



Review

# Exploring DNA Damage and Repair Mechanisms: A Review with Computational Insights

Jiawei Chen <sup>1</sup>, Ravi Potlapalli <sup>2</sup>, Heng Quan <sup>3</sup>, Lingtao Chen <sup>2</sup>, Ying Xie <sup>2</sup>, Seyedamin Pouriyeh <sup>2</sup>, Nazmus Sakib <sup>2</sup>, Lichao Liu <sup>4</sup> and Yixin Xie <sup>2,\*</sup>

- <sup>1</sup> College of Letter and Science, University of California, Berkeley, CA 94720, USA; jc01@berkeley.edu  
<sup>2</sup> College of Computing and Software Engineering, Kennesaw State University, Marietta, GA 30060, USA; lchen25@students.kennesaw.edu (L.C.); rpotlapa@students.kennesaw.edu (R.P.); yxie2@kennesaw.edu (Y.X.); spouriyeh@kennesaw.edu (S.P.); nsakib1@kennesaw.edu (N.S.)  
<sup>3</sup> Department of Civil and Urban Engineering, New York University, New York, NY 11201, USA; hq322@nyu.edu  
<sup>4</sup> Stanford Cardiovascular Institute, Stanford University School of Medicine, Palo Alto, CA 94304, USA; lliu@stanford.edu  
\* Correspondence: yxie11@kennesaw.edu

**Abstract:** DNA damage is a critical factor contributing to genetic alterations, directly affecting human health, including developing diseases such as cancer and age-related disorders. DNA repair mechanisms play a pivotal role in safeguarding genetic integrity and preventing the onset of these ailments. Over the past decade, substantial progress and pivotal discoveries have been achieved in DNA damage and repair. This comprehensive review paper consolidates research efforts, focusing on DNA repair mechanisms, computational research methods, and associated databases. Our work is a valuable resource for scientists and researchers engaged in computational DNA research, offering the latest insights into DNA-related proteins, diseases, and cutting-edge methodologies. The review addresses key questions, including the major types of DNA damage, common DNA repair mechanisms, the availability of reliable databases for DNA damage and associated diseases, and the predominant computational research methods for enzymes involved in DNA damage and repair.



**Citation:** Chen, J.; Potlapalli, R.; Quan, H.; Chen, L.; Xie, Y.; Pouriyeh, S.; Sakib, N.; Liu, L.; Xie, Y. Exploring DNA Damage and Repair Mechanisms: A Review with Computational Insights. *BioTech* **2024**, *13*, 3. <https://doi.org/10.3390/biotech13010003>

Academic Editor:

Ioannis Michalopoulos

Received: 29 October 2023

Revised: 21 November 2023

Accepted: 29 December 2023

Published: 16 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** cancer; aging-related diseases; DNA damage; DNA repair mechanisms; uracil-DNA glycosylase; computational biology; DNA database

**Key Contribution:** This paper overviewed DNA damage and repair mechanisms, and it summarizes useful computational methods and databases for researchers to study in this field. Also, it points out potential research directions in DNA-related fields.

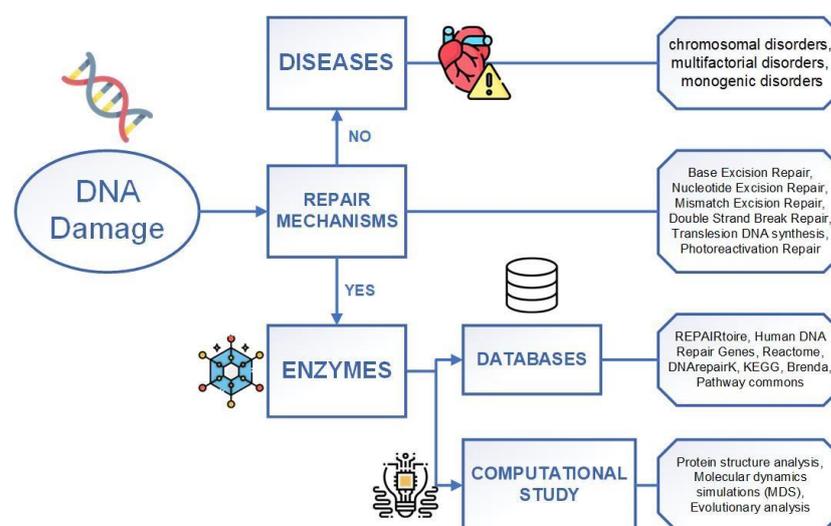
## 1. Introduction

DNA damage occurs at a high rate per second, and it causes a change in the genetic information [1–4]. This may cause cell loss or even the transformation of normal cells to cancer cells. Each cell suffers ten thousand to one million DNA lesions per day [5]. There are two significant resources of DNA damage: exogenous resources, including X-rays, toxins, viruses, bacteria, etc., and endogenous resources, including reactive oxygen species (ROS) [6]. As DNA damage is harmful to normal cells and a significant threat to human health, various mechanisms in DNA repair fix the damage and errors that occur in different cell processes [7,8]. An example of DNA damage is the uracil replacement of cytosine caused by spontaneous deamination, usually excised from DNA by the enzyme uracil-DNA glycosylase. There are several DNA repair pathways to fix DNA damages, including nucleotide excision repair (NER), base excision repair (BER), and mismatch repair (MMR), which are active in different cell cycle stages [9,10]. In addition to those three, homologous

recombination (HR) and non-homologous end joining (NHEJ) are discussed frequently in research on DNA damage and repair [11].

The main reason why we care about DNA damage and repair is that DNA damage and sub-optimal DNA damage response (DDR) events lead to diseases, including neurodegenerative diseases that are classified into three types, chromosomal disorders, multifactorial disorders, monogenic disorders: (1) chromosomal disorders, such as Cockayne syndrome; (2) multifactorial disorders, such as Alzheimer's disease; (3) cancers, such as breast cancer; and (4) monogenic disorders, like ataxia-telangiectasia, age-related macular degeneration, heart disease, etc. [12–14]. Endogenous or exogenous cellular processes cause all these diseases mentioned above. The oxidation of nitrogen bases and generation of reactive oxygen species disrupt DNA strands, the alkylation of bases, and hydrolysis, including deamination, depurination, and depyrimidination. The development of bulky adducts is an example of endogenous biological activities, the mismatch of bases because of errors in DNA replication, and monoadduct damage due to a change in the mononitrogen base. DNA adduct damage also results in diseases like diabetes, Parkinson's disease, heart disease, and atherosclerosis [15,16]. Industrial chemicals such as vinyl chloride, polycyclic aromatic hydrocarbons, and hydrogen peroxide lead to diseases like hereditary diseases, macular degeneration, and sporadic cancer [15]. Research on DNA repair enzymes has been performed and studied since the 1970s [17,18]. The enzymes concerned in DNA restoration are methylguanine methyltransferase, uracil-DNA glycosylase, DNA polymerase  $\beta$ , poly (ADP-ribose) polymerase-1 [6,19], and DNA ligase.

