SINGLE MOLECULE STUDIES OF DAMAGE RECOGNITION BY THE HUMAN UV-DAMAGED DNA-BINDING PROTEIN (UV-DDB)

by

Harshad Ghodke

M.S in Chemical Engineering, Carnegie Mellon

Submitted to the Graduate Faculty of
the School of Medicine in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh
2013

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Harshad Ghodke

It was defended on

November 22, 2013

and approved by

Bennett Van Houten, Ph.D., Professor

Guillermo Romero, Ph.D., Associate Professor

Neil Kad, Ph.D., Lecturer

Roger Hendrix, Ph.D.,Professor

Marcel Bruchez, Ph.D., Associate Professor

Dissertation Director: Bennett Van Houten, Ph.D., Professor

SINGLE MOLECULE STUDIES OF DAMAGE RECOGNITION BY THE HUMAN UV-DAMAGED DNA-BINDING PROTEIN (UV-DDB)

Harshad Ghodke, PhD

University of Pittsburgh, 2013

Nucleotide excision repair (NER) is a DNA repair pathway that processes helix distorting lesions in DNA. In humans, lesions such as UV-induced photoproducts are recognized by the UV-damaged DNA binding protein (UV-DDB). How human DNA repair proteins survey the genome for UV-induced photoproducts remains a poorly understood aspect of the initial damage recognition step in nucleotide excision repair (NER). Specifically, the transport mechanisms employed by UV-DDB, as well as, the stoichiometry of UV-DDB on physiologically relevant DNA substrates containing DNA damage remain unclear. To understand damage recognition by UV-DDB, we performed single molecule experiments, which revealed that the human UV-damaged DNA binding protein (UV-DDB) samples damage in DNA primarily via a three dimensional search mechanism. We found that UV-DDB displays a remarkable heterogeneity in the kinetics of damage recognition. Our results indicate that UV-DDB examines sites on DNA in discrete steps prior to forming long-lived, non-motile $(DDB1\text{-}DDB2)_2$ dimers at sites of damage. To understand structure-function relationships governing DNA damage recognition by UV-DDB, we tested the xeroderma pigmentosum group E (XP-E) causing K244E mutant of DDB2 found in patient XP82TO. We found that K244E DDB2 supported UV-DDB dimerization but was found to slide on DNA and failed to stably engage lesions. These findings provide molecular insight into the loss of damage discrimination observed in this XP-E patient. Here we propose a framework for a conformational proofreading mechanism for specific damage recognition by UV-DDB.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Hazaaron khwaishein aisi, ke har khwahish pe dam nikle,
bahut nikale mere armaan, lekin phir bhi kam nikle.
-Ghalib

Thousands of desires, each worth dying for,
many of them I have realized, yet I yearn for more.

The last few years have made me realize how incredibly fortunate I am. Everyday, I am reminded that there are vast swathes of humanity who have neither the freedoms nor the opportunities to pursue their dreams. I've been fortunate to have good mentors, dear friends and family by my side to help me achieve my dreams and ambitions.

I'm thankful for the opportunity to work in Ben's lab. Over the years, I have come to appreciate that the practice of good science requires a certain amount of fearlessness - to boldly go where no man has gone. Ben has been a great example of this, as he has taken his core strengths, to ask fundamental questions about DNA damage recognition, using new techniques. I would not be overstating it, if I said that Ben's visions and a great scientific team have led the way toward understanding nucleotide excision repair at the single molecule level. With this dissertation, I hope that my efforts will take this effort at least a little bit further and enhance our understanding of damage recognition in humans. I've also realized that good scientists are rare and good mentors are rarer still. Ben has been with me through this entire journey, every single step of the way. His emphasis on always asking the most important question and using the right techniques to answer them, is something I've had the good fortune to be part of. I'm also thankful to him for having the confidence in me and providing me with opportunities to attend some excellent conferences over the last few years.

I've also been fortunate to have had the opportunity to work with Vesna Rapic-Otrin -

undoubtedly one of the prevailing experts on UV-DDB. It is not often that one has access to such expertise and I've thoroughly enjoyed our interactions over the years. Vesna's unwavering interest and guidance in my work over the course of this dissertation have been a constant source of encouragement for me, even in tough times.

I also thank Simon Watkins and the team at CBI (Greg Gibson and Donna Stolz), for without their help, I would not be able to perform much of the research in this dissertation. Everyday, I am surprised at Simon's generosity in giving me the key to the CBI. Many of the long nights and weekends that I have spent at the CBI would simply have been impossible without it. Without Greg's help, the microscope could not have been reconfigured to enable the imaging that I needed for this work.

I thank my thesis committee - Guillermo Romero, Neil Kad, Marcel Bruchez and Roger Hendrix for the guidance over the course of this dissertation which has undoubtedly made the work better. Along those lines, I would like to thank Jennifer Walker and Susanna Godwin for ensuring that I was on top of things as far as academic requirements go.

Additionally, I've been fortunate to have good colleagues in the lab - both past and present - especially Hong Wang, Ye Peng and Amy Furda - from whom I have learnt a great deal about working at the bench. A special thanks to Vera Roginskaya for helping out with the practical aspects of research at the bench, and for the delicious cake that I will surely miss. Perhaps more than anything else in the lab, I've enjoyed training the undergrads in the lab, especially Selamawit Woldemeskel - an endeavour that has reminded me of a passion to teach that I'd shelved away since my days at IIT-B. My heartfelt gratitude goes out also to the members of the DNA repair group at the Hillman Cancer Center.

I've also been fortunate to find friends that I can keep close to me. Rakesh, John and Siva - admittedly between us will always have at least three continents to separate us - I thank each one of you for always being there for me. Grant, Bart and Yao - I will miss the Sunday brunches. My friends from undergrad, especially, Prarit Agarwal, Kirit Makwana and Prerit Jaiswal - I'm glad to have your lifelong friendship, and I look forward to many more getaways together. My time in Pittsburgh would have been bland without my Pittsburgh family - Rohan Manohar, Kalidasan Thambiayya, Shri Padwad, Brian Tumelty, Adhiraj Joshi, Lavanya Sharan, K-Ann Buela, Madhav Sankunny, Vidya Ramachandran and Mike

Freeman. Michelle Barbi and Lucas Santana dos Santos (and baby Lavinia), Elise Foquerel, Andrea Braganza, Advaitha Madireddy and Armin Gamper, it has been a pleasure knowing each one of you, and working at the Cancer Center has certainly been enriched by your presence.

Absolutely none of this would have been possible without the support of my family. Mom, Dad and Rohit - I don't say it nearly often enough - but thank you for your support over the years. Also, Sanjay mama and Arun mama, my stay in the United States would have been impossible if you did not have faith in me. Aji and Ajo, your wisdom and insistence that things always turn out for the best as long as one does the work and maintains faith, have undoubtedly helped me during tough times.

The last few months of my stay in Pittsburgh have been unusually eventful and I want to thank Dave for being with me in sickness (two ER visits and a broken foot!) and in health. Thank you for your patience in putting up with me.

1.0 INTRODUCTION

1.1 MAMMALIAN NUCLEOTIDE EXCISION REPAIR AND XERODERMA PIGMENTOSUM

Exposure of DNA to genotoxins results in the alteration of the integrity of the double helix of DNA (termed as ‘damage’) (Figure 1.1). Damage in DNA may correspond to base damage (including spontaneous loss of bases, oxidation of bases, formation of chemical adducts, intra- and inter- strand cross-links) or damage to the phosphodiester backbone of the DNA (including single or double strand breaks) [1]. DNA damage may be reversed, repaired or tolerated. Uncorrected damage in DNA can lead to a loss or alteration of the genetic information. Maintaining an undamaged copy of the genetic information is critical to survival and propagation of life, and various DNA repair pathways have evolved to process the different classes of damage [1]. Not surprisingly, these mechanisms are broadly conserved across all life forms.

DNA damage which results in the formation of adducts or intra-strand crosslinks due to chemical modification (henceforth referred to as ‘lesions’) is repaired by the Nucleotide Excision Repair pathway (NER). This pathway has specifically evolved to repair bulky, helix distorting lesions in the genome. In this repair pathway, the damage is recognized, verified and excised followed by restoration of the DNA by the coordinated and highly regulated activity of over 30 different polypeptides (repair factors) [3]. Deficiencies in the molecular functions of these repair factors lead to the imperfect processing of DNA lesions resulting in unrepaired lesions. Patients with mutations in the genes corresponding to some of these repair factors exhibit forms of the autosomal recessive disorder termed xeroderma pigementosum (XP), Cockayne syndrome (CS), combined XP/CS and Trichothiodystrophy (TTD)

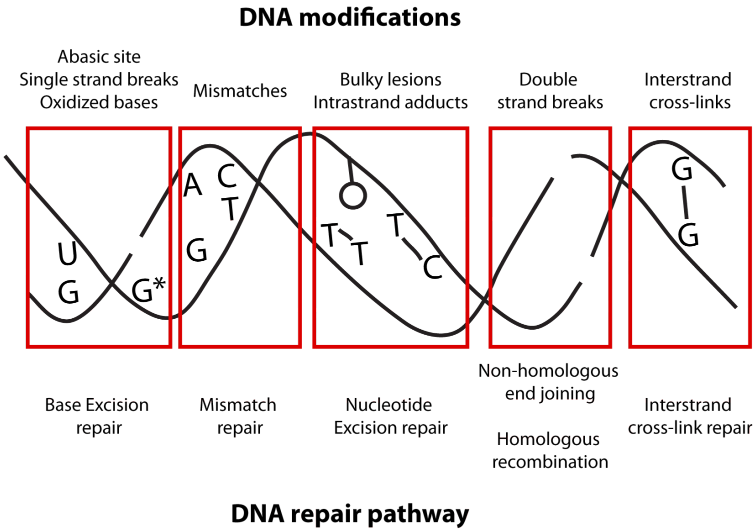


Figure 1.1: DNA damage and repair pathways

Adapted from [2]

[2]. These patients can exhibit photosensitivity, susceptibility to skin-cancers, neurological abnormalities and premature aging among other symptoms [4, 2].

Bulky DNA damage may be processed in three different ways. These are: global genomic repair (GG-NER), transcription coupled repair (TC-NER) and translesion synthesis [5, 6, 7]. The presence of lesions in the genome leads to the stalling of the replication fork in replicating cells. This is because replicative polymerases such as Polδ and Polϵ are high fidelity polymerases [8]. Upon encountering unfamiliar bases, these polymerases stall and are thought to be 'switched' out by the lower fidelity, low processivity polymerases which are capable of incorporating the correct bases against certain types of damaged bases [8, 7]. This process is called translesion synthesis. In this process, the lesion is not repaired, it is merely tolerated and the genome is replicated relatively accurately, giving the repair machinery another chance to detect it.

An alternate strategy to process damage is to detect it and to repair it. In this case, there are two ways by which damage may be recognized (see figure 1.2). In actively transcribed DNA, RNAP which reads the DNA to generate template RNA stalls at sites of damage.

These stalled RNA polymerases (RNAP) at sites of lesions in the template strand undergoing transcription serve to trigger the TC-NER pathway of NER [6]. These lesions are then handed over to the damage verification factor in NER, followed by lesion processing.

On the other hand, damage in genomic DNA is also recognized and repaired independently of the transcriptional state of the genome. This pathway of NER is known as Global Genomic Repair (GG-NER). In this pathway, DNA damage surveillance factors recognize damage in genomic DNA and then hand over the lesion to downstream damage verification and processing factors. In mammalian cells, these two pathways converge at the stage of damage verification and these reactions are orchestrated by over 30 different gene products (figure 1.2) [3].

1.2 SUBSTRATES OF THE NER PATHWAY

The NER machinery is capable of processing a wide variety of lesions. Some of these have been characterized and include: UV-induced photoproducts, polycyclic aromatic hydrocarbons, aromatic amines, cholesterol adducts and psoralen mono-adducts. For a comprehensive list, refer to the review by Gillet and Scharer [5].

1.3 DNA TRANSACTIONS IN GLOBAL GENOMIC REPAIR

In this work, we will discuss damage recognition during the GG-NER pathway of NER. The basic DNA transactions involved in NER broadly consist of the following steps: damage surveillance and recognition; damage verification; helix opening and stabilizing of the repair intermediates; dual incision of the DNA in the context of the lesion; repair synthesis and DNA ligation [9]. These DNA transactions are coordinated by a large number of repair factors which assemble and disassemble sequentially at sites of damage [10, 11, 12, 13, 14]. It follows thus, that the composition of the NER machinery is dynamic and its biological function is highly regulated both by the presence of available binding motifs/partners as well as post-

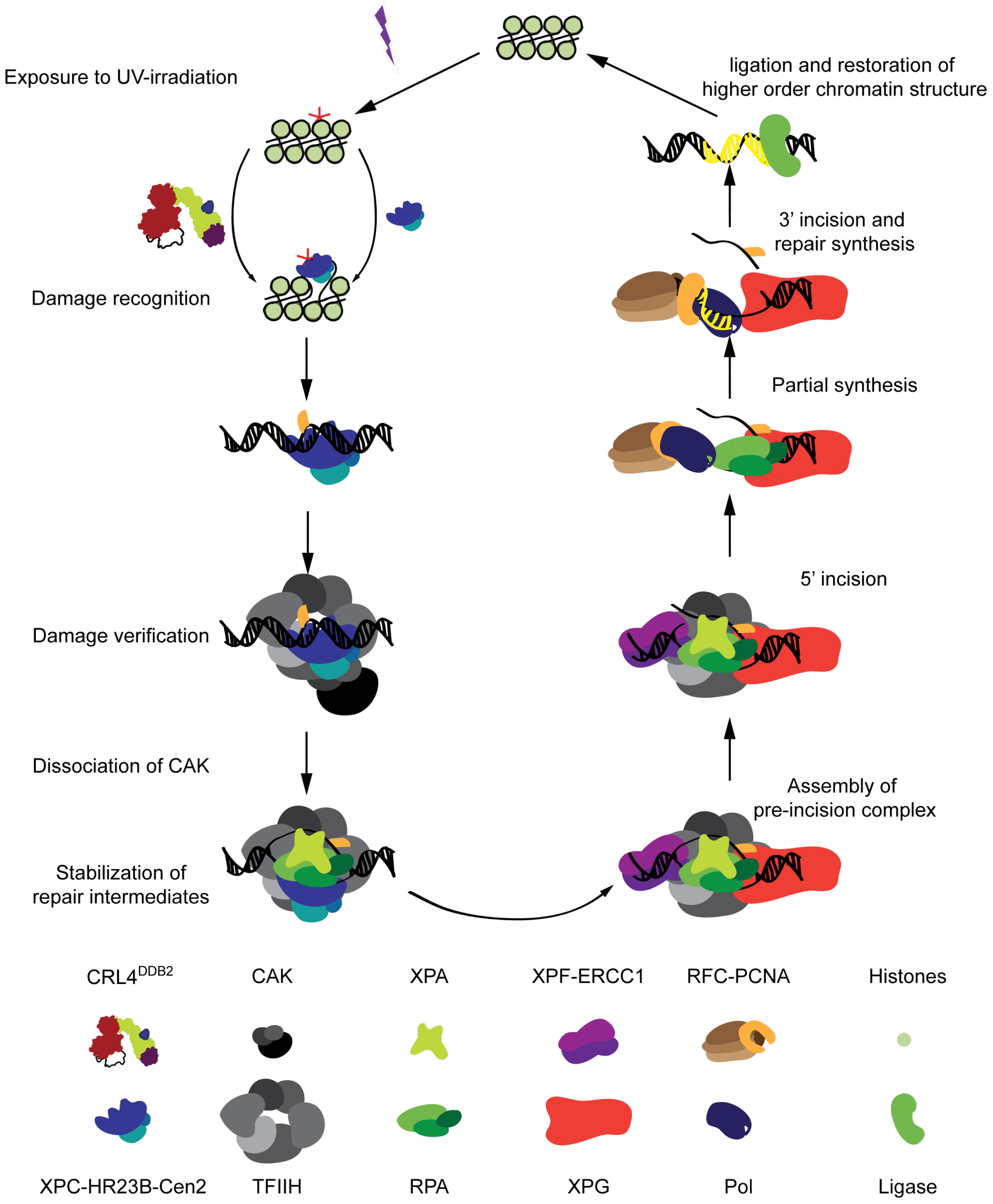


Figure 1.2: Global genomic nucleotide excision repair

translational modifications. With this concept of hand-offs in mind, here, we will outline the roles of the various gene products that co-ordinate the various DNA transactions discussed above, with an emphasis on known interactions and known roles in DNA processing.

Damage in DNA is recognized by two damage surveillance factors - the UV-DDB complex and the XPC complex (Figure 1.2) [15]. The UV-DDB complex consists of the heterodimer of the DDB1 (p127) and DDB2 (p48) peptides, which are tightly associated together [16]. Early studies failed to detect a significant role of UV-DDB in promoting NER in vitro, since the complete repair reaction could be recapitulated in the absence of UV-DDB using recombinant repair factors on naked damaged DNA substrates [3]. Hence, UV-DDB was considered to be an accessory factor in the recognition of photoproducts in vitro. However, in vitro studies provided evidence that UV-DDB efficiently recognizes certain types of damage such as cyclobutane pyrimidine dimers (CPD) [17, 18], compared to the global damage sensor XPC-HR23B-Centrin2 ('XPC') [19, 20]. This in vitro finding was recapitulated in vivo, where UV-DDB was found to greatly stimulate the repair of these lesions in cells [21, 22]. Further, it was found that the repair of lesions in nucleosomal DNA is greatly inhibited in the absence of UV-DDB and that this repair factor is not dispensable for the recognition of lesions in the context of chromatin [22, 23]. In vivo, DDB2 is found to be constitutively associated with the E3 ligase formed by DDB1, Cullin4A/B proteins and the E2 ligase RBX1/Roc1 ($CRL4^{DDB2}$) [24, 25]. DDB2 is considered to be a DDB-Cullin4 associated factor (DCAF) [24] specializing in damage recognition [16] and can be thought of as an adapter protein for targeting E3 ligase activity to sites of DNA repair (Figure 4.1) [25, 26, 27, 28, 29]. A survey of cell lines derived from XP-E patients has revealed that patients suffering from the XP-E phenotype exhibit defects in the DDB2 gene corresponding to mutations in the DNA binding interface, or result in a misfolded protein or premature stop codons [30].

The ubiquitination activity of $CRL4^{DDB2}$ is regulated by the Cop9 signalosome (CSN) [29, 25, 31]. Upon DNA damage binding, the CSN complex dissociates, enabling the NEDDylation of Cullin4, thus activating the E3 ligase [31, 25]. Activated $CRL4^{DDB2}$ ubiquitinates histones (H2A, H2B, H3 and H4) in the vicinity of the lesion, and this activity is required for chromatin relaxation [28, 27, 32]. This chromatin relaxation regulates access to the downstream repair factors such as the XPC complex [33]. In addition, XPC has been found to be a substrate for ubiquitination of $CRL4^{DDB2}$ [29]. This post-translational modification of XPC leads to an enhanced affinity for the lesion [29]. As final substrates, lysines in the N-terminus of DDB2 are auto-ubiquitinated resulting in a degradation signal for DDB2 completing the hand-off of the lesion from UV-DDB to XPC [29, 28, 26, 32, 25].

The XPC protein exists as a heterotrimer of XPC-hR23A/B-Centrin2 (XPC complex) [34, 35, 36]. In this complex, hR23A and hR23B play a functionally redundant role by enhancing XPC stability and protecting it against proteasomal degradation [37, 38, 34]. A recent report suggests that hR23A/B dissociates upon stable damage recognition by XPC in vivo [39]. Centrin2 has been shown to enhance the DNA binding activity of XPC to 6-4 PP in vitro and the binding of centrin2 leads to an enhanced rate of NER in vitro as well as in vivo [36]. The damage recognition activity of XPC is indispensable for NER, and patients suffering from molecular defects in XPC exhibit the xeroderma pigmentosum group C phenotype [40]. It is now clear that UV-DDB is the initiator of global genomic repair in mammalian cells, whereas, the XPC complex recognizes a wider set of damaged DNA substrates and functions as a global damage sensor [41]. Thus, between UV-DDB and XPC, these two damage recognition factors recognize a diverse set of structurally distinct bulky DNA adducts in DNA.

Damage recognized by upstream factors such as UV-DDB and XPC is verified by the basal transcription factor IIH (TFIIH) (Figure 1.2) - a ten subunit protein complex consisting of XPB, XPD, p62, p52, p44, p34 and p8/TTDA, which form the core complex, and the cdk-activating kinase (CAK) sub-complex, which consists of cdk7, MAT1 and cyclin H [42]. Mutations in TFIIH have been found in XPB, XPD and p8 and these have been linked to XP, XP/CS and TTD [43]. Disease causing mutations in the other subunits of TFIIH have not been discovered so far. This may be due to an extremely disruptive manifestation of these mutations leading to inviability because of the multiple known roles of TFIIH in repair as well as transcription. It is quite conceivable that pathogenic mutations in the other subunits which affect the molecular functions of TFIIH may be discovered in the future. TFIIH is the next factor to arrive at the site of damage and this recruitment is thought to occur through specific interactions of XPB and/or p62 subunits with XPC bound to a lesion [44, 45]. This interaction occurs via the C-terminus of XPC which is intrinsically disordered in the apo state [45, 46]. In addition to this previously described binding site of XPC, a new set of residues 76-115 have been identified for the Rad4/XPC protein which bind the PH domain of the Tfb1/p62 subunit of TFIIH [47]. This newly discovered interaction has implications in regulating the interactions with XPG and will be discussed later.

XPB is a 3'-5' helicase whose activity is modulated by the p52 subunit of TFIIH via a physical interaction between the two [48, 49, 50]. Interestingly, it has been demonstrated that its helicase activity is dispensible for NER, whereas its ATPase activity is not [51]. The ATPase activity of XPB is thought to induce a conformational change which promotes the loading of TFIIH at sites of damage [52]. The precise role of XPB in DNA melting during NER is not known, however, it is speculated that XPB acts as a wedge to keep the two DNA strands from re-annealing [51]. Defects in the ATPase activity of XPB helicase lead to the xeroderma pigmentosum group B phenotype [53].

TFIIH contains within it, another helicase, known as the XPD helicase [52]. XPD is a FeS containing 5'-3' helicase and is considered to be the active helicase in NER [54, 55]. Its helicase (and associated ATPase) activity is required to open up the DNA around a damaged site by about 25 nt [56, 51]. Emerging evidence points to its importance as a damage verification factor in NER and recent biochemical studies have revealed that XPD can discriminate the damaged strand from the undamaged strand and further, mutational studies have revealed the presence of residues which lack this ability [57]. Defects in XPD lead to the xeroderma pigmentosum group D phenotype, TTD and also result in combined XPD\Cockayne Syndrome [58]. XPD has been demonstrated to physically interact with CAK via MAT1 [59]. The helicase activity of XPD is highly regulated by protein-protein interactions with negative regulation by the CAK complex and positive regulation by its interactions with p44 [60, 61]. Mutations found in XP-D patients have been mapped to the p44 interacting domain of XPD and the ATPase domain. In vitro, these mutants were found to exhibit impeded DNA opening in the absence of ATPase (and hence helicase) activity and limited DNA opening for mutations in the p44 interacting domain [51].

Opening up of the helix by the XPD helicase in TFIIH is thought to generate repair intermediates which are synergistically stabilized by the XPA and RPA proteins (Figure 1.2) [59]. In this intermediate, XPA is thought to bind the kinked DNA [62]. The RPA proteins are single strand DNA binding proteins and are thought to stabilize the single strands generated by the unwinding action of TFIIH. Importantly, RPA recruitment is contingent to XPA recruitment to TFIIH [59]. The XPA protein has been demonstrated to bind residues in the N-terminus of XPC (156-325) [63]. The recruitment of XPA to TFIIH in turn, is dependent on successful helix unwinding by XPD, via interactions in its C-terminus and results in the dissociation of CAK from TFIIH [59, 64]. This dissociation of CAK acts as a switch that determines the participation of TFIIH in repair or transcription [59]. Successful damage recognition and verification leads to the formation of a pre-incision complex containing XPC, TFIIH (minus CAK), XPA and RPA (Figure 1.2).

Clearly, XPA is a central player mediating the various interactions in mammalian NER. Opening up of the lesion allows for recruitment of XPA which in turn recruits the endonucleases XPF-ERCC1 and XPG. In addition to interactions with XPC, RPA and TFIIH, XPA interacts with the structure specific, metal dependent nuclease (Mg^{2+}) XPF-ERCC1 which performs the 5' cut [65]. XPF is an obligate heterodimer with the ERCC1 protein and recognizes and incises various DNA structures including bubbles, flaps and loops [65, 66, 67, 68]. XPF-ERCC1 is known to interact with RPA and this interaction is thought to be responsible for the correct cellular localization of XPF in vivo [69, 70, 71]. Further, ERCC1 has been demonstrated to interact with XPA [72, 73, 74]. XPF-ERCC1 is thought to be recruited in an XPB dependent manner and binds the single strands in the open complex which are generated by unwinding [75, 76, 77].

The metal (Mg^{2+}) dependent, structure specific endonuclease XPG performs the 3' cut and has been demonstrated to incise 10-30 nt bubbles, splayed arms and release a mononucleotide from ssDNA in vitro [78, 75]. XPG interacts with RPA as well as with the XPB, XPD, p44, p62 in TFIIH [79, 80, 81]. Importantly, there is evidence that both Rad4/XPC and Rad2/XPG compete for binding to the Tfb1/p62 subunit of TFIIH in NER by occupying the same site on Tfb1 [47]. This finding provides evidence for the competition of XPC-HR23B and XPG for a common binding site on the PH domain of Tfb1, validating previous evidence which indicates that XPG competes off XPC-HR23B [82]. This suggests that recruitment of XPG to the open complex generated by TFIIH displaces XPC (Figure 1.2).

The incision by XPF-ERCC1 occurs after the DNA has been unwound by XPD [75]. In this function, XPG plays a structural role in promoting the 5' incision since, catalytically dead XPG is sufficient to induce incision by XPF-ERCC1, suggesting that the 5' incision occurs first [83, 84, 85]. On the other hand, recent research reveals that active XPF-ERCC1 is required for incision by XPG [84, 85]. Further, partial repair synthesis has been detected in the presence of catalytically dead XPG suggesting that repair synthesis triggers a conformational change in XPG which in turn induces its catalytic activity [85]. Since, incision by XPF-ERCC1 occurs in an XPB dependent manner [86], and XPG is recruited via interactions with TFIIH, it is likely that TFIIH dissociates after the 5' incision is performed by XPF-ERCC1 (Figure 1.2). It is known however, that recruitment of a minimal replication system consisting of RFC, PCNA and Polδ is dependent on the presence of catalytically active XPF-ERCC1 [85]. Polymerization by Polδ is believed to displace RPA and XPG from the site [86]. Finally, ligation is performed by ligase I and by ligaseIII-XRCC1 [86, 87] (Figure 1.2).

1.4 DAMAGE RECOGNITION AND VERIFICATION IN NER

The basic DNA transactions involved in NER broadly consist of the following steps: damage surveillance and recognition; damage verification; helix opening and stabilizing of the repair intermediates; dual incision of the DNA in the context of the lesion; repair synthesis and DNA ligation. At each step along this complex transformation and restoration of DNA, protein partners are recruited to the site of repair to perform these DNA transactions. Each of these transactions proceeds via combinations of highly regulated protein-protein and protein-DNA interactions that act upon specific substrate repair intermediates to transform them into a new set of product repair intermediates that can be recognized by subsequent players in the pathway. Here we examine some of the features of damage recognition and verification.

1.4.1 DNA damage influences the local dynamics and thermodynamics of DNA

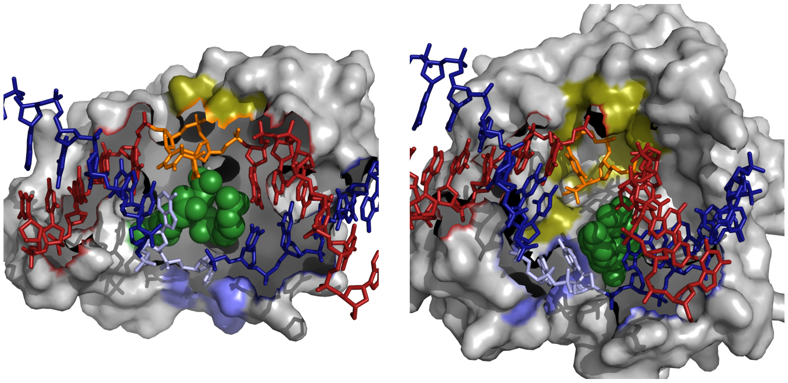
Chemical modification of the base pairs in DNA via the formation of bulky adducts results in the loss of base pairing information. These lesions lead to the formation of poorly paired or unpaired bases ('orphaned bases') in DNA. Modified bases in DNA may be extra- or intra-helical in nature arising from a reconfiguration of the local structure and dynamics of the DNA in the context of the lesion. For example, bulky polycyclic aromatic hydrocarbon (PAH) lesions intercalate in the DNA helix displacing the orphaned base and subsequently altering its dynamics and thermodynamics [88]. Further, the dynamics of the lesion are also affected by the specific sequence of the context in which the lesion is formed [89]. Indeed, the energy minimized rearrangement of the lesion in the context of the DNA influences both the local dynamics as well as the local thermodynamics of the DNA. This is achieved by maximizing favorable interactions (hydrogen bonding, van der Waals forces, aromatic stacking) of the lesion with the neighboring bases at the cost of unfavorable interactions (disruption of both hydrogen bonding and aromatic stacking of the orphaned bases via base eviction or distortion of the double helix). Experimental evidence has revealed that the repair of structurally diverse lesions proceeds at different rates and this is due to the differential detectability of the lesions by NER - a phenomenon which is crucially dependent on the structure and dynamics of the lesion in the context of the undamaged DNA [88, 90, 91].

1.4.2 Base-flipping is a universal mechanism in damage recognition

The impairment of Watson-Crick pairing may result in a greater ability of the orphaned base or the lesion to be flipped out. The base flipping transition involves two regimes, a regime in which the penalty of disrupting the Watson-Crick pairing rises quadratically around the mean position of the base in the intrahelical conformation until about a displacement by 25° on either side, and a second regime in which the energy rises linearly [92]. The energetic costs of flipping undamaged adenine (A), thymine (T), guanine (G) and cytosine (C) are estimated to lie in the range of 15-20 kcal mol^{-1} [93, 94]. Alteration of base chemistry results in perturbations in its base flipping energy. In contrast, the energetic cost of flipping a cis, syn-cyclobutane pyrimidine dimer (CPD) is estimated to be 6.25 kcal mol^{-1} [94]. Further, this estimated value was found to be sensitive to the sequence context with a lower ΔG_{flip} for the CPD in the context of A or T sequences by about 0.5-2 kcal mol^{-1} [95]. Similarly, the energetic cost of flipping another NER substrate 14R (+)-trans-anti-dibenzo[a,l]pyrine-N2-dG (14R-dG*) was found to be 10.4 kcal mol^{-1}, about 7.7 kcal mol^{-1} lower than the corresponding undamaged DNA [96]. It is important to note that even with these lower penalties of base flipping, neither the damaged nor the orphaned bases are predominantly extrahelical in nature. These values merely indicate that damaged bases have a greater propensity to make extrahelical excursions compared to undamaged bases, in a lesion dependent manner. The destabilization of the damaged and orphaned bases arising from the lower energetic penalty of base flipping of damaged bases may be exploited by damage recognition enzymes to recognize damage with high specificity. Given the significantly lower energetic barrier to flipping damaged bases, a first test for damage might involve the probing of the Watson-Crick base pairing (or lack thereof) between the bases being tested in the search for damage. Structural features on the protein which act as the probe and evict the damaged bases could interact with undamaged bases in the context of the DNA, resulting in the lowering of the barrier for base-flipping so that it would be accessible to thermal fluctuations in the presence of the repair factor. Damage recognition might involve probing the DNA for deformability via the energetically unfavorable transition of base flipping and this may represent a test for the presence of damage. Such a mechanism would confer damage specificity, enabling the recognition of rare damaged bases in a vast majority of undamaged bases upon rapid sampling. Indeed examination of the DNA bound states of various photoproduct recognizing repair factors reveals that base flipping is a common theme for the recognition of UV-induced photoproducts and likely other bulky lesions (see table 1.1).

1.4.3 Damage recognition factors utilize structural features to sense alterations in DNA dynamics

If damage recognition occurs spontaneously, the energetic cost of displacing the lesion must be compensated by enthalpic interactions which release energy upon binding or increase the entropy of the system upon binding. How might these constraints be met?

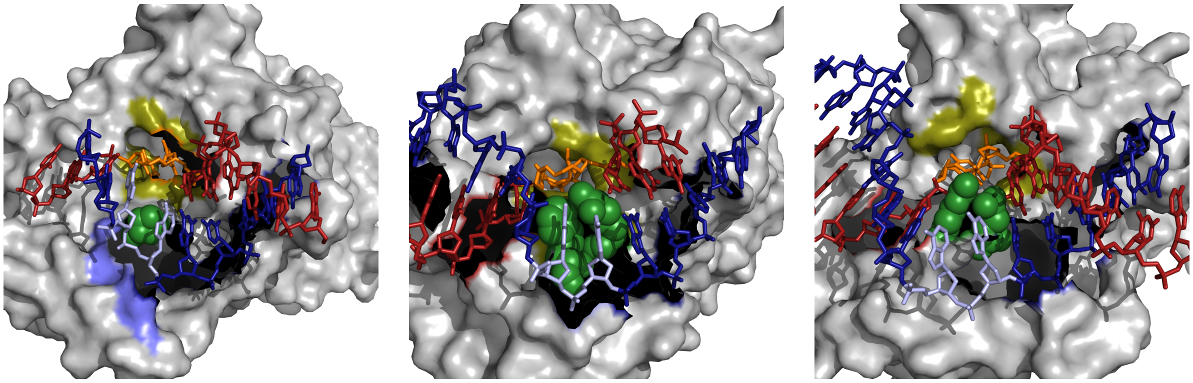


(a) *T4 bacteriophage* EndoV (b) *S. acidocaldarus* UVDE

Figure 1.3: UV-endonucleases

Two possible solutions involve providing enthalpic stabilization by providing interactions to the displaced lesion or to the orphaned bases. Thus, successful damage recognition would be bipartite: in the first step, the lesion is displaced to an extrahelical conformation and in the second step, the orphaned base or the lesion is stabilized via interactions with the protein.

An examination of crystal structures of damage recognition enzymes reveals that these ideas outlined above are general features of photoproduct recognition and variations on the combinations of these are present in nature. Table 1.1 represents a non-exhaustive, but illustrative survey of different classes of DNA repair proteins that recognize UV induced photoproducts as part of their biological functions.



(a) *A. nidulans* CPD Photolyase

(b) *M. mazei* CPD Photolyase

(c) *D. melanogaster* 6-4 Photolyase

Figure 1.4: Photolyases

Table 1.1: Survey of structural features in proteins recognizing UV-induced damage

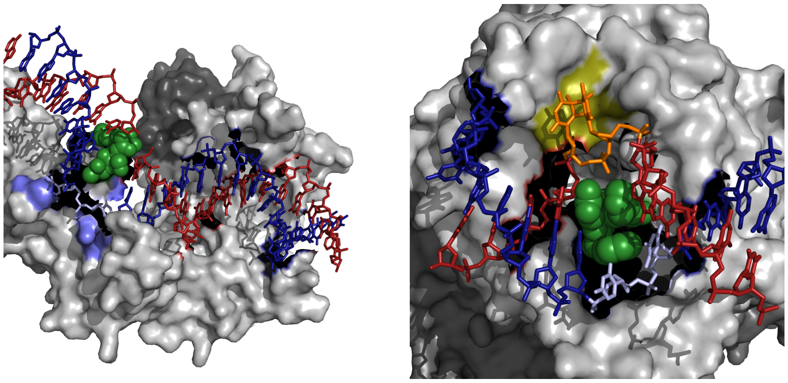
Repair factor	Activity	Organism	Lesion	Probe	Lesion Pocket	Orphaned base pocket	PDB	Ref
T4 endoV	endonuclease	T4 bacteriophage	Thymine dimer	R22, Q23, R26	K121, R117	Y21,D87, T89, Q91	1VAS	[97]
UVDE	endonuclease	Sulfolobus acidocaldarius	6-4 PP	Q103, Y104	K271, S7, K269, S60, R57, E266, H243,	L65, S67, H68, E121	4GLE	[98]
CPD photolyase	lyase	Anacystis nidulans	split TT dimer	P402	M353, N349, W392, E283, R232, W286	Y468, L403, K472	1TEZ	[99]
CPD photolyase	lyase	Methanosarcina mazei	CPD	R429, W431	M379, W305, W421, E301, D428, R441	None	2XRZ	[100]
6-4 photolyase	lyase	Drosophila melanogaster	6-4 PP	Q418, R421	Q299, W302, W409, P247, V294	None	3CVU	[101]
Rad4	damage recognition	Saccharomyces cerevisiae	CPD	E600, R601, G602, S603, T604	none	Q495, M498, R494,F556, V594, F597, F599	2QSG	[41]
UV-DDB	damage recognition	Homo sapiens DDB1, Danio rario DDB2	CPD	F371, Q372, H373	W239, I213, G192, R214	None	4A08	[25]

1.4.4 Combinations of structural features determine biological function

The substrate affinity, specificity and repertoire for DNA damage recognition enzymes arise from the steric constraints imposed by the lesion recognition pocket. Indeed, from the examples provided here (see table 1.1), it is evident that the lesion binding pocket has evolved specifically to recognize certain substrates and these direct the substrate specificity of the enzyme. Highly functionally specialized enzymes such as CPD or 6-4 photolyases demonstrate a high specificity for one substrate over another that arises from sterically selective interactions in the lesion binding pockets [98, 99, 100, 101]. Broader substrate specificity is obtained by more permissive lesion binding pockets. For example, the UV-DDB protein exhibits a lesion binding pocket that accommodates 6-4PP and CPD lesions, suggesting that this repair factor has evolved for the specific recognition of these lesions [25, 102, 16]. Other damage recognition factors possess even broader substrate specificity. The global damage recognition factor, Rad4 (orthologue of the human XPC protein) exhibits a binding pocket for the orphaned base while exhibiting no apparent binding pocket for the lesion [41]. This lack of a binding pocket for the lesion explains the broad substrate repertoire of Rad4 and the related XPC protein [103, 104, 29], while conferring the properties of a sensor of altered DNA dynamics arising from DNA damage.

1.4.5 Successful damage recognition generates a repair intermediate with a conserved topology and reduced conformational entropy

The NER pathway repairs structurally and dynamically heterogeneous lesions in DNA. DNA damage surveillance factors must recognize a wide variety of initial configurations of the DNA and manipulate them to generate a product repair intermediate that has little dependence on the exact chemical nature of the lesion. In this process, the repair machinery must modulate and accomodate both the structure as well as dynamics of the DNA containing the lesion to create this product repair intermediate. This product repair intermediate must be recognizable by downstream repair machinery so that manipulations may be performed on the damaged DNA with a minimal requirement for damage re-recognition or re-verification. As discussed above, the substrate repertoire is determined by the specific configurations of the



(a) *S. cerevisiae* Rad4-Rad23 (b) hsDDB1-drDDB2

Figure 1.5: NER damage recognition factors

lesion binding pocket. From the examples presented in table 1.1 and figures 1.3, 1.4 and 1.5, a remarkable convergence of the topology of the repair intermediate generated upon repair factor binding emerges - one in which the helix is probed by the factor, and the lesion or the orphaned bases or both are stabilized on the binding surface. In this process of binding, the repair factor converts highly dynamic damaged DNA to a highly stable DNA intermediate, reducing its conformational entropy in this process. This reduction in conformational entropy may be a pre-requisite for successful damage processing by downstream factors.

1.4.6 Intermediates generated by damage recognition may not be substrates for the pathway

It is important to note that this process of recognition discriminates and selects for DNA damage that may be stabilized in this manner. The set of substrates repaired by the NER machinery may be smaller than the set of substrates recognized by the repair factors and smaller yet than the set of bulky lesions formed in DNA. For example, DDB2 has been demonstrated to bind a variety of DNA structures including 6-4 photoproducts, abasic sites and two base mismatches with extremely high affinity and CPD lesions and cisplatin adducts with relatively lower affinity [105, 18], however, the repair of abasic sites and mismatches is not known to proceed via NER. In this first step, the recognition factors falsely identify other lesions that can be stabilized in the configuration of the intermediate recognized by subsequent NER machinery. On the other hand, there may exist a class of bulky lesions which might appear to be canonical NER substrates at first glance, but escape detection since these lesions may fail to participate in the recognition reaction or yield an unstable repair intermediate resulting from the binding of the repair factor. The efficiency of repair of lesions is highly determined by the lesion context, chemistry and stereochemistry as well as downstream events as observed [106].

1.4.7 Repair intermediates are further processed to verify damage

Given that the recognition factors generate repair intermediates that can be stabilized by intrahelical probing and lesion or orphaned base stabilization, the end products of this reaction are not guaranteed substrates for downstream processing in NER. This necessitates the requirement for a damage verification enzyme. The NER machinery must then transform the repair intermediate containing a candidate NER substrate to a repair intermediate containing a verified NER substrate. In the next step following damage recognition, the machinery must distinguish the damaged strand from the undamaged strand and identify the specific position of the lesion. Further, the repair machinery must interact with the lesion in a non-specific manner, so as to be able to process a diverse substrate repertoire. In the absence of structural evidence for the intermediates, identification of general structural principles is difficult; however, hints for the mechanism of action are available from genetic, molecular biology, biochemical and single molecule biophysical studies.

Following damage recognition by the XPC complex, damage verification is performed by TFIIH. TFIIH is recruited to the sites of damage via specific interactions with the XPB subunit. This finding suggests that loading of TFIIH at sites of damage occurs in a defined manner which might explain the ability of the complex to distinguish the non-damaged from the damaged strand. Binding of TFIIH also promotes the binding of XPA to the sites of damage. XPA is known to bind highly distorted DNA structures and perhaps the synergistic binding of XPA and RPA to the open complex generated by unwinding of the double helix by XPD might serve to form a damage verification intermediate. Since NER is known to progress via the sequential assembly of repair factors, this intermediate may function to ensure that all the protein partners are in conformations that enable the complex to progress along the correct reaction trajectory. Available data from single molecule studies of the helicase activity of XPD indicate that XPD is a poorly processive helicase with a step size of 1 bp, with frequent backtracking due to re-annealing [54]. This finding is consistent with its biological role where it unwinds short tracks of dsDNA in the context of lesions. However, these data must be interpreted cautiously since the biological function of XPD is manifested in the context of the other subunits of TFIIH, which are known to modulate its processivity. Biochemical evidence is available demonstrating that the XPD helicase stalls upon encountering CPD lesions in its ATP dependent translocation [107]. Further, mutants of XPD have also been identified which fail to stall at sites of CPD lesions, suggesting the presence of structural features on XPD which recognize damage in DNA [57]. The emerging model suggests that damage verification in NER, occurs in an ATP dependent manner by a helicase which is the ultimate gatekeeper of competency of the damage for NER. The roles of the other subunits of TFIIH and the interplay between them during damage verification by XPD remain largely un-elucidated.

1.5 PROOFREADING MECHANISMS FOR LESION RECOGNITION AND VERIFICATION

Damage recognition factors are faced with the task of rapid and accurate recognition of damage. In this process, recognition factors bind DNA and upon successful recognition, the complex transits to a stable intermediate in a process that requires the crossing of energy barriers. While some of these barriers may be overcome by the enthalpy of binding, in some steps, the energy of ATP hydrolysis may be required to successfully navigate the free energy landscape along the reaction coordinate. However, examination of the various DNA repair pathways reveals a general theme (not without exceptions) that damage recognition can occur in the absence of external energetic inputs, whereas damage verification relies on additional energy for accurate processing. Keeping with these differences, two mechanisms of proof-reading have been described, including kinetic proofreading [108] and conformational proofreading [109] for molecular recognition with high specificity. How is specificity achieved in molecular recognition in these two distinct mechanisms?

1.5.1 Kinetic proofreading

Kinetic proofreading was originally proposed to explain the low error rates of polymerases in replication. Systems employing kinetic proofreading are proposed to have three distinct features:

1. They exhibit a low specificity for substrates
2. An essentially irreversible reaction involving the conversion of a high energy cofactor to a low energy product is used to generate a high energy intermediate. Importantly, in the absence of this external energetic input, the high energy intermediate is rarely formed.
3. A branched enzymatic pathway converts the high energy intermediate with the correct substrate to the correct product along one pathway, and rejects the intermediate to generate the original enzyme configuration and a rejected, modified substrate.

Such a proofreading mechanism has been proposed to explain the ability of the NER pathway to utilize ATP hydrolysis to process a broad set of substrates which are recognized with low specificity by the damage recognition factors [110].

1.5.2 Conformational proofreading

An alternate mechanism for discrimination of substrates with high specificity in the absence of energy utilization has been proposed, termed as conformational proofreading [109]. In this mechanism, instead of a high energy intermediate, a structural mismatch is proposed to exist between the bound state and the unbound state for the enzyme. In this case, only the correct substrate is able to induce a conformational transition between the unbound state and a highly stabilized bound state. Binding of the incorrect substrate does not invoke the structural transition to the same stabilized state; enhancing the substrate specificity of the recognition factor. Conformational proofreading has been proposed to explain the specific recognition of cognate sequences by transcription factors and has been experimentally validated for homologous recombination where the dsDNA is proposed to be the recognizer [109, 111].

Conformational proofreading is an attractive alternate explanation for molecular recognition in the absence of energy inputs in the initial step of damage recognition. However, to date, conformational proofreading has not been proposed as a mechanism for damage recognition with high specificity in the binding of damage recognition factors to DNA damage in the initial stages of NER.

Here, we hypothesize that conformational proofreading is the predominant mechanism for recognition of damage in NER. Further, we propose that a complex interplay of conformational as well as kinetic proofreading is required to process NER competent substrates. While apparently operative in molecular recognition in homologous recombination, whether conformational proofreading is a universal mechanism in damage discrimination in other DNA repair pathways remains to be seen.

1.6 THE SUBSTRATE SEARCH PROBLEM

DNA binding proteins that bind specific targets (sequences or structures) in DNA must discriminate their cognate substrates in a noisy environment constituted of a large number of mis-substrates that resemble the cognate substrate. The search problem for DNA binding proteins locating their targets is conceptually very simply stated as follows: *what transport mechanisms do DNA binding proteins utilize to identify their targets in DNA?*

Various transport mechanisms have been proposed involving combinations of three dimensional diffusion and/or ‘jumping’ (see figure 1.6 a) based iterative cycles of binding and dissociation and one-dimensional sliding mechanisms [112]. Within the one dimensional sliding mechanism, there are several possible modes of facilitated diffusion including, hopping (figure 1.6 b), directed motion (figure 1.6 c), simple linear diffusion or linear diffusion coupled to rotational diffusion (figure 1.6 d), and inter-segmental transfer (figure 1.6 e). Not all of

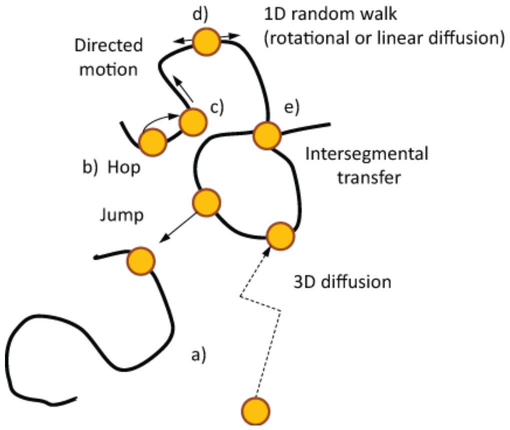


Figure 1.6: Transport mechanisms involved in target site location

these modes of motion may be accessible to all DNA binding proteins and this accessibility of search modes is influenced by protein structure. For instance, only proteins with ATPase domains may be able to participate in an energy driven directed motion, and only those proteins containing multiple DNA binding domains may perform inter-segmental transfer. Irrespective of the transport mechanism, proofreading is an indispensible aspect of the search mechanism. Efficient search relies on cycles of transport and proofreading.

1.7 OBJECTIVES OF THIS WORK

In recent years, the core set of NER factors which orchestrate repair has been identified and many features of repair have been reconstituted in vitro. Even with this understanding of interacting partners and a broad understanding of the roles of these factors, our understanding of NER remains incomplete. Much of the progress has arisen from investigation of this pathway using genetic, molecular biology, biochemical and cell biology approaches. However, our understanding of the heterogeneity of the dynamics, sequence and determinants (presence of protein partners, conformational changes and post-translational modification) of these protein-protein and protein-DNA interactions using targeted structure-function approaches at the single molecule level has been limited owing to tremendous challenges in purification of functionally active recombinant NER proteins.

The elucidation of crystal structures and the amenability to recombinant expression and purification of the UV-DDB heterodimer have enhanced our understanding of the initial stages of damage recognition in NER. In vivo, as well as, in vitro studies have greatly clarified the role of UV-DDB in initiating global genomic nucleotide excision repair. However, even as our understanding of the network of interactions of UV-DDB and its protein partners increases, fundamental questions regarding damage surveillance and substrate recognition remain. How does UV-DDB search for damage on DNA? How does it proofread DNA for damage? What are the structure-function relationships that govern this proofreading? What is the oligomeric state of UV-DDB in damage recognition? What is the basis for the specificity of UV-DDB for UV-damage in DNA?

In this work, we wanted to examine the interactions of UV-DDB with undamaged and damaged DNA substrates using single molecule techniques in order to understand some of these fundamental questions pertaining to the structural and dynamical bases of damage recognition. To that end, the objectives of this work were three fold:

1. Implement assays and algorithms to observe and analyze single particle tracking data to understand the diffusive and binding behaviors of single labeled molecules
2. To develop biological reagents and identify strategies to observe site specific interactions of quantum dot conjugated proteins at sites of repair substrates
3. To implement these methods to study the real-time interactions of the damage recognition enzyme UV-DDB in order to elucidate the kinetics of the initial steps of damage recognition

In the following chapters, we will treat these objectives sequentially. Chapter 2 will describe experimental methods used in the study. Chapter 3 will discuss the relevant theory and methods for analysis of data arising from single particle tracking experiments. Chapter 4 will describe the study of UV-DDB binding to various damaged DNA substrates. In Chapter 5 we will summarize the results of this study and discuss avenues for further investigation.

2.0 MATERIALS AND METHODS

2.1 CHARACTERIZATION OF DNA BINDING ACTIVITY

2.1.1 Electrophoretic mobility assays

2.1.1.1 Binding conditions

UV-DDB binding was performed by incubating the indicated amounts of UV-DDB with indicated DNA substrate for 20 min at RT. Samples were diluted in loading buffer containing glycerol prior to separation using native gels.

2.1.1.2 Agarose gel electrophoresis

Binding reactions were separated over a 1 % agarose gel for 50 min at 100 V on ice. Gels were dried for 1 h at 65 °C.

Abasic site substrates

36 mer DNA substrates for EMSA containing a defined lesion were created by hybridizing the top strand (AP36: 5'- Phos CCG AGT CAT TCC TGC AGC G/idSp/G TCC ATG GGA GTC AAA - 3' or UD36: 5' - Phos CCG AGT CAT TCC TGC AGC GAG TCC ATG GGA GTC AAA - 3') with its ^{32}P radiolabeled reverse complement. Hybridization was performed in 1X hybridization buffer at 1μM concentration of each strand.

2.1.2 Pulldown experiment

2.1.2.1 Preparation of DNA substrates

DNA substrates for pulldown experiment were designed with modifications on both ends. 517 bp DNA was created by amplifying pUC18 using the Dig1890A (5' - /5DigN/ GGT CTG

ACG CTC AGT GGA ACG - 3') and Bio1373s (5' - GGA ACC GGA GCT GAA TGA AG - 3') primers to create DNA substrate which can be blocked on either end with streptavidin or anti-digoxigenin antibody. PCR was performed using Pfu ultra with 10 ng of pUC18. Cycle conditions for PCR were: 94 °C for 5 min (hot start, add pol after 2 min), 94 °C for 30 s, 57 °C for 1 min, 72 °C for 1 min (30 cycles) and 72 °C for 7 min. PCR product was separated on a 1 % agarose gel to identify homogeneity of amplification. PCR reaction was then purified using Qiaquick PCR purification kit (Qiagen, cat. # 28104). Purified DNA was eluted in tris-Cl pH 8.0 and concentrations of DNA were measured using a Nanodrop. Typical yields for purification were found to be between 50 - 60 %.

2.1.2.2 UV-irradiation of 517 bp DNA

UV-damaged DNA was obtained by exposing the purified DNA to a dose of 20 Jm^{-2} at 50 $\mathrm{ng}\mu\mathrm{L}^{-1}$. DNA damage in 517 bp DNA was quantified using QPCR and was found to occur at a frequency of 1.2 polymerase blocking lesions per 517 bp fragment from two experiments (15 and 16 cycles).

2.1.2.3 Conjugation to beads

Bio-517-dig DNA (either mock or UV- irradiated) was then incubated with streptavidin coated magnetic beads in the presence (4X excess over DNA) or absence of anti-digoxigenin antibody (Roche, cat. # 11 333 062 910) to get undamaged or UV-damaged DNA bound to beads with either one free end or no free ends. Blocking was performed for 3 overnights at 4°C. WT UV-DDB or K244E was incubated with DNA bound on beads in 1X binding buffer for 30 min at RT in a reaction volume of 20 μL. In these experiments, the final concentration of the UV-DDB (WT or K244E mutant) was either 5 nM or 10 nM (as indicated) and that of DNA on beads was 12.5 nM in a total reaction volume of 20 μL. After binding, the beads were washed with 3 volumes and resuspended in 21 μL of 1X binding buffer and 7 μL of 4X LDS buffer. Samples were boiled for 10 min at 95°C and separated on a 4-12 % SDS gradient gel. Gel was subsequently transferred and Western blotting was performed on the various fractions which were probed with α-DDB1 and α-DDB2 antibody.

2.2 ATOMIC FORCE MICROSCOPY

2.2.1 Sample preparation

2.2.1.1 Proteins

Protein only samples were deposited on a freshly cleaved mica surface at concentrations in the range from 20-40 nM in 1X AFM deposition buffer (25 mM NaOAc, 10 mM $Mg(OAc)_2$ and 25 mM Hepes pH 7.5).

2.2.1.2 DNA

DNA substrates in the range of 450-600 bp were deposited under conditions of 0.5-2 nM in AFM deposition buffer on freshly cleaved mica surfaces.

2.2.1.3 Protein-DNA samples

50 nM WT or K244E UV-DDB was incubated with 25 nM DNA substrate for 10 minutes at 37°C in 1X binding buffer - glycerol -DTT. Samples were diluted 1:5 or 1:10 fold prior to deposition.

2.2.1.4 Deposition

Diluted samples were deposited on freshly cleaved mica (SPI Supply, West Chester, PA) followed by washing with MilliQ water and drying under a stream of nitrogen gas.

2.2.2 Data acquisition

AFM images were collected using a MultiModeV microscope (Bruker, Santa Barbara, CA) using an E scanner in tapping mode in air. Pointprobe plus non-contact/tapping mode silicon probes (PPP-NCL, Agilent) with spring constants of approximately 50 Nm^{-1} and resonance frequencies in the range from 150 - 200 kHz were used. Images were captured at a scan size of 1 μM × 1 μM with a resolution of 512 × 512 pix^2 at a scan rate of 4 Hz and a target amplitude of 300 mV.

2.2.3 Volume analysis

AFM volumes of proteins deposited on the surface were calculated as $V = S \times (H - B)$ where, S is the cross sectional area of the protein at a height H over the mean background B. Images were analyzed in Image SXM in order to obtain the values of S, H and B. Histograms of AFM volumes for a large number of particles were plotted and peaks were fit to Gaussians to identify the mean AFM volume. Mean AFM volumes and standard error of the mean were obtained from three depositions.

2.2.4 Calibration curve

A plot of the mean AFM volumes (nm^3) of the following proteins and their molecular weight in kDa {Pot1 (65 kDa), PcrA monomer (86.4 kDa), UvrA monomer (105 kDa), Taq MutS dimer (181 kDa), UvrA dimer (210 kDa) and Taq MutS tetramer (362 kDa)} was plotted and a linear relationship was found to best describe the data (see appendix B and Figure 2.1). The fit to the data is: $V(nm^3) = 1.471MW(kDa) - 7.294(Rsq = 0.9886)$.

