

ABSTRACT

Compilation and Structural Analyses of Disease-linked Mutations in the Xeroderma Pigmentosum C (XPC) DNA Repair Protein

Jennifer Le

Director: Jung Hyun Min, Ph.D.

The xeroderma pigmentosum C (XPC) protein complex recognizes various types of environmentally induced DNA damage from the genome to initiate the nucleotide excision repair (NER) process in eukaryotes. In this study, (1) I compiled a list of missense mutations/variants in the human *XPC* gene that are linked to diseases (e.g., cancer) or impair its functions *in vitro*, from three public databases and 13 research papers. (2) I have made a homology-directed 3D structural model of human XPC protein based on existing yeast Rad4 structures; and (3) I mapped the compiled mutations onto the homology model. This work provides an important starting point to understand the impact of the various mutations in the function of the XPC protein.

APPROVED BY DIRECTOR OF HONORS THESIS:

Dr. Jung Hyun Min, Department of Chemistry and Biochemistry

APPROVED BY THE HONORS PROGRAM:

Dr. Elizabeth Corey, Director

DATE: _____

Compilation and Structural Analyses of Disease-Linked Mutations in the Xeroderma
Pigmentosum C (XPC) DNA Repair Protein

A Thesis Submitted to the Faculty of
Baylor University
In Partial Fulfillment of the Requirements for the
Honors Program

By
Jennifer Le

Waco, Texas

May 2020

TABLE OF CONTENTS

<i>List of Figures</i>	<i>iii</i>
<i>List of Tables</i>	<i>iii</i>
CHAPTER ONE: Introduction on NER, XPC, and their association to human diseases...	1
1.1 DNA damage and repair	1
1.2 Nucleotide Excision Repair (NER)	2
1.3 Xeroderma Pigmentosum (XP)	3
1.4 Diagnosis and Classification of XP	4
1.5 The XPC-RAD23B-Centrin2 complex: in vivo function	6
1.6 Structure of the XPC	6
1.7 In vitro DNA binding of the Rad4-Rad23 complex	12
CHAPTER TWO: Literature survey on disease-associated mutations in XPC.....	13
CHAPTER THREE: Literature survey on in vitro mutations in XPC.....	25
CHAPTER FOUR: Homology Model of XPC	26
CHAPTER FIVE: Structural Analyses of XPC Genetic Variations.....	31
CHAPTER SIX: Summary	47
REFERENCES	48

List of Figures

Figure 1. UV light induced DNA damage	1
Figure 2. Nucleotide excision repair (NER) pathway.....	3
Figure 3. Classical presentation of xeroderma pigmentosum (XP) patients.....	4
Figure 4. UDS assay.	5
Figure 5. Sequence alignment of yeast Rad4 and orthologs.....	9
Figure 6. Homology between Rad4 and XPC on Molecular Surface.....	10
Figure 7. Overall Structures of the Rad4-Rad23-DNA complex.....	11
Figure 8. Ramachandran plot and MolProbity results of SWISS-M	30
Figure 9. Locations of clusters mapped on the overall XPC homology model structure.	40
Figure 10. Locations of clusters mapped on the overall XPC homology model structure.	42
Figure 11. Local structural views for Clusters 1-4.	43
Figure 12. Local structural views for Clusters 5-8.	44
Figure 13. Local structural views for Clusters 9-12.	45
Figure 14. Local structural views for Clusters 13-16.	46
Figure 15. Local structural views for Clusters 17-20.	46

List of Tables

Table 1. Missense mutations in the XPC gene compiled from Tumor Portal	15
Table 2. Missense genetic mutations in the XPC gene compiled from ICGC.....	16
Table 3. Single amino acid genetic variants in the XPC gene compiled from ClinVar. ..	21
Table 4. Missense genetic mutations in the XPC gene compiled from the literature.	23
Table 5. Functionally important XPC amino acid changes caused by in vitro mutagenesis.	25

CHAPTER ONE

Introduction on NER, XPC, and their association to human diseases

1.1 DNA damage and repair

Cells are continuously exposed to various environmental sources such as sunlight that can cause damage to DNA. Exposure to UV radiation, for instance, causes DNA to form stable dipyrимidine dimers. As shown in **Figure 1**, the major photoproducts are cyclobutane pyrimidine dimers (CPDs) of adjacent cytosines or thymines, as well as 6-4 pyrimidine pyrimidone photoproducts (6-4PPs) (1). The prevalence of each is affected by DNA sequence and UV wavelengths, but CPDs are generally the major source of mutation in mammalian cells (2). The yield of CPD is 3 to 4 times higher than 6-4PP upon UV-C and UV-B irradiation at ≤ 296 nm (2). These lesions differ in chemical composition and cause varying distortions of the DNA helix. If left unrepaired, CPDs and 6-4PPs promote genomic instability and inhibit replication and transcription.

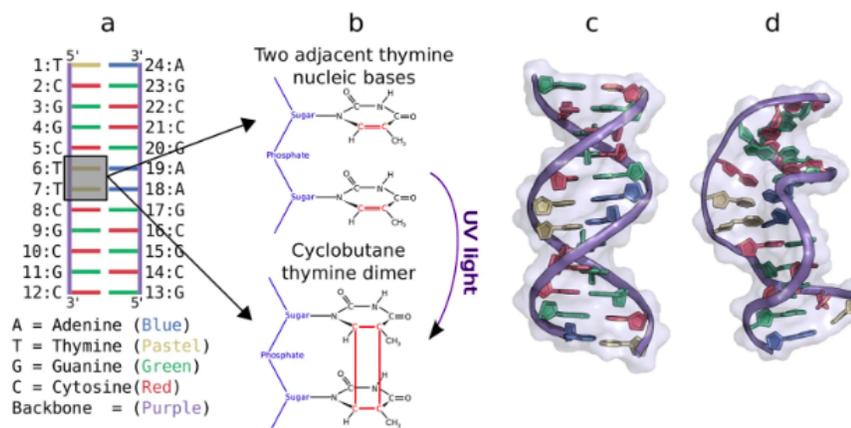


Figure 1. UV light induced DNA damage. (a) DNA segment with dipyrимidine dimer. (b) Structures of adjacent thymine bases and photoproduct. (c) Normal B-DNA structure compared to (d) deformed DNA structure containing CPD (pdb-entry:PDB:1TEZ5). Figure is modified from (3)

1.2 Nucleotide Excision Repair (NER)

Nucleotide Excision Repair (NER) repairs a variety of structurally diverse, bulky lesions caused by UV damage and other carcinogenic agents. It recognizes, excises, and replaces damaged bases on one DNA strand (**Figure 2**) (reviewed in ref. (4,5)). Two sub-pathways of NER recognize base lesions. Global genome repair (GGR) identifies photoproducts located in non-transcribed sites of the genome, while transcription coupled repair (TCR) is specialized to recognize photoproducts located only on transcribed strands of active genes (6). TCR is initiated by the presence of stalled RNA pol II and proteins CSA and CSB. In contrast, GGR is initiated by the XPC complex which specifically scans the genome to locate bulky UV lesions.

After the lesion is identified, both NER sub-pathways share a common mechanism that involves XPA, -B, -D, -F, and -G groups. First, the basal transcription factor IIIH (TFIIH) complex including the XPB and XPD helicases is recruited to unwind duplex DNA around the bulky lesion (7). Then the fully opened DNA complex, called the pre-incision complex, recruits additional factors. XPA protein stabilizes the complex. The endonucleases XPG and ERCC1-XPF excise a single-stranded oligonucleotide segment containing the DNA damage. Finally, DNA polymerase δ or ϵ synthesizes the correct bases, assisted by the proliferating cell nuclear antigen (PCNA) sliding clamp and replication factor C. The strands are rejoined by DNA ligase I (4,8).

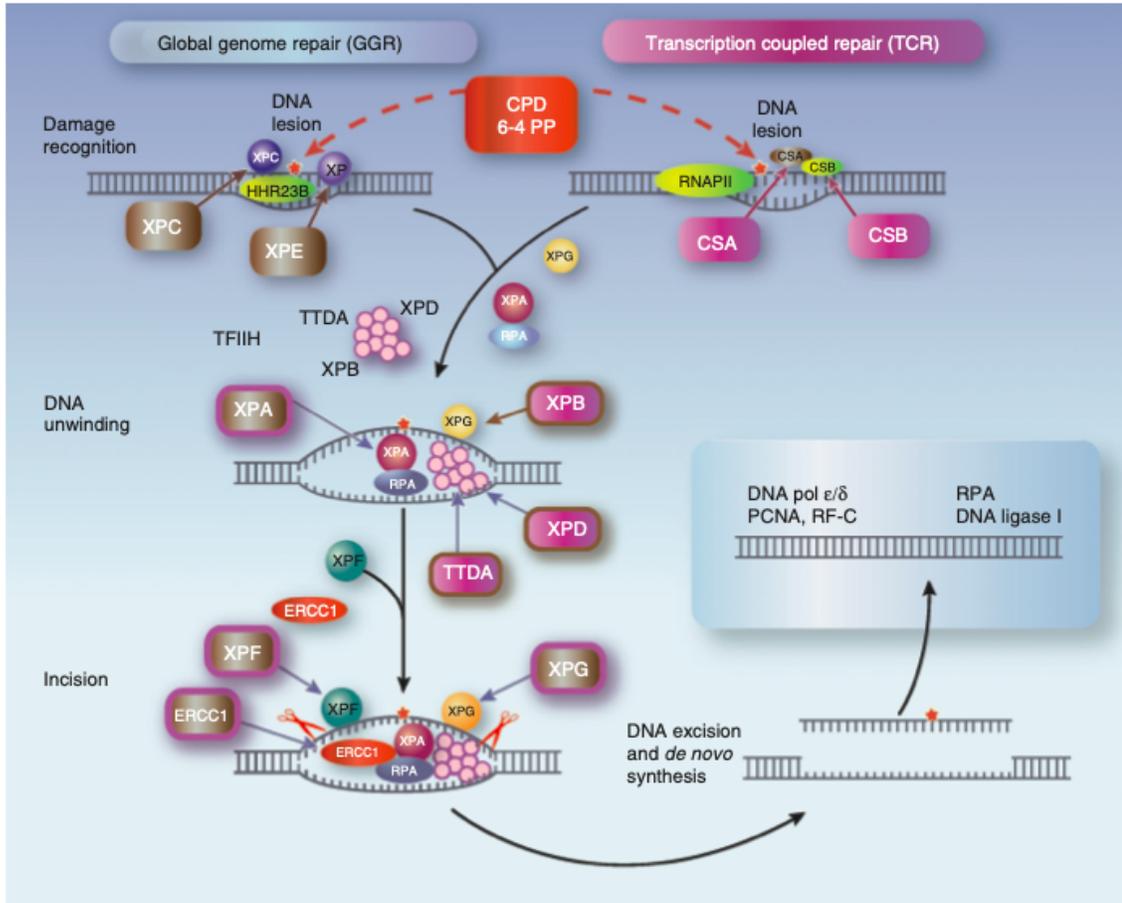


Figure 2. Nucleotide excision repair (NER) pathway. Figure is from (8).

1.3 Xeroderma Pigmentosum (XP)

The impact of unrepaired UV-induced photolesions in DNA were initially discovered by the studies on xeroderma pigmentosum patients. XP is an autosomal recessive cancer predisposition syndrome caused by genetic defects in key NER genes (reviewed in (8)). The decreased ability to repair UV-induced DNA lesions by NER leads to the accumulation of photo-damage. Such damage, if left unrepaired, leads to genetic mutations, which can eventually give rise to cancers especially in the sun-exposed organs such as skin. XP patients are estimated to have a 2,000-fold increased risk of melanoma before the age of 20 (8). Common XP symptoms include an increased sensitivity to

sunburn, abnormal lentiginosis, ocular lesions, photophobia, and neurological disorders.

The skin of XP patients appears dry, thin, and wrinkled with pigmentation (**Figure 2**). XP has been found in all racial groups and continents (4). The severity of clinical symptoms mostly depends on the amount of sun exposure, followed by the type and location of mutation and level of melanin in the skin (4). Areas of the body that are more exposed to sunlight have a higher risk of cancer (9). There is currently no cure for XP. However, symptoms are lessened by minimizing sun exposure and surgically removing tumors.



Figure 3. Classical presentation of xeroderma pigmentosum (XP) patients. **(a)** Symptoms of severe skin blistering after brief sun exposure. **(b)** Symptoms of skin hyperpigmentation, lesions, and cell. Figure modified from (7).

1.4 Diagnosis and Classification of XP

The diagnosis of XP can be definitively confirmed using cellular tests for defective DNA repair (4). The most common cellular test is a UDS (unscheduled DNA synthesis) assay. UDS measures the level of NER in the cell, and a reduction in UDS confirms the diagnosis of XP. Primary dermal fibroblasts are collected from a punch biopsy, then irradiated with UV light. UDS is the measurement of incorporated nucleotides by irradiated cells. In **Figure 4**, the irradiated cells are incubated with ^3H -

thymidine, a DNA radioactive precursor. Autoradiography is used to show the labelling pattern in G1 and G2 nuclei. The number of grains indicates the amount of ^3H -thymidine incorporated during repair synthesis and directly measures cellular ability to perform UDS. Compared to normal cells in the left panel, the cells of an XP patient in the right panel show fewer grains and a reduced level of UDS.

