UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

ÉDERSON SALES MOREIRA PINTO

Characterizing the Biophysical Effects of Mutations on the XPF-ERCC1 Complex

Thesis presented in partial fulfillment
of the requirements for the degree of
Doctor of Cell and Molecular Biology

Advisor: Prof. Dr. Márcio Dorn
Coadvisor: Prof. Dr. Bruno César Feltes

Porto Alegre
December 2024

CIP — CATALOGING-IN-PUBLICATION

Sales Moreira Pinto, Éderson

Characterizing the Biophysical Effects of Mutations on the XPF-ERCC1 Complex / Éderson Sales Moreira Pinto. – Porto Alegre: PPGBCM da UFRGS, 2024.

74 f.: il.

Thesis (Ph.D.) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Biologia Celular e Molecular, Porto Alegre, BR–RS, 2024. Advisor: Márcio Dorn; Coadvisor: Bruno César Feltes.

1. Molecular Dynamics Simulation. 2. Nucleotide Excision Repair. 3. Protein Biophysics. 4. Essential Dynamics. 5. Principal Component Analysis. 6. Dynamic-Cross Correlation. 7. XPF. 8. Protein Mutation. 9. ERCC1. I. Dorn, Márcio. II. César Feltes, Bruno. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitora: Prof. Marcia Barbosa
Vice-Reitor: Profª. Pedro Costa
Pró-Reitora de Pós-Graduação: Profª. Claudia Wasserman
Diretor do Centro de Biotecnologia: Profª. Alexandre José Macedo
Coordenador do PPGBCM: Prof. Charley Christian Staats
:

"I worry that, especially as the Millennium edges near, pseudoscience and superstition will seem year by year more tempting, the siren song of unreason more sonorous and attractive. Where have we heard it before? Whenever our ethnic or national prejudices are aroused, in times of scarcity, during challenges to self-esteem or nerve, when we agonize about our diminished cosmic place and purpose, or when fanaticism is bubbling up around us - then, habits of thought familiar from ages past reach for the controls. The candle flame gutters. Its little pool of light trembles. Darkness gathers. The demons begin to stir." — "The Demon-Haunted World" by Carl Sagan

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my parents, Éverson Bastos Pinto and Nildes Sales Moreira Pinto, for their unconditional love and unwavering support. Their encouragement to pursue education has been instrumental in guiding me toward the path of science.

To Prof. Márcio Dorn and Prof. Bruno César Feltes, my advisors, whose guidance, expertise, and inspiration have shaped me in countless ways—I am truly grateful.

I extend my heartfelt thanks to the examining board: Prof. Rodrigo Ligabue-Braun, who has supported my research since I joined the graduate program at UFRGS; Prof. Paulo Augusto Netz; and Dr. Conrado Pedebos. I deeply appreciate the time, attention, and critical evaluation you have dedicated to this work.

My sincere thanks to all my colleagues and friends from the SBCB lab. A special mention goes to Fernando Guimarães Cavatão for his invaluable support during challenging times. Also, to Arthur Tonietto Magini and Lorenzo Chaves Costa Novo, the structural bioinformatics team who have walked alongside me throughout this journey.

I am also profoundly thankful to Prof. Bruno Iochins Grisci, for the inspirational discussions and steadfast friendship over the years.

To my family and friends, whose names are too numerous to mention but who all hold a special place in my heart—thank you for your constant support and love.

I owe immense gratitude to all the teachers and professors who have taught me throughout my life. Without their dedication and effort, I would not have come this far.

This work was supported by grants from the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (`CAPES`), *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (`CNPq`), and Alexander von Humboldt-Stiftung (`AvH`). I am deeply grateful for their commitment to supporting Brazil's national scientific programs despite the challenges.

ABSTRACT

Characterizing the Biophysical Effects of Mutations on the XPF-ERCC1 Complex. UV radiation generates lesions on DNA, such as cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs), which must be repaired to preserve genetic information. The nucleotide excision repair (NER) pathway plays a critical role in locating, excising, and repairing these lesions. Central to the NER pathway is the XPF-ERCC1 heterodimer, a 5' endonuclease that cleaves DNA near the damaged site. Mutations in XPF have been linked to severe disorders, including xeroderma pigmentosum (XP), Cockayne syndrome (CS), and XFE progeroid syndrome, yet the molecular mechanisms underlying these effects are not fully understood.

This thesis employs molecular dynamics simulations to investigate the structural and dynamic properties of wild-type (WT) XPF and its disease-associated mutants (R153P, C236R, R589W, and R689S) in their unbound states and as part of the XPF-ERCC1 complex. Using an all-atom approach, the XPF structure was accurately modeled, and its ion cluster geometry optimized to support microsecond-scale simulations. Analyses revealed that upon ERCC1 binding, XPF transitions from a flexible unbound state to a catalytically competent conformation with restricted domain motions, enhanced inter-domain communication, and stabilization of the catalytic core. The conjunction employment of diverse analyses elucidated the allosteric regulation and activation process.

Mutant analyses highlighted the structural and dynamic disruptions caused by each variant. R153P changes in flexibility in the unbound state, leading to mislocalization, while C236R reduced flexibility and disrupted inter-domain interactions. R589W caused severe structural instability and disorganized dynamics, whereas R689S retained partial functionality but impaired inter-domain connectivity and delayed release kinetics, affecting interstrand crosslink (ICL) repair.

By linking structural dynamics to functional outcomes, this thesis provides a molecular-level understanding of XPF-ERCC1 activation and the pathogenic effects of mutations. These findings bridge computational and experimental observations, offering insights into potential therapeutic strategies for conditions caused by DNA repair deficiencies. This work advances our understanding of DNA repair mechanisms and the critical role of XPF-ERCC1 in maintaining genomic stability.

Keywords: Molecular Dynamics Simulation. Nucleotide Excision Repair. Protein Biophysics. Essential Dynamics. Principal Component Analysis. Dynamic-Cross Correlation. XPF. Protein Mutation. ERCC1.

Caracterizando os efeitos biofísicos de mutações no complexo XPF-ERCC1

RESUMO

A radiação UV gera lesões no DNA, como dímeros de pirimidina ciclobutano (CPDs) e 6-4 fotoprodutos (6-4PPs), que precisam ser reparados para preservar a integridade da informação genética. A via de reparo por excisão de nucleotídeos (NER) desempenha um papel fundamental na localização, excisão e reparo dessas lesões. Central para a via NER está o heterodímero XPF-ERCC1, uma endonuclease que realiza a clivagem do DNA na extremidade 5' próxima ao local danificado. Mutações na XPF estão associadas a doenças graves, incluindo xeroderma pigmentosum (XP), síndrome de Cockayne (CS) e síndrome progeroide XFE, embora os mecanismos moleculares subjacentes a esses efeitos ainda não sejam completamente compreendidos. Esta tese utiliza simulações de dinâmica molecular para investigar as propriedades estruturais e dinâmicas da XPF selvagem (WT) e de seus mutantes associados a doenças (R153P, C236R, R589W e R689S) em seus estados não ligados e como parte do complexo XPF-ERCC1. Com uma abordagem *all-atom*, a estrutura do XPF foi modelada integralmente, e a geometria de seu *cluster* iônico foi otimizada para otimizar simulações na escala de microssegundos. As análises revelaram que, após a ligação à ERCC1, o XPF transita de um estado não ligado, mais flexível, para uma conformação cataliticamente ativa, com movimentos entre domínios restritos, comunicação interdomínios aprimorada e estabilização do núcleo catalítico. A aplicação de diferentes análises permitiu elucidar os processos de regulação alostérica e ativação.

As análises dos mutantes destacaram as perturbações estruturais e dinâmicas causadas por cada variante. A mutação R153P alterou a flexibilidade no estado não ligado, levando à deslocalização do complexo; C236R reduziu a flexibilidade e prejudicou as interações entre domínios; R589W causou uma severa instabilidade estrutural e desorganização dinâmica; enquanto R689S manteve funcionalidade parcial, mas comprometeu a conectividade interdomínios, impactando o reparo de ligações cruzadas entre cadeias (ICL).

Ao vincular dinâmicas estruturais a resultados funcionais, esta tese oferece uma compreensão em nível molecular da ativação do XPF-ERCC1 e dos efeitos pato-

gênicos das mutações. Estes modelos conectam observações computacionais e experimentais, fornecendo insights para potenciais estratégias terapêuticas voltadas a condições causadas por deficiências no reparo de DNA. Este trabalho contribui para o avanço do conhecimento sobre os mecanismos de reparo de DNA e o papel crítico do XPF-ERCC1 na manutenção da estabilidade genômica.

Palavras-chave: Simulações de Dinâmica Molecular, NER, Biofísica de Proteínas, Dinâmica Essencial, Análise de Componentes Principais, Matriz Dinâmica de Correlação Cruzada, XPF, Mutação de Proteínas, ERCC1.

LIST OF ABBREVIATIONS AND ACRONYMS

5mC	5-methylcytosine
6-4PPs	(6-4) Photoproducts
8-oxoG	8-oxo-7,8-dihydroguanine
CETN2	Centrin 2
CM	Catalytic Module
CMM	CheckMyMetal
CPDs	Cyclobutane Pyrimidine Dimers
CS	Cockayne Syndrome
CSA	Cockayne syndrome group A
CSB	Cockayne syndrome group B
CpG	5′-C-phosphate-G-3′
Cryo-EM	Cryogenic Electron Microscopy
DCCM	Dynamics-Cross Correlation Matrix
DDB1	Damage Binding Protein 1
DDB2	Damage Binding Protein 2
DNA	Deoxyribonucleic acid
DSBs	Double-Strand Breaks
DSSP	Dictionary of Secondary Structure in Proteins
ERCC1	Excision Repair Cross Complementation Group 1
FA	Fanconi Anemia
FF	Force Field
GG-NER	Global Genomic-NER
HLM	Helicase-like Module
ICL	Interstrand Crosslink

LINCS	LINear Constraint Solver
MD	Molecular Dynamics
MM/GBSA	Molecular Mechanics Generalized Born and Surface Area
MM/PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
NER	Nucleotide Excision Repair
NLD	Nuclease-Like Domain
NPT	Isothermal–isobaric ensemble
NVT	Canonical Ensemble
PCA	Principal Component Analysis
PCNA	Proliferating Cell Nuclear Antigen
PCs	Principal Components
PDB	Protein Data Bank
PME	Particle Mesh Ewald
Rg	Radius of Gyration
RAD23B	Human Rad23 Homolog B
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNAPII	RNA polymerase II
ROS	Reactive Oxygen Species
RPA	Replication Protein A
SASA	Solvent-Accessible Surface Area
SSBs	Single-Strand Breaks
TC-NER	Transcription-Coupled-NER
TFIIH	Transcription Factor IIH
TLS	Translesion Synthesis
TTD	Trichothiodystrophy

USP7	Ubiquitin-Specific Protease 7
UV	Ultraviolet
UV-DDB	UV-Damage Binding Protein
UVSSA	UV-Stimulated Scaffold Protein A
XP	*Xeroderma pigmentosum*
XPA	*Xeroderma pigmentosum* group A Protein
XPB	*Xeroderma pigmentosum* group B Protein
XPC	*Xeroderma pigmentosum* group C Protein
XPD	*Xeroderma pigmentosum* group D Protein
XPE	*Xeroderma pigmentosum* group E Protein
XPF	*Xeroderma pigmentosum* group F Protein
XPG	*Xeroderma pigmentosum* group G Protein
ds/ss	Double-Stranded/Single-Stranded
sMD	Steered Molecular Dynamics

LIST OF SYMBOLS

$\sum$	Summation
$\frac{a}{b}$	a divided by b
e	Euler's constant
k_B	Boltzmann's constant
T	Temperature
Z	Partition function
$P(E)$	Probability of the system being in a state with energy, E
Δ	Variation between final and initial states
S	Entropy
G	Gibbs free energy
G_{nonpolar}	Nonpolar solvation free energy
G_{polar}	Polar solvation free energy
G_{bind}	Free energy of binding
G_{complex}	Free energy of the complex
G_{receptor}	Free energy of the receptor
G_{ligand}	Free energy of the ligand
G_{solv}	Solvation free energy
G_{SA}	Nonpolar contribution to the solvation free energy
G_{PB}	Polar contribution to the solvation free energy based on the Poisson-Boltzmann equation
E_{MM}	Molecular mechanics energy
E_{int}	Internal energy
E_{elec}	Energy of electrostatic interactions
E_{vdw}	Energy of van der Waals interactions
γ	Surface tension coefficient

q_i	Charge of atom i
q_j	Charge of atom j
r_{ij}	Distance between atoms i and j
R_i	Effective Born Radii of atom i
R_j	Effective Born Radii of atom j
C	Covariance matrix
A^T	Transpose of a matrix
Λ	Diagonal matrix containing the eigenvalues of C
V	Orthogonal matrix whose columns are the eigenvectors of C

LIST OF FIGURES

Figure 1.1 DNA lesions 20
Figure 1.2 NER pathways 26
Figure 1.3 XPF-ERCC1 Complex 31
Figure 1.4 XPF Mutations 32

LIST OF TABLES

CONTENTS

1 Introduction **18**

1.1 Overview of UV Radiation 18

1.2 Photochemistry of UV-Induced DNA Damage 19

1.2.1 Cyclobutane Pyrimidine Dimers 19

1.2.2 Pyrimidine (6-4) Photoproducts 21

1.2.3 Dewar Valence Isomers 21

1.2.4 Oxidative DNA Damage 22

1.3 Biological Consequences of UV-Induced DNA Damage 23

1.3.1 Disruption of DNA Replication and Transcription 23

1.3.2 Clinical Evidence from Genetic Disorders 23

1.4 Nucleotide Excision Repair (NER) Pathway 25

1.4.1 Global Genome NER (GG-NER) 25

1.4.1.1 Damage Recognition 25

1.4.1.2 DNA Unwinding 27

1.4.1.3 Scaffolding and Damage Verification 27

1.4.1.4 Dual Incision 27

1.4.1.5 Repair Synthesis 28

1.4.2 Transcription-Coupled NER (TC-NER) 28

1.4.2.1 Damage Detection 28

1.4.2.2 Recruitment of TFIIH 28

1.4.2.3 Scaffolding and Damage Verification 29

1.4.2.4 Dual Incision 29

1.4.2.5 Repair Synthesis 29

1.4.3 Key Differences Between GG-NER and TC-NER 29

1.5 XPF-ERCC1 Structure and Function 30

1.5.1 Structural Organization 30

1.5.2 DNA Binding and Catalytic Activation 31

1.6 Clinical Relevance of XPF-ERCC1 and DNA Repair Deficiencies 32

1.6.1 R153P Mutation 32

1.6.2 C236R Mutation 33

1.6.3 R589W Mutation 33

1.6.4 R689S Mutation 34

1.6.5 Broader Clinical Implications 34
1.7 Molecular Dynamics Simulations 34
1.8 Chapter Summary 52

2 Objectives **53**
2.1 General Objective 53
2.2 Specific Objectives 53
2.3 Chapter Summary 54

3 Methods and Concepts **55**
3.1 Protein Structure Preparation and Modeling 55
3.2 Molecular Dynamics Simulations of XPF and Mutants 56
3.3 Modeling and Simulation of XPF-ERCC1 Interactions 56
3.4 Principal Component Analysis 57
3.5 Dynamic Cross-Correlation Matrix 57
3.6 Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) and Generalized Born Surface Area (MM/GBSA) 58
3.6.1 MM/PBSA 58
3.6.2 MM/GBSA 59
3.7 Chapter Summary 61

4 Conclusions **62**
4.1 Chapter Summary 63

1 INTRODUCTION

Evolution follows a path where organisms strive to preserve and transmit their genetic material across generations. The concept of "survival of the fittest" encompasses numerous biological mechanisms to ensure the reproduction process's fidelity and quality. Whether considering a unicellular organism or a multicellular one, specialized molecular systems are required to continuously monitor and repair the genetic material when it becomes compromised.

The DNA molecule serves as the primary repository of genetic information. The maintenance of DNA integrity is crucial for cell function, development, and inheritance. However, DNA is susceptible to mutations, which can disrupt replication and alter the genetic code's reading frame, particularly when exposed to environmental stressors. A well-known mutagen that induces such DNA damage is ultraviolet (UV) radiation, which can trigger the formation of various lesions that compromise DNA stability.

1.1 Overview of UV Radiation

UV radiation is a segment of the electromagnetic spectrum ranging from 100 to 400 nm. It is classified into three main types based on wavelength: UVC (100–280 nm), UVB (280–315 nm), and UVA (315–400 nm) (DIFFEY, 2002; DÖHRING, 2017), but this subdivision can vary between environmental and dermatological photobiologists (DIFFEY, 2002). UVC, the most energetic form, is almost entirely absorbed by atmospheric ozone, preventing it from reaching the Earth's surface. Additionally, there is a weak absorption band known as the Herzberg continuum in the 200–242 nm wavelength range. When oxygen is absorbed at wavelengths shorter than 242 nm, it photo-dissociates into two O atoms. Ozone can be formed by adding the generated O atoms to other molecular oxygen molecules. (DÖHRING, 2017). Although comprising only about 5% of solar UV radiation, UVB is particularly effective in causing DNA damage due to its wavelength proximity to the peak DNA absorption at 260 nm (CADET; DOUKI, 2018). UVA, which constitutes 95% of solar UV radiation, penetrates deeper into tissues and contributes to DNA damage primarily through indirect mechanisms such as reactive oxygen species (ROS) generation (FRAIKIN, 2024).

Moreover, a study showed that UVA could be carcinogenic when considering the double-strand nature of the DNA molecule instead of DNA monomers typically used in UV absorption studies (MOURET et al., 2010; CADET; DOUKI, 2018). The depletion of the stratospheric ozone layer, driven by anthropogenic activities, has led to increased UV exposure. This increase is associated with a higher incidence of skin cancers and photoaging. This widespread exposure highlights the critical need to understand the mechanisms of UV-induced DNA damage and the cellular responses that counteract it.

1.2 Photochemistry of UV-Induced DNA Damage

UV radiation can induce DNA damage through direct absorption by nucleic acids and indirect mechanisms involving photosensitization. The direct absorption of UV photons by DNA bases, particularly under UVB and UVC, leads to the formation of various DNA photoproducts, which include cyclobutane pyrimidine dimers (CPDs), pyrimidine (6-4) photoproducts (6-4PPs), and Dewar valence isomers (RASTOGI et al., 2010; DOUKI; SAGE, 2016). UVA, while less aggressive in DNA damage, contributes to oxidative stress and secondary DNA lesions through the generation of ROS (CADET et al., 2012), facilitates the isomerization of UVB-induced 6-4PPs into Dewar isomers (CADET; SAGE; DOUKI, 2005; MARKOVITSI, 2016), and has been shown to induce CPD formation in both keratinocytes and melanocytes in human skin (MOURET et al., 2006; KHAN; TRAVERS; KEMP, 2018). Figure 1.1 shows the different DNA lesion types generated by UV radiation.

1.2.1 Cyclobutane Pyrimidine Dimers

CPDs are the most common and mutagenic UV-induced DNA lesions (BESARATINIA et al., 2011). The formation of CPDs involves a $[2+2]$ cycloaddition reaction between the C5 and C6 carbon atoms of adjacent pyrimidine bases, typically thymine or cytosine (CADET; DOUKI, 2018). This ultrafast reaction occurs within picoseconds after photon absorption and results in the formation of a cyclobutane ring (MARKOVITSI, 2016; SCHREIER et al., 2007). The CPD lesion, although not

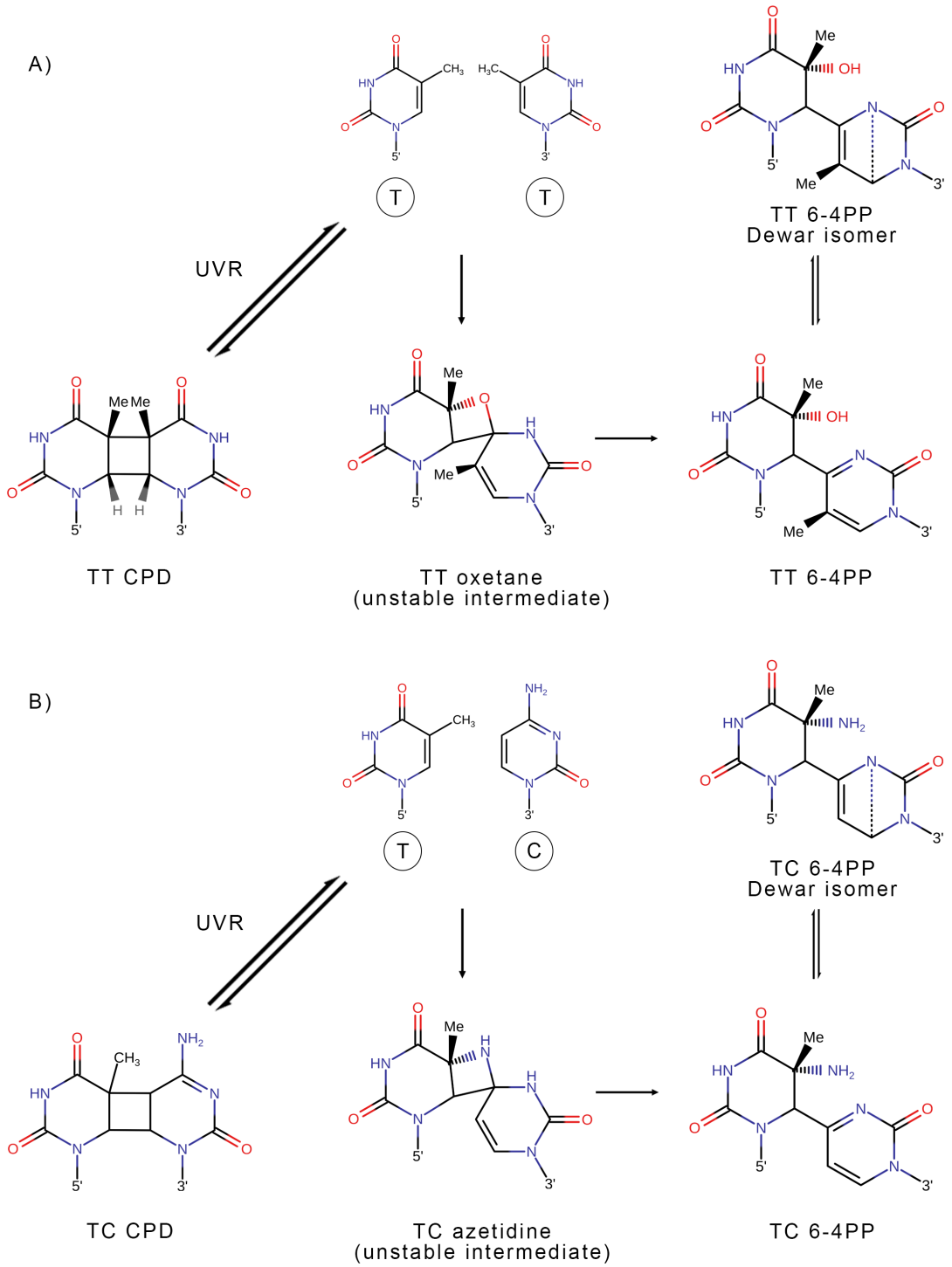


Figure 1.1 – DNA lesions. A) Thymine dimer reactions under UV radiation; B) Thymine and cytosine reactions under UV radiation.

particularly helix-distorting, can disrupt base pairing and alter the helical conformation, obstructing DNA replication and transcription (MOIRANGTHEM; GAMAGE; ROKITA, 2023).

CPDs are known to cause C→T transition mutations, a signature of UV-induced mutagenesis observed in skin cancers. The frequency of CPD formation is influenced by the sequence context, with dipyrimidine sites (TT, TC, CT, CC) being the most susceptible, with the observed preference for dimerization of TT > TC > CT > CC detected both *in vitro* and *in vivo* (PREMI et al., 2019; MOIRANGTHEM; GAMAGE; ROKITA, 2023). The presence of 5-methylcytosine (5mC) further increases the likelihood of CPD formation due to enhanced absorption properties, particularly in 5′-C-phosphate-G-3′ (CpG) islands, which are mutational hotspots in skin cancer genomes (CADET; DOUKI, 2018).

1.2.2 Pyrimidine (6-4) Photoproducts

6-4PPs are formed via a covalent bond between the C6 atom of a 5′ pyrimidine and the C4 atom of a 3′ pyrimidine, typically involving thymine and cytosine (CADET; DOUKI, 2018). The formation of 6-4PPs is slower than CPDs, taking place over milliseconds and involving an oxetane or azetidine intermediate (MARGUET; MARKOVITSI, 2005; RASTOGI et al., 2010). These lesions introduce greater helical distortion compared to CPDs, making them more readily recognized by DNA repair machinery (LUKIN; SANTOS, 2006).

The mutagenic potential of 6-4PPs is high due to their ability to block DNA polymerases during replication (HUNG et al., 2020), but NER-proficient cells can efficiently repair this type of lesion (LIMA-BESSA et al., 2008). Therefore, 6-4PPs become really health-threatening for individuals presenting mutations in any protein that somehow impairs Nucleotide Excision Repair (NER) function.

1.2.3 Dewar Valence Isomers

Dewar valence isomers are photoisomerization products of 6-4PPs, formed upon further exposure to UVA or UVB radiation (DOUKI; SAGE, 2016). These lesions involve an electrocyclization reaction between the N3 and the C6 in the pyrim-

idine ring structure of the 3′ base of the 6-4PP (BERENS; MOLINIER, 2020). Dewar isomers, although less common, exhibit mutagenic properties, especially at TC dinucleotides which are mutational hotspots in UV-exposed cells (DOUKI; SAGE, 2016). Their biological relevance is highlighted in contexts of prolonged UVA exposure, such as in artificial tanning practices, where their formation is significantly enhanced (COELHO et al., 2015).

1.2.4 Oxidative DNA Damage

Oxidative DNA damage arises primarily from interacting reactive oxygen species (ROS) with DNA. ROS are chemically reactive molecules derived from oxygen and can be generated both endogenously and through external factors like UVA radiation. UVA photons penetrate cells and are absorbed by chromophores, such as flavins, porphyrins, and melanin precursors (HALLIWELL; GUTTERIDGE, 2015). This absorption excites these molecules, leading to the production of ROS through energy transfer to molecular oxygen. The primary ROS generated include superoxide anions ($O_2^{\cdot -}$), hydrogen peroxide (H_2O_2), and hydroxyl radicals ($^{\cdot}OH$) (CADET; DAVIES, 2017).

Once produced, ROS interact with DNA, causing structural and chemical modifications. Guanine is particularly susceptible to oxidation, resulting in the formation of 8-oxo-7,8-dihydroguanine (8-oxoG), one of the most studied oxidative lesions (FRIEDBERG et al., 2005; CADET et al., 2005). This mutagenic lesion mispairs with adenine during DNA replication, leading to G→T transversions. Beyond base modifications, ROS can attack the sugar-phosphate backbone, causing single-strand breaks (SSBs). When such damage occurs on both DNA strands in close proximity, double-strand breaks (DSBs) can result, which are among the most cytotoxic forms of DNA damage. Additionally, ROS can lead to DNA-protein crosslinks, disrupting essential processes like replication and transcription, and generate abasic sites where DNA bases are lost, further destabilizing the DNA molecule (BOITEUX; RADICELLA, 1999).

1.3 Biological Consequences of UV-Induced DNA Damage

UV-induced DNA damage has profound biological consequences, affecting replication, transcription, and overall genomic stability. The lesions caused by UV radiation, including CPDs, 6-4PPs, and oxidative damage, disrupt DNA structure and function, triggering a cascade of molecular events. If left unrepaired, these lesions can result in mutagenesis, apoptosis, or cancer. The consequences of impaired DNA repair are particularly evident in genetic disorders where repair pathways are compromised (TIWARI; WILSON, 2019).

