

In silico* interactome analysis of cross-link repair proteins (RPA32b and RPA2) in *Arabidopsis thaliana

Karić Amina^{1*}, Mohamed Ragab Abdel Gawwad²

^{1,2}Genetics and Bioengineering, International University of Sarajevo, Bosnia and Herzegovina

*Corresponding author: 210301013@student.ius.edu.ba.

© The Author
2021.
Published by
ARDA.

Abstract

DNA molecule is found in every cell in a living organism and function as an instruction manual for growth, development, as well as any other processes that happen. As the cell divide, it needs to copy its DNA. During the copying, an error can occur, which can lead to a disease. However, there are different error repair mechanisms that are searching for an error, and repairing it. This study focused on 2 proteins involved in DNA repair mechanism, RPA32B and RPA2 from *Arabidopsis thaliana*. Throughout multiple in silico approached, multiple sequence alignment, phylogenetic tree construction, 3D visualisation, combination and interaction, we confirmed that the RPA2 and RPA 32b are involved in DNA replication, meiotic recombination and DNA repair mechanism.

Keywords: DNA repair mechanism; RPA32B; RPA2; structural analysis; *Arabidopsis thaliana*

1. Introduction

Every cell in our body contains Deoxyribonucleic Acid (DNA). In order for cells to divide, they need to copy their DNA, so that every daughter cell has the same copy of DNA. During the copying process (replication), an error can occur, which can make the product of that part of the copied DNA, non-functional. Also, errors can occur from different agents from outside, such as UV light. However, cell has various mechanism by which it repairs it errors [1]. It differs in terms of the structure and complexity of organism, as well as the type of damage made, and thus the they are divided into several types: BER – base excision repair, MMR – mismatch repair, DSBR – double-strand break repair and NER – nucleotide excision repair [2].

1.1. RPA32B and RPA2 proteins

Replication protein A (RPA) is single stranded DNA binding protein. It is important for many DNA processing pathways like DNA replication, meiotic recombination and for some DNA repair mechanisms. RPA is found in every eukaryote [3]–[6]. RPA protein in *Arabidopsis thaliana* is divided into five phylogenetically distinct RPA1 subunits (RPA1A-E), two RPA2 (RPA2A and B) subunits, and two RPA3 (RPA3A and B) subunits [7]. In this paper, the focus will be on two proteins that are part of DNA replication factor A complex (RPA): RPA2 and RPA32B.

1.2. *Arabidopsis thaliana*

Arabidopsis thaliana is a plant from the family of mustard. It is a small plant with a small genome, which is used as a model organism for research. Its genome is organised into 5 chromosomes and contains about 20,000 genes [8].

This work is licensed under a [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>) that allows others to share and adapt the material for any purpose (even commercially), in any medium with an acknowledgement of the work's authorship and initial publication in this journal.



2. Material and methods

2.1. Multiple sequence alignment

Multiple sequence alignment is used to find similarities or relationships between sequences [9]. It gives insights into sequence-structure-function relationships of nucleotide or protein sequence families. It can also give information about evolutionary and functional relationships of protein families [10]. In this paper, Clustal Omega tool was used.

2.2. Phylogenetic tree construction

The second step, which takes place after multiple sequence alignment, is phylogenetic tree construction in Clustal Omega. Phylogenetic tree construction is a representation of the evolutionary relationship between organisms and diagrammatic form which shows different species, organisms, their ancestors. Through its branching, it represents a prediction about relationships between sequences and their possible common ancestors [11]. Phylogeny Fr. was used for the Phylogenetic tree for RPA32B and RPA2 proteins [33].

2.3. 3D structure prediction, visualisation and validation

3D structure of RPA32B and RPA2 proteins will be presented using PyMol Molecular Graphic System [37].

2.4. Domain identification

Tolls SMART software and PFAM database was used to identify the domains of RPA32B and RPA2 proteins. SMART is an online resource used for annotation and identification of protein domains, as well as architecture of protein domains analysis [12]. Pfam database is a database for protein classification into domain and families [13].

2.5. Protein interactome prediction

Interactome is a complete set of molecular interaction of a certain molecule in a given cell or some other biological environment [14]. For protein-protein interaction prediction, STRING database; which includes both physical and functional interactions; is used [15].

2.6. Metabolic pathway mapping

KEGG pathway map is a graphical diagram drawn manually, used to show signalling, metabolic and other molecular reaction and interaction networks. It also shows information such as how genes or their products interacts in pathways [16]. In this paper, KEGG tool was used for determining the metabolic pathways of RPA32B and RPA2 proteins.

2.7. Docking site prediction

For docking site prediction, tools SWISS-MODEL and ClusPro are used[24]-[30]. Docking site prediction shows us 3D visualization of protein-protein interaction between RPA32B/RPA2 and 10 proteins found by STRING.