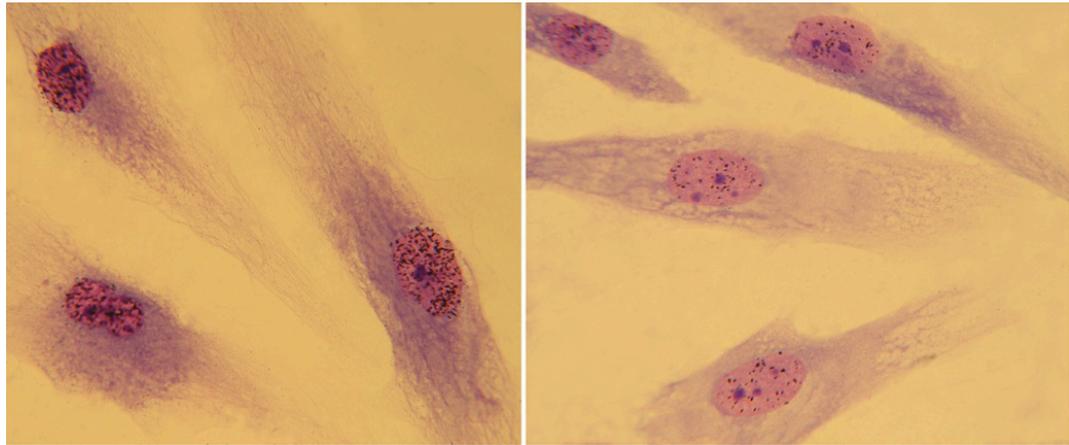


Figure 4. UDS assay. Left panel shows cells with normal UDS levels and relatively higher number of grains. Right panel shows XP cells with reduced UDS levels and fewer grains. Figure is from (4).

The classification of XP patients into specific complementation groups (XP-A through XP-G) is based on which XP gene is defective. This classification developed from an understanding of the heterogeneity of the XP disease (10). When fibroblasts of different XP patients were fused, the hybridized binuclear cells (also called heterodikaryons) displayed normal levels of DNA synthesis. The results indicated that each of the XP fibroblasts had a mutation located in a different gene; and a wild-type allele masked each recessive XP allele, showing genetic complementation. Eight complementation groups have been identified. They correspond to eight genes that are named XPA through G and V. The genes XPA through G code for NER proteins, while

XPV codes for DNA pol η that replicates unrepaired DNA (10). A mutation within these eight complementation groups causes XP.

1.5 The XPC-RAD23B-Centrin2 complex: in vivo function

The XPC-RAD23B-Centrin2 complex recognizes a wide variety of chemically unrelated DNA lesions. It is an important component in nucleotide excision repair that recognizes bulky lesions and recruits additional NER factors to the open complex. The XPC heterotrimeric complex consists of XPC, RAD23B and CETN2. The DNA damage binding property of XPC resides solely in the XPC protein. RAD23B is a 26S proteasome-interacting factor, and CETN2 is an EF-hand calcium-binding protein (11–13). RAD23B and CETN2 support the proper folding of XPC to increase DNA-binding affinity, and protect XPC from degradation (11). The XPC-RAD23B-Centrin2 complex scans the genome to identify thermodynamically destabilizing DNA lesions including sunlight-induced 6-4PPs and CPDs, as well as various bulky organic adducts (14). Lesion-bound XPC then recruits transition factor IIIH (TFIIH) to the lesion site. The helicase activity of TFIIH subunits XPB and XPD enables the subsequent steps of NER, damage excision and gap-filling synthesis and ligation (4,13). The ability to recognize chemically and structurally diverse lesions is possible due to its unique DNA binding mechanism, which is further discussed below.

1.6 Structure of the XPC

The *XPC* gene consists of 16 exons on the chromosomal location of 3p25.1 and encodes a protein of 940 residues (4). No high-resolution crystal image of human XPC-RAD23B has been solved, but an evolutionarily conserved homolog from *Saccharomyces*

cerevisiae (Rad4-Rad23 complex) appears to have similar binding properties to DNA, and function equivalently in NER (2,12). Furthermore, a strong sequence homology is shared by human RAD23B and yeast Rad23 (**Figure 5A, Figure 6**), as well as human CETN2 and CETN2 yeast homolog Rad33 (**Figure 5C**) (12). A low resolution cryo-EM image of the human XPC complex also shares homologous features with the crystal structure of yeast Rad4 (12).

Human XPC and yeast Rad4 share sequence homology of key domains that are essential for interaction with both DNA and protein cofactors (**Figure 7B**) (12). The structure of yeast Rad4 contains a ~45-residue core transglutaminase fold (TGD in **Figure 7A**) near the N-terminal (14). This is followed by three consecutive 50-90-residue [alpha]/[beta] domains, each containing a long [beta]-hairpin (BHD1, BHD2 and BHD3, for beta-hairpin domain) (14). The TGD and BHDs fold together consecutively, with the TGD forming a large lobe, and the three BHDs folding away from it as a curved protrusion (14).

Similarly, the human XPC sequence contains two regions (154–331, 517–657) that align with members of the transglutaminase superfamily, but it is separated by a ~175 residue linker region (**Figure 7A**) (13). This is followed by the predicted sequence of three consecutive beta-hairpin domains (**Figure 7B**) (12). The predicted human XPC structure contains a transglutaminase fold composed of a highly α -helical N-terminal subdomain that packs with the β -strands of the C-terminal subdomain (13,14). **Figure 8** shows the similar homologous structures of yeast Rad4 and human XPC (14).

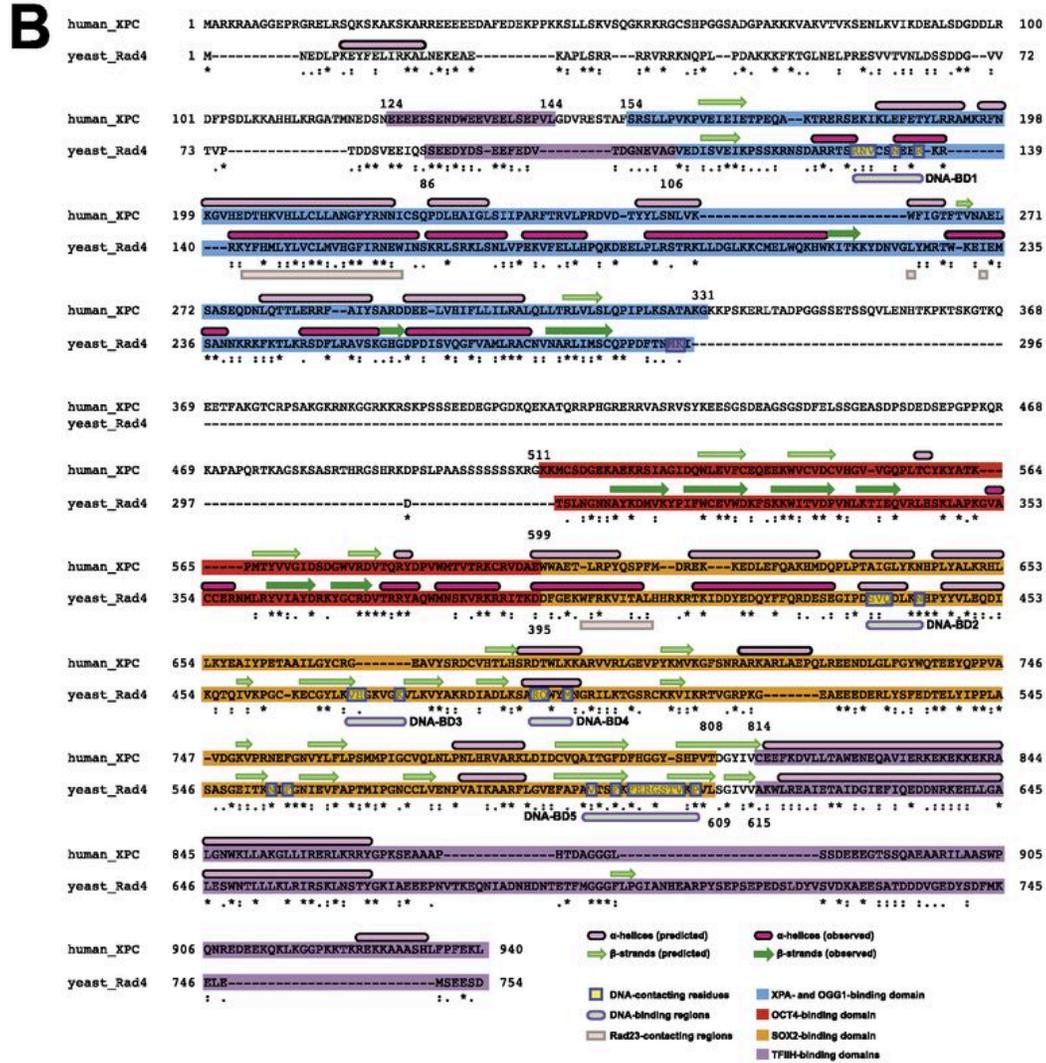
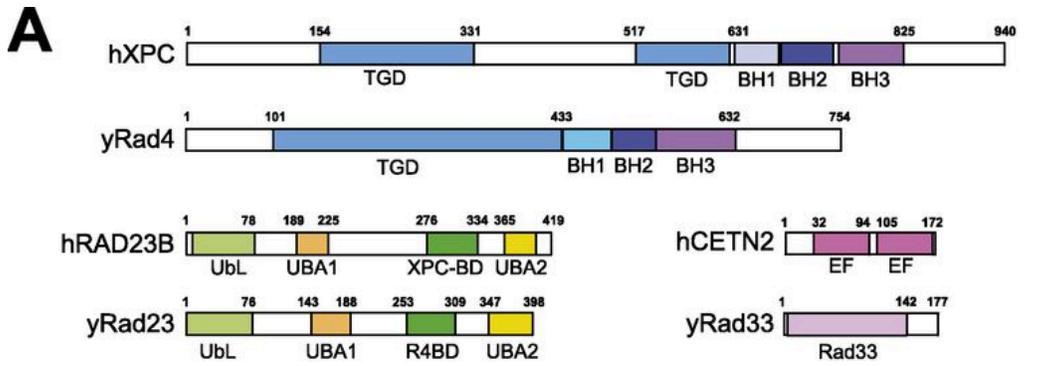


Figure 5. Conservation analysis via comparison of the human XPC complex and the yeast Rad4 complex. **(a)** Schematic representation of the subunits and domains of the human XPC complex and the yeast Rad4 complex. **(b)** Sequence alignment between human XPC and yeast Rad4. Figure is modified from (12)

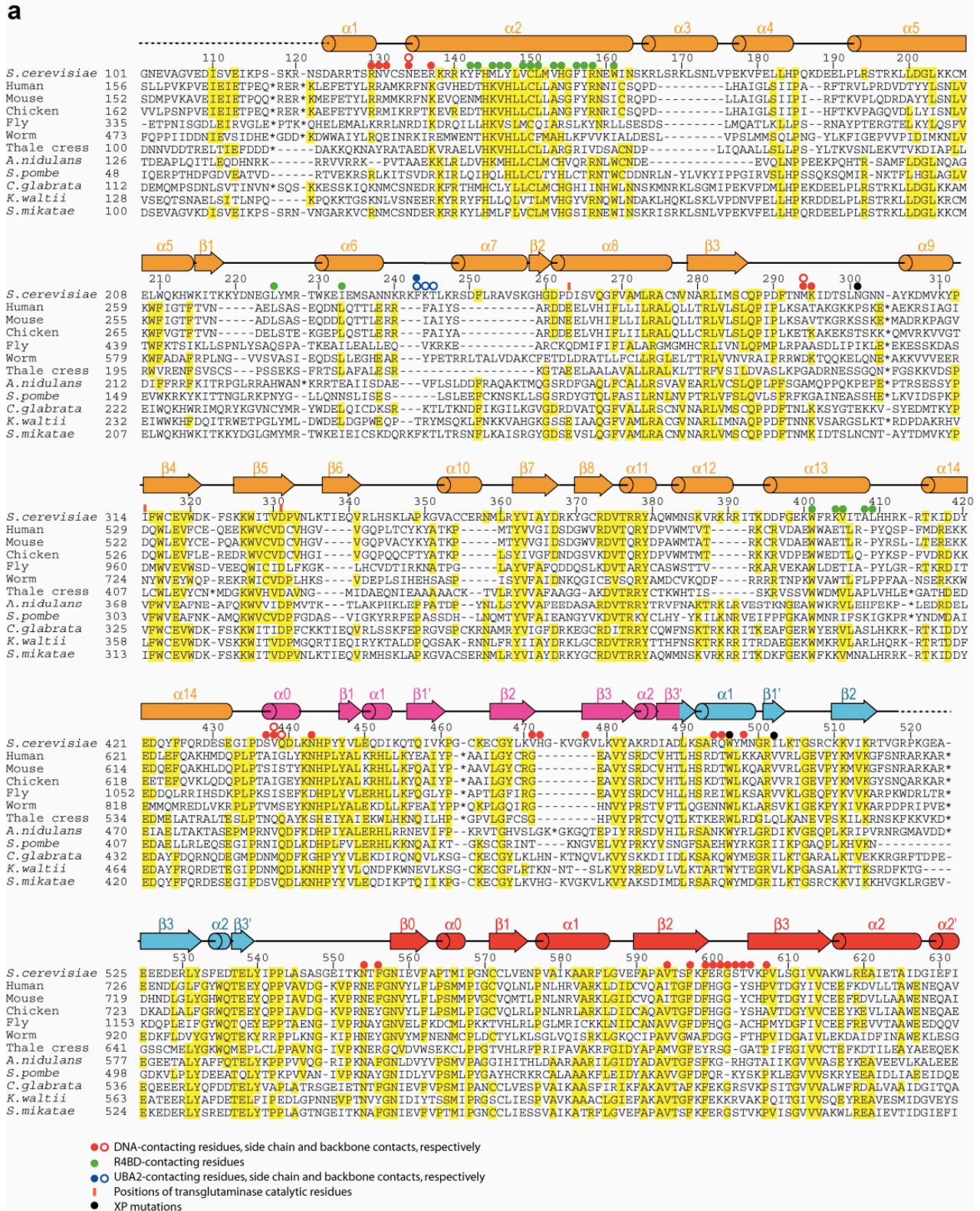


Figure 6. Sequence alignment of yeast Rad4 and orthologs (14).

