JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Computational analysis of a Glutamate decarboxylase (GAD) protein from *Selaginella moellendorffii*

Ashutosh Mukherjee and Asis Kumar Pal*

Department of Botany,
Vivekananda College, 269, Diamond Harbour Road, Thakurpukur, Kolkata - 700 063, West Bengal, India.
*Author for correspondence: asispal66@gmail.com

Abstract : **In this study, a glutamate decarboxylase (GAD) protein from *Selaginella moellendorffii* was analyzed *in silico* to predict its physicochemical properties and three-dimensional structure. The computed heoretical isoelectric point (pI) was found to be less than 7 indicating the acidic nature of this protein. The aliphatic index of 91.64 indicates the thermal stability of the protein. Grand average hydropathy (GRAVY) was predicted to be -0.224 which shows the possibility of better interaction of this protein with water. Secondary structure analysis revealed the predominance of alpha helix followed by random coil, extended strand and beta turns. It is also predicted that the protein is localized on the endoplasmic reticulum membrane. Homology modeling revealed the three-dimensional structure of the protein which shows that the protein contains 19 alpha helices, 3 beta sheets with 13 strands and 3 beta alpha beta units. This study will benefit further laboratory based evaluation of this protein and will also benefit evolutionary study of GAD proteins as a whole.**

Index Terms **- disorder, GABA, GRAVY, template, Homology modeling, sequence alignment.**

I. INTRODUCTION

γ-Aminobutyrate (GABA) is a ubiquitous four-carbon, non-proteinogenic amino acid which is one of the major inhibitory neurotransmitters in mammals (Erlander and Tobin, 1991). In plants, GABA functions as both metabolite and signal in response to abiotic and biotic stresses (Shelp *et al*., 2021). Abiotic and biotic stresses result in elevated production of reactive oxygen species (ROS) which, in turn, stimulates GABA production (Shelp *et al*., 2021). The accumulated GABA can bind directly to the aluminum-activated malate transporter and the guard cell outward rectifying K^+ channel, thereby improving drought and hypoxia tolerance, respectively (Shelp *et al*., 2021). Thus, study of production and metabolism of GABA in plants can be of great interest for development of climate resilient crops. GABA is synthesized from alpha-decarboxylation of glutamate (Glu) catalyzed by glutamate decarboxylase (GAD, EC 4.1.1.15) enzyme in the cytosol; this process is called the GABA shunt (Bouché and Fromm, 2004; Yin *et al*., 2018). GAD has been identified and purified in various higher plants, and its enzymatic properties have been thoroughly studied (Baum *et al*., 1993; Oh *et al*., 2005; Yang *et al*., 2013). Several plant GADs possess a C-terminal domain that binds the Ca^{2+}/calmodulin (CaM) complex, activating GAD activity at neutral pH (Shelp *et al*., 2012; Shelp *et al*., 2021). *Arabidopsis* has five GAD genes in total, but only one (AtGAD1,2,4) possess the C-terminal domain (Shelp *et al*., 2012; Shelp *et al*., 2017). Thus, stress-induced increases in cytosolic Ca^{2+}/CaM complexation or H^+ can activate/stimulate GAD activity (Kinnersley and Turano, 2000; Knight *et al*., 1997; Bose *et al*., 2011; Behera *et al*., 2018). Physicochemical, structural and functional analysis of plant GADs is, thus, pivotal for the understanding of the GABA production of plants. In this regard, analysis of GADs from plants other than angiosperms can be helpful for studying evolutionary aspect of this protein family. The genome of *Selaginella moellendorffii* Hieron. (Selaginellaceae) with a size of 212.6 Mbp (with two haplotypes of ~106 Mbp) has been sequenced as it is a valuable resource for studying the early evolution of developmental and metabolic processes specific for vascular plants (Banks *et al*., 2011). Thus, analysis of GAD proteins from this species can be useful in this regard.

The major drawbacks of experimental methods used for protein characterization is involvement of high cost and time consumption. On the other hand, *in silico* approaches provide a viable solution to these problems. The amino acid sequence of a protein provides most of the information required for determining and characterizing the protein's physicochemical properties and structural aspects. Biological function of a protein is the manifestation of its three dimensional structure and knowledge of the structural organization of the protein is a prerequisite for understanding its functional aspects (Paital *et al*., 2011). However, three-dimensional structure of any GAD protein from *S. moellendorffii* is not known so far. Thus, it would be useful to recognize the 3D structure of GAD protein for the understanding of its structural aspects. In absence of crystal structure, homology modeling, which is done *in silico*, provides a faster way to obtain structural insight into the protein (Dolan *et al*., 2012).