3. Results

3.1. Multiple sequence alignment

Cluster Omega tool was used to get multiple sequence alignment results. The results are shown in Fig 1 and Fig.2.

In figure 1 and 2 we see the results represented in different symbols, their meaning is as following:

- “.” “ Shows the level of conservation at a particular position.
- “*” Means that there is exactly the same amino acid at that position. One protein has different amino acid, all the other have the same amino acid.
- “.” Represents lower degree of similarity
- “.” One protein has different amino acid, all the other have the same amino acid
- “.” Represents higher degree of similarity

3.2. Phylogenetic tree

Phylogeny Fr. was used for the Phylogenetic tree for RPA32B and RPA2 proteins. The results are shown in Fig 3.

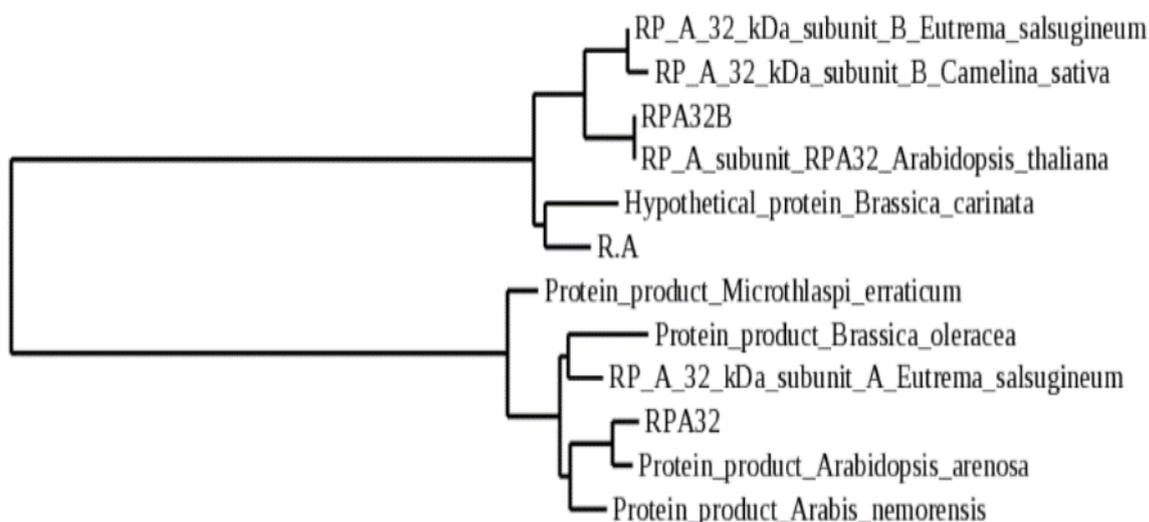


Figure 3. Phylogenetic tree of RPA32B and RPA2 proteins

3.3. 3D structure visualization and validation

For 3D structure prediction and validation, we used online tools SWISS-MODEL, PyMOL, PROCHECK and Ramachandran plot diagrams. Fig. 4 shows 3D structure visualization. Figures 5 and 6 represent ramachandran plot results.

