Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI - 10.26479/2017.0303.10**

IDENTIFICATION OF METALLOPROTEASES LIKE DOMAIN IN TRITICUM AESTIVUM α/β GLIADIN PROTEIN FROM BIOINFORMATICS PERSPECTIVE: A HYPOTHESIS

Selvaa Kumar C*, Sneha Dokhale

School of Biotechnology and Bioinformatics, D.Y. Patil University,

CBD Belapur, Navi Mumbai-400 614, India

ABSTRACT: Introduction: Gliadin intake causes celiac disease in selected population of *Homo sapiens*. Short motifs in α/β gliadin wheat protein were behind this food related disorder. **Materials and Methods:** Fold based template selection software GenTHREADER was considered for template selection and Modeller software was used for modelling wheat gliadin. T-Coffee software was used for the sequence alignment. Finally, antigenic region was identified. **Results:** Through *in silico* based study, we identified a novel motif 'HNxxH' in α/β gliadin wheat gliadin in C-terminal region which shows striking resemblance with the helical zinc binding motif 'HExxH' of metalloproteases. Interestingly, this short segment of gliadin is reported to be both an antigen and toxin which aligns well with the zinc binding site of the metalloproteases. Sequence analysis reports a spatial arrangement of the first two histidine residues from predicted motif of gliadin very much similar to the arrangement exhibited by the histidine residues in the selected metalloproteases like Astacin, Matrix metalloproteases and Predicted metal-dependent hydrolase. Additionally, glutamic acid, glycine and histidine from 2nd, 8th and 11th position were substituted by asparagine, isoleucine and glutamine, respectively in this domain. **Conclusion:** A wet lab based analysis could certainly confirm the evolutionary relatedness of both gliadin and metalloproteases from sequential, structural and immunological perspective.

KEYWORDS: Metalloproteinase, wheat gliadin, T-Coffee, Zinc, Zincins, HExxH

***Corresponding Author: Dr. Selvaa Kumar C** Ph.D.

School of Biotechnology and Bioinformatics, D.Y. Patil University,

CBD Belapur, Navi Mumbai-400 614, India. * Email Address: selvaakumarc@gmail.com

1. INTRODUCTION

Celiac disease is a food sensitive autoimmune disorder in *Homo sapiens* caused due to gliadin intolerance. Basically, this protein is reported in cereals like wheat (*Triticum aestivum*), rye (*Secale cereal*) and barley (*Hordeum vulgare*).[1-3] They generally exist in the form of monomeric subunit in α, β, γ, and ω gliadin proteins. Of these four types, α, β, and γ were structurally related to each other with repetitive N-terminal and non-repetitive C-terminal domains.[4] Symptoms associated with gliadin intake in children include growth inhibition, absorbtional disorder, chronic diahorrea and abdominal distension.[5, 6] In adults it causes fatigue, diahorrea and weight loss due to malabsorption.[7] Most importantly, this protein is also reported to bring in severe intestinal damage.[8] Literature survey confirms the presence of excess repeats of proline and glutamine residues in them.[9, 10] Their partial digestion with gastric and pancreatic enzyme provides prolline and glutamate rich stable and resistant short motifs which exhibits intolerance in *Homo sapiens* (PFPQPQLPY,PQPQLPYPQ(threecopies);PYPQPQLPY(twocopies); 33mer (LQLQPFPQPQLPY PQPQLPYPQPQLPYPQPQPF).[11-13] All these short stretches has an enormous toxic potency against the small intestinal mucosa with their excess amino acid repeats.[14, 15]. In this study α/β gliadin was investigated from sequence and structure perspective to identify novel motif, their antigenic and toxic regions in comparison with metalloprotease. Regarding metalloprotease, they are the most diverse form of proteins with more than 50 families reported till date. They were further classified based on their preference to metal binding residues. Majority of them have the conserved signature motif 'HExxH' zincins fold essential for zinc binding.[16] This happens to be the core active site helix wherein two histidine residues serve as zinc ligands and the additional glutamic acid assists in polarizing a water molecule actively involved in nucleophilic attack. They were first reported in thermolysin with diversified functions like cell proliferation, differentiation, inflammation, tissue remodeling, neurogenesis, angiogenesis, and apoptosis, wound healing and blood coagulation. [17-22] (Figure 1). Even though this active site is highly conserved, still there were sporadic reports of amino acid substitutions reported in the histidine (with arginine and lysine) and glutamic acid (with alanine, leucine and glutamate).[23] Such mutations were reported in Asp f2-Aspergillus fumigates (agent of allergic bronchopulmonary aspergillosis in *Homo sapiens*) [24]; CLCAs (Calcium-activated chloride channel regulator family) the novel zincins-like protease family which causes asthma, [25]; and ADAM family members (LWxxL or HQxxH) which were actively involved in sperm-egg interactions in the fertilization process.[26, 27]