1.3.1 Disruption of DNA Replication and Transcription

DNA replication and transcription are sensitive to UV-induced lesions due to the distortion of the DNA helix caused by CPDs and 6-4PPs. During replication, these lesions block the progression of DNA polymerases, leading to fork stalling and replication stress. Persistent stalling can result in fork collapse and the formation of DSBs, which, if misrepaired, contribute to chromosomal aberrations and genomic instability (CANNAN; PEDERSON, 2016). To bypass lesions, translesion synthesis (TLS) polymerases such as Pol η and Pol κ are recruited. However, these polymerases are error-prone and can introduce mutations, particularly at sites of CPDs, contributing to UV-induced mutagenesis (SALE, 2013).

1.3.2 Clinical Evidence from Genetic Disorders

Impaired DNA repair mechanisms highlight the critical importance of maintaining genomic integrity. Clinical evidence of the consequences of defective repair is observed in several genetic disorders:

Xeroderma Pigmentosum (XP): XP is caused by mutations in genes encoding NER proteins, going from *xeroderma pigmentosum* group A protein (XPA) to *xeroderma pigmentosum* group G protein (XPG) (PICCIONE et al., 2021). Patients with XP exhibit extreme sensitivity to UV radiation, resulting in severe sunburns, freckling, and a 2,000- to 10,000-fold increased risk of developing skin cancers (BAYKAL et al., 2021). XP patients often develop basal cell carcinoma, squamous

cell carcinoma, and melanoma at an early age. Neurological degeneration is also observed in some XP subtypes due to accumulated oxidative DNA damage in neuronal cells (LEUNG et al., 2022).

Cockayne Syndrome (CS): CS is caused by mutations in the *CSA* or *CSB* genes, which are critical for transcription-coupled NER (TC-NER) (KARIKKINETH et al., 2017). Patients with CS do not have a predisposition to cancer but suffer from severe developmental defects, progressive neurodegeneration, growth failure, and premature aging (LEHMANN, 2003). CS cells cannot repair transcription-blocking lesions, leading to impaired gene expression and cellular dysfunction. UV exposure exacerbates these symptoms due to increased accumulation of transcriptional damage (STEFANINI et al., 1996).

XFE Progeroid Syndrome: This disorder is caused by mutations in the XPF gene, which encodes an endonuclease critical for NER and interstrand crosslink repair (ZHAO et al., 2023). Patients with XFE exhibit features of premature aging, including graying hair, osteoporosis, and skin atrophy, as well as heightened sensitivity to UV radiation. These symptoms underscore the role of NER in preventing cellular senescence and maintaining tissue homeostasis (MORI et al., 2018).

Trichothiodystrophy (TTD): TTD results from mutations in genes encoding components of the TFIIH complex, such as XPB and XPD. Unlike XP, TTD is not associated with cancer but involves brittle hair, intellectual disability, developmental delays, and photosensitivity. The disease is linked to defective transcription and repair, particularly in response to UV-induced DNA damage (STEFANINI et al., 2010).

The diverse clinical features observed in disorders such as XP, CS, XFE progeroid syndrome, and TTD illustrate the vital role of an evolutive adapted mechanism for protecting against UV-induced DNA damage and preserving cellular function. NER pathway is a highly specialized repair mechanism that removes bulky, helix-distorting lesions from DNA, operating through two distinct sub-pathways: global genomic NER (GG-NER) and transcription-coupled NER TC-NER. GG-NER is responsible for detecting lesions across the entire genome, while TC-NER focuses on actively transcribed genes to prevent transcriptional stalling. Next, we delve into the intricacies of the NER pathway.

1.4 Nucleotide Excision Repair (NER) Pathway

NER is a critical mechanism for maintaining genome integrity by removing bulky DNA lesions that distort the DNA helix. These lesions interfere with essential cellular processes such as replication and transcription. NER is divided into two sub-pathways: global genome NER (GG-NER) and transcription-coupled NER (TC-NER) (Figure 1.2). Both pathways share the same downstream repair steps but differ in their mechanism of lesion recognition. The NER pathways have been extensively reviewed over the years (KUMAR et al., 2020; HEUVEL et al., 2021; D'SOUZA; BLEE; CHAZIN, 2022; KUPER; KISKER, 2023; MORENO; OLTHOF; SVEJSTRUP, 2023; D'SOUZA et al., 2024), and this section summarizes the current state of knowledge of NER.

1.4.1 Global Genome NER (GG-NER)

1.4.1.1 Damage Recognition

GG-NER operates throughout the genome, scanning for DNA helix distortions caused by lesions. This process begins with the XPC-RAD23B-CETN2 complex, which detects bulky DNA lesions, especially 6-4PP (D'SOUZA et al., 2024; FITCH et al., 2003; SUGASAWA et al., 2002). XPC recognizes DNA regions with helical distortions without directly interacting with the lesion. RAD23B (human Rad23 homolog B) stabilizes XPC, enhancing its DNA-binding ability (MASUTANI et al., 1994), while CETN2 (Centrin 2), a calcium-binding protein, further stabilizes the complex (ARAKI et al., 2001).

For lesions such as CPDs, which minimally distort the helix, the UV-DDB complex (UV-damage binding protein) facilitates damage recognition (WITTSCHIEBEN; IWAI; WOOD, 2005). The UV-DDB complex comprises DDB1 and DDB2 (the product of the *XPE* gene). DDB2 binds directly to CPDs, while DDB1 is part of the CRL4 ubiquitin ligase complex that ubiquitinates histones and repair proteins, promoting chromatin remodeling and enabling XPC recruitment (GROISMAN et al., 2003; FISCHER et al., 2011). If DDB2 does not leave the lesion site, the recruitment of the NER machinery is delayed or halted (RIBEIRO-SILVA et al., 2020).

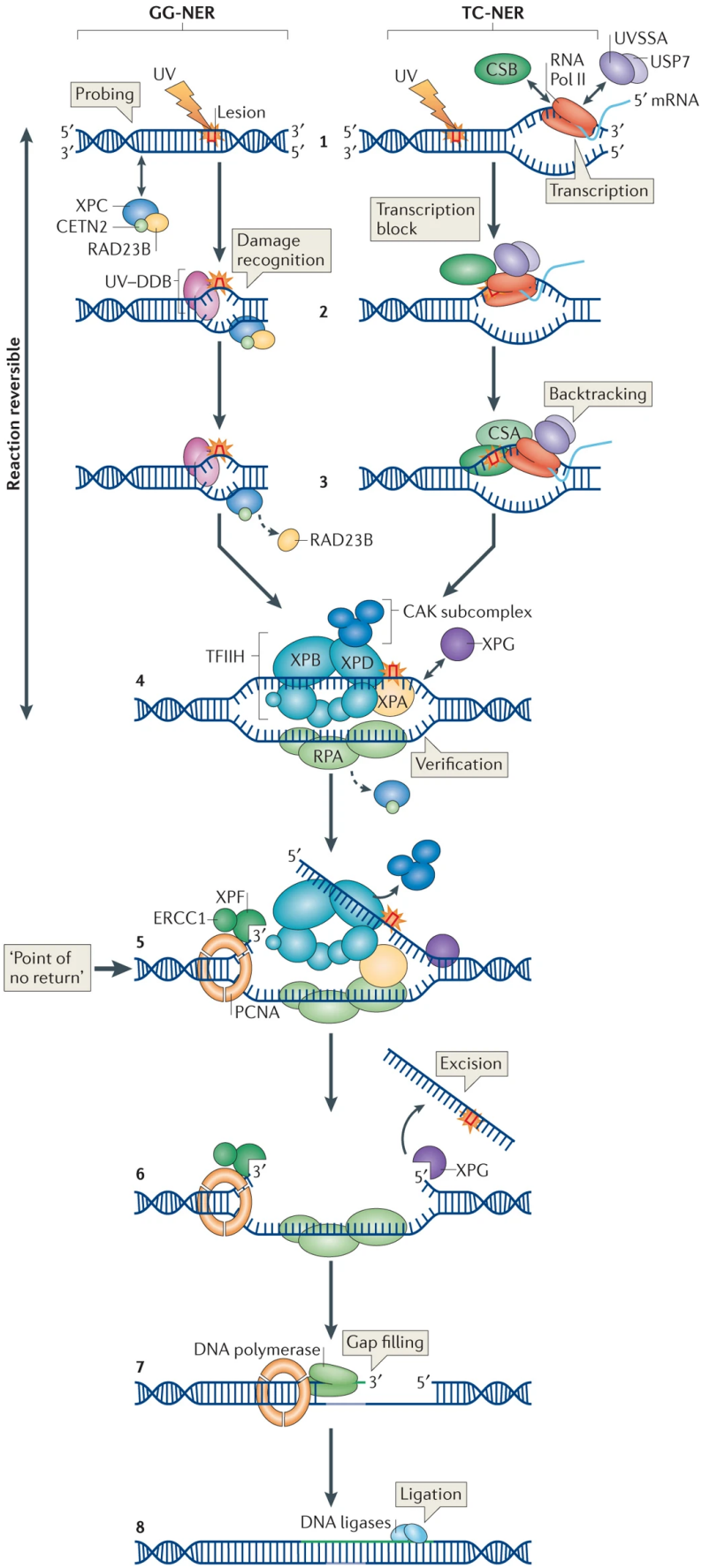


Figure 1.2 – Overview of the GG-NER and TC-NER. Image adapted from (MARTEIJN et al., 2014).

1.4.1.2 DNA Unwinding

Once XPC binds to the lesion, it recruits the transcription factor IIH (TFIIH), a multi-subunit complex essential for DNA unwinding and lesion verification (OKUDA et al., 2015). TFIIH consists of ten subunits, including the helicases XPB (*ERCC3*) and XPD (*ERCC2*). XPB initiates unwinding upstream of the lesion (FAN et al., 2006), while XPD translocates along the damaged strand in a 5′ to 3′ direction, confirming the lesion by stalling at the site (LIU et al., 2008).

The other TFIIH subunits (p8, p34, p44, p52, and p62) stabilize the helicase complex and regulate its activity. To facilitate helicase engagement, the CAK module of TFIIH (composed of CDK7, Cyclin H, and MAT1) dissociates, allowing the conformational changes required for efficient DNA unwinding (COMPE; EGLY, 2012; EGLY; COIN, 2011).

1.4.1.3 Scaffolding and Damage Verification

Following DNA unwinding, the scaffold protein XPA and the single-stranded DNA-binding protein RPA (replication protein A) stabilize the DNA bubble and verify the lesion (FANNING; KLIMOVICH; NAGER, 2006). XPA binds to the single-stranded region near the lesion and interacts with TFIIH, RPA, and the endonucleases. RPA, a heterotrimer composed of RPA70, RPA32, and RPA14, binds to the undamaged strand, protecting it from degradation and secondary structure formation (SAIJO; TAKEDACHI; TANAKA, 2011).

1.4.1.4 Dual Incision

The lesion-containing oligonucleotide is excised by dual incisions performed by the endonucleases XPG (GONZÁLEZ-CORROCHANO et al., 2020) and XPF-excision repair cross complementation group 1 (ERCC1) (JONES et al., 2020). XPG makes the 3′ incision and interacts with TFIIH (via the p44 and p62 subunits) and RPA. XPF-ERCC1, a heterodimeric endonuclease, makes the 5′ incision. ERCC1 interacts with XPA to position the complex correctly, while XPF performs the catalytic cleavage (STARESINCIC et al., 2009).

1.4.1.5 Repair Synthesis

After excision, the resulting gap is filled by DNA polymerase δ, ϵ, or κ, which synthesizes new DNA using the undamaged strand as a template (OGI et al., 2010). Proliferating cell nuclear antigen (PCNA) acts as a sliding clamp to coordinate repair synthesis (GILLJAM et al., 2012). DNA ligase I seal the final nick in the DNA backbone, completing the repair. DNA ligase III has also been shown to take part in the end of NER (PAUL-KONIETZKO et al., 2015).

1.4.2 Transcription-Coupled NER (TC-NER)

1.4.2.1 Damage Detection

TC-NER targets lesions on the transcribed strand of active genes. Damage detection begins when RNA polymerase II (RNAPII) stalls at a lesion during transcription elongation (DUAN et al., 2021). CSB (Cockayne syndrome group B protein, encoded by *ERCC6*) binds to the stalled RNAPII and facilitates its dissociation or remodeling to expose the lesion. CSA (Cockayne syndrome group A protein, encoded by *ERCC8*) is part of an E3 ubiquitin ligase complex and is recruited by CSB to stabilize the repair complex and regulate the ubiquitination of repair factors (WEEGEN et al., 2020).

UVSSA (UV-stimulated scaffold protein A) interacts with CSB and CSA, stabilizing the repair complex. UVSSA recruits USP7 (ubiquitin-specific protease 7), which removes inhibitory ubiquitin modifications from CSB and RNAPII, maintaining their activity during repair (ZHANG et al., 2012).

1.4.2.2 Recruitment of TFIIH

The core NER machinery is recruited once the lesion is detected and RNAPII is displaced. TFIIH unwinds the DNA around the lesion, with XPB and XPD helicases playing the same roles as in GG-NER. XPB unwinds the DNA upstream of the lesion, and XPD translocates along the damaged strand, verifying the lesion (KOKIC et al., 2019).

1.4.2.3 Scaffolding and Damage Verification

XPA and RPA stabilize the DNA bubble created by TFIIH. XPA binds near the lesion and interacts with RPA, TFIIH, and endonucleases. RPA binds the undamaged strand, protecting it and preventing secondary structures (TOPOLSKA-WOŚ et al., 2020).

1.4.2.4 Dual Incision

Dual incisions are carried out by XPG and XPF-ERCC1, as in GG-NER. XPG performs the 3′ incision and interacts with TFIIH and RPA, while XPF-ERCC1 makes the 5′ incision with the help of XPA (STARESINCIC et al., 2009).

1.4.2.5 Repair Synthesis

The repair synthesis step is identical to GG-NER. DNA polymerase δ or ϵ synthesizes new DNA, coordinated by PCNA, and DNA ligase I or III seals the nick, restoring the integrity of the DNA (PAUL-KONIETZKO et al., 2015).

1.4.3 Key Differences Between GG-NER and TC-NER

GG-NER and TC-NER differ primarily in their mechanism of lesion recognition. GG-NER uses the XPC-RAD23B-CETN2 complex to detect helix distortions across the entire genome, with UV-DDB assisting in recognizing less-distorting lesions like CPDs. This makes GG-NER a genome-wide surveillance system that repairs both transcribed and non-transcribed regions. Conversely, TC-NER targets lesions that block transcription, initiating repair when RNAPII stalls. CSB facilitates RNAPII dissociation or remodeling, while CSA and UVSSA regulate repair dynamics through ubiquitination and stabilization of key proteins. The functional focus of GG-NER is broad, preserving genomic integrity by repairing all regions, including non-coding DNA. TC-NER, however, prioritizes active genes, rapidly restoring transcriptional activity to maintain gene expression.

Building on the distinct mechanisms of lesion recognition and repair in GG-NER and TC-NER, the efficiency of these pathways relies on the coordinated actions of specialized proteins to excise damaged DNA and restore genomic integrity.

Central to the excision step is the XPF-ERCC1 complex, a structure-specific endonuclease that performs the critical 5′ incision near the DNA lesion. This complex plays a pivotal role in ensuring the accuracy and efficiency of the repair process, not only in NER but also in other DNA repair pathways. Understanding the structural features and functional dynamics of XPF-ERCC1 is essential to elucidate how it interfaces with other NER components and executes its catalytic role, as well as how mutations in XPF can disrupt repair mechanisms and contribute to disease. The following section delves into the architecture and molecular functions of XPF-ERCC1, highlighting its significance in DNA repair and genomic stability.

1.5 XPF-ERCC1 Structure and Function

XPF-ERCC1 is a structure-specific endonuclease critical for the repair of DNA damage through multiple pathways, including NER and interstrand crosslink (ICL) repair. This heterodimeric complex comprises the catalytically active XPF subunit and the non-catalytic ERCC1 subunit, which together form a functional unit that recognizes and incises DNA at specific double-stranded/single-stranded (ds/ss) junctions. These actions are pivotal in resolving DNA distortions caused by UV and bulky chemical adducts. This section is based on the work of Jones et al. that elucidated the partial structure of the XPF-ERCC1 complex through cryo-EM (JONES et al., 2020).

1.5.1 Structural Organization

The XPF-ERCC1 is an endonuclease that cleaves DNA at the 5′ end of damaged strands. Its structure features a well-organized modular architecture that enhances its functional versatility. The structure of XPF is divided into five domains: the RecA1, Helicase, and RecA2 domains, which include the Helicase-like module (HLM), as well as the Nuclease and HhH_2 domains, forming the catalytic module (CM).

The XPF-ERCC1 heterodimer has a modular architecture that includes a HLM and a CM in XPF, alongside structurally complementary domains in ERCC1. XPF's HLM, which lacks canonical ATPase activity, serves as a platform for DNA

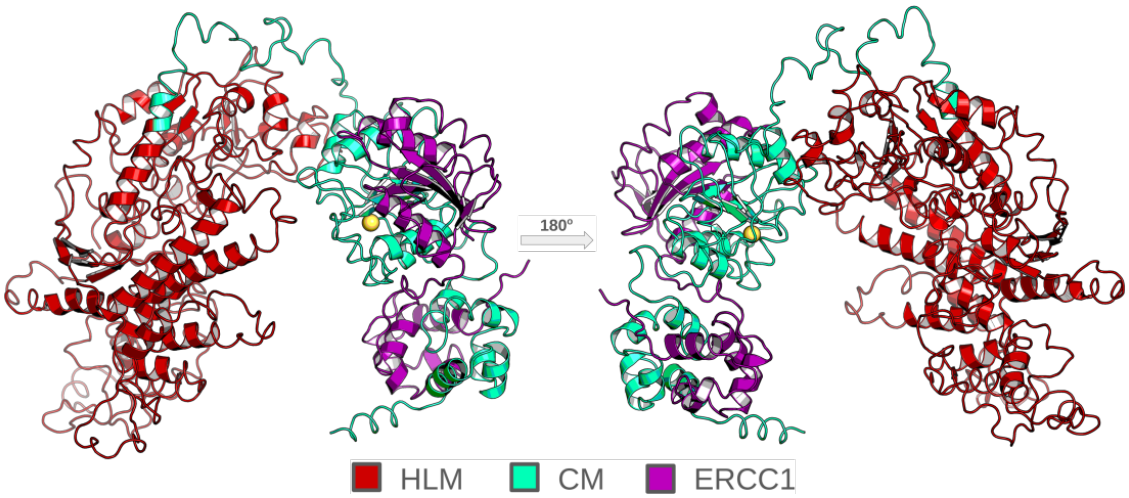


Figure 1.3 – Overview of the XPF-ERCC1 Complex. HLM is red, CM is green cyan, the ERCC1 protein is purple, and the Mg^{2+} ion is yellow.

binding and interaction with regulatory factors such as SLX4. The CM includes a metal-dependent nuclease domain with a conserved GDX_nERKX_3D motif and a helix-hairpin-helix (HhH_2) domain. ERCC1 complements this structure with a nuclease-like domain (NLD) and an additional HhH_2 domain, which synergizes to recognize and stabilize DNA substrates. Figure 1.3 shows the structure of the XPRF ERCC1 complex.

Moreover, the study revealed that XPF-ERCC1 exists in an auto-inhibited conformation in the absence of DNA, with the XPF helical domain occluding the catalytic site. This conformational state prevents unintended cleavage of non-specific DNA substrates, highlighting a sophisticated regulatory mechanism.

1.5.2 DNA Binding and Catalytic Activation

Upon encountering a DNA junction, the XPF-ERCC1 complex undergoes a conformational shift that activates its nuclease activity. This involves the release of the ERCC1 HhH_2 domain from the XPF helical domain, enabling cooperative engagement of both dsDNA and ssDNA elements by the HhH_2 domains of XPF and ERCC1. This interaction positions the scissile bond of the DNA substrate within the catalytic site of XPF, initiating the cleavage reaction.

The architecture and flexibility of XPF-ERCC1 are tailored to its functions in diverse repair pathways. For example, in the NER pathway, XPF-ERCC1 nicks the damaged strand upstream of the lesion, facilitating excision and subsequent

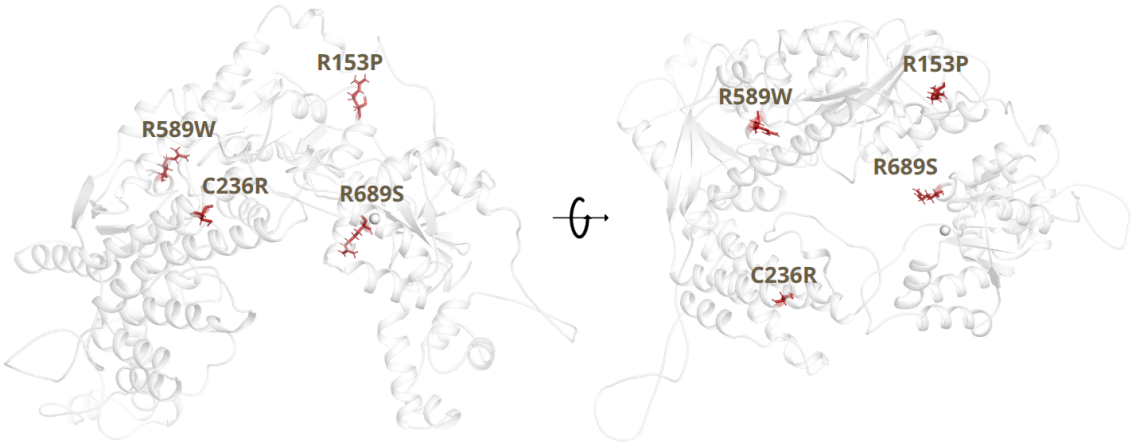


Figure 1.4 – Structure of the WT XPF. Residues that undergo mutations are depicted in red. This highlights some areas of the protein that contribute to functional changes due to genetic variations.

repair. In ICL repair, its activity is regulated by SLX4, which recruits XPF-ERCC1 to crosslinked DNA structures. For more details about the structure and function of the XPF-ERCC1 complex, please refer to (JONES et al., 2020).

1.6 Clinical Relevance of XPF-ERCC1 and DNA Repair Deficiencies

Mutations in the XPF and ERCC1 genes are associated with several genetic disorders, including XP, CS, and Fanconi anemia (FA). These mutations can disrupt the protein complex's structural stability or catalytic function. For an in-depth discussion of known mutations that affect the XPF protein, please refer to this review (FELTES; BONATTO, 2015). In this discussion, we will explore the clinical significance of specific mutations in XPF, including R153P, C236R, R589W, and R689S, which are central to this work. Figure 1.4 depicts the position of the residues from the WT XPF that undergo mutations.

1.6.1 R153P Mutation

The R153P mutation in the XPF protein carries significant clinical implications, particularly regarding DNA repair and disease manifestation. This mutation is closely associated with XP and is also linked to XFE progeroid syndrome, a condition that leads to accelerated aging (NIEDERNHOFER et al., 2006). The R153P mutation causes the XPF-ERCC1 complex to mislocalize to the cytoplasm instead of

transporting it to the nucleus. Despite this mislocalization, the XPF-ERCC1 complex maintains its catalytic activity, allowing it to incise DNA *in vitro* (AHMAD et al., 2010). However, this mislocalization inhibits its ability to repair DNA damage *in vivo* effectively. XP-F patients, especially those with the R153P mutation, often present with milder clinical symptoms, including a later onset of skin cancer. This phenomenon is thought to result from the residual, although reduced, endonuclease activity of the XPF-ERCC1 complex (MATSUMURA et al., 1998). Ultimately, the specific molecular consequences of this mutation on the XPF protein and its impact on proper folding for nuclear transport remain unclear.

1.6.2 C236R Mutation

The C236R mutation in the XPF protein has notable clinical implications, particularly concerning XP, CS, and FA (HASHIMOTO et al., 2015). This specific mutation in XPF is primarily linked to XP. However, individuals with the C236R mutation typically exhibit milder clinical features, including a delayed onset of skin cancer. Unlike other mutations within the SLX4-interacting region, the C236R mutation does not induce instability in the XPF protein. Furthermore, it does not significantly hinder the recruitment of XPF to sites of DNA damage, which is crucial for effective NER. Thus, it is unsolved why the C236R mutation causes the combined phenotype of FA, CS, and XP, even though it does not cause protein instability or impede protein-protein interaction (KASHIYAMA et al., 2013).

1.6.3 R589W Mutation

The R589W mutation in the XPF protein results in a structurally unstable protein, leading to deficiencies in DNA repair. This mutant protein demonstrates impaired folding dynamics, which compromises its catalytic function (HASHIMOTO et al., 2015). Clinically, this particular mutation is associated with XP, characterized by a delayed onset of symptoms and a milder severity compared to other XPF mutations. Patients carrying the R589W mutation often experience a later development of skin cancer, likely due to residual NER activity mediated by partially functional XPF-ERCC1 complexes. However, there is currently no mechanism that

explains how the R589W mutation can retain some level of function despite the structural changes that lead to loss of stability.

1.6.4 R689S Mutation

The R689S mutation in the XPF protein has been linked to Fanconi anemia, a disorder characterized by genomic instability and impaired DNA ICL repair (BOGLIOLO et al., 2013). Located within the nuclease active site of XPF, this mutation significantly disrupts the protein's functionality in ICL repair, though it retains some partial activity in NER. This functional imbalance contributes to the clinical features of Fanconi anemia, which include hypersensitivity to ICL-inducing agents, progressive bone marrow failure, developmental abnormalities, and an elevated risk of cancer. Despite its residual activity in NER, the R689S variant is inadequate to fully restore DNA repair mechanisms, resulting in ongoing genomic instability and the associated clinical symptoms, including severe susceptibility to DNA-damaging agents (OSORIO et al., 2013).

1.6.5 Broader Clinical Implications

Mutations in XPF, such as R153P, C236R, R589W, and R689S, highlight the importance of this protein in preserving genomic stability. Deficiencies in XPF-ERCC1-mediated repair lead to disorders such as XP, CS, and FA. These conditions are characterized by symptoms ranging from increased cancer risk and accelerated aging to neurodegeneration and developmental abnormalities. Understanding the molecular impacts of specific mutations provides critical insights into disease mechanisms and potential therapeutic strategies, such as using targeted molecules to stabilize mutant proteins.

1.7 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are based on the principles of molecular mechanics and statistical mechanics, which provide a framework for linking the microscopic motion of particles to macroscopic thermodynamic properties. In

the study of biological macromolecules, MD simulations enable us to investigate the molecular mechanisms underlying the functions of proteins. Additionally, these simulations allow us to explore the configurational space and energy landscapes associated with the processes being studied. Since MD simulations are central to this work, we have conducted an extensive review of the literature on their use in investigating proteins related to NER. Following this, we will present a comprehensive review of the topic, entitled "The nucleotide excision repair proteins through the lens of molecular dynamics simulations."

DNA Repair 127 (2023) 103510

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

Review Article

The nucleotide excision repair proteins through the lens of molecular dynamics simulations

Éderson Sales Moreira Pinto [a], Mathias J. Krause [b], Márcio Dorn [a,c,d], Bruno César Feltes [c,*,1]

[a] Center for Biotechnology, Federal University of Rio Grande do Sul, RS, Brazil
[b] Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, Karlsruhe, Germany
[c] Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
[d] National Institute of Science and Technology - Forensic Science, Porto Alegre, RS, Brazil

ARTICLE INFO

Keywords:
NER
Molecular dynamics
Xeroderma pigmentosum
Docking
Protein structure
Bioinformatics

ABSTRACT

Mutations that affect the proteins responsible for the nucleotide excision repair (NER) pathway can lead to diseases such as xeroderma pigmentosum, trichothiodystrophy, Cockayne syndrome, and Cerebro-oculo-facio-skeletal syndrome. Hence, understanding their molecular behavior is needed to elucidate these diseases' phenotypes and how the NER pathway is organized and coordinated. Molecular dynamics techniques enable the study of different protein conformations, adaptable to any research question, shedding light on the dynamics of biomolecules. However, as important as they are, molecular dynamics studies focused on DNA repair pathways are still becoming more widespread. Currently, there are no review articles compiling the advancements made in molecular dynamics approaches applied to NER and discussing: (i) how this technique is currently employed in the field of DNA repair, focusing on NER proteins; (ii) which technical setups are being employed, their strengths and limitations; (iii) which insights or information are they providing to understand the NER pathway or NER-associated proteins; (iv) which open questions would be suited for this technique to answer; and (v) where can we go from here. These questions become even more crucial considering the numerous 3D structures published regarding the NER pathway's proteins in recent years. In this work, we tackle each one of these questions, revising and critically discussing the results published in the context of the NER pathway.