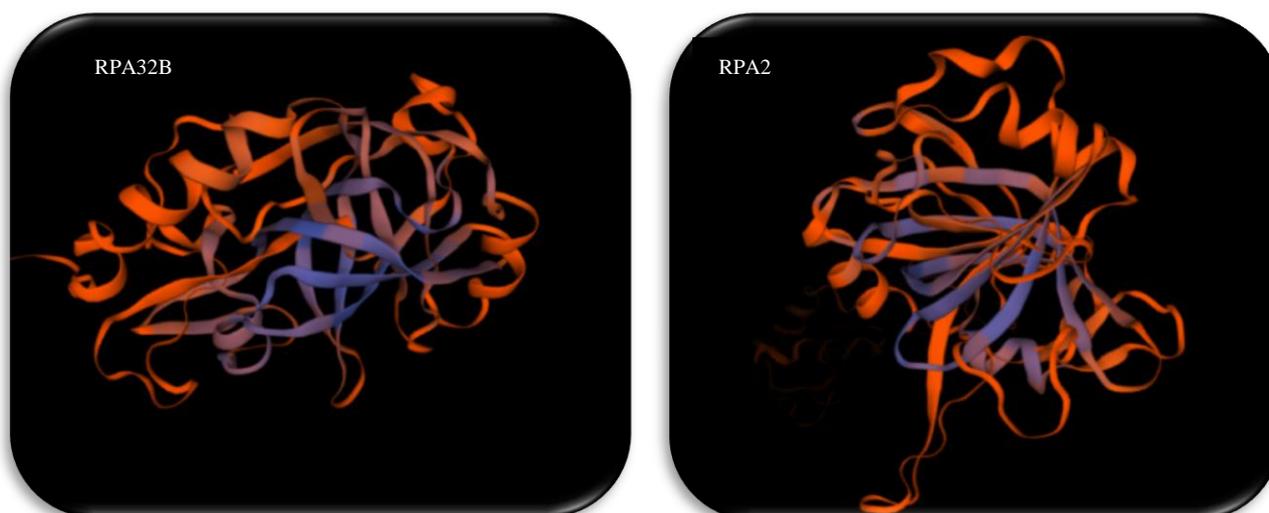


Figure 4. RPA32B and RPA2 3D models

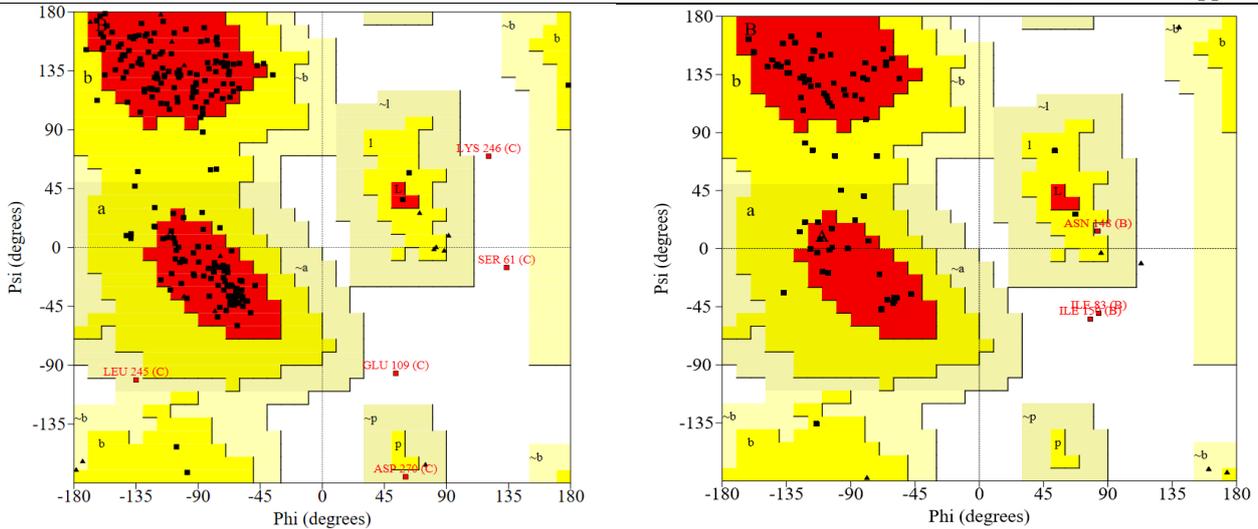


Figure 5. Ramachandran plot result of RPA2 protein. Figure 5. Ramachandran plot result of RPA32B protein.

3.4. Domain identification

Domain identification and analysis was done using SMART software and PFAM database. Table 1. shows position, and start and end points of RPA32B and RPA2 protein’s domains.

Table1. Positions of RPA32B and RPA2 domains

RPA32B protein				RPA2 protein			
Domain	Description	Start	End	Domain	Description	Start	End
RPA_C	Replication protein A C terminal	161	270	RPA_C	Replication protein A C terminal	163	272
tRNA_anti-codon	OB-fold nucleic acid binding domain	70	144	tRNA_anti-codon	OB-fold nucleic acid binding domain	71	148

3.5. Interactome prediction

Using STRING software, interactome prediction for proteins RPA32B and RPA2 was analyzed.

