Identification of an binding interaction between *Curcumin derivative* and *PERK13- Proline-rich receptor-like protein kinase* protein using *In-silico* docking techniques – TO HELP PLANTS TACKLE SALT WATER

Prakash Vaithyanathan*

---

* Science Teacher, SACROSANCT SERVICES, Chennai-600041, India.

Corresponding Email: pvaithyanathan@gmail.com

Abstract

Keywords: PERK13, Protein modelling, small molecule Docking, Curcumin

Introduction

Arabidopsis thaliana, mouse-ear cress or thale cress are small flowering plants included in the cruciferae family. It comprises of various characteristics such as diploid genetics, small genome size, rapid growth cycle, and relatively low repetitive DNA content, making it a perfect model for plant genome projects.

Objective

The aim of this present *Insilico* research study is to carry out molecular drug docking studies between PERK13- Proline-rich receptor-like protein kinase and De-O-Acetylated Curcumin Di-Galactose or di–galactosylated curcumin, a derivative of curcumin . PERK13 protein is considered to play a significant role in helping plants tackle salinity levels in water .

Methods

In this study, protein modelling tools and servers are used to model the 3D structure and the same is validated using Protein structure validation tools. Automated drug docking servers were used to dock the modelled protein with the chemical compound to analyse the electrostatic (H-bond) interaction between PERK13 and De-O-Acetylated Curcumin Di-Galactose, a better water-soluble compound than curcumin. The docked structure was visualized using an advanced molecular visualization tool.

Results and Discussion

The overall results obtained from this study on De-O-Acetylated Curcumin Di-Galactose and PERK13 protein shows that De-O-Acetylated Curcumin Di-Galactose , directly binds with the active site and other potentially binding regions of PERK13. Hence, it is concluded that De-O-Acetylated Curcumin Di-Galactose could potentially play a vital role in future research related to the problem of helping the plants to tackle increased saline levels in water.

Keywords: *A. thaliana***,** Drug Docking**,** *Insilico*, De-O-Acetylated Curcumin Di-Galactose , PERK13, Protein Modelling.

Introduction

Arabidopsis is an efficient model to carry out research. Since many years, *Arabidopsis* has been used to understand about gravitropism (Kiss *et al.*, 2000),[1] genetics (Zheng 2006)[2], and genomics (Brooks *et al.* 2011)[3]. Mutants can be obtained for reverse-genetics projects from *Arabidopsis* Biological Resource Center (https://abrc.osu.edu/) or the Nottingham *Arabidopsis* Stock Centre (http://www.arabidopsis.info), which includes sequence-indexed collections of above 30,000 homozygous T-DNA insertional lines (O'Malley and Ecker 2010)[4,5] from many thousands of insertion events (Alonso *et al.*, 2003)[6]. Full-length complementary DNAs (Yamada *et al.*, 2003)[7] can also be obtained for most genes from the *Arabidopsis* Biological Resource Center. Additionally, collections of *Arabidopsis* expression vectors (Curtis and Grossniklaus, 2003; Earley *et al.* 2006)[8,9] and yeast two-hybrid vectors (Trigg *et al.*, 2017)[10] are available which can be made use of to conduct overexpression, localization, and interaction studies.

The Arabidopsis proline-rich extensin-like receptor kinase (PERK) family includes 15 predicted receptor kinases. A comprehensive analysis of expression was performed for identifying the overlapping and unique expression patterns within this family related to their phylogeny. PERK13 protein is believed to play a significant role to help plants tackle salinity levels in water.

Methodology

Protein target analysis: In this study, the gene-coded protein sequence of **PERK13 , (UniProt ID:** Q9CAL8) was retrieved from UniProt database (https://www.uniprot.org/uniprot/Q9CAL8) for performing sequence analysis and protein modelling. The sequence was converted into FASTA format. **Sequence analysis:** The retrieved protein was analysed using an UniProt Proteomics tool named MobiDB

(https://mobidb.bio.unipd.it/) for identifying the functional domains present in PERK13. **Molecular modelling**: The selected amino acid sequence of the target protein (PERK13) was converted into 3D structure using an automated protein-modelling server, SWISS MODEL (https://swissmodel.expasy.org/). The modelled structure was validated using ProCheck server (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) to analyse the Ramachandran Plot. **3D structure visualization**: The modelled protein was visualized in 3D form using an advanced molecular visualization tool called Discovery Studio software in order to clearly view the intramolecular structural details of PERK13. **Molecular drug docking**: In this study, the drug compound De-O-Acetylated Curcumin Di-Galactose was selected and was docked with the modelled PERK13 protein using an automated molecular modelling and drug docking server called PatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/) for analysing the drug binding score and the interaction between De-O-Acetylated Curcumin Di-Galactose and PERK13.

