Structural modeling and analyses of genetic variations in the human XPC nucleotide excision repair protein

Jennifer Le[1*±,] and Jung-Hyun Min[1†]

[1]Department of Chemistry & Biochemistry, Baylor University, Waco, TX 76798, USA

[†] To whom correspondence may be addressed

[±] Present address: College of Pharmacy, University of Texas at Austin, Austin, TX 78712

Tel: (254)710-2095

E-mail: JungHyun_Min@baylor.edu

ABSTRACT

Xeroderma pigmentosum C (XPC) is a key initiator in the global genome nucleotide excision repair pathway in mammalian cells. Inherited mutations in the XPC gene can cause xeroderma pigmentosum (XP) cancer predisposition syndrome that dramatically increases the susceptibility to sunlight-induced cancers. Various genetic variants and mutations of the protein have been reported in cancer databases and literature. The current lack of a high-resolution 3-D structure of human XPC makes it difficult to assess the structural impact of the mutations/genetic variations. Using the available high-resolution crystal structure of its yeast ortholog, Rad4, we built a homology model of human XPC protein and compared it with a model generated by AlphaFold. The two models are largely consistent with each other in the structured domains. We have also assessed the degree of conservation for each residue using 966 sequences of XPC orthologs. Our structure- and sequence conservation-based assessments largely agree with the variant's impact on the protein's structural stability, computed by FoldX and SDM. Known XP missense mutations such as Y585C, W690S, and C771Y are consistently predicted to destabilize the protein's structure. Our analyses also reveal several highly conserved hydrophobic regions that are surface-exposed, which may indicate novel intermolecular interfaces that are yet to be characterized.

Keywords: xeroderma pigmentosum; XPC; Rad4; nucleotide excision repair; X-ray crystallography; homology model; gene mutation; polymorphism

List of abbreviations:

BHD1: β-hairpin domain 1

BHD2: β-hairpin domain 2

BHD3: β-hairpin domain 3

H-bond: hydrogen bond

NER: nucleotide excision repair

R4BD: Rad4-binding domain

TGD: transglutaminase domain

VdW: van der Waals

XPC: xeroderma pigmentosum C

INTRODUCTION

Humans are exposed to many environmental genotoxins that induce DNA damage, including UV radiation from sunlight and chemicals from pollution. DNA damage, if not repaired, can interfere with essential cellular functions such as replication and transcription. Nucleotide excision repair (NER) is an evolutionarily conserved mechanism to repair a wide variety of such DNA damage (1-4). Failure in NER can cause the accumulation of DNA lesions that lead to mutations and cancers, as manifested in the human xeroderma pigmentosum (XP) cancer predisposition syndrome.

The versatile NER is carried out in a stepwise manner involving ~30 protein factors. The first step is lesion recognition, which can occur in two distinct mechanisms. In transcription-coupled NER (TC-NER), lesions are recognized by an RNA polymerase stalled at a lesion during the transcription (5,6). Thus TC-NER repairs lesions only on actively transcribed strands. In global-genome NER (GG-NER), the lesions anywhere in the genome can be sensed by specialized damage sensors such as the UV-damaged DNA binding protein (UV-DDB) complex and the xeroderma pigmentosum C (XPC)-RAD23B-CETN2 complex. Once the lesions are recognized, both TC- and GG-NER proceed through the same subsequent steps.

UV-DDB is important in sensing cyclobutane pyrimidine dimer UV lesions and helps XPC to recognize the lesion in the chromatin context (7-10). For many other helix-distorting and -destabilizing bulky DNA adducts, XPC functions as the primary sensor. Using thermal energy alone, XPC surveys DNA within chromatin and locates diverse target lesions by specifically binding to them (11-14). DNA lesions that XPC recognizes for NER include various intra-strand crosslinks and helix destabilizing/distorting adducts, which are formed by ultraviolet (UV) light, air and water pollutants, and toxins (15). Once bound to a lesion, XPC recruits the 10-subunit

general transcription factor II H complex (TFIIH), which then verifies the presence of a bulky lesion and recruits other subsequent factors (16-20); eventually the lesion-containing single-stranded DNA is excised by XPF-ERCC1 and XPG endonucleases, and the DNA is restored by repair synthesis and nick sealing by DNA polymerases and DNA ligases, respectively (15,21).

The versatility of NER in repairing diverse lesions is largely attributed to its unique lesion recognition mechanism involving the XPC damage sensor. Studies have shown that human XPC and yeast Rad4 share similar DNA binding and lesion recognition characteristics as well as their role in the NER initiation (22,23). Though there is no high-resolution XPC structure available, there are crystal structures of its yeast ortholog, Rad4. The crystal structures of Rad4 bound to UV-induced 6-4 photoproduct (6-4PP) or mismatched DNA showed that Rad4 (and by analogy XPC) recognizes the damaged DNA site by inserting a β-hairpin from its β–hairpin domain 3 (BHD3) into the DNA duplex and flipping out damage-containing nucleotide pairs to form an "open" conformation (24,25). In this conformation, Rad4 interacts exclusively with the two nucleotides on the undamaged strand without making specific contacts with the two damaged residues. This indirect recognition thus can allow Rad4/XPC to bind to DNA lesions with varied structures (24,25). Recent structural and biophysical studies have also indicated that the lesion recognition by XPC/Rad4 may rely on a 'kinetic gating' mechanism (26-32). In this mechanism, damaged DNA may be selectively opened over undamaged DNA by a freely diffusing protein, not necessarily because the protein is inherently unable to open undamaged DNA, but because the free energy barrier for opening is high and the protein does not linger long enough at an undamaged site to efficiently open that site before diffusing away (26). On the other hand, damaged DNA, since already destabilized, has a lower free energy barrier for opening, thus increasing the probability that the protein can open that site and not diffuse away.

Such a kinetically controlled, indirect readout mechanism enables rapid, yet reliable recognition of diverse lesions dispersed in the predominantly undamaged genomic DNA.

Genetic mutations in human XPC can cause XP cancer syndrome, marked by extreme sun sensitivity and >1,000-fold higher risk of sunlight-induced skin cancers (33). Polymorphisms of XPC contribute to a variety of cancers, including lung, prostate, and bladder cancers (34-40). Recent studies also implicated XPC in functions other than NER, such as transcription and base excision repair, but without clear mechanistic explanation (41-48). However, it has been difficult to assess how the mutations and variations in the XPC protein may affect its structure or key interactions to lead to the phenotypes, as mainly hampered by the lack of a high-resolution structure of human XPC.

To aid the structural and mechanistic assessment of known XPC mutations and genetic variations, we built a homology model of human XPC based on the crystal structures of Rad4, which remains the only XPC ortholog with an available high-resolution structure. The homology model was largely consistent with a model generated by a recent, artificial intelligence-based algorithm, AlphaFold. Subsequently, potential structural impact of each variant was assessed based on the structural models and sequence conservation, as well as by computational analyses using FoldX and SDM. The assessments generally agreed with one another. Notably, known XP missense mutations such as Y585C, W690S, and C771Y were predicted to destabilize the protein's structure. Our analyses also revealed several highly conserved hydrophobic regions that are surface-exposed, which may indicate novel intermolecular interfaces that are yet to be characterized. Structural models and variant analyses are also available as downloadable files as Supplementary Information.

MATERIALS and METHODS

Homology modeling and refinement of the human XPC protein

Homology modeling was done in the following steps, largely as described by Petrovic et al. (49): template selection, sequence and structural alignment, model building, and refinement. For template selection, the Basic Local Alignment Search Tool (BLAST) was used to choose the protein exhibiting the greatest sequence identity with the human XPC isoform 1 (UniProtKB/Swiss-Prot**:** Q01831-1), within the Protein Data Bank (PDB) database. Rad4 (PDB: 2QSH_A), the yeast ortholog of XPC exhibited the highest sequence identity (26.4%) and thus chosen as the template. The sequence alignment between the two sequences was generated using Clustal Omega (50,51). Next, the homology model structure was constructed by SWISS-MODEL (52-55) (https://swissmodel.expasy.org) using the Rad4 crystal structure from the Rad4-Rad23-mismatched DNA structure (PDB: 2QSH) (24) (**Figure S1**). Several loop regions containing Ramachandran outliers in the initial model (residues 291-294, 597-600, 611-616, 631-637, 718-728 and 742-751) were selected and further refined using MODELLER loop refinement (version 10.4 in Chimera (version 1.16)) (56-58). This refinement increased the Ramachandran favored residues (82% to 86%) and decreased the Ramachandran outlier residues (8% to 4%) (**Figures S2-S3**). The final model has 95% of all modeled residues in Ramachandran favored/allowed with 86% in Ramachandran favored. Though 4% were Ramachandran outliers, all of these except for one residue, Arg293, belonged to the long low-complexity linker region within TGD that lacked a structural template from Rad4 (see Results). The final model encompasses residues 191-829, including the flexible internal loop corresponding to residues 331-520 in human XPC. More detailed validation for the SwissModel as well as the model generated by AlphaFold was carried out using ProCheck (59), and the results are included as SI

(**SI Files S2 and S3**). Structural analyses and visualization were done using PyMOL (version 2.5.4).

Multiple sequence alignments of XPC orthologs using Clustal Omega and ConSurf

Multiple sequence alignment (MSA) of representative XPC orthologs was generated with Clustal Omega using default parameters and an output format 'ClustalW with character counts', and the sequences from UniProtKB/Swiss-Prot, Genbank, and NCBI retrieved in a FASTA format. The sequence alignment figures were made by Jalview (version 2.11.0) by importing the output from Clustal Omega. The accession numbers of the ortholog sequences used are: human XPC **(**NP_004619.3), *S. cerevisiae* Rad4 (XP_033765979.1), mouse XPC (NP_033557.2), chicken XPC (UniProtKB/Swiss-Prot**:** E1BUG1), fly XPC (*D. melanogaster* UniProtKB/Swiss-Prot**:** Q24595), worm *C. elegans* (UniProtKB/Swiss-Prot**:** Q9N4C3), thale cress *A. thaliana* (UniProtKB/Swiss-Prot**:** Q8W489), fugu *T. rubripes* XPC (NCBI Reference Sequence: XP_003963083.2), zebrafish *D. rerio* XPC (NCBI Reference Sequence: NP_001038675.1), frog *X. tropicalis* XPC (NCBI Reference Sequence: XP_004914219.2), yeast *A. nidulans* Rad4 (UniProtKB/Swiss-Prot**:** Q5B6E0), yeast *S. pombe* Rhp41 (UniProtKB/Swiss-Prot**:** Q10445), yeast *C. glabrata* Rad4 (UniProtKB/Swiss-Prot**:** A0A0W0CU84), and yeast *K. lactis* Rad4 (Genbank QEU62520.1).

Next, we aimed to analyze the sequence conservation on a larger scale. To achieve this, BLAST was used to retrieve 1000 sequences of XPC orthologs in a FASTA format using human XPC as a search query (UniProtKB: Q01831). The search was done using default parameters, except that the number of maximum target sequences was set to 1000 and that homo sapiens was excluded from search organism. Subsequently, we refined the selection of 1000 orthologous sequences by limiting the sequence identity from 35% to 95% to remove sequences that were too

different or too similar (60,61), which resulted in 964 sequences. The sequences of human XPC (UniProtKB: Q01831) and *S. cerevisiae* Rad4 (UniProtKB: P14736) were manually added to the previous result in FASTA format, as reference sequences for this study. Clustal Omega was then used to align total 966 XPC homolog sequences and the result was saved in an .aln format. The file including all sequences used for our large-scale MSA is included as SI (**SI File S4)**. Finally, to generate the conservation scores, the ConSurf server (https://consurf.tau.ac.il/) was used: the human XPC homology model generated above was uploaded as the reference structure, and the 966-sequence MSA was used as the input sequences. The degree of conservation for each residue position is indicated in scores ranging from 1 to 9: 1 being most variable to 9 being most conserved (62) (**Figure S5**).

Assessing the impact of the gene variants on the protein stability (ΔΔG) using FoldX and Site Directed Mutator (SDM)

The putative effect of each gene variant on the structural stability of human XPC was analyzed with FoldX (63) and Site Directed Mutator (SDM) (64). The FoldX version 5.0 was run in YASARA version 22.9.24 (65). In FoldX, our human XPC model from SwissModel was uploaded, and each gene variant was analyzed using default parameters (1 run, pH 7, 298 K, Ionic strength 5, VdW design 2). In SDM, the Mutation List option was used to analyze multiple gene variants per run. The human XPC model from SwissModel was uploaded and gene variants were analyzed in a set of 20 each time. The statistical correlation coefficients between our scores (see Results), FoldX, and SDM were calculated using IBM SPSS Statistics version 29.0.0.0.

RESULTS

1. XPC homology model and the domain arrangement

The human *XPC* gene consists of 16 exons on the chromosomal location of 3p25.1 and encodes a protein of 940 residues (66,67). The XPC protein exists as a heterotrimeric complex together with RAD23B and CETN2/Centrin 2 in cells (67-71). High-resolution 3D structures that include the human XPC protein are limited to the structures of its C-terminal peptide (residues 847-863) in a complex with CETN2 (72-74) and the structure of an N-terminal acidic peptide (residues 124-143) in complex with TFIIH p62 (75). A low-resolution cryo-EM architecture of the human XPC complex is also available (76). Therefore, to build a structural model of human XPC, we used one of the highest-resolution structures available for its yeast ortholog, Rad4 (24). NER machinery including XPC is evolutionarily conserved in all eukaryotes from yeast to humans. Yeast Rad4 is the ortholog of mammalian XPC with 27% overall sequence identity. Both form a heterotrimeric complex in cells with two other proteins, RAD23B and CETN2 (Rad23 and Rad33 in yeast) and exhibit similar DNA binding and lesion recognition properties for NER (22,23,77,78). Prior computational study have also suggested a conserved mode of DNA binding between Rad4 and human XPC (79).

Figures 1 and **2** show the homology model of human XPC and sequence alignment of select XPC orthologs, respectively. The structured domains of XPC/Rad4 contain 4 key DNA-binding domains: a transglutaminase domain (TGD) followed by three β-hairpin domains (BHD1-3) (**Figure 1c**). TGD at its N-terminus contains a ~45-residue transglutaminase fold (24). Most in the transglutaminase superfamily are enzymes containing a catalytic triad comprised of a histidine, a cysteine, and an aspartate, but Rad4 and XPC lack the catalytic triad and have no catalytic activity (24,80). The ensuing three BHDs form a curved protrusion from the globular TGD lobe. Each span ~50-90-residues to form a winged helix-turn-helix (wHTH)-like domain with a long β-hairpin that participates in DNA binding (81). Upon binding to damaged DNA,

TGD-BHD1 binds to ~10-bp undamaged portion of the DNA on the 3' side of the lesion. BHD2-BHD3 binds to ~4-bp DNA segment where 2 nucleotide pairs containing the lesion are flipped out from both strands (25). Notably, BHD3 β-hairpin inserts through the DNA double helix, filling the gap resulting from the two flipped-out nucleotide pairs. The groove between BHD2–BHD3 also holds the backbone of the undamaged strand and binds the two complementary flipped-out nucleotides. It is noteworthy that the BHD3's key contacts involving the inserted hairpin (e.g., F799 in human (F599 in Rad4)) and the binding pockets for the flipped-out nucleotides remain the same as the involved residues are identically conserved (e.g., F797 (F597), P806 (P607), F756 (F556), N754 (N554), N758 (N558) in human XPC (yeast Rad4)).

One of the key differences between yeast Rad4 and human XPC is the presence of a ~175 residue insertion within the TGD domain in the higher organisms (human XPC residues 331-520). This 'linker' region is enriched with basic residues and is predicted to be flexible and unstructured. In our model, the linker faces the DNA on the same side as the α2 helix of TGD which contacts the undamaged portion of the DNA together with the BHD2 β-hairpin on the opposite side of the DNA. A recent study shows that this region interacts with the deacetylated histone H3 tail and participates in promoting GG-NER in DDB-independent manner (82). Human XPC is also predicted to have a shorter β-hairpin in BHD1 and a longer β-hairpin in BHD2 compared with those in the yeast Rad4 (**Figure 2**). A shorter β-hairpin in BHD1 may be allowed because of the presence of the long linker region in TGD that could have a compensatory role in DNA binding. A longer BHD2 may also allow more extensive contact with the minor groove around the lesion site (79). These features may contribute to the higher affinity of human XPC to DNA than that of Rad4 (22,23,83,84).

