Turkish Journal of Agriculture and Forestry

Volume 48 | Number 2 Article 2

4-1-2024

In-silico analysis for Structural and Functional Characterization of Phosphorus-Starvation Tolerance 1 (PSTOL1) Gene

Quratulain Mehdi KHAN

Muhammad AQEEL

Maryam MURTAZA

Wajya AJMAL

Muhammad UZAIR

See next page for additional authors

Recommended Citation

KHAN, Quratulain Mehdi; AQEEL, Muhammad; MURTAZA, Maryam; AJMAL, Wajya; UZAIR, Muhammad; FIAZ, SAJID; ATTIA, Kotb A.; ABUSHADY, Asmaa M.; KIMIKO, Itoh; KHAN, Muhammad Ramzan; and ALI, Ghulam Muhammad (2024) "In-silico analysis for Structural and Functional Characterization of Phosphorus-Starvation Tolerance 1 (PSTOL1) Gene," *Turkish Journal of Agriculture and Forestry*: Vol. 48: No. 2, Article 2. https://doi.org/10.55730/1300-011X.3172
Available at: https://journals.tubitak.gov.tr/agriculture/vol48/iss2/2

In-silico analysis for Structural and Functional Characterization of Phosphorus-Starvation Tolerance 1 (PSTOL1) Gene

Authors

Quratulain Mehdi KHAN, Muhammad AQEEL, Maryam MURTAZA, Wajya AJMAL, Muhammad UZAIR, SAJID FIAZ, Kotb A. ATTIA, Asmaa M. ABUSHADY, Itoh KIMIKO, Muhammad Ramzan KHAN, and Ghulam Muhammad ALI

Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2024) 48: 169-182

doi:10.55730/1300-011X.3172

In silico analysis of the structural and functional characterization of the *phosphorus-starvation tolerance 1 (PSTOL1)* gene

Quratulain Mehdi KHAN[1], Muhammad AQEEL[1], Maryam MURTAZA[1], Wajya AJMAL[1], Muhammad UZAIR[1], Sajid FIAZ[2,*], Kotb A. ATTIA[3], Asmaa M. ABUSHADY[4,5], Itoh KIMIKO[6], Muhammad Ramzan KHAN[7], Ghulam Muhammad ALI[7]

[1]Bioinformatics Laboratory, National Institute for Genomics and Advanced Biotechnology, National Agricultural Research Centre (NARC), Park Road, Islamabad, Pakistan

[2]Department of Plant Breeding and Genetics, The University of Haripur, Haripur, Pakistan

[3]Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia

[4]Biotechnology School, Nile University, 26th of July Corridor, Sheikh Zayed City, Giza, Egypt

[5]Department of Genetics, Agriculture College, Ain Shams University, Cairo, Egypt

[6]Institute of Science and Technology, Niigata University, Ikarashi-2, Nishi-ku, Niigata, Japan

[7]National Institute for Genomics and Advanced Biotechnology, National Agricultural Research Centre (NARC), Park Road, Islamabad, Pakistan

Received: 01.08.2023 • **Accepted/Published Online:** 04.12.2023 • **Final Version:** 01.04.2024

Abstract: As an important macro element for all living cells, phosphorus is essential in agricultural production systems and is required in large quantities by elite varieties of crops to maintain yields. Approximately 70% of the worldwide cultivated land suffers from phosphorous deficiency, and it has recently been estimated that the worldwide phosphorous resources will be shattered by the end of this century, thereby increasing the need to develop phosphorus-efficient crops. A greater understanding of how plants can maintain yield with lower phosphorous availability is highly desirable to both breeders and farmers. For this research, we focused on the phosphorus-starvation tolerance 1 (*PSTOL1*) gene, which is known to be involved in enhancing early root growth, thereby enabling plants to acquire more phosphorus and other nutrients and enhancing grain yield in phosphorus-deficient soil. As there is no reported structure and function of the *PSTOL1* gene, this project involves a distinct set of opportunities and challenges and requires different approaches to model the interaction between *PSTOL1* and the gene phosphorous uptake 1 (*PUP1*). This article covers the modeling, docking, and simulation of *PSTOL1* to check the protein stability and its behavior over time. The physiochemical properties were ascertained, a phylogenetic tree was constructed to find the evolutionary relationship, and the conserved domains were analyzed for functional annotation. This study reports that the advancement of the *PSTOL1*-mediated phosphorous uptake 1 (*PUP1*) signaling cascade using structural bioinformatics is a potent biological mechanism against phosphorous starvation in wheat.

Key words: wheat, protein, bioinformatics, crop, evolution

1. Introduction

Wheat (*Triticum aestivum*) is a dominant crop in temperate countries being used for human food and livestock feed. Its success depends partly on its adaptability and high yield potential. Wheat contributes essential amino acids, minerals, vitamins, beneficial phytochemicals, and dietary fiber components to the human diet, and these are particularly enriched in whole-grain products. Worldwide, wheat is counted among the 'big three' cereal crops, with over 600 million tons harvested annually (Shewry, 2009). Out of 121 wheat-growing countries, Pakistan ranks eighth in terms of production in the world but 29th in terms of yield per unit area and is presently producing over 25 million tons due to the consolidated efforts of farmers ; however, it is still low compared to other countries. Wheat is a leading food grain of Pakistan and occupies a central position in the agricultural policies of the country. It contributes 9.1% to value added in agriculture and 1.7% to GDP (Mehmood et al., 2023).

There are many physiological, agronomic, socioeconomic, political, and management factors responsible for the low wheat yields in Pakistan. Favorable weather, meaning sufficient rainfall and cool temperatures, are essential requirements for grain formation and grain maturity. Also, research has found that an increase in nitrogen and phosphorus and a decrease in the use

* Correspondence: muhammadaqeel@narc.gov.pk

of tillage and irrigation enhances wheat productivity (Bashir et al., 2005). Research conducted in Punjab, Pakistan concluded that the fertilizers nitrogen (urea) and phosphorus (diammonium phosphate) are the two main factors involved in increasing yield (Mehmood et al., 2018). Therefore, phosphorus (P) is considered to be a vital nutrient required by wheat crops. P supports plant development throughout its growth from seedling to maturity and helps in photosynthesis, energy storage, and cell division. Wheat takes up 0.5–0.6 pounds of P_2O_5 (an available form of P for the plant) per metric ton (Jin et al., 2020), while a deficiency of P in the soil will cause plants to suffer from mineral shortage. The obvious signs of P deficiency in wheat include stunted growth, stems and leave turning purple, reduced root systems, and poor tillering (López-Bucio et al., 2002). As time passes, it is becoming more and more urgent to be able to grow wheat crops in P deficient soil.

The *PSTOL1* and *PUP1* genes are necessary for P solubilization and better root systems in wheat. Successful *PSTOL1* gene simulation and docking with *PUP1* is vital for the availability of P in plants; thus, a comprehensive understanding of the *PSTOL1* gene using bioinformatics studies can help disclose its functions. For this study, bioinformatics tools were used to investigate the 3-dimensional structure and docking of the *PSTOL1* gene along with a molecular dynamics (MD) simulation and its evolutionary relationship in wheat.

2. Material and methods

2.1. Sequence retrieval

The amino acid sequences of the *PSTOL1* gene (QKY76792.1) and the *PUP1* gene were downloaded from the NCBI database (*https://www.ncbi.nlm.nih.gov/protein*) in FASTA format. The downloaded sequences were used for the in silico analysis.

2.2. Physiochemical characterization

After sequence retrieval, the protein was subjected to physiochemical characterization analysis. Physiochemical characterization of protein in the design and development phase is significant as it helps in subsequent wet lab studies (Asad et al., 2018). This was accomplished by calculating several parameters, including grand average of hydropathicity (GRAVY) index, instability index, theoretical isoelectric point (pI) value, average aliphatic index value and molecular weight (Mw). The analytics were carried out using the Protparam online server (*https://web.expasy.org/protparam/*) (Gasteiger et al., 2005).

2.3. Subcellular localization predictions

Further, a comparative subcellular localization was performed using the 4 following online tools: (1) CELLO v2.5 (*http://cello.life.nctu.edu.tw/*) (Yu et al., 2006), (2) CELLO2GO (*http://cello.life.nctu.edu.tw/cello2go/*) (Yu et al., 2014), (3) EuLoc (*http://euloc.mbc.nctu.edu.tw/*) (Chang et al., 2013), and (4) Plant-mSubP (*http://bioinfo.usu.edu/Plant-mSubP/*) (Sahu et al., 2020).