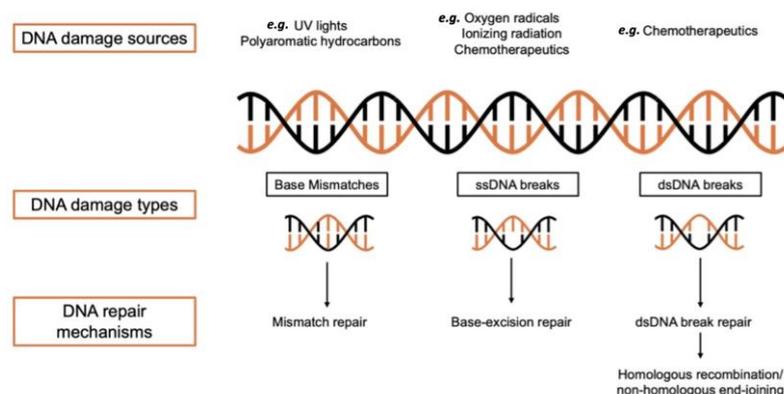
In addition to wet lab research approaches, such as an information management system for clinical genome sequencing [18,19], computational studies are also essential in investigating DNA damage and repair. Computational studies have become effective and efficient with the fast development of computer technologies. They show extraordinary abilities and potential in dealing with large-scale data in DNA study, and they fasten the discovery of biological mechanisms. In computational studies, a database is important in managing and organizing the data, which can provide a structured framework of data to help researchers better understand and access them more efficiently [20]. In DNA repair studies, databases or datasets of human diseases correlate with gene mutations relevant to DNA integrity, stability, and information about DNA damage caused by mutagenic agents [21]. So, in this review article, we also included DNA-related databases and computational methods that have been used for decades in DNA research (Figure 1).



**Figure 1.** A structural representation of this review paper. It describes how repair mechanisms can repair DNA damage by utilizing enzymes.

## 2. DNA Repair Mechanisms

The effective response to DNA damage necessitates the coordinated involvement of numerous factors. The integrity of the genome must be preserved, and any potentially harmful mutations that could lead to cellular damage or tumor formation must be prevented. It is crucial to establish a background context that facilitates efficient repair by signaling the presence of DNA lesions. Since DNA repair mechanisms can also be employed as anti-cancer treatments in medical practice, various genotoxic chemicals have been used for several years to cause DNA damage [21]. DNA repair can occur through multiple mechanisms, such as BER, NER, MMR, and DSBR. The categories of DNA damage and repair are shown in Figure 2.



**Figure 2.** DNA damage types and repair mechanisms. Three major DNA repair mechanisms with their key factors and associated damage types are shown.

### 2.1. BER Mechanism

The primary repair process to eliminate DNA damage is the base excision repair pathway, or BER [22,23]. The BER (base excision repair) process is used when DNA is damaged by reactive oxygen species, single-strand breaks, or alkylating agents through oxidation. The key steps in the BER process are as follows: (1) recognition of damage, which utilizes DNA glycosylase enzyme, and each DNA glycosylase is specific to particular types of base damage; (2) removal of damaged base: the DNA glycosylase enzyme cleaves the bond between the damaged base and the sugar phosphate backbone, leaving the apurinic/aprimidinic (AP) site, which is also known as an abasic site; (3) AP endonuclease: which recognizes the AP site and makes an incision in the DNA strand and creates a single-strand break with a 3'-OH and a 5'-deoxyribosephosphate termini; (4) DNA polymerase: the enzyme fills in the gap by adding the correct base complementary to the undamaged strand; and (5) DNA ligase: as the last step, DNA ligase seals the nick in the DNA backbone, completing the repair process [20,21,24]. The BER pathway is crucial for maintaining the integrity of the genome by fixing common forms of DNA damage. It is a versatile and efficient repair mechanism, addressing a wide range of DNA lesions to ensure the stability and functionality of the genetic material.

### 2.2. NER Mechanism

The NER (nucleotide excision repair) mechanism is employed to repair the damage by creating large adducts and intra-strand crosslinks when UV light and polycyclic aromatic hydrocarbons damage DNA. The nucleotide excision respirasome, a multi-protein complex, performs the NER process in mammals [25–27]. The excision of about twenty-eight nucleotide DNA segments furnishing the damaged site is the primary process in eukaryotic NER [27,28]. The two different sub-pathways of global genome repair (GGR) and transcription-coupled repair (TCR) make up NER in mammalian cells [29–31]. The XPC-hHR23 complicated is the primary DNA damage key factor in GGR. Another GGR DNA damage binding factor (DDB) [32,33] is a DNA damage sensor. The TFIIHp62 subunit

interacts with the damage recognition complex XPC-HR23B in GGR to transport it to the damaged area [34,35]. In terms of liberating damaged DNA, XPB seems to have less helicase activity than XPD. GGR and TCR need this 36 kDa protein [36,37]. RPA, ERCC1, and TFIIH are all known to interact with damaged DNA. RPA and ERCC1 investigations have shown that XPA preferentially binds to damaged DNA. The TFIIH helicase subunits' generated ssDNA intermediate is stabilized by this protein's ssDNA binding activity [38]. Following DNA synthesis, a twin insertion occurs on the damaged surface because of the sequential recruitment of the XPG and XPF-ERCC1 nuclear. XPG and XPF-ERCC1 are structurally specific nucleases that prefer to hydrolyze double-stranded substrates near ssDNA and sDNA junctions [39], ensuring the proper localization of these proteins to the site of injury and stimulating their junction-breaking endonuclease activity [40]. The DNA substrate bladder, valves, arms, and stem loops are only a few of the DNA substrates that XPG, a 133 kDa protein, affects [41]. The XPG protein possesses two highly conserved nucleic acid motifs spaced apart by a region that aids protein interactions. A protein of 37 kDa in molecular weight, PCNA belongs to the DNA sliding clamp family. In an ATP-dependent process, RFC assembles PCNA on the DNA template by ideally aligning with the 3'-hydroxyl ends of the DNA primer. Polymerase interactions with PCNA and RFC make it possible to synthesize DNA accurately and effectively [42]. POL  $\eta$  and POL  $\iota$ , in the polymerase  $\beta$  family, exhibit intrinsic exonuclease activity (3'-5') to correct for reading. The four subunits of mammals form the POL complex are 50, 12, 68, 125, and 68 kDa [43].

