986. **192**(1): p. 39-47.
370. Bebenek, K., et al., *5'-Deoxyribose phosphate lyase activity of human DNA polymerase iota in vitro.* Science, 2001. **291**(5511): p. 2156-9.
371. Smith, C.M., et al., *Sequence analysis and mapping of the Salmonella typhimurium LT2 umuDC operon.* J Bacteriol, 1990. **172**(9): p. 4964-78.
372. Thomas, S.M., et al., *Structural characterization of the Salmonella typhimurium LT2 umu operon.* J Bacteriol, 1990. **172**(9): p. 4979-87.
373. Sundin, G.W., et al., *Resistance to ultraviolet light in Pseudomonas syringae: sequence and functional analysis of the plasmid-encoded rulAB genes.* Gene, 1996. **177**(1-2): p. 77-81.
374. Ennis, D.G., et al., *Analysis of recA mutants with altered SOS functions.* Mutat Res, 1995. **336**(1): p. 39-48.