# Comparative *in-silico* Study of The Major Prolamin Protein of Wheat and Rice in Relation to Celiac Disease

**Mouli Nahar, Semanti Ghosh***

Department of Biotechnology, School of Life Sciences, Swami Vivekananda University, Telinipara, Barasat-Barrackpore Rd. West Bengal - 700121

*Corresponding author: semantig@svu.ac.in

**Abstract**

Celiac disease, an autoimmune disease, practically an immune response to consuming gluten protein popularly present in wheat, rye and barley, resulting "leaky gut "an inflammation of small intestine inner lining. Since wheat is enough infamous for being responsible for either celiac disease or gluten sensitivity, rice has become the only reliable source of staple around the world among people with gluten sensitivity. Glutenin and gliadin are thought to be responsible for the problem and glutenin is thought to be non-toxic. Rice proteins is in rich glutelin (oryzenin) and prolamins. The most common type of glutelin is glutenin and more recent studies reveal that in primary amino acid structure, there's great similarities and homogeneities between glutelin and gliadin. Not only that the comparison between the prolamin structure can be an indicator how it is related to even avenin sensitivity. Hence, having celiac disease can be triggered by consuming rice. Sequence alignment of proteins and phylogenetic tree has been done with these proteins to understand their homology and prediction of 3D structure. The superimposed structures of the proteins revealed the structural similarities and deviations as well as other enigmatic links with other allergens like cocosin.

**Keywords:** Celiac disease, Glutenin, Gliadin, Homology modeling, protein-protein interaction network

## Introduction

Celiac disease (CD), an autoimmune condition usually specified by one or more certain immunological and histological profiles generated through gluten ingestion in hereditary predisposed human organism (Fassano&Catassi;2012). Gluten, the alcohol-soluble protein found in different types of cereals, including wheat, rye, barley etc. (Fassano & Catassi; 2012). Among the existing maximum populations, roughly 1% of the population is affected by this disease and among 97% of American population is suffering from celiac disease. The condition manifests from being highly severe to moderate, having the overlapping symptoms like stomach cramps, indigestion, pain in joints similar to artharitis, nausea, foggy brain and abnormality in cognitive functions. The most common and major symptom is inflammation in the intestinal lining further resulting to the complete damage of the villies, commonly known as leaky gut. This kind of inflammation inhibits the complete absorption of certain nutrients and leads to abnormal immune response. Two seed storage proteins or prolamins – Glutenin and Gliadin, found in wheat (*Triticum aestivum*) majorly responsible for this problem, although glutenin reported to be non-toxic. When baking wheat products basically these two proteins come closer forming a interlinked situation (Wieser, 2007) and thus causing immune response to the consumer. Similar kind of prolamin found in rice is Glutelin (Oryzenin). The most basic difference between glutelin and glutenin is their difference in molecular weight. Another similar prolamin found in oats called Avenin. Since people with celiac disease cannot consume wheat products so major portion of the patients among the globe do rely on rice and rice products where some portions of the people also rely on oats. So, one of the major objectives of this study was to find whether people with such condition can blindly consume rice and rice products and if the similar prolamin-glutelin, present rice could trigger any such inflammation or any other kind of immune response. Also, we tried to investigate whether, this kind of protein had any structural similarities with any

other kind of allergen. The study also aimed to do the case study of structural comparison of proteins found in wheat and rice, respectively, related to celiac disease.

## Materials And Methods

### Sequence extraction

*In-silico* evaluation of plant proteins related to celiac disease and similar kind of other plant proteins were obtained from various literature study (Cornell & Townley,1974 ; Malalgoda & Simsek, 2017; Fassano *et al.* 2019).For the sequence study, we used the available data in FASTA format from Uniprot-KB/SWISSPROT(https://www.uniprot.org/).BLAST has been done for the protein sequence homology study.

### Homology modeling

The three-dimensional structure of Glutelin protein wasn't available in PBD, hence a homology model of this protein has been built using SWISS MODEL(https://swissmodel.expasy.org/) tool. CLUSTER-OMEGA version 1.2.4 (https://www.ebi.ac.uk/Tools/msa/clustalo) has been used to multiple sequence alignment of the concerned sequences. The superimposition was done by using the template PDB structure and as well as newly built glutelin structure by SUPERPOSE (http://superpose.wishartlab.com; Zhang & Wishart *et.al*, 2004) server.

### Protein network mapping

The proteins were uploaded into STRING (Snel *et al*. 2003,2005) database with species option set to the particular for inspecting the protein interaction network of the target proteins. STRING network(https://string-db.org) were used in medium (0.40) confidence interaction (0.70) and not more than 20 interactors. Concerned proteins with accession no. and their size in the table 1 below.