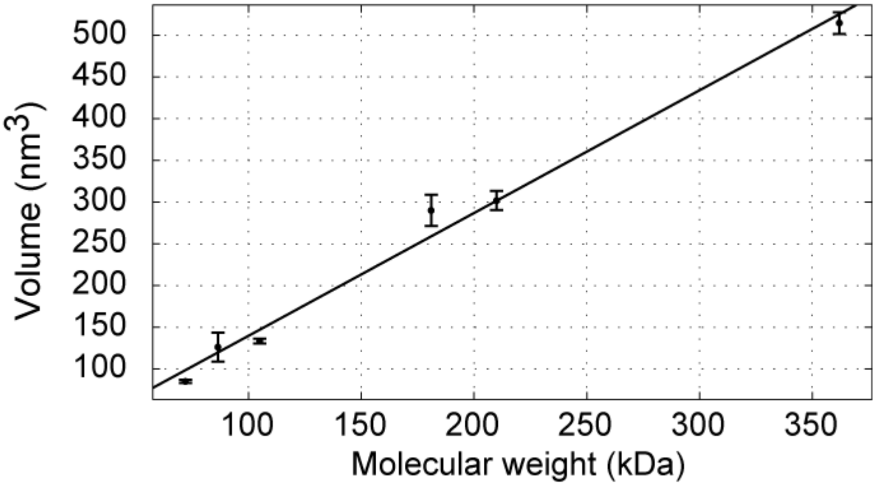


Figure 2.1: AFM calibration curve

2.3 DNA TIGHTROPE ASSAY

2.3.1 DNA substrates for tightrope assay

2.3.1.1 Undamaged DNA

λ-DNA (New England Biolabs, Catalog # N3013L N^6-methyladenine-free) at 500 μgmL^{-1} was diluted 10 fold in TE buffer (DNase, RNase and Protease Free, 10 mM Tris pH 8.0 ± 0.1, EDTA 1 mM, Fisher Scientific Catalog # BP2473-500) and its concentration was verified after dilution by measurement using a Nanodrop. Efficient stringups were obtained when 1 μg of λ-DNA (corresponding to 20 μL of 50 ngμL^{-1} DNA) was used.

2.3.1.2 UV-damaged DNA

In order to study binding of UV-DDB to long DNA substrates in the DNA tight rope assay two strategies were employed. In the first strategy, undamaged λ-DNA tightropes were elongated between beads in the flowcell and this was followed by exposure to UV-C irradiation. This was followed by examination of binding of CPD antibody conjugated to IgG coated QD. However, after several unsuccessful attempts to detect binding of CPD antibody, an alternate approach was employed to create UV-damaged DNA tightropes.

In this approach, a UV-C lamp was used to generate randomly UV-damaged λ-DNA. UV-induced damage in DNA was quantified by qPCR essentially as previously described [113], using the GeneAmp XL PCR Kit (Applied Biosystems cat. # N8080193). A 12.5kb fragment of λ-DNA was amplified between nucleotides 26,890 and 39,488 (Forward primer: 5 CCA ACC ATC TGC TCG TAG GAA TGC 3; reverse primer: 5AGT TGG GTC CAC TTA TCG CGG AGT 3). Cycle conditions for amplification of 15 ng of λ-DNA template were: 1:30 s for 75°C, followed by addition of polymerase, 94°C for 1:00 min, 94°C for 15s, 64°C for 12 min (11 cycles), final extension: 72°C 10 min. Amplification of damaged DNA was measured relative to mock irradiated λ-DNA that was treated similarly. Final product was visualized by gel electrophoresis with ethidium bromide staining. PCR product concentrations were measured using PicoGreen fluorescent DNA binding dye (Molecular Probes, Quant-iT PicoGreen dsDNA assay kit, cat. # P7581). 3μL of PCR product were diluted in a 1:20 dilution of PicoGreen in TE and relative fluorescence units were calculated for each sample. Further, these samples were converted to an absolute DNA concentration by using a standard curve created from measuring the relative fluorescence units for different concentrations of λ-HindIII standards. 50% controls were performed with undamaged λ-DNA. A dose response curve relating the number of UV lesions in the λ-DNA fragment as a function of UV-dose was obtained as: Lesions/λ-DNA = 1.0958 UV-dose(Jm^{-2}) 0.0136; Rsq = 0.9997, n = 2 with each experiment performed in duplicate.

2.3.1.3 Defined lesion substrates for tightrope assay

Oligonucleotides with custom chemical modification were introduced into the pSCW01 plasmid essentially as described before [114]. pSCW01 was amplified in DH5α and purified using the Qiafilter Maxiprep kit. Typically, 400 μg of pSCW01 were incubated with 60 μL^{-1} of Nt.BstNbI (New England Biolabs, 10 $U\mu L^{-1}$) for 4h at 55°C in the presence of 100X complementary displacer oligonucleotides to nick the plasmid at the four adjacent nickase sites. Following this the reaction was inactivated by heating at 85°C for 10 minutes followed by annealing to displace the nicked oligonucleotides. An equal volume of the reaction mixture was then added to a 2X solution containing 26% PEG-8000 in 20 mM $MgCl_2$ followed by centrifugation for 1 h at 4°C. Precipitated DNA pellet was then washed with ethanol followed by resuspension of the gapped plasmid. Purified gapped plasmid DNA was then incubated with the desired oligo in 3X excess followed by annealing in NEB4 buffer. After annealing, the reaction was supplanted with 8 mM ATP and ligation was performed by the addition of 5 μL of T4 DNA ligase (New England Biolabs, M0202, 2000 U μL^{-1}) to a 400 μL reaction containing 400 ng μL^{-1} of purified gapped plasmid DNA for 18 h at 16°C. The reaction was then heated to 65°C for 20 mins to inactivate the T4 DNA ligase and supplemented with XhoI (New England Biolabs) to digest the plasmid DNA containing custom oligonucleotide. Restriction was performed for 2 h at 37°C followed by inactivation at 85°C for 20 minutes. Linearized plasmid DNA containing custom chemical modification (monomer) was then stored at -20°C until further use. To obtain long DNA substrates for the DNA tight rope assay, 1μg of DNA monomers was ligated for 15 mins at RT in Quickligase buffer with 2 μL of T4 DNA ligase (2000 U μL^{-1}). Using this strategy we were able to make DNA substrates with a tetrahydrofuran moeity (abasic site analog) by inserting the following oligonucleotide:

/5Phos/ CCG AGT CAT TCC TGC AGC G/idSp/G TCC ATG GGA GTC AAA /3BiodT/

2.3.2 Conjugation strategies for QD conjugation to UV-DDB

2.3.2.1 Using the BT-NTA compound

SA QD were incubated with the biotinylated nitrilotriaceticacid (BT-NTA, Jacob Piehler) compound in 1X binding buffer (20 mM Hepes 7.5, 150 mM NaCl, 2 mM BME, 5% glycerol and 1 mM MgCl2) in a molar ratio of 1:2 for 30 min at RT to obtain the BT-NTA-QD conjugate. ^{FLAGHis}DDB1-DDB2 was incubated with the BT-NTA-QD conjugate to obtain a final ratio of 1:2:1 corresponding to UV-DDB:BT-NTA:QD for an additional 30 min at RT.

2.3.2.2 Using the penta-His biotin conjugate

The strategy for conjugating the penta-His-biotin conjugate (His-Ab, Qiagen, cat # 34440) was adapted from a previously published strategy. His-Ab was incubated with SA QD in the molar ratio of 5:1 for at least 30 min at RT. 2 μL of the His-Ab were incubated with 2 μL of SA QD for an additional 30 min at RT to obtain a final molar ratio of 1:5:1 corresponding to UV-DDB:His-Ab:QD.

2.3.3 Flow cells for DNA tightrope assay

Flow cells were constructed essentially as described before [115], but with changes described here. Cleaned glass slides (Thermo scientific, precleaned 25×75×1mm microscope slides, cat. # 4951-001) were drilled with two holes, 15 mm apart using 1.25 mm Cylinder Diamond drill bit (Dads rock shop). Inlets and outlets were created by passing PFA tubing (Idex, Tub PFA Nat 1/16×0.03 ×5ft, cat # 1502) which was glued in place using epoxy (BONDiT B-45TH) followed by curing for 36-48 h. Prior to gluing, the ends of the tubing were sanded using a sanding sponge (3M PRO GRADE, Sanding sponge) to create a greater surface area to enable efficient adhesion. Following tubing affixation, the excess tubing was cut off using a scalpel. The flowcell was assembled by using a double sided sticky gasket (created from

3M double sided sticky tape, cat #. B1079). The gasket consists of a rectangular ring of outer dimension: 0.775”×1.195” with an inner hole of dimension: 0.395”×0.77”). The flow chamber is assembled by adhering a treated cover slip (Corning, cat. # 2940-244, 24×40 mm). The inlet and outlet from the flow cells are connected to the syringe and reservoir with a union assembly (Idex, cat. # P-702) and flangeless ferrules (Idex, cat. # P-200X). Coverslips were treated with N-(2-Aminoethyl)-3-aminopropyltrimethoxysilane (APTES; UCT, cat. # A0700-KG) followed by pegylation (Laysan Bio, Inc. cat. # MPEG-SVA-5000-1 gram).

2.3.3.1 Setting up the DNA tightropes

Flow cells constructed as described above were first passivated by incubating with a blocking buffer (50 mM Hepes 7.5, 50 mM NaCl, 1 mg mL^{-1} BSA(Roche cat. # 10 711 454 001). Blocking is performed for at least 10 minutes. Next, 5 μm poly-L-lysine (Wako Chemicals USA, Inc cat. # 163-19091) coated silica beads (Polysciences Inc., Uniform Silica Microspheres, cat. # 24332-15) are deposited in the flow cell on the inner surface of the cover slip. The flow chamber is then washed with water to get rid of excess poly-L-lysine followed by elongation of the DNA substrates in 1X imaging buffer (50 mM Hepes 7.5, 150 mM NaCl, 100 mM DTT, 10 mg mL^{-1} BSA) between these beads by a continuous to-and-fro flow for 30 min.

2.3.3.2 Visualizing the DNA

The DNA may be visualized by staining with YOYO-1 dye (Molecular Probes, YOYO-1, cat. # Y-3601). DNA tightropes were visualized by incubating the flowcell with up to 1 nM YOYO-1 dye in 1X imaging buffer.

2.3.4 Oblique angle fluorescence microscopy

Oblique angle fluorescence imaging: Oblique angle fluorescence imaging was performed using a Nikon Ti eclipse base with a Nikon 100X oil emersion TIRF objective with 1.45 NA. 488 nm light was used to excite the sample after passing through a 498 nm excitation filter. Emissions from quantum dots were separated using emission filters: 655 nm (640/20 or 700/75), 705 nm (700LP or 700/75), 605 nm (600/50), 585 nm (600/50) and 565 nm (535/50) and 520 nm using a 520/40 (see Figure 2.2). Images were acquired using Nikon Elements Ar 4.11.00 with a temporal resolution of 100 ms and a laser power of 1-2 mW using an Andor Neo sCMOS camera (DC-1520-C00-Fl).

2.3.5 Quantum dots for single molecule imaging

Quantum dots for fluorescence imaging are available from Invitrogen with two types of surface modifications including streptavidin and goat F(ab')$_2$ anti-mouse IgG Conjugate (H+L) conjugates. Streptavidin conjugated QDs are available in two formats 'classic' (c) or 'vivid' (v).

Quantum dots for fluorescence imaging possess several attractive properties. Among these are high extinction coefficients (and consequent brightnesses), a size based tunable emission spectrum, strong biocompatibility (inertness, non-toxicity, and solubility) and photoactivity under a wide range of biologically relevant conditions.

However, quantum dots suffer from stochastic excursions to 'off' states that are described by a power-law model [116]. This behavior is controlled by the presence of reducing agents such as DTT and appears to depend on the size of the QD [117]. Recent versions of commercially available (Invitrogen) Qdot Streptavidin conjugates have exhibited poor photostability, specifically demonstrating a blue-shift in the emission spectrum and photobleaching. Such spectral diffusion of QDs has been described previously [118] and was found to occur for certain QD conjugates in our hands, under certain conditions.

Biocompatible SA QDs offer multiple possibilities to develop orthogonal QD conjugation strategies, however, QDs suffer from problems of multivalency arising from a limited ability to estimate the number of accessible binding sites [119]. Achieving a homogenous population of 1:1 labeled proteins remains a challenge.

The Qdot525 streptavidin conjugate (Invitrogen, Q10141MP, QD525) was used for single color imaging with a 520/40 nm filter. QD525 was found to exhibit poor brightness and its use was limited in single color experiments (see Figure 2.2).

The Qdot565 streptavidin conjugate (Invitrogen, Q10131MP, QD565) was used for single color imaging with a 530/50 nm filter (see Figure 2.2). QD565 was found to exhibit high lifetime of 'on' states, good brightness. Photostability was not characterized.

The Qdot585 streptavidin conjugate (Invitrogen, 'vivid', Q10111MP, QD585) was used for single color imaging with a 600/50 nm filter (see Figure 2.2). QD585 was found to exhibit high lifetime of 'on' states, good brightness but poor photostability. This QD was found to photodim under mild imaging conditions and long acquisition times.

The QD605 streptavidin conjugate (Invitrogen, 'classic', Q10101MP, QD605) was used for single color imaging with a 600/50 nm filter (see Figure 2.2). QD605 was found to exhibit high lifetime of 'on' states, excellent brightness and photostability however, it suffered from slow photobrightening, making it unsuitable for studying lifetimes of highly transient species on DNA.

The Qdot655 streptavidin conjugate (Invitrogen, 'classic', Q10121MP, QD655) was used for single color imaging with a 700/75 nm filter (see Figure 2.2). QD655 was found to exhibit high lifetime of 'on' states, excellent brightness and photostability. This QD was found to be good for studying the lifetimes of highly transient interactions.

The QD705 streptavidin conjugate (Invitrogen, 'classic', Q10161MP, QD705) was used for single color imaging with a 700/75 nm or a 700 LP filter (see Figure 2.2). QD705 was found to exhibit poor lifetime of 'on' states, poor photostability in the absence of DTT but possessed excellent brightness. These properties make it unsuitable for studying lifetimes of highly transient species on DNA.

2.3.5.1 Choice of QDs for Single color imaging

For single color imaging either the QD655 or QD605 were used depending on the application. QD655 was used to study the transient behavior of proteins on DNA whereas, QD605 was used to mark the positions of introduced lesions in defined DNA substrates.

2.3.5.2 Choice of QDs for Multi-color imaging

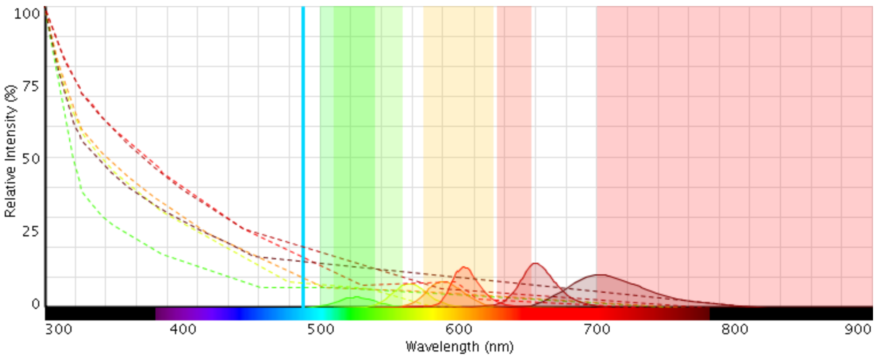


Figure 2.2: Normalized emission spectra for Invitrogen Qdot conjugates with overlay of emission filters and excitation source

Table 2.1: QD combinations for two color imaging

Sr. no	QD_1	$Filter_1$	QD_2	$Filter_2$	Notes
1	525 nm (c)	520/40	565 nm (c)	530/50	Avoid. Low SNR for 525 nm.

2	525 nm (c)	520/40	585 nm (v)	600/50	Avoid. Low SNR for 525 nm. Photoconversion and photobleaching of 585 under low (<1 mM) DTT and low brightness under high (>10 mM) DTT.
3	525 nm	520/40	605 nm	600/50	Avoid. Low SNR for 525 nm. Delayed photobrightening for 605.
4	525 nm	520/40	655 nm (c) or (v)	640/25	Avoid. Low SNR for 525 nm.
5	525 nm	520/40	705 nm	700 LP or 700/75	Avoid. Low SNR for 525 nm.

6	565 nm	530/50	585 nm (v)	600/50	Usable in the presence of 1-10 mM DTT.
7	565 nm	530/50	605 nm (c)	600/50	Usable in the presence of 1-100 mM DTT.
8	565 nm	530/50	655 nm (c)	640/25 or 700/75	Usable in the presence of 1-100 mM DTT.
9	565 nm	530/50	655 nm (v)	640/25 or 700/75	Usable in the presence of 1-10 mM DTT.
10	565 nm	530/50	705 nm (c)	700LP or 700/75	Usable in the presence of 1-100 mM DTT.
11	585 nm (v) or 605 (c)	600/50	655 (c) nm	700/75	Usable in the presence of 1-100 mM DTT. Ensure no spectral overlap.

12	585 nm (v) or 605 nm (c)	600/50	705 nm (c)	700/75 or 700LP	Usable in the presence of 1-100 mM DTT.
13	585 nm (v)	600/50	605 nm (c)	600/50	Do not use.
14	655 nm (c)	640/25	705 nm (c)	700 LP	Usable in 1-100 mM DTT. Ensure no spectral overlap or spectral diffusion.
15	655 nm (v)	640/25	705 nm (c)	700 LP	Usable in 1-10 mM DTT. Ensure no spectral overlap or spectral diffusion.

Table 2.2: Combinations of QDs for three-color single molecule imaging.

Sr. no	QD_1	$Filter_1$	QD_2	$Filter_2$	QD_3	$Filter_3$	Notes
1	525 nm (c)	520/40	565 nm (c)	530/50	605 nm (c)	600/50	Low SNR for 525 nm.
2	525 nm (c)	520/40	585 nm (c)	600/50	655 nm (c)	640/25	Low SNR for 525 nm.
3	525 nm (c)	520/40	585 nm (c)	600/50	705 nm (c)	700 LP or 700/75	Low SNR for 525 nm. Low on time for 705 nm.
4	565 nm (c)	530/50	605 nm (c)	600/50	655 nm (c)	640/25	Slow photo-brightening of 605 nm.
5	565 nm (c)	530/50	605 nm (c)	600/50	705 nm (c)	700 LP or 700/75	Slow photo-brightening of 605 nm. Low on time for 705 nm.
6	585 nm (v)	600/50	655 nm (c)	640/25	705 nm (c)	700 LP or 700/75	Low on time for 705 nm. Ensure fresh DTT.
7	605 nm (c)	600/50	655 nm (c)	640/25	705 nm (c)	700 LP or 700/75	Low on time for 705 nm. Slow photo-brightening. Ensure fresh DTT.

3.0 SINGLE PARTICLE TRAJECTORIES: ACQUISITION AND ANALYSES

Single particle tracking (SPT) is the basis of understanding the interactions of single biomolecules. Here, we describe the methods employed for acquisition of single particle data from observed protein-DNA interactions in the DNA tightrope assay described previously (2.3). In these experiments, real time video microscopy of the interactions of fluorescently tagged biomolecules (particles) yields a sequence of spatial positions for the interacting particles at successive times.

3.1 DATA ACQUISITION

3.1.1 Digitization of the spatial signal

If the density of single fluorophores is sufficiently low, then an individual fluorophore may be identified as a diffraction limited spot (corresponding to the Airy disk) in the image. The radius of the Airy disk is given by equation 3.1:

$$r = \frac{0.61\lambda}{NA} \tag{3.1}$$

Two proximal fluorophores may be resolved in a microscope if they satisfy the Rayleigh criterion, which defines the resolution as equation 3.2:

$$Resolution_{x,y} = \frac{0.61\lambda}{NA} \tag{3.2}$$

The Nyquist sampling theorem states that, for optimal resolution of the fluorophore, the sampling should be performed at half the resolution equation 3.3:

$$Nyquistsampling = \frac{0.61\lambda}{2NA} \tag{3.3}$$

In order to resolve single molecules, the data should be spatially sampled at atleast the Nyquist sampling rate. When the pixel size is larger than the Nyquist sampling rate, the features of the fluorophore may not be resolved accurately. Pixel sizes which are smaller than the Nyquist sampling result in oversampling of the data. For 525 nm emission, this rate corresponds to 110 nm × 110 nm pixels, with progressively higher pixel sizes for higher emissions. In this work, the pixel sizes were in the range from 46 × 46 $\text{nm}^2\text{pix}^{-1}$ or 60 × 60 nm^2 pix^{-1}.

3.1.2 Determining the positional accuracy

In practice, the resolution criterion described above does not determine the precision with which the center of the Airy disc may be located. This is because, the PSF of the emitted light may be approximated as a 2-dimensional Gaussian function with high accuracy. Positional accuracy then depends on the signal, the quality of the Gaussian fit to the diffraction limited spot given (at least) optimal Nyquist sampling and the shot noise. The accuracy with which a single fluorophore in a frame may be localized is termed as the positional accuracy. The positional accuracy (σ_i) of a fluorophore in one dimension may be calculated as [120]:

$$\sigma_i = \sqrt{\frac{s^2 + \frac{a^2}{12}}{N_{photons}} + \frac{4s^3b^2\sqrt{\pi}}{aN_{photons}^2}} \tag{3.4}$$

where

s is the standard deviation of the peak position

a is the pixel size

b is the standard deviation of the background intensity

$N_{photons}$ is the photon count

To calculate the positional accuracy, first, a line was drawn across the center of the fluorophore to obtain the intensity profile. The intensity profile was then fit to a Gaussian fit, to obtain the mean position and its standard deviation (s). The Gaussian fit to the intensity profile was then used to calculate the number of photons per pixel N as:

$$N = \frac{(Counts - bias\, offset) \times sensitivity}{Quantum\, Efficiencyatdesiredwavelength} \tag{3.5}$$

Here, N was considered to be a smooth function. In reality, because of finite pixel size, both the intensity and N is a step function. The total number of photons $N_{photons}$ was obtained by integrating the area under the curve corrected for the background.

Using this formula and the procedure described here, the positional accuracy was calculated for 605 nm quantum dots conjugated to biotin on DNA damage arrays. For this purpose, individual QDs were identified from a 200 frame video with an exposure time of 100 ms. For each individual frame in the video, the intensity profile was fit to a Gaussian function and the corresponding total number of photons was calculated. Then, using $a = 46$ nm, the standard deviation of the background (b) and the number of photons ($N_{photons}$), the positional accuracy was calculated using equation 3.4 for each individual frame, and the average positional accuracy was calculated for the particle in the entire video. The positional accuracies for 16 such DNA bound QDs were calculated and found to lie in the range from 2 - 11 nm with an average positional accuracy of 6 nm ± 3 nm (mean ± standard deviation).

3.1.3 Determining the localization precision

The positional accuracy does not account for the underlying fluctuations of the DNA tightropes. In order to account for thermal fluctuations and background stage drift, the uncertainty in the mean position of the fluorophore over the entire duration of observation needs to be identified. This is known as the localization precision. The localization precision of a stationary fluorophore can be measured from a time lapse video of the fluorophore.

Consider a stack of N frames corresponding to a video of a stationary fluorophore, with N_{on} being the subset of frames in which the fluorophore is in the fluorescent ‘on’ state. The Gaussian fit to the PSF of the particle can be used to obtain the coordinates (x_i, y_i) of its mean position in the i^{th} frame in the series of N_{on} frames. Then, the mean position of the fluorophore during the observation window can be calculated as (x_c, y_c) given by equation 3.6:

$$\{\langle x_c \rangle, \langle y_c \rangle\} = \frac{1}{N_{on}} \sum_{i=1}^{N_{on}} \{x_i, y_i\} \tag{3.6}$$

and, the standard deviation along x and y may be calculated using equation 3.7

$$\{\triangle_x^2, \triangle_y^2\} = \frac{1}{N_{on} - 1} \sum_{i=1}^{N_{on}} \{(x_i - \langle x_c \rangle)^2, (y_i - \langle y_c \rangle)^2\} \tag{3.7}$$

Then, the localization precision is calculated using equation 3.8

$$\delta r = (\triangle_x^2, \triangle_y^2)^{\frac{1}{2}} \tag{3.8}$$

Evidently, an accurate measurement of the localization precision depends on the size of N_{on}.

In order to identify the localization precision of a quantum dot, we followed a previously described method [117]. Estimates for the localization precision of QD605 conjugated to DNA were found to be 33 nm ± 2 nm and 36 nm ± 3 nm along orthogonal directions. This value corresponds to a length of 100 bp of DNA. This indicates that two colocalized quantum dots on DNA may be resolved with a resolution of 100 bp.

3.2 ANALYSES OF SINGLE PARTICLE TRAJECTORIES

The interaction of single particles on DNA is complex, with theoretical predictions as well as experimental observations of extremely complex modes of motion ranging from immobile complexes to particles participating in diffusive (unrestricted, confined) and directed transport. Single particle tracking thus enables the distinction between these various types of transport mechanisms. Single particle trajectories can be analyzed to measure two physical characteristics of the interactions: the diffusion constant D and the lifetime (τ) of interaction. These two quantities can then be used to identify the nature of the motion and also the existence and affinities of multiple biomolecular intermediates.

3.2.1 Measurement of the diffusion constant

For particles which perform a 1-dimensional diffusion, for example, DNA binding proteins diffusing on stretched DNA, the analysis of the data involves identification of the trajectory to identify the nature of the diffusion. It is important to note that, in the analysis of diffusive particle trajectories, the measurement of the diffusion constant does not require high temporal or spatial resolution.

The analysis of the single particle trajectory depends on the experimental set up in question. In this work, we will consider the diffusion of DNA bound proteins on DNA tightropes stretched between beads in the absence of flow. The diffusion constant of a particle diffusing along a single line may be calculated in two ways.

In the first method, the position of the fluorescent probe is measured by fitting the point spread function to a two dimensional Gaussian function for each frame in a stack of frames. This approach yields the localization of the probe in 2-D space as a function of time. Let (x_i, y_i) represent the mean position of the probe in the i^{th} frame. Let n represent the number of frames skipped between frames used in the calculation of the diffusion constant and let N represent the total number of frames acquired. Then, the mean square displacement of the particle is calculated as:

$$MSD(n, N) = \frac{1}{N-n} \sum_{i=1}^{N-n} \{(x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2\} \tag{3.9}$$

Alternately, making use of the configuration of the DNA in the assay, and taking into account that the DNA binding protein is restricted to 1-D motion on the DNA, the data may be transformed into a kymograph. In a kymograph, the data are presented such that the y-axis represents the position of the probe along a vector with an arbitrary origin (r_i) in the i^{th} frame, aligned along the length of the DNA and time is represented along the x-axis. Thus, a kymograph represents the time dependent position of the particle along the length of the DNA molecule. In this case, the mean square displacement (Δr) of the particle is calculated as:

$$MSD(n, N) = \frac{1}{N-n} \sum_{i=1}^{N-n} \{(r_{i+n} - r_i)^2\} \tag{3.10}$$

For a particle exhibiting a purely diffusive behavior, the MSD demonstrates a linear dependence on the time step (Δt) as:

$$MSD(n, N) = \Delta r^2 = 2D\Delta t(n) \tag{3.11}$$

For true Brownian motion, the relationship between the MSD and time is linear. This means, that the displacement of a diffusing particle is a linear function of $\sqrt{t}$. However, it has been observed that in practice, the displacement of a particle does not always exhibit a linear relationship with $\sqrt{t}$, prompting the development of new models for fitting the experimental data. One alternate model suggested in the literature describes a sub-diffusive phenomenon. In this case, the MSD is related to the diffusion constant as:

$$MSD(n, N) = 2D\Delta t(n)^{\alpha} \tag{3.12}$$

where α represents the anomalous diffusive exponent.

3.2.1.1 Factors affecting the estimation of the diffusion constant

The estimate of the diffusion constant depends on several factors. Some of these are discussed here.

1. The accuracy of measurement of the diffusion constant depends on the accuracy with which the particle is localized. This in turn, depends on the quality of the Gaussian fit to the point spread function.

2. The ability to localize a diffusing fluorophore in turn depends on the length of the exposure time.

3. The estimate of the diffusion constant also depends on the length of the acquisition window. The observation window must be large enough to accommodate the heterogenous diffusive behaviors of single particles. Importantly, when fitting the mean square displacement to a Brownian diffusion model, it is important to check whether the sum of consecutive displacements is zero.

3.2.2 Lifetime analysis

3.2.2.1 Describing a Poisson process for a system shuttling reversibly between two states

Consider a system which reversibly shuttles between two observable states S_i and S_k. Here, we define a 'state' as a configuration of the biomolecule in consideration with a unique Gibbs free energy. The differences in these states arise from differences in the spatial positions of the constituent atoms in the macromolecule wrt each other. In this work we will consider the reversible interaction of DNA bound proteins with DNA and so the two states may be thought of as the DNA bound state and the DNA free state. Assume that the decay from the i^{th} state to the k^{th} state (and vice versa) is experimentally measurable.

For a Poisson process (T_i), the escape from the S_i state to the S_k state is given by equation (3.13)

$$T_i(t) \equiv exp\{-k_{d,i}t\} \tag{3.13}$$

where, $k_{d,i}$ represents the rate constant for the decay process and the mean lifetime for the S_i state is obtained as $\tau_i = \frac{1}{k_{d,i}}$. A similar T_k process can be described for the reverse reaction.

3.2.2.2 Cumulative residence time distribution analysis for transients on DNA

Dissociation kinetics of a population of proteins bound to DNA as a single intermediate may be described by a Poisson process. For a heterogeneous population of intermediates, let the number of the molecules in the population participating in the i^{th} kinetic intermediate (described by the i^{th} Poisson process T_i) be given by:

$$n_i(t) = a_i T_i(t) \tag{3.14}$$

Here, $n_i(t)$ represents the number of molecules of the i^{th} kinetic intermediate remaining on the DNA as a function of time, starting with an initial population of a_i observations.

In general, the dissociation kinetics of populations decaying according to N_p distinct Poisson processes are described by:

$$n(t) = \sum_{i=1}^{N_p} a_i T_i \tag{3.15}$$

Equation 3.15 describes the total number of molecules which remain associated with the DNA as a function of time for all the different kinetic intermediates taken together. The normalized fraction $f(t)$ represents the cumulative residence time distribution (CRTD).

3.2.2.3 CRTD analysis of particles dissociating from DNA

DNA bound proteins which dissociate during observation represent another class of molecules. For this class of molecules, we can count the number of particles that dissociate as a function of time and performed a similar CRTD analysis. The CRTD was parametrized by fitting a number of Poisson processes essentially as described by equation 3.15.

3.2.2.4 On the detectability of decay processes

Importantly, the measurement of Poisson processes should be invariant under the type of observation. That is, both types of analyses (the analysis of transient and of dissociating molecules), should reveal the same Poisson processes, if all processes are represented adequately. The ability to detect a decay process depends on at least the following factors:

1. the frequency with which an event occurs (ν)
2. the decay constant (τ)
3. the time resolution (t_s) and

4. the length of the observation window (L).

Our ability to measure a Poisson process depends directly on ν and L, and inversely on τ and t_s. Processes which occur with low ν and $\tau = O(t_s)$ or $\tau = O(L)$ may not be measured accurately.

3.2.3 Statistical analysis of experimental data

Pairs of parameters were tested for statistical significance by performing Welch t-tests (Student's t-test with unequal sample size and unequal variances) on the mean values and the standard deviations (obtained from the 95%CI) of the parameters obtained from fitting. Fitting was performed using the 'cftool' package in MATLAB by fitting the data to appropriate equations and using Occam's razor to choose between comparable fits.

4.0 FROM STRUCTURE TO DYNAMICS: HOW DOES UV-DDB SEARCH FOR AND RECOGNIZE DAMAGE IN DNA?

UV-DDB is the initiator of GG-NER. The recognition of CPD lesions in UV-damaged chromatin is mediated by the UV-damaged DNA binding protein (UV-DDB), composed of the tightly associated heterodimer of DDB1 (p127) and DDB2 (p48) [16, 17]. Following surveillance and CPD recognition by UV-DDB, NER proceeds via lesion handover to XPC-HR23B-centrin2 followed by damage verification, helix opening and stabilizing of the repair intermediates, dual incision of the DNA in the context of the lesion, repair synthesis and DNA ligation [9, 25, 121, 29, 5].

In vivo, UV-DDB is constitutively associated with Cullin4A or 4B and RBX1, forming the $CRL4^{DDB2}$ E3 ligase complex [25, 31, 122]. In this complex, DDB2 is a DNA damage recognition factor and functions as an adapter protein which targets the E3 ligase activity to sites of UV-induced photoproducts, promoting chromatin relaxation, and enabling access to subsequent repair factors [28, 26, 18]. Recombinant DDB2 has been demonstrated to bind a variety of DNA structures including 6-4 photoproducts, abasic sites, two base mismatches with extremely high affinity and CPD lesions and cisplatin adducts with relatively lower affinity [105, 18]. Molecular defects in DDB2 lead to a slower loss of UV-induced photoproducts and presentation of the skin cancer prone xeroderma pigmentosum complementation group E (XP-E) [123, 30]. Molecular analysis of DDB2 from XP-E patients revealed genetic defects which lead to structural defects in DDB2 such as truncations, misfolding or a modification of the DNA binding interface [123, 30]. In the case of the XP82TO patient, a point mutation was observed in lysine 244 to glutamate (K244E) in DDB2 which results in significantly reduced DNA binding activity and damage specificity [30, 18, 121].

A current working model for damage recognition in GG-NER is that UV photoproducts are first recognized by UV-DDB (Figure 4.1) [29]. Stable binding of UV-DDB to sites of damage activates the ubiquitination activity of $CRL4^{DDB2}$ which targets histones, primarily H2A, and enables nucleosome disassembly and subsequent recruitment of the XPC-HR23B-centrin2 (XPC) complex, which is also a ubiquitination substrate [29, 25, 121, 28, 26, 27, 124]. Lesion handover between $CRL4^{DDB2}$ and XPC is thought to be achieved by the autoubiquitination of DDB2 at lysines in the intrinsically disordered N-terminus of DDB2 [29, 25]. This autoubiquitination of DD2 serves to flag the repair factor for degradation [124].

Recent reports have identified that the N-terminus of DDB2 is also PARylated in vivo in response to UV-damage, resulting in stabilization of UV-DDB in damaged chromatin [125, 126]. While the molecular mechanisms of these two post-translational modifications remain to be fully elucidated, it appears that these modifications compete for substrates in the N-terminus of DDB2 and oppositely regulate each other, with implications on the residence of UV-DDB on DNA at sites of damage. The N-terminus of DDB2 also plays an important role in damage binding. This flexible domain has been recently implicated in dimerization of UV-DDB at sites of damage. In our previous work, we provided structural and biophysical evidence that UV-DDB dimerizes at sites of damage [102].

The identification of this dimeric UV-DDB complex on DNA reveals a previously unanticipated complexity in damage recognition and raises several important questions in the initial damage recognition step of human GG-NER: How do approximately 180,000 molecules of UV-DDB interrogate four billion base pairs of genomic DNA to find relatively rare lesions in DNA? How does UV-DDB proofread the DNA to achieve remarkable specificity in damage discrimination? How does dimeric UV-DDB modulate the specificity of damage discrimination? How do mutations in the DNA binding interface found in the K244E mutant of DDB2 influence the kinetics of DNA binding and damage recognition?

To address these questions, we have conjugated WT and K244E mutant of UV-DDB to streptavidin coated quantum dots (SA-QD) and observed their interactions in real time with elongated damaged DNA substrates in a DNA tightrope assay [115, 127].

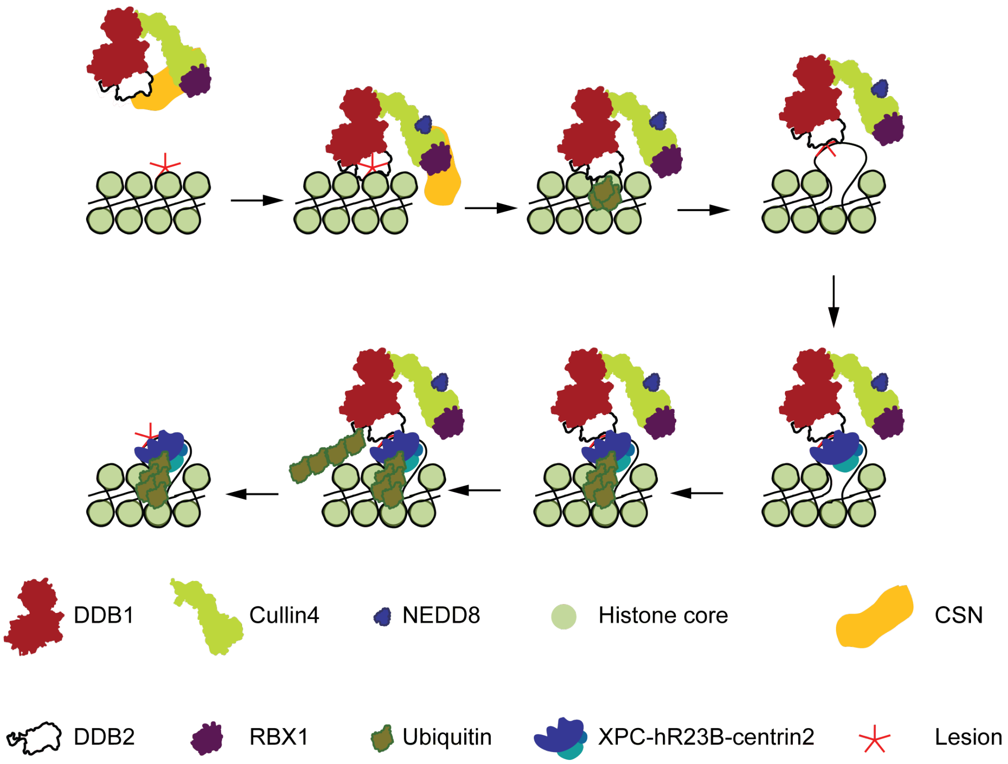


Figure 4.1: Model for lesion handover between UV-DDB and XPC

4.1 VISUALIZING THE SEARCH MECHANISM FOR UV-DDB

The question of target site search by DNA binding proteins has received extensive attention previously using biochemical methods and more recently, several search modes have been proposed (section 1.6, Figure 1.6), experimentally validated and visualized using single molecule methods [128, 129, 127, 130, 131]. We have previously developed a DNA tightrope assay which enables the visualization of DNA mediated QD conjugated protein-protein and protein-DNA interactions [115, 127]. Briefly, in this assay, λ-DNA tight ropes are elongated between five micron poly-L-lysine coated beads that are deposited on a pegylated cover slip. Biomolecular interactions on DNA tightropes in the absence of buffer flow are visualized by oblique angle fluorescence microscopy imaging. This strategy also enables observations in the absence of surface interactions. A schematic of the flow cell and a YOYO-1 stained image of a typical string-up of DNA obtained using oblique angle fluorescence microscopy are shown in Figure 4.2.

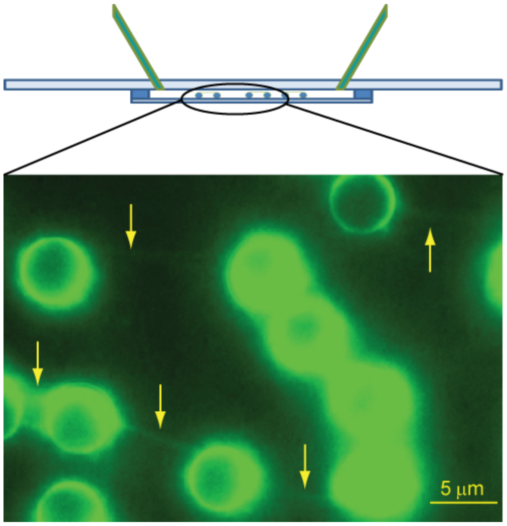


Figure 4.2: Schematic of flow cell used in DNA tightrope assay

Schematic of flow cell used in the DNA tightrope assay and oblique angle fluorescence microscopy image of YOYO-1 stained λ-DNA tightropes between poly-L-lysine coated beads. Arrows indicate DNA between beads.

4.2 QD CONJUGATION STRATEGIES

The goal of these studies was to understand the dynamics of UV-DDB on damaged DNA substrates. To achieve this goal, we first systematically tested three strategies for conjugating QDs to UV-DDB.

4.2.1 Using the biotinylated trisnitrilotriaceticacid compound

We first tested the conjugation strategy involving a biotinylated trisnitrilotriaceticacid (BT-NTA) compound [132]. We assayed DNA binding of QD conjugated UV-DDB ($^{FLAG-His}$DDB1-DDB2) by the use of an agarose based EMSA with a 5' ^{32}P radiolabeled AP36 substrate (Figure 4.3). Using the BT-NTA compound, we found that under the conditions tested, UV-DDB exhibits a super shift in DNA binding upon conjugation with BT-NTA (Lane 4, Figure 4.3); however, this super shift is lost upon conjugation to QDs (Lane 5, Figure 4.3). Indeed, BT-NTA conjugated QD-UV-DDB exhibits a small super shift, but also leads to a significant loss of DNA damage binding ability, suggesting that the short length of the BT-NTA linker to the QD inhibits the stable binding of UV-DDB to DNA through steric hindrance.

4.2.2 Using the FLAG Ab sandwich approach

We next tested an antibody sandwich approach [133] using a primary anti-FLAG antibody and IgG coated QDs (Figure 4.4) to conjugate $^{FLAG-His}$DDB1-DDB2 to IgG coated QDs. We found that QD-FLAG Ab-UV-DDB exhibits a measurable super shift in the agarose EMSA corresponding to DNA bound QD-FLAG Ab-UV-DDB complexes on 177bp UV-irradiated DNA (Figure 4.4). Evidence for these complexes was also observed by AFM when QD-FLAG Ab-UV-DDB was incubated with UV-irradiated 517 bp DNA containing on average 1 lesion per fragment (Figure 4.5). This experiment was performed by Dr. Hong Wang.

4.2.3 Using the His-Ab strategy

We performed electrophoretic mobility shift assays (EMSAs) to measure the binding of His-Ab [131] conjugated QD-UV-DDB to AP36 (Figure 4.6) [18]. We found that His-Ab conjugated UV-DDB ($^{FLAG-His}$DDB1-DDB2) demonstrated a measurable super-shift corresponding to UV-DDB-His-Ab-QD:DNA complexes (Figure 4.6, lane 6). Additionally, no non-specific binding of either the His-Ab or QDs binding to DNA was detected (Figure 4.6: lanes 3 and 4). We also tested other UV-DDB constructs (^{His}DDB1-^{FLAG}DDB2 and DDB1-$^{FLAG-His}$DDB2) and found that in all cases, DNA binding activity was retained (Figure 4.6: lanes 7-9 and 10-12 resp.).

Further, we tested different UV-DDB:His-Ab:QD ratios such as 1:5:1, 1:5:2.5 and 1:5:5 and found super-shifted bands for all the conditions tested (Figure 4.7). However, we used a ratio of 1:5:1 for our experiments, since under these conditions, the biotin binding sites on the streptavidin coated quantum dots are completely saturated [131].

The His-Ab conjugation strategy was found to be robust and we therefore decided to proceed with this strategy for observing the interactions of UV-DDB with DNA in real time in the DNA tightrope assays. Importantly, UV-DDB is a tightly associated complex of DDB1

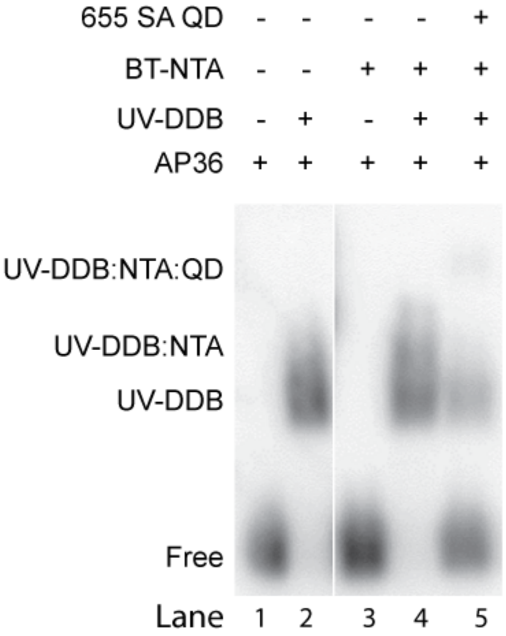


Figure 4.3: Agarose EMSA to assay activity of QD-BT-NTA-UV-DDB

Agarose EMSA showing supershift of 2.5 nM AP36 substrate (lane 1) upon binding of UV-DDB ($^{FLAG-His}$DDB1-DDB2; Lane2). Lane 3 demonstrates lack of detectable interaction of DNA with BT-NTA. Lane 4 shows biotinylated NTA compound conjugated to UV-DDB ($^{FLAG-His}$DDB1-DDB2), and subsequent loss of DNA binding upon conjugation to QDs (lane 5).

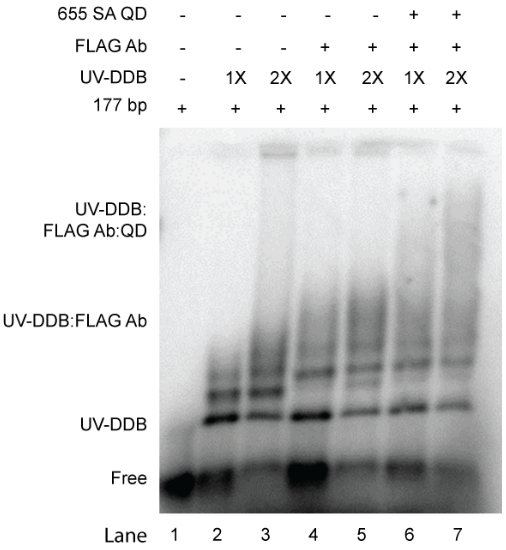


Figure 4.4: Agarose EMSA to assay QD-FLAG Ab-UV-DDB activity

Agarose based electrophoretic mobility shift assay showing supershift of 7.5 nM 177 bp UV-irradiated radiolabeled probe (lane 1) when bound to UV-DDB $^{FLAG-His}$DDB1-DDB2 alone (lanes 2 and 3, [UV-DDB] = 28.5 nM and 57 nM resp) or when conjugated using the primary anti-FLAG antibody (lanes 4 and 5 [UV-DDB] = 28.5 or 57 nM and [Ab] = 66 nM) to goat anti-mouse IgG secondary antibody coated QDs ([UV-DDB]=57 nM, [Ab] = 66 nM and [QD] = 100 nM or 500 nM lanes 6 and 7 resp). Higher bands in the EMSA correspond to either end-binding, higher oligomeric states of UV-DDB or binding to distinct photoproducts.

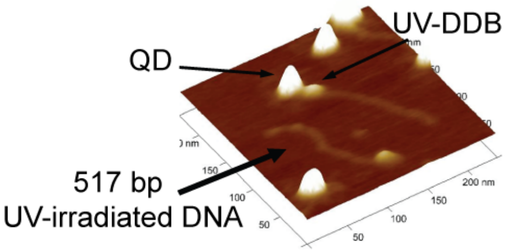


Figure 4.5: AFM image of QD-FLAG Ab-UV-DDB bound to UV-irradiated 517 bp DNA

AFM image (250 nm x 250 nm x 3nm) with 3D rendering, showing goat anti-mouse IgG coated QDs co-localized to UV-irradiated 517 bp PCR product in the presence of the anti-FLAG antibody conjugated to FLAG-HisDDB1-DDB2.

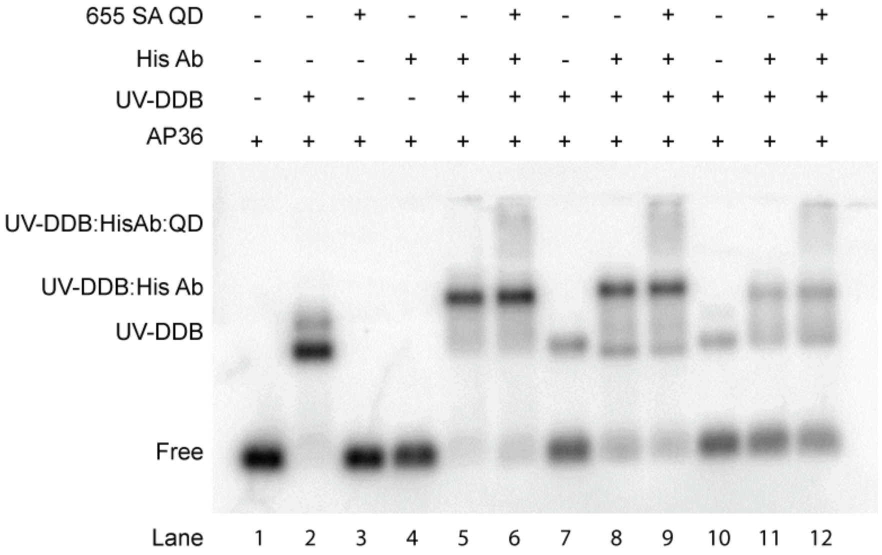


Figure 4.6: Agarose EMSA to assay QD-His Ab-UV-DDB binding

Agarose EMSA showing that 5' ^{32}P 36mer containing AP site (lane1) can be bound by UV-DDB (lane2). Lanes 3 and 4 are negative controls indicating that neither the His-Ab nor the QD non-specifically bind 36 mer DNA substrate. Lane 5 indicates a complete shift in the presence of His-Ab and lane 6 indicates a super shift in the presence of streptavidin coated QD. For these experiments, 2.5 nM dsDNA substrate containing an abasic site was incubated with either UV-DDB only (FLAG-HisDDB1/DDB2: 50 nM, HisDDB1/FLAGDDB2: 29 nM, DDB1/FLAG-HisDDB2: 37 nM, Lanes 2, 7 and 10 resp) and Penta-His Ab ([UV-DDB]: [Penta-His Ab] $\equiv$ 1:5, Lanes 5, 8 and 11 resp) and 655 nm streptavidin coated Qdots ([UV-DDB]: [Qdots] $\equiv$ 1 :1, Lanes 6, 9 and 12 resp).

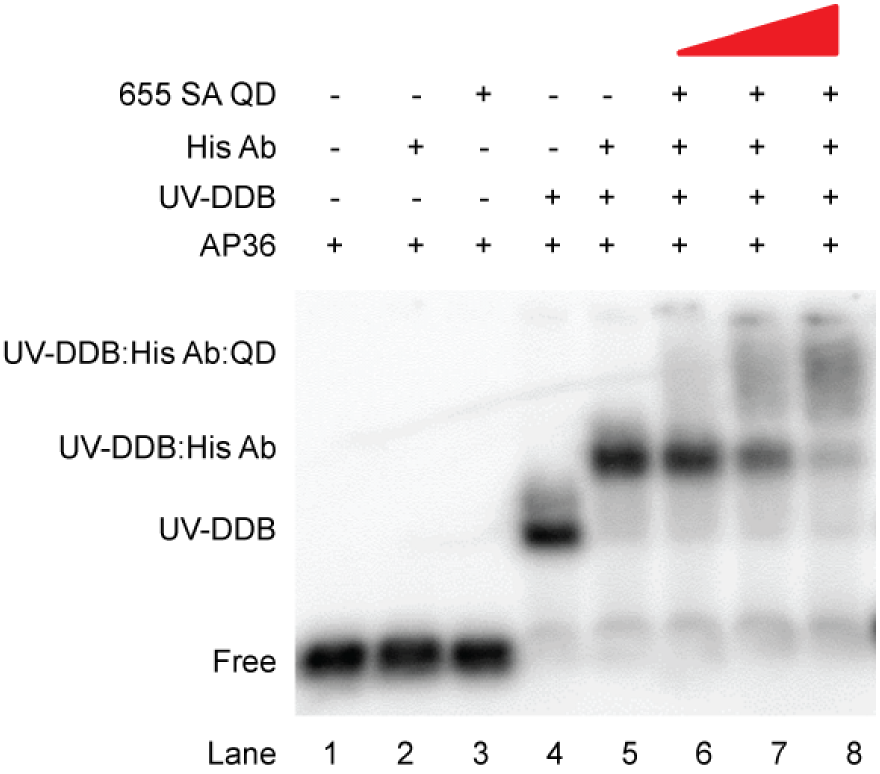


Figure 4.7: Agarose EMSA assaying QD-His Ab-UV-DDB activity as a function of QD ratio

Agarose EMSA showing that 5’ ^{32}P labeled 36mer containing AP site (2.5 nM) alone (lane 1) or in the presence of His-Ab (lane 2) or 655 SAQD (lane 3) or incubated with FLAG-HisDDB1/DDB2 (lane 4) and antibody (lanes 5-8) and QD (lanes 6-8) ratios were varied as indicated in the lanes keeping UV-DDB concentration constant (Lane 6: P:Ab:QD≡1:5:1, Lane 7: P:Ab:QD≡1:5:2.5 and Lane 8: P:Ab:QD≡1:5:5).

and DDB2 and previous reports have reported similar diffusion constants for the complex containing DDB1 and DDB2 when either subunit was fluorescently tagged inside living cells [134]. Given these results, we believe that conjugating the DDB1 subunit with QDs serves as an accurate reporter for the behavior of DDB2 on DNA.

4.3 IMAGING IN THE ABSENCE OF YOYO-1 DYE

Consistent with a recent report, we observed shattering of DNA tightropes incubated with YOYO-1 during imaging, presumably via the formation of single strand breaks followed by double strand breaks [135]. In order to enable long observation times for UV-DDB binding to DNA with minimal DNA damage (which could create binding sites for UV-DDB), the use of YOYO-1 was avoided in these experiments.

4.4 QD-UV-DDB PERFORMS A 3D SEARCH ON UNDAMAGED AS WELL AS UV-DAMAGED DNA

We next wanted to characterize the binding of QD-UV-DDB to long undamaged DNA substrates, as well as, DNA containing UV-induced lesions using the DNA tightrope assay. Upon incubating QD-UV-DDB with undamaged or UV-damaged λ DNA tightropes containing on average 1 photoproduct in 2200 bp of DNA (see figure 4.9), we were able to observe the binding of UV-DDB to the DNA tightropes (Figure 4.8, mov 1 and mov 2).

For both undamaged, and UV-damaged DNA, four classes of binding events were identified in a typical 900 s observation window (Figure 4.10); those that associate, dissociate, persist and are transient. In the majority of cases (>98%), upon incubation with either undamaged or UV-damaged DNA, QD-UV-DDB molecules associated with the DNA primarily as non-movers (see mov1 and mov2). Less than 2% of observations (12 out of 990 events) exhibited one dimensional sliding of UV-DDB within the limits of our spatial and temporal resolutions (36 nm, 100 ms). Since these assays were performed in the absence of flow, we

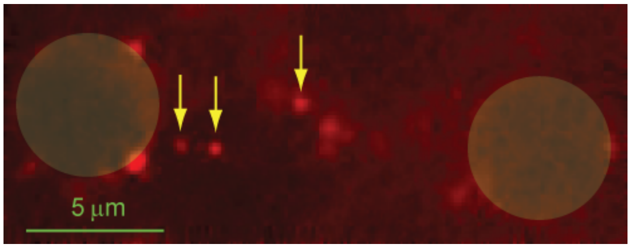


Figure 4.8: Still image of QD-UV-DDB bound to undamaged λ-DNA tightrope

Image of QD-UV-DDB binding to unstained DNA tightropes between beads. Arrows indicate bound QD-UV-DDB.

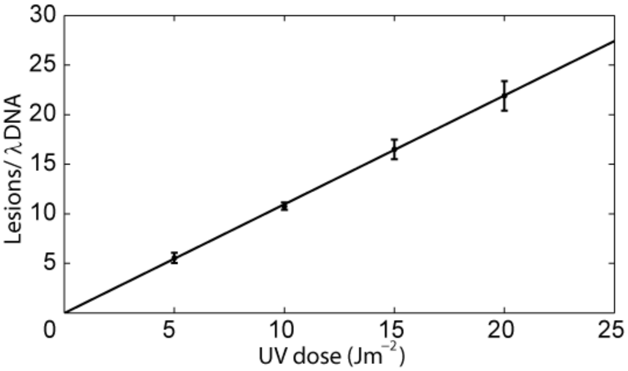


Figure 4.9: UV-induced photoproduct frequency per λ-DNA is a linear function of UV-dose as measured by qPCR.

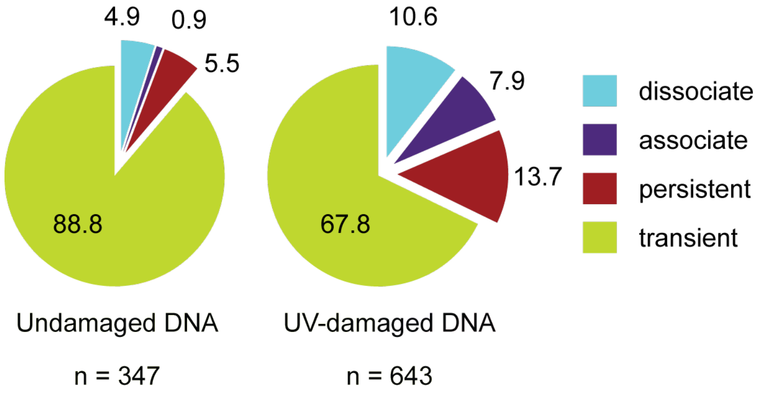


Figure 4.10: Four types of binding events observed in the DNA tightrope assay

Kymographs of individual QD-UV-DDB molecules that: A) dissociate from DNA during observation B) associate with DNA during observation C) are present during the entire observation window D) both associates and dissociates during observation. E) and F) pie chart of percentage of each of the observations described above for undamaged DNA (n = 347) and UV-damaged DNA (n = 643), respectively.

were able to observe macroscopic dissociation and re-association behavior of UV-DDB on vicinal DNA molecules, a phenomenon which is characterized as jumping (Figure 4.11; mov1: an example of transient binding of QD-UV-DDB to DNA) [115].