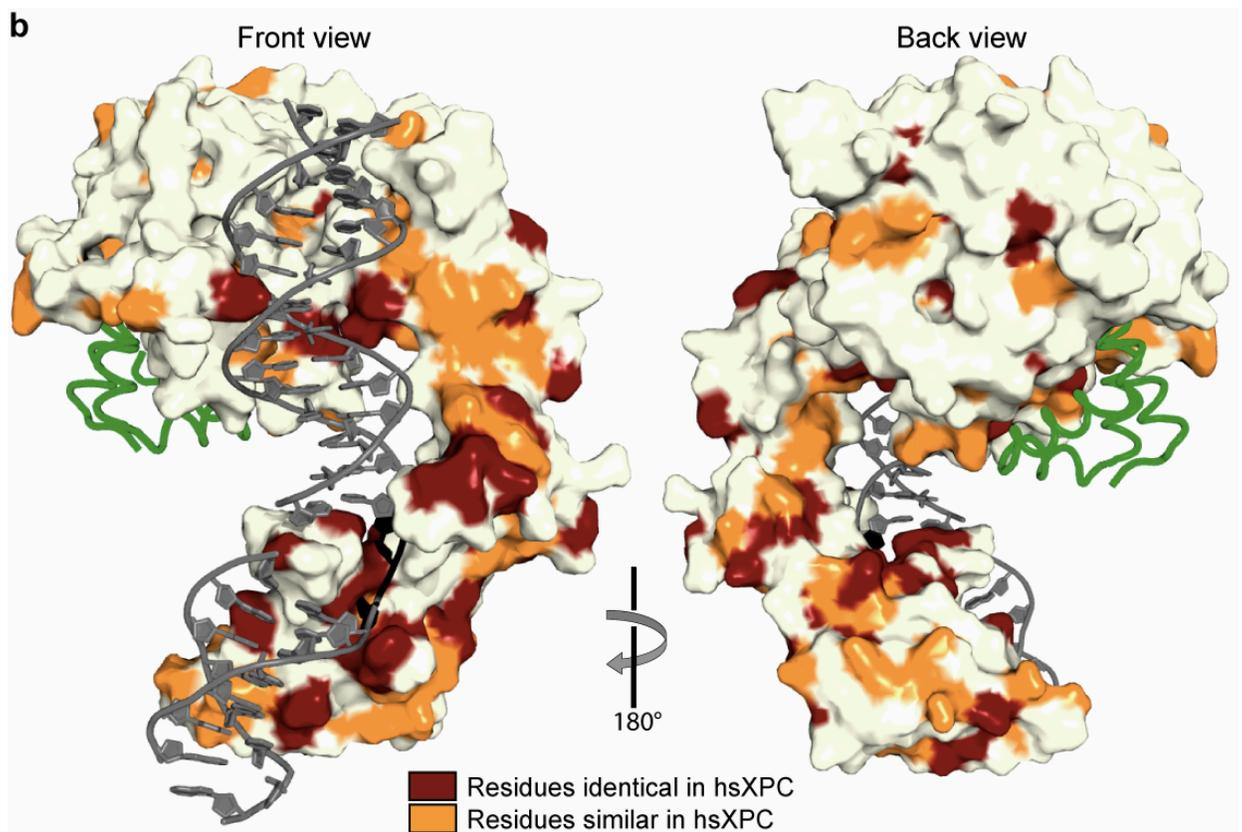


Figure 7. Homology between Rad4 and hsXPC depicted on a molecular surface representation of Rad4 in two orientations. Red indicates identical residues, and orange indicates similar residues. DNA and Rad23's R4BD are shown in cartoon representations colored in gray and green, respectively. Front view is rotated 90° about the vertical axis from Fig.1a and the back view is rotated 180° from the front view. Figure is from (14).

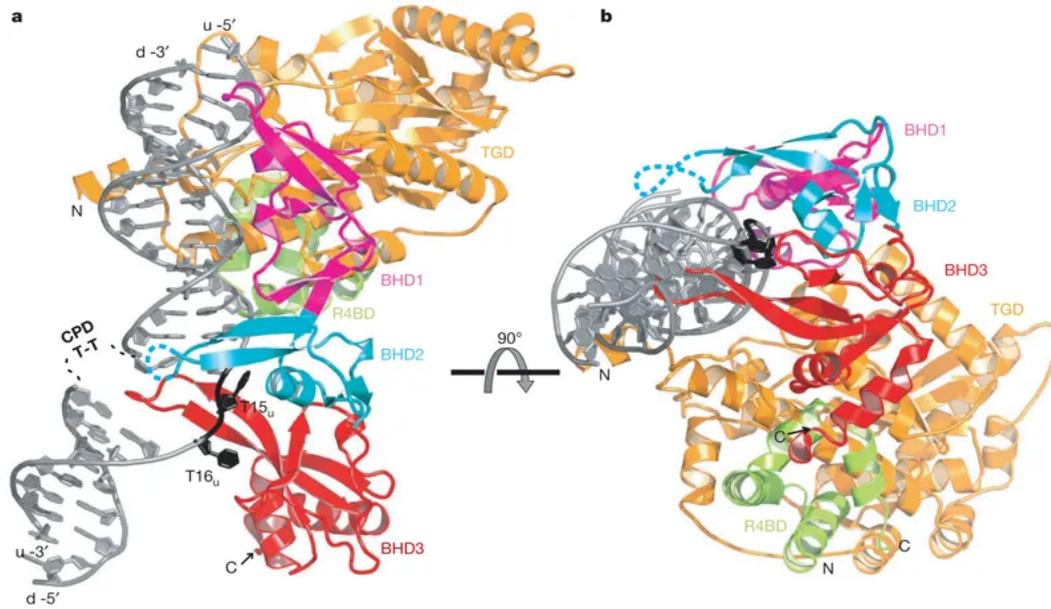


Figure 8. Overall structures of the Rad4–Rad23–DNA complex. **(a)** Ribbon diagram of the Rad4–Rad23–DNA complex: TGD, gold; BHD1, magenta; BHD2, cyan; BHD3, red; R4BD, green; and DNA, grey. **(b)** Ribbon diagram of the Rad4–Rad23–DNA complex. Figure is modified from (14).

One key domain of XPC/Rad4 that enables DNA binding is the transglutaminase-like domain (TGD). The majority of enzymes in the transglutaminase superfamily contain a catalytic triad comprised of a histidine, a cysteine, and an aspartate (15). RAD4 and XPC proteins are proposed to be an inactive homolog of this superfamily. In both Rad4 and XPC, TGD lacks the catalytic triad and is predicted to only possess the protein interaction function in DNA repair (15). The transglutaminase fold in yeast and the remnant transglutaminase sequence in XPC also share a strong resemblance (15).

Beta-hairpin domains (BHD) are also essential. Rad4/XPC use three BHDs to form and stabilize the recognition complex. Structural studies show that the beta-hairpin domains, especially BHD3, are critical for Rad4/XPC to recognize DNA lesions (16).

1.7 In vitro DNA binding of the Rad4-Rad23 complex

The crystal structure of yeast Rad4-Rad23 bound to duplex DNA shows that Rad4 uses the four core domains mentioned previously to bind to damaged DNA (**Figure 8**) (14).

Rad4 binds to DNA in two parts. In the first part, the transglutaminase-like domain (TGD) and a β -hairpin domain (BHD1) nonspecifically bind to an 11-base pair segment on the 5' side of the lesion (14).

In the second part, two β -hairpin domains (BHD2 and BHD3) recognize the damaged site and bind to a 4-nucleotide segment containing the lesion (14). The BHD2 and BHD3 domains continuously hold the nucleotides in a hand-like structure, which is an interaction with two central features (14). One feature is the insertion of a long β -hairpin finger protruding from BHD3 into the DNA, causing the CPD lesion to flip out of the double helix completely (14). The other feature is the BHD2-BHD3 binding groove that binds to the undamaged, flipped-out nucleotides (14). There are thus no direct contacts to the lesion.

CHAPTER TWO

Literature survey on disease-associated mutations in XPC

Here we compiled single-amino acid missense variants in the human XPC gene reported in three different databases: Tumor Portal (TP), International Cancer Genome Consortium (ICGC), and ClinVar (CL).

TumorPortal (TP) catalogues mutated genes in tumor samples (17). TP identifies possible cancer genes by using genomic analysis to detect somatic mutations that occur at a statistically significant rate in comparison to normal tissue samples. Somatic point mutations in each of the 18,388 genes from 4,742 cancers are shown for 21 different cancer types (17). For the *XPC* gene, there were 31 mutations reported from 16 types of tumor samples (<http://www.tumorportal.org/view?geneSymbol=xpc&sort=TumorType>).

The International Cancer Genome Consortium (ICGC) Data Portal presents genomic abnormalities in 50 types of cancer managed by collaborative institutions ((18), (19)). The data is taken from projects such as the Tumor Sequencing Project and includes somatic mutations, structural rearrangements, and germline variations. For the *XPC* gene, there were 173 mutations reported from 174 donors, of which 4 were high functional impact and 169 was low impact

(https://dcc.icgc.org/search/m?filters=%7B%22gene%22:%7B%22id%22:%7B%22is%22:%5B%22ENSG00000154767%22%5D%7D%7D,%22mutation%22:%7B%22consequenceType%22:%7B%22is%22:%5B%22missense_variant%22%5D%7D%7D%7D&mutations=%7B%22from%22:1%7D).

ClinVar (CL) is a public archive of reported relationships among human gene variants and phenotypes (20). CL organizes reports found in patient samples, assertions of their clinical significance, and other supporting evidence. The complexity of CL entries vary, from providing the presentation and interpretation of an allele to structure observation or experimental evidence about the effect of a genetic variant. For the *XPC* gene, there were 49 missense mutations reported from germline clinical testing, of which 2 were likely pathogenic, 28 were uncertain, 6 were likely benign, 9 were benign, and 3 had conflicting interpretations (<https://www.ncbi.nlm.nih.gov/ezproxy.baylor.edu/clinvar>).

The literature and databases report gene variants that are associated with or cause disease in the human *XPC* gene. The clinical significance of gene variants is classified according to (21). The classification ranges from “likely benign” to “likely pathogenic”, which is strongly expected to cause diseases such as the XP phenotype. Benign sequence variants may be single nucleotide polymorphisms (SNPs) or the result of genomic instability. **Table 1** shows the position and, if stated, symptoms and cellular effects of missense mutations found in XP-C patients in the literature.

Table 1. Missense mutations in the XPC gene compiled from Tumor Portal. The amino acid residue numbers are for human XPC gene. The types of single amino acid mutations are indicated under description. Corresponding residues in Rad4 are indicated only if the residue falls into the structured region that were observed in the the 'open' crystal structure of DNA-bound Rad4 (PDB: 2QSH).

Position	Mutation	Corresponding Rad4 Residue
6	p.Ala6Gly	
67	p.Ala67Ser	
103	p.Pro103Gln	
107	p.Lys107Asn	
121	p.Asp121Glu	
147	p.Val147Met	
224	p.Cys224Tyr	I162
270	p.Glu270Asp	G224
293	p.Arg293Gln	G260
353	p.Val353Phe	In linker region
372	p.Phe372Leu	in linker region
387	p.Lys387Thr	in linker region
393	p.Arg393Leu	in linker region
459	p.Asp459Val	in linker region
550	p.Gly550Cys	L335
578	p.Val578Phe	C372
589	p.Trp589Gln	M383
641	p.Tyr641Cys	L441
645	p.Pro645His	P445
662	p.Glu662Lys	G462
673	p.Glu673Lys	L479
715	p.Arg715Gln	K521 (doesn't appear in 6CFI structure)
743	p.Pro743Leu	P542
746	p.Ala746Asp	A545
749	p.Gly749Trp	S548
757	p.Gly757Arg	G587
848	p.Trp848Cys	
849	p.Lys849Thr	
861	p.Leu861Met	
927	p.Lys927Asn	
932	p.Ser932Pro	

Table 2. Missense genetic mutations in the XPC gene compiled from ICGC. The amino acid residue numbers are for human XPC gene. The types of single amino acid mutations are indicated under description. Corresponding residues in Rad4 are indicated only if the residue falls into the structured region that were observed in the the ‘open’ crystal structure of DNA-bound Rad4 (PDB: 2QSH).