Thus, the objectives of this study were to study and physicochemical characters as well as the three-dimensional structure of a GAD protein of *S. moellendorffii* with the help of several computational methods.

II. METHODOLOGY

Sequence retrieval and physicochemical characterization

Protein sequence of one glutamate decarboxylase from *S. moellendorffii* (gene symbol: LOC9656977; GeneID: 9656977; NCBI Reference Sequence: XP_002980886.1) was obtained from NCBI gene database (https://www.ncbi.nlm.nih.gov/gene). This protein will be called smGAD in this study. Physicochemical characters like theoretical pI, total number of positive and negative residues, instability index, aliphatic index and grand average hydropathy (GRAVY) were computed using the Expasy's ProtParam server (Gasteiger *et al*., 2005).

Secondary structure prediction and subcellular localization

SOPMA (Geourjon and Deléage 1995) was employed for calculating the secondary structural features of the selected protein sequence considered for this study. LocTree3 (Goldberg *et al*., 2014) was used to predict the subcellular localization of the protein. Protein disorder was predicted using PrDOS (Ishida and Kinoshita, 2007).

Model building, energy minimization and evaluation

The modeling of the three-dimensional structure of the proteins was performed by Swiss model (Waterhouse *et al*., 2018). Sequence alignment of the template and smGAD has been visualized with ESPript (Robert and Gouet, 2014). Hydrogens were added to the predicted model using ChimeraX (Pettersen *et al*., 2021) and then the model was energy minimized by YASARA energy minimization server (http://www.yasara.org/minimizationserver.htm) (Krieger *et al*., 2009) and the energy minimized structure was obtained by YASARA View (Krieger and Vriend, 2014). The overall stereochemical property of the energy minimized protein was assessed by Ramachandran plot analysis with PROCHECK (Laskowski *et al*., 1993). The validation was also performed by using ERRAT (Colovos and Yeates, 1993) and Verify3D (Bowie *et al*., 1991, Lüthy *et al*., 1992) using SAVESv6.0 structure validation server (https://saves.mbi.ucla.edu/). For an at-a-glance overview of the modeled protein, PDBsum (Laskowski, 2009) web server was used (http://www.ebi.ac.uk/pdbsum/). The modeled protein was visualized with ChimeraX (Pettersen *et al*., 2021).

III. RESULTS AND DISCUSSION

Physicochemical properties and subcellular localization of smGAD

The physicochemical properties of the glutamate decarboxylase (GAD) from *S. moellendorffii* (smGAD) was predicted using Expasy's ProtParam server (https://web.expasy.org/protparam/) using the protein sequence and the results are shown in table 1. The protein is 513 amino acids long. The most frequent amino acid present in the sequence was found to be leucine (48 residues, 9.4%) and the least was that of cysteine (6 residues, 1.2%). The total number of negatively charged residues (Asp + Glu) was 70 and the total number of positively charged residues (Arg + Lys) was 62 which indicate the protein to be intracellular as intracellular proteins have higher fraction of negatively charged residues (Cedano *et al*., 1997). The calculated isoelectric point (pI) is useful for the fact that at isoelectric point, the solubility is the least and the mobility in an electric field is zero. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. The calculated isoelectric point (pI) was computed to be 5.62 which indicates that the protein is acidic. The high aliphatic index (91.64) indicates that this protein is stable for a wide range of temperature range. The instability index (29.78) also provides the evidence that the protein in stable. The Grand Average Hydropathicity (GRAVY) value is negative (-0.224) which indicates better interaction of the protein with water. LocTree3 predicted that the protein is localized on the endoplasmic reticulum membrane.

Table 1: Different physicochemical properties of smGAD as obtained with ProtParam

Parameters	Value	Explanation
pI	5.62	Indicates that the protein is acidic.
Total number of negatively charged residues (Asp + Glu)	70	Total number of negatively charged residues is greater than total number of positively charged residues. This indicates that the protein is intracellular.
Total number of positively charged residues (Arg + Lys)	62	
Instability index (II)	29.78	This classifies the protein as stable.
Aliphatic index	91.64	Indicates that this globular protein is thermostable.
Grand average of hydropathicity (GRAVY)	-0.224	A negative GRAVY score indicates that the protein is hydrophilic.

