Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 13, 2022, page 107-113

Structural Analysis of Calpain-Type Cysteine from *Nicotiana Benthamiana* in Regulation of Leaf Development

Pua Lei Wen, Huszalina Hussin*
Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia
*Corresponding author: huszalina@utm.my

Abstract
The *Nb*DEK gene produces a calpain-type cysteine protease that affects *Nicotiana benthamiana* leaf development. This enzyme's mechanism is unknown. This study aims to characterize the structure and mechanism of calpain-type cysteine protease (*Nb*DEK) from *Nicotiana benthamiana*. UniProtKB sequence query for *Nb*DEK from *Nicotiana benthamiana* (Q6SSJ2) ExPASy's ProtParam, PeptideMass, JPred 4, and GOR V were used to predict physicochemical primary and secondary structures. *Nb*DEK from *Nicotiana benthamiana* was thermally stable with a 100.12 index. It has 45.19% random coil and 43.88% -helix, indicating a globular, compact, and stable protein. Multiple sequence alignment using Clustal Omega confirmed that *Nb*DEK from *Nicotiana benthamiana* has the conserved catalytic triad Cys-71, His-229, and Asn-249, along with conserved residues Gln-65, Trp-72, Gly-160-161, Pro-250, and Trp-251. 3D structure prediction was made using homology modeling (SWISS-MODEL). The highly conserved active site was seen in a 3D superimposition of *Nb*DEK from *Nicotiana benthamiana* and DEK1. AutoDock Vina confirmed the molecular interaction efficiency between *Nicotiana benthamiana*'s *Nb*DEK and E-64 protease inhibitor (-7.4 kcal/mol). Overall, the *Nb*DEK from *Nicotiana benthamiana* has potential function as the well-characterized DEK1 and may be inhibited by the E-64 protease inhibitor.
Keywords: *Nb*DEK; Molecular Docking; Structural Analysis

Introduction

The tobacco relative *Nicotiana benthamiana* is used extensively in molecular agriculture for transient expression of recombinant proteins (Tsekoa *et al.*, 2020). In addition, its big leaves are easily penetrated to produce the required protein, and its reactivity to *Agrobacterium tumefaciens* is poor, allowing for high quantities of transgenic transcripts (Bombarely *et al.*, 2012). The calpain-type cysteine protease produced by the *Nb*DEK gene has been thought to play a function in *Nicotiana benthamiana* leaf development.

Calpain-type cysteine protease directs cell fate and specification during leaf development. The molecular mechanism of DEK action in plant cell fate determination has not been elucidated in dicotyledonous plants. Plant calpain-type cysteine protease has no recorded protein-ligand interaction. Research suggests E-64 inhibits cysteine protease (Matsumoto *et al.*, 1999). Research on calpain-type cysteine protease in leaf development will increase molecular farming production.

This study uses bioinformatics to predict and examine calpain-type cysteine protease from *Nicotiana benthamiana* as a leaf-development enzyme. Homology modeling is used to create a 3D model of the *Nicotiana benthamiana* calpain-type cysteine protease, then superimposed with the reference protein structure model. To understand the protein's putative catalytic function, the features

of the predicted protein 3D structure and its binding affinity to the E-64 are investigated.

Materials and methods

As the query sequence, the calpain-type cysteine protease from *Nicotiana benthamiana* with accession number Q6SSJ2 is retrieved from UniProtKB. The sequence of the *Arabidopsis thalia*na calpain-type cysteine protease with accession number Q8RVL2 is retrieved from UniProtKB as the reference protein utilized in the superimposition. ProtParam and PeptideMass are then used to find the essential physicochemical characteristics in the query sequence for the physicochemical properties study. Following that, the secondary structure prediction is performed using the JPred and GOR V algorithms, which predict the query protein's alpha- helices, beta-sheets, coils, and turns.