AA position	Description	Corresponding Rad4 Residue
2	A2D	
2	A2T	
6	A6V	
16	L16V	
21	S21C	
24	K24E	
27	A27D	
34	E34G	
42	P42T	
61	S61C	
62	H62N	
63	P63L	
67	A67E	
67	A61S	
70	P70T	
85	L85I	
101	D95E	
103	P103L	
103	P97Q	
105	D105H	
106	L106F	
106	L106I	
110	H110Q	
112	L112P	
121	D115E	
132	D132A	
147	V141M	
148	R142S	
150	S150G	
152	A152V	
152	A152S	
155	R149Q	
157	L151V	
168	E162G	
169	T169K	

169	T169M	
171	E171V	
173	A167V	
175	T175K	
178	R141I	
191	R191Q	R129
193	A156V	V131
195	K195T	S133
200	G163E	K138
214	L177I	M152
215	A215S	V153
218	F218V	F156
220	R220Q	R158
224	C187Y	I162
227	P190Q	K165
235	S235P	L182
236	I236V	L183
260	W223R	L209
270	E270D	G224
293	R256Q	G260
296	E296K	D263
307	R270W	R274
307	R270Q	R274
321	P321Q	P288
332	K332N	S299
348	E348K	linker
353	V316F	linker
361	K324N	linker
368	Q368K	linker
372	F335L	linker
380	S343R	linker
393	R356Q	linker
398	S398F	linker
402	D365H	linker
406	G369E	linker
407	D370H	linker
409	Q372H	linker
415	R415Q	linker
416	R416H	linker
422	R385W	linker
423	R386W	linker

432	E432Q	linker
437	D437V	linker
439	A439V	linker
440	G440S	linker
442	G442D	linker
454	D454A	linker
455	P455S	linker
459	D422V	linker
469	K469N	linker
435	A435T	linker
472	A472V	linker
479	G479V	linker
480	S443F	linker
482	S482I	linker
482	S482R	linker
488	R488C	linker
492	R455H	linker
499	A499V	linker
500	A463P	linker
501	S464A	linker
505	S505L	linker
508	K508T	linker
516	D479V	linker
518	E518D	linker
557	T557N	K346
562	A525T	G351
565	P565T	C354
567	T567N	R361
575	D538E	K369
580	D580Y	D374
589	W552S	M383
594	R557H	T393
605	E605K	K404
612	S612I	R412
615	M615T	T415
620	K620T	Y420
621	E621Q	E421
633	P633T	G434
633	P633L	G434
633	P633S	G434
646	L609M	L446

658	A621D	I458
662	E662K	G462
671	R634H	K470
671	R634C	K470
673	E673K	L479
689	T689K	Q495
691	L691P	Y497
715	R678Q	N/A
723	Q723K	N/A
729	D692V	D528
742	Q742L	I541
742	Q742E	I541
743	P743T	P542
746	A709D	A545
747	V747M	S546
748	D748G	A547
750	K750N	E550
752	P715T	T552
753	R753L	K553
757	G757W	G557
757	G757R	G557
771	C771G	C571
773	Q736K	L573
779	L779V	A579
781	R744H	K581
782	V745M	A582
783	A746D	A583
786	L749M	L586
789	D789Y	E589
793	A756T	A593
798	D798H	K598
804	S804F	V605
805	H805N	K606
806	P769L	P607
807	V807M	V608
808	T808N	L609
812	I775V	V613
815	E815K	K616
819	D782E	E620
820	V820M	A621
848	W848C	

849	K812T	
851	L814M	
864	R827G	
864	R864C	
867	P867T	
878	A878E	
884	S884F	
891	S854I	
893	Q856E	
905	P868L	
905	P868S	
908	R871Q	
927	K890N	
929	A892T	
939	Q939K	

Table 3. Single amino acid genetic variants in the XPC gene compiled from ClinVar. The amino acid residue numbers are for human XPC gene. The types of single amino acid mutations are indicated under description. Corresponding residues in Rad4 are indicated only if the residue falls into the structured region that were observed in the ‘open’ crystal structure of DNA-bound Rad4 (PDB: 2QSH).

AA position	Description	Corresponding Rad4 Residue
1	M1L	
1	M1V	
13	G13R	
16	L16V	
18	S18R	
45	K45R	
48	L48F	
68	D68V	
160	V160L	
182	I88V	
199	K199N	R137
236	I230V	L183
240	R240C	E189
246	P246S	S195
252	Y252C	D201
287	F287C	K244
291	S291C	R248
334	P334H	N301
334	P334S	N301
341	A341G	linker
393	R393W	linker
410	E410G	linker
423	R423Q	linker
437	D244G	linker
481	K481N	linker
490	S490N	linker
492	R492H	linker
499	A499V	linker
513	M513I	linker
513	M513I	linker
539	E539G	S324
594	R594C	T393
608	R608K	T407
614	F614S	R414
641	Y641H	L441

662	E662K	G462
671	R671C	K470
674	A481V	K480
689	T689M	Q495
713	R713H	N/A
721	E721K	N/A
802	G802S	G602
839	K839E	
846	G846A	
864	R864H	
874	P874L	
878	A878G	
928	K928Q	
939	Q939K	

Table 4. Missense genetic mutations in the XPC gene compiled from the literature. The amino acid residue numbers are for human XPC gene. The types of single amino acid mutations are indicated under description. Corresponding residues in Rad4 are indicated only if the residue falls into the structured region that were observed in the the ‘open’ crystal structure of DNA-bound Rad4 (PDB: 2QSH).

Position	Amino Acid Change	Clinical Symptoms/ Notes	Reference
48	p.Leu48Phe	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
61	p.Gly61Ser	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
86	p.Lys86Arg	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
302	p.Phe302Ser	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(23)
314	p.Arg314Gln	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
334	p.Pro334His	Pathogenic variant. Protein has active <i>in vitro</i> NER ability but alters TFIIH interaction and delays XPA recruitment.	(22,24–26)
393	p.Arg393Trp	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(23)
492	p.Arg492His	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
499	p.Ala499Val	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
511	p.Lys511Gln	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
513	p.Met513Ile	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
632	p.Gln632Glu	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
671	p.Arg671Gln	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
689	p.Thr689Met	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer	(22)
690	p.Trp690Ser	Pathogenic variant. Turkish boy with multiple cancers died at 10 years of age. Also had p.R579stop mutation. This mutation drastically lowers <i>in vivo</i> stability, lowers DNA binding affinity for both undamaged and damaged DNA, and fails to recruit NER transcription factors IIH and XPA in a stable complex	(22,27,28)

703	c.2108 C>T, P703L, in exon 11	83 year old French woman with multiple melanomas, severe DNA repair deficiency in one of the oldest living patients with XP reported worldwide. Unusual long-term survival may be due to avoidance of sun exposure. Also had a frameshift mutation in exon 15 (c.2544delG, p.W848X).	(29)
738	c.2212A>G (p.Thr738Ala) missense	Healthy Caucasian male diagnosed with multiple primary melanomas. Also had c.2287delC (p.Leu763Cysfs*4) frameshift. p.Thr738Ala missense mutation may enable partial functionality that causes unusual late onset XP	(30)
928	p.Lys928Gln	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer.	(22)
939	A to C change at the first base in AAG. p.Lys939Gln	Patient XP8BE, diagnosed with XP at birth, rigorously avoided sunlight and had no malignant neoplasms as of 13 years of age. Cell line cells showed near-normal level of XPCC mRNA. Also had insVal after Val580.	(26)
939	p.Lys939Gln	Polymorphism associated with DNA repair deficits that may cause an increased risk of skin cancer. Present in 10% of population.	(22)

CHAPTER THREE

Literature survey on in-vitro functionally important mutations in XPC

In this Chapter, we have compiled human XPC gene mutations and variants that have been studied and reported before with various cellular, molecular and biochemical approaches. The mutations and their effects are listed in Table 5.

Table 5. Functionally important XPC amino acid changes caused by *in vitro* mutagenesis. The amino acid residue numbers are for human XPC gene. The types of single amino acid mutations are indicated under description. Corresponding residues in Rad4 are indicated only if the residue falls into the structured region that were observed in the the ‘open’ crystal structure of DNA-bound Rad4 (PDB: 2QSH).

Position	Mutation	Biochemical Effect	Reference
531	W→A	Slightly diminishes repair activity and slightly impairs DNA binding.	(31)
542	W → A	Slightly diminishes repair activity and slightly impairs DNA binding.	(31)
690	W→S	Diminishes repair activity and impairs DNA binding.	(32)
733	F → A	Diminishes repair activity and impairs DNA binding.	(32)
754	N → A	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; reduces localization at DNA damaged foci.	(33)
755	E → K	Reduces nuclear mobility and impairs repair activity	(31)
756	F → A	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; reduces localization at DNA damaged foci.	(33)
797	F → A	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; reduces localization at DNA damaged foci; decreases recruitment of TFIIH complex to lesion sites.	(33)
799	F → A	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; greatly reduces localization at DNA damaged foci; decreases recruitment of TFIIH complex to lesion sites.	(33)
848	W → A	Reduces NER activity and abolishes interaction with CETN2; when associated with A-851 and A-855.	(34)
851	L → A	Reduces NER activity and abolishes interaction with CETN2; when associated with A-848 and A-855.	(34)
855	L → A	Reduces NER activity and abolishes interaction with CETN2; when associated with A-848 and A-851.	(34)

CHAPTER FOUR

Homology Model of XPC

Since a high resolution crystal image of human XPC has not been solved, we have constructed a homology model using the procedure published by Petrovic et al (35). The structure was obtained by homology model software, SWISS-MODEL (36–41) using the existing crystal structure of yeast XPC ortholog, Rad4 (14). The overall sequence identity between yeast Rad4 and human XPC was 24%. Sequence alignment and the homology modeling results are below. 82% of all modeled residues were Ramachandran favored; 8.0% were Ramachandran outliers. The clash score was 4.27 and MolProbity score was 2.08 (**Figure 9**). Therefore, the model is not geometrically optimized but was deemed of sufficient quality to indicate the positions of the residues and consider the functional impacts of the mutations.

Model Building Report

This document lists the results for the homology modelling project "homology model" submitted to SWISS-MODEL workspace on April 15, 2020, 8:24 p.m.. The submitted primary amino acid sequence is given in Table T1.

If you use any results in your research, please cite the relevant publications:

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46(W1), W296-W303 (2018). [doi>](#)
- Guex, N., Peitsch, M.C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 30, S162-S173 (2009). [doi>](#)
- Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res.* 45, D313-D319 (2017). [doi>](#)
- Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, G., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. *Bioinformatics* 36, 1765-1771 (2020). [doi>](#)
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* 7 (2017). [doi>](#)

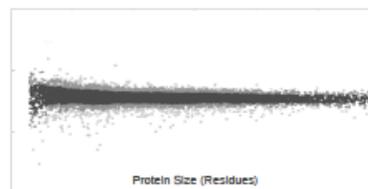
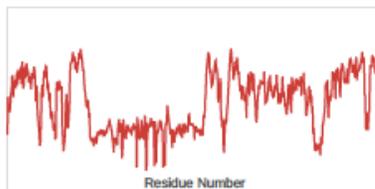
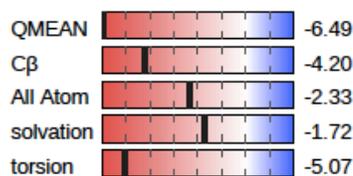
Results

The user uploaded a template structure to use for the modelling process.

Models

The following model was built (see Materials and Methods "Model Building"):

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN
	PDB	ProMod3 3.0.0	monomer	None	0.44	-6.49



Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
template_upload.1.A	24.19	monomer	0.00	HHblits	Unknown	-	0.33	1 - 637	0.68	Polypeptide

The template contained no ligands.

Target template_upload.1.A	RRAMKRFNKGVHEDTHKVVHLLCLLANGFYRNNICSQPDLHAIGLSIIPARFTRV- RNVCSNEERKRRKYFHMLYLVLCLMVHGFIRNEWINSKRLSRKLSNLVPEKVFELL
Target template_upload.1.A	LPRD---V-----DTYYLSNLVKWFIGTFTVNAEL---SA---S----- HPQKDEELPLRSTRKLLDGLKCKMELWQKHWKITKKYDNEGLYMRTWKEIEMSAN
Target template_upload.1.A	-----EQDNLQTTLERRFAIYSARDDEELVHIFLLILRALQLLTRLVLSLQP NKRKFKTLKRSDFLRAVS---KGGHDPDI--SVQGFVAMLRACNVNARLIMSCQP
Target template_upload.1.A	IPLKSATAKGGKPSKERLTADPGGSSETSSQVLENHTKPKTSKGTQEEETFAKGT PDFTNMKIDT-----S-----
Target template_upload.1.A	CRPSAKGKRNGGRKKRSKPSSEDEEGPGDKQEKATQRRPHGRERRVASRVSYK -----L-----N-----G-----N-----
Target template_upload.1.A	EESGSDEAGSGDFELSSGEASDPSEDESEPQKQKAPAPQRTKAGSKSASRT ---N-----A-----YKD-----
Target template_upload.1.A	HRGSHRKDPSLPAASSSSSSSKRGKMKCSDGEKAEKRSIAGIDQWLEVFCEQEEK -----MVKYPIFWCEVWDKFSKK
Target template_upload.1.A	WVCVDCVHG-VVGQPL---TC-----YKYATKPMYVVGIDSDGWVRDVTQRYDP WITVDPVNLKTIQVRLHSLKAPKGVACCERNMLRYVIAYDRKYGCRDVTTRYAQ
Target template_upload.1.A	VWMTVTRKCRV-----DAEWWAETLRPYQSPF-MDREKKEDLEFQAKHMDQPLPT WMNSKVRKRRITKDDFGKWFVRKVVITALHHRKRTKIDDYEDQYFFQRDESEGIPD
Target template_upload.1.A	AIGLYKNHPLYALKRHLKYEAIYPETAAILGYCRGE-----AVYSRDCVHTL SVQDLKNHPYVLEQDIKQTQIVKPGC-KECGYLKVHGKVGKVLKVYAKRDIADL
Target template_upload.1.A	HSRDTWLKKARVVRLEGEVYPYKMKVGFNSRARKARLAEPQLREENDLGLFGYWQTE KSARQWYMNGRILKTGSRCKKVIKRTV-----EEDERLYSFEDTE
Target template_upload.1.A	EYQPPVAV-DGKVPVPRNEFGNVYLFPSMMPIGCVQLNLPNLHRVARKLDIDCVQA LYIPPLASASGEITKNTFGNIEVFAPTMIPGNCCLVENPVAIKAARFLGVEFAPA
Target template_upload.1.A	ITGFDFH-GGYSHPVTDGYIVCEEFKDVLTTAWENEQ VTSFKFERGSTVKPVLVGIWAKWLRREAIETAIDG--

Materials and Methods

User Template Alignment

The user entered their own target sequence together with an uploaded a template structure file in PDB format.

Model Building

Models are built based on the target-template alignment using ProMod3. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodelled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modelling with ProMod3 fails, an alternative model is built with PROMOD-II ([Guex et al.](#)).

Model Quality Estimation

The global and per-residue model quality has been assessed using the QMEAN scoring function (Studer et al.).

