Article

In silico interactome and docking site study of DNA repair proteins (APE1 and APE2) and their role in base excision repair in *Arabidopsis thaliana*

Esma Kurtanović, **Mohamed Ragab Abdel Gawwad**

International University of Sarajevo, Faculty of Engineering and Natural Sciences, Sarajevo, 71210, Bosnia and Herzegovina

E-mail: esma@kurtanovic.net, mragab@ius.edu.ba

Received 10 May 2021; Accepted 20 June 2021; Published 1 September 2021

Abstract

Flowering time is a life history trait of adaptiveness. Over many generations, phenotypes happen to emerge as mutations or spontaneous damage accumulates in the plastome. Thus, it is of great importance to investigate DNA repair mechanism roles of some proteins. Specifically, this study aims to determine potential targets that are part of base excision repair mechanism in *Arabidopsis thaliana*. To do so, bioinformatic methods are implemented in order to shed light on the functioning of our protein homologs. Their structural and functional similarities are confirmed by multiple sequence alignment, 3D structure prediction, phylogenetic tree construct and interactome analysis. The results indicate that interaction between two proteins is strong evidence that the proteins are involved in the same biological process. This study can be seen as a valuable data resource of predicted cellular functions of proteins and the evolutionary conservation of AP endonuclease families, which again, portrays the divergence of activities and biological contributions.

Keywords DNA damage and repair; interactome; docking sites; APE homologs; *A. thaliana*; base excision repair.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Plants are constantly exposed to endogenous cellular processes and exogenous environmental events that can compromise genome integrity through DNA damage. Cells have evolved a significant and evolutionarily conserved defense mechanism known as DNA damage response (DDR) to combat the adverse effects of these processes and events (Anderson et al., 2008).

DNA found in the cellular compartments of both the nuclear and mitochondrial compartments is susceptible to damage from different sources. Mitochondria and chloroplast are more prone to damage due to

reactive oxygen species (ROS). These organelles possess a base excision repair (BER) which may remove oxidative lesions. Therefore, we focus on pathway crosstalk events primarily mediated through protein-protein interactions and extend the analysis of proteins that could affect BER activity and pathway crosstalk in response to DNA damage through multiple mechanisms (Limpose et al., 2017)..

The essential DNA repair enzymes involved in BER pathway are apurinic/apyrimidinic (AP) endonucleases, which cleave DNA at AP sites. Our model organism is *Arabidopsis thaliana*; its genome encodes three AP endonuclease-like proteins: APE1L, APE2 and ARP, which is the focus of our work (denoted as APE1-1, APE1-2, APE2, respectively).

The aim of the present study is to better understand DNA repair mechanisms through identifying potential targets of BER proteins and analyzing the structure and function of APE homologs in *A. thaliana*.

2 Materials and Methods

We performed bioinformatics analysis, which is crucial before any downstream experiments. Firstly, we retrieved the BLAST sequences of APE proteins using The Arabidopsis Information Resource (TAIR) database. These were aligned using Multiple Sequence Alignment (MSA) via ClustalW software. Afterwards, *phylogeny.fr* was used to construct the phylogenetic tree considering *A. thaliana* close relatives. Prediction of protein 3D structure was performed using SWISS model. These were further validated by RAMPAGE software. Furthermore, two additional software were implemented, STRING and ClusPro, for the prediction of the interactome and docking sites, respectively. PyMol was used to represent the interacting residues. Finally, KEGG software was utilized for generating the BER pathway, which included our APE proteins of interest.

3 Results

According to the general rule of thumb for the homology of the proteins with ≥100 amino acids, proteins with 20% or more identity are homolgous. Thus, APE1-1, APE1-2 and APE2 are homologous. This is shown as tabular representation in Table 1.

Table 1 Homology score.

	APE2	APE1-1	APE1-2
APE2	100.00	30.77	22.14
APE1-1	30.77	100.00	31.31
APE1-2	22.14	31.31	100.00

Phylogenetic tree shows three groups of proteins with respect to their similarity/divergence. Displayed close relatives of *A.thaliana* are enriched with APE homologues. APE1-1 and APE1-2 are highly similar among themselves and more distant from APE2. This is shown in Fig. 1A below.