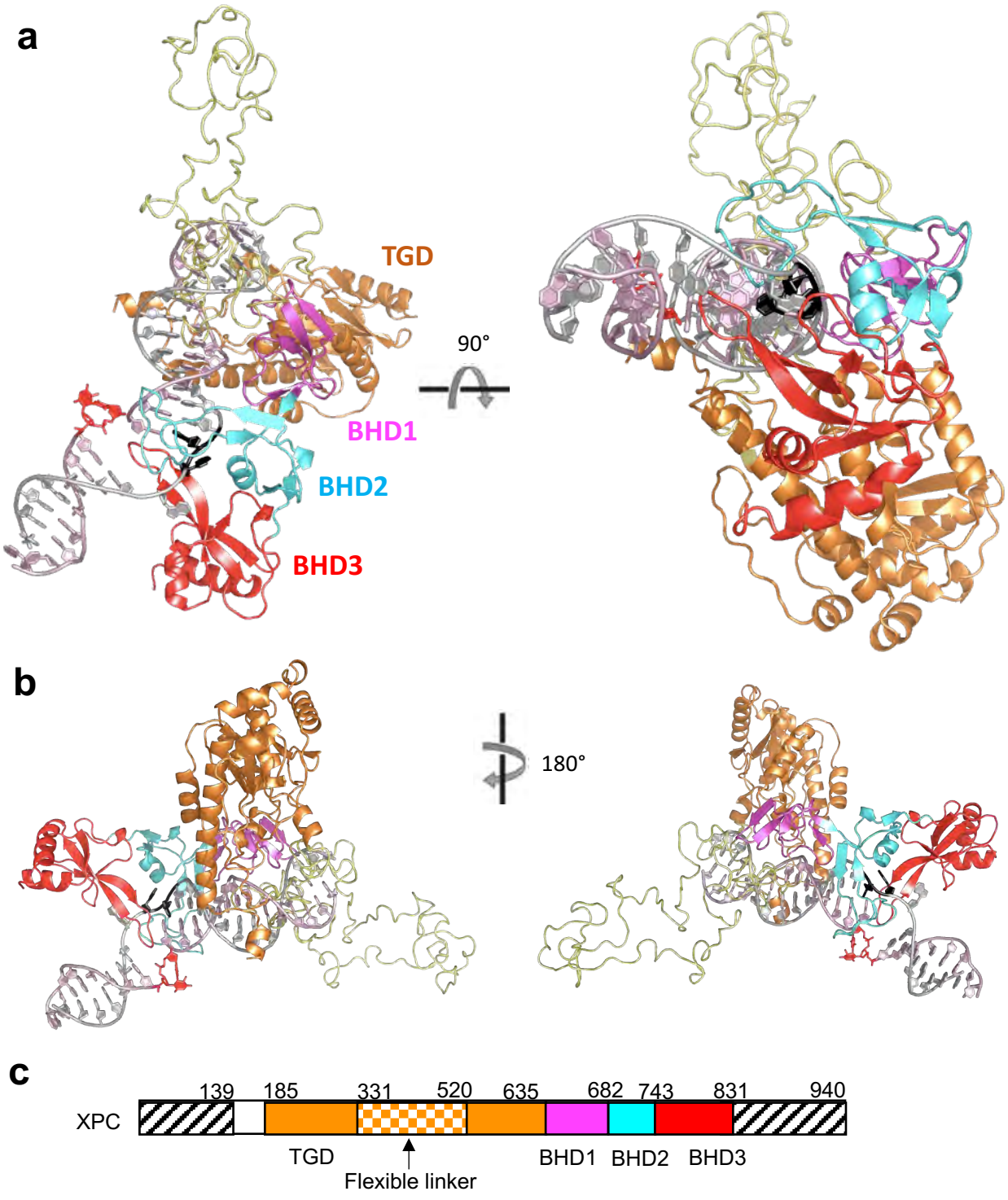
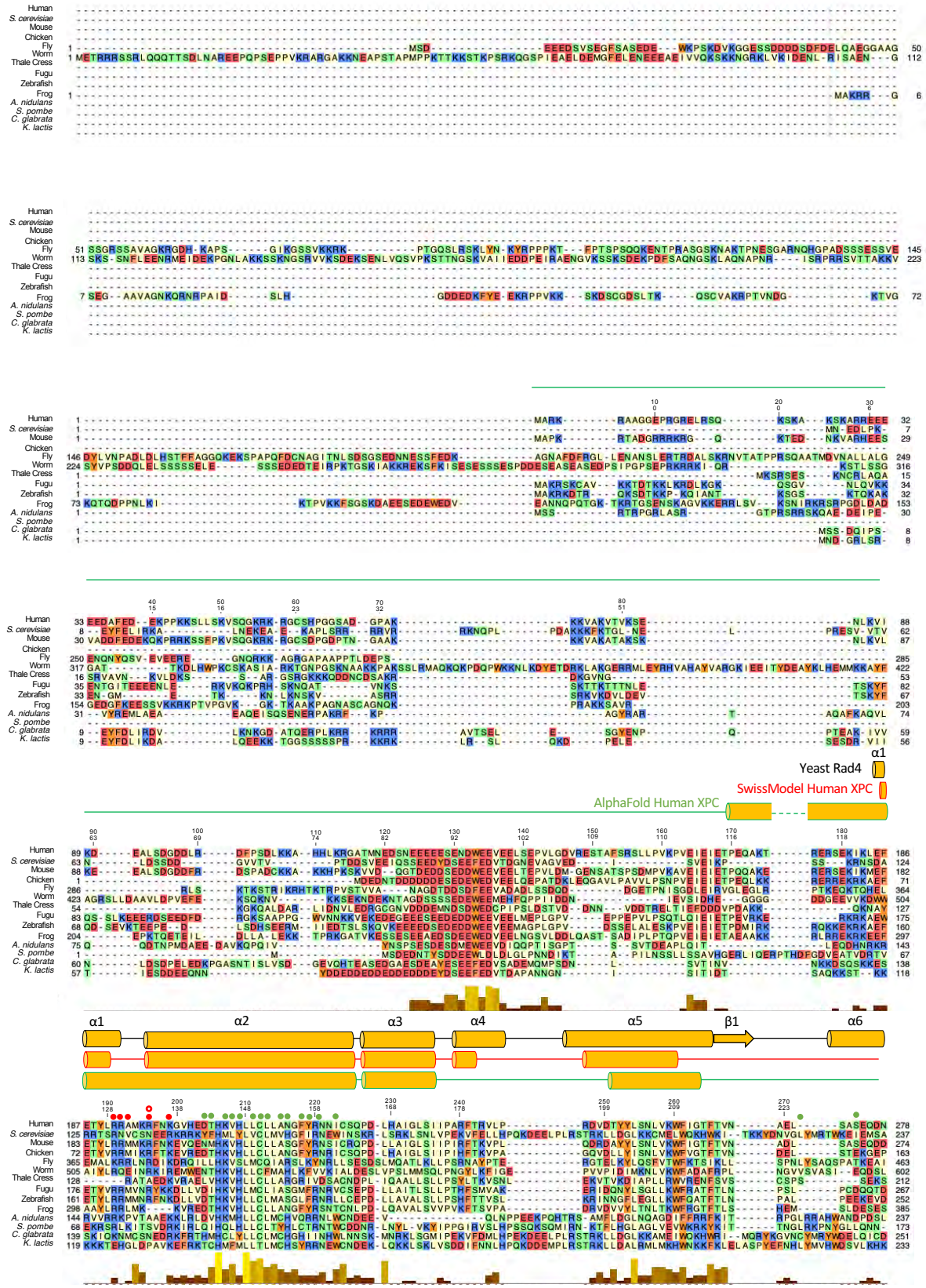
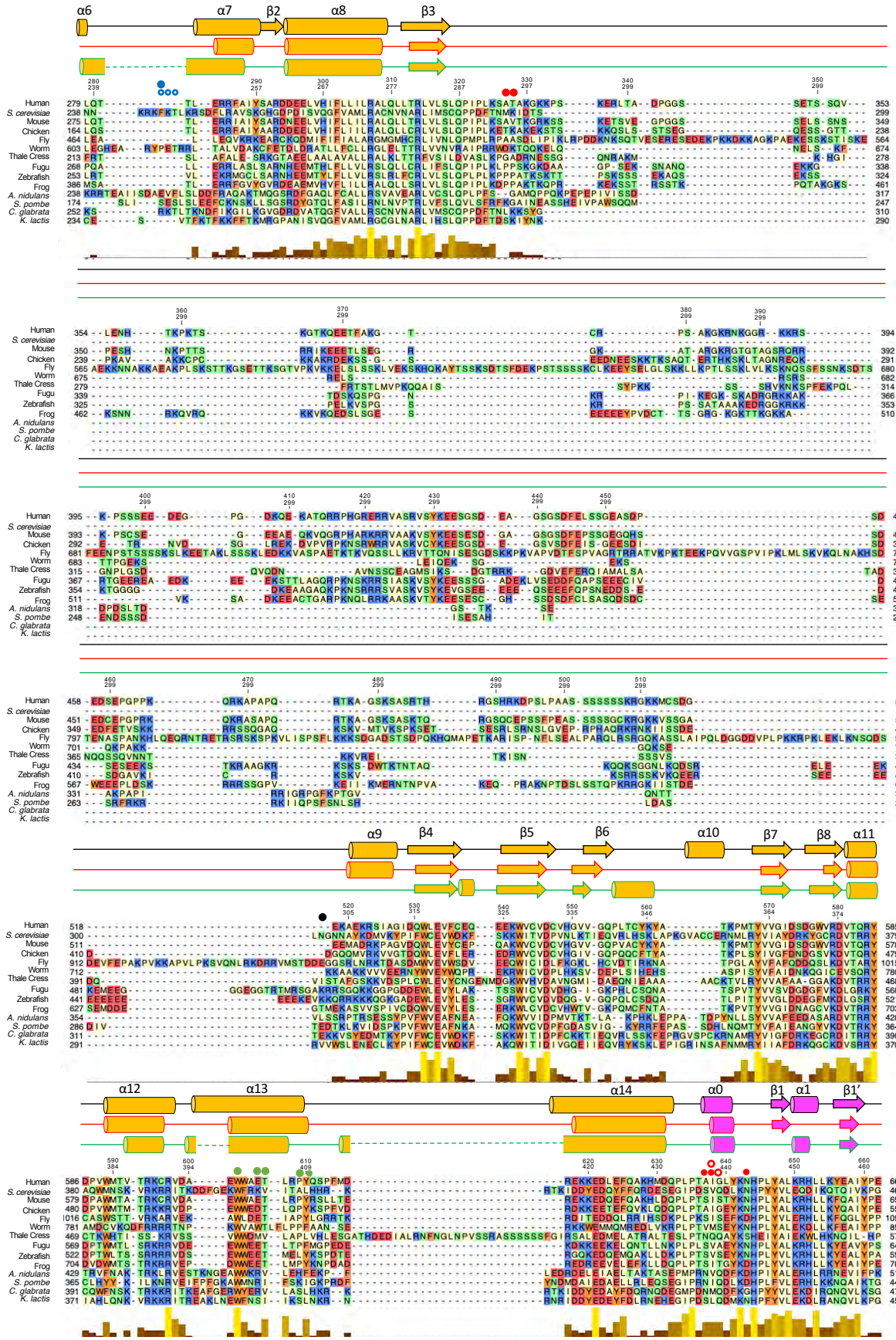


Figure 1. XPC homology model from SwissModel bound to DNA shown in cartoon representation. The transglutaminase domain (TGD) is colored orange, β-hairpin domain 1 (BHD1) magenta, BHD2 cyan, and BHD3 red, and the disordered linker region is light yellow. In DNA of **(a-b)** the 6-4PP lesion is shown in red on the damaged strand (light pink). XPC directly contacts bases (black) on the undamaged strand (grey), opposite of the 6-4PP lesion. **(c)** The domain arrangement of human XPC. Striped/checkered regions correspond to the N- and C-termini and the TGD linker; these regions are absent in the template Rad4 structure.


Human
S. cerevisiae
Mouse
Chicken
Fly
Worm
Thale Cress
Fugu
Zebrafish
Frog
A. nidulans
S. pombe
C. glabrata
K. lactis
α1
Yeast Rad4
SwissModel Human XPC
AlphaFold Human XPC
α1
α2
α3
α4
α5
β1
α6


α6
α7
β2
α8
β3
Human
S. cerevisiae
Mouse
Chicken
Fly
Worm
Thale Cress
Fugu
Zebrafish
Frog
A. nidulans
S. pombe
C. glabrata
K. lactis
α9
β4
β5
β6
α10
β7
β8
α11
α12
α13
α14
α0
β1
α1
β1'

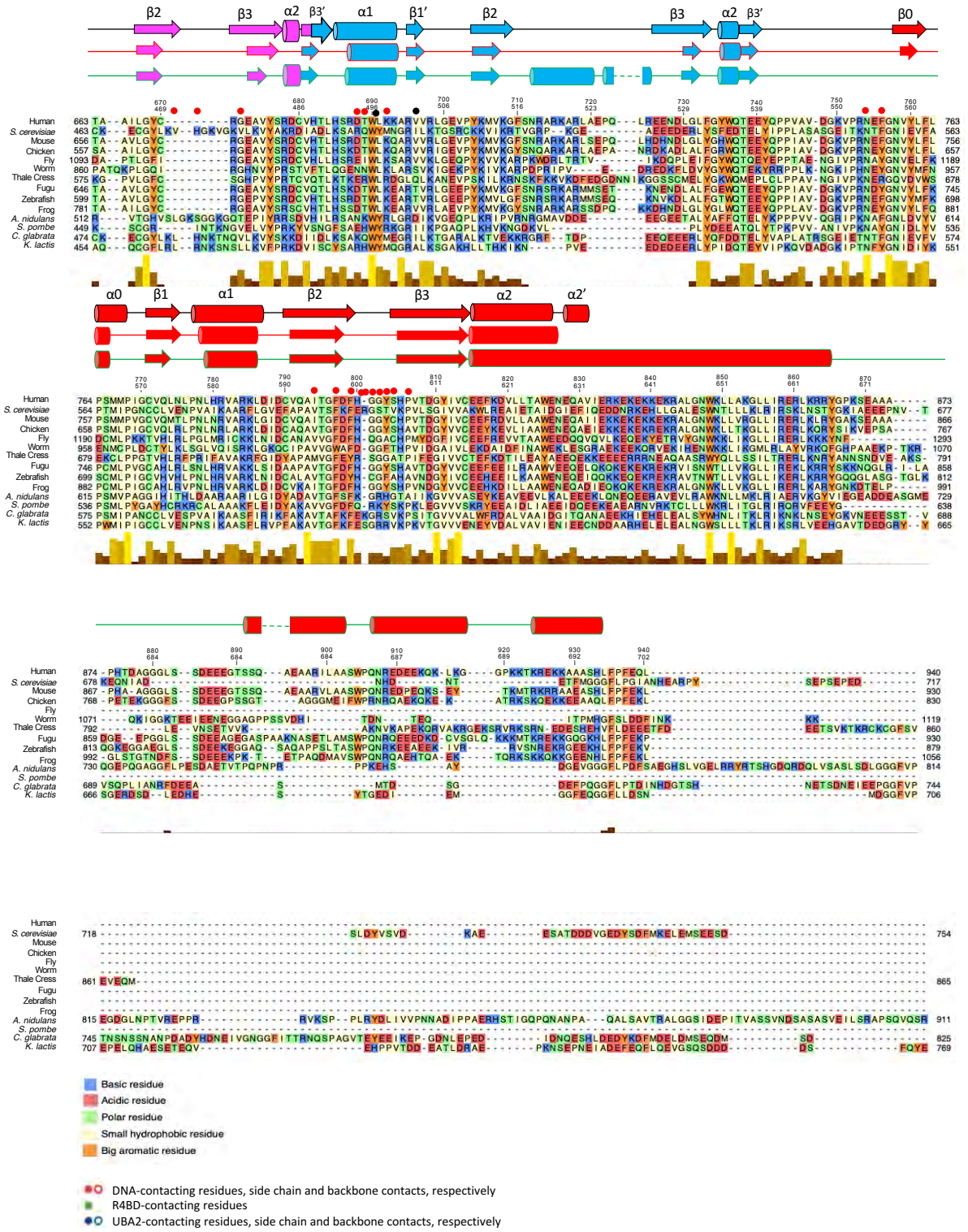


Figure 2. Sequence alignment of XPC orthologs and secondary structures from Rad4/XPC models. The secondary structures of the following structures are shown above the alignment: yeast Rad4 (PDB:2QSH, black outline), the SwissModel of human XPC (red outline), and the AlphaFold model of human XPC (green outline).

2. Comparison between the structural models by SwissModel and AlphaFold

Recent development in structural prediction programs using artificial intelligence such as AlphaFold (https://alphafold.ebi.ac.uk/) have garnered much attention (85). AlphaFold uses machine learning that combines physical and biological knowledge about protein structure, multi-sequence alignments with deep learning algorithms.

Overall, the AlphaFold model of human XPC is highly similar to the model generated by SwissModel. The RMSD for all atoms is 2.618 Å (2367 to 2367 atoms), and for C_a is 2.165 Å (288 to 288 atoms) within the structured domains (residues 184-330 and 522-831) (**Figure 3**). The residue-by-residue confidence estimate as well as the predicted aligned error plot of the AlphaFold model are shown in **Figure S6.** The conformation of the AlphaFold model resembles that of the lesion-bound Rad4 more than that of the apo-Rad4, similarly as the SwissModel of human XPC (**Figure S7**). However, it differs in several ways in the regions including the N- and C- termini and the flexible linker within TGD, compared with the SwissModel. In general, we find that AlphaFold tends to predict continuous α-helices in place of two separate helices as in the SwissModel modeled after the yeast Rad4 crystal structures. AlphaFold also predicted additional helices where it is modeled or predicted to be unstructured in the SwissModel. Major specific differences between the two models are described below.

First, AlphaFold predicts a single contiguous α-helix in the N-terminus of XPC (residues 170-225). In the SwissModel of XPC, the residues before residue 184 are not included as this region in the Rad4 crystal structures is disordered, and the α-helical region after residues 185 is broken into two helices, α1 (residues 185-190) and α2 (196-225), analogously to those in the Rad4 crystal structures (residues 123-128 and 134-163). Second, AlphaFold also predicts a ~50-residue, contiguous α-helix for residues 815-865 in BHD3 that includes residues 847-863 that

bind to CETN2 (72-74), instead of the two shorter helices, α2 (residues 815–826 corresponding to Rad4 residues 616-627) and α2' (residues 828-831 corresponding to Rad4 629-632) in the SwissModel. Third, AlphaFold predicts a loop for residues 520-526 instead of an α-helix in the SwissModel. AlphaFold also predicts an α-helix for residues 555-562 instead of a loop in the SwissModel. Fourth, AlphaFold models the loop spanning residues 266-281 differently from the SwissModel. This region in yeast Rad4 is significantly longer (residues 215-246) and is adjacent to its Rad23 binding site. Fifth, AlphaFold predicts two short α-helices (residues 712-720 & 722-725) between the β2 and β3 of BHD2. Sixth, the AlphaFold model also predicts that the conserved hydrophobic residues, Y660, W736 as well as Y704, pack against one another to form a hydrophobic interface between BHD1 and BHD2. Lastly, AlphaFold includes three additional α-helices beyond the end of (α2') in BHD3 (residues 891-903, 906-918 and 924-933) albeit with a low confidence level (pLDDT < 70) (**Figure S6**). Not surprisingly, the flexible linker within TGD is also modeled with very low confidence (pLDDT <50).