Ligand Modelling

Ligands present in the template structure are transferred by homology to the model when the following criteria are met: (a) The ligands are annotated as biologically relevant in the template library, (b) the ligand is in contact with the model, (c) the ligand is not clashing with the protein, (d) the residues in contact with the ligand are conserved between the target and the template. If any of these four criteria is not satisfied, a certain ligand will not be included in the model. The model summary includes information on why and which ligand has not been included.

Oligomeric State Conservation

The quaternary structure annotation of the template is used to model the target sequence in its oligomeric form. The method (Bertoni et al.) is based on a supervised machine learning algorithm, Support Vector Machines (SVM), which combines interface conservation, structural clustering, and other template features to provide a quaternary structure quality estimate (QSQE). The QSQE score is a number between 0 and 1, reflecting the expected accuracy of the interchain contacts for a model built based a given alignment and template. Higher numbers indicate higher reliability. This complements the GMQE score which estimates the accuracy of the tertiary structure of the resulting model.

References

- **BLAST**
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421-430 (2009). [doi](#)
- **HHblits**
Remmert, M., Biegert, A., Hauser, A., Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 9, 173-175 (2012). [doi](#)

Table T1:

Primary amino acid sequence for which templates were searched and models were built.

```
RRAMKRFNGKGVHEDTHKVLHLLCLLANGFYRNINCSQPDLHAIGLSIIPARFTRVLPDVTYYLSNLVKWFIGTFTVNAELSAEQDNLQTTLERRFAIY  
SARDEELVHIFLLI LRALQLLTRLVLSLQPIPLKSATAKGGKPSKERLTADPGSSSETSSQVLENHTKPKTSKGTQKQETFAKGTCPKSAKGRNKGGR  
KKRSKPSSSEEDGPGDKQEKATQRRPHGRERRVASRVSYKEESGSDAEGSGDFELSSGEASDPSDEDSEPGPPKQRKAPAPQRTKAGSKSASRTHRGS  
HRKDPSLPAASSSSSSSKRGKMCSDGEKAEKRSIAGIDQWLEVFCEQEEKWVDCVHGVVGGPLTCYKYATKPMTYVVGIDSDGWVRDVTQRYDPVVM  
TVTRKCRVDAEWAETLRPYQSPFMDREKEDLEFQAKHMDQPLPTAIGLYKNHPLYALKRHLLKYEAITYPETAAILGYCRGEAVYSRDCVHTLHSRDTW  
LKKARVRLGEVPYKMKVGFNSRARKARLAEPQLREENDLGLFGYWQTEEQPPVAVDGGKVPVPRNEFGNVYLFPSMMPICGVQLNLPNLHRVARKLDIDC  
VQAITGDFHGGYSHPVTDGYIVCEEKDVLLTAWENEQ
```

Table T2:

Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Coverage	Description
template_upload.1.A	24.19	monomer	-	HHblits	Unknown	NA	0.33	0.68	Polypeptide
template_upload.1.A	29.35	monomer	-	BLAST	Unknown	NA	0.35	0.46	Polypeptide

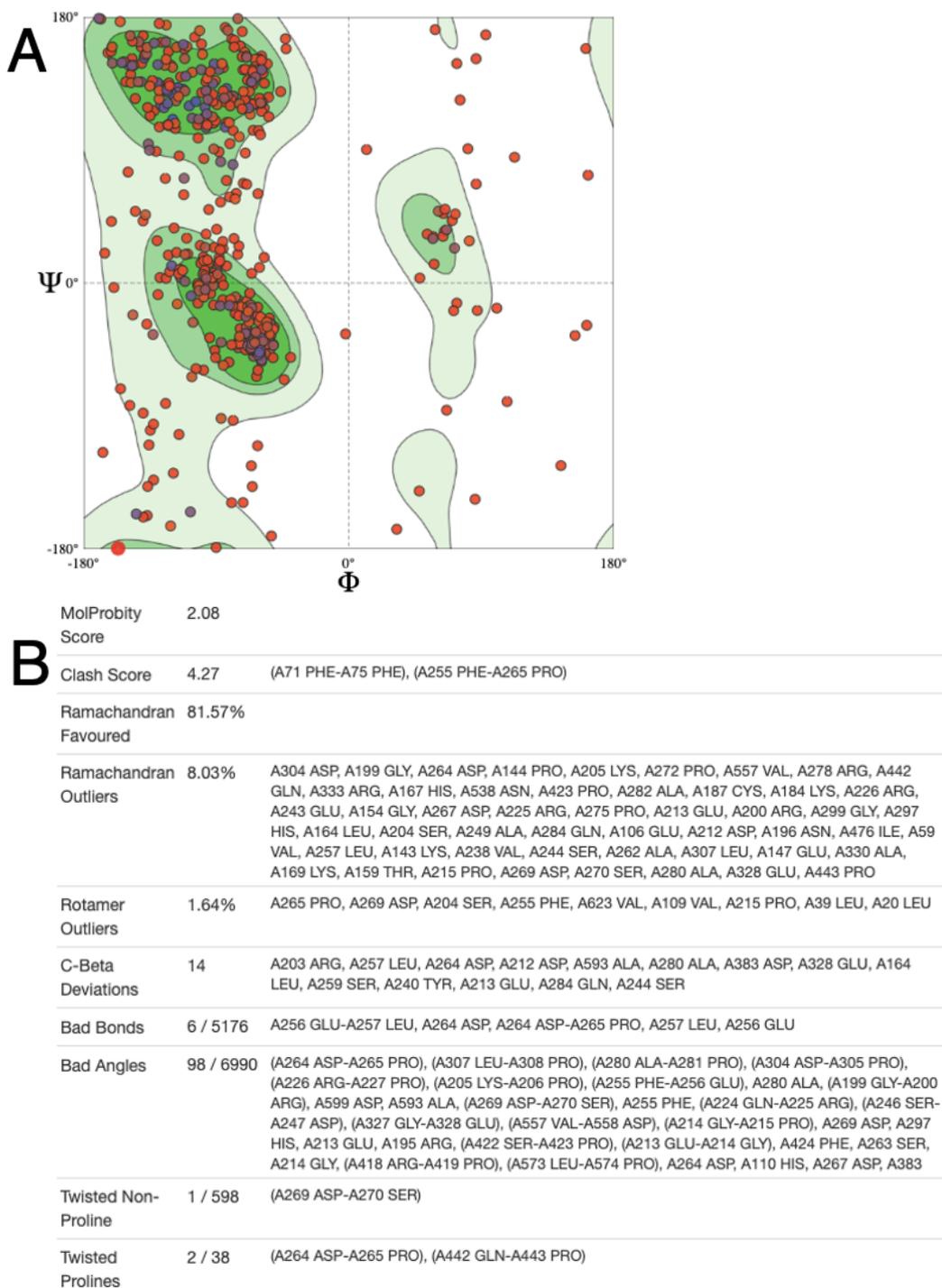


Figure 9. Ramachandran plot and MolProbity results of SWISS-MODEL Homology Model. (a) the Ramachandran plot of the human XPC homology model. **(b)** The MolProbity results.

CHAPTER FIVE

Structural Analyses of XPC Genetic Variations

Here, we have mapped the XPC mutations (compiled in Chapters 2-3) onto the 3-D homology model of human XPC protein constructed (Chapter 4). The total of 288 mutations were spread over the four domains of the protein: 55 were in the TGD domain, 11 in BHD1, 12 in BHD2, 38 in BHD3 and 64 were in the unstructured loop region within TGD (residues 331-517). To aid visual inspections and characterizations, we have grouped the mutations into 20 clusters. The clusters are made up of mutated residues that are physically close to one another within the protein structure; the proximity may indicate functional connection among the residues in the cluster. **Figures 10-11** shows the approximate locations of each cluster in the context of the overall structure of XPC. **Figures 12-16** show close-up views of each cluster along with the corresponding lists of mutations. Using the structural models as well as the sequence conservation among various orthologs (**Figure 6**), we discuss the putative roles of each residue and the impacts of their mutations below.

Cluster 1 is located around alpha1 and alpha2, and contains R191Q, A193V, K195T, K199N and G200E (**Figure 12**). **R191** in XPC corresponds to R129 in Rad4 which is situated in the minor groove of the DNA; its positively charged guanidino group forms ionic interactions with the negatively charged phosphate backbone. The mutation R191Q likely weakens this interaction as the positive charge is absent in Gln. **A193** corresponds to Rad4 V131 which also directly contact with the minor groove of the

DNA. As A193V is likely to preserve (if not strengthen) this interaction, it is not clear if the mutation would impact XPC function by altering its interaction with DNA based on this structural model. **K195** corresponds to Rad4 S133 which is at the junction before the alpha2 starts. While it is in proximity with a phosphate group of DNA, it does not make contact using its side chain and most of the orthologs contain a polar residue albeit various identities (Lys, Ser, Asn and Thr) for this position (**Figure 6**). Therefore, K195T also may not have significant structural impact although it is conceivable that it may weaken the interaction with the DNA in case K195 can interact with DNA phosphates in the human XPC structural contexts. **\$\$\$K199** corresponds to Rad4 R137 and belongs to alpha 2; its side chain is flanked between two phosphates groups of the two DNA strands resulting in narrowing of the minor groove. Notably, the amino acid type at residue 199 is predominantly Lys/Arg indicating its functional importance in XPC. It is conceivable that **K199N** mutation may have significant impact on the interaction of the protein with DNA due to the loss of the positive charge in the residue. **G200** corresponds to Rad4 K138 near the DNA backbone and belongs to alpha 2. The residue type at position 200 can be varied (G/Glu/Asp/Val) in other organisms. It is not clear what impact G200E may have on the protein function.

Cluster 2 is located near the alpha2 helix and includes L214I, A215S, F218V, V578F, and D580Y. **L214** (Rad4 M152) is found in alpha2 and forms hydrophobic interaction with W603 (F402), T606 (V405) and Y610 (L409) in alpha13 and V570 (I364) of beta7 (numbers in () indicate residue number in Rad4). This hydrophobic interface is largely conserved among different orthologs: L214 is represented as M/L/I/T in other orthologs; W603 as F/W/L/V/Y/M; T606 as T/V; Y610 as L/P/F/I; V570 as

I/V/F in other orthologs. In Rad4, V405 and L409 also form van der Waals (vdw) contacts with residues in Rad23's R4BD, indicating that T606 and Y610 in human XPC may also participate in inter-molecular interactions. As L214I variant would largely retain the hydrophobic nature and size of the residue side chain, it is not clear how it would impact the protein function.

A215 (V153) is conserved as Ala or Val in most orthologs; In Rad4, V153 makes hydrophobic contact with V302, L305, L306 of R4BD (**L328**xxxx in human RAD23B). A serine in this position (A215S) may disrupt this hydrophobic interface.

F218 (F156) is variably represented as F/L/K/R/Q/C/H/Y/S in various orthologs and is positioned in the interface between alpha xx of TGD, beta.. of TGD and xx of R4BD of Rad23 (in particular L306 (**N329** in human RAD23B)). The benzyl group is half exposed to solvent if not bound to Rad23 indicating that it can in principle accommodate polar/charged residues, as indicated by low conservation. The impact of F218V on the XPC structure is thus not entirely clear.

V578 (C372) is represented as C/V/L/I/A in orthologs and is in beta8 to form a hydrophobic interface between alpha 2/13???, beta 8.. of TGD. The contacting residues include V570(xxx), I572 (Y366; Y/I/F), Y610 (L409, L/P/F/I) as well as L214 (xxx). While it is possible that a phenylalanine (**V578F**) may be too large to fit in, it is unclear what impact it may have as the hydrophobic contacts would be largely preserved.

D580 is conserved as D374 in Rad4. The carboxy group of D580 forms a hydrogen bond with the backbone amide of Thr582 and the NH indole of Trp603 in the alpha13 helix. D580, V581, and T582 are conserved in most orthologs, while W603 is not conserved. The D374 in Rad4 also forms a salt bridge with R391 (R597???? in

human) where it is partially embedded in the protein core. The R597 in the human XPC homology model was positioned differently from R391 of Rad4, but it is likely that the salt bridge relationship as seen in Rad4 may exist in human XPC. Taken together, the mutation \$\$\$D580Y is likely to have a distinct impact and destabilize the integrity of the TGD domain as Tyr will not be able to preserve these interactions formed by an Asp.

Cluster 3 is located near the end of the alpha2 helix and the beginning of the alpha3 helix, and contains R220Q, C224Y, and P227Q. **R220** is in alpha2 and is conserved in Rad4 as R158, which contacts R4BD. This position is conserved in nearly all orthologs. Its positively charged guanidino group is close to the negatively charged E297 (Rad4 D263 or I264?). Rad4's R158 hydrogen-bonds with the carbonyl groups of Rad23's Ser270 backbone (human Q338) as well as of Rad4 Gln267 side chain. Gln267 is conserved as Gln in most orthologs. **It is conceivable that R220Q lacks the positive charge and may weaken this interaction.** **C224** (I162 in Rad4) in alpha2 is conserved as Cys in most orthologs. The side chain of **C224** packs against H300, L303 and L304 in alpha8. **L303 (Rad4 A271) is conserved as Ala in most orthologs, while H300 (Rad4 G268) and L304 (Rad4 M272) are not conserved.** \$\$\$C224Y may be too bulky and destabilize the packing between these alpha helices within TGD. **P227** (Rad4 K165) is positioned near the beginning of alpha3 and is exposed to solvent. This position is not conserved in orthologs (Pro, Ser, Glu, Asn). The impact of P227Q is unclear.

Cluster 4 is located where the alpha4 and 5 helices interact, and contains S235P, I236V, and W260R. **S235** (Rad4 L182) is in alpha4 and is conserved as Ser in most orthologs. It is in contact with solvent. S235P may destabilize alpha4. I235V (Rad4 L183) is conserved as Leu in most orthologs.