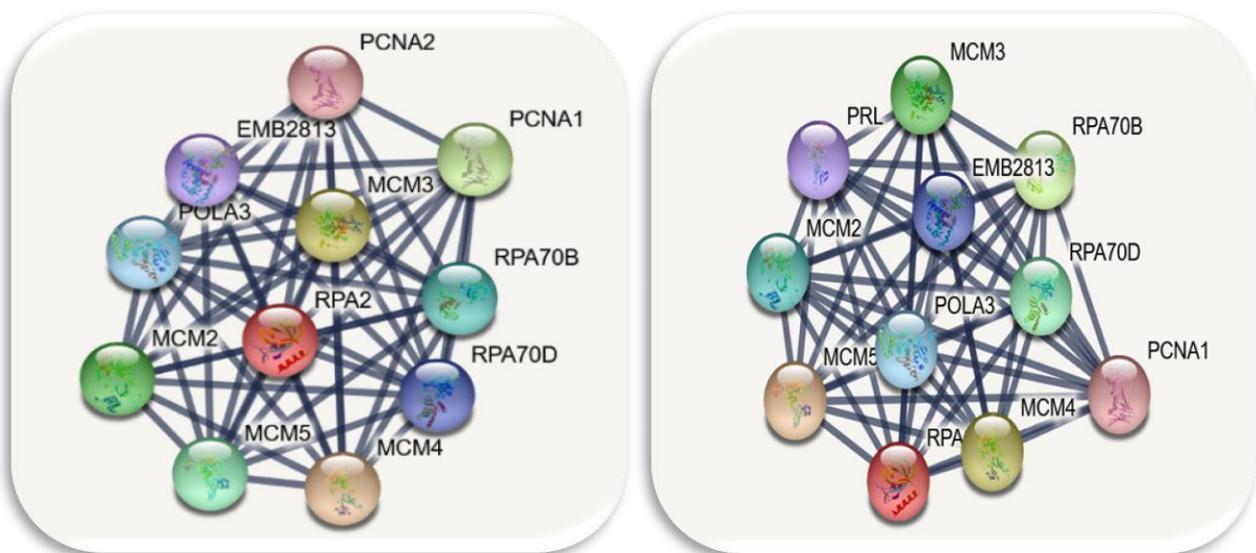


Figure 6. Interactome prediction representation for RPA2 and RPA32B proteins

Table 2. List of 10 proteins that interact with RPA32B and RPA2 proteins

Protein	Full name	ID	Function
MCM3	Minichromosome maintenance (MCM2/3/5) family protein	AT5G46280.1	Function as a DNA helicase and which is essential to undergo a single round of replication initiation and elongation per cell cycle in eukaryotic cells.
PRL	Minichromosome maintenance (MCM2/3/5) family protein	AT4G02060.2	Involved in DNA replication initiation.
MCM2	Minichromosome maintenance (MCM2/3/5) family protein	AT1G44900.1	Function as a DNA helicase and which is essential to undergo a single round of replication initiation and elongation per cell cycle in eukaryotic cells. Plays a crucial role in the control of de-differentiation and cell proliferation processes required for lateral root formation. Is essential for embryo development.
MCM5	Minichromosome maintenance (MCM2/3/5) family protein	AT2G07690.1	Involved in DNA replication initiation.
MCM4	Minichromosome maintenance (MCM2/3/5) family protein	AT2G16440.1	Function as a DNA helicase and which is essential to undergo a single round of replication initiation and elongation per cell cycle in eukaryotic cells.
PCNA1	Proliferating cellular nuclear antigen 1	AT1G07370.1	Involved in the control of eukaryotic DNA replication.
RPA70D	Replication protein A 70 kDa DNA-binding subunit D	AT5G61000.1	Required for DNA recombination, repair and replication. Involved in repair of double-strand DNA breaks (DSBs) induced by genotoxic stresses (By similarity).
RPA70B	Replication protein A 70 kDa DNA-binding subunit B	AT5G08020.1	Component of the replication protein A complex (RPA) required for DNA recombination, repair and replication.
EMB2813	Probable DNA primase large subunit	AT1G67320.3	Synthesizes small RNA primers for the Okazaki fragments made during discontinuous DNA replication.
POLA3	DNA primases	AT5G41880.1	Its function is described as DNA primase activity; Involved in DNA replication, synthesis of RNA primer, DNA replication

From table 2 we can clearly see that all interactome proteins are involved in DNA replication mechanisms, DNA recombination, initiation, repair, and control.

3.6. Metabolic pathway mapping

To find metabolic pathway mapping, through the KEGG tool, showed that of RPA32B and RPA2 proteins are involved in the DNA replication, nucleotide excision repair and mismatch repair, respectively.

3.7 Docking site prediction

In figure 7 we present the predicted docking models of all interactome partners, visualized in PyMol software.

