ISSN 2744-1636
Bioengineering Studies *Original Research*
Vol. 5, No. 1, July 2024, pp.15-28
https://doi.org/10.37868/bes.v5i1.id267

In silico functional analysis of DNA repair proteins (OGG1, MYH) in *Arabidopsis thaliana*

Annissa Rachel Van Wieren[1*], Mohamed Ragab Abdel Gawwad[2]
[1] Molecular Biology, Genetics and Bioengineering, Sabanci University, İstanbul, Turkey
[2] Genetics Bioengineering, International University of Sarajevo, Sarajevo, Bosnia and Herzegovina

*Corresponding author: annissa.vanwieren@sabanciuniv.edu

Received Jun. 30, 2024
Revised Jun. 15, 2024
Accepted Jun. 30, 2024

Abstract

Deoxyribonucleic acid (DNA) is a hereditary material found in every living organism. However, every organism is exposed to endogenous and exogenous an agent which causes damage to DNA molecules. Even though genetic variation is important for evolution, every organism requires genetic stability for survival. Not only it is needed for DNA to replicate accurately but also several DNA repair mechanisms are used to fix the damage which occurred in DNA. A detailed structure analysis along with docking site and metabolic pathway prediction were used to confirm that these proteins are involved in DNA repair mechanisms. This study investigated OGG1 and MYH proteins from *Arabidopsis thaliana*, and showed their interaction with predicted partners, focusing on their detailed interactome and 3D structure predictions. This study concluded that OGG1 and MYH are not only found in DNA repair mechanisms but also show strong functions in DNA replication and the development of plant and cell cycle.

Keywords: DNA, DNA repair mechanism, *Arabidopsis thaliana*, OGG1, MYH

1. Introduction

Deoxyribonucleic acid (DNA) is a type of nucleic acid which has the genetic instructions for an organism to develop and function properly. It is present in all living organisms and in some viruses [1]. DNA needs to be stable because it needs to store all genetic information for a long time, and it is passed down from parents to offspring. Mutation is a change in DNA sequence that happens randomly due to DNA replication errors and oxidative stress. Oxidative stress is when free radicals and antioxidants are not properly balanced in our body. Every living thing is exposed to exogenous and endogenous gents. This may include ionizing radiation, ultraviolet (UV) radiation, etc. Luckily, there are several types of DNA repair mechanisms that will correct the replication error which include: base excision repair (BER), mismatch repair (MMR), double-strand break repair (DSBR), and nucleotide excision repair (NER) [2] [3]. *Arabidopsis thaliana*, also named as mouse-ear cress, is a model plant which is used in this research, and it is a part of the Brassicaceae or mustard family [4]. It contains five pairs of chromosomes which it has in total ten chromosomes (2n = 10) [5]. In this paper, the focus will be on two proteins that are part of DNA glycosylases: 8-Oxoguanine-DNA Glycosylase 1 (OGG1) and mutY DNA glycosylase (MUTYH) or (MYH).

2. Material and methods

2.1. Retrieving sequences of *OGG1* and *MYH* proteins

The sequences of OGG1 and MYH proteins are retrieved from The Arabidopsis Information Resource (TAIR) and National Center for Biotechnology Information (NCBI) websites. TAIR has all the information about all genes present in the genome of *A. thaliana* [5]. NCBI is used to give information based on biomedicine and the genomes of the organisms [6].

2.2. Multiple Sequence Alignment

Multiple sequence alignment (MSA) provides three or more sequences that will be aligned. The sequences which were retrieved are now aligned with several homologous proteins. The homologous proteins from OGG1 and MYH are found using Basic Local Alignment Sequence tools (BLAST). BLAST tools are used to find similar regions between nucleotide sequence [7]. In this research, Clustal Omega is used and also conserved regions will be detected between these sequences.

2.3. Phylogenetic Tree Construction

The phylogenetic is constructed using Clustal Omega. This type of tree is used to show how these proteins evolved, and which proteins are more closely related and less related. The branching of phylogenetic trees presents how these proteins may have evolved from a common ancestor. It is possible that all organisms evolved from one common ancestor [8] [9].

2.4. 3D structure prediction, visualization and validation

The molecular Graphics System (PyMol) was used to predict and visualize the 3D structure of OGG1 and MYH [10]. The 3D structure of OGG1 and MYH were validated using PDBsum and PROCHECK software [11].

2.5. Domain search

The domains of OGG1 and MYH proteins were found using ExPASy - PROSITE. Using ExPASy - PROSITE tool, where more than 1000 domain proteins can be found. This website provides information, orthologs, and architecture about domains and how domains will interact with other proteins. Not only domains are predicted in proteins, but also their motifs are predicted [12].

2.6. Finding subcellular localization of protein

The subcellular localization of OGG1 and MYH proteins were found using the Subcellular localization database for Arabidopsis proteins 4 (SUBA4). The function of proteins is influenced by subcellular localization due to having all types of molecular interaction partners involved. There are specific mechanisms present inside our cells that ensure a needed protein is found in that certain region which is known as subcellular localization. With knowing the subcellular localization of the protein, the functions of recently discovered proteins can be characterized [13].

2.7. Protein Interactome Prediction

Protein Interactome represents a protein interacting with all other proteins. The word ‘interactome’ is used to describe the molecular interactions found in each cell [14]. To predict which protein does interact with other proteins, a bioinformatics tool, known as Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), was used [15].

2.8. Docking Site Prediction

Molecular docking refers to how a small molecule or ligand interacts with a protein at the atomic level. The behavior, binding affinity and conformation of ligands in the protein's binding site is determined [16]. To predict docking sites, ClusPro server will be used to visualize 3D interactome between OGG1/MYH and 10 proteins found by STRING [17].