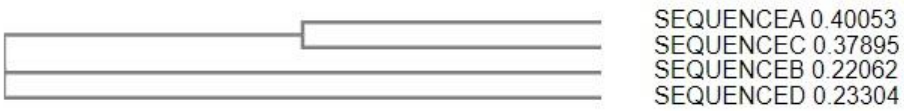
**Table 1: Proteins of concern, their accession id and size**

| Protein name | Accession ID | Amino acid Seq Length |
|---|---|---|
| Glutelin type-A 2 precursor [*Oryza sativa Japonica Group*] | NP_001390895.1 | 499 |
| Gliadin – *Triticum aestivum* (bread wheat) | PIR: A27319 | 296 |
| Glutenin - *Triticum aestivum* (Wheat) | P10388 | 838 |

## Results and Discussion

Pairwise sequence analysis(Fig 1A)of the concerned proteins- glutelin and gliadin showed less similarities, whereas alignment between the glutenin-glutelin and glutenin-avenin was very low (data not shown), diminishing the further case study of them. The obtained phylogenetic tree showed possible similar homology for glutelin and gliadin(Fig 1B).


A




B

**Fig 1. Multiple sequence alignment (Fig 1A) &Phylogenetic tree (Fig 1B) of glutelin (seq A), gliadin (seq C), glutenin (seq B) and avenin (seq D)**

A B

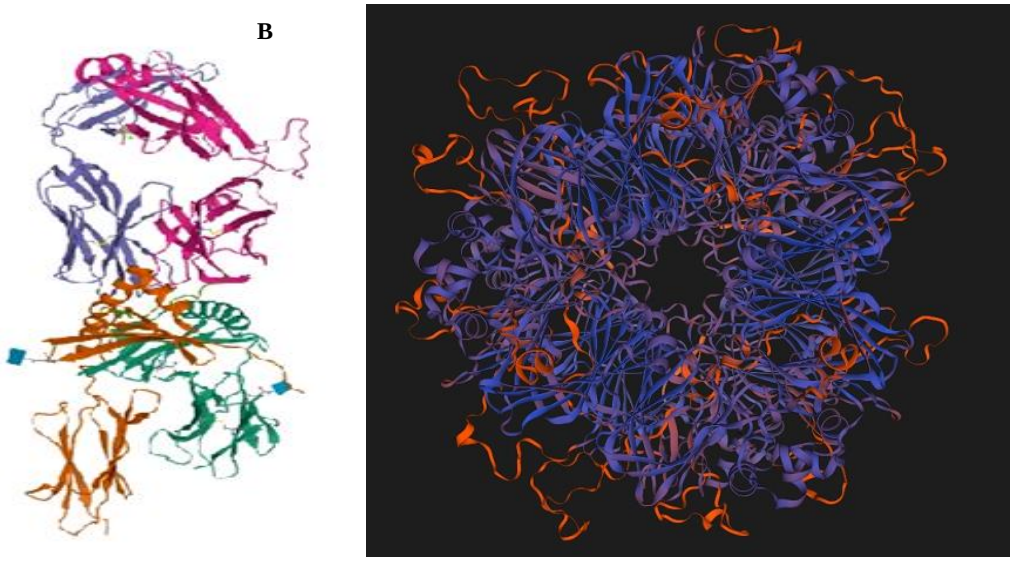
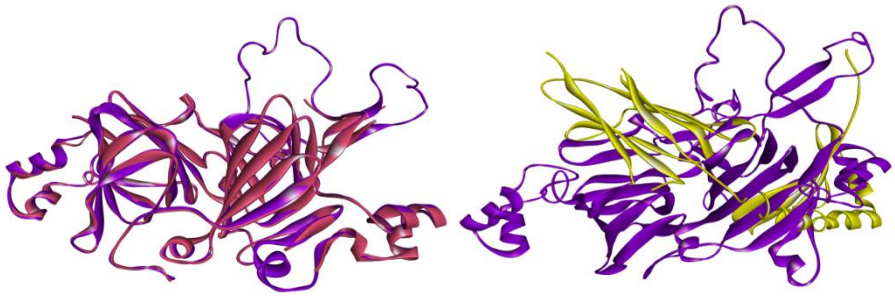


**Fig 2. Three-dimensional structure of gliadin PDB ID: 5KS9(2A) and glutelin (2B)**

While building the homology model for glutelin(Fig 3), the query coverage between the target and template obtained was 87%, having e-value 1e-158 with 51% sequence identity. The best template was found from coconut allergen known as cocosin.