Using NCBI BLASTp, the top ten homolog proteins of the query protein are then chosen. Clustal Omega is then used to do multiple sequence alignment for these top ten homologous sequences with the query sequence and reviewed reference protein sequence. Following that, the tertiary structure prediction of the query protein sequence is performed using homology modeling by SWISS-MODEL. The predicted 3D structures are then verified using the MolProbity and Saves v6.0 services.

3D structural superimposition is performed using PyMOL for comparison with the reference protein. The last stage is molecular docking, which is performed between the query protein and E-64 protease inhibitor to examine binding affinity using AutoDock Vina. PyMOL is used to show the data for comparison and analysis.

Results and discussion

The automatically annotated unreviewed query amino acid sequence of calpain type cysteine protease from *Nicotiana benthamiana* (*Nb*DEK) with accession number Q6SSJ2 was retrieved from UniProtKB/TrEMBL in FASTA sequence format. The sequence of the manually annotated calpain type cysteine protease from *Arabidopsis thaliana* (DEK1) was acquired from UniProtKB/Swiss-Prot in FASTA format.

Table 1: The comparison of the outputs of ProtParam for query *Nb*DEK with accession number Q6SSJ2 and reviewed DEK1 with accession number Q8RVL2

Parameters	***Nb*DEK**	**DEK1**
No. of amino acids	2142	2151
Molecular weight (Da)	236791.67	238264.34
Theoretical Pi	6.09	6.15
Negatively charged residue (Asp+Glu)	223	227
Positively charged residue (Arg+Lys)	203	210
Extinction coefficient (M^{-1} cm^{-1}) at 280 nm	405775	404160
Instability index	40.16	42.27
Aliphatic index	100.12	96.88
Grand average of hydropathicity (GRAVY)	0.127	0.056

ExPASy's ProtParam and PeptideMass services were used to conduct investigation of the query *Nb*DEK, accession number Q6SSJ2, for its physicochemical parameters. ProtParam was also used to analyse the physicochemical characteristics of the reference protein DEK1 (accession number

Q8RVL2). The outputs were tabulated in Table 1. The molecular weight (Da) and theoretical Pi of the *Nb*DEK are 236791.67 Da and 6.09 respectively. The theoretical pI of *Nb*DEK is slightly lower than 7, meaning that the *Nb*DEK is slightly acidic (Sahay *et al.*, 2020). The negatively charged residues (Asp+Glu) of *Nb*DEK is 223, while the positively charged residues (Arg+Lys) of *Nb*DEK is 203. Its negatively charged residues are more than the positively charged residues. This suggested that *Nb*DEK is an intracellular protein based on the negative-outside rule (Baker *et al.*, 2017). The extinction coefficient of *Nb*DEK at 280 nm was 405775 M^{-1} cm^{-1} which is higher than the extinction coefficient of DEK1, 404160 M^{-1} cm^{-1} due to the higher concentration of tyrosine residues, Tyr (2.8%) than DEK1, in agreement with the report of Satyanarayana *et al.* (2018). The instability index of *Nb*DEK (40.16) that is greater than 40 as shown in Table 1 indicates the *Nb*DEK protein may be unstable (Mukesh Kumar *et al.*, 2018). The high aliphatic index of *Nb*DEK (100.12) suggests the *Nb*DEK is thermally stable. The positive GRAVY of *Nb*DEK (0.127) suggest that the protein is generally hydrophobic, similar to the DEK1 which have positive GRAVY score 0.056.

Table 2: Prediction of secondary structure of query *Nb*DEK with accession number Q6SSJ2

	JPred 4		**GOR IV**	
	Length	**Percentage (%)**	**Length**	**Percentage (%)**
Alpha helix (Hh)	940	43.88	707	33.01
Extended strand (Ee)	234	10.92	476	22.22
Random coil (Cc)	968	45.19	959	44.77