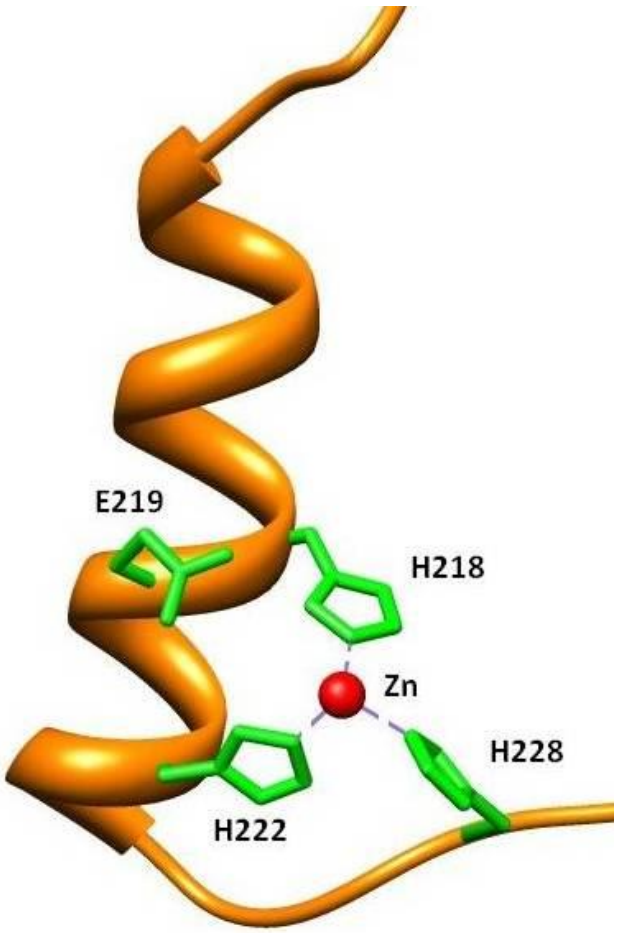


Figure 1. The ribbon structure of metalloproteinase active site with histidine residues interacting with zinc ion. The glutamic acid is also shown here.

An overall sequence and structure based comparison was instigated to identify novel motifs in gliadin protein in comparison with the members of metalloproteinase which shed light on their overall antigenicity and their toxicity from *in silico* perspective. The flow chart for the prediction of novel motif is listed in Figure 2.

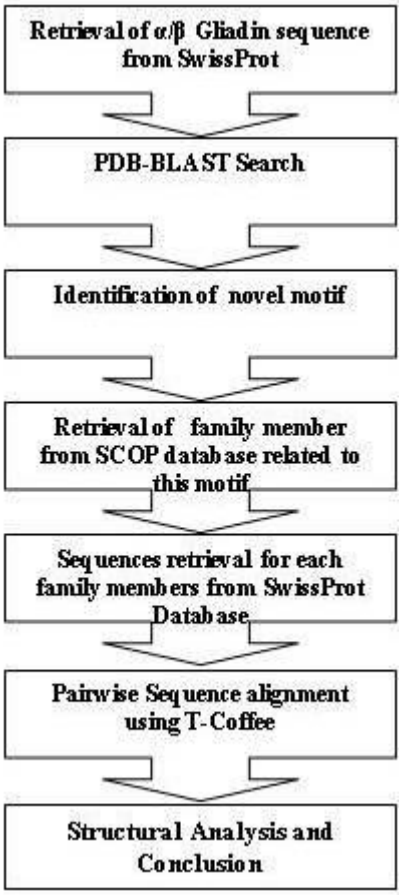


Figure 2. Flow chart for the identification of novel motif in gliadin protein.