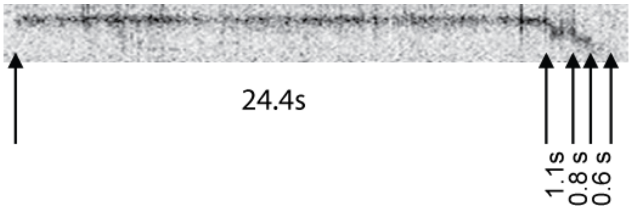


Figure 4.11: Kymograph of UV-DDB jumping between vicinal DNA molecules in field of view.

4.5 UV-DAMAGE IN DNA SHIFTS THE BINDING EQUILIBRIUM TO LONGER LIVED STATES

A comparison of the percentage of observations that are represented in each of the four classes of observations on undamaged (Figure 4.10 E, n = 347) and UV-damaged DNA and (Figure 4.10 F, n = 643) reveals that the fraction of molecules that:

1. dissociate is 2.2 fold higher for UV-damaged DNA (compare 4.9% vs. 10.6%)
2. associate is 8 fold higher for UV-damaged DNA (7.9% vs. 0.9%)
3. persist is 2.5 fold higher for UV-damaged DNA (13.7% vs. 5.5%)
4. are transient is 1.3 fold lower (88.8% vs. 67.8%) for UV-damaged DNA.

These data reveal that across all these classes of binding events, UV-DDB associates with and persists on DNA for longer time periods when the DNA tightropes contain UV-damage, while showing lower transient behavior.

4.6 SINGLE MOLECULE VISUALIZATION REVEALS MULTIPLE KINETIC INTERMEDIATES OF UV-DDB

We then analyzed the kinetics of transient binding on both undamaged and UV-damaged DNA to understand the heterogeneity in dwell times and to extract dissociation rate constants. Dissociation of UV-DDB complexes on DNA can be modeled as a first order decay process (Poisson process) from the DNA bound intermediate state (S_i) to the DNA free state (S_0). With the objective of quantifying this process, we performed a cumulative residence time distribution analysis (CRTD; see 3.2.2) [136, 137]. The CRTD may be interpreted as a type of survival curve representing the fraction of the population of DNA bound proteins remaining on DNA as a function of time. Fitting the CRTD to a Poisson process ($T_i \equiv exp\{-k_{d,i}t\}$) yields the decay rate (here dissociation rate ($k_{d,i}$) and consequently the mean lifetime (τ_i) of particles participating in the Poisson process describing the escape from the S_i state into the S_0 state ($S_i \leftrightarrow S_0$; here describing the reversible and spontaneous two state transition of a DNA binding protein from the DNA bound state to the DNA free state). For systems with multiple intermediates, the CRTD may be fit to a number of exponential terms, each of which corresponds to a measurable intermediate. This yields the total number of Poisson processes, that describe the population.

The CRTDs describing the data for transient binding to undamaged DNA and damaged DNA are presented in Figure 4.12. A log rank test indicated that the survival functions are not significantly different (p = 0.075), suggesting that that the aggregate survival functions for UV-DDB interacting with undamaged DNA and damaged DNA are similar. Next, we proceeded to parameterize these CRTDs by fitting a sum of exponentials to the data (red line, Figure 4.12 A and B). In each case, to fit all the features of the data it was necessary to fit three exponentials corresponding to three independent decay processes (see section 4.6.1).

4.6.1 Best fit of the experimental data

The CRTD was fit to a sum of N_p Poisson processes starting with n = 1 , . . . , N_p. Fit parameters for double ($N_p = 2$) and triple ($N_p = 3$) are provided in the table 4.1. Log-log plots for the experimental data and double exponential fit (dashed blue line) and triple exponential fits (solid, red) are presented in figure 4.13 for undamaged and damaged DNA. From these plots, it is evident that all the features of the data are only captured when N_p = 3. Since, SSE/DOF was found to be smaller for N_p = 3, we proceeded with fitting the population to 3 Poisson processes (Table 4.1).

We denote the three independent Poisson processes for transient binding to undamaged DNA as $T_{1,ud}$, $T_{2,ud}$, and $T_{3,ud}$ (where $T \equiv exp\{-kt\}$). These processes describe the dissociation of UV-DDB from the DNA bound state to the DNA free state (corresponding to the transitions $S_{1,ud} \rightarrow S_0$, $S_{2,ud} \rightarrow S_0$ and $S_{3,ud} \rightarrow S_0$; subscript 'ud' corresponds to undamaged DNA, see table 4.2, 3.2.2). Similarly, dissociation from UV-damaged DNA was found to follow the three processes $T_{1,d}$, $T_{2,d}$, and $T_{3,d}$ (corresponding to the transitions $S_{1,d} \rightarrow S_0$, $S_{2,d} \rightarrow S_0$ and $S_{3,d} \rightarrow S_0$; subscript 'd' corresponds to damaged DNA, see table 4.2, 3.2.2). From the data displayed in table 4.2, it is evident that in each case, the three lifetimes of interactions differ from each other by an order of magnitude. In summary, our data reveals

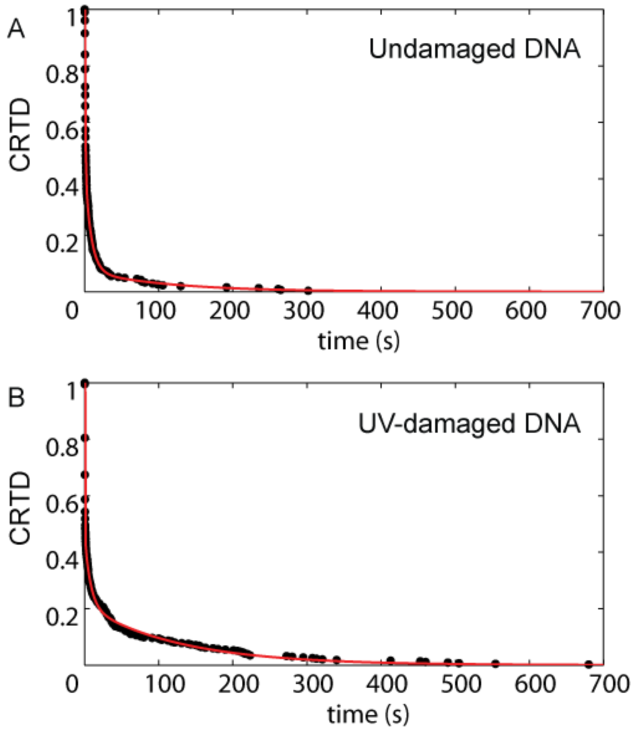


Figure 4.12: CRTD analysis of transient UV-DDB binding

Cumulative residence time distribution plots for transient UV-DDB binding to A) undamaged DNA and B) UV-damaged DNA with triple exponential fits (red).

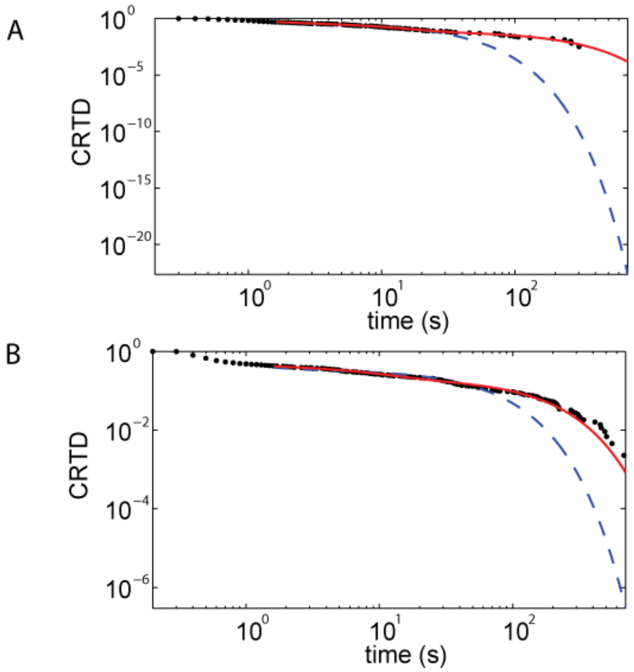


Figure 4.13: Log-Log CRTD plots

Log-Log Cumulative residence time distribution plots for transient UV-DDB binding to A) undamaged DNA and B) UV-damaged DNA with triple exponential fits (red) and double exponential fits (dashed, blue).

Table 4.1: Fit parameters for double and triple exponential fits to the CRTD

	DOF	R-sq	SSE	SSE/DOF	DOF	R-sq	SSE	SSE/DOF
Undamaged DNA	343	0.9929	3.82 $\times 10^{-2}$	1.11 $\times 10^{-4}$	341	0.9972	1.61 $\times 10^{-2}$	4.72 $\times 10^{-5}$
UV-damaged DNA	639	0.9658	1.59 $\times 10^{-1}$	2.49 $\times 10^{-4}$	637	0.9938	2.91 $\times 10^{-2}$	4.57 $\times 10^{-5}$

the presence of three decay processes (T_i for intermediates on undamaged, as well as, UV-damaged DNA) and four intermediate states (a DNA free state S_0 and three DNA bound states for dissociation from undamaged and damaged DNA each).

Table 4.2: Estimates for kinetic parameters from analysis of transient molecules

	N	a_1	$k_{d,1}(s^{-1})$	τ_1	a_2	$k_{d,2}(s^{-1})$	τ_2	a_3	$k_{d,3}(s^{-1})$	τ_3
Undamaged DNA	307	0.7863	1.217	0.8	0.3884	0.1233	8.1	0.07354	0.008853	113.0
LB of 95%CI		0.7587	1.127	0.8	0.677	0.1071	7.2	0.05296	0.004502	75
UB of 95%CI		0.8138	1.306	0.9	0.409	0.1396	9.3	0.09413	0.0132	222.1
Damaged DNA	436	1.107	3.033	0.3	0.251	0.1249	8.0	0.2119	0.007895	126.7
LB of 95%CI		1.029	2.782	0.3	0.2369	0.1087	7.1	0.2001	0.007125	115.4
UB of 95%CI		1.185	3.248	0.4	0.265	0.1412	9.2	0.2238	0.008664	140.4

4.6.2 The $T_{1,ud}$ process is significantly different from $T_{1,d}$ process

To identify the relationships between the pairs of processes $\{T_{1,ud}, T_{1,d}\}$, $\{T_{2,ud}, T_{2,d}\}$ and $\{T_{3,ud}, T_{3,d}\}$, we next examined if the rate constants for each pair were significantly different. For $T_{1,ud}$ we obtained $\tau_{1,ud} = 0.3$s (0.3s, 0.4s) (see table 4.2; reported as mean (LB, UB), where LB and UB represent the lower and upper bound of the 95% confidence interval on the estimate of the mean value; values are rounded to the temporal resolution of the instrument) compared to $\tau_{1,d} = 0.8$s (0.8s, 0.9s) for $T_{1,d}$ (table 4.2). Using a Welch T-test, we identified that the rate constants derived from the fits for these two processes were significantly different with $p < 0.0001$, indicating that the kinetics of the $T_{1,ud}$ process are significantly different compared to $T_{1,d}$. This suggests that the $T_{1,ud}$ and $T_{1,d}$ processes are unique. The fraction of the S_1 intermediate for both undamaged and UV-damaged DNA transient binding, which does not decay according to the T_1 process, must transition to a more stable intermediate (possibly the S_2 intermediate). It is noteworthy that the T_1 process identified here does not correspond to non-productive collisions that are found to occur on an approximately 100X faster timescale ($\approx$ 5 ms) [130].

4.6.3 T_2 represents a damage verification process; whereas T_3 represents dissociation from a stable damage bound state

We next examined the T_2 and T_3 processes for both undamaged and UV-damaged DNA and found that the rate constants for each process were not significantly different (table 4.2, $p > 0.05$), indicating that $\tau_{2,ud} = \tau_{2,d} = \tau_2 = 8$s (7.2s, 9.2s) and $\tau_{3,ud} = \tau_{3,d} = \tau_3 = 126$s (115.4s, 140.4s). To better understand how the presence of photoproducts in the DNA changes the kinetics of UV-DDB interaction, we first identified that the estimates for the amplitudes of the pairs of the decay processes were significantly different from each other ($p < 0.0001$). We then normalized the fitted data to identify the relative fractions of the kinetic intermediates that decay according to the three different processes for transient binding to undamaged versus UV-damaged DNA (Figure 4.14 A and B).

In the population of molecules bound to undamaged DNA, 58% were found to dissociate with a lifetime of $\tau_{1,ud}$ compared to 57% for UV-damaged DNA with a lifetime of $\tau_{1,d}$.

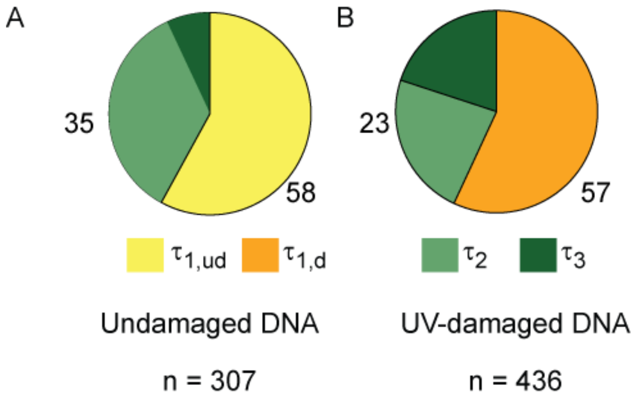


Figure 4.14: Pie chart showing the percentage of molecules participating in decay processes

Pie chart showing the percentage of molecules participating in each of the Poisson processes observed in the transient binding events for binding to undamaged DNA (A; n = 307) or UV-damaged DNA(B; n=436), respectively.

Similarly, 35% (1.5 fold higher) were found to dissociate from undamaged DNA compared to 23% for UV-damaged DNA with a lifetime of τ_2 and 7% of the molecules dissociated with a lifetime of τ_3 compared to 20% (2.9 fold higher) for binding to UV-damaged DNA (Figure 4.14 C and D). From these data, we infer a greater fraction of the $S_{2,ud}$ population dissociates from DNA according to the T_2 process compared to the $S_{2,d}$ population. We conclude that the increase in the $S_{3,d}$ population which comes with an accompanying loss of the $S_{2,d}$ population is a consequence of UV-DDB interacting with damaged DNA. This suggests that the T_2 process represents a damage verification process. This model suggests that the fraction of the S_2 intermediate that does not decay according to the S_2 intermediate is converted to the S_3 intermediate. Further, these data reveal a dependence of the stabilization of kinetic intermediates of UV-DDB on the presence of damage on DNA: upon introduction of UV-damage, we find that the $S_{3,d}$ population is enriched, indicating that a greater fraction of the UV-DDB is stably bound to UV-damaged DNA.

4.6.4 Analysis of dissociating molecules reveals an additional long lived intermediate

Further evidence for the presence of long lived complexes was the class of observations that are present at the beginning of data acquisition and dissociate during observation (Figure 4.10 A). We performed a CRTD analysis for this class of particles and obtained the survival curves presented in figure 4.15 for undamaged DNA (n = 16, blue) and damaged DNA (n = 67, red).

For the case of undamaged DNA, the CRTD was found to best fit a double exponential (Figure 4.16A), one of which was statistically indistinguishable from the $T_{2,ud}$ process (p = 0.0749) and the other exponential fit ($T_{4,ud}$) was statistically different from the $T_{3,ud}$ (p <0.0001) process. Parameterization of the CRTDs revealed the presence of the $T_{2,ud}$ process and a new $T_{4,ud}$ process for dissociation from undamaged DNA.

Similarly, from the CRTDs of UV-DDB dissociating from UV-damaged DNA, a new process ($T_{4,d}$) was identified which was consistent with the $T_{4,ud}$ process. In the case of UV-damaged DNA, the CRTD was dominated by a single exponential (Figure 4.16B) corresponding to a single process ($T_{4,d}$). Although this exponential was found to be distinct from the $T_{4,ud}$ process (p <0.0001) it was found to possess a decay rate which was very similar to that of the $T_{4,d}$ process. It is likely that the $T_{4,ud}$ and $T_{4,d}$ processes are identical, and the differences in the values arise from poor estimation of the of the $T_{4,ud}$ process due to low sample size (n = 11).

We wondered if it was possible that the T_4 and T_3 processes were identical. The data set of dissociating molecules represents those molecules for which $t_{assoc} \geq t_{meas}$ (where t_{assoc} represents the total time of association; and t_{meas} represents the measured time of association. Note that $t_{assoc} > t_{meas}$). We examined the limiting case that $t_{assoc} = t_{meas}$ (i.e. the molecules associated with the DNA at the instant prior to the start of observation) and plotted the decay of 16 molecules present for undamaged DNA (or 67 in case of UV-damaged DNA) molecules for the cases where $\tau = \{\tau_{3,ud}, \tau_{3,ud+}, \tau_{3,ud-}\}$ and $\tau = \{\tau_{4,d}, \tau_{4,d+}, \tau_{4,d-}\}$, where $\tau_{i,j+}$ represents the upper limit of the 95% CI and $\tau_{i,j-}$ represents the lower limit of the 95% CI (see Figure 4.17). In all cases we found no overlap of the curves, consistent with the results

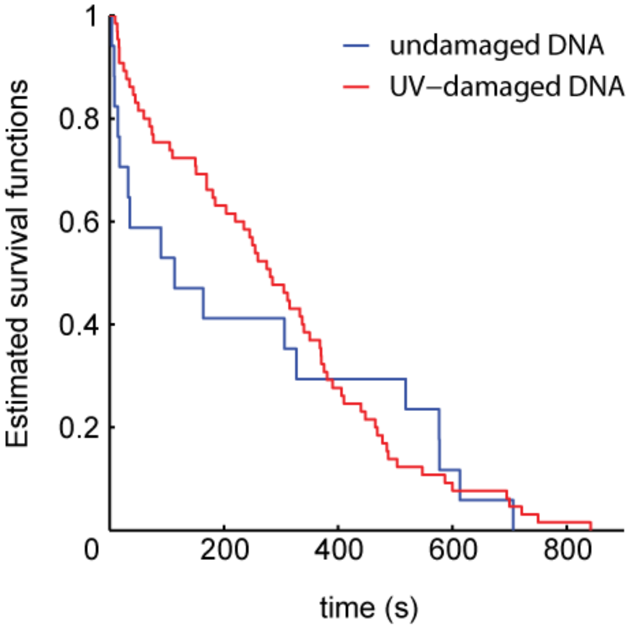


Figure 4.15: Survival curves for dissociating UV-DDB

Estimated survival functions for dissociating molecules from undamaged DNA (blue, n = 16) and UV-damaged DNA (red, n = 67).

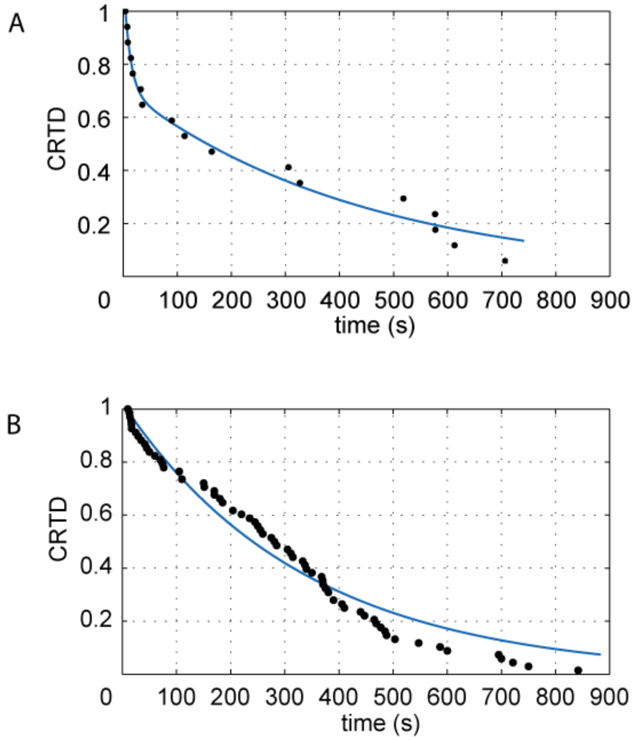


Figure 4.16: Exponential fits to the CRTDs of population of dissociating molecules

CRTD vs. t plot for dissociating UV-DDB from undamaged DNA A. and UV-damaged DNA B. with exponential fits (blue) to the experimental data (closed black circles).

of the hypothesis indicating that the T_3 and T_4 processes represent distinct intermediates. The fraction of molecules participating in the T_4 process was 3.3 fold higher and was calculated as fraction decaying according to T_4 process for damaged DNA (67/643)/ fraction decaying according to T_4 process for damaged DNA (11/347). Importantly, the T_4 process is statistically distinguishable from the corresponding T_3 process suggesting the presence of a fourth kinetic intermediate $S_{4,ud}$ and $S_{4,d}$ that binds DNA as a function of damage.

4.7 WT UV-DDB IS PERSISTENT ON UV-DAMAGED DNA AND SLIDES AT HIGH IONIC STRENGTH

In addition to the intermediates detected above, the tightrope platform also revealed a population of molecules which persisted during the entire observation window of 900 seconds (Figure 4.10 C, mov 2). For this class of molecules, the association time was indeterminate since individual persistent molecules were found to persist up to 90 minutes (data not shown).

4.7.1 Persistent UV-DDB are enriched on damaged DNA

Importantly, these persistent molecules represent a distinct, stable intermediate (S_5) that does not decay according to any of the T_1, T_2, T_3 or T_4 processes and is not a remnant of the slow T_3 or T_4 process.

From our previous analyses, we identified that the number of molecules that participate in T_3 and T_4 decay processes for undamaged DNA is: 21 and 16 respectively, while that for UV-damaged DNA is: 87 and 67 respectively. Using the values of the rate constants measured for these processes, the number of molecules that will survive for 900 s can be calculated to be <1 for the first three cases and 4 for the last case. In comparison, the total number of persistent molecules was found to be 20 (out of 347 observations) for binding to undamaged DNA, and 88 (of 643 observations) for UV-damaged DNA. This tremendous enrichment in the number of persistent molecules strongly suggests that these molecules participate in a process (assume Poisson; T_5 - with a mean lifetime of τ_5) which is far slower

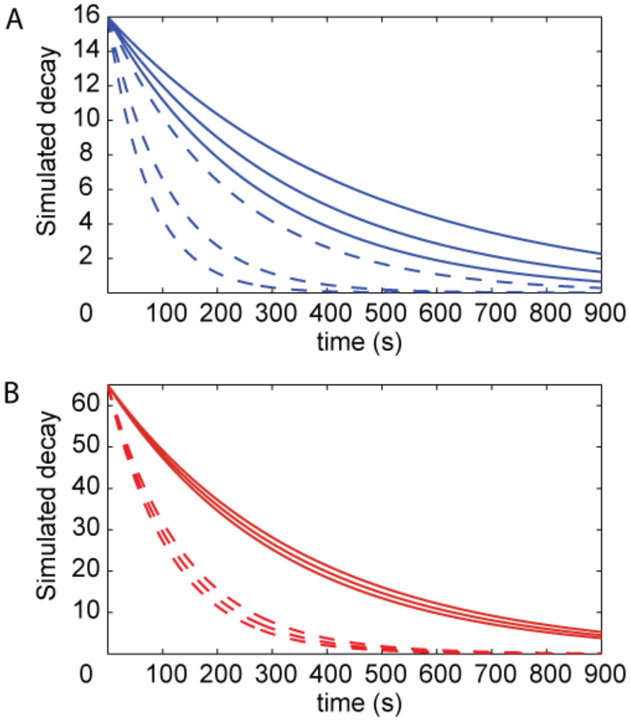


Figure 4.17: Simulated decay of longlived UV-DDB

Simulated decay of population of molecules dissociating with mean lifetimes of τ_3 (red) or τ_4 (blue) on A. undamaged DNA and B. UV-damaged DNA. Each set of three curves corresponds to the simulated decay assuming a time constant of τ or tau_+ corresponding to the upper limit of the 95% CI or τ_- corresponding to the lower limit of the 95% CI.

than any of the processes that are measureable in our experiments.

4.7.2 Persistent UV-DDB molecules exhibit salt dependent mobility

To eliminate the possibility that the persistent UV-DDB molecules observed in the DNA tightrope assay were irreversible aggregates of UV-DDB on DNA, we washed the chamber with four chamber volumes of a high salt buffer. After the buffer exchange, we observed a significant loss (>90%) of DNA binding events due to dissociation in the presence of high ionic strength. Of the few molecules that remained on the tightropes, we observed that previously stationary UV-DDB on DNA now performed a rapid 1D random walk on DNA (mov 3, Figure 4.18 A) with an average diffusivity of 0.21 (±2.1) $\mu\text{m}^2\text{s}^{-1}$ (geometric mean ± geometric standard deviation, where D is the diffusivity, equivalently this represents the geometric mean; n=31, Figure 4.18 B). Importantly, the mean uncorrected diffusion constants we measured were an order of magnitude greater than the upper limit for rotational diffusion proposed by Schurr [138], suggesting that UV-DDB does not rotationally track the DNA when sliding under conditions of high ionic strength. Further, all of these complexes were found to be stable on DNA for at least the duration of acquisition (120 s), suggesting that these complexes possess a molecular topology which constrains them to the DNA.

The percentage of these observations increased 2.5 fold from 5.5 % for undamaged DNA to 13.7 % for UV-damaged DNA. This dependence on the presence of UV-damage in DNA indicates that these persistent molecules are on a kinetic pathway toward stable damage discrimination.

4.8 CREATION OF DNA DAMAGE ARRAYS: UV-DDB CO-LOCALIZES AND IS PERSISTENT AT SITES OF LESIONS

In our work, we have identified five kinetic intermediates S_1, S_2, S_3, S_4, S_5; in increasing order of stability in the interaction of UV-DDB with undamaged and UV-damaged DNA. Additionally, we have measured the kinetics of dissociation from each of these states to the

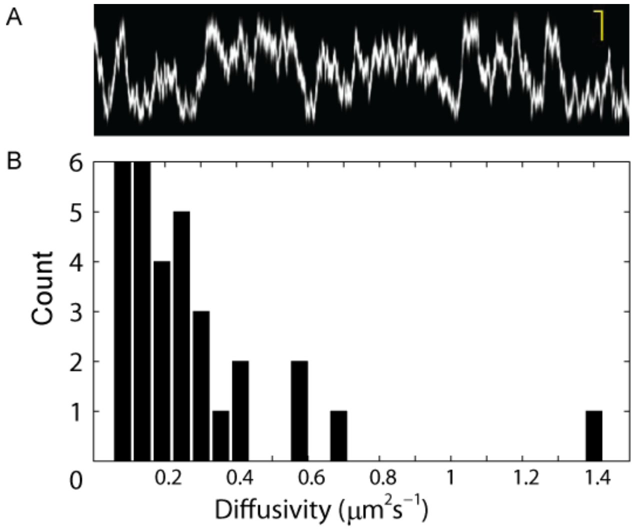


Figure 4.18: Effect of high ionic strength on UV-DDB interaction with DNA

A. Representative kymograph of UV-DDB sliding on DNA at high ionic strength. B. Histogram of diffusion constants for UV-DDB sliding on DNA at high ionic strength. Horizontal scale bar represents 1s and vertical scale bar represents 2 μm (n = 31).

DNA-free state S_0. The experiments so far have not revealed the molecular configurations of UV-DDB bound to DNA in each of these states. We next probed whether dimeric UV-DDB participated in the kinetic pathway was represented in any of these intermediates. Further we wanted to directly test whether long-lived UV-DDB arises from interactions with sites of damage.

To better understand the nature of the long-lived UV-DDB complexes we created substrates with engineered lesions at defined sites (see section 2.3.1.3). To this end, we introduced an oligonucleotide containing an abasic site analog with a 3' biotin modification ('APbiodT') within a plasmid using a previously described protocol [114]. Linearized plasmids were tandemly ligated to form long damage arrays (>40 kb) which were strung up between beads in the DNA tight-rope assay. Upon incubation with SA QDs, we were able to observe QD arrays on these substrates marking sites of lesions (Figure 4.19A).

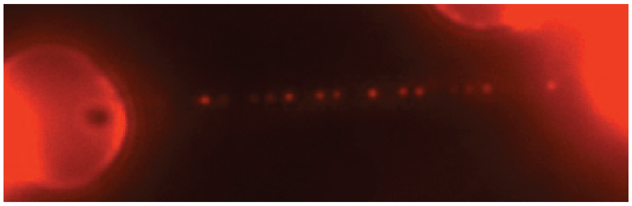


Figure 4.19: Use of defined lesion substrates in DNA tightrope assay

Oblique angle fluorescence microscope image of array of quantum dots on DNA tightrope of a long DNA molecule containing a defined abasic site analog with a proximal biotin marking the site of the lesion.

A histogram of pair-wise distances between QDs revealed a series of peaks centered at lengths consistent with integral multiples of the linearized plasmid length of 0.65 μm (Figure 4.20 A). We were able to fit the histogram of inter-QD distances for pairs of adjacent QDs in a DNA damage array to a multiple Gaussian fit (up to three terms, see Figure 4.20 B) and found that 46.7% of inter-QD distances on these DNA damage arrays were 0.697 μm apart (95% CI: (0.6899, 0.7044)), 37.8% were 1.348 μm (CI:(1.336, 1.359)) and 8.9% were 2.076 μm (95%CI: (2.022, 2.131)) apart (Figure 4.20 C). The measured distances between adjacent QDs on DNA agree very well with the predicted distances between two QDs if they are one (0.65 μm) or two (1.30 μm) or three (1.95 μm) plasmid lengths apart. Thus, 85% of the quantum dots were either one or two plasmid lengths apart (Figure 4.20 C). We believe that the missing sites reflect a lack of QD conjugation or dark quantum dots rather than sites where the oligonucleotide is not incorporated as restriction digests of the parental damage engineered plasmid indicated that >99 % had the oligonucleotide correctly incorporated (data not shown). These data demonstrate that DNA damage can be introduced and specifically marked in these DNA tight-rope substrates efficiently.

To investigate whether long-lived UV-DDB molecules bound DNA at sites of DNA damage, we performed dual color experiments involving the incubation of QD-UV-DDB conjugates with QD conjugated APbiodT DNA tightropes. We found persistent UV-DDB co-localized to sites of damage (n = 21, Figure 4.21, mov4). Significantly, this general method of creating long DNA substrates with site specific modifications provides new possibilities for

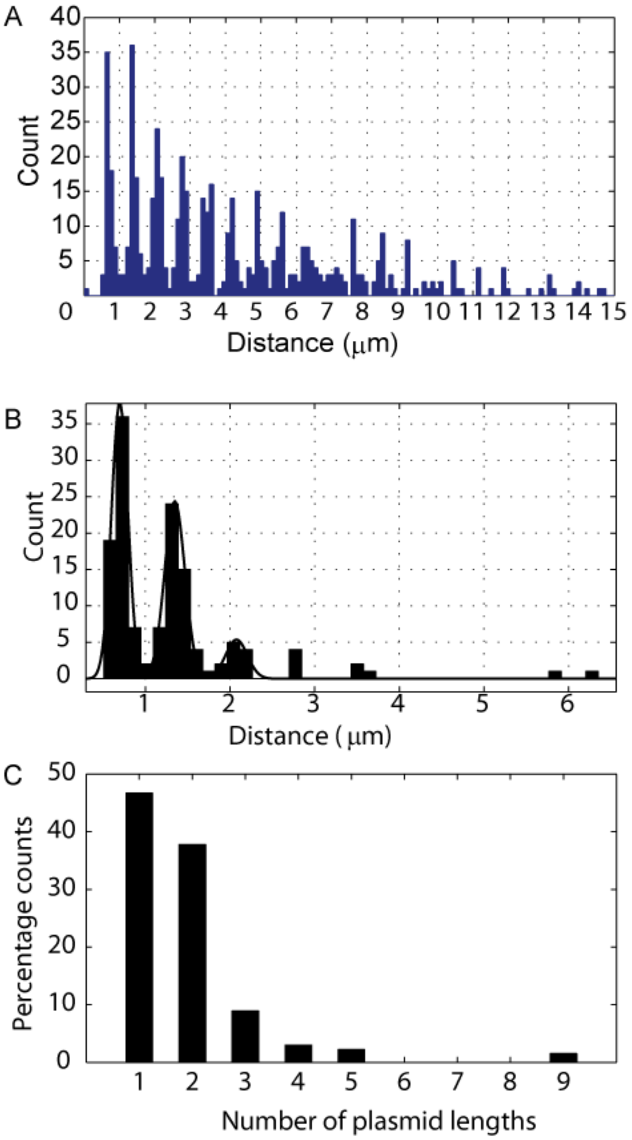


Figure 4.20: Characterization of defined DNA damage arrays

A) Histogram of pairwise distances between QDs marking sites of introduced lesion. B. Gaussian fit to the histogram of measured distances between adjacent QDs on DNA substrate. C. Histogram of percentage counts of plasmid lengths between adjacent QDs on DNA tightrope containing defined lesion.

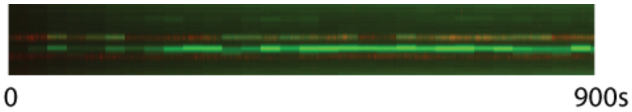


Figure 4.21: Persistent co-localization of QD-UV-DDB with QD marking sites of DNA damage

Kymograph of UV-DDB (red) co-localized to sites of defined lesion (green) for 900 s (see mov4) indicating the formation of long lived UV-DDB at sites of damage. 1 pix = 40 nm along y-axis.

studying site specific interactions of DNA repair proteins in the DNA tight-rope platform.

4.9 DIMERIC UV-DDB IS PERSISTENT AT SITES OF DAMAGE

We next sought to understand the molecular nature of the kinetic intermediates found for UV-DDB binding to DNA. To address the nature of UV-DDB stoichiometry, we incubated together UV-DDB molecules which had been separately labeled with two differently colored QDs with UV-damaged DNA tight-ropes. We found that, co-localization of both colors could be observed, representing dimerization of WT-UV-DDB (mov5, Figure 4.22 A). Further for 72% of observations (n = 18), co-localization was found to persist during the entire observation window of 900s, suggesting that, UV-DDB dimers are long lived on UV-damaged DNA. Further evidence for persistent UV-DDB dimers was obtained from experiments involving the incubation of dual colored UV-DDB with APbiodT substrates. Even in this case, we were able to observe long lived UV-DDB dimers (n = 18, mov5, Figure 4.22 B).

4.10 UV-DDB (K244E) DIMERIZES AND SLIDES ON DNA

To gain insight into the nature of the complex binding kinetics of WT UV-DDB, we turned our attention to a XP causing mutant of UV-DDB containing a lysine to glutamate mutation

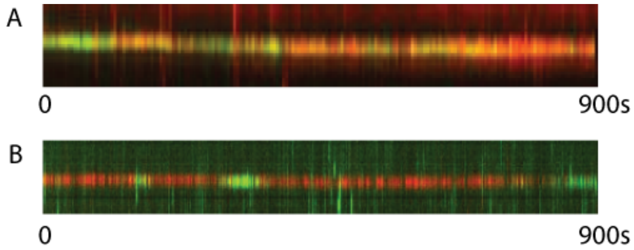


Figure 4.22: Co-localization of UV-DDB monomers on DNA substrates containing lesions

Dual color kymographs of long-lived UV-DDB dimer on A) UV-damaged DNA and B) APbiodT substrate respectively. 1 pix = 40 nm along y-axis.

in position 244 in DDB2. The K244E mutation affects a crucial DNA binding residue and results in a greatly reduced affinity and specificity for damage [18, 121].

We probed the DNA binding ability of recombinant UV-DDB (K244E) in a pull-down experiment (see Figure 4.23). In this experiment, UV-DDB or UV-DDB (K244E) were incubated with UV-irradiated or undamaged bio-517-dig bp DNA (this is a fragment of DNA containing a biotin on one end and a digoxigenin on the other end, see Figure 4.23) bound to streptavidin coated magnetic bead and blocked (or not) on the other end with an anti-dig antibody (Figure 4.23). Pulled-down UV-DDB or UV-DDB (K244E) was then probed using α-DDB1 and α-DDB2 antibodies.

We found that, both WT and mutant UV-DDB (K244E) exhibited strong end binding, which was lost upon blocking with the anti-dig antibody (compare lanes 1,2 and 5,6 and 3, 4 and 7, 8, Figure 4.24). Further, WT UV-DDB exhibited a 4 fold increased binding to UV-damaged DNA containing on average 1 photoproduct per fragment (compare lane 5 vs. lane 6, Figure 4.24). In contrast, UV-DDB (K244E) failed to exhibit any specific binding for UV-damaged DNA (compare lane 7 vs. 8, Figure 4.24). These results are consistent with a previous report [121], which reported that the UV-DDB (K244E) mutant lacks the ability to discriminate UV-damage in DNA.

This finding was further confirmed by atomic force microscopy experiments in which UV-DDB (K244E) was incubated with undamaged 517 bp DNA fragments (Figure 4.25). Consistent with the results of the pull-down experiment, we found that, UV-DDB (K244E)

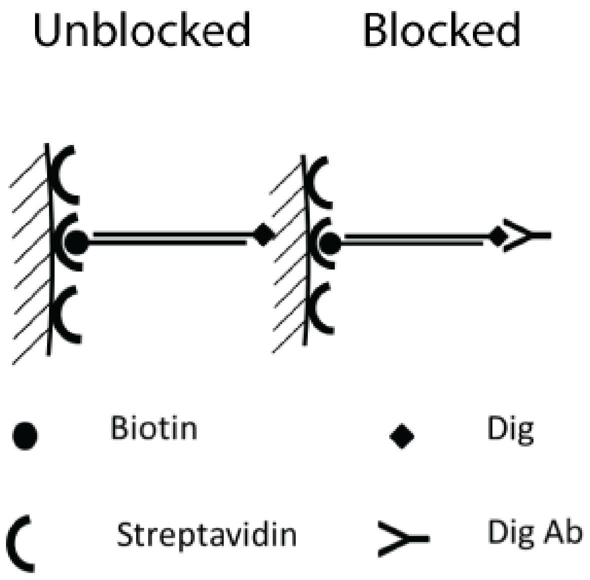


Figure 4.23: Schematic of pulldown experiment.

Schematic of pulldown experiment indicating either undamaged or UV-damaged 517 bp DNA containing a biotin and a digoxigenin on either end, bound to streptavidin coated beads. This experiment allows for the pulldown of WT or mutant UV-DDB from undamaged or UV-damaged DNA with free or blocked ends.

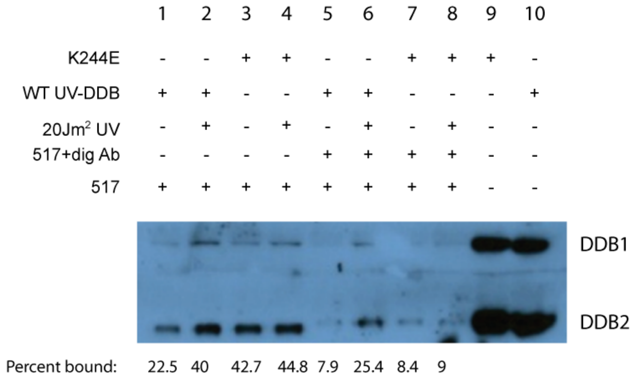


Figure 4.24: Western blot of pulldown experiment.

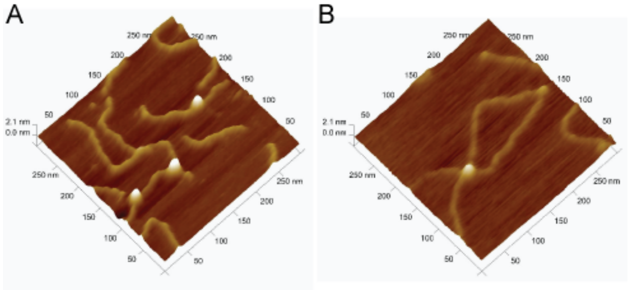


Figure 4.25: AFM images of UV-DDB (K244E) bound to undamaged DNA

AFM image of UV-DDB containing K244E mutant of DDB2 bound to A. 517 bp undamaged DNA B. pUC18 plasmid DNA.

was capable of binding DNA at ends and non-specifically over the length of the DNA. A comparison of percentage of binding events to DNA ends vs. internal sites revealed marginal preference for ends over internal sites (55% vs. 45%, n = 410 binding events, Figure 4.26). We observed that the UV-DDB (K244E) could bind a single DNA molecule non-specifically (20%), a single DNA end (6%), two DNA ends (16%), one DNA end and an internal site forming a three-way junction (50%), two DNA molecules at internal sites forming four-way junctions (9%). 74% of the bound UV-DDB (K244E) (see below) participated in binding to two DNA molecules compared to 26% bound to single DNA molecules (n = 216 UV-DDB (K244E)-DNA complexes). Interestingly, dimeric WT UV-DDB was found to bind two DNA helices only 20% of the time [102].

To obtain dynamic information describing the binding of UV-DDB (K244E) to DNA, we incubated QD-UV-DDB (K244E) with undamaged λ-DNA in the DNA tightrope assay and found that, in stark contrast to the WT UV-DDB, UV-DDB (K244E) showed sliding behavior. We performed a CRTD analysis on UV-DDB (K244E) and found that 69% of UV-DDB (K244E) dissociated from DNA with a mean lifetime of 1.9s (1.7s, 2.1s), whereas 31% dissociated from DNA with a mean lifetime of 34.7 s (30.4s, 40.4s) (n = 99, Figure 4.27). UV-DDB (K244E) was also found to slide on APbiodT DNA (mov6, Figure 4.28). We calculated diffusion constants for UV-DDB (K244E) molecules which were found to slide on DNA for the entire observation window of 120s. Sliding QD-UV-DDB (K244E) exhibited heterogeneity

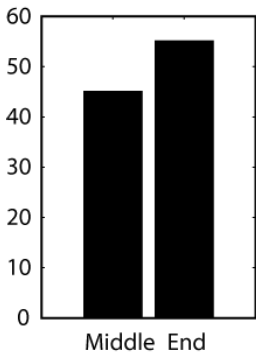


Figure 4.26: Quantification of end vs. non-end binding of UV-DDB

Plot of end binding vs. non-end binding of UV-DDB (K244E) on 517 bp DNA.

in its diffusive behavior spanning three orders of magnitude (Figure 4.29). Importantly, the diffusive behavior of UV-DDB (K244E) on DNA (both undamagedλ-DNA and APbiodT) was found to have a mean diffusion constant of 0.11 ± 0.2 $\mu m^2 s^{-1}$, with an anomalous diffusive exponent (α) of 0.5 ± 0.22 (n = 31, Figure 4.29). This subdiffusive behavior of UV-DDB (K244E) with DNA suggests that UV-DDB (K244E) retains interactions with the DNA as it diffuses on the DNA.

Unexpectedly, the DNA tightrope assay provided further insight into the interaction of UV-DDB (K244E) with DNA. We identified, albeit rarely, events which suggest that dimerization of UV-DDB (K244E) is DNA dependent and proceeds via random collisions of UV-DDB (K244E) molecules on DNA (mov7, Figure 4.30). This observation prompted us to examine the stoichiometry of UV-DDB (K244E) bound to DNA. We have previously used AFM to identify the stoichiometry of WT UV-DDB bound to DNA using a calibration curve relating the AFM volume of the complex to its molecular weight (Figure 4.31) [102]. Volume analysis of DNA bound UV-DDB (K244E) revealed a peak at 564.3 ± 10.1 nm^3 corresponding to a molecular weight of 388.6 ± 11.8 kDa, consistent with that of dimeric UV-DDB (K244E) bound to DNA (n = 171, Figure 4.32). These data demonstrate that a mutation in the DNA binding interface of DDB2 does not inhibit dimerization of UV-DDB. However, in contrast to WT UV-DDB [102], dimeric UV-DDB (K244E) is more likely to be stabilized on two DNA molecules in a damage independent manner. In summary, our

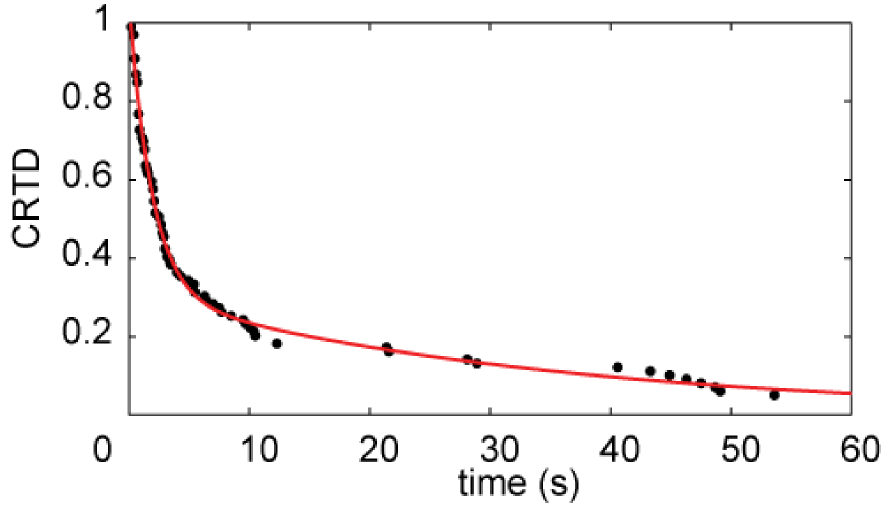


Figure 4.27: CRTD of UV-DDB (K244E) on DNA

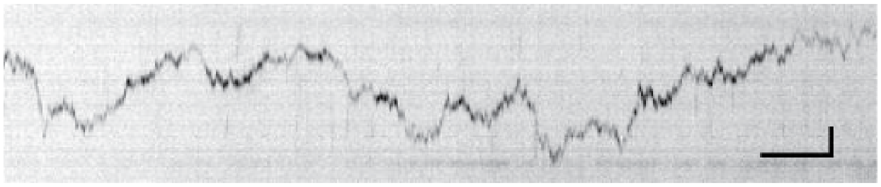


Figure 4.28: Kymograph of diffusing UV-DDB (K244E) on APbiodT

UV-DDB (K244E) slides on AP-biodT DNA substrate without apparent pausing at damage sites. Scale bars represent 5 s along the x-axis and 1 μm along y-axis.

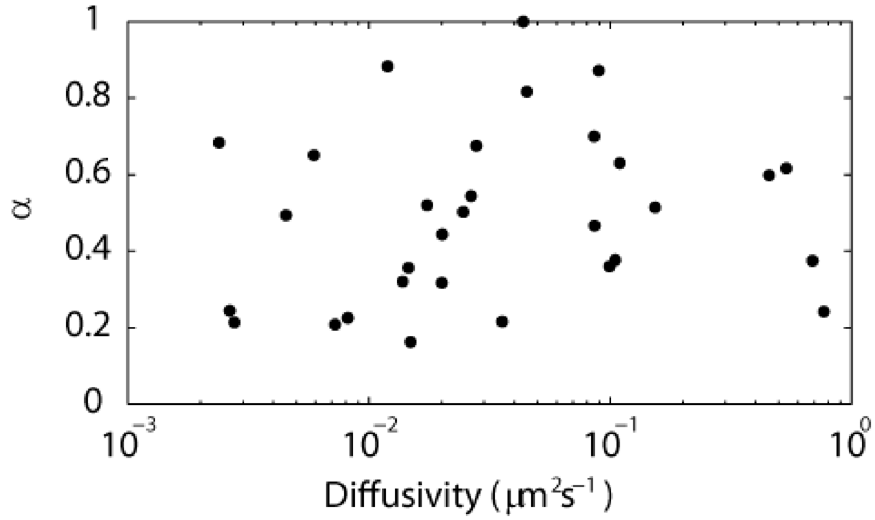


Figure 4.29: Plot of anomalous diffusive exponent (α) vs. Diffusivity for UV-DDB (K244E) sliders

study indicates that UV-DDB (K244E) retains the ability to dimerize on DNA, but lacks specificity for lesions and consequently slides on DNA. This dimerization proceeds via the formation of kinetic intermediates which are distinct from those that describe dimerization of WT UV-DDB.

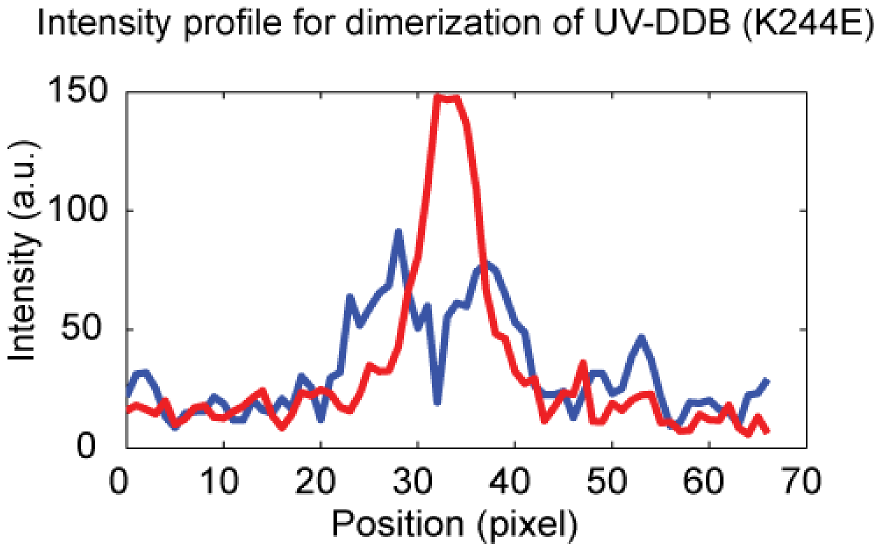


Figure 4.30: Dimerization of UV-DDB (K244E)

DNA dependent dimerization of UV-DDB (K244E) showing the intensity of individual QD-UV-DDB (K244E) before and after dimerization.

4.11 DISCUSSION

In this study, we sought to understand the initial recognition step of NER by using single molecule techniques to examine the dynamics of UV-DDB interactions at sites of damage without ensemble averaging. We have identified that UV-DDB consisting of DDB1 and DDB2, performs a 3D search for damage sites in DNA. Using DNA substrates containing UV-induced photoproducts or AP sites, we discovered a complex kinetic pathway of damage recognition by UV-DDB that culminates in the formation of long-lived dimers of UV-DDB at sites of damage. Additionally, we studied the stoichiometry and dynamics of the XP causing K244E mutant of DDB2 on DNA and identified that UV-DDB (K244E) dimerizes and slides on DNA, but does not associate to damaged sites.

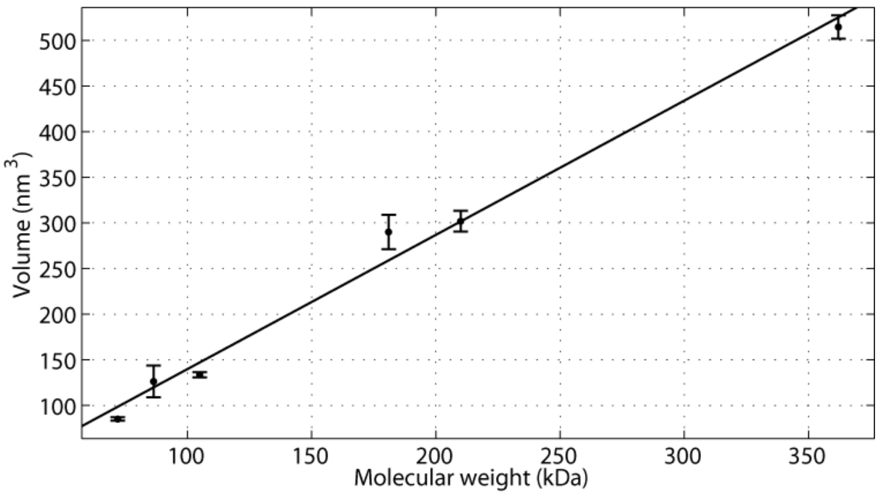


Figure 4.31: Calibration curve relating the molecular mass of a protein to its measured AFM volume

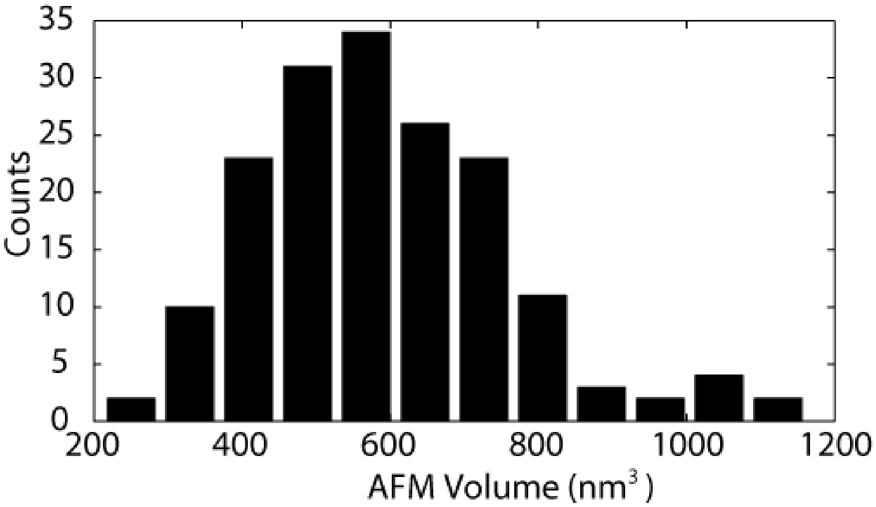


Figure 4.32: Volume histogram of UV-DDB containing the K244E mutant of DDB2

4.11.1 Jumping as a mechanism for target search implications for search

In the work presented here, we have directly visualized jumping of UV-DDB from one λ-DNA molecule to another in real time using a single molecule DNA tightrope assay (mov1). Of the total DNA bound WT UV-DDB molecules observed, less than 2% showed perceptible diffusion. Our work provides direct evidence for a 3D search mechanism and rules out 1D sliding mechanisms as the primary means for target search. Multiple reports have demonstrated that in vivo, UV-DDB exists as part of a larger E3 ligase complex bound to the COP9 signalosome ($CSN - CRL4^{DDB2}$) [31, 122]. In this complex, DDB2 is the DNA damage binding factor and damage recognition is not known to be determined by the other factors. Based on this work, we propose that the $CSN - CRL4^{DDB2}$ complex also performs a 3D search in vivo. The presence of other DNA binding proteins and higher order chromatin organization may serve as a barrier for search if a sliding mechanism were the primary mode of target site location. Rapid sampling of exposed DNA by a 3D search mechanism minimizes the need to overcome such obstacles. Indeed, it has been demonstrated that nucleosomal core particles containing site specific UV-photoproducts exhibit spontaneous unwrapping of the DNA via a mechanism that is thought to provide access to damage recognition enzymes [139]. Our findings on the association of UV-DDB with naked DNA are consistent with a recent report on the promoter search employed by *E. coli* RNA polymerase which posits that target search is favored via a 3D diffusion mode, over a facilitated mode of diffusion in a concentration dependent manner [130]. Since UV-DDB exists in about 180,000 copies per human cell nucleus [18], we propose that UV-DDB rapidly surveys the genome employing 3D diffusion and examines the DNA for damaged bases in discrete kinetic steps in an excess of undamaged DNA.

4.11.2 Damage recognition is a multi-step kinetic cascade culminating in UV-DDB dimerization

Analysis of data describing the association lifetimes of UV-DDB with either undamaged or damaged DNA substrates revealed five different repair intermediates $\{S_{1,ud}$, $S_{2,ud}$, $S_{3,ud}$, $S_{4,ud}$, $S_{5,d}$; and their damaged DNA counterparts$\}$, which decay to the DNA free state

$\{S_0\}$, according to five different decay processes $\{T_{1,ud}, T_{2,ud}, T_{3,ud}, T_{4,ud}, T_{5,ud}$; and their damaged DNA counterparts$\}$ on each of the DNA substrates. In our experiments, the $\{T_{1,ud}, T_{2,ud}, T_{3,ud}, T_{4,ud}$; and their counterparts on damaged DNA$\}$ decay rates were measurable whereas the T_5 decay rate was found to be slower than the time scale of data acquisition. These decay processes represent the set of measurable decay processes in our setup. It is possible that there exist intermediates outside our detection range.

Of the measurable processes, the T_1 process was found to possess a significantly different decay rate for dissociation of UV-DDB from undamaged DNA compared to UV-damaged DNA. This means that the $S_{1,ud}$ state decays to the S_0 state at a slower rate than that at which the $S_{1,d}$ state decays into the S_0 state. Statistical analyses revealed that the T_2, T_3 and T_4 decay rates were identical for dissociation from undamaged or damaged DNA. This means that each of the pairs of decay processes $\{S_{2,ud} \to S_0, S_{2,d} \to S_0\}, \{S_{3,ud} \to S_0, S_{3,d} \to S_0\}$ and $\{S_{4,ud} \to S_0, S_{4,d} \to S_0\}$ occur at the same rate. However, the number of molecules participating in these processes was very different for undamaged vs. damaged DNA (Figure 4.14). 0.65X fewer complexes participated in the T_2 decay process for damaged DNA compared to undamaged DNA (Figure 4.14). On the contrary, 3X more UV-DDB-DNA complexes decayed according to each the T_3 and T_4 decay process in the presence of damage respectively (Figure 4.10). Since the decay rates for the pairs of process $\{S_{2,ud}, S_{2,d}\}, \{S_{3,ud}, S_{3,d}\}$ and $\{S_{4,ud}, S_{4,d}\}$ show no dependence on the presence of damage, we can assume that they arise from isoenergetic repair intermediates. Under this assumption, we find that the 12% loss of molecules in the S_2 state on damaged DNA corresponds to an enrichment of the S_3 state (Figure 4.14), indicating that the T_2 process is a damage verification process, while the T_3 process is a damage stabilizing process. Some of these longer lived intermediates may arise from endogenous damage in commercially available λ-DNA (for example from depurination) that is recognized by UV-DDB [18].