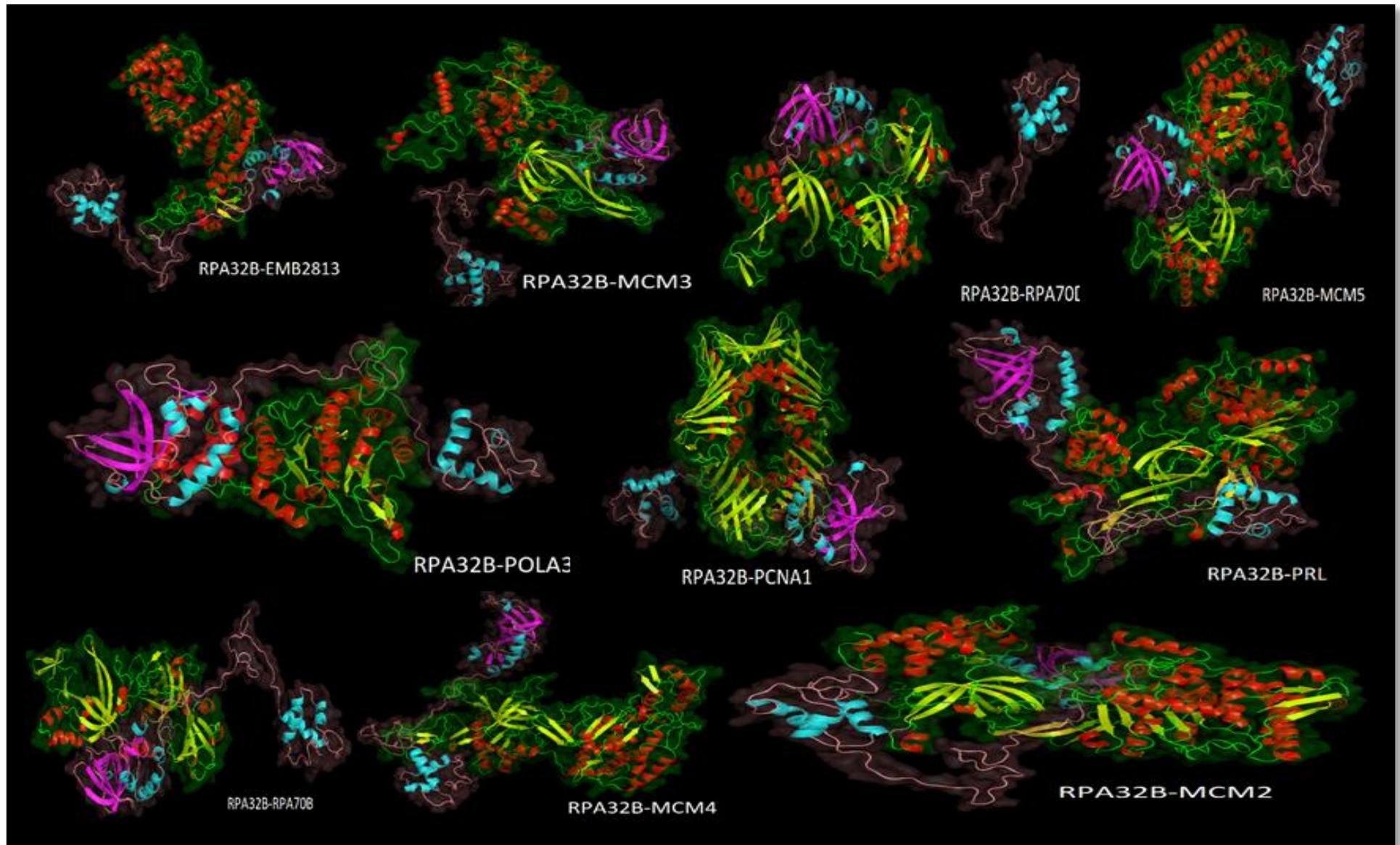


Figure 7. 3D visualization of protein-protein interaction

4. Discussion

The RPA32B and RPA2 proteins were analyzed via multiple sequence alignment, phylogenetic tree, 3Dstructure and visualization, subcellular localization, domain identification, interactome prediction and metabolic pathway mapping. Using the TAIR database, with its integrated tools, we searched for homologs needed for multiple sequence alignment and comparison. Homolog organisms are derived from the common ancestor and they have some similarities [17]. Multiple sequence alignment is a process used to analyze related proteins, for example homolog from common ancestor, in order to find relationships between proteins or similar function or structure [18]. Using Cluster Omega tool, we compared the following homologous organisms: *Capsela rubella*, *Camelina sativa*, *Brassica carinata* and *Euterna salsugineum*. In Fig. 1 we see the multiple sequence alignment results of RPA32B and RPA2 proteins from the above, mentioned organisms, annotating the conserved region. We obtained in this analysis the level of conservation at a particular position, looking at the same or different amino acids withing the sequence, concluding that in all organisms they have the same amino acid sequence at some point, representing higher and lower degree of similarity.

The phylogenetic tree is seen in Fig.2, showing connection between organisms based on evolutionary connection. It can be observed that RPA23B in *Arabidopsis thaliana* is closely related with RPA32B in *Camelina sativa* and in *Euterna salsugineum*.

From Fig. 3, the 3D models of RPA2 and RPA32B were shown, including the alpha helix and beta sheet locations.

Ramachandran plot analysis shows us prediction of amino acids types. Every black dot on Ramachandran plot represents one amino acid. By the location of these dots in psi and phi coordinates, we can find out which secondary structure each of our amino acid has. Based on our Ramachandran plot analysis, we found out that RPA2 proteins' amino acids have right-handed alpha helix, anti-parallel beta sheets and parallel beta sheet secondary structure. Few amino acids have collagen triple helix and only two of them have left-handed alpha helix. RPA32B proteins' amino acids mostly have antiparallel *b sheets* and parallel *beta sheet* secondary structure. In addition, we observe few right-handed alpha helix amino acids.

Domain identification and analysis was done using SMART software and PFAM database. Protein domains are the basic functional units of protein [19]. From Fig. 4, we see that both protein's polypeptides have 2 domains. Using PFAM database amino acid residues lengths as well as start and endpoints were found. In the table 1 we can see that RPA32B domain RPA_C has start point at 161 residue and end point at 270 residue. tRNA anti-codon binding domain has start point at 70 residue and shows the end point at 144 residue. RPA32 protein domain RPA_C has start point at 163 residue and end point at 272 residues. TRNA anti-codon binding domain has start point at 71 residue and end point at 148 residue [13].

