

**The Structure and Mechanism of DNA Damage Recognition by XPC/Rad4
Nucleotide Excision Repair Complex**

BY

XUEJING CHEN

B.S., China Agricultural University, 2009

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry (Biochemistry)
in the Graduate College of the
University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

Dr. Jung-Hyun Min, Chair and Advisor
Dr. Wonhwa Cho, Department of Chemistry
Dr. Lawrence Miller, Department of Chemistry
Dr. Tim Keiderling, Department of Chemistry
Dr. Xiaojing Yang, Department of Chemistry
Dr. Anjum Ansari, Department of Physics

*The thesis is dedicated to my family, colleagues and friends,
without them it would never have been accomplished.*

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Jung-Hyun Min for all the guidance, encouragement, and inspiration throughout the years. I thank all the Min group members, especially Dr. Beomseok Park for training me, and Yoonjung Shim.

I would like to thank my committee members: Dr. Wonhwa Cho, Dr. Lawrence Miller, Dr. Tim Keiderling, Dr. Xiaojing Yang and Dr. Anjum Ansari for their help. And especially Dr. Ansari and her group members, particularly Yogambigai Velmurugu, for the wonderful collaboration.

Lastly I thank my friends for their support.

CONTRIBUTION OF AUTHORS

Chapter 1 introduces my research topic and its significance.

Chapter 2 represents part of a published paper (*Nat Commun* **6**, 5849 (2015)) for which I was co-first author. Dr. Beomseok Park and Yoonjung Shim contributed to protein purification and crystallization. Guanqun Zheng and her advisor Dr. Chuan He synthesized the DNA strand for crosslinking. Dr. Jung-Hyun Min contributed to crystal diffraction data collection, and built and refined the model with contribution from Dr. Youngchang Kim. Dr. Lili Liu and her advisor Dr. Bennett Van Houten performed AFM experiments.

The first section of chapter 3 represents the rest of the paper mentioned above. Yogambigai Velmurugu performed equilibrium and Tjump experiments and analyzed the data. Dr. Anjum Ansari and Dr. Jung-Hyun Min were advisors and wrote the manuscript.

The second section of chapter 3 represents a paper for which I am co-first author. It is currently in press for *PNAS*. Yogambigai Velmurugu performed T-jump experiments and analyzed the data. Phillip Slogoff Sevilla performed equilibrium experiments. Dr. Anjum Ansari and Dr. Jung-Hyun Min were advisors and wrote the manuscript.

Chapter 4 describes my unpublished work towards solving the structure of Rad4-Rad23-Rad33 trimeric complex.

Chapter 5 describes my unpublished work to label Rad4 with a fluorophore at specific site for single-molecule microscopic studies. Dr. Xinghua Shi and his advisor Dr. Taekjip Ha provided protocol and some reagents.

TABLE OF CONTENTS

<u>Chapter</u>		<u>Page</u>
I.	Introduction	1
A.	DNA damage and maintaining genome stability	1
B.	Nucleotide excision repair (NER)	5
C.	XPC, the initiator of NER	8
	1. The DNA substrates of XPC	8
	2. The domains of XPC complex.....	11
	3. Damage recognition by XPC	13
	4. Regulation of XPC	17
	5. XPC beyond NER	17
II.	The crystal structure of Rad4-Rad23 in complex with undamaged	
DNA	19	
A.	Introduction	19
B.	Materials and Methods.....	19
	1. Cloning.....	19
	2. Virus production	20
	3. Protein expression and purification	20
	4. DNA synthesis	21
	5. Crosslinking reaction and purification	21
	6. Crystallization and structure determination	22
	7. EMSA.....	23
	8. DNA sequence.....	23
C.	Results.....	27

D.	Discussion	29
III.	The DNA opening and twisting kinetics induced by Rad4–Rad2337	
A.	Introduction	37
B.	Materials and Methods.....	40
1.	Cloning.....	40
2.	Virus production	40
3.	Protein expression and purification	40
4.	Dynamic light scattering (DLS).....	41
5.	Equimolar electrophoretic mobility shift assay	41
6.	Electrophoretic mobility shift assay with excess protein.....	41
7.	Equilibrium temperature scan experiment.....	42
8.	T-jump kinetics measurement.....	43
9.	Melting temperature (T _m) measurement.....	43
10.	DNA sequence.....	44
C.	Result.....	53
1.	Examination of Rad4-Rad23 complex and DNA bound state in solution	53
2.	The fluorescence intensity of 2AP constructs	57
3.	The kinetics of Rad4 bound 2AP constructs	58
4.	The kinetics of Rad4 bound tC ^O /tC _{nitro} constructs	65
D.	Discussion	79
IV.	Structural study of Rad4-Rad23-Rad33 trimeric complex.....	85

A.	Introduction	85
B.	Materials and Methods.....	86
1.	Cloning of Rad4-Rad23 complex and Rad33/Cdc31	86
2.	Cloning of trimeric complex in MultiBac system	86
3.	Virus production	88
4.	Expression test using NE-PER kit (Thermo Scientific) and GST pull-down	88
5.	Large scale protein purification	89
6.	Dialysis of protein–DNA complex.....	90
7.	Crystallization.....	90
C.	Preliminary results and Discussion	97
V.	Site-specific fluorescence labeling of Rad4-Rad23	101
A.	Introduction	101
B.	Materials and Methods.....	103
1.	Cloning.....	103
2.	Virus production	104
3.	Expression and purification of XC65	104
4.	Expression of FGE and co-expression of DinB and FGE.....	109
5.	Cy3 labeling	110
C.	Results.....	111
1.	Labeling of DinB.....	111
2.	Labeling of Rad4-Rad23	114

D. Discussion	118
VI. Conclusion	121
Appendix	123
A. Protocols	123
1. PCR	123
2. Restriction enzyme digestion and ligation	124
3. Transformation and colony identification	125
4. Transposition and bacmid extraction	126
5. Transfection and Virus production	128
6. Protein expression	129
7. Large scale protein purification	130
8. Electrophoretic mobility shift assay and K_d determination	132
B. Copyright Permissions	135
Cited Literature	137
VITA 143	

LIST of TABLES

<u>Table</u>	<u>Page</u>
Table 1. DNA damage and repair pathways	4
Table 2. A summary of protein constructs used in chapter 2.....	20
Table 3. The sequence of DNA constructs used in chapter 2.	23
Table 4. Crystal diffraction data collection and refinement.	33
Table 5. The sequence of DNA constructs containing 2AP used in chapter 3.	44
Table 6. The sequence of tC ^O / tC _{nitro} DNA constructs used in chapter 3.	45
Table 7. A summary of measured and predicted FRET for all DNA and complexes.	73
Table 8. A summary of protein constructs used in chapter 4.....	87

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 1. Eukaryotic GG-NER pathway.	7
Figure 2. Examples of different substrates of XPC.	10
Figure 3. The domains of hXPC/yRad4 and hRad23B/yRad23.	12
Figure 4. The structure of Rad4 bound to damaged DNA [36]	15
Figure 5. Lysis, affinity column and SourceQ purification of SC32.	24
Figure 6. Purification of SC32 by SourceS.	25
Figure 7. Purification of SC32 by Superdex200.	26
Figure 8. Purification of crosslinked complex and crystallization.	27
Figure 9. EMSA of full-length, truncated Rad4 with point mutation, and the mutants....	31
Figure 10. The structure of Rad4 crosslinked to match DNA.	32
Figure 11. Comparison of Rad4 bound to damaged DNA and undamaged DNA.	34
Figure 12. AFM study of Rad4 non-covalently bound to match DNA.	35
Figure 13. The FRET pair tC^O and tC_{nitro}	39
Figure 14. Lysis, affinity column and SourceQ purification of <108>.	46
Figure 15. Purification of <108> by SourceS.	47
Figure 16. Purification of <108> by Superdex200.	48
Figure 17. Lysis and Ni pull-down of <140>.	49
Figure 18. Purification of <140> by SourceQ.	50
Figure 19. Purification of <140> by SourceS.	51
Figure 20. Purification of <140> by Superdex200.	52
Figure 21. Dynamic light scattering (DLS) of Rad4-Rad23-DNA complex.	55
Figure 22. EMSA of Rad4-Rad23 binding to DNA showing 1:1 protein:DNA ratio.	56

Figure 23. The summary of intensity of all the single strand DNA containing 2AP.....	60
Figure 24. The summary of intensity of all constructs containing 2AP.	61
Figure 25. The fluorescence spectra, equilibrium, and T-jump trace of AN3 and AN4 bound to Rad4.	62
Figure 26. The fluorescence spectra, equilibrium, and T-jump trace of AN3 and AN4 bound to Rad4 mutants.	63
Figure 27. The fluorescence spectra, equilibrium, and T-jump trace of AN21 bound to Rad4.	64
Figure 28. The relaxation rates of Rad4 bound DNA.	65
Figure 29. Melting temperature of AN12 and AN14 constructs.	69
Figure 30. EMSA of Rad4 binding to AN12 constructs.	70
Figure 31. EMSA of Rad4 binding to AN14 constructs.	71
Figure 32. FRET of AN12 and AN14.	72
Figure 33. The equilibrium and T-jump trace of AN12 alone and bound to Rad4.	74
Figure 34. The equilibrium and T-jump trace of AN12 bound to Rad4 mutants.	75
Figure 35. The equilibrium and T-jump trace of AN12u alone and bound to Rad4.	76
Figure 36. The equilibrium and T-jump trace of AN14 and AN14u bound to Rad4.	77
Figure 37. The equilibrium and T-jump trace of AN14 alone and AN14u alone.	78
Figure 38. Lysis, affinity column and SourceQ purification of XC54.	91
Figure 39. Purification of XC54 by SourceS.	92
Figure 40. Purification of XC54 by Superdex200.	93
Figure 41. Purification of XC54 after thrombin digestion by SourceS.	94
Figure 42. Purification of XC54 without UBL by Superdex200.	95

Figure 43. Purification of XC54 with UBL by Superdex200.	96
Figure 44. Constructing Rad4 trimeric complex and crystallization.	100
Figure 45. Site-specific fluorescence labeling of Rad4-Rad23 for single-molecule FRET study.	103
Figure 46. Lysis and Ni pull-down of XC65.	105
Figure 47. Purification of XC65 by SourceQ.	106
Figure 48. Purification of XC65 by SourceS.	107
Figure 49. Purification of XC65 by Superdex200.	108
Figure 50. Co-expression of DinB and FGE, and test of FGE expression.	112
Figure 51. Cy3 labeling of DinB.	113
Figure 52. Expression of XC65 and Cy3 labeling.	116
Figure 53. Validation of Cy3 labeling XC65.	117

LIST OF ABBREVIATIONS

2AP	2-aminopurine
6-4 PP	6-4 photoproduct
AAF	2-acetylaminofluorene
AFM	atomic force microscopy
AP	apurinic/apyrimidinic
AGT	O ⁶ -alkylguanine-DNA alkyltransferase
BER	base excision repair
BHD	β-hairpin domain
Cen2	Centrin 2
CPD	cyclobutane pyrimidine dimer
CS	Cockayne syndrome
DSB	double-strand break
FA	Fanconi Anemia
FGE	formyl-glycine generating enzyme
EMSA	electrophoretic gel mobility shift assay
FRET	Förster resonance energy transfer
GG-NER	global-genome NER
HR	homologous recombination
ICL	interstrand crosslinking
MMR	mismatch repair
NER	nucleotide excision repair
NHEJ	non-homologous end joining
IR	ionizing radiation
OGG1	8-oxoguanine glycosylase
PAH	polycyclic aromatic hydrocarbon
QY	quantum yield
R4BD	Rad4-binding domain
ROS	reactive oxygen species
SCID	Severe combined immunodeficiency
tC	tricyclic cytosine
T _m	melting temperature
TFIIH	transcription factor IIH
TGD	transglutaminase-homology domain
T-jump	temperature jump
TLS	translesion synthesis
TC-NER	transcription-coupled NER
UBA	ubiquitin-associated domain
UBL	ubiquitin-ike domain

UNG	uracil-DNA glycosylase
UV	ultra-violet
XP	Xeroderma Pigmentosum
XPC	Xeroderma Pigmentosum Complementation Group C

SUMMARY

Nucleotide excision repair (NER) is an important pathway that repairs many structurally diverse DNA lesions in the cell, to maintain genome stability. XPC-Rad23B-Cen2 (yeast ortholog Rad4-Rad23-Cdc31/Rad33) complex initiates NER by finding and binding to lesions in the genome, and then recruits downstream NER factors. The crystal structure of Rad4-Rad23 bound to damaged DNA revealed that Rad4 recognizes lesions in an indirect manner, and forms a stable ‘open’ complex with DNA. But there is still much unknown about the function and mechanism of XPC/Rad4 complex, such as how it is able to efficiently distinguish lesions embedded in large excess of normal DNA.

In this dissertation, the mechanism of damage recognition of Rad4 was investigated first by solving crystal structure of Rad4 complex bound to undamaged DNA. To prevent Rad4 binding non-specifically in multiple registers which inhibits crystallization, the protein was covalently tethered to DNA. This crystal structure revealed the same ‘open’ conformation as the damaged DNA bound structure.

Next, the kinetics of DNA base opening and twisting were studied by temperature jump spectroscopy. First, using 2-aminopurine (2AP) as a probe, a relaxation rate of ~7 ms was observed, which represented full nucleotide flipping as the open conformation forms. This was specific to damaged DNA, and required the β -hairpin which was seen to insert into DNA duplex in the crystal structure (β -hairpin3). Second, using a novel FRET pair tC^O/tC_{nitro} as probes, two distinct phases of kinetics were observed: The slow phase, which overlapped on the same time scale as the nucleotide flipping kinetics captured by 2AP, was observed only when bound to specific, damaged DNA, but did not require β -

hairpin3, thus representing the rate-limiting step during the formation of the open conformation. On the other hand, the fast phase did not require a DNA lesion or β -hairpin3, and represents a nonspecific interrogation step preceding open conformation formation.

Taken altogether, a “kinetic gating” mechanism for lesion recognition has been proposed, where Rad4/XPC complex interrogates the DNA for distortion in duplex structure by fast twisting, and becomes “trapped” at lesion site, possibly due to an increase in residence time, leading to fully opening the DNA and stable protein-DNA complex formation. This hypothesis could explain how Rad4/XPC is able to detect thermodynamically destabilized lesions without recognizing specific structure of the lesion.

I. INTRODUCTION

A. DNA damage and maintaining genome stability

One of the most crucial tasks of a cell is to cope with the large amount of DNA damage caused by various sources. The number of lesions that appear in one human cell per day is approximately on the order of 10^5 [1]. They could be generated by exogenous/environmental sources, such as sunlight, ionizing radiation (IR), and carcinogenic chemicals, as well as endogenous/physiological sources, such as the erroneous activities of DNA polymerases, topoisomerases, and products of metabolism such as reactive oxygen species (ROS). If the lesions are not repaired, they cause mutations or block important biological processes. The resulting instability of the genome is tightly linked to aging and cancer (Reviewed by Hoeijmakers [2]).

The cell has developed several repair pathways to maintain genome stability. Table 1 briefly summarizes the repair pathways and the major activities of each. The diseases caused by failure in DNA repair clearly demonstrate the biological importance of these pathways.

Mismatch repair pathway (MMR) : During DNA synthesis, about 1 error arises in 10^8 bases. Insertion/deletion loops can be generated from microsatellite repeats. These are repaired by the **mismatch repair pathway (MMR)** (Reviewed by Jiricny [3] and Modrich & Lahue [4]). A homodimer composed of MutS and MutL recognizes the newly synthesized strand and detects the mismatch. Then the new strand is digested by exonuclease ExoI, and the gap is filled by DNA polymerase δ and DNA ligase.

Base excision repair (BER): BER is responsible for fixing damaged bases generated by many spontaneous reactions such as oxidation, deamination, alkylation, hydrolysis or depurination (reviewed by Krokan & Bjoras [5]). In mammalian systems, these lesions are recognized by 11 DNA glycosylases, each with certain specificity. For example, 8-oxoguanine glycosylase (OGG1) detects 8-oxoguanine, an oxidative product; uracil-DNA glycosylase (UNG) recognizes uracil, which is a product of deamination. These glycosylases flip out and excise the damaged base, leaving an abasic site, which is then processed by apurinic/aprimidinic (AP)-endonuclease, and followed by gap filling and ligation.

Direct reversal: Some base damage can be repaired by **direct reversal**. For example, UV lesions can be directly repaired by photolyase, although placental mammals do not have this enzyme (Reviewed by Sancar [6]). For humans, O⁶-alkylguanine-DNA alkyltransferase (AGT) and ABH family proteins repair alkylated bases (Reviewed by Mishina *et al.* [7]). These enzymes bind to specific substrate and also flip out the base [8] [9], then directly reverse the base to normal without cutting the DNA backbone.

Nucleotide excision repair (NER): NER repairs bulky lesions generated by ultraviolet (UV) light and environmental toxins, as well as DNA intra/interstrand crosslinking, generated by cisplatin and other anti-cancer drugs used in chemotherapy. This process will be discussed further in detail in the following section.

Translesion synthesis (TLS): When a replicating DNA polymerase is blocked by damage, it can be switched out by certain special polymerases to carry out **translesion synthesis (TLS)**, which continue to incorporate nucleotides on the opposite strand

ignoring the lesion (reviewed by Lehmann *et al* [10]). These are done by several polymerases from the Y-family and B-family, each with their own specificity for tolerating certain lesions.

Double-strand break repair: Double strand break (DSB), caused by ionizing radiation (IR) and other factors, is the most deleterious lesion, due to the absence of template for repair (reviewed by Kanaar *et al.* [11] and Khanna *et al.* [12]). When a sister chromatid is available, it could be used as template to repair the DSB by **homologous recombination (HR)**; in other cases, the repair is carried out by **non-homologous end joining (NHEJ)**, which simply joins the two ends back together.

Table 1. DNA damage and repair pathways

Type of damage	Source of damage	DNA repair pathway	Examples of human disease/symptom caused by defect in repair
Mismatch, insertion/deletion loops	DNA polymerase errors	Mismatch repair (MMR)	Predisposition to non-polyposis colorectal cancer and endometrial carcinomas
Base damage such as oxidation, deamination, alkylation, hydrolysis	Spontaneous decay of DNA, reactive oxygen species (ROS)	Direct damage reversal	Predisposition to sporadic colorectal cancer
		Base excision repair (BER)	Premature aging, spontaneous tumor
Bulky helix-distorting lesions	Ultraviolet (UV) light, environmental toxins	Nucleotide excision repair (NER)	Xeroderma pigmentosum (XP), Cockayne syndrome (CS), trichothiodystrophy
Intra/interstrand crosslinking (ICL)*	Cisplatin, mitomycin C and some other anticancer drugs		Fanconi anemia (FA)
Base damage, bulky lesions encountered during replication	Spontaneous decay of DNA, UV	Translesion synthesis (TLS)	Xeroderma pigmentosum-variant
Double-strand break (DSB)	Ionizing radiation (IR)	Nonhomologous end joining (NHEJ)	Severe combined immunodeficiency (SCID), radiosensitivity
		Homologous recombination (HR)	Werner syndrome, Bloom syndrome, breast cancer, ovarian cancer

*Intrastrand crosslinking is repaired by NER. Interstrand is repaired by a more complicated process involving NER proteins, Fanconi anemia complementation group (FA) proteins and HR.

B. Nucleotide excision repair (NER)

NER is a unique repair pathway, able to correct a wide variety of structurally diverse damaged DNA substrate. Eukaryotic NER is a highly conserved process and can be categorized into two sub-pathways.

During transcription, an RNA polymerase II blocked by lesion will trigger NER, which is called **transcription-coupled NER (TC-NER)** (reviewed by Vermeulen & Foustieri [13]). Deficiency in this sub-pathway can cause Cockayne syndrome, which manifests symptoms such as UV sensitivity, stunted physical and mental development.

When the lesion is not on actively transcribed strand, it is repaired by **global-genome NER (GG-NER)** (reviewed by Gillet & Scharer [14]). More than 30 proteins are involved in the process, and six of them are considered essential. With loss of function of any one of them, NER will be abolished, which can cause diseases such as xeroderma pigmentosum (XP) syndrome. XP patients are extremely sensitive to sunlight, due to the failure to repair UV lesions by NER. They suffer high mutation rates in the genome, and 1000-fold higher chance to develop skin cancers [15]. Hence the six core NER proteins are named xeroderma pigmentosum complementation groups.

XPC is the initiator of GG-NER and is responsible for the substrate versatility of NER (discussed in detail below). In complex with Rad23B and Centrin2 (Cen2), XPC searches in the genome and binds to the lesion. Then it recruits transcription factor IIH (TFIIH), whose helicase subunits, XPD and XPB, further unwind the DNA using ATP hydrolysis energy. This is irreversible and acts as a proofreading step. The verification is followed by the assembly of pre-incision complex, including: RPA, a single-strand DNA

binding protein, which binds the undamaged strand; XPA, which works with RPA to stabilize the pre-incision complex; XPG, an endonuclease, recruited by TFIIH. At this point XPC leaves the pre-incision complex. Lastly ERCC1-XPF complex is recruited by XPA. XPF and XPG make incision on 5' and 3' of the lesion, respectively, producing a single-strand fragment of characteristic 24 – 32 nucleotides long. Finally the gap is filled by polymerase δ/ϵ and sealed by DNA ligases. (Figure 5)

Prokaryotic NER uses similar mechanism but different components, mainly UvrABC proteins (reviewed by Truglio *et al.* [16]). In bacteria, UvrA and UvrB form a heterotrimer or heterotetramer and scan the genome. UvrA recognizes the lesion and load UvrB onto DNA. Then UvrA dissociates, leaving UvrB forming pre-incision complex with DNA. UvrC is recruited to the complex and makes incision on both sides.

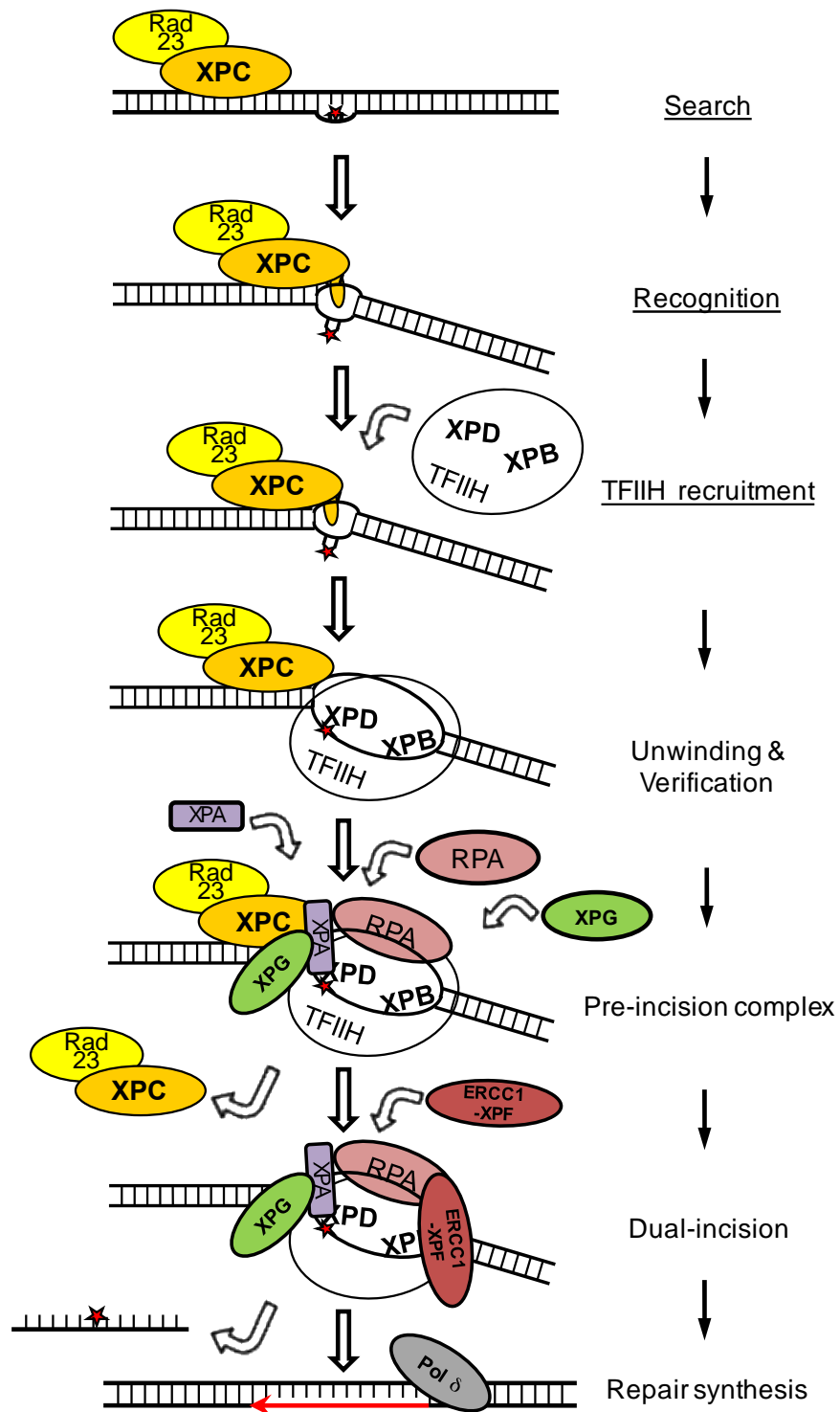


Figure 1. Eukaryotic GG-NER pathway.

C. XPC, the initiator of NER

1. The DNA substrates of XPC

DNA crosslinking agents, such as cisplatin and its derivatives, are one of the most effective and common chemotherapy drugs. About 90% of the lesions generated by cisplatin is **intrastrand crosslinks** [17] (formed between 2 bases on the same strand), which is a substrate of XPC and NER. On the other hand, interstrand crosslinks (formed between 2 bases on the complementary strands) only takes up a small percentage, and their repair is done by a distinct interstrand crosslink repair (ICL) pathway (reviewed by McHugh *et al.* [18], Deans & West [19]). The current model involves the coordination of FA proteins, NER proteins, HR and translesion synthesis.

UV is a common source of DNA lesions. A cell exposed in strong sunlight receives $\sim 10^5$ lesions in 1 hour [20]. UV produces cyclobutane pyrimidine dimer (CPD) ($\sim 75\%$) and 6-4 photoproduct (6-4 PP) ($\sim 25\%$) [21], formed between 2 neighboring pyrimidines, both of which could disrupt transcription or replication. NER is the main pathway, and for human, the only pathway to repair them (reviewed by Sinha & Hader[22]).

NER is also responsible for removing many other bulky lesions. For example, **polycyclic aromatic hydrocarbons (PAH)**, present in motor vehicle emission, industrial exhaust, cigarette, and other environmental pollutant, form adducts with DNA. Benzo[*a*]-pyrene, a potent carcinogen from cigarette smoke, has been extensively studied as a PAH prototype (reviewed by Gelboin [23]). Another example is DNA adduct of **aromatic**

amine (reviewed by Turesky [24]), found in house-cooked meat and has been suggested to contribute to several types of cancers. A widely studied representative of these compounds is 2-acetylaminofluorene (AAF).

XPC binds to 3–5 bp mismatch bubble in DNA, although a mismatch will not pass the verification step and can not be excised by NER [25]. Finally, XPC also binds to undamaged DNA strongly (nanomolar affinity) although XPC prefer damaged DNA to normal DNA.

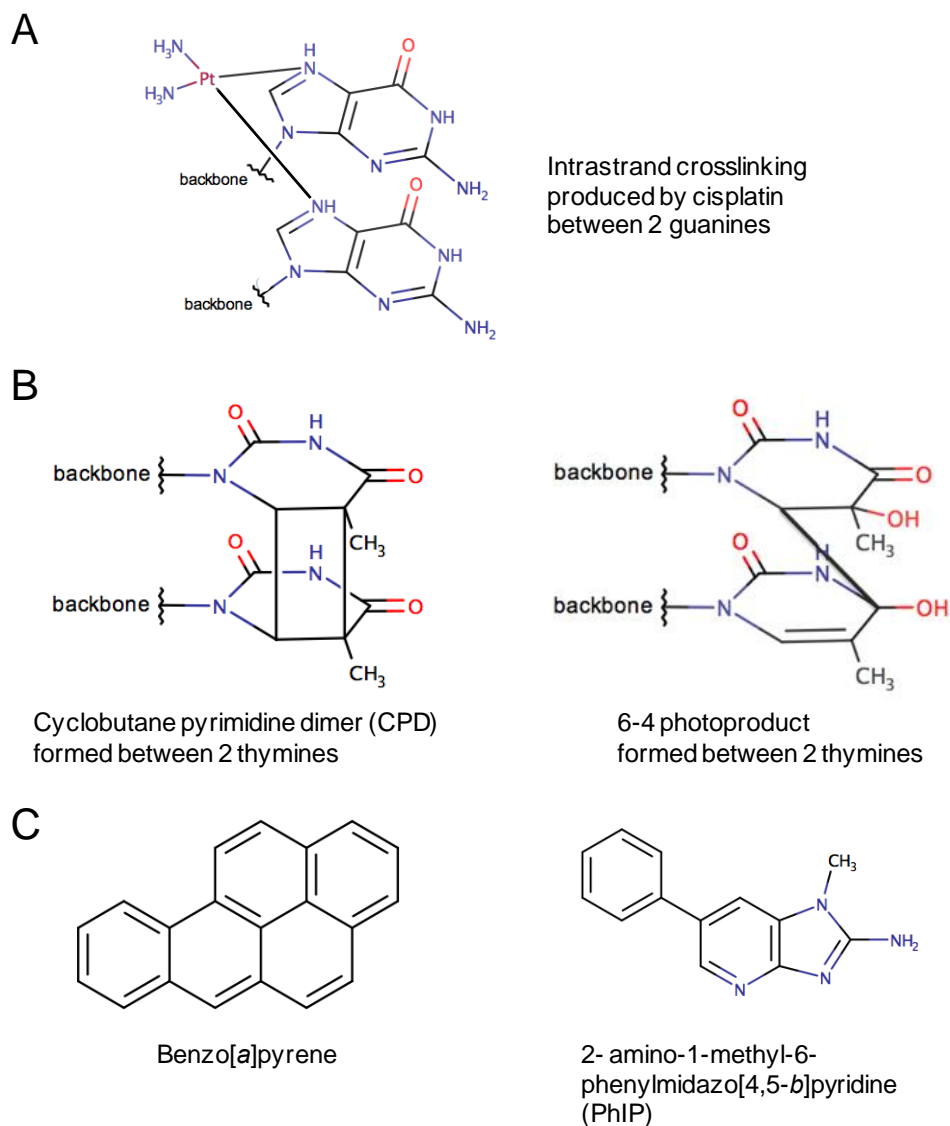


Figure 2. Examples of different substrates of XPC.

A. Intrastrand crosslinking produced by cisplatin.

B. UV lesions.

C. Examples of other chemicals that can form bulky DNA adduct: benzo[a]pyrene, a member of polycyclic aromatic hydrocarbons; 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP), a member of aromatic amines.

2. The domains of XPC complex

The human XPC is total 940-residues long. The N-terminus consists of transglutaminase-fold with a long loop in the middle [26]. Region 154 – 331 in the N-terminus was shown to interact with XPA [27]. The C-terminus contains many important functional domains: residues 495 – 734 is required to bind Rad23B at least in humans; residues 606 – 742 is required to bind to DNA; residues 816 – 940 is responsible for recruiting and binding TFIIH [28]. The Cen2 binding site is mapped to 847 – 863, which is within TFIIH binding site [29]. The C-terminal domain is highly conserved [30] and shares a high level of homology with Rad4, the yeast ortholog of XPC [31]. Homology in the TGD domains in XPC and Rad4 was also detected [36].

Due to the difficulty of producing stable XPC protein, yeast Rad4 is used in most of the studies reported here. The N-terminus of Rad4 contains a transglutaminase-homology domain (TGD), and the C-terminus contains 3 consecutive β -hairpin domains (BHD1 – 3). In the studies reported here, the flexible segments on the extreme of both termini have been trimmed for best stability.

The binding partner, human Rad23B and yeast Rad23 are very similar. On the N-terminus is the ubiquitin-like domain (UBL), which interacts with proteasome [32]. The 2 ubiquitin-associated domains (UBA1 and UBA2) are suggested to inhibit proteolysis by competing with proteasome to bind ubiquitin chain [33]. They flank a XPC or Rad4 binding domain (R4BD) [34]. Through these domains Rad23 could protect Rad4 from degradation [35]. In the constructs used here, UBA1 is removed, and UBL is removed for crystallization but retained for other studies.

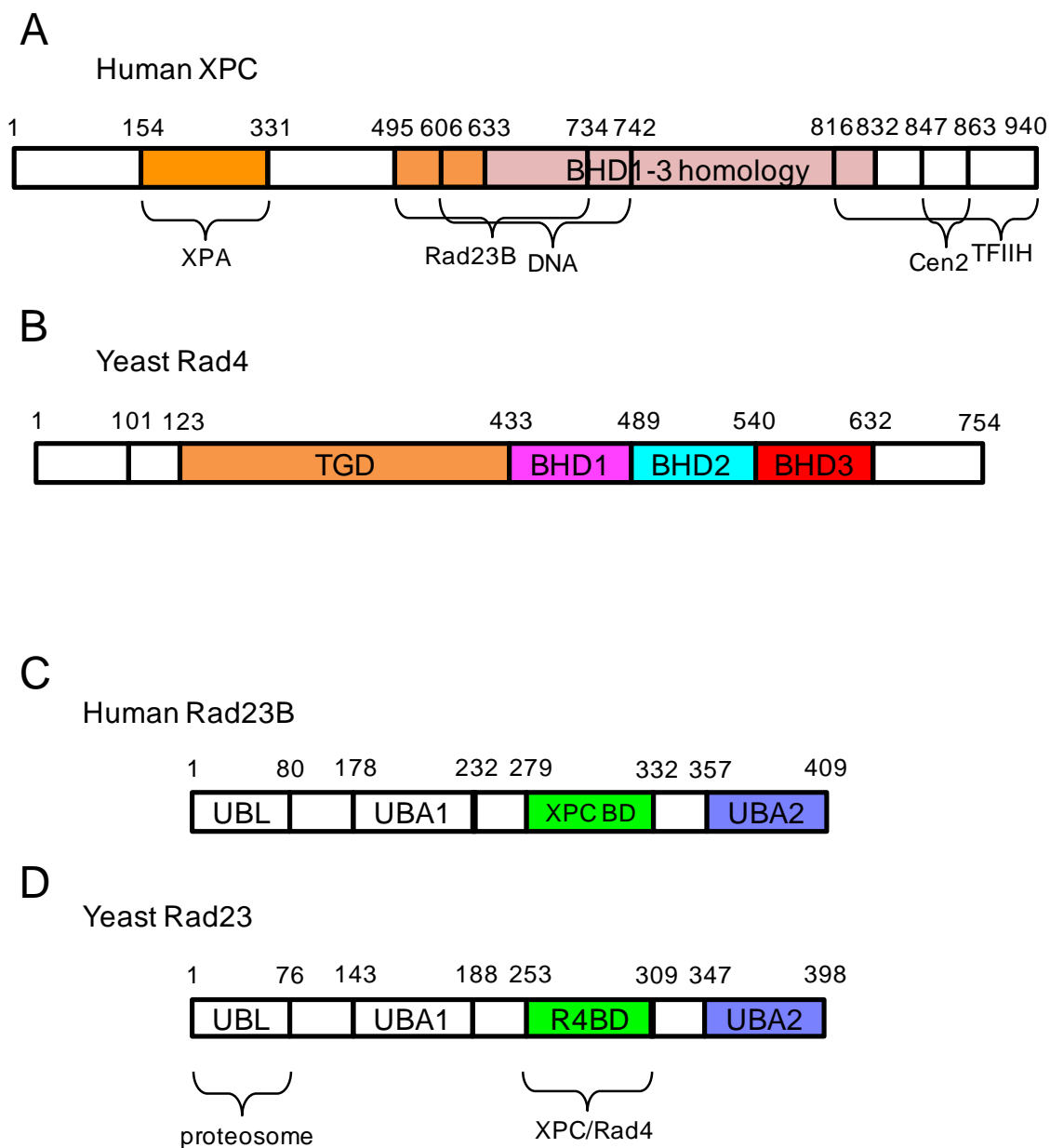


Figure 3. The domains of hXPC/yRad4 and hRad23B/yRad23.

3. Damage recognition by XPC

The mechanism of lesion binding was revealed by the crystal structure of yeast Rad4-Rad23 complex bound to a 24-bp DNA with TTT/TTT mismatch bubble and CPD lesion between T16 and T17 on the bottom strand [36]. Whereas CPD is a poor substrate for Rad4, incorporating CPD into a mismatch site can enhance Rad4's binding specificity [25]. Compared with apo-protein structure (Figure 4A), the β -hairpin domains in the DNA bound structure rotate to come to contact the DNA. TGD and BHD1 bind the backbone of undamaged portion non-specifically on 3' side of the lesion, while BHD2 and BHD3 interacts with the lesion site specifically by binding the opposite strand of the CPD lesion. The tip of β -hairpin in BHD3 (β -hairpin 3) inserts into the DNA duplex and flips out the two nucleotides containing the lesion. BHD2 and BHD3 make extensive interaction with the flipped out bases on the undamaged strand, but do not specifically contact the damaged bases (Figure 4B). The β -hairpin is a common strategy used by several NER proteins including Rad4, DDB2 [37], UvrB [38] and XPA [39], underlining its importance in damage detection. This structure explained how Rad4 is able to bind diverse DNA lesion structures by interacting with the undamaged strand, rather than the lesion itself.

It has been shown that XPC has low affinity to the lesions that do not distort DNA significantly. For example, XPC almost does not distinguish CPD from undamaged DNA, but binds to 6-4PP with much higher affinity [40, 41], and 6-4PP bends the DNA by $\sim 44^\circ$ while CPD only bends DNA slightly by $\sim 9^\circ$ [42]; a benzo[a]pyrene diol epoxide adduct embedded in DNA is recognized with high affinity, but the same lesion with an

opposite abasic site, which stabilizes the duplex structure, evades NER completely [43]. These evidences indicate that XPC/Rad4 binds lesions by detecting structural distortion or destabilization induced by the lesion, albeit to a varying degree [44].

Another damage detector of NER, UV-damaged DNA-binding protein (UV-DDB), a complex of DDB1- DDB2 (XPE), is able to bind UV lesions with higher affinity than XPC [45]. Unlike XPC, DDB2 flips out the two damaged bases into its binding pocket, thus recognizing the lesion specifically [37]. Loss of DDB2 causes mild NER deficiency and slower removal of UV lesion [46], and it was postulated that DDB2 facilitates damage detection by binding to lesions such as CPD that XPC has poor affinity to then hands off the lesion to XPC [47], especially in the context of chromatin [48].

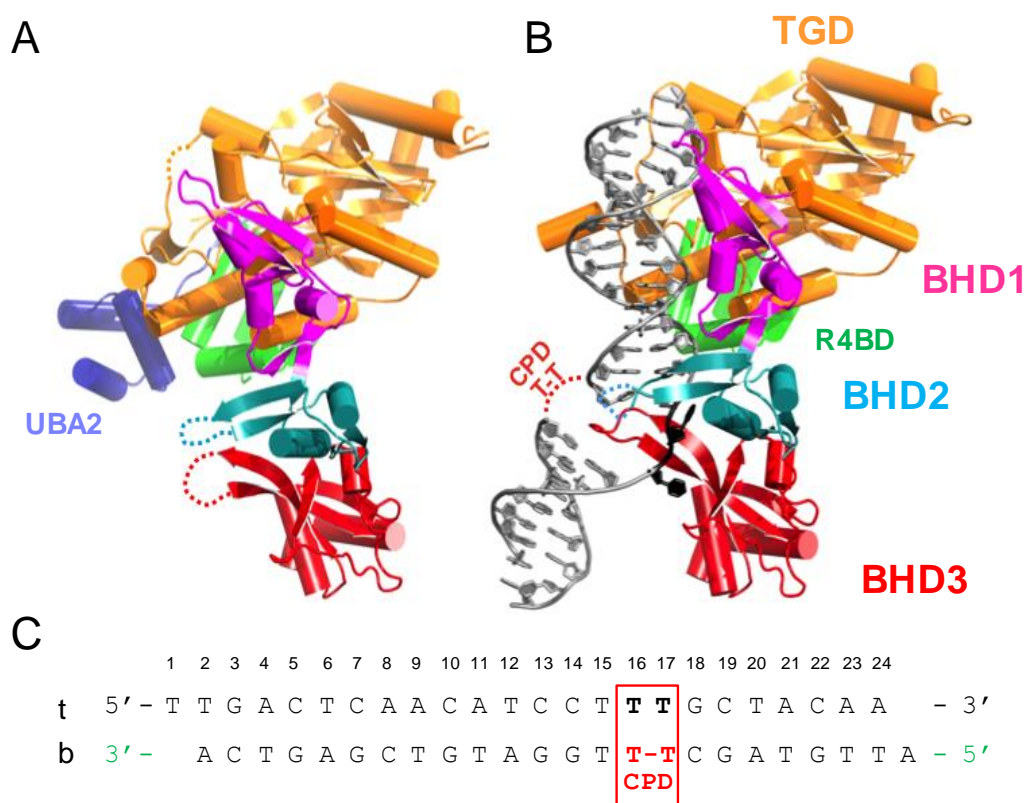


Figure 4. The structure of Rad4 bound to damaged DNA [36] .
 A. The structure of Rad4-Rad23. B. The structure of Rad4-Rad23 bound to 24-bp DNA containing a TTT/TTT mismatch with CPD. C. The sequence of DNA in the structure. Red box indicates the flipped out bases.