The significance of these differences between the models would be a worthy subject for future investigations. Most genetic variants discussed in this study fall into the regions where AlphaFold and SwissModel are largely similar. We used the SwissModel which has better geometry as our primary model for further discussion and included a comparison with the AlphaFold model as necessary.

3. Mutations and variants in the XPC gene from databases and clinical literature.

Single-amino acid missense variants in the human XPC gene were compiled from three different databases: Tumor Portal (TP), International Cancer Genome Consortium (ICGC), and ClinVar (CL). Though many known pathogenic mutations in the *XPC* gene (e.g., in XP patients) encode nonsense or frameshifts that encode truncated proteins with undetectable protein levels or

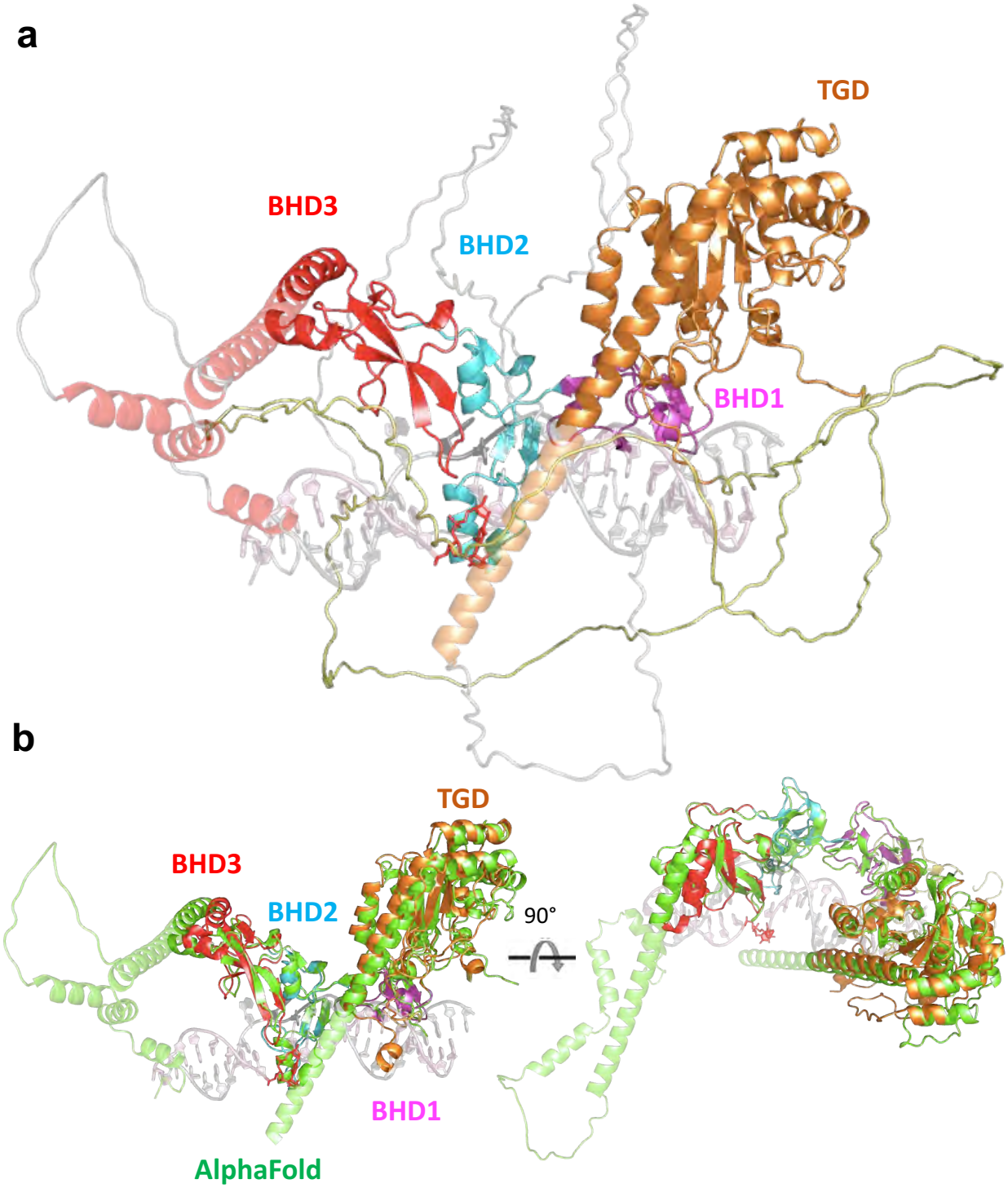


Figure 3. Comparison between the AlphaFold model and the SwissModel. (a) The AlphaFold model of full-length human XPC. The domains are colored the same as the SwissModel in Figure 1. The N- and C-terminal regions not included in the SwissModel are shown in semi-transparent form. DNA is also shown semi-transparent to help orient the viewer. **(b)** Superposition between the AlphaFold model and the SwissModel. The AlphaFold model (green) is superposed on the Swiss Model (colored by the domains as in Figure 1). The flexible linker within TGD and the N-terminal and C-terminal regions (shown semi-transparent in panel **a**) are omitted for clarity.

activities in cells (86), we focus our analyses on missense mutations/variants whose biological consequence may be more directly influenced by the residue's specific role for the protein's structure and interactions.

TumorPortal (TP) catalogues mutated genes in tumor samples (87). 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. For the *XPC* gene, there were 31 mutations reported from 16 types of tumor samples (http://www.tumorportal.org/). (**Table S2**; see **SI File S5**).

The International Cancer Genome Consortium (ICGC) data portal catalogs genomic abnormalities in 50 types of cancer managed by collaborative institutions (88). The data are taken from projects such as the Tumor Sequencing Project and include 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 were low impact (https://dcc.icgc.org/) (**Table S3**; see **SI File S5**).

ClinVar (CL) is a public archive of reported relationships among human gene variants and pathological phenotypes (86). CL organizes reports found in patient samples, assertions of their clinical significance, and other supporting evidence. The complexity of CL entries varies, from providing the presentation and interpretation of an allele to structure observation or experimental evidence about the effect of a genetic variant. The clinical significance of gene variants is classified according to Richards *et al.*(89). The classification ranges from "likely benign" to "likely pathogenic", which is strongly expected to cause diseases. Benign sequence variants may be single nucleotide polymorphisms (SNPs) or the result of genomic instability. 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/clinvar/ (**Table S4;** see **SI File S5**).

Finally, we compiled single-amino acid missense variants from clinical literature. The analysis provided for each mutation varies, from describing a patient case study to a meta-analysis of polymorphisms. We include the clinical impact and notes. For the *XPC* gene, there were 12 missense mutations reported from patient genotyping (**Table 1**).

Table 1. Missense genetic variants/mutations in the *XPC* gene compiled from clinical literature. The residue numbers are for the human XPC protein. Corresponding residues in Rad4 are indicated only if the residue falls into the structured regions in the crystal structure of DNA-bound Rad4 (PDB: 2QSH). Mutations or variants that fall outside the core region (TGD-BHD3) of XPC were not included. Residue positions that appear in our model and correspond to XPC domains are shown as orange for TGD, green for the flexible linker within TGD, purple for BHD1, blue for BHD2, and red for BHD3.

Position	Amino Acid Change	Corresponding Rad4 Residue	Clinical Symptoms/ Notes	Reference
302	F302S	F269	Rare variant identified as associated in bladder cancers. Showed defect in XPC recruitment to oxidative damage and increased risk of bladder cancers	(90)
314	R314Q	R281	Polymorphism	(91)
334	P334H	No corresponding residue	Pathogenic variant. Identified in cell line from an XP-C patient with neurological symptoms, which is rare. Also associated with increased risks in hereditary breast/ovarian and colon cancers.	(91-95)
393	R393W	No corresponding residue	Rare variant identified as associated in bladder cancers. Showed defect in XPC recruitment to oxidative damage and increased risk of bladder cancers	(90)
499	A499V	No corresponding residue	Somatic substitution in 4 out of 19 (21% frequency) prostate tumors in African American men Polymorphism associated with DNA repair deficits associated with increased overall cancer risk including bladder and ovarian cancers, but not lung cancer specifically In acute myeloid leukemia treated with induction chemotherapy, this variant was linked with lower overall disease-free survival, especially when combined with an XPD codon 751 AC/CC variant Associated with variable responses to imatinib in BCR-ABL driven chronic myelogenous leukemia	(34,37,40,91,96-104)

			Does not alter response to platinum-based chemotherapy Protective variant against prostate cancer development compared to controls. Is independently protective from smoking-associated colorecteral cancer and adenomas (CRC) PharmGKB: Genotypes AA (Ala/Ala) + AG (Ala/Val) are associated with decreased risk of gastrointestinal toxicity, Hematologic Diseases and Leukopenia when treated with Platinum compounds in people with Carcinoma, Non-Small-Cell Lung as compared to genotype GG (Val/Val).	
585	Y585C	Y379	Pathogenic variant. A 33 year-old Spaniard with an unusual distribution of lentigines limited to his upper back and scalp but sparing his face and forearms, despite excess UVR exposure in his youth. Melanoma developed on his scalp. UDS in his skin fibroblasts was ~40% of normal after a UVR dose of 10 Jm^{-2} The UDS level suggests that the XPC protein was partially functional.	(105)
689	T689M	Q495	Polymorphism	(91)
690	W690S	W496	Pathogenic variant identified in XP patient. The XPC transcript was present at a normal level and the XPC protein was detectable, although at a level lower than normal.	(91,96)
703	P703L	C509	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). Also has a loss-of-function variant of melanocortin 1 receptor (MC1R).	(106)
738	T738A	T537	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	(107)
771	C771Y	C571	Brazilian Caucasian male, diagnosed with XP at 10 mo. Symptoms of lentigines and photophobia. Multiple malignant skin tumors. C771Y missense is compound heterozygous with XPC c.780-1G>A intronic variant.	(108)

4. Literature survey on functionally important mutations in XPC *in vitro*.

Here, 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 2.

Table 2. Functionally important *XPC* mutations reported from *in vitro* studies. The residue numbers are for the human XPC protein. The types of single amino acid mutations are indicated along with corresponding residues in Rad4, which are indicated only if the residue falls into the structured regions in the crystal structure of DNA-bound Rad4 (PDB: 2QSH). Mutations or variants that fall outside the core region (TGD-BHD3) of XPC were not included. Residue positions that appear in our model and correspond to XPC domains are shown as orange for TGD, green for the flexible linker within TGD, purple for BHD1, blue for BHD2, and red for BHD3.

Position	Mutation	Corresponding Rad4 Residue	Biochemical Effect	Reference
334	P334H	No corresponding residue	Variant has active *in vitro* NER ability but impairs interaction with TFIIH p62 and delays XPA recruitment; weakens XPC interaction with Ogg1. Fails to fully stimulate XPB ATPase activity. Has no effect on XPC/KAT2A (a histone acetyltransferase) interaction	(47,109)
531	W531A	W316	Repair defective mutation causes moderate loss of DNA binding. Has a higher nuclear mobility compared to WT controls.	(110)
542	W542A	W327	Repair defective mutation causes moderate loss of DNA binding. Has a higher nuclear mobility compared to WT controls.	(110)
690	W690S	W496	This mutation drastically lowers *in vivo* stability, lowers DNA binding affinity for both undamaged and damaged DNA, and fails to recruit TFIIH and XPA in a stable complex	(13,110-113)
733	F733A	Y532	Repair defective mutation causes near-complete loss of DNA binding with poor ability to concentrate at the site of the lesion. Has a higher nuclear mobility compared to WT controls.	(110,112)
754	N754A	N554	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; severely reduces localization at DNA damaged foci.	(114)
755	E755K	T555	Causes significant GG-NER defect. Has a significantly lower nuclear mobility compared to WT controls.	(110)
756	F756A	F556	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; reduces localization at DNA damaged foci.	(114)
797	F797A	F597	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; reduces localization at DNA	(114)

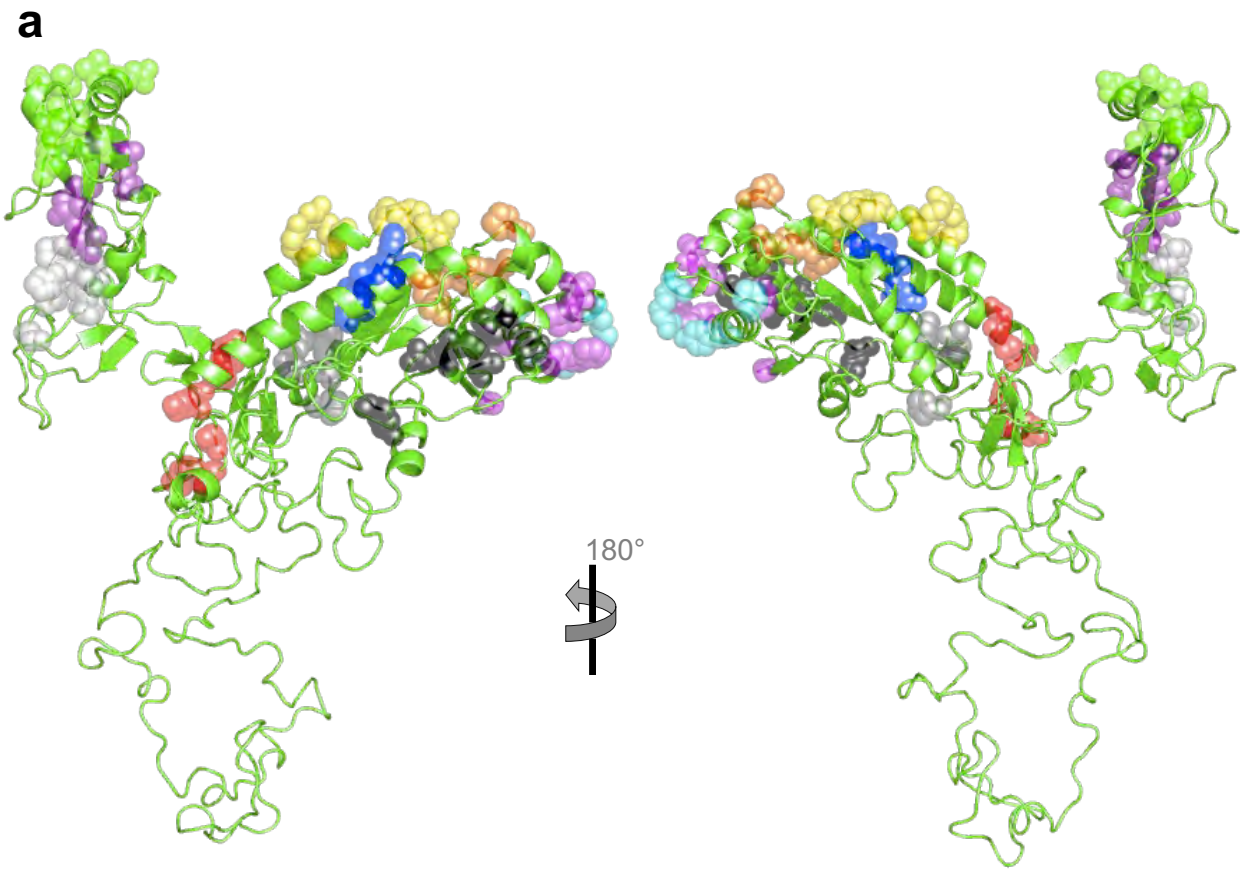
			damaged foci; decreases recruitment of TFIIH complex to lesion sites.	
799	F799A	F599	Reduces DNA repair activity; abolishes single-stranded DNA binding; reduces binding to homoduplex DNA; greatly reduces localization at DNA damaged foci; decreases recruitment of TFIIH complex to lesion sites.	(114)
848	W848A	W649	When combined with L851A and L855A, it abolishes interaction with CETN2 and reduces NER activity in vitro.	(70)
851	L851A	L652	When combined with W848A and L855A, it abolishes interaction with CETN2 and reduces NER activity in vitro.	(70)
855	L855A	L656	When combined with W848A and L851A, it abolishes interaction with CETN2 and reduces NER activity in vitro.	(70)

5. Structural analyses of XPC genetic variations.

Here, we have mapped the XPC variants compiled above (**Tables 1-2 and S1-S5** in **SI File S5**) onto the 3-D models of human XPC constructed using SwissModel as well as AlphaFold. Out of 246 unique variants, 168 variants were spread over the central domains of the protein: 53 were in the TGD domain, 11 in BHD1, 12 in BHD2, 39 in BHD3 and 53 were in the linker region within TGD (residues 331-520).

To aid visual inspections and characterizations, we have grouped the mutations into 21 clusters. The clusters are made up of mutated residues that are physically close to one another within the protein structure; the proximity may indicate a functional connection among the residues in the cluster. **Figure 4** shows each cluster in the context of the overall structure of XPC. **Figures 5-11** show close-up views of each cluster along with the corresponding lists of mutations. A Pymol pse file in which the clusters and residues can be easily found as individual selections is also available (see **SI File S6**).