Results

```
>sp|Q9CAL8|PEK13_ARATH Proline-rich receptor-like protein kinase PERK13
OS=Arabidopsis thaliana OX=3702 GN=PERK13 PE=1 SV=1
MSDSPTSSPPAPSADSAPPPDTSSDGSAAPPPTDSAPPPSPPADSSPPPALPSLPPAVFS
PPPTVSSPPPPPLDSSPPPPPDLTPPPSSPPPPDAPPPIPIVFPPPIDSPPPESTNSPPP
PEVFEPPPPPADEDESPPAPPPPEQLPPPASSPQGGPKKPKKHHPGPATSPPAPSAPATS
PPAPPNAPPRNSSHALPPKSTAAGGPLTSPSRGVPSSGNSVPPPANSGGGYQGKTMAGFA
IAGFAVIALMAVVFLVRRKKKRNIDAYSDSQYLPPSNFSIKSDGFLYGQNPTKGYSGPGG
YNSQQQSNSGNSFGSQRGGGGYTRSGSAPDSAVMGSGQTHFTYEELTDITEGFSKHNILG
EGGFGCVYKGKLNDGKLVAVKQLKVGSGQGDREFKAEVEIISRVHHRHLVSLVGYCIADS
ERLLIYEYVPNQTLEHHLHGKGRPVLEWARRVRIAIGSAKGLAYLHEDCHPKIIHRDIKS
ANILLDDEFEAQVADFGLAKLNDSTQTHVSTRVMGTFGYLAPEYAQSGKLTDRSDVFSFG
VVLLELITGRKPVDQYQPLGEESLVEWARPLLHKAIETGDFSELVDRRLEKHYVENEVFR
MIETAAACVRHSGPKRPRMVQVVRALDSEGDMGDISNGNKVGQSSAYDSGQYNNDTMKFR
KMAFGFDDSSDSGMYSGDYSVQDSRKGSNGASSEFTRNETENRNFNNRRY
```

Fig. 1 FASTA sequence of PERK13

The above picture shows the amino acid sequence of PERK13 protein

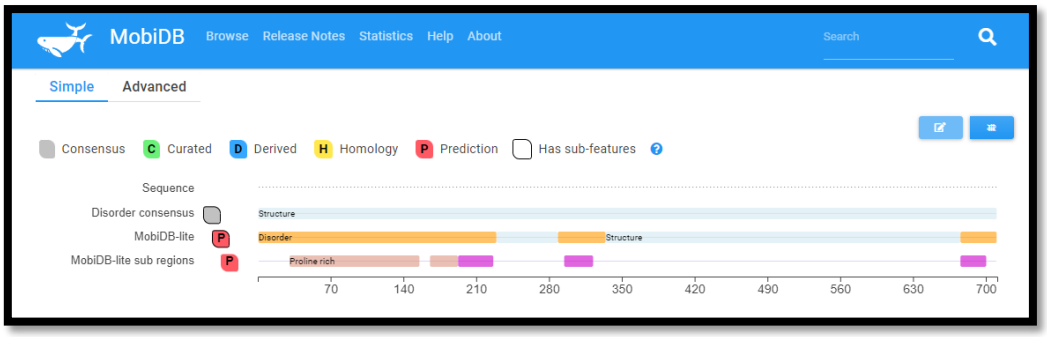
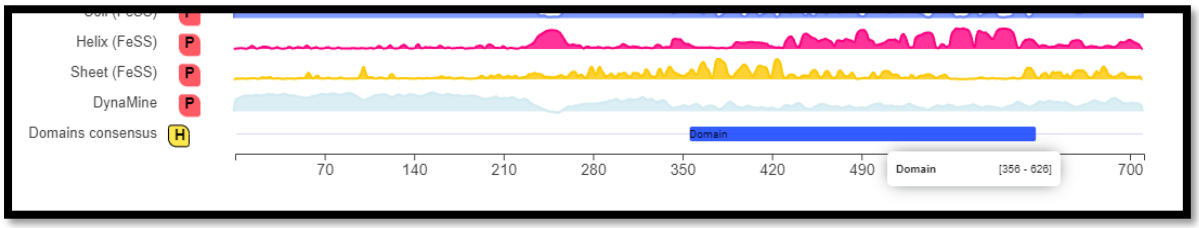


Fig. 2 Sequence analysis of PERK13 using MobiDB Database

In the above picture, the coloured region indicate the various functional domains present in PERK13.

Fig. 3 Sequence analysis of PERK13 using MobiDB Database



In the above picture, the blue coloured region represents predominant functional domains of PERK13 (356-626aa).

