Comparative Structural Analysis of MAP kinase Protein in Some Seed Monocot Plants

Narges Gorjizadeh, Milad Moeini and Samin Seddigh

***Abstract*—Mitogen-activated protein kinases (MAPKs) cascades coexist in eukaryotic cells and facilitate appropriate cellular responses to distinct environmental inputs. They are the family of kinases that transduce signals from the cell membrane to the nucleus in response to a wide range of stimuli. The MAPK signaling pathways modulate gene expression, mitosis, proliferation, motility, metabolism, and programmed death apoptosis. In this study, the MAPK RefSeq (protein reference sequences) belonging to six seed monocot plants species from various families were retrieved from NCBI (the National Center for Biotechnology Information) in FASTA format. They were included Ananas comosus (Bromeliaceae), Asparagus officinalis (Asparagaceae), Phalaenopsis equestris (Orchidaceae), Dioscorea cayenensis (Dioscoreaceae), Elaeis guineensis (Arecaceae), and Musa acuminata (Musaceae). Structural analyses of MAPK protein were performed on the EXPASY website. Primary and secondary structures were estimated using ProtParam and SOPMA, respectively. The tertiary structure prediction and visualization of MAPK were accomplished by PHYRE2 server and WebLab ViewerLite 4.2, respectively. MAPK protein in all plant samples ranged from 1103 to 1294aa in length. The theoretical isoelectric points and molecular weight were calculated at a range of 5.18-5.40 and 121543.41-143439.70Da, respectively. High aliphatic index in MAPKs indicated structural stability of this protein in these monocot plant species. The GRAVY values of this protein designated these families to be globular (hydrophilic protein) in nature. The Instability index analysis of these plant MAPK revealed that all the plants samples were unstable proteins. The secondary structure analysis showed that the most abundant structural elements in all the samples were random coils and alpha helices while extended strands and beta turns were occasionally distributed in the proteins. 3D structure prediction revealed that the tertiary structure of seed monocot plants of these families are very similar to each other which confirmed structural similarity of this protein.**

***Keywords*— Expasy, MAPK protein, Primary structure, Secondary structure, Tertiary structure, OPMA, Seed monocot**

I. Introduction

Cells constantly need to integrate external stress signals, so they either die permanently or survive depending on their cell fate. These fate decisions are driven by a variety of signaling pathways controlled by kinases. Developmental programs and environmental factors trigger evolutionarily conserved distinct kinases that transduce signals that mediate proliferation, survival, death, or cell cycle arrest [1]. An increasing number of studies have indicated that the temporal dynamics of MAPK activity could encode environmental information and may mediate cell fate decisions [2-4]. The mitogen-activated protein kinases (MAPKs) are a family of kinases that transmit signals from the cell membrane to the nucleus in response to a wide range of stimuli, including stress [5-8]. MAPK are protein kinases that phosphorylate their own dual serine and threonine residues (autophosphorylation), or those found on their substrates, to activate or de-activate their target [7, 9]. Consequently, MAPKs regulate essential cellular processes such as proliferation, stress responses, apoptosis and immune defense [10, 11]. MAPKs are ubiquitously expressed and evolutionarily conserved in eukaryotes [9, 12]. MAPK cascades coexist in eukaryotic cells and mediate appropriate cellular responses to different environmental stimuli [13, 14]. The MAPK cascade consists of a MAPK, a MAPK kinase (MAPKK), and a MAPKK kinase (MAPKKK). In Saccharomyces cerevisiae, four complete MAPK cascade modules have been identified [15, 16]. The pheromone response pathway is activated by peptide pheromones and prepares cells for mating [17].

MAPKs are serine/threonine kinases that phosphorylate specific substrates on serine and/or threonine residues upon stimulation. Such phosphorylation events can either positively or negatively control substrates and hence entire signaling cascade activity. Therefore, the MAPK signaling pathways modulate gene expression, programmed death 'apoptosis', mitosis, proliferation, motility and metabolism [5-8].

The aim of the present study was to deliver structural analyses on a key gene in plants to determine the biochemical characterization in order to gain a better understanding of resistance of plants. Analyses of molecular mass, isoelectric point, instability and aliphatic index, grand average of hydropathicity, secondary and spatial structure, were performed to provide insights into the evolutionary mechanisms of the MAPK protein family.

II. Material and Methods

A. MAPK protein data received

The MAPK protein Reference Sequences (RefSeq) belonging to different plants species were downloaded from NCBI (http://www.ncbi.nlm.nih.gov) in FASTA format (date received April 2022). Among the plant sequences, those with the range of 1000–1300 amino acids were selected by the NCBI filters. Due to the large number of sequences available for MAPK in plants, it is not possible to analyze all the sequences.