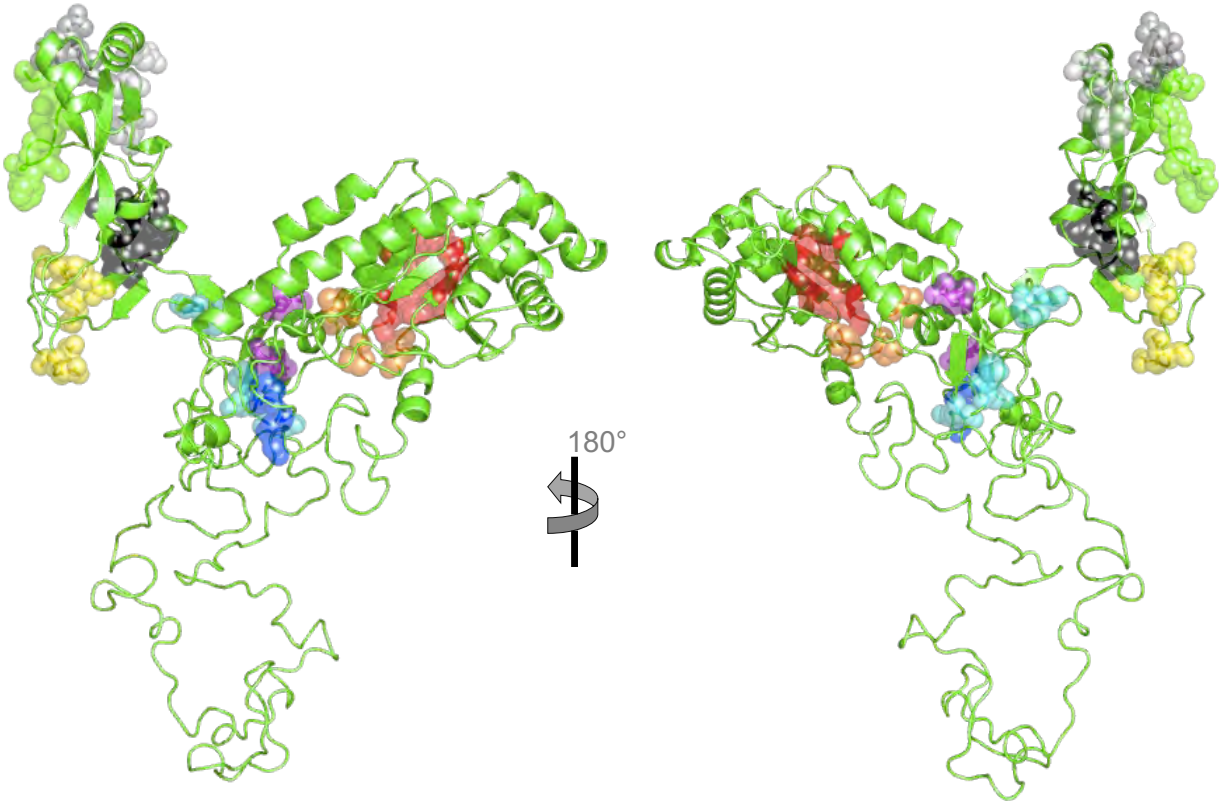
Next, we used the structural models together with the sequence conservation among various orthologs (**Figures 2 and S5**) (62) to qualitatively assess the putative roles of each residue and the impacts of their mutations. The degree of sequence conservation computed by



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

Figure 4. Locations of clusters mapped on the overall XPC homology model structure. (a-b) The overall structure is in ribbon representation. The residues discussed in each cluster are shown in spheres with colors indicated as in the table.

b



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

ConSurf is also indicated in **Table S1** in **SI File S5**. Generally, a score of +3 (high impact) is assigned for variants at strongly conserved positions that have a published indication to be pathological or disruptive of the structure/function; +2 (moderate impact) for variants in conserved positions that are likely to be structurally impactful; +1 (low impact) for variants in conserved positions yet whose impact is likely small; 0 (uncertain impact) for those in most conserved positions whose structural impact is not obvious (e.g., solvent-exposed) or likely to be tolerated; -1 (no impact) for the variants that synonymous mutations while being conserved or those that are not conserved with low predicted impacts. These scores were also used to indicate our predictions for residues in **Figures 5 – 10**.

Additionally, we quantitatively assessed the potential impacts of the variants using FoldX (63) and SDM (64). FoldX predicts ΔΔG (kcal/mol) based on polar and hydrophobic desolvation, hydrogen bond energy, and electrostatic energy for residues located at the interface of oligomeric proteins. Here, $\Delta\Delta G = \Delta G_{variant} - \Delta G_{wildtype}$. Mutations that are predicted to have a negative ΔΔG are generally considered stabilizing, while positive ΔΔG are considered destabilizing (115). The SDM server predicts ΔΔG based on local structural environment (solvent accessibility, secondary structure, hydrogen bond energy, packing density, and residue depth). These probability scores are derived from environment-specific substitution tables (ESSTs) calculated from protein family sequence alignments. Here, $\Delta\Delta G = \Delta G_{wildtype} - \Delta G_{variant}$. Mutations with a negative ΔΔG are considered destabilizing, while positive ΔΔG are considered stabilizing (64). The ΔΔG's calculated by FoldX and SDM showed a statistically significant and a moderate correlation which is negative as expected ($r = -0.496$, $N=118$, $p<0.001$). The correlation between our scores and those of FoldX were also significant with a moderately positive correlation as expected ($r = 0.490$, $N=118$, $p<0.001$). The correlation between our

assessment and SDM-calculated ΔΔG values were also statistically significant with a moderately negative correlation as expected, albeit with a bit lower correlation coefficient (r= -0.366, N=118, p<0.001) (**Figure S8**). The scores from our assessment, FoldX and SDM for each residue are included in **Table S1** in **SI File S5**.

Below, we discuss individual variants by the clusters, primarily focusing on the variants that are predicted or shown to be of high or moderate impact and those that are likely to form solvent-exposed hydrophobic surfaces, suggesting novel inter-molecular interfaces. The low- or uncertain-impact residues within a cluster are denoted with a dagger (†) and are described in Supplementary Information (**Text S1** in **SI File S1**). The residue numbers in () indicate the corresponding residue number in Rad4. Residues that are before (/) are conserved in a majority of orthologs.

Cluster 1 is located around α1 and α2 of TGD, and contains R191Q, A193V†, K195T†, K199N and G200E† (**Figure 5**).

The conserved **R191** (R129; R in most orthologs: R(K/E)) contacts the minor groove of the DNA: its positively charged guanidino group forms ionic interactions with the negatively charged phosphate backbone. The mutation R191Q may weaken this interaction as the positive charge is absent in Gln.

K199 (R137) belongs to α2; 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 K/R indicating its functional importance in XPC. **K199N** mutation may weaken the interaction between α2 and the DNA due to the loss of the positive charge in the residue.

Cluster 2 is located near the α2 and β8 helices of TGD and includes L214I†, A215S†, F218V†, V578F†, and D580Y (**Figures 5 and S4**).

D580 (D374) is located at the end of β8 and is conserved as D/E in all orthologs. The D580 forms a salt bridge with another highly conserved residue, R597 (R391 in Rad4) in a partially embedded environment. The carboxyl group of D580 also forms a hydrogen bond with the backbone amide of T582 of α11 and the N1-H in the indole ring of W603 of α13. T582 and W603 are conserved in most orthologs (T(S/G) and W(Y/F/M/L/V). Taken together, the mutation D580Y likely have a negative impact on the conserved packing of α11 and α12 in the TGD domain.

Cluster 3 is located near the end of the α2 helix and the beginning of the α3 helix of TGD, and contains R220Q†, C224Y, P227Q†, R307W, and R307Q (**Figure 5**).

C224 (I162) in α2 is represented as Cys in most orthologs: C(L/I/A). The side chain of **C224** packs against H300, L303 and L304 in α8 and is also close to R307 (see below), which are all relatively well conserved: H300 (G268) is H/Q(/Y/F/T/A), L303 (V270) is L(/V/I/F/C/A), and L304 (A271) is L(/A/I/C/S). C224Y may be too bulky and destabilize the packing between these α2 and α8 within TGD.

R307 (R274) is located in α8, partially buried, and identically conserved in all orthologs. R307 forms H-bonds with N221 (N159) and S225 (N163) of α2, and the carbonyl groups of I572 and D573 of β7. N221 is identically conserved in most orthologs N(/S/V/D); the solvent-exposed S225 is represented as polar residues: N/S(/D/C/L/R). R307W may be too bulky and hydrophobic, disrupting the packing of α2, α8 and β7. R307Q lacks the guanidino group to maintain H-bonds and is shorter, which may destabilize the structure.

Cluster 4 is located where the α3, 4, and 5 helices of TGD interact, and contains S235P†, I236V†, W260R, F302S, and G550C† (**Figure 5**).

W260 (L209) in α5 faces the loop between α3 and α4. Rad4's L209 makes hydrophobic contacts to stabilize the packing. This position is conserved mostly as W(L/I/V/H) in most orthologs and is facing solvent along with other highly conserved aliphatic/hydrophobic residues (P238 and F241), which suggests that it may form a binding interface with other proteins in cells (**Figure 12**). AlphaFold models these residues as more buried but still a significant part of them are solvent-exposed. W260R has the potential to maintain the hydrophobic packing, and thus its destabilizing impact could be relatively small. FoldX and SDM predicts it to be slightly stabilizing. If W260 in mammalian XPC were to participate in forming an intermolecular hydrophobic interface with other proteins in cells, we envision that W260R can have a more significant impact.

F302 (F269) is located in α8 and conserved as F in almost all orthologs: F(/L/A). F302 forms hydrophobic contact with L257 (C206), V258 (M207) and F261 (W210) of α5, L306 (L273) of α8, L532 (C317) of β4, and C547 (P332) of β5. All the contact residues are also highly conserved as hydrophobic. F302S may disrupt the hydrophobic packing of α8 with α5, β4, and β5. F302S was shown to impair the recruitment of XPC to 408 nm laser-induced damage spots in cells (90). This residue also falls in the region (residues 154-331) that was previously reported to interact with XPA (111).

Cluster 5 is located near α4 and α5 of TGD, and contains R240C, P246S†, and Y252C (**Figure 6**).

R240 (K178) is in α4 and exposed to solvent. This position is represented by a basic residue R/K/H/(G/Y/D) in most orthologs. R240 is within a H-bond distance with Y252 (D201;

see below) where both are solvent-exposed. The impact of R240C is not clear for the structural integrity and DNA binding function of XPC, but it may be important in other protein-protein interactions as discussed below with Y252. FoldX and SDM predict it to be slightly destabilizing.

Y252 (D201) located in α5, is exposed to solvent, and not conserved (represented as D/Y/L/V/N/G/K) in orthologs) although it tends to be hydrophobic in higher organisms. Y252 forms a contiguous solvent-exposed interface along with W260 (cluster 4), P238 (P176), and F241 (V179) which are all highly conserved, along with R240 (K178) and K259 (E208) which are conserved as basic residues in higher eukaryotes. This hydrophobic/basic interface may function in other protein-protein interactions in higher eukaryotes (**Figure 12**). Although the impact of Y252C on the XPC structural integrity is unclear in our analysis, FoldX and SDM predict it to be slightly destabilizing.

Cluster 6 spans α7, β2, and α8 of TGD. It contains E270D†, F287C†, S291C†, R293Q†, and E296K† (**Figure 6**). No residue in this cluster is predicted to have a significant impact on protein structure or function.

Cluster 7 is located in a hydrophobic interface that packs α11 and α14 with the central β sheet (especially β3, β4 and β5) in the TGD domain. It contains R314Q, W531A, W542A, Y585C, and E621Q as well as E539G† and K620T† (**Figures 6 and 11**).

Notably, **R314**, **W531**, **W542, Y585, E621,** as well as nearby residues E533, R584 (/K) are nearly identically conserved in all orthologs as they pack in a hydrophobic core laced with multiple embedded salt bridges and H-bonds. **R314** (R281) is located in β3, **W531** (W316) and E533 (E318) in β4, **W542** (W327) in β5, R584 (R378) and **Y585** (Y379) in α11, and **E621** (E421) in α14. Altogether, they stabilize the packing between β3/β4/β5 and α11/ α14:

R314 and R584 make multiple salt bridges with E533 and **E621**. R314 also forms an Arg-Trp cation-pi stacking between the Arg guanidino group and **W542** aromatic ring. R584 also forms a salt bridge with E624 (Y424) and packs with **Y585**. **Y585** H-bonds with E533 and **W531** (see below). R314Q (91) lacks the guanidino group to weaken its interaction with W542 and could disrupt the salt bridge network E533 and E621. W542A would also disrupt the Arg-Trp stacking with R314. Functional studies show that W542A moderately lowers accumulation at the lesion and DNA binding (110). Y585C may be too small and likely destabilizes the protein and results in partial lack of function, as manifested in some XP-C patients (105). E621Q may weaken Glu's salt bridges interaction with R314 and R584.

W531 (W316) is located in β4 and is identically conserved in all orthologs. W531 makes hydrophobic contact with **Y585** (Y379) of α11, I528 (P313) of the loop between α9 and β4, M566 (L360) of the loop between α10 and β7, V569 (V363) of β7, V316 (I283) of β3 and Y559 (S345) between β6/α10. The amine group in the indole group of W531 forms an H-bond with the hydroxyl group of Y585 (Y379) of α11. Y585 is identically conserved in all orthologs, as mentioned above. I528 is conserved as Pro in lower orthologs P(/V/T/C/D/A/S/R). Both M566 and V569 are nonpolar in all orthologs as M(/L/V/I) and V(/I), respectively. V316 is also nonpolar in all orthologs V(/I). Y559 is not conserved (Y/S/F/R/H/E), and the contact between W531 and Y559 is not retained in the AlphaFold model, which modeled the region around residues 553-566 differently. Functional studies show that W531A moderately lowers accumulation at the lesion and DNA binding (110). W531A may be too small and weaken its hydrophobic interactions.

Cluster 8 is located in the long, unstructured linker between β3 and β4 of TGD, and contains K332N†, P334H, and P334S (**Figure 6**).

P334 (L300/N301) is not conserved (P/S/T/Q/A/L/I). Yet, P334H is reported as a pathogenic variant in patients (47,91-95,109). Functional studies show p.Pro334His mutation prevents the stimulation of Ogg1 glycosylase because it impairs the interaction of XPC with Ogg1 as well as with the p62 subunit of TFIIH (109). A recent study shows that the linker region (residues 331-520) containing P334 interacts with the deacetylated histone H3 tail and participates in promoting GG-NER in DDB-independent manner (82). The impact of P334H or P334S on the structure of Rad4 itself or on its DNA binding is likely to be mild, but it may have different functions through other inter-molecular interactions.

Cluster 9 is located near the loop connecting β6 and β7 of TGD, and contains P321Q, T557N†, A562T†, P565T†, and T567N† (**Figure 7**). This loop is longer in lower eukaryotes and contains a short α helix in Rad4. **P321** (P288) is located at the loop after β3 and conserved as a Pro in most orthologs P(/A/V). P321 resides in the hydrophobic core of TGD, forming hydrophobic contacts with Y568 (Y362) of β7, H206 (H144) of α2 and L210 (L148) and L213 (L151) of α2. Y568, H206 and L210 are identically conserved in all orthologs. L213 is also highly conserved as L(/Q/F). P321Q may disrupt the hydrophobic packing with β3, β7 and α2.

Cluster 10 is located near β8 and α13 of TGD. It contains D575E†, E605K†, R608K†, S612I†, F614S†, and M615T† (**Figure 7).** No residue in this cluster is predicted to have a significant impact on protein structure or function.

Cluster 11 is located around α12 and α14 of TGD near the TGD-BHD1 hinge. It contains W589Q, R594H, P633T, P633L, and P633S (**Figure 7**). **W589** (M383) is located in α12 and is conserved as aromatic in most orthologs W(/Y/F/M/C/L). **R594** (R388) is located in α12, and conserved as R in most orthologs: R(/S/L). W589 and R594 interact with each other while both are partially exposed to solvent. In the SwissModel, W589 also forms cation-pi interaction with

R594. W589Q and W589S may weaken these interactions. Additionally, R594 is in close distance from P645 (P445; Cluster 12) in the loop connecting α0 and β1 of BHD1. P645 is Pro in all but one ortholog (/E). R594H or R594C may disrupt the key interactions important for the hinge interface between TGD and BHD1.

P633 (G433) is located at the end of TGD domain before the α0 of BHD1 starts. It is exposed to solvent and is conserved as P in most orthologs (/G/S). P633 may play a role in coupling the TGD and BHD1 domains. In yeast, there is a Gly, but the junction instead presents W382 (in place of V588 in humans) to pack α12 and α0. P633L/T/S may disturb the coupled movement of TGD and BHD1 in DNA binding.

Cluster 12 is located around α0 and β1 of BHD1 near the TGD-BHD1 hinge. It contains Y641C, Y641H, P645H, and L646M† (**Figure 7**). **Y641** (L441) is located in α0 of BHD1 and is conserved as Y or F in most orthologs Y(/F/L/M). Y641 forms highly conserved hydrophobic interactions with L634 (I434) and P635 (P435) between α14 of TGD and α0 of BHD1, L649 (L449) of β1 of BHD1, C670 (L469) of β2, and V675 (V481) of β3. L634 and L649 are conserved as hydrophobic in all orthologs L(/I/M). P635 is Pro in all orthologs. C670 is C(/L/I/V). V675 is V(/I). Since Y641 is involved in a highly conserved hydrophobic interface that stabilizes the BHD1 core (connecting α0/β1, β2, and β3) as well as the interaction between BHD1 and TGD, Y641C and Y641H may be disruptive for the protein structure.

P645 (P445) and **L646** (Y446)† are is located in the highly conserved β-turn (XPC 642-646, Rad4 443-446) connecting α0 and β1 that is part of the inter-domain hinge between TGD and BHD1. P645 is partially exposed to solvent and conserved as Pro in all but one ortholog (/E). In XPC, P645 (P445) forms hydrophobic contact with M590 (N384) and potentially also with R594 (R388; Cluster 11) of α12 of TGD, similarly as in Rad4. P645H may disrupt the β–turn

and destabilizes the packing between the α0-β1 loop of BHD1 and α12 of TGD necessary for DNA binding.

Cluster 13 is near β3 and α2 of BHD1. It includes A658D, E662K†, R671Q†, R671H†, R671C†, E673K†, and A674V (**Figure 8**). **A658** (I458) is conserved as nonpolar in all orthologs A(/V/I/G). A658 (I458) is located in β1' of BHD1 and packs against β3' of BHD1 in the BHD1-BHD2 interdomain hinge through hydrophobic interactions. Most of the residues in this interface are highly conserved as hydrophobic/aromatic in higher eukaryotes compared with lower eukaryotes, consisting of Y660 (K460), Y704 (K510), W736 (E535), H682 (A488), L684 (L490) as well as A658 (I458). The interdomain packing is especially extensive in the AlphaFold model, indicating that the BHD1-BHD2 interaction may be tighter in human XPC than in Rad4. A658D may disrupt the packing of the interdomain hinge.