The secondary structure of *Nb*DEK was predicted by using the JPred 4 and GOR IV server. The outputs from both JPred 4 and GOR IV are tabulated in Table 2 for comparison. For *Nb*DEK, the random coil was determined to be the most prevalent secondary structural component. This indicated the stability of the model of *Nb*DEK and its enzyme functionality, in agreement with the report of Satyanarayana *et al.* (2018) that flexibility and structural variations of proteins, such as enzyme turnover, are largely due to the presence of random coils. The model's great conservation and stability were highlighted by the prominence of its coiled regions (Hasan *et al.*, 2015). Besides, the second highest secondary structure for *Nb*DEK is the alpha helix. The *Nb*DEK could be assumed to be present in the transmembrane region due to its coiling nature and globular structure (Satyanarayana *et al.*, 2018). This is in conformity with the report of Guzzi *et al.* (2017) where the most common protein structure element that crosses biological membranes in the form of transmembrane proteins are the helices.

Multiple sequence alignment was performed using Clustal Omega to compare the similarities and conserved regions of the query *Nb*DEK between *Nicotiana benthamiana* and other organisms selected. The amino sequences of DEK proteins in the multiple sequence alignment are highly conserved as shown in Figure 4.2. This is supported by Wang *et al.*, (2003) that found DEK1 proteins from various plant species share 70–98% sequence similarity, indicating that this family of proteins is highly conserved. The active sites of DEK proteins are well conserved as highlighted in yellow in Figure 1. This is in agreement with the Wang *et al.*, (2003) that found the catalytic triad Cys-71, His-229 and Asn-249, with conserved residues Gln-65, Trp-72, Gly-160–161, Pro-250, and Trp-251 in DEK1 calpain domain II and domain III, whereas calpain domain II&III Lys-1 corresponds to Lys-1699 in the full-length DEK1 protein sequence (Wang et al., 2003).

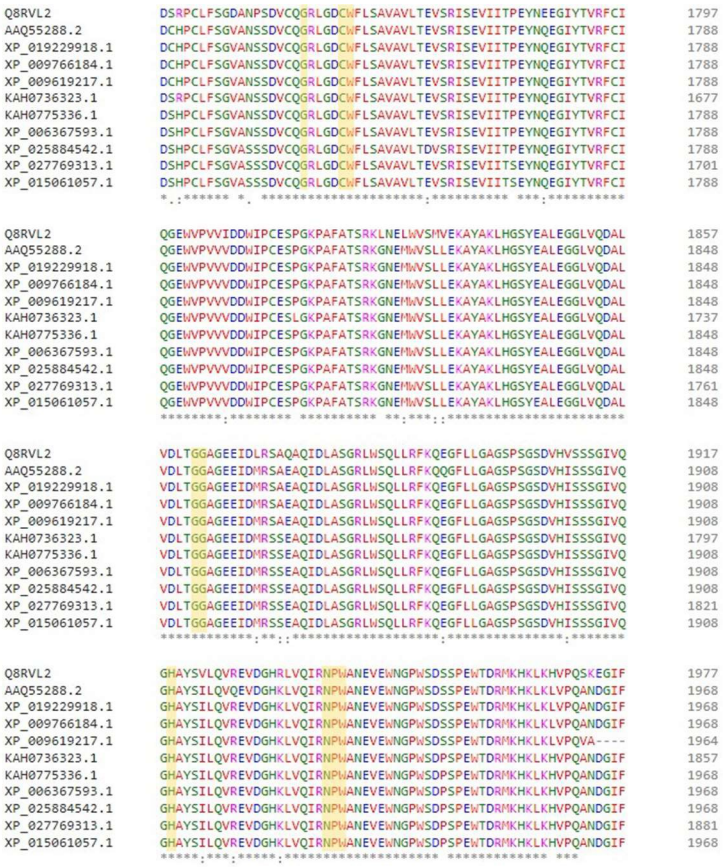

Figure 1 Highly conserved active sites are highlighted in yellow in the multiple sequence alignment of DEK from different organisms, reference DEK1 and target *Nb*DEK