2. MATERIALS AND METHODS

2.1 Sequence retrieval and Blast search

α/β gliadin protein from *Triticum aestivum* was retrieved from SWISSPROT database [28] (Swissprot id: P02863 (gliadin) and subjected to PDB-BLAST search [29] to identify potential template and domain details from Conserved Domain Database (CDD).[30] This data search has provided us with potential 3D structure of wheat gliadin from Protein Data Bank.[31] Additionally, from Uniprot database, 100 sequences consisting of 12 reviewed and 88 unreviewed sequences of α/β gliadin were downloaded and considered for multiple sequence alignment using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) which were further used to generate a WebLogo (http://weblogo.berkeley.edu/) [32] In SCOP database [33] "metalloprotease" keyword search was instigated to identify different members of this family. Of these, selected sequences with better sequence identity with gliadin were downloaded and considered for sequence analysis during the latter part of our study.

2.2 T-Coffee based multiple alignment

All the representative sequences were downloaded and aligned individually against gliadin protein using T-Coffee software (Tree-based Consistency Objective Function For alignment Evaluation).

[34] Subsequently, the overall residue identity, similarity and positional conservedness of motifs among these family members were studied individually.

2.3 Antigen Prediction

Antigenic regions were predicted for the gliadin and selected metalloprotease proteins using Predicted antigenic peptide (http://imed.med.ucm.es/Tools/antigenic.pl) [35,36], Antibody epitope prediction method (http://tools.immuneepitope.org/tools/bcell/iedb_input) and SYFPEITHI: Epitope prediction (http://www.syfpeithi.de/bin/MHCServer.dll/ EpitopePrediction.htm). [37]

2.4 Fold based template identification

Based on the sequence relatedness between gliadin and metalloproteinase, all four crystal structures viz. Astacin (1AST) [38], Matrix metalloproteases (1HOV) [39], Predicted metal-dependent hydrolase (1XM5) [40] and gliadin (2NNA) [41] were downloaded from Protein Data Bank. All three of them have complete crystal structures except for gliadin protein. Thus, we opted for fold based GenTHREADER (Rapid Fold Recognition) software [42] for template identification for wheat gliadin.

2.5 Homology modelling

As per the GenTHREADER report, the highest scored template was considered for gliadin modelling using the Modeller9.10 software.[43] Thus, twenty structures were generated out of which the structure with least DOPE (Discrete optimized protein energy) score was selected and further validated using PROCHECK [44] and Verify3D tools.[45]

2.6 Structural superimposition analysis

Modelled gliadin protein was superimposed against the downloaded crystal structures of Astacin, Matrix metalloproteases and Predicted metal-dependent hydrolase using CHIMERA software [46] with the zinc binding (HEXXH) active site. All these crystals have ions intact within the active site (1AST-*Astacus astacus* has bound zinc ion [47]; 1HOV-MMP-2 with bound zinc ion [48]; 1XM5-metal-dependent hydrolase ybeY from E.coli has bound with Nickel.[49]

3. RESULTS AND DISCUSSION

3.1 Sequence retrieval

Gliadin based PDB-BLAST search has listed PDB id: 2NNA [41] as the crystal structure which is a 18mer. This is an antigen associated with T-cell receptor which shows a conserved 'HNVVH' motif which resembles with the metalloprotease domain "HExxH" (Figure 3a). This motif was also observed in the WebLogo generated using the representative α/β gliadin dataset (Figure 3b). This has a striking resemblance with the metalloprotease domain "HExxH". To generate the complete structure here we opted for fold based template identification. Of the listed hits from GenTHREADER, 2S albumin Ber e1, the major allergen from Brazil nut was considered (PDB id: 2LVF) [50] as a potential template which shows an identity score of 22.8%. Furthermore, 19 sequences were extracted through SCOP database search (Table 1), of which only three sequences

has better sequence alignment with the C-terminal region of the gliadin protein. They are Astacin, Matrix metalloproteases (MMP2) and Predicted metal-dependent hydrolase.