The model was validated using SAVES (https://saves.mbi.ucla.edu/; Bowie *et al*,1991; Luethy *et al*,1992). The validation result showed 84.38% of residues having average 3D-1D score was >= 0.2, indicating a good quality model. The model protein contained 2637 aa residues and the obtained result of Ramachandran Plot showed 85.3% in the core, 12.8% as allowed region and 0.6% as disallowed region. The quality factor has been obtained as 87.706 by using program ERRAT2.

The structural comparison between the glutelin and template(5WPW) in superimposed condition revealed an ideal RMSD value i.e*;* 1.29 Å giving a greater imposed similarity (Fig 3A). The comparison between the superimposed structure of glutelin and gliadin (Fig 3B) gave a higher RMSD value *i.e;* 24.94 Å indicating very low structural similarities that agree with the low sequence identity also.



A B

**Fig 3. The superimposition of proteins. Glutelin vs template (5WPW) (Fig 3A); Gliadin (5KS9) vs. Glutelin (Fig 3B).** [Colour code: Red = template(5WPW); violet = Glutelin model &Gliadin (5KS9) yellow]

**Protein network mapping:**

No interaction between gliadin and other protein in wheat were found in the protein interaction network The protein interaction network has nodes and edges. The edges indicate both functional and physical protein associations and the nodes with highest average degree of connectivity to other gene symbols are the network's hub. From this network mapping we got an idea about the most and the least connected target proteins of the glutelin protein of rice and the glutenin protein of the wheat. Glutelin interaction shows 41 interacting nodes along with 392 edges combining both first and second shell interactors with known 3D structure. The PPI enrichment p-value is <1.0e-16, with average local clustering co-efficient of using STRING. While a huge and complicated protein network system in case glutelin in rice (Fig4) and glutenin in wheat (Fig 5) were observed.

0.836. The interaction network shows 20 proteins as first shell interactors and the rest of the interactors consider as second shell interactor. Glutenin interaction shows 38 interacting nodes along with 295 edges, having average local clustering co-efficient 0.658, with similar PPI p-value similar of glutelin (Table 2). The interaction network shows 20 proteins as the first shell interactors while 15 nodes among them having unknown 3D structure. 18 proteins act as second shell interactors while maximum of them lack known 3D structure.

**Table 2: Proteins of concern, source, symbol and no. of interacting proteins**

| Protein name | Organism | Gene symbol | No. of interacting proteins |
|---|---|---|---|
| Glutelin | *Oryza sativa* | Osj 06044 | 41 |
| Glutenin | *Triticum aestivum* | GluD3 | 38 |