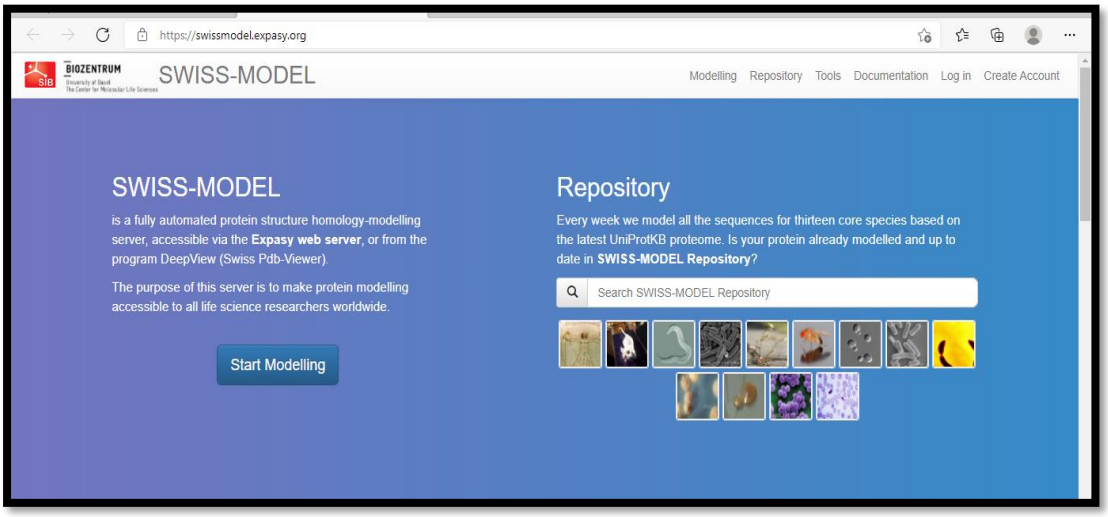


Fig. 4 Protein modelling – SWISS-MODEL server

The above picture is the Home Page of SWISS-MODEL server.

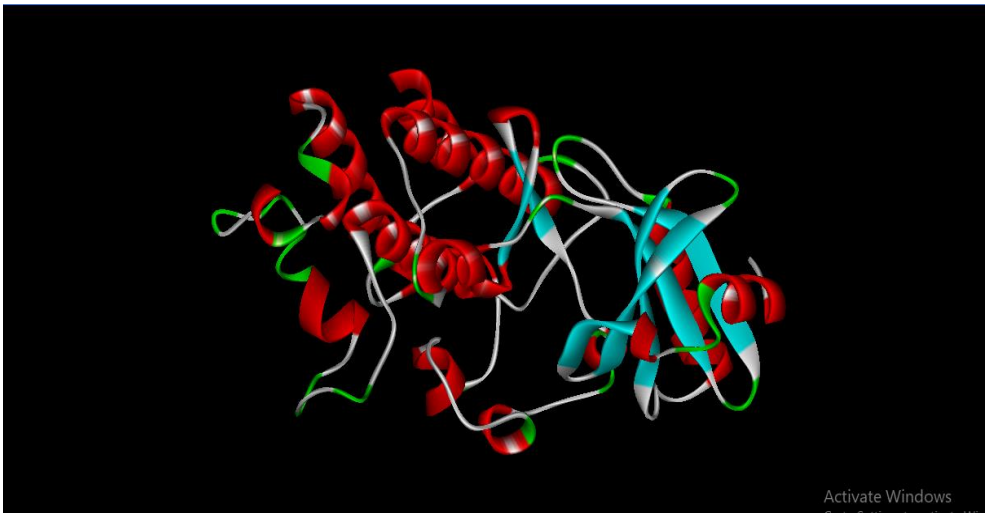


Fig. 5 3D structure visualization of PERK13

The above image shows the 3D structure of PERK13 viewed in “Solid Ribbon Model” using Discovery Studio Software.

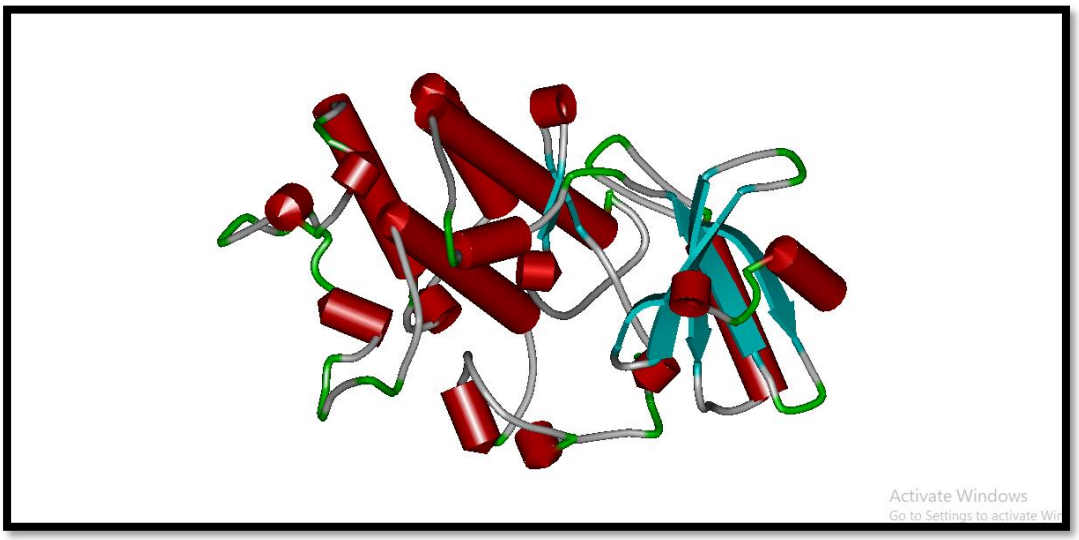


Fig. 6 3D structure visualization of PERK13

The above image shows the 3D structure of PERK13 viewed in “Cartoon Model” using Discovery Studio Software.