Milad Moeini and Narges Gorjizadeh, Department of Plant Pathology, Varamin-Pishva Branch, Islamic Azad University, Varamin, Iran.

Samin Seddigh Department of Plant Protection, Varamin-Pishva Branch, Islamic Azad University, Varamin, Iran

So, sequences from different plants belonging to different seed monocot plants were retrieved for structural analyses. Total data consisted of six plants species which are listed in Table 1 with their families and accession numbers.

TABLE I. MAPK PROTEIN IN SIX PLANT SPECIES

Index	Scientific name of the plant	Family	Order	Abbreviation	Accession number
1	*Ananas comosus*	Bromeliaceae	Poales	*Ac*-MAPK	XP_020079688
2	*Asparagus officinalis*	Asparagaceae	Asparagales	*Ao*-MAPK	XP_020244178
3	*Dioscorea cayenensis subsp. rotundata*	Dioscoreaceae	Dioscoreales	*Dcr*-MAPK	XP_039130050
4	*Elaeis guineensis*	Arecaceae	Arecales	*Eg*-MAPK	XP_029124494
5	*Musa acuminata subsp. malaccensis*	Musaceae	Zingiberales	*Mam*-MAPK	XP_018683570
6	*Phalaenopsis equestris*	Orchidaceae	Asparagales	*Pe*-MAPK	XP_020593371

B. Structural analyses

Several online web services and software were used for the analyses of MAPK in plants. Comparative and proteomics analyses were carried out online at the EXPASY website (http://expasy.org). Total number of atoms, instability index, molecular weight, isoelectric point, number of charge residues, aliphatic index, and grand average of hydropathicity (GRAVY) were predicted by ProtParam (http://web. expasy.org/protparam/) [18]. Secondary structures of seed plant samples were predicted using SOPMA (http://npsa-pbil.ibcp.fr/cgibin/npsa_automat.pl?page¼npsa_s opma.html) [19]. The tertiary structure prediction analysis of MAPK protein was done by Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) [20] using profile-profile matching and secondary structure in representative seed monocot plant species. WebLab ViewerLite 4.2 was used for 3D structure visualization.

III. RESULTS AND DISCUSSION

In order to investigate the structures of MAPKs, six species from different plant families were analyzed by bioinformatics tools. The primary structure of selected MAPKs was analyzed by ProtParam. This server computes various physico-chemical properties that can be deduced from a protein sequence including the, instability and aliphatic index, and grand average of hydropathicity (GRAVY) (Table 2). MAPK protein in all plants ranged from 1103-1294 amino acids in length. The theoretical isoelectric points and molecular weight were calculated at a range of 5.18-5.40 and 121543.41-143439.70Da, respectively (Table 2). According to Kyte and Doolittle [21] integral membrane proteins normally have higher GRAVY scores than do globular proteins. Gravy index score of proteins below 0 are more likely globular (hydrophilic protein), however scores above 0 are more likely membranous (hydrophobic protein). Values of GRAVY ranged from -0.490 (Dcr-MAPK) to -0.604 (Eg-MAPK) in plant samples which showed MAPK as a hydrophilic protein.

According to Table 2, the least aliphatic index belonged to Eg-MAPK while the most was observed in Dcr-MAPK. The aliphatic index of a protein is known as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be considered as a positive factor for the increase of thermostability of globular proteins. The instability index is one of the protein primary structure-dependent methods to predict protein stability in a test tube. Instability index less than 40 predicts a stable protein, though values higher than 40 represent a potentially unstable protein [22]. Computed values of instability index of MAPK were between 41.78 (Pe-MAPK) and 51.35 (Ac-MAPK) (Table 2) and showed that as an unstable protein.

TABLE II. PRIMARY STRUCTURE OF MAPK PROTEIN IN SIX PLANT SPECIES

index	Samples	Number of amino acids	Molecular weight	Theoretical pI	Aliphatic index	Instability index	Gravy
1	*Ac*-MAPK	1103	121543.41	5.21	73.19	51.35	-0.493
2	*Ao*-MAPK	1106	122102.06	5.18	72.20	51.34	-0.559
3	*Dcr*-MAPK	1103	122089.92	5.19	76.87	47.57	-0.490
4	*Eg*-MAPK	1294	143439.70	5.25	69.01	50.90	-0.604
5	*Mam*-MAPK	1138	126744.13	5.40	74.98	44.49	-0.534
6	*Pe*-MAPK	1267	139629.30	5.11	74.28	41.78	-0.525

The secondary structure of plant MAPK proteins was predicted by Self-Optimized Prediction Method with Alignment (SOPMA) server which calculates the percentage of alpha helix, extended strand, beta turn, and random coil. The prediction of the secondary structure of MAPK based on hierarchical neural network analysis [23] revealed that the most plentiful structural elements of the secondary structure were random coils in all the species, While and alpha helices and extended strands were occasionally distributed in the proteins. Also, the least structures belonged to beta turns in protein's secondary structure (Table 3).