2.9. Metabolic Pathway Prediction

Kyoto Encyclopedia of Genes and Genomes (KEGG) is used to determine the metabolic pathway of OGG1 and MYH as it provides information about details of biological systems and the building blocks. It studies the interaction and relations between OGG1/MYH and the genes or their products [18] [19].

3. Results

3.1. Retrieved sequences of *OGG1* and *MYH*

The sequences of OGG1 and MYH were retrieved from TAIR and NCBI webpage. The accession numbers and sequence length of OGG1 and MYH are shown in table 1.

Table 1: Accession numbers of OGG1 and MYH

Protein	TAIR ID	PBD code	Sequence length
OGG1	*AT1G21710*	1lwy	1486aa
MYH	*AT4G12740*	3n5n	1893aa

Cluster Omega tool was used to get MSA results which shows differences and similarity between 10 different proteins. In figure 1, MYH and OGG1 will be compared to other homologous proteins.

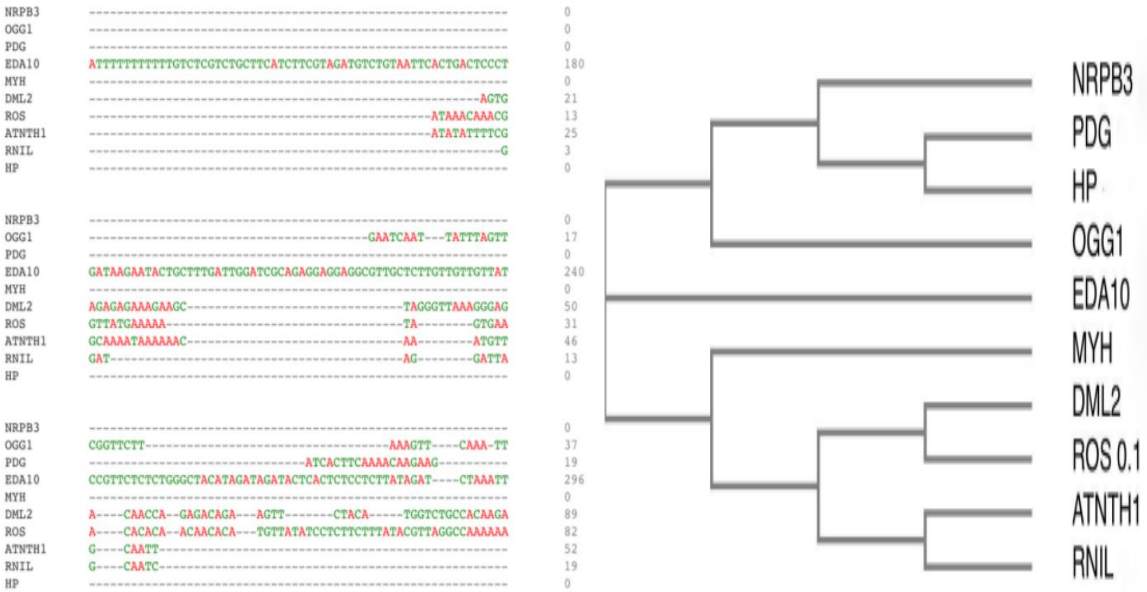

Figure 1: MSA and phylogenetic tree of 10 proteins

3D structure of OGG1 and MYH proteins are predicted, and visualized using PyMol and PROCHECK (figure 2 and 3). The scores of these proteins are validated by Ramachandran Plot assessment (RAMPAGE) and it is found that residues of OGG1 and MYH are in favored regions and they are both 90%.