### 2.3. MMR Mechanism

The MMR (mismatch excision repair) process is employed when a mismatch develops between bases, such as the T-C and A-G pairs. This is accomplished by removing a strand, which is digested and replaced. DNA mismatch repair (MMR) is the main post-replicative DNA repair mechanism that can increase replication fidelity by 1000 ss [44,45]. Cells exposed to external chemicals and physical agents over time develop DNA damage. Multiple ways exist inside cells to repair DNA damage and stop mutations [44,46]. The DNA resynthesis and MMR initiation processes are thought to involve PCNA [47]. Localizing MutS and MutS to mismatches in freshly duplicated DNA may be made easier by PCNA. Both 5' and 3' directed MMR involve the 5'-3' exonuclease EXO1. High flexibility group box 1 protein (HMGB1 (High Mobility Group Box 1)), RPA, RFC, DNA pol  $\delta$ , and HMGB1 are other proteins connected to MMR [48]. RPA participates in the entire process of MMR since it attaches to crushed heteroduplex DNA before MutS and MutL, promotes mismatch-provoked excision, guards the ssDNA-gapped region generated after excision, and facilitates DNA synthesis. Additionally, MMR proteins have been associated with homoeologous recombination, immunoglobulin elegance switching, hypermutation, interstrand-crosslink restoration, and trinucleotide repeat (TNR) expansion [49,50]. The MMR employs double-strand DNA breaks produced using uracil DNA glycosylase to restore the AID-triggered G-U mismatches in a strand-indiscriminate manner [51].

## 3. DNA-Related Database

When studying DNA, it is crucial to have databases and datasets to learn the correlation of human diseases with gene mutations relevant to DNA integrity, stability, and information about DNA damage caused by mutagenic agents. Except for some databases that are no longer, here are some valuable databases and datasets in the DNA repair area. Among all the available databases, REPAIRtoire, Reactome, and the KEGG are the most commonly used databases. Table 1 shows the examples of commonly used databases.

**Table 1.** Examples of DNA repair-related databases.

Database	Size	Feature	Function	Last Update
REPAIRtoire	Proteins and DNA damage Diseases: 429	Multi-organism support; gene search	Identify and analyze pathways involved in DNA repair	October 2010

**Table 1.** *Cont.*

Database	Size	Feature	Function	Last Update
Human DNA Repair Genes	Genes: 256	Categorized DNA repair datasets	Related gene activity and chromosome location	June 2020
Reactome	Curated human protein: 11,350	Visualize biological processes online	Visualization, interpretation, and analysis	November 2022
DNArepairK	Proteins: 72	Animation of DNA repair protein kinetics	Dynamics of DNA repair proteins at the sites of DNA lesions	Unknown
KEGG	45, 822, 810	Large-scale integrated database	Molecular networks and network variants	November 2020
Brenda	Enzymes: 8423	Collection of gene data and enzymes	Searching enzymes	January 2023
Pathway commons	5772 Pathways; 2.3 million interaction data	Multiple databases to collect data	Data downloads, BioPAX web services, and data visualization	January 2020

### 3.1. REPAIRtoire

REPAIRtoire is a database of repair pathways for protein-coding genes. It provides a comprehensive and curated collection of genetic and epigenetic events that lead to the restoration of normal gene function in human cells. The database includes information on various types of repair mechanisms, such as DNA repair, RNA repair, and protein repair, and it can be used to aid in understanding disease mechanisms and developing new therapeutic strategies [52]. Researchers can search data through the following five sections in this database: (i) proteins: by searching the protein name, you can find the alternative names of the protein, the species of the protein, repair activities, the families of the proteins, and its related diseases; (ii) damage: by searching the name of the DNA damage, you can find the sources of the DNA damage and its effects, and it recognizes proteins; (iii) disease: you can find the related proteins with the disease name; (iv) pathways: the pathways section allows you to access to data through eight pathways from three species (*homo sapiens*, *saccharomyces cerevisiae*, and *escherichia coli*); and (v) publications: this section gives you the literature references to entries in the PubMed database. This database lets you quickly search data by entering protein sequences, profile searches, and browsing keywords to find the protein. The links button will give you access to other DNA repair-related databases (REACTOME, KEGG, etc.). REPAIRtoire can be accessed at <https://repairtoire.genesilico.pl/> (accessed on 3 January 2023).

### 3.2. Human DNA Repair Genes

Human DNA Repair Genes is a database of a table of human genes. The structure of this database is a table with columns of “Gene Name”, “Activity”, “Chromosome location”, and “Accession number” [53]. This table also has clearly distinguished different sections; it categorized different components and processes involved in DNA repair mechanisms, such as base excision repair (BER) and strand break joining factors, poly (ADP-ribose) polymerase (PARP) enzymes, the section of mismatch excision repair (MMR), chromatin structure, etc. The Human DNA Repair Genes database can be accessed at <https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html> (accessed on 3 January 2023).

### 3.3. Reactome

Like an online laboratory, Reactome is a user-friendly database of huge human pathways and sub-pathways. The main functions of this database are the pathway browser, analysis tools, Reactome FIViz, and documentation. By searching the name, ID, or the location of the gene in the search engine, this database will show you an overview of the topic of DNA repair; after clicking on the event, it will guide you through the pathway browser and show you a mind map, where you can find the description,

molecules, structures, and analysis [54]. Here is the link to the DNA repair section: <https://reactome.org/content/detail/R-DRE-73894> (accessed on 3 January 2023).

#### 3.4. DNAREPAIRK

A database called DNAREPAIRK tracks the kinetics of 70 fluorescently titled DNA repair proteins' recruitment and clearance from complicated DNA damage sites in vivo in HeLa Kyoto cells. It offers some simple analyses of the dynamics of proteins involved in different DNA repair processes using an interactive graph complemented with live cell imaging movie facilities. Most DNA repair proteins are represented by their kinetics in cells that have not been treated and cells that have been treated with the PARP1/2 inhibitor. This gives an unprecedented overview of how anti-cancer medications affect the regular dynamics of the DNA damage response. Scientists may investigate the DNA damage response using the unique dataset in DNA repair, which will also help them develop and test new anti-cancer medications that target DNA repair [55] DNAREPAIRK can be accessed at <http://dnarepair.bas.bg/index.php/dnarepairk/> (accessed on 3 January 2023).