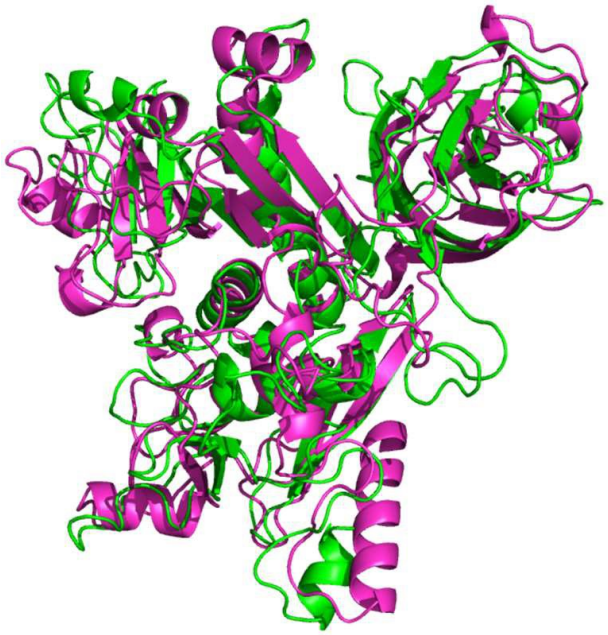

Figure 2 The matched regions of the superimposed structure of *Nb*DEK (green) and DEK1 (magenta, UniProtKB Entry: Q8RVL2)

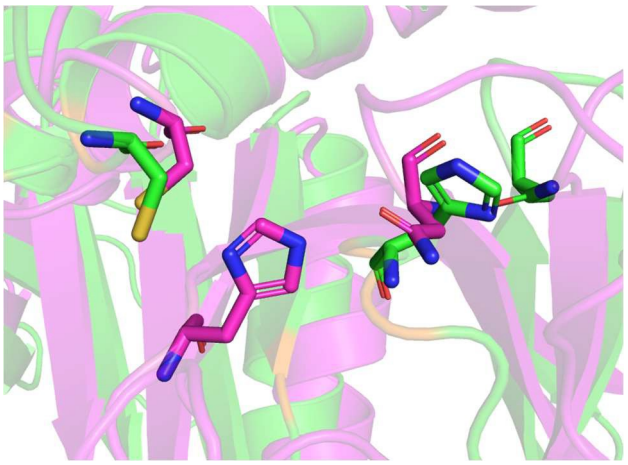

Figure 3 The active site (sticks) of the superimposed structure *Nb*DEK (green) and DEK1 (magenta)

The superimposition of the *Nb*DEK structure over the well-characterized DEK1 structure indicated that the two structures are significantly equivalent, as shown in Figure 2. The *Nb*DEK structure only matched the downward sections of DEK1 including the Peptidase C2 calpain catalytic domain and the Calpain III domain. However, the *Nb*DEK structure is not exactly matched to the DEK1 structure in location. The *Nb*DEK active site (Cys-1752, His-1910, and Asn-1930) is close to the DEK1 active site (Cys-1761, His-1919, And Asn-1939), as illustrated in Figure 3. These active sites may provide a similar function.

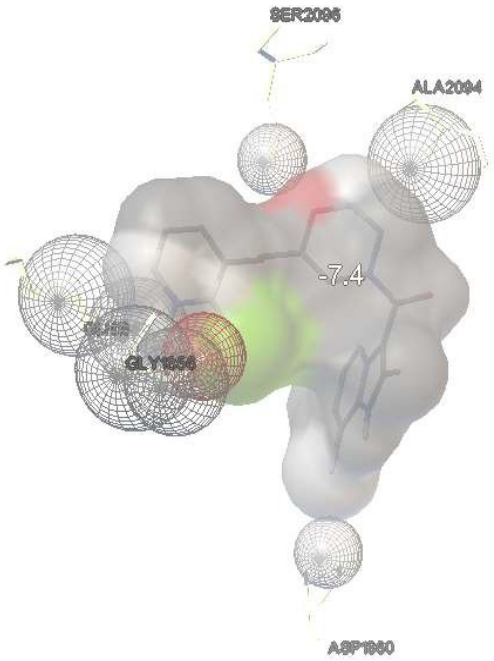

Figure 4 The ligand conformation of E-64 with the highest negative value of binding affinity of -7.4 kcal/mol