2.4. Topological analysis

The number of transmembrane helices in the target protein was evaluated using TMHMM v2.0 (*http://www.cbs.dtu.dk/services/TMHMM/*) (Krogh et al., 2001) and HMMTOP v2.0 (*http://www.en.zim.hu/hmmtop/*) (Tusnady & Simon, 2001). Computing transmembrane helices is vital as protein harboring values less than or equal to 2 are more feasible for cloning and expression analysis (Käll et al., 2004).

2.5. Secondary structure prediction

A protein's secondary structure acts as an information bridge between the primary sequence and the tertiary structure. An accurate 8-state secondary structure prediction gives more precision and higher resolution for structure-based property analyses. Also, the biological activity of a protein largely depends on its folding pattern and amino acid sequence. PSIPRED v4.0 was used for secondary structure prediction (*http://bioinf.cs.ucl.ac.uk/psipred/*) (McGuffin et al., 2000).

2.6. Tertiary structure prediction

An in silico approach was used to determine the structure of structurally unknown proteins. BlastP (*https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp*) was used for finding sequence similarities in the Protein Data Bank (PDB) database (Sussman et al., 1998). The sequence with the lowest e-value, the maximum coverage, and the highest sequence identity was selected as a template for the query sequence. The protein tertiary structure was built using several tools: (1) phyre2 v2.0 (Protein Homology/analogy Recognition Engine version 2), which predicts 3D protein models by using advanced remote homology detection methods (*http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index*) (Kelley et al., 2015), (2) the SWISS MODEL (*https://swissmodel.expasy.org/*), a server that generates protein secondary and tertiary structures, functions, contact maps, solvent accessibility, and binding sites (Schwede et al., 2003), (3) the Intfold2 (*https://www.reading.ac.uk/bioinf/IntFOLD/*) server, which generates a protein's 3D structure, identifies disordered regions, evaluates the quality of 3D models, detects disordered regions, identifies structural domain boundaries, and detects likely ligand binding site residues (McGuffin et al., 2015), (4) I-Tasser v5.1, a comprehensive platform for structure and function prediction based on iterative structural assembly simulations and multiple threading alignments (*http://zhanglab.ccmb.med.umich.edu/I-TASSER/*) (Yang et al., 2015), and (5) CPHmodels v3.0, an easy protein homology modeling (HM) server in which template recognition is based on profile–profile alignment guided by secondary structures and exposure predictions. (*http://www.cbs.dtu.dk/services/CPHmodels/*) (Nielsen et al., 2010).

2.7. Model validation and optimization

Each of the predicted structures underwent a structure evaluation analysis using the following tools: (1) PROCHECK (PDBSum generate) (*http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html*) (Laskowski, 2001), (2) ProSA v4.0 (*https://prosa.services.came.sbg.ac.at/prosa.php*) (Wiederstein & Sippl, 2007), (3) ERRAT (*https://www.doe-mbi.ucla.edu/errat/*) (Colovos & Yeates, 1993), and (4) Verify3D (*https://www.doe-mbi.ucla.edu/verify3d/*) (Eisenberg et al., 1997). The structure with the best results was selected and energetically optimized through UCSF Chimera v1.17 (*https://www.cgl.ucsf.edu/chimera/*) (Pettersen et al., 2004). Gasteiger charges were applied and the steepest descent and conjugate gradient were set to 750 steps with a step size of 0.02 Å under a tripos force field. The backbone conformation of the protein structure was estimated by using a Ramachandran plot. To determine the compatibility of the atomic model (3D) with its amino acid sequence (1D), Verify3D was used, in which higher scores indicate a better quality structure. Nonbonded interactions between different atom types were analyzed using ERRAT.

2.8. Phylogenetic analysis

An evolutionary analysis of the *PSTOL1* protein was carried out. In this analysis, the *PSTOL1* protein of *Triticum aestivum* was compared with the *PSTOL1* protein sequences of other species of *Triticum* and *Oryza sativa.* These sequences were taken from Uniprot (*https://www.uniprot.org/*). The comparison was made by using multiple sequence alignment tools from Clustal Omega (*https://www.ebi.ac.uk/Tools/msa/clustalo/*) (Sievers & Higgins, 2014). The analysis involved 5 sequences, including the target protein. The *Triticum* species that were selected for the *PSTOL1* protein sequence comparison were *Triticum dicoccoides* (QKY76794.1), *Triticum monococcum subsp. aegilopoides* (QKY76797.1), *Triticum monococcum* (QKY76796.1), and *Triticum dicoccoides var. namuricum* (QKY76795.1). In addition, MEGA7 (molecular evolutionary genetic analysis) (*https://www.megasoftware.net/*) (Kumar et al., 2016) was used for phylogenetic tree construction by the neighbor-joining method.

2.9. Functional analysis and protein-protein interaction study

For functional annotation, the conserved domain of the *PSTOL1* protein sequence was identified using the following 3 tools: (1) NCBI CDD v3.20 (*https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi*), which identifies the conserved domain by employing multiple sequence alignment of ancient domains and full length proteins (Marchler-Bauer et al., 2015), (2) DeepGOWeb v1.0.15 (*https://deepgo.cbrc.kaust.edu.sa/deepgo/*), a web server for predicting protein functions based on protein sequence using the DeepGOPlus v1.0.1 (*https://deepgo.cbrc.kaust.edu.sa/deepgo/*) method that uses deep convolutional neural networks to learn sequence features and combines predictions with sequence similarity-based predictions (Kulmanov et al., 2021), and (3) ScanProsite (*https://prosite.expasy.org/scanprosite/*), which consists of manually created rules that can automatically generate annotation in the UniProtKB/Swiss-Prot format based on PROSITE (*https://prosite.expasy.org/*) motifs (De Castro et al., 2006). In addition, the STRING (search tool for the retrieval of interacting genes/proteins) v10.0 (*http://string-db.org/*) database was used to identify the interacting network of prioritized proteins at the cellular level. The STRING database is designed to provide critical assessment and integration of protein–protein interactions, including direct (physical) as well as indirect (functional) associations (Szklarczyk et al., 2015).

2.10. Phosphorylation site prediction

The phosphorylation sites of *PSTOL1* were predicted using MusiteDeep software, which is a deep-learning framework for protein posttranslational modification site prediction (*https://www.musite.net/*) (Wang et al., 2017).

2.11. Docking and simulation study

Molecular interactions, including protein–protein, enzyme–substrate, protein–nucleic acid, drug–protein, and drug–nucleic acid, play important roles in many essential biological processes, such as signal transduction, transport, cell regulation, gene expression control, enzyme inhibition, antibody–antigen recognition, and even the assembly of multidomain proteins. These interactions very often lead to the formation of stable protein–protein or protein–ligand complexes that are essential to perform the necessary biological functions. Knowledge of suitable cavities in the protein structure capable of binding to other molecules is required to perform docking. The binding pocket and active residues within binding sites were predicted using ClusPro v2.0 (Kozakov et al., 2017). The docked complex with the best binding interactions was selected and evaluated further for its dynamics. The MD simulation of the target protein was performed with the help of AMBER14 (assisted model building with energy refinement) v14 to reveal the mechanistic, dynamics, and stability details of the target enzyme inhibitor complex (Ahmad et al., 2017). The initial parameters for the complex were prepared using Visual Molecular Dynamics (VMD) v1.9.4 (Humphrey et al., 1996). Additionally, root mean square fluctuation (RMSF), root mean square deviation (RMSD), and principal component analysis (PCA) were visualized using Bio3D v2.4-1.9000 (Grant et al., 2021).

3. Results

In this study, the protein sequence of wheat was explored for its structural and functional evaluation. Docking and simulation were carried out to check the stability of *PSTOL1*. The step-wise procedure followed in the study

is presented in Figure 1. P, an essential macronutrient, is a prerequisite for various plant-growth mechanisms including root establishment/development, early/late vegetative stage development, and reproductive stage development. Wheat usually takes up 0.5–0.6 pounds of P per metric ton, and a deficient supply of P can lead plants to suffer.

