ISSN 2744-1636
Bioengineering Studies *Original Research*
Vol. 4, No. 2, December 2023, pp.46-70
https://doi.org/10.37868/bes.v4i2.id315

In silico analysis of MAPKK5 variants in *Brassica Oleracea*

Majed Alwis*1
[1] Genetics and Bioengineering, International University of Sarajevo

*Corresponding author: majd2112001@gmail.com

Received Oct. 3, 2023
Revised Nov. 29, 2023
Accepted Dec. 30, 2023

Abstract

Brassica Oleracea is a well-known heavy metal accumulator and one of the most economically important crops. Besides its industrial usages, it could be utilized for phytoremediation of contaminated soil. In this study, we tried to analyze the four variants of MAPKK5 genes, mainly correlated to heavy metal stress, and to correlate their functions with other heavy metal proteins related to heavy metal absorption. This was done through a detailed *in silico* analysis, by using various bioinformatics tools like Clustal Omega, Phelogeny.fr, Phyre2, PyMOL, Loctree3, SMART, STRING, and Cluspro2. We found that MAPKK5 variants have high sequence similarity, resulting in their 3D structure similarity. The results indicate that MAPKK5 proteins interact with the same protein members and that all proteins are located in the cytoplasm. Eventually, when we concluded and confirmed that MAPKK5 proteins might be involved in the MAPK cascade process, regulating many cellular activities, demonstrating that mitogen-activated protein kinase cascades are quickly activated when a plant is attacked by pathogens.

Keywords: Brassica Oleracea; *in silico*; MAPKK; heavy metals; phytoremediation

1. Introduction

1.1. Brassica Oleracea

Brassica Oleracea is a morphologically diversified species in the *Brassicaceae* family that includes common headed *cabbage (B. Oleracea ssp. Capitata l.), cauliflower (B. Oleracea ssp. Botrytis l.), broccoli (B. Oleracea ssp. Italica l.), kohlrabi (B. Oleracea ssp. Gongylodes l.), kale (B. Oleracea ssp. Medullosa thell.),* and *brussels sprouts (B. Oleracea ssp. Gemmifera dc),* among other nutritious vegetable crops. This diversity, combined with its phylogenetic membership in a group of three diploid and three tetraploid species and the recent availability of genome sequences within *Brassica*, offers a once-in-a-lifetime opportunity to study intra- and interspecific divergence and evolution in this species and its relatives [1].

The *Brassicaceae* (*cruciferae*) family contains 350 genera and over 3500 species, including some economically important crops such as *Brassica napus L.*, *Brassica rapa L.,* and *Sinapis alba L*. These species are utilized as food, spices, and a source of vegetable oil. *Brassicaceae* vegetables are an essential element of the human diet and are enjoyed by people all over the world. They are important food crops in China, Japan, India, and Europe. *Brassica* production has significantly increased during the last three decades, becoming an important source of

plant-based oil and protein for animal and human nutrition, respectively. *Rapeseed* (*canola*) is currently the third most important source of vegetable oil (after *soy* and *palm*) and the third most important source of oil meal after *soy* and *cotton*. *Brassica* is a low-cost, high-nutritional-value vegetable that contains nutrients and health-promoting phytochemicals such as phenolic compounds, vitamins, phytic acid, fibre, soluble sugars, glucosinolates, minerals, polyphenols, lipids, and carotenoids. The food industry is currently extremely interested in identifying phytochemicals with helpful biological activity in food, and any important finding relating to the existence of valuable substances in *Brassica* species will be appreciated [2].

1.2. Heavy metals in plants

Wastes are the major source of soil pollution, which comes from the mining, chemical, and metal processing industries, as well as other related businesses. These wastes contain a wide range of compounds, including heavy metals, phenolic, organics, and non-metals. Heavy metals are metals and metalloids with atomic densities greater than 4 g/cm3, or 5 times or more that of water. These metals and metalloids include copper (Cu), manganese (Mn), lead (Pb), cadmium (Cd), nickel (Ni), cobalt (Co), iron (Fe), zinc (Zn), chromium (Cr), iron (Fe), arsenic (As), and silver (Ag). They are generally found in a distributed form in rock formations [3]. Increasing industrialization and urbanization increased the anthropogenic contribution of heavy metals in the biosphere, with the greatest availability in soil and aquatic ecosystems and a relatively lesser fraction in the atmosphere as particulate or vapours. Its toxicity in plants varies with plant species, individual metal, concentration, chemical form, soil composition, and PH, while many heavy metals are thought to be required for plant growth. Heavy metal toxicity refers to an excess of required concentrations or unwanted metals that were naturally found on the earth and became concentrated as a result of human-caused activities, enter plant, animal, and human tissues via inhalation, diet, and manual handling, and can bind to and interfere with the functioning of vital cellular components. For ecological, evolutionary, nutritional, and environmental reasons, heavy metal toxicity is a growing concern [4].

1.3. Known HMA (heavy metal ATPase) genes and their roles in metal accumulation

In terms of tissue distribution, subcellular localization, and metal specificity, P1B-type heavy metal ATPases (HMAs) are diverse. HMAs can be split into two groupings based on their metal-substrate specificity, according to functional studies: a copper (Cu)/silver (Ag) group and a zinc (Zn)/cobalt (Co)/cadmium (Cd)/lead (Pb) group [5]. Despite the fact that ions are involved in cell and chloroplast metabolism, little is known about ion transport across the chloroplast envelope. HMA1 was discovered to be a metal-transporting P1B-type ATPase utilizing a proteomic method specific to the *Arabidopsis* chloroplast membrane. HMA1 is primarily expressed in green tissues, and its chloroplast envelope localisation was confirmed. HMA1 is implicated in copper homeostasis, and deletion of its N-terminal His-domain affects metal transport to some extent, according to yeast expression tests. HMA1 *Arabidopsis* mutants were shown to have decreased chloroplast copper concentration and lower overall chloroplast superoxide dismutase activity. There was no effect on the plastocyanin content in these lines. In ordinary conditions, the HMA1 insertional mutants developed like WT plants, but showed a photosensitivity phenotype under high light. Finally, direct biochemical ATPase experiments on purified chloroplast envelope membranes revealed that copper exclusively stimulates HMA1 ATPase activity [6]. HMA2 and HMA4, the Zn/Cd-transporting ATPases required for root-to-shoot Zn translocation, may also transport Cd [7]. The plant *Arabidopsis thaliana* HMA2 is a Zn2+ transporting P1B-type ATPase that is necessary for plant metal homeostasis. The conserved N- and C-terminal sequences of HMA2 and all eukaryotic Zn2+-ATPases distinguish them from other P1B-type ATPases [8]. Plant Zn2+-ATPases differ from P1B-type ATPases in that they contain extended C-terminal sequences rich in Cys and His [8]. The HMA4 gene in *Arabidopsis halleri* is required for Zn/Cd hypertolerance and hyperaccumulation by regulating metal transfer from root to shoot [9]. Furthermore, AtHMA4 in *Arabidopsis thaliana* is a P1B-type ATPase that belongs to the Zn/Cd/Pb/Co subgroup. Previously, heterologous expression and the analysis of AtHMA4 knockout or overexpressing lines indicated that AtHMA4 is involved in zinc homeostasis and cadmium tolerance in *Arabidopsis* [10]. AtHMA3 in *Arabidopsis thaliana* is a protein that belongs to the P1B-2 subgroup of the P-type ATPase family, which is

involved in heavy metal transport. AtHMA3 most likely contributes to the detoxification of biological (Zn) and non-biological (Cd, cobalt, and Pb) heavy metals by engaging in their vacuolar sequestration, which is an unusual function for a P1B-2 ATPase in a multicellular eukaryote [11].

Table 1. Heavy Metal ATPase gene (HMA1, HMA2, HMA3, HMA4, and HMA5) transporters in Different Plants [12]

Gene	Plant	Metal	References
HMA 1	*Arabidopsis (chloroplast membrane)*	HMA1 plays a role in copper homeostasis and deleting its N-terminal His domain influences metal transport.	[13]
HMA 2	*Arabidopsis thaliana*	The N and C-terminal domains of HMA2 have a strong affinity for Zn ions.	[7] [14]
	Hordeum vulgare		
	Oryza sativa	During the vegetative stage, HMA2 is a carrier of Zn and Cd from root to shoot and selectively transports Zn and Cd to the grain during the reproductive stage.	[15]
	Triticum aestivum	Overexpression of TaHMA2 promoted root-shoot Cd and Zn translocation in rice.	[16]
HMA 3	*Glycinax*	Cd levels at high speeds. Glycine is required for GmA3 function, and seed Cd levels are low.	[17]
	Arabidopsis thaliana	To reduce Cd transfer to the xylem, Cd is sequestered in vacuoles.	[11]
	Oryza sativa	Cd amounts in grains and shoots are high.	[18]
	O. sativa	Low amounts of Cd in grains and shoots.	[19]
HMA 4	*Arabidopsis. thaliana (Wassilewskija and Columbia)*	In Col-0, HMA4 expression was induced.	[20] [21]
	Arabidopsis. thaliana (Columbia)	AtHMA4's C-terminal domain functions as a Zn2+ and Cd2+ sensor as well as a regulator for efficient Zn2+ and Cd2+ export.	
	Thlaspi caerulescens	In the case of excessive Cd tolerance, this protein plays a crucial function. Cd tolerance appears to be due to HMA4's effective translocation to the shoots and decreased absorption.	[22]
	Populus trichocarpa	HMA4 reduces some of the harmful effects of Zn during the early stages of development.	[23]
	Lycopersicon esculentum	Transgenic and wild-type plants were subjected to low and high Zn levels.	[24]
	Noccaea caerulescens	At the root, there is a Zn and Cd tolerance gene.	[25]
HMA5	*Arabidopsis thaliana*	Cu activates this gene, which is mostly expressed in the roots and is strongly and specifically activated in complete plants.	[26]

Because copper (Cu) is required in important physiological oxidation activities, organisms have evolved mechanisms for dealing with Cu while avoiding its potentially harmful effects. A network of Cu homeostasis factors, including Cu-transporting P-type ATPase, which play a vital role in transmembrane Cu transport, is one of the instruments that have developed to deal with Cu. Furthermore, we describe the functional characterisation of an *Arabidopsis* Cu-transporting P-type ATPase known as heavy metal ATPase 5 (HMA5), as well as its interaction with *Arabidopsis metallochaperones*. HMA5 is expressed largely in the roots and is strongly and specifically activated by Cu in the entire plant. We discovered and analysed plants with two distinct T-DNA insertion alleles, HMA5-1 and HMA5-2. Both mutants are hypersensitive to copper but not to iron, zinc, or cadmium. Root tips from Cu-treated HMA5 mutants exhibit a wavelike pattern early on, and later, main root growth entirely stops, although lateral roots sprout near the crown. As a result, under Cu excess, these lines deposit more Cu in their roots than wild-type plants. Moreover, yeast two-hybrid experiments show that the metal-binding domains of HMA5 interact with *Arabidopsis* ATX1-like Cu chaperones, implying a regulatory role for the CCH Cu chaperone's plant-specific domain. Based on our findings, we propose that HMA5 has a role in Cu compartmentalization and detoxification [26].

1.3.1. MAPK gene Expression in the Brassica family

Plants are frequently subjected to a variety of biotic and abiotic stressors in nature, such as pathogen infection, drought, salt, cold, and oxidative stress. These unfavourable circumstances have a negative impact on plant development and production, resulting in a significant loss of agricultural production. Plants have evolved a complex network of signalling networks and regulatory systems that provide protection at the molecular, cellular, organ, plant, and population levels [27]. The cell division, development, hormone signalling, and biotic and abiotic stress responses in *Brassica* family plants are all controlled by mitogen-activated protein kinase (MAPK) cascades [28].