TABLE III. SECONDARY STRUCTURE OF MAPK PROTEIN IN SIX PLANT SPECIES

index	Samples	Alpha helix%	Beta turn%	Extended strand%	Random coil%
1	*Ac*-MAPK	27.56	7.07	16.86	48.50
2	*Ao*-MAPK	19.71	4.43	14.65	61.21
3	*Dcr*-MAPK	23.39	5.08	13.78	57.75
4	*Eg*-MAPK	22.18	4.79	11.44	61.59
5	*Mam*-MAPK	25.04	4.31	10.63	60.02
6	*Pe*-MAPK	20.60	3.79	12.08	63.54

The key objective of protein structure prediction is the tertiary structural modeling, which is essential for full understanding of protein function. Also structures are much more conserved than sequences, so their evaluations allow us to look even further back into biological prehistory [22, 24]. So, three-dimensional modeling of six plant samples was predicted by Phyre2. This server uses profile–profile matching and secondary structure for tertiary structure prediction [20] and uses the alignment of hidden Markov models through HHsearch [25]. Based on these analyses the tertiary structure of seed monocot plants was very similar to each other which confirmed structural similarity of this protein (Figure 1).

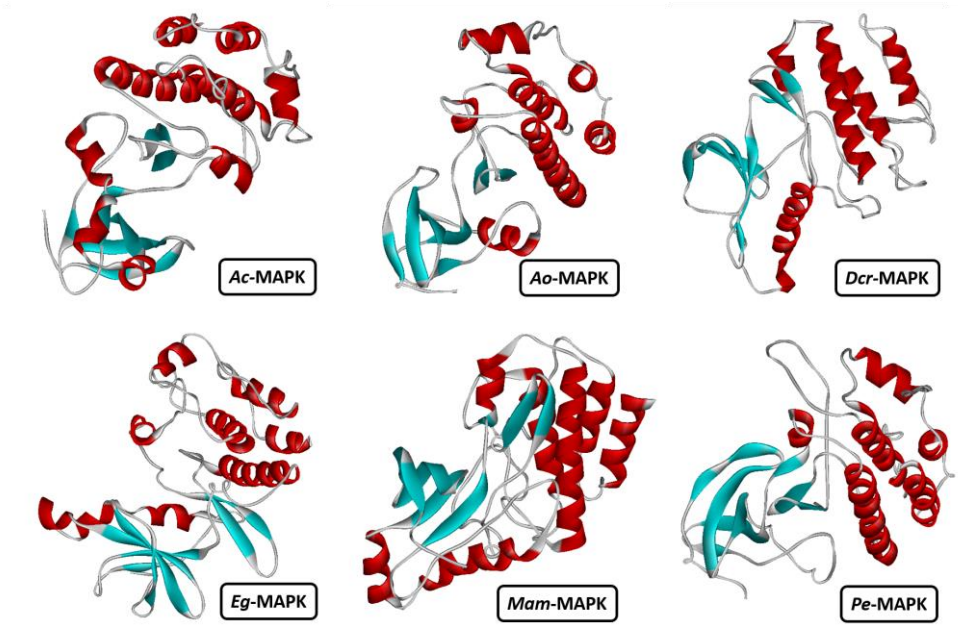

Fig. 1. Tertiary structure prediction of MAPK protein in six seed monocot plants. The α-helix is shown helix-shaped in red, the beta sheet wide ribbon-shaped in blue, and the random coil line-shaped in gray. The tertiary structure is shown as a solid ribbon model by Weblab Viewerlite 4.2.

REFERENCES

[1] C. Widmann, S. Gibson, M. B. Jarpe, and G. L. Johnson, "Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human," Physiological reviews, vol. 79, pp. 143-180. 1999. https://doi.org/10.1152/physrev.1999.79.1.143

[2] M. Behar, and A. Hoffmann, "Understanding the temporal codes of intra-cellular signals," Current opinion in genetics & development, vol. 20, pp. 684-693, 2010. https://doi.org/10.1016/j.gde.2010.09.007

[3] J. E. Purvis, and G. Lahav, "Encoding and decoding cellular information through signaling dynamics," Cell, vol. 152, pp. 945-956, 2013. https://doi.org/10.1016/j.cell.2013.02.005

[4] J. G Albeck, G. B. Mills, and J. S. Brugge, "Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals," Molecular cell, vol. 49, pp. 249-261, 2013. https://doi.org/10.1016/j.molcel.2012.11.002