Some of the slower decay processes are consistent with previously available bulk estimates. Bulk estimates of the kinetics of dissociation of UV-DDB from UV-damaged DNA in cellular chromatin are difficult to obtain because of the complex and unclear interaction between UV-DDB and XPC and its dependence on the ubiquitylation of each of these factors. Kinetic studies of C-terminally tagged murine DDB2-EYFP in XP20MA cells (XP-C cells) revealed a half-life of 110s for dissociation from UV-damaged chromatin [122]. Previous biochemical analysis of the dissociation kinetics of recombinant UV-DDB from photoproducts were found to be lesion dependent with a slow off rate for a high affinity lesion such as the 6-4 PP ($8.1 \times 10^{-4}s^{-1}$) and faster off-rates for the Dewar isomer of the 6-4 PP and T[t,s]T (2.9×10^{-3} and $3.7 \times 10^{-3}s^{-1}$ resp.) and even faster off-rates for the dissociation from T[c,s]T photoproducts ($4.6 \times 10^{-2}s^{-1}$) [140]. However, in these studies, the estimated rate constants represent bulk dissociation rates, from not just the site-specific lesion, but also end binding, irrespective of the oligomeric state of UV-DDB bound to the DNA substrate or in the presence of other interacting partners or post-translational modifications that may influence these estimates.

These kinetic intermediates progress along a kinetic cascade and culminate in the formation of an enriched and distinct population of persistent molecules on UV-damaged DNA. Additionally, we found that dimeric UV-DDB was also persistent and that the persistent molecules were not irreversible aggregates of UV-DDB on DNA. We believe that the dimeric molecules observed in this work are structurally consistent with the dimeric UV-DDB structure we elucidated recently and represent specific binding at sites of lesions [102]. While the crystal structure suggests that UV-DDB requires two DNA molecules to dimerize on DNA, our AFM data reveal that dimeric UV-DDB can form even on single DNA molecules containing damage [102]. In fact, AFM experiments revealed that only 1 in 5 complexes were composed of dimeric WT UV-DDB mediating the binding of two DNA molecules [102]. In contrast, 76% of UV-DDB complexes containing the K244E mutant of DDB2 on DNA were found to be dimeric, and bound to two molecules of DNA(Figure 4.25). These findings indicate that while the presence of two DNA molecules in the UV-DDB DNA complex is sufficient for dimerization of UV-DDB, it is not a necessary requirement.

We predict that these long-lived dimers of UV-DDB at sites of lesions could inhibit the progression of NER if these highly stable UV-DDB dimers are not dismantled. In support of this hypothesis, introduction of excess recombinant UV-DDB to in vitro reconstituted NER reactions resulted in inhibition of repair of 6-4 PPs [17]. Studies of fluorescently tagged UV-DDB binding in vivo have reported immobile binding of UV-DDB to DNA for up to 4 h in XP-A cells [134]. In a study involving siRNA knock-down of Cullin4A, fluorescently tagged UV-DDB exhibited delayed disappearance from CPD foci in HeLa cells [141]. Inhibition of the proteasome using MG132 also resulted in inhibited recruitment of XPC to sites of CPD lesions in mammalian cells [142]. We believe that the highly stable, persistent, dimeric UV-DDB complexes detected in our studies represent a distinct species on the kinetic pathway to recognize damage with high specificity and affinity. How post translational modifications of DDB2 such as phosphorylation [143], SUMOylation [144], ubiquitylation [29] and PARylation [125, 126], as well as, interacting partners such as XPC and XPA [145] influence the intermediates of UV-DDB and the transition rates measured in our system will be investigated in future experiments.

4.11.3 Damage recognition involves dynamic conformational changes in both UV-DDB and DNA

What might be the physical basis of this heterogeneity observed in the lifetimes of the repair intermediates? Crystal structures of UV-DDB in the apo and DNA damage bound forms reveal some of these states (PDB ID: 3EI1, 4A0A, 4A0K, 4E54). The protein in the apo state (PDB ID: 3EI4) upon binding to damaged DNA undergoes an FQH-hairpin (F334-Q335-H336) transition which probes the major groove of the DNA for the presence of damage (see Figure 4.33; PDB ID: 4E45, see mov8) [25, 16, 102].

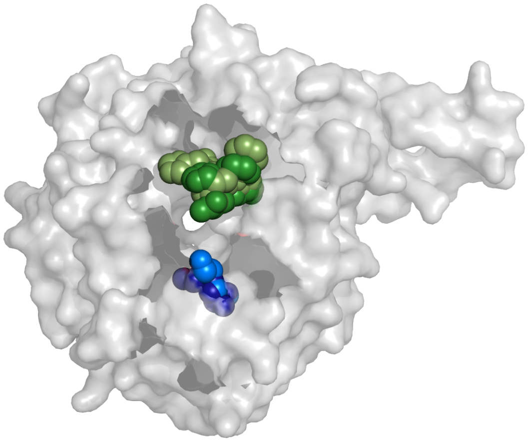


Figure 4.33: Model of DDB2 in the apo and DNA bound state. Here the FQH motif is shown in light green (apo structure, PDB ID: 3EI4) and dark green (DNA bound conformation, PDB ID: 4E54). Similarly, the K244 residue is shown in light blue (apo structure, PDB ID: 3EI4) and dark blue (DNA bound structure, PDB ID:4E54).

At sites of damage, this conformational change in DDB2 is accompanied by base-flipping of the damaged bases in DNA (consisting of the photoproducts in the case of UV damage, alternatively the abasic site and the adjacent 3' base) to an extrahelical conformation and stabilization in the lesion binding pocket of DDB2 (PDB ID: 3EI1, 4A0A, 4A0K; see Figure 4.34) [25, 16]. Stable damage recognition is thought to induce folding of the intrinsically disordered N-terminus an α-paddle structure, which along with the β-wing, forms a winged helix structure upon DNA binding (PDB ID: 4E54; see figure 4.35) [102]. Damage recognition may be considered to progress along a reaction coordinate which describes a series of dynamically interconverting structural intermediates. Some of the highly transient, short lived binding intermediates observed here might reflect abortive attempts at damage recognition by UV-DDB. These species may correspond to metastable intermediates that participate in varying extents of lesion engagement, failing to stabilize at sites of damage. Indeed, previous work has demonstrated that the assembly and disassembly of subunits of large macromolecular complexes such as the spliceosome proceeds via a kinetic pathway which rejects non-productive sub-complexes along the reaction coordinate [146].

Heterogeneous lifetimes may arise from the nature of the lesion or the repair intermediate. Indeed, different lesions are differentially recognized by UV-DDB [140, 18] and this may explain some of the multiple long-lived populations observed here. Future work will focus on examining the interaction of UV-DDB with damaged DNA substrates containing defined 6-4PPs and CPD lesions using the DNA damage arrays and the tightrope assay described here.

4.11.4 Specific damage recognition depends on K244 switching

We have tested a human disease causing mutant of UV-DDB (K244E) and found that this mutant still maintained the ability to dimerize and displayed residual DNA binding. However, as opposed to WT UV-DDB, UV-DDB (K244E) mutation exhibited sliding on the DNA. This finding suggests that proofreading of damaged DNA is contingent upon the successful sandwiching of the undamaged base 3' to the two damaged bases, between the FQH hairpin and K244 (see figure 4.33, 4.34).

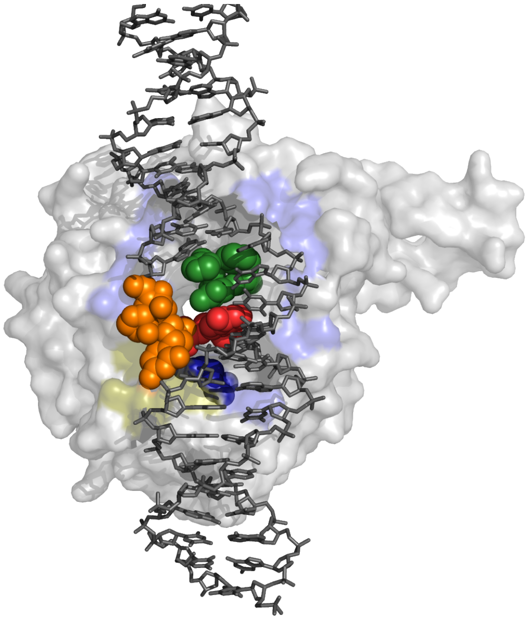


Figure 4.34: Model of DDB2 in the DNA bound state. Here the abasic site analog and the adjacent base to it constitute the lesion and are shown in orange. The DNA is bound to DDB2 via a constellation of electrostatic interactions shown in light blue. The FQH lesion (green) occupies the interhelical space vacated by the flipped out lesion (orange). The undamaged base adjacent to the lesions is shown in red and stacks with the glutamine residue and is pinned between the FQH motif and the K244 (dark blue spheres).

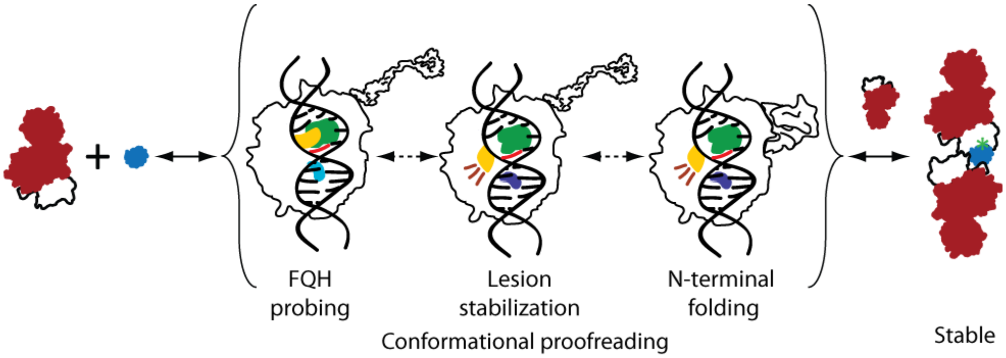


Figure 4.35: Model of damage recognition by DDB2. See description in text.

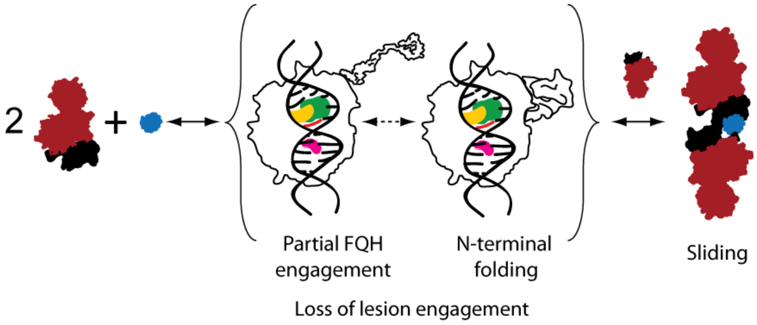


Figure 4.36: Model of UV-DDB (K244E) binding to DNA. UV-DDB (K244E) dimerizes on DNA but fails to stably engage the lesion due to the positive charge of the glutamate residue. See description in text.

Dimerization of UV-DDB (K244E) probably occurs by rapid 3D diffusion of one UV-DDB molecule colliding with another the long-lived UV-DDB molecule sliding on the DNA, trapping this dimer in a topologically constrained complex on the DNA, which is not actively engaged in a damage detection conformation (Figure 4.36). The sub-diffusive nature of the sliding indicates that this complex performs a constrained Brownian walk on the DNA, suggestive of interactions with it [131]. This explains the observations by us and others, that UV-DDB (K244E) retains weak DNA binding ability, but lacks specificity for damage [121, 18].

4.11.5 Conformational proofreading is a candidate mechanism for damage recognition

As evidenced from the crystal structures, both UV-DDB and the DNA undergo a series of concerted conformational changes that ensure successful damage recognition (PDB ID: 3EI4, 3EI1, 4A0A, 4A0K, 4E54). This problem of damage recognition by UV-DDB falls under a category of problems in molecular recognition in which proteins utilize conformational changes to achieve highly specific recognition in a noisy environment. This mechanism termed conformational proofreading [109] utilizes a 'structural mismatch' between the protein binding pocket and the ligand such that binding of the correct ligand facilitates a conformational change in the protein which stabilizes the binding, while the incorrect ligand is unlikely to allow this conformation change and is therefore rejected. Conformational proofreading operates in the absence of energy consumption and is an alternative to kinetic proofreading [109].

A working model for damage recognition by UV-DDB is presented here (figure 4.37). Each of the states S_i can be considered to be comprised of a protein configuration P_i and a DNA configuration D_i. Here, the apo protein and the DNA interconvert between ensembles of conformers $\{P_i\}$ and $\{D_i\}$. Collisions between the protein and DNA result in the formation of interconvertible repair intermediates $\{S_i\}$, some of which are relatively stable and observable (such that their decay to the DNA-free state $\{S_0\}$ is measurable). In this work, we have provided evidence for the existence of five such intermediates and measured the decay from four of these states. We propose that during the initial stages of damage recognition by UV-DDB, target specificity arises from the ability of the repair intermediate to cross energy barriers between the various states, whereas additional affinity arises from the dimerization of UV-DDB, which then locks the repair factor to the site of damage.

In our system, both UV-DDB and the DNA exhibit candidate structural mismatches - the FQH open-close transition and the intrahelical-extrahelical base flipping transitions respectively. While the FQH transition is necessarily required for probing, the specificity for damage arises from the deformability and stabilization of the damaged bases in the lesion binding pocket. We propose that the critical ‘structural mismatch’ corresponds to the ability of the DNA to deform. Given the high penalties of flipping undamaged DNA bases compared to damaged DNA bases, an undamaged patch of DNA which is being proofread by UV-DDB is unlikely to access the base-flipped, deformed DNA configuration, and this is the structural mismatch which inhibits stable binding by UV-DDB at sites of undamaged DNA.

Our work reveals that even a relatively simple step of damage binding is highly regulated even in the absence of protein partners and supports the hypothesis that multiple layers of damage recognition and verification are needed before the final commitment to repair DNA is made [9]. Previous work has suggested that damage verification in NER proceeds via an ATP-dependent kinetic proofreading mechanism performed by the XPD helicase in TFIIH

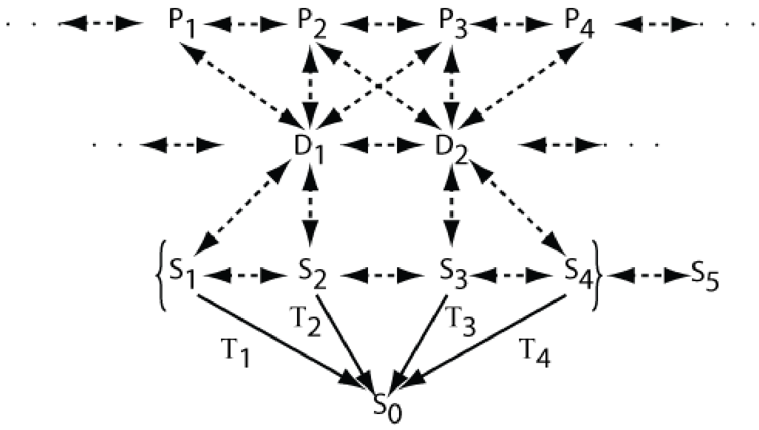


Figure 4.37: Conformational proofreading model for molecular recognition by UV-DDB. P_i represent conformations of protein, D_i represent conformations of DNA and S_i represent protein-DNA complexes. T_i represent the decay processes that are measured in this work. See description in text.

[107, 110]. Here, we propose that prior to kinetic proofreading by TFIIH, damage recognition by UV-DDB proceeds via the formation of multiple repair intermediates in a kinetic cascade, via a mechanism which resembles conformational proofreading [109]. Further, we hypothesize that this conformational proofreading is a common feature of damage recognition in the absence of energy consumption and is also employed by the XPC complex to discriminate damage. In this regard it is interesting to note that conformational and kinetic proofreading mechanisms have been found to operate together for highly specific recognition of homologous sequences during homologous recombination [147, 111]. We believe that this synergy of damage detection mechanisms is required for the successful navigation of the complex kinetic and thermodynamic landscape of DNA damage recognition while achieving high specificity by rejecting non-optimal repair intermediates. Future studies will help reveal whether the combination of proofreading mechanisms is a universal feature of DNA damage recognition.

5.0 CONCLUSION AND FUTURE WORK

This work has identified a previously unknown and unanticipated complexity in damage recognition by the UV-DDB protein. We found that damage recognition by UV-DDB proceeds along a kinetic pathway comprised of multiple intermediates and culminates in the formation of long-lived dimeric UV-DDB on DNA at sites of lesions. These findings raise several questions and testable hypotheses that can elucidate the initial stage of damage recognition by UV-DDB.

1. What is the contribution of lesion chemistry on heterogeneity observed in the intermediates? We identified at five measurable intermediates of UV-DDB on DNA. Our assays do not reveal the nature of these intermediates. Some of the heterogeneity in the mean lifetimes of these intermediates may be explained by the various structural intermediates formed by UV-DDB during damage recognition. Alternately, some of this heterogeneity may be explained by the nature of the lesion. Previous research has identified that UV-DDB recognizes 6-4 PPs more efficiently and binds more tightly than CPD lesions [18]. We hypothesize that the longer lived intermediates reflect lesion specific interactions. This hypothesis may be tested using DNA damage arrays containing UV-DDB substrates in the DNA tightrope assay and measuring the mean lifetimes of long-lived intermediates.
2. Previous work has revealed that the N-terminus of DDB2 is intrinsically disordered [25] and is structured upon binding to DNA [102]. How does the N-terminus of DDB2 determine the stability of UV-DDB on DNA? Based on the crystal structure of dimeric UV-DDB, we hypothesize that interactions of the folded N-terminus of DDB2 contribute additional affinity for the DNA at sites of lesion. This hypothesis may be tested by using

a mutant version of DDB2 involving a deletion of the N-terminus of DDB2 [UV-DDB (Δ N-terminus)]. A prediction of this hypothesis is that UV-DDB (Δ N-terminus) will exhibit a reduced fraction of long-lived intermediates compared to WT UV-DDB.

3. Our work reveals that UV-DDB forms long-lived dimers on DNA at sites of lesions. This dimerization is known to involve the N-terminus of DDB2 [102]. We hypothesize that deletion of the N-terminus of DDB2 abrogates dimerization of UV-DDB. A prediction of this hypothesis is that UV-DDB (Δ N-terminus) would form primarily monomers on DNA. Evidence for this hypothesis may be obtained from AFM experiments which have the ability to reveal stoichiometry of UV-DDB on DNA. Further evidence may be a loss of the persistent population of UV-DDB on DNA at sites of lesions, when assayed in the DNA tightrope assay.
4. The N-terminus of DDB2 is heavily post-translationally modified [143, 144, 29, 125, 126]. Ubiquitination is known to destabilize the UV-DDB DNA interaction [29], whereas, PARylation is shown to promote it [125, 126]. Further, these modifications are known to occur at lysines in the N-terminus of DDB2 [25]. How do post-translational modifications affect the stability of the DNA bound intermediates identified in this work? The effect of PARylation on UV-DDB dwell times on DNA, may be tested using in-vitro PARylated DDB2 followed by direct observation in the DNA tightrope assay with appropriate controls.
5. Testing the effect of ubiquitination on the dwell times of UV-DDB on DNA, requires QD conjugated CRL4^{DDB2} which retains ubiquitination activity. QD-CRL4^{DDB2} may then be directly visuallized in the DNA tightrope with appropriate controls. Experiments in the absence of ATP will reveal the influence of Cullin4/RBX1 on the stability of intermediates formed by CRL4^{DDB2} compared to WT UV-DDB tested in this work. The binding of CRL4^{DDB2} in the presence of ATP may reveal a dependence of ubiquitination of the N-terminus of UV-DDB on DNA binding. However, several control experiments need to be performed, including the UV-DDB (Δ N-terminus) mutant described previously. Further, an orthogonal labeling strategy to identify the presence of ubiquitin on DDB2 is required.
6. UV-DDB is known to hand-over the lesion to the global damage sensor XPC. However,

observation of a ternary complex involving UV-DDB, XPC and DNA has been elusive. It is likely that this complex is transient in the presence of functional CRL4^{DDB2} or that existing assays may not be able to trap detectable amounts of this ternary complex. The single molecule DNA tightrope assay used in this work has the ability to reveal intermediates which may escape detection by bulk methods. Indeed preliminary data suggests that UV-DDB and XPC colocalize within the resolution of the microscope.

7. Recent research has provided direct evidence that UV-DDB recognizes lesions in nucleosomal DNA [25]. However, several questions remain. Does UV-DDB dimerize on nucleosomes containing damage? Does damage recognition in nucleosomal DNA proceed via the formation of the kinetic intermediates described in this work? This hypothesis may be tested using 'chromatinized' long DNA substrates in the DNA tightrope assay. Long DNA substrates containing nucleosomes may be created by reconstituting nucleosomes on arrays of the 601 positioning sequence. An orthogonal strategy to label histone cores should be used to identify colocalization of QD-UV-DDB with nuclesomes.

In this work, we have implemented a single molecule DNA tightrope assay which serves as a platform to study interactions of biomolecules with DNA. Further, we have identified reagents and strategies to permit multicolor observation of QD conjugated DNA repair proteins. We have also developed and characterized new DNA substrates that permit observations of site specific interactions of DNA repair proteins with introduced custom chemical modifications at defined lesions. These reagents overcome significant limitations in the family of DNA stretching assays and provide new possibilities in probing site specific interactions. We believe that the assays developed here in combination with mutational analyses will enable the characterization of structure-function relationships that govern specific protein-DNA interactions. Further, the use of single molecule methods used here enables the observation of the stochastic nature of protein-DNA interactions, as proteins assemble and disassemble at sites of custom chemical modifications in DNA in real time. Ultimately, we believe that these technologies will be useful in the study of other DNA repair pathways beyond nucleotide excision repair.

APPENDIX A

ABBREVIATIONS

6-4 PP - 6-4 photoproduct

Ab - antibody

CPD - Cyclobutane pyrimidine dimer

CRL4^{DDB2} - Cullin4A/B-RBX1-DDB1-DDB2 E3 ligase

CS - Cockayne syndrome

CSN - COP9 signalosome

DNA - Deoxyribonucleic acid

ERCC1 - Excision Repair cross complementation protein 1; obligate heterodimer of XPF

GG-NER - Global genomic nucleotide excision repair

NER - Nucleotide excision repair

nt - nucleotide

Polδ - Eukaryotic DNA polymerase δ

Polϵ - Eukaryotic DNA polymerase ϵ

PSF - Point spread function

RPA - Replication protein A; single strand DNA binding protein; eukaryotic homolog of bacterial single strand DNA binding protein (SSB)

TC-NER - Transcription coupled nucleotide excision repair

TFIIH - General transcription factor IIH

TTD - Trichothiodystrophy

UV-DDB - DNA damage recognition factor; Ultraviolet induced DNA damage binding pro-

tein, consisting of DDB1 and DDB2
XP - Xeroderma pigmentosum
XP-A - Xeroderma pigmentosum group A phenotype
XPA - Repair factor defective in XP-A
XP-B - Xeroderma pigmentosum group B phenotype
XPB - 3'-5' helicase defective in XP-B
XP-C - Xeroderma pigmentosum group C phenotype
XPC - Repair factor found to be defective in XP-C
XPC complex - DNA damage recognition factor; comprised of XPC, HR23A/B and Centrin2
XP-D - Xeroderma pigmentosum group D phenotype
XPD - 5'-3' helicase defective in XP-D
XP-E - Xeroderma pigmentosum group E phenotype
XP-F - Xeroderma pigmentosum group F phenotype
XPF - Structure specific 5' endonuclease deficient in XP-F
XP-G - Xeroderma pigmentosum group G phenotype
XPG - Structure specific 3' endonuclease deficient in XP-G

APPENDIX B

AFM CALIBRATION CURVE

Enzymology:

Single Molecule Studies of Physiologically Relevant Telomeric Tails Reveal POT1 Mechanism for Promoting G-quadruplex Unfolding

Hong Wang, Gerald J. Nora, Harshad Ghodke
and Patricia L. Opresko

J. Biol. Chem. 2011, 286:7479-7489.
doi: 10.1074/jbc.M110.205641 originally published online December 23, 2010

Supplemental material:
http://www.jbc.org/content/suppl/2011/01/03/M110.205641.DC1.html

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 9, pp. 7479–7489, March 4, 2011

Single Molecule Studies of Physiologically Relevant Telomeric Tails Reveal POT1 Mechanism for Promoting G-quadruplex Unfolding*[S]

Received for publication, November 24, 2010, Published, JBC Papers in Press, December 23, 2010, DOI 10.1074/jbc.M110.205641

Hong Wang[‡§1,2], **Gerald J. Nora**[¶‖1], **Harshad Ghodke**[§‖], **and Patricia L. Opresko**[¶‖3]

From the [‡]*Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine and* [§]*University of Pittsburgh Cancer Institute, Hillman Cancer Center, Pittsburgh, Pennsylvania 15213, the* [¶]*Department of Environmental and Occupational Health, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania 15219, and the* [‖]*Molecular Biophysics and Structural Biology Graduate Program, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, Pennsylvania 15260*

Human telomeres are composed of duplex TTAGGG repeats and a 3′ single-stranded DNA tail. The telomeric DNA is protected and regulated by the shelterin proteins, including the protection of telomeres 1 (POT1) protein that binds telomeric single-stranded DNA. The single-stranded tail can fold into G-quadruplex (G4) DNA. Both POT1 and G4 DNA play important roles in regulating telomere length homeostasis. To date, most studies have focused on individual quadruplexes formed by four TTAGGG repeats. Telomeric tails in human cells have on average six times as many repeats, and no structural studies have examined POT1 binding in competition with G4 DNA folding. Using single molecule atomic force microscopy imaging, we observed that the majority of the telomeric tails of 16 repeats formed two quadruplexes even though four were possible. The result that physiological telomeric tails rarely form the maximum potential number of G4 units provides a structural basis for the coexistence of G4 and POT1 on the same DNA molecule, which is observed directly in the captured atomic force microscopy images. We further observed that POT1 is significantly more effective in disrupting quadruplex DNA on long telomeric tails than an antisense oligonucleotide, indicating a novel POT1 activity beyond simply preventing quadruplex folding.

Cells with linear chromosomes must solve the following two problems: the progressive lagging strand shortening with each cycle of DNA replication and the need to protect the ends of linear chromosomes from unwanted DNA damage responses (1). As a solution to both these problems, telomeres stand at the junction between aging, genomic stability, and cancer. Telomeres are composed of the "shelterin complex" of proteins and TTAGGG repeats of duplex DNA along with an ssDNA overhang or "tail" of 50–500 nucleotides (1). The ssDNA tail can fold into G-quadruplex DNA (G4 DNA),[4] which consists of three tetrads of four guanines that form Hoogsteen base pairs with each other (Fig. 1*A*). These tetrads are in a square planar conformation and are stacked atop one another with the TTA sequences forming linker loops (2, 3). The formation of G4 DNA has been shown to inhibit the telomere-lengthening enzyme complex telomerase *in vitro* (4), although a recent *in vivo* study of *Saccharomyces cerevisiae* telomerase found that G4 DNA can promote the activity of yeast telomerase (5).

Protection of telomeres 1 (POT1) is part of the shelterin protein complex and binds to single-stranded telomeric TTAGGG repeats (6, 7). POT1 protects mammalian chromosome ends from the ataxia telangiectasia mutated and Rad3-related (ATR)-dependent DNA damage response, inhibits 5′ end resection at telomere termini, and regulates telomerase-mediated telomere extension (8). Although POT1 was shown to trap an oligonucleotide with four telomere repeats in an unfolded state to prevent G4 formation (4), the biological significance of this result is unclear. First, POT1 could not bind the short four telomere repeat substrate when the oligonucleotide was prefolded into G4 DNA (4), and second, the telomeric tail has upwards of 30 tandem repeats in human cells (1). Thus, these studies imply that POT1 cannot actively load on telomeric tails *in vivo* unless the G4 structures are melted by a helicase, yet POT1 cellular function is not reported to depend on G4 unwinders and helicases. On the contrary, we reported that POT1 pre-loading on telomeric DNA regulates the unwinding activity of WRN helicase (9–12). At the late G_2 phase of the cell cycle, POT1 levels at the telomeres decrease, and the telomeres are temporarily unprotected and recognized as DNA damage before POT1 relocalizes to the telomeres (13). Because the unprotected tail can spontaneously fold into G4 DNA and block POT1 binding, the mechanism of POT1 reloading on the exposed telomeric tail is unknown.

Studying POT1 loading on physiological telomeric tails is complicated by a lack of information on G4 DNA formation

---

* This work was supported, in whole or in part, by National Institutes of Health Grants ES0515052 (to P. L. O.), 1K99ES016758-01 (to H. W.), and F30AG032861 (to G. J. N.). This work was also supported by the Ellison Medical Foundation (to P. L. O.).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1, Figs. S1–S7, and additional references.

[1] Both authors contributed equally to this work.

[2] To whom correspondence may be addressed: 5117 Centre Ave., Research Pavilion, Ste. 2.6, Pittsburgh, PA 15213. Tel.: 412-623-7803; Fax: 412-623-7761; E-mail: wangh4@upmc.edu.

[3] To whom correspondence may be addressed: Bridgeside Point, 100 Technology Dr., Ste. 326, Pittsburgh, PA 15219. Tel.: 412-624-8285; Fax: 412-624-9361; E-mail: plo4@pitt.edu.

[4] The abbreviations used are: G4 DNA, G-quadruplex DNA; AFM, atomic force microscopy; ssDNA, single-stranded DNA; oligo, oligonucleotide.

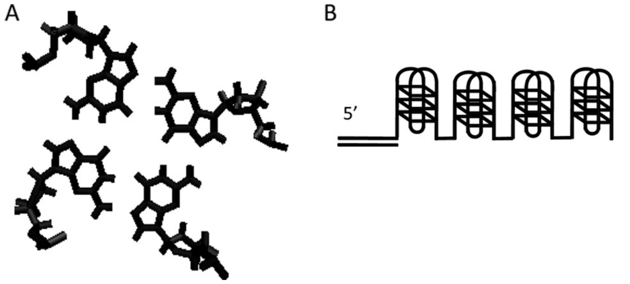


FIGURE 1. **Base pairing in G-quadruplex DNA and the beads-on-a-string model.** *A*, planar tetrad of guanines bound by Hoogsteen base pairing. Image was created on Visual Molecular Dynamics (Urbana, IL) using the Protein Bank 2JPZ structure (20). *B*, schematic illustration of the beads-on-a-string model (18, 19). In this model, long single-stranded telomeric DNA form a beads-on-a-string G4 assembly in which individual quadruplexes are connected by an ssDNA linker.

and distribution on long ssDNA strands. X-ray crystallographic and NMR studies of G4 DNA have focused on individual quadruplexes formed from four TTAGGG repeats (3, 14–18). Possible heterogeneity of the long telomeric ssDNA substrates makes them unamenable to conventional crystallographic and NMR studies (19). Furthermore, bulk biochemical assays, such as native gel electrophoresis, circular dichroism, and UV melting analysis, can only provide a mean value. Results from thermal melting assays support the hypothesis that longer telomeric ssDNAs form a beads-on-a-string G4 assembly in which individual quadruplexes are separated from each other by a TTA linker (Fig. 1*B*) (19), although some data and extrapolations from an NMR structure of individual G4 support a "stacked" arrangement of quadruplexes (20, 21). The discrepancies between these studies underscore the need to examine the formation of G4 structures on physiologically relevant telomeric tails.

Atomic force microscopy (AFM) offers a powerful single molecule approach that allows one to examine distinct nucleic acid structures (single-, double-, and triple-stranded) and their distribution within a heterogeneous population (22, 23). Previous AFM studies established the visualization of human telomeric single G4 DNA units by AFM (24). However, the quantitative distributions of various quadruplex numbers and arrangement ensembles of individual molecules within a potentially heterogeneous population of long single-stranded telomeric molecules have not been addressed. Even more importantly, POT1 coats the 3′ ssDNA tail of the telomere (6, 25, 26). However, the potential modulation of G4 folding by POT1 on physiologically relevant telomeric tails has not been investigated, and whether G4 DNA and POT1 can coexist on a telomeric tail is unknown. AFM has been used extensively to study protein-DNA interactions (27, 28), validating its application for the visualization of telomeric tail structures in the presence and absence of POT protein at the single molecule level.

First, to visualize the formation of G4 DNA on realistic telomeric tails, we performed single molecule AFM imaging of defined DNA substrates with a duplex stem followed by single-stranded TTAGGG repeats (4, 8, or 16) and conducted detailed quantitative analysis of the length and height of the G4 structures. At physiological salt concentrations, the majority of $(TTAGGG)_{16}$ molecules form only two G4 structures, instead of the maximum of four, so that not all the POT1-binding sites are occluded. Consistent with this, the AFM images revealed that POT1 coexists with G4 DNA on some 3′ tails. We report that POT1 addition shifts the population distribution toward telomeric molecules that have fewer G4 units or that are completely unfolded. Importantly, POT1 was significantly more effective in disrupting G4 DNA on $(TTAGGG)_{16}$ molecules than an antisense oligonucleotide, indicating an activity beyond simply preventing G4 folding as proposed previously (4). Our data are consistent with a model in which POT1 acts as a "steric driver" on long telomeric ssDNA to promote unfolding of neighboring G4 structures.

MATERIALS AND METHODS

DNA Substrates—All oligonucleotides were purchased from Integrated DNA Technologies and were purified using PAGE by the manufacturer. The sequences of the oligos are listed in supplemental Table S1. DNA substrates that contain a 5′ duplex region and a 3′ ssDNA tail were formed by incubating equal molar amounts of oligonucleotides in 1× phosphate buffer (10 mM potassium phosphate and 150 mM KCl) or 1× POT1 buffer (40 mM Hepes, pH 7.5, and 50 mM KCl) at 85 °C for 5 min, followed by slow cooling to room temperature. Linear dsDNA substrate, PCR517, used as an internal size standard was made by PCR amplification of nucleotides 1374–1890 on pUC18 plasmid and purification using Illustra GFX^{TM} PCR DNA and a gel band purification kit (GE Healthcare).

Protein Purification—Recombinant GST-tagged and untagged POT1 proteins were purified using a baculovirus/insect cell expression system and an AKTA Explorer FPLC (GE Healthcare) as described previously (10). Protein concentrations were determined using Coomassie staining along with a standard of known concentration. Proteins used in this study are more than 90% pure based on SDS-PAGE and Coomassie staining (supplemental Fig. S2*F*).

AFM Sample Preparation and Imaging—All DNA substrates and POT1 protein were diluted in 1× POT1 buffer containing additional 10 mM $MgCl_2$ for AFM imaging. All buffers were heated at 65 °C for 15–30 min to dissolve small salt particles that may have accumulated during storage. Samples of DNA with and without POT were prepared using the same buffer. POT1 and DNA were incubated at 37 °C for 10 min before deposition onto mica. The G-wire solution was prepared by incubating a 270 μM solution of $G_4T_2G_4$ monomer in 100 mM potassium phosphate buffer, pH 7, at 90 °C for 10 min and slow cooling to room temperature, followed by incubation at 4 °C for 12 h. For experiments using the antisense oligo, C-oligo (supplemental Table S1) was incubated with Tel16 DNA (prepared by annealing Tel16 top and bottom oligos, supplemental Table S1) at 37 °C for 10 min. All samples for AFM imaging were prepared by depositing samples onto a freshly cleaved mica (SPI Supply, West Chester, PA), followed by washing with Milli-Q water and drying under a stream of nitrogen gas. All images were collected using a MultiMode V microscope (Veeco Instruments, Plainview, NY) using E scanners in tapping mode. Pointprobe® plus noncontact/tapping mode silicon probes (PPP-NCL, Agilent) with spring constants of ~50 newtons/m and resonance frequencies of ~190 kHz were used. Images were captured at a scan size of 1 × 1 μm, a scan rate of 2–3 Hz, a target amplitude of 0.30 to 0.35 V, and a resolution of 512 × 512 pixels.

Combinatoric Model for G4 Formation—Statistical analyses of G4 formation on Tel8, Tel13, Tel14, Tel15, and Tel16 were calculated by treating them as a sequence of 8 and 13–16 lattices, respectively. It was assumed that G4 structures can form by four consecutive TTAGGG repeats and that individual G4s can fold randomly along the entire length of the lattice. The number of possible arrangements of the *h* items (G4s and unstructured repeats) can be described as shown in Equations 1 and 2,

$$^{h}C_{i} = \frac{h!}{i!(h-i)!} \qquad \text{(Eq. 1)}$$

where *i* is the number of G quadruplexes. For example, for Tel8,

$$^{5}C_{1} = \frac{5!}{1!(5-1)!} = 5 \qquad \text{(Eq. 2)}$$

there are five ways to arrange a single quadruplex, and only one way to arrange two quadruplexes.

Statistical Analysis of AFM Images—The length measurement was done using the Nanoscope7.30 software; unless stated otherwise, structures over 1 nm were noted as G4 DNA on Tel4, Tel8, and Tel16. G4 length was measured along the longest axis at the cutoff height. On Tel16, ~92% of the G4 structures form straight lines, whereas 8% of the molecules display a curvature of less than 30°. For the latter molecules, two intersecting lines were drawn following the center line of the contours. Consequently, the alignment of multimers of G4 on Tel16 does not significantly affect the measurement of G4 length. Two discernable G4 peaks on Tel16 were defined as the presence of two local maxima over 1 nm with a trough in-between that was at least 0.2 nm lower than the shorter peak. When using PCR517 fragments as internal standards for the height and length measurements, at least 20 measurements were done of peak height or full-width at half-maximum height on 517PCR. The adjusted peak height or G4 length was calculated as $F = D \times R$, where F is the adjusted value for height or G4 length; D is the value from direct measurement, and R is the ratio of the mean value measured from multiple depositions of PCR517 alone ($n = 20$) using different imaging probes to the mean value of the PCR517 internal standards ($n = 20$). The mean values of height and full-width at half-maximum height for PCR517 are 0.44 and 10 nm, respectively. For AFM volume analysis, the dimensions of proteins were measured using Image SXM software (28–30). The AFM volume of a particle was calculated as $V = S \times (H - B)$, where V is the AFM volume; S is the area generated at the base of a protein using "density slice" function of the SXM software; H is the average height, and B is the background height. Two-tailed Student's *t* test was conducted for statistical analysis of the height measurement.

RESULTS

Physiological Telomere Tails Rarely Form the Maximum Number of Quadruplexes—Prior to studying POT1 modulation of G4 DNA on physiological telomere tails, we set out to elucidate G4 DNA structures on these molecules in the absence of POT1. Previous AFM studies of G4 DNA used either short telomeric sequence (four repeats), 3′ tails of unknown lengths, or did not provide quantitative or distribution analysis of the images (31–33). Consequently, detailed information regarding the distribution and types of conformations of physiological telomeric tails was lacking. We designed a series of defined DNA substrates that have a 34-bp duplex stem at the 5′ end followed by a 3′ ssDNA overhang of 4, 8, or 16 TTAGGG repeats (Tel4, Tel8, and Tel16, respectively, supplemental Table S1). Tel4, Tel8, and Tel16 can potentially form a maximum of 1, 2, and 4 G4 units, respectively. We reasoned that comparison of G4 structures formed on these substrates as visualized through AFM imaging would provide quantitative information regarding the number of G4 units present on each molecule. AFM field view image and surface plots of Tel4, Tel8, and Tel16 show that all three telomeric substrates formed structures with heights between 1 and 2 nm (Figs. 2, *A–C*, and 3 and supplemental Fig. S1), which were not observed in images of duplex DNA or an ssDNA substrate that lacks G4-forming sequences (supplemental Fig. S2*A* and Fig. 2*E*, respectively). The heights of the peaks observed for the Tel4, Tel8, and Tel16 substrates are consistent with the height measurements from previous AFM studies of single G4 units (31). Evaluation of the AFM height at different target amplitudes indicated that within the range of target amplitudes used in this study (0.30 to 0.35 *V*), the height variation in our AFM images is ~15% of the total height (supplemental Fig. S3*A*). Because the height difference between G4 (1.32 ± 0.22 nm) and duplex DNA (0.44 ± 0.11 nm) exceeds the possible variation in height measurement, we used 1 nm as the height cutoff to measure the length of DNA with G4 character (Fig. 3). A previous AFM study reported a very similar average and standard deviation of G4 peak height on nontelomeric G4-forming sequences (1.30 ± 0.07) (34).

The number of G4 units formed on Tel4, Tel8, and Tel16 molecules was delineated by comparing lengths of G4 regions. To standardize the length measurement, we measured the full-width at half-maximum height of the PCR fragments (517 bp) deposited along with the telomeric DNA substrates (supplemental Fig. S2). The standardized G4 lengths of Tel4, Tel8, and Tel16 (see under "Materials and Methods") are shown in Fig. 2*D*, and yielded similar patterns as the nonstandardized lengths (supplemental Fig. S1*D*). The mean standardized lengths of G4 DNA at 1-nm height of Tel4 and Tel8 are 10 nm. The mean length of DNA with G4 character on Tel16 (20 nm) is only about twice that of Tel4, even though Tel16 could theoretically form a maximum of four quadruplexes as compared with Tel4 which can only form one G4. Further analysis of G4 DNA at higher salt and DNA concentrations and incubation times of up to 2 days did not yield an increase in G4 DNA formation, as judged by the AFM G4 DNA length and volume of Tel16 (data not shown). Together, our data indicate that the majority of molecules with 8 or 16 telomeric repeats only fold into one and two G4 units, respectively, which is 50% of the expected number.

To investigate the mechanism underlying the underfolding (*i.e.* formation of less than the maximum number of quadruplexes) for Tel8 and Tel16, we constructed a first-principles

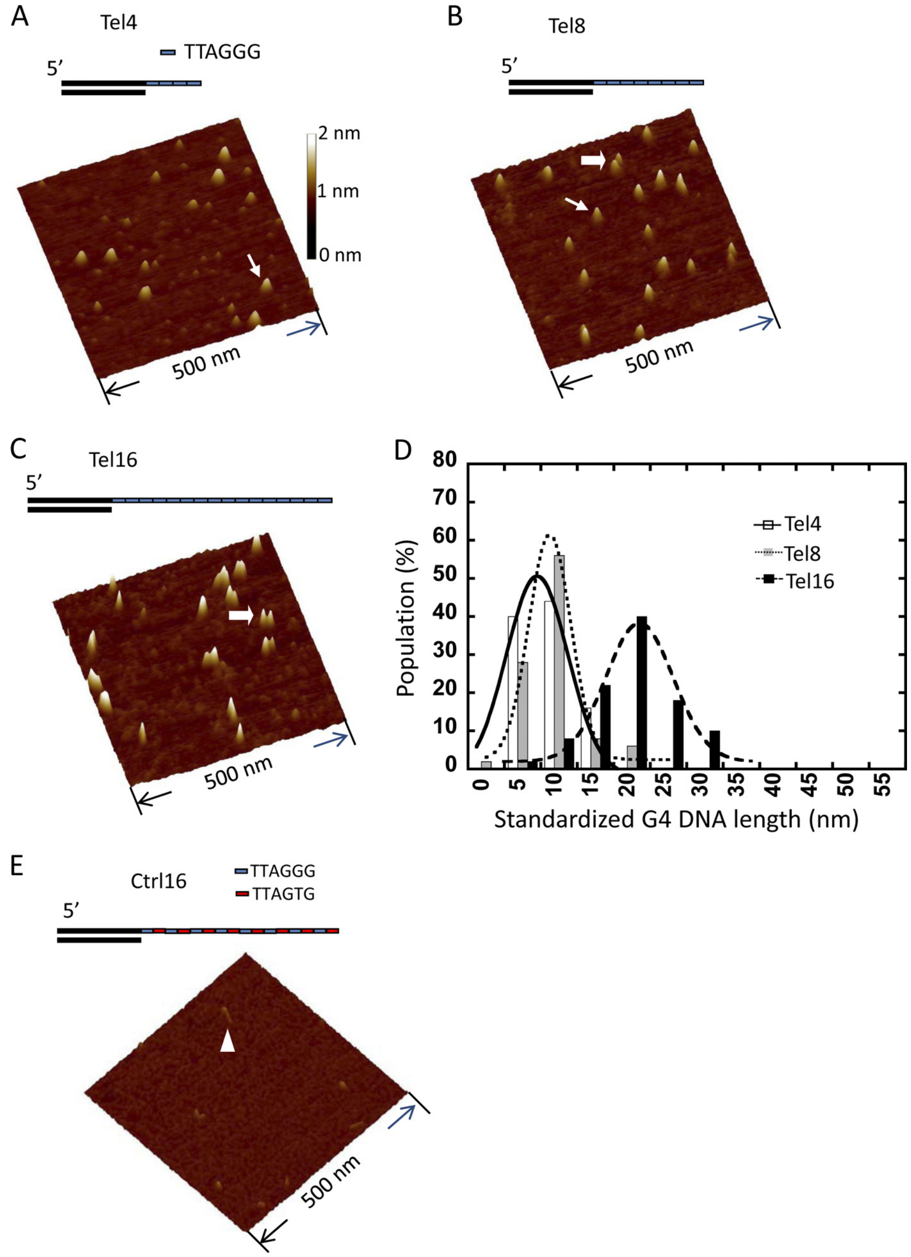


FIGURE 2. **Quantification of the number of G4 structures formed on Tel4, Tel8, and Tel16.** *A–C,* schematic drawings and representative AFM surface plots of Tel4 (*A*), Tel8 (*B*), and Tel16 (*C*) DNA substrates. *Thin arrows* point to single G4 structures, and *wide arrows* denote two distinct G4 structures on individual Tel8 or Tel16 molecules. See supplemental Table S1 for sequences. All DNA substrates were incubated in a buffer containing 150 mM KCl and deposited at 500 nM concentration (see under "Materials and Methods"). Minor particles in *A* are likely contaminants in the Tel4 preparation (*i.e.* acrylamide from the gel purification) rather than unfolded molecules because these images differ from unfolded Ctrl16 structures. *D, histogram* of G4 length (cross-section at 1-nm height) standardized using the mean full-width at half-maximum height of PCR fragments from AFM images of Tel4 (*open bars*, $n = 50$ molecules), Tel8 (*gray bars*, $n = 50$ molecules), and Tel16 (*black bars*, $n = 50$ molecules). The *black lines* represent the Gaussian fit to the data ($R^2 > 0.93$), which are centered at 10 nm (Tel4 and Tel8) and 20 nm (Tel16), respectively. *E,* representative AFM surface plot of Ctrl16 DNA, which contains eight TTAGGGTTAGTG repeats (supplemental Table S1) and does not form G4 structures. The *triangle* points to an individual Ctrl16 molecule. All images are 500 × 500 nm, and the *color bar* corresponds to height from 0 to 2 nm (from dark to bright).

combinatoric model (see "Materials and Methods") considering each telomeric repeat as a lattice point which can either be extended or folded into G4 DNA (Fig. 4*A*). The model shows that the formation of a single G4 in Tel8 is nearly five times more probable than two G4 structures. For Tel16, the most striking insight from the combinatoric model is that formation of four G4 structures on Tel16 is a rare event, which is consistent with our experimental data. In addition, the folding of two quadruplexes was the most probable conformation, but three quadruplexes were almost as probable as two (Fig. 4*B*). This did not fit the normalized experimental data in which the lengths of G4 regions on Tel16 were divided by the mean G4 length from the Tel4 data (Fig. 4*B*). Similarly, a previous study suggested an oligonucleotide with 13 telomeric repeats formed only two quadruplexes based on circular dichroism spectra with a G4 ligand (35). To assess whether the combinatoric model was consistent with our data, we calculated the probability distributions for DNA containing 13–15 repeats. Tel13 and Tel14 both exhib-

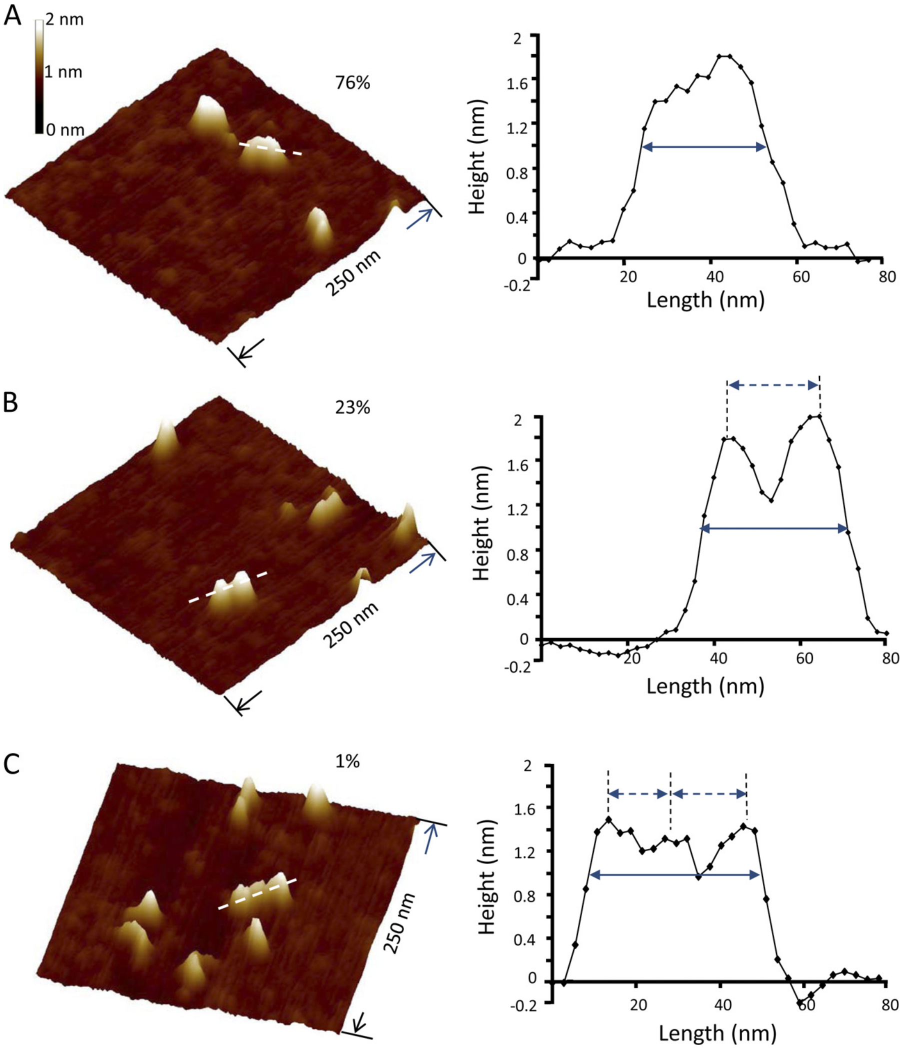


FIGURE 3. **Subpopulations of Tel16 molecules display structures that resemble beads-on-a-string.** A representative AFM image (*left panel*) and sectional analysis (*right panel*) of a Tel16 molecule in which individual G4 structures cannot be resolved (*A*), a molecule which contains two distinct peaks (*B*), or one with three distinct peaks (*C*) are shown. The *white lines* in AFM images indicate the lines drawn for section analysis. The *solid lines* with *arrows* in the section analysis indicate the length of G4 measured at 1-nm height; the *dashed lines* with *arrows* indicate the interpeak distances. The *number* at the *top right corner* of each image indicates the percent of each molecule conformation in the total population of Tel16 molecules. The AFM images are 250 × 250 nm, and the *color bar* corresponds to a height from 0 to 2 nm (from dark to bright).

ited maxima for two G4s, but for Tel15 three G4s was highly probable as well (Fig. 4*B*).

Physiologically Relevant Telomeric Tails Form Structures That Resemble Beads-on-a-String—Different models have been proposed to describe the intra-molecular assembly of multiple G4 units on long telomeric ssDNA (19, 36, 37). In a beads-on-a-string model, two G4 units are connected by one linker without stacking interactions between the units (Fig. 1*B*). In the stacking model, every G4 unit stacks onto adjacent G4, with residues on the TTA loops interacting with each other (19, 21, 38). Among all the Tel16 molecules observed, 23% displayed two distinct peaks in the AFM images (Fig. 3*B*). Although the height difference between the two distinct peaks on individual Tel16 molecules is 0.3 nm, the heights of the lower peaks are still above 1 nm at 1.3 (± 0.3) nm. The mean interpeak distance of Tel16 molecules with two distinct peaks is 20 nm, which corresponds to ~7 TTAGGG repeats between the individual quadruplexes (supplemental Fig. S4). In the AFM images of Tel16 molecules, a small population (1%) of molecules exhibited three distinct peaks (Fig. 3*C*). The assembly of multiple defined peaks resembles individual beads-on-a-string. It is worth noting that because of limitations in the AFM resolution, results from AFM imaging could underestimate the number of Tel16 molecules forming the beads-on-a-string structure (see supplemental calculations).

To further differentiate between the beads-on-a-string and the stacking models, we imaged G-wires that are long complexes of highly ordered self-assembly of inter-molecular G4 units (Fig. 5*A*). G-wires are long, uniformly quadruplectic structures with heights greater than 1 nm in AFM images (39). AFM images of G-wires formed by the short oligonucleotides $G_4T_2G_4$ are shown in Fig. 5, *B* and *C*. Because the G-wires involve stacking of the adjacent G4 units, regular well separated peaks were not apparent in the AFM images as expected, even for G-wires that were the same length as Tel16 molecules (Fig. 5, *C* and *D*). In addition, G-wires exhibited a statistically significant ($p < 0.008$) greater average height (1.63 ± 0.17 nm) compared with the Tel16 structures (1.32 ± 0.22 nm) (nonstan-

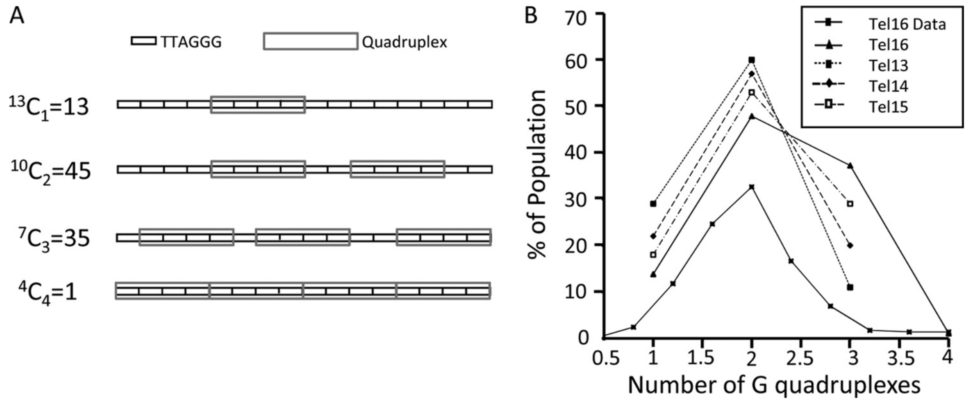


FIGURE 4. **Longer telomeric tails rarely form the maximum potential number of quadruplexes.** *A,* schematic examples and calculations of the number of possible arrangements of G4 DNA on Tel16 (see under "Materials and Methods" for equations). *B,* probability of forming 1–4 G4 structures on substrates with 13–16 (Tel13–16, respectively) based on the combinatoric calculations detailed under "Materials and Methods." The Tel16 data curve is based on the length of G4 regions on Tel16 molecules normalized using the length of single G4 measured from AFM images of Tel4 (supplemental Fig. S1*D*).