#### 3.5. KEGG

The KEGG is a comprehensive database for computer representation of biological systems from cell to organism and the ecosystem; the information is from the genomic to molecular level. This integrated database has been categorized by different information into 16 databases. The molecular networks include the interaction between molecular, reaction, and relation networks featured in the KEGG. The infrastructure of this database is also focused on keeping different organisms' genes, genomes, and their variations. The database also includes additional types of generalization, such as reaction classes and drug groups [56]. The KEGG can be accessed at <https://www.genome.jp/kegg/> (accessed on 3 January 2023).

#### 3.6. Brenda

Brenda is a crucial database for primary enzyme functional data collection. The creator of this database extracted data from the primary literature by scientists. The enzymes are categorized by the Enzyme Commission's list of enzymes. To sort enzyme functional data, you can search enzymes by their EC number, enzyme name, and protein. Some common enzymes with very different properties will share the same EC number, and for more detailed information, you will have to go to the primary literature [57]. This database can be accessed at <https://www.brenda-enzymes.info/oldstart.php> (accessed on 3 January 2023).

#### 3.7. Pathway Commons

Pathway Commons is a public database of public pathway data from several organisms. It contains detailed data on biochemical processes, transport, catalysis activities, complex assembly, and physical interactions involving DNA, complexes, proteins, small molecules, and RNA. This meta-database compiles data from additional databases, like Reactome and Bio Grid. Users have access to a variety of accessible public pathway databases where they may browse and search for routes. Download a comprehensive set of Bio PAX format pathways for a more in-depth study. Additionally, it offers programmers a method to design software for complex investigations [58]. This database can be accessed at <http://www.pathwaycommons.org/> (accessed on 3 January 2023).

Like DNA databases, DNA-related databases also have ethical issues. Since the DNA repair databases are publicly accessible, it benefits law enforcement for forensic evidence [59]. Still, the problems of related privacy and human rights may arise at the same time. We should restrict the use of DNA databases for research, investigation, and study purposes, and we need to ask for the data contributor's acknowledgment to conduct further studies beyond restrictions.

However, most of the above-mentioned databases may contain outdated information. The reasons include that few people are working on building a database, and the data are

not easy to collect and analyze. As we noticed in Table 1, the largest amount of data in the KEGG is 45,822,810 because the KEGG is not only a DNA repair-related database but also a vast comprehensive DNA database, as mentioned in Section 3.5. However, some databases (DNArepairK, Human DNA Repair Genes) in Table 1 only contain a few hundred data. As an alternative plan, if a DNA repair database in need is not comprehensive, researchers can refer to other DNA databases; for example, the NCBI GenBank, which is maintained by the National Center for Biotechnology Information, and researchers worldwide use GenBank for a wide range of studies [60].

It is also worth drawing our attention to future database developments. It will be convenient if the database can be frequently updated and provide an easier way to analyze new data. With the help of artificial intelligence, we may have a clever way to automatically collect, analyze data, and manage data in the future.

#### 4. DNA Repair Computational Research Methods

Experimental methods tend to be more reliable, time-consuming, and complex to implement. Computational methods offer more cost-effective and efficient ways to explore various scenarios. For example, AlphaFold is a revolutionary advance in the history of protein research that for the first time, offers the practical ability to accurately predict the three-dimensional structure of a protein using amino acid sequences as inputs [61]. While computational methods may not be able to replace laboratory experiments fully, they can help identify and prioritize a selected group of promising candidates from large datasets. Computational studies in biomolecular interactions have proven effective in different topics [59–62]. This article will introduce three categories of *in silico* methods based on protein structure, interactions, and evolution.

##### 4.1. Protein Structure Analysis

Early efforts to investigate DNA repair enzymes focused on utilizing information from protein structures determined by experimental research. For example, Wang and Moulton analyzed missense mutations that cause disease and developed rules based on protein structure stability that could predict the effects of these mutations on molecular function [62]. These rules included the loss of interaction pairs, including hydrogen bonds and salt bridges, the basis of a buried polar residue, the proline insertion into an alpha helix, and the breakage of a disulfide bond. Other researchers also used similar rules [63,64].

Despite significant experimental efforts, the structures of only about 100,000 unique proteins have been resolved, which is a small portion of the known human protein sequences [65]. AlphaFold2 is a tool that has demonstrated success in predicting protein structures. In the 14th Critical Assessment of Protein Structure Prediction (CASP14), a blind test, AlphaFold2, outperformed other prediction methods [66]. While the new prediction algorithm does not fully explain the relationship between a protein's three-dimensional structure and its sequence, it can accurately predict the structure from the sequence in many cases, making it a practical solution to the protein folding problem [67]. The "SWISS-MODEL" server offers a range of options for constructing homology models, including fully automatic construction through its web interface [68]. In addition to model construction, the server also helps users locate suitable templates and alignments. It is especially useful for modeling proteins highly related to experimentally determined structures, as these relationships can be identified using tools, like "BLAST" [69]. Later, BLAST evolved to BLAST+, and the latest version is "BLAST+ 2.13.0" with advanced features that have made protein structure analysis easier.

##### 4.2. Molecular Dynamics Simulations

Molecular dynamics (MD) has become essential in studying DNA repair, allowing for detailed structural and dynamic insights [70]. CHARMM (Chemistry at Harvard Molecular Mechanics), AMBER (Assisted Model Building with Energy Refinement), NAMD (Nanoscale Molecular Dynamics), and GROMACS (Groningen Machine for Chemical

Simulations) are well-known software tools that can be used for molecular modeling to study DNA repair and its relationship to proteins and diseases. Many researchers have predicted analyses of protein-DNA and protein-protein contacts by performing molecular dynamics (MD) simulations to support their studies [71–74], including but not limited to:

- Performing molecular dynamics simulations to study the dynamics of proteins involved in DNA repair processes, such as simulating the movement and interactions of DNA repair enzymes;
- Studying the effects of genetic variations on the structures and functions of DNA repair proteins, such as simulating the impact of SNPs on the structure and function of DNA repair enzymes;
- Predicting the binding process between small molecules and proteins involved in DNA repair to identify potential drug candidates for treating DNA repair-related diseases;
- Identifying potential drug candidates for treating and studying the interactions between proteins involved in DNA repair processes, such as simulating the interactions between DNA repair enzymes and DNA damage response proteins.