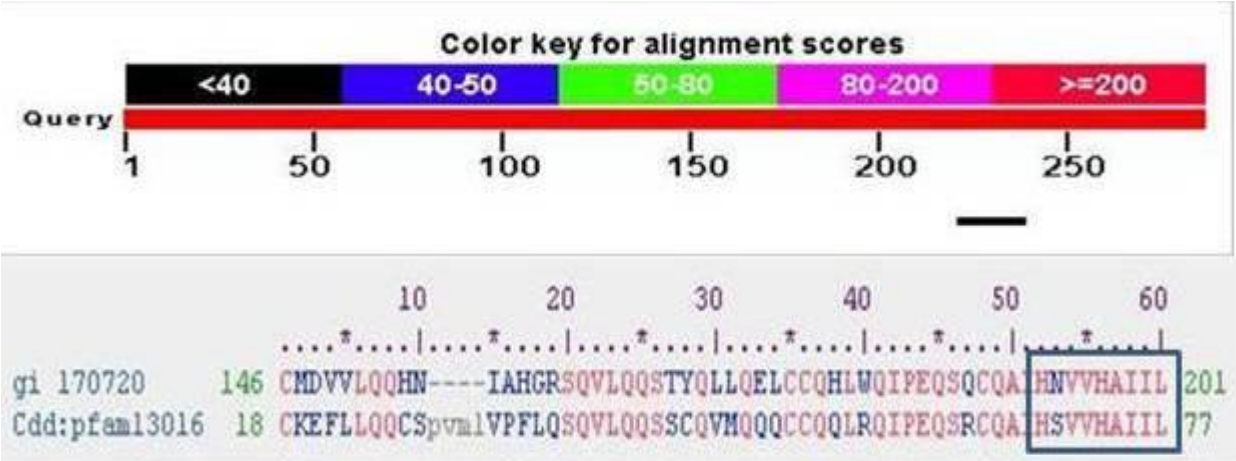


Figure 3(a). PDB-BLAST based search hit for α/β wheat gliadin. The listed PDB file has reported residue identity below 40% with black color code against the query sequence. The conserved domain database reports a 56mer domain for gliadin. This has a conserved 'HNxxH' (boxed) with no reported function.

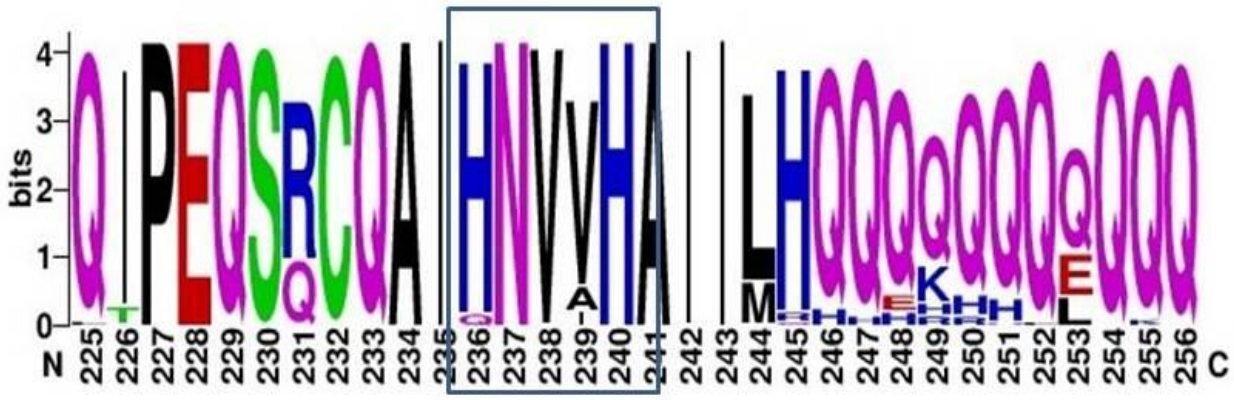


Figure 3 (b). Conserved motif 'HNxxH' in selected gliadin protein.

Family	Accession number	Protein	Organisms
Zinc protease	P56406	Extracellular small neutral protease	*Streptomyces caespitosus*
Fungal zinc peptidase	P46076	Neutral protease 2	*Aspergillus oryzae*
Thermolysin-like	P00800	Thermolysin	*Bacillus thermoproteolyticus*
Leukotriene A-4 hydrolase catalytic domain	P09960	Leukotriene A-4 hydrolase	*HUMAN*
NEP_HUMAN (Neprilysin)	P08473	NEP_HUMAN Neprilysin	*Homo sapiens*
Neurolysin, mitochondrial	P42676	Neurolysin, mitochondrial	*Rattus norvegicus*
Anthrax toxin lethal factor, N- and C-terminal domains	P15917	Lethal factor	*Bacillus anthracis*
Leishmanolysin	P08148	Leishmanolysin	*Leishmania major*
Serralysin-like metalloprotease, catalytic (N-terminal) domain	Q03023	Serralysin	*Pseudomonas aeruginosa*
Clostridium neurotoxins, catalytic domain	P10845	Botulinum neurotoxin	*Clostridium botulinum*
Astacin	P07584	Astacin	*Astacus astacus*
Reprolysin-like	P34179	Snake venom metalloproteinase adamalysin-2	*Crotalus adamanteus*
	Q9BZ11	Disintegrin and metalloproteinase domain-containing protein 33	*Homo sapiens*
TNF-alpha converting enzme, TACE, catalytic domain	P78536	Disintegrin and metalloproteinase domain-containing protein 17	*Homo sapiens*
Matrix metalloproteases, catalytic domain	P08253	72kDa type IV collagenase	*Homo sapiens*
Predicted metal-dependent	O67367	Endoribonuclease YbeY	*Aquifex aeolicus*