A674 (K480) is located in β3 and conserved as nonpolar A(/L/N/P/K). A674 forms hydrophobic contact with I666 (E465) of β2. A674 and I666 are nonpolar in higher orthologs in contrast to yeasts. Together with nearby Y676 (Y482) and Y669 (Y468) that are highly conserved (ConSurf score of 9's), these residues seem to form a solvent-exposed hydrophobic surface that may play a role in intermolecular interactions in higher organisms (**Figure 12**). The impact of A674V is unclear.

Cluster 14 is located near α1, β3, and β3' of BHD2. It contains T689M, T689K, W690S, L691P, P703L, F733A, and T738A (**Figure 8**). **T689** (Q495) is located in α1 and conserved as polar in most orthologs T(/Q/N/R/K/H/I). Q495 can form H-bond with S492 of α1 and its amide group hydrogen bond to the phosphate groups of dA15 (W15). Similarly, in XPC, T689 could interact with the phosphate backbone of DNA as well as S686 (S492) of α1. S686 is conserved as Ser in almost all orthologs (/T/G). T689M (91) may be too nonpolar for this position though

FoldX and SDM predict it to be slightly stabilizing. The impact of T689K is not clear and likely benign.

W690 (W496) is located in α1 and identically conserved in all orthologs. W690 forms extensive hydrophobic interactions with multiple, highly conserved residues: T738 (T537) of BHD2 β3', H685 (K491) of BHD2 β3'/α1 loop, F733 (Y532) of β3/α2 loop, and R695 (R501) of β1'. The amine of the indole group H-bonds with the backbone carbonyl of S686 of α1. T738 is identically conserved in all but one ortholog (/M). H685 is conserved as basic in most orthologs H(/K/R/Q/F/Y). F733 is conserved as aromatic in all orthologs Y(/F). R695 is identically conserved in all but one ortholog (/L). W690S is reported as a pathogenic mutation in humans that cause XP (91,96). 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 (110-112) . The W690S mutation is shown to increase the mobility of XPC in cells (13) and drastically lower the *in vivo* stability and the *in vitro* DNA binding affinity for both undamaged and damaged DNA (113). Recently, it was reported that this mutation negatively impacts the association of XPC with Sox2 transcription factor but not with Oct4 (45). Ser in W690S may be too small and polar and disrupt hydrophobic packing that stabilizes BHD2 core involving β3', β3, α1, and β1'. These findings may also underscore the general importance of BHD2 in the DNA recognition function of XPC.

L691 (Y497) is located in α1 and is conserved as Leu in all higher orthologs and Tyr in all yeasts. While L691 does not make significant contacts with other residues in our current model, it is located in an interface facing a group of highly conserved, hydrophobic residues in BHD3, including Y741 (Y540), Y760 (E560), F762 (F562), L763 (A563), and M766 (M566). This packing is more extensive in the AlphaFold model. In Rad4, Y497 is located in α1 of BHD2

and packs against α0 and β0 of BHD3 in the interdomain hinge. Y497 forms an H-bond with E560 of β0 and hydrophobic contact with M566 of α0 and F562 of β0 of BHD3. E560 is conserved as acidic in all yeasts E/D and Tyr in other orthologs. M566 is Met in all but one ortholog. F562 is conserved as F(/W/Y) in all orthologs. It may be possible that these interactions formed with Y497 in Rad4 are not possible with L691 in higher organisms. Nevertheless, L691P may destabilize the α1 helix and thus interfere with the hydrophobic packing that interfaces BHD2 and BHD3. FoldX and SDM predict L691P to be destabilizing.

P703 (C509) is located in the β1'-β2 loop, partially exposed to solvent, and conserved as Pro in most orthologs except yeasts: P(/C/A/H). P703 packs against a conserved F733 (Y532, see below) of β3. The impact of P703L (106) is not clear though FoldX and SDM predict it to be slightly destabilizing.

F733 (Y532) is located in β3 and conserved as Tyr or Phe in all orthologs. F733 forms hydrophobic contact with W690 (W496; see above) of α1 and P703 (C509; see above). W690 is identically conserved in all orthologs. P703 is also highly conserved as P(/C/A/H). F733 also makes hydrophobic contacts with E701 (S507) in the SwissModel or V697 (L503) in the AlphaFold model. E701 is conserved as Glu in all higher orthologs (/S/A), and V697 is also conserved as small aliphatic residues, V(L/I), in all orthologs. The site-directed mutagenesis of F733A results in an almost complete loss of DNA binding, reduced interaction with single-stranded DNA, and lowered ability to localize to sites of DNA damage (110,112). F733A may be too small and weaken the hydrophobic packing.

T738 (T537) is located in β3' and identically conserved in all but one ortholog (/M). T738's methyl group forms hydrophobic contact with W690 (W496; see above) of α1 and its hydroxy group H-bonds with R695 (R501) of β1'. R695 is identically conserved in all but one

ortholog (/L). T738A (107) may destabilize the hydrophobic core of BHD2 and have an impact similar to W690S. We note that this missense mutation is also speculated to allow partial functionality of XPC, which may explain the unusual late onset XP observed in the patient (107).

L691, P703 and F733 in the Cluster 14 form a conserved, solvent-exposed hydrophobic surface in BHD3 together with V697, L699 and V702. This surface may function as a inter-molecular interface (**Figure 12**).

Cluster 15 is located in the β2-β3 loop of BHD2. It includes R713H†, R715Q†, E721K†, Q723†, and D729† (**Figure 8**). No residue in this cluster is predicted to have a significant impact on protein structure or function.

Cluster 16 is located near the β turn-β0 loop and the α1-β2 loop of BHD3. It includes A746D†, V747M†, D748G†, G749W, K750N†, and D789Y† (**Figure 8**). **G749** (G549) located at the BHD3 β-turn is conserved as Gly in all but one ortholog (/N). G749W may be too bulky and disrupt the β-turn located in the beginning of BHD3.

Cluster 17 is located near the β-turn region before the β0 of BHD3 and includes P752T†, R753L, N754A, E755K, F756A, G757W, and G757R (**Figure 9**).

R753 (K553) is partially exposed to solvent and conserved as basic in most orthologs R(/K/H/T). R753 forms an H-bond with the carbonyl backbone of C790 (F590) of β2 in the SwissModel whereas it interacts with D789 (E589) in the AlphaFold model. D789 is mostly conserved as Asp in most orthologs: D(/E/Q/K/P). R753L may thus destabilize the interaction of β2 and the β-turn region.

N754 (N554), **E755** (T555), **F756** (F556), and **G757** (G557) are highly conserved and form a β-turn motif. This β-turn has been proposed to play a role as a sensor for DNA damage (110). N754 and F756 are predicted to contact DNA together similarly (114) by interacting with

one face of the flipped out dA17 (W17) adenine group opposite of F797 (F597), P806 (P607) and I794 (V594). **N754** (N554) is identically conserved in all orthologs, while **F756** (F556) is conserved as aromatic in all but one ortholog (/R). In Rad4, N554 Nδ2 group also forms a single H-bond with the dA17 (W17) N6 group. This H-bond could also be found when the flipped-out nucleotide was a thymine as seen in the structure of Rad4 bound to TTT/TTT mismatches (PDB ID: 2QSG and 2QSH). N754's backbone carbonyl group also H-bonds with the backbone amide of **G757** (G557) to form the β-turn. Functional studies indicate that N754 is important to mobility and the ability to accumulate at the lesion (114). N754A may abolish these interactions and thus impact Rad4 function. N754A may disrupt interaction with the flipped out dA17 (W17) adenine group. F756A may be too small and weaken interactions with the flipped out dA17 (W17). G757W/R may lack flexibility and disrupt the β-turn. **E755** (T555) is conserved as Glu in higher eukaryotes (/D/T/A/F) and exposed to solvent. E755 has been suggested to have DNA-repellent properties. thereby providing the necessary mobility for dynamic lesion recognition (110). In functional studies, E755K reduces nuclear mobility and causes a significant defect in GG-NER (110) while F756A is only partially impaired in localizing to UV lesions but fully able to recruit TFIIH (114). E755K may have DNA-attractive properties and disrupt the mobility of the β-turn.

Cluster 18 is located near the β-turn-β0 loop, β1, and α2 of BHD3. It includes Q742L†, Q742E†, P743L, P743T, C771G, C771Y, and E815K† (**Figure 9**).

P743 (P542) is conserved as Pro in most orthologs: P(/R/L/A) together with the adjacent P744 (P543). P743 forms van der Waals contact with the Y741 (Y540) which is also highly conserved. P743L/T may disrupt the packing.

C771 (C571) is located in β1 and conserved as Cys in most orthologs: C(/T/G/A; ConSurf score 9). C771 forms hydrophobic contact with conserved hydrophobic residues such as I812 (V613), P768 (P568) and V751 (I551). This packing is more extensive in the AlphaFold model than in the SwissModel. C771G and C771Y (108) may destabilize these interactions. C771Y missense mutation is found in an XP patient who was compound heterozygous in the XPC gene with another intronic variant (XPC c.780-1G>A).

Cluster 19 is located near α1 and α2 of BHD3. It includes L779V, R781H†, V782M†, A783D, L786M, D819E†, and V820M (**Figure 9**). **L779** (A579) is located in α1 and conserved as nonpolar in all but one ortholog L(/A/I/S). L779 forms hydrophobic packing contacts to stabilize the core of the BHD3 domain. The contacting residues are L774 (V574) and L776 (N576) of β1, Y811 (I612) of β3, and L821 (I622) of α2. These residues are mostly nonpolar in orthologs even if some of them are partially exposed to solvent: L774 L(/V/M/I/C); L776 L(/N/F/V/H/K/S/N); Y811 Y(V/I/A/F); L821 L(/I/V/F) and also potentially W825 W(Y/I/L/A) according to the AlphaFold model. L779V may impact the BHD3 core packing. FoldX and SDM also predict this to be destabilizing. Surrounding residues (V772, L774, L776, L822 and W825) are likely to form a hydrophobic interface for other intermolecular interactions (**Figure 12**).

A783 (A583) is located in α1 of BHD3 and conserved as Ala in most orthologs A(/C/S). A783 forms hydrophobic contact with Y811 (I612) of β3 and L821 (I622) of α2; the carbonyl backbone of A783 H-bonds with the amide backbone of I788. Y811 is aromatic Y(/F) in some higher orthologs and nonpolar in lower orthologs V(/I/A). L821 is nonpolar in most orthologs L(/I/V/F). In the AlphaFold model, A783 contacts C790 (F590) of β2, instead of Y811, which is also highly conserved. A783D may disrupt the hydrophobic packing of α1, β3, and α2. FoldX and SDM predict it to be destabilizing.

L786 (L586) is located in α1 of BHD3, partially exposed to solvent, and conserved as Leu in most orthologs: L(/I/F). L786 forms hydrophobic contact with V820 (A621) and A824 (A625) of α2, as well as with I788 (V588). V820 is nonpolar in most orthologs (V/A/I/D/T). A824 is Ala in most orthologs (/V/I). I788 is Ile in most orthologs, I(/V/K). The impact of L786M is unclear. FoldX and SDM predict it to be slightly destabilizing.

V820 (A621) is located in α2, exposed to solvent, and conserved as nonpolar in most orthologs V/A/I/(D/T). It also contacts L786 (L586) as described above. The impact of V820M is unclear. We note that **V820** (A621), **L786** (L586), I788 (V588) and F817 (L618) form a solvent-exposed hydrophobic surface, which may participate in an intermolecular interaction with another protein (**Figure 12**).

Cluster 20 includes F797A, D798H†, F799A, G802S, S804F, H805N†, and P806L (**Figure 10**). These residues constitute the tip of a long β-hairpin in BHD3 which inserts into the DNA duplex to 'open' the lesion site in the Rad4-DNA crystal structures (24-26). In most other orthologs, this tip is one residue shorter than in yeast (e.g., human XPC (797)FDFHGGYSHP(806) vs. Rad4 (597)FKFERGSTVKP(607)). The β-hairpin approaches the DNA from the major groove and extends through the width of the Watson-Crick double helix. The aliphatic portions of side-chain and backbone groups of the β-hairpin pack with the base pairs flanking the flipped-out ones (e.g., dA15 & dG18 in the undamaged strand (W) and dC8 and dT11 in the damaged strand (Y)) with the plane of the hairpin at a steep angle to the base pair planes that are still a part of duplexed DNA.

F797, **F799**, and **P806** are all identical in Rad4 (F597, F599, and P607, respectively) and located in the BHD3 β-hairpin. These residues in Rad4 primarily form van der Waals contacts with the undamaged strand of the flipped-out ('open') DNA segment. **F797** (F597) and **F799**

(F599) are both located in β2 and conserved as Phe in almost all orthologs (/W and /Y, respectively). **P806** (P607) is located in β3 and conserved as Pro in most orthologs (/A). In Rad4, F597 and P607 pack with one face of the flipped out dA17 (W17) adenine group, together with V594 (I794 in humans); both F597 and F599 side chains pack with the dA15 (W16) ribose group. Functional studies indicate that both F797A and F799A impaired the ability to recruit TFIIH to the lesion site and caused severe GG-NER defect (114). F797A, F799A and P806L may impact these DNA binding roles.

G802 (S603) is conserved as Gly in most orthologs: G(/S/R/K) but is a serine in Rad4 and is located at the very tip of the β-hairpin (β2- β3) of BHD3, making up tight a β-turn. G802S is likely to disrupt this β-turn, especially when the β-hairpin length is one residue short in higher organisms than in yeast. FoldX and SDM predict G802S to be destabilizing.

S804 (V605) is located in β3 and conserved as a small residue in most orthologs (S/C/T/V/A/G). In Rad4, V605 packs with the ribose groups of the flipped-out dA16 (W16) and dA17 (W17) nucleotides. S804 may interact with DNA in a similar manner as V605. S804F may be too bulky for this position. FoldX and SDM predict it to be destabilizing.

Cluster 21 is located near β1, β2, and β3 of BHD3, and includes Q773K†, A793T†, V807M†, T808N†, and I812V† (**Figure 10**). No residue in this cluster is predicted to have a significant impact on protein structure or function.

DISCUSSION

Here, we have compiled the XPC variants from various tumor databases (ClinVar, International Cancer Genome Consortium,and TumorPortal) and clinical/biochemical literature. Out of a total of 246 unique variants, 115 variants were found in the central, structured domains of the protein:

Cluster 1

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	R191Q	R129	8
ICGC	A193V	V131	8
ICGC	K195T	S133	8
CL	K199N	R137	9
ICGC	G200E	K138	7

Cluster 2

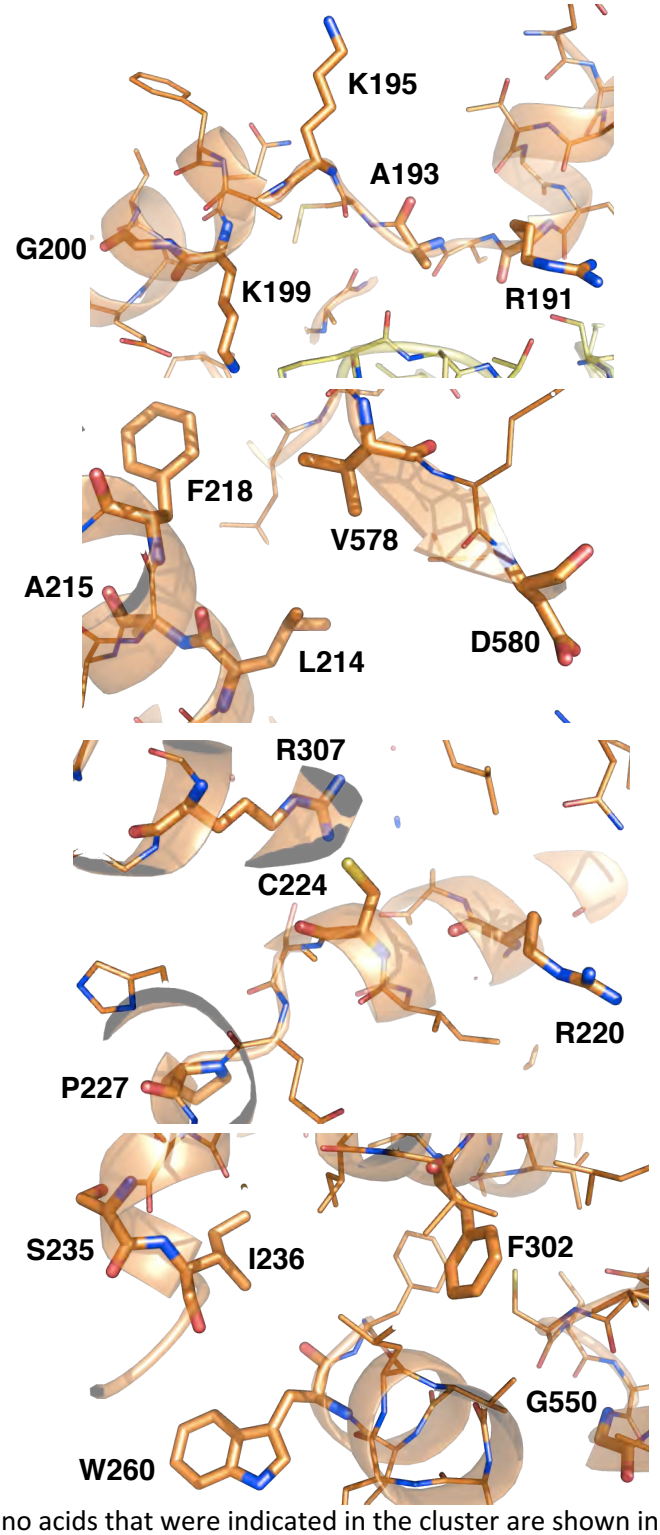
Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	L214I	M152	8
ICGC	A215S	V153	9
ICGC	F218V	F156	7
TP	V578F	C372	7
ICGC	D580Y	D374	9

Cluster 3

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	R220Q	R158	8
TP, ICGC	C224Y	I162	8
ICGC	P227Q	K165	7
ICGC	R307W	R274	9
ICGC	R307Q	R274	9

Cluster 4

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	S235P	N173	9
ICGC, CL	I236V	L174	7
ICGC	W260R	L209	9
L	F302S	F269	8
TP	G550C	L335	8



Figures 5. Local views for Clusters 1-21. The amino acids that were indicated in the cluster are shown in stick representation and all other residues are in line representation. Residues are colored according to their domains (TGD in orange, BHD1 in magenta, BHD2 in turquoise, and BHD3 in maroon) except nitrogen (blue) and oxygen (red). In tables, the XPC mutations predicted to have a high impact (+++; 3) are shown as red bold; moderate impact (++; 2) red; low impact (+; 1) orange; uncertain impact (+/-; 0) purple; and no impact (-) black (see **Table S1**; also see **File S6.** XPC models.pse). Corresponding Rad4 residues and ConSurf scores for evolutionary conservation are also indicated for each residue.