Structural properties of smGAD

Table 2 presents the results of secondary structure prediction analysis by SOPMA from which it is clear that alpha helix is predominantly present (41.52%), followed by random coil (38.60%), extended strand (14.23%) and beta turn (5.65%).

Table 2: Secondary structure prediction of smGAD by SOPMA server

Type of secondary structure	Number of residues	Percentage of residues
Alpha helix	213	41.52
3_{10} helix	0	0.00
Pi helix	0	0.00
Beta bridge	0	0.00
Extended strand	73	14.23
Beta turn	29	5.65
Bend region	0	0.00
Random coil	198	38.60
Ambiguous states	0	0.00
Other states	0	0.00

PrDOS showed that total disordered amino acid residues were 89 (17.35%). However, they were spread over the protein in 4 regions (Met1-Val26, Ser114-Ala117, Leu450-Ala497 and Lys503-Cys513). The longest disordered region was spread from Leu450 to Ala497. Disordered regions differ considerably from the well-structured regions of a protein (Wright and Dyson, 1999). These disordered regions play significant roles in protein interaction (Patil and Nakamura, 2006). From these results, it seems that these disordered regions of smGAD may interact with other proteins.

Homology modeling and validation of the structure of smGAD

Template search against the Protein Data Bank (PDB) (Berman *et al*., 2000) by Swiss model revealed 3HBX.1 (chain A) as the best template for homology modeling. 3HBX is the crystal structure of GAD1 from *Arabidopsis thaliana* (Gut *et al*., 2009). The sequence alignment between the template and smGAD is shown in figure 1.

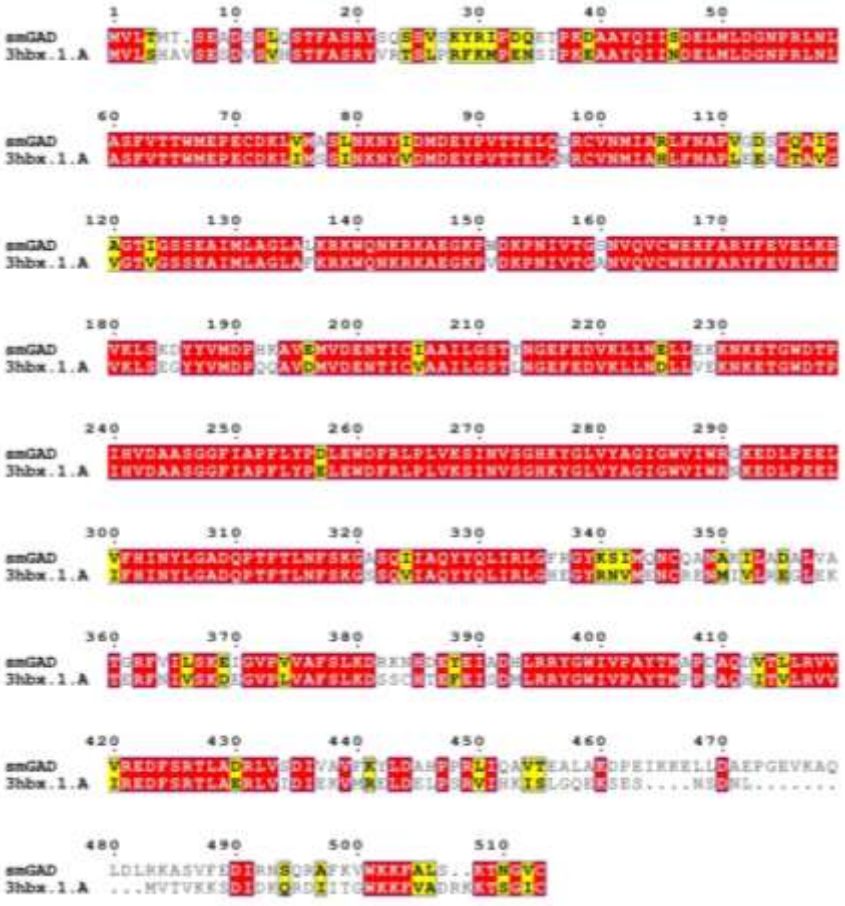

Figure 1: Sequence alignment between the template (3hbx.1.A) and modeled smGAD as revealed by ESPript. Red boxes indicate strict identity while yellow background indicates similarity across groups.