1. Introduction

DNA repair-related diseases, such as xeroderma pigmentosum (XP), trichothiodystrophy (TTD), Cockayne syndrome (CS), Cerebro-oculo-facio-skeletal syndrome (COFS), which are associated with impairments in the nucleotide excision repair (NER) pathway [1–5], are all related to specific mutations that affect the proteins associated to NER, influencing protein-protein interactions (PPI), catalysis, and/or DNA-binding. Hence, understanding the molecular behavior of NER-related proteins becomes essential to uncover how these diseases cause their phenotypes and how NER is organized.

However, describing a protein's structural information goes beyond identifying domains and residues that play indispensable roles in their functions. Proteins are dynamic macromolecules that undergo conformational changes in response to different stimuli and binding partners. Although multiple methods exist to elucidate a protein 3D structure, observing their dynamic molecular behavior in a continuous timescale outside bioinformatics is still unattainable. Molecular dynamics (MD) simulations have been broadly employed to witness the molecular behavior of proteins and the study of proteins bound to different DNA conformations, mainly those containing mutated sequences, shedding light on lesion recognition. For these and other reasons, MD analyses have become an indispensable part of the ``structural biologist'' toolkit. However, as popular as they are, MD studies focusing on DNA repair have yet to be more widespread.

While some works provide explanations of their strengths [6,7] and/or limitations [8], there are no reviews articles compiling the advancements made in molecular dynamics approaches applied to NER and discussing: (i) how this technique is currently employed in the field of DNA repair, focusing on NER proteins; (ii) which technical setups are being employed, their strengths and limitations; (iii) which insights or information are they providing to understand the NER pathway or

* Correspondence to: Instituto de Informática – Laboratório de Bioinformática Estrutural e Biologia Computacional, Departamento de Informática Teórica, Universidade Federal do Rio Grande do Sul – UFRGS, Avenida Bento Gonçalves 9500 - Prédio 72, Porto Alegre, Rio Grande do Sul 91501-970, Brazil.
[1] Contract/grant sponsor: CNPq.

https://doi.org/10.1016/j.dnarep.2023.103510
Received 26 October 2022; Received in revised form 7 April 2023; Accepted 23 April 2023
Available online 26 April 2023
1568-7864/© 2023 Elsevier B.V. All rights reserved.

NER-associated proteins; (iv) which open questions would be suited for this technique to answer; and (v) where can we go from here. In this work, we will tackle these questions, revising and critically discussing the results published in the context of NER proteins. We will focus on the core canonical proteins described as essential protagonists: the XPA to XPG proteins and their major molecular partners (RAD23B, CETN2, DDB1, TFIIH, ERCC1, RPA70, RPA32, and RPA14). MD works performed exclusively in DNA models were not considered. We hope that the debate conducted in this work aid in the decision-making process of new research applying macromolecule simulation in DNA repair-associated diseases.

2. Molecular dynamics: proteins in motion

Analysis of multiple structural conformations is directly pertinent to the protein's function. In this sense, wet-lab techniques such as nuclear magnetic resonance (NMR) and cryo-Electron Microscopy (cryo-EM) bring valuable information regarding the structural polymorphism while in solution [9,10]. Cryo-EM has been employed to investigate the implication of proteins' structural polymorphism to a variety of biological processes, such as allostery [11], subunit cooperativity [12], and function plasticity [13]. Moreover, the development of time-resolved cryo-EM allows obtaining an ensemble of structures in a non-crystal configuration while retaining time-progress information [14]. While cryo-EM is an excellent method for structural biology studies, an alternative approach that allows the investigation of structural features, molecular mechanisms, and kinetic and thermodynamic properties is MD [15,16].

In our context, MD analyzes all variations in orientation and interaction between residues, the content of secondary structures, and information about biomolecule interactions as a function of a predefined time. These conformational properties are based on Newton's second law, also called the Fundamental Principle of Dynamics. Thus, the force employed on each atom changes as a function of its position and the position of the other atoms with which it interacts [17]. MD analyses depict a variety of structural phenomena whose level of information is not possible in the stationary view of the structures obtained, for example, by crystallography, such as flexibility, which is achieved through the temporal variation of atoms coordinates of physicochemical state functions, such as the temperature.

MD is influenced by choice of the force fields (FFs), which contain a set of parameters that reproduce different experimental data to determine the intra and intermolecular forces between atoms. FFs assume that the energy of any system can be described by the sum of different physical potentials, such as bond deformation, bond angle, and electrostatic, among others [18]. Additionally, FFs assume that energy potentials applied to a small group of molecules with similar chemical groups can be used for a broad array of molecules with the same characteristics. However, not all FFs have this transferability. For more on the different types of FF (all atoms, united atoms, and coarse-grained) and their specificity, refer to [18,19].

The strengths of MD rely on their versatility, where they can be applied to a wide array of isolated biomolecules, protein complexes, protein-ligand systems, nucleic acids, carbohydrates, peptides, and so on. Its versatility also lies in the choice of FF, which not only can be chosen based on available computational resources but can also be parameterized (modified) for distinct molecules [20]. Finally, MD allows adjusting different environmental parameters. For more on the various aspects and applications of MD, please see [17,21].

Generally, every MD work has three interdependent core bottlenecks: (i) obtaining a reliable structure for input and correct FF parameters; (ii) employing the appropriate MD method to investigate the phenomenons of interest; and (iii) understanding the biochemical nuances of the system. The first is pivotal for the entire work; the most meticulous analyses and diacritic observations become meaningless once the input is erroneous. The second one depends on the biological question. Typically, the central quandary is the trade-off between computational cost and exactness. Various MD techniques have been developed, each with advantages in specific situations; however, most techniques are not straightforward, demanding a higher level of expertise, thus, limiting their application. The third bottleneck must lead the decision-making. Often, strictly computational methods provide only a "good enough" input to answer specific questions. However, experimentally resolved structures and multimolecular systems and their interactions provide information that eliminates the need to use techniques with inherent high inaccuracy, such as protein modeling (especially *de novo* approaches) and molecular docking.

For example, modeling techniques to obtain unresolved protein structures may successfully pass through validation metrics, and molecular docking protocols may provide consistent low-binding energy between ligands. However, rather than being seen as conclusive, those results should be appreciated as indicative of the molecular mechanism. Hence, understanding the biochemical characteristics of the system should be the major factor in decision-making. This issue is more distinguishable in molecular docking cases. A ligand artificially docked at a protein's pocket may present the lowest energy. Yet, that pocket might be unavailable under biological conditions, like in the context of complex formation and/or competitive ligands. Furthermore, obtaining reliable protein structures and FFs parameters in protein modeling also demands a deeper understanding of biochemistry. Comparative modeling tends to generate the most reliable structures [22]. Fortunately, recent significant breakthroughs in the *de novo* protein modeling field emerged with AlphaFold 2 [23], and RoseTTAFold [24], allowing researchers to obtain workable structures without demanding resolved, comparable structures to serve as templates. For more on the latest advancements in protein prediction, please refer to [25–27]. Moreover, automated modeling software usually lists possible models, ranking them by their energy score, but the lowest energy structures might not accurately describe folding - the biggest reason being the difficulty of spotting erroneous output as it is for molecular docking. Lastly, without the correct structure as input, including folding and protonation state, the results will not be as accurate [28].

The reader must keep in mind the take-home message: understanding what physical-chemical information each method gives and what information it omits as a trade-off for computational performance should be considered according to the physical-chemical and kinetics parameters pertinent to the phenomenon of interest. For more information on this topic, please refer to [29,30].

However, one major limitation of MD lies in the fact that FFs ignore electronic movements in a system - thus, being based solely on atomic localization [17]. Therefore, being incapable of analyzing catalytic reactions and quantum effects.

3. NER: a brave new world of opportunities

The NER pathway focuses on removing bulky DNA lesions generated by exposure to non-ionizing radiation, majorly 6–4 photoproducts (6–4PP) and cyclobutene pyrimidine dimers (CPD) (Fig. 1**A-B**) [31,32]. The core NER machinery is coordinated by the proteins XPA to XPG, which closely work together with other proteins to promote the recognition, strand-opening, damage excision, and polymerization of a new strand in damaged areas. NER has two branches, global-genome NER (GG-NER), which acts on the entire genome (Fig. 2), and transcription-coupled NER (TC-NER), which is restricted to removing lesions in transcriptionally active areas. The only difference between the two branches is that in GG-NER, the lesions are recognized by XPC and DDB2 (the product of the *XPE* gene), whereas in TC-NER, it is made majorly by CSA, CSB, and XAB2 [31,33,34]. In GG-NER, XPC has a higher affinity for 6–4PP, whereas DDB2 displays an increased affinity for CPD [35–37]. Nevertheless, XPC is still recruited to the lesion site for subsequent steps after the recognition by DDB2 to aid the proper positioning of the TFIIH complex [35–37]. For more on the structural

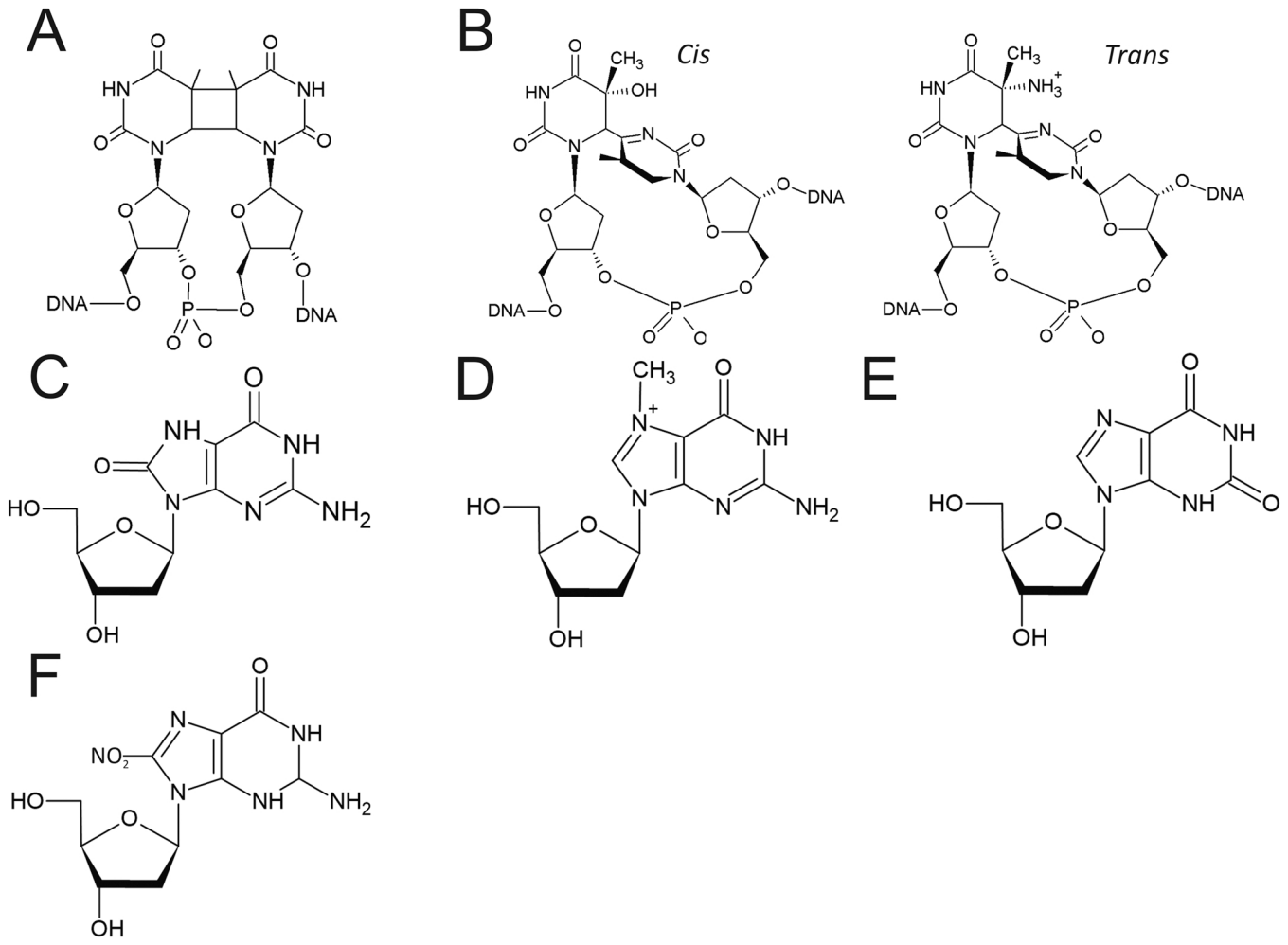


Fig. 1. Examples of common DNA lesions. A) Cyclobutane pyrimidine dimer (CPD) between two adjacent thymines; B) 6,4-photoproducts (6–4PP) between two adjacent thymines in *cis* and *trans*. C) The 8-oxoguanine (8-oxoG) oxidative damage; D) The N7-Methyl-2′-deoxyguanosine (N7-mG), an example of DNA damage by alkylation; E) A xanthine, a type of DNA damage by deamination of guanine; F) The 8-nitroguanosine (8-NO_2G), a type of nitration DNA damage.

features of XPA to XPG, refer to [38,39], and the major molecular partners that act together with the XP proteins (RAD23B, CETN2, DDB1, RPA70, RPA32, RPA14, ERCC1, and the TFIIH complex) refer to [38–40]. For detailed information about GG and TC-NER pathways, refer to [34,41].

For a long time, the only available complete structure of a human XP protein was DDB2 in complex with DDB1, like the one obtained by [42], and the core domain of XPA, such as obtained by [43]. However, lately, large models of XPF in complex with ERCC1 [44] and the TFIIH complex containing XPB and XPD were acquired [45]. A slightly large area of XPA was also published [46], as well as the catalytic domain of XPG [47]. Hence, the application of MD was never as relevant as they are now for NER.

The following section aims not to list all possible parameters used by authors but to critically discuss how the systems are designed and how these choices might impact the knowledge of NER. We delegated to Table 1 the summary of some technical decisions that influenced the discussion.

3.1. Unraveling the lesion recognition mechanisms of XPC (Rad4) using MD

XPC is a 940 amino acid protein that recognizes bulky DNA adducts, especially 6–4PP [37], in NER. XPC has nine distinguishable domains and PPI regions, most located in the C-terminal half [39]. Although we currently lack the complete 3D structure of the human XPC, a large model of its yeast homolog, Rad4, is available (Fig. 3A). There are no MD works on XPC, but numerous works on Rad4.

A study by Pitta and Krishnan explored the free energy profile of Rad4-associated CPD base flipping in two different systems, bound to DNA and DNA-free [49]. The authors performed a modeling step to fill a missing loop in the β-hairpin domain 2 (BHD2) of Rad4 (PDB 2QSH) and two missing disordered thymine in the damaged strand (Table1). MD simulations were performed for 20 ns, followed by umbrella sampling. Umbrella sampling is a technique that samples conformational dynamics along reaction coordinates to improve ergodicity and calculate the free energies of different states. For more, refer to [62]. The authors employed it to quantify the energy-revolving base-flipping in Rad4-free and Rad4-bound DNA. The pseudodihedral angles of a pair of vicinal, mismatched thymines were chosen as collective variables.

By comparing the lowest and higher energy states of DNA-bound Rad4, the authors identified key residues that take part in the interaction. From the lowest energy profiles (stable) and highest (transition), Asn554 (hs Asn754*), Phe556 (hs Phe756*), Asn558 (hs Asn758*), Phe597 (hs Phe797*), Phe599 (hs Phe799*), and Pro607 (hs Pro806*) appears as conserved residues with the strongest interaction. However, non-conserved residues, such as Arg494 (hs Asp688) and Arg515 (hs Gly719), were the ones that displayed vital roles in stabilizing the extrahelical conformation of the CPD. In fact, the work of Cai et al., which investigated the free-energy profiles in base flipping of

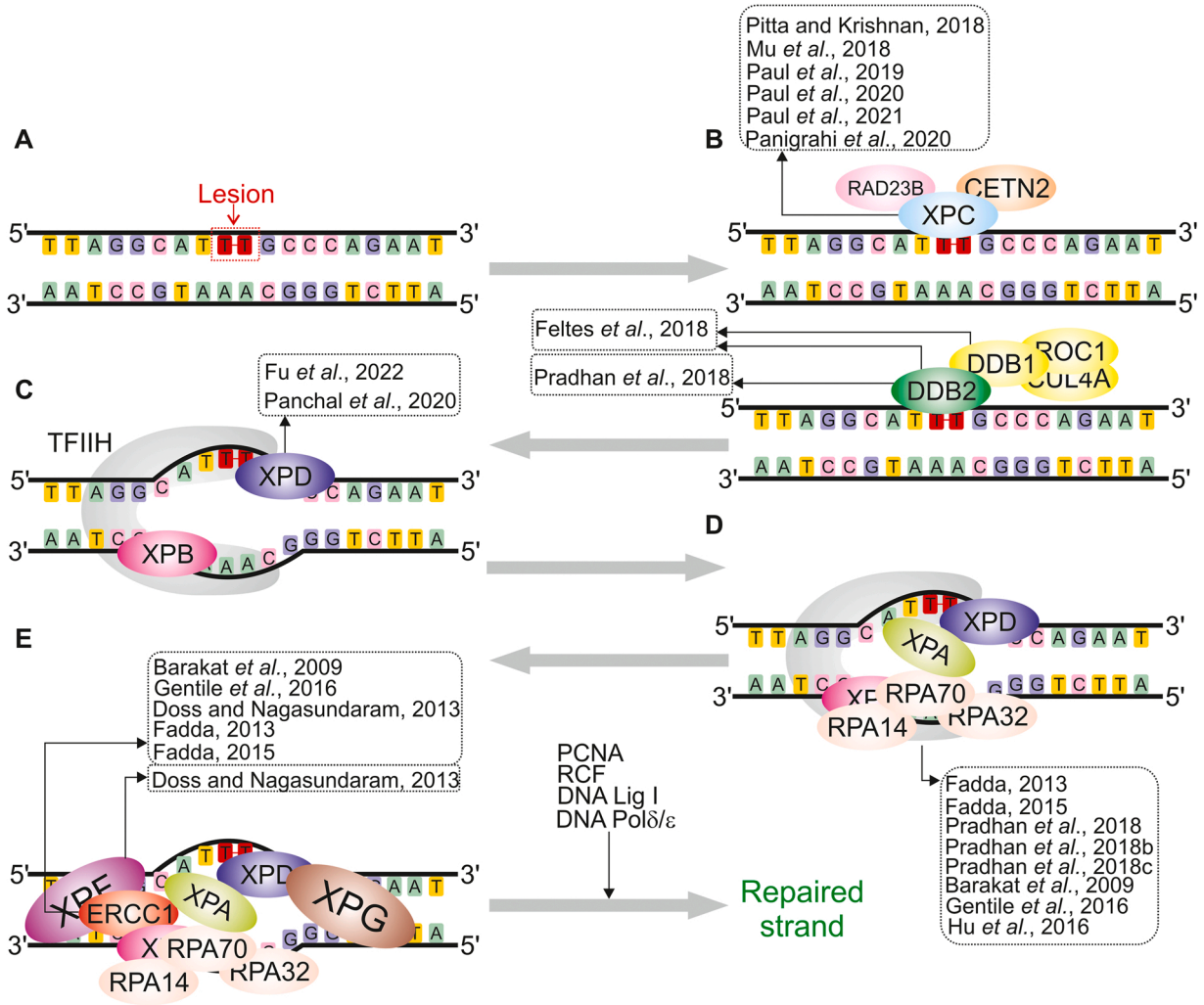


Fig. 2. Global-genome NER. A) Hypothetical DNA with a CPD lesion; B) Recognition step by either the XPC-CETN2-RAD23B or the DDB2-DDB1-CUL4A-ROC1 complexes; C) Strand opening by the TFIIH complex; D) Pre-excision complex formation by XPA-RPA; E) Lesion excision by the endonucleases XPF and XPG. The dotted squares list all MD studies reviewed in this work.

carcinogenic polycyclic aromatic hydrocarbons (PAH), and their relationship with Rad4 recognition, also observed that Phe599 in BHD3 would take part in the insertion and stabilization of Rad4 in the PAH adduct region. Their model also indicates that Arg519 in BHD2, which is in close vicinity of Arg515, forms a hydrogen bond with the DNA backbone, potentially partaking in the stabilization of Rad4 during the recognition of PAH [63]. However, Arg519 lies in a modeled region; thus, the accuracy of its proper coordination might not be precise.

The work of [49] presented an interesting setup, exemplifying that meaningful results can be obtained from MD analyses not focused on global structural behavior. From our perspective, one interesting aspect of this work is the identified CPD-interacting residues regarding their equivalent in the human XPC. Although some of the before-mentioned residues are conserved between *H. sapiens* and *Saccharomyces cerevisiae*, Arg494 (hs Asp688) and Arg515 (hs Gly719) are not. Nevertheless, Asp688 is next to an Arg residue. It is highly possible that, in humans, Asp688, a negatively charged residue, is repelled by the flipping CPD, allowing the adjacent Arg to make contact with the lesion. No missense mutation in the *H. sapiens* corresponding residues is found in XP-C patients, but several are located in the corresponding domains (BHD1–3) [39]. Still, one in-Frame (deletion) mutation was reported in this region (Glu740del), which would partially abolish the XPC-DNA interaction [64].

Panigrahi et al. investigated the flexibility and flipping of the mismatched bases, β-hairpin insertion into the DNA duplex, and the energy related to the Rad4-DNA association process [50]. The authors constructed the Rad4-DNA models based on the crystal PDB 2QSH. MD simulations of 50 ns were performed for each system, one Rad4-free and three Rad4-bound mismatched DNA. Lastly, umbrella sampling was used to compute the potentials of mean force using the weighted histogram analysis method.

The pseudo-dihedral angles between bases in different DNA strands (CCC/CCC, TTT/TTT, and TAT/TAT) were selected. For the β-hairpin insertion analysis, the reaction coordinates were described by the distance between the center of mass of the backbone heavy atoms of three residues (Ser603, Thr604, and Val605) and the center of mass of the ring heavy atoms of the nucleotide. Lastly, to calculate the energy for the association between Rad4 and the DNA, the distance between the center of mass of the backbone heavy atoms of the two mismatched paired bases of interest and the center of mass of the backbone heavy atoms of the residues that constitute Rad4's binding pocket for the flipped pair bases were chosen.

Free energy analysis of the base flipping process indicates that, in the absence of Rad4, mismatched bases favor intra-helical conformation. When Rad4 binds to DNA, the mismatched bases undergo a conformational shift to an extrahelical configuration where Rad4 binding pocket residues stabilize them. Compared to the other systems, CCC/CCC showed more conformational variability and quicker mean base flipping

Table 1
Summary of the technical background of the reviewed works.

Molecule	Modeling tool	Validation	PDB ID and obtention method*	Software	FF	Simulation time	Free energy calculation	Refs
Rad4 DNA (6–4PP) DNA (CPD)	NA	NA	6CFI (X-ray crystallography)	AMBER 16	Amberff14SB (charge correction)	4 μs each	NA	[48]
Rad4 DNA (CPD)	Modeller PSFGEN VMD	NA	2QSH (X-ray crystallography)	Gromacs 4.5.5	Amber99bsc1	20 ns	Umbrella sampling	[49]
Rad4 DNA (mismatch)	Modeller	NA	2QSH (X-ray crystallography)	Gromacs 2016.3	Amber99bsc1	4 systems (50 ns each)	Umbrella sampling	[50]
Rad4 DNA (WT)	NA	NA	2QSF and 2QSG (X-ray crystallography)	AMBER 16	Amberff14SB (charge correction)	1.5 μs each	NA	[51]
Rad4 DNA (matched)	NA	NA	4YIR, 6UIN, 6UBF (X-ray crystallography)	AMBER 16	Amberff14SB	3 μs each	NA	[6]
Rad4 DNA (matched)	NA	NA	6UG1 (X-ray crystallography)	AMBER 18	Amberff14SB	2 μs each	NA	[52]
DDB1 DDB2	Modeller PHYRE 2	PROCHECK Anolea, Qmean	3EI4 (X-ray crystallography)	Gromacs 4.5.1	GROMOS 54a7	9 systems (600 ns each)	Metadynamics	[53]
XPA DDB2	I-TASSER NA	RAMPAGE, ProSA, Molprobity NA (DDB2)	4E5Z (X-ray crystallography)	AMBER 14	Amberff99	3 systems (40 ns each)	No	[54]
XPA ERCC1	No	No	2JNW (NMR)	NAMD	Amberff99SB	50 ns	MM-PBSA	[55]
XPA ERCC1	No	No	2JNW (NMR)	Gromacs 4.6.3	Amberff99SB-ILDN Amberff99SB	5 systems (1 μs each)	No	[56]
XPA	NA	NA	1XPA (NMR)	AMBER 12	Amberff99SB-ILDN (ion parameter correction)	3 systems (100 ns each)	Free-energy landscape	[57]
XPA DNA	No	No	1XPA (NMR)	AMBER 12	Amberff99SB (XPA free) Amberff99bsc0 (with DNA)	2 systems (40 ns each)	No	[58]
XPA (homo)	Partial (RaptorX)	No	1XPA (NMR)	AMBER 12	Amberff99	4 systems (60 ns each)	MM-PBSA	[59]
XPD	NA	NA	6RO4 (Electron microscopy)	Gromacs 5.0.1	OPLS	2 systems (25 ns each)	NA	[60]
XPD DNA (WT) DNA (6–4PP)	NA	NA	6RO4 (Electron microscopy)	AMBER 18	Amberff14SB (charge correction)	28.5 μs total	NA	[7]
XPF ERCC1	No	No	2A1I and 2A1J (X-ray crystallography)	Gromacs 4.5.3	GROMOS96 43a1	4 systems (20 ns each)	No	[61]

* for studies that used (or generated) more than one PDB, we listed the ones we mentioned in the discussion. NA: Not applicable.

times. Regarding β-hairpin insertion into the DNA duplex, the estimated free energy cost of β-hairpin insertion was relatively lower for CCC/CCC compared to the other systems. This might result from the increased level of DNA distortion brought on by mismatches in CCC/CCC, making it easier for favorable β-hairpin insertion. Lastly, the energy related to the Rad4-DNA association process reveals that BHD2/BHD3 domains in association with the mismatched site have an overall role in regulating the stability of the Rad4-DNA complex. Furthermore, Rad4 interacts favorably with CCC/CCC mismatched bases considerably earlier during association than with TAT/TAT or TTT/TTT mismatched bases.

This work is a fine example of how MD coupled with robust energy of interaction analyses can aid in understanding the lesion recognition process. The setup used by the authors could be an example for further studies, especially using different DNA templates. Lesion recognition does not occur with similar efficiency for all types of damage. For example, it is known that XPC has a low specificity for CPD, whereas the UV-DDB complex has a higher affinity for it but a lower for 6–4PP [42, 65–67]. Therefore, studying the mechanistic process behind lesion recognition for different DNA sequences is highly advisable. However, researchers must focus on their research question when dealing with energy calculations because the rationale behind the choice of collective variables to study DNA-protein interactions can be distinct from the study of PPI.