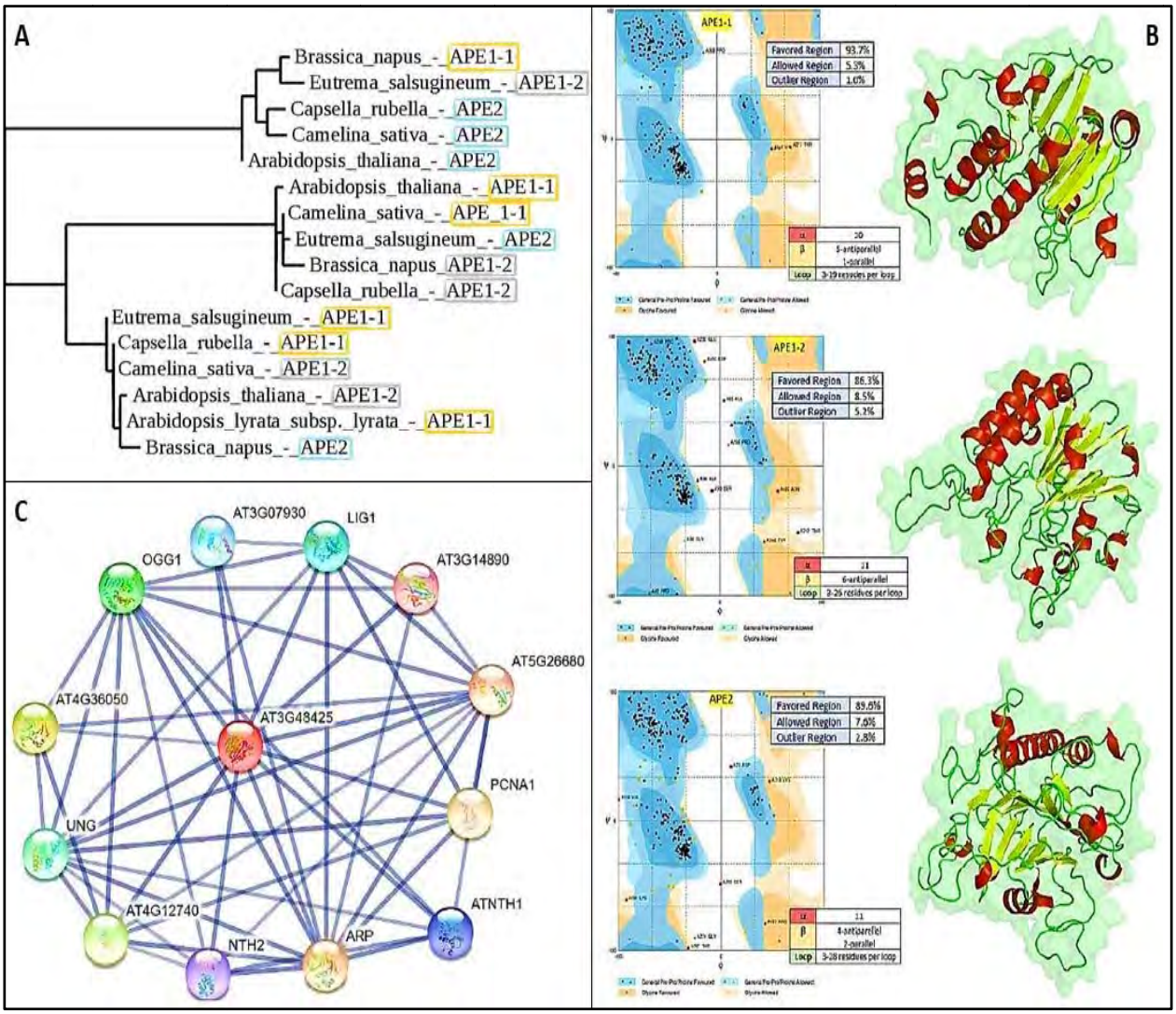

Fig. 1 A. Phylogenetic tree; **B**. 3D prediction and validation; **C**. Interactome analysis.

Furthermore, the predicted 3D structures of our target proteins confirm their structural similarities. APE1-1, APE1-2 and APE2 have alpha-helices, beta sheets and loops in them with the difference being that APE1-1 contains 10, while APE1-2 and APE2 contain 11 alpha-helices. Additionally, loop length is visibly shorter in APE1-1 and APE1-2, whereas the APE2 protein consists of longer loops. Quality of submitted models was validated based on the phi/psi angle arrangement. This is represented by Ramachandran plot as well as PyMol in Fig. 1B.

Fig. 1C shows the Interactome analysis which generated following functional partners of our target proteins: AT4G12740, OGG1, UNG, AT3G07930, LIG1, ATNTH1, NTH2, AT5G26680, AT3G14890, PCNA1.

Representation of interacting residues data is listed in Table 2.

Table 2 Interacting residues.