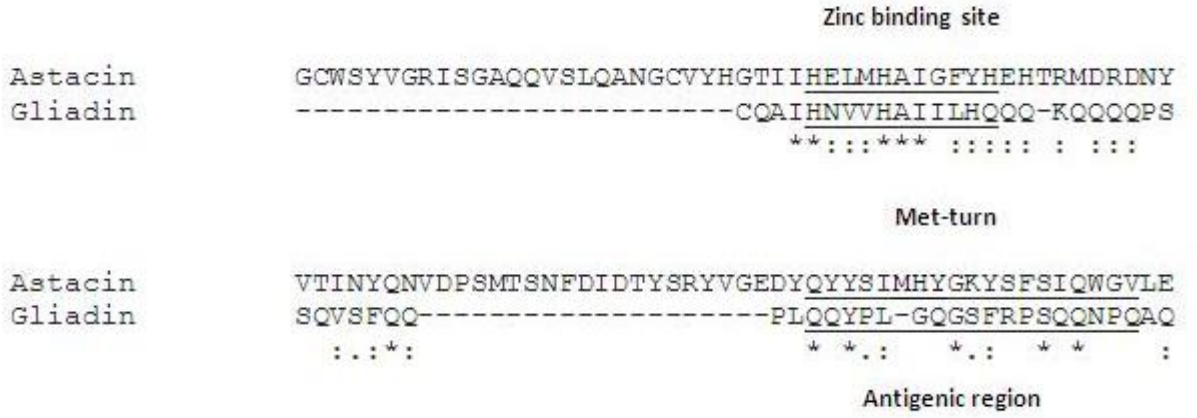
hydrolase	Q9X1J7	Endoribonuclease YbeY	*Thermotoga maritima*
	P0A898	Endoribonuclease YbeY	*Escherichia coli*
Cauro0242-like	A9WCA2	Uncharacterized protein CAUR 0242-like	*Chloroflexus aurantiacus*
TTHA0227-like	Q5SLR6	Uncharacterized protein TTHA00227-like	*Thermus thermophilus*

Table 1. The family members from metalloproteases reported from SCOP database with Family, Accession number, name of the protein and source organism details.