4. Regulation of XPC

Post-translational modification is extensively involved in the regulation of XPC. Upon UV irradiation, XPC is ubiquitinated by a ubiquitin ligase (E3) complex containing cullin4A, UV-DDB (DDB1-DDB2) proteins. Instead of degradation, however, the ubiquitination enhances the DNA binding affinity of XPC [49]. XPC is also modified with SUMO (small ubiquitin-related modifier) proteins. Failure of SUMOylation leads to DDB2 abnormally retained at damage site and NER defect *in vivo* [50]. In addition, SUMOylated XPC can be ubiquitinated by RNF111, a SUMO-targeted ubiquitin ligase, promoting NER efficiency [51]. There is also evidence indicating XPC may be phosphorylated by ATM/ATR, but details about XPC phosphorylation is yet to be discovered [52]. Although XPC and DDB2 have to bind to DNA lesion, if they are not removed readily by p97 segregase after recognition, the retention of the damage sensors on chromatin causes impaired NER, suggesting that XPC is tightly and temporally regulated [53].

5. XPC beyond NER

XPC has been heavily implicated in BER. *Xpc* defective cells displayed increased sensitivity for oxidative stress, but *Xpa* defective cells did not, indicating a role for XPC outside NER [54]. XPC has been found to physically interact with TDG, and stimulate its release from abasic site [55]. XPC also stimulates OGG1, possibly by promoting loading and turnover of OGG1 [56]. In addition, XPC, but not other NER protein, was shown to

co-localized with non-NER substrate 8-oxoG [57]. These strongly suggest that XPC may function in BER as a damage sensor or co-factor.

DNA damage triggers DNA damage response (DDR), a pathway to maintain genome stability by inducing DNA repair, cell cycle arrest and apoptosis. Lie in the center of this signaling network are the kinases ATM (Ataxia telangiectasia mutated) and ATR (Ataxia telangiectasia- and Rad3-related). XPC participates in DDR by interacting with ATM and ATR, and plays crucial role in ATM/ATR recruitment to damage site and subsequent phosphorylation of ATM/ATR substrates [58].

The yeast ortholog Rad4-Rad23 was shown to interact with SWI/SNF chromatin remodeling complex, and the interaction was enhanced upon UV irradiation. Based on the evidence that SWI/SNF increased accessibility to nucleosomal DNA, it was hypothesized that Rad4-Rad23 recruits SWI/SNF to facilitate repair proteins accessing damage site [59]. Outside DNA damage signaling and repair, XPC complex was identified as coactivator of transcriptional activators Oct4 and Sox2, which was required for embryonic cell pluripotency [60].

II. THE CRYSTAL STRUCTURE OF RAD4-RAD23 IN COMPLEX WITH UNDAMAGED DNA

A. Introduction

The task of XPC/Rad4 is to find DNA lesions within the whole genome. The human genome contains over 3 billion nucleotide pairs, while there are only $\sim 4 - 8 \times 10^4$ XPC molecules in one cell [61], which means each molecule has to find the lesions embedded within a large excess of undamaged DNA. To understand how XPC can quickly distinguish lesion versus normal base while scanning the genome, we solved a crystal structure of Rad4 bound to undamaged DNA.

Rad4 binds to undamaged DNA non-specifically, which would cause the protein to bind in multiple registers on undamaged DNA. Such a heterogeneous mixture of protein-DNA would prevent crystallization. To overcome this problem, we introduced the strategy of covalently tethering protein on DNA [62]. The desired base of DNA is modified with a thiol-linker, which forms a disulfide bond with a cysteine residue in the protein of interest, covalently trapping the protein to a specific position on DNA.

B. Materials and Methods

1. Cloning

Rad4₁₀₁₋₆₃₂ was cloned from yeast genomic library, with V131 mutated to C and C132 mutated to S by PCR, and inserted between BssHII and NotI in pFBD vector (Invitrogen), with a His tag engineered between BamHI and BssHII. Full-length Rad23

was cloned from yeast genomic library, with amino acids 135–299 deleted and replaced by a thrombin recognition sequence (Rad23_1-398_Δ135-299/Thr), and inserted between NheI and SphI in the same vector. This construct is called SC32.

Some protein constructs with variations in Rad23 sequence were also made. A thrombin recognition sequence was inserted between amino acids 346 and 347 in Rad23_1-398_Δ135-299/Thr to remove UBA2 domain after purification. This makes construct SC37. Alternately, Rad23_1-398_Δ135-299/Thr was truncated at amino acid 366 to delete UBA2 after its first helix. This makes construct SC49.

Table 2. A summary of protein constructs used in chapter 2.

Construct name	Insert 1 (ph promoter)	Insert 2 (p10 promoter)	Vector
SC32	(His-)Rad4_101-632_V131C_C132S	Rad23_1-398_Δ135-229/Thr	pFastBac Dual
SC37	(His-)Rad4_101-632_V131C_C132S	Rad23_1-398_Δ135-229/Thr Thr(346-347)	pFastBac Dual
SC49	(His-)Rad4_101-632_V131C_C132S	Rad23_1-366_Δ135-229/Thr	pFastBac Dual

2. Virus production

See protocol d and e.

3. Protein expression and purification

See protocol f and g. The complex eluted around 190 – 240 mM NaCl on Source Q. The fraction containing the protein was digested with thrombin. Then the complex without UBL domain eluted around 260 – 300 mM NaCl on Source S. The fraction

containing the protein was concentrated to around 20 mg/ml, and eluted around 12.75 ml on Superdex 200. The sample was finally concentrated to around 20 mg/ml.

4. DNA synthesis

The top strand was made by solid-phase synthesis in collaboration with Prof. Chuan He and Dr. Guanqun Zheng at the University of Chicago. For the cross-linking nucleotide, first a 2-Fluoro-2'-deoxyInosine (Glen Research) was incorporated, then converted by reacting with cystamine, and deprotected with 1,8-diazabicycloundec-7-ene (http://www.glenresearch.com/Technical/TB_2-F-dI.pdf). Finally the strand was purified by denaturing gel electrophoresis. The bottom strand was purchased from IDT and annealed with top strand to make the DNA duplex.

5. Crosslinking reaction and purification

First nitrogen gas was passed through all the buffer used in the reaction to expel oxygen, for about 10 min. A Zeba spin desalting column 40K MWCO (Thermo Scientific) was equilibrated with SD buffer (5 mM bis-tris propane (BTP), 800 mM sodium chloride (NaCl), pH 6.8) without DTT, and purified protein was passed through the column to remove DTT from its storage buffer. Then the DNA was added to the protein at a 1:1 molar ratio, and MQ-A (5 mM BTP pH 6.8, 10% glycerol) buffer was slowly added into the mixture to lower the concentration of NaCl to finally 100 mM. The crosslink reaction mixture was incubated at 4 °C for overnight.

On the next day, 10 µl was drawn from the mixture and mixed with 1 µl of 1 mM MMSF (S-Methylmethanethiosulfonate) (Sigma-Aldrich), incubated at room temperature

for 10 min, to quench all unreacted DNA. Then 1 μ l of 10X non-reducing SDS-PAGE loading dye (200 mM Tris-HCl pH 6.8, 50% glycerol, 10% SDS, 0.2% Bromophenol Blue) was mixed in, and the mixture was separated on 15% SDS-PAGE at 200 V for 60 min. A higher band than the free protein indicated successful crosslinking. Usually the reaction efficiency is about 50% or higher.

The reaction mixture was separated on Mono Q 5/50 GL (GE Healthcare) at 0 – 2 M NaCl gradient. Separated peaks indicating free protein, cross-linked complex and free DNA could be seen. The complex usually eluted around 400 – 450 mM NaCl. (Figure 8) The fractions containing cross-linked complex were combined and concentrated in Amicon stirred cell (EMD Millipore) to around 30 μ M.

6. Crystallization and structure determination

The crystallization drop contained 1 μ l of the complex mixed with 1 μ l of crystallization buffer, and were set up with hanging drop diffusion. Good crystals usually appeared in 50 mM BTP pH 6.8, 100 – 200 mM NaCl, 10 – 16% isopropanol and 50 – 100 mM CaCl_2 . When the crystal was ready to be harvested, the stabilizing buffer (25 mM BTP pH 6.8, 130 mM NaCl, 14% isopropanol, 20 mM CaCl_2) was added to the drop. The crystal was transferred to stabilizing buffer with 20% PEG (polyethylene glycol) 200. After about 1 min, the crystal was scooped out and immediately plunged in liquid nitrogen to freeze. The diffraction data was collected at Advance Photon Source LS-CAT 21IDD or in SBC-CAT beamlines. The data were collected with HKL3000 suite [63]. The structure was determined by molecular replacement using the structure of Rad4-

Rad23 as appears in the damaged DNA-bound complex (PBD 2QSH), and refined by Phenix [64] .

7. EMSA

See protocol h.

8. DNA sequence

Table 3. The sequence of DNA constructs used in chapter 2.

DNA construct name	Single strand name	Sequence
CH6	CH6_top	5'-TTGACTC G*ACAT CCGGGGC TACAA -3'
	CH6_btm	3'- ACTGAG C TGTA GGCCCCG ATGTTA -5'
CH7	CH7_top	5'-TTGACTC G*ACAT CCCCCGC TACAA -3'
	CH7_btm	3'- ACTGAG C TGTA GGGGGCG ATGTTA -5'
CH8	CH8_top	5'-TTGACTC G*ACAT ATATATA TACAA -3'
	CH8_btm	3'- ACTGAG C TGTA TATATAT ATGTTA -5'
CH9a	CH9a_top	5'-TTGACTC G*ACAT CGCGCGC TACAA -3'
	CH9a_btm	3'- ACTGAG C TGTA GCGCGCG ATGTTA -5'
CH9b	CH9b_top	5'-TTGACTC G*ACAT CGCGCGC TACAAA -3'
	CH9b_btm	3'- ACTGAG C TGTA GCGCGCG ATGTTTA -5'
CH9c	CH9c_top	5'-TTGACTC G*ACAT CGCGCGC TACA -3'
	CH9c_btm	3'- ACTGAG C TGTA GCGCGCG ATGTA -5'
CH9d	CH9d_top	5'-TTGACTC G*ACAT CGCGCGC TAC -3'
	CH9d_btm	3'- ACTGAG C TGTA GCGCGCG ATGA -5'

* indicates the nucleotide with thiol-linker.

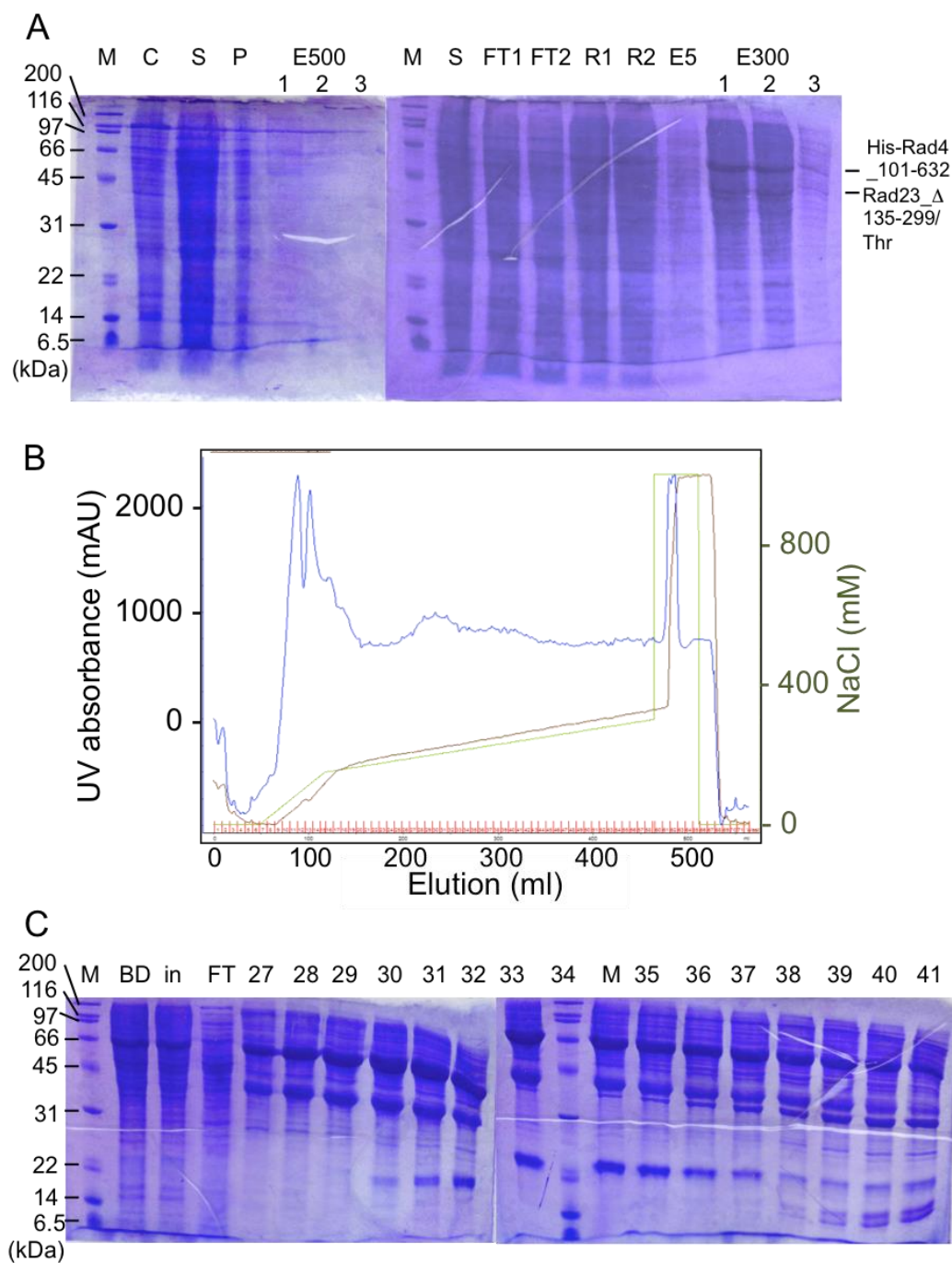


Figure 5. Lysis, affinity column and SourceQ purification of SC32.

A. Lysis and Ni pull-down of SC32. See protocol g for labels. B. An exemplary SourceQ (24 ml) chromatogram for SC32. C. The fractions from B examined on SDS-PAGE. #27 – 36 and #37 – 42 were collected.

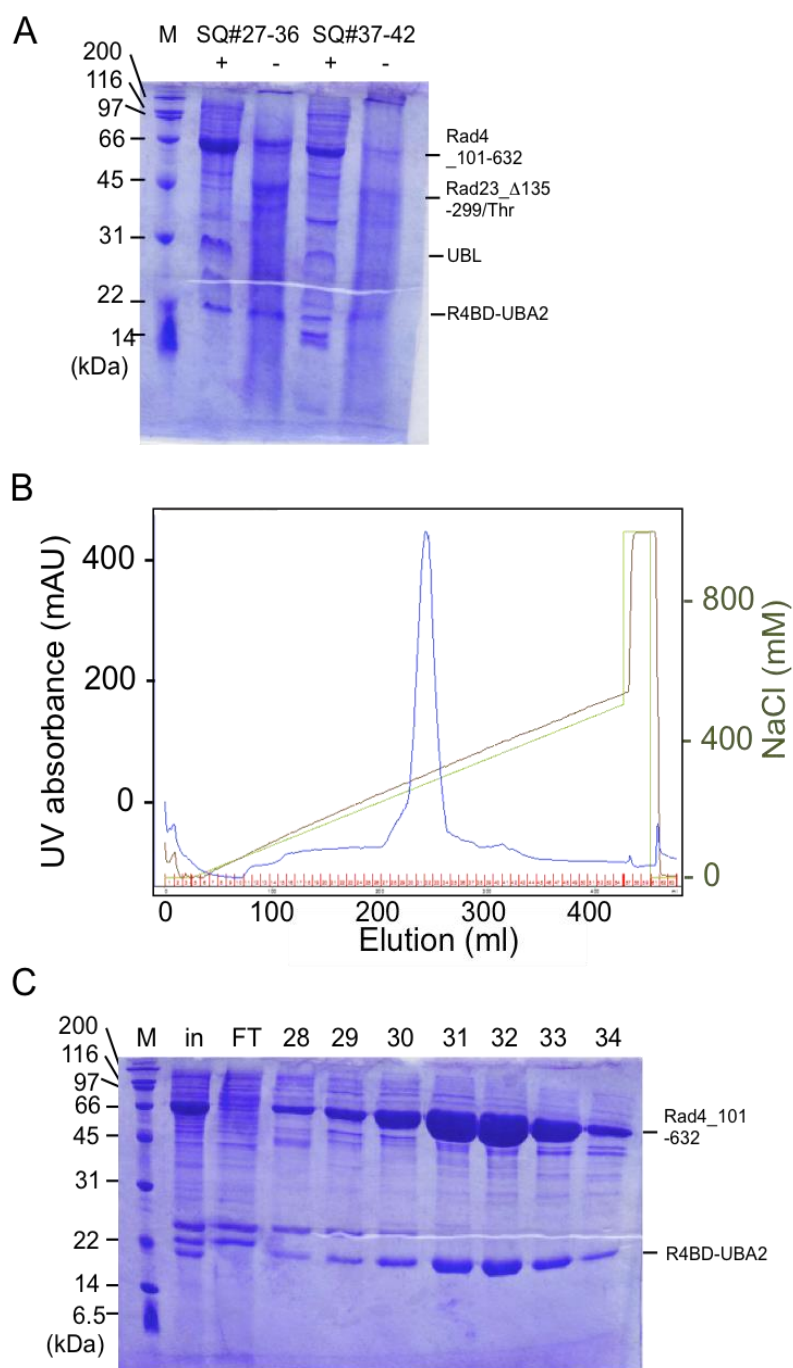


Figure 6. Purification of SC32 by SourceS.

A. The collected fractions from SourceQ (Figure 5) were digested with thrombin. B. An exemplary SourceS (8 ml) chromatogram for digested SC32. C. The fractions from B examined on SDS-PAGE. # 31 – 34 were collected.

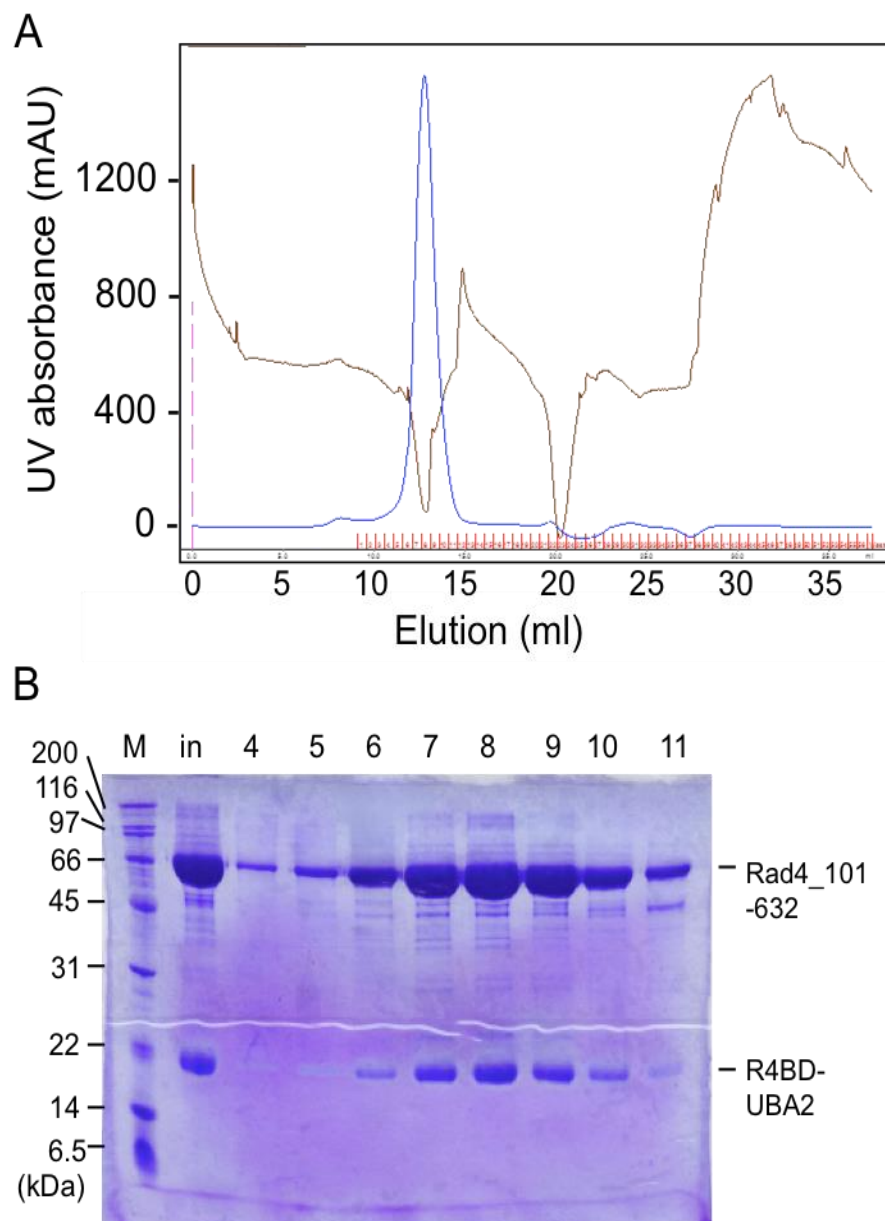


Figure 7. Purification of SC32 by Superdex200.

A. An exemplary Superdex 200 (24 ml) chromatogram for digested SC32.

B. The fractions from B examined on SDS-PAGE. # 6 – 10 were collected.

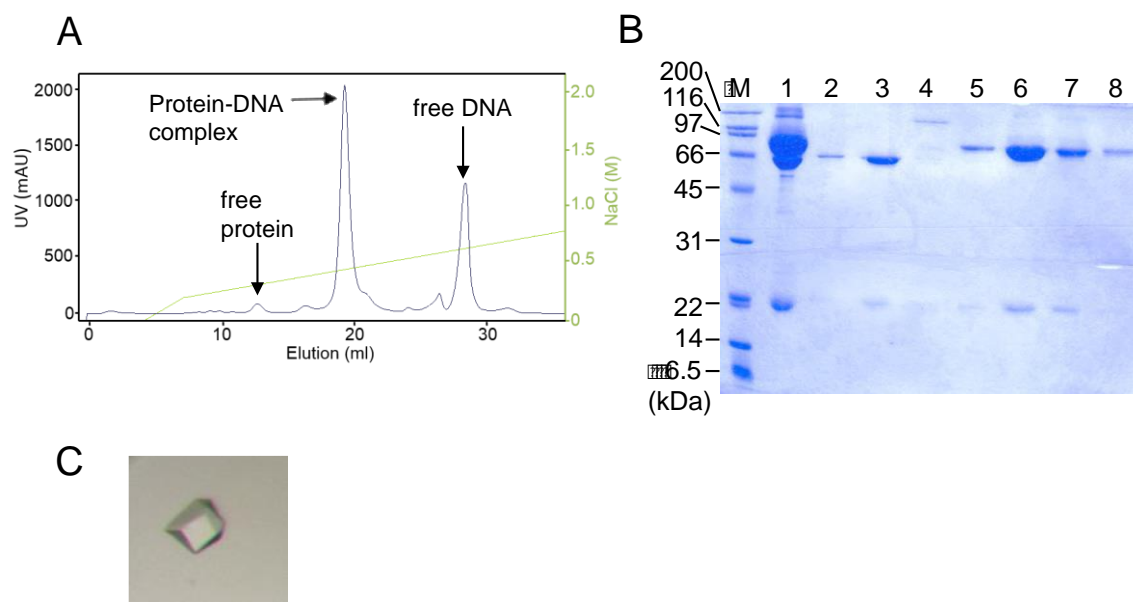


Figure 8. Purification of crosslinked complex and crystallization.

A. An exemplary chromatogram showing crosslinking reaction separated on MonoQ column. Free protein elutes ~300 mM NaCl; protein-DNA complex at ~400 mM and free DNA at ~600 mM. B. The elution fractions from MonoQ examined on 15% SDS-PAGE. M: molecular weight. Lane 1: input, lane 2 – 3: free protein, lane 5 – 8: successfully cross-linked complex. C. A crystal of crosslinked complex.

C. Results

To modify the protein and DNA for crosslinking, a serine in TGD domain of Rad4, which was seen in the Rad4-damaged DNA structure to interact with DNA non-specifically but not involved in DNA opening and lesion recognition, was mutated to cysteine, and a nearby cysteine was also mutated to serine to prevent undesired heterogeneous crosslinking. These two point mutations did not affect the undamaged or damaged DNA binding affinity of Rad4 (Figure 9). On the other hand, position 8 of the DNA, which was shown to be in close vicinity in previous structure, was modified with a

thiol-bearing linker. The cysteine in turn forms a disulfide bond with the linker on G8, trapping the protein at that specific position (Figure 10A).

We tested several DNA constructs with different bubble sequence (CH6 – CH9a), or different length (CH9b – CH9d), as well as different proteins. SC32-CH6 produced only very small crystals in MPD (2-methyl-2,4-pentanediol) and isopropanol, therefore was not pursued. CH8 could not crosslink to Rad4 successfully. SC32 crosslinked to CH9a, CH9b, CH9c and CH9d all produced very small crystals in isopropanol, and were not pursued either. SC37-CH2 produced crystals in isopropanol, but could not diffract near 4Å. SC49-CH2 also produced crystals in isopropanol, but could not diffract beyond 5Å. SC49 crosslinked to CH6, CH7, CH9b, CH9c and CH9d all produced very small crystals in isopropanol, and were not pursued. SC49-CH9a produced crystals in isopropanol and diffracted to 4.2Å. Among these SC32-CH7 produced best crystals. The structure was solved by molecular replacement using the structure of Rad4 bound to damaged DNA as model, and refined to 3.05 Å (Figure 10B).

The presence of disulfide bond between protein and DNA was confirmed by electron density data (Figure 10C). Surprisingly, this structure did not reveal an “intermediate” state between closed and open Rad4 complex, but was mostly indistinguishable from the open form when overlaid (RMSD = 0.99 Å) (Figure 11). The undamaged DNA was bent to a similar angle as the damaged DNA (~ 42°), the β-hairpin inserted into the duplex, and the nucleotides, although being undamaged/matched, were also opened and flipped out as in the structure of Rad4 bound (not tethered) to DNA lesions.

To further assure that crosslinking or crystallization did not cause artifact or distort the observed structure, the 514-bp undamaged DNA non-covalently bound by Rad4 in solution was examined by atomic force microscopy (AFM). Among the internal bound proteins, the position of binding was mostly random, confirming Rad4 binding to undamaged DNA non-specifically (Figure 12C). The data also revealed that DNA was still bent by Rad4. The distribution of bent angle gave an average $48.4 \pm 34.2^\circ$ (Figure 12D), similar to the bent angle in the crystal structure.

D. Discussion

Many DNA damage-sensing proteins in the cell distort the DNA and flip out the bases, like Rad4. Some structures have been solved showing DNA repair proteins interrogating DNA in an intermediate step of recognition by crosslinking the protein on undamaged DNA [65, 66]. Hence we used a similar strategy hoping to trap Rad4 in an intermediate state. But interestingly we obtained a crystal structure of Rad4 covalently tethered to undamaged DNA that is essentially the same as Rad4 bound to damaged DNA, supported by the AFM image of undamaged DNA being bent by untethered Rad4 in solution. And especially the opened site sequence is GGG/CCC, which is stable and rigid, presumably not easy to open. This indicates that unlike other direct damage-sensing proteins, Rad4 is able to form open conformation with undamaged DNA, therefore the mechanism of lesion searching by Rad4 must not rely on directly distinguishing the structural difference between lesion and normal DNA. This is consistent with the fact that

Rad4 is not interacting with the lesion itself, and it is able to recognize many structurally distinct lesions.

But if Rad4 opens every matched site to interrogate the genome, it would be too time-consuming, and impossible to discriminate lesion from normal DNA. There must be another factor that determines whether Rad4 should open a certain site. The fact that it opens matched DNA when it was tethered (crosslinked) to it suggests that keeping it at a specific site long enough may allow the protein to open the undamaged DNA as it does damaged DNA. Therefore we proposed that the mechanism of lesion recognition may lie in the kinetics of Rad4 on DNA.

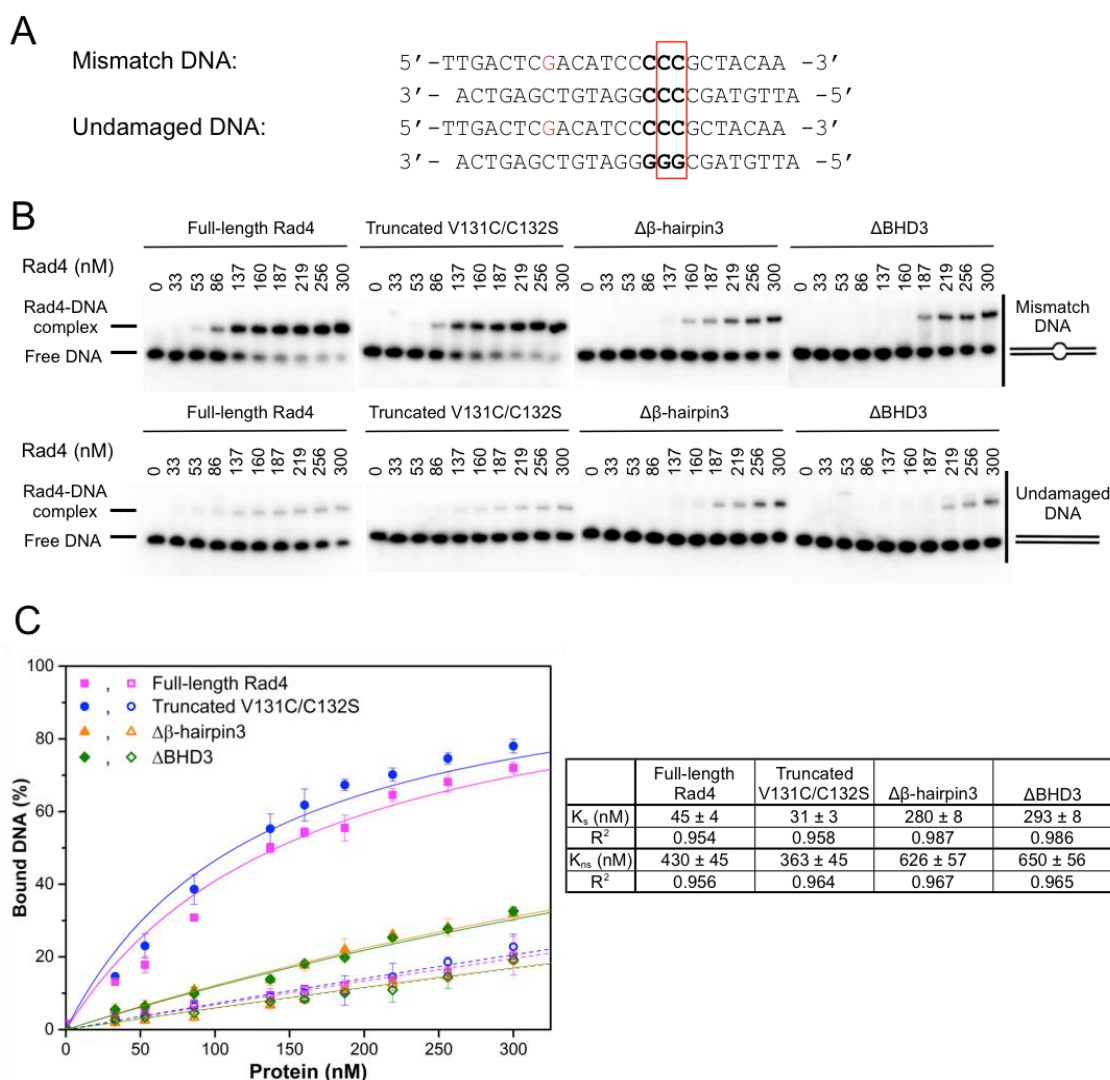


Figure 9. EMSA of full-length, truncated Rad4 with point mutation, and the mutants.

A. The sequence of mismatch and undamaged DNA used in the assay. G in red indicates the crosslinking position in the crosslinking version. These sequences do not contain thiol linker. Letters in bold indicate the bubble sequence. Letters in red box are the flipped out nucleotides. B. The exemplary gels of different Rad4 constructs bound to mismatch or undamaged DNA. C. (left) The binding curve of different Rad4 constructs. Solid symbol: mismatch DNA; Open symbol: undamaged DNA; Solid line: fitted binding curve for mismatch DNA; Dotted line: fitted binding curve for undamaged DNA. The error bars indicate standard deviation from 3 sets of measurements. (right) Calculated K_d for each construct binding to mismatch DNA (K_s) and undamaged DNA (K_{ns}).

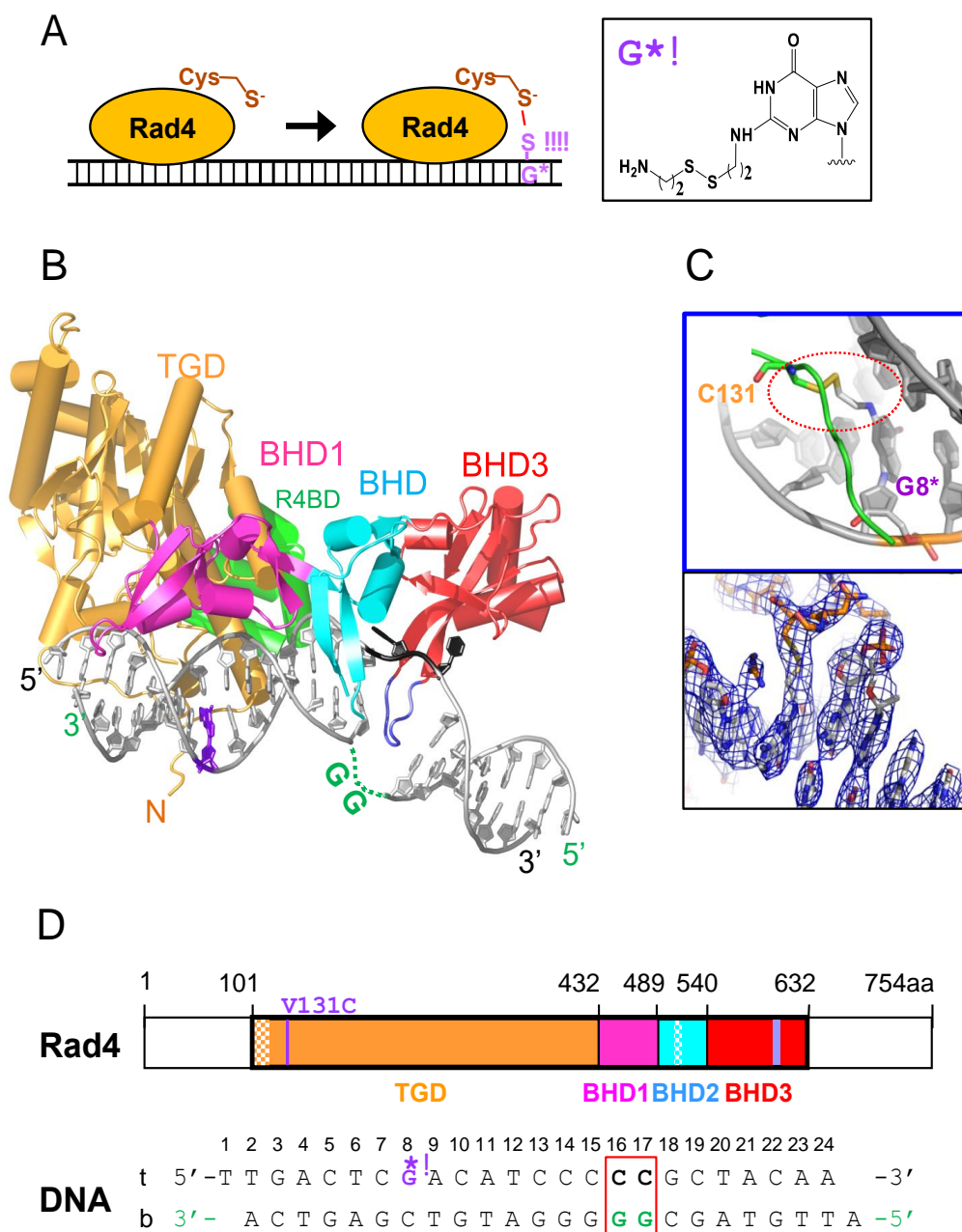


Figure 10. The structure of Rad4 crosslinked to match DNA.

A. The scheme of crosslinking strategy. The structure of G with thiol-linker is shown on the right. B. The structure of the crosslinked complex. C. Close-up view of the crosslinking region showing disulfide bond was formed. D. (top) The domains of Rad4 and (bottom) the sequence of DNA, red box indicates the flipped out nucleotides, star indicates the nucleotide bearing the linker.

Table 4. Crystal diffraction data collection and refinement.

Data collection	
Space group	P 4 ₁ 2 ₁ 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	79.405, 79.405, 404.366
α , β , γ (°)	90, 90, 90
Resolution (Å)	50.00-3.05 (3.10-3.05)
<i>R</i> _{sym} or <i>R</i> _{merge}	7.1% (79.2%)
<i>I</i> / σ <i>I</i>	24.56 (2.3)
Completeness (%)	99.9% (100%)
Redundancy	6.7 (6.9)
Refinement	
Resolution (Å)	39.7-3.05
No. of reflections	25974
<i>R</i> _{work} / <i>R</i> _{free} (%)	20.41/25.66
No. of atoms	5447
Protein	4508
DNA	939
water	0
B-factors (Å ²)	68.00
Protein	63.30
DNA	90.60 (91.67)
water	0
Root mean squared deviations	
Bond lengths (Å)	0.017
Bond angles (°)	1.94

Values in parentheses are for highest-resolution shell.

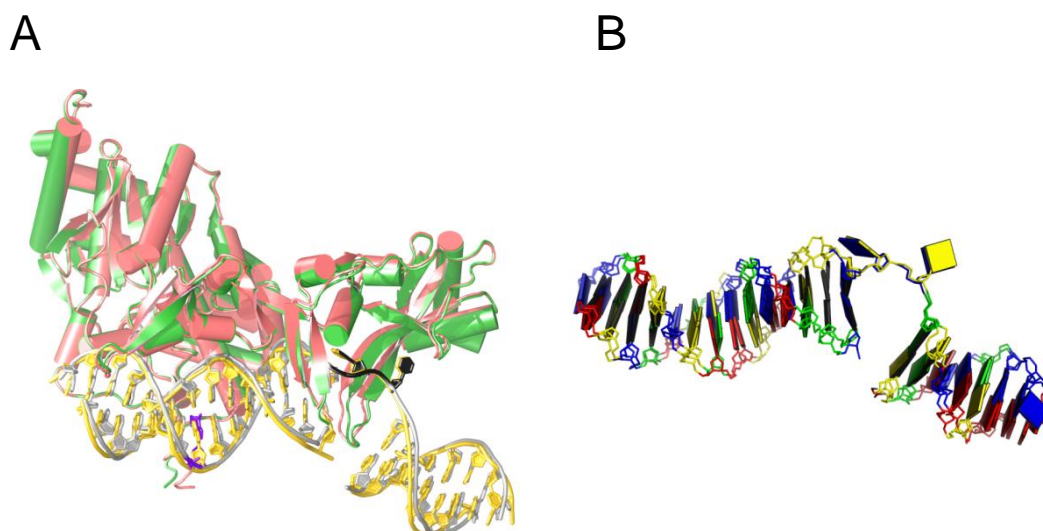


Figure 11. Comparison of Rad4 bound to damaged DNA and undamaged DNA.