Also, using the Rad4 protein structure, Mu et al. explored its binding to a library of DNA lesions caused by carcinogenic PAH [51]. The initial model was constructed by taking the apo Rad4 (PDB 2QSF) and positioning the transglutaminase domain (TGD) and BHD1 according to PDB 2QSG; the enzyme opened state. The structure was docked with lesion-containing DNA (Table1). MD simulations were carried out for 1.5 μs each. Measurement of twist angles was based on an ``untwist'' score, which is calculated as Untwist = $Twist_{start}$ - $Twist_{\textit{initial binding state}}$, where $Twist_{start}$ is the ensemble average twist angle of the lesion-containing 6-mer during the first 1 ns of simulation and; $Twist_{\textit{initial binding state}}$ is the twist angle of the lesion-containing 6-mer for the structures in the initial binding state ensemble (1–1.5 μs). Positive values for the Untwist score indicate an increase in the untwisting process, whereas negative values depict twisting. The volumes and occupancy of the binding pockets in the DNA were investigated, and the total occupied volume was calculated to quantify the extent of BHD2 binding into the DNA minor groove.

The authors argue that, generally, in well-repaired lesion-containing duplexes, the pocket between BHD2 and BHD3 is responsible for capturing the lesions, in which the BHD2 is primarily responsible for binding the minor groove of the DNA molecule. In repair-resistant lesion-containing duplexes, BHD3 does not participate in base capturing, and BHD2 shows only marginal interaction and low binding energy with the DNA backbone. Finally, in the moderately repaired lesion-containing duplexes, most interactions rely on BHD2 to stabilize the lesion, displaying diverging affinities for the different lesions. Thus, BHD2 appears as a significant protagonist in overall lesion recognition.

This work allowed the understanding of the mechanistic process behind Rad4/XPC lesion recognition, being an example of how MD can

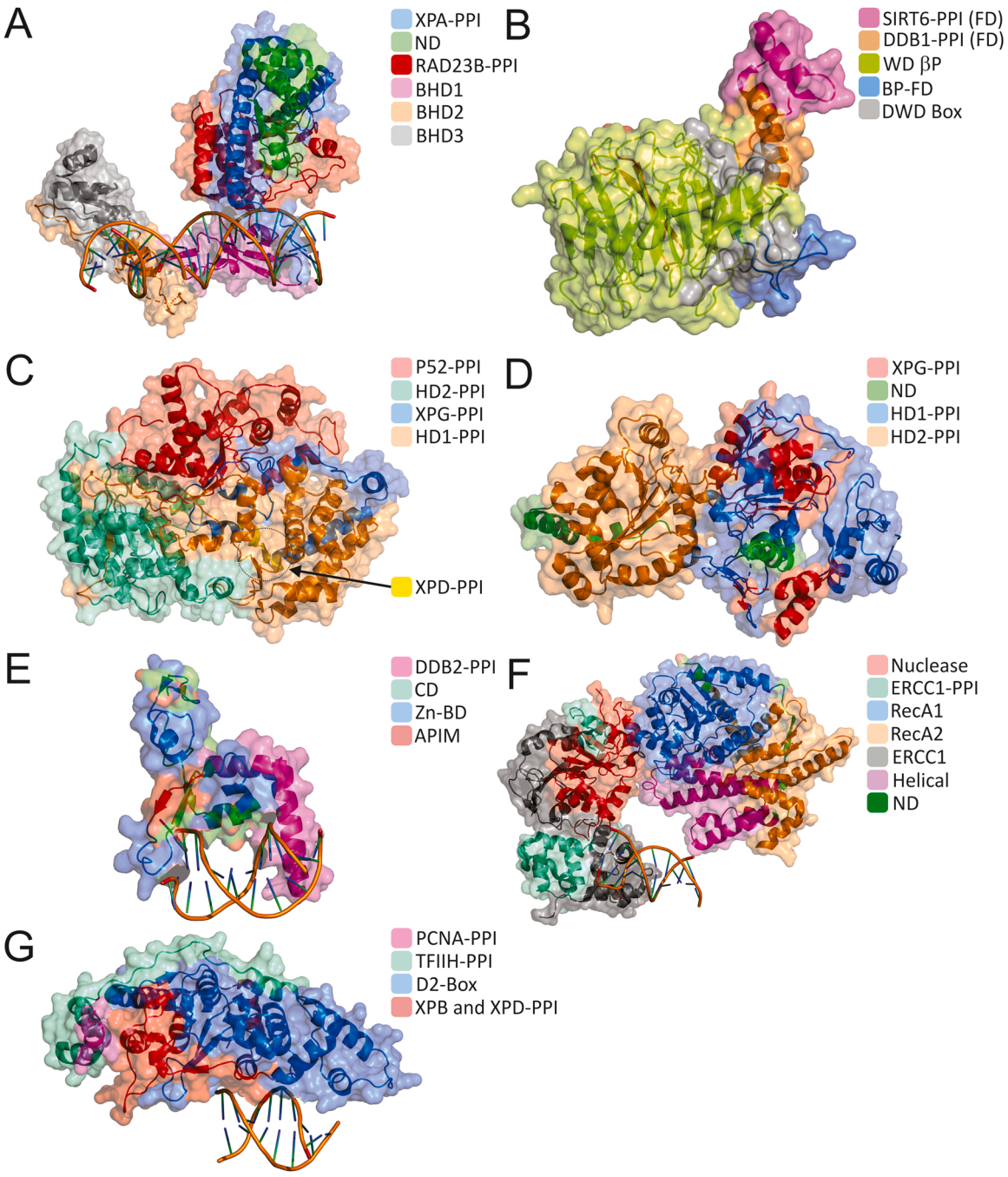


Fig. 3. The structure of the xeroderma pigmentosum proteins. All domains were colored according to [38,39]. A) The 3D model of the yeast homolog (Rad4) of the human XPC protein (PDB:6UBF) [73]. The *Saccharomyces cerevisiae* protein was depicted because no model of the human XPC is currently available; B) Structure of the human DDB2 protein (PDB: 3EI4) [42]; C-D) Models of the human XPB and XPD proteins, respectively, extracted from PDB 5OF4) [45]; E) The 3D structure of the human XPA (PDB: 6LAE) [74]; F) Structure of the human XPF protein bound to ERCC1 (PDB: 6SXB) [44]; G) The model of the catalytic domain of XPG (PDB:6TUX) [47].

aid the understanding of DNA repair-associated phenomena. It explores the roles of NER in dealing with other DNA adducts, which are known to trigger NER and are far less studied than their interaction with 6–4PP and CPD.

Paul et al. obtained the crystal structure of the Rad4–Rad23, equivalent to human XPC-RAD23B, bound to a 6–4PP DNA lesion and performed MD to investigate the mechanism of lesion binding [48]. Based on the obtained crystal, the author performed MD to investigate the initial binding of Rad4 to DNA containing 6–4PP and CPD.

They retrieved the apo-Rad4 (PDB 2QSF) and docked it to the mismatched DNA by positioning the TGD and BHD1 domains in their open form based on the structure PDB 2QSG (Table1). The base pairs at the lesion sites were modeled by replacing the lesion according to other PDB structures: TT 6–4PP partnered with GA (PDB 1CFL); CPD with normal AA partners (PDB 1T4I). Thus, the 5′ G of 1CFL was replaced with an A, resulting in a new DNA template of 6–4PP partnered with AA (PDB 6CFI), which was used for the MD analyses. While some parameters could be retrieved from different works, partial charges for the 6–4PP lesion were assessed by Hartree Fock quantum mechanical calculations with 6–31G* basis set. The docked protein-DNA complex was submitted to MD with a restraint on the distance between BHD2-BHD3 domains and the lesion site, providing an initial configuration before the

engagement between BHD2, BHD3, and the lesion. Simulations were conducted for up to 4 μs each.

While crystallography depicts a frozen recognition state, MD simulations reveal the complex formation process. This way, it is possible to understand the conformational selection behind the interaction path and determine the kinetics and thermodynamics parameters, such as k_{on}, k_{off}, binding energies, solvent contribution, etc.

The results presented evidence that 6–4PP-DNA untwists and bends while, at the same time, the pair of damaged nucleotides flips out upon its initial binding to Rad4. Aside from corroborating substantial differences in the binding process between Rad4 and the different photolesions, the authors observed that the BHD2 had more extensive interactions with the minor groove in 6–4PP than with the CPD, where Val517 played a significant role. This is also in agreement with biochemical studies that revealed an affinity of XPC for 6–4PP [66,68]. This notion was further explained by the MD analysis, which showed that the CPD could resist its unwinding by the BHD2, which is required for extrusion - thus, avoiding recognition. Likewise, Rad4 made no contact with the lesions, relying exclusively on the interaction with non-damaged residues, suggesting it can recognize a broad range of substrates.

Following the notion of Rad4 recognizing a broad range of substrates, Paul et al. studied the effect of Rad4 mutants lacking the hairpin from BHD2 or BHD3 domains in binding to CCC/GGG DNA [6]. This study used crystallography, fluorescence-based techniques, and MD simulations to examine the mechanism by which Rad4/XPC unwinds, opens, and flips nucleotides, whether from mismatched (CCC/CCC) or matched (CCC/GGG) DNA. Focusing on the MD perspective, a simulation was performed of three different systems: WT, mutated BHD2, and mutated BHD3 (lacking regions 515–527 and 599–605, respectively), all of them with CCC/GGG DNA. The simulation time required to achieve stable BHD2 conformations demanded roughly 1 μs; thus, 2 μs of simulation time was carried out for each system based on their initial structures. Afterward, ten simulations of 100 ns for each system were carried out based on the best representative structures, comprising a total of 3 μs for each system (Table1). Specific analyses, such as DNA untwisting upon the binding of the Rad4 mutants, DNA bend angles and bend direction pseudo-dihedral angles, and base extrusion pseudo-dihedral angles, were performed following previous setups [48].

Fluorescence-based techniques demonstrated that mutated BHD2 and BHD3 had slightly less affinity for matched CCC/GGG but significantly decreased affinity for mismatched sequences, indicating that Rad4 could recognize such mismatch. Crystallography coupled with a site-specific disulfide tethering strategy using the WT and mutant Rad4 proved that although required for recognition, the process of creating the open DNA conformation (i.e., with flipped-out nucleotides) is facilitated by, but majorly does not require the full BHD2 or BHD3. Using the obtained crystals, MD simulations indicated that the mutated BHD2 domain could form stable complexes with CCC/GGG during the first 1 μs. Still, it failed to bind the minor groove of the DNA, previously discussed to be a key role of this domain [51]. However, full BHD3 still engaged the major groove, where Phe599, Arg601, and Ser603 interacted closely with the potential flipping cytosines. Moreover, bending directions also changed in the investigated systems: mutated BHD3 showed DNA bending toward the major groove, whereas mutations in BHD2 displayed the opposite. Twist and untwist angles revealed that BHD2 facilitates the untwisting, whereas BHD3 is more associated with base extrusion. Thus, this work greatly expands the results of [51] and [48], revealing how the conformation by which Rad4 binds to other DNA substrates aside from bulky DNA lesions.

Finally, another study by Paul et al. studied how Rad4 is bound to CG/CG-containing duplexes [52]. The setup for the MD simulations was done according to [6]. Each system, on the other hand, was simulated for 2 μs each.

The first notable result was that the BHD2–3 in the CGC/CGC construct, which was the one that diffracted the best, displayed a reversed 180° orientation, where the C-terminal of BHD2–3 faced the short end of the DNA, thus, not bound to the opened region in CCC/GGG like observed in their previous study [6]. In contrast, the TGD and BHD1 were attached to the large end portion, where the DNA sustained a closed form. The authors argued that such a striking difference is that the CGC/CGC posed a stronger resistance to its open state, modifying the way Rad4 engaged DNA. MD results revealed that CGC/CGC exhibited resistance to base flipping, similar to CPD [48]. Most importantly, BHD2 was repelled from the minor groove, now residing partially in the major groove, impeding BHD3 from properly binding its preferred site.

As extracted from the discussed MD studies, XPC appears to be a versatile recognition factor, adaptable to a wide range of DNA lesions and templates. In this sense, MD consistently observed how the BHD2 interacts with the minor groove of the DNA and stabilizes Rad4/XPC in the template. Likewise, the cooperation between BHD2 and BHD3 in the stabilization and lesion recognition was also observed in more than one study, reinforcing the existence of this molecular behavior. A particular highlight must be given to the works of [6,48,51], and [52], which tell a single narrative. They revealed peculiarities of the BHD2–3 in the recognition process of PAH-derived lesions, 6–4PP, CPD, and different DNA-matched and mismatched sequences. Their works clearly show that the BHD2 and BHD3 are crucial in stabilizing Rad4/XPC in the recognition site, but how this process occurs strongly depends on the substrate. These studies are also a great example of how MD aids in the validation and elucidation of experimental data in better detail. Due to their sound technical choices, the MD setup followed by [6,48,51,52] could be used as examples for further similar studies.

3.2. *Elucidating the PPI dynamics of DDB2: a case of structural context*

DDB2 is a DNA-binding protein composed of 427 amino acids and is responsible for recognizing non-destabilizing lesions, such as CPD, in NER [37]. DDB2 is divided into four known functional and PPI regions, where most of its functions, including the residues that interact with the DNA, lie in its WD-β-propeller (Fig. 3B) [38,39]. The N-terminus of DDB2 is strongly associated with its binding to DDB1 and also contains multiple posttranslational modification sites related to its fine-tuning in NER, such as its ubiquitination sites, which are necessary for the displacement of DDB2 from the lesion so that the TFIIH complex could adequately bind the region [39]. Recent data also indicates that DDB2 is required for proper 8-oxoG lesion recognition during BER by aiding OGG1 recruitment [69]. As previously stated, XPC has a higher affinity for 6–4PP and helix-distorting lesions. Since CPD does not display the same level of DNA topological distortion, XPC inefficiently binds to CPD with slow kinetics. The UV-DDB complex, on the other hand, can confer the necessary chromatin modifications to the CPD site to allow XPC to identify the helix-distorting region and bind it with higher affinity for subsequent NER steps [37].

Based on the earlier work of Wakasugi et al. [67], Pradhan et al. used MD to characterize the interface between $XPA_{185–226}$ and DDB2 [54]. To the date of this review, we have no model of the XPA-DDB2 interface that has been previously proved to exist in vivo and in vitro [67]. Thus, the authors docked the described XPA region in the DDB2 model, where they tested various conformations, generating an artificial PPI model, and selected the ones with the lowest energy for a 40 ns MD analysis. They described a stable XPA-DDB2 complex, where $XPA_{185–226}$ would bind the flexible N-terminal region of DDB2 (Fig. 3**B**) and listed the interface residues between those proteins, shedding light on a PPI connection which we know little about. However, some limitations must be addressed, extending to other MD works. First is the simulation time. DDB2 is a β-propeller, an evolutionarily conserved protein family widely known to have high stability due to the numerous intra-molecular interactions [70,71]. Hence, large, highly stable proteins such as DDB2 might not show any meaningful structural changes in 40 ns, especially if bound to another partner. Secondly, although they studied the complex in three different models, each system had no replicates, a practice

gradually becoming normative in MD studies, offering robust comparisons and a more accurate depiction of molecular phenomena [72]. Due to the dynamic nature of proteins, they can undergo some structural changes considered a low probability phenomenon. In other words, the conformational changes might not be frequent, thus, not characterizing a WT behavior. Third, there are some concerns about how this PPI dynamic between XPA and DDB2 might occur based on new data, as explained further.

We characterized the conformational changes associated with the impact of three XP-related naturally occurring mutations on the DDB2-DDB1 complex [53]. Using the available 3D structure of the human DDB2-DDB1 complex [42], we employed MD and dynamic-residue interaction networks (DRIN) to uncover how the naturally-occurring mutants Arg273His, Leu350Pro, and Lys244Glu influenced the DDB-complex. We simulated DDB2 and its mutants alone and complexed to DDB1 using replicates for each system during 200 ns each, comprising 600 ns per system (Table1). MD analyses revealed unique conformations for DDB2, lost in all three mutations: (i) all DDB2 mutants impacted a new region (354–371), which lost flexibility; (ii) a new clamp-like movement was described for DDB1; and (iii) unique molecular behavior was detected for the uncomplexed WT DDB2 compared to its mutants, where the WT form clearly showed greater mobility. Finally, DRINs revealed a broad loss of flexibility in the mutant proteins and allowed to detail the interface residues between DDB2 and DDB1 that were conserved and altered in each mutation.

We can also draw limitations from our approach. The first is the use of a united-atom FF. Although GROMOS 54a7 displayed excellent results in the benchmark [75], we cannot exclude other discrete structural changes that could have been noted if an all-atom FF had been used. Secondly, even though we utilized replicates with an overall extended simulation time (600 ns total), we must recognize that for a large structure such as the DDB-complex, a longer simulation time would have revealed more clues on the mutation's effects. For example, Leu350Pro displayed the most aggressive impact, impairing the interaction between DDB2 and the β-propellers A and C of DDB1. A more extended simulation might have better investigated this abrogation of PPI [53]. Finally, the last limitation would be the calculation of interaction energy using only GROMACS and not another more robust approach. However, we indeed calculated the energy of interaction of DDB2 and DDB1 using metadynamics, which corroborated with the interaction energy calculated by GROMACS (unpublished results but accepted by the reviewers).

DDB2 flexible domain has a tail form inserted between the BPA and BPC of DDB1, whereas the BPB is solely dedicated to interacting with CUL4A [40]. The work of [67] identified an association between XPA and DDB2, but although the region in which this PPI appeared was mapped in XPA, it was not in DDB2. Recent data showed that the DDB2 must leave the lesion site shortly after its recognition so that XPC can properly bind the DNA and recruit the subsequent machinery [76]. Additionally, the recently obtained TFIIH-XPA model [46] indicates that XPA is oriented opposed to TFIIH, clamping it to the DNA. In that case, it is not viable that XPA would bind the flexible domain (res 1–101) because it would be occluded between two other β-propellers of DDB1 and could also impair the ubiquitination of DDB2 by the ubiquitin ligase complex. Considering the new TFIIH data and the PPI interface between DDB1 and DDB2, it is reasonable to infer that XPA is expected to be positioned opposite DDB2, making contact with its β-propeller instead of the flexible domain near DDB1. This conformation would also allow the dissociation of XPA and DDB2 to occur faster, as it appears to be necessary.

Therefore, the artificial PPI made by Pradhan et al. is unlikely to happen strictly as reported, as the flexible domain of DDB2 is not fully available for XPA binding. In agreement, most of the residues with the higher interacting score, which appeared in all three authors' models, were from this portion. In this sense, some models generated by Pradhan et al. listed some residues along DDB2 β-propeller, which could be considered potential XPA-binding residues. Nevertheless, they were not observed in all models, and many displayed a low interacting score.

3.3. Studying different aspects of the XPA structure and the enigmatic XPA-ERCC1 interaction

XPA is a 273 amino acid protein (Fig. 3E), present in the NER bubble from lesion recognition to damage excision, making PPI with a broad range of targets, a fact reflected in the amount of PPI regions it displays [38,39]. As previously discussed, XPA probably interacts with the lower region of DDB2's β-propeller. It also forms PPI with the TFIIH complex [46]. One of these PPI regions is with ERCC1, the primary molecular partner of the endonuclease XPF in NER [77]. While some XPA artificially-induced mutations in this region are associated with NER deficiency [78], no naturally-occurring mutations in the XPA-ERCC1 PPI region have been described so far [38,39].

The work of Barakat et al. [55] explored the XPA-ERCC1 interaction aiming to identify new potential compounds that abrogate their PPI. The work is based on three pillars: (i) XPA is associated with the recruitment of the XPF-ERCC1 heterodimer to the lesion site; (ii) XPF-ERCC1 overexpression was related to chemoresistance to cisplatin, and; (iii) patients with testicular cancer underexpressing either XPA or ERCC1 showed sensitivity to cisplatin treatment.

No modeling step was executed, and only 11 residues from XPA (out of 273) and 115 residues from ERCC1 (out of 297) were utilized. MD simulations during 50 ns were carried out for the XPA-ERCC1 complex [79] (Table1), and Root Mean Square Deviation (RMSD) clustering was used to extract representative complexes, as well as Principal Component Analysis (PCA). Molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) was used to investigate the binding energy, and the screening of potential inhibitory ligands from NCDIS and DrugBank was carried out. Ligands with rare elements were excluded, and different protonation states were considered. Lastly, pharmacophore construction was executed. The authors then manually determined patterns of interaction and chemical features associated with residues in the ERCC1 binding site.

This work set up for molecular docking is robust, exploring representative structures of ERCC1 instead of using the structure straight out of PDB. The adjunct usage of MD simulations to include receptor flexibility and different ways to assess solvation effects result in a solid docking protocol. Excluding ligands from databases with rare elements is adequate and avoids false positive hits resulting from incorrect parameterization. Regarding pharmacophore construction, a variety of software that is fully automated exists [80]. The concern to manually investigate patterns of interaction and main residues' chemical features is a good practice over blinding trusting software outputs.

The authors found a cryptic pocket in the ERCC1 structure, where the top 30 compounds displayed higher interaction. A total of 13 residues, including some associated with the experimentally described XPA PPI, such as Asn110, Asp129, and Ser142, were responsible for such interaction. The same group in 2016 remade the experiments using the same XPA-ERCC1 region and described new compounds that could affect their interaction. The cryptic pocket region remained similar [81]. However, employing only fragments of XPA and ERCC1 for simulations might introduce conformational disturbances that are impractical to evaluate and could compromise the results. Since this paper has been published for over a decade, the limitations are understandable. Moreover, screening ligand databases for drug discovery purposes should accompany ADMET [82] analyses to avert selecting compounds with low chances of being employed as an actual drug.

There is little to dissect from a NER perspective since the goal was not to observe the molecular behavior of the complex. Nevertheless, recent data still points to XPA being related to cisplatin resistance. For example, XPA was associated with cisplatin chemoresistance in melanoma cells by activating PARP1 [83]. Therefore, although data points to ERCC1 and XPA being associated with chemoresistance to cisplatin, experimental studies still need to confirm if the XPA-ERCC1 interaction is related to

the phenomenon. For example, the lack of ERCC1 was already correlated to the loss of XPF activity [84,85], an endonuclease with roles in the interstrand crosslink (ICL) pathway [86,87], which removes, as the name suggests, interstrand crosslinks, one of the lesions caused by cisplatin [88]. Hence, the inhibition of ERCC1 might be more related to errors in the ICL pathway than NER. However, XPA was also correlated to the ICL pathway [89,90]. More importantly, in human cells, the ERCC1-XPF complex requires XPA to properly be recruited to psoralen–DNA interstrand crosslinking sites [91]. Thus, although the XPA-ERCC1 PPI may be one of the culprits behind cisplatin resistance, it is plausible that the effects of the inhibition of XPA and ERCC1 in cisplatin resistance are not directly correlated to their PPI, but the disruption of other PPI performed by these proteins.

Moreover, Fadda explored the conformational preference of $XPA_{67–80}$ by the ERCC1 central region [56,92]. Due to their similarities, we will focus on [56] since we consider it the most relevant. The author devised different peptides to study the WT $XPA_{67–80}$: (i) a model where Ala replaced $XPA_{78–80}$: (ii) two other peptides containing point mutations to enhance the interaction with ERCC1 on XPA (Phe75Trp and Ile68Ly); (iii) and an $XPA_{67–69}$ triple Ala mutant. By simulating each peptide during 1 μs in 10 different systems, the author reported that all mutants increased the peptide's disorder, increasing the number of possible conformations it could achieve (Table1). The author concluded that there is a preference for ERCC1 for a given conformation of XPA corresponding to its WT state.

Comparing conformational populations is a practical approach that can produce valuable information for interactions we know little. However, [56] did not perform any energy of interaction analysis. Binding free energy calculations provide a quantitative understanding of the PPI and the mechanism behind the binding process [93], while distance-based metrics, such as RMSD and radius of gyration (R_g), contribute to a more qualitative understanding of the global process. Van der Waals interactions, electrostatic interactions, hydrogen bonding, solvent exclusion, and entropic changes are the processes that govern protein association [94]. Binding free energy calculations allow investigating the primarily physical processes driving the conformational changes and discrete interaction forces contributing to each step. For instance, the driving force in protein coupling is hydrophobic interactions, which leads to a gain in free binding energy [95]. Nevertheless, the distribution of this energy across the interface is uneven, and a small number of residues present at the PPI interface are the leading contributors to the free binding energy [96,97].

From a DNA repair point of view, not much was added to understand the XPA-ERCC1 interaction aside from ERCC1 having a conformational preference for the WT $XPA_{67–80}$. The significant limitation is studying the interaction using only small portions of the proteins since protein conformation in the cell environment is influenced by their complete structures. Excluding protein regions from simulations is a practical protocol when that portion has unknown functions or was proven unnecessary for a given process. However, that is not the case for the remaining structure of XPA. At the time, however, the structures of XPA and ERCC1 were not resolved.

Hu et al. [57] elucidated the role of the Zn^{2+} ion on the XPA structure and the impact of changing the canonical ion for (Ni^{2+}). The workflow is based on comparing MD simulation results among three distinct systems, $XPA_{98–210}$-Zn, $XPA_{98–210}$-Ni, and $XPA_{98–210}$-apo. The first resolved XPA structure (PDB 1XPA) was used as the model [98]. A parameterization step was conducted with Metal Center Parameter Builder for the systems containing a metal ion. First, it substitutes the protein backbone C atoms with methyl groups, where the B3LYP/6–31G* method is used to calculate a stable and convergent side-chain model. Next, each methyl group is capped with acetylamide (ACE) and N-methylamide (NME), allowing quantum chemistry calculations and the restrained electrostatic potential charge parameters (RESP). MD simulations were carried out for 100 ns for each of the three systems. Basic MD analyses, PCA, and free-energy landscape analyses were executed.

The author's concern with using quantum chemistry calculation to obtain optimized parameters for Zn^{2+} and Ni^{2+} is appropriate. These calculations may be laborious and demand expertise to perform. Yet, they are essential when working with structural transition metals with no parameters available for one specific ion center coordination. In this work, the authors focused on the impact of removing the Zn^{2+} or replacing it with Ni^{2+} in the zinc-finger domain. The results align with the initial question because the loss of proper geometry and coordination caused by the Ni^{2+} ion will likely result in functional biological changes.

From a NER perspective, this work answered some engaging questions. The authors determined that the exchange of the Zn^{2+} by Ni^{2+} abrogated residue contacts in the Zn-finger domain (98–129), one of the regions described to bind RPA70, altering the native conformation [99]. As expected, the absence of the Zn^{2+} ion also influenced the native conformation of the domain due to the effect of the solvent net, exhibiting a distinguishable molecular behavior and free energy landscape when compared to the Zn-containing domain.

Two missense mutations are described in the Zn-binding domain of XPA, both on residues that coordinate the Zn^{2+} (Cys108Phe and Cys126Trp) [39]. We argue that one explanation for the biological effects of these mutations is that the loss of the Zn^{2+} coordination could abolish not only DNA and RPA70 binding but also the preferable conformation orientation of XPA's central domain, restricting its DNA and protein interactions.

Moreover, Pradhan et al. [58] studied the interaction of $XPA_{98–210}$ with a thymine dimer. The authors selected a 112 residues length structure of XPA (PDB 1XPA), which was previously described to contain the Zn^{2+} [98,100]. However, the authors removed the Zn^{2+}, alleging that it would not impact XPA's structural behavior. A DNA decamer resolved by crystallography was retrieved (PDB 1N4E) and docked with $XPA_{98–210}$ (Table1). MD simulations were carried out considering two systems, one of free $XPA_{98–210}$, and the other composed of the $XPA_{98–210}$-DNA interaction. Simulations were performed for 40 ns in each system, and analyses, such as RMSD, R_g, Solvent Accessible Surface Area (SASA), and B-Factor, were executed.