3.1. Physiochemical characterization

The *PSTOL1* protein was physiochemically characterized using Protparam (*https://web.expasy.org/protparam/*). The GRAVY value is the measure of the hydrophobic or hydrophilic nature of the protein. For most proteins, the GRAVY values range from –2 to +2, in which the negative value indicates a hydrophobic nature. For the *PSTOL1* protein, the GRAVY and instability indices were –0.178 and 30.31, respectively, indicating that it is a hydrophobic and highly stable protein. The instability index of proteins is predicted by calculating the weighted sum of dipeptides that occur more frequently in unstable proteins when compared to stable proteins. An instability index greater than 40 indicates an unstable protein. The calculated theoretical pI value was 6.24, indicating the acidic nature of the target proteins. The average aliphatic index value for the protein was 81.97, indicating high thermal stability. The calculated Mw of the target protein was 32109.69 (Table 1).

3.2. Subcellular localization prediction

The *PSTOL1* protein was analyzed for subcellular localization using 4 online tools, as described in section 2.3. Among these tools, CELLO, CELLO2GO, and EuLoc showed that the protein is located in the plasma membrane while Plant-mSubP indicated that the protein is located in the cell membrane (Table 2).

3.3. Structure prediction and model validation

The *PSTOL1* protein was investigated using HM to determine its molecular structure and functional properties. At first, the availability of experimental structures was explored, but the protein was found to have no experimentally determined structure. In such cases, a comparative structure modeling approach can be useful for structure modeling. Suitable templates for HM were available for nominated proteins. The 6ctha protein from *Chloroherpeton thalassium* was used as a template on the basis of having the highest Z-score to predict structure. The selected template is involved in signal transduction, ADP binding, and microtubule motor activity. The 3D structure of the protein was predicted using the comparative structure prediction approach. The procedure to find the dimer structure of the generated model is shown in Figure 1. To select the optimum model, all of the structures predicted by the different structure prediction tools were evaluated. The most reliable models were created by I-TASEER, as shown in Figure 2.

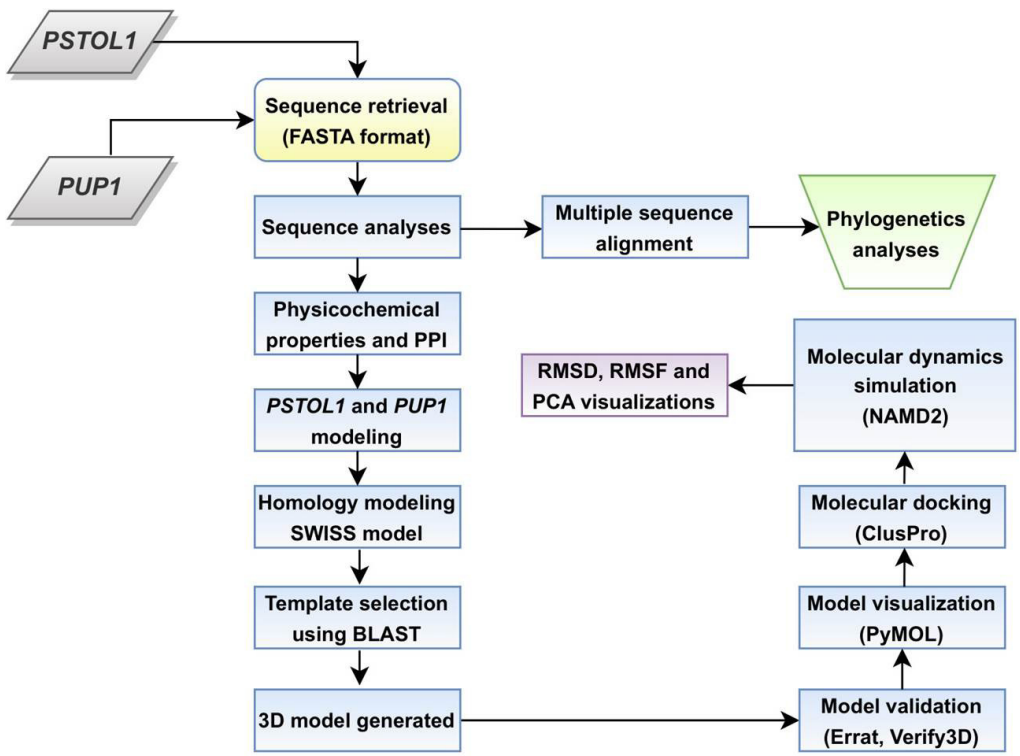

Figure 1. Schematic workflow highlighting the major steps employed.

Table 1. Physiochemical characteristics of the *PSTOL1* protein.

No. of amino acids	Molecular weight (Da)	Theoretical pI	Instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
289	32109.69	6.24	30.31	81.97	–0.178

Table 2. Prediction of the subcellular localizations of the *PSTOL1* protein.

Tool	**Subcellular localization**
CELLO	Plasma membrane
CELO2GO	Plasma membrane
EuLoc	Plasma membrane
Plant-mSubP	Cell membrane

The protein was subjected to transmembrane helices analysis using TMHMM and HMMTOP. From the prediction of both tools, it was revealed that the protein has no transmembrane helices (Table 3). Computing the number of transmembrane helices was important, as proteins having a large number of transmembrane helices are difficult to clone and express.

3.4. Phylogenetic/function analysis and protein interaction study

The functional annotation of hypothetical proteins was analyzed by searching protein families based on conserved domains. It was observed that these proteins were present in different functional categories, including catalytic activity, transferase activity, phosphorus metabolic process, protein modification process, response to stimulus, and metabolic process. The *PSTOL1* protein was found to have a histone H4 domain, as shown in Figure 3A. It was observed that *PSTOL1* belongs to the major facilitator superfamily (H4), which is one of four histones, with H2A, H2B, and H3, that form the eukaryotic nucleosome core, and with H3, plays a central role in nucleosome formation; histones bind to DNA and wrap the genetic material into "beads on a string". H4 plays a vital role in the inheritance of specialized chromosome structures and the control of gene activity. Evolutionary studies were performed using multiple sequence alignment and the construction of a phylogenetic tree. From the multiple sequence alignment, it can be understood that the protein sequences are 100% conserved among *Triticum aestivum* (QKY76792.1), *Triticum monococcum* (QKY76796.1), and *Triticum monococcum subsp. Aegilopoides* (QKY76797.1), but the protein sequences of *Triticum dicoccoides* (QKY76794.1) and *Triticum dicoccoides* var. *namuricum* (QKY76795.1) showed a single residue difference in the alignment, as shown in Figure 3B. The constructed phylogenetic tree illustrates that *Triticum aestivum* (QKY76792.1), *Triticum monococcum* (QKY76796.1), and *Triticum monococcum subsp. Aegilopoides* (QKY76797.1) fall in the same clade, showing their phylogenetic closeness, while *Triticum dicoccoides* (QKY76794.1) and *Triticum dicoccoides var. namuricum* (QKY76795.1) are expressed as a separate branch of the tree, indicating they are phylogenetically distant from the other 3 sequences. The constructed phylogenetic tree has a branch length of 0.00693895. The evolutionary distances were computed using the Poisson correction method and were in the units of the number of amino acid substitutions per site (Figure 4).

To understand the cellular interactome of the target protein, a STRING analysis was carried out. The *PSTOL1* protein (*A0A3BGV69*, highlighted with red in Figure 5) showed 10 interacting partners in its protein network. Furthermore, *PSTOL1* protein possesses one transmembrane helix, which additionally favors easy purification and expression.

3.5. Phosphorylation site prediction

The phosphorylation site prediction revealed 8 phosphorylation sites for the *PSTOL1* protein, as shown in Figure 6A. According to the predicted results, phosphorylation occurred at residues 4 (serine), 75 (tyrosine), 82 (serine), 117 (tyrosine), 163 (serine), 170 (threonine), 173 (tyrosine), and 180 (serine), as shown graphically in Figure 6B.

3.6. Molecular docking

The molecular docking was then applied to the best-modeled structure. A hypothetical protein was chosen as the domain protein for the protein–protein docking. The putative protein was downloaded to participate in a type of kinase processes. This potential *PUP1* protein possessed one domain involved in the P uptake process.

The model with the most precise binding-site interface and the model with the largest percentage of nativelike contacts, out of 180 and 230 submissions, respectively, were both predicted by ClusPro, the first completely automated, web-based software for the computational docking of protein structures. Utilizing solely the information provided by

Table 3. Prediction of the transmembrane helices in the *PSTOL1* protein.