MAPK cascades relay and amplify signals via three types of reversibly phosphorylated kinases, resulting in the phosphorylation of substrate proteins, the changed activities of which influence a wide range of responses, including changes in gene expression. MAPKK kinases (MAPKKKs, MAP3Ks, or MEKKs), MAPK kinase (MAPKKs, MAP2Ks, MKK or MEKs), and MAPK/MPKs are the types of MAPK cascade. Although these Cascades share kinase components, their signalling specificity is maintained by spatiotemporal limitations and dynamic protein-protein interactions, as well as processes including as cross inhibition, feedback regulation, and scaffolding. Stress and hormonal responses, innate immunity, and developmental programs are all regulated by plant MAPK cascades [29].

1.3.2. MAPK cascade

MAPK (Mitogen-activated protein kinase) cascades are critical signalling pathways that control a wide range of biological functions [30]. The activation of extracellular signal-regulated kinases 1 and 2 by the mitogen-activated protein kinase cascade is important for the differentiation of some cell types and the proliferation of others. However, several recent studies have linked this cascade to the adult brain's control of synaptic plasticity. Extracellular signal-regulated kinase signalling appears to be required for known neuronal transcriptional activities and may potentially adjust synaptic targets to control plasticity. Another scenario that has recently emerged is the involvement of a parallel but distinct kinase cascade that leads to the activation of p38 MAPK, which may control different types of synaptic plasticity [31].

1.3.3. MAPK cascade in *Brassica napus*

Brassica napus is one of China's most important oil crops, accounting for more than half of all edible oil. However, various stressors, such as cold, salt, and drought [32]. Plant growth and development, as well as biotic and abiotic stress responses, rely on MAPK-mediated cellular signalling. Various conditions created a danger to the production of *oilseed rape*, prompting scientists to increase resilience using transgenic technology, using MAPK as the target gene. Several physiological indicators were assessed in the current work, which used transgenic lines overexpressing *Brassica napus* MAPK1 as the major experimental materials. The expression

of associated genes involved in drought resistance was studied during natural drought and artificial stress mimicked with PEG8000 to determine the role of *Brassica napus* MAPK1 in increasing drought tolerance of *Brassica napus*. The results showed that *Brassica napus* over-expressing BnMAPK1 had increased leaf water content, which might sustain metabolic capacity and reduce drought stress damage [32].

1.3.4. MAPK cascade in Brassica Rapa

Brassica rapa MAPK cascade gene retention, phylogeny, evolution, and expression patterns were investigated to identify how the MAPK cascade genes are responsive to various biotic and abiotic stresses [33]. Furthermore, the MAPK signalling cascade leads to the synthesis of camalexin, which induces defensive responses in *B.rapa.* While exposed to the most devastating necrotrophic fungus [34].

1.4. Phytoremediation

Polluted soil and water have an impact on the quality of food and nutrients available to human and animal biota. Soil and water are mostly polluted by industrial effluent discharges, which are generically classed as metallic and non-metallic pollutant-bearing effluent. To resolve this issue, a plant-based approach known as phytoremediation is employed to clear contaminated soils. Phytoremediation relies on a number of processes, including phytodegradation, phytovolatilization, phytoaccumulation, and phytoextraction. These techniques are effective, environmentally friendly, and cost-effective [35].

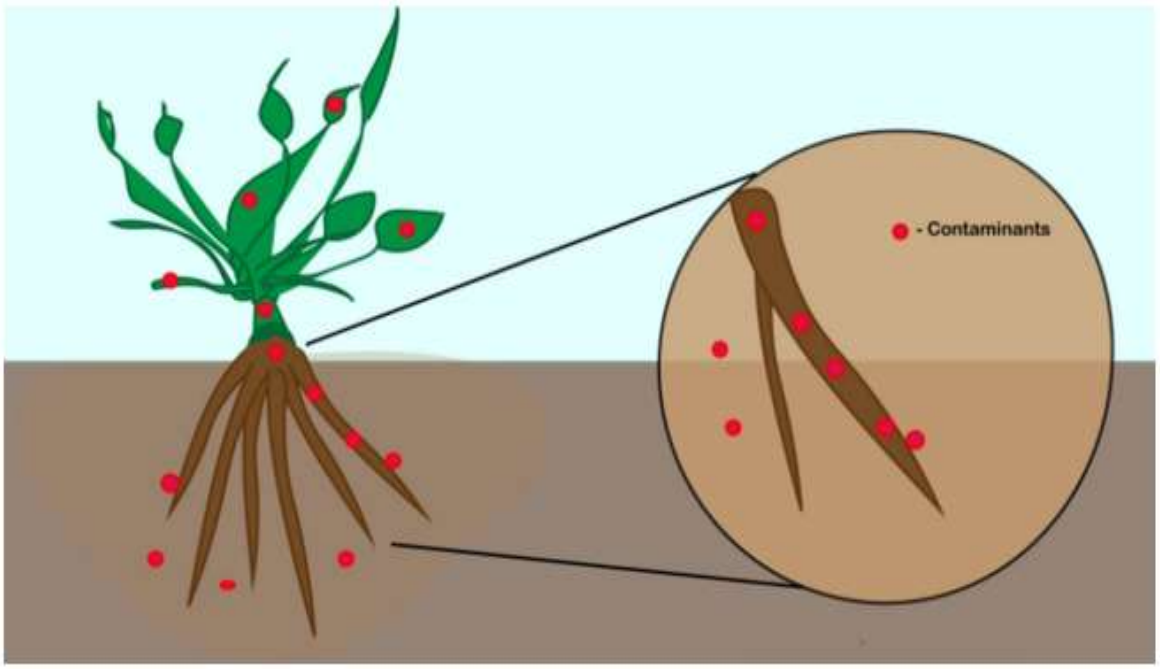


Figure 1. Phytoaccumulation mechanism, Contaminants, which appear as red spots, are absorbed by plants and accumulate in shoots, leaves, and other plant parts [35].

A pot experiment was carried out to screen four different *Brassicaceae* family species (*Brassica juncea, Brassica napus, Brassica Rapa, and Brassica campestris*) for the remediation of HMs-polluted soil in Lakki Marwat city, Pakistan, irrigated with municipal wastewater. Plants were examined for morpho-physiological, biochemical, and phytoextraction variables such as bioaccumulation and translocation factor. The results showed that *B. juncea* had the best morpho-physiological responses in terms of seed germination, chlorophyll content, root fresh and dry weights, and shoot fresh and dry weights, followed by *B. napus, B. campestris*, and *B. rapa*. Plant biochemical tests of antioxidant enzymes such as superoxide dismutase, peroxidase, and catalase revealed that *B. juncea* had the highest activity in both control and contaminated soils, followed by *B. napus, B. campestris, and B. rapa* [36].

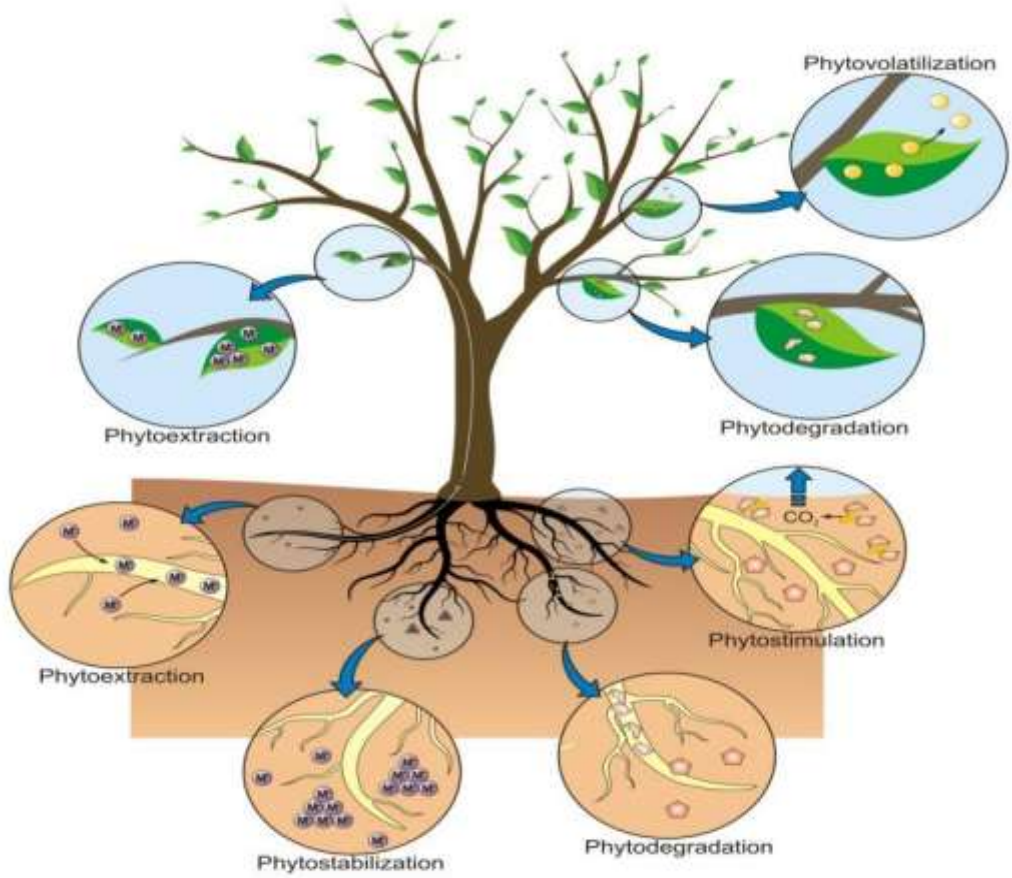


Figure 2. Phytoremediation techniques are divided into six modalities, based on the chemical composition and properties of the pollutant [37].

1.5. *In Silico* studies – Advantages of bioinformatics

Bioinformatics, often known as *in silico* biology, is a rapidly expanding area that involves the theory and use of computer approaches to modelling, predicting, and explaining biological activity at the molecular level. To take advantage of the fast-rising amount of sequence, expression, and structure information in public and private databases, this information-rich field needs new abilities and a new understanding of genome-scale studies. Toxicologists are preparing to make use of massive public databases in order to decipher the molecular basis of toxicity. With the introduction of high-throughput sequencing and computational approaches, expressed sequences in target tissues may be quickly discovered and quantified by database searching. Novel genes can also be extracted *in silico,* and their function predicted and defined using sequence homology to other known proteins. Based on single-nucleotide polymorphism data, genomic DNA sequence data can be used to identify target genes and their modes of regulation, as well as identify susceptible genotypes. Furthermore, high parallel gene expression profiling technologies will enable toxicologists to mine enormous databases of gene expression data for molecular biomarkers and other diagnostic and prognostic genes or expression patterns [38]. In addition, the *in silico* technique has numerous benefits. These are (i)the versatility that results from independence from specific databases/tools or software, (ii)relative algorithm simplicity (ability to avoid complex computational/statistical methods), and (iii)applicability to molecular cytogenetic data because of the chromosome-centric nature [39].

1.6. Aim of the study

Using an *in silico* approach, we aim to understand the evolutionary history of our target genes, predict their 3D structures, and analyse their protein sequences for localization, domains, and interactomes. We concentrated our efforts in this study on further researching the roles of four variants of the MAPKK5 genes in *Brassica*

Oleracea. The stated goal is achieved by the usage of numerous common bioinformatics approaches in order to reach the most accurate result.

2. Materials and methods

2.1. Retrieving MAPKK5 sequence and Multiple Sequence Alignment

The National Center for Biotechnology Information (NCBI) was used to obtain the sequences of mitogen-activated protein kinase 5-like. The accession numbers for the sequences are shown in the Table2.

Table 2. Accession numbers for the four variants of MAPKK5

MAPKK5 variants	NCBI Reference Sequence	Protein sequence ID (NCBI)
MAPKK5x1	XM_013773841.1	XP_013629295.1
MAPKK5x2	XM_013773842.1	XP_013629295.1.1
MAPKK5x3	XM_013773843.1	VDC94097.1
MAPKK5x4	XM_013773844.1	VDD01053.1

Clustal programs are commonly used for making automatic multiple sequence alignment of nucleotide or amino acid sequences [40]. Clustal Omega is a completely updated and upgraded version of the widely used Clustal series of programs for multiple sequence alignment that was released in 2011 [41]. In addition, MSA was performed using Clustal Omega via the European Bioinformatics Institute (EBI) website with default parameters.