#### 4.3. Evolutionary Analysis

During human population screenings, it was predicted that many amino acid substitution variants found in genes related to DNA repair could reduce protein function and activity [75], potentially leading to reduced repair capacity and an increased risk factor of cancer due to elevated genetic susceptibility. Using evolutionary structure-based methods and sequences makes it possible to differentiate variants depending on a score. Tools such as SIFT, PolyPhen 2.0, and SNAP are commonly used. SIFT predicts the effect of variants as neutral or deleterious using a normalized probability score based on sequence homology. It assumes that important amino acids will be reserved in the protein family, so various changes at well-reserved positions are often predicted to be harmful [76]. PolyPhen 2.0 combines sequence and structure-based attributes and utilizes a naive Bayesian classifier to identify the effect of an amino acid substitution [77]. SNAP is a neural network-based tool that accurately predicts the functional effects of nonsynonymous single nucleotide polymorphisms (nsSNPs) by combining evolutionary information (such as residue conservation within sequence families), predicted aspects of protein structure (such as secondary structure and solvent accessibility), and other pertinent data [78]. These algorithms correctly identified approximately 80% of amino acid substitutions that were assumed to significantly decrease the activity of the variant protein, as tested in the set [78,79].

For ethical considerations, as researchers, we should maintain the integrity and accuracy of data and show respect for the dignity, rights, and privacy of research participants. To avoid risks like privacy leaks, shortly research, we will mainly focus on using a protein data bank, open-source data, for training purposes. We will also implement robust methods to avoid such risks.

Computational methods have their own limitations. They may simplify the representation of biological systems [80]. For example, computational resources often limit molecular dynamics (MD) simulations, restricting the timescales and length scales that can be realistically simulated [81]. Many biological processes, such as protein folding or large-scale conformational changes, occur on timescales beyond the reach of current MD simulations [82–86]. In the protein structural prediction area, the accuracy of computational models heavily relies on the quality and quantity of the data used for training [86–88]. Biological systems are dynamic, with molecules constantly moving and interacting. Computational models, including AlphaFold, may not accurately capture these temporal dynamics. Computational models are constrained by existing knowledge and databases. Therefore, if a biological mechanism is poorly understood or novel, the model may struggle to predict the structures or functions associated with that mechanism accurately. Overfitting occurs when a model becomes too specific to the training data, performing poorly on new, unseen data, especially when the training data are limited.

Here are some future developments that may be helpful. We can enhance models to capture the temporal dynamics of biological systems better. Also, future models should be more adaptable to unknown or poorly understood biological mechanisms. We also need to make efforts to enhance data quality and quantity. The future models should be more resource-efficient and enhance generalization across diverse biological systems.

## 5. Summary

DNA damage and repair mechanisms have been studied for decades and will still be considered an important topic in future research related to diseases, aging, cancers, etc. The mutations in DNA repair genes can influence and regulate individual cancer susceptibility, and polymorphism screening has recently become a research area in molecular epidemiology with high potential. Targeted gene therapy is one of the possible methods since it can selectively repair drug sensitivity in cancer cells with drug sensitivity abnormalities. Understanding these mechanisms will aid in the development of new therapeutic approaches for patients with defective tumors, as well as in the choice of treatment and prognosis for ovarian and breast cancer patients. Several forms of cancer are affected by both intrinsic and acquired resistance mechanisms, which is a significant area for the development of new medications. In our review article, we summarized DNA repair enzymes that are responsible for fixing DNA damage, DNA repair databases that are helpful for researchers to study DNA damage and repair mechanisms using data science and data analysis techniques, and DNA repair computational research methods that are preferable for scientists to perform research in different ways. By conducting research in environment chemical research, toxicological genotoxic drugs can lead to a reduction in DNA damage. The numerous clinical trials that evaluate the possibility of making tumors more responsive to chemotherapy by inhibiting RSR signaling may provide directions for future efforts to create tractable drugs that specifically target tumor cells.

**Author Contributions:** Conceptualization, Y.X. (Ying Xie) and R.P.; resources, H.Q. and Y.X. (Yixin Xie); data curation, L.L. and J.C.; writing—original draft preparation, R.P., H.Q. and J.C.; writing—review and editing, L.L., Y.X. (Yixin Xie), S.P., L.C. and N.S.; visualization, R.P. and Y.X. (Yixin Xie); supervision, S.P., N.S., Y.X. (Ying Xie) and Y.X. (Yixin Xie). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by OVPR Research Seed Grants from Kennesaw State University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** REPAIRtoire can be accessed at <https://repairtoire.genesilico.pl/> (accessed on 3 January 2023). The Human DNA Repair Genes can be accessed at <https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html> (accessed on 3 January 2023). DNAREPAIRK can be accessed at <http://dnarepair.bas.bg/index.php/dnarepairk/> (accessed on 3 January 2023). KEGG can be accessed at <https://www.genome.jp/kegg/> (accessed on 3 January 2023). The Brenda database can be accessed at <https://www.brenda-enzymes.info/oldstart.php> (accessed on 3 January 2023). The pathway database can be accessed at <http://www.pathwaycommons.org/> (accessed on 3 January 2023).

**Acknowledgments:** The authors would like to sincerely thank Wen Xin, in the English Department at the University of Kansas, for his invaluable assistance in improving the writing quality of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the study's design, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

## References

1. Chen, Y.; Yang, Z.; Zhou, X.; Jin, M.; Dai, Z.; Ming, D.; Zhang, Z.; Zhu, L.; Jiang, L. Sequence, structure, and function of the Dps DNA-binding protein from *Deinococcus wulumuqiensis* R12. *Microb. Cell Fact.* **2022**, *21*, 132. [[CrossRef](#)]