The authors listed several residues that could be part of the XPA interaction with the lesion and compared the molecular behavior of the peptide-free and DNA-bound. However, the technical limitations follow the same pattern explained before about docking scores, lack of replicates in small simulations, and the lack of full protein models. From a biochemical perspective, removing the zinc ion before docking and MD analyses must be highlighted.

Removal of metal ions from a protein structure is not an adequate practice for MD. Metal ion coordination is precise, and depending on the nature of the metal ion and what residues it interacts with, the ion-residues cluster changes its geometry [101,102]. Due to charge transfer and polarization effect, metal ions as co-factors exert a structural role and coordination that standard amino acids cannot [103]. Zinc-residue clusters show this same behavior. Zinc presents the electronic configuration of $[Ar]3d^{10}\ 4s^2$; thus, it is not susceptible to ligand field stabilization effects since the orbital d remains complete even after ionization. However, the coordination polyhedron of zinc ion varies in number, charge, structure, and amino acid composition accordingly to its biological role [104,105], something that applies to XPA protein [100].

From an evolutionary perspective, a preserved motif, such as the zinc finger, has a vital role in functionality, structural stability, and/or recognition (for more, refer to [106]). Last, the relevance of structural zinc for MD is notorious, as well as the importance of correct parameterization, for which the AMBER group developed a FF to specifically help in the modeling of zinc ion center, called Zinc Amber Force Field (ZAFF) [107,108].

The assumption that Zn^{2+} ion could be excluded from the XPA structure in the work of [58] was based on the work of Hu et al. [57]. However, upon comparative evaluation, we noticed that the work did not show that Zn^{2+} is structurally nonessential, only that a few

distance-based metrics of simulations with XPA- Zn^{2+} and XPA-apo were similar. Thus, stating that the Zn^{2+} ion can be removed to investigate phenomenons that are dependent on the whole protein structure is questionable. Additionally, the original work recognizes that some results might be uncertain, whether because of the missing part of XPA or because the RESP charges calculated for the Zn^{2+} center required further correction. Hence, the possibility of some of the listed interactions being artifacts is highly possible, primarily because numerous DNA interactions were described in the Zn^{2+} binding region.

Finally, Pradhan et al. [59] studied another XPA system, this time to evaluate the molecular behavior of the XPA homo-dimer by comparing two different XPA peptides: $XPA_{98–210}$ and $XPA_{98–239}$.

The authors constructed the initial structures of the DNA binding domain of XPA based on PDB 1XPA (98–210) constituted one of the systems. A modeling step was performed to construct the $XPA_{98–239}$ system. The XPA sequence ranging from residue 208–239 was submitted to template-based modeling and chosen based on several metrics. The selection of homo-dimers (Table1) was based on lower energy, highest geometric score, and maximum surface area. MD simulations of two independent systems, $XPA_{98–210}$ and $XPA_{98–239}$, were carried out for 60 ns each. Besides basic MD analyses, binding free energy was calculated employing MM/PBSA and MM/Generalized Born Surface Area (MM/GBSA). Analyses of each isolated monomer, as well as the composing of the homo-dimer, were performed to close the thermodynamic cycle.

The authors reported residues that would take place in both XPA homo-dimer conformations and concluded that the extra 29 residues of $XPA_{98–239}$ enhanced the homo-dimeric form of XPA and its stability. The work presented an interesting research question not yet addressed in an MD study, mainly because XPA dimeric form appears to be its dominant conformation to bind the DNA molecule and RPA [109,110]. However, it must be noted that there are three RPA PPI regions on XPA, two on its central domain to RPA70 (res 98–129 and res 153–179) and one on its N-terminal (res 16–46) [39], indicating that potentially disordered N-terminus of XPA might be essential for the PPI. Additionally, no information about the zinc ion was mentioned in this work, implying that it could have been removed from the structure, imposing severe bias on the results because any interaction involving the XPA zinc-finger domain cannot be assumed independent of the Zn^{2+} ion, as demonstrated previously by Hu et al. (discussed above). No systems replicates were performed, and 60 ns of simulation is relatively short to sample the configuration space. Furthermore, attempting to determine stability based on RMSD is not adequate. RMSD measures the variation of a structure from its initial state; thus, it does not imply why such a phenomenon is happening. Therefore, RMSD alone is unsuitable for inferring outcomes such as “stability” or “perturbations”. Nonetheless, a positive prospect is the addition of a binding free energy calculation as enrichment for discussing PPI stability.

3.4. Studying XPD interaction with DNA and mutation effects

XPD is the ATP-dependent 5′→3′ helicase of the TFIIH complex (Fig. 3D), composed of 760 amino acids, which, together with XPB (Fig. 3D), is accountable for the double-strand opening after damage recognition. The structure of XPD was uncovered by Greber et al. in 2017 [45]. The structure of XPD shows PPI with at least five other proteins directly associated with NER, where most XP-causing mutations are located in its helicase domain 2 (HD2) [39].

The work of Fu et al. [7] studied the effects of 6–4PP in the stalling of XPD. By employing multiple long all-atom simulations, the authors observed how XPD behaved in the presence and absence of 6–4PP (Table1). To achieve their goals, the authors: (i) retrieved the Cryo-EM structure of the XPD-ssDNA complex (Fig. 3D); (ii) prepared three systems consisting of normal DNA, mutated DNA outside XPD’-’s pore, and mutated DNA inside XPD’s pore; (iii) excluded residues 273–235 from XPD due to hindrance of DNA entry pore; (iv) FeS cluster parameters were obtained separately from another work, and; (v) simulated each XPD system with production step ranging from 2 to 3.5 μs.

This work provides a substantial number of observations. First, RMSD of Cα atoms of the XPD pore indicates that after 1.5 μs, XPD from all systems exhibits a correct fold and behaves inside one metastable state. Before reaching 1.5 μs, XPD with mutated DNA inside the pore displays higher RMSD values. This observation occurs due to the mobility of residues 128–133 from the FeS α-helix, which is partially flipped away from the DNA. However, after 1.5 μs, the residues flip back toward the DNA, and the XPD shows similar dynamics compared to the other systems. Next, RMSD analysis of the heavy atoms and the backbone P atom of the 6–4PP lesion reveals that, in contrast to the lesion-containing systems, the undamaged DNA at the XPD pore exhibits a one-nucleotide backbone translocation (dT1) in the 3′→5′ direction. The undamaged dT1 nucleotide deviates from its initial position and exhibits increased mobility compared to the 6–4PP, suggesting that XPD tethers more firmly to the backbone of the lesion-containing DNA than to the undamaged DNA.

Furthermore, the authors carried out eight extra independent 2 μs simulations for lesion-free XPD combined with the previous simulation of the XPD with undamaged DNA (a total of 9 simulations). The observation around the extended single-stranded DNA protruding outside the pore elucidated the role of residues situated in the pore’s vicinity acting like lesion sensors. Three simulations showed the undamaged dT1 flipping into the pore via base or backbone translocation. However, six simulations showed dT1 blocked outside the pore, impeded from translocating via entrapment of its base. Hence 6 out of 9 simulations consistently observed the T base being trapped by two base sensor regions in the vicinity of the entry pore. In this sense, His135, Ala139, Ser140, Tyr141, Tyr158, Phe161, Tyr192, Phe193, and His384 from the FeS pocket, and the residues Pro215 to Val221 constituting the α-helix from the ATPase lobe 1, form the base sensor regions. Moreover, atomistic simulation revealed the similarity in the mechanisms by which the ATPase lobe 1 helix captures the analogous unmodified dT1 and the 6–4PP lesion in the XPD with mutated DNA outside the pore. ATPase lobe 1 helix captures the nucleotide atoms O2 and N3, but while the native base keeps handicapped from translocation only temporarily, the bulky 6–4PP is completely obstructed from translocating. Overall, these results indicate that the legion sensor regions act in the initial recognition of modified bases. The least observed phenomenon (three out of 9 simulations), where the base or backbone translocation occurs, suggests the sensors generate a kinetic barrier that should be bypassed.

Although a potential critique would be the absence of any other component of the TFIIH complex since they are required to position XPD in the DNA molecule, the authors focused on understanding how the structure of XPD behaves in the presence of a 6–4PP and how it scans the lesion. Thus, the context of TFIIH might aid in understanding how XPD approaches and is positioned onto the DNA molecule but would not impact the elucidation of how the lesion interacts with a particular region of XPD.

Moreover, this work used the XPD from PDB 6RO4, lacking a small region from the Arch domain (res 273–325), which the authors did not model, and performed a capping of residues 272 and 326 with a C-terminal N-methyl amide and an N-terminal acetyl group, respectively, to avoid steric hindrance with the ssDNA molecule. From a DNA repair perspective, that could be detrimental. Although XP proteins display high structural and folding conservation, the Arch domain was described to be 60% larger in eukaryotes than in other species, suggesting a structural role in interacting with different subunits of the TFIIH complex [111]. In fact, the MAT1 and XPG PPI regions encompass part of the Arch domain [39,45,112]. However, the lack of this region would not affect the area containing the pocket studied by the authors, nor would it cloud any knowledge of the discussed phenomena.

A reasonable debate regarding this work is that although the authors employed RMSD, van der Waals energies, and methyl-π interactions, protein-ligand inferences of stability can only be drawn by employing

multiple additional measures, especially the free energy of interaction calculations, which were not used in their work. The free-energy state function is pertinent for conclusions regarding stability due to enthalpy-entropy compensations once van der Waals energies do not consider the entropy effect in the system.

Using this work as an example, a consideration that researchers should be aware of is dealing with structural metallic ions because it should be done on a case-by-case basis. Transition metals have a strong electronic contribution to their properties since they are highly polarizable and prone to charge transfer, becoming structurally unique depending on their interaction. However, MD deal with atoms based on classical mechanics, which is unsuitable for dealing with variations in point charges during simulation [101]. For this reason, the cluster FeS present in the XPD structure needs to be parameterized from ab initio quantum calculations. In this work, the authors wisely retrieved the parameters for the FeS cluster from another work that had already parameterized it before [113], which is a standard procedure in the field. This work also reinforced the use of replicates in MD. The observation of the base blockage occurred in six out of nine simulations and only after 1.5 μs of simulation, illustrating that few replicas might induce conclusions based on lower probability events, and short simulation time will prevent the observation of relevant phenomena.

Lastly, Panchal et al. [60] used a series of databases and software to obtain, filter, and predict potential pathogenic variants associated with *XPD* and later applied MD for 25 ns in the most preeminent result. The final mutation was a Lys48Arg, previously identified to affect the ATP binding site of XPD [114]. Overall, this workflow shows how basic MD analyses can be effective tools for assessing the impact of mutations on protein structures. When the results from RMSD, RMS Fluctuation (RMSF), R_g, SASA, hydrogen bonds, Dictionary of Secondary Structure of Proteins (DSSP), and PCA agree, they explain the process under investigation. The authors demonstrated that the selected mutant displayed an increase in coils and loss of α-helices that, together with higher RMSD, RMSF, and R_g when compared to the WT, suggested a loss of proper conformation for XPD, especially an emerging flexible region in the central portion (between residues 300–400). Likewise, PCA analyses indicated that the conformational distribution of Lys48Arg was utterly distinguishable from the WT.

This work starts with prospecting variants from databases and goes a step further by investigating potential structural effects using MD - a practice that is not the standard in the field. Although the authors applied sufficient tools to achieve their goals, the limitations still lie in the short simulation time and lack of sampling. Additionally, the energy term considered for the stability assessment is based on software that does not regard the protein's dynamics, which denotes a high inherent error.

The flexible region noted by the authors comprised part of the Arch domain and the XPG PPI region, suggesting that this mutation could jeopardize the XPD-XPG interaction. However, the location of Lys48Arg is more likely to affect the XPG PPI at the beginning of the N-terminus [39]. Because XPD is a large protein displaying a considerable packed structure, an extended simulation time is required to infer instability or any potential loss of proper conformation. Another limitation is that XPD possesses a 4FeS cluster, which is absent in the author's model. From a biological perspective, the negative effects observed by the authors may be an artifact of this technical issue. As a matter of fact, this mutation is near the 4FeS cluster in the XPD structure [39].

3.5. Investigating the XPF-ERCC1 interaction

XPF is a 5′ endonuclease, comprised of 916 amino acids, partially responsible for the removal of the DNA lesion (Fig. 3**F**). The human XPF has at least eight distinct known domains and PPI regions [39], where region 670–905 is associated with PPI with its primary molecular partner, ERCC1, a 297 amino acids protein [40]. Cells lacking ERCC1 display a loss of XPF activity, and recently, Jones et al. [44] showed that ERCC1 plays a role in inhibiting XPF's catalytic function in the absence of DNA. ERCC1, on the other hand, is associated with COFS, CS, and mild XP/CS [115].

The work of Doss and Nagasundaram sought to investigate how mutations in ERCC1 affect the establishment of the ERCC1-XPF complex [61]. The protocol was based on predicting deleterious variants in ERCC1 and five artificially made point mutations from the native ERCC1 structure. Since ERCC1 and XPF proteins were retrieved from different PDB files, a docking step was performed using a docking protocol, generating WT and mutant ERCC1-XPF complexes (Table1). Thus, by docking a small 63 residues portion of XPF in a 146 amino acids structure of ERCC1 [116], the authors performed calculations in multiple software to obtain a profile of potentially deleterious mutations in ERCC1 and chose Val116Met, Arg156Gln, and Ala199Thr for MD simulations during 20 ns. The three mutants displayed minor deviations in RMSD and RMSF from the native complexes. Additionally, the authors showed that these mutations decreased the number of hydrogen bonds compared to the WT complex and increased the SASA in the Arg156Gln, and Ala199Thr systems. The authors concluded that the selected mutations generate an unstable XPF-ERCC1 complex that might be associated with carcinogenesis.

The first limitation is the simulation time. Although the system comprises only 209 amino acids, a time frame of 20 ns is still reasonably small. Adding to the use of a united-atom FF, numerous molecular phenomena could have been clouded. Secondly, the lack of replicates for the systems makes it impossible to know if the molecular behaviors observed are not only low-probability phenomena, including the identified XPF-ERCC1 interface residues. Furthermore, docking scoring functions and standard MD are inappropriate for suggesting the affinity between binding proteins. Molecular docking is unsuitable for directly evaluating protein binding affinity since entropy changes are neglected for the solvent component, and the H-bond description is faulty [117]. A collection of enhanced sampling methods has been developed over the years to overcome the limitations of binding energy calculations [118].

From a DNA repair perspective, the limitation lies in some conclusions. Previous works from 2005 mapped numerous residues associated with the XPF-ERCC1 interaction [119,120], setting the XPF-ERCC1 PPI region to be scattered along residues 670–905 in XPF. Therefore, it is unlikely that the analyzed XPF region (~837–898) would be solely responsible for stabilizing, sustaining, and coordinating the XPF-ERCC1 heterodimer. Thus, it is impossible to conclude that the mutations generate unstable complexes by considering only a small portion of XPF.

4. MD and NER: additional considerations

MD has been consistently employed to unravel mechanisms and interactions behind the NER pathway. We reviewed several articles intersecting MD and NER, which elucidated molecular mechanisms other methodologies could not readily assess, corroborating previous experiments' observations and offering new insights. For those paving the way for incorporating MD simulations into their NER/DNA repair research, this technique is a valuable tool to comprehend and forecast molecular systems' characteristics, structure, properties, and function. These articles are recommended as a starting point for understanding and employing MD simulation [17,121].

As debated, MD results are sensitive to: (i) the input files (i.e., quality of the initial structure and FF parameters); (ii) the appropriate MD method to answer the phenomenon of interest; and (iii) the understanding of the system's biochemical nuances. Based on the reviewed articles, some practices should be considered by future MD studies focused on NER. One of them is protein modeling followed by model quality assessment, ligand charge correction, and structural metal ion correct parameterization, which some articles did not perform. This is a significant issue broadly discussed elsewhere [122–124]. Several reviewed protein structures that presented missing segments were modeled but not validated. Employing incomplete structures limits the

range of possible analyses as well as the level of confidence in the results. For instance, studies proposing to describe PPIs and binding mechanisms may depend on the missing portion of the protein structure for the interaction, or the missing part might be relevant for the enclosing process. A missing protein segment might change the internal configurational entropy, primarily if the missing piece is located in a central portion. Next, validation metrics are pivotal in estimating the reliability of the model. Nowadays, various validation software are available and easily accessible [125–129].

Moreover, the conformation changes that proteins undergo throughout the PPI process and the accompanying free energy should be considered. Conformation changes indicate gain or loss of secondary structure, approximation or distancing between interacting chemical groups, increase or decrease of structural flexibility, increase or decrease of solvent contact, opening or closing of cryptic pockets, and the enclosing or disengagement around the binding element. On the other hand, free energy changes express the affinity between proteins and ligands. The relationship between free energy and affinity lies in the connection between free energy and the equilibrium constant (K_{eq}), mathematically represented by $\Delta G^{\circ} = -RT\ln K_{eq}$. For PPIs, the essential equilibrium is the association/dissociation process, where the dissociation constant (K_D) describes the affinity between proteins. Therefore, studies regarding affinity in PPIs, protein-DNA, or between inhibitory compounds and proteins from the NER pathway should carry out binding free energy analysis.

The solvent effect around the hydration/dehydration of the proteins' surface is another relevant property to understanding the association/dissociation of protein-ligands or PPI, protein folding, and the overall energy of each process [130]. Water molecules are structurally and energetically different when solvating groups with distinct chemical properties or when comparing water molecules in bulk with those near interfaces or yet constrained in coordination with protein's residues [131]. Water molecules near interfaces differ chemically from bulk environments due to chemical asymmetries, most notably the termination of bulk water hydrogen bonding networks [132]. Water closely bound to proteins behaves as if it were part of the structure, exhibiting a decrease in the degrees of freedom compared to bulk waters. Another type of water molecule is one that briefly contacts the protein surface to stabilize it and control its interaction with other ligands. These water molecules, commonly referred to as hydration shells, are a network of shifting water molecules that are dynamically modulated by the physical-chemical properties of the protein's surface. The hydration shell bridges the protein interfaces during PPI, compensating for eventual contact fluctuation between interacting residues [133]. In this sense, the contribution of water molecules to biological processes can be assessed through MD simulations. Even though this analysis has not yet been considered for the NER proteins, this different angle of investigation can provide new insights for the still open questions regarding the NER pathway. For more information about the effect of hydration shells on biomolecules and their impact, please refer to this review [134].

The most questionable issue we notice in the reviewed paper is the conclusions based on basic MD analysis, such as RMSD, RMSF, and R_g. As previously discussed, RMSD denotes the difference in position between equivalent atoms in two proteins; the lower the RMSD value, the more similar the two structures are. In a trajectory, RMSD is calculated as an average of the distance variation of all atoms through the simulation compared to the initial set of coordinates. Since the results revolve around an average, increasing a loop deviation moiety will impact the overall RMSD result, even though more rigid structures, such as α-helix and β-sheet, keep considerably constant through time. Therefore, RMSD is an adequate metric to assess if a structure is approaching an equilibrium during a trajectory. Oppositely, trying to infer structural stability solely based on the RMSD is inadequate because increasing RMSD may be associated with a large movement of flexible regions, while functional domains remain shifting within their natural range of native conformations. Thus, increases in RMSD do not always mean loss of secondary structure, a typical interpretation that leads many to assume a decline in structural stability. Future studies should carefully consider the essence of RMSD to discuss the outcome of their simulations.

Fig. 4 summarizes some important bottlenecks and analyses that should be considered in the context of this review.

5. Conclusion: what lies ahead?

From the discovery of numerous PPI sites, critical amino acids, and post-translational modification (PTM) [39] to the newly resolved protein models [44,46,47], the understanding of NER at the protein level was never this detailed. From a structural perspective, the 3D models of human XPC and RAD23B are still to be elucidated. Likewise, we still lack the complete structure of human XPG, XPA, and RAD70. Nevertheless, the core NER machinery's current structural repertoire allows the application of sophisticated computational analysis to uncover more about the dynamics of these proteins. The most prosperous niche for investigation is understanding how the multitude of missense naturally occurring mutations that cause XP, TTD, CS, and COFS affect the molecular behavior of these proteins. We previously listed all published naturally-occurring disease-causing mutations associated with the XP proteins [38,39]. In total, 104 mutations lead to single amino acid changes, where three were discovered in XPA, five in XPB, 16 were listed for XPC, six in DDB2, 27 in XPF, and 18 in XPG. However, despite this being a fertile ground for new works, creating these systems is challenging. The first significant bottleneck is the lack of a 3D model for the human XPC and the complete models of XPA and XPG. Even though the catalytic portion of XPG was obtained, around 80% of the protein is still missing. Likewise, only around 50% of XPA is available. Studying mutations only in these portions is not optimal because conformation is influenced by the entire structural context, creating biased results if analyzed otherwise. Conversely, we have full models of DDB2 in the context of the whole UV-DDB complex and almost complete models for XPD, XPB, and XPF, allowing in-depth investigation of these proteins. Simulation of some truncated mutants to comprehend how complex formation is analyzed is also a potential base for discussion.

Yet, not all open questions are meant to be answered by MD. As mentioned before, MD is unable to investigate catalytic processes as it does not account for quantum effects in a system. Therefore, this approach cannot observe the nuclease activity of XPF and XPG. Moreover, we argue that a complex phenomenon to observe in MD is how a mutation abrogates PPI. Numerous mutations were described to abolish complex formation [38,39], but the probability of observing such disruption of PPI when using 3D models is almost negligible. However, it is possible to identify aberrant molecular behavior, which can infer protein stability. Therefore, studies must be careful when suggesting abrogation of PPI, especially considering which metrics were accounted to infer such phenomena.

Another open field of research using MD to study DNA repair is the effect of PTM. No MD works so far have examined the impact of different PTM on NER proteins. Most XP-protein PTMs are related to translocation, recruitment, and degradation, which are not meant to be observed by MD as they do not necessarily influence protein structure. It will always generate reasonable doubts if they are caused by meaningful structural changes or impairments in signaling pathways or other cellular functions. However, some PTM were studied to impact some level of interaction, such as the phosphorylation of RPA32 to single-stranded DNA [135] and the acetylation of RPA70 and its possible PPI with XPA [136]. Finally, we highlight that studying protein-DNA interactions of all XP proteins, either with regular or mutated DNA, can provide a detailed knowledge of NER, which is challenging and costly to observe outside MD.

Based on the revised works, future studies using MD to uncover the dynamics of DNA-repair-associated proteins should be careful with: (i) the applied docking protocols and inferences; (ii) presenting the results of different validation metrics when generating models generated by

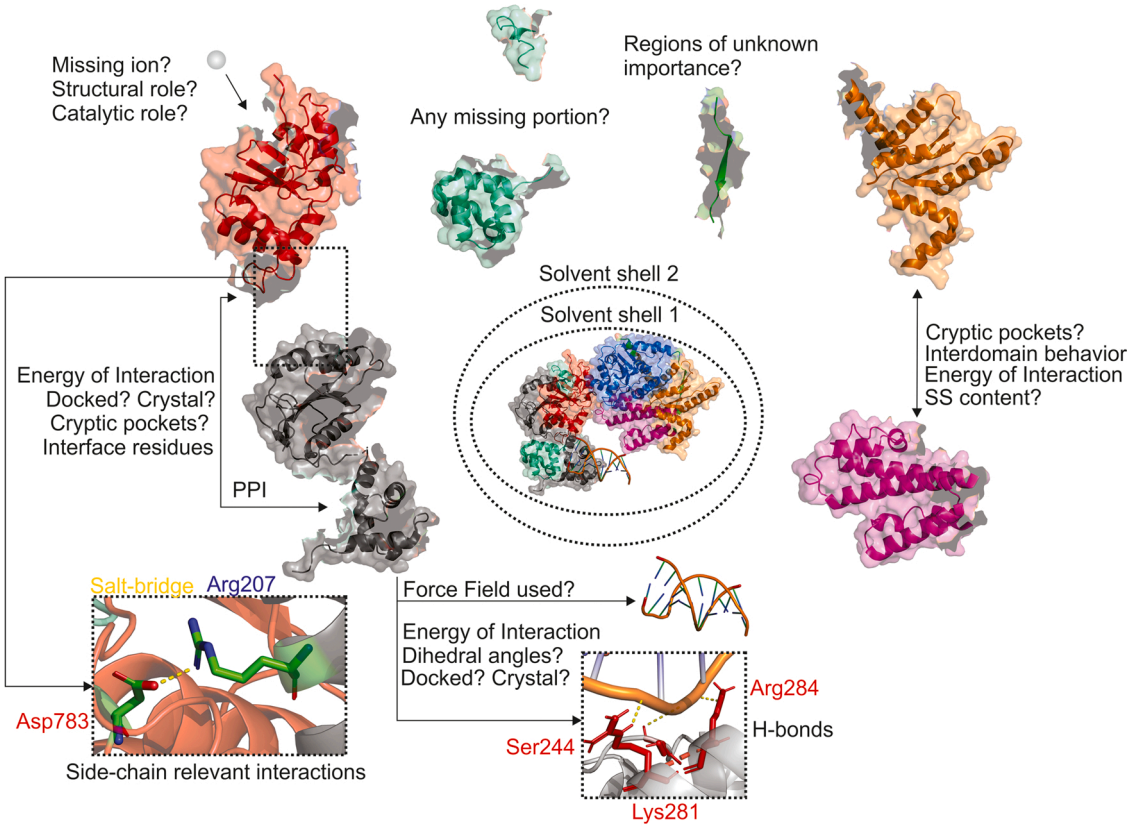


Fig. 4. Summary of the important bottlenecks, structural information, and analyses discussed throughout the review. The XPF (PDB: 6SXB) was used as the example [44]. SS = secondary structure.

software; (iii) the presence and parameterization of structural metal ions; (iv) simulation time and the use of replicates. Although this particular point heavily relies on available computational infrastructure, based on the reviewed articles, we argue that to observe the molecular behavior of large proteins complexes needs an overall extended simulation time, together with at least three replicates; (v) how the final results are interpreted, considering the background thematic. In this case, protein stability should not be inferred based solely on RMSD, RMSF, R_g, and short simulation time. Basic analyses must be accompanied by energy of interaction or other complementary analysis; and (vi) previous structural works detailing complex organization, protein position, and known functional behaviors. In this case, an insufficient literature survey may lead to a flawed system design, incomplete methodological workflow, and incorrect inference from the final results.

Declaration of Competing Interest

We declare no conflicts of interest of any kind.