A. Overlay of the structure of Rad4 bound to damaged DNA (pink and gold) and the structure of Rad4 crosslinked to undamaged DNA (green and silver). B. Overlay of the DNA from the above two structure.

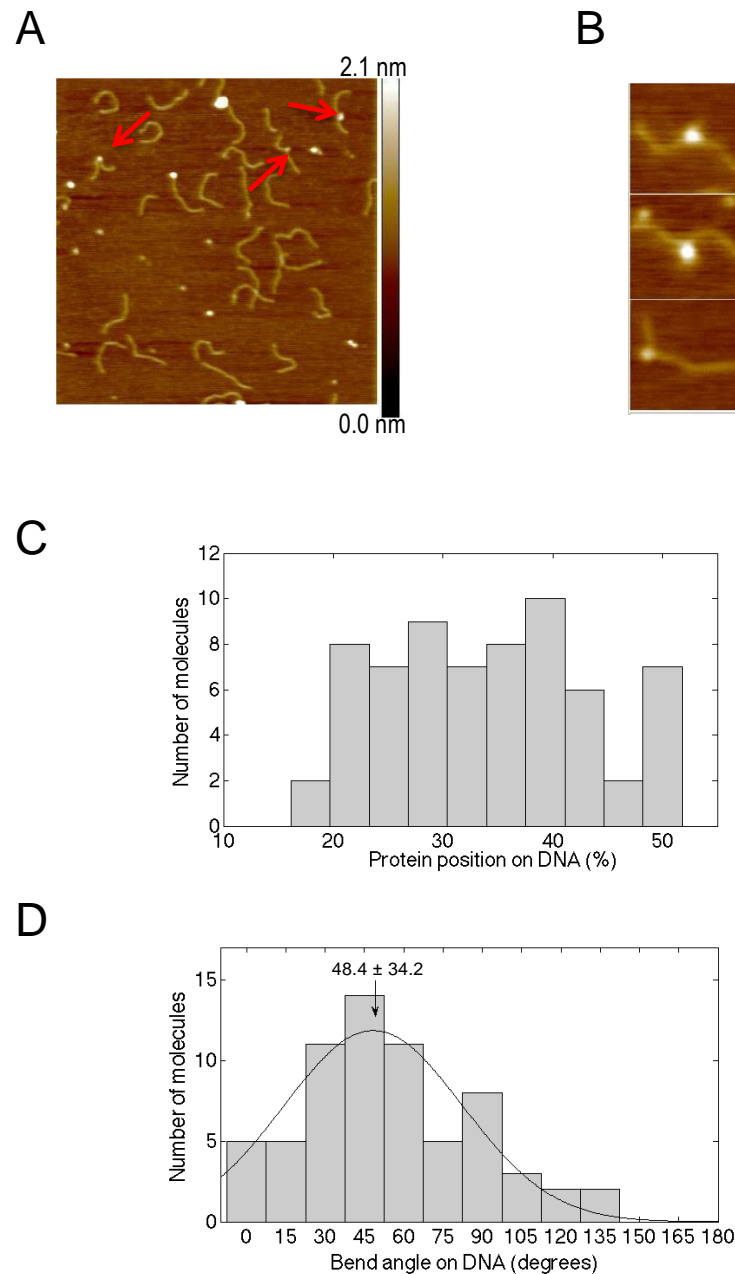


Figure 12. AFM study of Rad4 non-covalently bound to match DNA. Data was collected by Dr. Lili Liu in collaboration with Bennett Van Houten's group at University of Pittsburg.

A. An exemplary AFM image of Rad4 bound to 514-bp undamaged DNA. Arrows indicate bound protein. B. Zoomed images showing Rad4 bound to undamaged DNA. C. The position of bound Rad4 on DNA, shown as percentage of DNA length from the nearest end. The proteins bound to the end are excluded. D. The distribution of bent angle of DNA bound by Rad4. The solid line indicates Gaussian fit.

III. THE DNA OPENING AND TWISTING KINETICS INDUCED BY RAD4–RAD23

A. Introduction

The study of kinetics reveals the mechanism of biological process. Stopped-flow is a method involving rapid mixing of reactants and stopping the flow, that can detect kinetics as fast as a few milliseconds. For molecular events faster than that, laser temperature-jump spectroscopy (T-jump) is a powerful tool that covers nanosecond to millisecond resolution (reviewed by Kubelka [67]). In T-jump, laser absorption of a pulse can result in nearly instantaneous heating of the sample, which triggers a disruption of the equilibrium. The conformational change during the shift of equilibrium can be probed by a fluorophore or other probes, and monitored in real time. If the fraction of molecules undergoing conformational change in response to the T-jump is enough to be observed, the relaxation rate can be obtained, which represents the net rate of the conformational change. Evidence of whether there will be significant conformational change in response to T-jump can be obtained from equilibrium experiments, where the sample is slowly heated and its signal change is observed.

To probe the DNA opening kinetics by T-jump, a base analog fluorophore, 2-aminopurine (2AP), was selected. 2AP, an adenine analog, is able to take part in base-pairing like adenine. More importantly, its fluorescence intensity is highly dependent on local environment: The quantum yield decreases by ~100 fold when it is stacked in DNA duplex as compared to in solution [68](also see Figure 23, Figure 24). Because of this

property, 2AP has been widely utilized to probe DNA structure, including DNA base flipping induced by protein, where 2AP was placed at or next to the flipping site and the flipping motion was detected by increase in 2AP fluorescence [69] [70]. In our study, 2AP was incorporated in DNA duplex in or near the mismatch bubble, where the bases should be flipped out by Rad4 as seen in the crystal structure. The rate of base flipping was measured by T-jump and its results are published [71].

Next we examine the motions beyond the flipping out of the nucleotides at the damage sites. Wilhelmsson *et al.* discovered that tricyclic cytosine, a base analog, had a very high quantum yield, which was not affected by environment, such as in free form, in single strand, or double strand DNA [72]. When tC replaced regular cytosine in DNA duplex, the DNA still adopted a normal B form with minimal perturbation on the structure, and the thermal stability was slightly enhanced due to better base stacking [73]. A derivative of tC, tC^O, retained all the above properties, and was even brighter [74]. A non-fluorescent tC derivative, tC_{nitro}, could be used as a quencher for tC^O making them a FRET pair. Förster resonance energy transfer (FRET) is a powerful technology that detects structural or distance change, relying on the efficiency of energy transfer between a donor and an acceptor. The FRET efficiency of tC^O/ tC_{nitro} pair decreases as the distance increases between them, while also shows local periodicity as their position in the DNA duplex changes (Figure 13), indicating their FRET is sensitive to both distance and the orientation of the dipoles [75]. These properties make tC^O/ tC_{nitro} an ideal FRET pair to probe the twisting motion, which involves change in the orientation of nucleotides.

A series of DNA constructs were designed, with sequence similar to the DNA used in crystal structure, and tC^O/tC_{nitro} incorporated at different positions (from 4 bp to 7 bp apart). Each construct has a match and a mismatch version, with the mismatch site being TTT/TTT or TAT/TAT. The Rad4 induced kinetics of these constructs were studied by T-jump.

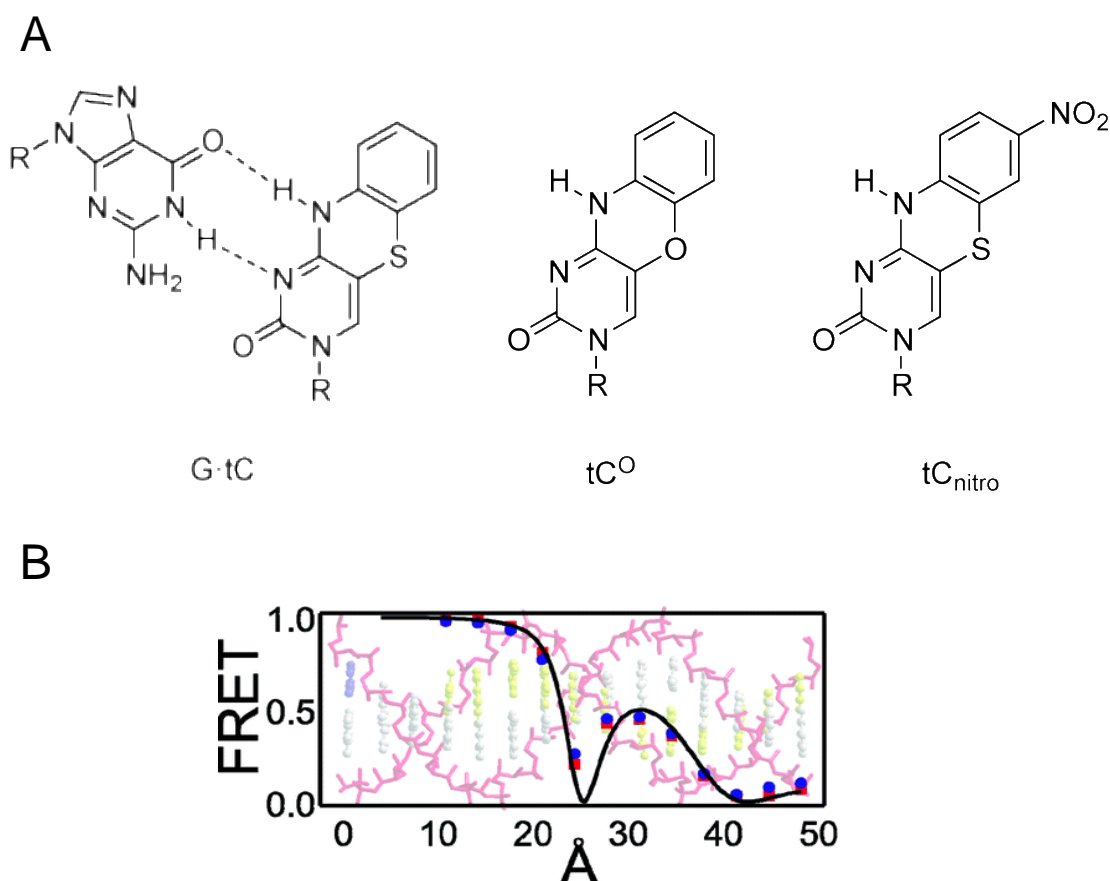


Figure 13. The FRET pair tC^O and tC_{nitro}
 A. The structure of G pairing with tC , tC^O and tC_{nitro} . B. The change of FRET efficiency of tC^O and tC_n pair in B-DNA as the distance changes between them in Å. Adapted from Borjesson *et al* [75].

B. Materials and Methods

1. Cloning

Rad4_101-632, Rad4_101-632_Δ599-605 (lacking β -hairpin3) and Rad4_101-540 (lacking BHD3 domain) were cloned from yeast genomic library, inserted between BssHII and NotI in pFBD vector (Invitrogen), with a His tag engineered between BamHI and BssHII: These constructs are named as <108>, <137>, <140> respectively. For all 3 constructs, Rad23 with amino acids 135 – 299 deleted and replaced by thrombin recognition sequence was inserted between NheI and SphI in the same vector.

2. Virus production

See protocol d and e.

3. Protein expression and purification

See protocol f and g. The heterodimeric complex <108> eluted around 200 – 250 mM NaCl on 24 ml Source Q, <140> eluted around 180 – 240 mM NaCl. The fractions containing the protein were not digested with thrombin. Then <108> eluted around 210 – 270 mM NaCl on Source S, <140> eluted around 240 – 300 mM NaCl. The fractions containing the complex were concentrated to around 20 mg/ml. Then they were purified on Superdex 200: <108> eluted around 12.1 ml, <140> eluted around 13.5 ml. The sample was finally concentrated to around 20 mg/ml.

4. Dynamic light scattering (DLS)

The samples used were 10 mM of Rad4-Rad23, or 10 mM each of protein and DNA in 1X PBS buffer with 1 mM DTT, prepared in the same manner as equilibrium samples. The measurements were taken with DynaPro-801 Dynamic Light Scattering / Molecular Sizing Instrument (Protein Solutions). The calculation was generated by Dynamics v4.0 software.

5. Equimolar electrophoretic mobility shift assay

10 μ M of Rad4-Rad23 and 10 μ M of DNA of interest was mixed, and serially diluted to 5, 2.5, 1, 0.5, 0.25, 0.1 mM. Then the mixture was mixed with 5 nM of the same DNA labeled with 32 P in binding assay buffer (5 mM BTP-HCl, 75 mM NaCl, 5 mM DTT, 5% glycerol, 0.74 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 500 μ g ml⁻¹ bovine serum albumin, pH 6.8). Mixed samples were subsequently incubated at room temperature for 20 min and separated on 4.8% non-denaturing polyacrylamide gels as described for the gel shift assay with competitor.

6. Electrophoretic mobility shift assay with excess protein

2 μ M of Rad4-Rad23 was serially diluted to 1, 0.5 μ M in 5X binding buffer (25 mM BTP pH 6.8, 375 mM NaCl, 25% glycerol, 3.7 mM 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 2.5 mg/ml bovine serum albumin, 25 mM DTT). 4 μ l of each protein solution was mixed with 1 μ l of milliQ water and 5 μ l of 400 nM cold DNA in milliQ water. Then 10 μ l of 10 nM of the same DNA labeled with 32 P in milliQ water was added to the mixture, and incubated at room temperature for 20

min and separated on 4.8% non-denaturing polyacrylamide gels as described for the gel shift assay with competitor.

7. Equilibrium temperature scan experiment

The protein and DNA were mixed at 1:1 molar ratio, to a final 10 μ M complex in 1X PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4) with 1 mM DTT. The fluorescence intensity was measured in FluoroMax4 spectrofluorimeter (Horiba Scientific, NJ) over the temperature range 10 – 40 °C or 5 – 30 °C with 2.5 °C increment. At each temperature the sample was equilibrated for 3 min, and finally one more measurement was taken at 25 °C after equilibrating for 8 min. The 2AP samples were excited at 314 nm and the emission intensity from 330 – 500 nm was recorded. The slit width for both excitation and emission was 4 nm. For the tC⁰/tC_{nitro} samples, the donor was excited at 365 nm. Because the acceptor does not fluoresce in aqueous solution, the donor emission intensity from 375 – 550 nm was recorded. The slit width for both excitation and emission was 4 nm.

The emission intensity at each temperature was plotted. For 2AP samples the area under 330 – 390 nm was taken as the intensity for the sample, and plotted against temperature. For normalized intensity plot, the intensity of both DNA and complex samples at the lowest temperature were normalized to 1.

For the tC⁰/tC_{nitro} samples, the DNA with only donor (_D) and the DNA with both donor and acceptor (_DA) were always measured on the same day. The donor fluorescence intensity at each temperature was plotted. The area under 425 – 515 nm was

taken as the intensity for the sample (I_D for donor only DNA, I_{DA} for DNA with both donor and acceptor), and plotted against temperature. For normalized intensity plot, the intensity of both DNA and complex samples at the lowest temperature were normalized to 1. For FRET, the FRET efficiency (E) at each temperature is defined as $E = 1 - \frac{I_{DA}}{I_D}$.

8. T-jump kinetics measurement

The protein and DNA were mixed at 1:1 molar ratio, to a final 60 μM complex. The sample was dialyzed in 10kDa Spectrum Spectra/Por7 Membrane Tubing (Fisher Scientific) against 1 L of 1X PBS with 1 mM DTT at 4 $^{\circ}\text{C}$ for $\sim 3\text{h}$.

The kinetics were measured on a home-built T-jump spectrometer [76]. A 10 ns laser pulse with 1550 nm wavelength rapidly heated a 1 mm spot of the sample, generating 5 – 10 $^{\circ}\text{C}$ temperature jump. The probe was excited by Hg-Xe lamp with a broadband filter, selecting the range of 300 – 330 nm for 2AP, and the range of 335 – 375 nm for tC^O, focusing on a 300 μm spot inside the heated volume. The emission intensity was measured perpendicular to the excitation beam.

9. Melting temperature (T_m) measurement

The duplex DNA was diluted to 1.5 μM in 1X PBS buffer. The DNA absorbance at 260 nm was measured with a Cary 300 Bio UV-Visible spectrophotometer as the sample was gradually heated from 25 – 90 $^{\circ}\text{C}$ in the accompanying temperature controller (Varian). The absorbance was recorded every 0.5 $^{\circ}\text{C}$. The temperature was derived. The absorbance was derived and smoothed over 5 data points. The fit and T_m were determined using a two-state van't Hoff transition.

10. DNA sequence

Table 5. The sequence of DNA constructs containing 2AP used in chapter 3.

DNA construct name	Note	Single strand name	Sequence
AN3	33-mer	AN3_Top	5'-CGT GAC TCA ACA TCC TAT GCT ACA ACT CAG TGC-3'
		AN3_Btm	3'-GCA CTG AGT TGT AGG T2T CGA TGT TGA GTC ACG-5'
AN4	33-mer, matched version of AN3	AN4_Top	5'-CGT GAC TCA ACA TCC ATA GCT ACA ACT CAG TGC-3'
		AN3_Btm	3'-GCA CTG AGT TGT AGG T2T CGA TGT TGA GTC ACG-5'
AN3b	AN3 with crosslinking G*	AN3b_Top	5'-CGT GAC TCG* ACA TCC TAT GCT ACA ACT CAG TGC-3'
		AN3b_Btm	3'-GCA CTG AGC TGT AGG T2T CGA TGT TGA GTC ACG-5'
AN4b	AN4 with crosslinking G*	AN4b_Top	5'-CGT GAC TCG* ACA TCC ATA GCT ACA ACT CAG TGC-3'
		AN3b_Btm	3'-GCA CTG AGC TGT AGG T2T CGA TGT TGA GTC ACG-5'
AN16	CH7 sequence context with 2AP at opening site	CH7_top	5'- TTGACTC G*ACAT CCCCCGC TACAA -3'
		AN16_btm	3'- ACTGAG C TGTA GG2CG ATGTTA -5'
AN21	mismatch version of AN16	CH7_top	5'- TTGACTC G*ACAT CCCCCGC TACAA -3'
		AN21_btm	3'- ACTGAG C TGTA GG2CG ATGTTA -5'
AN18	CH7 sequence context with 2AP in a nonflipping position	CH7_top	5'- TTGACTC G*ACAT CCCCCGC TACAA -3'
		AN18_btm	3'- ACTG2G C TGTA GGGGCG ATGTTA -5'
AN24	mismatch version of AN18	CH7_top	5'- TTGACTC G*ACAT CCCCCGC TACAA -3'
		AN24_btm	3'- ACTG2G C TGTA GG2CG ATGTTA -5'
AN19	Hybrid of CH7 and TA10[77] sequence with 2AP at opening site	AN19_top	5'- TTGACTC G*GATC TGTCTA TTGC A -3'
		AN19_btm	3'- ACTGAG C CTAG ACA2GAT AACG TA-5'
AN22	AN19 with 2AP after the bubble	AN19_top	5'- TTGACTC G*GATC TGTCTA TTGC A -3'
		AN22_btm	3'- ACTGAG C CTAG ACAAG2T AACG TA-5'
AN23	mismatch version of AN22	AN19_top	5'- TTGACTC G*GATC TGTCTA TTGC A -3'
		AN23_btm	3'- ACTGAG C CTAG ACATC2T AACG TA-5'
AN22a	AN19 with 2AP before the bubble	AN19_top	5'- TTGACTC G*GATC TGTCTA TTGC A -3'
		AN22a_btm	3'- ACTGAG C CTAG AC2AGAT AACG TA-5'
AN23a	mismatch version of AN23	AN19_top	5'- TTGACTC G*GATC TGTCTA TTGC A -3'
		AN23a_btm	3'- ACTGAG C CTAG AC2TCAT AACG TA-5'

* indicates the nucleotide with thiol linker.

“2” indicates 2AP.

Table 6. The sequence of tC⁰ / tC_{nitro} DNA constructs used in chapter 3.

DNA construct name	Note	Single strand name	Sequence
AN12u_DA	Matched version of AN12	AN11_Top	5' - TTGACTCGACATC P A A AGGTACAA -3'
		AN11, 12_Btm	3' - ACTGAGCTGTAGG T T T C DATGTTA -5'
AN12_DA		AN12_Top	5' - TTGACTCGACATC P T T GGTACAA -3'
		AN11, 12_Btm	3' - ACTGAGCTGTAGG T T T C DATGTTA -5'
AN13u_DA	Matched version of AN13	AN13m_Top	5' - TTGACTCGACGGGC A T AGGGACAA -3'
		AN13_Btm	3' - ACTGAGCTG C D C G T A T C P CTGTTA -5'
AN13_DA		AN13_Top	5' - TTGACTCGACGGGC A T AGGGACAA -3'
		AN13_Btm	3' - ACTGAGCTG C D C G T A T C P CTGTTA -5'
AN14u_DA	Matched version of AN14	AN13m_Top	5' - TTGACTCGACGGGC A T AGGGACAA -3'
		AN14_Btm	3' - ACTGAGCTG C D C G T A T C C P TGTTA -5'
AN14_DA		AN13_Top	5' - TTGACTCGACGGGC A T AGGGACAA -3'
		AN14_Btm	3' - ACTGAGCTG C D C G T A T C C P TGTTA -5'
AN15u_DA	Matched version of AN15	AN13m_Top	5' - TTGACTCGACGGGC A T AGGGACAA -3'
		AN15_Btm	3' - ACTGAGCTG C C D G T A T C P CTGTTA -5'
AN15_DA		AN13_Top	5' - TTGACTCGACGGGC A T AGGGACAA -3'
		AN15_Btm	3' - ACTGAGCTG C C D G T A T C P CTGTTA -5'

“D” indicates donor tC⁰, “P” indicates acceptor tC_{nitro}.

All the DNA used in this chapter were ordered from Trilink. Only the DNA containing both donor and acceptor are listed. For donor only DNA (_D), the acceptor is replaced by C while the rest of the sequence is the same. For DNA without labels (_NL), both the donor and acceptor are replaced by C while the rest of the sequence is the same.

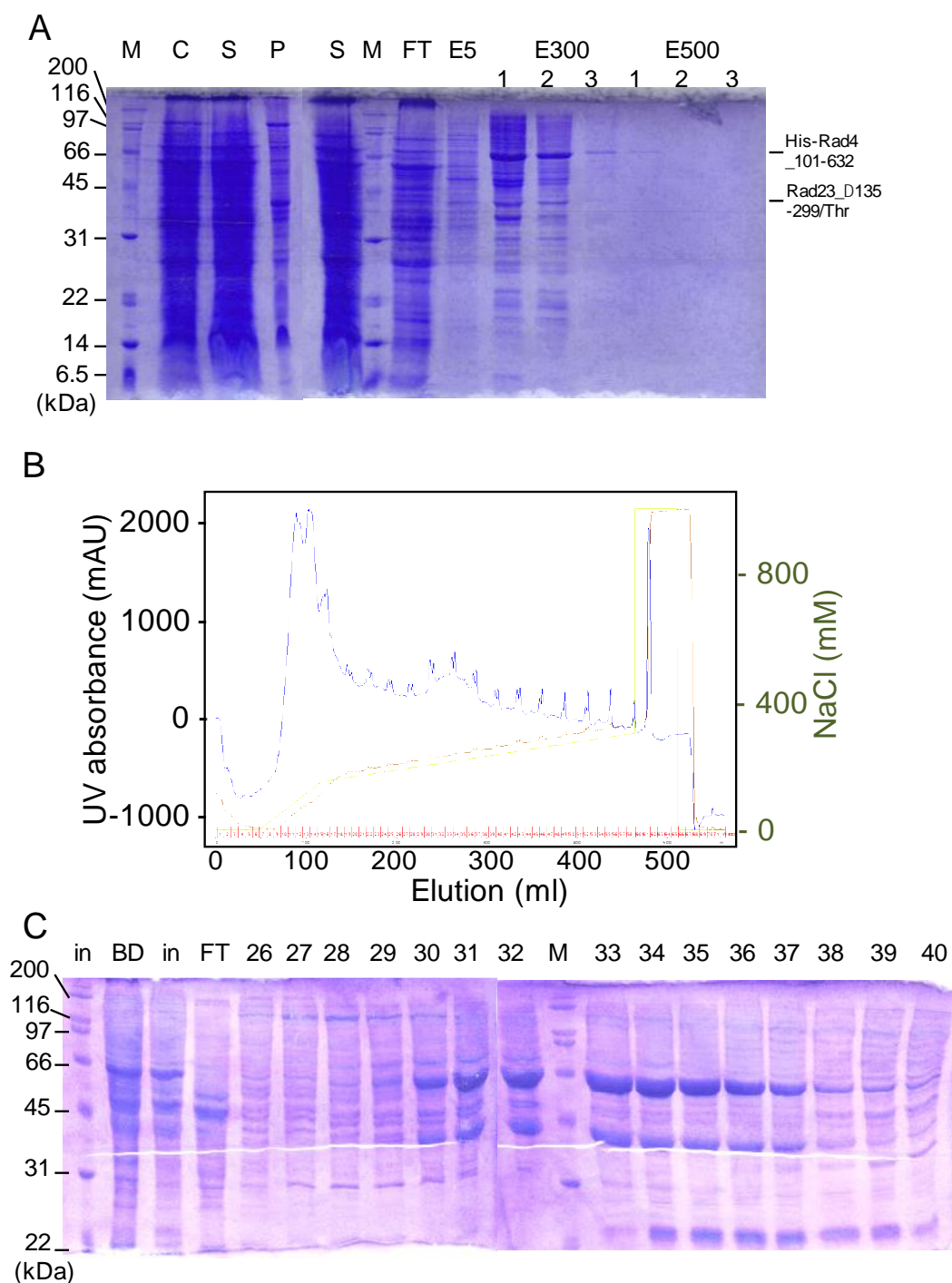


Figure 14. Lysis, affinity column and SourceQ purification of <108>.

A. Lysis and Ni pull-down of <108>. See protocol g for labels. B. An exemplary SourceQ (24 ml) chromatogram for <108>. C. The fractions from B examined on SDS-PAGE. #30 – 33 were collected.

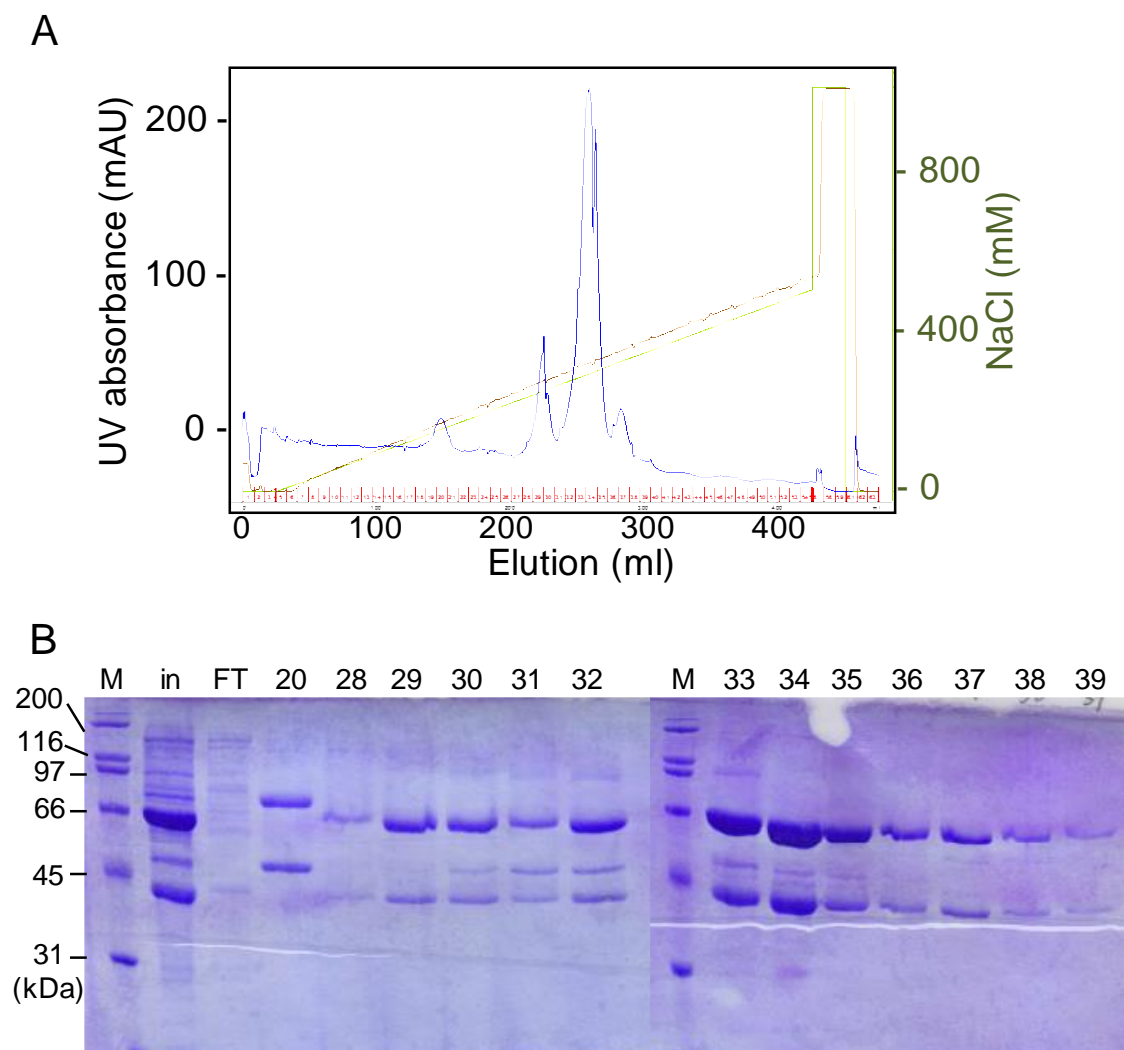


Figure 15. Purification of <108> by SourceS.

A. An exemplary SourceS (8 ml) chromatogram for <108>. B. The fractions from A examined on SDS-PAGE. #33 – 35 were collected.

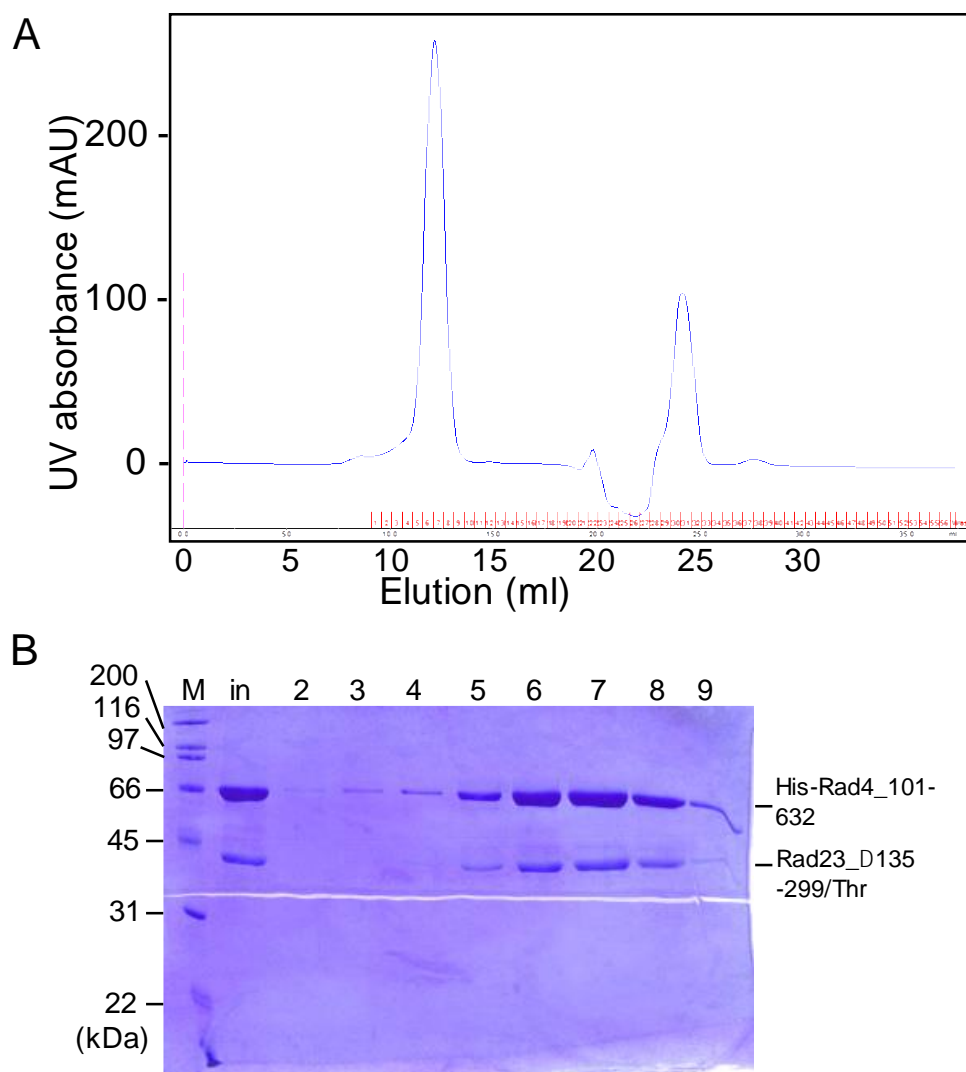


Figure 16. Purification of <108> by Superdex200.

A. An exemplary Superdex 200 (24 ml) chromatogram for <108>. B. The fractions from A examined on SDS-PAGE. # 5 – 8 were collected.

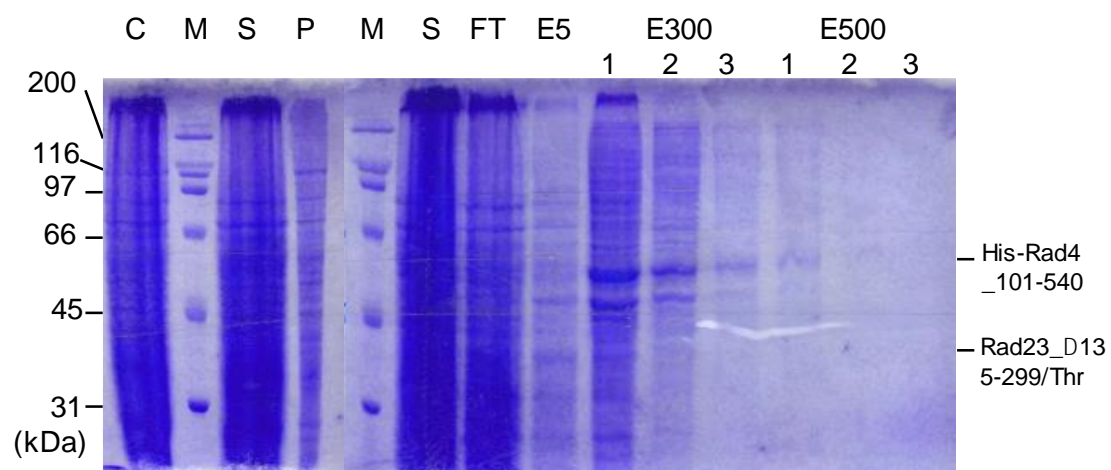


Figure 17. Lysis and Ni pull-down of <140>.
See protocol g for labels.

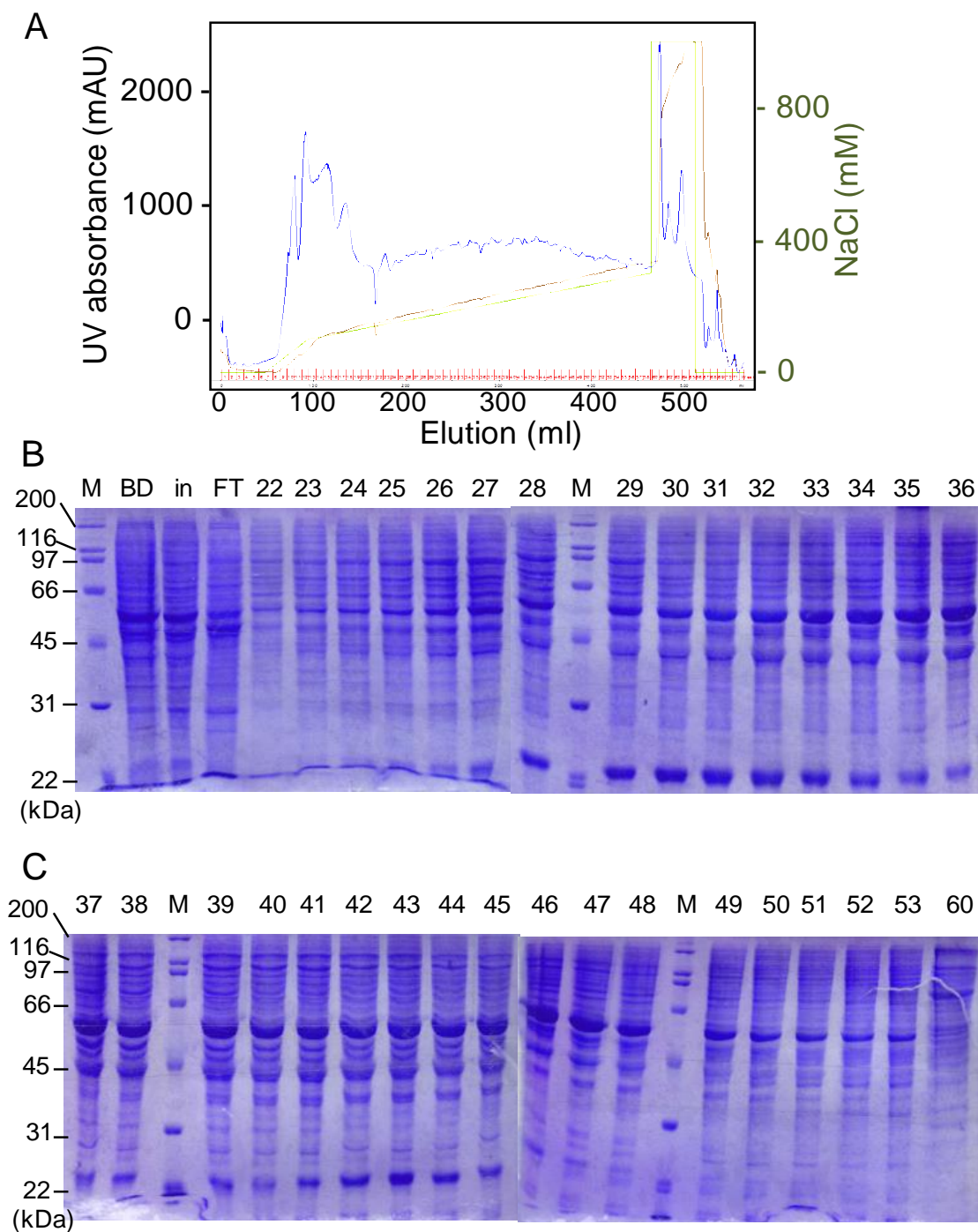


Figure 18. Purification of <140> by SourceQ.

A. An exemplary SourceQ (24 ml) chromatogram for <140>. B. The fractions from A examined on SDS-PAGE. # 29 – 45 were collected.