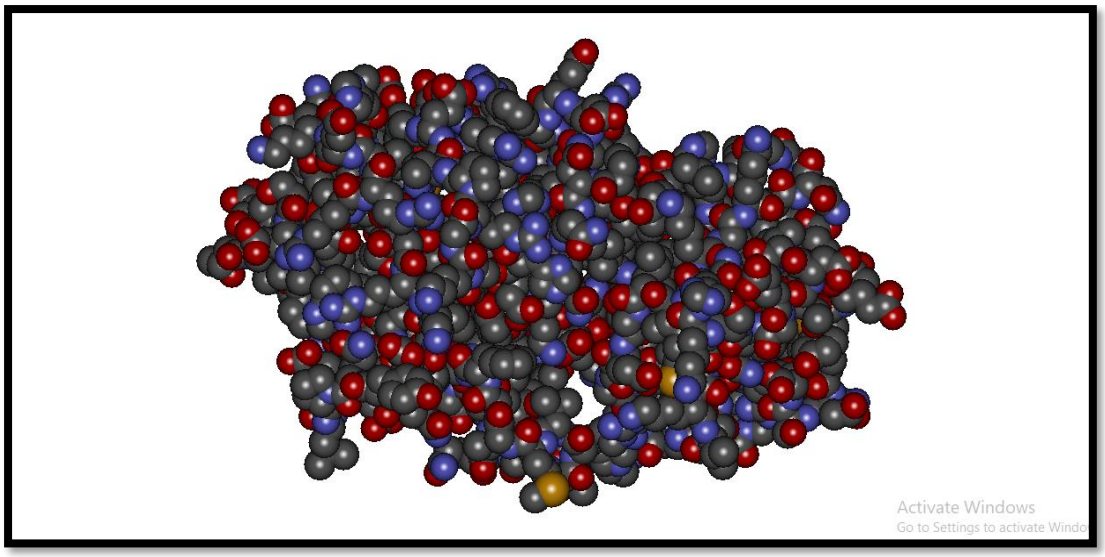


Fig 7 3D structure visualization of PERK13

The above image shows the 3D structure of PERK13 viewed in "Spacefill model" using Discovery Studio Software.

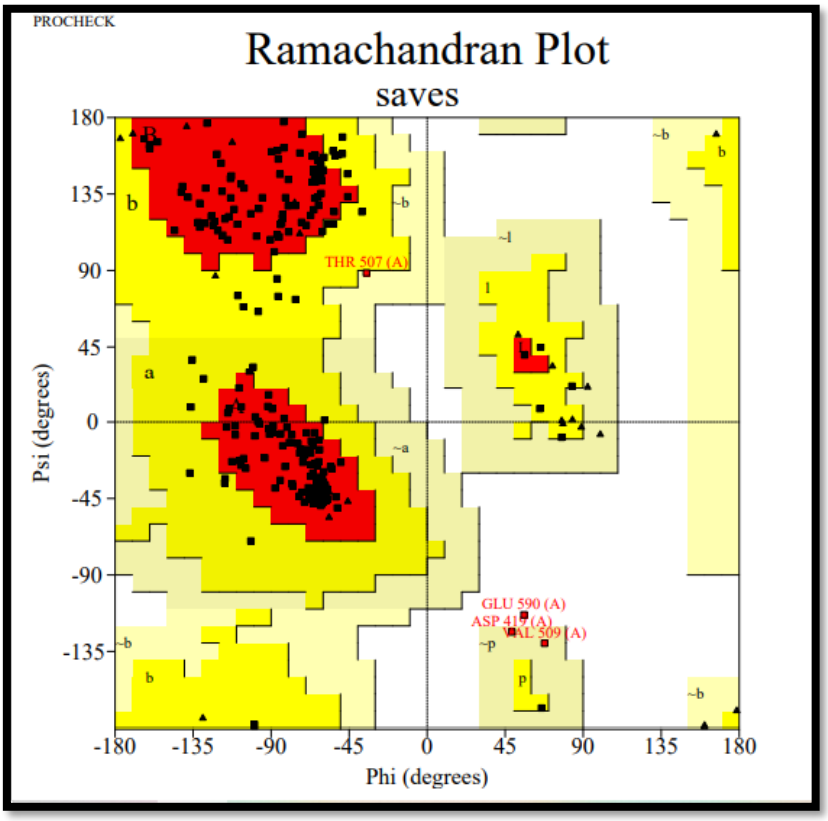


Fig. 8 3D structure validation- ProCheck server

Plot statistics

Residues in most favoured regions [A,B,L]	222	87.1%
Residues in additional allowed regions [a,b,l,p]	29	11.4%
Residues in generously allowed regions [~a,~b,~l,~p]	3	1.2%
Residues in disallowed regions	1	0.4%
	---	------
Number of non-glycine and non-proline residues	255	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	24	
Number of proline residues	9	

Total number of residues	290	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

The above image shows the assessment of Ramachandran Plot for the modelled PERK13 protein.