YASARA (Yet Another Scientific Artificial Reality Application) server uses its own YASARA force field for energy minimization which is an integration of AMBER (Assisted Model Building with Energy Refinement) all-atom force field equation, multi-dimensional knowledge-based torsion angle energy potentials, and a consistent set of force field parameters (Krieger *et al*., 2009). The energy minimized structure as prepared by YASARA was assessed for its quality with various tools. Overall Quality Factor obtained with ERRAT was 97.5787 which is good. Ramachandran plot with PROCHECK revealed that 90.3% residues was in most favoured region, 9.4% residues was in additional allowed regions, only 0.3% residues was in generously allowed regions and none (0.0%) was under disallowed region (figure 2). A protein structure is considered good if it has more than 90% of residues in the most favored regions on the main-chain Ramachandran plot (Laskowski *et al*., 1993). Thus the modeled smGAD protein structure is acceptable.

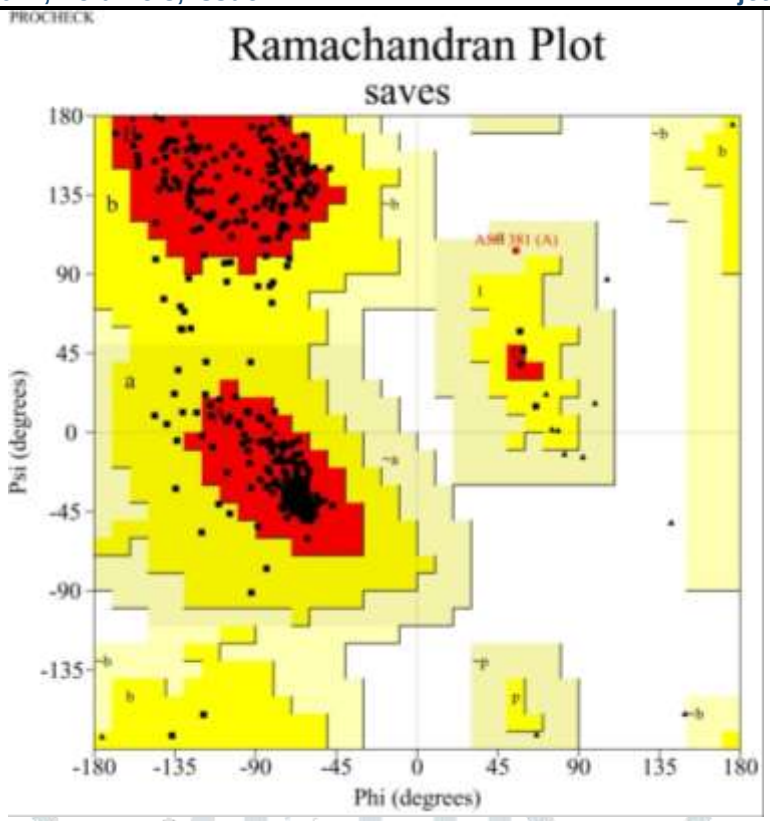

Figure 2: Ramachandran plot of smGAD homology model.

VERIFY3D revealed that 89.02% of the residues have averaged 3D-1D score >= 0.2 which also assessed the structure to be acceptable. The 3D structure of smGAD is shown in figure 3. PDBSum showed that the protein contains 19 alpha helices, 3 beta sheets with 13 strands and 3 beta alpha beta units (figure 4).