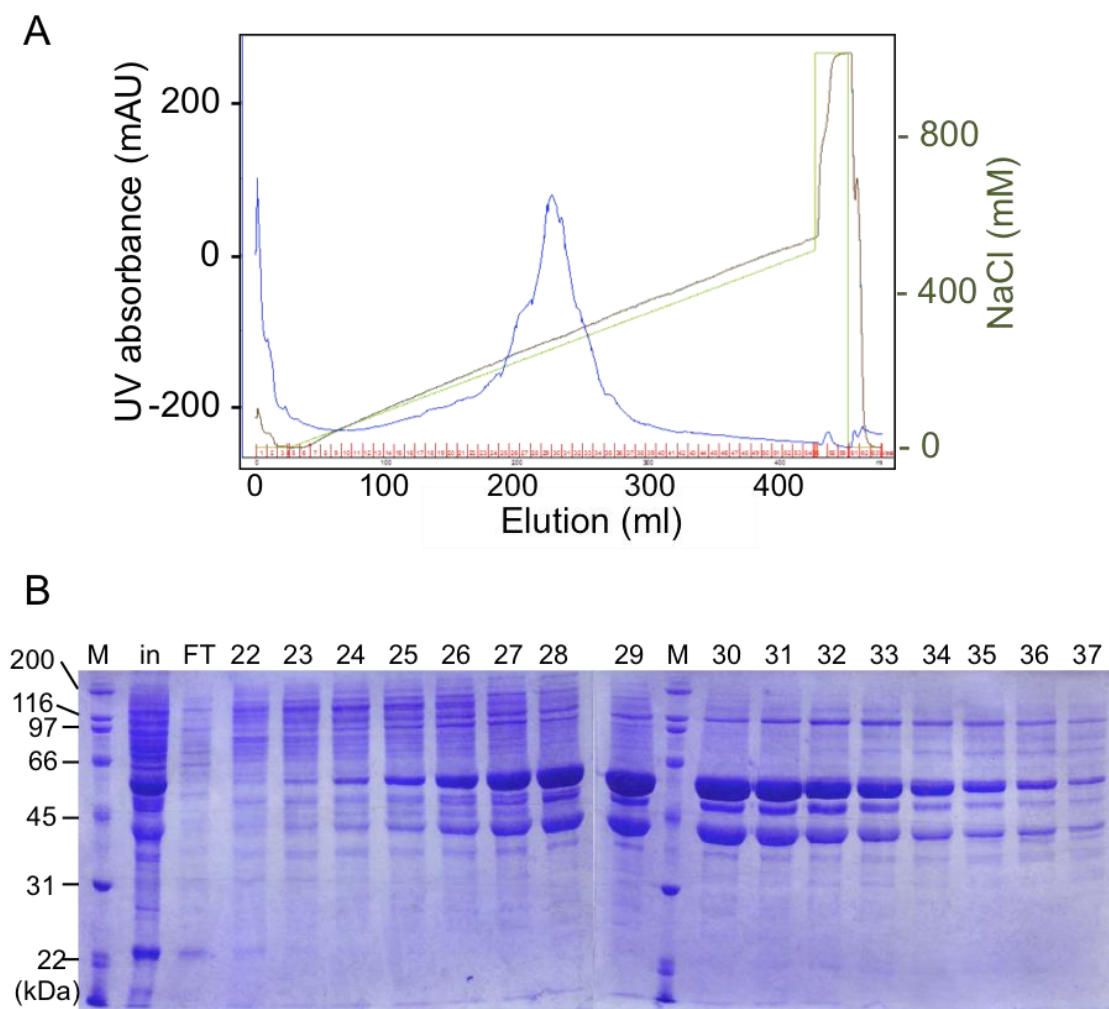


Figure 19. Purification of <140> by SourceS.

A. An exemplary SourceS (8 ml) chromatogram for <140>. B. The fractions from A examined on SDS-PAGE. # 29 – 34 were collected.

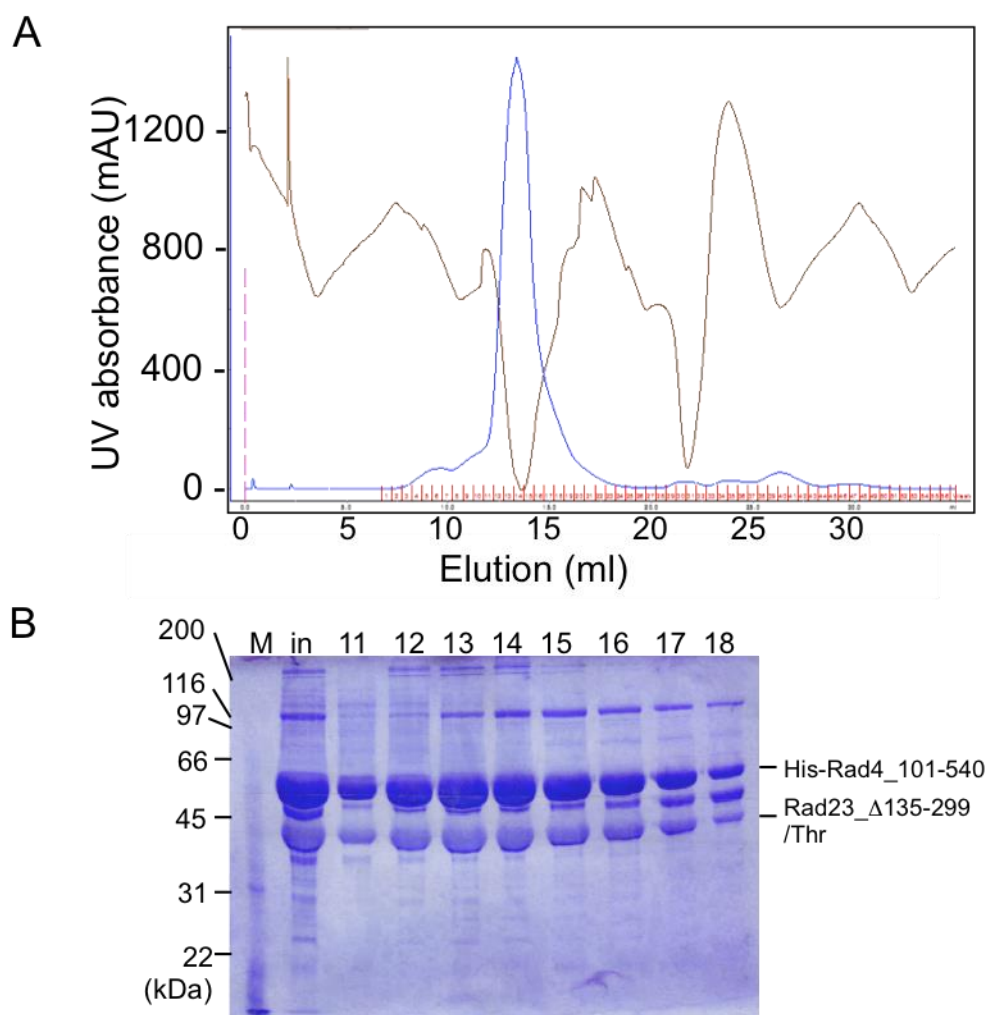


Figure 20. Purification of <140> by Superdex200. A. An exemplary Superdex 200 (24 ml) chromatogram for <140>. B. The fractions from A examined on SDS-PAGE. # 11 – 16 were collected.

C. Result

1. Examination of Rad4-Rad23 complex and DNA bound state in solution

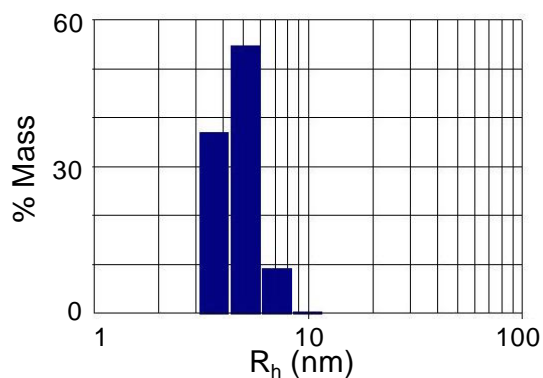
To show that the Rad4-Rad23 sample is made up of homogeneous 1:1 complexes, the protein sample was examined by dynamic light scattering (DLS) at 10 μ M, the concentration of equilibrium experiment. The mass distribution of R_h showed a single peak near 5 nm, indicating Rad4-Rad23 formed a complex that is monodispersed (Figure 21A). The data of Rad4-Rad23 bound to DNA was very similar to the protein itself, indicating the protein binds to DNA at 1:1 ratio. The average R_h did not change significantly, because the size of DNA was small (calculated molecular weight 247 kDa versus 255 kDa for Rad4-Rad23 protein complex) and did not affect the overall size of the complex when bound (Figure 21B).

To show that the protein is saturated with DNA at the concentration used in equilibrium and T-jump experiments, equimolar gel shift assay was performed with concentration up to 10 μ M. The gels showed more than 80% of DNA was bound to protein at 1 μ M and beyond, proving that the samples used in kinetic studies (at least 10 μ M) were protein-DNA complexes (Figure 22A, B).

When the protein was in excess compared to DNA in solution, higher bands were seen on the gels (Figure 22C, red arrows), representing the species with more than one protein molecules per duplex DNA bound, while the samples mixed at 1:1 ratio only had one band (black arrow). Combining this with the equimolar gel shift data, DLS data, and SD200 chromatograms (Figure 16) it can be concluded that Rad4-Rad23, as well as

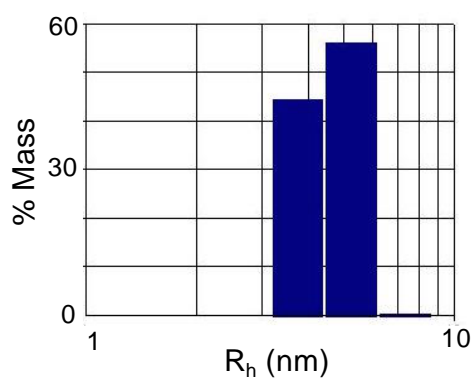
Rad4-Rad23-DNA, form monodispersed, equimolar, stable complexes under the conditions used in the equilibrium and kinetic fluorescence spectroscopy measurements.

A



Average R _h (nm)	5.59
Polydispersity	25.3%
Estimated MW	255 kDa

B



Average R _h (nm)	5.51
Polydispersity	23.1%
Estimated MW	247 kDa

Figure 21. Dynamic light scattering (DLS) of Rad4-Rad23-DNA complex.

A. Mass distribution of hydrodynamic radius (R_h) of Rad4-Rad23 protein complex. B.

Mass distribution of hydrodynamic radius (R_h) of Rad4-Rad23-AN12 complex. The tables show average R_h of the sample, polydispersity, which is indicated by percentage error of R_h , and estimated molecular weight of the particle calculated from R_h assuming spherical shape.

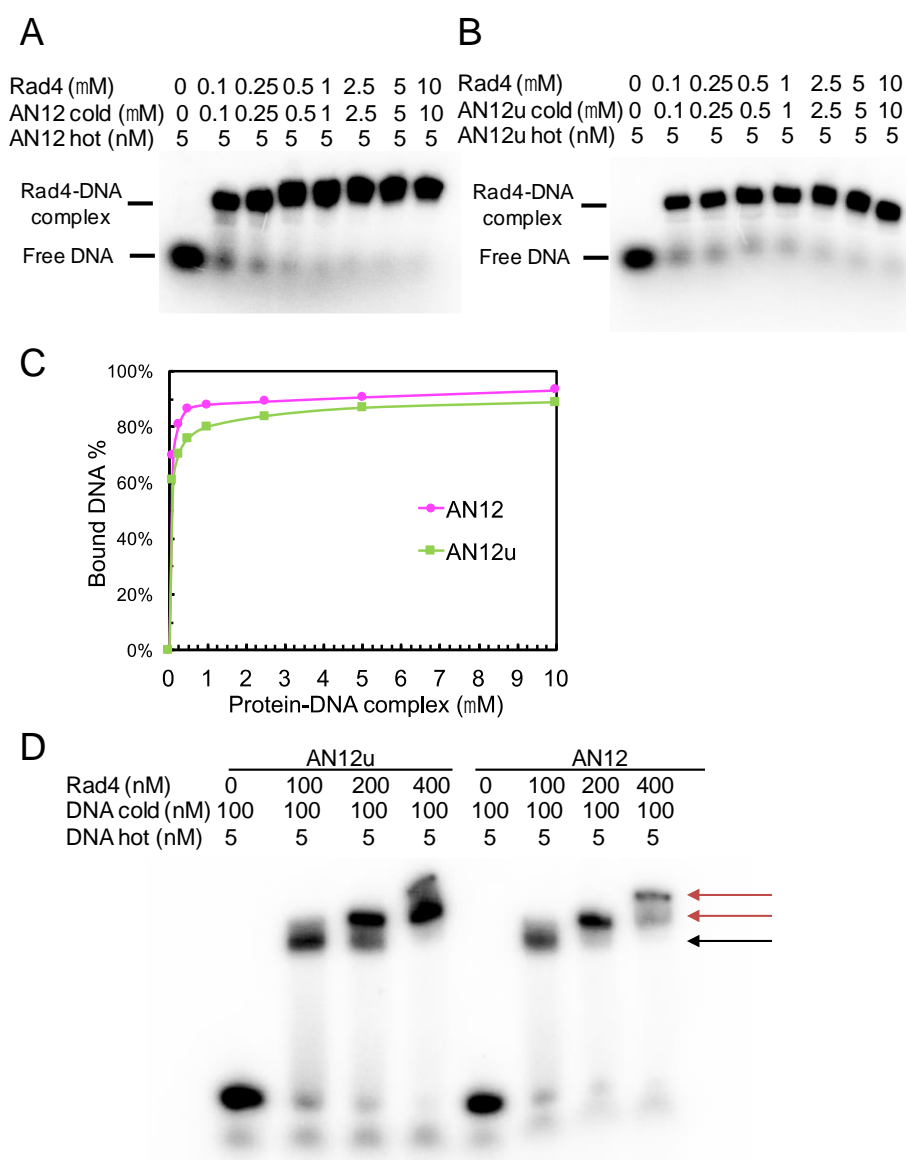


Figure 22. EMSA of Rad4-Rad23 binding to DNA showing 1:1 protein:DNA ratio.

A. Exemplary gels of Rad4 complex bound to AN12 mismatch or (B) AN12u match DNA at 1:1 molar ratio. The amount of hot DNA is very low and negligible. C. Binding curve for the gels in A and B. D. An exemplary gel of Rad4 complex bound to DNA, with protein:DNA molar ratio 0:1, 1:1, 2:1, 4:1.

2. The fluorescence intensity of 2AP constructs

Various constructs were designed with different sequence, or with 2AP placed at different position. The fluorescence of 2AP should be quenched when it is stacked, therefore the intensity of 2AP in single strand DNA should be lower than free 2AP in solution, and 2AP in duplex DNA should be even lower than single strand. When the single strands containing 2AP were examined, it appeared that the fluorescence was severely quenched in all of them, but the signature emission peak of 2AP centered near 375 nm could be seen. The quantum yield varied from 1% to 7% of the free 2AP signal (Figure 23).

In Figure 24, the fluorescence intensity of each duplex construct and corresponding single strand is summarized, also showing the change in fluorescence when protein is bound. Comparing to single strands containing 2AP, the fluorescence of duplex DNA all decreased, to different extent. Among them AN18 and AN22 were extremely quenched so that they did not even have the 375 nm peak. For this reason, these two constructs were excluded from the experiments. The rest of the duplexes emit ~ 35 – 90% compared to their corresponding single strands.

When bound to truncated (<108>), or full-length Rad4-Rad23 (<144>), all the mismatch DNA (AN3, AN3b, AN21, AN23) showed increase in intensity, from 1.5 – 7.4 fold. Considering 2AP in these constructs are either inside the mismatch bubble or right next to it, it is consistent with the crystal structure showing Rad4 opening the mismatch site, and the conformational change would expose 2AP in a more aqueous environment.

When DNA is bound to mutant lacking BHD3 (<140>) or the β -hairpin (<137>), there was essentially no increase, consistent with the low mismatch binding specificity of these mutants (Figure 9). When protein was added to match DNA, there was only slight increase, no more than 2 fold, consistent with the non-specificity of Rad4 binding to match DNA.

All the mismatch DNA that bear the thiol linkers (AN3b, AN21, AN23) showed increase in fluorescence when crosslinked to Rad4, the fold of increase is roughly similar to when they were not crosslinked. However most of the match DNA (AN16, AN19, AN22a) showed drastic drop in intensity when crosslinked, even lower than DNA itself, except for AN4b, which had similar intensity either crosslinked or not-crosslinked, although still lower than the mismatch version (AN3b). The results thus indicate that the fluorescence of 2AP is significantly affected by the sequence context in addition to the conformation context.

3. The kinetics of Rad4 bound 2AP constructs

Before measuring the T-jump relaxation kinetics of Rad4-DNA complexes, we performed equilibrium experiments by slowly heating the sample, in order to detect any change in distribution of different conformation as temperature was raised, which provided hints what to expect in T-jump, as well as assuring the sample stability within T-jump temperature range.

Because of the quantum yield of 2AP decreases as temperature increases, the intensity of DNA alone decreased (Figure 25B). When bound to Rad4, the matched DNA

AN4 only showed minimal increase in intensity, and the trend did not deviate from DNA. On the other hand, AN3-Rad4 complex not only had substantial increase in intensity compared to AN3 alone, but also increased further as temperature increased, indicating a larger population of the DNA were opened at higher temperature (Figure 25B).

When subjected to 5 – 10 °C temperature jump, the signal dropped immediately due to the lower quantum yield of 2AP at higher temperature. Then as the AN3-Rad4 complexes equilibrated to the higher temperature, that is, more DNA were opened by Rad4, the signal increased, and the relaxation rate can be obtained from fitting the kinetic trace. By plotting relaxation rate at different temperature, the rate at 25 °C can be extrapolated to be 5.5 ± 0.5 ms, representing the base opening rate at the mismatch bubble in AN3 induced by Rad4. No kinetics was observed for AN4-Rad4 complex in the 5 μ s – 50 ms time window of the instrument (Figure 25C). When bound to mutant Rad4 lacking BHD3 (<140>) or the β -hairpin (<137>), the intensity of AN3 only increased slightly (Figure 26A), and no difference from DNA alone was seen as temperature was raised (Figure 26B). No kinetics was observed for either complex (Figure 26C).

The mismatch DNA AN21, with 2-bp bubble, showed more pronounced increase when bound to Rad4 (Figure 28A). Unlike AN3 or AN4, the intensity of AN21 increased at higher temperature, which could be DNA unstacking due to AN21 being a shorter construct (Figure 28B). Nonetheless, the AN21-Rad4 complex had different trend as DNA alone, and the relaxation rate was 7.7 ± 2.0 ms, very similar to AN3 (Figure 26).

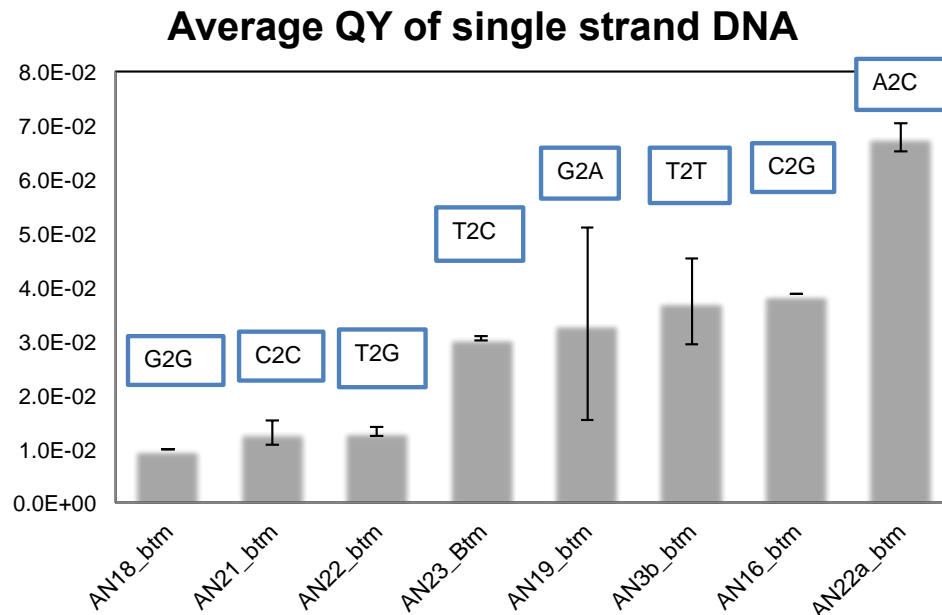


Figure 23. The summary of intensity of all the single strand DNA containing 2AP.

From the lowest on the left, to the highest on the right. The relative quantum yield (QY) is the ratio of the intensity of the sample at 370 nm and the intensity of free 2AP of same concentration at 370 nm. The bases flanking 2AP are shown above each column. “2” indicates 2AP.

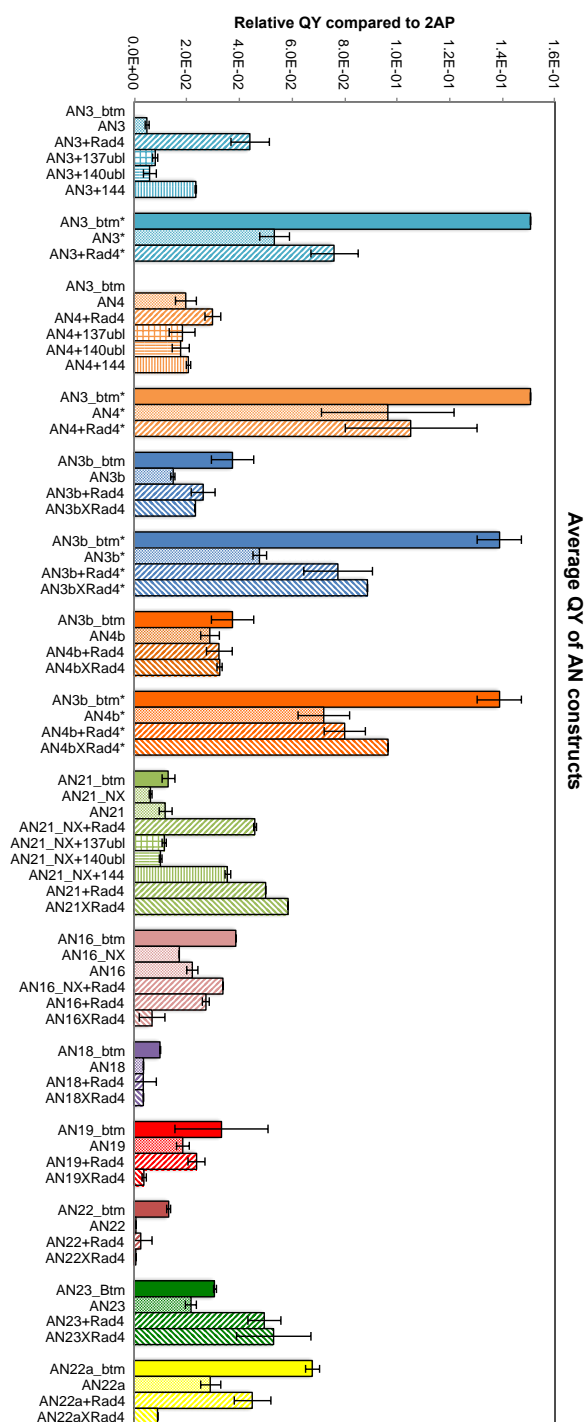


Figure 24. The summary of intensity of all constructs containing 2AP. The samples labeled with * were measured with slit width of 4 nm, while the others were measured with slit width of 5 nm. “+” indicates non-crosslinked, “X” indicates crosslinked complex. “Rad4” refers to <108>.

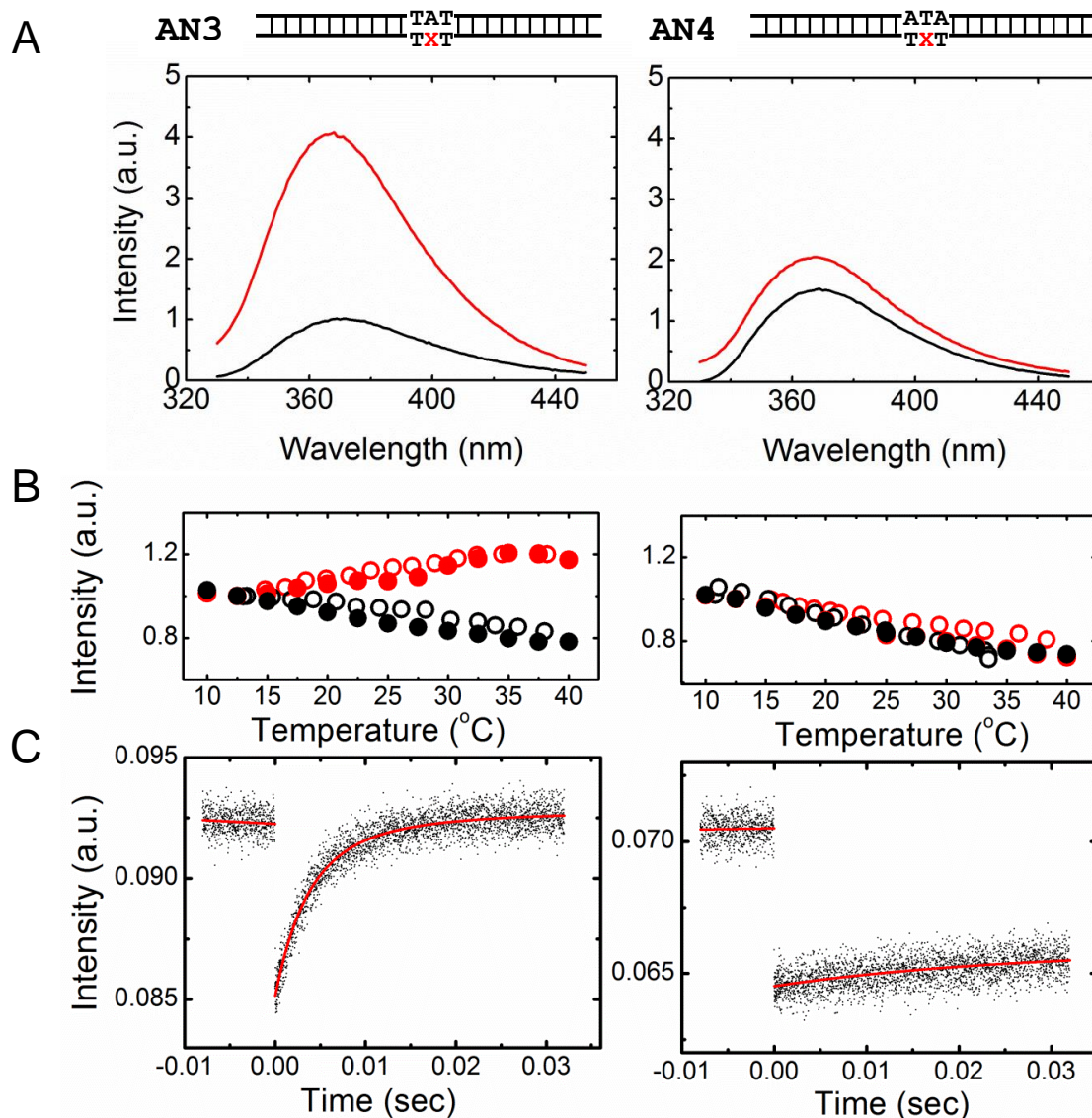


Figure 25. The fluorescence spectra, equilibrium, and T-jump trace of AN3 and AN4 bound to Rad4.

A. (top) Diagram of AN3 and AN4 sequence. (bottom) The emission spectra of each construct at 25 °C. Black: DNA only. Red: DNA bound with <108> protein. B. The change of emission intensity as a function of temperature for (left) AN3 and (right) AN4. Black: DNA only. Red: DNA bound with <108> protein. Open and filled symbols indicate two separated experiments. All the intensity at 10 °C is normalized to 1. C. The kinetics trace of (left) AN3 bound with <108> and (right) AN4 bound with <108>.

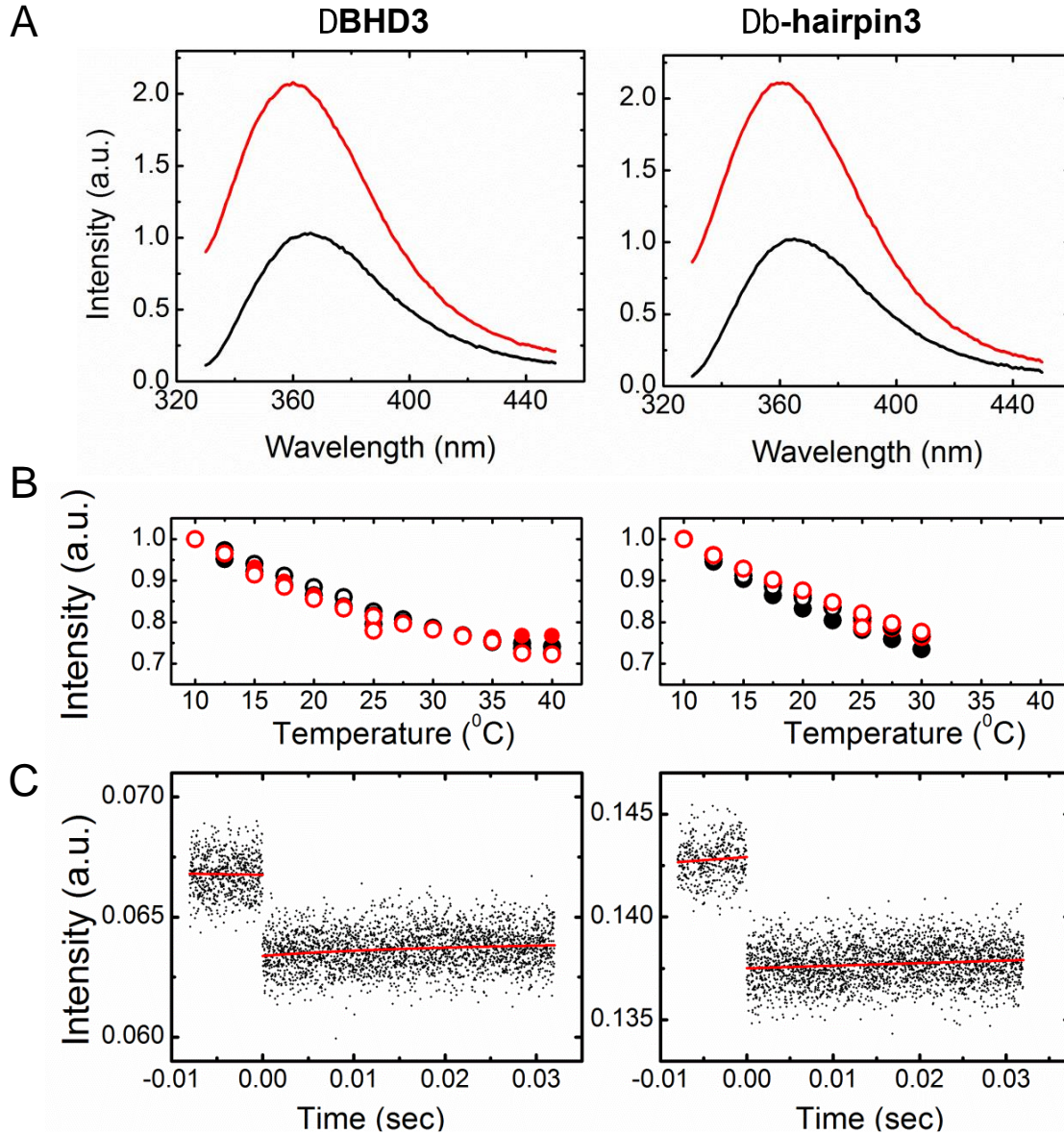


Figure 26. The fluorescence spectra, equilibrium, and T-jump trace of AN3 and AN4 bound to Rad4 mutants.

A. The emission spectra at 25 °C of AN3 bound to (left) <140>, the mutant lacking BHD3 and (right) <137>, the mutant lacking the tip of β -hairpin3. Black: DNA only. Red: DNA bound with protein. B. The change of emission intensity as a function of temperature for AN3 bound to (left) <140> and (right) <137>. Black: DNA only. Red: DNA bound with protein. Open and filled symbols indicate two separated experiments. All the intensity at 10 °C is normalized to 1. C. The kinetics trace of AN3 bound to (left) <140> and (right) <137>.

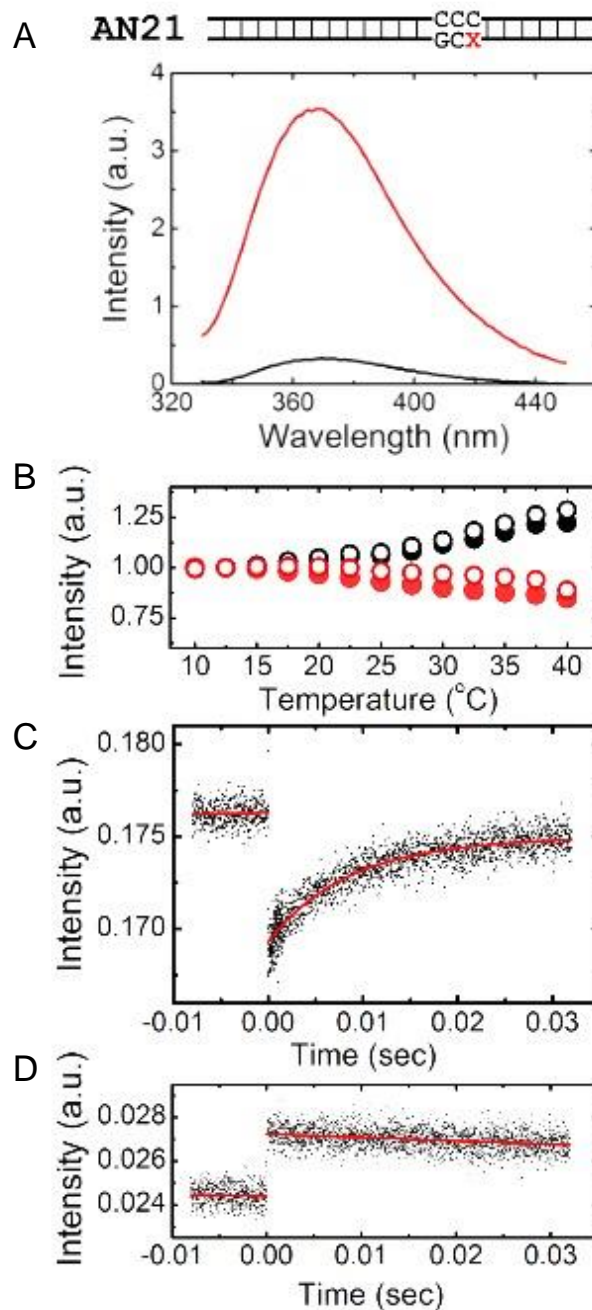


Figure 27. The fluorescence spectra, equilibrium, and T-jump trace of AN21 bound to Rad4.

A. (top) Diagram of AN21 sequence. (bottom) The emission spectra at 25 °C. Black: DNA only. Red: DNA bound with <108> protein. B. The change of emission intensity as a function of temperature. Black: DNA only. Red: DNA bound with <108> protein. Open and filled symbols indicate two separated experiments. C. The kinetics trace of AN21 bound with <108>. D. The kinetics trace of AN21 DNA.

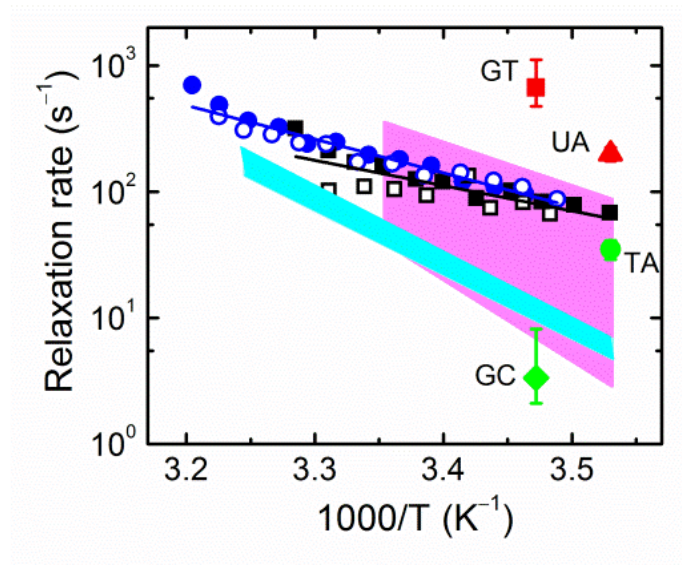


Figure 28. The relaxation rates of Rad4 bound DNA.

Blue symbol represents the rates of AN3. Black symbol represents the rates of AN21. Open and filled symbols of the same color represent 2 independent measurements. Solid lines of the same color represent the Arrhenius fit. The spontaneous base opening rates are shown as pink area for A/T, and cyan for G/C [77]. Other measured opening rate for different base pair are also shown: G/C and G/T pair [78], and T/A and U/A pair [79]. Matched pairs are shown in green and mismatched pairs in red.

4. The kinetics of Rad4 bound tC^O/tC_{nitro} constructs

(AN13 and AN15 did not showed any kinetics, possibly due to the position of the probes not being able to pick up any conformational change. Therefore, they will be excluded from the discussion.)

First, the melting temperature of all the labeled and unlabeled DNA were examined. The T_m of AN12 and AN14 were 7 – 8 °C lower than their matched counterparts due to the mismatch (Figure 29A, B). Except for AN12, the T_m of all

constructs increased slightly when the donor was incorporated, and the acceptor slightly increased their T_m further (Figure 29C-F). These showed that tC derived probes do not diminish the DNA stability, and this is consistent with previous reports which indicate that tC stacks better than C and therefore slightly increase the DNA stability [73].

The binding affinity of Rad4 to AN12 and AN14 was examined by electrophoretic mobility shift assay (EMSA), with CH7_NX (the DNA used for crystallization in Chapter 2, without the thiol linker) and its mismatch version CH10_NX as controls. AN12u showed non-specific binding similar to CH7_NX as expected, and AN12, with TTT/TTT bubble, showed slightly weaker (about 2 fold) binding affinity than CH10_NX (with CCC/CCC bubble), but still significantly stronger (about 4 fold) than match DNA (Figure 28). AN14u also behaved as CH7_NX, while AN14 (with TAT/TAT bubble) did not show any specificity in binding at all, indicating the mismatch binding affinity of Rad4 is sequence dependent (Figure 31). Moreover, none of the labeled versions differed significantly in K_d from non-labeled versions, meaning that the labels were not recognized as lesions.