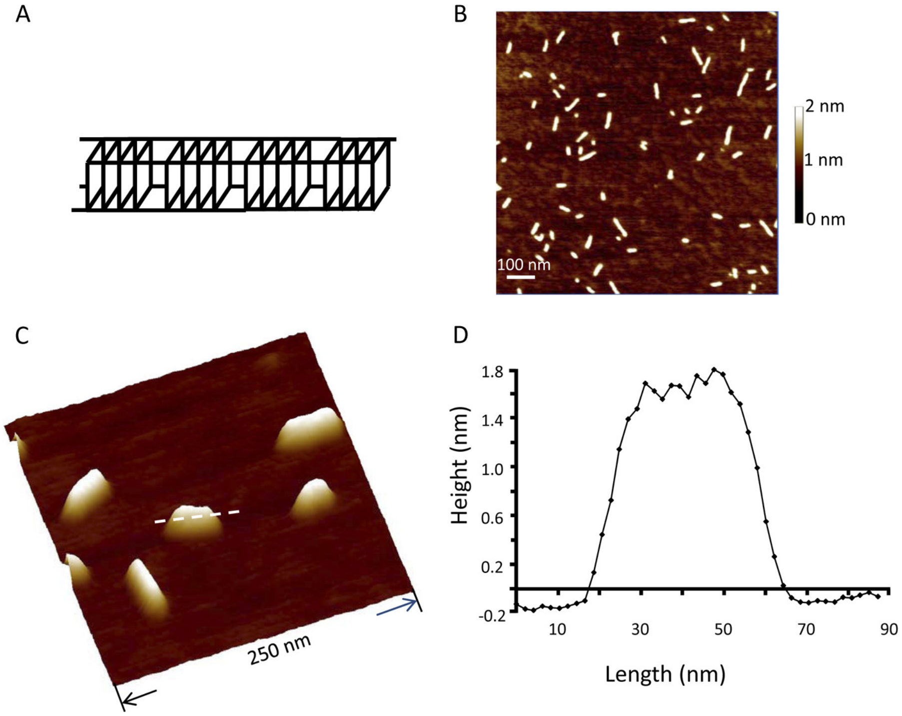
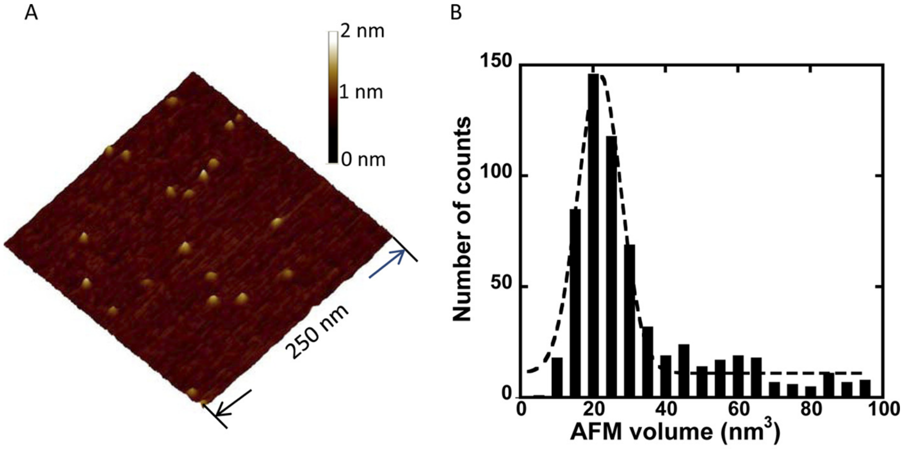


FIGURE 5. **AFM imaging of G-wires reveals a smooth surface without distinct peaks.** *A,* proposed model for G-wire formation. *B,* AFM field view image of G-wires. The image is 1 × 1 μm at 2-nm height scale. *C,* AFM surface plot of G-wires. The image is 250 × 250 nm at 2-nm height scale. The *white line* denotes the line for section analysis. *D,* section analysis of G-wire highlighted in *C.*

dardized). These results suggest that the G-wires appeared to be more rigid possibly because of the direct stacking interactions between adjacent G4 units, which lead to less compression by the mechanical AFM imaging process. The distinctly different structure of the G-wires compared with the Tel16 molecules revealed by AFM imaging suggest that G4 structures on Tel16 molecules are inconsistent with a stacked model of multiple G4 units.

Oligomeric State of POT1—A key issue in understanding the mechanism of action by POT1 is its oligomeric state. Despite evidence showing a monomeric state for the N-terminal domain of human POT1 (7), information on the oligomeric state of full-length human POT1 proteins was lacking. To evaluate the oligomeric state of full-length POT1, we measured the volume of POT1 in AFM images compared with other known proteins of various sizes. AFM-derived volumes of proteins can be correlated to their molecular masses, permitting determination of oligomeric states (see under "Materials and Methods") and protein-protein interactions (28, 30). Purified POT1 protein after removal of the GST tag appeared as monodispersed particles in the AFM images (Fig. 6*A*). At three different concentrations (20, 200, and 1000 nM), the distribution of the calculated AFM-derived volumes of POT1 is Gaussian and centered at ~22 nm^3 (for 200 nM POT1, see Fig. 6*B*, other data not shown), which is consistent with the expected value for a POT1 monomer based on the calibration curve for globular proteins (supplemental Fig. S5). These results demonstrate that POT1 exists as a monomer in solution under the AFM imaging conditions tested. In contrast, AFM images of GST tagged POT1 protein (GST-POT1) revealed particles consistent with GST-



C

Samples	Standardized height (nm)	Standard deviation (nm)
Tel4	1.36	0.30
Tel16	1.40	0.18
Pot1	0.65	0.14

FIGURE 6. **Full-length POT1 is a monomer in solution and exhibits a height in AFM images distinct from G4 DNA.** *A*, representative AFM image of untagged POT1 protein at 200 nM concentration. The image is 250 × 250 nm at 2-nm height scale. *B*, AFM volume distribution of POT1 from images of POT1 at 200 nM concentration. The *dashed line* represents the Gaussian fit to the data ($n = 664$ molecules, $R^2 = 0.96$), which is centered at 22 nm^3 and corresponds to POT1 monomer based on the standard calibration curve (supplemental Fig. S5). *C*, comparison of the standardized peak heights of Tel4, Tel16, and POT1 molecules ($n = 50$ each) in AFM images. The peak height was standardized using PCR517 DNA fragments as internal standards (see under "Materials and Methods").

POT1 dimers and tetramers (data not shown). Therefore, only untagged POT1 was used in all the imaging experiments with the DNA substrates. Importantly, the standardized height of POT1 (0.65 ± 0.14 nm) is significantly different from the standardized height for G4 DNA on Tel4 (1.36 ± 0.30 nm) and Tel16 (1.40 ± 0.18 nm) (Fig. 6*C*). The nonstandardized heights showed the same result (supplemental Fig. S3*B*). Thus, height measurement provides a robust criterion to differentiate between POT1 and G4 structure when POT1 and Tel16 are mixed together.

POT1 Binding Competes with G4 Formation on Physiologically Relevant Telomeric Tails—To study the binding of POT1 to physiological telomeric tails using AFM, we utilized two DNA substrates, Tel16 and Ctrl16 (supplemental Table S1). Ctrl16 is the same length as the Tel16 DNA substrate, but every other TTAGGG sequence in Ctrl16 is changed to TTAGTG, which eliminates G4 folding (Fig. 2*E*). The minimum DNA sequence that is required for high affinity binding of human POT1 *in vitro* is TTAGGGTTAG (7). Accordingly, both Tel16 and Ctrl16 substrates have a maximum of eight POT1 DNA binding sites. Electrophoresis mobility shift assays (EMSA) showed that under the same conditions POT1 binds Tel16 and Ctrl16 substrates to a similar extent (supplemental Fig. S6*B*). The appearance of more than one shifted band suggests that multiple POT1 molecules can bind to the Tel16 or Ctrl16 substrates.

In the AFM images of Ctrl16 with POT1, arrays of tandem POT1 proteins were observed (*thin arrow*, Fig. 7*A*), which were not present in the POT1-alone images (Fig. 6*A*). The mean height of these POT1 arrays is statistically similar to the POT1 height in the protein-alone images (supplemental Fig. S3*B*). We used the statistically significant height difference between POT1 and G4 DNA to differentiate between POT1 and G4 structures (Fig. 6*C* for standardized and S3B for nonstandardized heights). When POT1 (200 nM) was incubated with a 5-fold molar excess of Tel16 (1 μM), the percent of molecules that exhibited G4 DNA structures (peak heights >1 nm) was greatly reduced from 100% of the Tel16-alone molecules, to 24% (98:405) of the molecules visualized after coincubating Tel16 with POT1 (Fig. 7*C*). The majority of molecules (76%, 307:405) showed only structures that were characteristic of POT1. Importantly, of the G4 DNA structures observed (98: 405), 23 molecules displayed multiple peaks with differing heights that were consistent with G4 DNA and bound POT1 on the same molecule (compare Fig. 7*D* for POT1 + Tel16 and Fig. 7*B* for POT1 + Ctrl16). The height of the lower peaks is 0.7 (± 0.1) nm ($n = 23$ complexes), which is statistically different from the lower peaks on Tel16 molecules displaying two or more peaks in the absence of POT1 (1.3 ± 0.3 nm) and very closely matches the standardized peak for POT1 alone (Fig. 6*C*). These images indicate that G4 DNA and POT1 can coexist on the same molecule. The length distributions of POT1-bound regions for Ctrl16 and Tel16 (supplemental Fig. S6*C*) both exhibited a long right-sided "tail" representing similar numbers of POT1 proteins bound to Tel16 and Ctrl16 molecules. The length of longer POT1 arrays (45–60 nm) is consistent with the length of ssDNA (48 nm, assuming ssDNA as 0.5 nm/base) on fully extended Tel16 molecules. Together, these data indicate that POT1 binding can successfully compete with G4 DNA folding on telomeric ssDNA.

Previous work suggested that POT1 and an antisense 13-mer oligonucleotide, which base pairs with telomeric ssDNA, share the same mechanism of trapping a short oligonucleotide GGG(TTAGGG)$_3$ in an unfolded state to prevent G4 formation (4). To further investigate the mechanism of G4 disruption on long telomeric ssDNA, we quantified the G4 structures on the Tel16 substrate after incubation with the antisense oligonucleotide (C-oligo, supplemental Table S1) for comparison with

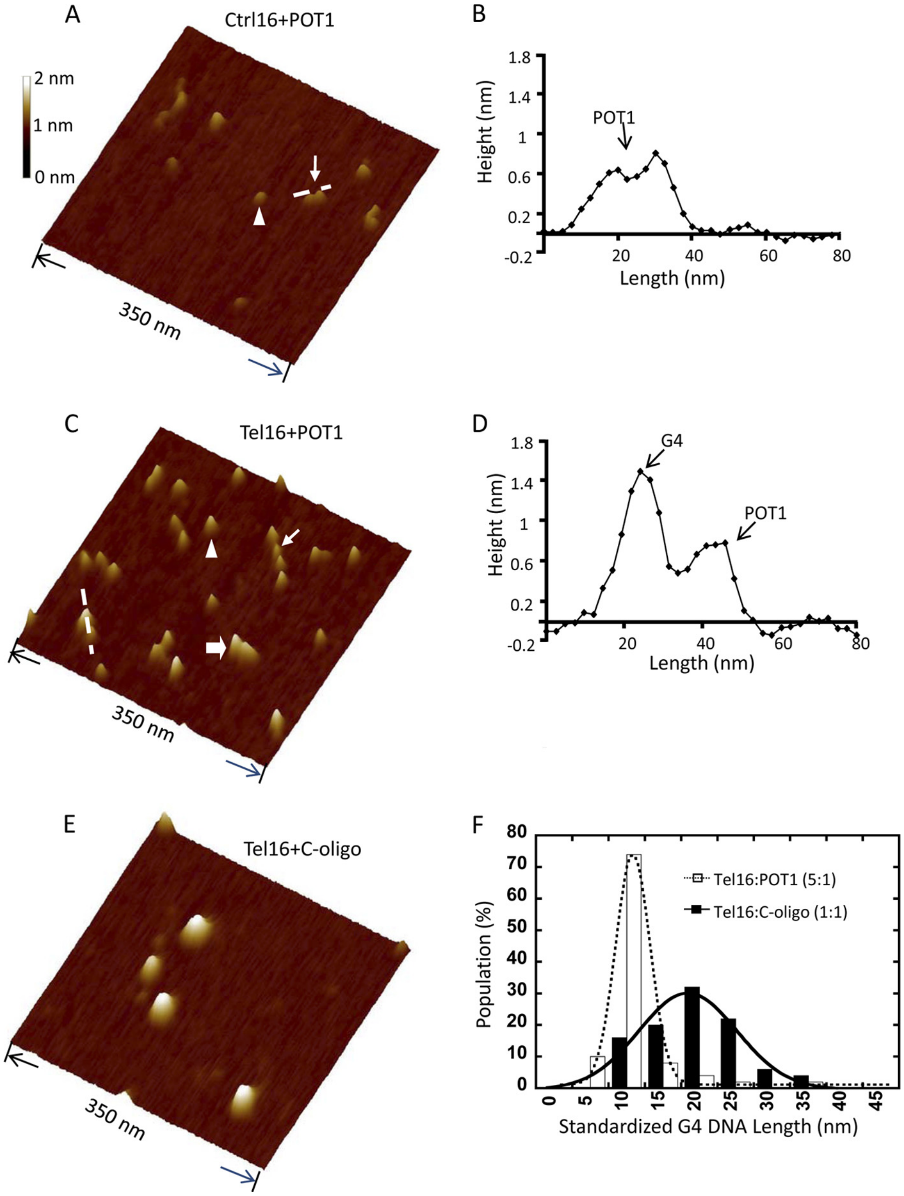


FIGURE 7. **POT1 is more effective at disrupting G4 DNA on Tel16 molecules than an antisense oligonucleotide.** *A,* representative AFM surface plot of the non-G4 forming Ctrl16 (1 μM) substrate in the presence of POT1 (200 nM). The *triangle* points to individual POT1 molecules. The *thin arrow* points to a structure with multiple POT1 proteins. *B,* cross-section of the molecule highlighted in *A* by the *dotted line* showing two POT1 molecules on the same telomeric tail. *C,* representative AFM surface plot of Tel16 (1 μM) in the presence of POT1 (200 nM). The *thick arrow* points to a structure with folded G4. The *triangle* points to an individual POT1 molecule. The *thin arrow* points to a structure with multiple POT1 proteins. *D,* cross-section of the molecule highlighted in *C* by the *dotted line* demonstrating that G4 (*left peak*) and POT1 (*right peak*) coexist on the same molecule. *E,* representative AFM surface plot of Tel16 (1 μM) in the presence of C-oligo (1 μM). *F, histogram* of the length of G4 DNA regions (stretch of DNA with peaks >1 nm) measured from AFM images of Tel16 (1 μM) in the presence of POT1 (200 nM, *open bar, n* = 50 molecules) and C-oligo (1 μM, *black bar, n* = 50 molecules). The *black lines* represent the Gaussian fit to the data ($R^2 > 0.96$), which are centered at 10 (POT1) and 20 nm (C-oligo), respectively. The G4 length values were standardized using PCR517 DNA fragments as internal standards (see "Materials and Methods"). The AFM images are 350 × 350 nm at 2-nm height scale.

the images of POT1 added to Tel16. When Tel16 and C-oligo were incubated at a 1:1 molar ratio, most (92%) of the molecules displayed peaks at a height consistent with G4 structures (>1 nm). Thus, POT1 was more effective in decreasing the population of molecules with G4 character (24%), even though POT1 was present at lower stoichiometric amounts (5-fold less) compared with the C-oligo. An excess of C-oligo over Tel16 (5:1) is required to fully trap the G4 structures in an unfolded state (supplemental Fig. S7), which indicates that C-oligo can bind the Tel16 ssDNA. However, at this ratio the disruption of G4 structure by C-oligo is through elimination of consecutive single-stranded TTAGGG repeats that can form G4. At a 5-fold molar excess, if the oligo is evenly distributed, the distance between individual C-oligos is ~5 nucleotides.

One caveat of our experiment is that a fraction of the Tel16 molecules that lack G4 character (76%) upon POT1 addition may represent POT1 unbound to DNA. This is unlikely because Tel16 is present at a 5-fold excess over POT1, which represents a 40-fold excess of POT1-binding sites. However, for a more rigorous analysis, we measured the length of the G4 regions on the Tel16 molecules that showed G4 peaks in the presence of POT1 (24%) or C-oligo (92%). For the C-oligo, the majority of the G4 structure lengths were consistent with the existence of two G4 units similar to Tel16 alone (15–20 nm, standardized lengths, Fig. 7*F* and Fig. 2*D*). It is worth noting that the peaks for two G4 units on Tel16 with the C-oligo are less well defined compared with Tel16 alone, perhaps because of the oligo annealing to the region (~7 repeats or ~40 nucleotides, supplemental Fig. S4) between the G4 units. In stark contrast, the lengths of the G4 regions remaining on Tel16 after POT1 addition were about half as long as G4 regions on Tel16 with or without C-oligo (Fig. 7*F* and supplemental Fig. S6*D*). This is consistent with POT1 inducing a shift from two to one G4 unit on those Tel16 molecules that retain G4 folds. In summary, our data indicate that contrary to results with short telomeric tails (4), POT1 is much more effective at disrupting G4 DNA on long telomeric tails, compared with an antisense oligonucleotide.

DISCUSSION

POT1 binding to $(TTAGGG)_4$ substrates prevents G4 DNA folding (4, 40). However, the arrangement of G4 DNA and the competition with POT1 binding on long, physiologically realistic telomeric tails were unknown. In this study we used single molecule imaging to examine the assembly of G4 units on DNA substrates containing 4 (Tel4), 8 (Tel8), and 16 (Tel16) TTAGGG repeats, with the latter representing the mid range of the telomeric overhang length in human cells (1). Telomeric DNA with well defined lengths allowed us to study the length-dependent formation of G4 structures at the single molecule level. We demonstrated that G4 DNA assemblies on physiologically relevant telomeric tails rarely form the maximum potential number of G4 units. We observed via AFM imaging that full-length POT1 is monomeric and stabilizes the ssDNA, driving the $(TTAGGG)_{16}$ structural equilibrium toward an extended protein-bound state. This study is the first to report that bound POT1 can coexist with G4 DNA on the same Tel16 molecule. Compared with an antisense oligo that statically binds the telomeric ssDNA, POT1 is much more effective in disrupting G4 structures on long telomeric tails. Our results are consistent with a novel and more dynamic mechanism of POT1 G4 disruption, in contrast to a simple static trapping of unfolded DNA.

We applied a first principles combinatoric approach to understand the mechanism underlying the underfolding, and we found that the model prediction for ssDNA with 13 repeats (Fig. 4) is consistent with a bulk circular dichroism study that suggested oligonucleotides with 13 telomeric repeats formed on average only two quadruplexes (35). However, the normalized G4 distributions of Tel16 images demonstrated a sharp peak at two quadruplexes, whereas the probabilistic model based on the first-principles combinatoric approach predicted a nearly equal quantity of molecules with three quadruplexes as well (Fig. 4). The discrepancy between our experimental observations and the probabilistic models may be explained by differences in the probability of forming G4 at different positions along the length of Tel16 and that the model does not take into account free energy of folding. A previous study using dimethyl sulfate footprinting and exonuclease hydrolysis with $T_{24}(TTAGGG)_7$ DNA substrates revealed that the probability of forming G4 rapidly decreases toward the 5′-flanking sequence (41), from 55.8% at the 3′ end (0 position) to 21.8, 14.5, and 7.9% at the first, second, and third positions (next to 5′-flanking sequence), respectively. Our model (Fig. 4) presumes that probabilities of forming G4 along the 3′ G-rich tail of Tel16 are the same. The dramatic decrease in the probability of forming G4 units when the repeat positions are close to the 5′-flanking region effectively shortens the number of available repeats for G4 folding on Tel16. This explains the close agreement of the normalized G4 distributions from the experimental data with the theoretical G4 distributions of two shorter substrates with 13 and 14 repeats (Fig. 4*B*). A previous report indicated that $GGG(TTAGGG)_3$ forms the most stable G4, and as repeat number increases ($n = 7–16$), the quadruplex molecules become less thermostable (42). The presence of loops with various lengths on the tetraplex sides can potentially lead to irregularities in G4 structure and consequently cause structure destabilization. Current literature suggests that loop length and composition strongly influence the quadruplex stability, and quadruplexes formed by $(TTAGGG)_5$ with a 9-nucleotide loop were less stable than quadruplexes formed from four consecutive repeats (41).

The arrangement of G4 DNA on longer physiological telomeric tails has been controversial. One thermal melting study supported a beads-on-a-string conformation whereby long telomeric substrates fold into the maximum number of quadruplexes that do not directly interact with each other (19). Another study found support for a stacked model whereby individual quadruplexes fold in a way that their loop reactions interact, and a more rigid superstructure is formed (21, 38). Direct visualization of individual molecules in our study revealed that 23% and 1% of the measured Tel16 molecules had two and three discernable peaks, respectively. These results support a beads-on-a-string model whereby the quadruplexes form as individual G4 units separated by stretches of ssDNA, creating a more flexible structure with discernable peaks (Figs. 2 and 3, for interpeak distance distribution see supplemental Fig. S4). Although not all the molecules displayed distinct peaks, this was likely due to the resolution limits of the AFM under the current imaging conditions. If two quadruplexes are linked by a TTA linker, the AFM cannot resolve two individual peaks; roughly 1.5 telomeric repeats are required to resolve two peaks (for the calculation of AFM resolution see supplemental material). Also, although the average nonstandardized height of the Tel16 molecules was 1.32 (± 0.22) nm, the average height of the G-wires was 1.63 (± 0.17) nm, suggesting that Tel16 G4 DNA is more flexible, corroborating a beads-on-a-string arrangement.

Previous studies indicated that POT1 binding to substrates with four repeats trapped the molecules in an extended state, shifting the equilibrium from a folded G4 unit to an extended conformation (4, 40). However, POT1 binding to physiologically relevant telomeric tails had not been examined. Our finding that the majority of Tel16 molecules only form two G4 structures has important implications for POT1 loading on realistic telomeric tails. POT1 cannot bind the short $GGG(TTAGGG)_3$ substrates until the equilibrium shifts from G4 structure to an extended state (4). In contrast, on the physiologically relevant Tel16 substrates, an underfolded Tel16 molecule constantly has multiple ssDNA sites available for POT1

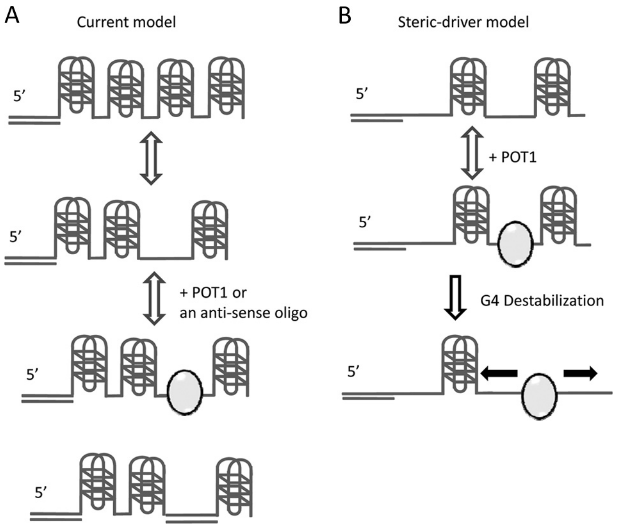


FIGURE 8. **Static passive and dynamic steric driver models of POT1 modulation of G4 DNA at telomeric tails.** A 3′ telomeric tail with 16 TTAGGG repeats is shown as an example. POT1 is shown as *gray ovals*. *A*, passive model whereby G4 DNA is arranged as a beads-on-a-string, maximally folded telomeric tail. POT1 or an antisense oligo could not bind the telomeric sequence until the G4 thermally melted to an extended state. Then POT1 or the antisense oligo traps the molecule in the unfolded state without altering adjacent G4 folds. *B*, in a steric driver model, at equilibrium the telomeric tails rarely form the maximum number of G4 units. POT1 is able to bind unfolded telomeric repeats and destabilize existing G4 DNA on the same molecule possibly through two nonmutually exclusive mechanisms (represented by the *black arrows*) as follows: dynamic movements on DNA including one-dimensional sliding, hopping, and jumping and/or its ability to destabilize adjacent G4 structures.

binding (Fig. 3), and thus, POT1 loading does not require thermal melting of existing G4 DNA.

We propose that POT1 promotion of G4 disruption on long telomeric DNA is not simply by trapping thermally melted G4 structures, as described for short substrates (4). This is because POT1 is more effective in disrupting G4 DNA than a 13-mer antisense oligonucleotide on long telomere tails (Fig. 7*F*) but not on short tails (4). At equal concentrations of antisense oligo and Tel16, the length of the majority of the G4 structures is consistent with two G4 units (Fig. 7*F*). This suggests that similar to the proposed passive model (4), the 13-mer antisense oligo can bind to the unfolded ssDNA on Tel16, but it cannot significantly influence the adjacent remaining G4 folds. On the contrary, for POT1 at a much lower protein to Tel16 ratio (1:5), the majority of molecules were unfolded, and the distribution of G4 length was shifted to one G4 unit. Our results clearly demonstrate that POT1 can disrupt G4 structures more efficiently than the antisense oligo (Fig. 7*F*).

We propose that POT1 binds to the unfolded ssDNA regions and sterically impairs adjacent telomeric repeats from folding into G4 DNA, thereby promoting unfolding into extended ssDNA (Fig. 8*B*). This is in contrast to the previous passive model based on experiments using short oligos, in which POT1 and the antisense oligo share the same ability to trap the short telomeric DNA in an unfolded form (Fig. 8*A*). We propose a steric driver model for the mechanism of G4 disruption by POT1 at 3′ telomeric tails based on the following two nonmutually exclusive mechanisms. First, POT1 binding can destabilize adjacent G4 structures. Recently, it was demonstrated using an isothermal differential hybridization method that binding of a 46-kDa antidigoxin antibody fragment adjacent to a G4 fold dramatically destabilized the G4 structure (43). Another possible mechanism of G4 disruption by POT1 is through dynamic one-dimensional sliding and/or microscopic dissociation and re-association to adjacent sequences. Precedent for one-dimensional diffusion of single-stranded DNA-binding proteins has been described for *Escherichia coli* single-stranded DNA-binding protein based on the single molecule studies (44). The steric driver model is consistent with results from AFM imaging of Ctrl16 and Tel16 with POT1 (Fig. 7 and supplemental Fig. S6). Specifically, upon addition of POT1, the equilibrium shifts from a majority of Tel16 molecules forming two quadruplexes to one quadruplex and/or multiple POT1 monomers bound (Fig. 7*F* and supplemental Fig. S6). Importantly, multiple POT1 molecules bind Tel16 and the non-G4-forming Ctrl16 substrate to similar extents, leading to protein arrays of roughly equal length distributions (supplemental Fig. S6*C*). If POT1 can only capture the ssDNA when the G4 DNA thermally melts, then we would expect a greater number of molecules with long POT1-bound arrays for Ctrl16 relative to Tel16, because POT1 does not need to compete with G4 folding to bind Ctrl16.

In summary, we propose a model whereby POT1 acts not as an active DNA unwinder but rather as a steric driver by binding to underfolded telomeric tails and thereby destabilizing the adjacent remaining G4 structures on the molecule (Fig. 8*B*), as evidenced by the reduction of G4 DNA structures upon POT1 addition (Fig. 7*F* and supplemental Fig. S6*C*). Our results demonstrate that on a long telomeric substrate, the mechanism of action by POT1 is different from the simple static trapping mechanism utilized by an antisense oligo. POT1 binding competition with G4 DNA folding on physiologically relevant 3′ telomeric tails suggests an important mechanism for preserving telomere stability. Because a telomeric tail that is exposed during replication of the telomere can spontaneously fold into G4 DNA, this raises the issue of how POT1 reloads on the telomeric tail to promote telomerase activity or telomere remodeling into a capped structure (13). Another study demonstrated that a G4-stabilizing agent induced an ATR-dependent DNA damage response but that POT1 levels at the telomere ends remained unchanged (45), implying that G4 DNA and POT1 may coexist at telomere ends. The AFM images in this study show that the underfolding (*i.e.* less than the maximum number of G4 units) of long telomeric ssDNA provides a route for POT1 binding and a mechanism for POT1 and G4 DNA coexistence on the same molecule. The direct visualization of single molecules that resemble physiologically relevant telomeric tails provide a mechanistic basis for understanding the modulation of telomere structure and function by POT1 and G4 DNA.

Acknowledgments—We thank Drs. Bennett Van Houten (University of Pittsburgh), Dorothy Erie (University of North Carolina, Chapel Hill), and Ingrid Tessmer (University of Wurzburg) for critical reading of the manuscript and suggestions. We thank Dr. Fujun Liu and Rama Rao Damerla (University of Pittsburgh) for assistance with manuscript preparation. We thank Dr. Peggy Hsieh (NIDDK, National Institutes of Health) and Drs. Karen Thickman and Saleem Khan (University of Pittsburgh) for providing Taq MutS and PcrA proteins.

REFERENCES

1. Palm, W., and de Lange, T. (2008) *Annu. Rev. Genet.* **42,** 301–334
2. Li, J., Correia, J. J., Wang, L., Trent, J. O., and Chaires, J. B. (2005) *Nucleic Acids Res.* **33,** 4649–4659
3. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) *Nucleic Acids Res.* **34,** 5402–5415
4. Zaug, A. J., Podell, E. R., and Cech, T. R. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 10864–10869
5. Zhang, M. L., Tong, X. J., Fu, X. H., Zhou, B. O., Wang, J., Liao, X. H., Li, Q. J., Shen, N., Ding, J., and Zhou, J. Q. (2010) *Nat. Struct. Mol. Biol.* **17,** 202–209
6. Baumann, P., and Cech, T. R. (2001) *Science* **292,** 1171–1175
7. Lei, M., Podell, E. R., and Cech, T. R. (2004) *Nat. Struct. Mol. Biol.* **11,** 1223–1229
8. Palm, W., Hockemeyer, D., Kibe, T., and de Lange, T. (2009) *Mol. Cell. Biol.* **29,** 471–482
9. Opresko, P. L., Mason, P. A., Podell, E. R., Lei, M., Hickson, I. D., Cech, T. R., and Bohr, V. A. (2005) *J. Biol. Chem.* **280,** 32069–32080
10. Sowd, G., Lei, M., and Opresko, P. L. (2008) *Nucleic Acids Res.* **36,** 4242–4256
11. Sowd, G., Wang, H., Pretto, D., Chazin, W. J., and Opresko, P. L. (2009) *J. Biol. Chem.* **284,** 34682–34691
12. Opresko, P. L., Sowd, G., and Wang, H. (2009) *PLoS One* **4,** e4825
13. Verdun, R. E., Crabbe, L., Haggblom, C., and Karlseder, J. (2005) *Mol. Cell* **20,** 551–561
14. Smith, F. W., and Feigon, J. (1992) *Nature* **356,** 164–168
15. Wang, Y., and Patel, D. J. (1993) *Structure* **1,** 263–282
16. Balagurumoorthy, P., and Brahmachari, S. K. (1994) *J. Biol. Chem.* **269,** 21858–21869
17. Parkinson, G. N., Lee, M. P., and Neidle, S. (2002) *Nature* **417,** 876–880
18. Maizels, N. (2006) *Nat. Struct. Mol. Biol.* **13,** 1055–1059
19. Yu, H. Q., Miyoshi, D., and Sugimoto, N. (2006) *J. Am. Chem. Soc.* **128,** 15461–15468
20. Dai, J., Carver, M., Punchihewa, C., Jones, R. A., and Yang, D. (2007) *Nucleic Acids Res.* **35,** 4927–4940
21. Petraccone, L., Trent, J. O., and Chaires, J. B. (2008) *J. Am. Chem. Soc.* **130,** 16530–16532
22. Hansma, H. G., Revenko, I., Kim, K., and Laney, D. E. (1996) *Nucleic Acids Res.* **24,** 713–720
23. Lyubchenko, Y. L., and Shlyakhtenko, L. S. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94,** 496–501
24. Kan, Z. Y., Lin, Y., Wang, F., Zhuang, X. Y., Zhao, Y., Pang, D. W., Hao, Y. H., and Tan, Z. (2007) *Nucleic Acids Res.* **35,** 3646–3653
25. Hockemeyer, D., Sfeir, A. J., Shay, J. W., Wright, W. E., and de Lange, T. (2005) *EMBO J.* **24,** 2667–2678
26. Price, C. M. (2006) *Nat. Struct. Mol. Biol.* **13,** 673–674
27. Janiæijeviæ, A., Ristic, D., and Wyman, C. (2003) *J. Microsc.* **212,** 264–272
28. Wang, H., Yang, Y., and Erie, D. A. (2007) in *Protein Interactions, Biophysical Approaches for the Study of Complex Reversible Systems* (Schuck, P., ed) pp. 39–78, Springer Science+Business Media, LLC, Boston
29. Ratcliff, G. C., and Erie, D. A. (2001) *J. Am. Chem. Soc.* **123,** 5632–5635
30. Yang, Y., Wang, H., and Erie, D. A. (2003) *Methods* **29,** 175–187
31. Zhang, X. Y., Cao, E. H., Zhang, Y., Chou, C., and Bai, C. (2003) *J. Biomol. Struct. Dyn.* **20,** 693–702
32. Yoshimura, S. H., Maruyama, H., Ishikawa, F., Ohki, R., and Takeyasu, K. (2004) *Genes Cells* **9,** 205–218
33. Xu, Y., Ishizuka, T., Kurabayashi, K., and Komiyama, M. (2009) *Angew. Chem. Int. Ed. Engl.* **48,** 7833–7836
34. Neaves, K. J., Huppert, J. L., Henderson, R. M., and Edwardson, J. M. (2009) *Nucleic Acids Res.* **37,** 6269–6275
35. Chang, C. C., Chien, C. W., Lin, Y. H., Kang, C. C., and Chang, T. C. (2007) *Nucleic Acids Res.* **35,** 2846–2860
36. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) *Nucleic Acids Res.* **34,** 2723–2735
37. Renciuk, D., Kejnovská, I., Skoláková, P., Bednárová, K., Motlová, J., and Vorlícková, M. (2009) *Nucleic Acids Res.* **37,** 6625–6634
38. Petraccone, L., Garbett, N. C., Chaires, J. B., and Trent, J. O. (2010) *Biopolymers* **93,** 533–548
39. Marsh, T. C., Vesenka, J., and Henderson, E. (1995) *Nucleic Acids Res.* **23,** 696–700
40. Torigoe, H., and Furukawa, A. (2007) *J. Biochem.* **141,** 57–68
41. Tang, J., Kan, Z. Y., Yao, Y., Wang, Q., Hao, Y. H., and Tan, Z. (2008) *Nucleic Acids Res.* **36,** 1200–1208
42. Vorlícková, M., Chládková, J., Kejnovská, I., Fialová, M., and Kypr, J. (2005) *Nucleic Acids Res.* **33,** 5851–5860
43. Wang, Q., Ma, L., Hao, Y. H., and Tan, Z. (2010) *Anal. Chem.* **82,** 29469–29475
44. Roy, R., Kozlov, A. G., Lohman, T. M., and Ha, T. (2009) *Nature* **461,** 1092–1097
45. Rizzo, A., Salvati, E., Porru, M., D'Angelo, C., Stevens, M. F., D'Incalci, M., Leonetti, C., Gilson, E., Zupi, G., and Biroccio, A. (2009) *Nucleic Acids Res.* **37,** 5353–5364

APPENDIX C

UV-DDB STOICHIOMETRY

Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair

Joanne I. Yeh[a,b,1], Arthur S. Levine[c,d], Shoucheng Du[a], Unmesh Chinte[a], Harshad Ghodke[e], Hong Wang[d,e], Haibin Shi[a], Ching L. Hsieh[c,d], James F. Conway[a], Bennett Van Houten[d,e], and Vesna Rapić-Otrin[c,d]

[a]Departments of Structural Biology, [b]Bioengineering, [c]Microbiology and Molecular Genetics, [e]Pharmacology and Chemical Biology, and [d]University of Pittsburgh Cancer Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260

Edited by Lorena S. Beese, Duke University School of Medicine, Durham, NC, and approved January 17, 2012 (received for review June 24, 2011)

UV light-induced photoproducts are recognized and removed by the nucleotide-excision repair (NER) pathway. In humans, the UV-damaged DNA-binding protein (UV-DDB) is part of a ubiquitin E3 ligase complex (DDB1-CUL4A^{DDB2}) that initiates NER by recognizing damaged chromatin with concomitant ubiquitination of core histones at the lesion. We report the X-ray crystal structure of the human UV-DDB in a complex with damaged DNA and show that the N-terminal domain of DDB2 makes critical contacts with two molecules of DNA, driving N-terminal-domain folding and promoting UV-DDB dimerization. The functional significance of the dimeric UV-DDB [(DDB1-DDB2)$_2$], in a complex with damaged DNA, is validated by electron microscopy, atomic force microscopy, solution biophysical, and functional analyses. We propose that the binding of UV-damaged DNA results in conformational changes in the N-terminal domain of DDB2, inducing helical folding in the context of the bound DNA and inducing dimerization as a function of nucleotide binding. The temporal and spatial interplay between domain ordering and dimerization provides an elegant molecular rationale for the unprecedented binding affinities and selectivities exhibited by UV-DDB for UV-damaged DNA. Modeling the DDB1-CUL4A^{DDB2} complex according to the dimeric UV-DDB-AP24 architecture results in a mechanistically consistent alignment of the E3 ligase bound to a nucleosome harboring damaged DNA. Our findings provide unique structural and conformational insights into the molecular architecture of the DDB1-CUL4A^{DDB2} E3 ligase, with significant implications for the regulation and overall organization of the proteins responsible for initiation of NER in the context of chromatin and for the consequent maintenance of genomic integrity.

UV damage | ubiquitin-proteosome system | X-ray crystallography

Genome integrity is under constant challenge and various cellular mechanisms exist to maintain DNA fidelity. In human cells, the nucleotide-excision repair (NER) pathway is responsible for the repair of a variety of DNA lesions (1). Although the mechanism of damage detection in chromatin is not well-understood, various studies have identified two principal initiators of the global-genome branch of NER (GG-NER), XPC-human RAD23B (XPC-HR23B) (2), and the UV-damaged DNA-binding protein complex UV-DDB (3–5). UV-DDB is composed of two proteins, a 127-kDa protein (DDB1) and a 48-kDa protein (DDB2) encoded by the *DDB1* and *DDB2* genes, respectively (6). Mutations in DDB2 cause a cancer prone autosomal recessive disease, xeroderma pigmentosum (XP) complementation group E (XP-E), and are associated with a partial deficiency in GG-NER (7–9). Through the DDB2 subunit, UV-DDB binds avidly to fragments of DNA containing various types of damage, such as UV-induced 6-4 pyrimidine-pyrimidone (6-4PP) and cyclobutane pyrimidine dimers (CPD) (10, 11). Detection of CPD in nontranscribed DNA by XPC is inefficient, indicating that the UV-DDB complex plays a primary and crucial role in the detection and repair of CPD in the context of chromatin (4, 12).

The ubiquitination pathway has recently been shown to play an important regulatory function in the initiation of NER (13, 14). The DDB1 protein is part of the substrate-recruiting module for two closely related types of E3 ligases, the cullins CUL4A and CUL4B, which target proteins for ubiquitination (15, 16). The DDB1-CUL4A complex belongs to a superfamily of cullin-RING ligases (CRL) (17–19), which participate in various aspects of the UV-damage response for maintaining genome stability (20–22). DDB2 is both a binding partner and a substrate receptor for the DDB1-CUL4A-based E3 ligase, DDB1-CUL4A^{DDB2} (11, 21, 23, 24). Following UV exposure of cells, DDB2 recruits the DDB1-CUL4A^{DDB2} complex to the site of damaged chromatin, regulating the initiation of GG-NER by modifying core histones around the site of the lesion (13, 24, 25). Available data show a connection between DDB1-CUL4A^{DDB2} and the monoubiquitination of the core histones (i.e., H2A, H3, and H4) in the cellular response to UV-irradiation (23, 24). Following initial damage recognition, the DDB1-CUL4A^{DDB2} E3 ligase ubiquitinates XPC and auto-ubiquitinates DDB2 (13), however with different consequences. Ubiquitination stabilizes XPC, increasing its affinity for damaged DNA, whereas polyubiquitination of DDB2 reduces its affinity for damaged DNA and ultimately leads to its degradation (13, 26). This paradoxical UV-dependent degradation of a protein [i.e., DDB2 (27–29)] that is intrinsically involved in the recognition of radiation-induced DNA damage is not fully understood. It has been speculated that this sequence of events is necessary for the accessibility of repair factors at the lesion site—i.e., for reducing the affinity between DDB2 and DNA to facilitate the handover of the damaged DNA from the DDB1-CUL4A^{DDB2} E3 ligase complex to XPC-Rad23 and for regulation of the cellular response to DNA damage (26, 30). It is currently unknown how DDB2 interacts with the substrate when E3 is anchored to damaged DNA nor how DDB2 targets multiple substrates of various sizes for mono- or polyubiquitination.

Recent progress in understanding the structural basis of NER initiation came from crystal structures of the yeast XPC orthologue Rad4 (31) and of the zebrafish UV-DDB bound to UV-damaged DNA (11). In these structures, these DNA-binding

---

Author contributions: J.I.Y., A.S.L., B.V.H., and V.R.-O. designed research; J.I.Y., S.D., U.C., H.G., H.W., H.S., C.L.H., and J.F.C. performed research; J.I.Y. contributed new reagents/analytic tools; J.I.Y., S.D., U.C., H.G., H.W., H.S., J.F.C., B.V.H., and V.R.-O. analyzed data; and J.I.Y., A.S.L., J.F.C., B.V.H., and V.R.-O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4E54 and 4E5Z).

[1]To whom correspondence should be addressed. E-mail: jiyeh@pitt.edu.

See Author Summary on page 16408 (volume 109, number 41).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1110067109/-/DCSupplemental.

proteins (i.e., XPC, DDB2) appear to recognize conserved perturbations to the DNA topology induced by the lesions. UV-irradiation-induced modifications, such as CPD and 6-4 PP, are believed to disrupt the dynamics and helical topology through DNA bending, altering base-pairing interactions, and widening the major groove, features that are recognized by the NER apparatus through a bidentate recognition process (11, 31–34). The means by which UV-DDB can efficiently scan DNA for damage, while at the same time binding damaged DNA with the highest affinity of any damaged DNA-binding proteins (10, 35) remain unknown. The crystal structures of UV-DDB bound to DNA containing 6-4 PP or an abasic site showed contacts between DDB2 and DNA to be limited to the β-loops, exhibiting largely identical interactions (11).

We report here the crystal structure of full length human UV-DDB bound to damaged DNA, revealing the unique structural motif of the N-terminal helical domain of DDB2. Using biophysical methods of analysis to monitor and characterize the changes in molecular associations and dynamics initiated upon damaged DNA binding, we propose that this helical domain participates in forming the high-affinity binding state of UV-DDB. Mechanistically, the conformational dependence of the N-terminal domain of DDB2 on damaged DNA binding illuminates how UV-DDB can efficiently scan the genome to detect DNA damage, while enabling high-affinity DNA interactions to be formed once damage is detected. In the context of DNA repair, modeling the cullin-RING E3 ligase nucleosome complex on the dimeric UV-DDB-AP24 architecture facilely aligns the numerous molecular components, revealing spatial orientations likely significant in substrate ubiquitination. These results support the role of oligomerization in modulating molecular flexibility, affinities, and specificities in cullin-RING E3 ligase receptor-substrate complexes.

Results

Electron Microscopy and X-ray Crystal Structure Reveal a Dimer of Human UV-DDB in a Complex with Damaged DNA. EM characterization of the full-length human UV-DDB in the presence of varying amounts of DNA identified solution conditions that stabilized the dimeric state of the complex. A central apyrimidic lesion was generated by introducing a tetrahydrofuran moiety at position 11 (THF11) in a 24-basepair oligodeoxynucleotide (AP24). Several ratios of AP24 were incubated with UV-DDB before EM imaging (Fig. 1) (additional details can be found in *SI Appendix*). Prior to DNA binding, our negative stain EM studies revealed predominantly spherical particles with a minor fraction exhibiting elliptical profiles (Fig. 1 and *SI Appendix*, Fig. S1). In the absence of DNA, the distribution of the projected protein surface area yielded a well-defined peak at approximately 3,600 Å^2, corresponding to a spherical particle of approximately 70 Å in diameter, consistent with the dimension of a monomer of UV-DDB (i.e., a heterodimer of DDB1 and DDB2; Fig. 1). However, in the presence of damaged DNA substrate (i.e., AP24), a second peak appeared with an area that is consistent with that predicted for dimeric UV-DDB [(DDB1-DDB2)$_2$], approximately 7,200 Å^2 (Fig. 1 and *SI Appendix*, Fig. S1). Notably, particle size distributions shifted dramatically from monomeric to dimeric UV-DDB at a molar ratio of 3 AP24 to 1 UV-DDB (3:1)—the ratio used for crystallization screening (described below).

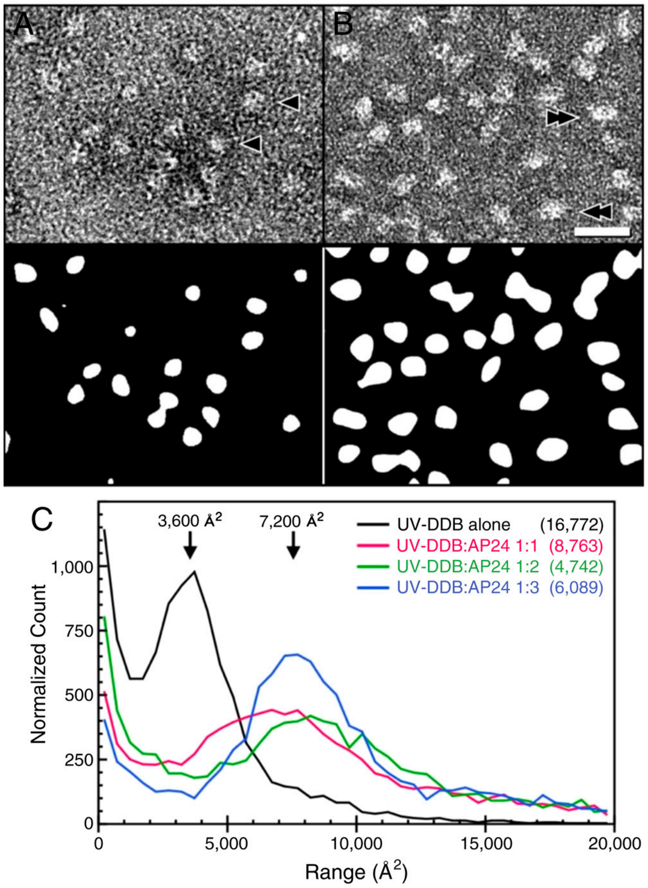


Fig. 1. Visualization and size estimation of UV-DDB particles by negative stain electron microscopy. Representative areas are shown in *A* without DNA and in *B* with AP24 oligodeoxynucleotide at a ratio of 1:3. (*Top*) Images from electron micrographs and (*Bottom*) after global and local filtering and thresholding to yield countable particle areas. (*C*) Histograms collected from micrographs of particle areas for different ratios of UV-DDB to AP24 oligodeoxynucleotide, as indicated, and normalized by particle count (in parentheses). The peak at approximately 36 nm^2 evident in the absence of DNA corresponds to a circle of diameter approximately 7 nm that is consistent with a monomer of the UV-DDB1-DDB2 complex. Increasing concentrations of AP24 oligodeoxynucleotide causes the peak shifts to approximately 72 nm^2 consistent with a population of dimers. Examples of monomer-sized areas are indicated with arrowheads in *A* and dimers with double-arrowheads in *B*.

Crystallization protocols were devised using an analytical approach (36) to systematically identify chemical and additive conditions that stabilized conformational states of UV-DDB in solution. Single crystals of native or selenomethionine (SeMet)-substituted human DDB1 and DDB2 proteins, in complex with the same damaged DNA substrate analogue used in the EM analysis, AP24 (at a molar ratio of 1:3 UV-DDB:AP24), were grown. UV-DDB-AP24 crystallized in monoclinic and orthorhombic lattices, depending on the crystallization condition. The unit cell parameters of the orthorhombic (referred to as “ortho-UV-DDB”) crystal form are very similar to the monoclinic (“mono-UV-DDB”) (Table 1), except for a doubling along one axis in the orthorhombic dataset. The early native and anomalous datasets used to phase and refine the dimeric UV-DDB-AP24 complex model were most favorably processed in a monoclinic $P2_1$ space group setting. Similarly, the highest resolution dataset used to build missing regions in DDB2, add nucleotides to the DNA substrate, modify loop conformations of DDB1, and fully refine the human UV-DDB-AP24 complex structure was also most favorably processed in monoclinic lattice setting (Table 1, first column). In the final cross-validation stage, orthorhombic data collected from crystals optimized using alternative additive conditions were used to independently verify the overall backbone tracing and subunit configuration of the dimeric UV-DDB-AP24 crystal structure.

Combinations of bromide and selenomethionine (SeMet) anomalous dispersion methods, in tandem with partial model molecular replacement approach, were applied for initial phasing and refinement. The quality of the early maps was significantly improved by combining the phases calculated from the SeMet heavy atom positions together with phases calculated from the coordinates of human DDB1 and a partial poly-Ala model of the zebrafish DDB2 (residues 100–400; PDB ID code 3EI2).

Table 1. Refinement Statistics for the Human UV-DDB Complexes

	UV-DDB-AP24 'monomeric form'	UV-DDB-AP24 'orthorhombic dimeric form'
Bravais Lattice	Monoclinic	Orthorhombic
Resolution (Å)	31.74–2.85	41.09–3.22
$R_{\text{work}}/R_{\text{free}}$	0.22/0.24	0.25/0.26
Number of atoms	13010	13010
Protein	12033	12033
Ligand/ion (DNA)	977	977
Water	0	0
$\langle$*B*-factors$\rangle_{\text{average}}$		
All atoms	35.4	39.4
Proteins	32.2	35.2
DNA	33.5	35.7
Water	—	-
R.m.s deviations		
Bond lengths (Å)	0.0052	0.0059
Bond angles (°)	1.363	1.012

Although the double-stranded (ds) AP24 oligodeoxynucleotide substrate molecule was deliberately omitted from the initial phasing model, strong contiguous densities at the surface of the β-barrel domain of human DDB2 were apparent in solvent-flattened, positively contoured difference Fourier maps, verifying the damaged-DNA bound state of the DDB2 subunit in UV-DDB crystals. Molecular features evident even in the initially phased electron density maps permitted the structure of the central approximately 18 bases in both the damaged and undamaged strands of AP24 molecule to be built according to map densities. Iterative cycles of model building to incorporate the sequence of human DDB2, to adjust regional conformational differences in DDB1 and DDB2, and to extend the AP24 oligodeoxynucleotide molecule gradually improved map and model quality. Once approximatley 90% of the structure of the complete human UV-DDB-AP24 complex was modeled and refined, phase combination utilizing a native monoclinic dataset increased data completeness, intensities, and redundancy of wide angle reflections enhancing the overall data quality, resulting in more distinct electron densities radiating from N-terminal region of DDB2 truncated in the UV-DDB model. Importantly, regions of UV-DDB that were missing or altered in conformation could be progressively modeled as map definitions steadily improved commensurate with data extension to 2.85 Å. Distinct regions of contiguous electron densities radiating from residue 100 of DDB2 allowed additional approximately 80 residues at the N-terminal domain of DDB2 to be traced, monitoring R factors and other statistical factors until refinement converged. To validate the human UV-DDB-AP24 structure we also solved an orthorhombic UV-DDB SeMet dataset to 3.2 Å resolution by ab-initio SAD phasing. The structure of the orthorhombic crystal form independently confirmed the N-terminal-domain fold and subunit organization in the dimeric UV-DDB (refinement statistics for both crystal structures are shown in Table 1). In both monoclinic and orthorhombic datasets, an elongated configuration of the dimer is recapitulated, mirroring the molecular envelope of UV-DDB seen in EM images taken in the presence of damaged DNA. Altogether, the EM and crystal data support the substrate-dependent dimerization of UV-DDB (Fig. 2).

Crystal Structure of the Dimeric UV-DDB Complex. DDB1 is a large tri-β-propeller substrate adaptor protein. Following nomenclature defined previously (19), the DDB1 β-propeller domains are denoted as BPA, BPB, and BPC, with a C-terminal helical domain referred to as CTD (37). The structure of the human DDB2 substrate receptor is composed of a large seven-bladed WD40 β-propeller domain (residues 103–421), preceded by an N-terminal domain (residues 1–102) (Fig. 2 *A* and *B*) (11). The dimeric DDB2 forms the core of the UV-DDB complex, with a twofold axis located close to blade 6 of the major seven-bladed β-propeller domain of DDB2 (Fig. 2 *A* and *B*; yellow), the conserved WD40 structural motif.

Distinct Topological Motifs Mediate Associations Between DDB1 and DDB2. In the human UV-DDB dimeric complex structures, the previously missing N-terminal helical domain of DDB2 (11) has been built by modeling into experimentally phased electron density maps. The N-terminal region preceding the β-propeller domain of DDB2 is composed of approximately the first 102 residues and topologically distinguished by predominantly helical features. The first 66 residues fold into three helical segments arranged into a triangular topology (α-paddle, in red, Fig. 2 *A*–*C*), followed by an extended helix-turn-helix (residues 67–102) that inserts into the BPA-BPC double-propeller cleft (Fig. 2 *A* and *B*). The variations in conformation and domain organization in the dimeric relative to the monomeric states of UV-DDB are primarily centred at the DDB2 component.

The interface formed between BPA-BPC double propellers of DDB1 displays significant hydrophobic characteristics, concentrated mainly on the surface of the BPA domain (*SI Appendix*, Fig. S2*A*; hydrophobic surfaces in white, defined by a dotted yellow oval) facing the cleft where the β-propeller domain of DDB2 docks. Upon initial complex formation, extensive hydrophobic contacts are formed between residues on the BPA domain and aliphatic loop residues extending from the β-propeller domain of DDB2 (*SI Appendix*, Fig. S2*B*; hydrophobic surfaces in white, defined by a dotted blue rectangle). In contrast, interactions between the BPC domain of DDB1 (*SI Appendix*, Fig. S2*A*; negative electrostatic surfaces in red, positive in blue, defined by a dotted violet oval) and the N-terminal-α-helical region preceding the β-propeller domain of DDB2 are largely electrostatic in nature (*SI Appendix*, Fig. S2*B*; negative electrostatic surfaces in red, positive in blue, defined by a dotted green rectangle). The helical topology of the N-terminal domain of DDB2 is important for aligning clusters of acidic/basic residues on surfaces facing the BPA domain, enabling charge complementation at the intermolecular interface (*SI Appendix*, Fig. S2*C*). The α-paddle helical fold of DDB2 segregates aromatic residues to the opposite face of the N-terminal domain of DDB2, juxtaposing hydrophobic patches to form favorable contacts to the BPC domain of DDB1. The helical conformation of the N-terminal domain of DDB2 is a key topological feature that enables the precise spatial alignment of residues at intermolecular interfaces, mediating multiple molecular associations in the context of the complete complex.

Damaged DNA Binding Induces Helical Folding of the N-terminal Domain of DDB2. Another novel structural feature found in the dimeric UV-DDB-AP24 complex is at the region encompassing residues 356–370 of DDB2, which forms a well-ordered loop that extends from blade 6 of the seven-bladed β-propeller domain of DDB2, forming a fold defined herein as β-wing (Fig. 2 *A*–*C*). The closest contacts between two DDB2 subunits within the dimer are at the β-wing regions of DDB2, centred at Asn360. The twofold axis is located between two Asn360 side chains, which form of H-bonds with favorable geometries and distances [Fig. 2*A*; 2.8 Å between neighboring twofold related Oδ1 (red) and Nδ2 (blue) atoms of Asn360].