2.2. Phylogenetic tree construction

Phylogenetic analyses commonly incorporate the determination of homologous successions, different arrangements of those groupings, phylogenetic reconstruction, and graphical representation of the estimated tree [42]. After analyzing multiple sequence alignments, a phylogenetic tree was constructed in order to represent the evolutionary relationships among the four variants of mitogen-activated protein kinase 5-like. Phylogenetic trees are the most fundamental visual representations in evolutionary biology and are frequently employed in the biological sciences [43]. The tree construction was performed using Phlogeny. fr, an online platform designed for both non-specialists and experts. All settings are configured to suit the majority of research, and users simply need to supply their input sequences to acquire a ready-to-print tree. The use of a computerized method minimizes the need for manually revising many alignments to some extent, allows for the automation of phylogenetic analysis of enormous data sets, and promotes the reproduction of the final alignment by other scholars [44].

2.3. 3D structure prediction and validation

The prediction of three-dimensional protein structures from amino acid sequences has become a valuable tool in computational structural biology [45]. Three-dimensional protein structures provide essential insights into the molecular basis of protein activity, enabling more effective experiment design, such as site-directed mutagenesis, disease-related mutation studies, or the structure-based design of particular inhibitors [46]. In this study, the 3D structure of four MAPKK5 variants was predicted *in silico* using Phyre2 protein homology modelling server. This is a web-based tool that predicts protein structure, ligand-binding sites, and the effects

of amino acid variations using remote homology detection techniques. Homology modeling's major goal is to predict a structure from its sequence with an accuracy that is similar to the best experimental results [47].

When experimental structures are unavailable, homology modelling is used to create 3D models of proteins. Many proteins are simply too big for NMR analysis and X-ray diffraction [48].

Additionally, a practical and widely cited molecular visualization tool, PyMOL, was used for structure visualization and validation. PyMOL, a cross-platform molecular graphics tool, has been widely used for three-dimensional (3D) visualization of proteins, nucleic acids, small molecules, electron densities, surfaces, and trajectories [49]. The validation included Verify3D analysis, based on evaluating the compatibility of a protein's 3D structure with its corresponding amino acid sequence. The scores range from -1 (bad score) to +1 (good score)[50]. Next, we used Ramachandran plots, which show the dihedral angles phi and psi of amino acid residue positions in the protein structure.

2.4. Localization of proteins and identification of protein domains

The computational prediction of a protein's subcellular localization is a critical step in determining and explaining its function and interactions with other proteins in the cell. We employed the newly developed tool LocTree3 for subcellular localization.

Furthermore, the identification of domains in different variants of mitogen-activated protein kinase 5-like was performed using the Simple Modular Architecture Research Tool (SMART), available on the European Molecular Biology Laboratory (EMBL) website. SMART can identify and annotate various mobile domains, as well as do further domain architectural analysis. The software can identify over 500 domain families in chromatin-associated, extracellular, and signaling proteins. These domains are thoroughly interpreted in terms of functional class, significant residues, and tertiary structures [51] [52].

2.5. Interactome prediction

Interactome of the mitogen-activated protein kinase 5-like was determined by using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins). The primary goal of the STRING database is to display and offer information about specific protein-protein interactions, including both physical (direct) and functional (indirect) interactions. The STRING database incorporates data from four independent sources (genomic context, published studies, co-expression, and high-throughput experiments) to produce accurate results [53].

2.6. Prediction of docking sites

Computational docking is a set of methods that predicts the placement and binding affinity of a ligand (protein) in the binding site of another protein. Docking methods are mostly dependent on specialized algorithms that are responsible for determining the position of the ligand in the active site, as well as calculating the scoring function, which shows the binding affinity (the strength of the contact between the ligand and the receptor)[54]. For docking prediction of mitogen-activated protein kinase 5-like with some other chosen proteins, an online protein-protein docking tool was used, and it is ClusPro 2.0 [55]. It is a fully automated docking server that evaluates billions of protein conformations. ClusPro generates 70 000 rotations for the ligand protein, of which 1000 have the lowest score, ultimately clustering them based on C-alpha RMSD radius of nine angstroms. The largest cluster represents the most probable docking conformation [56].

3. Results

3.1. Multiple sequence alignment

Multiple sequence alignment was performed by the Clustal Omega online tool. The results of the sequence alignment are in the following Table 3.

Table 3. Percent identity score of MAPKK5 proteins calculated by Clustal Omega

Sequence A	Name	Length(AA)	Sequence B	Name	Length(AA)	Clustal Omega score (%)
1	MAPKK5X4	353	2	MAPKK5X3	331	77.44
1	MAPKK5X4	353	3	MAPKK5X1	319	73.04
1	MAPKK5X4	353	4	MAPKK5X2	390	74.16
2	MAPKK5X3	331	3	MAPKK5X1	319	88.71
2	MAPKK5X3	331	4	MAPKK5X2	390	95.30
3	MAPKK5X1	319	4	MAPKK5X2	390	92.48

Comparing each sequence with all the remaining three from MAPKK5 sequences and generating a percent identity matrix, which shows different scores that present similarity between sequences. Thus, the biggest similarity is noted between MAPKK5x3 and MAPKK5x4 homologues, followed by the score of the alignment between MAPKK5x1and MAPKK5x2.

Figure 3. MSA visualization of MAPKK5 proteins retrieved from Clustal Omega (some conserved regions are shown)

3.2. Phylogenetic tree construction

The phylogenetic tree for the four proteins was constructed using Phylogeny.fr, and the cladogram of the results is visible in the following Figure 4:

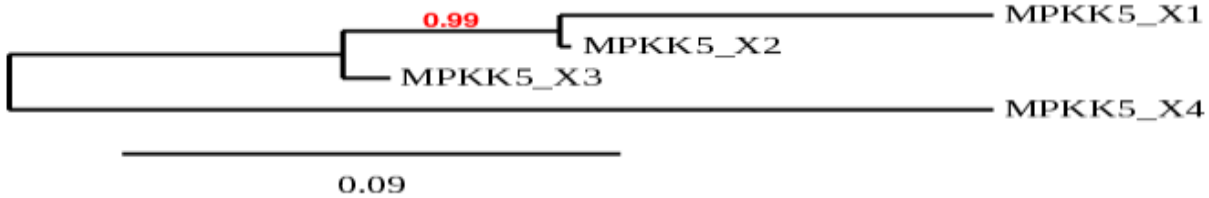


Figure 4. Phylogenetic tree of MAPKK5 proteins in Brassica Oleracea

A cladogram is a visualization that does not take the element of time (genetic mutations accumulated) into account during the building of the tree. As is visible from the tree, MAPKK5x1 and MAPKK5x4 are most similar, and MAPKK5x2 and MAPKK5x3 are on the other branch, with the complete tree having two separate branches, as is expected from the MSA results previously presented.

3.3. 3D structure visualization and validation

Protein structure determination is critical for a complete understanding of the function, interactions, and potential ligands, conserved domains and their homologues, and many other applications. However, because determining the 3D structure experimentally is difficult and time-consuming, bioinformatics tools like these are employed to estimate the structures *in silico*. As shown below in Figure 5, Figure 6, Figure 7, and Figure 8, 3D structures of MAPKK5 proteins are predicted by the Phyre2 tool and visualized by PyMOL as well as Ramachandran plot diagrams for all three structures. In the Ramachandran plot, the green dots represent the highly preferred regions, the orange dots represent the preferred regions, and the red dots represent the Questionable regions.

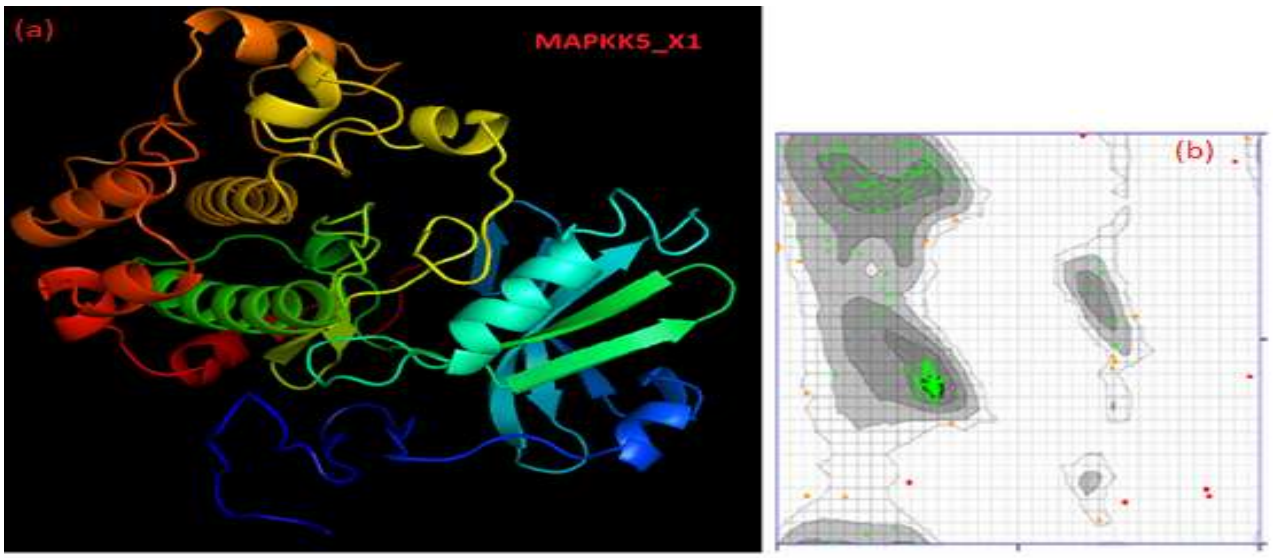


Figure 5. Predicted 3D structures (a) and respective Ramachandran plots for MAPKK5x1 (b)

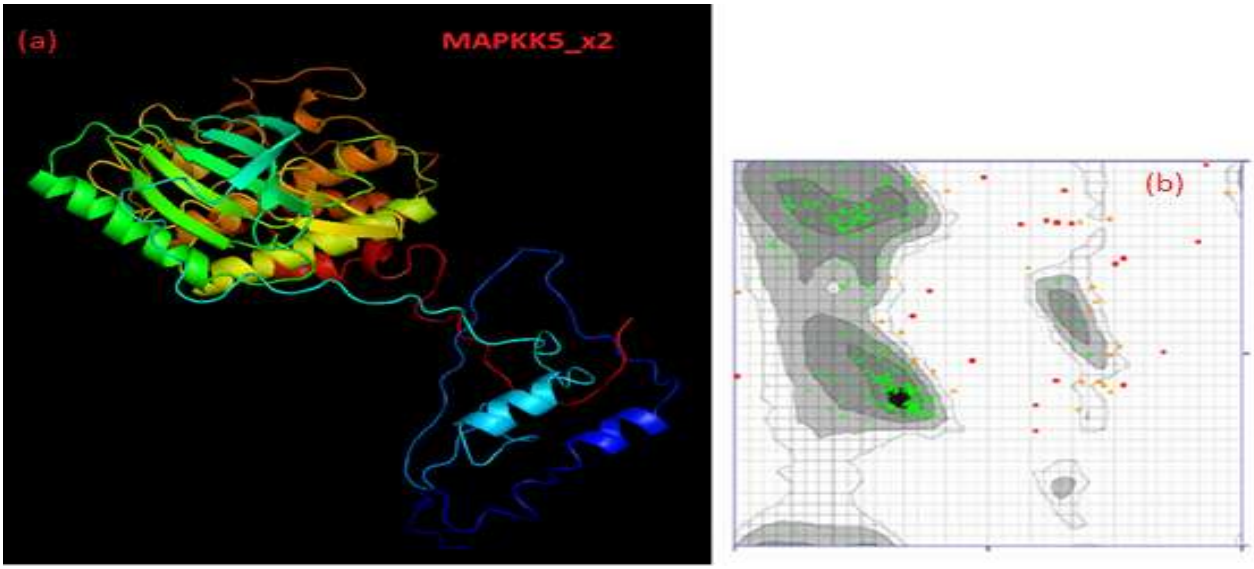


Figure 6. Predicted 3D structures (a) and respective Ramachandran plots for MAPKK5x2 (b)