The measured FRET value of AN12 and AN12u, where the donor and acceptor were placed on opposite strands and 4 nucleotides apart spanning the bubble, are consistent with the predicted value. When Rad4 was bound to AN12, there was a slight drop in FRET, but not with AN12u, indicating conformational change induced by the protein binding to the mismatch, but the change was much smaller than predicted value based on crystal structure. When mutants were bound, neither AN12 nor AN12u showed significant change. (Table 7, Figure 32A)

As temperature increased, I_D of DNA decreased due to the quantum yield of tC^O , while I_{DA} increased, indicating certain conformational change, possibly caused by local unstacking similar to the observation of AN21. Therefore the FRET of AN12, defined as $E = 1 - \frac{I_{DA}}{I_D}$, decreased at higher temperature. The change of FRET in AN12-Rad4 complex (0.08 ± 0.02) was slightly more pronounced than DNA only (0.05 ± 0.01), suggesting the contribution from protein. The FRET curve of mutant bound AN12 essentially overlapped with AN12. For AN12u with or without protein, the extent of decrease was all the same. (Figure 32B)

The FRET of AN14, AN14u and the complexes are all in good agreement with predicted value, with the donor and acceptor on the same strand and 7 nucleotides apart spanning the bubble. There was a significant drop in FRET when Rad4 was bound to both match and mismatch DNA, indicating conformational change induced by non-specifically bound Rad4. But the mutants did not induce any change in FRET. (Table 7, Figure 32C)

In equilibrium experiments, I_{DA} of AN14 did not increase like AN12 (Figure 37A), possibly because of AN14 being more stable than AN12 (Figure 29) and the fluorophores are further apart, therefore not showing local unstacking. As a result, the decrease in FRET over the measured temperature range was small. The extent of decrease in AN14 and AN14u were almost the same as their complexes, suggesting the change in the distribution of different conformation was very small, if there was any. (Figure 32D)

In T-jump experiments, all the DNA with only donor showed recovery trace as expected. None of the DNA constructs had kinetics, and the immediate change of signal (either increase or decrease) at the temperature jump was always consistent with the direction of signal change observed in equilibrium. We observed kinetics in AN12 bound to Rad4, with relaxation rate of 7.7 ± 0.8 ms at 25 °C (Figure 33), which fell in the range of base opening kinetics detected by 2AP, but no kinetics were present in AN12u bound to Rad4 (Figure 35), indicating this motion is only specific to mismatch DNA. The $\Delta\beta$ -hairpin3 mutant bound AN12 revealed kinetics that were never seen previously (227.5 ± 18.2 μ s at 25 °C), and about 30 fold faster than the ms range kinetics. It was only detected near 30 °C, the highest temperature this mutant could endure (Figure 34). Interestingly, both fast and slow kinetics were observed in Δ BHD3 mutant bound AN12, only at high temperature above 30 °C. The slow phase (11.1 ± 7.1 ms at 32 °C) was consistent with Rad4 bound AN12, and the fast phase (118.2 ± 35.2 μ s at 32 °C) was consistent with $\Delta\beta$ -hairpin3 mutant bound AN12. (Figure 34)

For AN14, kinetics was observed in both match and mismatch version bound to Rad4, 448.4 ± 98.8 μ s at 25 °C with mismatch DNA, 292.4 ± 173.9 μ s at 25 °C for match DNA, both fell in the range of fast phase observed with AN12. No kinetics was detected with mutant bound AN14 or AN14u. (Figure 36)

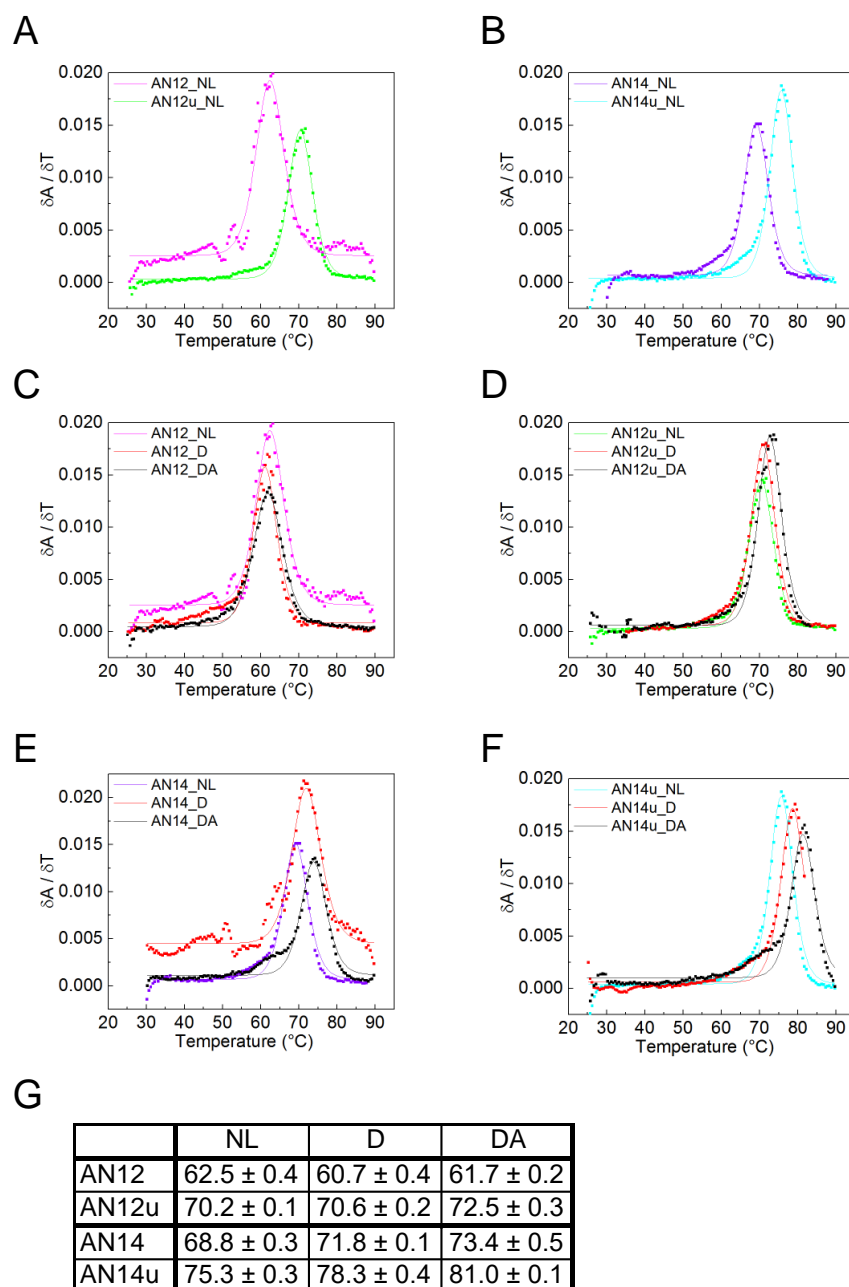


Figure 29. Melting temperature of AN12 and AN14 constructs.

A – F. The first derivative of DNA absorbance at 260 nm as a function of temperature. Solid lines are the fit curves. “_NL” indicates DNA without fluorophores; “_D” indicates DNA with tC^0 ; “_DA” indicates DNA with both tC^0 and tC_{nitro} . G. A summary of T_m in °C for all the AN12 and AN14 variants. The error is the standard deviation from the method assuming linear baselines and the derivative method.

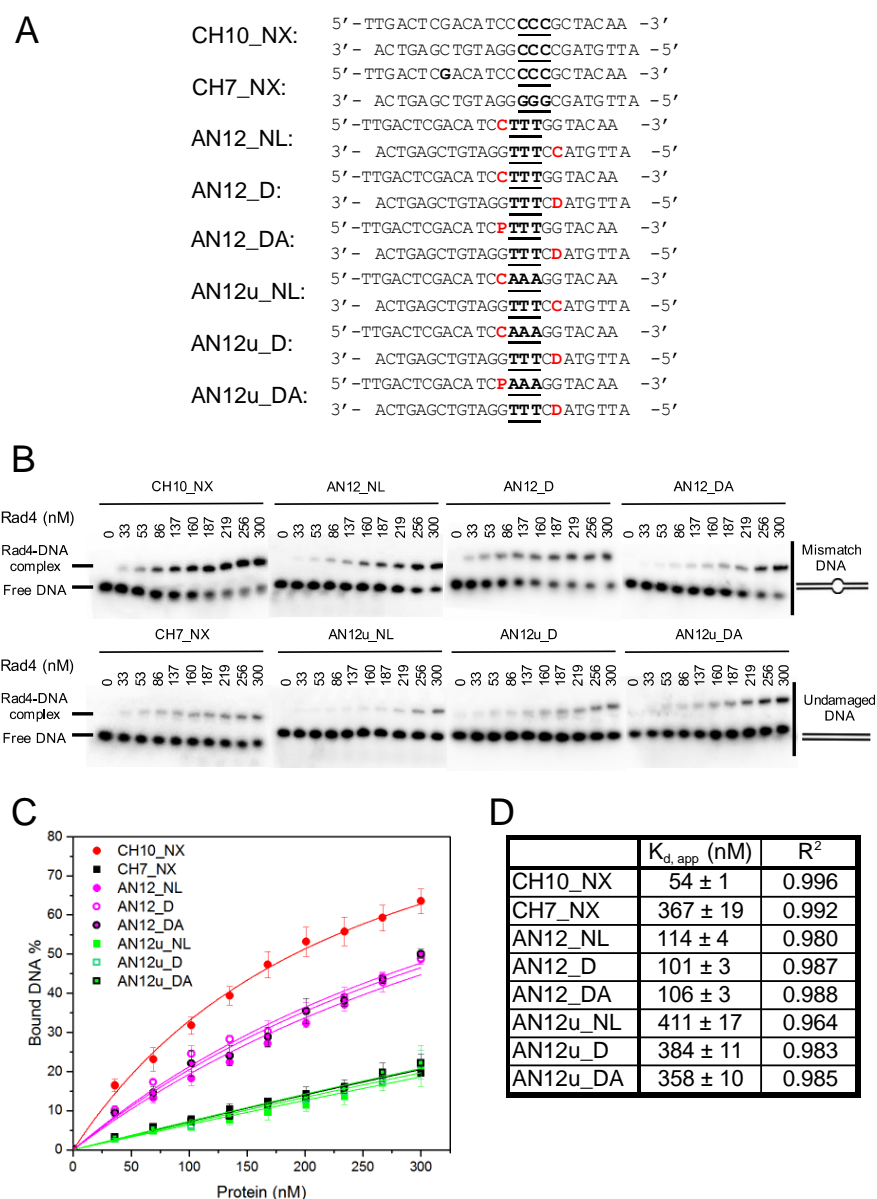


Figure 30. EMSA of Rad4 binding to AN12 constructs.

A. The sequence of the constructs used in this figure. The mismatch bubble and the corresponding position in matched version are indicated by bold and underline. “D” represents tC^O , “P” represents tC_{nitro} . Letters in red indicate the position of fluorophores. B. Exemplary gels of each construct bound to Rad4. C. Binding curve for each construct. Symbols represent the average percentage of DNA bound, and the error bars are standard deviation from 3 measurements. The solid lines are the fitted curve. D. A summary of calculated K_d . R^2 is the error of the fitting.

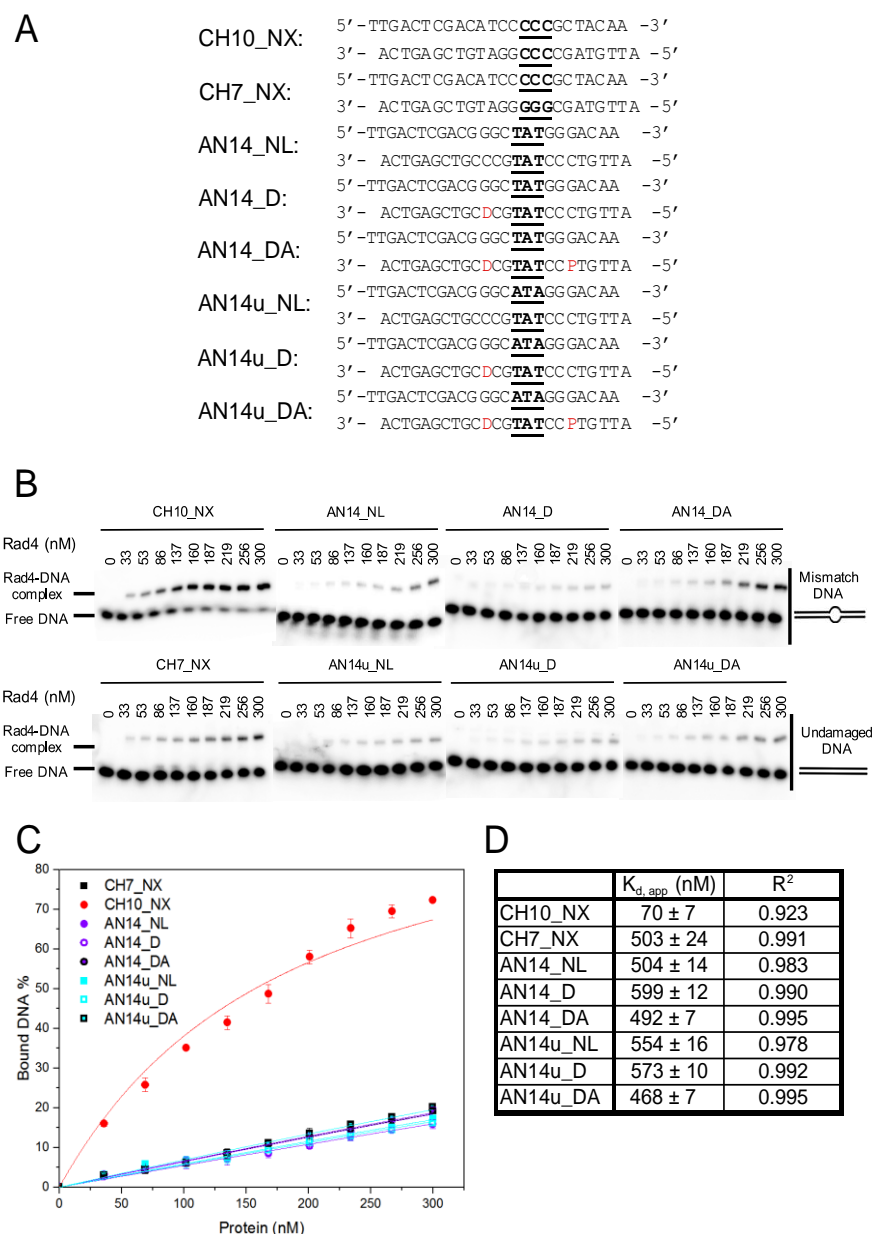


Figure 31. EMSA of Rad4 binding to AN14 constructs.

A. The sequence of the constructs used in this figure. The mismatch bubble and the corresponding position in matched version are indicated by bold and underline. “D” represents tC^O, “P” represents tC_{nitro}. Letters in red indicate the position of fluorophores. B. Exemplary gels of each construct bound to Rad4. C. Binding curve for each construct. Symbols represent the average percentage of DNA bound, and the error bars are standard deviation from 3 measurements. The solid lines are the fitted curve. D. A summary of calculated K_d . R^2 is the error of the fitting.

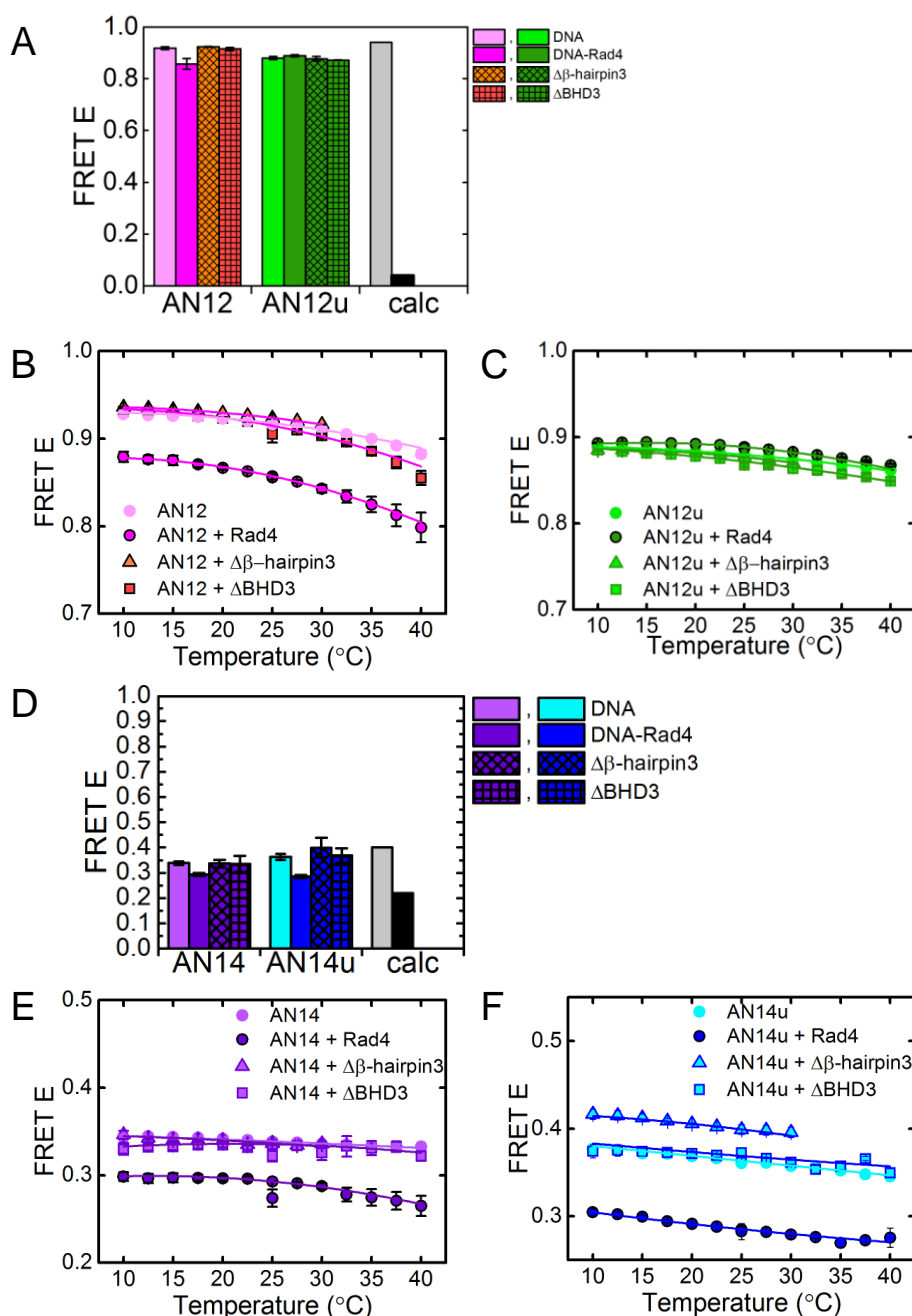


Figure 32. FRET of AN12 and AN14.

A, D. FRET efficiency of donor and acceptor labeled DNA bound to Rad4 or mutants. Grey column: predicted FRET of match DNA; black column: predicted value of match DNA bound to Rad4. B, C, E, F. The change in FRET efficiency as temperature increases. The columns and symbols represent average FRET value from at least 2 equilibrium experiments; the error bars are standard error of the mean.

Table 7. A summary of measured and predicted FRET for all DNA and complexes.

The protein names indicate the complexes formed with the protein and DNA. The predicted FRET for DNA is based on matched sequence, the one for complex is based on the crystal structure of Rad4 bound to mismatched sequence. SEM indicates standard error of the mean.

		DNA only	Rad4	$\Delta\beta$ -hairpin3	Δ BHD3	DNA predicted	Rad4 predicted
AN12	Average FRET	0.92	0.86	0.92	0.91		0.04
	SEM	0.01	0.02	0.00	0.01		
	Number of measurements	4	2	2	4		
AN12u	Average FRET	0.88	0.89	0.88	0.87	0.94	
	SEM	0.01	0.00	0.01	0.00		
	Number of measurements	4	2	2	2		
AN14	Average FRET	0.34	0.29	0.34	0.34		0.22
	SEM	0.01	0.00	0.01	0.03		
	Number of measurements	5	4	2	5		
AN14u	Average FRET	0.36	0.29	0.40	0.37	0.40	
	SEM	0.01	0.01	0.04	0.03		
	Number of measurements	5	2	2	3		

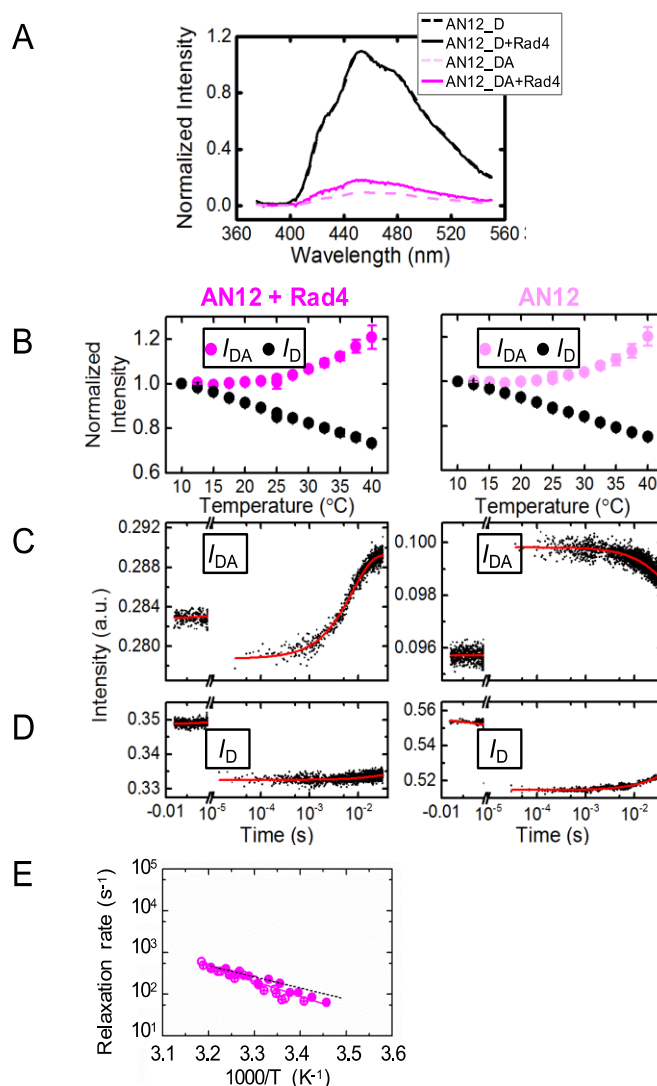


Figure 33. The equilibrium and T-jump trace of AN12 alone and bound to Rad4.

A. Exemplary emission spectra of AN12_D and AN12_DA with or without Rad4 at 25 °C. Each _D and _DA pair were normalized so that the _D spectra overlap. B. The change of donor intensity in the complex (left) and DNA alone (right) as temperature increases. I_{DA} indicates donor intensity in _DA, I_D indicates donor intensity in _D. C. The kinetics trace of AN12_DA bound to Rad4 (left) and by itself (right). D. The kinetics trace of AN12_D bound to Rad4 (left) and by itself (right). E. The Arrhenius plot of AN12_DA bound to Rad4. The different symbols indicate relaxation rates from 3 measurements; the pink line is the Arrhenius fit; the black dashed line is the relaxation rate of AN3 bound to Rad4 from Chapter 2.

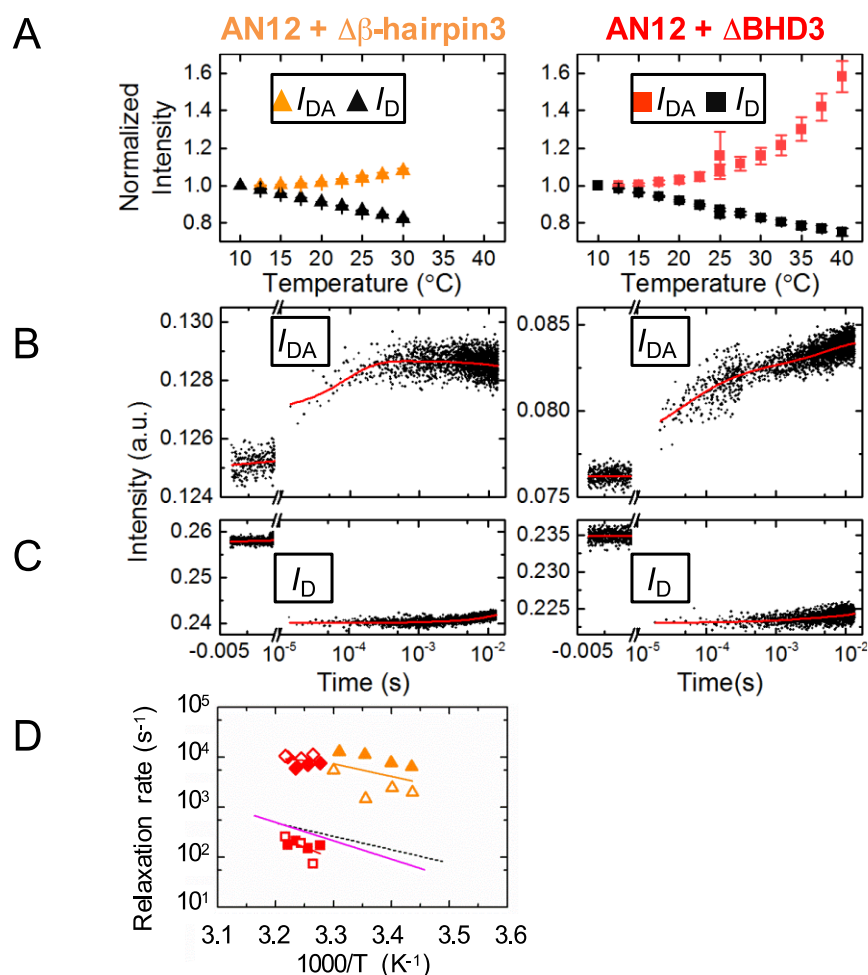


Figure 34. The equilibrium and T-jump trace of AN12 bound to Rad4 mutants.

A. The change of donor intensity in AN12 bound to Rad4 mutant lacking β -hairpin3 (left) and mutant lacking BHD3 (right) as temperature increases. I_{DA} indicates donor intensity in $_{DA}$, I_D indicates donor intensity in $_D$. B. The kinetics trace of AN12_{DA} bound to β -hairpin3 mutant (left) and BHD3 mutant (right). C. The kinetics trace of AN12_D bound to β -hairpin3 mutant (left) and BHD3 mutant (right). E. The Arrhenius plot of AN12_{DA} bound to Rad4 mutants. Orange symbol: β -hairpin3 mutant. Red symbol: BHD3 mutant; square indicates slow phase and diamond indicates fast phase. Open and filled symbols are 2 measurements. Solid lines are the Arrhenius fit for the rates of the same color. The pink line is the relaxation rate of AN12_{DA} bound to Rad4 from Figure 33; the black dashed line is the relaxation rate of AN3 bound to Rad4 from Chapter 2.

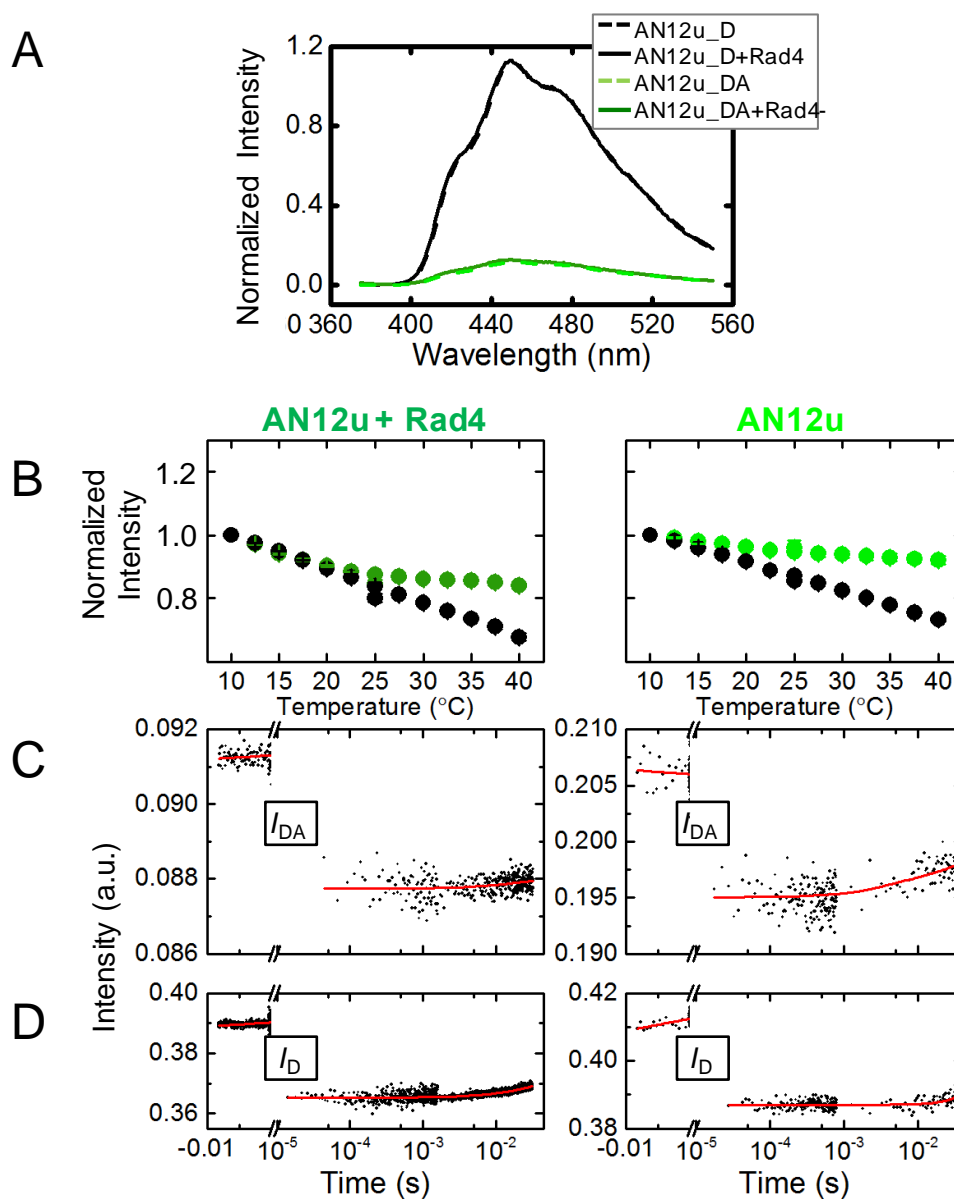


Figure 35. The equilibrium and T-jump trace of AN12u alone and bound to Rad4.

A. Exemplary emission spectra of AN12u_D and AN12u_DA with or without Rad4 at 25 °C. Each _D and _DA pair were normalized so that the _D spectra overlap. B. The change of donor intensity in the complex (left) and DNA alone (right) as temperature increases. I_{DA} indicates donor intensity in _DA, I_D indicates donor intensity in _D. C. The kinetics trace of AN12u_DA bound to Rad4 (left) and by itself (right). D. The kinetics trace of AN12u_D bound to Rad4 (left) and by itself (right).

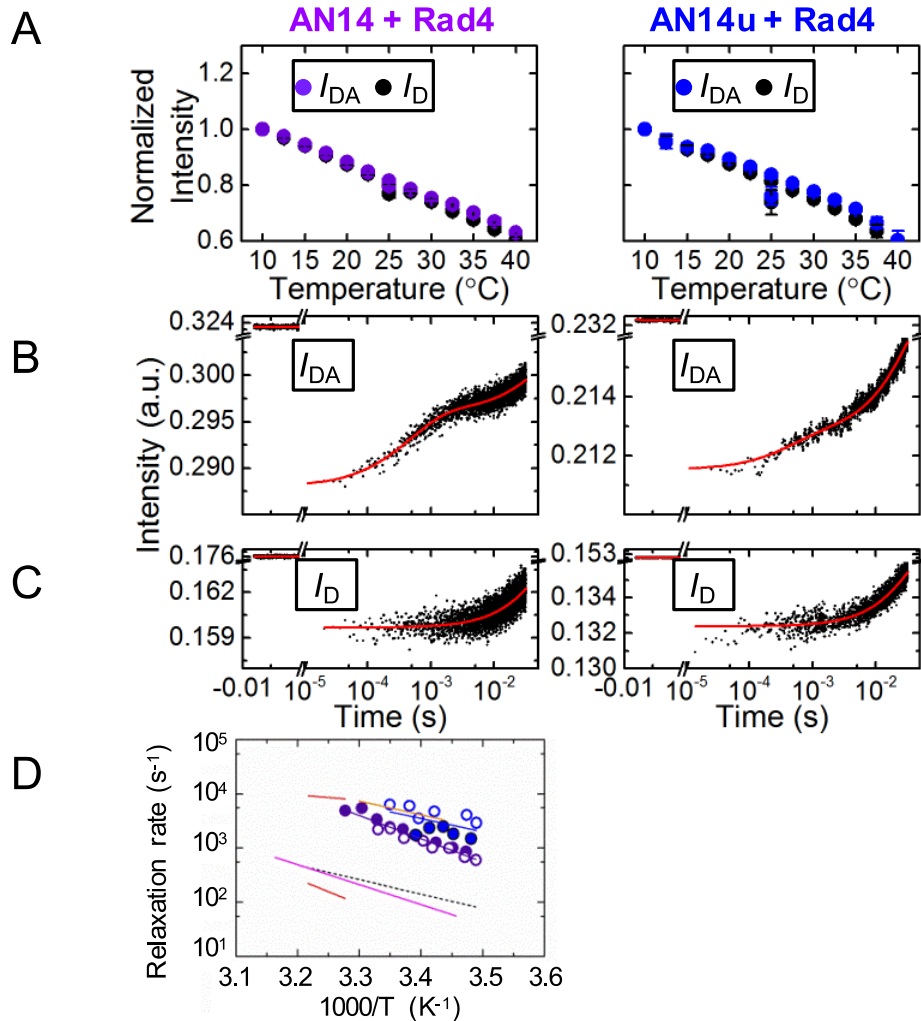


Figure 36. The equilibrium and T-jump trace of AN14 and AN14u bound to Rad4.

A. The change of donor intensity in the Rad4 complex with AN14 (left) and AN14u (right) as temperature increases. I_{DA} indicates donor intensity in $_{\text{DA}}$, I_{D} indicates donor intensity in $_{\text{D}}$. B. The kinetics trace of AN14_{DA}-Rad4 (left) and AN14u_{DA}-Rad4 (right). C. The kinetics trace of AN14_D-Rad4 (left) and AN14u_D-Rad4 (right). D. The Arrhenius plot of AN14_{DA} and AN14u_{DA} bound to Rad4. Purple: AN14_{DA}; Blue: AN14u_{DA}. The open and filled symbols indicate relaxation rates from 2 measurements; Solid lines are the Arrhenius fit for the rates of the same color. The pink line is the relaxation rate of AN12_{DA} bound to Rad4 from Figure 33; the orange and red lines are the relaxation rates of AN12_{DA} bound to mutants from Figure 34; the black dashed line is the relaxation rate of AN3 bound to Rad4 from Chapter 2.

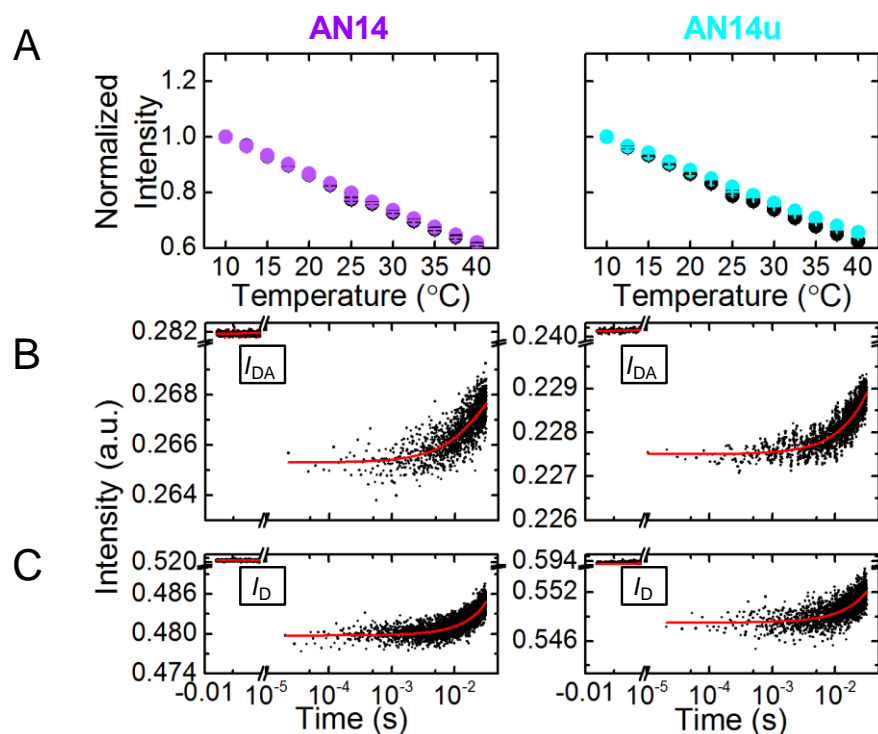


Figure 37. The equilibrium and T-jump trace of AN14 alone and AN14u alone.

A. The change of donor intensity in AN14 (left) and AN14u (right) as temperature increases. I_{DA} indicates donor intensity in $_{DA}$, I_D indicates donor intensity in $_D$. B. The kinetics trace of AN14 $_{DA}$ (left) and AN14u $_{DA}$ (right). C. The kinetics trace of AN14 $_D$ (left) and AN14u $_D$ (right).

D. Discussion

2AP has been used by many researchers to probe DNA conformational change and protein-DNA interaction, because of its sensitivity to local environment. The quantum yield of 2AP decreases when incorporated in DNA duplex, due to the stacking with the neighboring bases, although the molecular mechanism of the quenching is still unclear. Some reported that the quenching effect was similar for all bases [80], while some reported different quenching efficiency, with G being the highest, followed by T, A, C [81].

Examining the single strands containing 2AP used in our study, it seemed that the quenching was sequence-dependent: the fluorescence intensity of different strands varied up to 7 fold, and there did not appear to be a clear trend on which nucleotides affected the signal more than others. Nonetheless, all the single strands were quenched, showing less than 10% of the intensity of free 2AP of same concentration, and the duplex were further quenched, with all the DNA samples showing the characteristic emission peak at 370 nm as expected.

The increase in signal of mismatched DNA when protein was bound was always greater than their matched DNA counterpart, this is consistent with Rad4's higher affinity to mismatch site, and 2AP was able to detect the conformational change at mismatch site induced by Rad4. The extent of increase and the final intensity upon Rad4 binding varied for different samples, which was likely to be sequence dependent as well.

When Rad4-bound AN3 and AN21 were subjected to increasing temperature, the signal significantly deviated from DNA itself, indicating the population of different

conformation (opened and closed conformation) was shifted. Consistent with this, kinetics was observed for both complexes. Similar relaxation time of ~ 7 ms was observed, which should correspond to the time for the closed conformation to become opened, or, the time for the nucleotide to flip out. This is quite slower than the spontaneous base opening rate (Figure 28), indicating that Rad4 induced opening involves complex conformational change at the mismatch, and Rad4 has to overcome certain energy barrier to form the open complex.

The mutant complex and the match DNA-protein complex did not deviate from DNA itself, and no kinetics was observed. This indicated that the β -hairpin was required to induce the observed kinetics, consistent with the crystal structure showing the β -hairpin inserting into the DNA and pushing out the nucleotides. As for the matched DNA, Rad4 was shown to be able to form open conformation in the tethered crystal structure, but in solution fluorescence assays under non-tethered conditions, the opening event was likely very rare due to the non-specific binding in heterogeneous registers, or, the time it took to open a match site was too long to be detected in the T-jump time window.

Evidences have shown that Rad4 easily recognizes lesions that distort the helix structure, but has difficulty with helix-stabilizing lesions [82]. This indicated Rad4 detects weak points in the helix structure. The energy barrier of flipping out a helix-destabilizing, repair-susceptible lesion was estimated to be ~ 7.7 kcal/mol lower than undamaged DNA [83] .

These results combined lead to the conclusion that Rad4 has to overcome higher energy barrier, and require longer time to open a match site than a lesion. It is likely that

the residence time of Rad4 while searching at a certain site is not enough for it to induce full opening. On the other hand, a helix-destabilizing lesion stalls Rad4 and prolongs its residence time, plus the time (~ 7 ms) and energy it takes to open such lesion is smaller, helping Rad4 easily overcoming the energy barrier and induce full opening. The crosslinked complex is a case where Rad4 is permanently stalled at an undamaged site, therefore it was given more than enough time to induce opening. This “kinetic gating” mechanism revealed how Rad4 is able to efficiently distinguish between undamaged DNA and various structurally diverse lesions.

The DNA kinetics induced by Rad4 were also studied using novel FRET pair tC^O and tC_{nitro} . This is the first time tC fluorophores being applied to probe DNA-protein interaction. tC^O and tC_{nitro} behaved nicely and as expected: they did not distort the normal conformation of B-DNA; they were rigidly stacked in the duplex and not recognized as lesions by Rad4; they were also quite sensitive to the conformational change of DNA, and the change in FRET upon protein binding was the same direction as predicted, supporting that the motion they sensed was indeed DNA unwinding.

The FRET of AN12, AN14 and AN14u decreased when Rad4 was bound, indicating conformational change induced by Rad4, which we interpret to be unwinding. Mutant bound AN12 did not show noticeable change, but kinetics was observed later in T-jump, suggesting the conformation change in these complexes were subtle, and that T-jump is a very sensitive detection technique. Most of the FRET values of the DNA-only samples agreed with prediction (Table 7), except that AN12-Rad4 did not decrease as much as predicted. A possible reason is that the prediction was based on the crystal

structure of the complex, which is a snap shot of the possible conformations, while FRET measures an ensemble of molecules at equilibrium, therefore it represents an averaged conformation, and the crystal structure represents an extreme.

The difference in FRET of AN12-Rad4 and AN12 propagated as temperature increased, hinting the population distribution among different conformations shifted, probably from less to more unwound, which was then detected by T-jump. AN14-Rad4 or AN14u-Rad4 did not deviate noticeable from the trend of DNA, but their kinetics were still detected by T-jump, again demonstrating the sensitivity of this method.