[5] L. Chang, and M. Karin, "Mammalian MAP kinase signalling cascades," Nature, vol. 410, pp. 37-40, 2001. https://doi.org/10.1038/35065000

[6] R. J. Davis, "Signal transduction by the JNK group of MAP kinases," Inflammatory Processes, pp. 13-21, 2000. https://doi.org/10.1007/978-3-0348-8468-6_2

[7] G. L. Johnson, and R. Lapadat, "Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases," Science, vol. 298, pp. 1911-1912, 2002. https://doi.org/10.1126/science.1072682

[8] Werlen, G., B. Hausmann, D. Naeher, and E. D. Palmer, "Signaling life and death in the thymus: timing is everything," Science, vol. 299, pp. 1859-1863, 2003. https://doi.org/10.1126/science.1067833

[9] W. Peti, and R. Page, "Molecular basis of MAP kinase regulation," Protein science, vol. 22, pp. 1698-1710, 2013. https://doi.org/10.1002/pro.2374

[10] Y. Liu, E. G. Shepherd, and L. D. Nelin, "MAPK phosphatases—regulating the immune response," Nature Reviews Immunology, vol. 7, pp. 202-212, 2007.

https://doi.org/10.1038/nri2035
[11] J. S. C. Arthur, and S. C. Ley, "Mitogen-activated protein kinases in innate immunity," Nature Reviews Immunology, vol. 13, pp. 679-692, 2013.
https://doi.org/10.1038/nri3495
[12] J. M. Kyriakis, and J. Avruch, "Mammalian MAPK signal transduction pathways activated by stress and inflammation: a 10-year update," Physiological reviews, vol. 92, pp. 689-737, 2012.
https://doi.org/10.1152/physrev.00028.2011
[13] M. J. Robinson, and M. H. Cobb, "Mitogen-activated protein kinase pathways," Current opinion in cell biology, vol 9, pp. 180-186, 1997.
https://doi.org/10.1016/S0955-0674(97)80061-0
[14] F. Banuett, "Signalling in the yeasts: an informational cascade with links to the filamentous fungi," Microbiology and Molecular Biology Reviews, vol. 62, pp. 249-274, 1998.
https://doi.org/10.1128/MMBR.62.2.249-274.1998
[15] I. Herskowitz, "MAP kinase pathways in yeast: for mating and more," Cell, vol. 80, pp. 187-197, 1995.
https://doi.org/10.1016/0092-8674(95)90402-6
[16] D. E. Levin, and B. Errede, "The proliferation of MAP kinase signaling pathways in yeast," Current opinion in cell biology, vol. 7, pp. 197-202, 1995.
https://doi.org/10.1016/0955-0674(95)80028-X
[17] E. Leberer, D. Y. Thomas, and M. Whiteway, "Pheromone signalling and polarized morphogenesis in yeast," Current opinion in genetics & development, vol. 7, pp. 59-66, 1997.
https://doi.org/10.1016/S0959-437X(97)80110-4
[18] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, "Protein Identification and Analysis Tools on the ExPASy Server," in The Proteomics Protocols Handbook, J. M. Walker, Ed. Humana Press: Totowa, NJ., 2005, pp. 571-607.
https://doi.org/10.1385/1-59259-890-0:571
[19] C. Geourjon, and G. Deleage, "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments," Computer applications in the biosciences: CABIOS, vol. 11. pp. 681-684, 1995.
https://doi.org/10.1093/bioinformatics/11.6.681
[20] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," Nature protocols, vol. 10, pp. 845-858, 2015.
https://doi.org/10.1038/nprot.2015.053
[21] J. Kyte, and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," Journal of Molecular Biology, vol. 157, pp. 105-132, 1982.
https://doi.org/10.1016/0022-2836(82)90515-0
[22] S. Seddigh, "Proteomics analysis of two heat shock proteins in insects," Journal of Biomolecular Structure and Dynamics, vol. 37, pp. 2652-2668, 2019.
https://doi.org/10.1080/07391102.2018.1494632
[23] C. Combet, C. Blanchet, C. Geourjon, and G. Deleage, "NPS@: network protein sequence analysis," Trends in biochemical sciences, vol. 25, pp. 147-150, 2000.
https://doi.org/10.1016/S0968-0004(99)01540-6
[24] C. A. Orengo, F. M. Pearl, and J. M. Thornton, "The CATH domain structure database," Methods of Biochemical Analysis, vol. 44, pp. 249-272, 2003.
https://doi.org/10.1002/0471721204.ch13
[25] J. Söding, "Protein homology detection by HMM–HMM comparison," Bioinformatics, vol. 21, pp. 951-960, 2005.
https://doi.org/10.1093/bioinformatics/bti125