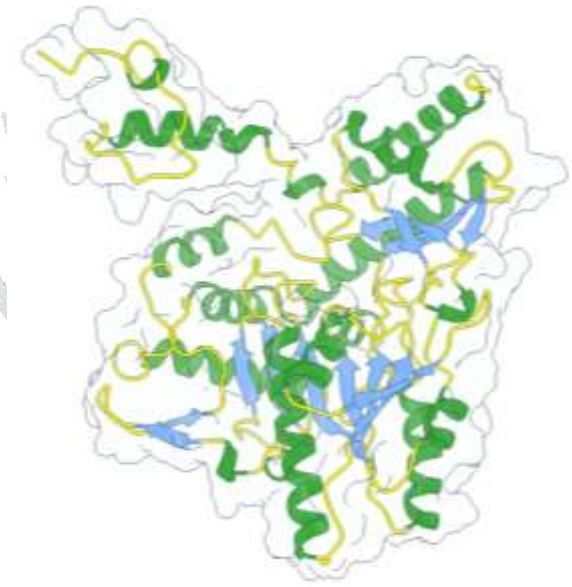

Figure 3: Three-dimensional model of smGAD. Helices and strands are shown in green and blue, respectively. The surface is shown with 90% transparency.

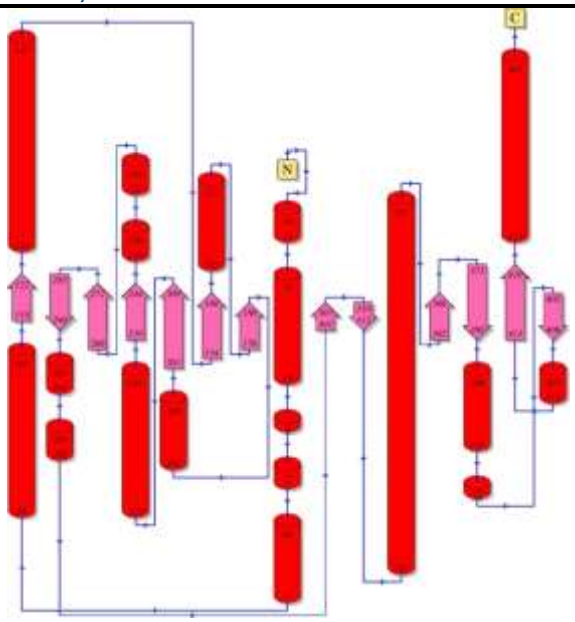

Figure 4: Topology of three-dimensional model of smGAD as predicted by PDBsum. Helices and strands are shown in red and pink, respectively.

IV. Conclusion

GABA as a plant metabolite is gaining attention in the scientific community. Detailed study of the glutamate decarboxylase (GAD) proteins will be pivotal for understanding GABA synthesis in plants. The information about physicochemical and structural properties of the GAD protein from *S. moellendorffii* (smGAD) by this study would be valuable for future purification and detailed *in vitro* and *in vivo* studies of this unexplored protein which can highlight the evolutionary aspects of this class of proteins.

V. Acknowledgment

The facility situated at the Department of Botany, Vivekananda College, Thakurpukur, Kolkata-700063 is gratefully acknowledged.

References

Banks J.A., Nishiyama T., Hasebe M., Bowman J.L., Gribskov M. *et al*. (2011) The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science 332(6032): 960-963.

Baum G., Chen Y., Arazi T., Takatsuji H., Fromm H. (1993). A plant glutamate decarboxylase containing a calmodulin binding domain. Cloning, sequence, and functional analysis, J. Biol. Chem. 268(26): 19610 -19617.

Behera S., Xu Z., Luoni L., Bonza M.C., Doccula F.G., De Michelis M.I., Morruis R.J., Schwarzländer M., Costa A. (2018). Cellular Ca^{2+} signals generate defined pH signatures in plants. Plant Cell. 30(11): 2704–2719.

Berman H.M, Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000). The Proten Data Bank. Nucleic Acids Res. 28(1): 235-242.

Bose J., Pottosin I.I., Shabala S.S., Palmgren M.G., Shabala S. (2011). Calcium efflux systems in stress signaling and adaptation in plants. Front. Plant Sci. 2: 85.

Bouché N., Fromm H. (2004). GABA in plants: Just a metabolite? Trends Plant Sci. 9(3): 110–115.

Bowie J.U., Lüthy R., Eisenberg D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. Science 253(5016): 164-170.

Cedano J., Aloy P., Pérez-Pons J.A., Querol E. (1997). Relation between amino acid composition and cellular location of proteins. J. Mol. Biol. 266(3): 594-600.

Colovos C., Yeates T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2(9): 1511-1519.