Tool	Number of transmembrane helices
HMMTOP	0
Phobius	0
TMHMM	0

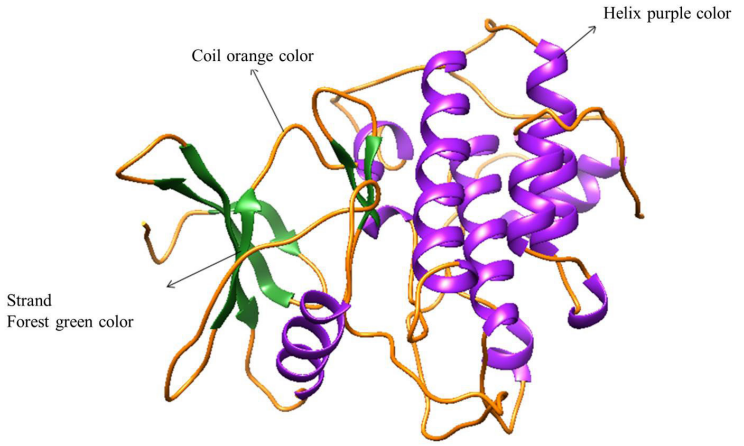

Figure 2. Predicted 3-dimensional structure of *PSTOL1* based on the top analysis result.

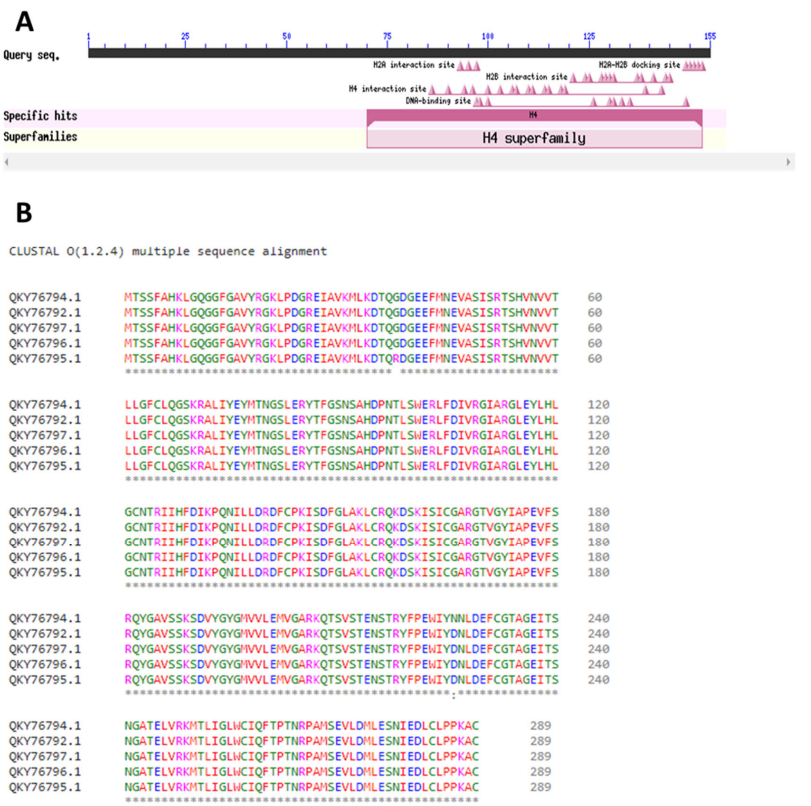

Figure 3. (A) The conserved domain prediction. (B) The multiple sequence alignment of *Triticum aestivum* (*PSTOL1* protein) with other Triticum species (*PSTOL1* protein).

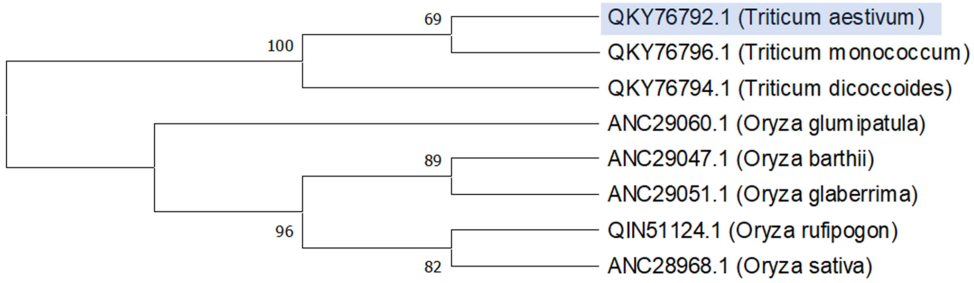

Figure 4. The phylogenetic tree constructed using MEGA7. The evolutionary history was inferred using the neighbor-joining method. The optimal tree has a sum of branch length of 0.00693895.

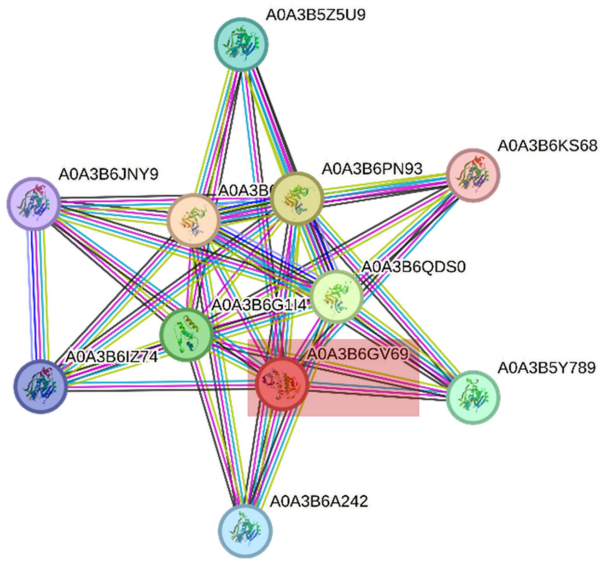

Figure 5. Interactome analysis of the *PSTOL1* gene. The *PSTOL1* protein (*A0A3BGV69*) highlighted in red shows 10 interacting partners in its protein network, including A0A3B5Y789, A0A3B5Z5U9, A0A3B6A242, A0A3B6G1I4, A0A3B6GV69, A0A3B6IZ74, A0A3B6JNY9, A0A3B6KS68, A0A3B6NN79, A0A3B6PN93, and A0A3B6QDS0.

the protein component structures and thermodynamic considerations, the ClusPro server quickly docks, filters, and ranks potential protein complexes. By default, the parameters were configured to investigate the protein's binding position within the designated active site residue, with a 10 area limit (Comeau et al., 2005). Despite the server being heavily utilized, runs typically take 4 h or less. With the use of a modified version of the particle swarm optimization algorithm, the SwarmDock technique reduces the interaction energy. The swarm optimization approach was coupled with a local search, where flexibility is represented by a linear combination of elastic network normal modes (Li et al., 2010). Depending on the size of the complex and whether the resources are shared, a typical server-based docking operation can last up to 36 h. Three straightforward requirements should be followed when uploading PDB structures of binding partners to SwarmDock: there should be a TER keyword after each chain, standard residues should be used, and there should preferably be no missing residues. The server will attempt to replace nonstandard residues and simulate missing residues or residues with missing atoms if the final two requirements cannot be met (Torchala et al., 2013). Prior to clustering, the structures are reduced, docked, minimized once more, and rescored using the centroid potential as described by Tobi (2010). These improvements significantly enhanced the algorithm's performance.