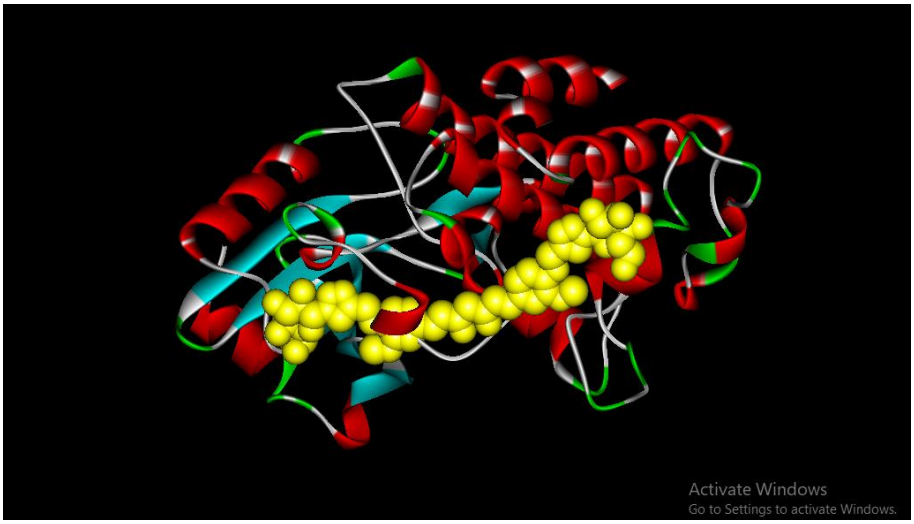
Fig 9 Molecular drug docking – PatchDock results

Receptor	Ligand	Complex Type	Cluste
C1.pdb	Q9CAL8_Protein.pdb	drug	4.0

Solution No	Score	Area	ACE
1	6318	756.90	-116.36
2	5784	797.30	-304.26
3	5742	796.60	-318.80
4	5698	844.00	-153.74
5	5570	1092.40	-431.21
6	5552	825.30	-234.33
7	5546	761.00	-179.79
8	5444	902.10	-217.50
9	5370	670.70	-124.91
10	5320	756.60	-90.53
11	5298	653.30	-60.10
12	4946	667.00	-117.20
13	4872	627.10	11.82
14	4860	782.20	-270.83
15	4858	622.70	-184.08
16	4828	726.00	-252.64
17	4788	778.80	-235.84
18	4776	597.20	-143.28
19	4746	589.40	-248.52
20	4740	530.10	-165.91

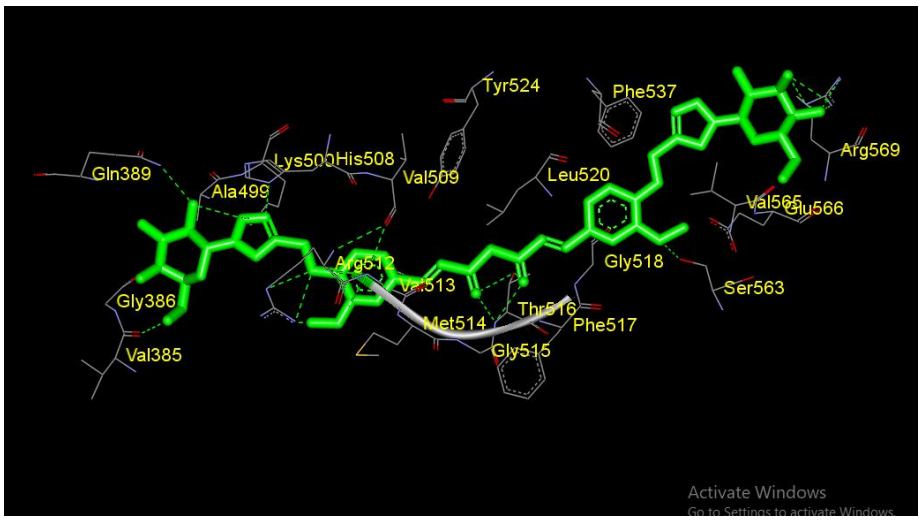
The above picture is the result page of PatchDock server. It shows the Atomic Contact Energy (ACE) value of De-O-Acetylated Curcumin Di-Galactose (37) with PERK13 protein complex (38) which is (negative 431.21).