No kinetics were observed with DNA not bound by protein, due to the lack of induced conformational change, or with the samples containing only donor, due to the absence of the quencher. These confirmed that the observed kinetics came from DNA conformational change induced by protein.

T-jump revealed two different phases of Rad4 induced kinetics. The “slow” phase, with relaxation time of $\sim 7 - 10$ ms, was observed in AN12-Rad4 complex, and in AN12- Δ BHD3 complex at higher temperature with lower amplitude. It was not detected in any match DNA samples, or AN14 samples, all of which are non-specific complexes. And it does not require β -hairpin3. The only exception was AN12- $\Delta\beta$ -hairpin3 not showing the slow phase, possibly because the mutant complexes only show slow kinetics at high temperature (as seen in AN12- Δ BHD3) due to their lower specific than wild type, and $\Delta\beta$ -hairpin3 mutant was not stable therefore was not able to be heated to such high temperature. This slow phase coincides with the $\sim 5 - 7$ ms base opening kinetics detected

by 2AP, indicating these motions occur on similar time scale, and both involve mismatch site in DNA. While the nucleotide flipping kinetics required β -hairpin3, the tC^O/tC_{nitro} kinetics did not, suggesting they may detect different motions with a common rate-limiting step. In the event of fully opening a mismatch site, the slow phase might represent Rad4 twisting the DNA open, and the 2AP kinetics may capture the β -hairpin3 insertion into DNA and flipping out the base, both contributing to a fully opened stable complex.

Interestingly, a “fast” phase was also observed, with relaxation time of $\sim 100 - 500 \mu\text{s}$, more than 10-fold faster than the slow phase, observed with AN12- Δ BHD3, AN12- $\Delta\beta$ -hairpin3, AN14-Rad4 and AN14u-Rad4. Evidently it does not require β -hairpin3 either, and the complexes showing fast phase are all non-specifically bound (either with specificity-compromised mutant or low-specificity substrate), indicating this is a non-specific unwinding motion. It was not observed in specific complex AN12-Rad4, because most of the protein were bound to the mismatch and they were able to fully open the site, causing the slow kinetics to dominate the signal.

AN14 and AN14u, although one is mismatch and the other is match, both behaved like non-specific substrate for Rad4, so only fast phase was detected. With the low signal amplitude of Rad4 complex, the mutant complexes were likely to be even lower, so it was not surprising that we were not able to observe kinetics.

AN12u, a non-specific substrate like AN14/AN14u, did not show fast kinetics with any protein. It was possible that the signal was too low to be detected, even for

AN12u-Rad4. Also, the probes in AN12u were placed closer to each other than in AN14/AN14u, as a result, they were only able to sense the conformational change in a smaller range, assuming the protein binding position is random on these DNA.

In summary, a newly-discovered kinetics was captured by tC^O/tC_{nitro} , a FRET pair which we interpret as a nonspecific interrogation through twisting. With time scales in hundreds of μs range, it preceded the event of fully opening the DNA. Also, this kinetics was not specific to mismatch, and did not require β -hairpin3 which is known to be essential for base opening. Based on these evidence, the fast phase revealed by T-jump likely represents a twisting motion along the DNA induced by Rad4, to interrogate and search for any structural weak point as it moves on DNA without slowing down. When a lesion is found, it increases the residence time of Rad4 so that it could open the lesion site and form a specific complex, including DNA unwinding and nucleotide flipping, leading to fully “recognition” of the lesion.

IV. STRUCTURAL STUDY OF RAD4-RAD23-RAD33 TRIMERIC COMPLEX

A. Introduction

Recently it was discovered that XPC complex contains a third protein, Centrin 2 (Cen2), which directly interacts with XPC [84]. Human centrin is a calcium binding EF-hand protein found in centrosome, with three isoforms: centrin 1, 2, and 3 [85]. Cen2 was found to further stabilize XPC in the presence of Rad23B, and slightly increases the NER activity of purified XPC-Rad23B both in XP-C complementation assay and cell-free NER incision [84, 86]. It was also shown to stimulate DNA binding of XPC [87], providing a probable explanation for its effect on NER activity. The region of XPC necessary and proficient to bind Cen2 was mapped to be C-terminal residues 847 – 866 [86], which lies within the TFIIH interaction region (816 – 940).

The yeast centrin, Cdc31, which is a homolog of Cen3, was found to co-purify with Rad4, and yeast strain lacking Cdc31 showed UV sensitivity [88]. On the other hand, a structural homolog of Cdc31 (based on structure prediction), Rad33, was also shown to directly bind Rad4 domain containing residues 274-667, where the conserved residues required for Cen2 binding were found to be essential for Rad33 interaction as well. And similar to *cdc31* mutant, *rad33* yeast cells showed elevated UV sensitivity [89].

In order to investigate the role of Cen2 in the Rad4 damage recognition complex and in NER, we plan to construct the trimeric complex of Rad4-Rad23-Rad33/Cdc31 and study its structure.

B. Materials and Methods

1. Cloning of Rad4-Rad23 complex and Rad33/Cdc31

Three different truncation of Rad4 101-667, 101-672, 101-754 were cloned from full-length Rad4 by PCR, and inserted between BssHII and NotI in pFBD vector (Invitrogen) with His tag engineered between BamHI and BssHII: They are named as constructs XC45, XC46, XC47 respectively. For all 3 constructs, Rad23_Δ135-299/Thrombin-site was inserted between NheI and SphI in the same vector.

Full-length Rad33 and full-length Cdc31 were cloned from yeast genome library, and each inserted between NheI and SphI in pFBD vector, with GST tag engineered between NcoI and XhoI: They are named as constructs KS1, XC52 respectively.

2. Cloning of trimeric complex in MultiBac system

Rad4_101-667 was inserted between BssHII and NotI in acceptor vector pFL (provided by Dr. Yuichiro Takagi, Indiana University School of Medicine) with His tag engineered between BamHI and BssHII. Then Rad23_Δ135-299/Thr was inserted between NheI and SphI in the same vector: construct XC50. Full-length Rad33 or Cdc31 was inserted between NheI and SphI in donor vector pUCDM (provided by Dr. Yuichiro Takagi): They are named as constructs XC51 and XC53 respectively.

Approximately same mass of donor (XC51 or XC53) and acceptor (XC50) plasmid (~ 500 ng each) were mixed with 1 unit of Cre recombinase in 1X Cre buffer (New England Biolabs) in a total 10 μ l reaction, incubated at 37 °C for 30 min (incubation for too long will not improve the efficiency, and could cause fusion of multiple copies of plasmid) [90]. When the mixture was separated on 1% agarose gel, faint bands that migrated slower than either plasmid could be seen, indicating successful recombination of the two. Then the mixture was transformed in DH5 α cells and plated on LB agar containing both ampicillin and chloramphenicol. The positive colony was amplified and the plasmid, which contained three genes, was extracted. Construct XC54 is the recombination of XC50 and XC51, and XC55 is the recombination of XC50 and XC53. See table 7 for a list of all the constructs used in this chapter.

Table 8. A summary of protein constructs used in chapter 4.

Construct name	Insert 1 (ph promoter)	Insert 2 (p10 promoter)	Insert 3	Vector
XC45	(His-)Rad4 101-667	Rad23 d135-229/Thr	--	pFastBacDual
XC46	(His-)Rad4 101-672	Rad23 d135-229/Thr	--	pFastBacDual
XC47	(His-)Rad4 101-754	Rad23 d135-229/Thr	--	pFastBacDual
XC49	(His-)Rad4 101-667	--	--	pFL
XC50	(His-)Rad4 101-667	Rad23 d135-229/Thr	--	pFL
KS1	--	(GST-)Rad33	--	pFastBacDual
XC51	--	--	Rad33	pUCDM
XC52	--	(GST-)Cdc31	--	pFastBacDual
XC53	--	--	Cdc31	pUCDM
XC54	(His-)Rad4 101-667	Rad23 d135-229/Thr	Rad33	pFL-pUCDM recombinant
XC55	(His-)Rad4 101-667	Rad23 d135-229/Thr	Cdc31	pFL-pUCDM recombinant

3. Virus production

See protocol d and e.

For co-infection, Hi5 cells were infected with a 1:1 mixture of the two viruses.

4. Expression test using NE-PER kit (Thermo Scientific) and GST pull-down

See protocol f for expression.

One Hi5 plate of cells ($\sim 3.5 \times 10^7$) were suspended in 200 μ l of hypotonic lysis buffer (25 mM Tris pH 7.5, 25 mM NaCl, 0.5 mM TCEP, 1 mM PMSF, 1 μ g/mL Leupeptin, 1 μ g/mL Pepstatin). The mixture was added to 1.2 ml of CERI buffer (10 mM Tris pH 8.0, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM TCEP), mixed on vortexer, and incubated on ice for 10 min. Next 66 μ l of CERII buffer (10% NP-40, 0.5 mM TCEP) was added, mixed on vortexer and incubated on ice for 1 min. After another short vortexing, the mixture was spun down at 13,000 rpm for 5 min. Then the supernatant, which contained cytoplasmic extract, was saved, and 0.6 ml of NER buffer (20 mM Tris pH 8.0, 1.5 mM MgCl₂, 0.42 M NaCl, 20% glycerol, 0.5 mM TCEP) was added to the pellet. The mixture was incubated on ice for 40 min, with mixing on vortexer for 15 s every 10 min. The mixture was spun down at 13,000 rpm for 10 min, separating the supernatant, which contained nuclear extract and the insoluble fraction in the pellet. All three fractions (cytoplasmic, nuclear, insoluble) were examined on SDS-PAGE.

200 μ l of glutathione agarose 50% slurry (Thermo Scientific) was briefly spun down. The storage buffer was removed and 100 μ l of binding buffer (50 mM Tris pH 8.0,

200 mM NaCl, 10 mM DTT) was mixed into the resin. The above step was repeated two more times, to equilibrate the resin in binding buffer. Then the resin slurry was mixed with cytoplasmic or nuclear fraction obtained from expression test, incubated with gentle agitation at 4°C for 2 h. After the protein binding to the resin, the mixture was spun down at 300 rpm briefly. The supernatant, and 10 µl of resin was saved. The resin was then washed with 100 µl of binding buffer twice, followed by eluting with 100 µl of elution buffer (50 mM Tris pH 8.0, 200 mM NaCl, 10 mM DTT, 10 mM reduced glutathione) twice. The supernatant, saved resin, wash fraction and elution were examined on SDS-PAGE.

5. Large scale protein purification

See protocol f and g.

The trimeric complex eluted at ~ 200 – 240 mM NaCl on Source Q. The fractions containing the complex were either cleaved with thrombin or not cleaved. On Source S the cleaved complex with UBL still bound eluted at ~ 250 – 280 mM NaCl, the one without UBL eluted at ~ 280 – 290 mM NaCl, while the uncut complex eluted at ~ 300 – 370 mM NaCl. After Source S, the protein was concentrated to around 14 mg/ml. The cut trimeric complex with UBL eluted around 11.7 ml on Superdex 200, the cut complex without UBL eluted around 12.2 ml, and the uncut complex eluted at 11.9 – 12.0 ml. The sample was finally concentrated to around 14 mg/ml.

6. Dialysis of protein–DNA complex

The purified protein and DNA were mixed at 1:1.1 molar ratio in Slide-A-Lyzer™ Dialysis Cassettes (Thermo Scientific), dialyzed against 1 L of dialysis buffer (5 mM BTP pH 6.8, 200 mM NaCl, 1 mM CaCl₂, 10 mM DTT) for 3 h at 4 °C.

7. Crystallization

The crystallization drop contained 1 µl of the protein or the protein-DNA complex mixed with 1 µl of crystallization buffer, and was set up with hanging drop diffusion.

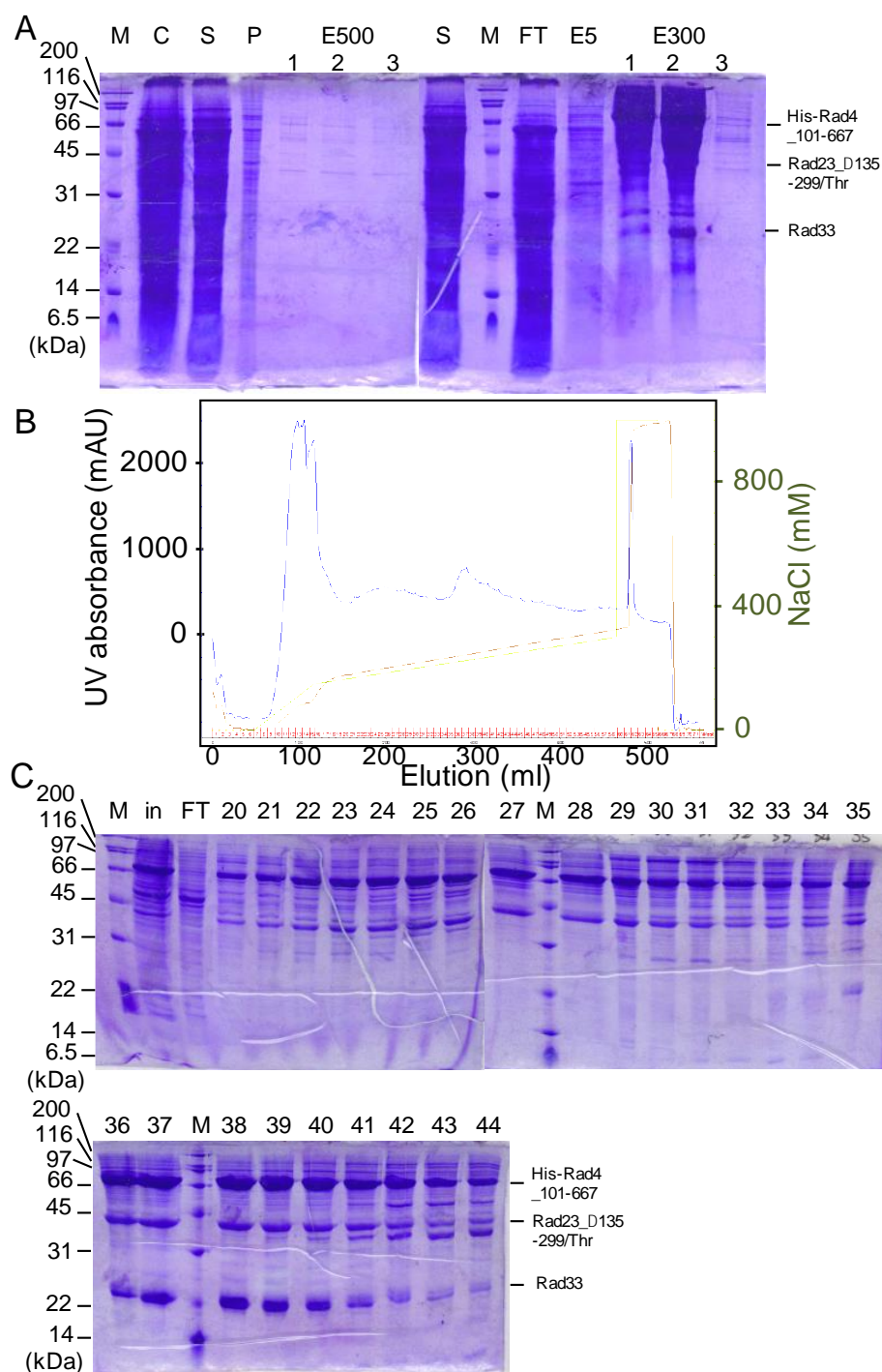


Figure 38. Lysis, affinity column and SourceQ purification of XC54. A. Lysis and Ni pull-down of XC54. See protocol g for labels. B. An exemplary SourceQ (24 ml) chromatogram for XC54. C. The fractions from B examined on SDS-PAGE. #35 – 40 were collected.

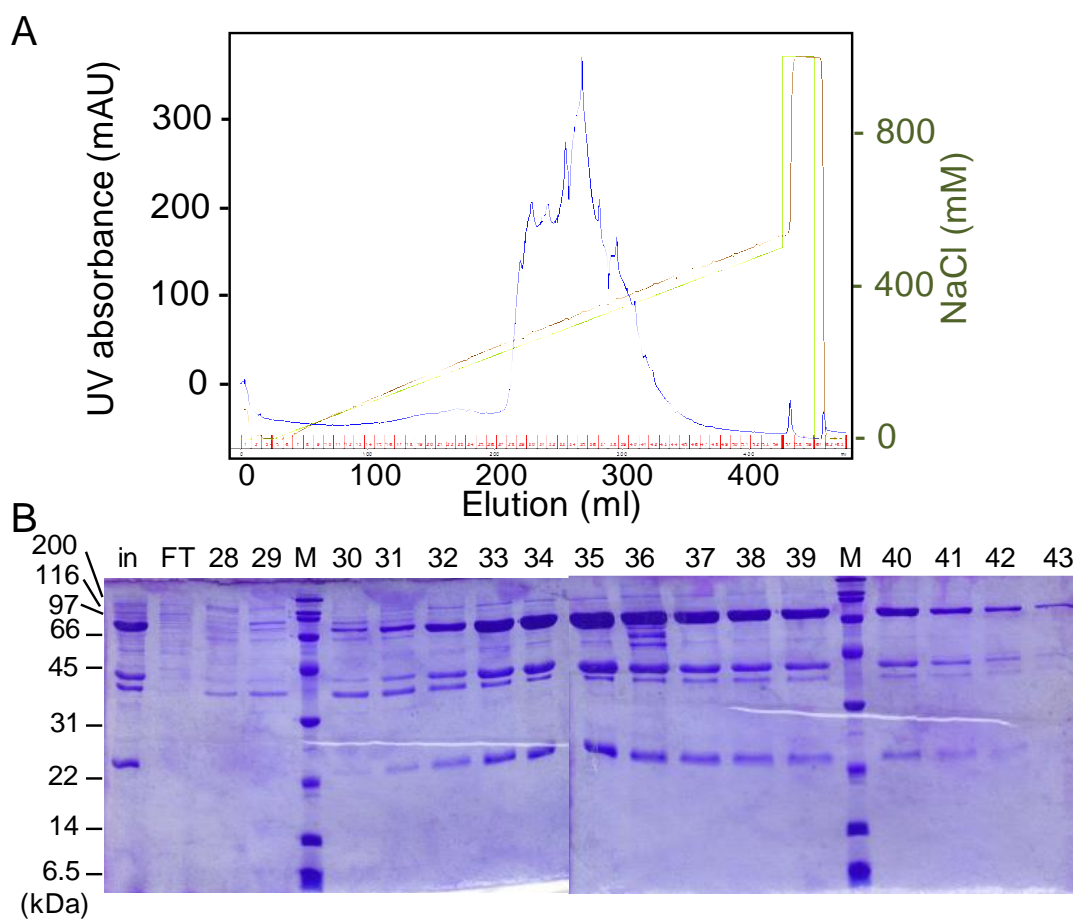


Figure 39. Purification of XC54 by SourceS.

A. An exemplary SourceS (8 ml) chromatogram for XC54 (uncut). B. The fractions from A examined on SDS-PAGE. #35 – 41 were collected.

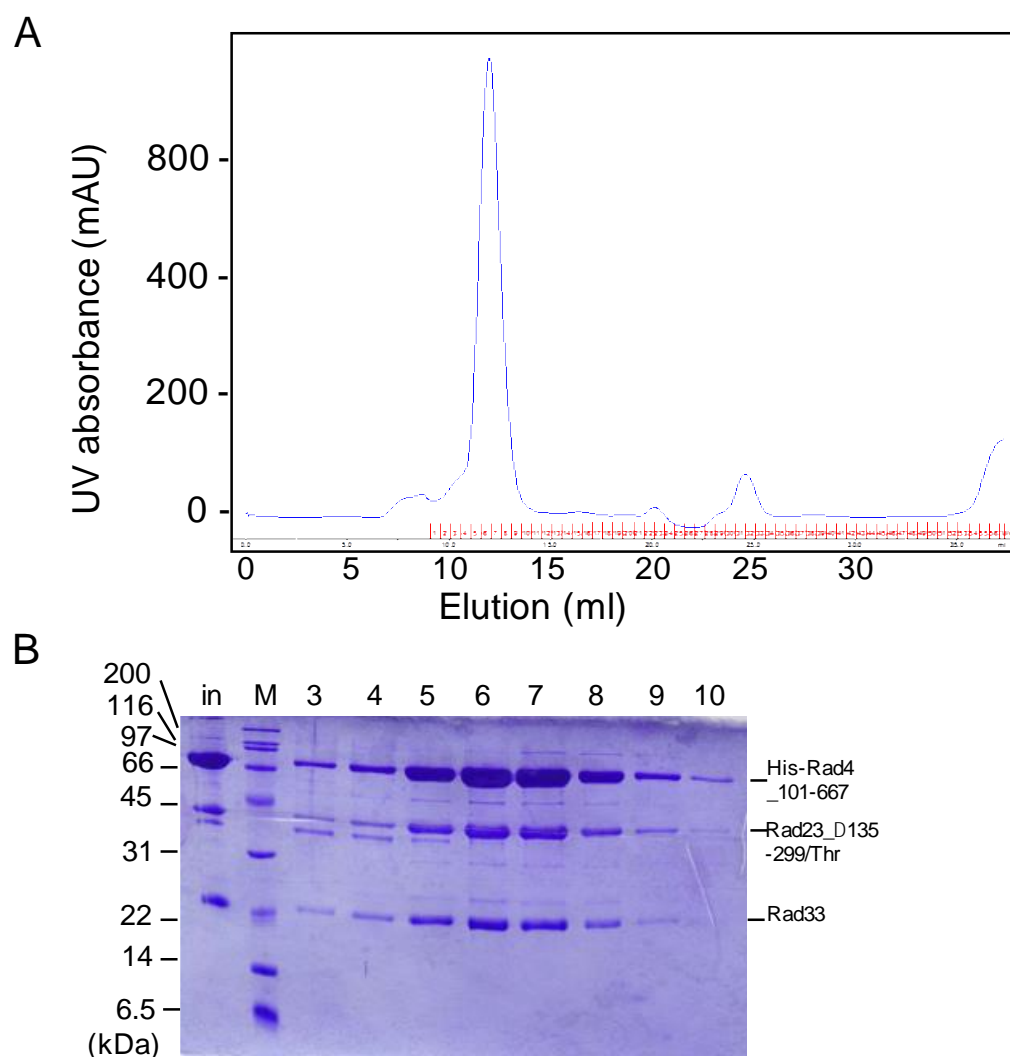


Figure 40. Purification of XC54 by Superdex200.

A. An exemplary Superdex 200 (24 ml) chromatogram for XC54 (uncut).

B. The fractions from A examined on SDS-PAGE. #5 – 8 were collected.

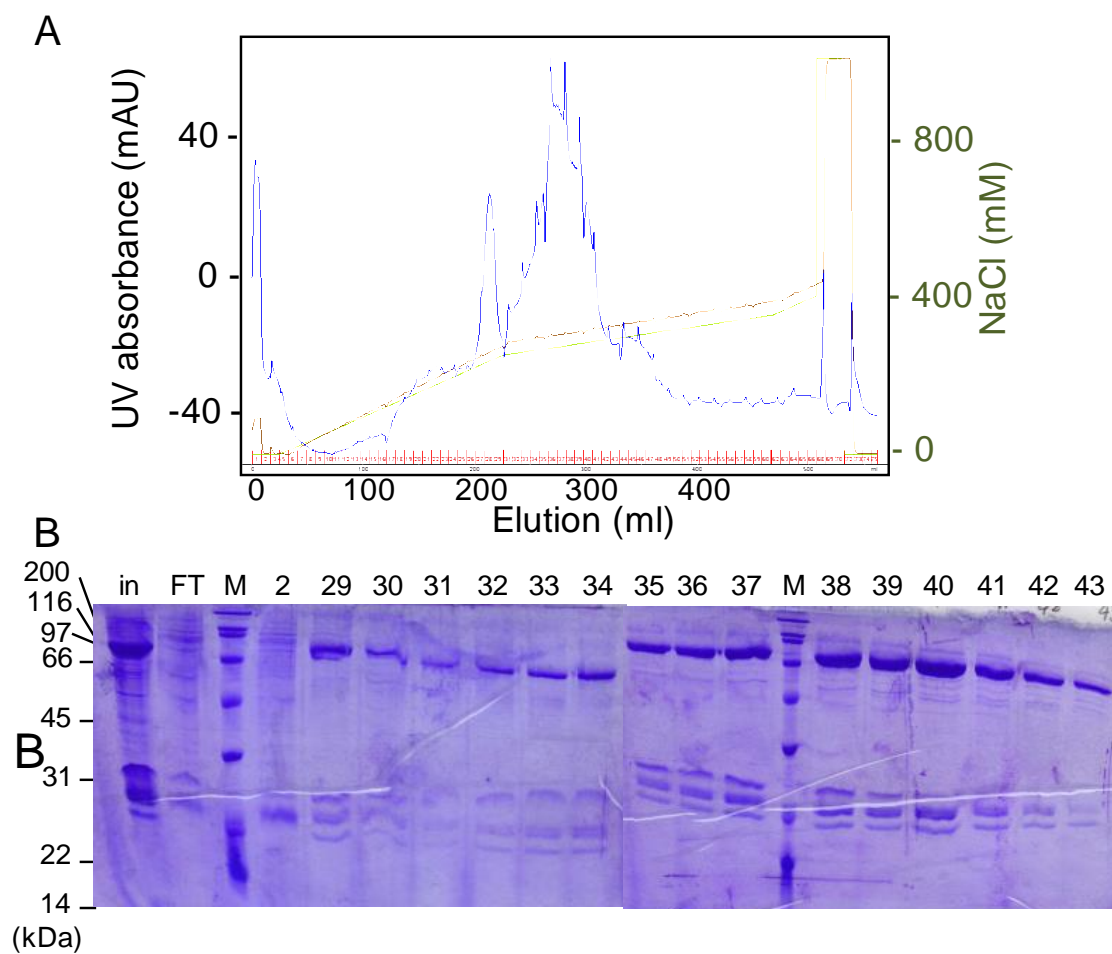


Figure 41. Purification of XC54 after thrombin digestion by SourceS.

A. An exemplary SourceS (8 ml) chromatogram for XC54 (cut). B. The fractions from A examined on SDS-PAGE. #34 – 39 contained UBL, #40 – 43 did not contain UBL.

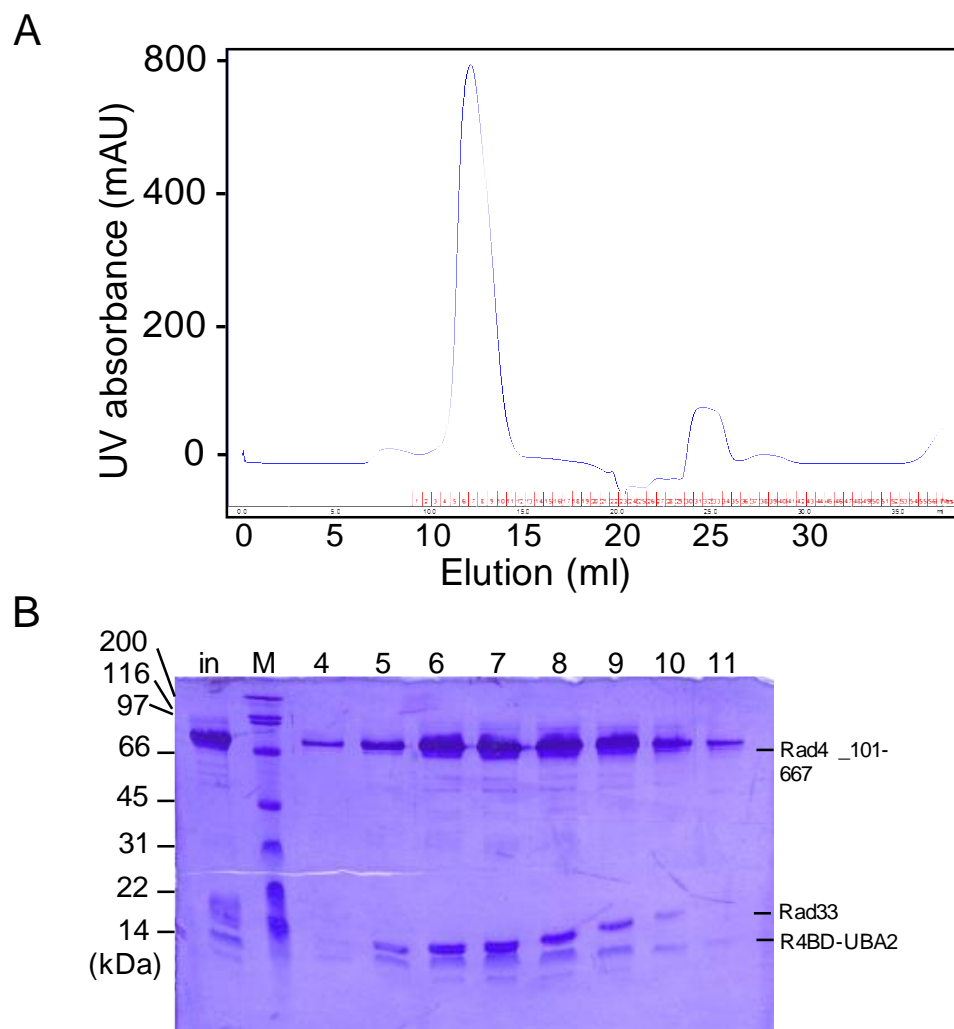


Figure 42. Purification of XC54 without UBL by Superdex200. A. An exemplary Superdex 200 (24 ml) chromatogram for XC54 (cut, without UBL). B. The fractions from A examined on SDS-PAGE. #6 – 9 were collected.

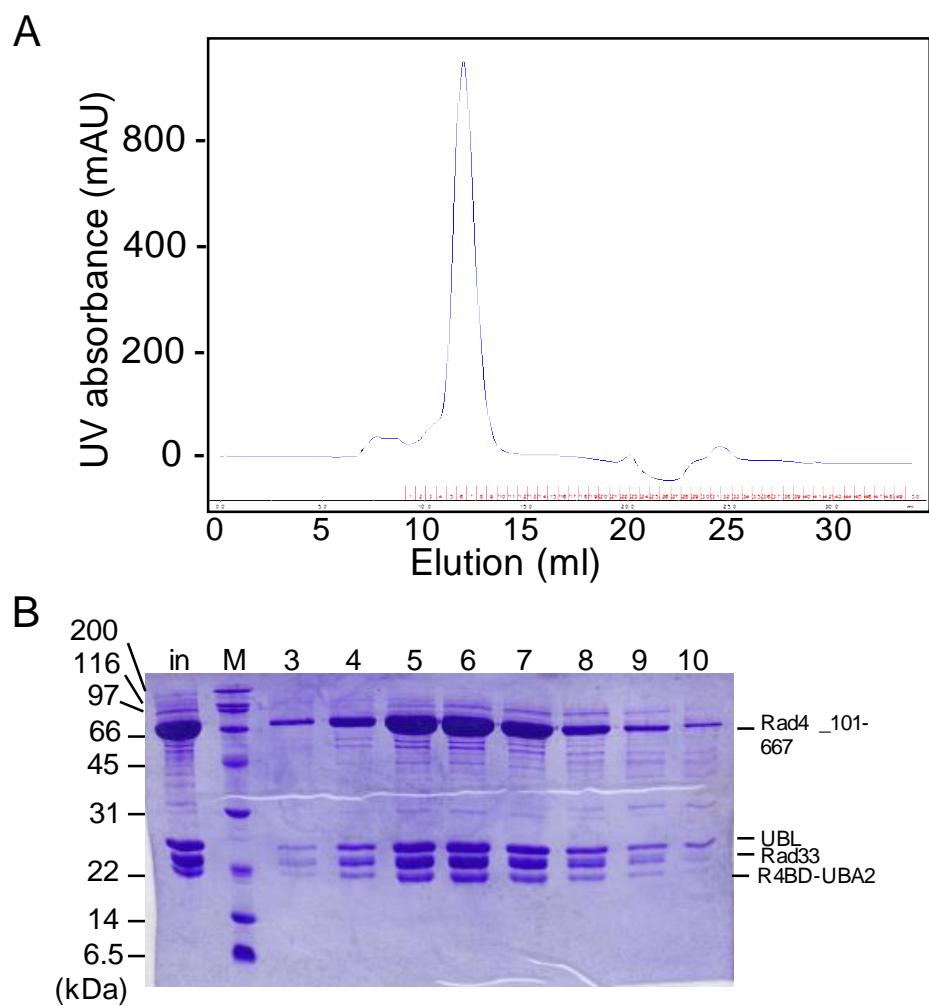


Figure 43. Purification of XC54 with UBL by Superdex200.
 A. An exemplary Superdex 200 (24 ml) chromatogram for XC54 (cut, with UBL). B. The fractions from A examined on SDS-PAGE. #4 – 8 were collected.

C. Preliminary results and Discussion

The expression of Rad33 was first tested in GST-tagged form. GST-Rad33 by itself expressed well, mostly in cytoplasmic fraction. Next, the trimeric complex containing Rad33 needed to be constructed.

The trimmed Rad4₁₀₁₋₆₃₂ used for crystallization and kinetic studies is not able to interact with Rad33, therefore several longer versions of Rad4 were constructed to determine the minimal length for Rad33 binding. The cells were co-infected with virus carrying His-Rad4-Rad23 and virus carrying GST-Rad33. The nuclear content of cells was extracted with NE-PER kit and subjected to GST pull-down. While Rad4₁₀₁₋₆₃₂ did not co-purify with GST-Rad33, Rad4₁₀₁₋₆₆₇ and all other longer versions including full-length Rad4 were co-purified with GST-Rad33 (Figure 44A), consistent with the results reported by den Dulk *et al.* [89]. Therefore Rad4₁₀₁₋₆₆₇ was selected to construct the trimeric complex.

In our current system relying on pFastBacDual, expressing all three proteins in large scale would require co-infecting large amount of cells with two different viruses, and different virus titer might result in imbalanced expression of the proteins. To eliminate this complication, we introduced MultiBac expression system [91]. Briefly, Rad33 was cloned into donor plasmid pUCDM, which carries *loxP* site and R6K γ conditional origin that only allows replication in bacteria strains expressing *pir* gene; while Rad4 and Rad23 were cloned into acceptor plasmid pFL, which contains *loxP* site and standard ColE1 replicon. Then the two plasmids were fused by Cre

recombinase recognizing *loxP* sites, producing a single plasmid carrying one copy of each of the three genes. (Figure 44B) These strategies consequently ensure the three proteins are expressed in same ratio.

Large scale of expression and purification of the trimeric Rad4-Rad23-Rad33 complex was carried out and crystal screens were set up. The complex not digested with thrombin (shown in Figure 38 to Figure 40) did not produce crystals. Alternatively, the complex was digested with thrombin after SourceQ purification, cutting Rad23 into R4BD-UBA2 and UBL domains. The complex with UBL bound and without UBL were separated on SourceS (Figure 41), and pooled separately. The cut complex without UBL domain did not yield crystals. But the digested protein complex with UBL domain (Figure 42C, D) formed crystals in 50 mM BTP pH 6.8 or MES (2-(N-Morpholino) ethanesulfonic acid) pH 6.0, 100 mM NaCl, 8 – 16% PEG 4,000, and with additives such as 40 – 80 mM MgCl₂, or 20 – 80 mM BaCl₂, or 10 – 40 mM trimethylamine HCl (Figure 43E). Then the mismatched DNA bound complex was prepared by dialysis. The DNA bound digested complex with UBL formed crystals in conditions such as 50 mM BTP pH 6.3 – 7.3 or HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 6.5 – 7.5, 100 mM NaCl, 3% – 6% PEG 3,350/PEG 4,000/PEG 6,000/PEG 8,000 (Figure 43E). Some crystals were harvested and tested, but none producing diffraction. Additional trimming and optimizing of Rad33 construct might be needed to improve the crystals.

GST-Cdc31 when expressed by itself was soluble and mostly in cytoplasmic fraction (Figure 43F lane 1, white arrow), similar to Rad33. But when Rad4, Rad23 and

Cdc31 were expressed together, all three proteins became insoluble and not able to be purified (Figure 43F lane 7, white arrows).

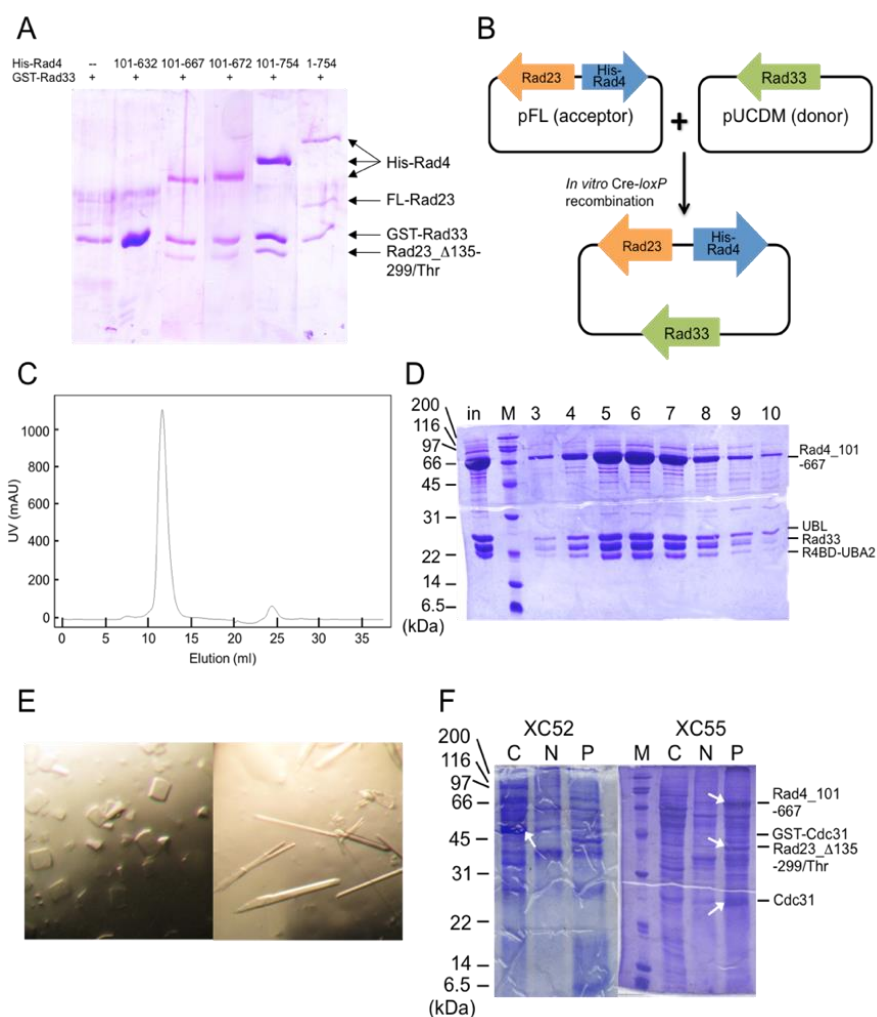


Figure 44. Constructing Rad4 trimeric complex and crystallization.

A. Hi5 cells were co-infected with GST-Rad33 and various versions of His-Rad4-Rad23. GST pull-down was carried out with the nuclear extract and separated on 15% SDS-PAGE, to determine which Rad4 was able to be co-purified with GST-Rad33. Rad4_101-667 was the minimal length to retain interaction with Rad33. B. A diagram showing the strategy of constructing 3-gene plasmid. C. A Superdex 200 chromatogram showing a single peak of purified trimeric complex expressed from 3-gene plasmid, eluting at 11.7 ml. D. The fractions from C were separated on 15% SDS-PAGE. in: input, M: molecular weight. Lane 4 – 8 were collected for crystal growing. E. (left) Crystal of the Rad4-Rad23-Rad33-damaged DNA complex, grown in HEPES pH 7.5, 100 mM NaCl, 5% PEG 4000. (right) Crystal of the Rad4-Rad23-Rad33 complex, grown in BTP pH 6.8, 100 mM NaCl, 16% PEG 4000. F. Expression test result for XC52 and XC55 using NE-PER kit. C: cytoplasmic fraction. N: nuclear fraction. P: insoluble fraction.

V. SITE-SPECIFIC FLUORESCENCE LABELING OF RAD4-RAD23

A. Introduction

Single molecule techniques have emerged as powerful tools to study biological problems. Fluorescence is a particularly popular detection method, aided by the development of fluorescence microscopes. One widely used technique is Förster resonance energy transfer (FRET), which requires a donor and an acceptor fluorophore attached to the molecules of interest. The resonance energy transfer from donor to acceptor, which is very sensitive to the distance between them, provides information on the interaction between molecules.