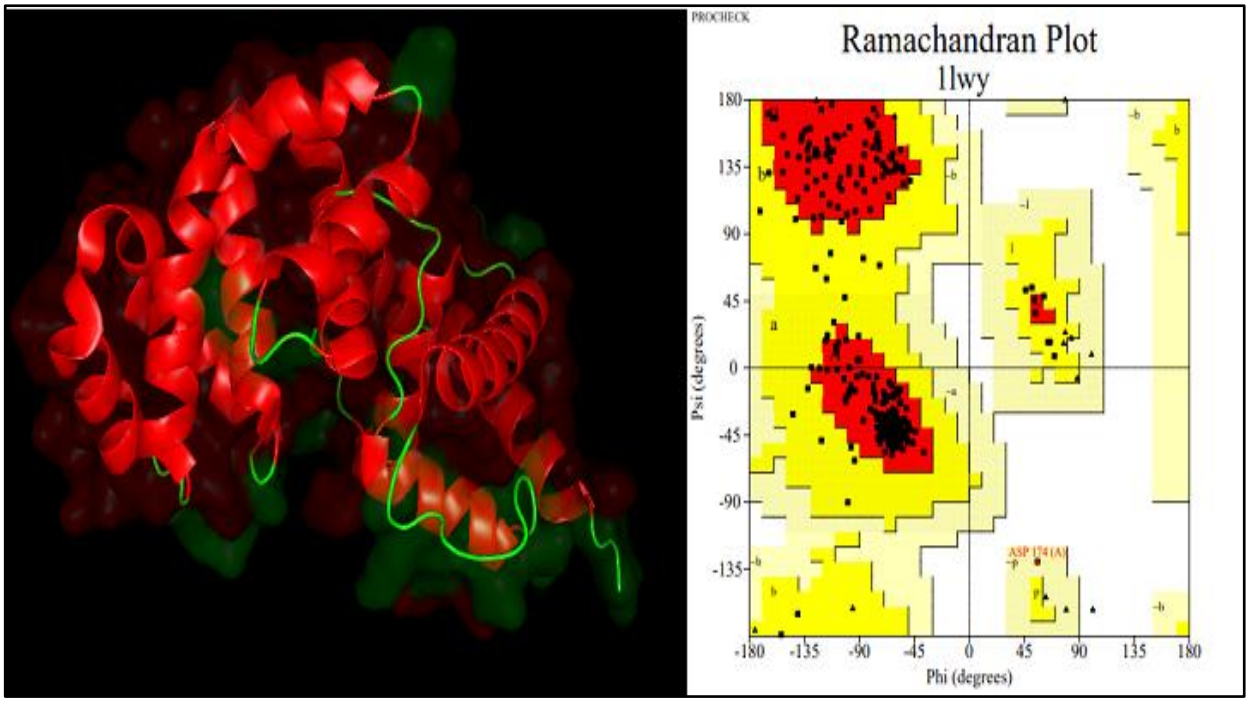

Figure 2: 3D structure of OGG1 on the left and Ramachandran plots on the right

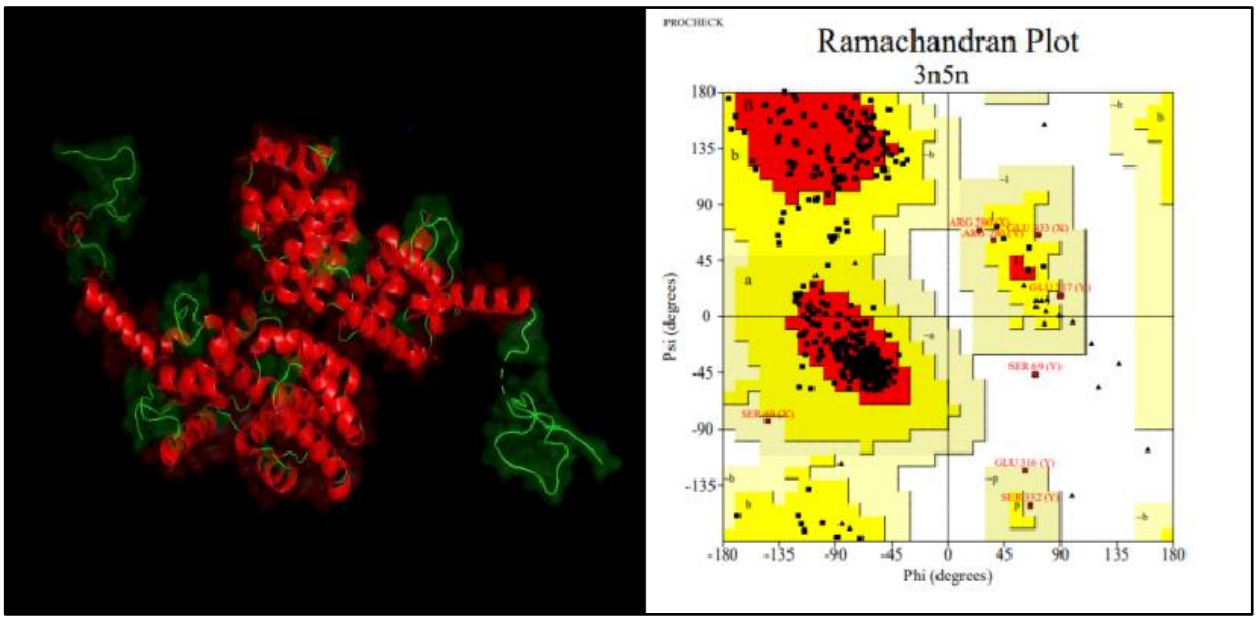

Figure 3: 3D structure of MYH on the left and Ramachandran plots on the right

Using ExPASy - PROSITE, it is found that OGG1 and MYH have in common four domains. But MYH has one more domain and it is called fungal hydrophobic domain (table 2).

Table 2: Positions of identified domains in OGG1 and MYH proteins

Domain	Protein	OGG1	MYH
Fungal hydrophobin domain	Start	/	815
	End		826
Araphylatoxin domain	Start	129	164
	End	1732	1765
VWFC domain	Start	139	950
	End	191	995
	Start	151	1013
	End	201	1086
	Start	1116	/
	End	1154	/
EGF-like domain 1	Start	256	807
	End	267	818
	Start	/	995
	End	/	1006
	Start	/	1044
	End	/	1055
2Fe-2S ferredoxin-type iron-sulfur domain	Start	529	5
	End	537	13
	Start	1230	1789
	End	1238	1797

Using SUBA4, OGG1 and MYH are located in plastid and with STRING software, 10 proteins are found to interact with OGG1 and MYH (figure 4).

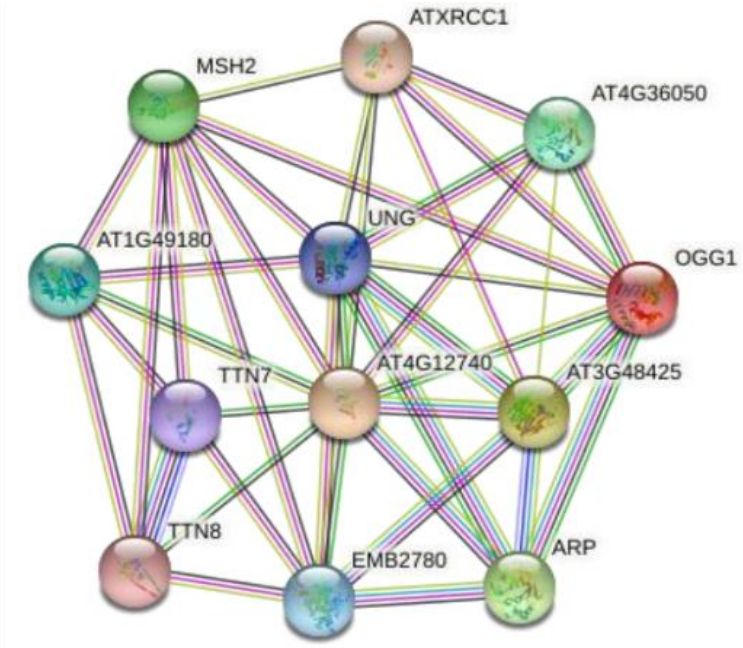

Figure 4: Interactome proteins of OGG1 and MYH

From table 3, we see that all interactome proteins are involved in various types of DNA replication and repair mechanisms.