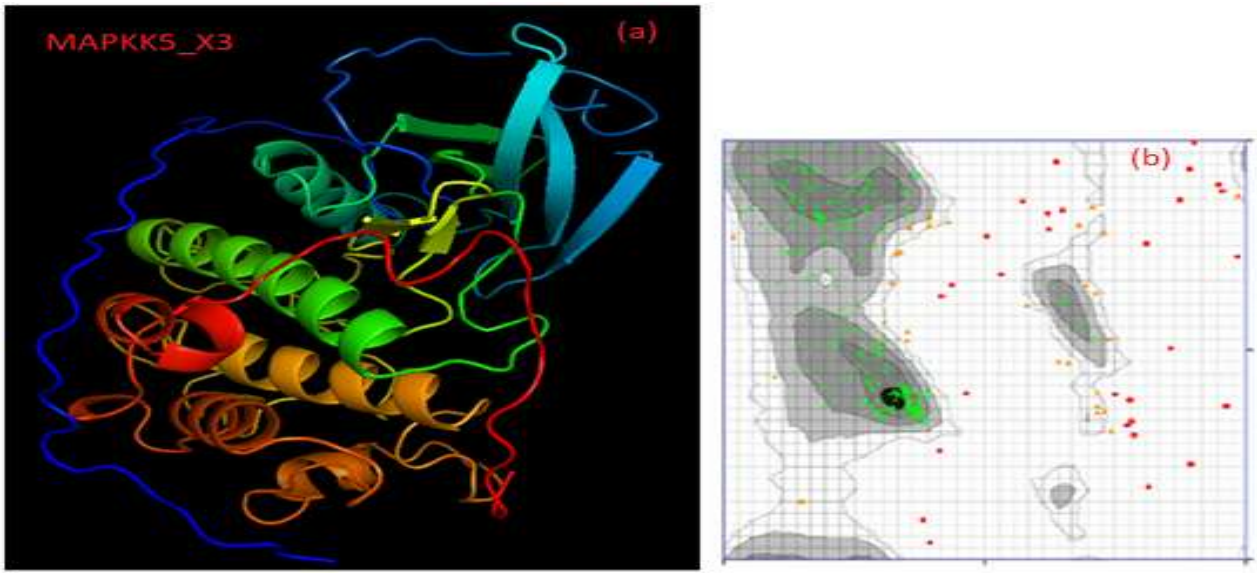


Figure 7. Predicted 3D structures (a) and respective Ramachandran plots for MAPKK5x3 (b)

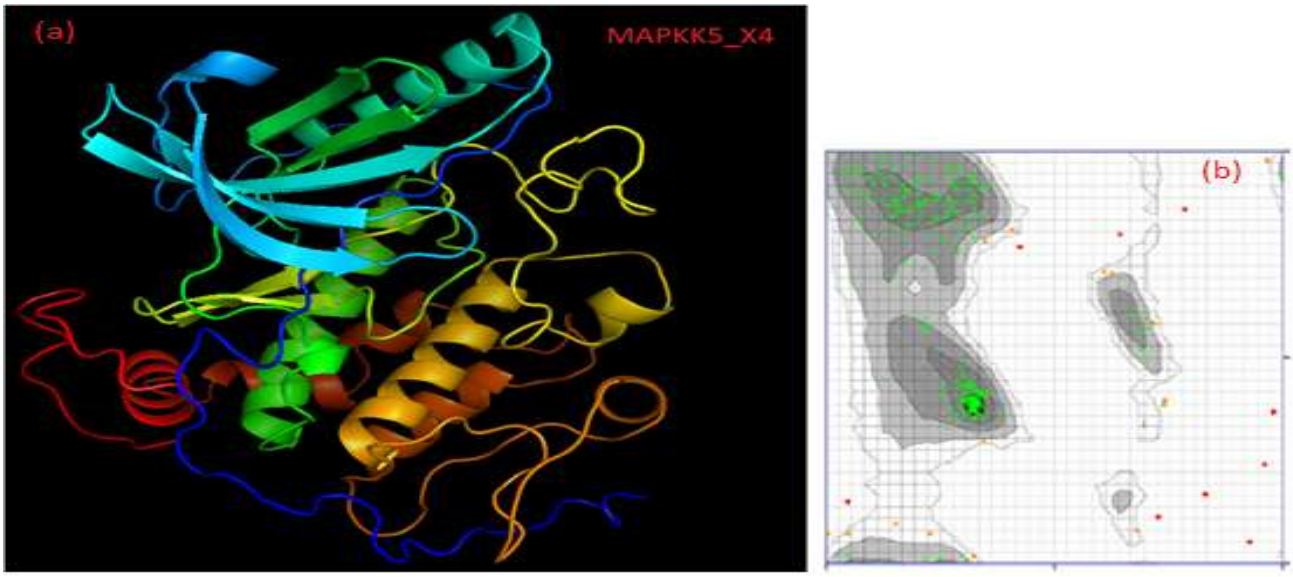


Figure 8. Predicted 3D structures (a) and respective Ramachandran plots for MAPKK5x4 (b)

After prediction, the 3D structures were validated by structure assessment tools. The findings are shown in the Table 4.

Table 4. Validation of predicted structures of the MAPKK5 protein

Protein Name	**Verify 3D score (%)**	**Ramachandran plot – favourable region**
MAPKK5x1	82.76% (0.8276)	91.513%
MAPKK5X2	60.26% (0.6026)	87.537%
MAPKK5X3	78.85% (0.7885)	79.715%
MAPKK5X4	86.97% (0.8697)	92%

3.4. Localization of protein

By loctree3, the proteins are localized to 18 possible locations, with a score from 0 to 100. The higher the score, the higher the confidence in the presence of the protein in a particular subcellular compartment. Results are shown for protein localization prediction in Table 5.

Table 5. Results of protein localization by Loctree 3

Protein ID	Score	Expected Accuracy	Localization Class
MAPKK5X1	20	83%	cytoplasm
MAPKK5X2	18	82%	cytoplasm
MAPKK5X3	18	82%	cytoplasm
MAPKK5X4	20	83%	cytoplasm

3.5. Domain identification

The domains of MAPKK5x1, MAPKK5x2, MAPKK5x3, and MAPKK5x4 were successfully identified by SMART software, and the results showed that some of the domains are shared between proteins. The domains are listed in Table 6, and the respective domains are shown in four variants of MAPKK5 proteins in Figure 9.

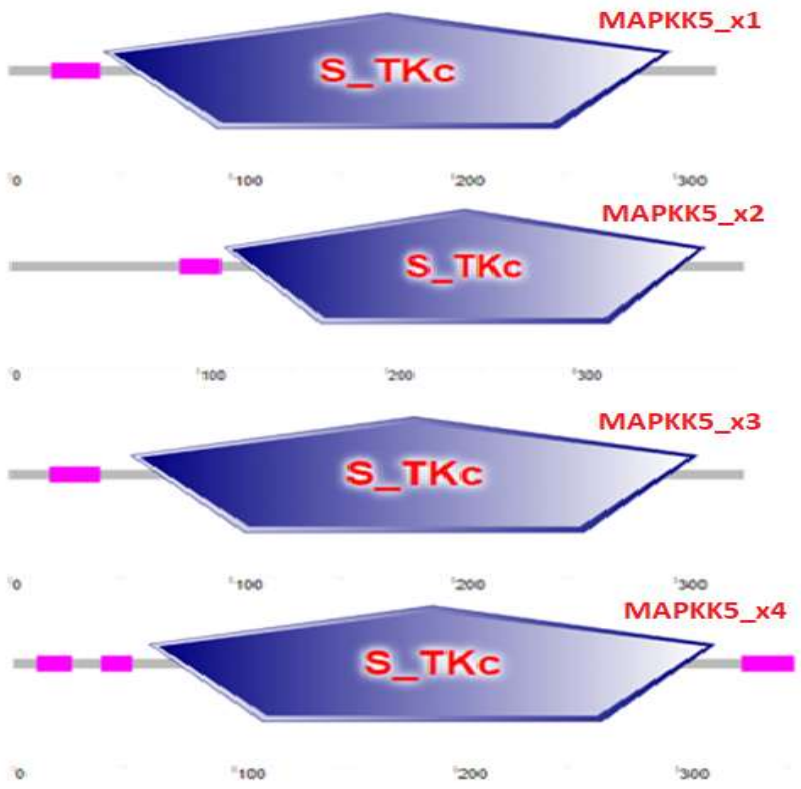


Figure 9. Diagram representations of domains of four variants of MAPKK5 by SMART

Table 6. Positions of identified domains in four variants of MAPKK5 proteins

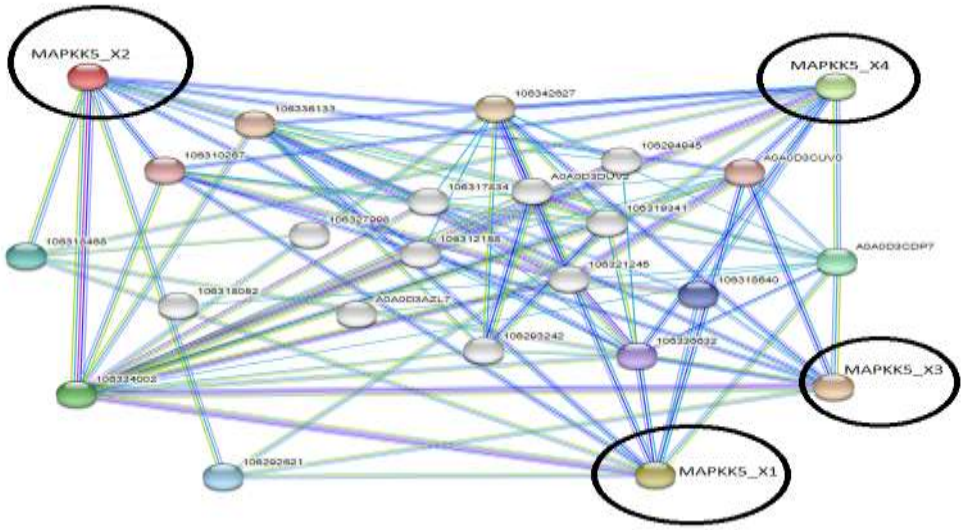
Protein name	Domains					
	S_TKc			Low complexity region		
	Start	End	E-value	Start	End	E-value
MAPKK5x1	43	298	1.32e-87	19	41	/
MAPKK5x2	114	369	1.22e-86	90	112	/
MAPKK5x3	55	310	7.09e-88	18	41	/
MAPKK5x4	62	317	6.99e-87	11	27	/
				40	54	/
				329	353	/

As is visible in Table 6, the proteins share only one domain, with different low complexity regions between them, starting from different residues and with a different number of residues in the domains. The domain is Serine/Threonine Protein Kinases, Catalytic Domains (S_TKc).

3.6. Interactome of MAPKK5 proteins

*The interacting proteins are classed as either physical or functional. We must define the organism by looking for our sequence before entering it in FASTA format into STRING. There are 10 proteins found interacting with MAPKK5 proteins (**Error! Reference source not found.**.). All the interacting proteins show high confidence scores. The results are presented in*

Table *7*.