2. Lin, L.; Zhang, Y.; Wei, Q.; Lin, H.; Li, X.; Yu, M.E.; Wang, J.; Huang, Z.; Xue, D. Structure and function encoding of a bidirectional activatable synergetic DNA machine for speeded and ultrasensitive determination of microRNAs. *Talanta* **2022**, *238 Pt 1*, 123037. [[CrossRef](#)]
3. Trasvina-Arenas, C.H.; Demir, M.; Lin, W.J.; David, S.S. Structure, function and evolution of the Helix-hairpin-Helix DNA glycosylase superfamily: Piecing together the evolutionary puzzle of DNA base damage repair mechanisms. *DNA Repair* **2021**, *108*, 103231. [[CrossRef](#)]
4. Said, H.; Kaji, I.; Kaunitz, J.D. Gastroduodenal mucosal defense mechanisms. *Curr. Opin. Gastroenterol.* **2015**, *31*, 486–491. [[CrossRef](#)]
5. Li, Y.; Tang, W.; Guo, M. The Cell as Matter: Connecting Molecular Biology to Cellular Functions. *Matter* **2021**, *4*, 1863–1891. [[CrossRef](#)]
6. Jackson, S.P.; Bartek, J. The DNA-damage response in human biology and disease. *Nature* **2009**, *461*, 1071–1078. [[CrossRef](#)]
7. Hoeijmakers, J.H. DNA damage, aging, and cancer. *N. Engl. J. Med.* **2009**, *361*, 1475–1485. [[CrossRef](#)]
8. Bian, K.; Lenz, S.A.P.; Tang, Q.; Chen, F.; Qi, R.; Jost, M.; Drennan, C.L.; Essigmann, J.M.; Wetmore, S.D.; Li, D. DNA repair enzymes ALKBH2, ALKBH3, and AlkB oxidize 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine in vitro. *Nucleic Acids Res.* **2019**, *47*, 5522–5529. [[CrossRef](#)]
9. Chatterjee, N.; Walker, G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **2017**, *58*, 235–263. [[CrossRef](#)]
10. Andrew, S.E.; Peters, A.C. DNA instability and human disease. *Am. J. Pharmacogenom.* **2001**, *1*, 21–28. [[CrossRef](#)]
11. Krejci, L.; Altmannova, V.; Spirek, M.; Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Res.* **2012**, *40*, 5795–5818. [[CrossRef](#)]
12. Rass, U.; Ahel, I.; West, S.C. Defective DNA repair and neurodegenerative disease. *Cell* **2007**, *130*, 991–1004. [[CrossRef](#)]
13. Mazurek, A.; Berardini, M.; Fishel, R. Activation of human MutS homologs by 8-oxo-guanine DNA damage. *J. Biol. Chem.* **2002**, *277*, 8260–8266. [[CrossRef](#)]
14. Best, B.P. Nuclear DNA damage as a direct cause of aging. *Rejuvenation Res.* **2009**, *12*, 199–208. [[CrossRef](#)]
15. Giacco, F.; Brownlee, M. Oxidative stress and diabetic complications. *Circ. Res.* **2010**, *107*, 1058–1070. [[CrossRef](#)]
16. Liou, G.Y.; Storz, P. Reactive oxygen species in cancer. *Free Radic. Res.* **2010**, *44*, 479–496. [[CrossRef](#)]
17. Meyenberg, M.; da Silva, J.F.; Loizou, J.I. Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing. *Front. Genet.* **2021**, *12*, 728520. [[CrossRef](#)]
18. Sudeshna, N.; Amy, B.; Andris, Z.; Peteris, R.; Karlis, P.; Martins, O.; Edgars, C.; Juris, V.; Maria, K.; McCarthy, M.I.; et al. PASSIM—An open source software system for managing information in biomedical studies. *BMC Bioinform.* **2007**, *8*, 52.
19. Grimes, S.M.; Ji, H.P. MendELIMS: A web-based laboratory information management system for clinical genome sequencing. *BMC Bioinform.* **2014**, *15*, 290. [[CrossRef](#)]
20. Alexandrov, N.; Alexandrov, V. Computational Science Research Methods for Science Education at PG Level. *Procedia Comput. Sci.* **2015**, *51*, 1685–1693. [[CrossRef](#)]
21. Milanowska, K.; Rother, K.; Bujnicki, J.M. Databases and bioinformatics tools for the study of DNA repair. *Mol. Biol. Int.* **2011**, *2011*, 475718. [[CrossRef](#)] [[PubMed](#)]
22. Krwawicz, J.; Arczewska, K.D.; Speina, E.; Maciejewska, A.; Grzesiuk, E. Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and DNA-end processors and their implication in mutagenesis and human disease. *Acta Biochim. Pol.* **2007**, *54*, 413–434. [[CrossRef](#)] [[PubMed](#)]
23. Klungland, A.; Lindahl, T. Second pathway for completion of human DNA base excision-repair: Reconstitution with purified proteins and requirement for DNase IV (FEN1). *EMBO J.* **1997**, *16*, 3341–3348. [[CrossRef](#)]
24. Zentout, S.; Smith, R.; Jacquier, M.; Huet, S. New Methodologies to Study DNA Repair Processes in Space and Time Within Living Cells. *Front. Cell Dev. Biol.* **2021**, *9*, 730998. [[CrossRef](#)] [[PubMed](#)]
25. Egly, J.M.; Coin, F. A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor. *DNA Repair* **2011**, *10*, 714–721. [[CrossRef](#)] [[PubMed](#)]
26. Lehmann, A.R. The molecular biology of nucleotide excision repair and double-strand break repair in eukaryotes. *Genet. Eng.* **1995**, *17*, 1–19.
27. Wood, R.D.; Shivji, M.K. Which DNA polymerases are used for DNA-repair in eukaryotes? *Carcinogenesis* **1997**, *18*, 605–610. [[CrossRef](#)] [[PubMed](#)]
28. Leibel, D.; Laspe, P.; Emmert, S. Nucleotide excision repair and cancer. *J. Mol. Histol.* **2006**, *37*, 225–238. [[CrossRef](#)]
29. Hanawalt, P.C. Subpathways of nucleotide excision repair and their regulation. *Oncogene* **2002**, *21*, 8949–8956. [[CrossRef](#)]
30. Gillet, L.C.; Scharer, O.D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.* **2006**, *106*, 253–276. [[CrossRef](#)]
31. Saxowsky, T.T.; Doetsch, P.W. RNA polymerase encounters with DNA damage: Transcription-coupled repair or transcriptional mutagenesis? *Chem. Rev.* **2006**, *106*, 474–488. [[CrossRef](#)] [[PubMed](#)]
32. Gaul, L.; Sveistrup, J.Q. Transcription-coupled repair and the transcriptional response to UV-Irradiation. *DNA Repair* **2021**, *107*, 103208. [[CrossRef](#)] [[PubMed](#)]
33. Kuper, J.; Kisker, C. Damage recognition in nucleotide excision DNA repair. *Curr. Opin. Struct. Biol.* **2012**, *22*, 88–93. [[CrossRef](#)] [[PubMed](#)]