Data Availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - FAPERGS [19/2551-0001906-8], Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq [314082/2021-2, 440279/2022-4 and 151591/2022-9], and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES/STICAMSUD [88881.522073/2020-01] and DAAD/CAPES PROBRAL [88881.198766/2018-01]. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

[1] K.H. Kraemer, N.J. Patronas, R. Schiffmann, B.P. Brooks, D. Tamura, J. J. DiGiovanna, Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype–phenotype relationship, Neuroscience 145 (2007) 1388–1396, https://doi.org/10.1016/j.neuroscience.2006.12.020.
[2] E. Bergmann, J.-M. Egly, Trichothiodystrophy, a transcription syndrome, TRENDS Genet. 17 (2001) 279–286.
[3] V. Natale, H. Raquer, Xeroderma pigmentosum-Cockayne syndrome complex, Orphanet J. Rare Dis. 12 (2017) 1–11, https://doi.org/10.1186/s13023-017-0616-2.
[4] C.F. Menck, V. Munford, DNA repair diseases: what do they tell us about cancer and aging? Genet. Mol. Biol. 37 (2014) 220–233. ⟨http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3983582&tool=pmcentrez&rendertype=abstract⟩.
[5] A.C. Karikkineth, M. Scheibye-Knudsen, E. Fivenson, D.L. Croteau, V.A. Bohr, Cockayne syndrome: clinical features, model systems and pathways, Ageing Res. Rev. 33 (2017) 3–17, https://doi.org/10.1016/j.arr.2016.08.002.
[6] D. Paul, H. Mu, A. Tavakoli, Q. Dai, X. Chen, S. Chakraborty, C. He, A. Ansari, S. Broyde, J.-H. Min, Tethering-facilitated DNA 'opening' and complementary roles of β-hairpin motifs in the Rad4/XPC DNA damage sensor protein, Nucleic Acids Res. 48 (2020) 12348–12364, https://doi.org/10.1093/nar/gkaa909.
[7] I. Fu, H. Mu, N.E. Geacintov, S. Broyde, Mechanism of lesion verification by the human XPD helicase in nucleotide excision repair, Nucleic Acids Res. 50 (2022) 6837–6853, https://doi.org/10.1093/nar/gkac496.
[8] C. Bowen, J.H. Ju, J.H. Lee, T.T. Paull, E.P. Gelmann, Functional activation of ATM by the prostate cancer suppressor NKX3.1, Cell Rep. 4 (2013) 516–529, https://doi.org/10.1016/j.celrep.2013.06.039.
[9] T.R. Alderson, L.E. Kay, Unveiling invisible protein states with NMR spectroscopy, Curr. Opin. Struct. Biol. 60 (2020) 39–49, https://doi.org/10.1016/j.sbi.2019.10.008.
[10] W.-H. Chang, S.-H. Huang, H.-H. Lin, S.-C. Chung, I.-P. Tu, Cryo-EM analyses permit visualization of structural polymorphism of biological macromolecules, Front. Bioinforma. 1 (2021), 788308, https://doi.org/10.3389/fbinf.2021.788308.

[11] S. Zhang, S. Zou, D. Yin, L. Zhao, D. Finley, Z. Wu, Y. Mao, USP14-regulated allostery of the human proteasome by time-resolved cryo-EM, Nature 605 (2022) 567–574, https://doi.org/10.1038/s41586-022-04671-8.
[12] S. Shibata, H. Matsunami, S.-I. Aizawa, M. Wolf, Torque transmission mechanism of the curved bacterial flagellar hook revealed by cryo-EM, Nat. Struct. Mol. Biol. 26 (2019) 941–945, https://doi.org/10.1038/s41594-019-0301-3.
[13] K. Zhang, D. Julius, Y. Cheng, Structural snapshots of TRPV1 reveal mechanism of polymodal functionality, Cell 184 (2021) 5138–5150.e12, https://doi.org/10.1016/j.cell.2021.08.012.
[14] M.-E. Mäeots, R.I. Enchev, Structural dynamics: review of time-resolved cryo-EM, Acta Crystallogr. Sect. D. Struct. Biol. 78 (2022) 927–935, https://doi.org/10.1107/S2059798322006155.
[15] S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for All, Neuron 99 (2018) 1129–1143, https://doi.org/10.1016/j.neuron.2018.08.011.
[16] E. Brini, C. Simmerling, K. Dill, Protein storytelling through physics, Science (80-.) 370 (2020) 1–23, https://doi.org/10.1126/science.aaz3041.
[17] S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for all, Neuron 99 (2018) 1129–1143, https://doi.org/10.1016/j.neuron.2018.08.011.
[18] P.S. Nerenberg, T. Head-Gordon, New developments in force fields for biomolecular simulations, Curr. Opin. Struct. Biol. 49 (2018) 129–138, https://doi.org/10.1016/j.sbi.2018.02.002.
[19] J. Yoo, D. Winogradoff, A. Aksimentiev, Molecular dynamics simulations of DNA–DNA and DNA–protein interactions, Curr. Opin. Struct. Biol. 64 (2020) 88–96, https://doi.org/10.1016/j.sbi.2020.06.007.
[20] X. He, B. Walker, V.H. Man, P. Ren, J. Wang, Recent progress in general force fields of small molecules, Curr. Opin. Struct. Biol. 72 (2022) 187–193, https://doi.org/10.1016/j.sbi.2021.11.011.
[21] A. Hospital, J.R. Goñi, M. Orozco, J. Gelpi, Molecular dynamics simulations: advances and applications, Adv. Appl. Bioinforma. Chem. 8 (2015) 37, https://doi.org/10.2147/AABC.S70333.
[22] M.T. Muhammed, E. Aki-Yalcin, Homology modeling in drug discovery: overview, current applications, and future perspectives, Chem. Biol. Drug Des. 93 (2019) 12–20, https://doi.org/10.1111/cbdd.13388.
[23] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
[24] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, J. Wang, Q. Cong, L.N. Kinch, R.D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J.H. Pereira, A.V. Rodrigues, A.A. van Dijk, A. C. Ebrecht, D.J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C.K. Yip, J.E. Burke, K.C. Garcia, N.V. Grishin, P. D. Adams, R.J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, Science (80-.) 373 (2021) 871–876, https://doi.org/10.1126/science.abj8754.
[25] V. Bongirwar, A.S. Mokhade, Different methods, techniques and their limitations in protein structure prediction: a review, Prog. Biophys. Mol. Biol. 173 (2022) 72–82, https://doi.org/10.1016/j.pbiomolbio.2022.05.002.
[26] A. Perrakis, T.K. Sixma, AI revolutions in biology, EMBO Rep. 22 (2021), e54046, https://doi.org/10.15252/embr.202154046.
[27] Z. Peng, W. Wang, R. Han, F. Zhang, J. Yang, Protein structure prediction in the deep learning era, Curr. Opin. Struct. Biol. 77 (2022), 102495, https://doi.org/10.1016/j.sbi.2022.102495.
[28] D.L. Mobley, M.K. Gilson, Predicting binding free energies: frontiers and benchmarks, Annu. Rev. Biophys. 46 (2017) 531–558, https://doi.org/10.1146/annurev-biophys-070816-033654.
[29] A. de Ruiter, C. Oostenbrink, Advances in the calculation of binding free energies, Curr. Opin. Struct. Biol. 61 (2020) 207–212, https://doi.org/10.1016/j.sbi.2020.01.016.
[30] S. Wan, A.P. Bhati, S.J. Zasada, P.V. Coveney, Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction, Interface Focus 10 (2020), https://doi.org/10.1098/rsfs.2020.0007, 20200007.
[31] O.D. Scharer, Nucleotide excision repair in eukaryotes, Cold Spring Harb. Perspect. Biol. 5 (2013), https://doi.org/10.1101/cshperspect.a012609 a012609–a012609.
[32] G. Spivak, Nucleotide excision repair in humans, DNA Repair 36 (2015) 13–18, https://doi.org/10.1016/j.dnarep.2015.09.003.
[33] W. Wang, J. Xu, J. Chong, D. Wang, Structural basis of DNA lesion recognition for eukaryotic transcription-coupled nucleotide excision repair, DNA Repair 71 (2018) 43–55, https://doi.org/10.1016/j.dnarep.2018.08.006.
[34] W. Vermeulen, M. Fousteri, Mammalian transcription-coupled excision repair, Cold Spring Harb. Perspect. Biol. 5 (2013) 1–16, https://doi.org/10.1101/cshperspect.a012625.
[35] R. Kusumoto, C. Masutani, K. Sugasawa, S. Iwai, M. Araki, A. Uchida, T. Mizukoshi, F. Hanaoka, Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro, Mutat. Res. Repair 485 (2001) 219–227, https://doi.org/10.1016/S0921-8777(00)00082-3.
[36] M.E. Fitch, S. Nakajima, A. Yasui, J.M. Ford, In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product, J. Biol. Chem. 278 (2003) 46906–46910.
[37] H. Mu, N.E. Geacintov, S. Broyde, J.-E. Yeo, O.D. Schärer, Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair, DNA Repair 71 (2018) 33–42, https://doi.org/10.1016/j.dnarep.2018.08.005.
[38] B.C. Feltes, D. Bonatto, Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications, Mutat. Res. - Rev. Mutat. Res. 763 (2015) 306–320, https://doi.org/10.1016/j.mrrev.2014.12.002.
[39] B.C. Feltes, Revisiting the structural features of the xeroderma pigmentosum proteins: Focus on mutations and knowledge gaps, Mutat. Res. Mutat. Res. 789 (2022), 108416, https://doi.org/10.1016/j.mrrev.2022.108416.
[40] B.C. Feltes, Every protagonist has a sidekick: structural aspects of human xeroderma pigmentosum-binding proteins in nucleotide excision repair, Protein Sci. 30 (2021) 2187–2205, https://doi.org/10.1002/pro.4173.
[41] J. a Marteijn, H. Lans, W. Vermeulen, J.H.J. Hoeijmakers, Understanding nucleotide excision repair and its roles in cancer and ageing, Nat. Rev. Mol. Cell Biol. 15 (2014) 465–481, https://doi.org/10.1038/nrm3822.
[42] A. Scrima, R. Konířková, B.K. Czyzewski, Y. Kawasaki, P.D. Jeffrey, R. Groisman, Y. Nakatani, S. Iwai, N.P. Pavletich, N.H. Thomä, Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex, Cell 135 (2008) 1213–1223, https://doi.org/10.1016/j.cell.2008.10.045.
[43] G.W. Buchko, G.W. Daughdrill, R. de Lorimier, K. Rao B, N.G. Isern, J. M. Lingbeck, J.S. Taylor, M.S. Wold, M. Gochin, L.D. Spicer, D.F. Lowry, M. a Kennedy, Interactions of human nucleotide excision repair protein XPA with DNA and RPA70 Delta C327: chemical shift mapping and ^{15}N NMR relaxation studies, Biochemistry 38 (1999) 15116–15128. ⟨http://www.ncbi.nlm.nih.gov/pubmed/10563794⟩.
[44] M. Jones, F. Beuron, A. Borg, A. Nans, C.P. Earl, D.C. Briggs, A.P. Snijders, M. Bowles, E.P. Morris, M. Linch, et al., Cryo-EM structures of the XPF-ERCC1 endonuclease reveal how DNA-junction engagement disrupts an auto-inhibited conformation, Nat. Commun. 11 (2020) 1–14.
[45] B.J. Greber, T.H.D. Nguyen, J. Fang, P.V. Afonine, P.D. Adams, E. Nogales, The cryo-electron microscopy structure of human transcription factor IIH, Nature 549 (2017) 414–417.
[46] G. Kokic, A. Chernev, D. Tegunov, C. Dienemann, H. Urlaub, P. Cramer, Structural basis of TFIIH activation for nucleotide excision repair, Nat. Commun. 10 (2019) 1–9.
[47] R. González-Corrochano, F.M. Ruiz, N.M.I. Taylor, S. Huecas, S. Drakulic, M. Spínola-Amilibia, C. Fernández-Tornero, The crystal structure of human XPG, the xeroderma pigmentosum group G endonuclease, provides insight into nucleotide excision DNA repair, Nucleic Acids Res. 48 (2020) 9943–9958, https://doi.org/10.1093/nar/gkaa688.
[48] D. Paul, H. Mu, H. Zhao, O. Ouerfelli, P.D. Jeffrey, S. Broyde, J.H. Min, Structure and mechanism of pyrimidine–pyrimidone (6-4) photoproduct recognition by the Rad4/XPC nucleotide excision repair complex, Nucleic Acids Res. 47 (2019) 6015–6028, https://doi.org/10.1093/nar/gkz359.
[49] K. Pitta, M. Krishnan, Molecular mechanism, dynamics, and energetics of protein-mediated dinucleotide flipping in a mismatched DNA: a computational study of the RAD4-DNA complex, J. Chem. Inf. Model 58 (2018) 647–660, https://doi.org/10.1021/acs.jcim.7b00636.
[50] A. Panigrahi, H. Vemuri, M. Aggarwal, K. Pitta, M. Krishnan, Sequence specificity, energetics and mechanism of mismatch recognition by DNA damage sensing protein Rad4/XPC, Nucleic Acids Res. 48 (2020) 2246–2257, https://doi.org/10.1093/nar/gkaa078.
[51] H. Mu, Y. Zhang, N.E. Geacintov, S. Broyde, Lesion sensing during initial binding by yeast XPC/Rad4: toward predicting resistance to nucleotide excision repair, Chem. Res. Toxicol. 31 (2018) 1260–1268, https://doi.org/10.1021/acs.chemrestox.8b00231.
[52] D. Paul, H. Mu, A. Tavakoli, Q. Dai, S. Chakraborty, C. He, A. Ansari, S. Broyde, J.-H. Min, Impact of DNA sequences on DNA 'opening' by the Rad4/XPC nucleotide excision repair complex, DNA Repair 107 (2021), 103194, https://doi.org/10.1016/j.dnarep.2021.103194.
[53] B.C. Feltes, C. Pedebos, D. Bonatto, H. Verli, Dynamics of DDB2-DDB1 complex under different naturally-occurring mutants in Xeroderma Pigmentosum disease, Biochim. Biophys. Acta - Gen. Subj. 1862 (2018) 2579–2589, https://doi.org/10.1016/j.bbagen.2018.08.007.
[54] S. Pradhan, P. Das, V.S.K. Mattaparthi, Characterizing the binding interactions between DNA-binding proteins, XPA and XPE: a molecular dynamics approach, ACS Omega 3 (2018) 15442–15454, https://doi.org/10.1021/acsomega.8b01793.
[55] K.H. Barakat, J.T. Huzil, T. Luchko, L. Jordheim, C. Dumontet, J. Tuszynski, Characterization of an inhibitory dynamic pharmacophore for the ERCC1–XPA interaction using a combined molecular dynamics and virtual screening approach, J. Mol. Graph. Model 28 (2009) 113–130.
[56] E. Fadda, The role of conformational selection in the molecular recognition of the wild type and mutants XPA67-80 peptides by ERCC1, Proteins Struct. Funct. Bioinforma. 83 (2015) 1341–1351, https://doi.org/10.1002/prot.24825.
[57] J. Hu, Z. Hu, Y. Zhang, X. Gou, Y. Mu, L. Wang, X.Q. Xie, Metal binding mediated conformational change of XPA protein:a potential cytotoxic mechanism of nickel in the nucleotide excision repair, J. Mol. Model 22 (2016) 1–33, https://doi.org/10.1007/s00894-016-3017-x.
[58] S. Pradhan, V.S.K. Mattaparthi, Structural dynamics and interactions of Xeroderma pigmentosum complementation group A (XPA 98–210) with damaged DNA, J. Biomol. Struct. Dyn. 36 (2018) 3341–3353, https://doi.org/10.1080/07391102.2017.1388285.
[59] S. Pradhan, H. Sarma, V.S.K. Mattaparthi, Investigation of the probable homo-dimer model of the Xeroderma pigmentosum complementation group A (XPA) protein to represent the DNA-binding core, J. Biomol. Struct. Dyn. 37 (2019) 3322–3336, https://doi.org/10.1080/07391102.2018.1517051.

[60] N.K. Panchal, A. Bhale, V.K. Verma, S.S. Beevi, Computational and molecular dynamics simulation approach to analyze the impactof XPD gene mutation on protein stability and function, Mol. Simul. 46 (2020) 1200–1219, https://doi.org/10.1080/08927022.2020.1810852.
[61] C. George Priya Doss, N. Nagasundaram, Molecular docking and molecular dynamics study on the effect of ERCC1 deleterious polymorphisms in ERCC1-XPF heterodimer, Appl. Biochem. Biotechnol. 172 (2014) 1265–1281, https://doi.org/10.1007/s12010-013-0592-5.
[62] J. Kästner, Umbrella sampling, Wiley Interdiscip. Rev. Comput. Mol. Sci. 1 (2011) 932–942, https://doi.org/10.1002/wcms.66.
[63] Y. Cai, H. Zheng, S. Ding, K. Kropachev, A.G. Schwaid, Y. Tang, H. Mu, S. Wang, N.E. Geacintov, Y. Zhang, S. Broyde, Free energy profiles of base flipping in intercalative polycyclic aromatic hydrocarbon-damaged DNA Duplexes: energetic and structural relationships to nucleotide excision repair susceptibility, Chem. Res. Toxicol. 26 (2013) 1115–1125, https://doi.org/10.1021/tx400156a.
[64] L. Tang, Y. Cheng, Y. Zhang, D. Ran, C. Yang, M. Gao, P. Wang, X. Zheng, S. Yang, Novel compound heterozygous variants in the XPC gene identified in a Chinese xeroderma pigmentosum group C patient with ovarian teratoma, J. Dermatol. 45 (2018) e300–e301, https://doi.org/10.1111/1346-8138.14340.
[65] Y. Fujiwara, C. Masutani, T. Mizukoshi, J. Kondo, F. Hanaoka, S. Iwai, Characterization of DNA recognition by the human UV-damaged DNA-binding protein, J. Biol. Chem. 274 (1999) 20027–20033, https://doi.org/10.1074/jbc.274.28.20027.
[66] K. Sugasawa, T. Okamoto, Y. Shimizu, C. Masutani, S. Iwai, F. Hanaoka, A multistep damage recognition mechanism for global genomic nucleotide excision repair, Genes\Dev. 15 (2001) 507–521.
[67] M. Wakasugi, A. Kawashima, H. Morioka, S. Linn, A. Sancar, T. Mori, O. Nikaido, T. Matsunaga, DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair, J. Biol. Chem. 277 (2002) 1637–1640.
[68] M.-R. Puumalainen, P. Rüthemann, J.-H. Min, H. Naegeli, Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation, Cell. Mol. Life Sci. 73 (2016) 547–566, https://doi.org/10.1007/s00018-015-2075-z.
[69] N. Kumar, A.F. Theil, V. Roginskaya, Y. Ali, M. Calderon, S.C. Watkins, R. P. Barnes, P.L. Opresko, A. Pines, H. Lans, W. Vermeulen, B. Van Houten, Global and transcription-coupled repair of 8-oxoG is initiated by nucleotide excision repair proteins, Nat. Commun. 13 (2022) 974, https://doi.org/10.1038/s41467-022-28642-9.
[70] C. Xu, J. Min, Structure and function of WD40 domain proteins, Protein Cell 2 (2011) 202–214, https://doi.org/10.1007/s13238-011-1018-1.
[71] C.K.-M. Chen, N.-L. Chan, A.H.-J. Wang, The many blades of the β-propeller proteins: conserved but versatile, Trends Biochem. Sci. 36 (2011) 553–561, https://doi.org/10.1016/j.tibs.2011.07.004.
[72] B. Knapp, L. Ospina, C.M. Deane, Avoiding false positive conclusions in molecular simulation: the importance of replicas, J. Chem. Theory Comput. 14 (2018) 6127–6138, https://doi.org/10.1021/acs.jctc.8b00391.
[73] X. Chen, Y. Velmurugu, G. Zheng, B. Park, Y. Shim, Y. Kim, L. Liu, B. Van Houten, C. He, A. Ansari, J.-H. Min, Kinetic gating mechanism of DNA damage recognition by Rad4/XPC, Nat. Commun. 6 (2015) 5849, https://doi.org/10.1038/ncomms6849.
[74] F.-M. Lian, X. Yang, Y.-L. Jiang, F. Yang, C. Li, W. Yang, C. Qian, New structural insights into the recognition of undamaged splayed-arm DNA with a single pair of non-complementary nucleotides by human nucleotide excision repair protein XPA, Int. J. Biol. Macromol. 148 (2020) 466–474, https://doi.org/10.1016/j.ijbiomac.2020.01.169.
[75] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W. F. van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, Eur. Biophys. J. 40 (2011) 843–856, https://doi.org/10.1007/s00249-011-0700-9.
[76] C. Ribeiro-Silva, M. Sabatella, A. Helfricht, J.A. Marteijn, A.F. Theil, W. Vermeulen, H. Lans, Ubiquitin and TFIIH-stimulated DDB2 dissociation drives DNA damage handover in nucleotide excision repair, Nat. Commun. 11 (2020) 1–14, https://doi.org/10.1038/s41467-020-18705-0.
[77] O.V. Tsodikov, D. Ivanov, B. Orelli, L. Staresincic, I. Shoshani, R. Oberman, O. D. Schärer, G. Wagner, T. Ellenberger, Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA, EMBO J. 26 (2007) 4768–4776, https://doi.org/10.1038/sj.emboj.7601894.
[78] L. Li, C.A. Peterson, X. Lu, R.J. Legerski, Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair, Mol. Cell. Biol. 15 (1995) 1993–1998, https://doi.org/10.1128/mcb.15.4.1993.
[79] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802, https://doi.org/10.1002/jcc.20289.
[80] D. Ramírez, Computational methods applied to rational drug design, Open Med. Chem. J. 10 (2016) 7–20, https://doi.org/10.2174/1874104501610010007.
[81] F. Gentile, J.A. Tuszynski, K.H. Barakat, New design of nucleotide excision repair (NER) inhibitors for combination cancer therapy, J. Mol. Graph. Model 65 (2016) 71–82, https://doi.org/10.1016/j.jmgm.2016.02.010.
[82] S. Kar, J. Leszczynski, Open access in silico tools to predict the ADMET profiling of drug candidates, Expert Opin. Drug Discov. 15 (2020) 1473–1487, https://doi.org/10.1080/17460441.2020.1798926.
[83] R. Ge, L. Liu, W. Dai, W. Zhang, Y. Yang, H. Wang, Q. Shi, S. Guo, X. Yi, G. Wang, T. Gao, Q. Luan, C. Li, Xeroderma pigmentosum group A promotes autophagy to facilitate cisplatin resistance in melanoma cells through the activation of PARP1, J. Invest. Dermatol. 136 (2016) 1219–1228, https://doi.org/10.1016/j.jid.2016.01.031.
[84] A.J. Van Vuuren, E. Appeldoorn, H. Odijk, A. Yasui, N.G. Jaspers, D. Bootsma, J. H. Hoeijmakers, Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F, EMBO J. 12 (1993) 3693–3701.
[85] M. Biggerstaff, D.E. Szymkowski, R.D. Wood, Co-correction of the ERCC1, ERCC4 and xeroderma pigmentosum group F DNA repair defects in vitro, EMBO J. 12 (1993) 3685–3692.
[86] J.J. Rahn, G.M. Adair, R.S. Nairn, Multiple roles of ERCC1-XPF in mammalian interstrand crosslink repair, Environ. Mol. Mutagen 51 (2010) 567–581, https://doi.org/10.1002/em.20583.
[87] D. Klein Douwel, W.S. Hoogenboom, R.A. Boonen, P. Knipscheer, Recruitment and positioning determine the specific role of the XPF - ERCC 1 endonuclease in interstrand crosslink repair, EMBO J. 36 (2017) 2034–2046, https://doi.org/10.15252/embj.201695223.
[88] M. Enoiu, J. Jiricny, O.D. Schärer, Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis, Nucleic Acids Res. 40 (2012) 8953–8964, https://doi.org/10.1093/nar/gks670.
[89] A. Mukherjee, K.M. Vasquez, HMGB1 interacts with XPA to facilitate the processing of DNA interstrand crosslinks in human cells, Nucleic Acids Res. 44 (2016) 1151–1160, https://doi.org/10.1093/nar/gkv1183.
[90] R.D. Wood, Mammalian nucleotide excision repair proteins and interstrand crosslink repair, Environ. Mol. Mutagen 51 (2010) 520–526, https://doi.org/10.1002/em.20569.
[91] M. Sabatella, A. Pines, J. Slyskova, W. Vermeulen, H. Lans, ERCC1–XPF targeting to psoralen–DNA crosslinks depends on XPA and FANCD2, Cell. Mol. Life Sci. 77 (2020) 2005–2016, https://doi.org/10.1007/s00018-019-03264-5.
[92] E. Fadda, Conformational determinants for the recruitment of ERCC1 by XPA in the nucleotide excision repair (NER) pathway: Structure and dynamics of the XPA binding motif, Biophys. J. 104 (2013) 2503–2511, https://doi.org/10.1016/j.bpj.2013.04.023.
[93] Z. Sun, Y.N. Yan, M. Yang, J.Z.H. Zhang, Interaction entropy for protein-protein binding, J. Chem. Phys. 146 (2017), 124124, https://doi.org/10.1063/1.4978893.
[94] T. Vreven, H. Hwang, B.G. Pierce, Z. Weng, Prediction of protein-protein binding free energies, Protein Sci. 21 (2012) 396–404, https://doi.org/10.1002/pro.2027.
[95] T. Tripathi, Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism, J. Proteins Proteom. 4 (2013) 85–91.
[96] H. Wang, C. Liu, L. Deng, Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting, Sci. Rep. 8 (2018) 14285, https://doi.org/10.1038/s41598-018-32511-1.
[97] D. Bukhsh, T.T. Singh, Frontiers in Protein Structure, Function, and Dynamics, 1st ed., Springer Singapore, Singapore, 2020. https://doi.org/10.1007/978–981-15–5530-5.
[98] T. Ikegami, I. Kuraoka, M. Saijo, N. Kodo, Y. Kyogoku, K. Morikawa, K. Tanaka, M. Shirakawa, Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA, Nat. Struct. Mol. Biol. 5 (1998) 701–706, https://doi.org/10.1038/1400.
[99] A.M. Topolska-Woś, N. Sugitani, J.J. Cordoba, K.V. Le Meur, R.A. Le Meur, H. S. Kim, J.-E. Yeo, D. Rosenberg, M. Hammel, O.D. Schärer, et al., A key interaction with RPA orients XPA in NER complexes, Nucleic Acids Res. 48 (2020) 2173–2188.
[100] E.H. Morita, T. Ohkubo, I. Kuraoka, M. Shirakawa, K. Tanaka, K. Morikawa, Implications of the zinc-finger motif found in the DNA-binding domain of the human XPA protein, Genes Cells 1 (1996) 437–442, https://doi.org/10.1046/j.1365-2443.1996.d01-252.x.
[101] M.R. Nechay, C.E. Valdez, A.N. Alexandrova, Computational treatment of metalloproteins, J. Phys. Chem. B 119 (2015) 5945–5956, https://doi.org/10.1021/acs.jpcb.5b00028.
[102] S.E.J. Bowman, J. Bridwell-Rabb, C.L. Drennan, Metalloprotein crystallography: more than a structure, Acc. Chem. Res. 49 (2016) 695–702, https://doi.org/10.1021/acs.accounts.5b00538.
[103] W. Li, J. Wang, J. Zhang, W. Wang, Molecular simulations of metal-coupled protein folding, Curr. Opin. Struct. Biol. 30 (2015) 25–31, https://doi.org/10.1016/j.sbi.2014.11.006.
[104] Y. ming Lee, C. Lim, Physical basis of structural and catalytic Zn-binding sites in proteins, J. Mol. Biol. 379 (2008) 545–553, https://doi.org/10.1016/j.jmb.2008.04.004.
[105] D. Neuhaus, Zinc finger structure determination by NMR: Why zinc fingers can be a handful, Prog. Nucl. Magn. Reson. Spectrosc. 130–131 (2022) 60–103, https://doi.org/10.1016/j.pnmrs.2022.07.001.
[106] K. Kluska, J. Adamczyk, A. Krężel, Metal binding properties, stability and reactivity of zinc fingers, Coord. Chem. Rev. 367 (2018) 18–64, https://doi.org/10.1016/j.ccr.2018.04.009.
[107] M.B. Peters, Y. Yang, B. Wang, L. Füsti-Molnár, M.N. Weaver, K.M. Merz, Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF), J. Chem. Theory Comput. 6 (2010) 2935–2947, https://doi.org/10.1021/ct1002626.
[108] Z. Yu, P. Li, K.M. Merz, Extended zinc AMBER force field (EZAFF), J. Chem. Theory Comput. 14 (2018) 242–254, https://doi.org/10.1021/acs.jctc.7b00773.
[109] Y. Liu, Y. Liu, Z. Yang, C. Utzat, G. Wang, A.K. Basu, Y. Zou, Cooperative interaction of human XPA stabilizes and enhances specific binding of XPA to DNA