Target	Partner	Amino Acid Residues		
2A	APE1-1	356 LYS	359 SER	368 GLY
	NTH2	166 GLU + 168 ASP	196 HIS	229 TYR
	APE1-1	365 GLY	367 SER	363 GLY + 365 GLY
	AT5G26680	102 TYR	105 ARG	102 TYR
2B	APE1-2	330 GLY	61 TRP	66 LYS
	OGG1	181 HIS	100 THR	38 PRO
	APE1-2	298 ASN	190 ASP	336 GLU
	ATNTH1	296 TRP	299 ARG	303 LYS
2C	APE2	183 LEU	188 ARG	333 ASP
	OGG1	103 LYS	100 THR	97 HIS
	APE2	146 HIS	279 ASP	283 HIS
	AT3G07930	323 HIS	442 ARG	442 ARG
2D	APE2	143 GLY	92 LEU	187 ARG
	NTH2	169 ARG	170 THR	165 VAL + 167 CYS + 168 ASP
	APE2	43 GLN	351 PRO	352 MET + 355 GLY
	PCNA1	253 PRO	132 GLU	130 ASP (2^{nd} fold group)
2E	APE2	289 HIS	185 GLN	336 ASP
	UNG	150 GLN	121 PHE	152 LEU
	APE2	184 ARG	185 GLN	90 THR
	ZDP	620 LYS	620 LYS	641 SER

APE1-1 showed interaction with all aforementioned functional partners. APE1-2 exhibited interaction with 8 of them. APE2 interacts with two of them, specifically AT4G12740 and OGG1. However, our results indicate fewmore potential interacting partners, i.e., OGG1, AT3G07930, NTH2, PCNA1, UNG, ZDP. This is shown in Fig. 2.

Generated KEGG pathway showed that our identified functional partners are a part of BER. The APE expression speeds up DNA repair mechanism, which is essential for protein survival.

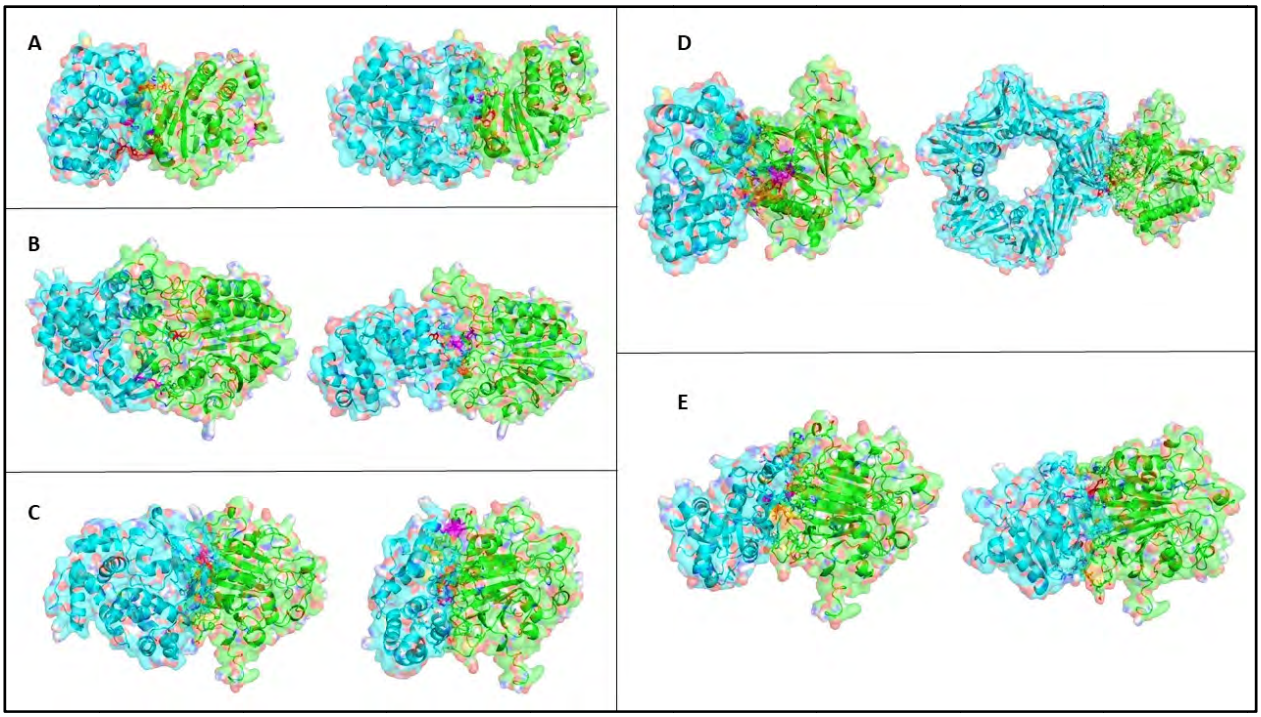

Fig. 2 Docking site prediction **A.** APE1-1 interacting with NTH2 (left) and AT5G26680 (right) **B.** APE1-2 interacting with OGG1 (left) and ATNTH1 (right) **C.** APE2 interacting with OGG1 (left) and AT3G07930 (right) **D.** APE2 interacting with NTH2 (left) and PCNA1 (right) **E.** APE2 interacting with UNG (left) and ZDP (right). Note that the green color represents our target proteins whereas the blue color represents functional partners.