Table 3: Protein Interactions of OGG1 and MYH proteins

Interactomes	Interactome name	Interactome function	Score
AT3G48425	DNA-AP lyase	It is part of DNA demethylation and imprints gene	0.994
ARP		It regulates transcription, part of BER in plastid and acts like a redox factor	0.992
MSH2	DNA MMR protein	It is involved in MMR after DNA replication	0.989
AT4G36050	Endonuclease/exonuclease /phosphatase family	It provides AP endonuclease activity *in vitro*	0.979
AT1G49180	Serine/threonine-protein kinase ATG1t	It is involved in enclosing autophagosome	0.947
EMB2780	DNA pol	It is synthesizing DNA and degrades 3'→5' ssDNA	0.946
UNG	Uracil-DNA glycosylase	It cleaves U when misincorporation during BER	0.937
TTN7	SMC family protein	It is involved in chromosome cohesion in which it's connected to DNA replication. This repairs DNA. It creates spindle pole during mitosis and mobilization of chromosomes	0.919
TTN8			
ATXRCCI	BRCT domain-containing DNA repair protein	It is involved DNA methylation and exchange of sister chromatid replacing damaged strand	0.918

3.8. Docking Site Prediction

Six proteins were used to provide docking site prediction. The results are obtained from ClusPro and represented in figure 5.

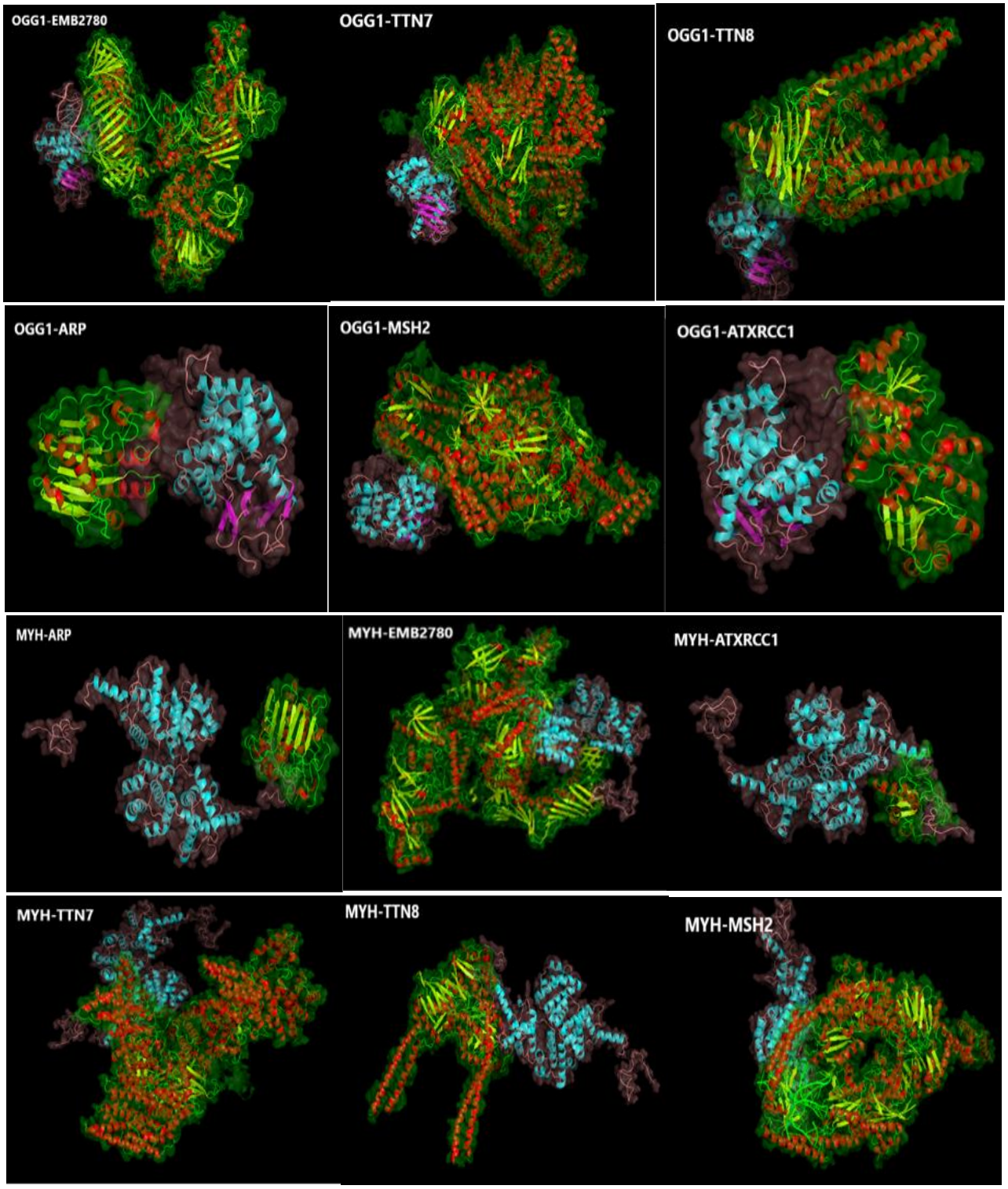

Figure 5: OGG1, *MYH* and interactome partners for docking site predictions

Using KEGG, it is found that MYH and OGG1 are mainly involved in BER (figure 6).