In order to study the movement of Rad4 on DNA by FRET, they must be first labeled with fluorophores. While labeling synthesized DNA is relative easy, site-specifically labeling proteins has been quite challenging. A fluorescent protein may be fused to the target protein, but it's bulky and might disturb the molecular event of interest. The current common labeling methods usually involve small organic dyes reacting with thiol group of cysteine or amine group of lysine in purified protein, and therefore can conjugate to multiple sites [92]. An easy, efficient and site-specific labeling method for these dyes is yet to be found.

It was discovered that human sulfatase bears a conserved motif. The cysteine in the motif is co- or post-translational converted to formylglycine [93]. Later the enzyme that recognizes the motif and modifies the cysteine was identified and named

formylglycine-generating enzyme (FGE) [94]. Taking advantage of the rare natural occurrence of aldehyde group in protein and its reactivity to common fluorophores such as hydrazides and alkoxyamines, the Bertozzi group developed a labeling strategy [95] where the conserved motif, called “aldehyde tag” (Ald tag), was fused to the target protein to achieve site specificity. When the tagged protein was co-expressed with prokaryotic FGE, the aldehyde conversion efficiency reached ~90%, and the protein was robustly labeled with commercially available fluorophores.

Recently the Ha group proposed a milder, physiological relevant labeling condition [96], because the harsh condition that Bertozzi group reported denatured the polymerase they tested. Changing the condition significantly decreased the labeling efficiency. But by increasing the concentration of fluorescent dye, they were able to achieve ~100% labeling for a prokaryotic polymerase.

Using the procedure reported by the Ha group, we label Rad4 with Cy3 hydrazide dye (FRET donor), which will allow us to monitor the interaction between Rad4 and Cy5 (FRET acceptor) -labeled DNA (Figure 45), and provide us the basis for studying Rad4 with single molecule fluorescence techniques. This is also the first time the aldehyde tag strategy being applied in insect cell expression system.

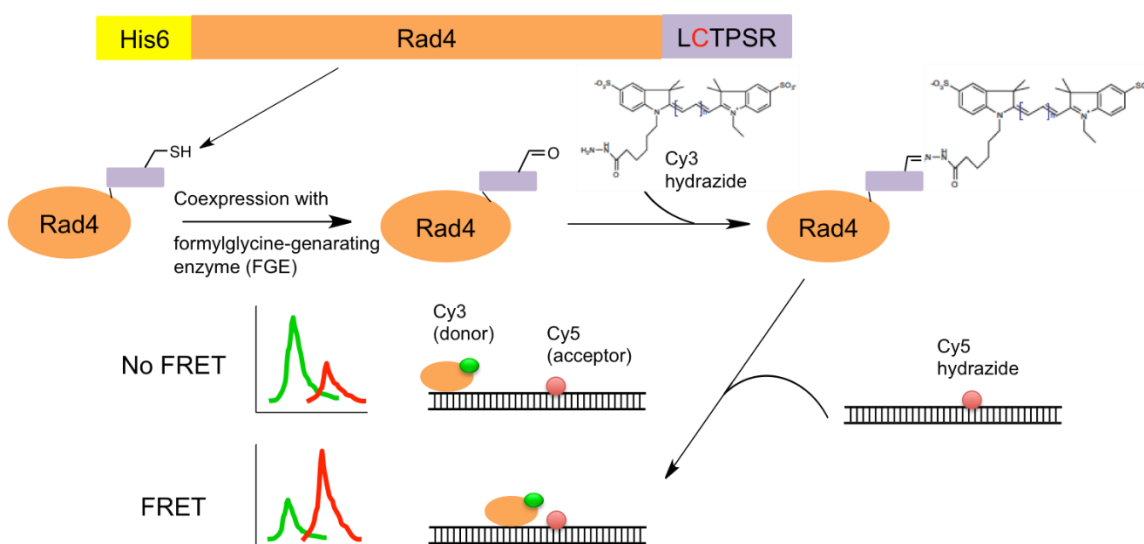


Figure 45. Site-specific fluorescence labeling of Rad4-Rad23 for single-molecule FRET study.

B. Materials and Methods

1. Cloning

The full-length *m. tuberculosis* FGE (formylglycine-generating enzyme) was cloned from the pBAD-His-Myc-FGE plasmid provided by Dr. Xinghua Shi (Dr. Taekjip Ha's group, University of Illinois at Urbana-Champaign) by PCR, and inserted between SacI and XbaI in pFBD vector with a His tag engineered between BamHI and BssHIII: construct XC56. And later inserted between same enzyme sites in pUCDM vector: construct XC57.

The stop codon of Rad4₁₀₁₋₆₃₂ was mutated to G, and the thrombin recognizing sequence LVPRGS, followed by the Ald tag LCTPSR were attached to the C-terminus of Rad4₁₀₁₋₆₃₂ by PCR, and inserted between BssHIII and NotI in pFL vector with His tag

engineered between BamHI and BssHII. Full-length Rad23, with amino acids 135–299 deleted and replaced by thrombin recognition sequence, was inserted between NheI and SphI in the same vector: construct XC64.

XC57 and XC64 were then fused in Cre-*loxP* recombination reaction to construct 3-gene plasmid (see chapter 5 for method): XC65.

2. Virus production

See protocol d and e.

3. Expression and purification of XC65

See protocol f and g. The complex eluted around 90 – 200 mM NaCl on Source Q. The fraction containing the protein was not digested with thrombin. Then the complex eluted around 160 – 270 mM NaCl on Source S. The fraction containing the protein was concentrated to around 20 mg/ml, and eluted around 12.85 ml on Superdex 200. The sample was finally concentrated to around 35 mg/ml.

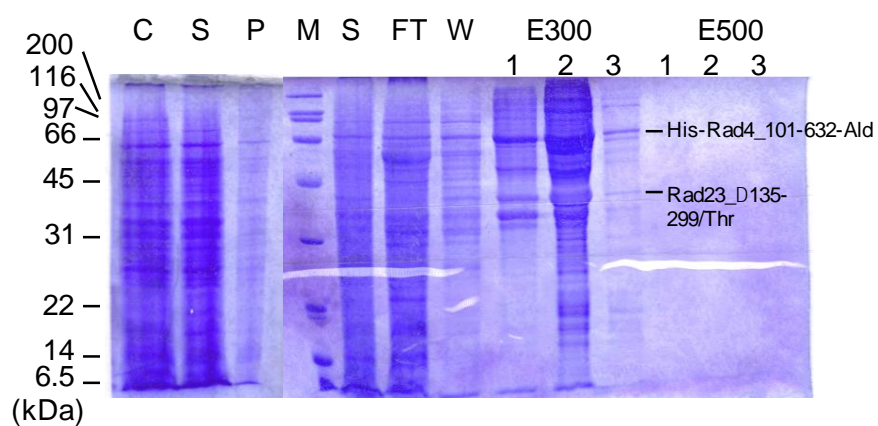


Figure 46. Lysis and Ni pull-down of XC65.
See protocol g for labels.

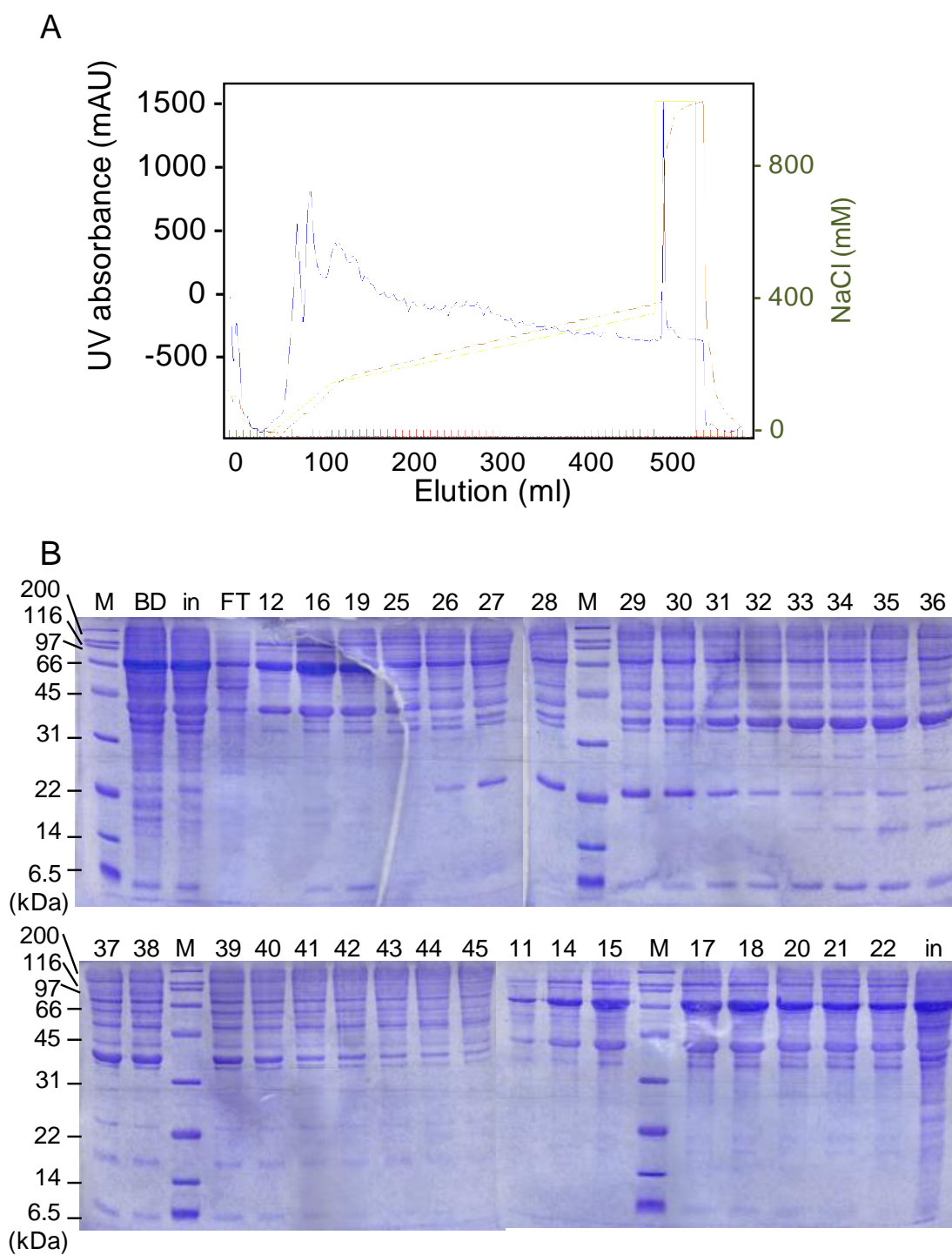


Figure 47. Purification of XC65 by SourceQ.

A. An exemplary SourceQ (24 ml) chromatogram for XC65. B. The fractions from A examined on SDS-PAGE. #12 – 15 were collected.

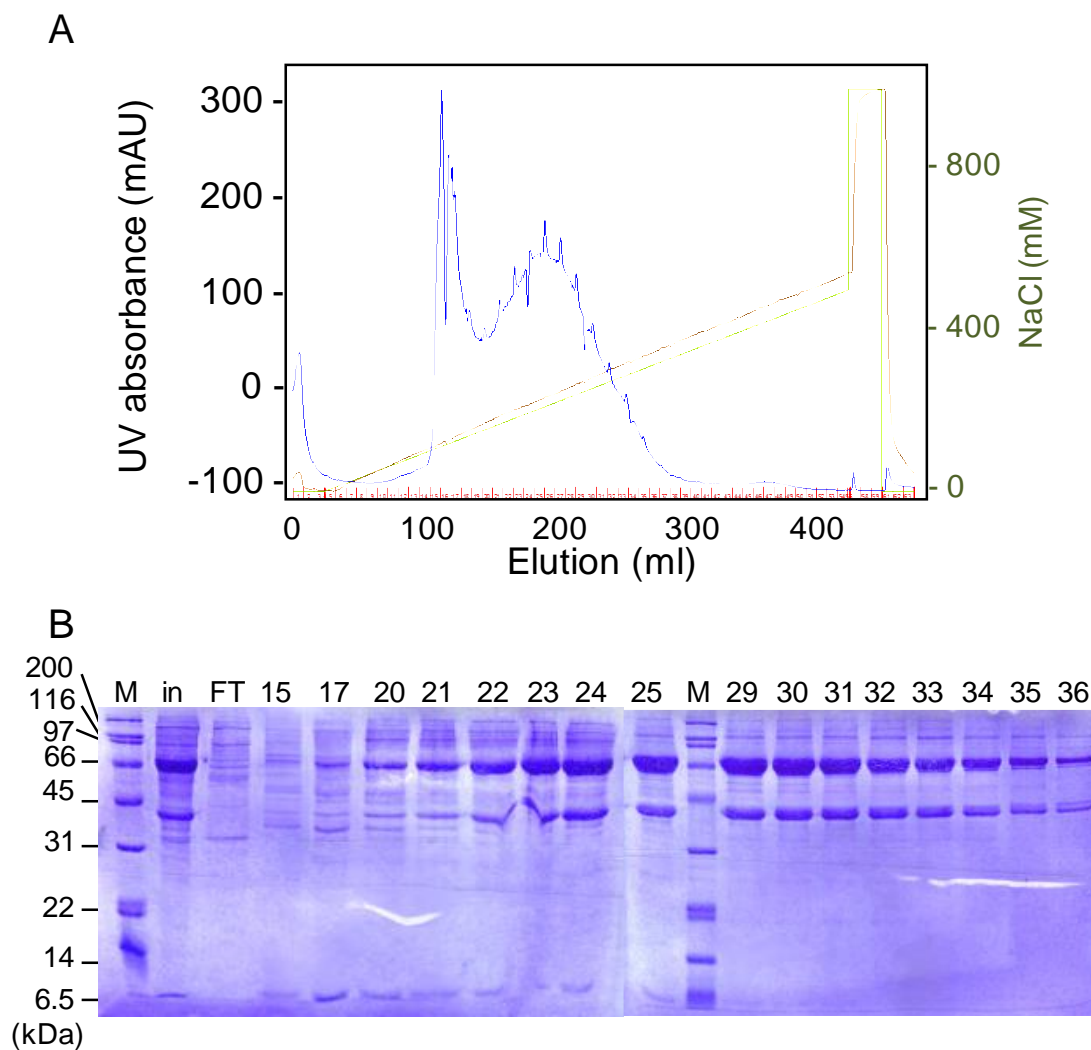
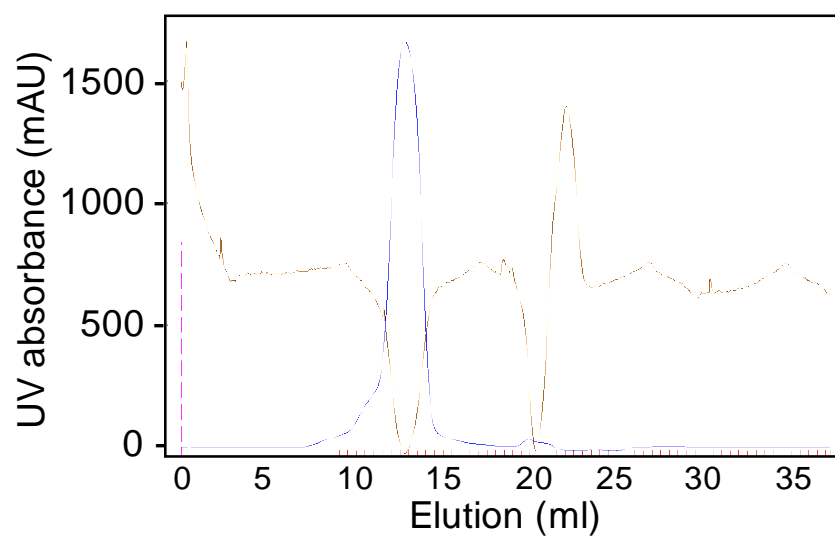


Figure 48. Purification of XC65 by SourceS.

An exemplary SourceS (8 ml) chromatogram for XC65. B. The fractions from A examined on SDS-PAGE. #21 – 31 were collected.

A



B

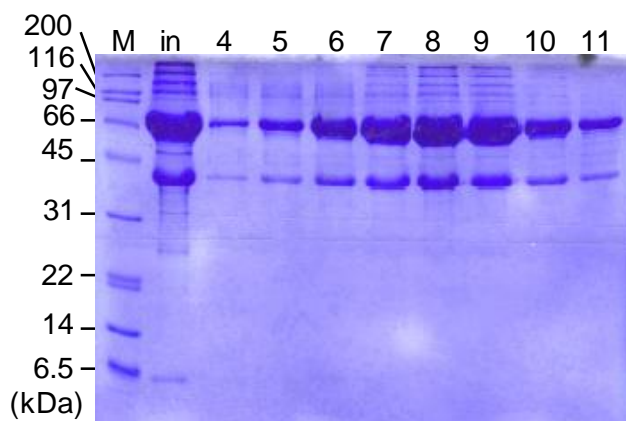


Figure 49. Purification of XC65 by Superdex200. A. An exemplary Superdex200 (24 ml) chromatogram for XC65. B. The fractions from A examined on SDS-PAGE. #5 – 10 were collected.

4. Expression of FGE and co-expression of DinB and FGE

The pET-28a(+) plasmid carrying archaeal polymerase DinB with FGE recognition sequence (Ald-) followed by His tag and thrombin recognition sequence at N terminus, and the pBAD plasmid carrying His-Myc-FGE (both provided by Dr. Xinghua Shi) were first amplified using Miniprep kit.

For co-expression, 1 μ l of each plasmid were added to 100 μ l of BL21(DE3) competent cells, incubated on ice for 30 min. Heat shock was applied by immersing the tube in 42 °C water bath for 1 min. The cells were returned to ice to recover for 2 min. 500 μ l of SOC medium was added to the cells, and the cells were allowed to grow with shaking for 1 h, then plated on LB agar with Ampicillin and Kanamycin.

LB medium containing Ampicillin and Kanamycin was inoculated with colonies from co-transformation plate and grown at 37 °C with shaking. When the OD₆₀₀ of the culture reached about 0.3, the cells were chilled on ice for 5 min, and L-(+)-arabinose was added to a final concentration of 0.2% (w/v) to induce the expression of FGE. 30 min later, the cells were chilled on ice for 5 min again, and IPTG was added to a final concentration of 0.1 mM to induce the expression of DinB. Then the culture was grown at 18 °C for overnight.

To express FGE alone, 1 μ l of pBAD-His-Myc-FGE plasmid miniprep was transformed in BL21(DE3) competent cells as described above, and plated on LB agar with ampicillin. The culture was grown as described above, but only induced with L-(+)-arabinose, or same volume of MQ H₂O as un-induced control.

5. Cy3 labeling

One tube of 1 mg Cy3 hydrazide (PA13121, GE Healthcare) was completely dissolved in 1 ml of methanol, and aliquotted to 100 μ l in 10 tubes (containing 0.1 mg each). The solution was dried completely in SpeedVac to remove methanol, and stored at -20 °C.

To change the protein storage buffer to labeling buffer (250 mM KH₂PO₄ pH 7.0, 500 mM KCl, 5 mM DTT), about 100 μ l of protein was diluted to 500 μ l with labeling buffer, and concentrated back to about 100 μ l in Amicon Ultra-0.5 mL Centrifugal Filter (EMD Millipore). The above step was repeated 3 times to ensure efficient buffer exchange. Then 3 μ l of protein was added to 1 tube of 0.1 mg Cy3 and incubated at 4 °C for 24 h. When the reaction was finished, 87 μ l of labeling buffer was added, and the mixture was passed through 2 Zeba desalting columns to remove the free Cy3 dye. The UV absorbance of the sample was examined, and compared to protein not incubated with the dye. Peak at 550 nm indicates the presence of Cy3 attached to the protein. The labeling efficiency was the ratio between the molar concentration of Cy3 (assuming all free dye was remove, the concentration of Cy3 was taken as the concentration of labeled protein) and the molar concentration of total protein. The concentration of Cy3 was calculated using UV absorbance at 550 nm, and the extinction coefficient $150000 \text{ M}^{-1}\text{cm}^{-1}$. The concentration of protein after labeling was calculated from UV absorbance at 280 nm, by first subtracting 8% of the A₅₅₀ value (the contribution to A₂₈₀ from Cy3) from A₂₈₀, then using the extinction coefficient of the protein.

C. Results

1. Labeling of DinB

First the labeling of DinB was first tested, following the procedure reported by Shi *et al.* [96], as a positive control for Rad4-Rad23 labeling. Ald-His-DinB and His-FGE were co-expressed in 6 L of BL21(DE3) cells. 500 ml of culture was lysed and subjected to Ni pull-down. DinB was seen expressed and purified by Ni resin, but FGE was not seen. (Figure 50A) To examine the expression of FGE, 250 ml of culture was prepared, either induced with L-(+)-arabinose or not induced. Both were subjected to Ni pull-down. By comparing the imidazole elution from the two samples, FGE (MW = 33kDa) was identified at the expected position on SDS-PAGE, although the expression amount was much less than DinB (Figure 50B, C). This confirmed that FGE was expressed, and the FGE recognition sequence of DinB should be modified.

Next, 3 μ l of DinB co-expressed with FGE was labeled, at the same time another 3 μ l was incubated at same condition but without Cy3 as control. Both input protein concentration was 0.401 mM (= 17.4 mg/ml). After incubation and passing through desalting columns, the UV absorbance spectrum was taken (Figure 51). A peak was seen in the sample with Cy3, but not in control sample, indicating a portion of the protein successfully incorporated the dye. The total protein concentration in control after incubation was 0.411 mM (= 17.9 mg/ml). The total protein of labeled sample was 0.184 mM (= 8.00 mg/ml), with ~ 50% protein lost. The concentration of Cy3 in labeled sample was 0.0247 mM, equivalent to 13.4% labeling efficiency.

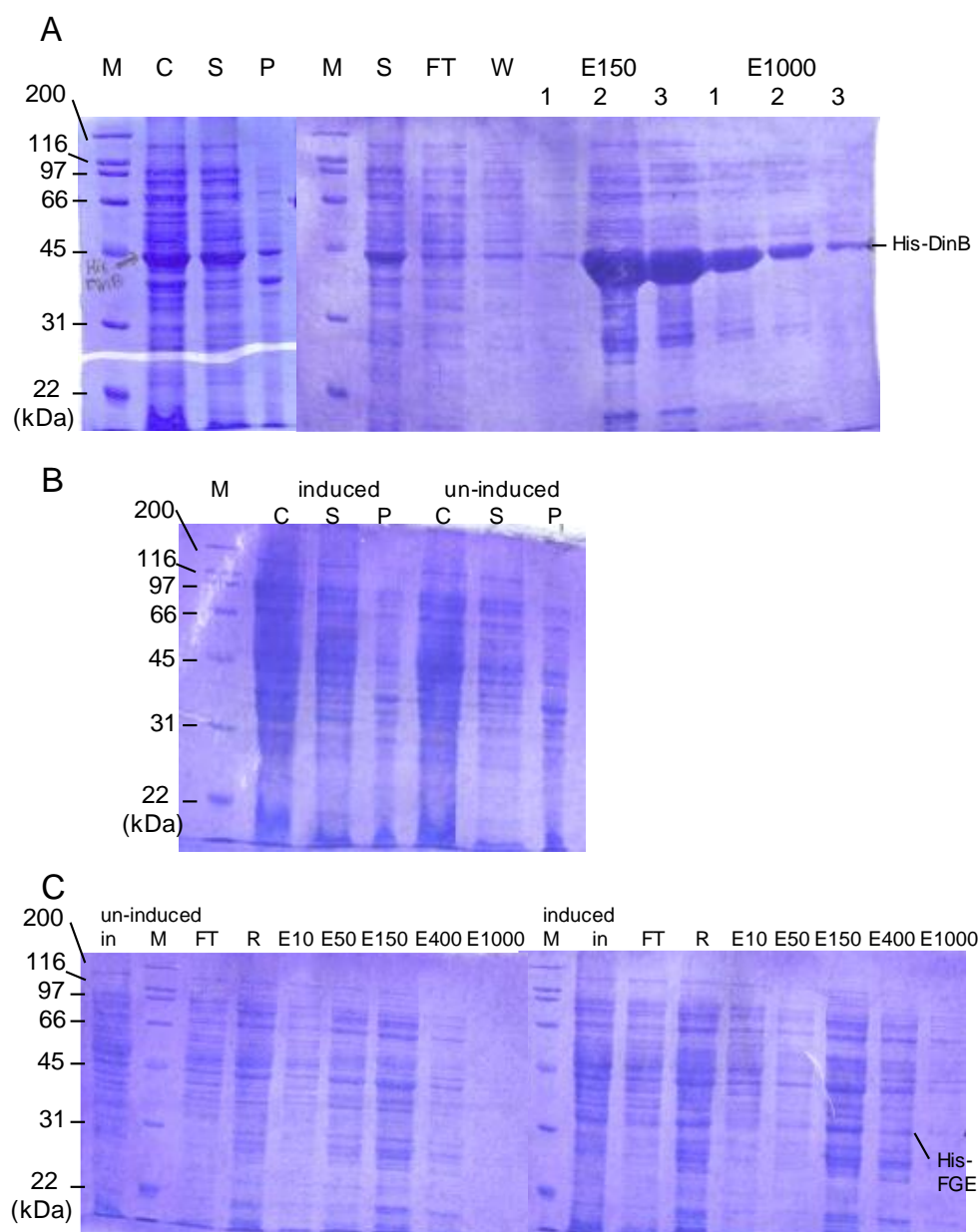


Figure 50. Co-expression of DinB and FGE, and test of FGE expression. A. Ni pull-down of co-expressed Ald-His-DinB and His-FGE. M: molecular weight marker, C: total cell, S: soluble, P: insoluble, FT: flow-through, W: wash, E150 and E1000: elution with 150 mM and 1000 mM imidazole, respectively, each for 3 column volumes. B. Lysis of 250 ml His-FGE culture, induced or un-induced. C. Ni pull-down of soluble fractions from B. in: input, R: denatured resin after sample binding, E10 – E1000: elution with 10, 50, 150, 400, 1000 mM imidazole, 1 column volume each.

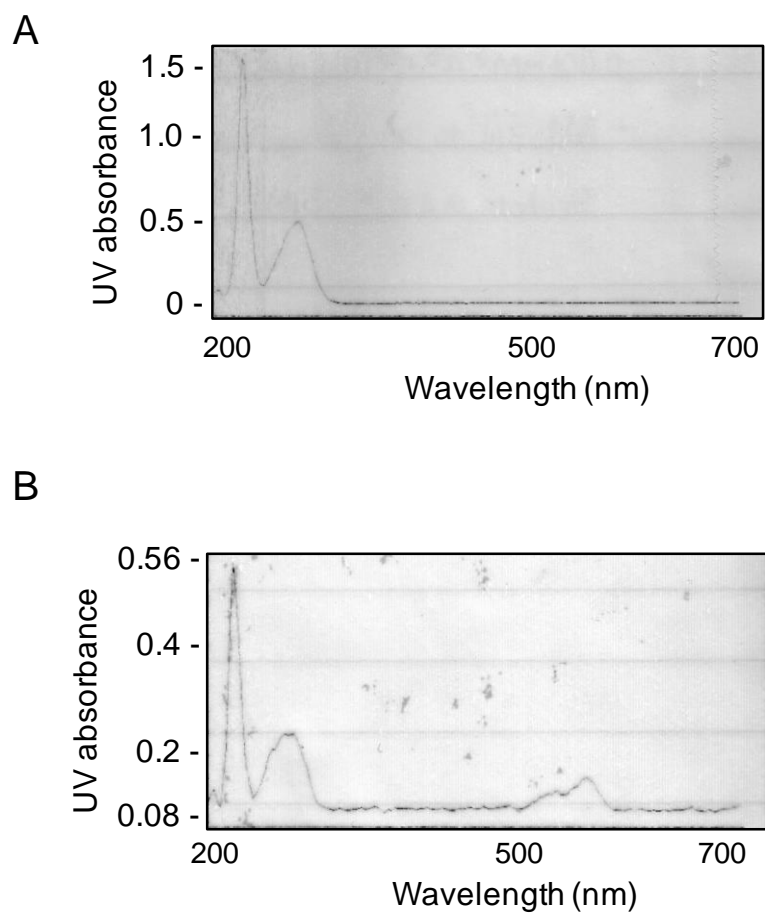


Figure 51. Cy3 labeling of DinB.

A. The UV spectrum of DinB before labeling, diluted by 30 fold in labeling buffer. B. The UV spectrum of DinB after labeling, diluted by 30 fold and passed through desalting columns.

2. Labeling of Rad4-Rad23

A plasmid bearing His-Rad4, Rad23 and FGE was made by recombination to ensure expression of all three proteins. The expression of Rad4-Rad23 was not affected, as shown in Ni pull-down (Figure 52A). The labeling of Rad4-Rad23 was then tested at a small scale. A UV absorbance spectrum was taken for the protein, and then the reaction was set up as described in previous section. The concentration of input protein was 0.278 mM (= 27.3 mg/ml). After 24 h of incubation, both samples were pass through two consecutive desalting columns steps. The concentration of protein only sample was 0.214 mM (= 21.0 mg/ml), which was 77% of input, with no peak seen at 550 nm. The total protein recovered in the labeling reaction was 0.132 mM (=13.0 mg/ml), which was a lower yield, 45.6%. The concentration of Cy3 was 0.0914 mM, giving a labeling efficiency of 69.2%. (Figure 52B, C)

To examine the specificity of the labeling, that is, whether the dye only attached to the Ald tag instead of non-specific residues, a labeled sample was thawed and the Ald tag was cleaved with thrombin to remove the conjugated dye. The previously labeled and frozen protein was thawed (0.334 mg/ml). First the stability of labeled protein after freezing and thawing was examined. Based on A280, the recovery of total protein was 58.7% after freezing and thawing, and the recovery of labeled protein (based on A550) was 67.7%. (Figure 53A)

The labeled sample, along with non-labeled protein as control (0.691 mg/ml), were both incubated with thrombin at 5:1 mass ratio at 4 °C for overnight. The spectrum

of the labeled protein still showed a relative tall peak at 550 nm, with Cy3 concentration 3.91 μM , and total protein concentration 6.17 μM (Figure 53B). Then both samples passed through desalting columns to remove cleaved dye, and spectra were taken (Figure 53C, D). The 550 nm peak in labeled sample significantly diminished after purified through 1 column, with only 30% of Cy3 molecules left compared to before purification. The second column was only able to lower the percentage to 22%, indicating 1 column was able to remove the majority of the free dye. Assuming all cleaved dye was removed, there was only 10% of protein remained labeled after digestion, possibly from incomplete thrombin digestion, or incomplete removal of free dye, or small amount of conjugation to non-specific sites. This suggested that most (~90%) of the dye was specifically attached to Ald tag.

To prepare labeled Rad4 complex for single molecule fluorescence assays, a large amount of sample was labeled. 6 reactions were set up at the same time, each as described in previous section. These reactions were then combined and purified. The labeling efficiency was calculated to be 40.1%.

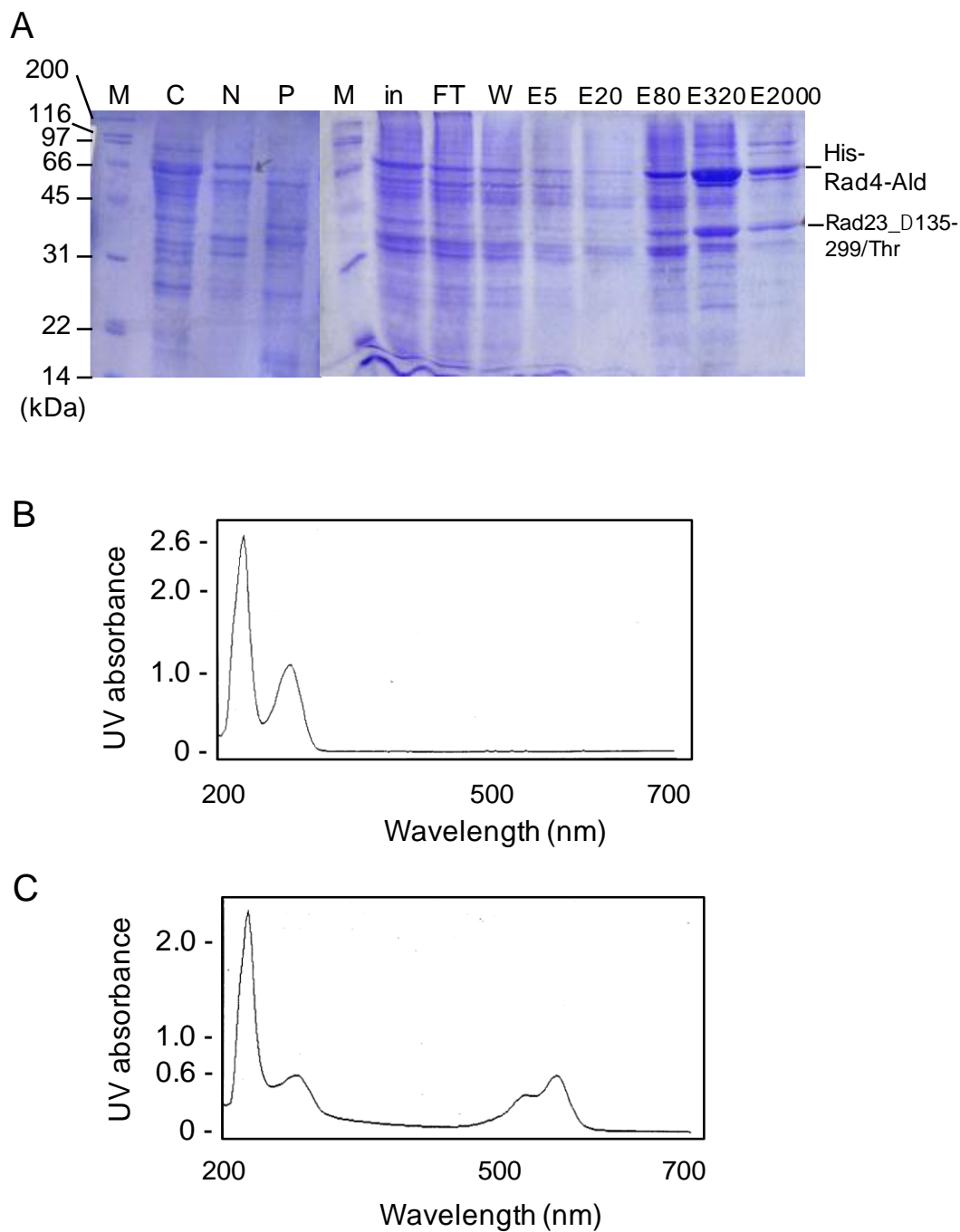


Figure 52. Expression of XC65 and Cy3 labeling.

A. Expression test of XC65 using NE-PER kit, and Ni pull-down. B. The UV spectrum of XC65 before labeling, diluted by 30 fold. C. The UV spectrum of XC65 after labeling, diluted by 30 fold and passed through desalting columns.

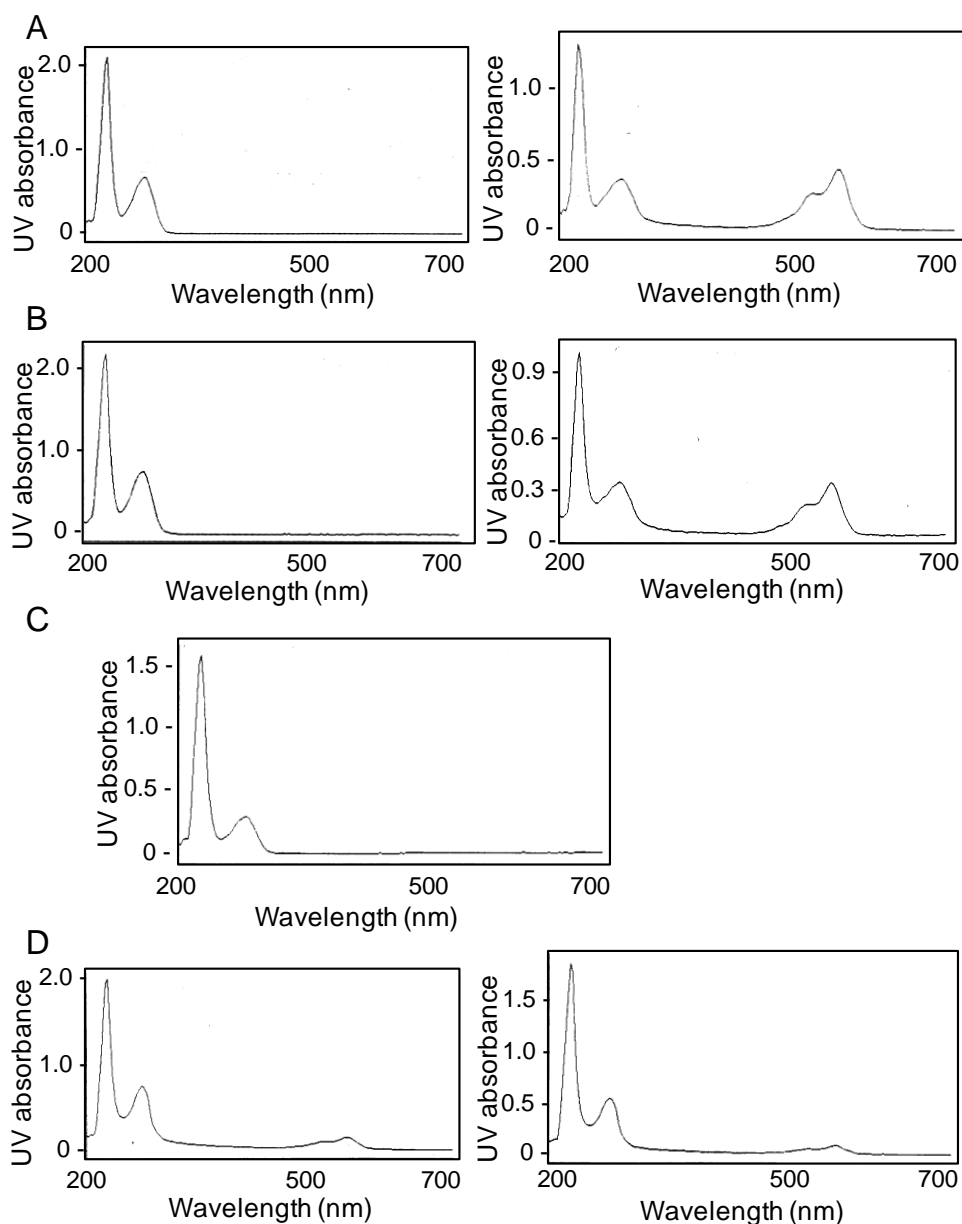


Figure 53. Validation of Cy3 labeling XC65.

A. The UV spectrum of (left) non-labeled XC65 and (right) previously labeled then frozen and thawed XC65. B. The UV spectrum of (left) non-labeled and (right) previously labeled XC65 after incubating with thrombin. C. The UV spectrum of non-labeled XC65 after thrombin digestion and desalting column purification. D. The UV spectrum of previously labeled XC65 after thrombin digestion and (left) after first desalting column purification, and (right) after second desalting column purification.

D. Discussion

In this chapter a novel method of fluorophore conjugation of protein is tested in insect cell expression system. A short peptide, the Ald tag, was fused to the protein at the desired position. Then the protein was co-expressed with FGE, which converts a cysteine in the tag to formylglycine, a residue that will conjugate with fluorophore.

This labeling method provides several advantages comparing to conventional methods: The insertion of Ald tag can be achieved by simple cloning, and the sequence can be incorporated into the cloning primers, thus does not require additional time or procedure. The use of Ald tag ensures that the fluorophore will attach to a specific position, and to only that position, which the researcher has the freedom to select. The aldehyde group reacts readily with commercially available fluorophores, under mild condition that doesn't compromise the integrity of the proteins.

Yeast does not express endogenous FGE, but by expressing bacterial FGE, yeast Rad4 with Ald tag was successfully modified. Despite being a challenging protein, the labeling of Rad4 was proved to be site-specific, with decent efficiency, and stable after freezing and thawing. This also shows that site-specific labeling of Ald tagged protein can be carried out in insect cell expression system, besides bacterial and mammalian cell [97], further expanding the field of application of this method.