The docked complex analysis results of both programs were compared. Both docking systems often indicated that the binding would occur in the same area. A complex was

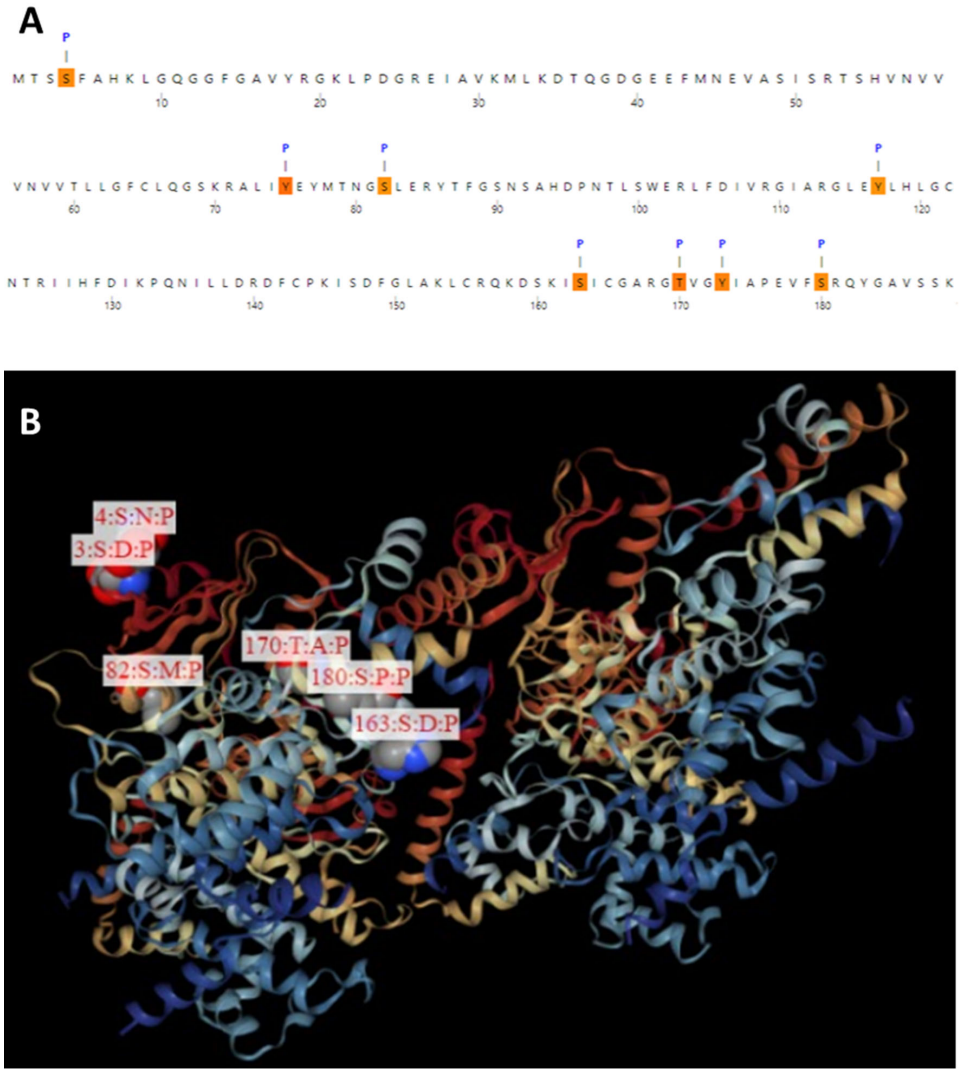

Figure 6. (A) The prediction of phosphorylation sites by the MusiteDeep software. (B) A graphical representation of the phosphorylation sites in the protein structure.

chosen based on a number of factors, including the robust protein–protein interactions. Simulated investigations were chosen to be performed on the top-scoring docked complex, as shown in Figure 7.

3.7. Molecular dynamics simulation

The docking study provides meaningful insights into the structural basis of the docked complex (Abbasi et al., 2016). However, the information provided by docking is limited to a single, static snapshot of the protein. The MD simulations were carried out to further characterize the protein–protein complex. The simulations show the dynamic behavior of the protein and also highlight the residues that play an important role in binding (Ahmad et al., 2017). Structural properties including RMSD, RMSF, and PCA were plotted as a function of time to understand the dynamics within a solvated environment. Analysis of the protein in the docked stage led to the evaluation of the structural transformation and the underlying atomic level transition (Ahmad et al., 2017). MD simulations of the selected docked complex were performed to reveal mechanistic, dynamic, and stability details of the target enzyme inhibitor complex (Stocchi et al., 2018). The initial parameters for the complex were prepared using AMBER14 (Pearlman et al., 1995). An antechamber was used to generate suitable parameters, and force field parameters were described for protein ff99SB (Özpınar et al., 2010). The MPI version of Sander within AMBER14 was used to carry out the simulations. The system was submerged in a three-point transferable intermolecular potential (TIP3P) water box with 12 Å padding distance between the protein and water box boundary, as illustrated in Figure 8. The overall charge of the system was

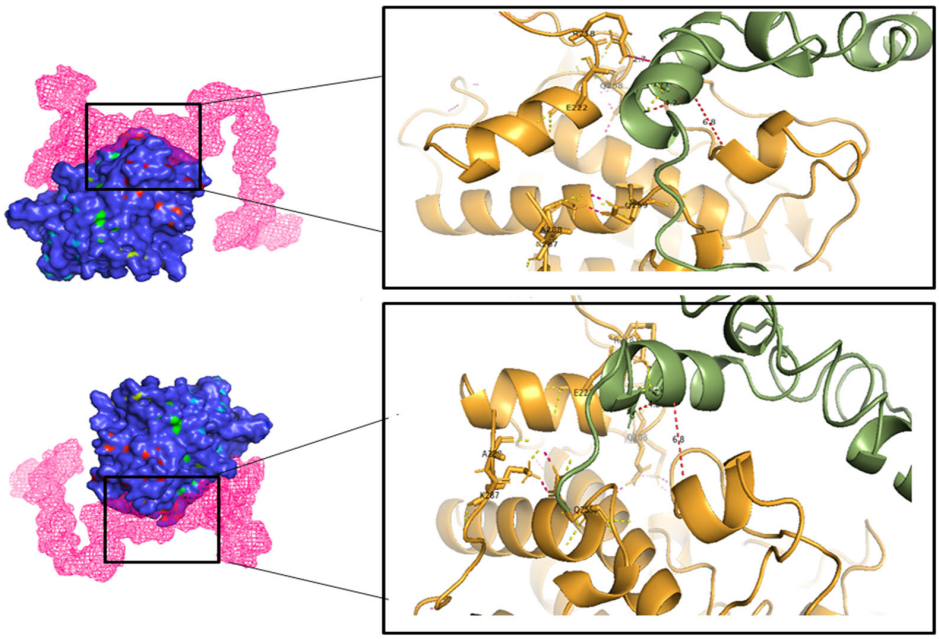

Figure 7. An illustration of different docked poses of the *PSTOL1* protein with a hypothetical protein.

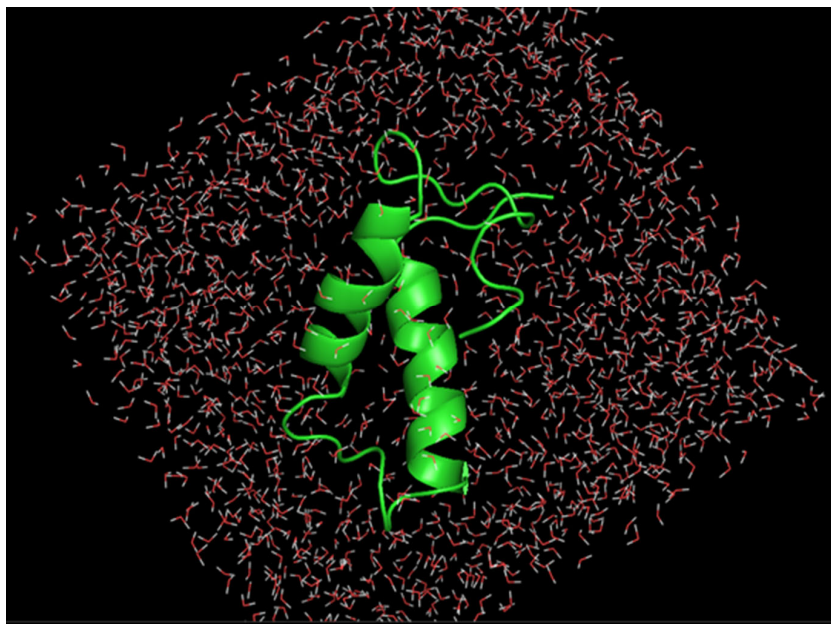

Figure 8. The solvation box surrounding the *PSTOL1* protein.

neutralized by the addition of 12 Na+ ions. The deviation of the backbone Cα atoms was monitored throughout the 10-ns simulation. The RMSD of the complex over the studied timescale is mostly stable with an average value of 3.2 Å, reaching a maximum value of 4 Å at 5 ns. Over the course of the simulation, minor structural changes were observed, as shown in Figures 9A and 9B.