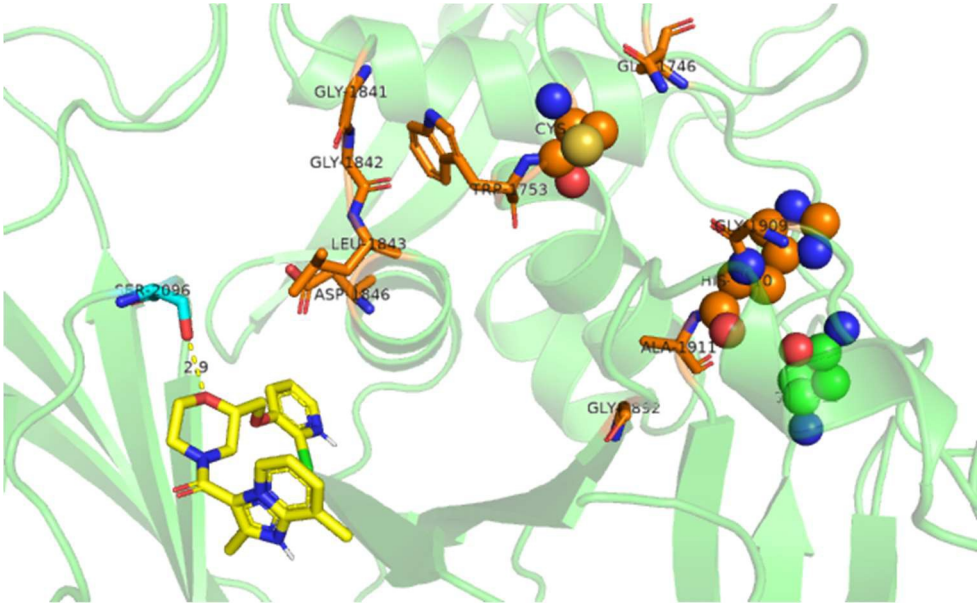

Figure 5 E-64 that is docked at the active site of *Nb*DEK binds with the non-conserved residues near the active site. The orange structure represents the predicted ligand binding site by I-TASSER. The sphere represents the catalytic triad of *Nb*DEK.

The interaction between the *Nb*DEK and E-64 protease inhibitor was investigated using molecular docking using AutoDock Vina. In this investigation, the ligand conformation with the largest negative value of binding affinity (kcal/mol) was chosen for examination. Protein–ligand binding only occurs spontaneously when the free energy change is negative, and the difference in ΔG levels of complexed and unbound free states is proportional to the stability of the protein–ligand interaction (Afriza *et al.*, 2018). From the result, the E-64 protease inhibitor showed high binding efficiency to the active site of *Nb*DEK with the highest affinity value of -7.4 kcal/mol, as shown in Figure 4. The high binding efficiency of E-64 toward *Nb*DEK revealed their interaction and association with the active site of *Nb*DEK. However, the E-64 protease inhibitor does not interact with the conserved residues of active sites and the predicted ligand binding site. The E-64 protease inhibitor has a polar interaction with non-conserved residue, Ser-2096, as shown in Figure 5. While the research from Matsumoto *et al.*, (1999) found a common action of these inhibitors for cysteine proteases is a covalent bond formation between the epoxy carbon atom of the inhibitor and the Sg atom of the active CYS residue.

Conclusion

In conclusion, *Nb*DEK was discovered to be conserved with all of the identified functional residues on well-characterized DEK1 that regulate leaf development. As a result, *Nb*DEK was proclaimed to have the potential to influence cell fate determination and cell specification during leaf development. The E-64 protease inhibitor has a significant affinity for *Nb*DEK in its catalytic domain. However, as compared to prior research with animal calpain type cysteine protease, E-64 binds to various residues for interaction, suggesting that E-64 may have unique effects on *Nb*DEK that require additional investigation.