In the dimeric DDB2, the β-wing loops are sandwiched between two DNA molecules, with β-wing residues forming contacts to both the DNA immediately bound and to its neighboring DNA bound to the second β-propeller domain (*SI Appendix*, Fig. S3). At one end of the β-wing turn, an electrostatic network stabilizes the undamaged DNA strand immediately opposite the lesion (*SI Appendix*, Fig. S2*A*; defined by a dotted yellow oval). Bonds between the β-wing residues to the DNA are predominantly electrostatic in nature, contacting anionic phospho-deoxyribose backbone atoms of the DNA, similar in nature to those formed

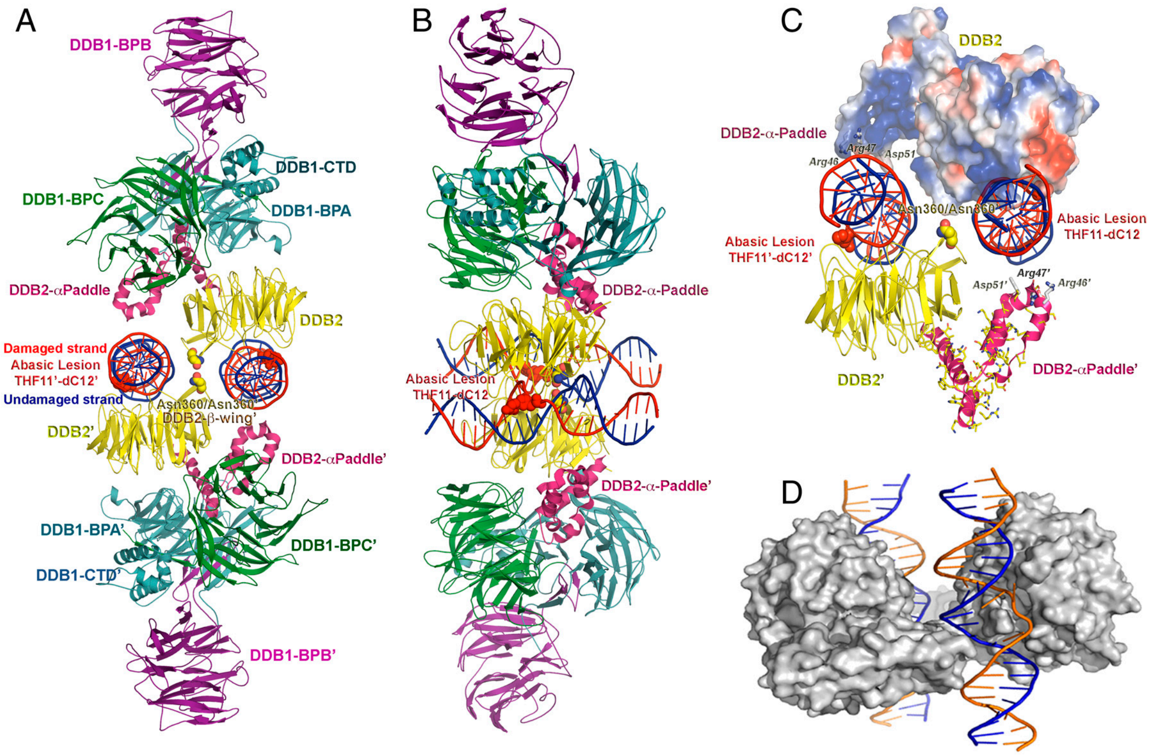


Fig. 2. Structure of the dimeric human UV-DDB in a complex with damaged DNA. (*A*) The dimeric UV-DDB subunit organization, shown in ribbon depiction, with each domain colored and labeled accordingly: yellow, DDB2 β-propeller; red, DDB2 N-terminal-α paddle; blue, DDB1 BPA; green, DDB1 BPC; and purple, DDB1 BPB. The 24-bp oligodeoxynucleotide (AP24) contains an abasic lesion site (THF11), with the phosphor-deoxyribose backbone of the damaged strand colored in red and the undamaged strand colored in blue. Each DDB2 subunit is bound to an AP24 oligonucleotide, with DDB2 residues Asn360/Asn360′ straddling the twofold symmetry axis, forming H bonds across the dimer interface. The surface of the Asn360/Asn360′ pair (colored using standard atom convention) is located in a loop spanning two antiparallel β-strands (β-wing). The abasic lesion site in AP24 is marked by the surface mesh drawn around nucleotides THF11/dC12 in their flipped, extra-helical configuration. The β-wing is sandwiched between the two AP24 oligodeoxynucleotides, astride of the twofold axis of rotation relating the monomer subunits in the dimeric DDB2. Residues on the leading β-strand and loop form contacts with the undamaged DNA strand whereas residues on the loop and the retreating β-strand form contacts with the neighboring undamaged DNA strand. Both sets of contacts are predominantly electrostatic in nature, thus largely sequence independent. (*B*) Same as *A* but rotated 90 degrees and tilted slightly to show both DNA molecules. (*C*) Electrostatic potential surfaces of the DDB2 N-terminal domain complement the charge characteristics of the DDB1 BPC domain and the DNA phosphor-deoxyribose interfaces, resulting in favorable electrostatic neutralization. Contacts between residues on the β-wing region form contacts with the DNA bound at its immediate active site and with the neighbouring DNA molecule bound to the second monomer of DDB2 in the dimer. Extensive interactions between residues on the N-terminal-helical domain (α-paddle) and the neighboring DNA molecule augment the intermolecular associations, contributing to the high affinity of damaged DNA binding. (*D*) The skewed positioning of the DNA binding surface can now be understood in terms of the DDB2 dimer interface, located adjacent to the DNA binding site, at a loop bridging blades 6 and 7 of the β-propeller (β-wing) of DDB2. To accommodate the steric constraints imposed through dimerization along with DNA binding, the two adjacent sites are positioned diametrically across one face of the molecular surface of DDB2, readily seen in the dimeric DDB2-DNA (AP24) complex.

by the N-terminal domain of DDB2. In comparison, contacts formed at the lesion site between the insertion loop of DDB2 and the damaged strands are nucleobase specific. Explicit contacts centred within a 3-nucleotide window on the duplex DNA are limited by atomistic and spatial constraints dictated by the DNA topology and chemical functionality of DDB2's insertion loop residues.

Located in the N-terminal-α-paddle domain of DDB2 are numerous arginines, lysines, glutamates, aspartates, and glutamines, aligned by the helical topology along a face of the α-paddle, presenting charged residues (Arg46, Arg47, Asp51; shown as sticks, Fig. 2*C*) to the phospho-deoxyribose backbone of the DNA. Additional charged residues contributed by the adjacent β-propeller domain of DDB2 further enhance the highly cationic electrostatic surface of the N-terminal domain of DDB2 (*SI Appendix*, Fig. S2*B*; positive electrostatic surfaces shown in blue). These form complementary electrostatic molecular interfaces for binding both the DNA and the BPC domain of DDB1 (*SI Appendix*, Fig. S2 *A*–*C*). As noted earlier, the helical motif of the N-terminal domain of DDB2 is important for aligning charged residues on one face of the helices and projecting hydrophobic residues on the other, enabling these to simultaneously interact with the hydrophobic BPA domain of DDB1 and the highly anionic nature of the DNA phosphor-deoxyribose backbone as well as the largely cationic surface features of the BPC domain of DDB1.

The dimer interface of DDB2 modulates multiple intermolecular contacts, providing a structural rationale for the remarkably high binding affinities to damaged DNA found in biochemical studies of UV-DDB (10, 35, 38). In the current studies, the 24-bp oligodeoxynucleotide (AP24) is substantially longer than the 14-and 16-bp oligodeoxynucleotides used in the earlier structural studies, affording unique insight to the interactions beyond those formed directly at the lesion site. Four distinct networks of contacts are formed between DDB2 and the DNA. At the lesion site, DDB2 residues (His333, Phe334, Gln335, His336, represented by orange spheres and rectangles; *SI Appendix*, Fig. S3 *A*–*C*) insert at the abasic site, resulting in flipping of the immediate upstream nucleotide, which is stabilized in an extra-helical conformation through a second group of DDB2 contacts (yellow spheres; *SI Appendix*, Fig. S3 *A*–*C*). Beyond the lesion, interactions between DDB2 and the deoxyribose-phosphate backbone atoms upstream (green spheres, rectangles; *SI Appendix*, Fig. S3 *A*–*C*) and downstream (purple spheres, rectangles; *SI Appendix*, Fig. S3 *A*–*C*) on the damaged DNA strand further serve to clamp the DNA. The distinct nature and multiple levels of interactions found between the DNA molecule and DDB2 in the crystal structure are highly suggestive that the additive contributions from the multiple sites of contacts are a major factor for the high overall avidity of binding exhibited by UV-DDB for damaged DNA.

DNA Binding Promotes Dimerization of the DDB1-DDB2 Heterodimer (DDB1-DDB2). To further probe the oligomeric states of UV-DDB under physiologically relevant solution conditions, atomic force microscopy (AFM) was used to characterize the molecular topology of UV-DDB and to monitor substrate-induced changes in intermolecular interactions. AFM-derived volumes have been used extensively in studies examining the oligomeric states of multicomponent complexes and to ascertain the nature of protein–protein interactions of globular proteins (39, 40). For our AFM studies, 517-bp PCR fragments were produced as the undamaged DNA substrate and the fragments were subjected to UV-irradiation to generate the damaged DNA (41, 42). AFM analyses of UV-DDB in the presence of undamaged DNA (*SI Appendix*, Fig. S4 *A* and *B*), UV-damaged DNA (Fig. 3), and in the absence of substrate binding (Fig. 3) found clearly distinguishable changes in the oligomeric states of UV-DDB. To quantitate the volumes and molecular mass derived from the AFM data, a standard curve was generated using proteins with well-defined oligomeric states, shapes, and molecular masses (additional experimental details are provided in *SI Appendix*, *SI Materials and Methods*). Converting the apparent AFM volumes of the UV-DDB molecules in the absence of DNA to molecular mass using the standard curve (*SI Appendix*, Fig. S5) shows that the peak at approximately 66 ± 10 nm^3 (three depositions) corresponds to a protein with a molecular mass of 184 ± 23 kDa, a value consistent with the size of monomeric UV-DDB (i.e., heterodimeric DDB1-DDB2, with a combined theoretical molecular mass of 175 kDa). A second peak at approximately 190 nm^3, corresponding to a molecular mass of approximately 505 kDa, represents a complex mixture of higher

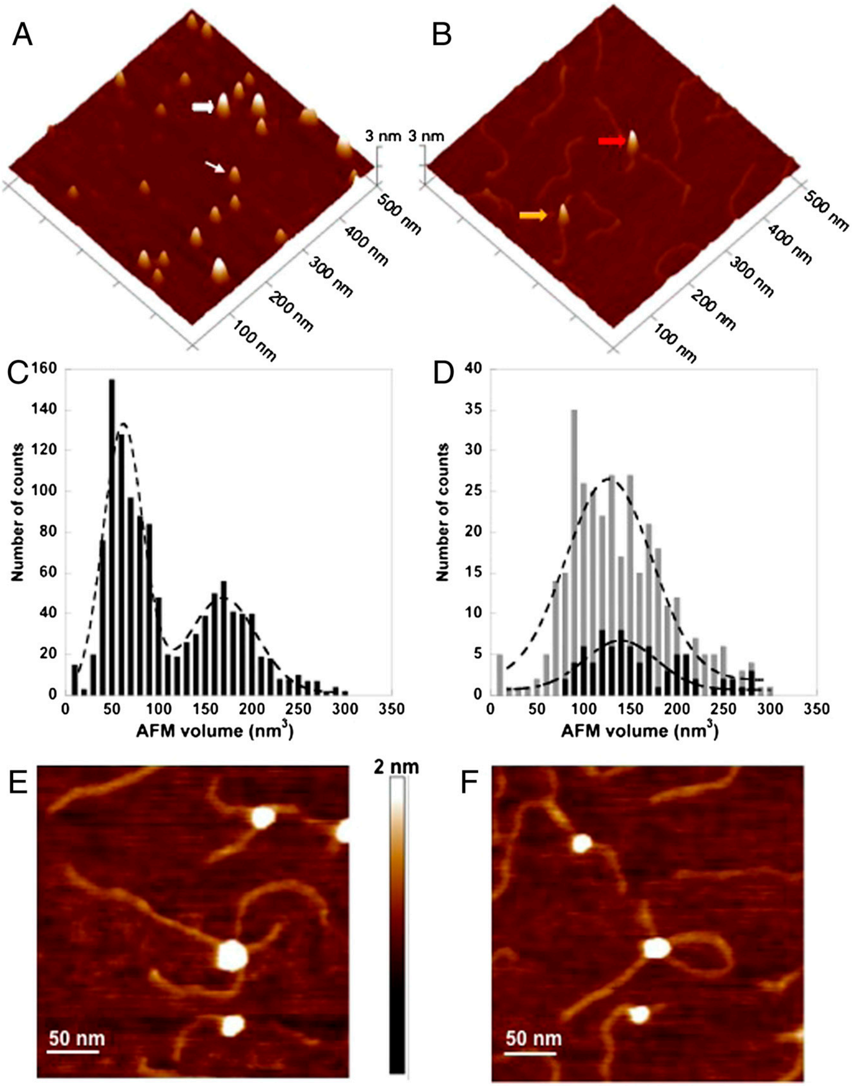


Fig. 3. AFM imaging shows that damaged DNA binding promotes the dimerization of the DDB1-DDB2 heterodimer. (*A*) A representative surface plot of UV-DDB (50 nM) in the absence of DNA. The thin and wide white arrows point to molecules consistent with the size of the UV-DDB monomer (DDB1-DDB2 heterodimer) and trimer of UV-DDB, respectively. (*B*) Representative surface plot of UV-DDB (50 nM) in the presence of UV-irradiated 517 bp PCR fragments (25 nM). The yellow and red arrows point to dimeric UV-DDB [(DDB1-DDB2)$_2$] binding to one and two molecules of duplex DNA, respectively. (*C*) AFM volume analysis of free UV-DDB ($n = 1,160$). (*D*) AFM volume analysis of UV-DDB on one strand (gray bars, $n = 339$) and two strands (black bars, $n = 79$) of duplex DNA. The images in *A* and *B* are at 500 nm × 500 nm and 3 nm in height. (*Bottom*) The dashed lines (*C*, free in solution, and *D*, bound to DNA) represent Gaussian fits to the data. Field view images of UV-DDB binding to separate DNA molecules (*E*) or two different regions of the same DNA molecule (*F*). The images are at 300 nm × 300 nm and 2 nm in height.

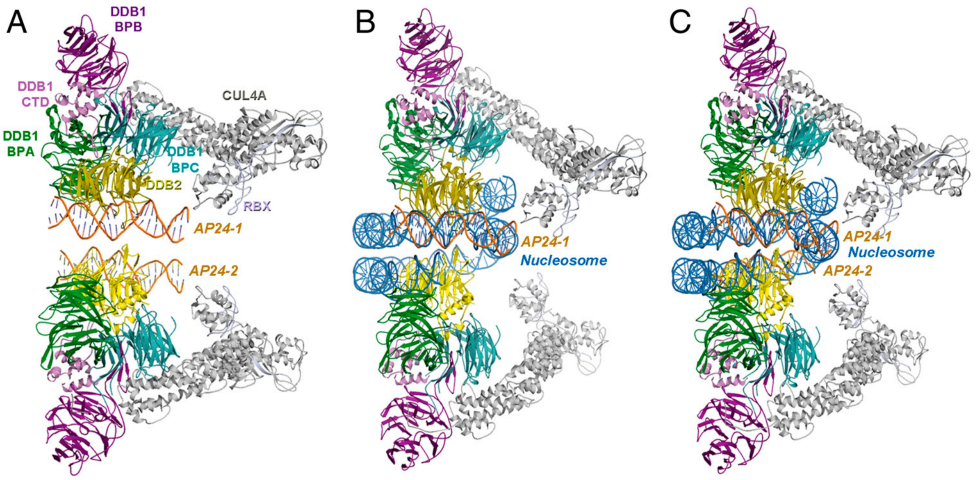


Fig. 4. Model of DDB1-CUL4A$^{\mathrm{DDB2}}$ ubiquitin ligase complexed to a nucleosome. (*A*) Modeling of the complex with CUL4A-Rbx (gray, light blue) onto the dimeric UV-DDB2 (domains colored as in Fig. 2); the region defined by two adjacent AP24 oligodeoxynucleotides (AP24-1 & AP24-2, in orange) used for the docking of a nucleosome; (*B*) Docking of the nucleosome in the dimeric UV-DDB, showing the fit of one AP24-1 (in orange) relative to the nucleosome (in blue); (*C*) fit of the nucleosome onto both oligomers showing that the distance between the two oligonucleotides can readily accommodate the nucleosome molecule (second DNA molecule, AP24-2, shown in orange) with minor adjustments of the second DDB2 component, as needed. The dimeric scaffold accommodates the numerous proteins that transiently assemble and disassemble on the DDB1-CUL4A$^{\mathrm{DDB2}}$ ubiquitin ligase complex at the vicinity of the lesion site in the subsequent DNA repair process. The dimeric architecture also spatially aligns the various molecular subunits in the reactions monoubiquitinylating histones and polyubiquitinylating substrate receptors (i.e., DDB2) for proteasomal degradation and verified by docking the E2 ubiquitin transferase enzyme onto the DDB1-CUL4A$^{\mathrm{DDB2}}$ ubiquitin ligase complex, resulting in E2 bridging distances to histones. In this figure, the histone and E2 proteins are omitted for clarity.

order oligomeric states (Fig. 3*C* and *SI Appendix*, Fig. S5). Consistent with the EM results, no distinct peak at a volume corresponding to a dimer of UV-DDB [(DDB1-DDB2)$_2$] is found in the absence of DNA.

When UV-DDB was incubated with UV-irradiated DNA at a molar ratio of 2∶1, approximately 96% of the UV-DDB molecules were found to be bound to DNA molecules (Fig. 3 *B* and *D* and *SI Appendix*, Fig. S4*C*). In addition, among all of the UV-DDB molecules bound to DNA, 18% of the molecules bound to two DNA molecules simultaneously. These binding events included both middle to middle (Fig. 3 *E* and *F*) and end to middle sites (Fig. 3*B*, red arrow) on two separate DNA molecules. The AFM-derived volume of UV-DDB bound to one molecule of DNA is Gaussian centred at approximately 133 nm^3 (Fig. 3*D*; gray histogram) which is consistent with the size of dimeric UV-DDB [(DDB1-DDB2)$_2$]. The AFM volume of UV-DDB simultaneously binding to two molecules of DNA is approximately 139 nm^3, which is slightly larger than UV-DDB binding to only one molecule (Fig. 3*D*; black histogram). These oligomeric states and substrate interactions found from the AFM analysis of UV-DDB in the presence of damaged DNA contrasts dramatically to those found when UV-DDB is in the presence of undamaged DNA substrate.

The interactions found between UV-DDB and the undamaged DNA fragment from the AFM analysis indicates that UV-DDB binds undamaged DNA to a significantly reduced extent (approximately 37% of the total UV-DDB was bound to DNA) compared to those formed when UV-damaged DNA is present (96%) (*SI Appendix*, Fig. S4*C*). These AFM results are consistent with those found earlier by EMSA analysis, which showed that small but measureable amounts of UV-DDB bound undamaged DNA (38, 43). Analysis of the volumes of these nonspecific UV-DDB complexes observed on DNA indicated that a majority (approximately 75%)[*] of the complexes were monomeric consisting of only one molecule of DDB1 and DDB2 (*SI Appendix*, Fig. S4*B*). These volume measurements are in striking contrast to the AFM-derived volume of UV-DDB when bound to UV-irradiated PCR fragments, inducing volume changes in UV-DDB that were consistent with a dimeric UV-DDB [(DDB1-DDB2)$_2$] bound to damaged DNA.

While EM and AFM studies involved different sample preparation procedures and different criteria for evaluating size (projected area vs. volume), these complementary techniques can provide insights into the molecular topologies, organization, and nature of interactions in multi-component complexes. Compounding the innate methods-related differences described above, the sizes of the damaged DNA substrates used in the EM and AFM studies represented different lesions types (THF in AP24 versus UV-induced lesions in 517-bp PCR fragments). Yet the EM and AFM studies provided corroborating data verifying that the shorter abasic site mimic (AP24) induced similar dimerization upon damaged DNA binding as found when UV-DDB bound the 517-bp UV-irradiated DNA fragments. Notably, the remarkable agreement between the dimensional values and molecular profile obtained by the negative-stained EM, AFM, and derived from the X-ray diffraction crystal structure supports our proposal that UV-DDB dimerizes as a function of damaged-DNA binding. To summarize, EM and AFM imaging revealed that (*i*) in solution, UV-DDB exists as a monomer (composed of DDB1-DDB2 heterodimer) and no significant dimer population of UV-DDB [(DDB1-DDB2)$_2$] was observed; (*ii*) binding to damaged DNA promotes the dimerization of UV-DDB, which can simultaneously bind to two DNA molecules; (*iii*) the organization and dimensions of the dimeric UV-DDB-damaged DNA complex found in the AFM and EM analyses are consistent with those found in the crystal structure of the dimeric complex. It is important to note that whereas the DDB1-CUL4A$^{\mathrm{DDB2}}$ ubiquitin ligase complexed to a nucleosome modeled according to the dimeric architecture captured in our crystal structure leads to a surprisingly rational organization of the individual molecular components (Fig. 4), the dimeric state does not necessarily constrain the number of lesions that can be simultaneously bound nor inform about the number of lesions required to induce dimerization; but the assumption that a single lesion can induce the dimeric form is reasonable.

Dimeric UV-DDB Binds Damaged DNA with Approximately Fourfold Higher Affinity than in the Monomeric State. The role of oligomer-

---

[*] Percentage value calculated from the integration of the number of molecules under two Gaussian fits after deconvoluting the peaks shown in (*SI Appendix*, Fig. S4*B*), *Inset*.

ization in mechanisms involving DCAF proteins and, more specifically, the dimerization of DDB2, has been proposed earlier (44, 45). In our study, numerous lines of structural evidence from crystallography, EM, and AFM results, when combined with earlier published results (10, 44) coherently implicates the involvement of dimeric UV-DDB $[(\mathrm{DDB1\text{-}DDB2})_2]$ in mediating molecular interactions at specific stages along the DNA repair pathway. To experimentally validate the premise derived from the structural results, that dimerization mediates DNA-binding activities, dynamic light scattering (DLS) and surface plasmon resonance (SPR) were used to characterize the interactions between monomeric and dimeric UV-DDB to AP24. Unlike EMSA or DNA footprinting gel-based assays that require labeling and are end points measurements that do not allow for kinetic analysis, DLS and SPR monitor molecular interactions in real time, permitting delineation of concentration dependencies and other solution effects on molecular interactions.

Characterization by DLS clearly shows that the binding of damaged DNA results in the formation of a distinct, monodisperse state of UV-DDB, with dimensions in agreement with the EM and AFM values for the dimeric UV-DDB $[(\mathrm{DDB1\text{-}DDB2})_2]$ (*SI Appendix*, Tables S2 and S4). Furthermore, dimerization of UV-DDB is readily promoted by binding AP24, even at dilute protein concentrations, whereas in the absence of damaged DNA binding, a mixture of monomers and dimers is found even at high protein concentrations. The kinetics and binding affinities of UV-DDB to AP24 differ in the monomeric versus dimeric states, according to the SPR data, which clearly show that the dimerization of UV-DDB is stimulated by damaged DNA binding and that, moreover, the kinetics of both the association and dissociation steps are modified, resulting in >four-fold[†] enhancement in the damaged DNA binding affinities in dimeric compared to the monomeric states (*SI Appendix*, Table S3). The combination of data from AFM, EM, and biophysical analysis (*SI Appendix*, Tables S2–S4) presented here supports the biological relevance of the dimeric state of UV-DDB as revealed in the crystal structure.

Discussion

Our comprehensive study has elucidated the structure of dimeric UV-DDB in a complex with damaged DNA, utilizing a combination of structural, biophysical, and biochemical approaches that collectively support the pivotal role that dimerization plays in modulating intermolecular associations and in organizing the architecture of the multi-component cullin-RING E3 ligase receptor-substrate complexes. The dimeric UV-DDB structure presented in this paper provides the first high-resolution views of a cullin-RING E3 ligase receptor-substrate complex captured in a high-affinity state, with direct mechanistic and functional implications.

Dimer Interface is Adjacent to the Damaged DNA Binding Site in DDB2. The primary damaged DNA binding site is located at the narrow end of the β-propeller, opposite to the DDB1 interaction surface in DDB2 (Fig. 2). The DNA spans the surface of the DDB2, surprisingly offset from the center of the seven-bladed β-propeller. This offset can now be explained in terms of the constraints arising from the dimerization interface. To accommodate these constrains along with DNA binding, the two neighboring binding sites are located diametrically across a molecular face of the β-propeller domain of DDB2, leading to the offset. The location of the dimer interface on the same molecular surface as the DNA binding site allows for cooperativity between DNA binding and dimerization (*SI Appendix*).

A High-Affinity DNA Binding Motif is Formed by Dimerization of UV-DDB. Given the overwhelming binding preference exhibited by UV-DDB to UV-induced lesions, a mechanism based on substrate-driven conformational folding of the N-terminal-domain of DDB2 would permit specificities and binding affinities to be tuned, optimizing interactions according to the specific chemical nature of the lesion site. This mechanism ensures that high affinity interactions are formed only when damage is found. Multiple unique sites of DNA contacts are found in the dimeric UV-DDB, interactions that are absent in the monomeric state and mediated by the N-terminal-α-helical paddle and the β-wing regions of DDB2.

The α-paddle helical fold adopted by the N-terminal domain of DDB2 aligns residues so contacts to the DNA immediately bound and to a neighboring DNA molecule are formed in tangent, promoting the dimerization of UV-DDB. The β-wing loop of DDB2 is located at the interface of two DNA molecules within the dimeric UV-DDB. These interactions independently augment DNA contacts but when analyzed together resemble a “winged helix” motif that has been found in numerous DNA-binding proteins (46). Analogous to those found in other winged-helix DNA binding proteins, the β-wings in UV-DDB form direct contacts to the backbone atoms of the DNA. Additionally, the conformation and the apparent function of the β-wings of DDB2 in dimeric UV-DDB resemble those shown in the transcription factors, Ets-1, and the tripartite factor X, RFX, by linking and modulating nucleotide binding affinities with dimerization (47, 48). In these winged-helix proteins, exposed patches of hydrophobic residues are displayed, causing conformational changes to present new protein–protein interaction surfaces and inducing dimerization as a function of nucleotide binding (*SI Appendix*). Thus, the N-terminal domain of DDB2 modulates molecular affinities, independently and in conjunction with its β–wing domain, while further coordinating dimerization. The specific α-paddle helical motif enables multi-molecular contacts to be formed without perturbing interactions with DDB1 and, presumably, without blocking the subsequent binding of proteins involved in DNA repair. The dimeric DDB1-CUL4A$^{\mathrm{DDB2}}$ ubiquitin ligase complexed to a nucleosome, modeled according to the molecular architecture of the UV-DDB-AP24 complex (Fig. 4), demonstrates that the multi-component complex can be accommodated within the dimeric framework, providing additional support for the plausibility that dimerization of UV-DDB regulates and modulates association to DNA lesions.

Dimerization Accommodates Spatial Constraints for Substrate Ubiquitination. Recently, the concept that dimerization is the key molecular determinant in enabling interactions with the vast and diverse set of proteins targeted by CRLs and their complexes has gained prominence. The dimeric state would be advantageous to monomeric E3 in targeting proteins of different sizes and in regulating auto-ubiquitination of the substrate receptor. The functional importance of CRL dimerization is supported by the observation that mutations of substrate-recognition regions retain their dimerization properties but act in a dominant-negative fashion, in vivo (45). Formation of higher-order oligomers can be initiated by receptor association or through another E3 component (49).

Consideration of the holocomplex containing the E3 ligase indicates that the molecular architecture of the DDB1-CUL4A$^{\mathrm{DDB2}}$ complex should complement the ubiquitination machinery in the assembled state. The cullin subunit is an elongated moiety in all cases, consisting of a long stalk and a globular domain RING finger adapter protein, RBX1, which docks through an intermolecular β-sheet, forming a two-subunit cataly-

---

[†] Determined by accounting for the presence of both monomeric and dimeric states, using the mass distributions found from the DLS measurements, conducted using identical protein concentrations and experimental conditions, and estimating respective contributions of monomeric and dimeric states to the kinetic and affinity parameters calculated from the SPR data.

tic core that recruits the ubiquitin-conjugating enzyme, E2. It is recognized that the cullin subunit (e.g., CUL4A/4B, CUL1, CUL5) serves as a rigid scaffold in organizing the various substrates for ubiquitination after complex formation. The distinct structural motif displayed by various cullin complexes results in a distance of over 100 Å between the RBX-E2-Ub proteins relative to the substrate protein (45), which poses a question as to how activated ubiquitin bridges the 100-Å gap. However, dimerization in conjunction with domain flexibility found in the dimeric UV-DDB complex, described here, appears to address the multitude of proteins and chemical variability while permitting the dynamic adjustments needed during the polyubiquitination reaction.

In the context of the holo-complex with CUL4A, the dimeric UV-DDB structure seems to be optimized to meet the spatial requirements of the elongated cullin architecture. Modeling the DDB1-CUL4A^{DDB2} E3 complex by superimposing CUL4A-RBX onto the BPB-domain of DDB1 in the dimeric UV-DDB complex shows that the 100-Å distance is readily bridged by the activated ubiquitin-E2 moiety (E2-Ub). Particularly striking is the overall alignment when the nucleosome is oriented onto the dimeric architecture, using the AP24 with the dimeric UV-DDB as the reference for placement of the nucleotide backbone (Fig. 4). The specific BPB domain conformation of the DDB1 subunit captured in the dimeric UV-DDB would present the activated Ub (CUL4A-RBX1-E2 –Ub complex) within 10 Å of several lysines of DDB2 that are candidate sites for auto-ubiquitination. The model also indicates that the resulting complex would additionally position the histone proteins, which are also ubiquitinated, to E2-Ub. Consequently, dimerization of the substrate-recognition subunits, as exemplified by UV-DDB, further supplements regulatory, fine-tuning activities so that a spectrum of ubiquitination can be moderated, possibly permitting the simultaneous modification of multiple substrates (e.g., XPC, histones and DDB2) and/or mono- versus poly-ubiquitination of substrates (e.g., poly-Ub of DDB2 and mono-Ub of H2A). The structural and biochemical findings reported here provide compelling evidence for the dimeric state as a critical organizational unit of UV-DDB. The dimeric associations found in the UV-DDB may be representative of those formed in other complexes based on the DDB1-CUL4 ligase platform.

Functional Implications of the Dimeric State of UV-DDB. The preference for binding to UV-damaged DNA by UV-DDB was verified by AFM analysis, which found significantly less binding of UV-DDB to an undamaged 517 bp PCR fragment, results consistent with the high specificity reported for UV-DDB (10, 35, 38, 43). The AFM analysis also shows the dimeric state of UV-DDB, bridging two duplexes of UV-damaged DNA, under conditions that reproduced UV-DDB's specificity for damaged DNA.

The functional significance of the dimeric state can be assessed by comparing the molecular regions identified as structurally significant to those reported by other approaches. In the dimeric UV-DDB complex structure, the β-wing represents the area forming closest intermolecular contacts and the functional importance of this region has strong genetic support. Four DDB2 variants, formed by alternative splicing, were identified in Hela cells (44). The D1 variant, with deletion of residues 153–341 that excluded part of the β-propeller domain of DDB2, but preserved Asn360 and the β-wing, could form dimers with DDB2-WT and itself. The variant D2, containing only the first 156 amino acid residues, could not form the dimeric complex. Interestingly, DDB2 splice variants are dominant negative inhibitors of NER when expressed in HeLa cells (44). The deleterious effects of these splice variants are difficult to reconcile from the respective locations of the residues or segments according to the structure of the monomeric UV-DDB complex. However, these resides map to regions at the β-wing of DDB2 in our dimeric structure of the UV-DDB-damaged DNA complex, residues centrally positioned to bridge both molecules of DNA in the dimeric configuration. Thus, residues located in the vicinity of the dimer interface (e.g., on the β-wing region according to the UV-DDB-AP24 dimer configuration in the crystal structure) may function in signalling the substrate complexed state of DDB2, leading to cooperative enhancement of DNA binding affinities upon stimulating helical folding of the N-terminal domain of DDB2.

Our studies suggest that the transition between disordered to ordered folding of the N-terminal domain of DDB2 may be intimately related to modulating the intermolecular associations formed subsequent to those primary contacts formed immediately upon docking of DDB2 to DDB1 and upon the binding of damaged DNA to DDB2 at the substrate binding site. Through a series of fine-tuning steps, secondary intermolecular contacts are formed between the damaged DNA substrate to DDB2 (i.e., at the β-wing and N-terminal domain of DDB2 to damaged DNA ± 4 nucleotides from the lesion site) (*SI Appendix*, Fig. S3*C*) and DDB1 to DDB2 (at the interface between the BPC domain of DDB1 to the N-terminal domain of DDB2) (*SI Appendix*, Fig. S3). These molecular interactions allow the orientation of the individual subunits within the multiprotein complex to be adjusted so that the plethora of reactions catalyzed by the DDB1-CUL4A^{DDB2} multiprotein complex can be accommodated for the monoubiquitination of histones, and polyubiquitination of DDB2, ultimately leading to DNA repair.

Regulation by oligomerization has been speculated for other proteins involved in binding various states of DNA. Our extensive studies commenced with the crystal structure elucidation, revealing the dimeric state of the UV-DDB-AP24 complex, and expanded to in-depth, multi-dimensional biophysical and structural characterization of the substrate-binding dependencies, are consistent with inducing distinct dimeric states of UV-DDB. These different lines of analysis consistently point to the roles played by dimerization and localized conformational changes in protein subunits which together modulate conformation of the multicomponent E3 ligase complex and influence catalytic efficiencies of specific reactions. The iterative cycles involve discrete modification of subunit intermolecular contacts that propagates to the overall complex and permits a spectrum of activities to be generated, centered on dimerization that additionally reduces spatial and molecular constraints while increasing the range of subunits and reactions that can be accommodated.

The binding of UV-damaged DNA initiates conformational changes at the N-terminal domain of DDB2, inducing helical folding in the context of the bound DNA to promote dimerization of the UV-DDB-substrate complex, to ensure that high affinity contacts are formed only when damage is found in DNA. This temporal and spatial interplay between domain ordering and dimerization provides an elegant molecular rationale for DDB2's enhanced UV-damaged DNA selectivity (10, 38). Based on the additional extensive contacts formed by the dimeric UV-DDB, oligomerization can modulate substrate affinities on multiple levels, serving to allosterically regulate the substrate-receptor complex.

XP-E Mutations Disrupt Key Intermolecular Contacts in Dimeric UV-DDB. The mutations found in XP-E patients have been mapped to their locations on the human UV-DDB complex structure (*SI Appendix*, Fig. S6; key amino acid mutations shown in space-filling depictions). Genetic mutations identified in XP-E patients (7, 9) are at residues that form either direct or key bridging interactions with the oligodeoxynucleotides (K244, D307) or DDB1 (R273, L350). Perturbation of these contacts is highly detrimental because these mediate both direct and secondary interactions with the DNA or DDB1. The effects for two of the mutations highlight the significance of the N-terminal helical domain and the dimer interface of DDB2. The L350P mutant would significantly perturb the stability of the DDB1-DDB2 complex as L350 aligns the long N-terminal-α-paddle of DDB2 that inserts into the BPA-BPC domain cleft of DDB1. The position of L350 is central to an aliphatic cluster at the DDB1 BPC interface with DDB2, so that mutation would cooperatively disrupt multiple associations. The position of D307 is at the DDB2 dimer interface, close to the β-wing loop, disturbing DNA binding as well as dimer formation. Notably, the mutated residues identified in XP-E patients are at sites in DDB2 that mediate multiple contacts, with the detrimental consequences amplified due to disruption of correlated interactions.

Conclusions

We describe here the 2.85-Å dimeric structure of the full-length human UV-DDB $(\mathrm{DDB1\text{-}DDB2})_2$ in a complex with damaged DNA. This new structure revealed the importance of the N-terminal 102 residues of DDB2 in mediating interactions with DDB1 and damaged DNA. The remarkable agreement on the molecular topology between the negative-stained EM, AFM, and crystal structure results (Figs. 1–4), further validated by DLS and SPR analysis, collectively supports the distinct dimeric state formed by UV-DDB upon binding damaged DNA. Taken together these multiple lines of evidence strongly support the existence of higher oligomeric states of UV-DDB, in vivo. These findings have direct regulatory and functional implications.

The dimeric UV-DDB acts as a molecular scaffold for aligning multiple protein partners, during the complex and dynamic process of damaged DNA detection and repair. DDB1-CUL4A$^{\mathrm{DDB2}}$ assists in transfer of ubiquitin from the E2 to the histones and repair proteins at the site of a lesion. UV-DDB thus has a unique role for the initiation of NER in the context of chromatin. The structural flexibility of the N-terminal domain of DDB2 suggests that this domain is conformationally adaptable, its precise domain-fold driven by substrate binding. This structural malleability in receptors enables recognition of a wide array of diverse protein and nucleic acid substrates. DDB2 presumably forms complexes with multiple substrates, including histones and XPC in addition to damaged DNA, displaying a range of binding affinities to different chemical lesions found in UV-irradiated DNA. We surmise that DDB2's selectivity—its ability to distinguish subtle discrete differences in chemical moieties within the framework of a nucleosome—is related to its domain-fold adaptability. Considered all together, substrate-induced N-terminal-domain folding endows molecular and conformational adaptability, features that are further enhanced and optimized by dimerization.

Our findings regarding the significance of the dimeric state of the UV-DDB-AP24 agree remarkably well with reports on other DCAF-family proteins, providing a molecular scaffold for integrating the assorted biochemical, genetic, and cellular observations into a coherent mechanism directing NER. A central tenet evolving from these multiple lines of evidence is the pivotal role that oligomerization plays in modulating specificities and affinities of associations in multi-component macromolecular complexes and, consequently, controlling rates of reactions. Recently, the concept of dimerization as a key molecular determinant in enabling recognition and interactions between the diverse set of proteins targeted by the family of cullin-RING E3 Ub ligases, has gained prominence. However, experimental evidence for the existence of dimeric substrate-receptor complexes has been largely missing. We believe that our structural and imaging studies of UV-DDB provide such experimental support.

The E3 ligase architecture derived from our crystal structure of the dimeric UV-DDB indicates that dimerization is a means of modulating intermolecular association parameters in cullin-RING E3 Ub ligase systems and is likely generalizable to other multi-component complexes with similar modular molecular architecture as UV-DDB. Homo-oligomerization is a fundamental step, allowing the individual components to be aligned in the context of the holo-complex, so that a multitude of reaction parameters can be spatially accommodated. Thus, a wide spectrum of functions can be regulated globally as biomolecular components dynamically assemble and disassemble along the NER and ubiquitin proteasome system pathways. Our study suggests that the rates of DNA binding and the high affinities for damaged DNA are a consequence of optimizing molecular associations in the holo-complex, an intrinsically basic mechanism for controlling substrate and protein–protein interactions, yet having profound effects on the overall efficiency of DNA repair.

Materials and Methods

Protein Expression and Purification. Native and SeMet-substituted proteins were expressed in Sf9 cells and purified as previously published (10) (*SI Appendix*).

Synthetic oligodeoxynucleotides, sequences, methods of purification, and analysis of the oligodeoxynucleotides described in this study are provided in *SI Appendix*.

Electron Microscopy. Uranyl acetate stained UV-DDB samples were prepared on grids and imaged on film in an FEI Tecnai T12 microscope operating at 120 kV and magnification of 30,000 ×. Micrographs were digitized with a Nikon Super CoolScan 9000 scanner and processed with the ImageJ software (50) to remove background variations, including uneven depth of stain and thickness of the carbon support film, and to estimate the particle size distributions (*SI Appendix*).

AFM Sample Preparation and Imaging. UV-DDB was incubated with nondamaged or UV-irradiated 517 bp PCR fragments and diluted 1:5- to 1:10-fold before deposition. All images were collected using a MultiMode V microscope (Veeco Instruments). Images were captured at a scan size of 1 μm × 1 μm, a scan rate of 2–4 Hz, a target amplitude of 0.3 V and a resolution of 512 × 512 pixels (additional experimental details and statistical analysis of AFM images are provided in *SI Appendix*).

Crystallography. Purified native UV-DDB mixed in a 1:3 molar ratio with AP24 oligodeoxynucleotides were prepared and immediately used in crystallization screening setups. Preliminary small crystals of the UV-DDB, with damaged DNA bound verified by gel electrophoresis, were obtained but diffracted weakly to 8 Å. Further optimization using additive screening protocols (36) generated single crystals in both monoclinic and orthorhombic lattices, and seeding eventually produced crystals that diffracted to 2.85–3.25 Å. A combination of a selenomethionine anomalous phasing approach in combination with partial model phasing yielded initial electron density maps clearly defining solvent and macromolecular boundaries. Solvent flattening and histogram matching improved the preliminary electron density maps (*SI Appendix*, Fig. S7), verifying the dimeric composition of the asymmetric unit. Iterative cycles of model building, rigid-body, molecular dynamics, simulated annealing, and grouped-B factor refinement monitoring R_{work} and R_{free} values throughout, improved model accuracy and map quality, permitting the DDB2 subunit to be traced and the AP24 oligodeoxynucleotide molecule built into its electron densities. The asymmetric unit is comprised of the full-length sequence of the human DDB1 (residues 1–1140), human DDB2 (residues 20–421), and 24-bp oligodeoxynucleotide duplex containing a central abasic site, refined to $R_{\mathrm{work}}/R_{\mathrm{free}}$ values of 0.22/0.24 (monoclinic) and 0.25/0.26 (orthorhombic). Data processing and refinement statistics are shown in (*SI Appendix*, Table S1), including details related to crystallization, data processing, structure determination, and refinement for both the monoclinic and orthorhombic forms of UV-DDB.

Dynamic Light Scattering (DLS) Analysis. A 20-μL UV-DDB sample was passed through a 0.2-μm filtering assembly into the sample chamber of a DynaPro (Wyatt Technology) molecular-sizing instrument equipped with a Plate Reader (Protein Solutions). Data collection and analysis utilized Dynamics 6.0 software package, as originally described (51). The particle sizes of UV-DDB, measured at six different protein concentrations were determined in the presence and absence of damaged oligodeoxynucleotide, AP24 (*SI Appendix*, Tables S2 and S4). Statistical analysis and additional details are included in *SI Appendix*.

Surface Plasmon Resonance (SPR). Immobilization of UV-DDB onto CM5 chip surfaces used standard EDC/NHS-mediated amine coupling procedures (52), using concentrations determined by the DLS results. The association and dissociation phases for the interaction of AP24 to UV-DDB were monitored on a BIAcore 3000 System (GE Healthcare), allowing the rates and binding affinities to be determined as a function of AP24 concentration. Data analysis (BIAevaluation software version 4.1) applied a Langmuir binding model to calculate the kinetics and affinity constants for the binding of AP24 to the monomeric UV-DDB (*SI Appendix*, Table S3).

Note Added in Proof. While this paper was under review a study appeared revealing the crystal structure of a single complex consisting of DDB1-DDB2-CUL4A-RBX1 (CRL4ADDB2) bound to a 12 bp DNA duplex containing a tetrahydrofuran (THF) lesion. While this structure differs significantly from the dimeric structure of UV-DDB bound to DNA in our study, their new structure helps explain the ubiquitin ligase substrate flexibility in damage recognition in chromatin (53).

ACKNOWLEDGMENTS. We thank Drs. John Rosenberg and Richard Wood for helping with the design of the AP oligodeoxynucleotides and Dr. Angela Gronenborn for early discussions. We appreciate the preliminary UV-DDB characterization done by Dr. Inja Byeon and the assistance of Dr. Jinwoo Ahn and Thomas Vu on early oligodeoxynucleotide purification. We thank Dr. Alexander Makhov for assisting with the electron microscopy and image analysis. We thank the staff members at the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) and the Southeast Regional Collaborative Access Team (SER-CAT), both at the Advanced Photon Source, Argonne National Laboratory, for access and technical assistance. Financial contributions toward the structural and biophysical methods applied in this work were provided by National Institutes of Health (NIH) Grants AI76121, GM82251, and GM66466 (to J.I.Y.). Financial support for this work were provided by NIH Grant ES019566 (to B.V.H.), the Commonwealth of Pennsylvania, Department of Health (to J.I.Y. and J.F.C.), and the Pittsburgh Foundation (to V.R.O.). GM/CA-CAT has been funded in whole or in part with federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). Use of the Advanced Photon Source is supported by the US Department of Energy, Office of Science, and Office of Basic Energy Sciences, under Contract W-31-109-Eng-38.

1. Friedberg EC, et al. (2005) *DNA Repair and Mutagenesis* (ASM Press, Washington, DC), 2nd Ed,, pp 1–1164.
2. Sugasawa K, et al. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* 2:223–232.
3. Wittschieben BO, Wood RD (2003) DDB complexities. *DNA Repair (Amst)* 2:1065–1069.
4. Moser J, et al. (2005) The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA Repair (Amst)* 4:571–582.
5. Wakasugi M, et al. (2002) DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *J Biol Chem* 277:1637–1640.
6. Dualan R, et al. (1995) Chromosomal localization and cDNA cloning of the genes (DDB1 and DDB2) for the p127 and p48 subunits of a human damage-specific DNA binding protein. *Genomics* 29:62–69.
7. Rapic-Otrin V, et al. (2003) True XP group E patients have a defective UV-damaged DNA binding protein complex and mutations in DDB2 which reveal the functional domains of its p48 product. *Hum Mol Genet* 12:1507–1522.
8. Itoh T, Mori T, Ohkubo H, Yamaizumi M (1999) A newly identified patient with clinical xeroderma pigmentosum phenotype has a non-sense mutation in the DDB2 gene and incomplete repair in (6-4) photoproducts. *J Invest Dermatol* 113:251–257.
9. Nichols AF, Ong P, Linn S (1996) Mutations specific to the xeroderma pigmentosum group E Ddb-phenotype. *J Biol Chem* 271:24317–24320.
10. Wittschieben BO, Iwai S, Wood RD (2005) DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *J Biol Chem* 280:39982–39989.
11. Scrima A, et al. (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* 135:1213–1223.
12. Fitch ME, Nakajima S, Yasui A, Ford JM (2003) In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product. *J Biol Chem* 278:46906–46910.
13. Takedachi A, Saijo M, Tanaka Ke (2010) DDB2 complex-mediated ubiquitylation around DNA damage is oppositely regulated by XPC and Ku and contributes to the recruitment of XPA. *Mol Cell Biol* 30:2708–2723.
14. Hannah J, Zhou P (2009) Regulation of DNA damage response pathways by the cullin-RING ubiquitin ligases. *DNA Repair (Amst)* 8:536–543.
15. Higa LA, et al. (2006) CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nat Cell Biol* 8:1277–1283.
16. Guerrero-Santoro J, et al. (2008) The cullin 4B-based UV-damaged DNA-binding protein ligase binds to UV-damaged chromatin and ubiquitinates histone H2A. *Cancer Res* 68:5014–5022.
17. Petroski MD, Deshaies RJ (2005) Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* 6:9–20.
18. Zheng N, et al. (2002) Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* 416:703–709.
19. Angers S, et al. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* 443:590–593.
20. Higa LA, et al. (2006) L2DTL/CDT2 interacts with the CUL4/DDB1 complex and PCNA and regulates CDT1 proteolysis in response to DNA damage. *Cell Cycle* 5:1675–1680.
21. Groisman R, et al. (2003) The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* 113:357–367.
22. Higa LA, Zhang H (2007) Stealing the spotlight: CUL4-DDB1 ubiquitin ligase docks WD40-repeat proteins to destroy. *Cell Div* 2:5.
23. Wang H, et al. (2006) Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol Cell* 22:383–394.
24. Kapetanaki MG, et al. (2006) The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. *Proc Natl Acad Sci USA* 103:2588–2593.
25. Luijsterburg MS, et al. (2007) Dynamic in vivo interaction of DDB2 E3 ubiquitin ligase with UV-damaged DNA is independent of damage-recognition protein XPC. *J Cell Sci* 120:2706–2716.
26. Sugasawa K, et al. (2005) UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* 121:387–400.
27. El-Mahdy MA, et al. (2006) Cullin 4A-mediated proteolysis of DDB2 protein at DNA damage sites regulates in vivo lesion recognition by XPC. *J Biol Chem* 281:13404–13411.
28. Rapic-Otrin V, McLenigan MP, Bisi DC, Gonzalez M, Levine AS (2002) Sequential binding of UV DNA damage binding factor and degradation of the p48 subunit as early events after UV irradiation. *Nucleic Acids Res* 30:2588–2598.
29. Nag A, Bondar T, Shiv S, Raychaudhuri P (2001) The xeroderma pigmentosum group E gene product DDB2 is a specific target of cullin 4A in mammalian cells. *Mol Cell Biol* 21:6738–6747.
30. Alekseev S, et al. (2008) Cellular concentrations of DDB2 regulate dynamic binding of DDB1 at UV-induced DNA damage. *Mol Cell Biol* 28:7402–7413.
31. Min JH, Pavletich NP (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449:570–575.
32. Maillard O, Camenisch U, Clement FC, Blagoev KB, Naegeli H (2007) DNA repair triggered by sensors of helical dynamics. *Trends Biochem Sci* 32:494–499.
33. Maillard O, Camenisch U, Blagoev KB, Naegeli H (2008) Versatile protection from mutagenic DNA lesions conferred by bipartite recognition in nucleotide excision repair. *Mutat Res* 658:271–286.
34. Camenisch U, et al. (2009) Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein. *EMBO J* 28:2387–2399.
35. Hwang BJ, Chu G (1993) Purification and characterization of a human protein that binds to damaged DNA. *Biochemistry* 32:1657–1666.
36. Yeh JI, Beale SI (2007) Calorimetric approaches to characterizing effects of additives on protein crystallization. *Cryst Growth Des* 7:2134–2139.
37. Li T, Chen X, Garbutt KC, Zhou P, Zheng N (2006) Structure of DDB1 in complex with a paramyxovirus V protein: Viral hijack of a propeller cluster in ubiquitin ligase. *Cell* 124:105–117.
38. Batty D, Rapic'-Otrin V, Levine AS, Wood RD (2000) Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. *J Mol Biol* 300:275–290.
39. Yang Y, Wang H, Erie DA (2003) Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods* 29:175–187.
40. Wang H, Yang Y, Erie DA (2007) Characterization of protein-protein interactions using atomic force microscopy. *Protein Interactions Biophysical Approaches for the Study of Complex Reversible Systems*, Protein Reviews, ed P Schuck (Springer Science+Business Media, LLC), Vol. 5, pp 39–78.
41. Kad NM, Wang H, Kennedy GG, Warshaw DM, Van Houten B (2010) Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single-molecule imaging of quantum-dot-labeled proteins. *Mol Cell* 37:702–713.
42. Wang H, Tessmer I, Croteau DL, Erie DA, Van Houten B (2008) Functional characterization and atomic force microscopy of a DNA repair protein conjugated to a quantum dot. *Nano Lett* 8:1631–1637.
43. Fujiwara Y, et al. (1999) Characterization of DNA recognition by the human UV-damaged DNA-binding protein. *J Biol Chem* 274:20027–20033.
44. Inoki T, et al. (2004) Human DDB2 splicing variants are dominant negative inhibitors of UV-damaged DNA repair. *Biochem Biophys Res Commun* 314:1036–1043.
45. Merlet J, Burger J, Gomes JE, Pintard L (2009) Regulation of cullin-RING E3 ubiquitin-ligases by neddylation and dimerization. *Cell Mol Life Sci* 66:1924–1938.
46. Rhodes D, Burley SK (2000) Protein–nucleic acid interactions. *Curr Opin Struct Biol* 10:75–77.
47. Nekrep N, et al. (2002) Mutations in a winged-leix DNA-binding motif causes atypical bare lymphocyte syndrome. *Nat Immunol* 3:1075–1081.
48. Lamber EP, et al. (2008) Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. *EMBO J* 27:2006–2017.
49. Bosu DR, Kipreos ET (2008) Cullin-RING ubiquitin ligases: Global regulation and activation cycles. *Cell Div* 3:7.
50. Abramoff MD, Magelhaes PJ, Ram SJ (2004) Image Processing with ImageJ. *Biophotonics Int* 11:36–42.
51. Moradian-Oldak J, Leung W, Fincham AG (1998) Temperature and pH-dependent supramolecular self-assembly of amelogenin molecules: A dynamic light-scattering analysis. *J Struct Biol* 122:320–327.
52. Murphy M, Jason-Moller L, Bruno J (2006) Using biacore to measure the binding kinetics of an antibody-antigen interaction. *Curr Protoc Protein Sci* 19:1–17.
53. Fischer ES, et al. (2011) The molecular basis of CRL4$^{DDB2/CSA}$ ubiquitin ligase architecture, targeting, and activation. *Cell* 147:1024–1039.

for ***PNAS***
Classification - Major: Biological Sciences
Classification - Minor: Biochemistry

Supplemental Information

Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair

Joanne I. Yeh[1,2,*], Arthur S. Levine[3,4], Shoucheng Du[1], Unmesh Chinte[1], Harshad Ghodke[5], Hong Wang[4,5], Haibin Shi[1], Ching L. Hsieh[3,4], James F. Conway[1], Bennett Van Houten[4,5], & Vesna Rapić-Otrin[3,4]

[1]Department of Structural Biology, [2]Department of Bioengineering, [3]Department of Microbiology and Molecular Genetics, [4]University of Pittsburgh Cancer Institute, and [5]Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh PA.

Address correspondence to*:* Joanne Yeh, Department of Structural Biology, University of Pittsburgh School of Medicine, 1036 BST3, 3501 5th Avenue, Pittsburgh, PA 15260, USA, Tel. (412) 648-9027, E-mail jiyeh@pitt.edu

Key Words

UV damage, ubiquitin-proteosome system, chromatin, X-ray crystallography

Materials and Methods

Protein expression and purification: native and SeMet labeled proteins

The protein components of the UV-DDB complex were co-expressed in Sf9 cells and purified through tandem-affinity chromatography with a His-tag on DDB1 and a FLAG-tag on DDB2, following the method that was developed previously (1, 2). Sf9 cells were co-infected with the viruses and incubated for 48 hours at 27 °C in SF900II media (Invitrogen), followed by centrifugation and snap freeze in liquid nitrogen. The frozen pellet was re-suspended in Sf9 lysis buffer 1 (500mM KCl, 50mM potassium phosphate pH 8.0, 10% glycerol, 0.5% NP-40), then rotated for 30 minutes at 4 °C for constant mixing. Centrifugation at 40,000 rpm in a Sorvall T-647.5 rotor for 45 minutes at 4 °C results in clear soluble lysate. The lysate was adjusted to 10mM imidazole and incubated overnight with Ni-NTA superflow resin (Qiagen). The protein bound Ni-resin was washed with 20 column volumes of Sf9 lysis buffer containing 10mM imidazole, then eluted with Sf9 lysis buffer containing 200mM imidazole. Peak fractions were collected and incubated overnight with anti-FLAG M2 affinity gel (Sigma). The protein bound anti-FLAG gel was washed with 20 columns of Sf9 lysis buffer, then eluted with Sf9 lysis buffer 2 (500mM KCl, 50mM potassium phosphate pH 8.0, 10% glycerol) containing 200µg/ml FLAG peptide (Sigma). Peak fractions were collected and concentrated in a 50KDa cut off concentrator (Pall Filtron) to remove the FLAG peptide. The yield of pure UV-DDB complex ranged from 1 to 2 mg/l of Sf9 cells. An EDTA-free protease inhibitor mix (Roche Applied Science) was added to all buffers in the protein purification except in Sf9 lysis buffer 2.

Seleno-L-methionine incorporation

Expression of seleno-L-methionine labeled DDB1 and DDB2 proteins was accomplished by adapting a bacterial expression protocol (3). Sf9 cells were co-infected with baculovirus encoding each UV-DDB subunit and incubated for 6 hours at 27 °C in SF900II media (Invitrogen). After centrifugation, cells were re-suspended in ESF-921 protein-free methionine-free cell culture medium (Expression Systems LLC) for 8 hours. 250mg/l final concentration of seleno-L-methionine (Acros Organics) was added to each liter of cells and further incubated for 36-40 hours at 27 °C. The same purification procedure was carried out as above except Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) at 1mM final concentration was added to all the buffers. The yield of pure seleno-L-methionine-UVDDB complex ranged from 200-250µg/l of Sf9 cells.

DNA Oligonucleotides: synthesis, purification, and annealing

Single stranded oligodeoxynucleotides were synthesized (Midland Certified Reagent Company Inc.; Midland, Texas) and further purified using anion-exchange chromatography (ProSphere P-WAX; 75x7.5 mm), eluting in a single peak during gradient purification with 25 mM Tris pH 8.5 with 0.02% sodium azide and 0-500mM NaCl. The molecular weights of deoxyoligonucleotides in the purified fractions were confirmed with MALDI-TOF-MS, combined, and concentrated. To form the double-stranded DNA, the purified complementary oligodeoxynucleotides were mixed in a 1:1 molar ratio at room temperature, heated to 90°C, and the samples gradually cooled to room temperature overnight to anneal. After annealing, the oligodeoxynucleotides were further purified using anion-exchange chromatography, as described for the single-strand DNA purification.

After purification, the double-strand DNA samples were buffer exchanged into 20mM Tris HCl, pH 7.5, 2mM $MgCl_2$, 1mM EDTA, 2mM TECP, 5% Glycerol, and 0.02% azide by passing through an Ultracel concentrator three times (Amicon). While several oligodeoxynucleotides containing a tetrahydrofuran (THF) moiety to mimic abasic lesions were synthesized and purified, the longest oligodeoxynucleotide duplex comprised of 24-base pairs (AP24) containing a central THF lesion site in complex to purified UV-DDB protein produced crystals with well-defined morphologies and better diffraction characteristics. Thus, the structural and functional results described in this study focuses on the UV-DDB-AP24 oligodeoxynucleotide complex. The sequences of the AP24 coding and complementary oligodeoxynucleotide strands are as follows:

AP24 coding strand: 5′-GTCAGCATCG(Abasic)CATCATACAGTCA-3′

Complementary: 5′-TGACTGTATGATGACGATGCTGAC-3′

In addition, for anomalous phasing and to verify positions of DNA strands, brominated AP24 oligodeoxynucleotide were synthesized by replacing all the deoxycytidine by 5-bromodeoxycytidine (^{Br}C) in the 24-mer DNA oligodeoxynucleotide (AP24Br). The brominated single strand oligodeoxynucleotide were purified, annealed, and then purified again, as described for the unhalogenated DNA. The sequences of AP24Br oligodeoxynucleotide are as follows:

AP24Br coding strand: 5′-GT^{Br}CAG^{Br}CAT^{Br}CG(_Abasic) ^{Br}CAT^{Br}CATA^{Br}CAGT^{Br}CA-3′

Complementary: 5′-TGA^{Br}CTGTATGATGA^{Br}CGATG^{Br}CTGA^{Br}C-3′

Electron Microscopy

3µl of sample were pipetted onto a freshly glow-discharge carbon-coated grid, blotted, washed on the surface of a 100µl drop of 1% uranyl acetate stain solution, blotted again and air-dried. Grids were imaged in an FEI T12 microscope operating at 120kV and magnification of 30,000x on film. Micrographs were digitized with a Nikon Super CoolScan 9000 scanner. The ImageJ software (4) was used for image processing of electron micrographs. Image processing to remove background variations, including uneven depth of stain and thickness of the carbon support film, allows the size distribution to be estimated (Figure S1, panels a-f). The dark areas are due to scatter from the uranyl acetate salt, surrounding white stain-excluding areas of proteins. Preliminary analysis to measure representative areas shows a well-defined peak at ~3500 Å^2, corresponding to a spherical particle of ~70Å diameter, consistent with a monomer of the UV-DDB complex. A shoulder is also apparent at twice the area (~7100 Å^2), corresponding to a dimer of the UV-DDB, representing ~5% of the particles in the absence of substrate DNA. However, this peak, at ~7100 Å^2, is enriched with the addition of damaged DNA (AP24) substrate and approaches 100% when an excess of damaged DNA is present (Fig S1, panels e,f).