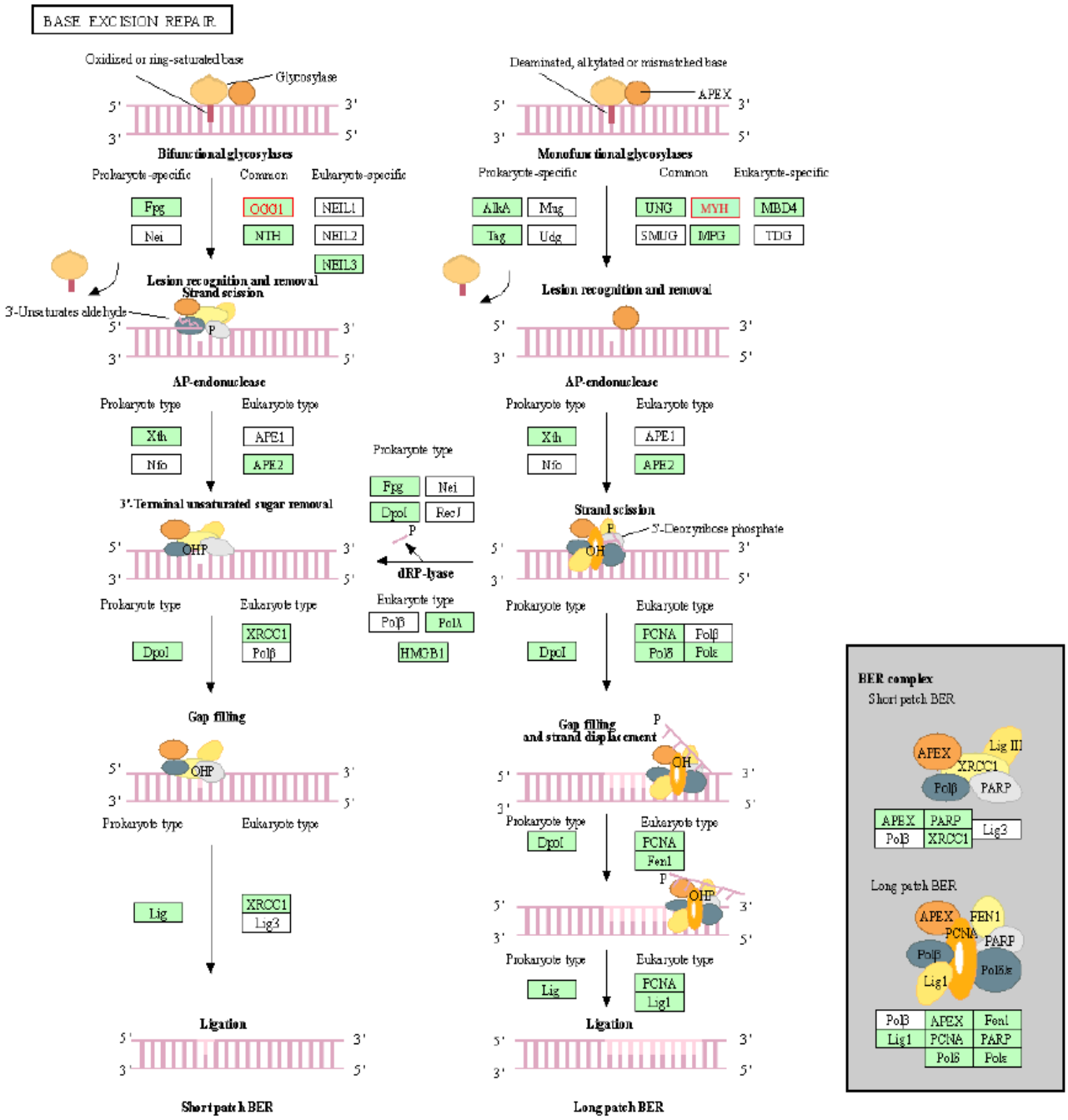

Figure 6: Base Excision Repair Mechanism with OGG1 and MYH proteins

4. Discussion

MYH and OGG1 were analyzed through MSA, phylogenetic tree, 3D structures validation and prediction, subcellular localization, finding their domains and interacting proteins, docking site and metabolic pathway prediction. The sequences of OGG1 and MYH were retrieved and the length of each protein was identified as OGG1 has 1486 amino acids while MYH has 1893 amino acids. Clustal Omega was used to compare ten

different proteins in which their similarities are detected and how proteins have evolved. In figure 1, OGG1 and MYH together with Embryo sac development arrest 10 (EDA10) are more distantly related compared to other proteins. Because OGG1 is more related to Hypothetical protein (HP), Plasmodesma germin-like protein (PDG) and Nuclear DNA-dependent Pol II (NRPB3). While MYH is more closely related to Demeter-like 2 (DML2), Repressor of Silencing (ROS), Nth homolog in Arabidopsis (ATNTH1) and Ras and Rab Interactor Like (RNIL).

The residues of OGG1 and MYH are in the favored region and they are both 90%. In OGG1, there are 344 residues including 20 glycine residues and 18 proline residues. In MYH, there are 543 residues including 42 glycine residues and 28 proline residues. The protein structure of OGG1 is smaller than the protein structure of MYH. As we can see, OGG1 and MYH have alpha (α) helices(red) and beta (β) loops (yellow and green) in their structure. This 3D structure prediction shows how atoms are arranged in amino acids whereas amino acid sequence is folded into secondary and tertiary structure. It helps us to understand the mechanism details that dictate the function(s) of this protein [10]. The 3D structure of OGG1 and MYH will be validated using PDBsum and PROCHECK. PROCHECK is a special option found in PDBsum which is used to construct Ramachandran plot. Ramachandran plot has the backbones of dihedral angles of phi (φ)and psi (ψ) that are visualized in amino acid residues. It is found that OGG1 and MYH proteins' amino acids have the right-handed alpha helix, and their secondary structure has parallel and antiparallel beta sheets [11].