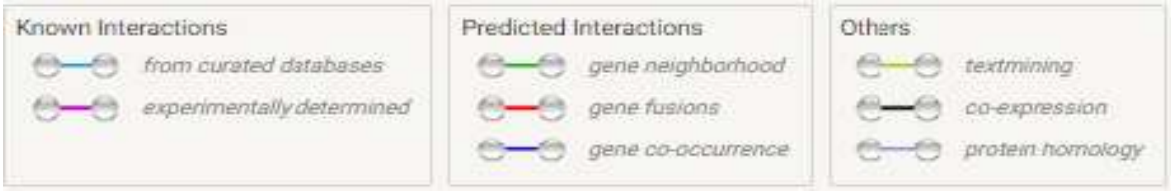


Figure 10. STRING interactome of Brassica Oleracea MAPKK5 four variants

Table 7. Interactome results of Brassica Oleracea MAPKK5 variants

ID	Name	Function	Score	References
106334002	Mitogen-activated protein kinase 3	ATP (Adenosine triphosphate) binding MAP kinase activity protein serine/threonine/tyrosine kinase activity	0.986	[57]
A0A0D3CDP7	Protein kinase domain-containing protein	ATP binding source protein kinase activity	0.885	[58]
106318468	Mitogen-activated protein kinase kinase YODA	ATP binding protein kinase activity	0.835	[59]
106292621	mitogen-activated protein kinase kinase kinase 3	ATP binding protein kinase activity	0.827	[60] [61]
106318640	Mitogen-activated protein kinase kinase 1-like isoform X1	ATP binding protein kinase activity	0.800	[62]
106336632	Mitogen-activated protein kinase homolog NTF4-like	ATP binding protein serine/threonine/tyrosine kinase activity	0.716	[63]
106310267	Mitogen-activated protein kinase homolog MMK1-like	ATP binding protein serine/threonine/tyrosine kinase activity	0.716	
A0A0D3CUV0	Mitogen-activated protein kinase homolog D5-like	ATP binding protein serine/threonine/tyrosine kinase activity	0.713	[58]

106336133	mitogen-activated protein kinase 6-like	ATP binding MAP kinase activity protein serine/threonine/tyrosine kinase activity	0.712	
106342627	Mitogen-activated protein kinase 6	ATP binding MAP kinase activity protein serine/threonine/tyrosine kinase activity	0.710	[57]

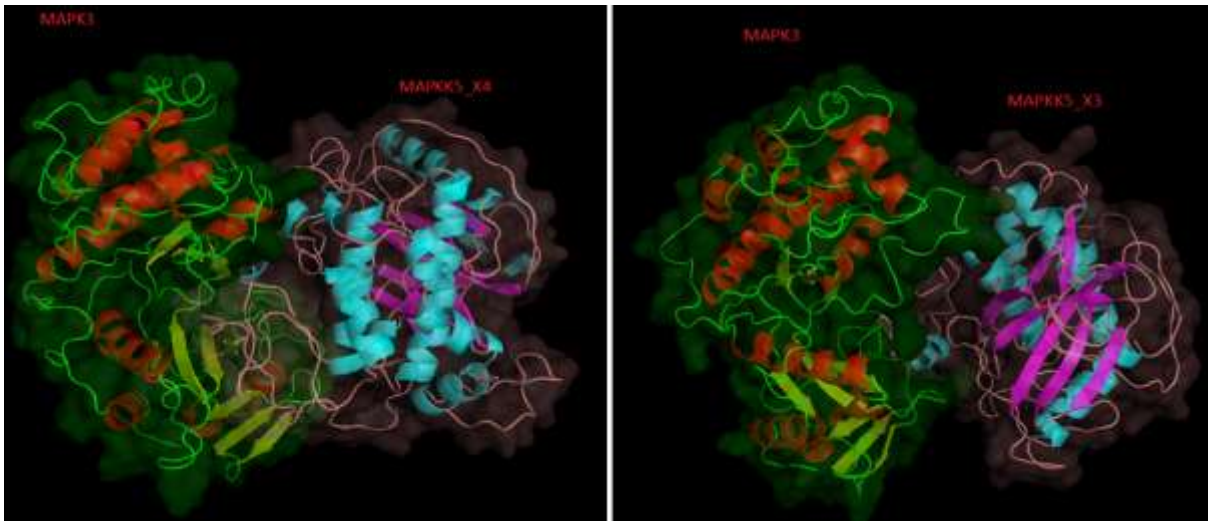
3.7. Docking site prediction

Using the STRING database, only four of the interacting proteins were discovered, and they were chosen for docking investigation. The proteins that were picked all have a high interaction score. As shown in Table 8 , those 4 proteins are presented. The models were retrieved from phyre2, and in the same table, we can see validation scores for the retrieved 3D model.

Table 8. Interacting proteins 3D models validation scores

Protein Name	Verify 3D score (%)
Mitogen-activated protein kinase 3	80.27% (0.8027)
Mitogen-activated protein kinase kinase Yoda	60.23% (0.6023)
Protein kinase domain-containing protein	56.15% (0.5615)
Mitogen-activated protein kinase kinase kinase 3	55.13% (0.5513)

All four proteins were picked for docking with the four varieties of MAPKK5, and those are Protein kinase domain containing protein, mitogen-activated protein kinase 3 (MAPK3), Mitogen-activated protein kinase kinase 3(MAPKKK3), and Mitogen-activated protein kinase kinase Yoda (MAPKKK Yoda). When docking analysis was performed in ClusPro2. In Figure 11, Figure 12, , Figure 14 docking results (3D models) for all four MAPKK5 variants are presented.



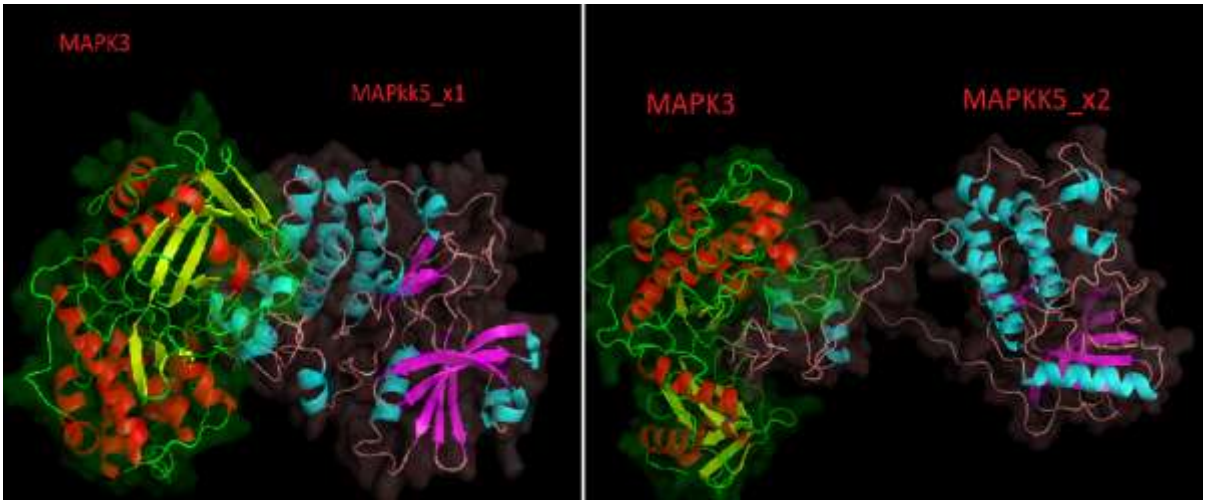


Figure 11. Docking site prediction of four variants of Mitogen-activated protein kinase 5 with Mitogen-activated protein kinase 3 from *Brassica Oleracea*

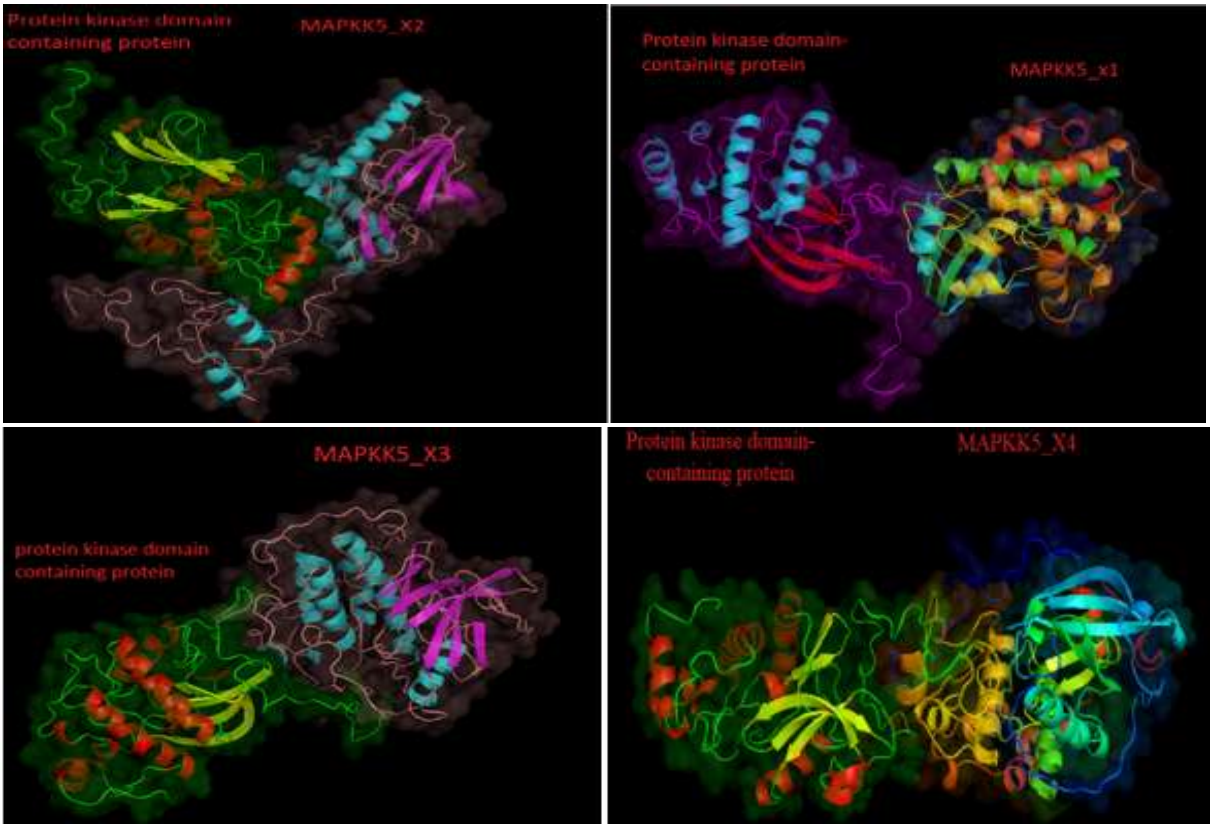


Figure 12. Docking site prediction of four variants of Mitogen-activated protein kinase 5 with protein kinase domain containing protein from *Brassica Oleracea*