Cluster 5

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
CL	R240C	K178	5
CL	P246S	H184	6
CL	Y252C	D201	3

Cluster 6

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
TP, ICGC	E270D	V223	5
CL	F287C	A254	8
CL	S291C	G258	8
TP, ICGC	R293Q	G260	7
ICGC	E296K	D263	7

Cluster 7

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
L	R314Q	R281	9
L	**W531A**	W316	9
CL	E539G	S324	7
L	**W542A**	W327	9
L	**Y585C**	Y379	9
ICGC	K620T	Y420	7
ICGC	E621Q	E421	9

Cluster 8

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	K332N	S299	5
L	**P334H**	N301	4
CL	**P334H**	N301	4
CL	P334S	N301	4

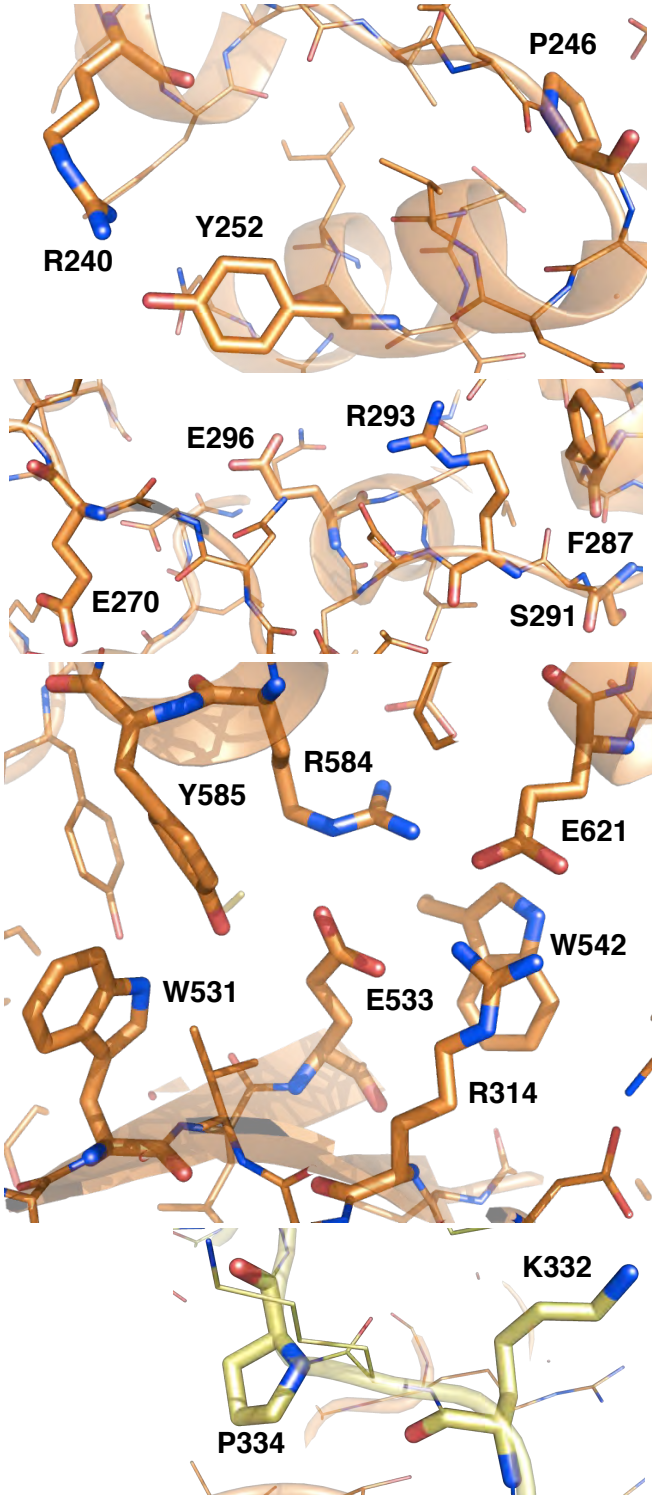


Figure 6. Local structural views for Clusters 5-8. In Cluster 7, E539 and K620 are omitted in the image for clarity.

Cluster 9

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	P321Q	P288	9
ICGC	T557N	L343	5
ICGC	A562T	A348	9
ICGC	P565T	M359	9
ICGC	T567N	R361	5

Cluster 10

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	D575E	K369	5
ICGC	E605K	K404	6
CL	R608K	T407	5
ICGC	S612I	R411	7
CL	F614S	N/A	4
ICGC	M615T	N/A	5

Cluster 11

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	W589S	M383	9
TP, ICGC	W589Q	M383	9
ICG	R594H	R388	9
CL	R594C	R388	9
ICGC	P633T	G433	9
ICGC	P633L	G433	9
ICGC	P633S	G433	9

Cluster 12

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
TP	Y641C	L441	8
CL	Y641H	L441	8
TP	P645H	P445	9
ICGC	L646M	Y446	8

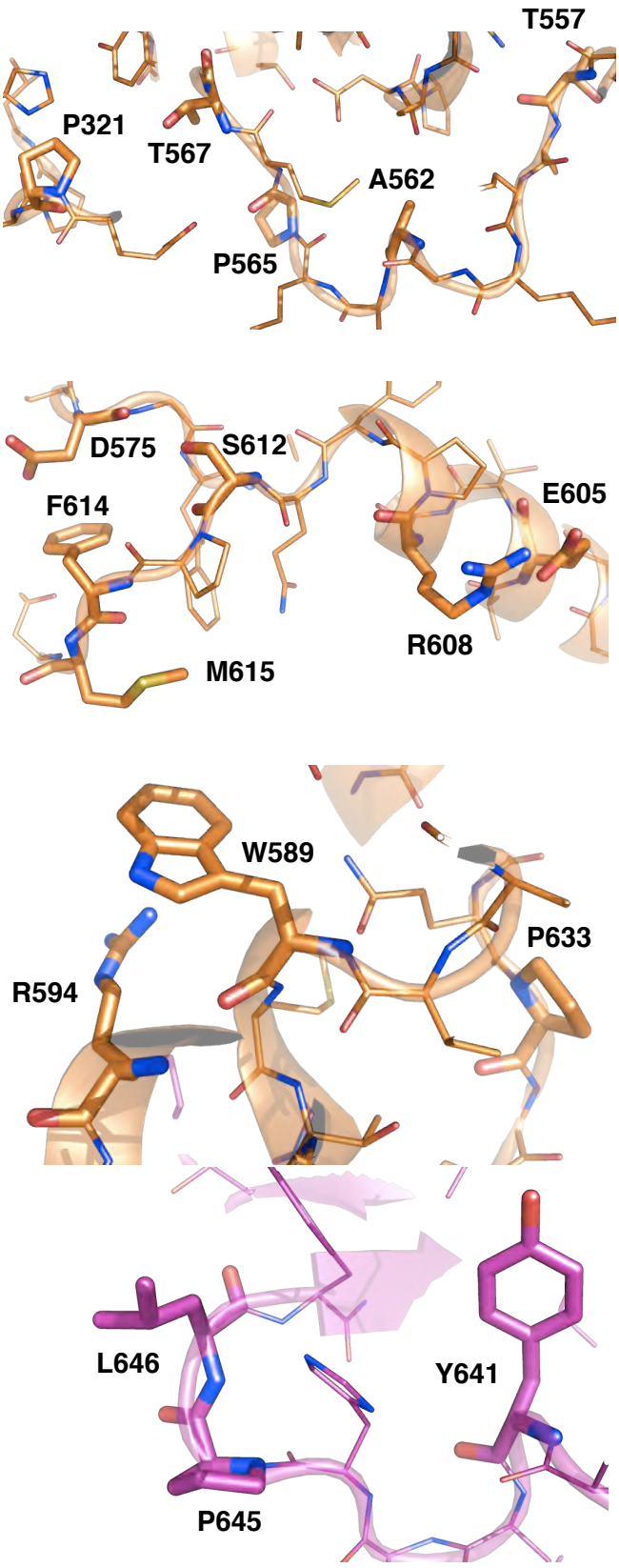
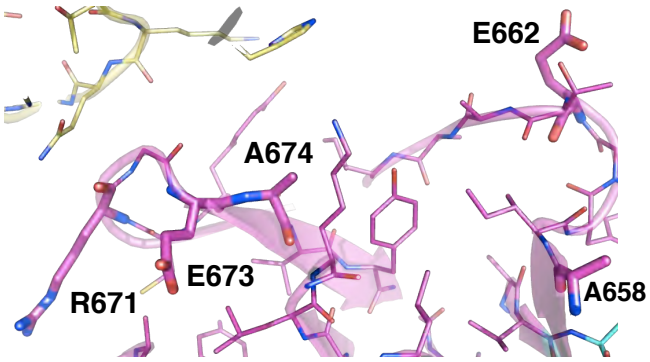


Figure 7. Local structural views for Clusters 9-12.

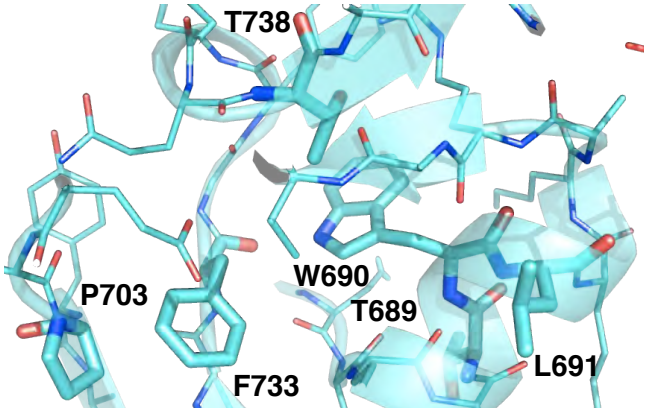
Cluster 13

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	A658D	I458	9
ICGC, TP, CL	E662K	G462	5
ICGC	R671H	K477	9
L	R671Q	K477	9
ICGC, CL	R671C	K477	9
TP, ICGC	E673K	L479	9
CL	A674V	K480	7



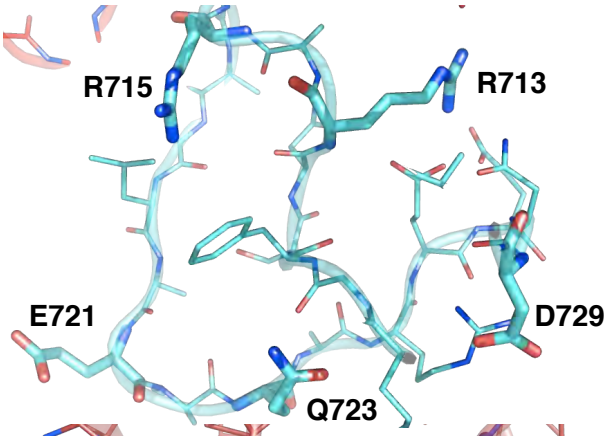
Cluster 14

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
CL, L	T689M	Q495	9
ICGC	T689K	Q495	9
L	**W690S**	W496	9
ICGC	L691P	Y497	9
L	P703L	C509	9
L	**F733A**	Y532	8
L	**T738A**	T537	8



Cluster 15

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
CL	R713H	R519	8
TP, ICGC	R715Q	K521	9
CL	E721K	N/A	7
ICGC	Q723K	N/A	6
ICGC	D729V	D528	9



Cluster 16

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC, TP	A746D	A545	9
ICGC	V747M	S546	8
ICGC	D748G	S548	8
TP	G749W	G549	9
ICGC	K750N	E550	8
ICGC	D789Y	E589	9

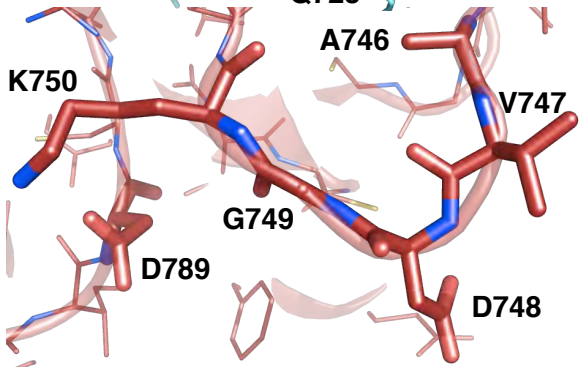
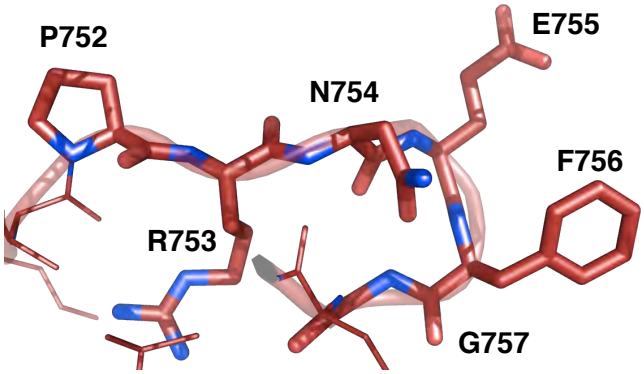


Figure 8. Local structural views for Clusters 13-16.

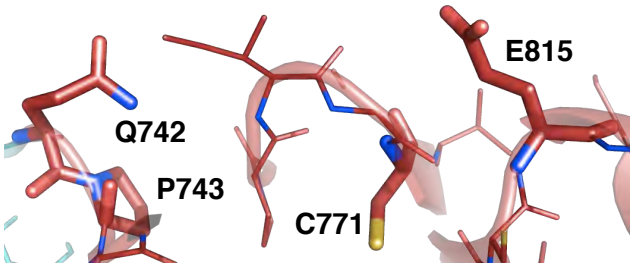
Cluster 17

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	P752T	T552	9
ICGC	R753L	K553	9
L	**N754A**	N554	9
L	**E755K**	T555	7
L	**F756A**	F556	7
ICGC	G757W	G557	9
ICGC, TP	G757R	G557	9



Cluster 18

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	Q742L	I541	9
ICGC	Q742E	I541	9
TP	P743L	P542	9
ICGC	P743T	P542	9
L	C771Y	C571	9
ICGC	C771G	C571	9
ICGC	E815K	K616	8



Cluster 19

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	L779V	A579	9
ICGC	R781H	K581	9
ICGC	V782M	A582	8
ICGC	A783D	A583	9
ICGC	L786M	L586	9
ICGC	D819E	E620	7
ICGC	V820M	A621	5

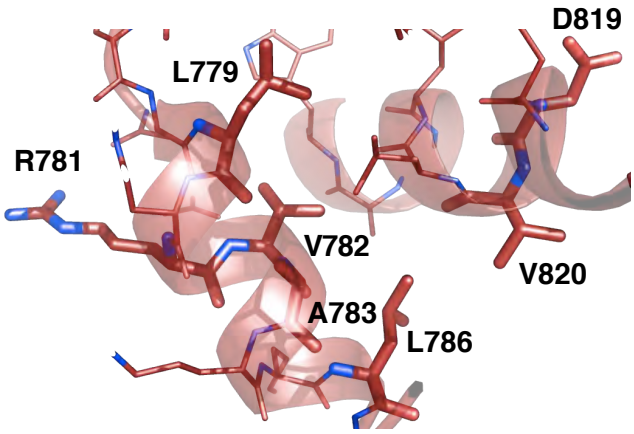
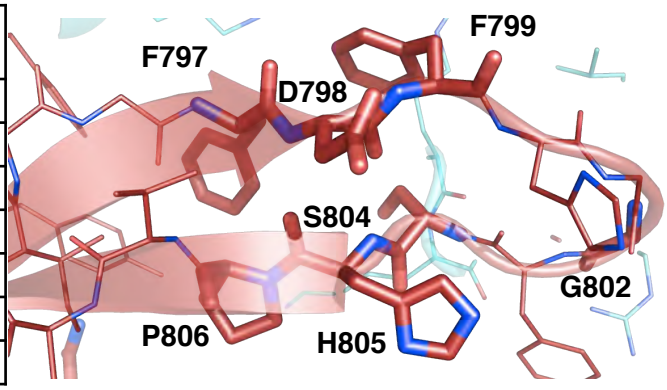


Figure 9. Local structural views for Clusters 17-19.

Cluster 20

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
L	**F797A**	F597	9
ICGC	D798H	K598	9
L	**F799A**	F599	8
CL	G802S	S603	9
ICGC	S804F	V605	9
ICGC	H805N	K606	9
ICGC	P806L	P607	8



Cluster 21

Source	XPC Mutation	Corresponding Rad4 residue	ConSurf Score
ICGC	Q773K	L573	8
ICGC	A793T	A593	9
ICGC	V807M	V608	8
ICGC	T808N	L609	6
ICGC	I812V	V613	5

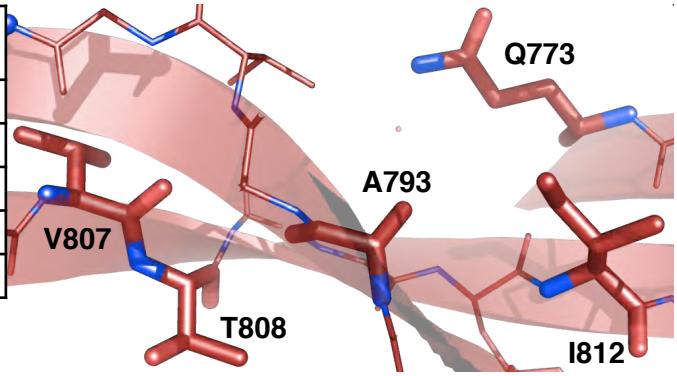


Figure 10. Local structural views for Clusters 20-21.