Using ExPASy – PROSITE, OGG1 and MYH have four domains in common, but MYH has Fungal hydrophobin domain as a fifth domain. Fungal hydrophobin domain has start point at 815 residues and end point at 826 residues. This domain has a lot of fungal spores covered by a hydrophobic sheath with 8 conserved cysteines and there is a rodlet layer in which is made up of rodlet protein [20] [21]. In MYH, iron-sulfur (2Fe-2S) cluster domain has start point at 5 residues and end point at 13 residues, and also has start point at 1789 residues and end point at 1797 residue. In OGG1, 2Fe-2S cluster domain has start point at 529 residues and end point at 537 residues, and also has start point at 1230 residues and end point at 1238 residues. 2Fe-2S cluster is a type of ferredoxin and acidic domain that transfers electrons in redox reaction. It has a β-sheet composed of four β-strands in which one α-helix flanks the sheet. The two iron atoms are arranged in tetrahedral structure by the two sulfur (S) atoms and four cysteinyl S atoms [22]. In OGG1, the epidermal growth factor (EGF)-like domain has a start point at 256 residues and end point at 267 residues. In MYH, an EGF-like domain has a start point at 807 residues and an end point at 818 residues, start point at 995 residues and end point at 1006 residues, and also has start point at 1044 residues and end point at 1055 residues. EGF is used to activate the tyrosine kinase in the receptor domain found in cytoplasm. This will initiate signal transduction which provides DNA synthesis and cell proliferation. This domain has 30-40 sequences of amino acids and has six cysteine residues which provide disulfide bridges. Its fold has 2 strands of β-sheet provided by a loop to a C-terminal short 2 sheets that are stranded [23] [24]. In OGG1, von Willebrand factor type C (VWFC) domain has start point at 139 residues and end point at 191 residues, and also has start point at 151 residues and end point at 201 residues, and also has start point at 1116 residues and end point at 1154 residues. In MYH, VWFC domain has start point at 950 residues and end point at 995 residues, and also has start point at 1013 residues and end point at 1086 residues. The name for the VWFC domain is given from the von Willebrand factor type C repeat that's two times present in this protein. Its domain has 70 amino acids that cover 10 conserved cysteines [25]. In OGG1, Anaphylatoxin domain start point at 129 residues and end point at 1732 residues. In MYH, Anaphylatoxin domain starts at 164 residues and ends at 1765 residues. This domain is involved in inflammatory processes and activation of smooth muscle contraction as it has six cysteine forming disulfide bonds with 75 amino acid residues [26] [27].

For subcellular localization, two proteins are present mainly in plastids as plastids can have their own circular DNA repair protein and like mitochondria, it may be descended from the same type of bacteria [28]. The whole interactome of OGG1 and MYH is represented by a graph in figure 4 with nodes and edges. Each node represents

a protein and edges represent protein-protein interactions [29]. Using STRING software, it can be seen that 10 proteins interact with OGG1 and MYH. In the obtained interactome, there are 12 nodes and 40 edges. The results are obtained with high confidence which enables a more detailed visualization. The scores, names and the functions of these 10 proteins are shown in table 3.

AT3G48425 or Apex1-like protein (APE1L), BReast CAncer 1 C-terminal (BRCT) domain-containing DNA repair protein (ATXRCC1), and endonuclease/exonuclease/phosphatase family protein (AT4G36050) are present in the nucleus while apurinic endonuclease redox protein (ARP) is found in plastid [30] [31]. ARP and AT4G36050 will bind to either manganese (Mn) or magnesium (Mg) ions while AT3G48425 binds to Mg ions. APE1L, AT4G36050 and ARP protein are used to exhibit AP endonuclease activity. Unlike AT3G48425, both APE1L and ARP protein speeds ups 3'-phosphor-alpha, beta unsaturated aldehyde (3'-PUA) being converted to 3'-hydroxyl (3'-OH) [31]. Uracil-DNA glycosylase (UNG) is found in mitochondria mechanism [32] and is involved with APE1L, ATXRCC1, and AT4G36050 in BER mechanism [33] [30] [31]. MutS homolog 2 (MSH2) is involved in the DNA MMR and SSBR mechanism and found in chromosomes [34]. Protein kinase family protein (At1g49180) is found in autophagosome where it speeds up the phosphorylation of serine/threonine and may also form a complex with Autophagy-related protein 13 (ATG13) where they regulate events needed for vacuolar delivery or/and enclosure of autophagosome [35]. DNA polymerase delta subunit 1 (EMB2780) found in the nucleus and used to synthesize DNA and degrade single-stranded (ss) DNA in the 3'- to 5'-direction. It is also involved in several DNA repair mechanisms such as BER and MMR. Structural maintenance of chromosome (SMC) proteins (TTN7 and TTN8) and are both part of the cohesin complex needed for the cell cycle [36].

The protein-protein docking sites with OGG1 and MYH are made using ClusPro and enables 3D visualization of protein-protein interaction between OGG1/MYH and 6 proteins obtained by STRING. In figure 5, OGG1 and MYH both bind with EMB2780, TTN7, MSH2, ARP, ATXRCC1 and TTN8. Using KEGG, the metabolic pathway of OGG1 and MYH was analyzed with other proteins (figure 6). Their metabolic pathway is mainly involved in BER which is useful to provide DNA stability and prevention of diseases in *A. thaliana*. BER mechanism uses DNA glycosylases which disconnects the impaired base by breaking the N-glycosidic bond which is linked to the base of the corresponding deoxyribose, generating Abasic site (AP). It basically repairs one base and can only repair non-bulky DNA damage [2]. There are two sub-pathways of BER shown in which these two proteins are involved in different sub-pathway. OGG1 and ATXRCC1, as a bifunctional glycosylase, is involved in formation of short patch BER in which one nucleotide will be replaced [30] [37]. MYH, EMB2780 and UNG, as mono functional glycosylase, is involved in formation of long-patch BER in which 2-13 nucleotides will be replaced [37] [32].