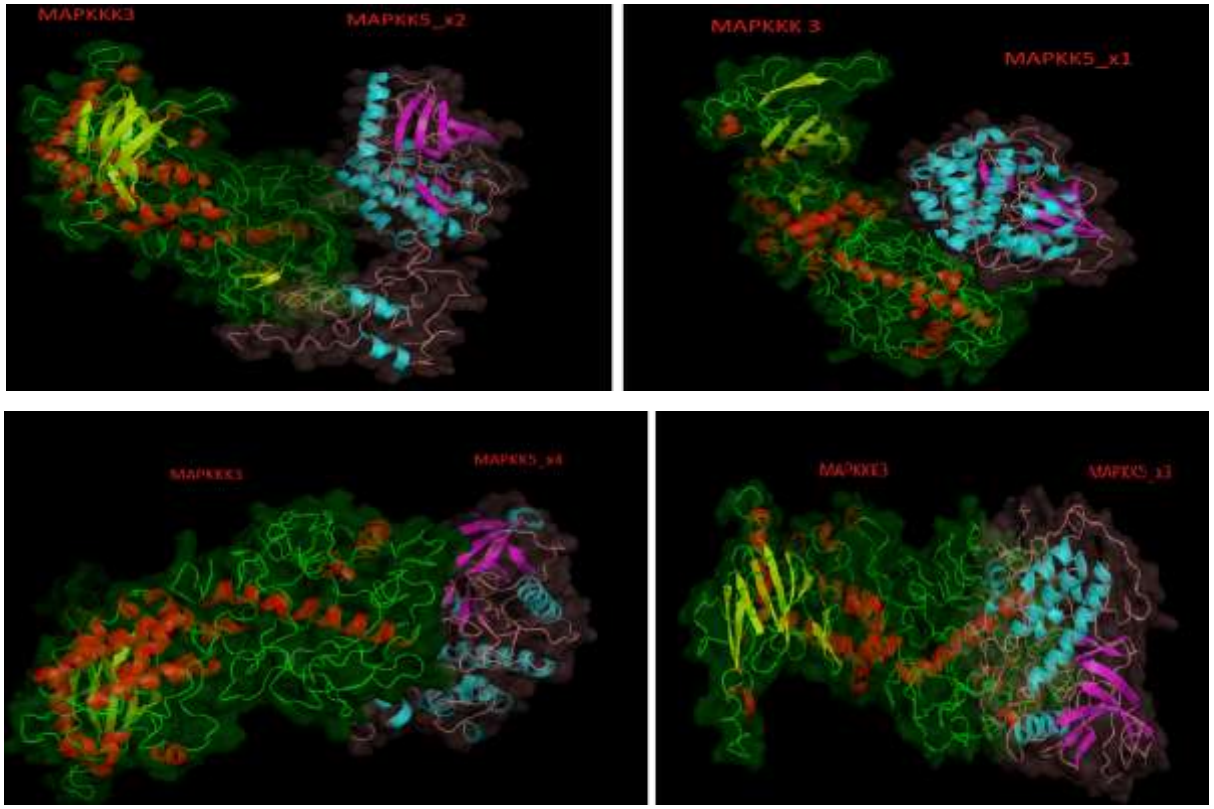


Figure 13. Docking site prediction of four variants of Mitogen-activated protein kinase 5 with Mitogen-activated protein kinase 3 from Brassica Oleracea

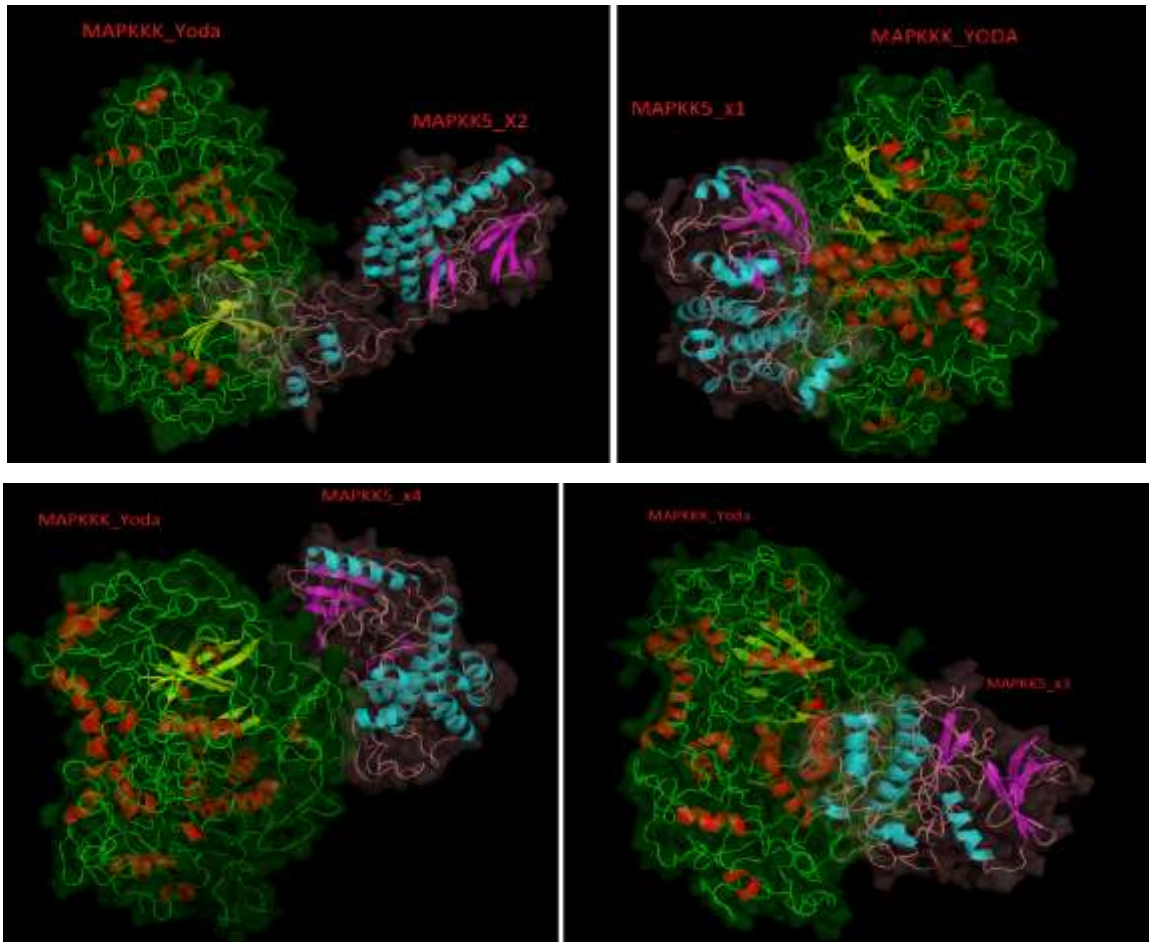


Figure 14. Docking site prediction of four variants of Mitogen-activated protein kinase 5 with Mitogen activated protein kinase Yoda from *Brassica Oleracea*

4. Discussion

Before performing MSA, the accession numbers and the sequences of our four variants of MAPKK5 gens were successfully retrieved as well as their protein sequences. From the sequences it was concluded that MAPKK5x2 has the longest amino acid sequences among the other three variants with 390 amino acids, then comes MAPKK5x4 with 353 amino acids, then MAPKK5x3 with 331 amino acids and the smallest sequence is for MAPKK5x1 with 319 amino acids.

After MSA was performed in Clustal Omega, as shown in Table 3.

Table *3*, the highest similarity score was observed between MAPKK5x3 and MAPKK5x2 with 95.30%. The second highest similarity is between MAPKK5x1 and MAPKK5x2 with 92.48% similarity. After that we obtain similarity with 88.71% between MAPKK5x3 and MAPKK5x1, and 77.44% similarity between MAPKK5x4 and MAPKK5x3, as well as 74.16% similarity between MAPKK5x2 and MAPKK5x4 and the smallest one is for MAPKK5x1 and MAPKK5x4 with 73.04% similarity. From these results, it has been concluded that the first three varieties of MAPKK5 show that they share significant similarity among each other with more than 80% similarity, thus Sequences with significant similarities are thought to be homologous and they have a common ancestor [64]. However, they also share close similarity scores with MAPKK5x4.

For further phylogenetic searches for MAPKK5 homologs from Brassica Oleracea, a phylogenetic tree was constructed (Figure 4.), demonstrating a distinct and strong evolutionary connection between the target species. According to the phylogenetic tree both MAPKK5x3 and MAPKK5x1 are attached with MAPKK5x2 have high similarity with it, however MAPKK5x4 sticks out of the group since it is on a separate branch. This is compatible with the MSA findings in Table 3.

Table *3*.

Predicting the three-dimensional structure of the protein can lead to a higher level of understanding of how a protein operates, allowing us to develop hypotheses about how to affect, modify and control the protein. When assessing protein function, structural information is frequently more valuable than having only sequences of the proteins, many factors influence the quality of the model created during each step of the homology modelling process, and optimal software selection can considerably increase the model's quality, Programs for homology modelling use many methodologies and methods to generate the final model [65]. As shown in Figure 5, Figure 6, Figure 7, Figure 8, Phyre2 online bioinformatics tool used to predict the homology modelling for our proteins. The degree of similarity between the model and the protein sequences determines the quality of the predicted structure through homology modelling. If the similarity is extremely poor, homology modelling of the query protein does not produce useful results [66]. Furthermore, the modelled structure was verified and confirmed using an online program called Verify3D.This program uses three-dimensional profiles to evaluate protein structures it examines a 3D model's compatibility with its own amino acid sequence. The scores range from the lowest score -1 to +1 which is the highest score [67].

Following the 3D structure modelling of all four variations of *Brassica Oleracea* MAPKK5 proteins, our structured models were validated using the above-mentioned tool and demonstrated appropriate quality for future study. Our findings show that the Verify3D score for MAPKK5x1 it is 82.76%, for MAPKK5x2 is the lowest score which is 60.26%, for MAPKK5x3 it is 78.85% and for MAPKK5x4 is the highest score which is 86.97%.

Ramachandran plot is a graph that shows which secondary structures include which amino acids based on where the amino acid residue falls. Each amino acid represented as a dot and they are in a phi and psi angels. Ramachandran plot has horizontal axis which represent the phi angel and the vertical axis which represent the psi angel. It is divided to four regions, the upper left-handed region represents the beta-sheet secondary structure proteins, the upper right-handed regions represent the left-handed alpha –helix secondary structure proteins and the lower left-handed region shows the right-handed alpha-helix secondary structure proteins, as shown in Figure 15.

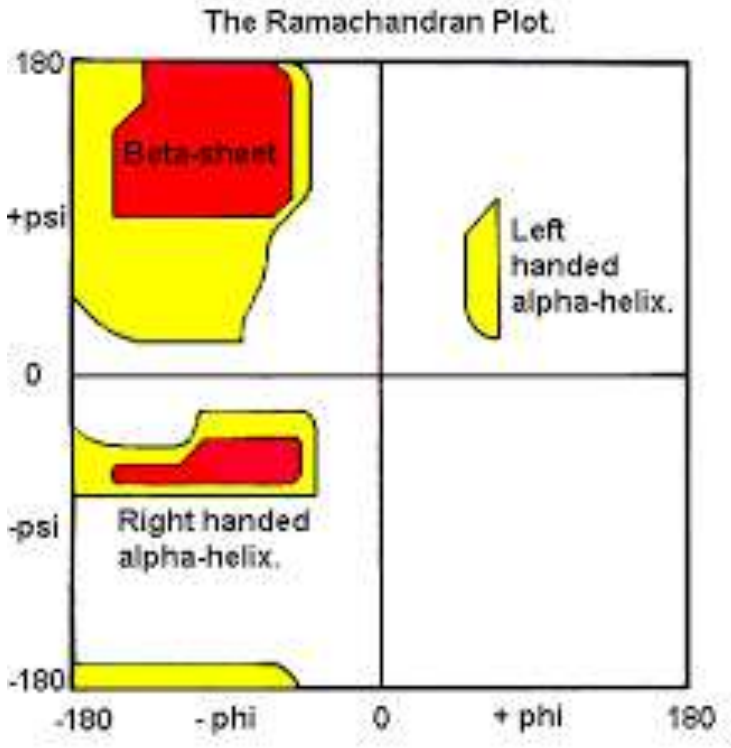


Figure 15. Ramachandran plot shows the torsion angles (phi and psi) of amino acid residues

As shown in Figure 5, Figure 6, Figure 7, Figure 8 and Table 4, the Ramachandran plot demonstrates that most amino acid residues in all four proteins are in the highly preferred regions. From the results obtained, MAPKK5x1 has 91.513% (248 AA residues) in the highly preferred region, 5.904% (16 AA residues) in the preferred region, and 2.583% (7 AA residues) in the questionable region. MAPKK5x2 has 87.537% (295 AA residues) in the highly preferred region, 6.825 % (23 AA residues) in the preferred region, and 5.638% (19 AA residues) in the questionable region. MAPKK5x3 has 79.715% (224 AA residues) in the highly preferred region, 9.964 % (28 AA residues) in the preferred region, and 10.320% (29 AA residues) in the questionable region. Furthermore, the best score is for the MAPKK5x4 model, 92% (276 AA residues) in the highly preferred region, 5% (15 AA residues) in the preferred region, and 3% (9 AA residues) in the Questionable region.