W260 is electrostatically stabilized by I236, bringing alpha4 and 5 to close proximity. S235 is between L and I in alpha4; its mutation S235L may be too bulky and nonpolar to preserve the alpha helix.

Cluster 5 is located near the N terminus of alpha5, and contains R240C, Y252C, and P246S. the OH of Y252 is electrostatically attracts the N's of R240 and N256. P246 is located in a loop near alpha5 and its side ring is stabilized by V240 and R243.

Cluster 6 spans alpha6, alpha7, beta2, and alpha8. It contains E270D, F287C, S291C, R293Q, and E296K. E296, D529, and D516 (Rad4 D263, I314, D331) correspond to the Cys–His–Asp catalytic triad characteristic of the transglutaminase superfamily, and have structure-stabilizing roles in hydrophobic core (14).

Cluster 7 is located where beta3, the junction of beta4 and 5, and alpha14 contact. It contains R314Q, E539G, W542A, K620T, and E621Q. R314 is embedded in the hydrophobic core; it makes a salt bridge with E621 (Rad4 E421), and surrounding residues are conserved. It also makes a Arg-Trp cation-pi stacking between guanidino ring and W542 (W327), which has residual DNA binding activity (31). Functional studies show that W542A moderately lowers accumulation at the lesion and DNA binding (31).

Cluster 8 is located where alpha8 and beta4 and 5 contact. It contains R307W, R307Q, G550C, F302S, and W531A. F302 shows hydrophobic packing inside the protein (L, P, C); serine would be too polar to be accommodated. W531 is located within beta4; its side chain NH attracts the OH of Y585; a mutation to A would be unable to interact with Y585. Functional studies show that W531A moderately lowers accumulation at the lesion and DNA binding (31).

Cluster 9 is located in the junction between beta3 and alpha9, and contains K332N, P334H, and P334S. Functional studies show p.Pro334His mutation prevents the stimulation of Ogg1 glycosylase because it stops the interaction between XPC and Ogg1 (20).

Cluster 10 is located around the short alpha10 helix, and contains P321Q, T557N, A562T, P565T, and T567N. A562 stabilizes the helical conformation of alpha10 by orienting more polar neighbors Y561 and T563 towards solvent. The OH of T557 H-bonds to the carbonyl of Q554. The N of P565 is attracted to the O of Q320. P321 packs with the alpha2 helix.

Cluster 11 is located where the junction of beta7 and 8, the junction of alpha 13 and 14, and alpha 13 meet. It contains D575E, E605K, R608K, S612I, F614S, and M615T.

Cluster 12 is located where alpha12, alpha14, and the junction of alpha14 and alpha0 meet. It contains W589Q, R594H, Q632E, P633T, P633L, and P633S. Q632 is final residue on alpha14, and its NH2 interacts with the ketone on Q583. The mutation Q632E may cause the carboxylic acid group of E to repel Q583 instead.

Cluster 13 is located around alpha0, beta1, beta2, and beta3. It contains Y641C, Y641H, P645H, L646M, R671Q, R671H, R671C, E673K, and A674V.

Cluster 15 is located around alpha1, beta2, beta3, and beta3'. It contains T689M, T689K, W690S, L691P, P703L, F733A, and T738A. T689, K692, and D688 could form van der Waals contacts with DNA. T689 is predicted to form phosphate DNA contact like Rad4 Q495 (14). Also, the OH of T738 H-bonds with F534's backbone C=O to pack the alpha2 helix on the beta3 strand. P703 is located immediately before beta2; its

pyrrolidine ring prevents hydrogen bonding between NH and CO to preserve edge of beta sheet. In a P703L mutation, leucine is too hydrophobic and allows H-bonding. Functional studies indicate F733 and W690 cooperate to detect and bind single stranded DNA on the undamaged side of the DNA duplex with high affinity (32). The W690S mutation is shown to drastically lower *in vivo* stability, lower DNA binding affinity for both undamaged and damaged DNA, and to recruit TFIIH and XPA in an unstable, short-lasting complex (22,27,28).

Cluster 16 is located where the junction between beta3' and beta0 contacts the junction between alpha1 and beta2. It includes A746D, V747M, D748G, G749W, K750N, P752T, R753L, N754A, E755K, F756A, G757W, G757R, and D789Y.

Cluster 17 is located near the junction of beta3' and beta0 and the junction of alpha1 and beta2. It includes A746D, V747M, D748G, G749W, K750N, P752T, R753L, N754A, E755K, F756A, G757W, G757R, and D789Y. N754 and F756 are identical to Rad4 N554 and F556, respectively, which contact a flipped-out nucleotide (14,33). Conserved in higher eukaryotes, E755 is within a β -turn motif and has DNA-repellent properties; it is suggested to provide the necessary mobility for dynamic lesion recognition (31). N754 is important to mobility and the ability to accumulate at the lesion (33).

Cluster 18 is located near the junction of beta3' and beta0, beta1, beta 3, and alpha2. It includes Q742L, Q742E, P743L, P743T, C771G, Q773K, I812V, and E815K.

Cluster 19 is located in alpha1 and alpha2. It includes L779V, R781H, V782M, A783D, L786M, D819E, and V820M.

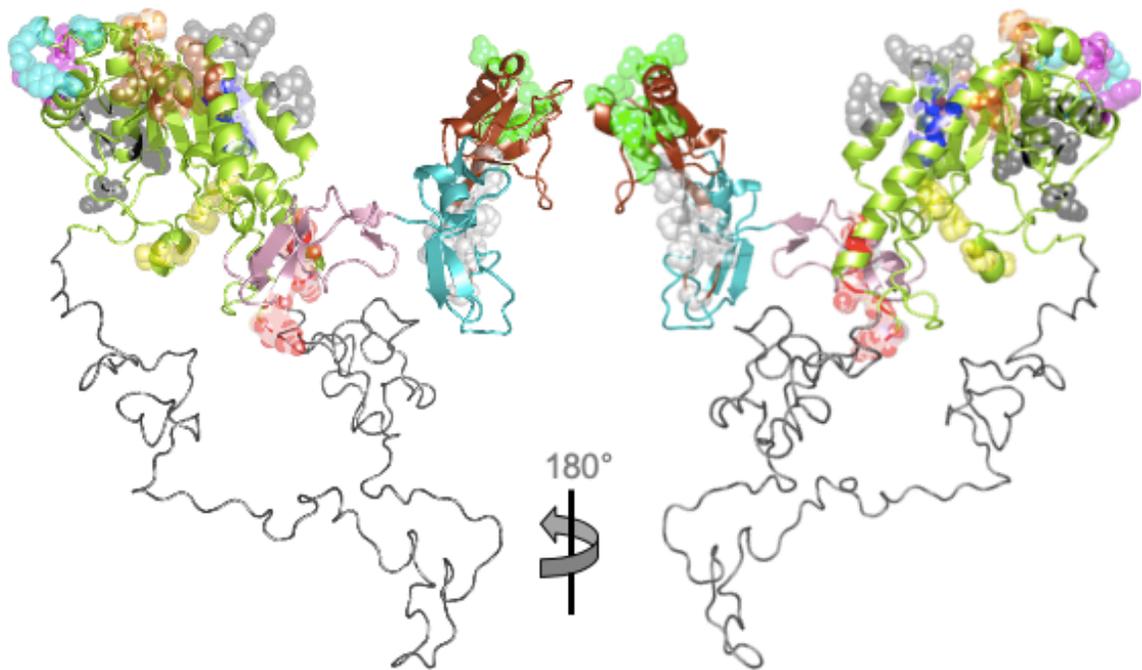
Cluster 20 is located near beta2 and beta3. It includes A793T, F797A, D798H, F799A, G802S, S804F, H805N, P806L, V807M, and T808N. F797, F799, and P806 are identical to Rad4 F597, F599, and P607, respectively, which contact flipped-out nucleotides (14). S804 is conservatively substituted with Rad4 V605, which also contacts a flipped-out nucleotide (14). *in vitro* studies suggest F797 and F799 are essential to recruit TFIIH (33).

Some mutations/variations correspond with Rad4 residues that contact DNA or the R4BD domain (14). The residues that correspond with Rad4 residues that have side-chain contact to DNA are human R191Q (yeast R129; Cluster 1), K199N (R137; Cluster x), A193V (V131), T689K (Q495), G802S (G602), S804F (V605), and P769L (P607). Those that correspond to Rad4 residues that contact R4BD are human A215S (V153), F218V (F156), R220Q (R158), and E605K (K404).

Certain mutations/variations appear repeated times in databases and the literature. C224Y is in TP and ICGC; I236V in ICGC and CL; E270D in TP and ICGC; R293Q in TP and ICGC; P334H in CL and L; W589Q in TP and ICGC; R594 in ICGC and CL; R671C in ICGC and CL; E673K in TP and ICGC; E662K in ICGC, TP, and CL; T689K in ICGC and CL; R715Q in TP and ICGC; A746 in ICGC and TP; G757R in ICGC and TP.

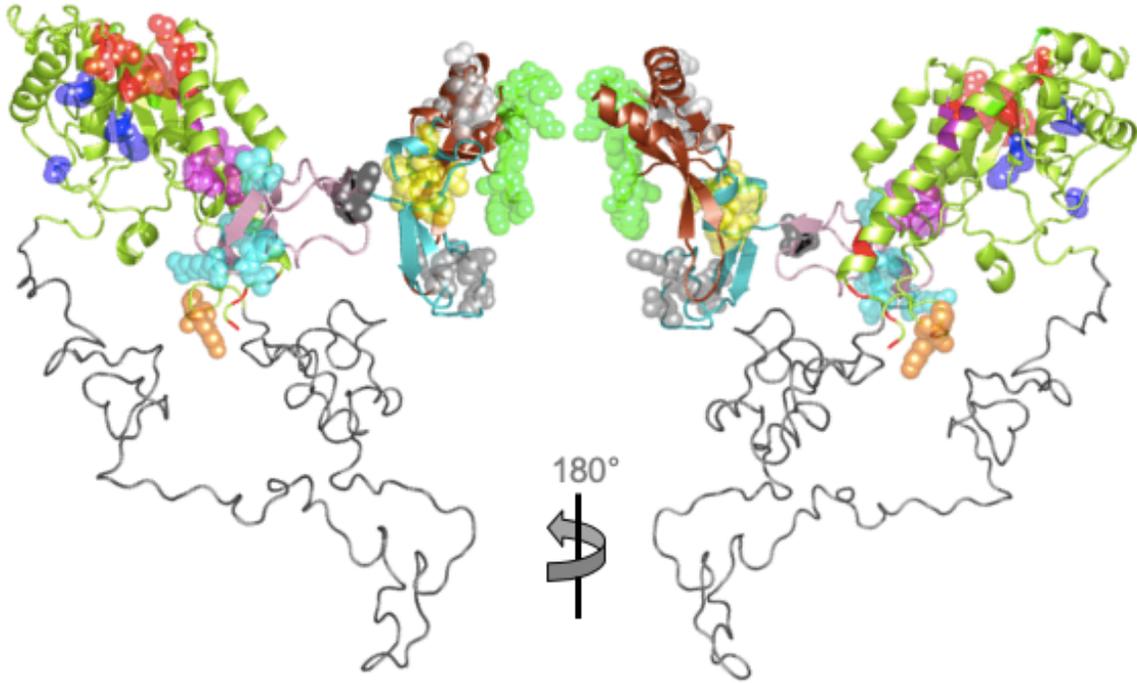
Certain residues in human XPC have multiple and different mutations/variants in databases and the literature. R307 is reported to change to W (ICGC) or Q (ICGC); P334 to H (L, CL) or S (CL); P633 to T (ICGC) or L (ICGC) or S (ICGC); Y641 to C (TP) or H (CL); R671 to Q (L) or H (ICGC) or C (ICGC, CL); T689 to M (L) or K (ICGC, CL);

G757 to W (ICGC) or R (ICGC, TP); Q742 to L (ICGC) or E (ICGC); P743 to L (TP) or T(ICGC).



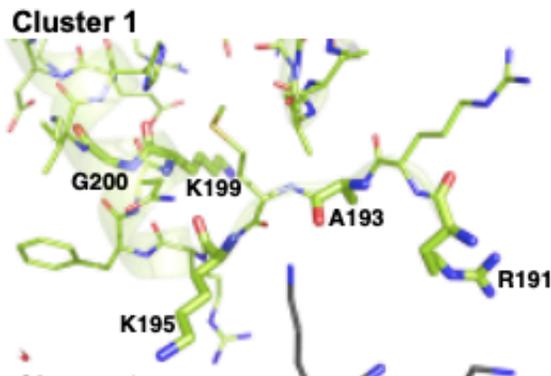
Domain	Color	Cluster
TGD	red	1
TGD	blue	2
TGD	orange	3
TGD	magenta	4
TGD	cyan	5
TGD	black	6
TGD	yellow	10
TGD	gray	11
BHD3	green	19
BHD3	white	20

Figure 10. Locations of clusters mapped on the overall XPC homology model structure. The overall structure is in ribbon representation and the residues for each cluster are shown in spheres with colors indicated as in the table below.