Interaction network is gene/protein graphical representation which contain nodes and edges. Each edge represents inaction between gene/protein, where a node is gene/protein [20]. Using STRING software, interactome prediction for proteins RPA32B and RPA2 was analysed. From Fig. 6 it can be seen that 10 proteins interact with RPA32B and RPA2 proteins. In the obtained interactome, 11 nodes and 55 edges are found. Proteins MCM5, MCM4, MCM3, MCM2, PRL, comes from MCM2/3/5 Minichromosome maintenance, family protein. Proteins that function as DNA helicase and which are needed for replication initiation are: MCM3, MCM4 and MCM2. PRL and MCM5 proteins are also involved in replication initiation. RPA32B and RPA2 show strong correlation to Proliferating cell nuclear antigen protein (PCNA1), known to be involved in the control of eukaryotic DNA replication by increasing the polymerase's processibility during elongation of the leading strand [44]. In this regard, these results confirm the strong correlation of the PCNA protein family with MCM family in DNA replication and repair. Further, the interaction of RPA70D and RPA70B proteins with EMB2813 is expected, since EMB2813 has a role in synthesizing RNA primers for Okazaki fragments and known to be connected to DNA primase function, so directly connected to the DNA replication mechanism strand [45]. DNA primase POLA3 has function as DNA primase activity. For our prediction, high confidence was used which means that visualization is better and more detailed.

For additional confirmation of RPA32B and RPA2 proteins function, the metabolic pathway mapping in KEGG program was used. Fig. 7, 8 and 9 show involvements of RPA32B and RPA2 proteins in DNA replication, nucleotide excision repair and mismatch repair, is shown, respectively. Nucleotide Excision Repair Mechanism (NER), is used to remove lesions which distort helix in more than one sequences of the DNA at

the same time. Such lesions can be caused by the excessive exposure to the sun or the ultraviolet light, carcinogenic chemicals and similar and they are usually bulky in shape [46]. RPA interacts with Xeroderma Pigmentosum group A (XPA) proteins, a response factor included in NER. If there is not enough XPA it can result in high sensitivity of cells in response to UV light. XPA needs protein complex, in this case RPA, because it does not have enzyme activity so to function properly it needs to bind to RPA. RPA32B and RPA2 bind to the central region of XPA where XPA binding is the main step in NER, after which process can continue. Further, RPA32B and RPA2 are included in the process of stimulating EXO-catalyzed excision, protecting the ssDNA gap generated during excision, facilitating the termination of MMR excision and repairing/synthesizing DNA [21].

Docking site prediction shows us 3D visualization of protein-protein interaction between RPA32B/RPA2 and 10 proteins found by STRING. Those are: MCM3, PRL, MCM2, MCM5, MCM4, PCNA1, RPA70D, RPA70B, EMB2813 and POLA3, its the first docking study predicted for RPA32B/RPA2 in correlation to the interactome proteins, as seen in figure 6. In addition, this study represents the first 3D modeling of RPA32B/RPA2 proteins in interaction with MCM3, PRL, MCM2, MCM5, MCM4, PCNA1, RPA70D, RPA70B, EMB2813 and POLA3, showing potential docking site and regions.

5. Conclusion

In this paper, in silico analysis of two proteins involved in DNA replication and repair mechanisms in *Arabidopsis thaliana* was done. Using different bioinformatics tools, RPA32B and RPA2 proteins were analyzed through the following steps: multiple sequence alignment, phylogenetic tree, 3D structure and visualization, subcellular localization, domain identification, interactome prediction and metabolic pathway mapping.

In regards to their structure, the RPA2 proteins' amino acids have right-handed alpha helix, *antiparallel b* sheets and *parallel b* sheet secondary structure. Few amino acids have collagen triple helix and only two of them have left-handed alpha helix. The RPA32B proteins' amino acids mostly have *antiparallel b* sheets and *parallel b* sheet secondary structure. There are also few right-handed alpha helix amino acids. Both of our protein's polypeptides have 2 domains. Domains are structural and functional units of proteins that are responsible for specific function. RPA2 and RPA32B proteins remove lesions which distort helix in more than one sequences of the DNA at the same time. This is the first study that predicted the interactome partners of RPA32B and RPA2 proteins, explained the docking options and assessed the 3D predicted models of all proteins. In general, we confirmed that RPA32B and RPA2 proteins have an important role in many processes and mechanism in *Arabidopsis thaliana*, especially in DNA replication, where they keep single stranded DNA unwound and DNA mismatch/nucleotide excision repair processes.