Fig. 10 Molecular drug docking (De-O-Acetylated Curcumin Di-Galactose with PERK13)



The above picture represents the 3D form of the De-O-Acetylated Curcumin Di-Galactose - PERK13 complex. The yellow coloured spacefill model represents De-O-Acetylated Curcumin Di-Galactose which binds with PERK13 and is viewed using Discovery Studio Software.

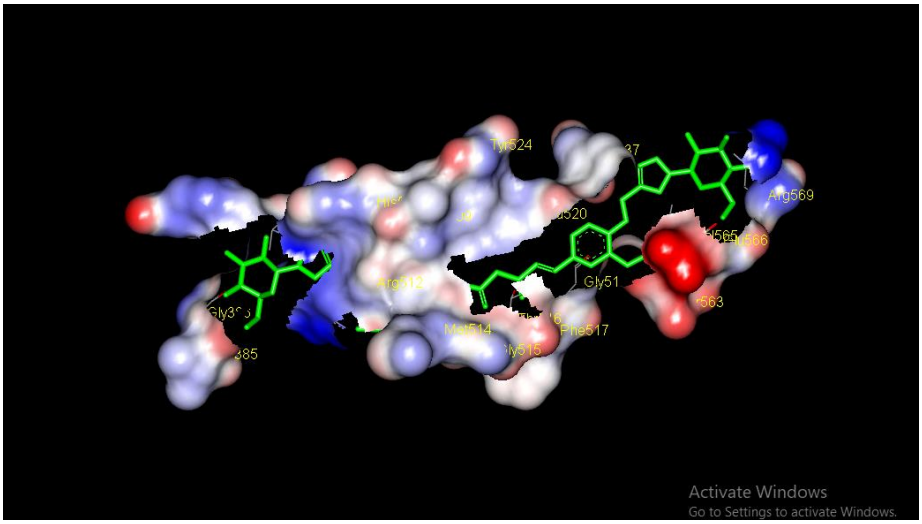
Fig. 11 Drug-Protein binding sites prediction (De-O-Acetylated Curcumin Di-Galactose with PERK13)



The above picture represents the 3D form of the De-O-Acetylated Curcumin Di-Galactose - PERK13 complex. Green coloured structure in Stick model represents De-O-

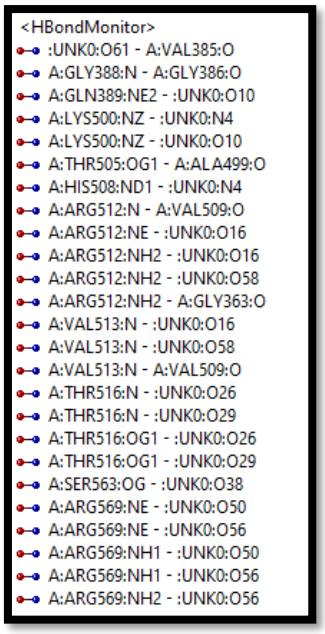
Acetylated Curcumin Di-Galactose which interacts with PERK13 at the amino acid binding regions viewed using Discovery Studio Software.

Fig. 12 Drug-Protein binding sites prediction (De-O-Acetylated Curcumin Di-Galactose with PERK13)



The above picture represents the 3D form of the De-O-Acetylated Curcumin Di-Galactose -PERK13 complex. Green coloured structure in Stick model represents De-O-Acetylated Curcumin Di-Galactose which interacts with PERK13 (Surface model) at the amino acid binding regions viewed using Discovery Studio Software.

Fig 12 H-bond interaction of De-O-Acetylated Curcumin Di-Galactose with PERK13



The above picture shows the Electrostatic and H-bond interaction between De-O-Acetylated Curcumin Di-Galactose and PERK13.

Table 1: Summary of Molecular Docking results

Target	**Compound :** *De-O-Acetylated Curcumin Di-Galactose*
PERK13 (UniProt ID: Q9CAL8 **)**	Negative 431.21

The above table shows the ACE (Atomic Contact Energy) value when De-O-Acetylated Curcumin Di-Galactose binds with PERK13.

Discussion

PERK13 protein was retrieved in FASTA format using UniProt database. The total length of the protein was 710 aa and Mass (Da):75,372 [Fig (1)]. In first step involved of protein analysis, we use UniProt Prosite scan database which clearly shows that the amino acid range 353 – 619 aa falls within the functional domains (protein kinases) of PERK13.