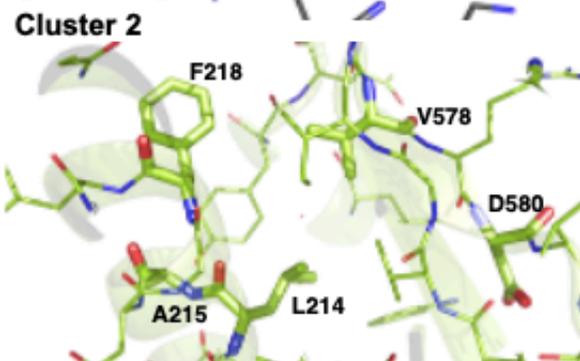


Domain	Color	Cluster
TGD	red	7
TGD	blue	8
TGD	orange	9
TGD	magenta	12
BHD1	cyan	13
BHD1	black	14
BHD2	yellow	15
BHD2	gray	16
BHD3	green	17
BHD3	white	18

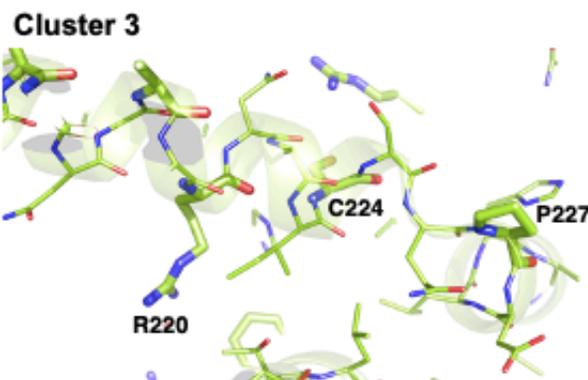
Figure 11. Locations of clusters mapped on the overall XPC homology model structure. The overall structure is in ribbon representation and the residues for each cluster are shown in spheres with colors indicated as in the table below.



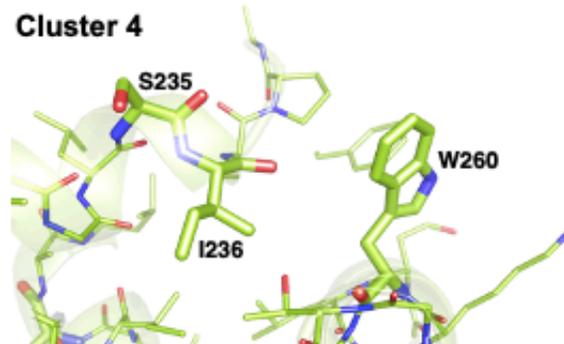
Reference	XPC Mutation	Corresponding Rad4 residue
ICGC	R191Q	R129
ICGC	A193V	V131
ICGC	K195T	S133
CL	K199N	R137
ICGC	G200E	K138



Reference	XPC Mutation	Corresponding Rad4 residue
ICGC	L214I	M152
ICGC	A215S	V153
ICGC	F218V	F156
TP	V578F	C372
ICGC	D580Y	D374



Reference	XPC Mutation	Corresponding Rad4 residue
ICGC	R220Q	R158
TP, ICGC	C224Y	I162
ICGC	P227Q	K165



Reference	XPC Mutation	Corresponding Rad4 residue
ICGC	S235P	L182
ICGC, CL	I236V	L183
ICGC	W260R	L209

Figure 12. Local structural views for Clusters 1-4. The amino acids that were indicated in the cluster are shown in stick representation and all other residues are in line representation. Green indicates carbon (in TGD), blue nitrogen, and oxygen red.

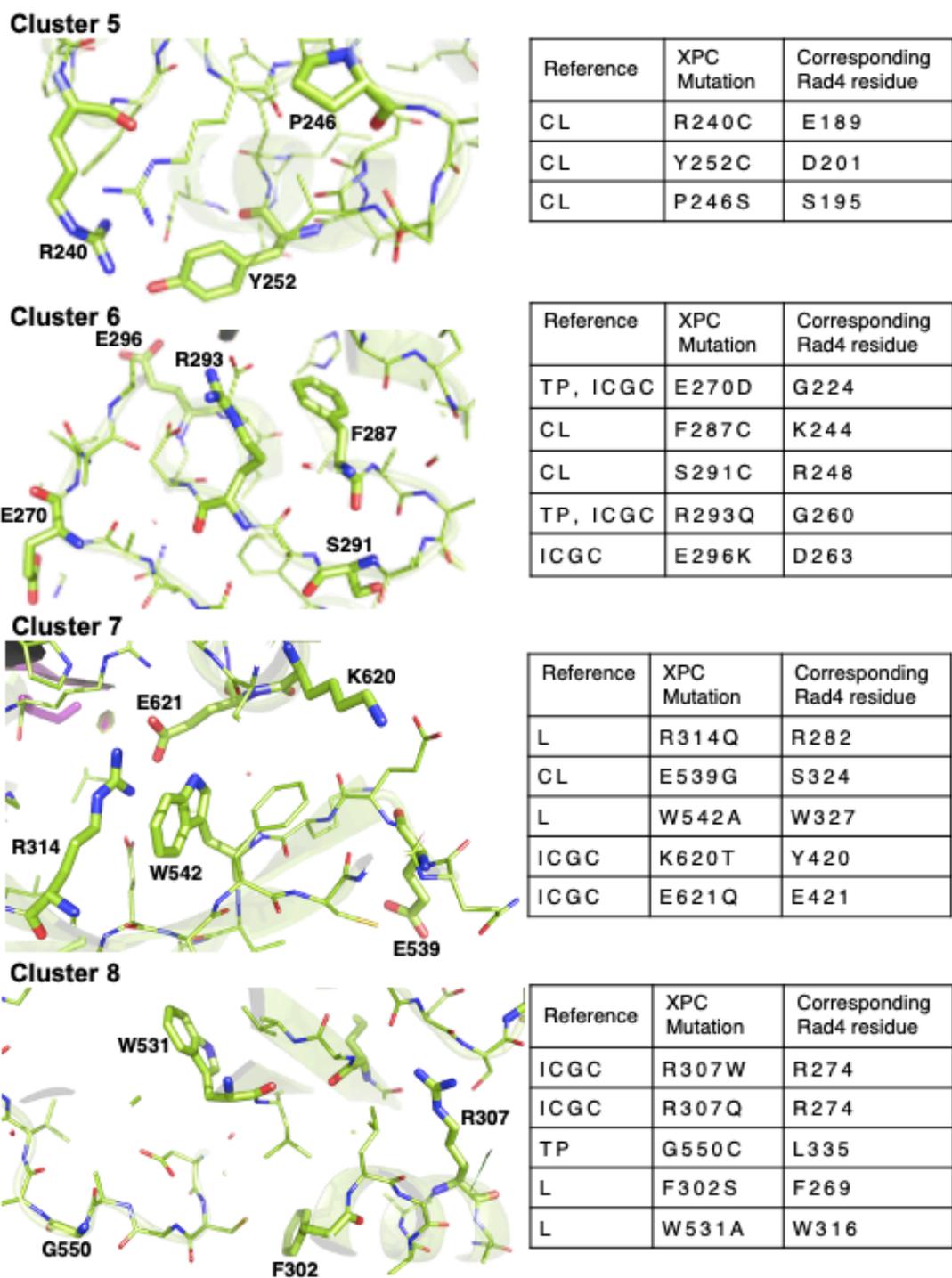


Figure 13. Local structural views for Clusters 5-8. The amino acids that were indicated in the cluster are shown in stick representation and all other residues are in line representation. Green indicates carbon (in TGD), blue nitrogen, and oxygen red.

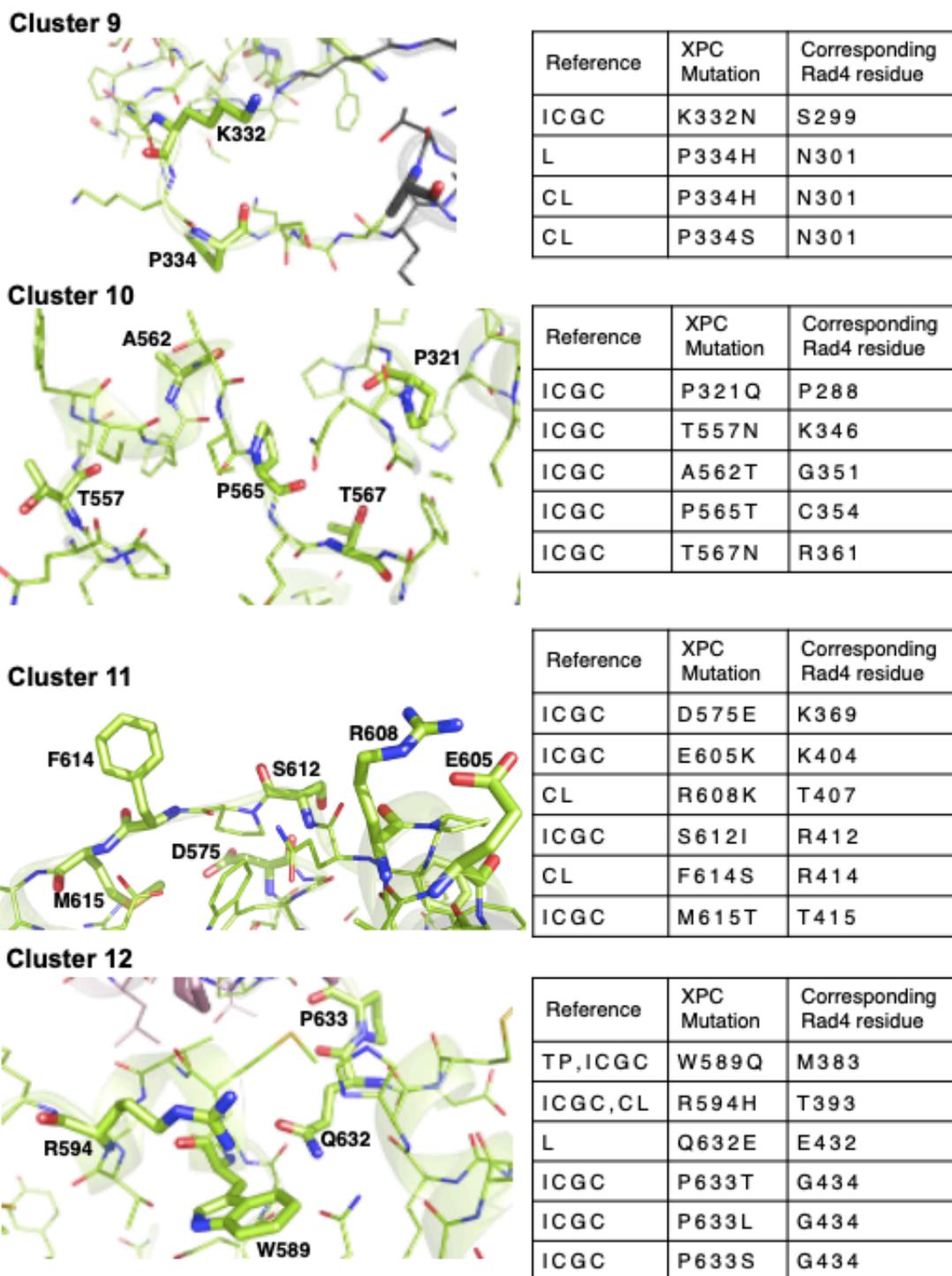


Figure 14. Local structural views for Clusters 9-12. The amino acids that were indicated in the cluster are shown in stick representation and all other residues are in line representation. Green indicates carbon (in TGD), blue nitrogen, and oxygen red.

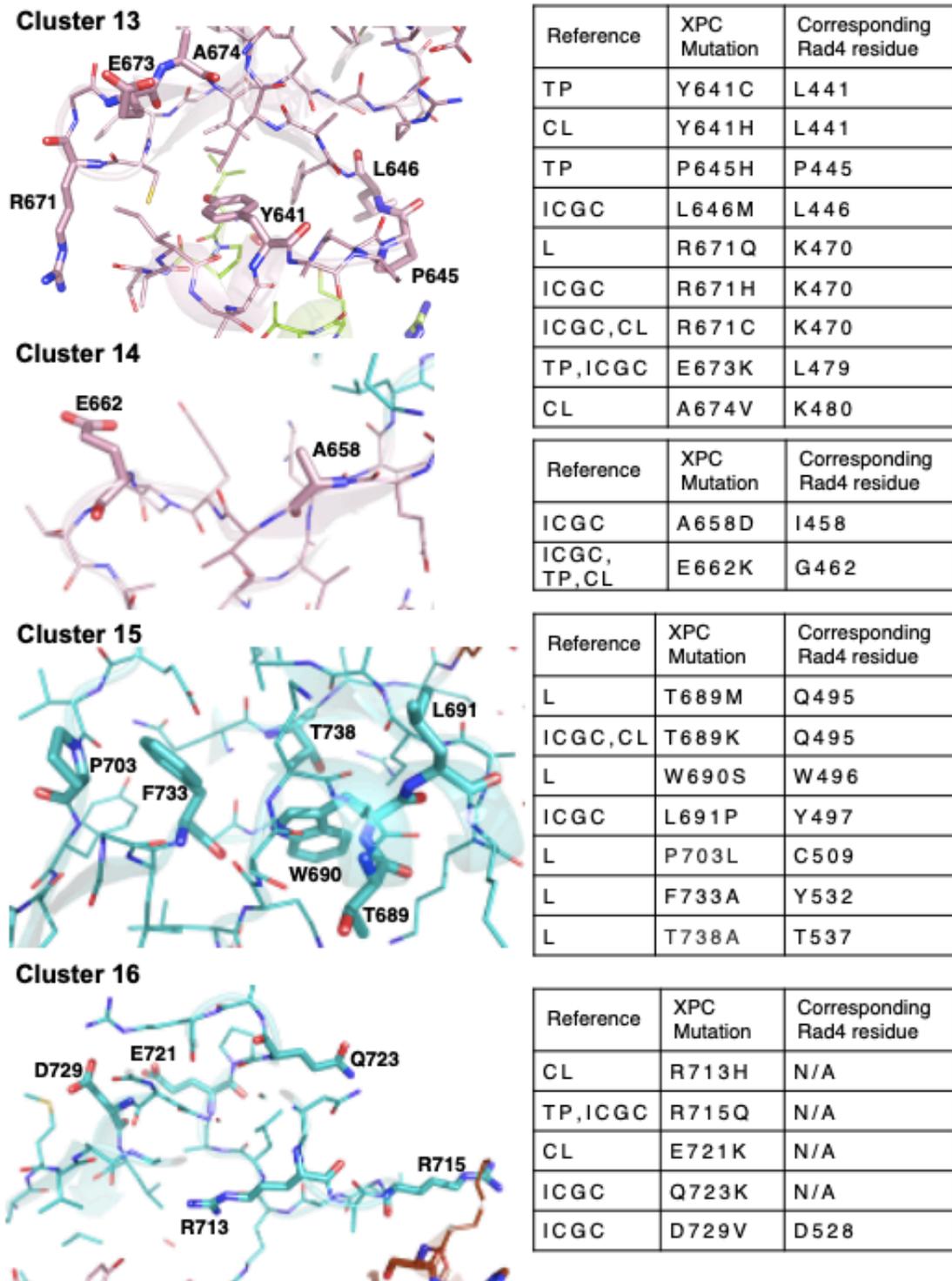


Figure 15. Local structural views for Clusters 13-16. The amino acids that were indicated in the cluster are shown in stick representation and all other residues are in line representation. Light purple indicates carbon (in BHD1), turquoise carbon (in BHD2), blue nitrogen, and oxygen red.