6. Abbreviations and acronyms

3D - Three-dimensional

BER - Base Excision Repair

DNA - Deoxyribonucleic Acid

DSBR - Double Strand Break Repair

KEGG - Kyoto Encyclopedia of Genes and Genomes

MMR - Mismatch Repair

NER - Nucleotide Excision repair system

PyMol - PyMOL Molecular Graphics System

RPA - Replication Protein A

SMART - Simple Molecular Architecture Research Tool

ssDNA - single-stranded Deoxyribonucleic acid

STRING - Search Tool for the Retrieval of Interacting Genes/Proteins

SUBA4 - Subcellular localization database for Arabidopsis proteins

SWISS - structural bioinformatics web-server

TAIR - the Arabidopsis Information Resource

XPA - Xeroderma Pigmentosum group A

References

- [1] B. [1] Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of*

- the cell 5E: Reference edition*, 5th ed. Boca Raton, FL: CRC Press, 2008.
- [2] L. A. Mathews, S. M. Cabarcas, and E. Hurt, Eds., *DNA repair of cancer stem cells*. Dordrecht, Netherlands: Springer, 2012.
- [3] M. S. Wold, "Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism," *Annu. Rev. Biochem.*, vol. 66, no. 1, pp. 61–92, 1997.
- [4] C. Iftode, Y. Daniely, and J. A. Borowiec, "Replication protein A (RPA): the eukaryotic SSB," *Crit. Rev. Biochem. Mol. Biol.*, vol. 34, no. 3, pp. 141–180, 1999.
- [5] E. Fanning, V. Klimovich, and A. R. Nager, "A dynamic model for replication protein A (RPA) function in DNA processing pathways," *Nucleic Acids Res.*, vol. 34, no. 15, pp. 4126–4137, 2006.
- [6] G. G. Oakley, "Replication protein A: directing traffic at the intersection of replication and repair," *Front. Biosci.*, vol. 15, no. 1, p. 883, 2010.
- [7] B. B. Aklilu *et al.*, "Functional diversification of replication protein A paralogs and telomere length maintenance in Arabidopsis," *Genetics*, vol. 215, no. 4, pp. 989–1002, 2020.
- [8] P. P. Pang and E. M. Meyerowitz, "Arabidopsis Thaliana: A model system for plant molecular biology," *Nat. Biotechnol.*, vol. 5, no. 11, pp. 1177–1181, 1987.
- [9] Y. Yu, L. A. Santat, and S. Choi, "Bioinformatics packages for sequence analysis," in *Applied Mycology and Biotechnology*, Elsevier, 2006, pp. 143–160.
- [10] P. Bawono, M. Dijkstra, W. Pirovano, A. Feenstra, S. Abeln, and J. Heringa, "Multiple sequence alignment," *Methods Mol. Biol.*, vol. 1525, pp. 167–189, 2017.
- [11] J. Gatesy, "Molecular evolution and phylogenetics (M. Nei and S. Kumar)," *Mol. Phylogenet. Evol.*, vol. 25, no. 3, pp. 567–568, 2002.
- [12] I. Letunic, T. Doerks, and P. Bork, "SMART 7: recent updates to the protein domain annotation resource," *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D302–5, 2012.
- [13] J. Mistry *et al.*, "Pfam: The protein families database in 2021," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D412–D419, 2021.
- [14] G. Maccarrone, J. J. Bonfiglio, S. Silberstein, C. W. Turck, and D. Martins-de-Souza, "Characterization of a protein interactome by co-immunoprecipitation and shotgun mass spectrometry," *Methods Mol. Biol.*, vol. 1546, pp. 223–234, 2017.
- [15] L. J. Jensen *et al.*, "STRING 8--a global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D412–6, 2009.
- [16] M. Kanehisa and Y. Sato, "KEGG Mapper for inferring cellular functions from protein sequences," *Protein Sci.*, vol. 29, no. 1, pp. 28–35, 2020.
- [17] Stryer, *Biochemistry 5ed*, 5th ed. New York, NY: W.H. Freeman, 2002.
- [18] Z. Zong, *Bioinformatics: (Volume 6, applied mycology and biotechnology)*. Elsevier Science & Technology, 2006.
- [19] J. E. Shim, J. H. Kim, J. Shin, J. E. Lee, and I. Lee, "Pathway-specific protein domains are predictive for human diseases," *PLoS Comput. Biol.*, vol. 15, no. 5, p. e1007052, 2019.
- [20] S. K. Miryala, A. Anbarasu, and S. Ramaiah, "Discerning molecular interactions: A comprehensive review on biomolecular interaction databases and network analysis tools," *Gene*, vol. 642, pp. 84–94, 2018.
- [21] T. Liu and J. Huang, "Replication protein A and more: single-stranded DNA-binding proteins in eukaryotic cells," *Acta Biochim. Biophys. Sin. (Shanghai)*, vol. 48, no. 7, pp. 665–670, 2016.
- [22] D. Szklarczyk *et al.*, "Correction to 'The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets,'" *Nucleic Acids Res.*, vol. 49, no. 18, p. 10800, 2021.
- [23] M. Biasini *et al.*, "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information," *Nucleic Acids Res.*, vol. 42, no. Web Server issue, pp. W252–8, 2014.
- [24] F. Kiefer, K. Arnold, M. Künzli, L. Bordoli, and T. Schwede, "The SWISS-MODEL Repository and associated resources," *Nucleic Acids Res.*, vol. 37, no. Database issue, pp.