References

Afriza, D., Suriyah, W. H., & Ichwan, S. J. A. (2018). In silico analysis of molecular interactions between the anti-apoptotic protein survivin and dentatin, nordentatin, and quercetin. *Journal of Physics: Conference Series*, *1073*(3), 032001. https://doi.org/10.1088/1742-6596/1073/3/032001

Baker, J. A., Wong, W. C., Eisenhaber, B., Warwicker, J., & Eisenhaber, F. (2017). Charged residues next to transmembrane regions revisited: “Positive-inside rule” is complemented by the “negative inside depletion/outside enrichment rule.” *BMC Biology*, *15*(1), 1–29.

https://doi.org/10.1186/S12915-017-0404-4/FIGURES/8

Bombarely, A., Rosli, H. G., Vrebalov, J., Moffett, P., Mueller, L. A., & Martin, G. B. (2012). A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. *Molecular Plant-Microbe Interactions: MPMI*, *25*(12), 1523–1530. https://doi.org/10.1094/MPMI-06-12-0148-TA

Guzzi, A. F., Oliveira, F. S. de L., Amaro, M. M. de S., Tavares Filho, P. F., & Gabriel, J. E. (2017). Predicting the effects of the single nucleotide polymorphism A122V on CXC chemokine receptor type 1 of Bos taurus (Artiodactyla: Bovidae) cattle by in silico analyses. *Biotemas*, *30*(4), 1–6. https://doi.org/10.5007/2175-7925.2017V30N4P1

Hasan, A. M., Mazumder, H. M. H., Chowdhury, S. A., Datta, A., & Khan, A. M. (2015). Molecular-docking study of malaria drug target enzyme transketolase in Plasmodium falciparum 3D7 portends the novel approach to its treatment. *Source Code for Biology and Medicine*, *10*(1). https://doi.org/10.1186/S13029-015-0037-3

Matsumoto, K., Mizoue, K., Kitamura, K., Tse, W.-C., Huber, C. P., & Ishida, T. (1999). *Structural Basis of Inhibition of Cysteine Proteases by E-64 and Its Derivatives*. https://doi.org/10.1002/(SICI)1097-0282(1999)51:1

Mukesh Kumar, G. B., Ram, R. K., Singh, D. K., & Udit, and U. K. (2018). Comparative Studies on Kisspeptin Receptor and their Physicochemical Characterization. *Article in International Journal of Current Microbiology and Applied Sciences*. https://doi.org/10.20546/ijcmas.2018.705.006

Sahay, A., Piprodhe, A., & Pise, M. (2020). In silico analysis and homology modeling of strictosidine synthase involved in alkaloid biosynthesis in catharanthus roseus. *Journal of Genetic Engineering and Biotechnology*, *18*(1). https://doi.org/10.1186/S43141-020-00049-3

Satyanarayana, S. D. V., Krishna, M. S. R., Pavan Kumar, P., & Jeereddy, S. (2018). In silico structural homology modeling of nif A protein of rhizobial strains in selective legume plants. *Journal of Genetic Engineering and Biotechnology*, *16*(2), 731–737. https://doi.org/10.1016/J.JGEB.2018.06.006

Tsekoa, T. L., Singh, A. A., & Buthelezi, S. G. (2020). Molecular farming for therapies and vaccines in Africa. *Current Opinion in Biotechnology*, *61*, 89–95. https://doi.org/10.1016/J.COPBIO.2019.11.005

Wang, C., Barry, J. K., Min, Z., Tordsen, G., Rao, A. G., & Olsen, O. A. (2003). The calpain domain of the maize DEK1 protein contains the conserved catalytic triad and functions as a cysteine proteinase. *The Journal of Biological Chemistry*, *278*(36), 34467–34474. https://doi.org/10.1074/JBC.M300745200