Dolan M.A., Noah J.W., Hurt D. (2012). Comparison of common homology modeling algorithms: application of user-defined alignments. In: A.J.W. Orry and R. Abagyan, (Eds.), Homology Modeling: Methods and Protocols, Methods in Molecular Biology, vol. 857, Humana Prerss, USA, pp: 399-414.

Erlander M.G., Tobin A.J. (1991). The structural and functional heterogeneity of glutamic acid decarboxylase: a review. Neurochem. Res. 16(3): 215–226.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In: J.M. Walker (Ed.), The Proteomics Protocols Handbook. Humana Press, Totowa, New Jersey, USA, pp: 571-607.

Geourjon C., Deléage G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput. Appl. Biosci. 11(6): 681-684.

Goldberg T., Hecht M., Hamp T., Karl T., Yachdav G. *et al*. (2014). LocTree3 prediction of localization. Nucleic Acids Res. 42(W1): W350-W355.

Gut H., Dominici P., Pilati S., Astegno A., Petoukhov M.V., Svergun D.I., Grutter M.G., Capitani G. (2009). A common structural basis for pH- and calmodulin-mediated regulation in plant glutamate decarboxylase. J. Mol. Biol. 392(2): 334-351.

Ishida T., Kinoshita K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 35(suppl_2): W460–W464.

Kinnersley A.M., Turano F.J. (2000). Gamma aminobutyric acid (GABA) and plant response to stress. Crit. Rev. Plant Sci., 19(6): 479–509.

Knight H., Trewavas, A.J., Knight M.R. (1997). Calcium signaling in *Arabidopsis thaliana* responding to drought and salinity. Plant J. 12(5): 1067–1078.

Krieger E., Joo K., Lee J., Lee J., Raman S., Thompson J., Tyka M., Baker D., Karplus K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins: Structure, Function, Bioinformatics 77(S9): 114–122.

Krieger E., Vriend G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics 30(20): 2981-2982.

Laskowski R.A. (2009). PDBsum new things. Nucleic Acids Res. 37(Database issue): D355-D359.

Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst. 26(2): 283-291.

Lüthy R., Bowie J.U., Eisenberg D. (1992). Assessment of protein models with three-dimensional profiles. Nature 356(6364): 83-85.

Oh S.-H., Choi W.-G., Lee I.-T., Yun S.-J. (2005). Cloning and characterization of a rice cDNA encoding glutamate decarboxylase. BMB Rep. 38(5): 595-601.

Paital B., Kumar S., Farmer R., Tripathy N.K., Chainy G.B.N. (2011). *In silico* Prediction and characterization of 3D structure and binding properties of catalase from the commercially important crab, *Scylla serrata*. Interdiscip. Sci. Comput. Life Sci. 3(2): 110-120.

Patil A., Nakamura H. (2006). Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. FEBS Lett. 580(8): 2041-2045.

Pettersen E.F., Goddard T.D., Huang C.C., Meng E.C., Couch G.S., Croll T.I., Morris J.H., Ferrin T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30(1): 70-82.

Robert X., Gouet P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42(W1): W320-W324.

Shelp B..J., Aghdam M.S., Flaherty E..J. (2021). γ-Aminobutyrate (GABA) Regulated Plant Defense: Mechanisms and Opportunities. Plants. 10(9):1939.

Shelp B.J., Bown A.W., Zarei A. (2017). 4-Aminobutyrate (GABA): A metabolite and signal with practical significance. Botany. 95(11): 1015–1032.

Shelp B.J., Bozzo G.G., Trobacher C.P., Chiu G., Bajwa V.S. (2012). Strategies and tools for studying the metabolism and function of γ-aminobutyrate in plants. I. Pathway structure. Botany 90(8): 651-668.

Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., de Beer T.A.P., Rempfer C., Bordoli L., Lepore R., Schwede T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1) W296-W303.

Wright P.E., Dyson H.J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J. Mol. Biol. 293(2): 321-331.

Yang R., Yin Y., Guo Q., Gu Z. (2013). Purification, properties and cDNA cloning of glutamate decarboxylase in germinated faba bean (*Vicia faba* L.). Food Chem. 138(2-3): 1945-1951.

Yin Y., Cheng C., Fang W. (2018). Effects of the inhibitor of glutamate decarboxylase on the development and GABA accumulation in germinating fava beans under hypoxia-NaCl stress. RSC Adv. 8(36): 20456-20461.