34. Sugasawa, K. The xeroderma pigmentosum group C protein complex and ultraviolet-damaged DNA-binding protein: Functional assays for damage recognition factors involved in global genome repair. *Methods Enzymol.* **2006**, *408*, 171–188.
35. Tang, J.Y.; Hwang, B.J.; Ford, J.M.; Hanawalt, P.C.; Chu, G. Xeroderma pigmentosum p48 gene enhances global genomic repair and suppresses UV-induced mutagenesis. *Mol. Cell* **2000**, *5*, 737–744. [[CrossRef](#)] [[PubMed](#)]
36. van den Heuvel, D.; van der Weegen, Y.; Boer, D.E.C.; Ogi, T.; Luijsterburg, M.S. Transcription-Coupled DNA Repair: From Mechanism to Human Disorder. *Trends Cell Biol.* **2021**, *31*, 359–371. [[CrossRef](#)]
37. Petrusseva, I.O.; Evdokimov, A.N.; Lavrik, O.I. Molecular mechanism of global genome nucleotide excision repair. *Acta Naturae* **2014**, *6*, 23–34. [[CrossRef](#)] [[PubMed](#)]
38. de Laat, W.L.; Appeldoorn, E.; Sugasawa, K.; Weterings, E.; Jaspers, N.G.; Hoeijmakers, J.H. DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes. Dev.* **1998**, *12*, 2598–2609. [[CrossRef](#)]
39. Sijbers, A.M.; de Laat, W.L.; Ariza, R.R.; Biggerstaff, M.; Wei, Y.F.; Moggs, J.G. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **1996**, *86*, 811–822. [[CrossRef](#)]
40. Matsunaga, T.; Park, C.H.; Bessho, T.; Mu, D.; Sancar, A. Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. *J. Biol. Chem.* **1996**, *271*, 11047–11050. [[CrossRef](#)]
41. Scherly, D.; Nospikel, T.; Corlet, J.; Ucla, C.; Bairoch, A.; Clarkson, S.G. Complementation of the DNA repair defect in xeroderma pigmentosum group G cells by a human cDNA related to yeast RAD2. *Nature* **1993**, *363*, 182–185. [[CrossRef](#)] [[PubMed](#)]
42. Kang, M.S.; Ryu, E.; Lee, S.W.; Park, J.; Ha, N.Y.; Ra, J.S. Regulation of PCNA cycling on replicating DNA by RFC and RFC-like complexes. *Nat. Commun.* **2019**, *10*, 2420. [[CrossRef](#)] [[PubMed](#)]
43. Liu, L.; Mo, J.; Rodriguez-Belmonte, E.M.; Lee, M.Y. Identification of a fourth subunit of mammalian DNA polymerase delta. *J. Biol. Chem.* **2000**, *275*, 18739–18744. [[CrossRef](#)]
44. Mennequier, S.; Coste, G.; Servant, P.; Bailone, A.; Sommer, S. Mismatch repair ensures fidelity of replication and recombination in the radioresistant organism *Deinococcus radiodurans*. *Mol. Genet. Genom.* **2004**, *272*, 460–469. [[CrossRef](#)] [[PubMed](#)]
45. Schofield, M.J.; Hsieh, P. DNA mismatch repair: Molecular mechanisms and biological function. *Annu. Rev. Microbiol.* **2003**, *57*, 579–608. [[CrossRef](#)]
46. Kunkel, T.A.; Erie, D.A. DNA mismatch repair. *Annu. Rev. Biochem.* **2005**, *74*, 681–710. [[CrossRef](#)] [[PubMed](#)]
47. Kadyrov, F.A.; Dzantiev, L.; Constantin, N.; Modrich, P. Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* **2006**, *126*, 297–308. [[CrossRef](#)]
48. Ramilo, C.; Gu, L.; Guo, S.; Zhang, X.; Patrick, S.M.; Turchi, J.J.; Li, G.M. Partial reconstitution of human DNA mismatch repair in vitro: Characterization of the role of human replication protein A. *Mol. Cell Biol.* **2002**, *22*, 2037–2046. [[CrossRef](#)]
49. Kolodner, R.D.; Putnam, C.D.; Myung, K. Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* **2002**, *297*, 552–557. [[CrossRef](#)]
50. Vogelstein, B.; Kinzler, K.W. Cancer genes and the pathways they control. *Nat. Med.* **2004**, *10*, 789–799. [[CrossRef](#)]
51. Schrader, C.E.; Guikema, J.E.; Linehan, E.K.; Selsing, E.; Stavnezer, J. Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *J. Immunol.* **2007**, *179*, 6064–6071. [[CrossRef](#)] [[PubMed](#)]
52. Milanowska, K.; Krwawicz, J.; Papaj, G.; Kosinski, J.; Poleszak, K.; Lesiak, J. REPAIRtoire—A database of DNA repair pathways. *Nucleic Acids Res.* **2011**, *39*, D788–D792. [[CrossRef](#)] [[PubMed](#)]
53. Wood, R.D.; Mitchell, M.; Sgouros, J.; Lindahl, T. Human DNA repair genes. *Science* **2001**, *291*, 1284–1289. [[CrossRef](#)] [[PubMed](#)]
54. Cutts, R.J.; Gadaleta, E.; Hahn, S.A.; Crnogorac-Jurcevic, T.; Lemoine, N.R.; Chelala, C. The Pancreatic Expression database: 2011 update. *Nucleic Acids Res.* **2011**, *39*, D1023–D1028. [[CrossRef](#)] [[PubMed](#)]
55. Babukov, Y.; Aleksandrov, R.; Ivanova, A.; Atemin, A.; Stoynov, S. DNAREPAIRK: An Interactive Database for Exploring the Impact of Anticancer Drugs onto the Dynamics of DNA Repair Proteins. *Biomedicines* **2021**, *9*, 1238. [[CrossRef](#)] [[PubMed](#)]
56. Li, J.; Wang, L.; Wan, J.; Dang, K.; Lin, Y.; Meng, S. Dynamic patterns of gene expression and regulatory variation in the maize seed coat. *BMC Plant Biol.* **2023**, *23*, 82. [[CrossRef](#)] [[PubMed](#)]
57. Schomburg, I.; Chang, A.; Hofmann, O.; Ebeling, C.; Ehrentreich, F.; Schomburg, D. BRENDA: A resource for enzyme data and metabolic information. *Trends Biochem. Sci.* **2002**, *27*, 54–56. [[CrossRef](#)]
58. Rodchenkov, I.; Babur, O.; Luna, A.; Aksoy, B.A.; Wong, J.V.; Fong, D. Pathway Commons 2019 Update: Integration, analysis and exploration of pathway data. *Nucleic Acids Res.* **2020**, *48*, D489–D497. [[CrossRef](#)]
59. Vollebergh, M.A.; Lips, E.H.; Nederlof, P.M.; Wessels, L.F.; Wesseling, J.; Vd Vijver, M.J. Genomic patterns resembling BRCA1- and BRCA2-mutated breast cancers predict benefit of intensified carboplatin-based chemotherapy. *Breast Cancer Res.* **2014**, *16*, R47. [[CrossRef](#)]
60. Xie, Y.; Du, D.; Karki, C.B.; Guo, W.; Lopez-Hernandez, A.E.; Sun, S. Revealing the mechanism of SARS-CoV-2 spike protein binding with ACE2. *Comput. Sci. Eng.* **2020**, *22*, 21–29. [[CrossRef](#)]
61. Xie, Y.; Karki, C.B.; Du, D.; Li, H.; Wang, J.; Sobitan, A. Spike Proteins of SARS-CoV and SARS-CoV-2 Utilize Different Mechanisms to Bind With Human ACE2. *Front. Mol. Biosci.* **2020**, *7*, 591873. [[CrossRef](#)] [[PubMed](#)]
62. Wang, Z.; Moulton, J. SNPs, protein structure, and disease. *Hum. Mutat.* **2001**, *17*, 263–270. [[CrossRef](#)] [[PubMed](#)]
63. Chasman, D.; Adams, R.M. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. *J. Mol. Biol.* **2001**, *307*, 683–706. [[CrossRef](#)] [[PubMed](#)]