After performing subcellular localization analysis using loctree 3, in the scores obtained in Table 5, it is shown that all four proteins are mostly located in the cytoplasm. Both MAPKK5x1 and MAPKK5X4 show that they have 83% expected accuracy with the same score (20), and both MAPKK5x2 and MAPKK5x3 show that they have 82% expected accuracy with the same score (18).

The following step was to use SMART software to search for protein domains. As seen in Table 6 and Figure 9, all proteins share only one domain that begins with various residues. The domain identified is the Serine or threonine-specific kinase subfamily (S_TKc). Protein phosphorylation is a reversible process mediated by protein kinases and phosphoprotein phosphatases that is essential for most cellular processes. Protein kinases catalyse the transfer of gamma phosphate from nucleotide triphosphates (often ATP) to one or more amino acid residues in the side chain of a protein substrate, resulting in a conformational change that affects protein

function. Furthermore, the reverse reaction is catalysed by phosphoprotein phosphatases [68]. In addition, we can observe from Table 6 and Figure 9 that all the proteins contain low complexity regions, which are rare areas in protein sequences that contain only a few diverse types of amino acids [69]. There are several domains worth highlighting that are not included in the results, because their scores are less significant than the required threshold as presented in Table 9.

Table 9. Specific domains – low-significant results

Domains that are not shown in the diagram	
RIO-like kinase, Helix-hairpin-helix DNA-binding motif class 1	DCP2
Tyrosine kinase	CULLIN, FANCL C-terminal domain, Defensin/corticostatin family
Catalytic domain (TyrKc),	FeoA
snRNP Sm proteins (Sm)	Mre11_DNA_bind
Domain in different transcription and chromosome remodeling factors (DDT),	Hr1
Ring finger	PKD domain,LamB
Calpain_III	UreE_N
Ribosomal_S15	Ubox

After locating certain domains in MAPKK5 proteins, the goal was to identify and find other proteins that interact with them. When this search was performed in STRING, ten interacting proteins were discovered, all of which interact with the four MAPKK5 protein variants.

As shown in

Figure 10. STRING interactome of Brassica Oleracea MAPKK5 four variants

Table *7*, the highest confidence interacting protein, mitogen-activated protein kinase 3 (MAPK3), and the lowest confidence interacting protein, mitogen-activated protein kinase 6 (MAPK6), show related functions in plants. Both are found in the nucleus. Because of the high confidence interaction of MAPK3, this protein was chosen for docking analysis with MAPKK5x1, MAPKK5x2, MAPKK5x3, and MAPKK5x4. The validation score of this protein was performed by verify3D, and we got 80.27% (Table 8). One interesting study shows that (MAPK) cascades, a highly conserved signalling pathway found in all eukaryotes, are a key mediator of hypersensitive response-associated PCD. Therefore, monitoring MAPK activation allows for the elucidation of the mechanisms behind HR-associated PCD. MAPK3 and MAPK6 are two of the most researched MAP kinases. Under pathogen infection, activation of these kinases is a key step and an early indication for full induction of defence responses and hypersensitive response-associated PCD. They are commercially available because they are conserved throughout eukaryotes. PCD is a vital part of development, as well as biotic and

abiotic responses. Furthermore, one of the most severe forms of PCD in plants is hypersensitive response-associated cell death caused by pathogen infection [57].

Protein kinase domain-containing protein is one of the interacting proteins, which is found in the nucleus. This protein was chosen for docking analysis with MAPKK5x1, MAPKK5x2, MAPKK5x3, and MAPKK5x4. The validation score of this protein was performed by verify3D, and we got 56.15%, as shown in Table 8.

The interaction protein mitogen-activated protein kinase YODA, found in the plasma membrane, docks with MAPKK5x1, MAPKK5x2, MAPKK5x3, and MAPKK5x4. The validation score of this protein was performed by verify3D, 60.23% (Table 8.). Mitogen-activated protein kinase cascades are important in plants because they convert developmental cues and environmental signals into cellular responses. Microbe-associated chemical patterns are among those recognized by pattern recognition receptors (PRRs), which activate immunity. Interesting research found that MAPK kinase YODA (YDA), which regulates various Arabidopsis developmental processes, including stomatal patterning, was discovered to control immunological responses. YDA regulates both immunity and stomatal patterning via the same route as ERECTA (ER) Receptor-Like Kinase [59].

Another interacting protein is mitogen-activated protein kinase 1-like isoform X1, which is found in the nucleus. Interesting research investigates the role of candidate plant MAP kinase homologs of human MEKK1 in pathogen-resistance pathways by using virus-induced gene silencing (VIGS). Their findings show that MAPKKK1 regulates N-, Bs2-, and Rx-mediated resistance responses and may participate in one or more MAPK cascades, regulating many cellular activities, as well as demonstrating that when plants recognize invading pathogens, mitogen-activated protein kinase cascades are rapidly triggered [62]. We can observe from the interactions that we got in

Figure 10. STRING interactome of Brassica Oleracea MAPKK5 four variants

Table *7* between MAPKK5and MAPKKK1 that our protein can also have related functions in resistance responses and regulation of cellular activities.

5. Conclusion

In nature, the *Brassicaceae* family is regularly subjected to a variety of unfavorable conditions caused by biotic and abiotic stressors, which have a negative impact on plant development and production. In *Brassica Oleracea*, four variants of mitogen-activated protein kinase5 (MAPKK5) genes were identified, as well as their protein sequences, MAPKK5x1, MAPKK5x2, MAPKK5x3, and MAPKK5x4. These genes all perform the same role in *Brassica* family plants; they control cell division, development, hormone signaling, and biotic and abiotic stress responses. In this research, we concentrated on further investigating the roles of MAPKK5 genes in *Brassica Oleracea*. This was accomplished via *in silico* analysis in regard to their sequences, structure, localization, domains, and interactions with other proteins. Through multiple sequence alignment, we discovered that all four variants of MAPKK5 share significant similarity among their sequences. For further analysis, we confirmed that these proteins are closely related and share the same ancestor, and strong evolutionary connection between them. Through three-dimensional structure prediction and validation, we revealed that these four proteins have quite a similar structure to each other. In addition, we discovered that these proteins have relatively good quality. Following that, we conducted domain analysis on the protein structures and found that these proteins share only one domain, and that is the Serine or threonine-specific kinase subfamily (S_TKc). Later, using a localization tool, we were able to predict the subcellular locations of these proteins, and they showed that all of them are in cytoplasm. Eventually, when we conducted interactome analysis, we discovered that all the MAPKK5 proteins interact with the same proteins, and through docking site analysis, we predicted the positioning and binding affinity of interacting proteins in the binding

site of our proteins. Since MAPKK5 is very important for signal transduction in plants, there should be more *in silico* research undertaken on its structures and interactions that are currently unknown.

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.

References

[1] J. Yu et al., "Bolbase: A comprehensive genomics database for Brassica oleracea," BMC Genomics, vol. 14, no. 1, 2013, https://doi.org/10.1186/1471-2164-14-664.

[2] M. Jahangir, H. K. Kim, Y. H. Choi, and R. Verpoorte, "Health-affecting compounds in Brassicaceae," 2009, https://doi.org/10.1111/j.1541-4337.2008.00065.x.

[3] S. Cheng, "Effects of heavy metals on plants and resistance mechanisms," 2003, https://doi.org/10.1065/espr2002.11.141.2.

[4] A. Asati, M. Pichhode, and K. Nikkil, "Effect of heavy metals on plants: An overview," International Journal of Application or Innovation in Engineering & Management, vol. 5, no. 3, 2016.

[5] J. M. Argüello, E. Eren, and M. González-Guerrero, "The structure and function of heavy metal transport P1B-ATPases," in BioMetals, 2007, https://doi.org/10.1007/s10534-006-9055-6.

[6] E. Eren, M. González-Guerrero, B. M. Kaufman, and J. M. Argüello, "Novel Zn2+ coordination by the regulatory N-terminus metal binding domain of Arabidopsis thaliana Zn2+-ATPase HMA2," Biochemistry, vol. 46, no. 26, 2007, https://doi.org/10.1021/bi7001345.

[7] C. K. E. Wong and C. S. Cobbett, "HMA P-type ATPases are the major mechanism for root-to-shoot Cd translocation in Arabidopsis thaliana," New Phytologist, vol. 181, no. 1, 2009, https://doi.org/10.1111/j.1469-8137.2008.02638.x.

[8] E. Eren, D. C. Kennedy, M. J. Maroney, and J. M. Argüello, "A novel regulatory metal binding domain is present in the C terminus of Arabidopsis Zn2+-ATPase HMA2," Journal of Biological Chemistry, vol. 281, no. 45, 2006, https://doi.org/10.1074/jbc.M605218200.

[9] C. Nouet et al., "Functional analysis of the three HMA4 copies of the metal hyperaccumulator Arabidopsis halleri," J. Exp. Bot., vol. 66, no. 19, 2015, https://doi.org/10.1093/jxb/erv280.

[10] F. Verret et al., "Heavy metal transport by AtHMA4 involves the N-terminal degenerated metal binding domain and the C-terminal His11 stretch," FEBS Lett., vol. 579, no. 6, 2005, https://doi.org/10.1016/j.febslet.2005.01.065.

[11] M. Morel et al., "AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in Arabidopsis," Plant Physiol., vol. 149, no. 2, 2009, https://doi.org/10.1104/pp.108.130294.

[12] K. Chaudhary, S. Jan, and S. Khan, "Heavy Metal ATPase (HMA2, HMA3, and HMA4) Genes in Hyperaccumulation Mechanism of Heavy Metals," in Plant Metal Interaction: Emerging Remediation Techniques, 2015, https://doi.org/10.1016/B978-0-12-803158-2.00023-0.

[13] D. Seigneurin-Berny et al., "HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions," Journal of Biological Chemistry, vol. 281, no. 5, 2006, https://doi.org/10.1074/jbc.M508333200.

[14] R. F. Mills, K. A. Peaston, J. Runions, and L. E. Williams, "HvHMA2, a P1B-ATPase from barley, is highly conserved among cereals and functions in Zn and Cd transport," PLoS One, vol. 7, no. 8, 2012, https://doi.org/10.1371/journal.pone.0042640.

[15] M. C. E. Van Dam-Mieras, A. J. Slotboom, W. A. Pieterson, and G. H. de Haas, "Interaction of Phospholipase A2 with Micellar Interfaces.The Role of the N-Terminal Region," Biochemistry, vol. 14, no. 25, 1975, https://doi.org/10.1021/bi00696a001.

[16] J. Tan et al., "Functional analyses of TaHMA2, a P1B-type ATPase in wheat," Plant Biotechnol. J., vol. 11, no. 4, 2013, https://doi.org/10.1111/pbi.12027.

[17] Y. Wang, K. F. Yu, V. Poysa, C. Shi, and Y. H. Zhou, "A single point mutation in GmHMA3 affects cadimum (Cd) translocation and accumulation in soybean seeds," 2012, https://doi.org/10.1093/mp/sss069.

[18] D. Ueno et al., "Gene limiting cadmium accumulation in rice," Proc. Natl. Acad. Sci. U. S. A., vol. 107, no. 38, 2010, https://doi.org/10.1073/pnas.1005396107.

[19] H. Miyadate et al., "OsHMA3, a P1B-type of ATPase affects root-to-shoot cadmium translocation in rice by mediating efflux into vacuoles," New Phytologist, vol. 189, no. 1, 2011, https://doi.org/10.1111/j.1469-8137.2010.03459.x.

[20] W. Park, K. H. Han, and S. J. Ahn, "Differences in root-to-shoot cd and Zn translocation and by HMA3 and 4 could influence chlorophyll and anthocyanin content in arabidopsis ws and col-0 ecotypes under excess metals," Soil Sci. Plant Nutr., vol. 58, no. 3, 2012, https://doi.org/10.1080/00380768.2012.684643.