However, Shi *et al.* reported that the labeling efficiency of DinB was only 60%, due to incomplete Cys to formylglycine conversion. In our lab this DinB-labeling efficiency was much lower (~13%), even though following same experimental procedure.

For Rad4 labeling, using high concentration of protein was helpful (reaching ~70% labeling efficiency), but still the dye is always in great excess (75.6 mM) and needs to be removed after the reaction. The desalting columns used in this step usually cause some loss of protein. And even with same batch of protein, the reaction set up on different days showed dramatically different efficiency. To improve the labeling efficiency, the aldehyde biosynthesis and the conjugation reaction require further investigation and validation. Quantitative labeling may not be easy to achieve for researchers newly adopting this technique.

Labeling Rad4 with Cy3 opens up the possibility to study Rad4 with single molecule techniques, such as FRET, which provide unique molecular information. FRET experiment has been proposed in collaboration with Ha group, to monitor the movement of Rad4 on Cy5-labeled DNA synthesized in situ by PolB.

VI. CONCLUSION

This dissertation is focused on the structure and mechanism of Rad4 complex, the initiator of NER.

First, the crystal structure of Rad4 complex tethered to undamaged DNA using disulfide crosslinking was solved. The structure of the tethered complex revealed to have the same open conformation as damaged DNA bound structure, indicating that Rad4 is able to open undamaged DNA, and that lesion discrimination is not based on structural difference between damaged versus undamaged DNA.

We then went on to perform the T-jump kinetics measurement using 2AP as probe revealed a ~ 7 ms relaxation rate of damaged DNA induced by Rad4, representing the nucleotide flipping ('opening') event when protein-DNA forms a stable open conformation. The opening also required β -hairpin3, consistent with its role in inserting into DNA duplex and flipping out the bases as seen in the structure. Next, using the probes tC^O/tC_{nitro} pair, Rad4 induced nonspecific kinetics was observed. A slow phase of twisting occurred on the same time scale as nucleotide flipping, but did not require β -hairpin3, indicating different motions were detected in spite of the common rate-limiting step of recognition. A fast phase was also observed, which did not require β -hairpin3 or the presence of a lesion, possibly representing a transient conformation used to interrogate weakened site in DNA duplex.

In chapter IV, the trimeric complex of Rad4-Rad23-Rad33 was purified, and crystals of the complex were obtained. Further optimization of the protein constructs is needed to solve this structure.

In chapter V, Rad4 was labeled with fluorophore using a novel labeling scheme: a short peptide was fused to Rad4, then the FGE protein co-expressed with Rad4 recognized and modified the peptide with aldehyde group, which could readily react with fluorophores, achieving site-specific labeling. These constructs will be used for further fluorescence dynamics studies in the future.

APPENDIX

A. Protocols

1. PCR

- Mix the PCR reaction as below:

	Item (stock concentration)	Volume (μl)
1	Autoclaved MQ H ₂ O	26.0
2	5X Phusion HF Buffer (New England Biolabs)	10.0
3	dNTP mix (2 mM, final 0.2 mM) (EMD Chemicals)	5.0
4	DMSO	2.5
5	Template (~10 ng/ml. Use 10 – 50 ng)	1.0
6	Primer1 (10 μM, final 0.5 μM)	2.5
7	Primer2 (10 μM, final 0.5 μM)	2.5
8	Phusion HF Polymerase (2 U/μl) (New England Biolabs)	0.5
	Total	50.0

- Perform PCR in MJ Mini thermal cycler (Bio-rad) as below:

	Temperature (°C)	Time (s)	
Initial denaturation	98°C	30	
Cycles	98°C	5 – 10	
	55°C*	10 – 30	
	72°C	(15 – 30)/kb	25 – 35 cycles
Final extension	72°C	5 – 10 min	
Storage	4°C	infinite	

$$\begin{aligned}
 \text{*Annealing temperature} &= T_m + 3\text{ }^{\circ}\text{C} && \text{for primers} < 20\text{ nt} \\
 &= T_m && \text{for primers} > 20\text{ nt}
 \end{aligned}$$

- Examine PCR product on 1% agarose gel.
- Load the successful reaction mix on agarose gel. Run for 10 – 15 min.
- Cut the target DNA band and purify using QIAquick Gel Extraction Kit (Qiagen).
Elute the DNA in 45 μL of autoclaved MQ H_2O .

2. Restriction enzyme digestion and ligation

- For insert digestion, add 5 μL of 10X buffer and 10 U of each enzyme to the gel extracted insert DNA. Incubate at the optimal temperature for the enzyme for 1 h. Purify the insert with QIAquick PCR Purification Kit (Qiagen).
- For vector digestion, add 10 U of one of the enzyme and 5 μL of 10X buffer optimal for the enzyme to $\sim 5\text{ }\mu\text{g}$ vector. Incubate at the optimal temperature for the enzyme for 1 h. Run the mixture as well as uncut vector on 1% agarose gel, cut the band containing digested vector (runs higher than uncut vector). Purify with QIAquick Gel Extraction Kit and elute with 45 μL of autoclaved MQ H_2O . Then add 10 U of the other enzyme and 5 μL of 10X buffer optimal for the enzyme. Incubate at the optimal temperature for the enzyme for 1 h. Add 1 μL of Alkaline Phosphatase Calf Intestinal (CIP) (New England Biolabs) and incubate at 37 $^{\circ}\text{C}$ for 1 h. Run the mixture on 1% agarose gel, cut out the band and purify with QIAquick Gel Extraction Kit.

- Estimate the concentration of digested vector and insert on 1% agarose gel, and set up the ligation reaction as below:

	Amount
Vector	30 – 50 ng
Insert	3 – 9 fold more than vector in moles
10x T4 ligase buffer	2 μ L
Autoclaved MQ H ₂ O	Calculated amount to make total 20 μ L
T4 DNA ligase	1 μ L
Total	20 μ L

- Incubate at room temperature for 2 h or 16 °C for overnight.

3. Transformation and colony identification

- Thaw 1 tube of 100 μ L DH5 α competent cells on ice. Add ligation reaction mixture. Tap the side of the tube to mix.
- Incubate on ice for 15 min.
- Apply heat shock by immersing the tube in 42 °C water bath for 60 s.
- Return to ice for 2 min.
- Add 0.5 mL of LB medium.
- Incubate at 37 °C with vigorous shaking for 60 min.
- Spin down briefly for 60 s.
- Discard 500 μ L of supernatant and re-suspend the cells in the remaining 100 μ L.
- Plate the cells onto LB agar with appropriate antibiotic and spread the cells using glass beads. Incubate at 37 °C for overnight.

- On the next day, inoculate 3 mL LB medium containing appropriate antibiotic with a single colony and grow at 37 °C for overnight with shaking.
- Purify plasmid from 2 mL of overnight culture using Qiagen Miniprep kit (Qiagen). Elute DNA with 50 µL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).
- Set up the following double digestion reaction:

	Volume
Miniprep DNA	2 µL
10X buffer compatible with both enzymes	1 µL
Autoclave MQ H ₂ O	7 µL
Restriction enzymes used for vector and insert	0.1 µL each
Total	10 µL

Incubate at 37 °C for 1 h.

- Examine the mixture on 1% agarose gel. Positive colonies should show a band consistent with the insert size.

4. Transposition and bacmid extraction

- Thaw the DH10Bac competent cells on ice.
- Dispense 50 µL of the cells into 15-ml round-bottom polypropylene tubes.
- Add 1 µL (~ 100 ng) of Miniprep plasmid into the cells. Mix well by tapping the side of the tube. Incubate on ice for 30 min.
- Apply heat shock by partially immersing the tube in a 42 °C water bath for 45 s.
- Chill on ice for 2 min.
- Add SOC media (MP Biomedicals) to 1 mL (i.e., 950 µL).

- Shake at 37°C with medium agitation (225 rpm) for 4 h.
- Plate 1/100-fold or 100 µL of a 1/10-fold diluted culture onto Bac-to-Bac plates (LB agar containing 50 µg/mL Kanamycin, 7 µg/mL Gentamicin, 10 µg/mL Tetracyclin, 100 µg/mL Bluo-gal and 40 µg/mL IPTG).
- Incubate the foil-covered plates for 36 to 48 h in dark. White colony indicates successful transposition.
- Inoculate a single, isolated **white** colony into 2 ml LB media containing 50 µg/mL Kanamycin, 7 µg/mL Gentamicin and 10 µg/mL Tetracycline. Grow at 37°C for 24 h (to a stationary phase) shaking at 250 to 300 rpm.
- Transfer 1.5 mL of the culture into a microcentrifuge tube. Spin down and remove the supernatant.
- Re-suspend the cells in 0.3 mL of cold P1 buffer from Plasmid Maxi Kit (Qiagen).
- Lyse the cells by adding 0.3 mL of room temperature P2 buffer from Plasmid Maxi Kit. Incubate at room temperature for 5 min.
- Neutralize by adding 0.3 mL of P3 buffer from Plasmid Maxi Kit. Incubate on ice for 5-10 min.
- Centrifuge at 14,000 rpm for 10 min.
- Prepare clean microcentrifuge tubes with 0.8 mL isopropanol.
- Transfer the supernatant to 0.8 mL isopropanol, mix.
- Centrifuge at 14,000 rpm for 15 min.
- Add 0.5 mL of 70% cold ethanol to the pellet. Invert the tube several times.

- Centrifuge at 14,000 rpm for 5 min.
- Remove as much of the supernatant as possible. Air dry the DNA pellet at room temperature for 5 – 10 min until all liquid has evaporated.
- Re-dissolve the bacmid DNA in 40 μ L TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

5. Transfection and Virus production

- Mix 12 μ L of each bacmid with 100 μ L of Grace's Insect Medium unsupplemented (Invitrogen).
- In another tube, for every Bacmid prepared, mix 6 μ L of Cellfectin II reagent (Invitrogen) with 100 μ L of Grace's Insect Medium unsupplemented.
- Combine the mixture from the above two steps, incubate for 30 min at room temperature.
- Add 800 μ L Grace's medium unsupplemented to the mixture.
- Seed the cells by splitting one confluent Sf9 plate (d =150 mm) into 1:12 into 6-well tissue culture plates (d = 35 mm). Each well should show about 50% confluency.
- Let it sit for at least one hour until all the cells have attached to the well.
- Wash cells with 2 mL Grace's medium unsupplemented. Remove the medium.
- Add the mixture and incubate at 27 °C for 5 h.
- Remove the transfection mixtures and add 2 ml of complete Grace's insect medium (Grace's insect medium supplemented (Invitrogen) containing 10% FBS (Gemini

Bio-Products), 250-fold diluted Pen/Strep/L-Glutamine (Fisher Scientific Cat. No. 17-718R), 10 µg/ml Gentamicin (Invitrogen) and 1000-fold diluted Fungizone Antimycotic (Invitrogen Cat. No. 15290-018).

- Incubate cells in a 27 °C incubator for 72 h.
- Harvest P0 virus at 72 h post-transfection by transferring the supernatant into a sterile 2 ml cryo tube. Store at -20 °C.
- To make P1 virus, aspirate the medium from a confluent Sf9 plate, and add 1 ml P0 virus. Incubate at room temperature for 1 h, swirl the plate every 20 min to prevent cells drying. After 1 h, add 20 ml of complete Grace's medium. Incubate at 27 °C.
- 72 h after infection, harvest P1 virus into 50 ml centrifuge tube and spin down at 2,000 rpm for 5 min to remove any cells. Filter the supernatant with 0.22 µm syringe filter into a sterile tube.
- P2 can be made from P1, and P3 from P2 in the same manner.

6. Protein expression

- For expression test, 1 plate (150 mm) of Hi5 cell is used.
- Aspirate the medium from a confluent Hi5 plate, and add 1 ml P1 virus. Incubate at room temperature for 1 h, swirl the plate every 20 min to prevent cells from drying. After 1 h, add 20 ml of complete Grace's medium. Incubate at 27°C.
- 48 h after infection, re-suspend all the cells with the medium from the plate, transfer to a centrifuge tube. Spin down at 4,000 rpm for 10 min. Discard the supernatant and store the cell pellet at -80 °C.

- For large scale Hi5 suspension culture, first determine the culture density by counting cells on hemacytometer. Infect when the cell count is around 4×10^6 /ml. Add 20 ml of P3 virus per 2×10^9 cells to the culture. And dilute the culture by adding same volume of complete Sf-900 medium (Sf-900 II SFM (Invitrogen) containing 100-fold diluted Penicillin-Streptomycin (Invitrogen Cat No. 15140-122), 1000-fold diluted Fungizone Antimycotic and 10 μ g/ml Gentamicin).
- 48 h after infection, pour the culture into centrifuge bottle. Spin down at 4,000 rpm for 10 min. Immediately discard the supernatant and store the cell pellet at -80 °C.

7. Large scale protein purification

- Add 50 ml of lysis buffer (25 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.5 mM TCEP-HCl, 1 mM PMSF, 1 μ g/ml Leupeptin hemisulfate, 1 μ g/ml pepstatin A) per 2×10^9 cells to the pellet. The cells were suspended in buffer and passed through EmulsiFlex-C3 homogenizer (Avestin) twice to be lysed (C), then spun down in Sorvall Discovery 90SE ultracentrifuge (Thermo Scientific) at 38,000 rpm for 1 h.
- The supernatant (S) was collected and incubated with 200 ml of Ni-NTA Agarose (50% slurry) (MCLab) equilibrated in lysis buffer for 1.5 h at 4 °C. The pellet was the insoluble fraction (P).
- Collect the supernatant by spinning down the mixture at 1,500 rpm in Sorvall RC 3BP+ centrifuge (Thermo Scientific) at 4 °C. Save the supernatant (flow-through, FT).

- Resuspend the resin in lysis buffer with 5 mM imidazole and transfer to a gravity column.
- Wash the resin with 1,000 ml of lysis buffer with 5 mM imidazole (E5), follow by 100 ml of lysis buffer with 300 mM imidazole for three times (E300) and 100 ml of lysis buffer with 500 mM imidazole for three times (E500).
- Examine 5 μ l of each C, S, P, FT, E5, E300, E500 on SDS-PAGE.
- The first two elution with 300 mM imidazole was collected, and dialyzed in Spectrum Spectra/Por 7 Membrane Tubing (Fisher Scientific) against 3.5 L of dialysis buffer (25 mM Tris-HCl pH 8.8, 50 mM NaCl, 4 mM DTT (dithiothreitol), 2 mM EDTA) for 3 h at 4 °C.
- The dialyzed sample was spun down at 4,000 rpm in Sorvall RC 3BP+ centrifuge at 4 °C. The supernatant was loaded onto 24 ml Source Q (GE Healthcare) resin, separated at 0 – 1 M NaCl gradient, and usually eluted around 200 mM NaCl.
- If the protein needs to be cleaved, the fractions containing the protein was digested with thrombin at 50:1 mass ratio in the presence of 10 mM CaCl_2 and 10 mM DTT at 4 °C for over night.
- Dilute the sample by three fold with S buffer (50 mM HEPES pH 7.0, 4 mM DTT, 2 mM EDTA) and loaded onto 8 ml Source S (GE Healthcare) resin, separated at 0 – 1 M NaCl gradient, and Rad4-Rad23 complex usually eluted around 260 mM NaCl.

- Add 4M NaCl stock to the fraction containing the protein, to make final NaCl concentration 800 mM, and concentrate to around 20 mg/ml in Amicon stirred cell (EMD Millipore).
- 1 or 2 ml of the concentrated sample is injected onto Superdex 200 10/300 GL (GE Healthcare), elute with SD buffer (5 mM BTP pH 6.8, 800 mM NaCl, 5 mM DTT). The complex usually eluted around 12.7 ml. The sample was finally concentrated to around 20 mg/ml.

8. Electrophoretic mobility shift assay and K_d determination

- Label the DNA to be probed with ^{32}P . Set up the labeling reaction as below:

	Item	Stock concentration	Volume (ml)	Final concentration
1	DNA oligo to be labeled	10 μM	2	400 nM
2	10X PNK Buffer (New England Biolab)	10X	5	
3	T4 Polynucleotide Kinase (New England Biolab)	10 U/ μl	1	0.02 U/ μl
4	adenosine 5'-triphosphate, [γ - ^{32}P] (MP Biomedicals)	24 pmol/ml	1.7	
5	Autoclaved MQ H_2O		40.3	
	Total		50	

- Incubate the reaction at room temperature for 2 h. Purify the labeled DNA with illustra MicroSpin G-25 Column (GE Healthcare). Store labeled DNA at $-20\text{ }^\circ\text{C}$.
- Pre-run 4.8% native gel at 150 V for 30 min.

- Measure the concentration of the protein, and dilute it to 1.5 μM in 5X binding buffer (25 mM BTP pH 6.8, 375 mM NaCl, 25% glycerol, 3.7 mM 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), 2.5 mg/ml bovine serum albumin, 25 mM DTT).
- Prepare 2.5 μM competitor DNA in autoclaved MQ H_2O , and 12.5 nM of ^{32}P labeled DNA to be probed in autoclaved MQ H_2O .
- Dilute the protein with 5X binding buffer as below:

tube	1	2	3	4	5	6	7	8	9	10
μl of 1.5 μM protein	0.00	3.60	6.90	10.20	13.50	16.80	20.10	23.40	26.70	30.00
μl of 5X binding buffer	30.00	26.40	23.10	19.80	16.50	13.20	9.90	6.60	3.30	0.00

- Take 2.5 μl from each tube described above, mix with 5 μl of competitor DNA, and 5 μl of labeled DNA. Incubate the mixture at room temperature for 20 min.
- Load 10 μl of the mixture on 4.8% native gel. Separate at 150 V for 15 min in 1x TBE buffer (89 mM Tris-HCl, 89 mM Boric acid, 2 mM EDTA, pH 8.0) at 4 $^{\circ}\text{C}$.

The final concentration of each lane is shown below.

Lane	1	2	3	4	5	6	7	8	9	10
Labeled DNA (nM)	5	5	5	5	5	5	5	5	5	5
Competitor DNA (μM)	1	1	1	1	1	1	1	1	1	1
Protein (nM)	0	36	69	102	135	168	201	234	267	300

- Fix the gel in fixing solution (30% methanol, 10% acetic acid) for 5 min.
- Expose the gel using Imaging Screen-K (Kodak) and the image is acquired by Personal Molecular Imager (PMI) System (Bio-rad) and Quantity One software (Bio-rad).

- The image is analyzed using Image Lab software (Bio-rad). The percentage of bound DNA is taken as the intensity of the upper band (the complex) over the total intensity of both upper and lower band (free DNA).
- The control labeled DNA in every set of experiment is CH7_NX, and the competitor DNA is also CH7_NX. The K_d for CH7_NX is calculated using the equation below:

$$y = \frac{1}{2} \left[(x_t + 1005 + K_d) - \sqrt{(x_t + 1005 + K_d)^2 - 4x_t \cdot 1005} \right]$$

where y is the amount of protein bound to DNA in nM, x_t is the total protein amount in the mixture in nM.

- For all other labeled DNA in the assay, the competitor DNA is always CH7_NX. First, the protein amount available to bind the labeled DNA of interest is calculated as below:

$$x_{ts} = x_t - \frac{1}{2} \left[(x_t + 1000 + K_{ns}) - \sqrt{(x_t + 1000 + K_{ns})^2 - 4x_t \cdot 1000} \right]$$

where x_{ts} is the protein amount available to bind the labeled DNA, K_{ns} is the K_d of CH7_NX measured in the same set of experiment.



- Then the K_d for the DNA of interest is calculated using the equation below:


$$y = \frac{1}{2} \left[(x_{ts} + 5 + K_d) - \sqrt{(x_{ts} + 5 + K_d)^2 - 4x_{ts} \cdot 5} \right]$$


where y is the amount of labeled DNA bound to protein.

- All data is taken from the average of three experiments. The fitted binding curve, K_d value and R^2 are generated in Origin.

B. Copyright Permissions



[Home](#) [Create Account](#) [Help](#)  [Live Chat](#)



Title: Kinetic gating mechanism of DNA damage recognition by Rad4/XPC
Author: Xuejing Chen, Yogambigai Velmurugu, Guanqun Zheng, Beomseok Park, Yoonjung Shim, Youngchang Kim
Publication: Nature Communications
Publisher: Nature Publishing Group
Date: Jan 6, 2015
Copyright © 2015, Rights Managed by Nature Publishing Group

[LOGIN](#)

If you're a [copyright.com user](#), you can login to RightsLink using your copyright.com credentials. Already a [RightsLink user](#) or want to [learn more?](#)

Author Use

Authors of NPG articles do not require permission to use content from their article in most cases as stated in the [author's guidelines](#).

Authors wishing to use their article for commercial purposes must request permission in the normal way.

For further questions, please contact NPG's permissions department: permissions@nature.com

[BACK](#)[CLOSE WINDOW](#)

For commercial reprints of this content, please select the Order Commercial Reprints link located beside the Rights and Permissions link on the Nature Publishing Group Web site.

Copyright © 2015 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#). Comments? We would like to hear from you. E-mail us at customercare@copyright.com

CITED LITERATURE

1. Lindahl, T., *Instability and decay of the primary structure of DNA*. Nature, 1993. **362**(6422): p. 709-715.
2. Hoeijmakers, J.H.J., *DNA damage, aging, and cancer*. New England Journal of Medicine, 2009.
3. Jiricny, J., *The multifaceted mismatch-repair system*. Nature reviews. Molecular cell biology, 2006. **7**(5): p. 335-346.
4. Modrich, P. and R. Lahue, *Mismatch repair in replication fidelity, genetic recombination, and cancer biology*. Annual review of biochemistry, 1996.
5. Krokan, H.E. and M. Bjørås, *Base excision repair*. Cold Spring Harbor perspectives ..., 2013.
6. Sancar, A., *Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors*. Chem. Rev., 2003. **103**(6).
7. Mishina, Y., E.M. Duguid, and C. He, *Direct reversal of DNA alkylation damage*. Chemical reviews, 2006.
8. Mees, A., et al., *Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair*. Science, 2004. **306**(5702): p. 1789-1793.
9. Yang, C.-G.G., et al., *Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA*. Nature, 2008. **452**(7190): p. 961-965.
10. Lehmann, A.R., *Translesion synthesis in mammalian cells*. Experimental cell research, 2006.
11. Kanaar, R., J.H.J. Hoeijmakers, and D.C. van Gent, *Molecular mechanisms of DNA double-strand break repair*. Trends in cell biology, 1998.
12. Khanna, K.K. and S.P. Jackson, *DNA double-strand breaks: signaling, repair and the cancer connection*. Nature genetics, 2001.
13. Vermeulen, W. and M. Foustéri, *Mammalian Transcription-Coupled Excision Repair*. Cold Spring Harbor Perspectives in Biology, 2013. **5**(8).
14. Gillet, L.C. and O.D. Schärer, *Molecular mechanisms of mammalian global genome nucleotide excision repair*. Chemical reviews, 2006. **106**(2): p. 253-276.
15. Robbins, J. and K.H. Kraemer, *Xeroderma pigmentosum: an inherited disease with sun sensitivity, multiple cutaneous neoplasms, and abnormal DNA repair*. Annals of Internal ..., 1974.
16. Truglio, J.J., et al., *Prokaryotic nucleotide excision repair: the UvrABC system*. Chemical reviews, 2006. **106**(2): p. 233-252.
17. Eastman, A., *Reevaluation of interaction of cis-dichloro (ethylenediamine) platinum (II) with DNA*. Biochemistry, 1986.
18. McHugh, P.J., V.J. Spanswick, and J.A. Hartley, *Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance*. The lancet oncology, 2001.
19. Deans, A.J. and S.C. West, *DNA interstrand crosslink repair and cancer*. Nature Reviews Cancer, 2011.

20. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease*. Nature, 2009.
21. Mitchell, D.L., *The relative cytotoxicity of (6-4) photoproducts and cyclobutane dimers in mammalian cells*. Photochemistry and photobiology, 1988. **48**(1): p. 51-57.
22. Sinha, R.P. and D.P. Häder, *UV-induced DNA damage and repair: a review*. Photochemical & Photobiological Sciences, 2002.
23. Gelboin, H.V., *Benzo [alpha] pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes*. Physiological reviews, 1980.
24. Turesky, R.J., *Heterocyclic aromatic amine metabolism, DNA adduct formation, mutagenesis, and carcinogenesis*. Drug metabolism reviews, 2002. **34**(3): p. 625-650.
25. Sugasawa, K., *A multistep damage recognition mechanism for global genomic nucleotide excision repair*. Genes & Development, 2001. **15**(5).
26. Anantharaman, V., E.V. Koonin, and L. Aravind, *Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold*. Human molecular genetics, 2001. **10**(16): p. 1627-1630.
27. Bunick, C.G., et al., *Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein*. Biochemistry, 2006. **45**(50): p. 14965-14979.
28. Uchida, A., et al., *The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH*. DNA repair, 2002. **1**(6): p. 449-461.
29. Popescu, A., et al., *Xeroderma pigmentosum group C protein possesses a high affinity binding site to human centrin 2 and calmodulin*. The Journal of biological chemistry, 2003. **278**(41): p. 40252-40261.
30. Karla, A.H., et al., *Cloning the Drosophila homolog of the xeroderma pigmentosum complementation group C gene reveals homology between the predicted human and Drosophila polypeptides and that encoded by the yeast RAD4 gene*. Nucleic Acids Research, 1994. **22**(3).
31. Legerski, R. and C. Peterson, *Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C*. Nature, 1992. **359**(6390): p. 70-73.
32. Schaubert, C., et al., *Rad23 links DNA repair to the ubiquitin/proteasome pathway*. Nature, 1998. **391**(6668): p. 715-718.
33. Raasi, S., *Rad23 Ubiquitin-associated Domains (UBA) Inhibit 26 S Proteasome-catalyzed Proteolysis by Sequestering Lysine 48-linked Polyubiquitin Chains*. Journal of Biological Chemistry, 2003. **278**.
34. Masutani, C., et al., *Identification and characterization of XPC-binding domain of hHR23B*. Molecular and cellular biology, 1997. **17**(12): p. 6915-6923.
35. Ortolan, T.G., et al., *Rad23 stabilizes Rad4 from degradation by the Ub/proteasome pathway*. Nucleic acids research, 2004. **32**(22): p. 6490-6500.

36. Min, J.-H.H. and N.P. Pavletich, *Recognition of DNA damage by the Rad4 nucleotide excision repair protein*. *Nature*, 2007. **449**(7162): p. 570-575.
37. Scrima, A., et al., *Structural Basis of UV DNA-Damage Recognition by the DDB1–DDB2 Complex*. *Cell*, 2008. **135**(7).
38. Truglio, J.J., et al., *Structural basis for DNA recognition and processing by UvrB*. *Nature structural & ...*, 2006.
39. Ikegami, T., et al., *Resonance assignments, solution structure, and backbone dynamics of the DNA-and RPA-binding domain of human repair factor XPA*. *Journal of ...*, 1999.
40. Hey, T., et al., *The XPC HR23B Complex Displays High Affinity and Specificity for Damaged DNA in a True-Equilibrium Fluorescence Assay*. *Biochemistry*, 2002. **41**(21).
41. Szymkowski, D.E. and C.W. Lawrence, *Repair by human cell extracts of single (6-4) and cyclobutane thymine-thymine photoproducts in DNA*. ... the National Academy ..., 1993.
42. Kim, J.K., D. Patel, and B.S. Choi, *Contrasting structural impacts induced by cis-syn cyclobutane dimer and (6–4) adduct in DNA duplex decamers: implication in mutagenesis and repair activity*. *Photochemistry and Photobiology*, 1995.
43. Hess, M.T., et al., *Base pair conformation-dependent excision of benzo[a]pyrene diol epoxide-guanine adducts by human nucleotide excision repair enzymes*. *Molecular and cellular biology*, 1997. **17**(12): p. 7069-7076.
44. Gunz, D., M.T. Hess, and H. Naegeli, *Recognition of DNA adducts by human nucleotide excision repair Evidence for a thermodynamic probing mechanism*. *Journal of Biological Chemistry*, 1996.
45. Wittschieben, B., *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*, 2005. **280**(48).
46. Hwang, B., et al., *Expression of the p48 xeroderma pigmentosum gene is p53-dependent and is involved in global genomic repair*. *Proceedings of the National Academy of Sciences*, 1999. **96**(2).
47. Fitch, M., et al., *In Vivo Recruitment of XPC to UV-induced Cyclobutane Pyrimidine Dimers by the DDB2 Gene Product*. *Journal of Biological Chemistry*, 2003. **278**(47).
48. Wang, Q.E., et al., *UV radiation-induced XPC translocation within chromatin is mediated by damaged-DNA binding protein, DDB2*. *Carcinogenesis*, 2004.
49. Sugasawa, K., et al., *UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex*. *Cell*, 2005.
50. Akita, M., et al., *SUMOylation of xeroderma pigmentosum group C protein regulates DNA damage recognition during nucleotide excision repair*. *Scientific reports*, 2015. **5**: p. 10984.
51. Poulsen, S., et al., *RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response*. *The Journal of Cell Biology*, 2013. **201**(6).

52. Matsuoka, S., B.A. Ballif, and A. Smogorzewska, *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage*. ..., 2007.
53. Puumalainen, M.-R., et al., *Chromatin retention of DNA damage sensors DDB2 and XPC through loss of p97 segregase causes genotoxicity*. Nat Comms, 2014. **5**.
54. Melis, J.P.M., S.W.P. Wijnhoven, and R.B. Beems, *Mouse models for xeroderma pigmentosum group A and group C show divergent cancer phenotypes*. Cancer research, 2008.
55. Shimizu, Y., et al., *Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase*. The EMBO Journal, 2003. **22**(1): p. 164-173.
56. D'Errico, M., E. Parlanti, and M. Teson, *New functions of XPC in the protection of human skin cells from oxidative damage*. The EMBO ..., 2006.
57. Menoni, H., J. Hoeijmakers, and W. Vermeulen, *Nucleotide excision repair-initiating proteins bind to oxidative DNA lesions in vivo*. The Journal of Cell Biology, 2012. **199**(7).
58. Ray, A., et al., *NER initiation factors, DDB2 and XPC, regulate UV radiation response by recruiting ATR and ATM kinases to DNA damage sites*. DNA repair, 2013. **12**(4): p. 273-283.
59. Gong, F., D. Fahy, and M.J. Smerdon, *Rad4-Rad23 interaction with SWI/SNF links ATP-dependent chromatin remodeling with nucleotide excision repair*. Nature structural & molecular biology, 2006. **13**(10): p. 902-907.
60. Fong, Yick W., et al., *A DNA Repair Complex Functions as an Oct4/Sox2 Coactivator in Embryonic Stem Cells*. Cell, 2011. **147**(1).
61. van der Spek, P.J., et al., *XPC and human homologs of RAD23: intracellular localization and relationship to other nucleotide excision repair complexes*. Nucleic acids research, 1996. **24**(13): p. 2551-2559.
62. He, C. and G.L. Verdine, *Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking*. Chemistry & biology, 2002. **9**(12): p. 1297-1303.
63. Otwinowski, Z. and W. Minor, *Processing of X-ray Diffraction Data Collected in Oscillation Mode*. Methods in Enzymology, 1997. **276**(Macromolecular Crystallography, part A): p. 307-326.
64. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 2): p. 213-221.
65. Yi, C., et al., *Duplex interrogation by a direct DNA repair protein in search of base damage*. Nature structural & molecular biology, 2012. **19**(7): p. 671-676.
66. Banerjee, A., et al., *Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA*. Nature, 2005. **434**(7033): p. 612-618.
67. Kubelka, J., *Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics*. Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology, 2009. **8**(4): p. 499-512.

68. Ward, D.C., E. Reich, and L. Stryer, *Fluorescence studies of nucleotides and polynucleotides I. Formycin, 2-aminopurine riboside, 2, 6-diaminopurine riboside, and their derivatives*. Journal of Biological Chemistry, 1969.
69. Holz, B., E. Weinhold, and S. Klimasauskas, *2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases*. Nucleic acids ..., 1998.
70. Stivers, J.T., K.W. Pankiewicz, and K.A. Watanabe, *Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase*. Biochemistry, 1999. **38**(3): p. 952-963.
71. Chen, X., et al., *Kinetic gating mechanism of DNA damage recognition by Rad4/XPC*. Nat Commun, 2015. **6**: p. 5849.
72. Wilhelmsson, L.M., et al., *A highly fluorescent DNA base analogue that forms Watson-Crick base pairs with guanine*. Journal of the American Chemical Society, 2001. **123**(10): p. 2434-2435.
73. Engman, K.C., et al., *DNA adopts normal B-form upon incorporation of highly fluorescent DNA base analogue tC: NMR structure and UV-Vis spectroscopy characterization*. Nucleic acids research, 2004. **32**(17): p. 5087-5095.
74. Sandin, P., et al., *Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue*. Nucleic acids research, 2008. **36**(1): p. 157-167.
75. Börjesson, K., et al., *Nucleic acid base analog FRET-pair facilitating detailed structural measurements in nucleic acid containing systems*. Journal of the American Chemical Society, 2009. **131**(12): p. 4288-4293.
76. Kuznetsov, S.V., et al., *Loop dependence of the stability and dynamics of nucleic acid hairpins*. Nucleic acids research, 2008. **36**(4): p. 1098-1112.
77. Coman, D. and I.M. Russu, *A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA*. Biophysical journal, 2005.
78. Moe, J.G. and I.M. Russu, *Kinetics and energetics of base-pair opening in 5'-d (CGCGAATTCGCG)-3' and a substituted dodecamer containing G. cntdot. T mismatches*. Biochemistry, 1992.
79. Parker, J.B., et al., *Enzymatic capture of an extrahelical thymine in the search for uracil in DNA*. Nature, 2007. **449**(7161): p. 433-437.
80. Rachofsky, E.L., R. Osman, and J.B. Ross, *Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence*. Biochemistry, 2001. **40**(4): p. 946-956.
81. Narayanan, M., et al., *Photoinduced electron transfer occurs between 2-aminopurine and the DNA nucleic acid monophosphates: results from cyclic voltammetry and fluorescence quenching*. The journal of physical chemistry. B, 2010. **114**(32): p. 10573-10580.
82. Geacintov, N.E., et al., *Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery*. Biopolymers, 2002. **65**(3): p. 202-210.
83. Zheng, H., et al., *Base flipping free energy profiles for damaged and undamaged DNA*. Chemical research in toxicology, 2010. **23**(12): p. 1868-1870.

84. Araki, M., *Centrosome Protein Centrin 2/Caltractin 1 Is Part of the Xeroderma Pigmentosum Group C Complex That Initiates Global Genome Nucleotide Excision Repair*. Journal of Biological Chemistry, 2001. **276**(22).
85. Friedberg, F., *Centrin isoforms in mammals. Relation to calmodulin*. Molecular biology reports, 2006. **33**(4): p. 243-252.
86. Nishi, R., et al., *Centrin 2 Stimulates Nucleotide Excision Repair by Interacting with Xeroderma Pigmentosum Group C Protein*. Molecular and Cellular Biology, 2005. **25**(13).
87. Krasikova, Y., et al., *Influence of centrin 2 on the interaction of nucleotide excision repair factors with damaged DNA*. Biochemistry (Moscow), 2012. **77**(4).
88. Chen, L. and K. Madura, *Centrin/Cdc31 is a novel regulator of protein degradation*. Molecular and cellular biology, 2008. **28**(5): p. 1829-1840.
89. den Dulk, B., et al., *The NER protein Rad33 shows functional homology to human Centrin2 and is involved in modification of Rad4*. DNA repair, 2008. **7**(6): p. 858-868.
90. Fitzgerald, D.J., et al., *Protein complex expression by using multigene baculoviral vectors*. Nature methods, 2006. **3**(12): p. 1021-1032.
91. Berger, I., D.J. Fitzgerald, and T.J. Richmond, *Baculovirus expression system for heterologous multiprotein complexes*. Nature biotechnology, 2004. **22**(12): p. 1583-1587.
92. Modesti, M., *Fluorescent labeling of proteins*. Methods in molecular biology (Clifton, N.J.), 2011. **783**: p. 101-120.
93. Schmidt, B., et al., *A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency*. Cell, 1995. **82**(2): p. 271-278.
94. Maria Pia, C., et al., *The Multiple Sulfatase Deficiency Gene Encodes an Essential and Limiting Factor for the Activity of Sulfatases*. Cell, 2003. **113**(4).
95. Carrico, I.S., B.L. Carlson, and C.R. Bertozzi, *Introducing genetically encoded aldehydes into proteins*. Nature chemical biology, 2007. **3**(6): p. 321-322.
96. Shi, X., et al., *Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging*. Nature methods, 2012. **9**(5): p. 499-503.
97. Wu, P., et al., *Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag*. Proceedings of the ..., 2009.

VITA

Education

Sep2005 – Jul2009 China Agricultural University, College of Resources and Environmental Sciences
Bachelor of Science in Ecology

Teaching Experience

Biochemistry Laboratory (CHEM455), 24 per class
Biochemistry I (CHEM 452), Discussion, 90 per class
Inorganic Chemistry (CHEM 314), Laboratory, 40 per class
Survey of Organic and Biochemistry (CHEM 130), Laboratory and Discussion, 40 per class
General Chemistry (CHEM 114), Laboratory and Discussion, 40 per class

Awards

- Paaren Graduate Fellowships, UIC Department of Chemistry, 2014 – 2015
- Women in Science and Engineering (WISE) UIC Graduate Student Research Award, 2015.

Publications

1. Chen, X.[¶], Velmurugu, Y.[¶], Zheng, G., Park, B. P., Shim, Y., Kim, Y., Liu, L., Van Houten, B., He, C., Ansari, A. and Min, J. H. Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat Commun.* 2015.6:5849. [¶] These authors contributed equally to this work.
2. Velmurugu, Y.[§], Chen, X.[§], Patel, A., Slogoff-Sevilla, P., Min, J.H., and Ansari, A. Twist-open mechanism of DNA damage recognition by Rad4/XPC nucleotide excision repair complex. Under revision of *PNAS*. 2016 [§]These authors contributed equally to this work. Shafirovich, V., Kropachev, K., Anderson, T., Liu, Z., Kolbanovskiy, M., Martin, B.D., Sugden, K., Shim, Y., Chen, X., Min, J.H., & Geacintov, N.E., Base and Nucleotide Excision Repair of Oxidatively Generated Guanine Lesions in DNA. *Journal of Biological Chemistry*. 2016. 291: 5309–19.
3. Lee, Y.C., Cai, Y., Mu H., Broyde, S., Amin, S., Chen, X., Min, J.H., and Geacintov, N.E. The relationships between XPC binding to conformationally diverse DNA adducts and their excision by the human NER system: is there a correlation? *DNA Repair*. 2014.19:55-63.
4. Krasikova, Y.S., Rechkunova, N.I., Maltseva, E.A., Pestryakov, P.E., Petruseva, I.O., Sugasawa, K., Chen, X., Min, J.H. and Lavrik, O.I. Comparative Analysis of Interaction of Human and Yeast DNA Damage Recognition Complexes with

- Damaged DNA in Nucleotide Excision Repair. *J Biol Chem*. 2013. 288: 10936-10947.
5. Shim, Y., Duan, M.R., Chen, X., Smerdon, M.J., and Min, J.H. Polycistronic coexpression and nondenaturing purification of histone octamers. *Analytical Biochemistry*.2012.427(2):190-192.
- *Kong, M., Liu, L., Chen, X., Driscoll, K., Böhm, S., Mao, P., Watkins, S., Wyrick, J., Bernstein, K., Min, J.H., and Van Houten, B., Single-molecule imaging reveals that Rad4 (XPC) employs a dynamic DNA damage recognition process in a lesion specific manner. Under revision for *Molecular Cell*. 2016

Presentations

1. Mechanism of DNA damage recognition by Rad4/XPC nucleotide excision repair complex. Keystone Symposia Conference: Genomic Instability and DNA Repair, Whistler, Canada, 2015.
2. Mechanism of DNA damage searching by XPC. International Year of Crystallography Symposium Chicago, Chicago, 2014.
3. Mechanism of DNA damage searching by XPC. UIC Cancer Center Cancer Research Forum, Chicago 2014.
4. Structure and Mechanism of the XPC DNA repair complex. University of Illinois at Chicago (UIC) Student Research Forum, Chicago 2012.
5. DNA Damage Search and Recognition by the XPC Nucleotide Excision Repair Complex. 14th Annual Midwest DNA Repair Symposium, Cincinnati 2012.