damage, Biochemistry 44 (2005) 7361–7368, https://doi.org/10.1021/bi047598y.
[110] Z. Yang, Y. Liu, L.Y. Mao, J.-T. Zhang, Y. Zou, Dimerization of human XPA and formation of XPA 2 −RPA protein complex, Biochemistry 41 (2002) 13012–13020, https://doi.org/10.1021/bi026064z.
[111] H. Liu, J. Rudolf, K.A. Johnson, S.A. Mcmahon, M. Oke, A. Mcrobbie, S.E. Brown, J.H. Naismith, M.F. White, Structure of the DNA repair helicase XPD, Cell 133 (2008) 801–812, https://doi.org/10.1016/j.cell.2008.04.029.Structure.
[112] N. Iyer, M.S. Reagan, K.-J. Wu, B. Canagarajah, E.C. Friedberg, Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and cockayne syndrome group B (CSB) protein, Biochemistry 35 (1996) 2157–2167, https://doi.org/10.1021/bi9524124.
[113] A.T.P. Carvalho, M. Swart, Electronic structure investigation and parametrization of biologically relevant iron–sulfur clusters, J. Chem. Inf. Model 54 (2014) 613–620, https://doi.org/10.1021/ci400718m.
[114] F. Coin, J.C. Marinoni, C. Rodolfo, S. Fribourg, A.M. Pedrini, J.M. Egly, Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH, Nat. Genet. 20 (1998) 184–188, https://doi.org/10.1038/2491.
[115] D. Ferri, D. Orioli, E. Botta, Heterogeneity and overlaps in nucleotide excision repair disorders, Clin. Genet. 97 (2020) 12–24, https://doi.org/10.1111/cge.13545.
[116] O.V. Tsodikov, J.H. Enzlin, O.D. Schärer, T. Ellenberger, Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF–ERCC1, Proc. Natl. Acad. Sci. USA 102 (2005) 11236–11241.
[117] T. Pantsar, A. Poso, Binding affinity via docking: fact and fiction, Molecules 23 (2018) 1899, https://doi.org/10.3390/molecules23081899.
[118] C.L. Brooks, D.A. Case, S. Plimpton, B. Roux, D. van der Spoel, E. Tajkhorshid, Classical molecular dynamics, J. Chem. Phys. 154 (2021), 100401, https://doi.org/10.1063/5.0045455.
[119] Y.-J. Choi, K.-S. Ryu, Y.-M. Ko, Y.-K. Chae, J.G. Pelton, D.E. Wemmer, B.-S. Choi, Biophysical characterization of the interaction domains and mapping of the contact residues in the XPF-ERCC1 complex, J. Biol. Chem. 280 (2005) 28644–28652, https://doi.org/10.1074/jbc.M501083200.
[120] K. Tripsianes, G. Folkers, E. Ab, D. Das, H. Odijk, N.G.J. Jaspers, J.H. J. Hoeijmakers, R. Kaptein, R. Boelens, The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair, Structure 13 (2005) 1849–1858, https://doi.org/10.1016/j.str.2005.08.014.
[121] J. Lemkul, From proteins to perturbed hamiltonians: a suite of tutorials for the GROMACS-2018 molecular simulation package [Article v1.0], Living J. Comput. Mol. Sci. 1 (2019), https://doi.org/10.33011/livecoms.1.1.5068.
[122] B. Kuhlman, P. Bradley, Advances in protein structure prediction and design, Nat. Rev. Mol. Cell Biol. 20 (2019) 681–697, https://doi.org/10.1038/s41580-019-0163-x.
[123] P. Li, K.M. Merz, Metal ion modeling using classical mechanics, Chem. Rev. 117 (2017) 1564–1686, https://doi.org/10.1021/acs.chemrev.6b00440.
[124] S. Riniker, Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: an overview, J. Chem. Inf. Model 58 (2018) 565–578, https://doi.org/10.1021/acs.jcim.8b00042.
[125] L. Samant, V.C. Sangar, A. Chowdhary, Online servers and offline tools for protein modelling, optimization and validation: a review, Int. J. Pharm. Sci. Rev. Res. 28 (2014) 123–127.
[126] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D. Biol. Crystallogr 66 (2010) 12–21, https://doi.org/10.1107/S0907444909042073.
[127] G. Studer, C. Rempfer, A.M. Waterhouse, R. Gumienny, J. Haas, T. Schwede, QMEANDisCo—distance constraints applied on model quality estimation, Bioinformatics 36 (2020) 1765–1771, https://doi.org/10.1093/bioinformatics/btz828.
[128] F. Baldassarre, D. Menéndez Hurtado, A. Elofsson, H. Azizpour, GraphQA: protein model quality assessment using graph convolutional networks, Bioinformatics 37 (2021) 360–366, https://doi.org/10.1093/bioinformatics/btaa714.
[129] G. Pagès, B. Charmettant, S. Grudinin, Protein model quality assessment using 3D oriented convolutional neural networks, Bioinformatics 35 (2019) 3313–3319, https://doi.org/10.1093/bioinformatics/btz122.
[130] T.V. Chalikian, Does the release of hydration water come with a Gibbs energy contribution? J. Chem. Thermodyn. 158 (2021), 106409 https://doi.org/10.1016/j.jct.2021.106409.
[131] S.N. Timasheff, Protein hydration, thermodynamic binding, and preferential hydration, Biochemistry 41 (2002) 13473–13482, https://doi.org/10.1021/bi020316e.
[132] D. Konstantinovsky, E.A. Perets, T. Santiago, L. Velarde, S. Hammes-Schiffer, E.C. Y. Yan, Detecting the first hydration shell structure around biomolecules at interfaces, ACS Cent. Sci. 8 (2022) 1404–1414, https://doi.org/10.1021/acscentsci.2c00702.
[133] F.C.L. Almeida, K. Sanches, R. Pinheiro-Aguiar, V.S. Almeida, I.P. Caruso, Protein surface interactions—theoretical and experimental studies, Front. Mol. Biosci. 8 (2021), 706002, https://doi.org/10.3389/fmolb.2021.706002.
[134] D. Laage, T. Elsaesser, J.T. Hynes, Water dynamics in the hydration shells of biomolecules, Chem. Rev. 117 (2017) 10694–10725, https://doi.org/10.1021/acs.chemrev.6b00765.
[135] X.V. Gomes, L.A. Henricksen, M.S. Wold, Proteolytic mapping of human replication protein a: evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA, Biochemistry 35 (1996) 5586–5595, https://doi.org/10.1021/bi9526995.
[136] H. He, J. Wang, T. Liu, UV-Induced RPA1 acetylation promotes nucleotide excision repair, Cell Rep. 20 (2017) 2010–2025, https://doi.org/10.1016/j.celrep.2017.08.016.

1.8 Chapter Summary

This chapter discusses the biological aspects of DNA damage repair and the molecular machinery responsible for overseeing genetic material and restoring damaged areas. The repair process is primarily carried out by the NER pathway. A key component of NER is the XPF-ERCC1 heterodimer, an endonuclease that cleaves damaged DNA at the 5′ end. Mutations in the XPF structure, such as R153P, C356R, R589W, and R689S, have been linked to serious disorders like XP, CS, FA, and XFE progeroid syndrome. However, the mechanisms by which these mutations affect NER remain unclear. In this context, MD simulations comprehend a valuable technique for exploring time-dependent mechanisms, allowing us to analyze the molecular behavior of each variant. The next chapter will outline the objectives of this work.

2 OBJECTIVES

2.1 General Objective

The principal aim of this study is to elucidate the molecular mechanisms by which specific mutations — R153P, C236R, R589W, and R689S — in the XPF protein exert deleterious effects on its structural and functional integrity. This research will investigate the alterations in the biophysical properties of XPF resulting from these mutations and assess their subsequent impact on the stability and functionality of the XPF-ERCC1 heterodimer. By integrating molecular dynamics simulations and other biophysical analyses, the study seeks to unveil how these mutations disrupt the normal function of the XPF protein, particularly focusing on its interaction dynamics with ERCC1. Through this investigation, we aim to enhance our understanding of the complex molecular interactions within the NER pathway and provide insights into the pathogenic mechanisms that underlie repair deficiencies associated with these mutations.

2.2 Specific Objectives

- Accurately model the XPF structure.
- Optimize the ion cluster geometry.
- Simulate the XPF-ERCC1 heterodimer using an all-atom model on the microsecond scale.
- Simulate the R153P, C236R, R589W, and R689S mutant variants of XPF, both in their unbound forms and in complex with ERCC1.
- Characterize the structural and dynamic features of WT XPF and the XPF-ERCC1 heterodimer.
- Investigate the effects of disease-related mutations - R153P, C236R, R589W, and R689S - on XPF structure and dynamics
- Examine the molecular mechanisms underlying the pathogenic effects of each mutation, connecting structural findings with experimental observations.
- Link molecular dynamics findings to the functional consequences of XPF-ERCC1 mutations in DNA repair

2.3 Chapter Summary

This chapter presents the objectives of this study. We face the challenge of starting with an incomplete structural model of the XPF-ERCC1 protein, and our aim is to draw conclusive insights into how various mutations affect its structure. Additionally, we seek to identify potential mechanisms that connect these mutations to the observed disruptive phenotypes. The following chapter will detail the methodology employed to achieve these objectives.

3 METHODS AND CONCEPTS

3.1 Protein Structure Preparation and Modeling

The XPF protein structure was retrieved from the Protein Data Bank[1] (PDB) entry 6SXA and resolved using cryogenic electron microscopy (Cryo-EM) (JONES et al., 2020). The structure lacked several residues: 1-8, 363-365, 441-550, 642-679, and 907-916. These missing segments were modeled using the Robetta server[2] (SONG et al., 2013; BAEK et al., 2021). Additionally, the native structure was devoid of a crucial magnesium ion. Thus, the ion cluster was modeled based on the homologous XPF structure from *Aeropyrum pernix* with PDB code 2BGW (NEWMAN et al., 2005), which includes the ion. We used the CheckMyMetal (CMM) tool[3] (ZHENG et al., 2017) to assess the ion geometry quality. Initial models of the magnesium ion cluster presented a trigonal planar geometry; however, an octahedral geometry was required for physiological relevance. To correct the ion geometry, a series of non-equilibrium steered molecular dynamics (sMD) simulations were conducted, followed by structure minimization, to achieve the desired octahedral configuration. A benchmarking procedure was undertaken to identify optimal conditions for the umbrella potential and pulling rate. The selection of specific atoms for applying the force included the O atom of residue D271, the OD2 atom of residue D715, the OE2 atom of residue E725, the O atom of residue R726, and the OD1 atom of residue D731. These choices yielded the most satisfactory results. The sMD simulations were performed under the NVT ensemble with a 1000 $kJ.mol^{-1}.nm^{-2}$ restraint applied to the protein backbone and a harmonic potential of the same magnitude pulling at a 0.5 pm/ps rate. This specific setup was critical to ensuring that the backbone structure remained stable, allowing only the necessary movements in the side chains. The slow pulling rate was essential to avoid disruptive forces, permitting the overall protein structure to adapt smoothly to the geometric adjustments. Following the sMD, the structure underwent energy minimization using the steepest descent algorithm, followed by a second minimization step employing the conjugate gradient method.

[1] (https://www.rcsb.org/)
[2] (https://robetta.bakerlab.org/)
[3] https://cmm.minorlab.org/

3.2 Molecular Dynamics Simulations of XPF and Mutants

The WT XPF protein and its mutants (R153P, C236R, R589W, and R689S) were simulated using GROMACS (VAN DER SPOEL et al., 2005). Each system (hereafter called XPF^{WT}, XPF^{R153P}, XPF^{C236R}, XPF^{R589W}, and XPF^{R689S}) was prepared in quintuplicate to mitigate the effects of stochastic variability in molecular simulations (PINTO et al., 2023). The chosen force field (FF) was Amber99SB-ILDN-phi-bsc0-CUFIX[4](YOU et al., 2020; YOO, 2021), with TIP3P water model. Neutralization was achieved with Na^+ and Cl^- ions at a concentration of 0.15 M. The MD protocol followed with energy minimization using the steepest descent algorithm, followed by equilibration in the canonical ensemble (NVT) using the V-rescale thermostat (BUSSI; DONADIO; PARRINELLO, 2007) at 310.5 K and then in the isothermal-isobaric ensemble (NPT) using the Parrinello-Rahman barostat (PARRINELLO; RAHMAN, 1981; NOSÉ; KLEIN, 1983) at 1 bar. Each equilibration phase lasted 1 ns with positional restraints applied to the protein backbone. Long-range electrostatics were handled using the Particle Mesh Ewald (PME) method (ESSMANN et al., 1995), and bond constraints were managed using the LINear Constraint Solver (LINCS) algorithm (HESS et al., 1997). The production phase for each replica extended to 2 μs, with a time step of 2 fs.

3.3 Modeling and Simulation of XPF-ERCC1 Interactions

To explore the interaction dynamics between XPF and ERCC1, both proteins were modeled using AlphaFold2-Multimer (JUMPER et al., 2021; EVANS et al., 2021) to predict their complex structure. Short MD simulations were conducted to refine the predicted complex structure. The XPF-ERCC1 complex, both in its WT form and containing the mutations above, was subsequently simulated under the same conditions as those used for the XPF protein alone.

[4]https://bionano.physics.illinois.edu/CUFIX

3.4 Principal Component Analysis

Principal Component Analysis (PCA) was used to identify the collective dynamics observed within our systems. This method leverages the eigenvectors derived from the covariance matrix of atomic positional fluctuations to reveal high-amplitude, concerted movements in MD trajectories (PAPALEO et al., 2009). The construction of the covariance matrix entailed concatenating the protein backbone data from all systems and replicas into a singular trajectory after removing translational and rotational movements. This preprocessing is critical as it isolates the principal axes that predominantly capture the intrinsic protein dynamics. Following the concatenation, the covariance matrix was subject to diagonalization, a mathematical process that decomposes the covariance matrix C into its constituent components $C = V\Lambda V^T$, where Λ is a diagonal matrix of eigenvalues, and V contains the corresponding eigenvectors. This procedure, executed within the framework provided by GROMACS, culminated in identifying protein motion through the projection onto the first two principal eigenvectors (PC1 versus PC2), containing the larger variance.

3.5 Dynamic Cross-Correlation Matrix

The Dynamic Cross-Correlation Matrix (DCCM) allows us to delineate regions within the protein where the movements of amino acid residues are synchronously coordinated with those in the distal areas. Correlation values within this matrix range from -1 to 1, where 1 signifies complete correlation, indicating that pairs of atoms move in the same direction. Conversely, a value of -1 represents complete anti-correlation, meaning that atom pairs move in opposite directions. A correlation value of zero suggests that the movements of two atoms are orthogonal, thereby implying the absence of correlation between their motions (YU; DALBY, 2020). The analytical process begins with constructing a covariance matrix, which is based on the alpha carbon atoms' position across the course of MD simulations. Each element in this matrix represents the covariance between the positional fluctuations of a pair of residues. Subsequently, this covariance matrix is transformed into a cross-correlation matrix by normalizing the covariance values by the variances of the respective residues' movements. This standardization ensures that the correla-

tion values are dimensionless, facilitating direct comparison across different pairs of residues. The resulting DCCM is represented as an n x n plot, where n denotes the number of residues in the protein.

3.6 Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) and Generalized Born Surface Area (MM/GBSA)

3.6.1 MM/PBSA

The Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method estimates the binding free energy (ΔG_{bind}) of biomolecular complexes by combining molecular mechanics energy, solvation free energy, and entropy contributions (TUCCINARDI, 2021). The total binding free energy is calculated as:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S \tag{3.1}$$

Here, ΔE_{MM} represents the gas-phase molecular mechanics energy, which includes internal energy (ΔE_{int}), electrostatic interactions (ΔE_{elec}), and van der Waals interactions (ΔE_{vdW}):

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} \tag{3.2}$$

The solvation free energy (ΔG_{solv}) is split into polar and nonpolar components. The polar contribution (ΔG_{PB}) is calculated using the Poisson-Boltzmann equation, treating the solvent as a continuum medium. The nonpolar contribution (ΔG_{SA}) is estimated based on the solvent-accessible surface area (SASA):

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{SA}} \tag{3.3}$$

Entropy contributions ($T\Delta S$) can be derived from normal mode analysis or similar methods, although they are often omitted due to computational cost (POLI et al., 2020).

The MM/PBSA approach evaluates these energy terms for the complex, receptor, and ligand states across a series of MD snapshots, with the final binding free energy obtained by averaging results over the selected frames. This method

is widely used for its balance of computational efficiency and accuracy, providing insights into the energetics of biomolecular interactions.

In this work, the `gmx_MMPBSA` was employed (VALDÉS-TRESANCO et al., 2021) for all calculations, and the parameters were adjusted to optimize protein-protein interaction calculations according to the original source. The system was prepared using the AMBER ff99SBildn force field to generate topology files, incorporating ion parameters and optimized radii (`PBRadii = 3`) to improve accuracy. The simulations were carried out at 310.15 K, representing physiological conditions. Solvated molecular dynamics trajectories were processed by removing solvent molecules and ions before the binding energy analysis, ensuring the trajectories were compatible with MM/PBSA calculations. Frames were extracted starting from frame 1, with a frame interval of 1000, up to the last frame available.

For electrostatic solvation energy, the linearized Poisson-Boltzmann (`ipb=2`) model was employed, with the internal and external dielectric constants set to $\epsilon_{\text{in}} = 1.0$ and $\epsilon_{\text{out}} = 80.0$, respectively. The ionic strength was fixed at 0.15 M to mimic physiological conditions, and a probe radius of 1.4 Å was used to define solvent accessibility. The nonpolar solvation energy was determined using solvent-accessible surface area (SASA), calculated with a probe radius of 0.557 Å. The nonpolar contributions incorporated a cavity offset of -0.5692 and a surface tension of 0.0378 kcal/mol/Å^2. Grid spacing for the finite difference solver was set to 0.5 Å to balance computational efficiency and accuracy, with dielectric smoothing applied at solute-solvent boundaries. An iterative solver with a convergence criterion of 10^{-3} ensured accurate results, and an electrostatic focusing calculation was performed to refine the grid for improved precision. Energy decomposition was conducted to evaluate the contributions of individual residues to the binding energy. Nonpolar interactions were calculated using van der Waals radii, and molecular volume was determined as the volume enclosed by the SASA. All parameters were carefully optimized to ensure reliable and reproducible results for the binding energy calculations.

3.6.2 MM/GBSA

The Molecular Mechanics Generalized Born Surface Area (MM/GBSA) method uses implicit solvent model to estimate the free energy of solvation in biomolecular systems. The binding free energy (ΔG_{bind}) is calculated as the difference between

the free energy of the complex (G_{complex}) and the sum of the free energies of the individual components (G_{receptor} and G_{ligand}):

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}})$$

The free energy for each component (G) is further decomposed into molecular mechanics energy (E_{MM}), polar solvation energy (G_{polar}), nonpolar solvation energy (G_{nonpolar}), and an optional entropic term ($-T\Delta S$):

$$G = E_{\text{MM}} + G_{\text{polar}} + G_{\text{nonpolar}} - T\Delta S$$

In MM/GBSA, the polar solvation energy (G_{polar}) is estimated using the Generalized Born (GB) model, which approximates the Poisson-Boltzmann equation. The GB energy is calculated as:

$$G_{\text{polar}} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j f_{\text{GB}}(r_{ij}, R_i, R_j)$$

where q_i and q_j are the atomic charges, r_{ij} is the distance between atoms i and j, R_i and R_j are the effective Born radii, and f_{GB} is a function that accounts for the dielectric properties of the solute and solvent.

The nonpolar solvation energy (G_{nonpolar}) is estimated using the solvent-accessible surface area (SASA) method:

$$G_{\text{nonpolar}} = \gamma \cdot \text{SASA} + b$$

where γ is the surface tension coefficient, SASA is the solvent-accessible surface area, and b is an empirical offset.

For this study, MM/GBSA calculations were also performed using the `gmx_MMPBSA` plugin integrated with GROMACS (VALDÉS-TRESANCO et al., 2021). The system topology was prepared using the AMBER ff99SBildn force field, with optimized radii (`PBRadii = 3`) to enhance accuracy. Simulations were carried out at 310.15 K to mimic physiological conditions. Trajectory frames were extracted starting from frame 1, with an interval of 1 frame, up to the final available frame. Solvated trajectories were used, but solvent and ions were stripped before the analysis.

The MM/GBSA model was applied with the Generalized Born model (`igb = 2`) to calculate the polar solvation energy. The internal dielectric constant was set

to $\epsilon_{\text{in}} = 1.0$, and the external dielectric constant was set to $\epsilon_{\text{out}} = 78.5$ to represent water as the solvent. Salt effects were accounted for using a salt concentration of 0.15 M. The nonpolar solvation energy was computed using SASA, with a surface tension coefficient of 0.0072 kcal/mol/Å^2 and a probe radius of 1.4 Å.

3.7 Chapter Summary

This chapter outlines the methodology used in this study. The methods are presented in a sequential manner to ensure clarity at each stage. Given that the original PDB was lacking several segments and a crucial Mg^{2+} ion, the structure was modeled, and sMD was utilized to optimize the geometry of the ion cluster. Additionally, the XPF-ERCC1 PPI and the XPF mutant structures were modeled. Following this, MD simulations were conducted for all the different systems in quintuplicate. Finally, MM/PBSA and MM/GBSA methods were employed to evaluate the binding affinity of the XPF-ERCC1 complex.

4 CONCLUSIONS

This work has provided a comprehensive investigation into the structural and dynamic features of the XPF protein and its heterodimeric complex with ERCC1, revealing critical insights into their roles in the NER pathway. Through MD simulations, we have characterized the baseline conformational behavior of the WT systems and elucidated the structural perturbations introduced by pathogenic mutations. These findings enhance our understanding of the molecular mechanisms underlying DNA repair and its disruption in disease contexts.

The study of WT XPF revealed its inherent flexibility in the unbound state, where it transitions between open and compact conformations, aligning with its role in preparing for ERCC1 binding. Upon heterodimerization, XPF undergoes significant structural reorganization, including restricted domain motions, decreased flexibility, and increased helical content, particularly in regions critical for catalytic activity. These conformational shifts are accompanied by the disruption of inhibitory domain interactions and the emergence of coordinated domain-domain communication, which stabilize the active conformation of the XPF-ERCC1 complex. PCA further confirmed these changes, highlighting the transition from a relaxed, equilibrium state in unbound XPF to a catalytically competent ensemble in the heterodimeric form.

The structural and dynamic differences between WT and mutant systems shed light on the molecular mechanisms of XPF-related pathologies. The R153P mutation, despite preserving catalytic potential in the XPF-ERCC1 complex, introduced increased flexibility and structural instability in unbound XPF, suggesting that its pathogenicity arises from cellular mislocalization rather than enzymatic dysfunction. The C236R mutation, associated with a severe clinical phenotype, displayed significant reductions in flexibility and altered correlation patterns, particularly within the helicase and nuclease domains. These changes likely impair the dynamic transitions required for efficient catalysis, aligning with experimental observations of reduced nuclease activity and hypersensitivity to UV damage.

In contrast, the R589W mutation exhibited the most severe structural disruptions, with heightened variability and fragmented inter-domain communication. The conformational landscape of XPF-ERCC1 R589W was disorganized, suggesting that this variant struggles to achieve and maintain active states, consistent with

its functional instability. The R689S mutation, while retaining partial functionality in NER, introduced deviations in inter-domain interactions and conformational sampling, which hindered the precise coordination required for ICL repair. These alterations likely contribute to delayed release kinetics and reduced efficiency in resolving complex DNA lesions.

Overall, the results emphasize the critical balance between structural stability and dynamic flexibility required for XPF-ERCC1 function. The ability of the WT system to dynamically transition between states highlights the finely tuned interplay of inter-domain interactions and allosteric regulation essential for its activity. In contrast, the mutations examined disrupt this balance to varying degrees, providing a molecular basis for their pathogenic phenotypes.

This work advances our understanding of the structural and dynamic mechanisms of XPF in DNA repair, bridging computational insights with experimental findings. The results also open avenues for further exploration, including enhanced sampling methods and Markov state models to investigate the energy landscapes governing conformational transitions, as well as QM/MM approaches to examine catalytic mechanisms. Experimental validation using time-resolved cryo-EM or fluorescence-based assays could complement these computational studies, providing a more complete picture of XPF-ERCC1 function and its disruption by pathogenic mutations.

4.1 Chapter Summary

This thesis successfully achieved all its specific objectives, providing comprehensive insights into the structural and dynamic mechanisms of the XPF protein and its heterodimer with ERCC1 in DNA repair. The XPF structure was accurately modeled, and the ion cluster geometry was optimized, ensuring a robust foundation for subsequent simulations. Using all-atom molecular dynamics simulations on the microsecond scale, we characterized the conformational dynamics of wild-type XPF and XPF-ERCC1, capturing the structural transitions essential for activation and catalytic function.

Simulations of the R153P, C236R, R589W, and R689S mutant variants, both unbound and in complex with ERCC1, revealed the distinct structural and dynamic perturbations introduced by each mutation. These findings explained their

pathogenic mechanisms, from mislocalization to impaired inter-domain communication, flexibility, and catalytic inefficiency. Structural and dynamic insights were directly linked to the functional consequences of these mutations in NER and ICL repair pathways, providing molecular-level explanations for their experimentally observed phenotypes.

In the following chapters (**??-??**), all additional works completed during my doctorate that are unrelated to the main research will be presented.