3.8. Principal components based on PSTOL1 and PUP1 residual positions

To aid interpretation, a PDB format trajectory of *PSTOL1* and *PUP1* can be produced that interpolates between the most dissimilar structures in the distribution along a given principal component. This involves dividing the difference between the conformers into a number of evenly spaced

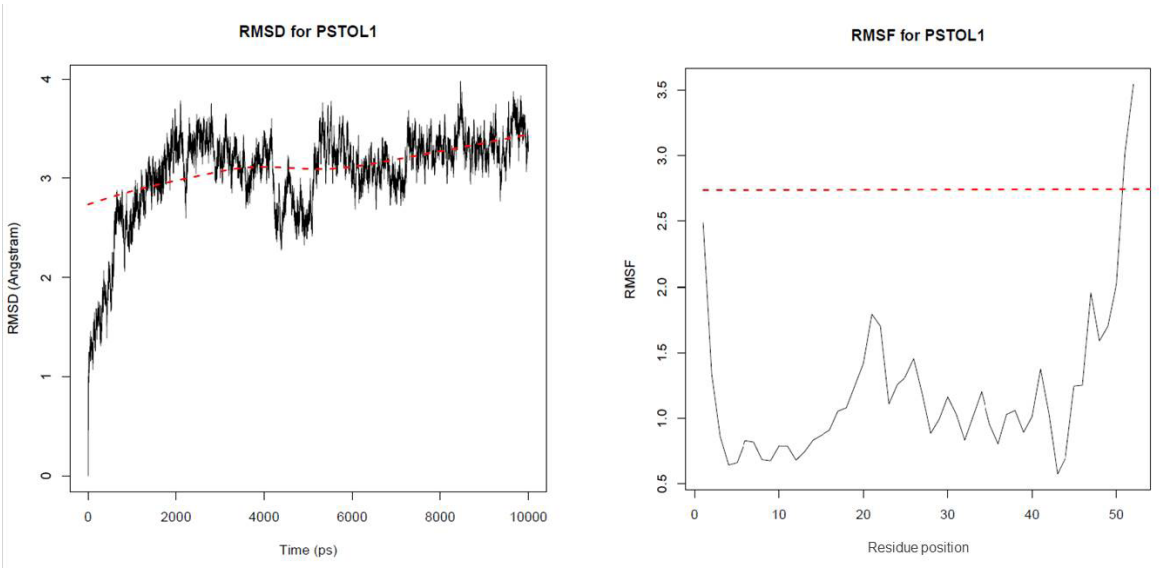

Figure 9. The statistical parameters for the top scoring docked complex over a simulation period of 10000-ps. (A) is RMSD and (B) is RMSF.

steps along the principal components, forming the frames of the trajectory (Figure 10). Such trajectories can be directly visualized in a molecular graphics program, such as VMD (Humphrey et al., 1996).

3.9. Trajectories analyses using PCA

Coarse-grained models enable us to carry out MD simulations starting from the unstabilized state and ending in the native state while consuming a fairly short CPU time. In this work, we present the results of the trajectories analyzed by PCA, as shown in Figure 10. All trajectories start with the same initial (extended) structure but with different velocities. The total time of all MD simulations was the same at ~10,000 ps. If the protein stabilizes before half of the full simulation time, the trajectory is called fast-stabilizing. If the protein spends half (or more) of the entire trajectory time to fold, the trajectory is called slow-stabilizing. If the system never folds during the entire 10000-ps MD simulation, the trajectory is called nonstabilizing.

In contrast to the fast and slow stabilization trajectories, in which the first principal component (PC) captures most of the behavior of the RMSD, the correlation between the PCs and RMSD is observed in the first three PCs in the nonstabilizing MD trajectory of (panel c). Also, the amplitudes of fluctuation along PC1 and PC2 (in panel c) are relatively similar to each other. Hence, the distribution of the captured parts of the overall fluctuations by the first few PCs is different for the nonstabilizing MD trajectory: PC1 is ~34.85%, PC2 is ~16.01%, and PC3 is ~15.13%, as shown in Figure 11. Thus, for the nonstabilizing trajectory, the first PC is not enough to depict the main features of the energy landscape.

As a measure of atomic fluctuation, the RMSF provides a means to identify structural flexibility of the rigid regions of the targets. The average Cα fluctuation for the ligand-bound protein was observed to be 0.9 Å. The maximum value of RMSF for the ligand-bound protein was 4 Å.

4. Discussion

Phosphorus deficiency is one of the factors that excessively increases farmers' costs because it must be applied as fertilizers, which are not cost effective. Plants try to survive under phosphorus deficiency by increasing their phosphorus uptake, which can be achieved by modifying their root system architecture, changing rhizospheres, interacting with microorganisms, or using the transport of internal phosphorus and mobilization (Bechtaoui et al., 2021). Over time, multiple genes have been identified in different plants that can influence their responses to many macronutrient deficiencies. With all the tools and genomes at our disposal, identifying genetic variants of many genes, even in very complex genomes such as wheat, has become more accessible. We have analyzed the interaction of *PSTOL1* and *PUP1* in wheat and highlighted their molecular dynamic behaviors involved in morphophysiological traits and root architecture.

Although the exact molecular mechanisms that lead to the occurrence, progression, and metastasis of P uptake are not yet fully understood, several genes, factors, and pathways are involved in the development of P uptake. *PSTOL1* is a member of the receptor-like kinases (RLK) family and is a calcium-independent serine-threonine kinase. The RLK family consists of 232 members (Yan et al., 2023). *PSTOL1* has a distinctive structure consisting

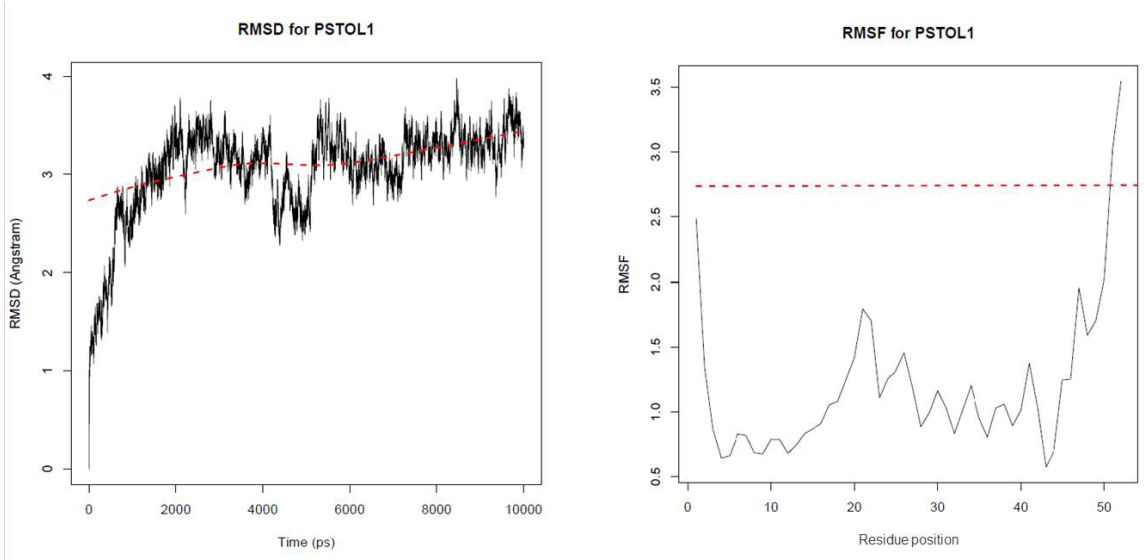

Figure 10. The PCA plot for the PSTOL1 simulation displays the relationships between different conformers, highlights the positions responsible for the major differences between PSTOL1 and PUP1 structures, and enables the interpretation and characterization of multiple interconformer relationships.