[21] D. Y. Chao et al., "Genome-Wide Association Studies Identify Heavy Metal ATPase3 as the Primary Determinant of Natural Variation in Leaf Cadmium in Arabidopsis thaliana," PLoS Genet., vol. 8, no. 9, 2012, https://doi.org/10.1371/journal.pgen.1002923.

[22] S. Ó Lochlainn et al., "Tandem quadruplication of HMA4 in the zinc (Zn) and cadmium (Cd) hyperaccumulator Noccaea caerulescens," PLoS One, vol. 6, no. 3, 2011, https://doi.org/10.1371/journal.pone.0017814.

[23] J. P. Adams et al., "Poplar maintains zinc homeostasis with heavy metal genes HMA4 and PCS1," J. Exp. Bot., vol. 62, no. 11, 2011, https://doi.org/10.1093/jxb/err025.

[24] A. Barabasz, U. Krämer, M. Hanikenne, J. Rudzka, and D. M. Antosiewicz, "Metal accumulation in tobacco expressing Arabidopsis halleri metal hyperaccumulation gene depends on external supply.," J. Exp. Bot., vol. 61, no. 11, 2010, https://doi.org/10.1093/jxb/erq129.

[25] P. Halimaa et al., "Gene expression differences between noccaea caerulescens ecotypes help to identify candidate genes for metal phytoremediation," Environ. Sci. Technol., vol. 48, no. 6, 2014, https://doi.org/10.1021/es4042995.

[26] N. Andrés-Colás et al., "The Arabidopsis heavy metal P-type ATPase HMA5 interacts with metallochaperones and functions in copper detoxification of roots," Plant Journal, vol. 45, no. 2, 2006, https://doi.org/10.1111/j.1365-313X.2005.02601.x.

[27] K. Lu et al., "Genome-wide survey and expression profile analysis of the mitogen-activated protein kinase (MAPK) gene family in brassica rapa," PLoS One, vol. 10, no. 7, 2015, https://doi.org/10.1371/journal.pone.0132051.

[28] Z. Wang et al., "Genome-wide identification and analysis of mkk and mapk gene families in brassica species and response to stress in brassica napus," Int. J. Mol. Sci., vol. 22, no. 2, 2021, https://doi.org/10.3390/ijms22020544.

[29] M. Cristina, M. Petersen, and J. Mundy, "Mitogen-activated protein kinase signaling in plants," Annu. Rev. Plant Biol., vol. 61, 2010, https://doi.org/10.1146/annurev-arplant-042809-112252.

[30] T. Sun and Y. Zhang, "MAP kinase cascades in plant development and immune signaling," EMBO Rep., vol. 23, no. 2, 2022, https://doi.org/10.15252/embr.202153817.

[31] G. M. Thomas and R. L. Huganir, "MAPK cascade signalling and synaptic plasticity," 2004, https://doi.org/10.1038/nrn1346.

[32] Y. Sun et al., "Identification and functional analysis of mitogen-activated protein kinase kinase kinase (MAPKKK) genes in canola (Brassica napus L.)," J. Exp. Bot., vol. 65, no. 8, 2014, https://doi.org/10.1093/jxb/eru092.

[33] P. Wu, W. Wang, Y. Li, and X. Hou, "Divergent evolutionary patterns of the MAPK cascade genes in Brassica rapa and plant phylogenetics," Hortic. Res., vol. 4, 2017, https://doi.org/10.1038/hortres.2017.79.

[34] M. Gaur, A. Tiwari, R. P. Chauhan, D. Pandey, and A. Kumar, "Molecular modeling, docking and protein-protein interaction analysis of MAPK signalling cascade involved in Camalexin biosynthesis in Brassica rapa," Bioinformation, vol. 14, no. 04, 2018, https://doi.org/10.6026/97320630014145.

[35] S. Muthusaravanan et al., "Phytoremediation of heavy metals: mechanisms, methods and enhancements," 2018, https://doi.org/10.1007/s10311-018-0762-3.

[36] I. Ali et al., "Screening of various Brassica species for phytoremediation of heavy metals-contaminated soil of Lakki Marwat, Pakistan," Environmental Science and Pollution Research, vol. 29, no. 25, 2022, https://doi.org/10.1007/s11356-021-18109-7.

[37] P. J.C., J. Pratas, M. Varun, R. DSouza, and M. S., "Phytoremediation of Soils Contaminated with Metals and Metalloids at Mining Areas: Potential of Native Flora," in Environmental Risk Assessment of Soil Contamination, 2014, https://doi.org/10.5772/57469.

[38] M. R. Fielden, J. B. Matthews, K. C. Fertuck, R. G. Halgren, and T. R. Zacharewski, "In silico approaches to mechanistic and predictive toxicology: An introduction to bioinformatics for toxicologists," 2002, https://doi.org/10.1080/20024091064183.

[39] I. Y. Iourov, S. G. Vorsanova, and Y. B. Yurov, "In silico molecular cytogenetics: A bioinformatic approach to prioritization of candidate genes and copy number variations for basic and clinical genome research," Mol. Cytogenet., vol. 7, no. 1, 2014, https://doi.org/10.1186/s13039-014-0098-z.

[40] J. D. Thompson, Toby. J. Gibson, and D. G. Higgins, "Multiple Sequence Alignment Using ClustalW and ClustalX," Curr. Protoc. Bioinformatics, vol. 00, no. 1, 2003, https://doi.org/10.1002/0471250953.bi0203s00.

[41] F. Sievers and D. G. Higgins, "The Clustal Omega Multiple Alignment Package," in Methods in Molecular Biology, vol. 2231, 2021, https://doi.org/10.1007/978-1-0716-1036-7_1.

[42] A. Dereeper et al., "Phylogeny.fr: robust phylogenetic analysis for the non-specialist.," Nucleic Acids Res., vol. 36, no. Web Server issue, 2008, https://doi.org/10.1093/nar/gkn180.

[43] J. Dees, J. L. Momsen, J. Niemi, and L. Montplaisir, "Student interpretations of phylogenetic trees in an introductory biology course," CBE Life Sci. Educ., vol. 13, no. 4, 2014, https://doi.org/10.1187/cbe.14-01-0003.

[44] J. Castresana, "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis," Mol. Biol. Evol., vol. 17, no. 4, 2000, https://doi.org/10.1093/oxfordjournals.molbev.a026334.

[45] Y. Zhang, "I-TASSER server for protein 3D structure prediction," BMC Bioinformatics, vol. 9, 2008, https://doi.org/10.1186/1471-2105-9-40.

[46] T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, "SWISS-MODEL: An automated protein homology-modeling server," Nucleic Acids Res., vol. 31, no. 13, 2003, https://doi.org/10.1093/nar/gkg520.

[47] L. Essen, "Structural Bioinformatics. Edited by Philip E. Bourne and Helge Weissig.," Angewandte Chemie International Edition, vol. 42, no. 41, 2003, https://doi.org/10.1002/anie.200385018.

[48] M. Biasini et al., "SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information," Nucleic Acids Res., vol. 42, no. W1, 2014, https://doi.org/10.1093/nar/gku340.

[49] S. Yuan, H. C. S. Chan, and Z. Hu, "Using PyMOL as a platform for computational drug design," 2017, https://doi.org/10.1002/wcms.1298.

[50] R. Lüthy, J. U. Bowie, and D. Eisenberg, "Assessment of protein models with three-dimensional profiles," Nature, vol. 356, no. 6364, 1992, https://doi.org/10.1038/356083a0.

[51] I. Letunic, T. Doerks, and P. Bork, "SMART 7: Recent updates to the protein domain annotation resource," Nucleic Acids Res., vol. 40, no. D1, 2012, https://doi.org/10.1093/nar/gkr931.

[52] 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, 1998, https://doi.org/10.1073/pnas.95.11.5857.

[53] D. Szklarczyk et al., "STRING v10: Protein-protein interaction networks, integrated over the tree of life," Nucleic Acids Res., vol. 43, no. D1, 2015, https://doi.org/10.1093/nar/gku1003.

[54] A. Dhanik and L. E. Kavraki, "Protein–Ligand Interactions: Computational Docking," in Encyclopedia of Life Sciences, 2012, https://doi.org/10.1002/9780470015902.a0004105.pub2.

[55] D. Kozakov et al., "The ClusPro web server for protein-protein docking," Nat. Protoc., vol. 12, no. 2, 2017, https://doi.org/10.1038/nprot.2016.169.

[56] J. J. Sutkovic, "An in silico approach for structural and functional analysis of Heavy Metal Associated (HMA) proteins in Brassica oleracea," Periodicals of Engineering and Natural Sciences (PEN), vol. 4, no. 2, 2016, https://doi.org/10.21533/pen.v4i2.63.

[57] Q. Wu and D. Jackson, "Detection of MAPK3/6 phosphorylation during hypersensitive response (HR)-associated programmed cell death in plants," in Methods in Molecular Biology, vol. 1743, 2018, https://doi.org/10.1007/978-1-4939-7668-3_14.

[58] A. Bateman et al., "UniProt: the universal protein knowledgebase in 2021," Nucleic Acids Res., vol. 49, no. D1, 2021, https://doi.org/10.1093/nar/gkaa1100.

[59] S. Sopeña-Torres et al., "YODA MAP3K kinase regulates plant immune responses conferring broad-spectrum disease resistance," New Phytologist, vol. 218, no. 2, 2018, https://doi.org/10.1111/nph.15007.

[60] R. Hasan et al., "Mitogen activated protein kinase kinase kinase 3 (MAP3K3/MEKK3) overexpression is an early event in esophageal tumorigenesis and is a predictor of poor disease prognosis," BMC Cancer, vol. 14, no. 1, 2014, https://doi.org/10.1186/1471-2407-14-2.

[61] P. M. Gilmore et al., "BRCA1 interacts with and is required for paclitaxel-induced activation of mitogen-activated protein kinase kinase kinase 3," Cancer Res., vol. 64, no. 12, 2004, https://doi.org/10.1158/0008-5472.CAN-03-4080.

[62] H. Jin et al., "NPK1, and MEKK1-like mitogen-activated protein kinase kinase kinase, regulates innate immunity and development in plants," Dev. Cell, vol. 3, no. 2, 2002, https://doi.org/10.1016/S1534-5807(02)00205-8.

[63] D. Ren, K. Y. Yang, G. J. Li, Y. Liu, and S. Zhang, "Activation of Ntf4, a tobacco mitogen-activated protein kinase, during plant defense response and its involvement in hypersensitive response-like cell death," Plant Physiol., vol. 141, no. 4, 2006, https://doi.org/10.1104/pp.106.080697.

[64] W. R. Pearson, "An introduction to sequence similarity ('homology') searching," Curr. Protoc. Bioinformatics, no. SUPPL.42, 2013, https://doi.org/10.1002/0471250953.bi0301s42.

[65] M. R. Pitman and R. I. Menz, "Methods for protein homology modelling," in Applied Mycology and Biotechnology, vol. 6, no. C, 2006, https://doi.org/10.1016/S1874-5334(06)80005-5.

[66] S. W. Robinson, A. M. Afzal, and D. P. Leader, "Bioinformatics: Concepts, Methods, and Data," in Handbook of Pharmacogenomics and Stratified Medicine, 2014, https://doi.org/10.1016/B978-0-12-386882-4.00013-X.

[67] D. Eisenberg, R. Lüthy, and J. U. Bowie, "VERIFY3D: Assessment of protein models with three-dimensional profiles," Methods Enzymol., vol. 277, 1997, https://doi.org/10.1016/S0076-6879(97)77022-8.

[68] S. K. Hanks, A. M. Quinn, and T. Hunter, "The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains," Science (1979)., vol. 241, no. 4861, 1988, https://doi.org/10.1126/science.3291115.

[69] S. M. Cascarina, M. R. Elder, and E. D. Ross, "Atypical structural tendencies among low-complexity domains in the protein data bank proteome," PLoS Comput. Biol., vol. 16, no. 1, 2020, https://doi.org/10.1371/journal.pcbi.1007487.