64. Sunyaev, S.; Ramensky, V.; Koch, I.; Lathe, W.; Kondrashov, A.S.; Bork, P. Prediction of deleterious human alleles. *Hum. Mol. Genet.* **2001**, *10*, 591–597. [[CrossRef](#)]
65. wwPDB Consortium. Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **2019**, *47*, D520–D528. [[CrossRef](#)]
66. Pereira, J.; Simpkin, A.J.; Hartmann, M.D.; Rigden, D.J.; Keegan, R.M.; Lupas, A.N. High-accuracy protein structure prediction in CASP14. *Proteins* **2021**, *89*, 1687–1699. [[CrossRef](#)]
67. Cramer, P. AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* **2021**, *28*, 704–705. [[CrossRef](#)]
68. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195–201. [[CrossRef](#)]
69. Mount, D.W. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc.* **2007**, *2007*, pdb.top17. [[CrossRef](#)]
70. Perez, A.; Luque, F.J.; Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.* **2012**, *45*, 196–205. [[CrossRef](#)]
71. Sampoli, B.B.; Barbati, Z.R.; Arora, K.; Bogdanovic, J.; Schlick, T. How DNA polymerase X preferentially accommodates incoming dATP opposite 8-oxoguanine on the template. *Biophys. J.* **2013**, *105*, 2559–2568. [[CrossRef](#)] [[PubMed](#)]
72. Salsbury, F.R.; Clodfelter, J.E.; Gentry, M.B.; Hollis, T.; Scarpinato, K.D. The molecular mechanism of DNA damage recognition by MutS homologs and its consequences for cell death response. *Nucleic Acids Res.* **2006**, *34*, 2173–2185. [[CrossRef](#)]
73. Minko, I.G.; Vartanian, V.L.; Tozaki, N.N.; Coskun, E.; Coskun, S.H.; Jaruga, P. Recognition of DNA adducts by edited and unedited forms of DNA glycosylase NEIL1. *DNA Repair* **2020**, *85*, 102741. [[CrossRef](#)] [[PubMed](#)]
74. Xie, Y.; Karki, C.B.; Chen, J.; Liu, D.; Li, L. Computational Study on DNA Repair: The Roles of Electrostatic Interactions Between Uracil-DNA Glycosylase (UDG) and DNA. *Front. Mol. Biosci.* **2021**, *8*, 718587. [[CrossRef](#)] [[PubMed](#)]
75. Xi, T.; Jones, I.M.; Mohrenweiser, H.W. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. *Genomics* **2004**, *83*, 970–979. [[CrossRef](#)]
76. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [[CrossRef](#)] [[PubMed](#)]
77. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [[CrossRef](#)]
78. Bromberg, Y.; Rost, B. SNAP: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* **2007**, *35*, 3823–3835. [[CrossRef](#)]
79. Ng, P.C.; Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* **2002**, *12*, 436–446. [[CrossRef](#)]
80. Zadok, E. Legislative and Ethical Questions regarding DNA and Other Forensic “Biometric” Databases. In Proceedings of the International Conference on Ethics and Policy of Biometrics, Hong Kong, China, 3–4 January 2010; Springer: Berlin/Heidelberg, Germany, 2010.
81. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2012**, *41*, D36–D42. [[CrossRef](#)]
82. Rothman, J.E. Starting at Go: Protein structure prediction succumbs to machine learning. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2311128120. [[CrossRef](#)] [[PubMed](#)]
83. Tian, H.; Jiang, X.; Trozzi, F.; Xiao, S.; Larson, E.C.; Tao, P. Explore protein conformational space with variational autoencoder. *Front. Mol. Biosci.* **2021**, *8*, 781635. [[CrossRef](#)] [[PubMed](#)]
84. Shaw, D.E.; Dror, R.O.; Salmon, J.K.; Grossman, J.; Mackenzie, K.M.; Bank, J.A. Millisecond-scale molecular dynamics simulations on Anton. In Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, Portland, OR, USA, 13–19 November 2009.
85. Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659. [[CrossRef](#)] [[PubMed](#)]
86. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
87. Carugo, O.; Djinić-Carugo, K. Structural biology: A golden era. *PLoS Biol.* **2023**, *21*, e3002187. [[CrossRef](#)]
88. Huang, B.; Kong, L.; Wang, C.; Ju, F.; Zhang, Q.; Zhu, J. Protein Structure Prediction: Challenges, Advances, and the Shift of Research Paradigms. *Genom. Proteom. Bioinform.* **2023**, *in press*. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.