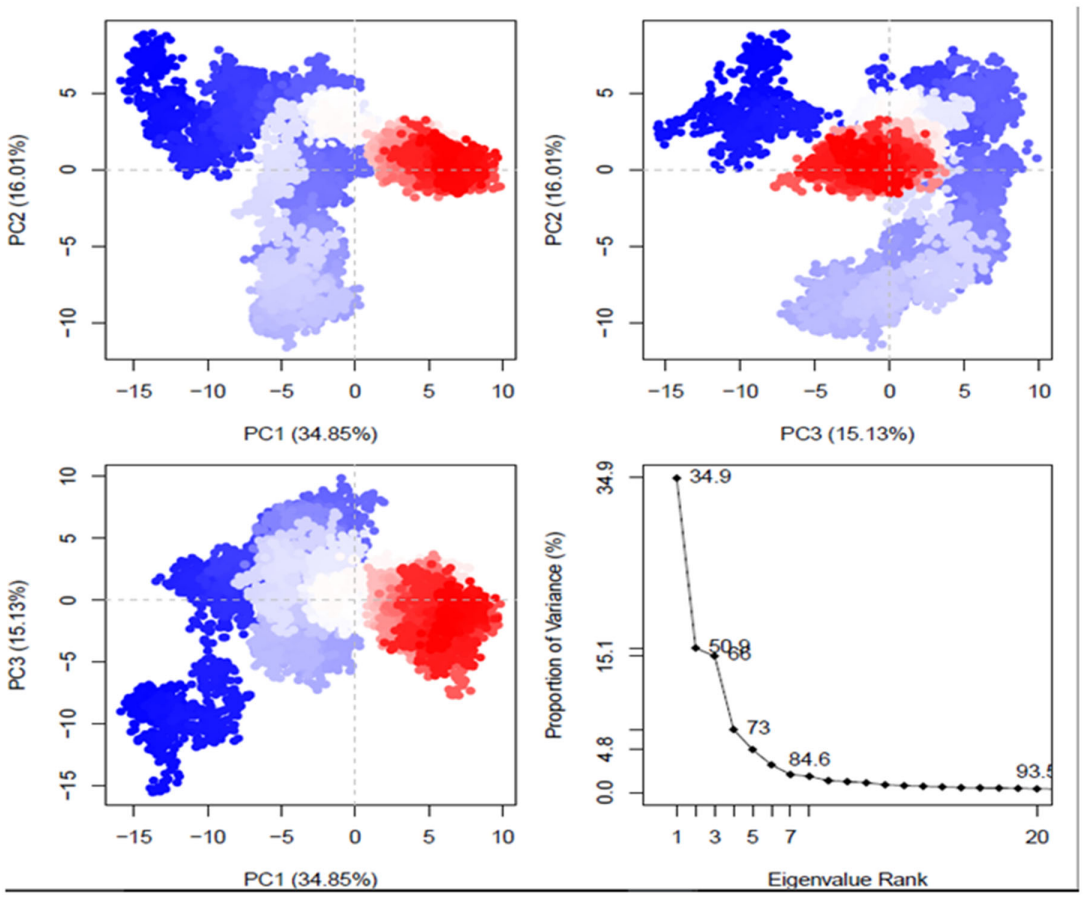

Figure 11. The PCA plot for a PSTOL1 simulation.

of one domain, namely the histone H4 domain, an important protein in the structure and function of chromatin. Previous research has suggested that *PSTOL1* overexpression is observed in P uptake and is correlated with deeper rooting. Similarly, *PUP1* is a kinase that promotes the sensing and signaling of P homeostasis, was discovered by Gamuyao et al. (2012), and was named as the *PSTOL1* gene in rice (*Oryza sativa*). It was observed that when the *PSTOL1* gene was overexpressed in rice plants grown in P-deficient soil, a significant increase in yield and biomass occurred. The *PSTOL1* gene acts by increasing the early growth of roots, enabling the plants to acquire more phosphorus from P-deficient soil (Gamuyao et al., 2012). The strategy of identifying *PSTOL1* orthologs in different crops and targeting their expression will help to develop crops that proliferate in P-deficient soil with more yield and biomass. However, the structural bioinformatics of *PSTOL1* in wheat is yet to be investigated.

The *PSTOL1* gene, in the RLK family, is responsible for early root growth in P-deficient soil and can increase grain yield exponentially. Computational biology helps to predict the binding positions of different potential ligand proteins within the active sites of target proteins as pockets. Modeling and molecular docking are often employed to discover new binding partners. Molecular docking methods aim to identify the correct position of a ligand protein within the binding or active site of the target or receptor protein to investigate the right position of the ligand with the receptor protein. The methods are also used for the prediction of the binding patterns of the ligand with the receptor protein. This study explored in silico *PSTOL1* and *PUP1* interaction predictions using the BLAST search, molecular modeling, docking, and simulation protocols.

We have identified *PUP1* as a potential complex of the main wheat kinase *PSTOL1* through large-scale in silico and focused simulation studies. P deficiency can reach alarming proportions due to the emergence of several environmental variations, such as low pH, dry soil, and high temperature, and thereby inhibit proper root system growth and efficient root architecture generation. The main kinase *PSTOL1*, which is reported as a key player among several wheat and rice species, provides an effective insight for P deficiency in crops. The naturally occurring biological protein *PUP1* present in wheat genomes has been found to be a potent activated complex with the *PSTOL1* gene. The protein was downloaded to participate in a variety of kinase processes containing one domain. The STRING network, phylogenetics, phosphorylation site prediction, molecular modeling, docking, MD simulations, RMSD, RMSF, and PCA were then applied to the *PUP1* and *PSTOL1*-docked and simulated structure involved in the root growth and architecture development. In the present study, P deficiency was positively associated with the overexpression of *PSTOL1* and *PUP1* in an activated complex state. In brief, our research highlights the potential role of *PSTOL1* genes in the efficient use of P in a deficient environment using structural bioinformatics approaches. Understanding the behaviors and dynamics of the interactions of proteins in the time frame of nanoseconds shows the stability of the protein. The binding free energy for the complex was calculated using both MGL tools. Further investigations are required to systematically characterize the function of all *PSTOL1* genes in wheat phosphorus use efficiency, but these findings are highly encouraging and *PSTOL1* appears to be a promising compound to enhance wheat crop growth. This in silico study can further assist in vivo studies of the *PSTOL1* protein.

Acknowledgments
The authors extend their appreciation for the Researchers Supporting Project (RSP-2024R369) of King Saud University, Riyadh, Saudi Arabia.

Funding
No specific funding has been received to conduct this research.

Author contributions
Quratulain Mehdi Khan and Muhammad Aqeel designed the study. Quratulain Mehdi Khan, Maryam Murtaza, and Wajya Ajmal performed the experiments and analysis. Quratulain Mehdi Khan, Muhammad Aqeel, Muhammad Uzair, and Wajya Ajmal wrote the original draft. Muhammad Ramzan Khan and Ghulam Muhammad Ali guided the whole study. Sajid Fiaz, Kotb A. Attia, Asmaa M. Abushady, and Itoh Kimiko provided technical expertise to improve the article and helped with funding acquisition. All authors review and edited the manuscript.

Data availability statement
All the relevant data are within the paper and can be obtained from the corresponding authors.

Conflicts of interest
The authors declare no conflict of interest.

References

Abbasi S, Raza S, Azam SS, Liedl KR, Fuchs JE (2016). Interaction mechanisms of a melatonergic inhibitor in the melatonin synthesis pathway. Journal of Molecular Liquids 221: 507-517. https://doi.org/10.1016/j.molliq.2016.06.034

Ahmad S, Raza S, Uddin R, Azam SS (2017). Binding mode analysis, dynamic simulation and binding free energy calculations of the MurF ligase from *Acinetobacter baumannii*. Journal of Molecular Graphics and Modelling 77: 72-85. https://doi.org/10.1016/j.jmgm.2017.07.024

Asad Y, Ahmad S, Rungrotmongkol T, Ranaghan KE, Azam SS (2018). Immuno-informatics driven proteome-wide investigation revealed novel peptide-based vaccine targets against emerging multiple drug resistant *Providencia stuartii*. Journal of Molecular Graphics and Modelling 80: 238-250. https://doi.org/10.1016/j.jmgm.2018.01.010

Bashir M, Khan D, Iqbal M (2005). An analysis of allocative efficiency of wheat growers in northern Pakistan. The Pakistan Development Review 44 (4): 643-657.

Bechtaoui N, Rabiu MK, Raklami A, Oufdou K, Hafidi M, Jemo M (2021) Phosphate-Dependent Regulation of Growth and Stresses Management in Plants. Frontiers in Plant Sciences 12: 679916. https://doi.org/10.3389/fpls.2021.679916

Chang TH, Wu LC, Lee TY, Chen SP, Huang HD et al. (2013). EuLoc: A web-server for accurately predict protein subcellular localization in eukaryotes by incorporating various features of sequence segments into the general form of Chou's PseAAC. Journal of Computer-Aided Molecular Design 27 (1): 91-103. https://doi.org/10.1007/s10822-012-9628-0

Colovos C, Yeates T (1993). ERRAT: An empirical atom-based method for validating protein structures. Protein Science 2 (9): 1511-1519.