- D387-92, 2009.
- [25] N. Guex, M. C. Peitsch, and T. Schwede, “Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective,” *Electrophoresis*, vol. 30 Suppl 1, no. S1, pp. S162-73, 2009.
- [26] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, “The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling,” *Bioinformatics*, vol. 22, no. 2, pp. 195–201, 2006.
- [27] S. R. Comeau, D. W. Gatchell, S. Vajda, and C. J. Camacho, “ClusPro: an automated docking and discrimination method for the prediction of protein complexes,” *Bioinformatics*, vol. 20, no. 1, pp. 45–50, 2004.
- [28] S. R. Comeau, D. W. Gatchell, S. Vajda, and C. J. Camacho, “ClusPro: a fully automated algorithm for protein-protein docking,” *Nucleic Acids Res.*, vol. 32, no. Web Server issue, pp. W96-9, 2004.
- [29] F. Sievers *et al.*, “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega,” *Mol. Syst. Biol.*, vol. 7, no. 1, p. 539, 2011.
- [30] N. Guex and M. C. Peitsch, “SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling,” *Electrophoresis*, vol. 18, no. 15, pp. 2714–2723, 1997.
- [31] J. Castresana, “Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis,” *Mol. Biol. Evol.*, vol. 17, no. 4, pp. 540–552, 2000.
- [32] A. Dereeper, S. Audic, J.-M. Claverie, and G. Blanc, “BLAST-EXPLORER helps you building datasets for phylogenetic analysis,” *BMC Evol. Biol.*, vol. 10, no. 1, p. 8, 2010.
- [33] A. Dereeper *et al.*, “Phylogeny.fr: robust phylogenetic analysis for the non-specialist,” *Nucleic Acids Res.*, vol. 36, no. Web Server issue, pp. W465-9, 2008.
- [34] P. Van Eijk, “Nucleotide excision repair in yeast,” Cardiff University, UK, 2021.
- [35] J. Castresana, “Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis,” *Mol. Biol. Evol.*, vol. 17, no. 4, pp. 540–552, 2000.
- [36] S. Guindon and O. Gascuel, “A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood,” *Syst. Biol.*, vol. 52, no. 5, pp. 696–704, 2003.
- [37] R. E. Rigsby and A. B. Parker, “Using the PyMOL application to reinforce visual understanding of protein structure: PyMOL Application to Understand Protein Structure,” *Biochem. Mol. Biol. Educ.*, vol. 44, no. 5, pp. 433–437, 2016.
- [38] J. Schultz, F. Milpetz, P. Bork, and C. P. Ponting, “SMART, a simple modular architecture research tool: identification of signaling domains,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 11, pp. 5857–5864, 1998.
- [39] I. Letunic, T. Doerks, and P. Bork, “SMART 7: recent updates to the protein domain annotation resource,” *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D302-5, 2012.
- [40] H. M. Berman *et al.*, “The Protein Data Bank, 1999–,” in *International Tables for Crystallography*, Chester, England: International Union of Crystallography, 2006, pp. 675–684.
- [41] M. Nielsen, C. Lundegaard, O. Lund, and T. N. Petersen, “CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles,” *Nucleic Acids Res.*, vol. 38, no. Web Server issue, pp. W576-81, 2010.
- [42] M. Hajduch *et al.*, “Systems analysis of seed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression,” *Plant Physiol.*, vol. 152, no. 4, pp. 2078–2087, 2010.
- [43] T. Igawa *et al.*, “Isolation and identification of ubiquitin-related proteins from Arabidopsis seedlings,” *J. Exp. Bot.*, vol. 60, no. 11, pp. 3067–3073, 2009.
- [44] N. H. Waseem, K. Labib, P. Nurse, and D. P. Lane, “Isolation and analysis of the fission yeast gene encoding polymerase delta accessory protein PCNA,” *EMBO J.*, vol. 11, no. 13, pp. 5111–5120, 1992.
- [45] A. Tkachenko *et al.*, “Transcriptomic Analysis of Radish (*Raphanus sativus* L.) Spontaneous Tumor,” *Plants*, vol. 10, no. 5, p. 919, 2021.

-
- [46] L. C. J. Gillet and O. D. Schärer, "Molecular mechanisms of mammalian global genome nucleotide excision repair," *Chem. Rev.*, vol. 106, no. 2, pp. 253–276, 2006.