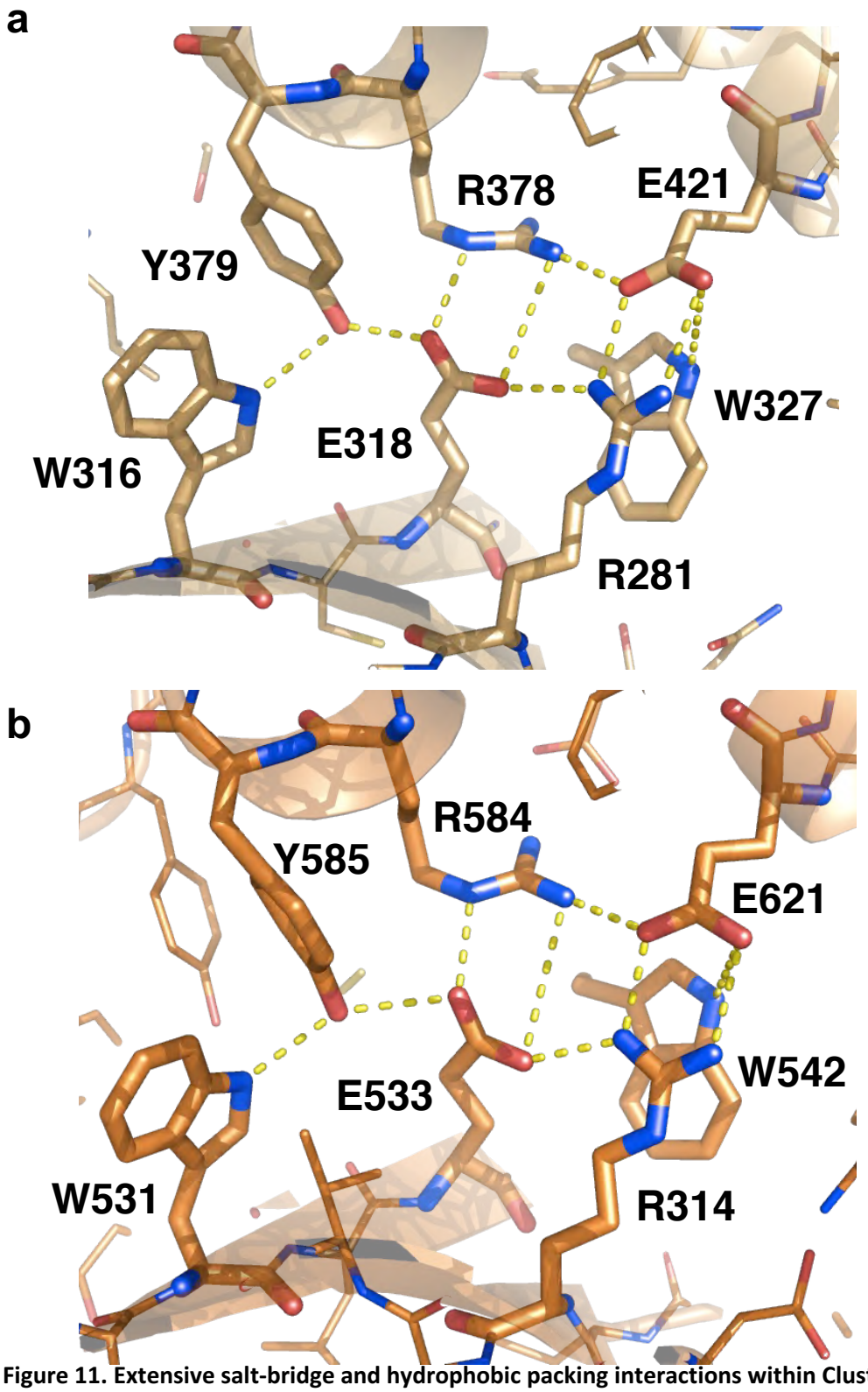


Figure 11. Extensive salt-bridge and hydrophobic packing interactions within Cluster 7. Packing of α11 and α14 with the central β sheet (β3, β4 and β5) in TGD is stabilized by highly conserved residues forming multiple salt bridges and H-bonds, as shown for **(a)** Rad4 and **(b)** XPC.

a

Site	Participating Residues
A	**W260**, P238, F241, **R240**, **Y252**
B	**A674**, Y669, Y676
C	**L691**, **P703**, **F733**, V697, L699, V702
D	**L779**, V772, L774, L776, L822, W825
E	**L786**, **V820**, I788, F817

b

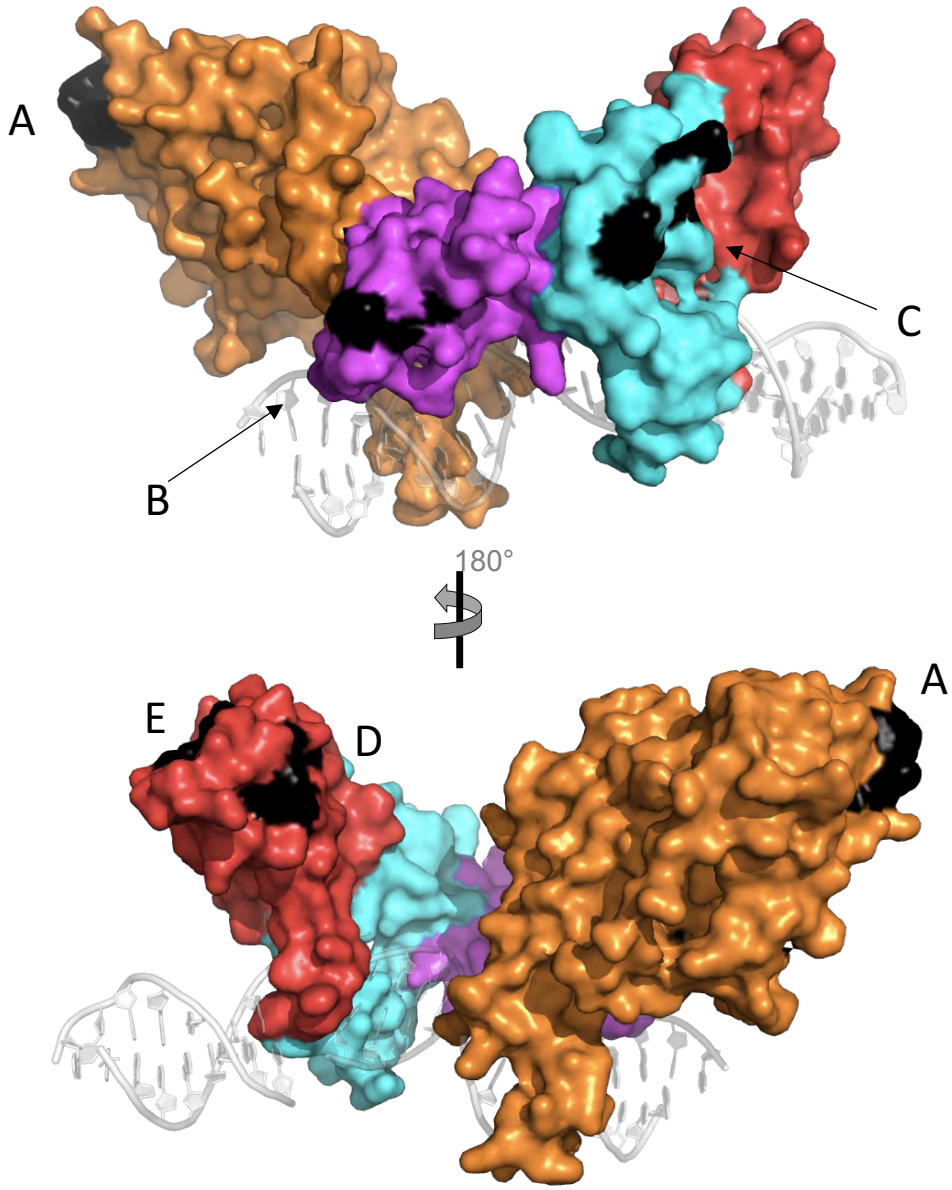


Figure 12. Putative intermolecular interfaces of XPC. (a) The residues that participate in each putative interface are listed. The residues whose variants were discussed in the text are bold; other highly conserved residues are also indicated. **(b)** Putative interfaces are indicated as black and labeled according to **(a)** on a surface representation of the XPC homology structure. TGD is in orange, BHD1 magenta, BHD3 cyan, and BHD3 red. A DNA model is shown in semi-transparent wires.

53 were in the TGD domain, 11 in BHD1, 12 in BHD2, 39 in BHD3 while 53 were in the flexible linker region within TGD (residues 331-520). We then mapped these variants on 3-D models of human XPC constructed by SwissModel as well as by AlphaFold, focusing on the ~115 variants in the structured regions where the models have higher confidence. In addition to visual structural inspections and assessments, the potential impact of each variant was quantitatively examined using ΔΔG calculations by FoldX and SDM. Evolutionary conservation for each residue was also assessed by multiple sequence alignments over 966 XPC orthologs using Clustal Omega and ConSurf.

Our analyses revealed that many of the known pathological mutations are at highly conserved positions and are likely destabilize the protein's structure (Y585C, W690S, and C771Y). One exception is P334H, one of XP mutations, which falls in the TGD linker and is not conserved even among mammals. The mutants that are shown to be impactful in biochemical assays also are in highly conserved, structural positions (W531A, W542A, W690S, F733A, N754A, E755K, F756A, F797A, F799A).

Most other variants lack relevant studies to probe their pathogenicity or biological impacts. However, the variants at highly conserved positions that are also predicted to destabilize the protein's structure are generally predicted to have high/moderate impacts. Among 41 variants that we predicted to have high/moderate impact (scores 2 or 3), 31 of them are predicted to be structure-destabilizing by both FoldX and SDM while 8 of them were predicted to be destabilizing by either FoldX or SDM. Only 1 variant, P806L, which we assessed to be impactful as it may participate in damaged DNA-binding, is not predicted to be structure-destabilizing by FoldX and SDM. However, we also note that the correlations among different scores are not

perfect, and certain variants deviate from the general trend. Our analyses help recognize which these residues/variants are and guide future studies.

Additionally, our analyses revealed that several highly conserved, hydrophobic regions are also solvent-exposed. These regions are formed around residues W260 (including P238, F241 as well as R240 and Y252), A674 (including Y669 and Y676), L691 (including P703, F733 as well as V697, L699 and V702), L779 (including V772, L774, L776, L822 and W825) and L786 (including V820, I788 and F817) and may indicate novel intermolecular interfaces that are yet to be characterized (**Figure 12**).

We hope that our report will help understand the function of each residue in biological processes, including NER while we wait for a high-resolution structure of human XPC protein along with more detailed structural, biochemical, cellular and clinical studies of the genetic variations.

Acknowledgements. This work was funded by National Science Foundation (NSF) (MCB-2131806 to J.-H.M) and National Institute of Health (1R15GM147899 to J.-H.M). We also thank the members of the Min group.

Declaration of interest statement. Authors declare no conflict of interest.

Appendix. Supplementary material is found online.

REFERENCES

1. Hanawalt, P.C. (2003) Four decades of DNA repair: from early insights to current perspectives. *Biochimie*, **85**, 1043-1052.
2. Friedberg, E.C., Aguilera, A., Gellert, M., Hanawalt, P.C., Hays, J.B., Lehmann, A.R., Lindahl, T., Lowndes, N., Sarasin, A. and Wood, R.D. (2006) DNA repair: from molecular mechanism to human disease. *DNA Repair (Amst)*, **5**, 986-996.
3. Scharer, O.D. (2013) Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol*, **5**, a012609.

4. Marteijn, J.A., Lans, H., Vermeulen, W. and Hoeijmakers, J.H. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol*, **15**, 465-481.
5. Fousteri, M. and Mullenders, L.H. (2008) Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res*, **18**, 73-84.
6. Lans, H., Hoeijmakers, J.H.J., Vermeulen, W. and Marteijn, J.A. (2019) The DNA damage response to transcription stress. *Nat Rev Mol Cell Biol*, **20**, 766-784.
7. Yeh, J.I., Levine, A.S., Du, S., Chinte, U., Ghodke, H., Wang, H., Shi, H., Hsieh, C.L., Conway, J.F., Van Houten, B. *et al.* (2012) Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair. *Proc Natl Acad Sci U S A*, **109**, E2737-2746.
8. Matsumoto, S., Fischer, E.S., Yasuda, T., Dohmae, N., Iwai, S., Mori, T., Nishi, R., Yoshino, K., Sakai, W., Hanaoka, F. *et al.* (2015) Functional regulation of the DNA damage-recognition factor DDB2 by ubiquitination and interaction with xeroderma pigmentosum group C protein. *Nucleic Acids Res*, **43**, 1700-1713.
9. Matsumoto, S., Cavadini, S., Bunker, R.D., Grand, R.S., Potenza, A., Rabl, J., Yamamoto, J., Schenk, A.D., Schubeler, D., Iwai, S. *et al.* (2019) DNA damage detection in nucleosomes involves DNA register shifting. *Nature*, **571**, 79-84.
10. Apelt, K., Lans, H., Scharer, O.D. and Luijsterburg, M.S. (2021) Nucleotide excision repair leaves a mark on chromatin: DNA damage detection in nucleosomes. *Cell Mol Life Sci*, **78**, 7925-7942.
11. Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D. and Hoeijmakers, J.H. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell*, **2**, 223-232.
12. Wood, R.D. (1999) DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochimie*, **81**, 39-44.
13. Hoogstraten, D., Bergink, S., Ng, J.M., Verbiest, V.H., Luijsterburg, M.S., Geverts, B., Raams, A., Dinant, C., Hoeijmakers, J.H., Vermeulen, W. *et al.* (2008) Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. *J Cell Sci*, **121**, 2850-2859.
14. Puumalainen, M.R., Ruthemann, P., Min, J.H. and Naegeli, H. (2016) Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation. *Cell Mol Life Sci*, **73**, 547-566.
15. Gillet, L.C. and Scharer, O.D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev*, **106**, 253-276.
16. Yokoi, M., Masutani, C., Maekawa, T., Sugasawa, K., Ohkuma, Y. and Hanaoka, F. (2000) The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *J Biol Chem*, **275**, 9870-9875.
17. Uchida, A., Sugasawa, K., Masutani, C., Dohmae, N., Araki, M., Yokoi, M., Ohkuma, Y. and Hanaoka, F. (2002) The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH. *DNA Repair (Amst)*, **1**, 449-461.
18. Li, C.L., Golebiowski, F.M., Onishi, Y., Samara, N.L., Sugasawa, K. and Yang, W. (2015) Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. *Mol Cell*, **59**, 1025-1034.
19. Kusakabe, M., Onishi, Y., Tada, H., Kurihara, F., Kusao, K., Furukawa, M., Iwai, S., Yokoi, M., Sakai, W. and Sugasawa, K. (2019) Mechanism and regulation of DNA damage recognition in nucleotide excision repair. *Genes Environ*, **41**, 2.
20. van Eeuwen, T., Shim, Y., Kim, H.J., Zhao, T., Basu, S., Garcia, B.A., Kaplan, C.D., Min, J.H. and Murakami, K. (2021) Cryo-EM structure of TFIIH/Rad4-Rad23-Rad33 in damaged DNA opening in nucleotide excision repair. *Nat Commun*, **12**, 3338.

21. Riedl, T., Hanaoka, F. and Egly, J.M. (2003) The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J*, **22**, 5293-5303.
22. 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. (2013) Comparative analysis of interaction of human and yeast DNA damage recognition complexes with damaged DNA in nucleotide excision repair. *J Biol Chem*, **288**, 10936-10947.
23. Krasikova, Y.S., Rechkunova, N.I., Maltseva, E.A., Anarbaev, R.O., Pestryakov, P.E., Sugasawa, K., Min, J.H. and Lavrik, O.I. (2013) Human and yeast DNA damage recognition complexes bind with high affinity DNA structures mimicking in size transcription bubble. *J Mol Recognit*, **26**, 653-661.
24. Min, J.H. and Pavletich, N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570-575.
25. Paul, D., Mu, H., Zhao, H., Ouerfelli, O., Jeffrey, P.D., Broyde, S. and Min, J.H. (2019) Structure and mechanism of pyrimidine-pyrimidone (6-4) photoproduct recognition by the Rad4/XPC nucleotide excision repair complex. *Nucleic Acids Res*, **47**, 6015-6028.
26. Chen, X., Velmurugu, Y., Zheng, G., Park, B., Shim, Y., Kim, Y., Liu, L., Van Houten, B., He, C., Ansari, A. *et al.* (2015) Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat Commun*, **6**, 5849.
27. Kong, M., Liu, L., Chen, X., Driscoll, K.I., Mao, P., Bohm, S., Kad, N.M., Watkins, S.C., Bernstein, K.A., Wyrick, J.J. *et al.* (2016) Single-Molecule Imaging Reveals that Rad4 Employs a Dynamic DNA Damage Recognition Process. *Mol Cell*, **64**, 376-387.
28. Velmurugu, Y., Chen, X., Slogoff Sevilla, P., Min, J.H. and Ansari, A. (2016) Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc Natl Acad Sci U S A*, **113**, E2296-2305.
29. Chakraborty, S., Steinbach, P.J., Paul, D., Mu, H., Broyde, S., Min, J.H. and Ansari, A. (2018) Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the Rad4 nucleotide excision repair complex. *Nucleic Acids Res*, **46**, 1240-1255.
30. Paul, D., Mu, H., Tavakoli, A., Dai, Q., Chen, X., Chakraborty, S., He, C., Ansari, A., Broyde, S. and Min, J.H. (2020) Tethering-facilitated DNA 'opening' and complementary roles of beta-hairpin motifs in the Rad4/XPC DNA damage sensor protein. *Nucleic Acids Res*, **48**, 12348-12364.
31. Paul, D., Mu, H., Tavakoli, A., Dai, Q., Chakraborty, S., He, C., Ansari, A., Broyde, S. and Min, J.H. (2021) Impact of DNA sequences on DNA 'opening' by the Rad4/XPC nucleotide excision repair complex. *DNA Repair (Amst)*, **107**, 103194.
32. Tavakoli, A. and Min, J.H. (2022) Photochemical modifications for DNA/RNA oligonucleotides. *RSC Adv*, **12**, 6484-6507.
33. DiGiovanna, J.J. and Kraemer, K.H. (2012) Shining a light on xeroderma pigmentosum. *J Invest Dermatol*, **132**, 785-796.
34. Francisco, G., Menezes, P.R., Eluf-Neto, J. and Chammas, R. (2008) XPC polymorphisms play a role in tissue-specific carcinogenesis: a meta-analysis. *Eur J Hum Genet*, **16**, 724-734.
35. Melis, J.P., Luijten, M., Mullenders, L.H. and van Steeg, H. (2011) The role of XPC: implications in cancer and oxidative DNA damage. *Mutat Res*, **728**, 107-117.
36. Fleming, N.D., Agadjanian, H., Nassanian, H., Miller, C.W., Orsulic, S., Karlan, B.Y. and Walsh, C.S. (2012) Xeroderma pigmentosum complementation group C single-nucleotide polymorphisms in the nucleotide excision repair pathway correlate with prolonged progression-free survival in advanced ovarian cancer. *Cancer*, **118**, 689-697.
37. Jin, B., Dong, Y., Zhang, X., Wang, H. and Han, B. (2014) Association of XPC polymorphisms and lung cancer risk: a meta-analysis. *PLoS One*, **9**, e93937.