REFERENCES

AHMAD, A.; ENZLIN, J. H.; BHAGWAT, N. R.; WIJGERS, N.; RAAMS, A.; APPLEDOORN, E.; THEIL, A. F.; HOEIJMAKERS, J. H. J.; VERMEULEN, W.; JASPERS, N. G. J. et al. Mislocalization of xpf-ercc1 nuclease contributes to reduced dna repair in xp-f patients. **PLoS genetics**, Public Library of Science San Francisco, USA, v. 6, n. 3, p. e1000871, 2010. 33

ARAKI, M.; MASUTANI, C.; TAKEMURA, M.; UCHIDA, A.; SUGASAWA, K.; KONDOH, J.; OHKUMA, Y.; HANAOKA, F. Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group c complex that initiates global genome nucleotide excision repair. **Journal of Biological Chemistry**, ASBMB, v. 276, n. 22, p. 18665–18672, 2001. 25

BAEK, M.; DIMAIO, F.; ANISHCHENKO, I.; DAUPARAS, J.; OVCHINNIKOV, S.; LEE, G. R.; WANG, J.; CONG, Q.; KINCH, L. N.; SCHAEFFER, R. D. et al. Accurate prediction of protein structures and interactions using a three-track neural network. **Science**, American Association for the Advancement of Science, v. 373, n. 6557, p. 871–876, 2021. 55

BAYKAL, C.; ATCI, T.; YILMAZ, Z.; BÜYÜKBABANI, N. Skin tumors in xeroderma pigmentosum: evaluation of a large series and a literature review. **Journal of Cutaneous Pathology**, Wiley Online Library, v. 48, n. 7, p. 884–895, 2021. 23

BERENS, P. J. T.; MOLINIER, J. Formation and recognition of uv-induced dna damage within genome complexity. **International journal of molecular sciences**, MDPI, v. 21, n. 18, p. 6689, 2020. 22

BESARATINIA, A.; YOON, J.-i.; SCHROEDER, C.; BRADFORTH, S. E.; COCKBURN, M.; PFEIFER, G. P. Wavelength dependence of ultraviolet radiation-induced dna damage as determined by laser irradiation suggests that cyclobutane pyrimidine dimers are the principal dna lesions produced by terrestrial sunlight. **The FASEB Journal**, The Federation of American Societies for Experimental Biology, v. 25, n. 9, p. 3079, 2011. 19

BOGLIOLO, M.; SCHUSTER, B.; STOEPKER, C.; DERKUNT, B.; SU, Y.; RAAMS, A.; TRUJILLO, J. P.; MINGUILLÓN, J.; RAMÍREZ, M. J.; PUJOL, R. et al. Mutations in ercc4, encoding the dna-repair endonuclease xpf, cause fanconi anemia. **The American Journal of Human Genetics**, Elsevier, v. 92, n. 5, p. 800–806, 2013. 34

BOITEUX, S.; RADICELLA, J. P. Base excision repair of 8-hydroxyguanine protects dna from endogenous oxidative stress. **Biochimie**, Elsevier, v. 81, n. 1-2, p. 59–67, 1999. 22

BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of chemical physics**, AIP Publishing, v. 126, n. 1, 2007. 56

CADET, J.; BERGER, M.; DOUKI, T.; RAVANAT, J.-L. Oxidative damage to dna: formation, measurement, and biological significance. **Reviews of Physiology Biochemistry and Pharmacology, Volume 131: Special Issue on Membrane-Mediated Cellular Responses: The Roles of Reactive Oxygens, NO, CO, II**, Springer, p. 1–87, 2005. 22

CADET, J.; DAVIES, K. J. Oxidative dna damage & repair: an introduction. **Free Radical Biology and Medicine**, Elsevier, v. 107, p. 2–12, 2017. 22

CADET, J.; DOUKI, T. Formation of uv-induced dna damage contributing to skin cancer development. **Photochemical & Photobiological Sciences**, Royal Society of Chemistry, v. 17, n. 12, p. 1816–1841, 2018. 18, 19, 21

CADET, J.; MOURET, S.; RAVANAT, J.-L.; DOUKI, T. Photoinduced damage to cellular dna: direct and photosensitized reactions. **Photochemistry and Photobiology**, Wiley Online Library, v. 88, n. 5, p. 1048–1065, 2012. 19

CADET, J.; SAGE, E.; DOUKI, T. Ultraviolet radiation-mediated damage to cellular dna. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, Elsevier, v. 571, n. 1-2, p. 3–17, 2005. 19

CANNAN, W. J.; PEDERSON, D. S. Mechanisms and consequences of double-strand dna break formation in chromatin. **Journal of cellular physiology**, Wiley Online Library, v. 231, n. 1, p. 3–14, 2016. 23

COELHO, S. G.; YIN, L.; SMUDA, C.; MAHNS, A.; KOLBE, L.; HEARING, V. J. Photobiological implications of melanin photoprotection after uvb-induced tanning of human skin but not uva-induced tanning. **Pigment cell & melanoma research**, Wiley Online Library, v. 28, n. 2, p. 210–216, 2015. 22

COMPE, E.; EGLY, J.-M. Tfiih: when transcription met dna repair. **Nature reviews Molecular cell biology**, Nature Publishing Group UK London, v. 13, n. 6, p. 343–354, 2012. 27

DIFFEY, B. L. Sources and measurement of ultraviolet radiation. **Methods**, Elsevier, v. 28, n. 1, p. 4–13, 2002. 18

DÖHRING, T. Critical discussion on the uv absorption properties of earth's atmosphere. In: SPIE. **Third International Conference on Applications of Optics and Photonics**. [S.l.], 2017. v. 10453, p. 128–137. 18

DOUKI, T.; SAGE, E. Dewar valence isomers, the third type of environmentally relevant dna photoproducts induced by solar radiation. **Photochemical & Photobiological Sciences**, Springer, v. 15, p. 24–30, 2016. 19, 21, 22

D'SOUZA, A.; BLEE, A. M.; CHAZIN, W. J. Mechanism of action of nucleotide excision repair machinery. **Biochemical Society Transactions**, Portland Press Ltd., v. 50, n. 1, p. 375–386, 2022. 25

DUAN, M.; SPEER, R. M.; ULIBARRI, J.; LIU, K. J.; MAO, P. Transcription-coupled nucleotide excision repair: new insights revealed by genomic approaches. **DNA repair**, Elsevier, v. 103, p. 103126, 2021. 28

D'SOUZA, A.; KIM, M.; CHAZIN, W. J.; SCHÄRER, O. D. Protein-protein interactions in the core nucleotide excision repair pathway. **DNA repair**, Elsevier, p. 103728, 2024. 25

EGLY, J.-M.; COIN, F. A history of tfiih: two decades of molecular biology on a pivotal transcription/repair factor. **DNA repair**, Elsevier, v. 10, n. 7, p. 714–721, 2011. 27

ESSMANN, U.; PERERA, L.; BERKOWITZ, M. L.; DARDEN, T.; LEE, H.; PEDERSEN, L. G. A smooth particle mesh ewald method. **The Journal of chemical physics**, American Institute of Physics, v. 103, n. 19, p. 8577–8593, 1995. 56

EVANS, R.; O'NEILL, M.; PRITZEL, A.; ANTROPOVA, N.; SENIOR, A.; GREEN, T.; ŽÍDEK, A.; BATES, R.; BLACKWELL, S.; YIM, J. et al. Protein complex prediction with alphafold-multimer. **biorxiv**, Cold Spring Harbor Laboratory, p. 2021–10, 2021. 56

FAN, L.; ARVAI, A. S.; COOPER, P. K.; IWAI, S.; HANAOKA, F.; TAINER, J. A. Conserved xpb core structure and motifs for dna unwinding: implications for pathway selection of transcription or excision repair. **Molecular cell**, Elsevier, v. 22, n. 1, p. 27–37, 2006. 27

FANNING, E.; KLIMOVICH, V.; NAGER, A. R. A dynamic model for replication protein a (rpa) function in dna processing pathways. **Nucleic acids research**, Oxford University Press, v. 34, n. 15, p. 4126–4137, 2006. 27

FELTES, B. C.; BONATTO, D. Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications. **Mutation Research/Reviews in Mutation Research**, Elsevier, v. 763, p. 306–320, 2015. 32

FISCHER, E. S.; SCRIMA, A.; BÖHM, K.; MATSUMOTO, S.; LINGARAJU, G. M.; FATY, M.; YASUDA, T.; CAVADINI, S.; WAKASUGI, M.; HANAOKA, F. et al. The molecular basis of crl4ddb2/csa ubiquitin ligase architecture, targeting, and activation. **Cell**, Elsevier, v. 147, n. 5, p. 1024–1039, 2011. 25

FITCH, M. E.; NAKAJIMA, S.; YASUI, A.; FORD, J. M. In vivo recruitment of xpc to uv-induced cyclobutane pyrimidine dimers by the ddb2 gene product. **Journal of Biological Chemistry**, ASBMB, v. 278, n. 47, p. 46906–46910, 2003. 25

FRAIKIN, G. Y. Role of various uv-induced dna lesions in human skin carcinogenesis. **Journal ISSN**, v. 2766, p. 2276, 2024. 18

FRIEDBERG, E. C.; WALKER, G. C.; SIEDE, W.; WOOD, R. D. **DNA repair and mutagenesis**. [S.l.]: American Society for Microbiology Press, 2005. 22

GILLJAM, K. M.; MÜLLER, R.; LIABAKK, N. B.; OTTERLEI, M. Nucleotide excision repair is associated with the replisome and its efficiency depends on a direct interaction between xpa and pcna. **PloS one**, Public Library of Science San Francisco, USA, v. 7, n. 11, p. e49199, 2012. 28

GONZÁLEZ-CORROCHANO, R.; RUIZ, F. M.; TAYLOR, N. M. I.; HUECAS, S.; DRAKULIC, S.; SPÍNOLA-AMILIBIA, M.; FERNÁNDEZ-TORNERO, C. The crystal structure of human xpg, the xeroderma pigmentosum group g endonuclease, provides insight into nucleotide excision dna repair. **Nucleic acids research**, Oxford University Press, v. 48, n. 17, p. 9943–9958, 2020. 27

GROISMAN, R.; POLANOWSKA, J.; KURAOKA, I.; SAWADA, J.-i.; SAIJO, M.; DRAPKIN, R.; KISSELEV, A. F.; TANAKA, K.; NAKATANI, Y. The ubiquitin ligase activity in the ddb2 and csa complexes is differentially regulated by the cop9 signalosome in response to dna damage. **Cell**, Elsevier, v. 113, n. 3, p. 357–367, 2003. 25

HALLIWELL, B.; GUTTERIDGE, J. M. **Free radicals in biology and medicine**. [S.l.]: Oxford university press, USA, 2015. 22

HASHIMOTO, K.; WADA, K.; MATSUMOTO, K.; MORIYA, M. Physical interaction between slx4 (fancp) and xpf (fancq) proteins and biological consequences of interaction-defective missense mutations. **DNA repair**, Elsevier, v. 35, p. 48–54, 2015. 33

HESS, B.; BEKKER, H.; BERENDSEN, H. J.; FRAAIJE, J. G. Lincs: A linear constraint solver for molecular simulations. **Journal of computational chemistry**, Wiley Online Library, v. 18, n. 12, p. 1463–1472, 1997. 56

HEUVEL, D. van den; WEEGEN, Y. van der; BOER, D. E.; OGI, T.; LUIJSTERBURG, M. S. Transcription-coupled dna repair: from mechanism to human disorder. **Trends in cell biology**, Elsevier, v. 31, n. 5, p. 359–371, 2021. 25

HUNG, K.-F.; SIDOROVA, J. M.; NGHIEM, P.; KAWASUMI, M. The 6-4 photoproduct is the trigger of uv-induced replication blockage and atr activation. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 117, n. 23, p. 12806–12816, 2020. 21

JONES, M.; BEURON, F.; BORG, A.; NANS, A.; EARL, C. P.; BRIGGS, D. C.; SNIJDERS, A. P.; BOWLES, M.; MORRIS, E. P.; LINCH, M. et al. Cryo-em structures of the xpf-ercc1 endonuclease reveal how dna-junction engagement disrupts an auto-inhibited conformation. **Nature communications**, Nature Publishing Group UK London, v. 11, n. 1, p. 1120, 2020. 27, 30, 32, 55

JUMPER, J.; EVANS, R.; PRITZEL, A.; GREEN, T.; FIGURNOV, M.; RONNEBERGER, O.; TUNYASUVUNAKOOL, K.; BATES, R.; ŽÍDEK, A.; POTAPENKO, A. et al. Highly accurate protein structure prediction with alphafold. **Nature**, Nature Publishing Group, v. 596, n. 7873, p. 583–589, 2021. 56

KARIKKINETH, A. C.; SCHEIBYE-KNUDSEN, M.; FIVENSON, E.; CROTEAU, D. L.; BOHR, V. A. Cockayne syndrome: Clinical features, model systems and pathways. **Ageing research reviews**, Elsevier, v. 33, p. 3–17, 2017. 24

KASHIYAMA, K.; NAKAZAWA, Y.; PILZ, D. T.; GUO, C.; SHIMADA, M.; SASAKI, K.; FAWCETT, H.; WING, J. F.; LEWIN, S. O.; CARR, L. et al. Malfunction of nuclease ercc1-xpf results in diverse clinical manifestations and

causes cockayne syndrome, xeroderma pigmentosum, and fanconi anemia. **The American Journal of Human Genetics**, Elsevier, v. 92, n. 5, p. 807–819, 2013. 33

KHAN, A. Q.; TRAVERS, J. B.; KEMP, M. G. Roles of uva radiation and dna damage responses in melanoma pathogenesis. **Environmental and molecular mutagenesis**, Wiley Online Library, v. 59, n. 5, p. 438–460, 2018. 19

KOKIC, G.; CHERNEV, A.; TEGUNOV, D.; DIENEMANN, C.; URLAUB, H.; CRAMER, P. Structural basis of tfiih activation for nucleotide excision repair. **Nature communications**, Nature Publishing Group UK London, v. 10, n. 1, p. 2885, 2019. 28

KUMAR, N.; MORENO, N. C.; FELTES, B. C.; MENCK, C. F.; HOUTEN, B. V. Cooperation and interplay between base and nucleotide excision repair pathways: From dna lesions to proteins. **Genetics and Molecular Biology**, SciELO Brasil, v. 43, n. 1 suppl. 1, p. e20190104, 2020. 25

KUPER, J.; KISKER, C. At the core of nucleotide excision repair. **Current Opinion in Structural Biology**, Elsevier, v. 80, p. 102605, 2023. 25

LEHMANN, A. R. Dna repair-deficient diseases, xeroderma pigmentosum, cockayne syndrome and trichothiodystrophy. **Biochimie**, Elsevier, v. 85, n. 11, p. 1101–1111, 2003. 24

LEUNG, A. K.; BARANKIN, B.; LAM, J. M.; LEONG, K. F.; HON, K. L. Xeroderma pigmentosum: an updated review. **Drugs in context**, BioExcel Publishing Ltd, v. 11, 2022. 24

LIMA-BESSA, K. M. de; ARMELINI, M. G.; CHIGANÇAS, V.; JACYSYN, J. F.; AMARANTE-MENDES, G. P.; SARASIN, A.; MENCK, C. F. M. Cpds and 6-4pps play different roles in uv-induced cell death in normal and ner-deficient human cells. **DNA repair**, Elsevier, v. 7, n. 2, p. 303–312, 2008. 21

LIU, H.; RUDOLF, J.; JOHNSON, K. A.; MCMAHON, S. A.; OKE, M.; CARTER, L.; MCROBBIE, A.-M.; BROWN, S. E.; NAISMITH, J. H.; WHITE, M. F. Structure of the dna repair helicase xpd. **Cell**, Elsevier, v. 133, n. 5, p. 801–812, 2008. 27

LUKIN, M.; SANTOS, C. de L. Nmr structures of damaged dna. **Chemical reviews**, ACS Publications, v. 106, n. 2, p. 607–686, 2006. 21

MARGUET, S.; MARKOVITSI, D. Time-resolved study of thymine dimer formation. **Journal of the American Chemical Society**, ACS Publications, v. 127, n. 16, p. 5780–5781, 2005. 21

MARKOVITSI, D. Uv-induced dna damage: the role of electronic excited states. **Photochemistry and photobiology**, Wiley Online Library, v. 92, n. 1, p. 45–51, 2016. 19

MARTEIJN, J. A.; LANS, H.; VERMEULEN, W.; HOEIJMAKERS, J. H. Understanding nucleotide excision repair and its roles in cancer and ageing.

Nature reviews Molecular cell biology, Nature Publishing Group UK London, v. 15, n. 7, p. 465–481, 2014. 26

MASUTANI, C.; SUGASAWA, K.; YANAGISAWA, J.; SONOYAMA, T.; UI, M.; ENOMOTO, T.; TAKIO, K.; TANAKA, K.; SPEK, P. Van der; BOOTSMA, D. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group c protein and a human homologue of yeast rad23. **The EMBO journal**, v. 13, n. 8, p. 1831–1843, 1994. 25

MATSUMURA, Y.; NISHIGORI, C.; YAGI, T.; IMAMURA, S.; TAKEBE, H. Characterization of molecular defects in xeroderma pigmentosum group f in relation to its clinically mild symptoms. **Human molecular genetics**, Oxford University Press, v. 7, n. 6, p. 969–974, 1998. 33

MOIRANGTHEM, R.; GAMAGE, M. N.; ROKITA, S. E. Dynamic accumulation of cyclobutane pyrimidine dimers and its response to changes in dna conformation. **Nucleic Acids Research**, Oxford University Press, v. 51, n. 11, p. 5341–5350, 2023. 21

MORENO, N. N.; OLTHOF, A. M.; SVEJSTRUP, J. Q. Transcription-coupled nucleotide excision repair and the transcriptional response to uv-induced dna damage. **Annual review of biochemistry**, Annual Reviews, v. 92, n. 1, p. 81–113, 2023. 25

MORI, T.; YOUSEFZADEH, M. J.; FARIDOUNNIA, M.; CHONG, J. X.; HISAMA, F. M.; HUDGINS, L.; MERCADO, G.; WADE, E. A.; BARGHOUTHY, A. S.; LEE, L. et al. Ercc4 variants identified in a cohort of patients with segmental progeroid syndromes. **Human mutation**, Wiley Online Library, v. 39, n. 2, p. 255–265, 2018. 24

MOURET, S.; BAUDOUIN, C.; CHARVERON, M.; FAVIER, A.; CADET, J.; DOUKI, T. Cyclobutane pyrimidine dimers are predominant dna lesions in whole human skin exposed to uva radiation. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 103, n. 37, p. 13765–13770, 2006. 19

MOURET, S.; PHILIPPE, C.; GRACIA-CHANTEGREL, J.; BANYASZ, A.; KARPATI, S.; MARKOVITSI, D.; DOUKI, T. Uva-induced cyclobutane pyrimidine dimers in dna: a direct photochemical mechanism? **Organic & biomolecular chemistry**, Royal Society of Chemistry, v. 8, n. 7, p. 1706–1711, 2010. 19

NEWMAN, M.; MURRAY-RUST, J.; LALLY, J.; RUDOLF, J.; FADDEN, A.; KNOWLES, P. P.; WHITE, M. F.; MCDONALD, N. Q. Structure of an xpf endonuclease with and without dna suggests a model for substrate recognition. **The EMBO journal**, John Wiley & Sons, Ltd Chichester, UK, v. 24, n. 5, p. 895–905, 2005. 55

NIEDERNHOFER, L. J.; GARINIS, G. A.; RAAMS, A.; LALAI, A. S.; ROBINSON, A. R.; APPELDOORN, E.; ODIJK, H.; OOSTENDORP, R.; AHMAD, A.; LEEUWEN, W. V. et al. A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. **Nature**, Nature Publishing Group UK London, v. 444, n. 7122, p. 1038–1043, 2006. 32

NOSÉ, S.; KLEIN, M. Constant pressure molecular dynamics for molecular systems. **Molecular Physics**, Taylor & Francis, v. 50, n. 5, p. 1055–1076, 1983. 56

OGI, T.; LIMSIRICHAIKUL, S.; OVERMEER, R. M.; VOLKER, M.; TAKENAKA, K.; CLONEY, R.; NAKAZAWA, Y.; NIIMI, A.; MIKI, Y.; JASPERS, N. G. et al. Three dna polymerases, recruited by different mechanisms, carry out ner repair synthesis in human cells. **Molecular cell**, Elsevier, v. 37, n. 5, p. 714–727, 2010. 28

OKUDA, M.; KINOSHITA, M.; KAKUMU, E.; SUGASAWA, K.; NISHIMURA, Y. Structural insight into the mechanism of tfiih recognition by the acidic string of the nucleotide excision repair factor xpc. **Structure**, Elsevier, v. 23, n. 10, p. 1827–1837, 2015. 27

OSORIO, A.; BOGLIOLO, M.; FERNÁNDEZ, V.; BARROSO, A.; HOYA, M. De la; CALDÉS, T.; LASA, A.; CAJAL, T. Ramon y; SANTAMARIÑA, M.; VEGA, A. et al. Evaluation of rare variants in the new f anconi anemia gene ercc 4 (fancq) as familial breast/ovarian cancer susceptibility alleles. **Human mutation**, Wiley Online Library, v. 34, n. 12, p. 1615–1618, 2013. 34

PAPALEO, E.; MEREGHETTI, P.; FANTUCCI, P.; GRANDORI, R.; GIOIA, L. D. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. **Journal of molecular graphics and modelling**, Elsevier, v. 27, n. 8, p. 889–899, 2009. 57

PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied physics**, AIP Publishing, v. 52, n. 12, p. 7182–7190, 1981. 56

PAUL-KONIETZKO, K.; THOMALE, J.; ARAKAWA, H.; ILIAKIS, G. Dna ligases i and iii support nucleotide excision repair in dt 40 cells with similar efficiency. **Photochemistry and photobiology**, Wiley Online Library, v. 91, n. 5, p. 1173–1180, 2015. 28, 29

PICCIONE, M.; FORTINA, A. B.; FERRI, G.; ANDOLINA, G.; BERETTA, L.; CIVIDINI, A.; MARNI, E. D.; CAROPPO, F.; CITERNESI, U.; LIDDO, R. D. Xeroderma pigmentosum: general aspects and management. **Journal of Personalized Medicine**, MDPI, v. 11, n. 11, p. 1146, 2021. 23

PINTO, E. S. M.; KRAUSE, M. J.; DORN, M.; FELTES, B. C. The nucleotide excision repair proteins through the lens of molecular dynamics simulations. **DNA repair**, Elsevier, p. 103510, 2023. 56

POLI, G.; GRANCHI, C.; RIZZOLIO, F.; TUCCINARDI, T. Application of mm-pbsa methods in virtual screening. **Molecules**, MDPI, v. 25, n. 8, p. 1971, 2020. 58

PREMI, S.; HAN, L.; MEHTA, S.; KNIGHT, J.; ZHAO, D.; PALMATIER, M. A.; KORNACKER, K.; BRASH, D. E. Genomic sites hypersensitive to ultraviolet radiation. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 116, n. 48, p. 24196–24205, 2019. 21

RASTOGI, R. P.; RICHA, n.; KUMAR, A.; TYAGI, M. B.; SINHA, R. P. Molecular mechanisms of ultraviolet radiation-induced dna damage and repair. **Journal of nucleic acids**, Wiley Online Library, v. 2010, n. 1, p. 592980, 2010. 19, 21

RIBEIRO-SILVA, C.; SABATELLA, M.; HELFRICHT, A.; MARTEIJN, J. A.; THEIL, A. F.; VERMEULEN, W.; LANS, H. Ubiquitin and tfiih-stimulated ddb2 dissociation drives dna damage handover in nucleotide excision repair. **Nature Communications**, Nature Publishing Group, v. 11, n. 1, p. 1–14, 2020. 25

SAIJO, M.; TAKEDACHI, A.; TANAKA, K. Nucleotide excision repair by mutant xeroderma pigmentosum group a (xpa) proteins with deficiency in interaction with rpa. **Journal of Biological Chemistry**, ASBMB, v. 286, n. 7, p. 5476–5483, 2011. 27

SALE, J. E. Translesion dna synthesis and mutagenesis in eukaryotes. **Cold Spring Harbor perspectives in biology**, Cold Spring Harbor Lab, v. 5, n. 3, p. a012708, 2013. 23

SCHREIER, W. J.; SCHRADER, T. E.; KOLLER, F. O.; GILCH, P.; CRESPO-HERNÁNDEZ, C. E.; SWAMINATHAN, V. N.; CARELL, T.; ZINTH, W.; KOHLER, B. Thymine dimerization in dna is an ultrafast photoreaction. **Science**, American Association for the Advancement of Science, v. 315, n. 5812, p. 625–629, 2007. 19

SONG, Y.; DIMAIO, F.; WANG, R. Y.-R.; KIM, D.; MILES, C.; BRUNETTE, T.; THOMPSON, J.; BAKER, D. High-resolution comparative modeling with rosettacm. **Structure**, Elsevier, v. 21, n. 10, p. 1735–1742, 2013. 55

STARESINCIC, L.; FAGBEMI, A. F.; ENZLIN, J. H.; GOURDIN, A. M.; WIJGERS, N.; DUNAND-SAUTHIER, I.; GIGLIA-MARI, G.; CLARKSON, S. G.; VERMEULEN, W.; SCHÄRER, O. D. Coordination of dual incision and repair synthesis in human nucleotide excision repair. **The EMBO journal**, John Wiley & Sons, Ltd Chichester, UK, v. 28, n. 8, p. 1111–1120, 2009. 27, 29

STEFANINI, M.; BOTTA, E.; LANZAFAME, M.; ORIOLI, D. Trichothiodystrophy: from basic mechanisms to clinical implications. **DNA repair**, Elsevier, v. 9, n. 1, p. 2–10, 2010. 24

STEFANINI, M.; FAWCETT, H.; BOTTA, E.; NARDO, T.; LEHMANN, A. R. Genetic analysis of twenty-two patients with cockayne syndrome. **Human genetics**, Springer, v. 97, p. 418–423, 1996. 24

SUGASAWA, K.; SHIMIZU, Y.; IWAI, S.; HANAOKA, F. A molecular mechanism for dna damage recognition by the xeroderma pigmentosum group c protein complex. **DNA repair**, Elsevier, v. 1, n. 1, p. 95–107, 2002. 25

TIWARI, V.; WILSON, D. M. Dna damage and associated dna repair defects in disease and premature aging. **The American Journal of Human Genetics**, Elsevier, v. 105, n. 2, p. 237–257, 2019. 23

TOPOLSKA-WOŚ, A. M.; SUGITANI, N.; CORDOBA, J. J.; MEUR, K. V. L.; MEUR, R. A. L.; KIM, H. S.; YEO, J.-E.; ROSENBERG, D.; HAMMEL, M.; SCHÄRER, O. D. et al. A key interaction with rpa orients xpa in ner complexes. **Nucleic acids research**, Oxford University Press, v. 48, n. 4, p. 2173–2188, 2020. 29

TUCCINARDI, T. What is the current value of mm/pbsa and mm/gbsa methods in drug discovery? **Expert opinion on drug discovery**, Taylor & Francis, v. 16, n. 11, p. 1233–1237, 2021. 58

VALDÉS-TRESANCO, M. S.; VALDÉS-TRESANCO, M. E.; VALIENTE, P. A.; MORENO, E. gmx_mmpbsa: a new tool to perform end-state free energy calculations with gromacs. **Journal of chemical theory and computation**, ACS Publications, v. 17, n. 10, p. 6281–6291, 2021. 59, 60

VAN DER SPOEL, D.; LINDAHL, E.; HESS, B.; GROENHOF, G.; MARK, A. E.; BERENDSEN, H. J. Gromacs: fast, flexible, and free. **Journal of computational chemistry**, Wiley Online Library, v. 26, n. 16, p. 1701–1718, 2005. 56

WEEGEN, Y. van der; GOLAN-BERMAN, H.; MEVISSEN, T. E.; APELT, K.; GONZÁLEZ-PRIETO, R.; GOEDHART, J.; HEILBRUN, E. E.; VERTEGAAL, A. C.; HEUVEL, D. van den; WALTER, J. C. et al. The cooperative action of csb, csa, and uvssa target tfiih to dna damage-stalled rna polymerase ii. **Nature communications**, Nature Publishing Group UK London, v. 11, n. 1, p. 2104, 2020. 28

WITTSCHIEBEN, B. Ø.; IWAI, S.; WOOD, R. D. Ddb1-ddb2 (xeroderma pigmentosum group e) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in dna. **Journal of Biological Chemistry**, ASBMB, v. 280, n. 48, p. 39982–39989, 2005. 25

YOO, J. On the stability of protein–dna complexes in molecular dynamics simulations using the cufix corrections. **Journal of the Korean Physical Society**, Springer, v. 78, n. 5, p. 461–466, 2021. 56

YOU, S.; LEE, H.-G.; KIM, K.; YOO, J. Improved parameterization of protein–dna interactions for molecular dynamics simulations of pcna diffusion on dna. **Journal of Chemical Theory and Computation**, ACS Publications, v. 16, n. 7, p. 4006–4013, 2020. 56

YU, H.; DALBY, P. A. A beginner's guide to molecular dynamics simulations and the identification of cross-correlation networks for enzyme engineering. In: **Methods in enzymology**. [S.l.]: Elsevier, 2020. v. 643, p. 15–49. 57

ZHANG, X.; HORIBATA, K.; SAIJO, M.; ISHIGAMI, C.; UKAI, A.; KANNO, S.-i.; TAHARA, H.; NEILAN, E. G.; HONMA, M.; NOHMI, T. et al. Mutations in uvssa cause uv-sensitive syndrome and destabilize ercc6 in transcription-coupled dna repair. **Nature genetics**, Nature Publishing Group US New York, v. 44, n. 5, p. 593–597, 2012. 28

ZHAO, Y.; SIMON, M.; SELUANOV, A.; GORBUNOVA, V. Dna damage and repair in age-related inflammation. **Nature Reviews Immunology**, Nature Publishing Group UK London, v. 23, n. 2, p. 75–89, 2023. 24

ZHENG, H.; COOPER, D. R.; POREBSKI, P. J.; SHABALIN, I. G.; HANDING, K. B.; MINOR, W. Checkmymetal: a macromolecular metal-binding validation tool. **Acta Crystallographica Section D: Structural Biology**, International Union of Crystallography, v. 73, n. 3, p. 223–233, 2017. 55