Comeau SR, Vajda S, Camacho CJ (2005). Performance of the first protein docking server ClusPro in CAPRI rounds 3–5. Proteins: Structure, Function, and Bioinformatics 60 (2): 239-244. https://doi.org/10.1002/prot.20564

De Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS et al. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Research 34 (S2): W362-W365. https://doi.org/10.1093/nar/gkl124

Eisenberg D, Lüthy R, Bowie JU (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods in Enzymology 277: 396-404. https://doi.org/10.1016/S0076-6879(97)77022-8

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S et al. (2012). The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. Nature 488: 535-539. https://doi.org/10.1038/nature11346

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR et al. (2005). Protein identification and analysis tools on the ExPASy server. In: Walker JM (editor). The Proteomics Protocols Handbook. Springer Protocols Handbooks. Humana Press. 571-607. https://doi.org/10.1385/1-59259-890-0:571

Humphrey W, Dalke A, Schulten K (1996). VMD: Visual molecular dynamics. Journal of Molecular Graphics 14 (1): 33-38. https://doi.org/10.1016/0263-7855(96)00018-5

Jin CH, Pang DW, Min JI, Luo YL, Li HY, Yong LI, Wang ZL (2020). Improved soil characteristics in the deeper plough layer can increase grain yield of winter wheat. Journal of Integrative Agriculture. 19(5):1215-26. https://doi.org/10.1016/S2095-3119(19)62679-1

Käll L, Krogh A, Sonnhammer EL (2004). A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology 338 (5): 1027-1036. https://doi.org/10.1016/j.jmb.2004.03.016

Kelley LA, Mezulis S, Yates CM, Wass MN (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10 (6), 845-858. https://doi.org/10.1038/nprot.2015.053

Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D et al. (2017). The ClusPro web server for protein–protein docking. Nature Protocols 12 (2): 255-278. https://doi.org/10.1038/nprot.2016.169

Krogh A, Larsson B, Von Heijne G, Sonnhammer EL (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology 305 (3): 567-580. https://doi.org/10.1006/jmbi.2000.4315

Kulmanov M, Zhapa-Camacho F, Hoehndorf R (2021). DeepGOWeb: Fast and accurate protein function prediction on the (semantic) web. Nucleic Acids Research 49 (W1): W140-W146. https://doi.org/10.1093/nar/gkab373

Kumar S, Stecher G, Tamura K (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33 (7): 1870-1874. https://doi.org/10.1093/molbev/msw054

Laskowski RA (2001). PDBsum: Summaries and analyses of PDB structures. Nucleic Acids Research 29 (1): 221-222. https://doi.org/10.1093/nar/29.1.221

Li X, Moal IH, Bates PA (2010). Detection and refinement of encounter complexes for protein–protein docking: Taking account of macromolecular crowding. Proteins: Structure, Function, and Bioinformatics 78 (15): 3189-3196. https://doi.org/10.1002/prot.22770

López-Bucio J, Hernández-Abreu E, Sánchez-Calderón L, Nieto-Jacobo MF, Simpson J et al. (2002.) Phosphate availability alters architecture and causes changes in hormone sensitivity in the *Arabidopsis* root system. Plant Physiology 129 (1): 244-256. https://doi.org/10.1104/pp.010934

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Research 43 (D1): D222-D226.

McGuffin LJ, Atkins JD, Salehe BR, Shuid AN, Roche DB (2015). IntFOLD: An integrated server for modelling protein structures and functions from amino acid sequences. Nucleic Acids Research, 43 (W1): W169-W173. https://doi.org/10.1093/nar/gkv236

McGuffin LJ, Bryson K, Jones DT (2000). The PSIPRED protein structure prediction server. Bioinformatics 16 (4): 404-405. https://doi.org/10.1093/bioinformatics/16.4.404

Mehmood Q, Riaz M, Sail MH, Moeen M (2018). Identifying key factors for maximizing wheat yield: A case study from Punjab (Pakistan). Pakistan Journal of Agricultural Research 31 (4): 361-367. https://doi.org/10.17582/journal.pjar/2018/31.4.361.367

Moal IH, Chaleil RA, Bates PA (2018). Flexible protein-protein docking with SwarmDock. Protein Complex Assembly: Methods and Protocols 1764: 413-428. https://doi.org/10.1007/978-1-4939-7759-8_27

Nielsen M, Lundegaard C, Lund O, Petersen TN (2010). CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. Nucleic Acids Research 38 (S2): W576-W581. https://doi.org/10.1093/nar/gkq535

Özpınar GA, Peukert W, Clark T (2010). An improved generalized AMBER force field (GAFF) for urea. Journal of Molecular Modeling 16 (9): 1427-1440. https://doi.org/10.1007/s00894-010-0650-7

Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham III TE et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications 91 (1-3): 1-41. https://doi.org/10.1016/0010-4655(95)00041-D

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. Journal of Computational Chemistry 25 (13): 1605-1612. https://doi.org/10.1002/jcc.20084

Sahu SS, Loaiza CD, Kaundal R (2020). Plant-mSubP: A computational framework for the prediction of single- and multi-target protein subcellular localization using integrated machine-learning approaches. AoB Plants 12 (3): plz068. https://doi.org/10.1093/aobpla/plz068

Schwede T, Kopp J, Guex N, Peitsch MC (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Research 31 (13): 3381-3385. https://doi.org/10.1093/nar/gkg520

Shewry PR (2009). Wheat, Journal of Experimental Botany 60(6):1537–1553, https://doi.org/10.1093/jxb/erp058

Sievers F, Higgins DG (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Multiple Sequence Alignment Methods 1079: 105-116. https://doi.org/10.1007/978-1-62703-646-7_6

Stocchi N, Revuelta MV, Castronuovo PAL, Vera DMA, Ten HA (2018). Molecular dynamics and structure function analysis show that substrate binding and specificity are major forces in the functional diversification of eqolisins. BMC Bioinformatics 19 (1): 1-16. https://doi.org/10.1186/s12859-018-2348-2

Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J et al. (1998). Protein Data Bank (PDB): Database of three-dimensional structural information of biological macromolecules. Acta Crystallographica Section D: Biological Crystallography 54 (6): 1078-1084. https://doi.org/10.1107/S0907444998009378

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D et al. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research 43 (D1): D447-D452. https://doi.org/10.1093/nar/gku1003

Tobi D (2010). Designing coarse grained-and atom based-potentials for protein-protein docking. BMC Structural Biology. 10(1):1-1. https://doi.org/10.1186/1472-6807-10-40

Torchala M, Moal IH, Chaleil RA, Fernandez-Recio J, Bates PA (2013). SwarmDock: A server for flexible protein–protein docking. Bioinformatics 29 (6): 807-809. https://doi.org/10.1093/bioinformatics/btt038

Tusnady GE, Simon I (2001). The HMMTOP transmembrane topology prediction server. Bioinformatics 17 (9): 849-850. https://doi.org/10.1093/bioinformatics/17.9.849

Wang D, Zeng S, Xu C, Qiu W, Liang Y et al. (2017). MusiteDeep: A deep-learning framework for general and kinase-specific phosphorylation site prediction. Bioinformatics 33 (24): 3909-3916. https://doi.org/10.1093/bioinformatics/btx496

Wiederstein M, Sippl MJ (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35 (S2): W407-W410. https://doi.org/10.1093/nar/gkm290

Yan J, Su P, Meng X, Liu P (2023). Phylogeny of the plant receptor-like kinase (RLK) gene family and expression analysis of wheat RLK genes in response to biotic and abiotic stresses. BMC genomics24(1): 1-7. https://doi.org/10.1186/s12864-023-09303-7

Yang J, Yan R, Roy A, Xu D, Poisson J et al. (2015). The I-TASSER suite: Protein structure and function prediction. Nature Methods 12 (1): 7-8. https://doi.org/10.1038/nmeth.3213

Yu CS, Chen YC, Lu CH, Hwang JK (2006). Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics 64 (3): 643-651. https://doi.org/10.1002/prot.21018

Yu CS, Cheng CW, Su WC, Chang KC, Huang SW et al. (2014). CELLO2GO: A web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PloS One 9 (6): e99368. https://doi.org/10.1371/journal.pone.0099368