38. Said, R., Bougatef, K., Setti Boubaker, N., Jenni, R., Derouiche, A., Chebil, M. and Ouerhani, S. (2019) Polymorphisms in XPC gene and risk for prostate cancer. *Mol Biol Rep*, **46**, 1117-1125.
39. Malik, S.S., Zia, A., Rashid, S., Mubarik, S., Masood, N., Hussain, M., Yasmin, A. and Bano, R. (2020) XPC as breast cancer susceptibility gene: evidence from genetic profiling, statistical inferences and protein structural analysis. *Breast Cancer*, **27**, 1168-1176.
40. Nasrallah, N.A., Wiese, B.M. and Sears, C.R. (2022) Xeroderma Pigmentosum Complementation Group C (XPC): Emerging Roles in Non-Dermatologic Malignancies. *Front Oncol*, **12**, 846965.
41. Le May, N., Mota-Fernandes, D., Velez-Cruz, R., Iltis, I., Biard, D. and Egly, J.M. (2010) NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Mol Cell*, **38**, 54-66.
42. Le May, N., Egly, J.M. and Coin, F. (2010) True lies: the double life of the nucleotide excision repair factors in transcription and DNA repair. *J Nucleic Acids*, **2010**.
43. Pascucci, B., D'Errico, M., Parlanti, E., Giovannini, S. and Dogliotti, E. (2011) Role of nucleotide excision repair proteins in oxidative DNA damage repair: an updating. *Biochemistry (Mosc)*, **76**, 4-15.
44. Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I. and Tjian, R. (2011) A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell*, **147**, 120-131.
45. Cattoglio, C., Zhang, E.T., Grubisic, I., Chiba, K., Fong, Y.W. and Tjian, R. (2015) Functional and mechanistic studies of XPC DNA-repair complex as transcriptional coactivator in embryonic stem cells. *Proc Natl Acad Sci U S A*, **112**, E2317-2326.
46. Ho, J.J., Cattoglio, C., McSwiggen, D.T., Tjian, R. and Fong, Y.W. (2017) Regulation of DNA demethylation by the XPC DNA repair complex in somatic and pluripotent stem cells. *Genes Dev*, **31**, 830-844.
47. Bidon, B., Iltis, I., Semer, M., Nagy, Z., Larnicol, A., Cribier, A., Benkirane, M., Coin, F., Egly, J.M. and Le May, N. (2018) XPC is an RNA polymerase II cofactor recruiting ATAC to promoters by interacting with E2F1. *Nat Commun*, **9**, 2610.
48. Zebian, A., Shaito, A., Mazurier, F., Rezvani, H.R. and Zibara, K. (2019) XPC beyond nucleotide excision repair and skin cancers. *Mutat Res*, **782**, 108286.
49. Petrović, D. and Zlatović, M. (2015) Modeling Human Serum Albumin Tertiary Structure To Teach Upper-Division Chemistry Students Bioinformatics and Homology Modeling Basics. *Journal of Chemical Education*, **92**, 1233-1237.
50. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*, **38**, W695-699.
51. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
52. Guex, N., Peitsch, M.C. and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*, **30 Suppl 1**, S162-173.
53. Benkert, P., Biasini, M. and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, **27**, 343-350.
54. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L. *et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, **46**, W296-W303.
55. Studer, G., Tauriello, G., Bienert, S., Biasini, M., Johner, N. and Schwede, T. (2021) ProMod3-A versatile homology modelling toolbox. *PLoS Comput Biol*, **17**, e1008667.
56. Sali, A. and Blundell, T.L. (1993) COMPARATIVE PROTEIN MODELING BY SATISFACTION OF SPATIAL RESTRAINTS. *Journal of Molecular Biology*, **234**, 779-815.

57. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.
58. Webb, B. and Sali, A. (2016) Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci*, **86**, 2 9 1-2 9 37.
59. Laskowski, R.A., Macarthur, M.W., Moss, D.S., & Thornton, J.M. . (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography.*, **26**, 283.
60. Rost, B. (1999) Twilight zone of protein sequence alignments. *Protein Eng*, **12**, 85-94.
61. Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.
62. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*, **44**, W344-350.
63. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res*, **33**, W382-388.
64. Pandurangan, A.P., Ochoa-Montano, B., Ascher, D.B. and Blundell, T.L. (2017) SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res*, **45**, W229-W235.
65. Krieger, E. and Vriend, G. (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics*, **30**, 2981-2982.
66. Legerski, R. and Peterson, C. (1992) Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C. *Nature*, **359**, 70-73.
67. Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P.J., Bootsma, D. *et al.* (1994) Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *EMBO J*, **13**, 1831-1843.
68. Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y. and Hanaoka, F. (2001) Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *J Biol Chem*, **276**, 18665-18672.
69. Popescu, A., Miron, S., Blouquit, Y., Duchambon, P., Christova, P. and Craescu, C.T. (2003) Xeroderma pigmentosum group C protein possesses a high affinity binding site to human centrin 2 and calmodulin. *J Biol Chem*, **278**, 40252-40261.
70. Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K. and Hanaoka, F. (2005) Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Mol Cell Biol*, **25**, 5664-5674.
71. Nishi, R., Sakai, W., Tone, D., Hanaoka, F. and Sugasawa, K. (2013) Structure-function analysis of the EF-hand protein centrin-2 for its intracellular localization and nucleotide excision repair. *Nucleic Acids Res*, **41**, 6917-6929.
72. Thompson, J.R., Ryan, Z.C., Salisbury, J.L. and Kumar, R. (2006) The structure of the human centrin 2-xeroderma pigmentosum group C protein complex. *J Biol Chem*, **281**, 18746-18752.
73. Yang, A., Miron, S., Mouawad, L., Duchambon, P., Bloquit, Y. and Craescu, C.T. (2006) Flexibility and plasticity of human centrin 2 binding to the xeroderma pigmentosum group c protein (XPC) from nuclear excision repair. *Biochemistry*, **45**, 3653-3663.
74. Charbonnier, J.-B., Renaud, E., Miron, S., Le Du, M.H., Blouquit, Y., Duchambon, P., Christova, P., Shosheva, A., Rose, T., Angulo, J.F. *et al.* (2007) Structural, thermodynamics, and cellular characterization of human centrin 2 interaction with Xeroderma pigmentosum group C protein. *Journal of Molecular Biology*, **373**, 1032-1046.

75. Okuda, M., Kinoshita, M., Kakumu, E., Sugasawa, K. and Nishimura, Y. (2015) Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor XPC. *Structure*, **23**, 1827-1837.
76. Zhang, E.T., He, Y., Grob, P., Fong, Y.W., Nogales, E. and Tjian, R. (2015) Architecture of the human XPC DNA repair and stem cell coactivator complex. *Proc Natl Acad Sci U S A*, **112**, 14817-14822.
77. Xie, Z., Liu, S., Zhang, Y. and Wang, Z. (2004) Roles of Rad23 protein in yeast nucleotide excision repair. *Nucleic Acids Res*, **32**, 5981-5990.
78. den Dulk, B., van Eijk, P., de Ruijter, M., Brandsma, J.A. and Brouwer, J. (2008) The NER protein Rad33 shows functional homology to human Centrin2 and is involved in modification of Rad4. *DNA Repair (Amst)*, **7**, 858-868.
79. Mu, H., Geacintov, N.E., Min, J.-H., Zhang, Y. and Broyde, S. (2017) Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo a pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chemical Research in Toxicology*, **30**, 1344-1354.
80. Anantharaman, V., Koonin, E.V. and Aravind, L. (2001) Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold. *Hum Mol Genet*, **10**, 1627-1630.
81. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *Fems Microbiology Reviews*, **29**, 231-262.
82. Kusakabe, M., Kakumu, E., Kurihara, F., Tsuchida, K., Maeda, T., Tada, H., Kusao, K., Kato, A., Yasuda, T., Matsuda, T. *et al.* (2022) Histone deacetylation regulates nucleotide excision repair through an interaction with the XPC protein. *iScience*, **25**, 104040.
83. Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S. and Hanaoka, F. (2001) A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev*, **15**, 507-521.
84. Hilton, B., Gopal, S., Xu, L., Mazumder, S., Musich, P.R., Cho, B.P. and Zou, Y. (2016) Dissociation Dynamics of XPC-RAD23B from Damaged DNA Is a Determining Factor of NER Efficiency. *PLoS One*, **11**, e0157784.
85. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.
86. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*, **46**, D1062-D1067.
87. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S. and Getz, G. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, **505**, 495-501.
88. Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D. and Ferretti, V. (2019) The International Cancer Genome Consortium Data Portal. *Nat Biotechnol*, **37**, 367-369.
89. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. *et al.* (2015) 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*, **17**, 405-424.

90. Qiao, B., Ansari, A.H., Scott, G.B., Sak, S.C., Chambers, P.A., Elliott, F., Teo, M.T., Bentley, J., Churchman, M., Hall, J. *et al.* (2011) In vitro functional effects of XPC gene rare variants from bladder cancer patients. *Carcinogenesis*, **32**, 516-521.
91. Rivera-Begeman, A., McDaniel, L.D., Schultz, R.A. and Friedberg, E.C. (2007) 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 (Amst)*, **6**, 100-114.
92. Li, L., Bales, E.S., Peterson, C.A. and Legerski, R.J. (1993) Characterization of molecular defects in xeroderma pigmentosum group C. *Nat Genet*, **5**, 413-417.
93. Martin-Morales, L., Rofes, P., Diaz-Rubio, E., Llovet, P., Lorca, V., Bando, I., Perez-Segura, P., de la Hoya, M., Garre, P., Garcia-Barberan, V. *et al.* (2018) Novel genetic mutations detected by multigene panel are associated with hereditary colorectal cancer predisposition. *PLoS One*, **13**, e0203885.
94. Pugh, J., Khan, S.G., Tamura, D., Goldstein, A.M., Landi, M.T., DiGiovanna, J.J. and Kraemer, K.H. (2019) Use of Big Data to Estimate Prevalence of Defective DNA Repair Variants in the US Population. *JAMA Dermatol*, **155**, 72-78.
95. Ramirez-Calvo, M., Garcia-Casado, Z., Fernandez-Serra, A., de Juan, I., Palanca, S., Oltra, S., Soto, J.L., Castillejo, A., Barbera, V.M., Juan-Fita, M.J. *et al.* (2019) Implementation of massive sequencing in the genetic diagnosis of hereditary cancer syndromes: diagnostic performance in the Hereditary Cancer Programme of the Valencia Community (FamCan-NGS). *Hered Cancer Clin Pract*, **17**, 3.
96. Chavanne, F., Broughton, B.C., Pietra, D., Nardo, T., Browitt, A., Lehmann, A.R. and Stefanini, M. (2000) Mutations in the XPC gene in families with xeroderma pigmentosum and consequences at the cell, protein, and transcript levels. *Cancer Res*, **60**, 1974-1982.
97. Huang, W.Y., Berndt, S.I., Kang, D., Chatterjee, N., Chanock, S.J., Yeager, M., Welch, R., Bresalier, R.S., Weissfeld, J.L. and Hayes, R.B. (2006) Nucleotide excision repair gene polymorphisms and risk of advanced colorectal adenoma: XPC polymorphisms modify smoking-related risk. *Cancer Epidemiol Biomarkers Prev*, **15**, 306-311.
98. Qiu, L., Wang, Z., Shi, X. and Wang, Z. (2008) Associations between XPC polymorphisms and risk of cancers: A meta-analysis. *Eur J Cancer*, **44**, 2241-2253.
99. Strom, S.S., Estey, E., Outschoorn, U.M. and Garcia-Manero, G. (2010) Acute myeloid leukemia outcome: role of nucleotide excision repair polymorphisms in intermediate risk patients. *Leuk Lymphoma*, **51**, 598-605.
100. Guillem, V.M., Cervantes, F., Martinez, J., Alvarez-Larran, A., Collado, M., Camos, M., Sureda, A., Maffioli, M., Marugan, I. and Hernandez-Boluda, J.C. (2010) XPC genetic polymorphisms correlate with the response to imatinib treatment in patients with chronic phase chronic myeloid leukemia. *Am J Hematol*, **85**, 482-486.
101. Dai, Q.S., Hua, R.X., Zeng, R.F., Long, J.T. and Peng, Z.W. (2014) XPC gene polymorphisms contribute to bladder cancer susceptibility: a meta-analysis. *Tumour Biol*, **35**, 447-453.
102. Zhao, Z., Zhang, A., Zhao, Y., Xiang, J., Yu, D., Liang, Z., Xu, C., Zhang, Q., Li, J. and Duan, P. (2018) The association of polymorphisms in nucleotide excision repair genes with ovarian cancer susceptibility. *Biosci Rep*, **38**.
103. Xie, C., Zhao, J., Hua, W., Tan, P., Chen, Y., Rui, J., Sun, X., Fan, J., Wei, X., Xu, X. *et al.* (2019) Effect of XPC polymorphisms on the response to platinum-based chemotherapy: a meta-analysis. *Onco Targets Ther*, **12**, 3839-3848.
104. Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C.F., Whaley, R. and Klein, T.E. (2021) An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther*, **110**, 563-572.

105. Fassihi, H., Sethi, M., Fawcett, H., Wing, J., Chandler, N., Mohammed, S., Craythorne, E., Morley, A.M., Lim, R., Turner, S. *et al.* (2016) Deep phenotyping of 89 xeroderma pigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect. *Proc Natl Acad Sci U S A*, **113**, E1236-1245.
106. Jacobelli, S., Soufir, N., Lacapere, J.J., Regnier, S., Bourillon, A., Grandchamp, B., Hetet, G., Pham, D., Palangie, A., Avril, M.F. *et al.* (2008) Xeroderma pigmentosum group C in a French Caucasian patient with multiple melanoma and unusual long-term survival. *Br J Dermatol*, **159**, 968-973.
107. Meneses, M., Chavez-Bourgeois, M., Badenas, C., Villablanca, S., Aguilera, P., Bennassar, A., Alos, L., Puig, S., Malvehy, J. and Carrera, C. (2015) Atypical Clinical Presentation of Xeroderma Pigmentosum in a Patient Harboring a Novel Missense Mutation in the XPC Gene: The Importance of Clinical Suspicion. *Dermatology*, **231**, 217-221.
108. Santiago, K.M., Castro, L.P., Neto, J.P.D., de Nobrega, A.F., Pinto, C.A.L., Ashton-Prolla, P., Pinto, E.V.F., de Medeiros, P.F.V., Ribeiro, E.M., Ribeiro, B.F.R. *et al.* (2020) Comprehensive germline mutation analysis and clinical profile in a large cohort of Brazilian xeroderma pigmentosum patients. *J Eur Acad Dermatol Venereol*, **34**, 2392-2401.
109. Bernardes de Jesus, B.M., Bjoras, M., Coin, F. and Egly, J.M. (2008) Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC. *Mol Cell Biol*, **28**, 7225-7235.
110. Camenisch, U., Traeutlein, D., Clement, F.C., Fei, J., Leitenstorfer, A., Ferrando-May, E. and Naegeli, H. (2009) Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein. *Embo Journal*, **28**, 2387-2399.
111. Bunick, C.G., Miller, M.R., Fuller, B.E., Fanning, E. and Chazin, W.J. (2006) Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. *Biochemistry*, **45**, 14965-14979.
112. Maillard, O., Solyom, S. and Naegeli, H. (2007) An aromatic sensor with aversion to damaged strands confers versatility to DNA repair. *PLoS Biol*, **5**, e79.
113. Yasuda, G., Nishi, R., Watanabe, E., Mori, T., Iwai, S., Orioli, D., Stefanini, M., Hanaoka, F. and Sugasawa, K. (2007) In vivo destabilization and functional defects of the xeroderma pigmentosum C protein caused by a pathogenic missense mutation. *Mol Cell Biol*, **27**, 6606-6614.
114. Clement, F.C., Kaczmarek, N., Mathieu, N., Tomas, M., Leitenstorfer, A., Ferrando-May, E. and Naegeli, H. (2011) Dissection of the xeroderma pigmentosum group C protein function by site-directed mutagenesis. *Antioxid Redox Signal*, **14**, 2479-2490.
115. Buß, O., Rudat, J. and Ochsenreither, K. (2018) FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Computational and Structural Biotechnology Journal*, **16**, 25-33.