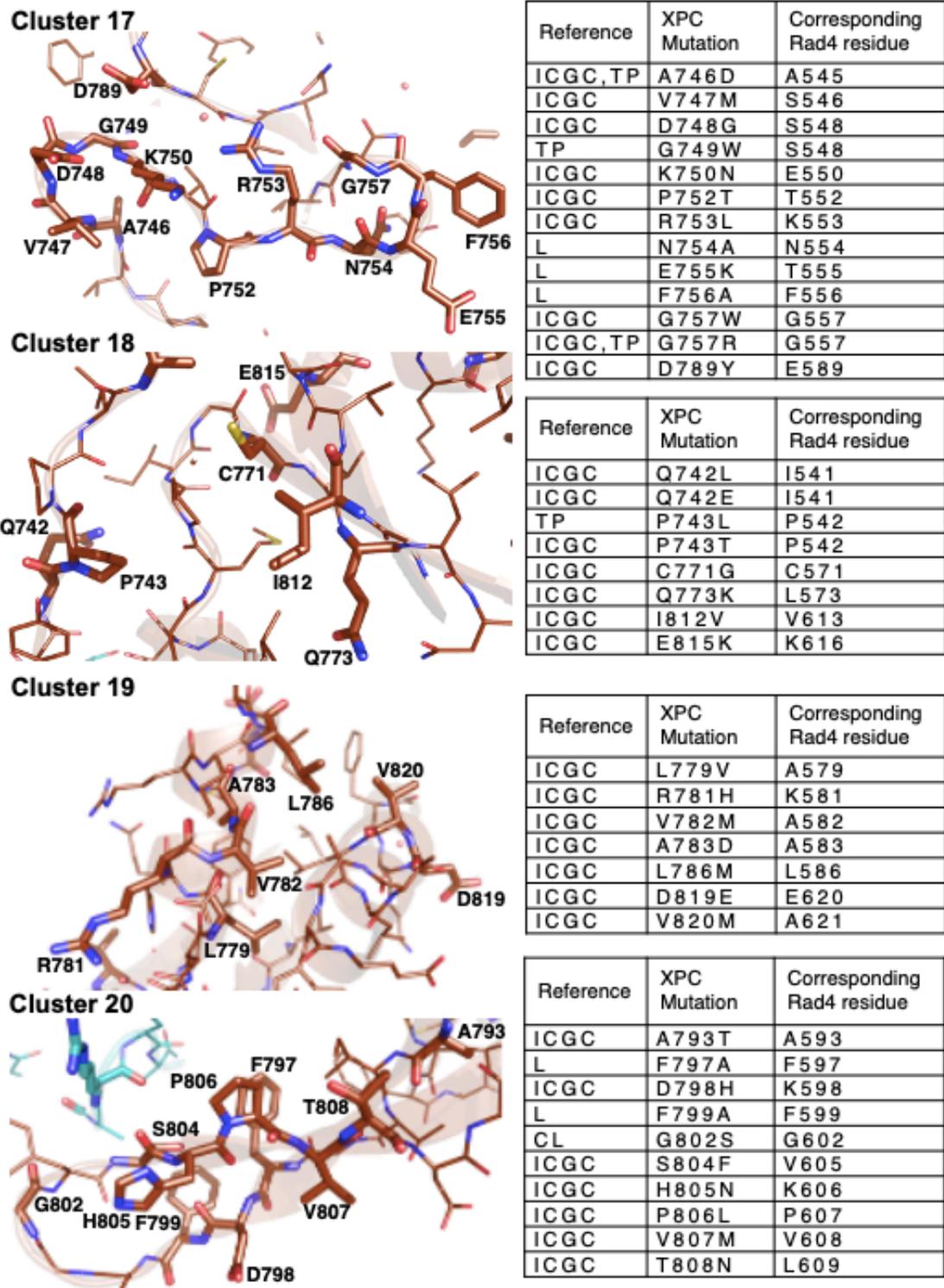


Figure 16. Local structural views for Clusters 17-20. The amino acids that were indicated in the cluster are shown in stick representation and all other residues are in line representation. Maroon indicates carbon (in BHD3), blue nitrogen, and oxygen red.

CHAPTER SIX

Summary

In this study, 288 genetic variations from ClinVar, International Cancer Genome Consortium, TumorPortal, and the literature were compiled and presented on to the homology model structure of human XPC. These studies will help understand the function and impact of each residue and present a compilation of genetic variations that may impact biological processes including nucleotide excision repair, as well as diseases such as cancer. Further research is needed to provide an accurate human XPC structure and examine the structural, biochemical, and clinical effects of little-known genetic variations.

REFERENCES

- Anantharaman V, Koonin EV, Aravind L. Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold. *Hum Mol Genet.* 2001 Aug 1;10(16):1627–30.
- Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 2011 Feb 1;27(3):343–50.
- Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports.* 2017 Sep 5;7(1):1–15.
- Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, et al. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D313–9.
- Bradford PT, Goldstein AM, Tamura D, Khan SG, Ueda T, Boyle J, et al. Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. *Journal of Medical Genetics; London.* 2011 Mar;48(3):168.
- Bunick CG, Miller MR, Fuller BE, Fanning E, Chazin WJ. Biochemical and Structural Domain-Analysis of XPC. *Biochemistry.* 2006 Dec 19;45(50):14965–79.
- Camenisch U, Träutlein D, Clement FC, Fei J, Leitenstorfer A, Ferrando-May E, et al. Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein. *The EMBO Journal.* 2009 Aug 19;28(16):2387–99.
- Chen X, Velmurugu Y, Zheng G, Park B, Shim Y, Kim Y, et al. Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat Commun [Internet].* 2015 Jan 6 [cited 2020 Jan 28];6. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4354021/>
- Clement FC, Naegeli H, Kaczmarek N, Mathieu N, Tomas M, Leitenstorfer A, et al. Dissection of the xeroderma pigmentosum group C protein function by site-directed mutagenesis. *Antioxidants and Redox Signaling.* 2011 Jun 15;14(12):2479–90.

- DiGiovanna JJ, Kraemer KH. Shining a Light on Xeroderma Pigmentosum. *Journal of Investigative Dermatology*. 2012 Mar 1;132(3, Part 2):785–96.
- ENSG00000154767 | ICGC Data Portal [Internet]. [cited 2020 Apr 14]. Available from: <https://icgcportal.genomics.cn/genes/ENSG00000154767>
- Espi P, Parajuli S, Benfodda M, Lebre A-S, Paudel U, Grange A, et al. Clinical and genetic characteristics of xeroderma pigmentosum in Nepal. *Journal of the European Academy of Dermatology and Venereology*. 2018;32(5):832–9.
- Gozukara EM, Khan SG, Metin A, Emmert S, Busch DB, Shahlavi T, et al. A Stop Codon in Xeroderma Pigmentosum Group C Families in Turkey and Italy: Molecular Genetic Evidence for a Common Ancestor. *J Invest Dermatol*. 2001 Aug 1;117(2):197–204.
- Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *ELECTROPHORESIS*. 2009;30(S1):S162–73.
- Jacobelli S, Soufir N, Lacapere JJ, Regnier S, Bourillon A, Grandchamp B, et al. Xeroderma pigmentosum group C in a French Caucasian patient with multiple melanoma and unusual long-term survival. *British Journal of Dermatology*. 2008;159(4):968–73.
- Jesus BMB de, Bjørås M, Coin F, Egly JM. Dissection of the Molecular Defects Caused by Pathogenic Mutations in the DNA Repair Factor XPC. *Molecular and Cellular Biology*. 2008 Dec 1;28(23):7225–35.
- Knips A, Zacharias M. Both DNA global deformation and repair enzyme contacts mediate flipping of thymine dimer damage. *Sci Rep*. 2017 Feb;7(1):41324.
- Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018 04;46(D1):D1062–7.
- Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014 Jan;505(7484):495–501.

- Lehmann et al. Xeroderma Pigmentosum [Internet]. Orphanet Journal of Rare Diseases. 2011 [cited 2019 Aug 13]. Available from: <https://ojrd.biomedcentral.com/track/pdf/10.1186/1750-1172-6-70>
- Li L, Bales ES, Peterson CA, Legerski RJ. Characterization of molecular defects in xeroderma pigmentosum group C. *Nature Genetics*. 1993 Dec;5(4):413.
- Maillard O, Solyom S, Naegeli H. An Aromatic Sensor with Aversion to Damaged Strands Confers Versatility to DNA Repair. *PLoS Biol* [Internet]. 2007 Apr [cited 2020 Apr 20];5(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1820611/>
- Martin-Morales L, Rofes P, Diaz-Rubio E, Llovet P, Lorca V, Bando I, et al. Novel genetic mutations detected by multigene panel are associated with hereditary colorectal cancer predisposition. *PLoS One*; San Francisco. 2018 Sep;13(9):e0203885.
- Meneses M, Chavez-Bourgeois M, Badenas C, Villablanca S. Atypical Clinical Presentation of Xeroderma Pigmentosum in a Patient Harboring a Novel Missense Mutation in the XPC Gene: The Importance of Clinical Suspicion. *Dermatology (Basel, Switzerland)*. 2015 Jan 1;231(3):217.
- Min J-H, Pavletich NP. Recognition of DNA damage by the Rad4 nucleotide excision repair protein [Internet]. *Nature*. 2007 [cited 2019 Aug 22]. Available from: <http://link.galegroup.com/apps/doc/A189796743/AONE?sid=lms>
- Min J-H, Pavletich NP. Recognition of DNA damage by the Rad4 nucleotide excision repair protein [Internet]. *Nature*. 2007 [cited 2019 Aug 22]. Available from: <http://link.galegroup.com/apps/doc/A189796743/AONE?sid=lms>
- Nishi R, Okuda Y, Watanabe E, Mori T, Iwai S, Masutani C, et al. Centrin 2 Stimulates Nucleotide Excision Repair by Interacting with Xeroderma Pigmentosum Group C Protein. *Mol Cell Biol*. 2005 Jul;25(13):5664–74.
- Paul D, Mu H, Zhao H, Ouerfelli O, Jeffrey PD, Broyde S, et al. Structure and mechanism of pyrimidine–pyrimidone (6-4) photoproduct recognition by the Rad4/XPC nucleotide excision repair complex. *Nucleic Acids Res*. 2019 Jul 9;47(12):6015–28.
- Petrović D, Zlatović M. Modeling Human Serum Albumin Tertiary Structure To Teach Upper-Division Chemistry Students Bioinformatics and Homology Modeling Basics [Internet].

- American Chemical Society and Division of Chemical Education, Inc.; 2015 [cited 2020 Apr 22]. Available from: <https://pubs.acs.org/doi/pdf/10.1021/ed500358f>
- Puumalainen M-R, Rütthemann P, Min J-H, Naegeli H. Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation. *Cell Mol Life Sci.* 2016 Feb 1;73(3):547–66.
- Qiao B, Ansari A-H, Scott GB, Sak SC, Chambers PA, Elliott F, et al. In vitro functional effects of XPC gene rare variants from bladder cancer patients. *Carcinogenesis.* 2011 Apr 1;32(4):516–21.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015 May;17(5):405–24.
- Rivera-Begeman A, McDaniel LD, Schultz RA, Friedberg EC. A novel XPC pathogenic variant detected in archival material from a patient diagnosed with Xeroderma Pigmentosum: A case report and review of the genetic variants reported in XPC. *DNA Repair.* 2007 Jan 4;6(1):100–14.
- Spivak G. Transcription-Coupled Repair: an update. *Arch Toxicol.* 2016 Nov;90(11):2583–94.
- Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics.* 2020 Mar 1;36(6):1765–71.
- Sugasawa K. *Carcinogenesis* vol.29 no.3 pp.455–465, 2008 doi:10.1093/carcin/bgm282 Advance Access publication January 3, 2008.
- SWISS-MODEL: homology modelling of protein structures and complexes | *Nucleic Acids Research* | Oxford Academic [Internet]. [cited 2020 Apr 14]. Available from: <https://academic.oup.com/nar/article/46/W1/W296/5000024>
- Weerd-Kastelein E a. D, Keijzer W, Bootsma D. Genetic Heterogeneity of Xeroderma Pigmentosum demonstrated by Somatic Cell Hybridization. *Nature New Biology.* 1972 Jul;238(81):80–3.

Yasuda G, Nishi R, Watanabe E, Mori T, Iwai S, Orioli D, et al. In Vivo Destabilization and Functional Defects of the Xeroderma Pigmentosum C Protein Caused by a Pathogenic Missense Mutation. *Molecular and Cellular Biology*. 2007 Oct 1;27(19):6606–14.

Zhang ET, He Y, Grob P, Fong YW, Nogales E, Tjian R. Architecture of the human XPC DNA repair and stem cell coactivator complex. *PNAS*. 2015 Dec 1;112(48):14817–22.

Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database (Oxford)*. 2011;2011:bar026.