5. Conclusion

In this paper, *in silico* experiment were done on two proteins, known to be involved in DNA repair mechanisms in *A. thaliana*. Using different bioinformatics tools, OGG1 and MYH were analyzed through the following steps: MSA, phylogenetic tree, 3D structures validation and visualization, subcellular localization, domain identification, interactome prediction, protein-protein docking site prediction, and metabolic pathway mapping. These bioinformatics tools bring us a lot of advantages. One of the main advantages is that function and structure of proteins can be analyzed and predicted by just using retrieved sequences. Another advantage is that we can see how these proteins have evolved when we compare this protein with other proteins. Overall, OGG1 and MYH are distantly related as mentioned earlier. Even though they are distantly related, they have similarities. However, OGG1 and MYH have four domains in common and can be found in the same subcellular locations. When it comes to subcellular locations, it is important to mention that DNA is not only found in the nucleus. So, this means that DNA repair enzymes can also be found in other organelles. OGG1 and MYH are found to interact with 10 other proteins and these 10 proteins are also involved in repairing DNA. Through findings of protein interactome, docking-site and KEGG, OGG1 and MYH are not only found in DNA repair mechanisms but also, they interact with proteins involved in DNA replication, development of plant, cell cycle, etc. Also,

OGG1 and MYH are involved with UNG, ARP, AT4G36050, EMB2780, and AT3G48425 in the BER mechanism.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

Abbreviations and acronyms

2Fe2S Cluster iron-sulfur cluster
3D Three-Dimensional
3'-OH 3'-hydroxyl
3'-PUA 3'-phosphor-alpha, beta unsaturated aldehyde
8-oxoG 7,8-dihdro-8-oxoGuanine
AP Apurinic/Apyrimidinic site
APE1L Apex1-like protein
ARP Apurinic endonuclease redox protein
At1g49180 Protein kinase family protein
AT4G36050 Endonuclease/exonuclease/phosphatase family protein
ATG13 Autophagy-related protein 13
ATNTH1 Nth homolog in Arabidopsis
ATXRCC1 BRCT domain-containing DNA repair protein
BER Base Excision Repair
BLAST Basic Local Alignment Sequence Tool
BRCT BReast CAncer 1 C-terminal
Cα-C Alpha Carbon and Carbon
DML2 Demeter-like 2
DNA Deoxyribonucleic acid
DSBR Double Strand Break Repair
EDA10 Embryo sac development arrest 10
EGF Epidermal growth factor
EMB2780 DNA polymerase delta subunit 1
HP Hypothetical protein
HR Homologous Recombination
KEGG Kyoto Encyclopedia of Genes and Genomes
Mg Magnesium
MMR Mismatch Repair
Mn Manganese
MSA Multiple Sequence Alignment
MSH2 MutS homolog 2
MSH2 MutS homolog 2
MUTYH/MYH mutY DNA glycosylase
NCBI National Center for Biotechnology Information
N-Cα Nitrogen and alpha Carbon
NER Nucleotide Excision Repair
NHEJ Non-homologous End-joining

NRPB3 Nuclear DNA-dependent Pol II
OGG1 8-Oxoguanine-DNA Glycosylase 1
PDB Protein Data Bank
PDG Plasmodesmal germin-like protein
PyMOL Molecular Graphics System
RAMPAGE Ramachandran Plot assessment
RNIL Ras and Rab Interactor Like
ROS Repressor of silencing
SMC Structural maintenance of chromosomes (TTN7 and TTN8)
ss Single stranded
STRING Search Tool for the Retrieval of Interacting Genes/Proteins
SUBA4 Subcellular localization database for Arabidopsis proteins 4
TAIR The Arabidopsis Information Resource
U Uracil
UNG Uracil-DNA glycosylase
UV Ultra Violet
VWFC von Willebrand factor type C

References

[1] Wessner, D. R. (2010). The origins of viruses. *Nature Education*, *3*(9), 37.
[2] Dexheimer, T. S. (2013). DNA Repair Pathways and Mechanisms. In L. A. Mathews, S. M. Cabarcas, & E. M. Hurt (Eds.), DNA Repair of Cancer Stem Cells (pp. 19–32). Springer Netherlands.
[3] Alberts, B., Johnson, A., Walter, P., Lewis, J., Raff, M., & Roberts, K. (2008). Molecular cell biology. *New York: Garland Science*.
[4] Krämer, U. (2015). Planting molecular functions in an ecological context with Arabidopsis thaliana. ELife, 4, e06100.
[5] Masson, P. H. (2001). Gravitropism in Arabidopsis thaliana. In Encyclopedia of Genetics (pp. 890–894). Elsevier.
[6] Auffray, C., Chen, Z., & Hood, L. (2009). Systems medicine: the future of medical genomics and healthcare. *Genome medicine*, *1*, 1-11.
[7] Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research, 31(13), 3497–3500.
[8] Cavalier-Smith, T. (2010). Deep phylogeny, ancestral groups and the four ages of life. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1537), 111– 132.
[9] Peretó, J., Bada, J. L., & Lazcano, A. (2009). Charles Darwin and the Origin of Life. Origins of Life and Evolution of the Biosphere, 39(5), 395–406.
[10] Reply, holly (2020). Why Structure Prediction Matters. DNASTAR.
[11] Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomolecular Concepts, 1(3–4), 271–283.
[12] Sigrist, C. J. A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A., & Hulo, N. (2010). PROSITE, a protein domain database for functional 30 characterization and annotation. Nucleic Acids Research, 38(Database issue), D161– D166.
[13] Scott, M. S., Calafell, S. J., Thomas, D. Y., & Hallett, M. T. (2005). Refining Protein Subcellular Localization. PLoS Computational Biology, 1(6).
[14] Plewczyński, D., & Ginalski, K. (2009). The interactome: Predicting the protein-protein interactions in cells. Cellular and Molecular Biology Letters, 14(1), 1–22.
[15] von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). STRING: A database of predicted functional associations between proteins. Nucleic Acids Research, 31(1), 258–261.