4 Discussion

MSA is a fundamental base for further evolutionary studies, such as constructing the phylogenetic tree. It was built solely from the sequences previously retrieved through TAIR database (Pearson, 2013). Even though APE1-1, APE1-2 and APE2 are homologous, APE1-1 and APE1-2 have the highest identity percentage.

Although alignment accuracy is positively associated with phylogenetic accuracy, we find that the amount of improvement in phylogenetic estimation resulting from an improved alignment can vary from very small to large. Based on this analysis, we show that *A.thaliana* relatives were founded by multiple individuals drawn from a diverse ancestral population closely related to extant *A.thaliana* (Brandvain et al., 2013). All species were closely related to *Arabidopsis thaliana.*

Tertiary structure is the most pivotal category of structure as it is the sole determinant of the function of the protein (StatPearls, 2021). What can be mentioned is the loop length in these proteins which is visibly shorter in APE1-1 and APE1-2, meaning that they are more durable at higher temperatures, whereas the APE2 protein consists of longer loops, therefore it affects its stability. On the other hand, the longer the alpha-helix, the more stable the protein is.

Ramachandran plot validates our 3D structure prediction by showing the high quality of targeted proteins having more than 86% residues in the favored region. Results confirm the shared structural and functional similarities between our targeted proteins.

Protein-protein interaction (PPI) is a highly specific physical form of interaction between two or more protein molecules, which is affected by electrostatic forces, hydrogen bonds and hydrophobic events. For an interaction to occur, ligand needs to be in close proximity to a charged residue. In order to find out the interacting residues, we need to know what our interacting partners are. The results show that the interactome of our three proteins mainly comprises of proteins with an already confirmed role in BER. This is another indicator of what we actually wanted to prove (Lu et al., 2020).

Moreover, the generated KEGG pathway provided us with more insight into this matter. All of the data we collected are pointing towards APE family of proteins being important contributors in plant DNA repair mechanisms.

5 Conclusion

Unified understanding of visualizations of repair proteins and their interactions with the DNA is a foundation of further discoveries. This helps to further recognize damaged bases, such as sequence-independent DNA recognition motifs, in other words the initial damage recognition and removal.

This study portrayed steps of bioinformatic analysis, which is critical before doing downstream experimental procedures. The functional similarity of our target proteins APE1-1 and APE1-2 is confirmed and validated, since both of them share the same interactome. Besides that, they share a significant identity score and have similar 3D structures. All of the data we collected is pointing to APE family of proteins being important contributors in plant DNA repair mechanisms.

For the future perspective, it would be of great significance to do experimental validation based on interactome analysis. Up until now, there is no definitive evidence that most of these proteins interact with each other, which can be determined by yeast-two-hybrid system. Although virtual screening/bioinformatics is a valuable tool, more effort should be put in endorsing novel alternatives to increase validity of these types of research.

References

Anderson HJ, Vonarx EJ, Pastushok L, Nakagawa M, Katafuchi A, et al. 2008. *Arabidopsis thaliana* Y-family DNA polymerase eta catalyses translesion synthesis and interacts functionally with PCNA2. The Plant Journal: For Cell and Molecular Biology, 55(6): 895-908. doi:10.1111/j.1365-313X.2008.03562.x

Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G. 2013. Genomic identification of founding haplotypes reveals the history of the selfing species *Capsella rubella*. PLoS genetics, 9(9): e1003754. doi:10.1371/journal.pgen.1003754

Limpose KL, Corbett AH, Doetsch PW. 2017. BERing the burden of damage: Pathway crosstalk and posttranslational modification of base excision repair proteins regulate DNA damage management. DNA Repair, 56: 51-64. doi:10.1016/j.dnarep.2017.06.007

Lu H, Zhou Q, He J, Jiang Z, Peng C, Tong R, Shi J. 2020. Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. Signal Transduction and Targeted Therapy, 5(1): 213. doi:10.1038/s41392-020-00315-3

Pearson WR. 2013. An introduction to sequence similarity ("homology") searching. In: Current Protocols in Bioinformatics (Chapter 3, Unit 3.1). Wiley, USA. doi:10.1002/0471250953.bi0301s42

StatPearls. 2021. StatPearls Publishing, USA