AFM sample preparation and imaging

UV-DDB (50 nM concentration of DDB1-DDB2 heterodimer) was incubated with 25 nM undamaged or UV-irradiated 517 bp PCR fragments (25 nM) for 5 mins at 37 °C in a buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl 0.2 mM $MgCl_2$, and 0.2 mM EDTA. UV-irradiation was done at 20 J/m^2 (254 nm wavelength) on the 517 bp PCR product at 50 µg/ml concentration. The protein-DNA mixtures were diluted 1:5 to 1:10 fold before deposition (25 mM NaOAc, 10 mM $MgOAc_2$, 25 mM Hepes pH 7.5). All samples for AFM imaging were prepared by depositing samples onto a freshly cleaved mica (SPI Supply, West Chester, PA), followed by washing with MilliQ water and drying under a stream of nitrogen gas. All images were collected using a MultiMode V microscope (Bruker, Santa Barbara, CA.) using E scanners in tapping mode. Pointprobe® plus noncontact/tapping mode silicon probes (PPP-NCL, Agilent) with spring constants of ~50 N/m and resonance frequencies of ~190 kHz were used. Images were captured at a scan size of 1 µm × 1 µm, a scan rate of 2-4 Hz, a target amplitude of 0.3 V and a resolution of 512 × 512 pixels.

Volume analysis of AFM images

For AFM volume analysis, dimensions of proteins were measured using Image SXM software (5, 6). AFM volume of a particle was calculated as V= S × (H - B), where V is the

AFM volume, S is the area generated at the base of a protein using “density slice” function of the SXM software (5, 7), H is the average height, and B is the background height. The standard equation relating the AFM volume (V) of a globular protein and its molecular weight (MW) is: V (nm^3) = 0.3856 MW (kDa) - 1.913. This equation is based on AFM volumes of the following proteins and their various oligomeric states: Pot1 (65kDa), PcrA monomer (86.4 kDa), UvrA monomer (105 kDa), *Taq* MutS dimer (181 kDa), UvrA dimer (210 kDa), and *Taq* MutS tetramer (362 kDa).

Quantitative PCR assays

In order to estimate the number of photoproducts induced upon UV exposure, we performed a quantitative PCR assay using untreated or treated template at a concentration of 0.1 ng/ml in a 20 μL reaction. A 517 bp DNA fragment was amplified as described previously (8). Briefly, the cycling conditions are as follows: 75°C for 90 s; 94°C for 5 min; 94°C for 30 s, 57°C for 1 min, and 72°C for 1 min (15 or 16 cycles).

Estimation of the possibility two independent monomeric binding events at vicinal lesions

Calculation of lesion frequency was performed as described before (9). We found that on average each UV irradiated 517 bp fragment of DNA has 1.2 lesions (0.6 photoproducts per strand). Since UV damage is distributed according to a Poisson distribution ($P(k) = \lambda^{k}e^{-\lambda}/k!$), the percentage of molecules with 0, 1, 2, 3 and 4 or more lesions is approximately 35.37, 35.37, 17.68, 5.89 and 5.69 %, respectively (9).

For the 17.68% of molecules with 2 UV induced lesions, we can calculate the probability of finding 2 lesions in a 30 bp contiguous stretch (corresponding to the footprint of DDB2 on DNA) of DNA as P_2. UV irradiated 517 bp DNA can be modeled as consisting of 517 nucleotides with damaged dinucleotides corresponding to the 6-4PP and CPD lesions. Thus, a DNA fragment with 2 lesions can be treated as a group of 3 objects with the first object being 513 undamaged identical bases and the two lesions corresponding to the other two objects. Therefore, the total number of combinations of 2 lesions is given by *Total outcomes* = ${}^{515}C_2$ = 132355. The number of combinations of 2 lesions where they are within 30 nucleotides of each other (corresponding to 30 nucleotides which is roughly the footprint of DDB2 dimer on DNA) is calculated as *Favorable outcomes* = $_{\{i=1\}}\Sigma\ {}^{1}C_1{}^{(515-i)}C_1$ number of combinations of the two lesions so that the number of nucleotides between them is 0, 1, 2 and so on until 27. Thus for a DNA fragment containing two lesions, the probability of both of those lesions lying within the footprint of DDB2 dimer is P_2 = favorable outcomes/total outcomes = 0.10. This corresponds to a total of about 1.77% of all molecules with 2 lesions within the foot print of the DDB complex.

Similarly, a 517 bp fragment of DNA containing 3 lesions can be treated as 4 objects – 511 non damaged bases, and 3 lesions. We can calculate the probability of finding at least 2 lesions within the footprint of DDB for DNA molecules containing 3 lesions as $P_3 = 1-\underline{P}_3$, where $\underline{P}_3$ is the probability of the 3 lesions always being outside of the footprint of DDB2 dimer. $\underline{P}_3$ = favorable outcomes where the lesions do not lie within 30 bases of each other/total outcomes. The total number of combinations of 3 lesions is *Total outcomes* = ${}^{513}C_3$ = 22369536. The number of favorable outcomes = $^{\{513-28\}}{}_{\{i=1\}}\Sigma\ {}^{i}C_1{}^{(513-28-i)}C_1$.Thus $\underline{P}_3 = 0.85$, and $P_3 = 0.15$. Thus, about 0.88% of all molecules have 3 lesions within the foot print of the DDB complex.

We can similarly calculate the probability of finding at least 2 lesions within the footprint of DDB for DNA molecules containing 4 lesions or more however, this is a sufficiently small fraction of molecules amounting to just 1.5% for molecules with 4 lesions (with the percentage of molecules with more than 4 lesions being even smaller) of molecules that we can neglect it without significantly affecting our analysis. Based on these numbers, we expect approximately 2.5% of all molecules to contain multiple lesions within the footprint of the DDB complex.

UV-DDB/DNA crystallization

For crystallization, the UV-DDB complex was mixed with the purified AP24 duplex DNA in a molar ratio of 1:3 (UV-DDB:DNA) in 20mM Tris pH 7.5, 2mM $MgCl_2$, 1mM EDTA, 2mM TECP, 5% Glycerol, and 0.02% azide. The sample was concentrated to about 2.5 mg/ml (UV-DDB) using an Ultracel concentrator. Numerous crystallization screening trays were set up at 4°C and the most promising conditions were optimized using a hanging drop diffusion method with a volume ratio of 1:1 (μl) of protein solution to reservoir. For data collection, crystals were transferred into a solution containing a cryoprotectant, typically comprised of the crystallization solution augmented with 22% ethylene glycol, then flash cooled in liquid nitrogen.

Preliminary small crystals were grown from a stock containing native human UV-DDB protein that was incubated with small excess of AP24 DNA oligomer. To ascertain that AP24 was stoichiometrically bound to UV-DDB in the crystallization condition, small crystals were harvested, washed, dissolved, and confirmed by gel-electrophoresis. The early crystals exhibited multiple morphologies and diffracted weakly to 8Å. Further optimization using additive protocols (10) resulted in crystals with better defined crystal habit and morphology and were subsequently used for streak- and micro-seeding. After iterative cycles of optimization, single prismatic crystals exhibiting gradually improved X-ray diffraction characteristics were obtained. These crystals were used as seeds for further optimization using the matrix seeding method (11-13), identifying several new promising conditions. The final crystallization condition yielding single crystals used for data collection contained 0.25M potassium thiocyanate, 0.08M Bis-Tris propane pH 6.5, 28% w/v PEG 3350, and 2% 1,5-pentanediol. The UV-DDB complex crystallized in a monoclinic lattice, with a screw axis along the unique two-fold rotation axis (space group $P2_1$) with unit cell dimensions of a=76.736 Å, b=70.877 Å, c=191.448 Å, α=90.00°, β=99.68°, γ =90.00°. APS synchrotron diffraction data to 2.85 Å resolution were collected on the 24ID beamline.

For phasing, Se-methionine substituted UV-DDB (DDB1-SeMet and DDB2-SeMet) were co-crystallized with AP24 (1:3 of UV-DDB to DNA) from a solution containing 0.35M di-ammonium tartrate and 30 %(w/v) PEG 3350, and nucleated using seeds transferred by the streak seeding method. The SeMet-substituted UV-DDB/AP24 complex crystallized in an orthorhombic lattice (spacegroup $P2_12_12_1$), with similar unit cell parameters for two of the axis while the third unique axis doubled in length. The *ab-inito* selenomethionine SAD data was phased to 3.2 Å (Table 1, Table S1). with a solvent content of ~56% with one molecule in the asymmetric unit. The DDB1-DDB2 sequence has a total of 37 methionine residues. For anomalous phasing, SAD datasets were collected on selenomethionine substituted crystals and the datasets were merged for high redundancy. SeMet SAD data collections were done at SER-CAT 22ID and 22BM beamlines at APS. The SAD datasets were collected at peak energy wavelength, optimizing redundancy (Table S1). For all datasets, the crystals were translated throughout the diffraction experiments, after collection of a small wedge of data, to minimize effects of radiation damage.

Brominated DNA (AP24Br) was also synthesized for phasing, substituting cytosine with 5-bromo-dC, purified, annealed, and purified as described earlier. Purified UV-DDB was co-crystallized with AP24Br. Similar to the crystallization of UV-DDB-AP24, UV-DDB-AP24Br initial crystallization hits also only grew tiny crystals with poor morphology. Further additive screening with a condition containing 0.2M sodium fluoride, 0.1M Bis Tris propane pH 6.5, and

33% w/v PEG 3350 produced single diffracting crystals. The best crystals were obtained with the additive, 0.08M GSH (L-Glutathione reduced) and GSSG (L-Glutathione oxidized). Data sets were collected at the Argonne Photon Source (Chicago, Illinois), on the SER-CAT and GM/CA beamlines and at the Stanford Synchrotron Radiation Lightsource (Palo Alto, CA), on beamlines BL7-1 and BL9-1. Due to the limited resolutions of diffraction of the AP24Br datasets (nominally 3.5-3.6 Å), the Br-datasets were used only in the early phase of model building, primarily to verify the placement and orientations of the AP24 DNA strands in the early models.

X-ray data collection, structure determination, and model building

Two datasets collected from crystals grown using SeMet-substituted UV-DDB protein were merged during processing, improving overall completeness and redundancy (Table S1). All datasets were processed using the HKL2000 suite (14). For the anomalous data, the intensity measurements for the Friedel pairs were separated at the scaling stage in HKL2000. To resolve the enantiomorphic ambiguity (i.e., hand of the substructure atom configuration) of the SAD phases, the phases were combined with those calculated from a model. Partial model phasing using the monoclinic P21 UV-DDB-AP24 structure was initially done in Phaser (15). Phase combinations of the calculated phases together with the SeMet SAD data were done in Phenix (16, 17), producing well-defined electron density maps. In one UV-DDB subunit, a total of 27 selenium sites were found, with 21 sites having occupancy greater than 0.6. The initial FOM from the combined phasing was 0.76 as compared to 0.27 using the SAD data alone in Solve (18). Attempts to phase directly by molecular replacement using the human DDB1 and zebrafish DDB2 (pdb accession numbers 3EI2-3EI4) resulted in poor rotation and translation function values, with FOM of ~0.25 and poorly defined electron density maps with large regions of discontinuous densities and the AP24 oligodeoxynucleotide. The maps calculated using combined phasing methods showed clear electron density for the AP24 DNA molecule, located close to the DDB2 domain (Figure S7). These maps were used for model rebuilding in Coot (19). Once the relative orientations of the DDB1, DDB2, and AP24 DNA were determined, these were cross-validated using the SeMet anomalous dataset. This orthorhombic crystal form (pdb ID code 4E5Z) independently confirmed the NT-domain fold and cross-validated the structure determined in the monoclinic lattice (pdb ID code 4E54) (refinement statistics for both crystal structures are shown in Table 1).

Extensive regions of the UV-DDB were traced and built *de-novo*, including loop regions connecting the beta sheets in DDB1, the β-wings, and terminal regions of both DDB1 and DDB2. The NT-domain of DDB2, comprised of residues 20-100, was gradually built into experimental electron density maps as the earlier published structures were missing this NT region. The NT-region of DDB2 (residues 20-100) was iteratively extended, locally then globally refined, validated by iterative composite omit maps using the 2.85Å dataset. Iterative refinement cycles included residue-by-residue fitting followed by energy minimization and grouped/isotropic B-factor refinement over the entire complex. In addition to the NT-domain of DDB2, several regions were verified for model accuracy by generating simulated annealing omit maps in Phenix; these included the newly extended NT-helical α-paddle and β-wing regions, several loops at the DDB1-DDB2 interface, and the BPB domain of DDB1.

For AP24 DNA, density was seen for all 24 bases for both strands. Also, for the damaged DNA strand no extra density was seen in the difference maps around the tetrahydrofuran (THF) legion, signifying the absence of a nucleobase. During DNA fitting, the planarities of complimentary bases were restrained initially, with loosening of restraints as refinement progressed except for a 2 base-pair window around the abasic site, where the planarity is disrupted. As seen in earlier structures of UV-DDB-DNA complexes, DDB2-induced DNA kinking is also seen in our UV-DDB-AP24 complex structures. The UV-DDB-AP24 complex has been refined to Rwork/Rfree values of 0.22/0.24 (monoclinic) and 0.25/0.26 (orthorhombic), with refinement statistics shown in Table 1 & Table S1.

Dynamic Light Scattering (DLS) Analysis

The DLS analysis was performed on a DynaPro (Wyatt Technology) molecular-sizing instrument equipped with a Plate Reader (Protein Solutions). A 20 μL UV-DDB sample was passed through a 0.20 μm filtering assembly into the sample chamber of the DynaPro. Data collection and analysis utilized Dynamics 6.0 software package, as originally described (20). For analyses, samples of UV-DDB were prepared using buffer (10 mM HEPES pH 7.4 and 0.15 M NaCl) to adjust total protein concentrations immediately before light scattering measurements. Data were collected under identical experimental conditions before and after the addition of damaged DNA substrate (AP24). The same molar ratio of 3:1 AP24:UV-DDB as that used for crystallization was maintained. The particle sizes and molecular weights (MW) of UV-DDB in the presence and absence of damaged DNA (AP24), calculated from the DLS data collected at six different protein concentrations, are summarized in Table S3.

Surface Plasmon Resonance (SPR) analysis

All SPR experiments were performed by using a BIAcore 3000 biosensor and Sensor Chip CM5 (GE Healthcare) at 6 ºC including maintaining the sample holders at this temperature with a circulating water-bath. All of the AP24 samples were diluted into running buffer (10 mM HEPES, pH 7.5, 0.15 M NaCl), with sample concentrations from 1.4 nM to 1.0 μM. All measurements were performed in series, with one channel dedicated as a control with its surface generated as similar as possible to the sample surface channel (i.e., channel immobilized with UV-DDB). Immediately following injection of AP24 (i.e., damaged DNA substrate) through the UV-DDB immobilized sample channel, AP24 was injected in the reference channel and sample channel, at flow rates of 30 μL/min for 2.5 min. Background signals caused by refractive index changes and non-specific surface interactions were recorded in tandem. The sensorgram signals from the reference channel were subtracted from the sample channel (i.e., UV- DDB immobilized channel) to obtain the overall signal corresponding to the binding of the AP24 to UV-DDB. At the end of the association phase, the flow rate in the absence of the AP24 substrate was maintained to monitor dissociation for duration of 5 min. A 2M NaCl buffered solution regenerated the surface.

Data analyses were performed using the BIAevaluation software to model the binding of AP24 to UV-DDB in the monomeric and dimeric forms. Preliminary assessment evaluating the feasibility of the SPR approach to detecting changes in protein-substrate (AP24) binding parameters were conducted under steady-state conditions. Additional optimization of SPR experimental conditions and immobilization protocols led to two separate methods for the immobilization of the dimeric form of UV-DDB (further elaborated below). All measurements were replicated in triplicate and cross-validated when possible for comparison and cross-validation of several data sets. SPR datasets collected for the binding of AP24 to the monomeric form of UV-DDB were modeled by fitting the data to a Langmuir adsorption model. Preliminary SPR data at high UV-DDB concentrations suggested that the binding of AP24 to the dimeric form of UV-DDB exhibited interactions that are more complex. Two different protocols for the immobilization of the dimeric UV-DDB were used for the SPR analysis, generating two sets of data, each replicated in triplicate, for cross-validation. To verify that predominantly monomeric or dimeric forms of UV-DDB were immobilized at specific protein concentrations, DLS analyses were done over a range of concentrations as described in the preceding section, in the absence and presence of AP24. Accordingly, at 20 μg/mL predominantly monomeric or at 800 μg/mL dimeric UV-DDB forms were immobilized in separate channels on research-grade carboxymethyl (CM5) chip surfaces via standard EDC/NHS-mediated amine coupling procedures (21). Briefly, the carboxymethyl dextran surface was activated using freshly a prepared aqueous solution containing 0.2 M EDC and 0.05 M NHS. Once the carboxymethyl (CM5) chip surface was activated, injection of purified UV-DDB covalently linkage formed between accessible nucleophilic groups (primary and secondary amines) of UV-DDB to the freshly formed succinamide groups on the chip surface. The same protocol was repeated by injecting purified and concentrated UV-DDB at 800 µg/mL, immobilizing the dimeric form in a separate channel. Unbound protein was removed and unreacted surface sites on the CM5 chip were capped by injecting 1M ethanolamine-HCl (pH 8.5). A series of AP24 association and dissociation experiments to monitor the interactions of AP24 to monomeric and dimeric UV-DDB. Contributions from non-specific interactions to the measured SPR signals were subtracted dynamically. Consistent experimental parameters and conditions were maintained between the reference and sample channels.

An alternative method to generate the dimeric form in-situ on the SPR surface was used for determining parameters for the binding of AP24 to monomeric and dimeric forms of UV-DDB in series. Using this approach, data from the association phase for the binding of AP24 to the monomeric form of UV-DDB was measured, and complete dissociation of AP24 using 2M NaCl to regenerate unbound (apo) monomers of UV-DDB. To form the dimeric form of UV-DDB in-situ, highly concentrated UV-DDB was injected at 3mM, a concentration several times greater than the estimated affinity of dimerization, generating dimeric UV-DDB through innate oligomerization tendencies. The dimerization was allowed to proceed for 150 s and the stability of the subsequent dimeric UV-DDB formed in-situ on the SPR chip surface was assessed under low flow rates while monitoring the sensorgram for dissociation. The submicromolar dissociation constant (KD) estimated for the monomer-dimer UV-DDB association is supported by the stable baseline that indicated that the dimeric form of UV-DDB was stably maintained. The binding of AP24 to dimeric UV-DDB was measured by injecting UV-DDB-ligand at 3.0, 1.0, 0.33, 0.11, 0.037, and 0.012 mM pre-incubated with AP24 (at the ratio of 1:3) through the UV-DDB affixed channel, at a flow rate of 30 µl/min. The association was allowed to proceed for 150 s, and dissociations of the complex monitored for 300 s and the data used to calculate the rates and affinities of AP24 binding (Table S3).

The values obtained for the dimeric form of UV-DDB generated in-situ to those calculated from the sets of data collected from immobilizing the dimeric UV-DDB at a high concentration (800 µg/mL) are comparable. Accounting for the heterologous population in the rates and affinity calculations gave generally comparable values and trends. The contributions from the mixed monomer-dimer population were accounted for by using the state distributions obtained from DLS analysis**.** Using BIAevaluation software (version 4.1) modeling according to heterogeneous ligand-parallel reaction to fit the data resulted in low residuals, indicating statistically valid fit. The kinetics and affinity constants estimated for AP24 in the monomeric and dimeric states of UV-DDB are tabulated in Table S3. It should be noted that additional experimental analysis are required to fully validate these preliminary results.

Expanded Discussion

Dimer Interface Adjacent to Damaged DNA Binding Site in DDB2

The primary DNA binding surface is located at the narrow end of the β-propeller, opposite the DDB1 interaction region of DDB2 (Figure 2). The bound DNA molecule spans the surface of DDB2, offset from the center of the 7-bladed β-propeller (Figure S2A). This offset can now be explained in terms of the spatial constraints arising from the dimerization interface. To accommodate the steric requirements allowing for dimerization of DDB2 along with binding of damaged DNA substrate (Figure S2B), the neighbouring intermolecular contact surfaces are located diametrically across a molecular face of the β-propeller domain of DDB2, leading to the offset (Figure S2C). The location of the dimer interface on the same molecular surface as the DNA binding site allows for cooperativity and coordination of oligonucleotide binding and dimerization.

Using the nomenclature introduced previously (22), the damage site is denoted as "D_{+1}" with the nucleotides 3' of the damaged site denoted as D_{+n} and 5' as D_{-n}, where the number refers to the location of the nucleotide relative to the damage site at D_{+1}. Overall, localized deformation at the abasic site forms a gap of ~15 Å and unwinds the DNA by ~18 degrees, radiating from the lesion (Figure S3). This perturbs the stacking of the bases immediately opposite the abasic site and the adjacent upstream nucleotide but limited in range to within two-nucleotides of the lesion.

DNA damage is detected by insertion of the β-loop extending from strand-5 of the β-propeller domain of DDB2 into a gap in the duplex formed by CPD, 6-4PP, or abasic lesions, which all produce similar nucleotide deformations in the duplex (22). This insertion loop is comprised of three highly conserved residues -- Phe334, Gln335, and His336 – with each residue forming specific contacts to the bases opposite the lesion (Figure S3C). The minor groove at the lesion site is occupied by the abasic ribose moiety, positioned through the extra-helical flipping of the adjacent nucleobase (dC12), which appears to drive the distortions at the lesion as no significant interactions to DDB2 are found. The β-loop of strand 3 of DDB2 defines the perimeter of the site occupied by the nucleobase (dC12) immediately upstream of the lesion.

This base is an extra-helical configuration, which enlarges the gap initially produced by the missing nucleobase (THF11). The resulting gap is filled by the imidazolium ring of DDB2 His336 on the 5' side and by Gln335 on the 3' side, limiting the lesion-induced distortion from being propagated. The extrahelical conformation of dC12 is maintained by a combination of H-bonds, hydrophobic, and π-stacking interactions between the cytosine base and DDB2 residues, particularly Ile200, Asn201, and Trp203 (Figure S3). The plane of the aromatic ring of Trp203 forms the shallow base-binding plateau with Met177 and Ile200 bordering the site, which can accommodate both pyrimidine and purine nucleobases. Additional sequence independent stabilization at the flipped nucleotide next to the lesion is afforded through electrostatic contacts between the phosphoribose-backbone of the DNA to Arg and Lys residues of DDB2.

Immediately 5' of the flipped dC12, dA13 remains in the duplex, stacking over the carboxamide moiety of Gln335, which replaces the nucleobase of dC12. Gln335 also forms H-bonds to His336, stacking the imidazole ring over the purine ring of dG10, so that the canonical base-stacking interactions 5' at the D_{+2} and 3' at the D_{-1} regions are maintained in the damaged strand. The positively charged guanidinium groups of Arg112 bridge the two strands at D_{-1}, D_{-2}, maintaining inter-strand base-pairing contacts on the 5' side of the lesion (Figure S3C).

Dimeric UV-DDB Binds Damaged DNA with Higher Affinity than in the Monomeric State

The DLS experiments clearly show that the binding of damaged DNA substrate (AP24) induces UV-DDB dimerization, even at the lowest protein concentration analyzed, 100 μg/mL concentration (which is 25X more dilute than the stock concentration used for crystallization). As the plots show (Figure S8), under the solution conditions for the DLS analysis (identical to SPR experimental conditions), in the absence of damaged DNA, a maximum of ~15% of UV-DDB was dimeric at the highest protein concentration studied, 1.25 mg/mL. In contrast, in the presence of AP24, only dimeric UV-DDB was found at 1.25 mg/mL and less than 5% were monomeric even at the most dilute protein concentration analyzed, 0.1 mg/mL. Taken together with the kinetics and binding affinities determined by SPR, both performed under identical solution conditions, the binding of damaged DNA induces dimerization of UV-DDB, which further enhances the tightness of binding, ~doubling the rates of association and reducing the dissociation rates by 50%. These results are consistent with the findings from X-ray crystallography, EM, and AFM, supporting the role that dimerization plays in modulating binding affinity, helping to assure fidelity of damage detection. The actual degree to which kinetic and affinity parameters are altered by dimerization need to be examined using UV-damaged DNA substrates – these analyses reflect the consistent trend found between solution and structural findings involving AP deoxyoligonucleotides.

Functional implications of the dimeric state of UV-DDB

Our analysis of dimeric UV-DDB provides conformational insights on potential associations between a substrate receptor and substrate, suggestive of dimerization functioning in regulating overall activity of the E3 ligase complex. Based on the additional extensive contacts formed by the dimeric UV-DDB, this new conformation can modulate substrate affinities on multiple levels, serving to allosterically regulate the substrate-receptor complex. Similar to the transcriptional factor, (Ets-1)2–S-EBS complex (Ets-1) (23), the dimeric state of UV-DDB is stabilized by DNA binding, inducing conformational ordering in the NT domain and further generating additional DNA-binding surfaces. In the Ets-1, dimerization is key to relieving auto-inhibition, whereby ternary complex formation of Ets-1 with itself initiates the conformational transition to a high-DNA-binding affinity state (23). In the UV-DDB, damaged DNA binding triggers the NT domain of DDB2 to adopt an α-helical paddle motif, presenting new DNA binding sites. Additional DNA-contact sites are generated once the β-wing conformation is induced, concomitant with dimerization (Figure 2).

The β-wing loop of DDB2 is located at the interface of two DNA molecules in the dimeric UV-DDB, so that multiple unique sites of DNA contacts are found in the dimeric UV-DDB, interactions that are absent in the monomeric state. These interactions are centered at the NT-α-helical paddle and β-wing regions of DDB2 and when analyzed together resemble a "winged helix" motif that has been found in numerous DNA-binding proteins (24).

Winged helix proteins share a related DNA-binding motif, combining beta loops ('wings'), alpha helices and beta-sheets to modulate DNA contacts. In these proteins, the helices form various levels of DNA contacts, from sequence-specific major groove insertions to base-independent electrostatically-mediated contacts to the deoxyribose-phosphate backbone of the DNA. Analogous to those found in other winged-helix DNA binding proteins, the β-wings in UV-DDB form direct contacts to the backbone atoms of the DNA (Figure 2). Additionally, the conformation and the apparent function of the β-wings of DDB2 in dimeric UV-DDB resemble those shown in the transcription factors, Ets-1, and the tripartite factor X, RFX, by linking and modulating nucleotide binding affinities with dimerization (23-25). In these winged-helix proteins, patches of hydrophobic residues are exposed causing conformational changes to present new protein-protein interaction surfaces and inducing dimerization as a function of nucleotide binding.

The winged-helix DNA-binding motif in the dimeric UV-DDB is obtained by combining the α-paddle with the β-wing motifs in the dimeric DDB2 (23-26). The orientation of the NT-α-paddle domain of DDB2 relative to the DNA is similar to those formed between the helical segments of Ets-1 to the DNA (PDB accession code 2nny). Both proteins contain clusters of acidic and basic amino acids that form charge and dipolar interactions on the helical surfaces facing the DNA backbone.

Using our structural information to align the NT domain of DDB2 to the H-T-H winged-type DNA-binding motif in Ets-1 and RFX results in a high conservation of sequence homology (>70% identity) and in the 3D-fold. When only the primary sequence of monomeric DDB2 was used in various prediction programs to identify protein function, including ProFunc (26), only the WD40 domain was identified whereas the programs were unsuccessful in recognizing other functional motifs in DDB2. However, defining specific domain boundaries using the dimeric UV-DDB structure and combining the sequences of these as composite sites, allowed the successful identification of DNA binding sites in DDB2. Thus, these structural analyses support the findings of the crystal structure, EM, and AFM studies, verifying that non-contiguous regions of DDB2 can adopt DNA-binding motifs when assembled in the dimeric state. Consequently, in DDB2, the NT-helical and β-wing domains contain conserved DNA binding sequences that allow DDB2 to adopt high-affinity DNA binding motifs. Segregating these topologies so that the high-affinity winged-helix motif is formed only upon dimerization modulates DNA affinities, enabling apparently incompatible kinetic and thermodynamic oligonucleotide-binding characteristics to be encoded within a single protein.

Supplementary Tables

Table S1: Data Statistics for SeMet UV-DDB

	UV-DDB-AP24 SeMet, Dimer Dataset 1
Data collection	
X-ray source	23ID-B, APS
Bravais Lattice	Orthorhombic
Cell Dimensions	
a, *b*, *c* (Å)	72.42, 76.50, 389.74
α, β, γ (°)	90, 90, 90
Resolution (Å)	3.22
# Unique reflections	30981
R_{merge} (3.2-3.31 Å)*	0.117 (0.358)
$I/\sigma I$ (3.2-3.31 Å)	10.9 (3.1)
Completeness (%) (3.2-3.31 Å)	93.8 (71.3)
Redundancy (3.2-3.31 Å)	5.6 (3.6)

*Values in parenthesis are for the highest resolution shell (3.2-3.31 Å)

Table S2. Dynamic Light Scattering: oligomeric states of UV-DDB in the presence and absence of damaged DNA (AP24)

Sample	**Molecular Weight of UV-DDB** (kDa)	**% Monomer** 200 kDa	**% Dimer** 400 kDa
No Damaged DNA (AP24) **UV-DDB** **[0.1 mg/mL]**	201	>98.5	<1.5
		Weighted <MW> kDa 205	
No Damaged DNA (AP24) **UV-DDB** **[0.4 mg/mL]**	209	~95	~5
		Weighted <MW> kDa 221.3	
No Damaged DNA (AP24) **UV-DDB** **[1.25 mg/mL]**	261	~85	~15
		Weighted <MW> kDa 245.0	
+ Damaged DNA (1:3 AP24*) **UV-DDB-AP24** **[0.1 mg/mL]**	422	<1	>99
		Weighted <MW> kDa 397.5	
+ Damaged DNA (1:3 AP24*) **UV-DDB-AP24** **[1.25 mg/mL]**	504	0	100
		Weighted <MW> kDa 400.0	

Table S3. Surface Plasmon Resonance: relative rates and affinities of damaged DNA (AP24) binding for monomeric and dimeric UV-DDB

Oligomeric State of UV-DDB	**Parameters**					
	k_a (1/Ms)	**k_d (1/s)**	**K_A (1/M)**	**K_D (M)**	**Chi^2**	**ΔK_D (Fold Increase ↑) Dimer vs Mono**
SPR Data Set 1						
AP24 to UV-DDB monomer	2.38×10^5	5.97×10^{-3}	3.98×10^7	2.5×10^{-8}	0.82	**4X (3.52)**
AP24 to UV-DDB dimer	4.52×10^5	3.21×10^{-3}	1.41×10^8	7.10×10^{-9}	1.01	
SPR Data Set 2						
AP24 to UV-DDB monomer	2.16×10^5	6.24×10^{-3}	3.46×10^7	2.88×10^{-8}	1.19	**4X (3.98)**
AP24 to UV-DDB dimer	4.33×10^5	3.13×10^{-3}	1.38×10^8	7.22×10^{-9}	1.10	
Summary Monomer Data Sets 1&2						
AP24 to UV-DDB monomer	2.38×10^5	5.97×10^{-3}	3.98×10^7	2.5×10^{-8}	0.82	
AP24 to UV-DDB monomer	2.16×10^5	6.24×10^{-3}	3.46×10^7	2.88×10^{-8}	1.19	
Monomer <Average>	*2.27×10^5*	*6.11×10^{-3}*	*3.72×10^7*	*2.69×10^{-8}*	--	
Summary Dimer Data Sets 1&2						
AP24 to UV-DDB dimer	4.52×10^5	3.21×10^{-3}	1.41×10^8	7.10×10^{-9}	1.01	
AP24 to UV-DDB dimer	4.33×10^5	3.13×10^{-3}	1.38×10^8	7.22×10^{-9}	1.10	
Dimer <Average>	*4.43×10^5*	*3.17×10^{-3}*	*1.40×10^8*	*7.16×10^{-9}*	-	
Overall Δ in Kinetics & Affinities of Damaged DNA (AP24) Binding Parameters in Dimeric Relative to Monomeric States of UV-DDB	***2X***	***½-1X***	***2-4X***	***¼-½X***	**--**	***Association rate 2X faster Dissociation rate 2X slower Overall ~4X Higher Affinity in dimeric state (relative to monomeric)***

** Molar ratio of UV-DDB to AP24 is 1:3, maintaining the stoichiometric ratios used for crystallization and EM analysis*

Table S4. Summary of UV-DDB molecular parameters

[UV-DDB] & Method	Radius (nm)	Area (nm^2)	Volume (nm^3)	Molecular Weight (kDa)	Oligomeric State(s)
No DNA or Bound to Undamaged DNA					
DLS: No DNA					
UV-DDB [UV-DDB]=0.5 µM	3.3			188	Monomer
UV-DDB [UV-DDB]=2.2 µM	3.7			202	Monomer
EM: No DNA					
UV-DDB [UV-DDB]=0.5 µM	3.5	35			Monomer
AFM: +Nondamaged DNA					
UV-DDB 25 nM undamaged DNA; [UV-DDB]=50 nM			~81 ± 10	184 ± 23	Monomer
Bound to Damaged DNA					
DLS: +AP24 Abasic DNA					
UV-DDB-AP24 1:3 AP24; [UV-DDB]=0.5 µM	6.7			403	Dimer
UV-DDB-AP24 1:3 AP24; [UV-DDB]=2.2 µM)	7.0			433	Dimer
EM: +AP24 Abasic DNA					
UV-DDB-AP24 1:3 AP24; [UV-DDB]=0.5 µM	~5	71			Dimer
AFM: +UV-irradiated 517 bp PCR DNA					
UV-DDB-AP24 1:2.5 UV-DNA; [UV-DDB]=50 nM			~133	349.9	Dimer+ 1 Duplex DNA
			~139	365.4	Dimer+ 2 Duplex DNA

Table S5: XPE Mutations

XP-E Patient	Mutation	Biochemical DDB2-DDB1	Biochemical DDB2-UV-DNA	References
XP2RO	R273H	affected	affected	(2, 27, 28)
XP25PV	D307Y	affected	affected	(2)
GM01389	L350P delN349	affected	affected	(2) (29)
XP82TO	K244E		affected	(2, 27, 28)

Supplemental References

1. Wittschieben BO, Iwai S, & Wood RD (2005) DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *J Biol Chem* 280(48):39982-39989.
2. Rapic-Otrin V*, et al.* (2003) True XP group E patients have a defective UV-damaged DNA binding protein complex and mutations in DDB2 which reveal the functional domains of its p48 product. *Hum Mol Genet* 12(13):1507-1522.
3. Cronin CN, Lim KB, & Rogers J (2007) Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci* 16:9.
4. Abramoff MD, Magelhaes PJ, & Ram SJ (2004) Image Processing with ImageJ. *Biophotonics International* 11(7):36-42.
5. Ratcliff GC & Erie DA (2001) A novel single-molecule study to determine protein--protein association constants. *J Am Chem Soc* 123(24):5632-5635.
6. Wang H, Yang Y, & Erie DA (2007) Characterization of protein-protein interactions using atomic force microscopy *Protein Interactions Biophysical approaches for the study of complex reversible systems,* ProteinReviews, ed Schuck P (Springer Science+Business Media, LLC), Vol 5, pp 39-78.
7. Wang QE*, et al.* (2007) Ubiquitylation-independent degradation of Xeroderma pigmentosum group C protein is required for efficient nucleotide excision repair. *Nucleic Acids Res* 35(16):5338-5350.
8. Wang H, Tessmer I, Croteau DL, Erie DA, & Van Houten B (2008) Functional characterization and atomic force microscopy of a DNA repair protein conjugated to a quantum dot. *Nano Lett* 8(6):1631-1637.
9. Ayala-Torres S, Chen Y, Svoboda T, Rosenblatt J, & Van Houten B (2000) Analysis of gene-specific DNA damage and repair using quantitative polymerase chain reaction. *Methods* 22(2):135-147.
10. Yeh JI, Beale SI (2007) Calorimetric Approaches to Characterizing Effects of Additives on Protein Crystallization. *Crystal Growth and Design* 7(11):2134-2139.
11. D'Arcy A, Villard F, & Marsh M (2007) An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr* 63(Pt 4):550-554.
12. Ireton GC & Stoddard BL (2004) Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr* 60(Pt 3):601-605.
13. Walter TS*, et al.* (2008) Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64(Pt 1):14-18.

14. Otwinowski Z, Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology: Macromolecular Crystallography, part A* 276:307-326.
15. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
16. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65(Pt 6):582-601.
17. Terwilliger TC*, et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1):61-69.
18. Terwilliger TC & Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55(Pt 4):849-861.
19. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
20. Moradian-Oldak J, Leung W, & Fincham AG (1998) Temperature and pH-dependent supramolecular self-assembly of amelogenin molecules: a dynamic light-scattering analysis. *J Struct Biol* 122(3):320-327.
21. Johnsson B, Lofas S, & Lindquist G (1991) Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Anal Biochem* 198(2):268-277.
22. Scrima A*, et al.* (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* 135(7):1213-1223.
23. Lamber EP*, et al.* (2008) Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. *EMBO* 27:2006-2017.
24. Rhodes D & Burley SK (2000) Protein–nucleic acid interactions. *Current Opinion in Structural Biology* 10:75–77.
25. Nekrep N*, et al.* (2002) Mutation in a winged-helix DNA-binding motif causes atypical bare lymphocyte syndrome. *Nat Immunol* 3(11):1075-1081.
26. Laskowski RA, Watson JD, & Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 33:W89-W93.
27. Nichols AF, Ong P, & Linn S (1996) Mutations specific to the xeroderma pigmentosum group E Ddb- phenotype. *J Biol Chem* 271(40):24317-24320.
28. Shiyanov P*, et al.* (1999) The naturally occurring mutants of DDB are impaired in stimulating nuclear import of the p125 subunit and E2F1-activated transcription. *Mol Cell Biol* 19(7):4935-4943.

29. Nichols AF, *et al.* (2000) Human damage-specific DNA-binding protein p48. Characterization of XPE mutations and regulation following UV irradiation. *J Biol Chem* 275(28):21422-21428.

Supplementary Figure Legends

Figure S1. Evidence of dimerization of UV-DDB from particle size analysis from negative stain electron micrographs.

A representative area of the UV-DDB sample without DNA is shown (a) raw; (b) band-pass filtered to preserve features from 5-18 nm in dimension; (c) after application of a local “rolling ball” filter (radius 9 nm) to further remove variations in the background gradient; and (d) thresholded to yield countable areas (white). This procedure was applied to all micrographs and results were combined into a total histogram for each sample. (e) A histogram of particle areas estimated from a single negative stain micrograph shows a peak at ~36 nm^2, corresponding to a circle of diameter ~7 nm that is consistent with a monomer of the UV-DDB1-DDB2 complex, and a shoulder at ~72 nm^2 that represents a minor population of dimers. Examples of monomer-sized areas are indicated with arrowheads, and dimers with double-arrowheads, in (d). The ImageJ software (4) was used for image processing. (f) A gallery of individual particles from different techniques of electron microscopy, as indicated. Some negative stain images are as large as monomers of UV-DDB (“1") while others are twice (“2") as large.

Figure S2: Electrostatic potential surfaces of dimeric UV-DDB2 in a complex with AP24 DNA

Electrostatic potential surfaces generated in PyMol, with positive potentials displayed in blue, negative in Electrostatic potential surfaces generated in PyMol, with positive potentials displayed in blue, negative in red, and uncharged in white. (A) The BPA domain of DDB1 displays a large hydrophobic region (white surfaces, outlined in a dotted yellow line) on the surfaces facing the β−propeller domain of DDB2. In contrast, the BPC domain of DDB1 presents strong negative charge clusters (surfaces in red, outlined in a purple dotted line) facing the NT α−paddle domain of DDB2. (B) DDB2 displays strong positive charge clusters (in blue) facing DDB1's BPC domain (outlined in a green dotted line) and at the DNA binding sites. These electrostatically neutralize the negative phospho-sugar backbone of the DNA, augmenting the DDB2-DNA binding interactions and contributing to the overall high-binding affinity for damaged DNA. (C) The DDB1-DDB2 molecular interface displays both aliphatic and charge characteristics, with the distribution of charged (in blue & red) and hydrophobic (in white) residues aligned to be complementary and neutralize the overall electrostatic potential upon complex formation.

Figure S3: Schematic representations of the multiple sites of interactions found at the abasic lesion between the dimeric DDB2 and the AP24 oligodeoxynucleotide .

(A) The dimeric DDB2, depicted by yellow ribbon representation, with an AP24 oligodeoxynucleotide mer bound at each site in the monomers. The NT region of DDB2 folds into a helical bundle ("α-paddle") that forms extensive interactions with a neighboring DNA molecule, depicted by stick-and-ball representation; the phospho-ribose backbone is colored in orange-red for the damaged strand and in blue for the undamaged strand. Defined sites of interactions between DDB2 and the damaged DNA can modulate binding affinities and optimize positioning of the damaged and undamaged DNA strands. Four distinct groups of interactions are found at the primary DNA binding site in DDB2 and mapped to the DDB2 structure, with a sphere representing the position of the DDB2 residue. At the lesion site on the damaged strand (D+1), residues at the β-loop are represented by an orange sphere; residues that stabilize the THF11 and dC12 (D+1) in their extrahelical configuration are denoted by yellow spheres; residues forming predominantly electrostatic interactions with the phosphodeoxyribose backbone of the DNA within a two basepair window upstream and downstream of the lesion sites are denoted by violet (D-1,D-2) and green (D+3,D+4) spheres. (B**)** The complete 24 base pair duplex containing an apurinic lesion mimicked at position 11 (THF11), with the same color scheme representing the four groups of DDB2-DNA interactions with the damaged DNA strand. Additionally, DDB2 residues that form interactions with the undamaged DNA strand are represented by tan rectangles, with the dimer junction denoted by Asn360, colored in red. **(**C) Additional DDB2-DNA interactions are found in the dimeric DDB2; the residues from one monomer subunit are denoted by ovals while residues from the other subunit are denoted as rectangles, using the same color scheme as in Figures 2A and 2B. DDB2 interactions with the undamaged strand are predominantly electrostatic in nature. A large cluster of lysines and arginines is found opposite the lesion site and bridges between the two strands. These charge clusters may denote poly-ubiquitination sites, especially as DDB2-DNA contacts are likely to be disrupted with a cascade effect, resulting in cooperative dissociation, based on the alignment of these charge residues.

Figure S4: Evidence for UV-DDB monomer to dimer transition when bound to damaged DNA as revealed by volume analysis using AFM.

A. Representative surface plot of UV-DDB (50 nM) in the presence of undamaged 517 bp PCR fragments (25 nM) showing free UV-DDB (white arrow) and bound to the undamaged DNA (yellow arrow). The image is at 500 nm by 500 nm and 3 nm height scale. **B.** Volume analysis of UV-DDB bound to undamaged DNA. Inset: Gaussian fits to the data in the range 0 – 175 nm^3. Data points outside this range were ignored to obtain the Gaussian fit. From the fit, the peak positions correspond to 81 and 133 nm^3 with an R^2= 0.9 consistent with the size of the heterodimer (DDB1-DDB2) and dimer of heterodimers of UV-DDB (DDB1-DDB2)$_2$. **C.** Summary of fraction of UV-DDB bound to DNA for undamaged (gray; 37%, N=163 and UV-irradiated (black; 96%, N=435 DNA. Fractions were calculated a (number of UV-DDB protein-DNA complexes)/ (number of UV-DDB protein-DNA complexes) + (number of free UV-DDB protein molecules on the mica). **D.** Overlay of percentage histograms for AFM derived volumes of UV-DDB molecules calculated on undamaged (dashed line) and UV-irradiated DNA (solid line). These data were generated by normalizing the count in each bin with respect to the total number of counts and converting to a percentage for each bin (undamaged DNA or UV-irradiated DNA). **E.** Histogram of the differences in the percentage histograms for UV-DDB bound to UV-irradiated DNA with respect to the undamaged DNA calculated bin wise.

Figure S5: Calibration curve for Nanoscope V.

The AFM volume of a protein (V) is a linear function of its molecular weight (MW). Calibration curve for Nanoscope V was constructed by performing AFM volume analysis of the following proteins: Pot1 (65 kDa), PcrA (86 kDa), UvrA monomer (105 kDa), Taq MutS dimer (181 kDa), UvrA dimer (210 kDa) and Taq MutS tetramer (362 kDa). The equation for the AFM volume is : V (nm^3) = 0.3856 MW (kDa) – 1.913 with R^2 = 0.9886. Based on this equation, the AFM volume of DDB1-DDB2 heterodimer (MW = 175 kDa) is 66 nm^3 (dotted arrows) and that of a dimer of the DDB1-DDB2 heterodimer (MW = 350 kDa) is 133 nm^3 (dashed arrows).

Figure S6: XP-E mutations mapped on to the dimeric human UV-DDB structure.

(A,B). Two views showing the locations of common DDB2 mutations found in XP-E patients *(also see Table S5)*. These residues (represented by atomic Van der Waals surfaces) form multiple, bridging contacts to DDB1 (Leu350, Arg273) or DNA (Lys244, Asp307). Disruption of these interactions presumably lead to more global destabilization in UV-DDB, as these residues appear to maintain multiple intermolecular associations in the holo-complex. (C,D). In the dimeric UV-DDB, the deleterious effects of two mutations in particular, Leu350 and Asn307, are more apparent, as these additionally modulate interactions to the NT α-paddle domain of DDB2. Thus, Leu350 and Asp307 of DDB2 form key intermolecular associations, at the junctions to both DDB1 and to the damaged DNA.

Figure S7: Representative electron density map of the dimeric UV-DDB-AP24 complex, **using combined SeMet anomalous phasing approaches combined with partial model phasing methods.**

Representative electron density map, using a combined SeMet anomalous phasing approach along with partial model phasing methods. The initial 2Fo-Fc map, contoured at 1.8σ, clearly showing the AP24 oligodeoxynucleotide bound at one surface of the β−propeller of DDB2.

Figure S8: Dynamic Light Scattering (DLS) analysis showing the particle sizes of UV-DDB and composition (% mass), determined at various protein complex concentrations and in the absence & presence of damaged DNA (AP24).

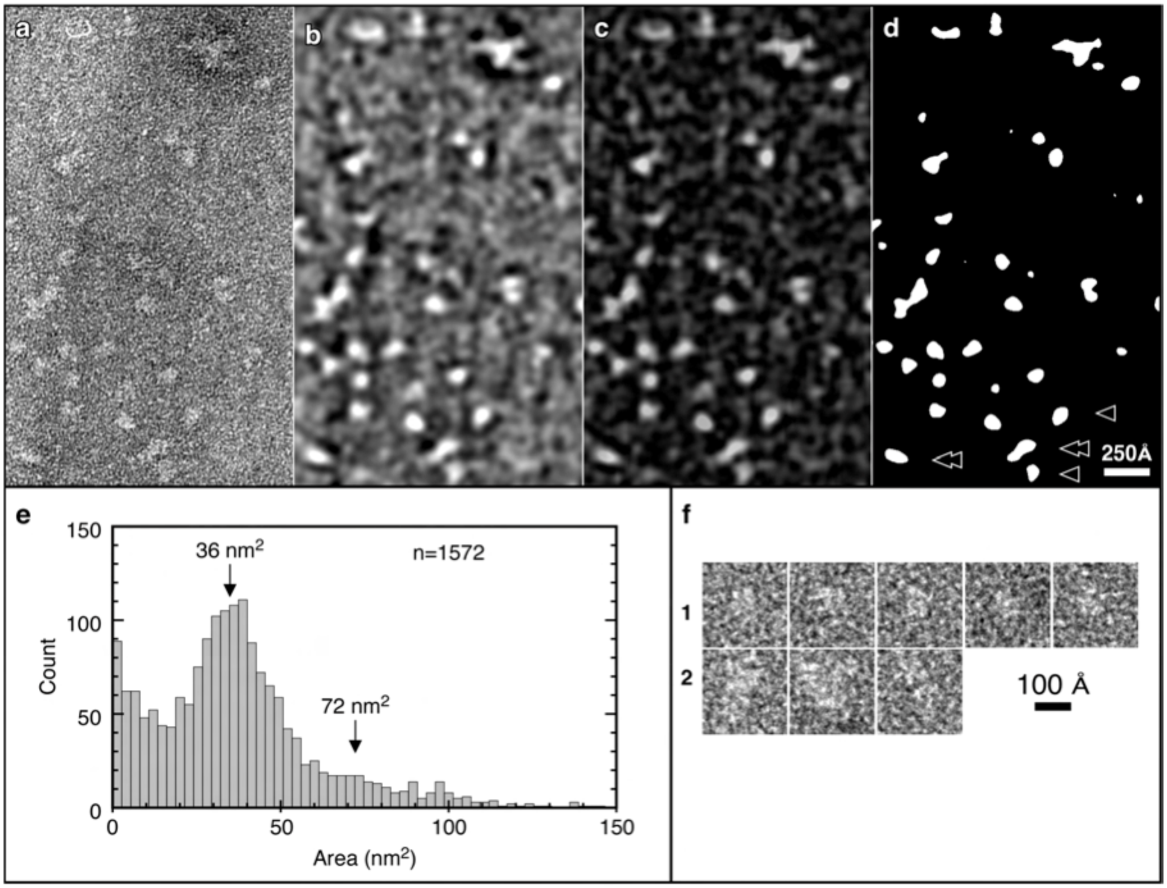

a
b
c
d
250Å
e
36 nm2
n=1572
72 nm2
150
100
50
0
Count
0
50
100
150
Area (nm2)
f
1
2
100 Å

Figure S1

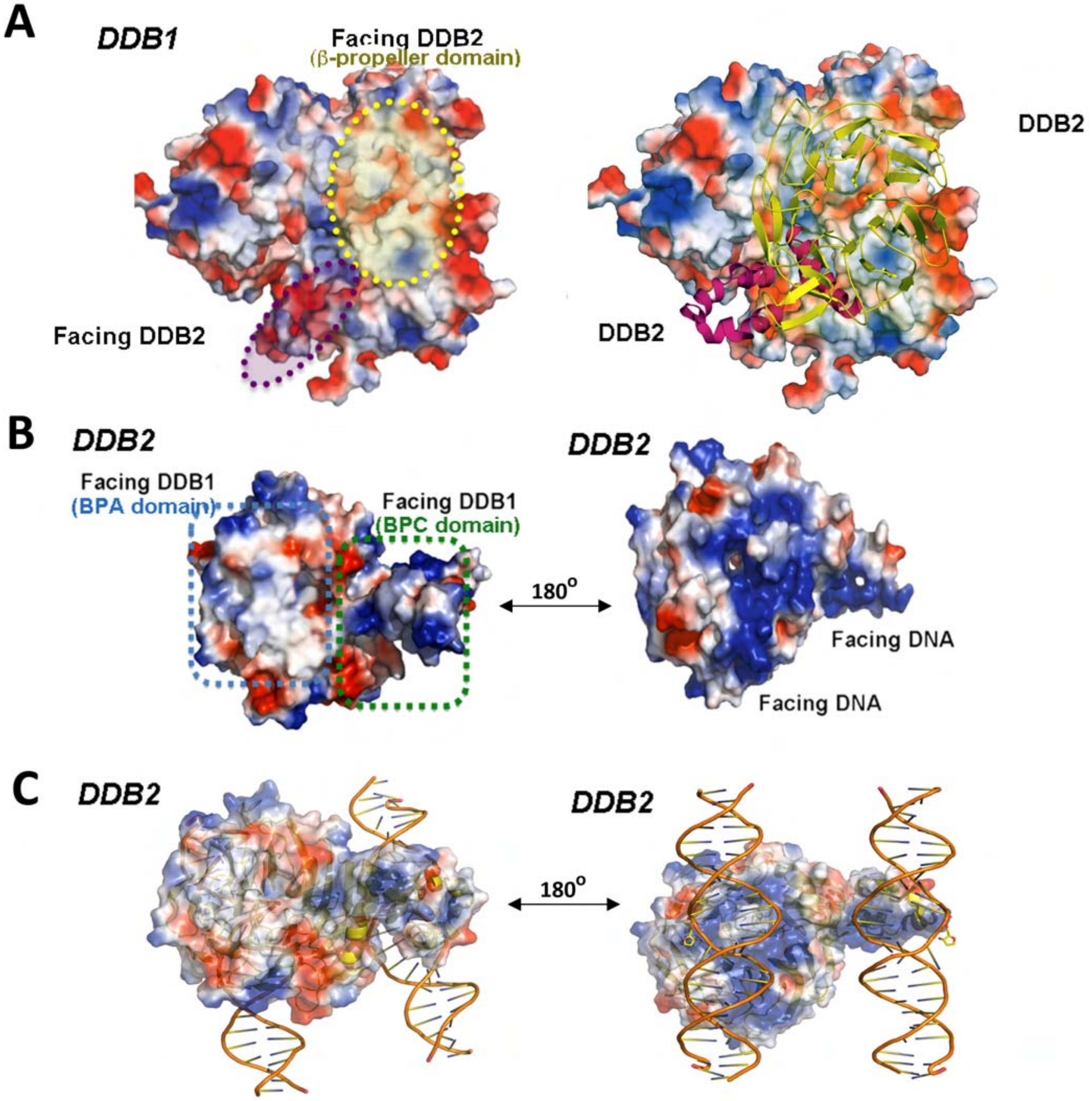

A
DDB1
Facing DDB2
(β-propeller domain)
Facing DDB2
DDB2
DDB2
B
DDB2
Facing DDB1
(BPA domain)
Facing DDB1
(BPC domain)
180°
DDB2
Facing DNA
Facing DNA
C
DDB2
180°
DDB2

Figure S2

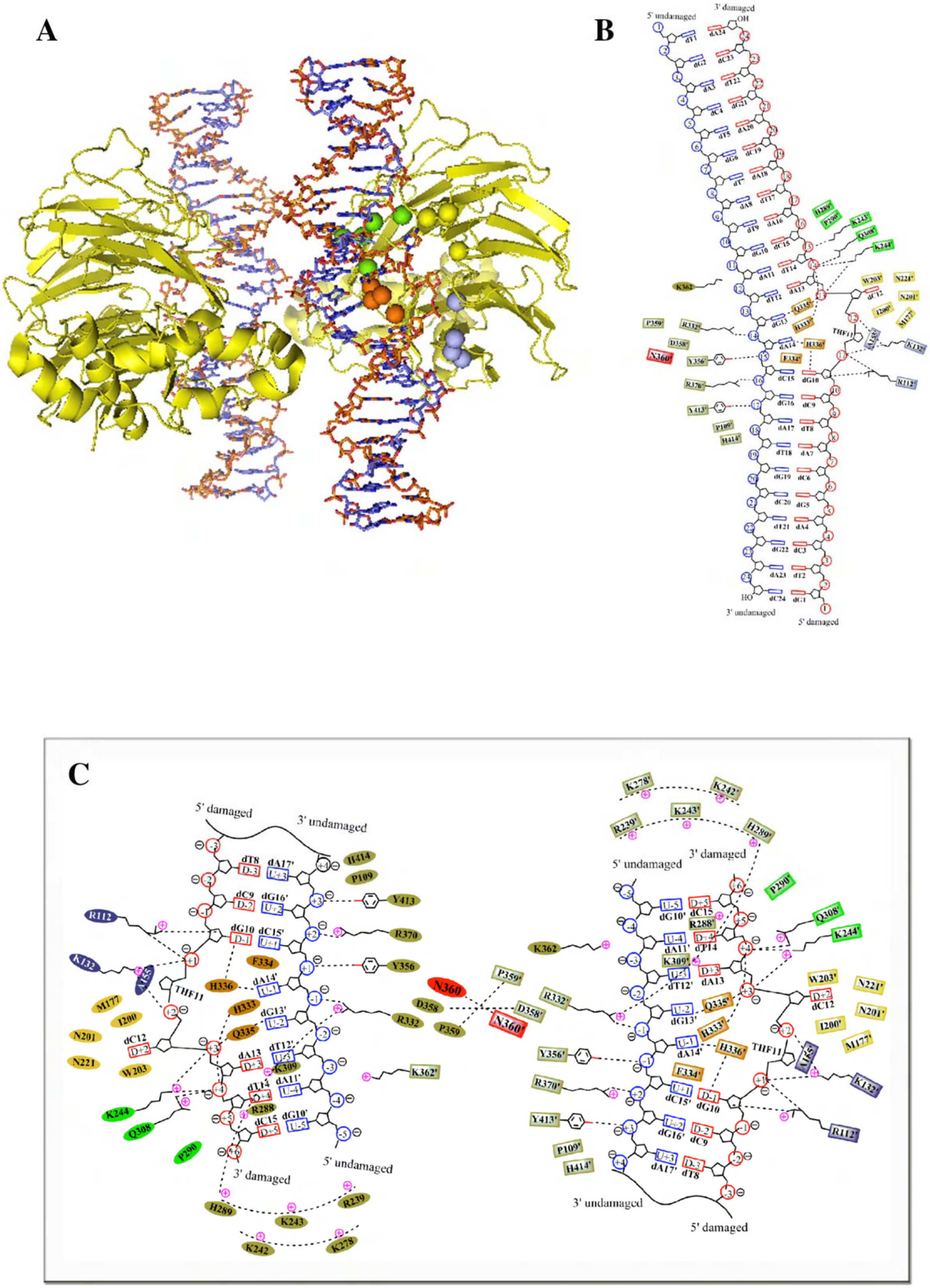

A
B
5' undamaged
3' damaged
3' undamaged
5' damaged
THF11
C
5' damaged
3' undamaged
3' damaged
5' undamaged
N360
D358
P359
N360'
D358'
P359'
K362
K362'
THF11