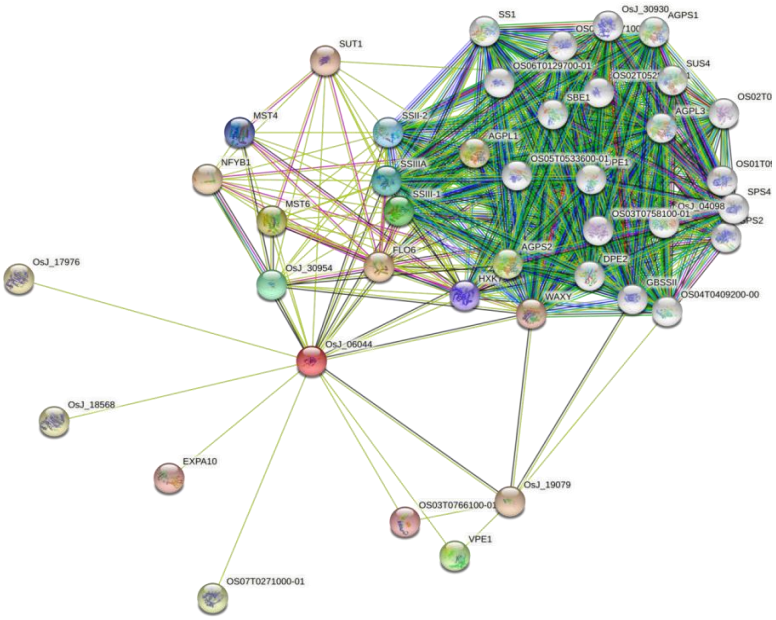



**Fig 4: Interaction between glutelin and other proteins in rice**

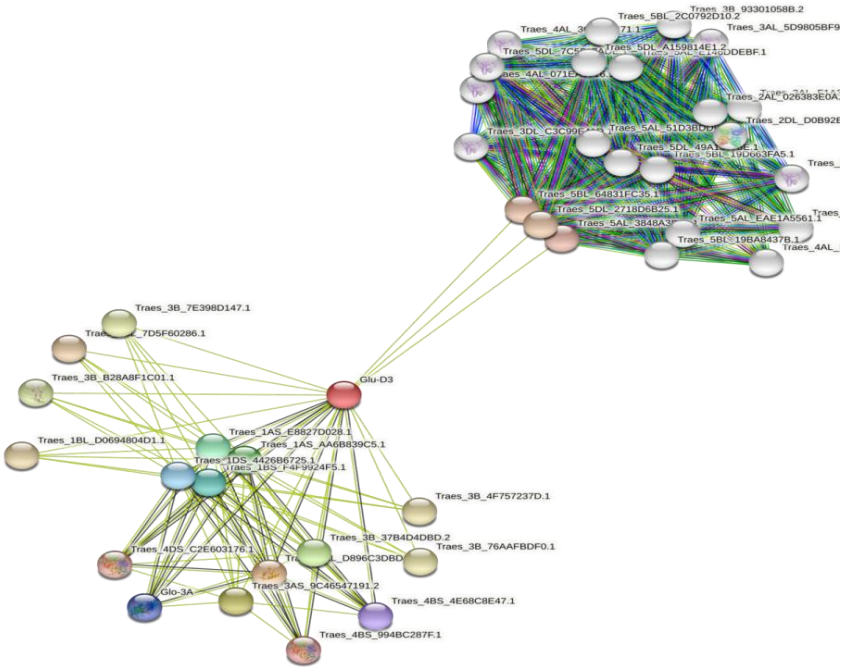



**Fig 5: Interaction between glutenin and other proteins in wheat**

## Conclusion

The study showed the sequence similarities between glutelin and gliadin is less, and the superimposition also stated the huge deviations between the two proteins. One of the sole responsible prolamin, i.e., glutenin although showed less similarities with glutelin, so does the avenin. During the homology model building, glutelin showed huge shared similarities with the coconut allergen cocosin.

Although rare, but some of the global population suffer from rice allergy or develop some kind of rice intolerance with time, the similarities between the rice prolamin and the cocosin might be the reason behind developing the sensitivity towards coconut for rice vulnerable people or vice versa. Since, coconut is considered as gluten free and the dissimilarities between the prolamins of wheat and rice, suggest people with celiac disease might consume rice except they have any intolerance towards rice and rice products.

**Conflict of interest**

The authors declare that they have no competing interests.

**Acknowledgement**

We sincerely acknowledge the facilities of Dept. of Biotechnology, School of Life Sciences, Swami Vivekananda University, for providing all the necessary facilities and support.

**References**

Bowie JU, Lüthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*,**253(5016):**164-70.

Caio G, Volta U, Sapone A, Leffler DA, De Giorgio R, Catassi C, Fasano A. 2019. Celiac disease: a comprehensive current review. *BMC Med*.**17(1):**142.

Catassi C, Fasano A. 2008.Celiac disease. *CurrOpin Gastroenterol*.**24:**687–691.

Cornell HJ, Townley RR. 1974. The toxicity of certain cereal proteins in coeliac disease. *Gut*.**15(11):**862-9.

Fasano A, Catassi C. 2012. Celiac disease. *N Engl J Med*. **367**:2419–2426.

Cohen IS, Day AS, Shaoul R. 2019. Gluten in Celiac Disease-More or Less? *Rambam Maimonides Med J*. **10(1):**e0007.

Lackner P, Koppensteiner WA, Sippl MJ, Domingues FS. 2000ProSup: a refined tool for protein structure alignment. *Protein Eng*. **13(11):**745-52.

Lüthy R, Bowie JU, Eisenberg D. 1992. Assessment of protein models with three-dimensional profiles. *Nature*.**356(6364)**:83-5.

Maneka M. Senay S. 2017, Celiac disease and other cereal proteins. *Food Hydrocolloids.***vol.68:**108-113

Orengo,C.A., Pearl,F.M. and Thornton,J.M. 2003. The CATH domain structure database. *Methods Biochem. Anal.*, **44, 249**–271.

RajarshiMaiti, Gary H. Van Domselaar, Haiyan Zhang, David S. Wishart, 2004.SuperPose: a simple server for sophisticated structural superposition, *Nucleic Acids Research*, **32(2):** W590–W594.

Snel B, Lehmann G, Bork P, Huynen MA. 2000. STRING: a web-server to retrieve and display therepeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.***28(18):**3442-4.

von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. 2003. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*.; **31:**258-61.

Wieser, H. 2007. Chemistry of gluten proteins. *Food Microbiol*. **24:**115-119.

Wieser, H., Antes, S., and Seilmeier, W. 1998. Quantitative determination of gluten protein types in wheat flour by reversed-phase high-performance liquid chromatography. *Cereal Chem*. **75:**644-650.