Protein kinases [11,12,13,14,] are enzymes that makeup a large family of proteins which share sequence homology. Their conserved catalytic core is common to serine/threonine

as well as to tyrosine protein kinases. There are several conserved regions present in the catalytic domain of protein kinases. Two of these regions were selected to construct signature patterns. The first region, which can be spotted in the N-terminal end of the catalytic domain, is a stretch of residues, which is rich in glycine, and which is present near a lysine residue. It has also been indicated to be concerned with ATP binding. Fig.(2)

In the second step, mobiDB Database (https://mobidb.bio.unipd.it/Q9CAL8) was used for domain analysis. Transmembrane (TM) α-helices comprise of a few residues which are strongly hydrophilic when compared to the general composition of proteins [15,16,17,18]. This can be partly explained by the high loss of energy when polar side chains are buried in a hydrophobic environment [19]. Polar residues are highly conserved, particularly in multispanning TM proteins, which suggests molecular interactions that favor these residues either functionally or structurally [20,21]. Functional roles of certain polar residues are noticed in protein structures, such as binding of prosthetic groups (H and E) in the photosynthetic reaction center [22], proton transport (D) and retinal binding (K) in bacteriorhodopsin [23,24], binding of hemes (H) in cytochrome *c* oxidase [25] and the cytochrome bc_1 complex, and Ca^{2+} binding (N, D, E, and T) in the Ca^{2+} ATPase [26]. The results obtained from MobiDB Database for PERK13 protein shows that another functional domain of PERK13 lies within 676-700 polar residues. Fig (13)

In this research study, SWISS-MODEL was used to convert the amino acid sequence of PERK13 into 3D structure [Fig (4)]. SWISS-MODEL [Bienert et al., 2017; Guex et al., 2009; Studer et al., 2020; Bertoni et al., 2017; Laskowski et al., 1993][27,28,29,30,31] was used to elaborately analyse the molecular and structural details of PERK13 for the purpose of docking. SWISS-MODEL is a server for automated comparative modelling of three-dimensional (3D) protein structures. Waterhouse et al. computed models by the SWISS-MODEL server homology modelling pipeline which relies on ProMod3, an in-house comparative modelling engine based on Open Structure. [Laskowski et al., 1996; Duhovny et al., 2002] [32,33][Fig. 5]. The modelled 3D protein was completely evaluated using ProCheck server for assessment of Ramachandran Plot. The results for PERK13 protein obtained from

ProCheck server show a Ramachandran plot score of 87.1% which indicates that there is no error present in the modelled structure.

After modelling the 3D structure of PERK13, the protein was viewed using Discovery Studio Software [Fig (6, 7, 8)]. Discovery Studio is a software which is currently used by the majority of Bioinformaticians and Cheminformaticians. It is used for simulating small molecule and macromolecule systems. It is developed and distributed by Dassault Systemes BIOVIA.

Molecular Drug Docking

The selected test compound, *De-O-Acetylated Curcumin Di-Galactose,* was docked with PERK13 and the molecular H-bond interaction between them was analysed. Proteins with small ligands are docked using Patchdock server [Duhovnyet al., 2005; Duhovny et al., 2003; Mashiach et al., 2010][34,35,36]. Hence, docking, for our study, was performed using this server. In molecular docking studies, the compound, *De-O-Acetylated Curcumin Di-Galactose* was docked with PERK13. H-bond interactions between the drug and the compound were observed along with binding affinity scores. Table 1 shows the drug-docking result with ACE value.

Fig.7 shows the PatchDock result scores of PERK13 with *De-O-Acetylated Curcumin Di-Galactose* and the 3D structure of H-bond interactions with the respective drug-binding amino acid pockets (VAL: 385,GLY:388,GLY:386,GLN:389,LYS:500,LYS:500,THR:505,ALA:499,HIS:508,ARG:512 ,VAL:509,ARG:512,ARG:512,ARG:512,ARG:512,GLY:363) (Fig 9,10,11,12).

The PatchDock results of *De-O-Acetylated Curcumin Di-Galactose* with PERK13 shows an ACE: Atomic Contact Energy value of (negative 431.21). According to theory, higher negative value indicates more binding affinities between the target and the compound. Here, it has been proved that *De-O-Acetylated Curcumin Di-Galactose* is an ideal candidate to interact with PERK13. When compared with the ACE values of interaction between the PERK13 protein and Curcumin (negative 221.73 kcal/mol), Neotame (artificial sweetener - negative 275.72 kcal/mol), pyrroloquinoline quinone (negative 170.33 kcal/mol), the water soluble test compound used in this study, digalactosylated curcumin[20] has higher binding affinity (negative 431.21 kcal/mol).

Interestingly, the results obtained from docking shows that *De-O-Acetylated Curcumin Di-Galactose* directly binds with the functional domain regions of PERK13 protein. The functional domains were analysed using ScanProsite and MobiDB. The following amino acids highlighted in yellow shows the positions where the drug *De-O-Acetylated Curcumin Di-Galactose* interacts with the PERK13 protein: "VAL: 385,GLY:388,GLY:386,GLN:389,LYS:500,LYS:500,THR:505,ALA:499,HIS:508,ARG:512 ,VAL:509,ARG:512,ARG:512,ARG:512,ARG:512,GLY:363. Database references mention that the Nucleotide binding regions of amino acids are from **359-367** positions (https://prosite.expasy.org/rule/PRU00159) and the Functional Domains residues of amino acids are from (https://mobidb.bio.unipd.it/Q9CAL8) **356-626.**

Conclusion

In this current *Insilico* drug docking research study, the selected compound, *De-O-Acetylated Curcumin Di-Galactose* efficiently binds with the target protein, PERK13 (Proline-rich receptor-like protein kinase). The current docking study clearly proves that the Domains and the polar residue amino acid positions (Drug Binding Pockets) of PERK13 exhibit H-bond interaction with *De-O-Acetylated Curcumin Di-Galactose*. Hence, the selected compound efficiently down-regulates PERK13 and can be used as an efficient antagonist agent in future studies to help plants tolerate saline levels.