[16] Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular Docking: A powerful approach for structure-based drug discovery. Current Computer-Aided Drug Design, 7(2), 146–157.
[17] Alekseenko, A., Ignatov, M., Jones, G., Sabitova, M., & Kozakov, D. (2020). Protein– Protein and Protein–Peptide Docking with ClusPro Server. In D. Kihara (Ed.), Protein Structure Prediction (pp. 157–174). Springer US.
[18] Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30
[19] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 27(1), 29–34.
[20] Lauter, F. R., Russo, V. E., & Yanofsky, C. (1992). Developmental and light regulation of eas, the structural gene for the rodlet protein of Neurospora. Genes & Development, 6(12A), 2373–2381.
[21] Stringer, M. A., Dean, R. A., Sewall, T. C., & Timberlake, W. E. (1991). Rodletless, a new Aspergillus developmental mutant induced by directed gene inactivation. Genes & Development, 5(7), 1161–1171.
[22] Harayama, S., Polissi, A., & Rekik, M. (1991). Divergent evolution of chloroplast-type ferredoxins. FEBS Letters, 285(1), 85–88.
[23] Davis, C. G. (1990). The many faces of epidermal growth factor repeats. The New Biologist, 2(5), 410–419
[24] Tamkun, J. W., DeSimone, D. W., Fonda, D., Patel, R. S., Buck, C., Horwitz, A. F., & Hynes, R. O. (1986). Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin. Cell, 46(2), 271–282.
[25] Zhou, Y.-F., Eng, E. T., Zhu, J., Lu, C., Walz, T., & Springer, T. A. (2012). Sequence and structure relationships within the von Willebrand factor. Blood, 120(2), 449–458
[26] Holbourn, K. P., Acharya, K. R., & Perbal, B. (2008). The CCN family of proteins: Structure–function relationships. Trends in Biochemical Sciences, 33(10), 461–473.
[27] Klos, A., Tenner, A. J., Johswich, K.-O., Ager, R. R., Reis, E. S., & Köhl, J. (2009). The Role of the Anaphylatoxins in Health and Disease. Molecular Immunology, 46(14), 2753–2766.
[28] Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Molecular cell biology 4th edition. *National Center for Biotechnology Information, Bookshelf*, *9*.
[29] Miryala, S. K., Anbarasu, A., & Ramaiah, S. (2018). Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene*, *642*, 84-94.
[30] Lee, J., Jang, H., Shin, H., Choi, W. L., Mok, Y. G., & Huh, J. H. (2014). AP endonucleases process 5-methylcytosine excision intermediates during active DNA demethylation in Arabidopsis. Nucleic Acids Research, 42(18), 11408–11418. 27.
[31] Li, Y., Córdoba-Cañero, D., Qian, W., Zhu, X., Tang, K., Zhang, H., Ariza, R. R., Roldán-Arjona, T., & Zhu, J.-K. (2015). An AP Endonuclease Functions in Active DNA Demethylation and Gene Imprinting in Arabidopsis. PLoS Genetics, 11(1), e1004905.
[32] CórdobaCañero, D., Dubois, E., Ariza, R. R., Doutriaux, M. P., & Roldán-Arjona, T. (2010). Arabidopsis uracil DNA glycosylase (UNG) is required for base excision repair of uracil and increases plant sensitivity to 5-fluorouracil. *Journal of Biological Chemistry*, *285*(10), 7475-7483.
[33] Akishev, Z., Taipakova, S., Joldybayeva, B., Zutterling, C., Smekenov, I., Ishchenko, A. A., Zharkov, D. O., Bissenbaev, A. K., & Saparbaev, M. (2016). The major Arabidopsis thaliana apurinic/apyrimidinic endonuclease, ARP is involved in the plant nucleotide incision repair pathway. DNA Repair, 48, 30–42.
[34] Culligan, K. M., & Hays, J. B. (2000). Arabidopsis MutS Homologs—AtMSH2, AtMSH3, AtMSH6, and a Novel AtMSH7—Form Three Distinct Protein Heterodimers with Different Specificities for Mismatched DNA. The Plant Cell, 12(6), 991.
[35] Suttangkakul A., Li F., Chung T., Vierstra R.D. The ATG1/ATG13 protein kinase complex is both a regulator and a target of autophagic recycling in *Arabidopsis*. *Plant Cell.* 2011;23:3761–3779.

[36] Lam, W. S. (2005). Characterization of Arabidopsis thaliana SMC1 and SMC3: Evidence that AtSMC3 may function beyond chromosome cohesion. Journal of Cell Science, 118(14), 3037–3048.
[37] Moen, M. N., Knævelsrud, I., Haugland, G. T., Grøsvik, K., Birkeland, N. K., Klungland, A., & Bjelland, S. (2011). Uracil-DNA glycosylase of Thermoplasma acidophilum directs long-patch base excision repair, which is promoted by deoxynucleoside triphosphates and ATP/ADP, into short-patch repair. *Journal of bacteriology*, *193*(17), 4495-4508.