Figure S3

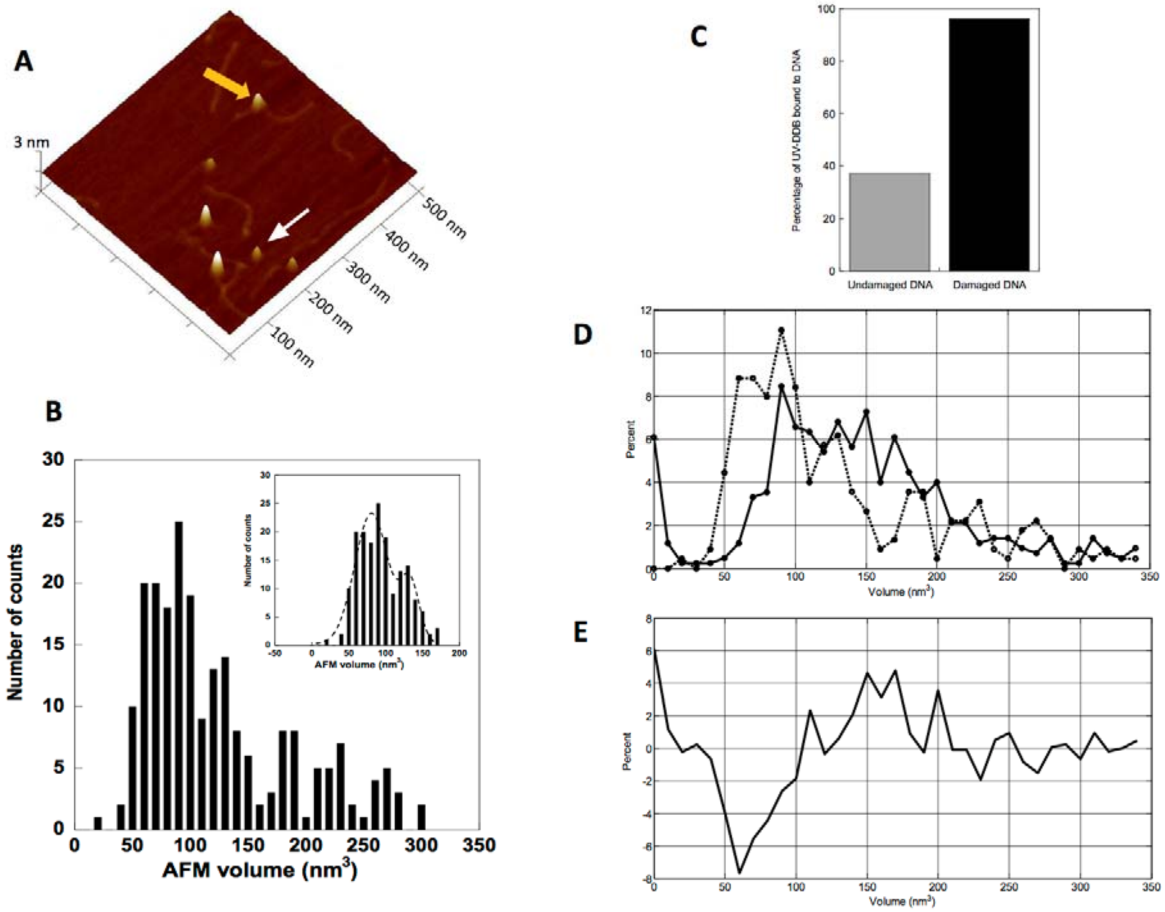
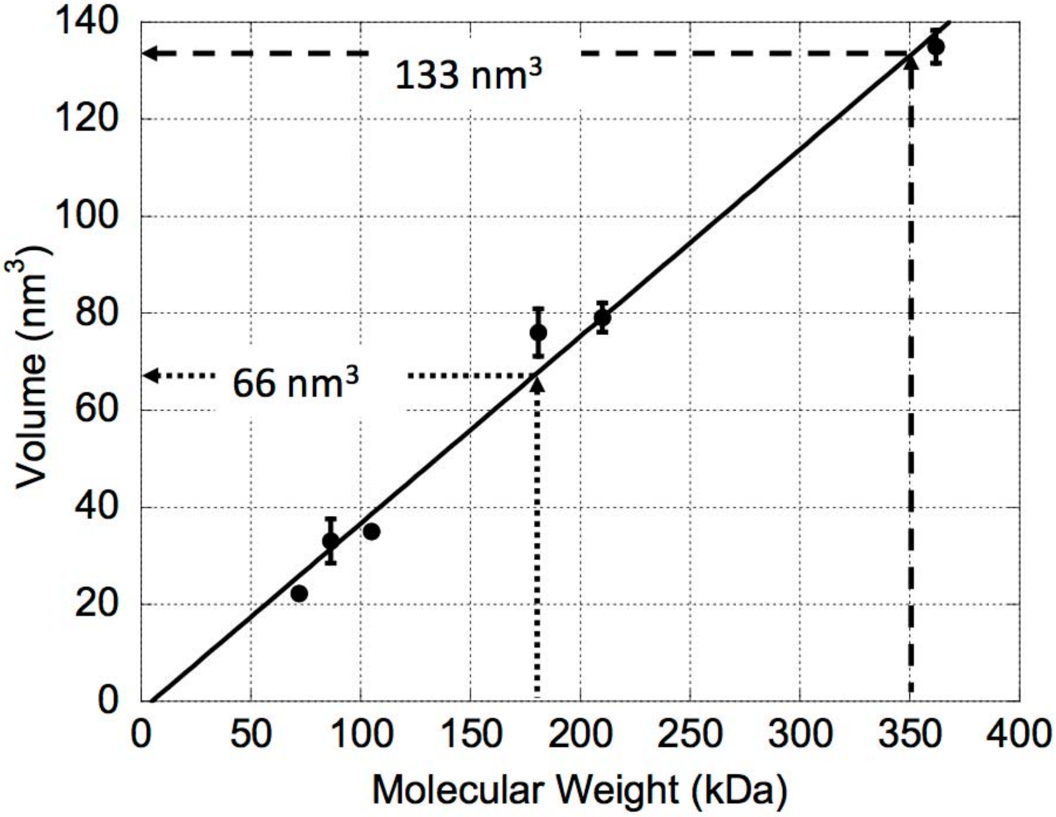

A
3 nm
500 nm
400 nm
300 nm
200 nm
100 nm
B
Number of counts
0
5
10
15
20
25
30
AFM volume (nm³)
0
50
100
150
200
250
300
350
Number of counts
AFM volume (nm³)
-50
0
50
100
150
200
C
Percentage of UV-DDB bound to DNA
0
20
40
60
80
100
Undamaged DNA
Damaged DNA
D
Percent
0
2
4
6
8
10
12
Volume (nm³)
0
50
100
150
200
250
300
350
E
Percent
-8
-6
-4
-2
0
2
4
6
8
Volume (nm³)
0
50
100
150
200
250
300
350

Figure S4



140
120
100
80
60
40
20
0
Volume (nm^3)
133 nm^3
66 nm^3
0
50
100
150
200
250
300
350
400
Molecular Weight (kDa)

Figure S5

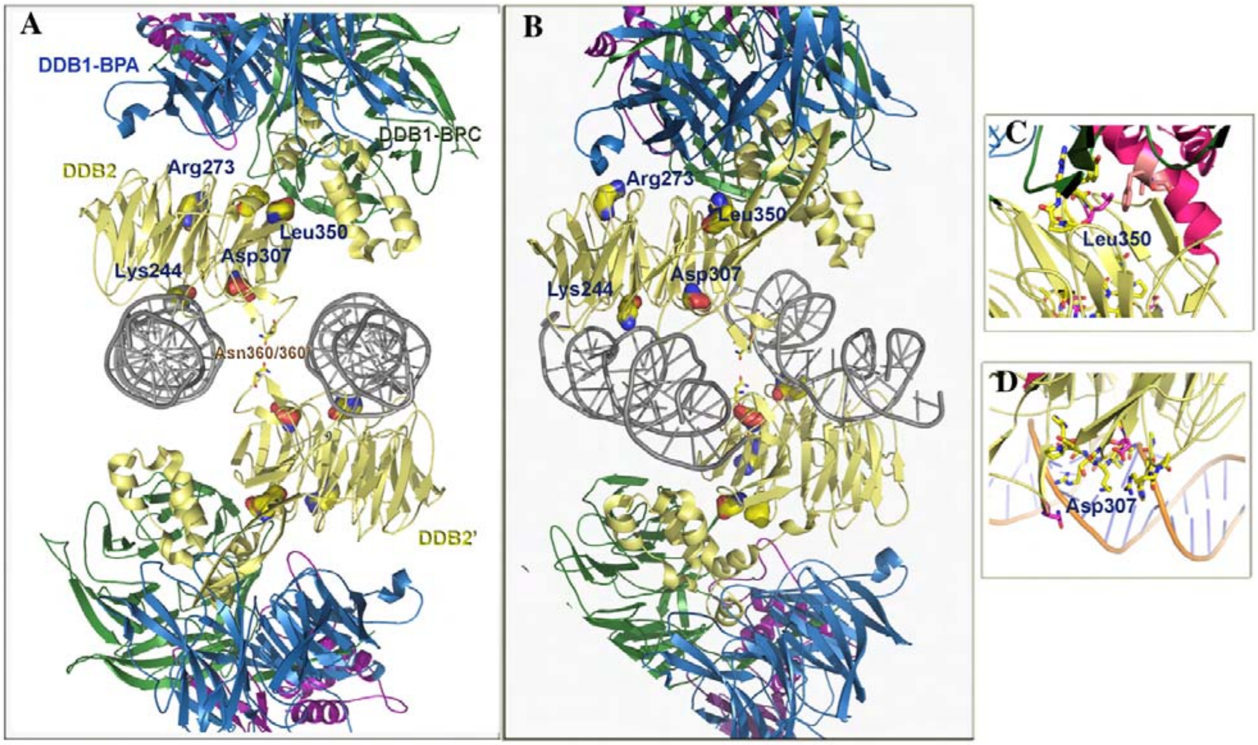

A
DDB1-BPA
DDB1-BPC
DDB2
Arg273
Leu350
Lys244
Asp307
Asn360/360'
DDB2'
B
Arg273
Leu350
Lys244
Asp307
C
Leu350
D
Asp307

Figure S6

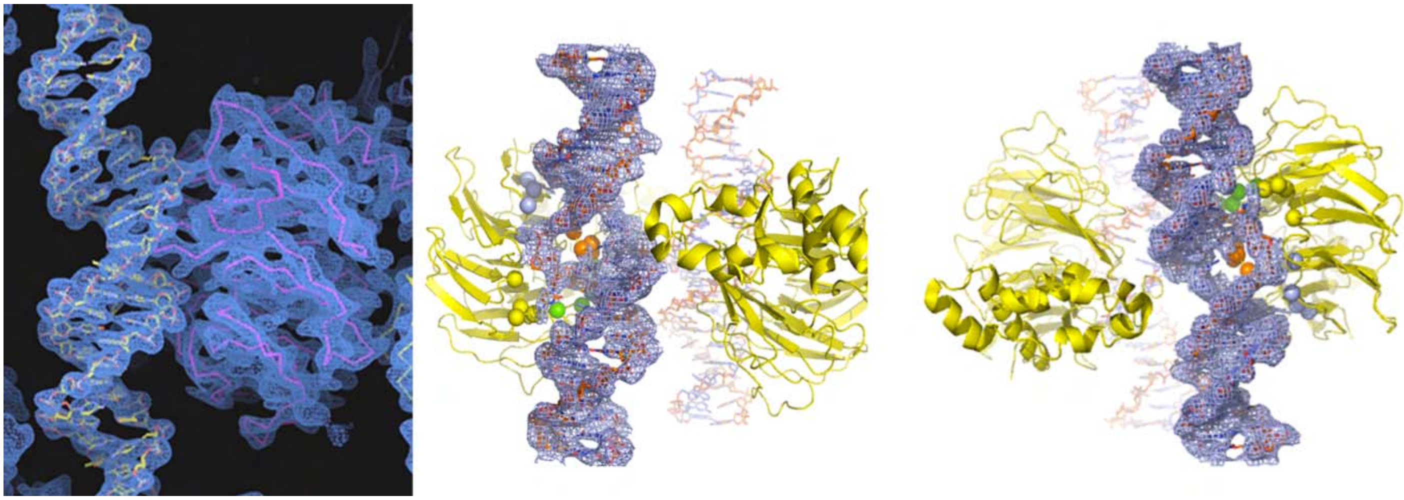


Figure S7

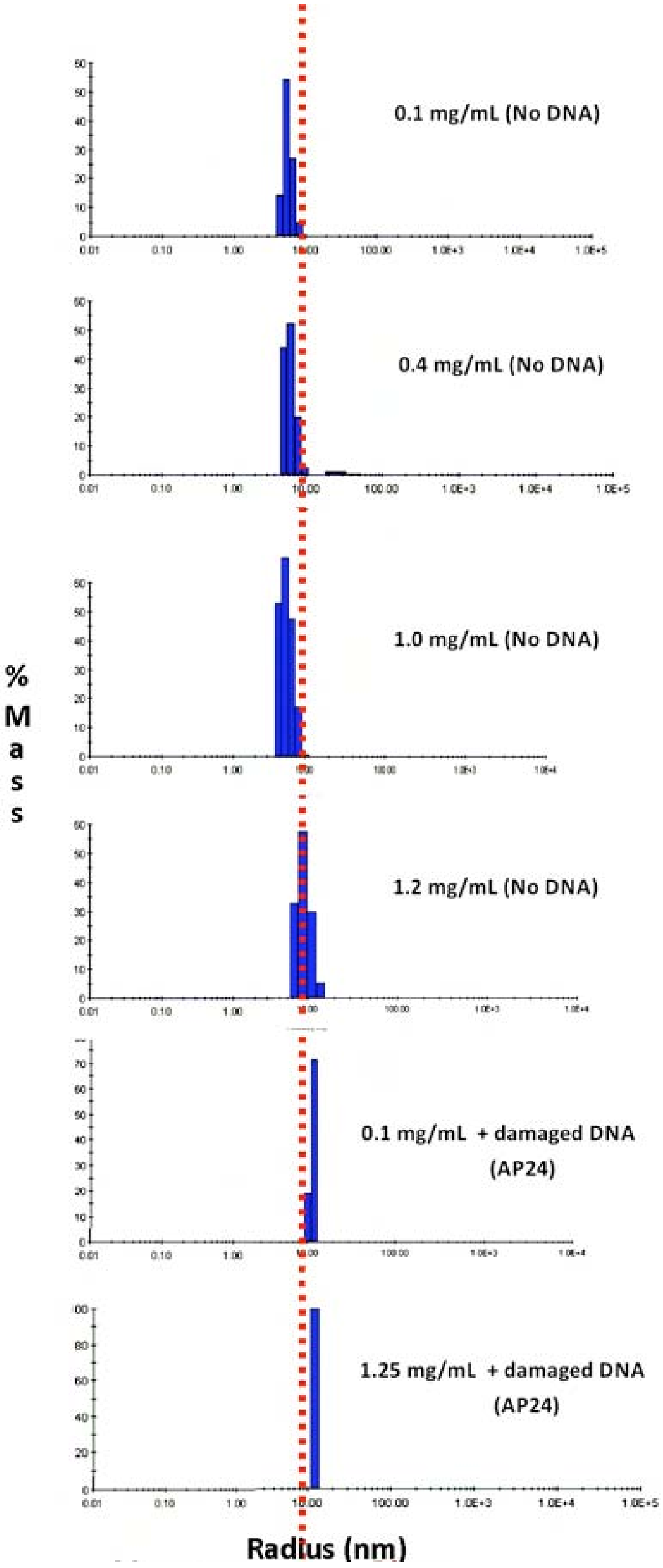

0.1 mg/mL (No DNA)
0.4 mg/mL (No DNA)
1.0 mg/mL (No DNA)
1.2 mg/mL (No DNA)
0.1 mg/mL + damaged DNA (AP24)
1.25 mg/mL + damaged DNA (AP24)
% Mass
Radius (nm)
0.01
0.10
1.00
10.00
100.00
1.0E+3
1.0E+4
1.0E+5

Figure S8

BIBLIOGRAPHY

[1] Jan HJ Hoeijmakers.
Genome maintenance mechanisms for preventing cancer.
Nature, 411(6835):366–374, 2001.

[2] Jan de Boer and Jan HJ Hoeijmakers.
Nucleotide excision repair and human syndromes.
Carcinogenesis, 21(3):453–460, 2000.

[3] A. Aboussekhra, M. Biggerstaff, M. K. Shivji, J. A. Vilpo, V. Moncollin, V. N. Podust, M. Protic, U. Hubscher, J. M. Egly, and R. D. Wood.
Mammalian dna nucleotide excision repair reconstituted with purified protein components.
Cell, 80(6):859–68, 1995.

[4] O. D. Scharer.
Hot topics in dna repair: the molecular basis for different disease states caused by mutations in tfiih and xpg.
DNA Repair (Amst), 7(2):339–44, 2008.

[5] L. C. Gillet and O. D. Scharer.
Molecular mechanisms of mammalian global genome nucleotide excision repair.
Chem Rev, 106(2):253–76, 2006.

[6] Philip C Hanawalt and Graciela Spivak.
Transcription-coupled dna repair: two decades of progress and surprises.
Nature Reviews Molecular Cell Biology, 9(12):958–970, 2008.

[7] Satya Prakash, Robert E Johnson, and Louise Prakash.
Eukaryotic translesion synthesis dna polymerases: specificity of structure and function.
Annu. Rev. Biochem., 74:317–353, 2005.

[8] Scott D McCulloch and Thomas A Kunkel.
The fidelity of dna synthesis by eukaryotic replicative and translesion synthesis polymerases.
Cell research, 18(1):148–161, 2008.

[9] Hanspeter Naegeli and Kaoru Sugasawa.
The xeroderma pigmentosum pathway: decision tree analysis of dna quality.
DNA repair, 10(7):673–683, 2011.

[10] Marcel Volker, Martijn J Moné, Parimal Karmakar, Anneke van Hoffen, Wouter Schul, Wim Vermeulen, Jan HJ Hoeijmakers, Roel van Driel, Albert A van Zeeland, and Leon HF Mullenders.
Sequential assembly of the nucleotide excision repair factors in vivo.
Molecular cell, 8(1):213–224, 2001.

[11] Sami N Guzder, Patrick Sung, Louise Prakash, and Satya Prakash.
Nucleotide excision repair in yeast is mediated by sequential assembly of repair factors and not by a pre-assembled repairosome.
Journal of Biological Chemistry, 271(15):8903–8910, 1996.

[12] Antonio Politi, Martijn J Moné, Adriaan B Houtsmuller, Deborah Hoogstraten, Wim Vermeulen, Reinhart Heinrich, and Roel van Driel.
Mathematical modeling of nucleotide excision repair reveals efficiency of sequential assembly strategies.
Molecular cell, 19(5):679–690, 2005.

[13] Vincent Mocquet, Jean Philippe Laine, Thilo Riedl, Zhou Yajin, Marietta Y Lee, and Jean Marc Egly.
Sequential recruitment of the repair factors during ner: the role of xpg in initiating the resynthesis step.
The EMBO journal, 27(1):155–167, 2007.

[14] Thilo Riedl, Fumio Hanaoka, and Jean-Marc Egly.
The comings and goings of nucleotide excision repair factors on damaged dna.
The EMBO journal, 22(19):5293–5303, 2003.

[15] K Sugasawa.
Multiple dna damage recognition factors involved in mammalian nucleotide excision repair.
Biochemistry (Moscow), 76(1):16–23, 2011.

[16] A. Scrima, R. Konickova, B. K. Czyzewski, Y. Kawasaki, P. D. Jeffrey, R. Groisman, Y. Nakatani, S. Iwai, N. P. Pavletich, and N. H. Thoma.
Structural basis of uv dna-damage recognition by the ddb1-ddb2 complex.
Cell, 135(7):1213–23, 2008.

[17] Mitsuo Wakasugi, Mayuko Shimizu, Hiroshi Morioka, Stuart Linn, Osamu Nikaido, and Tsukasa Matsunaga.
Damaged dna-binding protein ddb stimulates the excision of cyclobutane pyrimidine dimers in vitro in concert with xpa and replication protein a.
Journal of Biological Chemistry, 276(18):15434–15440, 2001.

[18] Birgitte Ø Wittschieben, Shigenori Iwai, and Richard D Wood.
Ddb1-ddb2 (xeroderma pigmentosum group e) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in dna.
Journal of Biological Chemistry, 280(48):39982–39989, 2005.

[19] Kaoru Sugasawa, Tomoko Okamoto, Yuichiro Shimizu, Chikahide Masutani, Shigenori Iwai, and Fumio Hanaoka.
A multistep damage recognition mechanism for global genomic nucleotide excision repair.
Genes & development, 15(5):507–521, 2001.

[20] Thomas Hey, Georg Lipps, Kaoru Sugasawa, Shigenori Iwai, Fumio Hanaoka, and Gerhard Krauss.
The xpc-hr23b complex displays high affinity and specificity for damaged dna in a true-equilibrium fluorescence assay.
Biochemistry, 41(21):6583–6587, 2002.

[21] Jean Y Tang, Byung Joon Hwang, James M Ford, Philip C Hanawalt, and Gilbert Chu.
Xeroderma pigmentosum¡ i¿ p48¡/i¿ gene enhances global genomic repair and suppresses uv-induced mutagenesis.
Molecular cell, 5(4):737–744, 2000.

[22] Maureen E Fitch, Satoshi Nakajima, Akira Yasui, and James M Ford.
In vivo recruitment of xpc to uv-induced cyclobutane pyrimidine dimers by the ddb2 gene product.
Journal of Biological Chemistry, 278(47):46906–46910, 2003.

[23] Vesna RapićOtrin, Isao Kuraoka, Tiziana Nardo, Mary McLenigan, APM Eker, Miria Stefanini, Arthur S Levine, and Richard D Wood.
Relationship of the xeroderma pigmentosum group e dna repair defect to the chromatin and dna binding proteins uv-ddb and replication protein a.
Molecular and cellular biology, 18(6):3182–3190, 1998.

[24] Stephane Angers, Ti Li, Xianhua Yi, Michael J MacCoss, Randall T Moon, and Ning Zheng.
Molecular architecture and assembly of the ddb1–cul4a ubiquitin ligase machinery.
Nature, 443(7111):590–593, 2006.

[25] E. S. Fischer, A. Scrima, K. Bohm, S. Matsumoto, G. M. Lingaraju, M. Faty, T. Yasuda, S. Cavadini, M. Wakasugi, F. Hanaoka, S. Iwai, H. Gut, K. Sugasawa, and N. H. Thoma.
The molecular basis of crl4ddb2/csa ubiquitin ligase architecture, targeting, and activation.
Cell, 147(5):1024–39, 2011.

[26] Maria G Kapetanaki, Jennifer Guerrero-Santoro, Dawn C Bisi, Ching L Hsieh, Vesna Rapić-Otrin, and Arthur S Levine.
The ddb1-cul4addb2 ubiquitin ligase is deficient in xeroderma pigmentosum group e and targets histone h2a at uv-damaged dna sites.
Proceedings of the National Academy of Sciences of the United States of America, 103(8):2588–2593, 2006.

[27] Hengbin Wang, Ling Zhai, Jun Xu, Heui-Yun Joo, Sarah Jackson, Hediye Erdjument-Bromage, Paul Tempst, Yue Xiong, and Yi Zhang.
Histone h3 and h4 ubiquitylation by the cul4-ddb-roc1 ubiquitin ligase facilitates cellular response to dna damage.
Molecular cell, 22(3):383–394, 2006.

[28] Jennifer Guerrero-Santoro, Maria G Kapetanaki, Ching L Hsieh, Ilya Gorbachinsky, Arthur S Levine, and Vesna Rapić-Otrin.
The cullin 4b–based uv-damaged dna-binding protein ligase binds to uv-damaged chromatin and ubiquitinates histone h2a.
Cancer research, 68(13):5014–5022, 2008.

[29] Kaoru Sugasawa, Yuki Okuda, Masafumi Saijo, Ryotaro Nishi, Noriyuki Matsuda, Gilbert Chu, Toshio Mori, Shigenori Iwai, Keiji Tanaka, Kiyoji Tanaka, et al.
Uv-induced ubiquitylation of xpc protein mediated by uv-ddb-ubiquitin ligase complex.
Cell, 121(3):387–400, 2005.

[30] Vesna Rapić-Otrin, Valentina Navazza, Tiziana Nardo, Elena Botta, Mary McLenigan, Dawn C Bisi, Arthur S Levine, and Miria Stefanini.
True xp group e patients have a defective uv-damaged dna binding protein complex and mutations in ddb2 which reveal the functional domains of its p48 product.
Human molecular genetics, 12(13):1507–1522, 2003.

[31] Regina Groisman, Jolanta Polanowska, Isao Kuraoka, Jun-ichi Sawada, Masafumi Saijo, Ronny Drapkin, Alexei F Kisselev, Kiyoji Tanaka, and Yoshihiro Nakatani.
The ubiquitin ligase activity in the ddb2 and csa complexes is differentially regulated by the cop9 signalosome in response to dna damage.
Cell, 113(3):357–367, 2003.

[32] Li Lan, Satoshi Nakajima, Maria G Kapetanaki, Ching L Hsieh, Matthew Fagerburg, Karen Thickman, Pedro Rodriguez-Collazo, Sanford H Leuba, Arthur S Levine, and Vesna Rapić-Otrin.
Monoubiquitinated histone h2a destabilizes photolesion-containing nucleosomes with concomitant release of uv-damaged dna-binding protein e3 ligase.
Journal of Biological Chemistry, 287(15):12036–12049, 2012.

[33] Q. E. Wang, Q. Zhu, G. Wani, J. Chen, and A. A. Wani.
Uv radiation-induced xpc translocation within chromatin is mediated by damaged-dna binding protein, ddb2.

Carcinogenesis, 25(6):1033–43, 2004.

[34] P. J. van der Spek, A. Eker, S. Rademakers, C. Visser, K. Sugasawa, C. Masutani, F. Hanaoka, D. Bootsma, and J. H. Hoeijmakers.
Xpc and human homologs of rad23: intracellular localization and relationship to other nucleotide excision repair complexes.
Nucleic Acids Res, 24(13):2551–9, 1996.

[35] Aurel Popescu, Simona Miron, Yves Blouquit, Patricia Duchambon, Petya Christova, and Constantin T Craescu.
Xeroderma pigmentosum group c protein possesses a high affinity binding site to human centrin 2 and calmodulin.
Journal of Biological Chemistry, 278(41):40252–40261, 2003.

[36] Ryotaro Nishi, Yuki Okuda, Eriko Watanabe, Toshio Mori, Shigenori Iwai, Chikahide Masutani, Kaoru Sugasawa, and Fumio Hanaoka.
Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group c protein.
Molecular and cellular biology, 25(13):5664–5674, 2005.

[37] Kaoru Sugasawa, JM Ng, Chikahide Masutani, Takafumi Maekawa, Akio Uchida, PJ Van der Spek, AP Eker, Suzanne Rademakers, Cecile Visser, Abdelilah Aboussekhra, et al.
Two human homologs of rad23 are functionally interchangeable in complex formation and stimulation of xpc repair activity.
Molecular and cellular biology, 17(12):6924–6931, 1997.

[38] J. M. Ng, W. Vermeulen, G. T. van der Horst, S. Bergink, K. Sugasawa, H. Vrieling, and J. H. Hoeijmakers.
A novel regulation mechanism of dna repair by damage-induced and rad23-dependent stabilization of xeroderma pigmentosum group c protein.
Genes Dev, 17(13):1630–45, 2003.

[39] S. Bergink, W. Toussaint, M. S. Luijsterburg, C. Dinant, S. Alekseev, J. H. Hoeijmakers, N. P. Dantuma, A. B. Houtsmuller, and W. Vermeulen.
Recognition of dna damage by xpc coincides with disruption of the xpc-rad23 complex.
J Cell Biol, 196(6):681–8, 2012.

[40] Randy Legerski and Carolyn Peterson.
Expression cloning of a human dna repair gene involved in xeroderma pigmentosum group c.
1992.

[41] J. H. Min and N. P. Pavletich.
Recognition of dna damage by the rad4 nucleotide excision repair protein.
Nature, 449(7162):570–5, 2007.

[42] P. Schultz, S. Fribourg, A. Poterszman, V. Mallouh, D. Moras, and J. M. Egly.
Molecular structure of human tfiih.
Cell, 102(5):599–607, 2000.

[43] J. M. Egly and F. Coin.
A history of tfiih: two decades of molecular biology on a pivotal transcription/repair factor.
DNA Repair (Amst), 10(7):714–21, 2011.

[44] M. Yokoi, C. Masutani, T. Maekawa, K. Sugasawa, Y. Ohkuma, and F. Hanaoka.
The xeroderma pigmentosum group c protein complex xpc-hr23b plays an important role in the recruitment of transcription factor iih to damaged dna.
J Biol Chem, 275(13):9870–5, 2000.

[45] A. Uchida, K. Sugasawa, C. Masutani, N. Dohmae, M. Araki, M. Yokoi, Y. Ohkuma, and F. Hanaoka.
The carboxy-terminal domain of the xpc protein plays a crucial role in nucleotide excision repair through interactions with transcription factor iih.
DNA Repair (Amst), 1(6):449–61, 2002.

[46] S. Miron, P. Duchambon, Y. Blouquit, D. Durand, and C. T. Craescu.
The carboxy-terminal domain of xeroderma pigmentosum complementation group c protein, involved in tfiih and centrin binding, is highly disordered.
Biochemistry, 47(5):1403–13, 2008.

[47] J. Lafrance-Vanasse, G. Arseneault, L. Cappadocia, P. Legault, and J. G. Omichinski.
Structural and functional evidence that rad4 competes with rad2 for binding to the tfb1 subunit of tfiih in ner.
Nucleic Acids Res, 41(4):2736–45, 2013.

[48] A. Jawhari, J. P. Laine, S. Dubaele, V. Lamour, A. Poterszman, F. Coin, D. Moras, and J. M. Egly.
p52 mediates xpb function within the transcription/repair factor tfiih.
J Biol Chem, 277(35):31761–7, 2002.

[49] R. Roy, L. Schaeffer, S. Humbert, W. Vermeulen, G. Weeda, and J. M. Egly.
The dna-dependent atpase activity associated with the class ii basic transcription factor btf2/tfiih.
J Biol Chem, 269(13):9826–32, 1994.

[50] S. N. Guzder, P. Sung, V. Bailly, L. Prakash, and S. Prakash.
Rad25 is a dna helicase required for dna repair and rna polymerase ii transcription.
Nature, 369(6481):578–81, 1994.

[51] F. Coin, V. Oksenych, and J. M. Egly.
Distinct roles for the xpb/p52 and xpd/p44 subcomplexes of tfiih in damaged dna opening during nucleotide excision repair.

Mol Cell, 26(2):245–56, 2007.

[52] V. Oksenych, B. Bernardes de Jesus, A. Zhovmer, J. M. Egly, and F. Coin.
Molecular insights into the recruitment of tfiih to sites of dna damage.
EMBO J, 28(19):2971–80, 2009.

[53] Geert Weeda, Reinier CA van Ham, Wim Vermeulen, Dirk Bootsma, Alex J van der Eb, and Jan HJ Hoeijmakers.
A presumed dna helicase encoded by¡ i¿ ercc-3¡/i¿ is involved in the human repair disorders xeroderma pigmentosum and cockayne's syndrome.
Cell, 62(4):777–791, 1990.

[54] Z. Qi, R. A. Pugh, M. Spies, and Y. R. Chemla.
Sequence-dependent base pair stepping dynamics in xpd helicase unwinding.
Elife, 2:e00334, 2013.

[55] J. Kuper, S. C. Wolski, G. Michels, and C. Kisker.
Functional and structural studies of the nucleotide excision repair helicase xpd suggest a polarity for dna translocation.
EMBO J, 31(2):494–502, 2012.

[56] R. A. Pugh, M. Honda, H. Leesley, A. Thomas, Y. Lin, M. J. Nilges, I. K. Cann, and M. Spies.
The iron-containing domain is essential in rad3 helicases for coupling of atp hydrolysis to dna translocation and for targeting the helicase to the single-stranded dna-double-stranded dna junction.
J Biol Chem, 283(3):1732–43, 2008.

[57] N. Mathieu, N. Kaczmarek, P. Ruthemann, A. Luch, and H. Naegeli.
Dna quality control by a lesion sensor pocket of the xeroderma pigmentosum group d helicase subunit of tfiih.
Curr Biol, 23(3):204–12, 2013.

[58] Alan R Lehmann.
The xeroderma pigmentosum group d (xpd) gene: one gene, two functions, three diseases.
Genes & Development, 15(1):15–23, 2001.

[59] F. Coin, V. Oksenych, V. Mocquet, S. Groh, C. Blattner, and J. M. Egly.
Nucleotide excision repair driven by the dissociation of cak from tfiih.
Mol Cell, 31(1):9–20, 2008.

[60] F. Coin, J. C. Marinoni, and J. M. Egly.
Mutations in xpd helicase prevent its interaction and regulation by p44, another subunit of tfiih, resulting in xeroderma pigmentosum (xp) and trichothiodystrophy (ttd) phenotypes.
Pathol Biol (Paris), 46(9):679–80, 1998.

[61] B. Sandrock and J. M. Egly.
A yeast four-hybrid system identifies cdk-activating kinase as a regulator of the xpd helicase, a subunit of transcription factor iih.
J Biol Chem, 276(38):35328–33, 2001.

[62] Ulrike Camenisch, Ramiro Dip, Sylvie Briand Schumacher, Benjamin Schuler, and Hanspeter Naegeli.
Recognition of helical kinks by xeroderma pigmentosum group a protein triggers dna excision repair.
Nature structural & molecular biology, 13(3):278–284, 2006.

[63] C. G. Bunick, M. R. Miller, B. E. Fuller, E. Fanning, and W. J. Chazin.
Biochemical and structural domain analysis of xeroderma pigmentosum complementation group c protein.
Biochemistry, 45(50):14965–79, 2006.

[64] C. H. Park, D. Mu, J. T. Reardon, and A. Sancar.
The general transcription-repair factor tfiih is recruited to the excision repair complex by the xpa protein independent of the tfiie transcription factor.
J Biol Chem, 270(9):4896–902, 1995.

[65] A. M. Sijbers, W. L. de Laat, R. R. Ariza, M. Biggerstaff, Y. F. Wei, J. G. Moggs, K. C. Carter, B. K. Shell, E. Evans, M. C. de Jong, S. Rademakers, J. de Rooij, N. G. Jaspers, J. H. Hoeijmakers, and R. D. Wood.
Xeroderma pigmentosum group f caused by a defect in a structure-specific dna repair endonuclease.
Cell, 86(5):811–22, 1996.

[66] T. Yagi, Y. Matsumura, M. Sato, C. Nishigori, T. Mori, A. M. Sijbers, and H. Takebe.
Complete restoration of normal dna repair characteristics in group f xeroderma pigmentosum cells by over-expression of transfected xpf cdna.
Carcinogenesis, 19(1):55–60, 1998.

[67] A. B. Houtsmuller, S. Rademakers, A. L. Nigg, D. Hoogstraten, J. H. Hoeijmakers, and W. Vermeulen.
Action of dna repair endonuclease ercc1/xpf in living cells.
Science, 284(5416):958–61, 1999.

[68] M. Bowles, J. Lally, A. J. Fadden, S. Mouilleron, T. Hammonds, and N. Q. McDonald.
Fluorescence-based incision assay for human xpf-ercc1 activity identifies important elements of dna junction recognition.
Nucleic Acids Res, 40(13):e101, 2012.

[69] T. Bessho, A. Sancar, L. H. Thompson, and M. P. Thelen.
Reconstitution of human excision nuclease with recombinant xpf-ercc1 complex.
J Biol Chem, 272(6):3833–7, 1997.

[70] T. Matsunaga, C. H. Park, T. Bessho, D. Mu, and A. Sancar.
Replication protein a confers structure-specific endonuclease activities to the xpf-ercc1 and xpg subunits of human dna repair excision nuclease.
J Biol Chem, 271(19):11047–50, 1996.

[71] L. A. Fisher, M. Bessho, M. Wakasugi, T. Matsunaga, and T. Bessho.
Role of interaction of xpf with rpa in nucleotide excision repair.
J Mol Biol, 413(2):337–46, 2011.

[72] C. H. Park and A. Sancar.
Formation of a ternary complex by human xpa, ercc1, and ercc4(xpf) excision repair proteins.
Proc Natl Acad Sci U S A, 91(11):5017–21, 1994.

[73] L. Li, S. J. Elledge, C. A. Peterson, E. S. Bales, and R. J. Legerski.
Specific association between the human dna repair proteins xpa and ercc1.
Proc Natl Acad Sci U S A, 91(11):5012–6, 1994.

[74] L. Li, X. Lu, C. A. Peterson, and R. J. Legerski.
An interaction between the dna repair factor xpa and replication protein a appears essential for nucleotide excision repair.
Mol Cell Biol, 15(10):5396–402, 1995.

[75] E. Evans, J. Fellows, A. Coffer, and R. D. Wood.
Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human xpg protein.
EMBO J, 16(3):625–38, 1997.

[76] O. V. Tsodikov, J. H. Enzlin, O. D. Scharer, and T. Ellenberger.
Crystal structure and dna binding functions of ercc1, a subunit of the dna structure-specific endonuclease xpf-ercc1.
Proc Natl Acad Sci U S A, 102(32):11236–41, 2005.

[77] D. Das, G. E. Folkers, M. van Dijk, N. G. Jaspers, J. H. Hoeijmakers, R. Kaptein, and R. Boelens.
The structure of the xpf-ssdna complex underscores the distinct roles of the xpf and ercc1 helix- hairpin-helix domains in ss/ds dna recognition.
Structure, 20(4):667–75, 2012.

[78] M. Hohl, F. Thorel, S. G. Clarkson, and O. D. Scharer.
Structural determinants for substrate binding and catalysis by the structure-specific endonuclease xpg.
J Biol Chem, 278(21):19500–8, 2003.

[79] Z. He, L. A. Henricksen, M. S. Wold, and C. J. Ingles.
Rpa involvement in the damage-recognition and incision steps of nucleotide excision repair.

Nature, 374(6522):566–9, 1995.

[80] N. Iyer, M. S. Reagan, K. J. Wu, B. Canagarajah, and E. C. Friedberg.
Interactions involving the human rna polymerase ii transcription/nucleotide excision repair complex tfiih, the nucleotide excision repair protein xpg, and cockayne syndrome group b (csb) protein.
Biochemistry, 35(7):2157–67, 1996.

[81] M. Hohl, I. Dunand-Sauthier, L. Staresincic, P. Jaquier-Gubler, F. Thorel, M. Modesti, S. G. Clarkson, and O. D. Scharer.
Domain swapping between fen-1 and xpg defines regions in xpg that mediate nucleotide excision repair activity and substrate specificity.
Nucleic Acids Res, 35(9):3053–63, 2007.

[82] T. Riedl, F. Hanaoka, and J. M. Egly.
The comings and goings of nucleotide excision repair factors on damaged dna.
EMBO J, 22(19):5293–303, 2003.

[83] M. Wakasugi, J. T. Reardon, and A. Sancar.
The non-catalytic function of xpg protein during dual incision in human nucleotide excision repair.
J Biol Chem, 272(25):16030–4, 1997.

[84] A. Tapias, J. Auriol, D. Forget, J. H. Enzlin, O. D. Scharer, F. Coin, B. Coulombe, and J. M. Egly.
Ordered conformational changes in damaged dna induced by nucleotide excision repair factors.
J Biol Chem, 279(18):19074–83, 2004.

[85] L. Staresincic, A. F. Fagbemi, J. H. Enzlin, A. M. Gourdin, N. Wijgers, I. Dunand-Sauthier, G. Giglia-Mari, S. G. Clarkson, W. Vermeulen, and O. D. Scharer.
Coordination of dual incision and repair synthesis in human nucleotide excision repair.
EMBO J, 28(8):1111–20, 2009.

[86] V. Mocquet, J. P. Laine, T. Riedl, Z. Yajin, M. Y. Lee, and J. M. Egly.
Sequential recruitment of the repair factors during ner: the role of xpg in initiating the resynthesis step.
EMBO J, 27(1):155–67, 2008.

[87] J. Moser, H. Kool, I. Giakzidis, K. Caldecott, L. H. Mullenders, and M. I. Fousteri.
Sealing of chromosomal dna nicks during nucleotide excision repair requires xrcc1 and dna ligase iii alpha in a cell-cycle-specific manner.
Mol Cell, 27(2):311–23, 2007.

[88] H. Mu, K. Kropachev, Y. Chen, H. Zhang, Y. Cai, N. E. Geacintov, and S. Broyde.
Role of structural and energetic factors in regulating repair of a bulky dna lesion with different opposite partner bases.

Biochemistry, 52(33):5517–21, 2013.

[89] Y. Cai, D. J. Patel, S. Broyde, and N. E. Geacintov.
Base sequence context effects on nucleotide excision repair.
J Nucleic Acids, 2010, 2010.

[90] Y. Cai, N. E. Geacintov, and S. Broyde.
Nucleotide excision repair efficiencies of bulky carcinogen-dna adducts are governed by a balance between stabilizing and destabilizing interactions.
Biochemistry, 51(7):1486–99, 2012.

[91] S. Ding, K. Kropachev, Y. Cai, M. Kolbanovskiy, S. A. Durandina, Z. Liu, V. Shafirovich, S. Broyde, and N. E. Geacintov.
Structural, energetic and dynamic properties of guanine(c8)-thymine(n3) cross-links in dna provide insights on susceptibility to nucleotide excision repair.
Nucleic Acids Res, 40(6):2506–17, 2012.

[92] N. K. Banavali and Jr. MacKerell, A. D.
Free energy and structural pathways of base flipping in a dna gcgc containing sequence.
J Mol Biol, 319(1):141–60, 2002.

[93] U Deva Priyakumar and Alexander D MacKerell.
Computational approaches for investigating base flipping in oligonucleotides.
Chemical reviews, 106(2):489–505, 2006.

[94] L. L. O'Neil, A. Grossfield, and O. Wiest.
Base flipping of the thymine dimer in duplex dna.
J Phys Chem B, 111(40):11843–9, 2007.

[95] L. L. O'Neil and O. Wiest.
Structures and energetics of base flipping of the thymine dimer depend on dna sequence.
J Phys Chem B, 112(13):4113–22, 2008.

[96] H. Zheng, Y. Cai, S. Ding, Y. Tang, K. Kropachev, Y. Zhou, L. Wang, S. Wang, N. E. Geacintov, Y. Zhang, and S. Broyde.
Base flipping free energy profiles for damaged and undamaged dna.
Chem Res Toxicol, 23(12):1868–70, 2010.

[97] Dmitry G Vassylyev, Tatsuki Kashiwagi, Yuriko Mikami, Mariko Ariyoshi, Shigenori Iwai, Eiko Ohtsuka, and Kosuke Morikawa.
Atomic model of a pyrimidine dimer excision repair enzyme complexed with a dna substrate: structural basis for damaged dna recognition.
Cell, 83(5):773–782, 1995.

[98] E. M. Meulenbroek, C. Peron Cane, I. Jala, S. Iwai, G. F. Moolenaar, N. Goosen, and N. S. Pannu.

Uv damage endonuclease employs a novel dual-dinucleotide flipping mechanism to recognize different dna lesions.
Nucleic Acids Res, 41(2):1363–71, 2013.

[99] A. Mees, T. Klar, P. Gnau, U. Hennecke, A. P. Eker, T. Carell, and L. O. Essen.
Crystal structure of a photolyase bound to a cpd-like dna lesion after in situ repair.
Science, 306(5702):1789–93, 2004.

[100] S. Kiontke, Y. Geisselbrecht, R. Pokorny, T. Carell, A. Batschauer, and L. O. Essen.
Crystal structures of an archaeal class ii dna photolyase and its complex with uv-damaged duplex dna.
EMBO J, 30(21):4437–49, 2011.

[101] M. J. Maul, T. R. Barends, A. F. Glas, M. J. Cryle, T. Domratcheva, S. Schneider, I. Schlichting, and T. Carell.
Crystal structure and mechanism of a dna (6-4) photolyase.
Angew Chem Int Ed Engl, 47(52):10076–80, 2008.

[102] J. I. Yeh, A. S. Levine, S. Du, U. Chinte, H. Ghodke, H. Wang, H. Shi, C. L. Hsieh, J. F. Conway, B. Van Houten, and V. Rapic-Otrin.
Damaged dna induced uv-damaged dna-binding protein (uv-ddb) dimerization and its roles in chromatinized dna repair.
Proc Natl Acad Sci U S A, 109(41):E2737–46, 2012.

[103] D. P. Batty and R. D. Wood.
Damage recognition in nucleotide excision repair of dna.
Gene, 241(2):193–204, 2000.

[104] J. E. Yeo, A. Khoo, A. F. Fagbemi, and O. D. Scharer.
The efficiencies of damage recognition and excision correlate with duplex destabilization induced by acetylaminofluorene adducts in human nucleotide excision repair.
Chem Res Toxicol, 25(11):2462–8, 2012.

[105] Shigenori Iwai, Toshimi Mizukoshi, Yoshie Fujiwara, Chikahide Masutani, Fumio Hanaoka, and Yoshihiro Hayakawa.
Benzimidazolium triflate-activated synthesis of (6–4) photoproduct-containing oligonucleotides and its application.
Nucleic acids research, 27(11):2299–2303, 1999.

[106] Yang Liu, Dara Reeves, Konstantin Kropachev, Yuqin Cai, Shuang Ding, Marina Kolbanovskiy, Alexander Kolbanovskiy, Judith L Bolton, Suse Broyde, Bennett Van Houten, et al.
Probing for dna damage with β-hairpins: Similarities in incision efficiencies of bulky dna adducts by prokaryotic and human nucleotide excision repair systems¡ i¿ in vitro¡/i¿.
DNA repair, 10(7):684–696, 2011.

[107] Nadine Mathieu, Nina Kaczmarek, and Hanspeter Naegeli.
Strand-and site-specific dna lesion demarcation by the xeroderma pigmentosum group d helicase.
Proceedings of the National Academy of Sciences, 107(41):17545–17550, 2010.

[108] John J Hopfield.
Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity.
Proceedings of the National Academy of Sciences, 71(10):4135–4139, 1974.

[109] Y. Savir and T. Tlusty.
Conformational proofreading: the impact of conformational changes on the specificity of molecular recognition.
PLoS One, 2(5):e468, 2007.

[110] J. T. Reardon and A. Sancar.
Thermodynamic cooperativity and kinetic proofreading in dna damage recognition and repair.
Cell Cycle, 3(2):141–4, 2004.

[111] Iwijn De Vlaminck, Marijn TJ van Loenhout, Ludovit Zweifel, Johan den Blanken, Koen Hooning, Susanne Hage, Jacob Kerssemakers, and Cees Dekker.
Mechanism of homology recognition in dna recombination from dual-molecule experiments.
Molecular cell, 46(5):616–624, 2012.

[112] O. G. Berg, R. B. Winter, and P. H. von Hippel.
Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory.
Biochemistry, 20(24):6929–48, 1981.

[113] Amy M Furda, Amanda Smith Bess, Joel N Meyer, and Bennett Van Houten.
Analysis of dna damage and repair in nuclear and mitochondrial dna of animal cells using quantitative pcr.
In *DNA Repair Protocols*, pages 111–132. Springer, 2012.

[114] Hui Geng, Chunwei Du, Siying Chen, Vincenzo Salerno, Candela Manfredi, and Peggy Hsieh.
In vitro studies of dna mismatch repair proteins.
Analytical biochemistry, 413(2):179–184, 2011.

[115] Neil M Kad, Hong Wang, Guy G Kennedy, David M Warshaw, and Bennett Van Houten.
Collaborative dynamic dna scanning by nucleotide excision repair proteins investigated by single-molecule imaging of quantum-dot-labeled proteins.
Molecular cell, 37(5):702–713, 2010.

[116] Rogier Verberk, Antoine M van Oijen, and Michel Orrit.
Simple model for the power-law blinking of single semiconductor nanocrystals.
Physical Review B, 66(23):233202, 2002.

[117] Eva C Arnspang, Jonathan R Brewer, and B Christoffer Lagerholm.
Multi-color single particle tracking with quantum dots.
PloS one, 7(11):e48521, 2012.

[118] Eva Arnspang Christensen, Pasad Kulatunga, and B Christoffer Lagerholm.
A single molecule investigation of the photostability of quantum dots.
PloS one, 7(8):e44355, 2012.

[119] Rowena Mittal and Marcel P Bruchez.
Biotin-4-fluorescein based fluorescence quenching assay for determination of biotin binding capacity of streptavidin conjugated quantum dots.
Bioconjugate chemistry, 22(3):362–368, 2011.

[120] Russell E Thompson, Daniel R Larson, and Watt W Webb.
Precise nanometer localization analysis for individual fluorescent probes.
Biophysical journal, 82(5):2775–2783, 2002.

[121] Arato Takedachi, Masafumi Saijo, and Kiyoji Tanaka.
Ddb2 complex-mediated ubiquitylation around dna damage is oppositely regulated by xpc and ku and contributes to the recruitment of xpa.
Molecular and cellular biology, 30(11):2708–2723, 2010.

[122] Martijn S Luijsterburg, Joachim Goedhart, Jill Moser, Hanneke Kool, Bart Geverts, Adriaan B Houtsmuller, Leon HF Mullenders, Wim Vermeulen, and Roel van Driel.
Dynamic in vivo interaction of ddb2 e3 ubiquitin ligase with uv-damaged dna is independent of damage-recognition protein xpc.
Journal of cell science, 120(15):2706–2716, 2007.

[123] Anne F Nichols, Phat Ong, and Stuart Linn.
Mutations specific to the xeroderma pigmentosum group e ddb- phenotype.
Journal of Biological Chemistry, 271(40):24317–24320, 1996.

[124] Vesna Rapić-Otrin, Mary P McLenigan, Dawn C Bisi, Martin Gonzalez, and Arthur S Levine.
Sequential binding of uv dna damage binding factor and degradation of the p48 subunit as early events after uv irradiation.
Nucleic acids research, 30(11):2588–2598, 2002.

[125] M. Robu, R. G. Shah, N. Petitclerc, J. Brind'Amour, F. Kandan-Kulangara, and G. M. Shah.
Role of poly(adp-ribose) polymerase-1 in the removal of uv-induced dna lesions by nucleotide excision repair.
Proc Natl Acad Sci U S A, 110(5):1658–63, 2013.

[126] Alex Pines, Mischa G Vrouwe, Jurgen A Marteijn, Dimitris Typas, Martijn S Luijsterburg, Medine Cansoy, Paul Hensbergen, André Deelder, Anton de Groot, Syota Matsumoto, et al.
Parp1 promotes nucleotide excision repair through ddb2 stabilization and recruitment of alc1.
The Journal of cell biology, 199(2):235–249, 2012.

[127] Craig D Hughes, Hong Wang, Harshad Ghodke, Michelle Simons, Atif Towheed, Ye Peng, Bennett Van Houten, and Neil M Kad.
Real-time single-molecule imaging reveals a direct interaction between uvrc and uvrb on dna tightropes.
Nucleic acids research, 41(9):4901–4912, 2013.

[128] Jason Gorman and Eric C Greene.
Visualizing one-dimensional diffusion of proteins along dna.
Nature structural & molecular biology, 15(8):768–774, 2008.

[129] Peter H von Hippel and OG Berg.
Facilitated target location in biological systems.
Journal of Biological Chemistry, 264(2):675–678, 1989.

[130] Feng Wang, Sy Redding, Ilya J Finkelstein, Jason Gorman, David R Reichman, and Eric C Greene.
The promoter-search mechanism of escherichia coli rna polymerase is dominated by three-dimensional diffusion.
Nature structural & molecular biology, 20(2):174–181, 2012.

[131] A. R. Dunn, N. M. Kad, S. R. Nelson, D. M. Warshaw, and S. S. Wallace.
Single qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along dna.
Nucleic Acids Res, 39(17):7487–98, 2011.

[132] Annett Reichel, Dirk Schaible, Natalie Al Furoukh, Mati Cohen, Gideon Schreiber, and Jacob Piehler.
Noncovalent, site-specific biotinylation of histidine-tagged proteins.
Analytical chemistry, 79(22):8590–8600, 2007.

[133] Hong Wang, Ingrid Tessmer, Deborah L Croteau, Dorothy A Erie, and Bennett Van Houten.
Functional characterization and atomic force microscopy of a dna repair protein conjugated to a quantum dot.
Nano letters, 8(6):1631–1637, 2008.

[134] Sergey Alekseev, Martijn S Luijsterburg, Alex Pines, Bart Geverts, Pierre-Olivier Mari, Giuseppina Giglia-Mari, Hannes Lans, Adriaan B Houtsmuller, Leon HF Mullenders, Jan HJ Hoeijmakers, et al.

Cellular concentrations of ddb2 regulate dynamic binding of ddb1 at uv-induced dna damage.
Molecular and cellular biology, 28(24):7402–7413, 2008.

[135] Michael A Tycon, Catherine F Dial, Keia Faison, Whitney Melvin, and Christopher J Fecko.
Quantification of dye-mediated photodamage during single-molecule dna imaging.
Analytical Biochemistry, 426(1):13–21, 2012.

[136] Mark Kastantin and Daniel K Schwartz.
Identifying multiple populations from single-molecule lifetime distributions.
ChemPhysChem, 14(2):374–380, 2013.

[137] Robert Walder, Mark Kastantin, and Daniel K Schwartz.
High throughput single molecule tracking for analysis of rare populations and events.
Analyst, 137(13):2987–2996, 2012.

[138] J Michael Schurr.
The one-dimensional diffusion coefficient of proteins absorbed on dna: Hydrodynamic considerations.
Biophysical chemistry, 9(4):413–414, 1975.

[139] M. R. Duan and M. J. Smerdon.
Uv damage in dna promotes nucleosome unwrapping.
J Biol Chem, 285(34):26295–303, 2010.

[140] JT Reardon, AF Nichols, S Keeney, CA Smith, JS Taylor, S Linn, and A Sancar.
Comparative analysis of binding of human damaged dna-binding protein (xpe) and escherichia coli damage recognition protein (uvra) to the major ultraviolet photoproducts: T [c, s] t, t [t, s] t, t [6-4] t, and t [dewar] t.
Journal of Biological Chemistry, 268(28):21301–21308, 1993.

[141] Mohamed A El-Mahdy, Qianzheng Zhu, Qi-en Wang, Gulzar Wani, Mette Prætorius-Ibba, and Altaf A Wani.
Cullin 4a-mediated proteolysis of ddb2 protein at dna damage sites regulates in vivo lesion recognition by xpc.
Journal of Biological Chemistry, 281(19):13404–13411, 2006.

[142] Qi-En Wang, Manzoor A Wani, Jianming Chen, Qianzheng Zhu, Gulzar Wani, Mohamed A El-Mahdy, and Altaf A Wani.
Cellular ubiquitination and proteasomal functions positively modulate mammalian nucleotide excision repair.
Molecular carcinogenesis, 42(1):53–64, 2005.

[143] Q. Zhao, B. M. Barakat, S. Qin, A. Ray, M. A. El-Mahdy, G. Wani, S. Arafa el, S. N. Mir, Q. E. Wang, and A. A. Wani.

The p38 mitogen-activated protein kinase augments nucleotide excision repair by mediating ddb2 degradation and chromatin relaxation.
J Biol Chem, 283(47):32553–61, 2008.

[144] M. Tsuge, Y. Masuda, H. Kaneoka, S. Kidani, K. Miyake, and S. Iijima.
Sumoylation of damaged dna-binding protein ddb2.
Biochem Biophys Res Commun, 438(1):26–31, 2013.

[145] Mitsuo Wakasugi, Hiromi Kasashima, Yuko Fukase, Mayumi Imura, Rika Imai, Saki Yamada, James E Cleaver, and Tsukasa Matsunaga.
Physical and functional interaction between ddb and xpa in nucleotide excision repair.
Nucleic acids research, 37(2):516–525, 2009.

[146] Aaron A Hoskins, Larry J Friedman, Sarah S Gallagher, Daniel J Crawford, Eric G Anderson, Richard Wombacher, Nicholas Ramirez, Virginia W Cornish, Jeff Gelles, and Melissa J Moore.
Ordered and dynamic assembly of single spliceosomes.
Science, 331(6022):1289–1295, 2011.

[147] Dror Sagi, Tsvi Tlusty, and Joel Stavans.
High fidelity of reca-catalyzed recombination: a watchdog of genetic diversity.
Nucleic acids research, 34(18):5021–5031, 2006.