Acknowledgement

I would like to extend my thanks to Dr. Balaji Munivelan, Ph.D., Senior Bioinformatician for providing me paid access to the commercial tools.I thank my student Dr. Harishchander Anandaram, Amrita Vishwa Vidyapeetham for all the basic lessons in bioinformatics.

References

1. Kiss, J. Z., Weise, S. E., Kiss, H. G. (2000). How can plants tell which way is up? *Am Biol Teach*., *62*(1):59-63.

2. Zheng, Z. L. (2006). Use of the gl1 Mutant & the CA-rop2 Transgenic Plants of Arabidopsis thaliana in the Biology Laboratory Course. *Am Biol Teach., 68*(9), e148-e153.
3. Brooks, E., Dolan, E., Tax, F. (2011). Partnership for Research & amp; Education in Plants (PREP): Involving High School Students in Authentic Research in Collaboration with Scientists, *Am Biol Teach., 73*(3).
4. O'Malley, R.C., Ecker, J. R. (2010). Linking genotype to phenotype using the Arabidopsis unimutant collection, *Plant J., 61*(6), 928-40.
5. *A*lonso, J. M., Stepanova, A. N., Leisse, T. J., et al., (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*, *Science, 301*(5633), 653-7.
6. Yamada, K., Lim. J., Dale, J. M., et al., (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome, *Science, 302*(5646), 842-6.
7. Curtis, M. D., Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in plants. *Plant Physiol., 133*(2*)*, 462-9.
8. Earley, K. W., Haag, J. R., Pontes, O., et al., (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J., 45*(4), 616-29.
9. Trigg, S. A., Garza, R. M., MacWilliams, A., et al., (2017). CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat Methods., 14*(8), 819-825.
10. Hanks, S. K., Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J. 9*, 576-596.
11. Hunter, T. (1991). Protein kinase classification. *Methods Enzymol., 200*, 3-37.
12. Hanks, S. K., Quinn, A. M. (1991). Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol. 200*, 38-62.
13. Hanks, S. K. (1991). *Curr. Opin. Struct. Biol., 1*, 369-383.
14. Hanks, S.K., Quinn, A. M., Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science, 241*, 42-52.
15. Landolt-Marticorena, C., Williams, K. A., Deber, C. M. & Reithmeier, R. A. (1993) J. Mol. Biol. 229, 602–608.
16. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1994) FEBS Lett. 339, 269–275.
17. Arkin, I. T. & Brunger, A. T. (1998) Biochim. Biophys. Acta 1429, 113–128.

18. Tourasse, N. J. & Li, W. H. (2000) Mol. Biol. Evol. 17, 656–664.

91. Engelman, D. M. & Steitz, T. A. (1984) in The Protein Folding Problem, ed. Wetlaufer, D. B. (Westview, Boulder, CO), pp. 87–113.

20. Gray, T. M. & Matthews, B. W. (1984) J. Mol. Biol. 175, 75–81.

21. Gromiha, M. M. (1999) Protein Eng. 12, 557–561.

22. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985) Nature (London) 318, 618–624.

23. Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990) J. Mol. Biol. 213, 899–929.

24. Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. (1999) J. Mol. Biol. 291, 899–911.

25. Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995) Nature (London) 376, 660–669.

26. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. (2000) Nature (London) 405, 647–655.

27. Waterhouse, A., Bertoni, M., Bienert, S., et al., (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res., 46*(W1), W296-W303.

28.Bienert, S., Waterhouse, A., de Beer, T. A. P., et al., (2017). The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res., 45*, D313-D319.

29. Guex, N., Peitsch, M. C., Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis, 30,* S162-S173.
30. Studer, G., Rempfer, C., Waterhouse, A.M., et al., (2020). QMEANDisCo - distance constraints applied on model quality estimation. *Bioinformatics, 36,* 1765-1771.
31. Bertoni, M., Kiefer, F., Biasini, M., et al., (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports,* 7, 10480.
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S., et al., (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.,* 26, 283-291.

33. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., et al., (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, *8*, 477-486.
34. Duhovny, D., Nussinov, R., Wolfson, H. J. Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002.
35. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., et al., (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucl. Acids. Res.*, *33*, W363-367.
36. Mashiach, E., Schneidman-Duhovny, D., Peri, A., et al., (2010). An integrated suite of fast docking algorithms. *Proteins*, *78* (15), 3197-204.
37. https://www.nature.com/articles/s41598-020-70921-2
38. https://patents.google.com/patent/WO2017217508A1/en