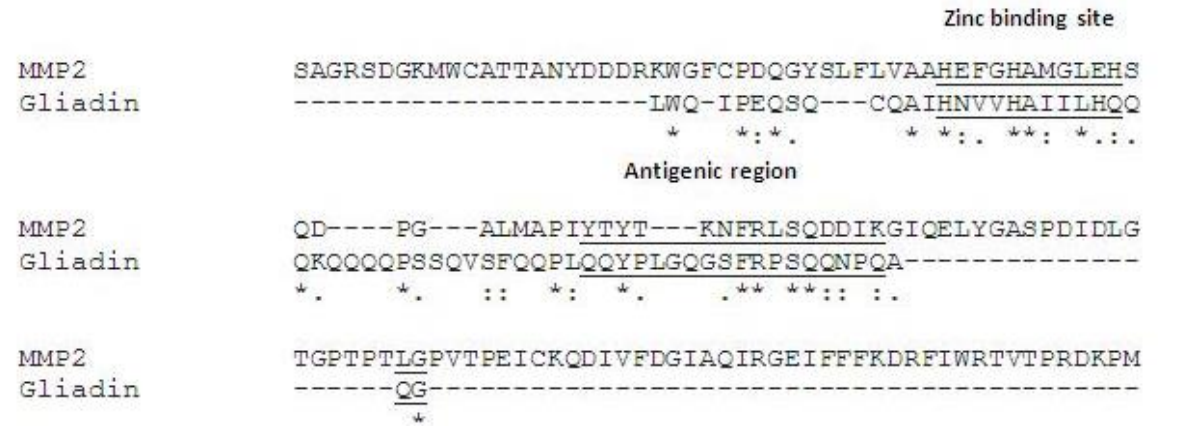
3.2 T-Coffee based alignment

Pairwise alignment of Gliadin protein with Astacin, Matrix metalloproteases (MMP2) and Predicted metal-dependent hydrolase was carried out using T-Coffee online tool. As per their alignment report, (a) zinc binding site of Astacin "HExxHxxGxxH" was aligned with the "HNxxHxxLxxQ" region of gliadin. Overall, first two histidine residues are strictly conserved in Astacin, predicted metal-dependent hydrolase and matrix metalloproteases. While the third histidine is substituted by glutamate due to the substitution of thymine with adenine (CAT->CAA). Even glutamic acid following third histidine is substituted by glutamate with the substitution of guanine with cytosine (GAA->CAA). Most importantly, glutamic acid in the 2nd position and glycine in the 8th position are not conserved in wheat protein. But there are four identical residues '*' and six conservative substitutions ':' were observed during this alignment. Regarding 'Met-turn' region, both methionine and tyrosine were deleted and substituted by glutamine respectively (Figure 4 a). (b) The zinc binding site of Matrix metalloproteases (zinc dependent endopeptidases) critically involved in extracellular matrix protein degradation [16] aligns with gladin. Here the zinc binding site 'HEFGHAMGLEH' aligns with gliadin region with four identical residues and three conservative substitutions. Histidine residues are highly conserved (Figure 4b).(c) Finally with respect to Predicted Metalloprotease (orthologous sequences of *Thermotoga maritima, Escherichia coli* and *Aquifex aeolicus*), the zinc binding site 'HGIVHLLGYDH' is partially conserved against the gliadin. The glutamic acid is substituted by glycine which is found to be asparagine in gliadin (Figure 4 c). Thus overall spatial arrangement of first two histidine residues in gliadin reciprocates well with metalloproteinase. (For complete sequence alignment please refer supplementary material).

4(a)



4(b)



4(c)

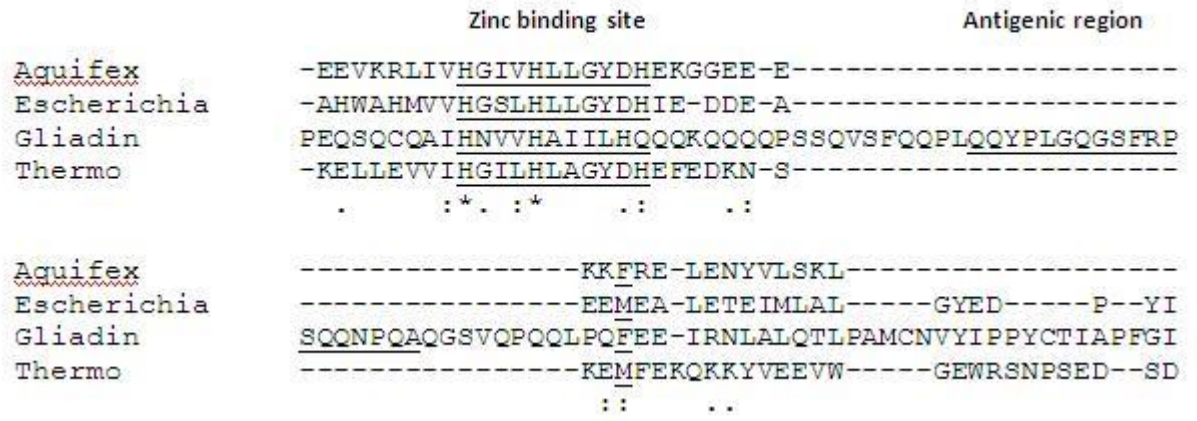


Figure 4. The pairwise alignment of gliadin against metalloprotease (a) Astacin and gliadin (b) MMP2 and gliadin (c) Predicted metal-dependent hydrolase comprising of sequences from *Aquifex aeolicus, Escherichia coli* and *Thermotoga maritima* against gliadin. Astacin protein with zinc

binding and the Met-turn regions highlighted. MMP2 and predicted metal-dependent hydrolase with their zinc binding sites. The '*' mark represents residue identity and the ':' represents conservative substitutions and the '·' signifies the non-conservative substitutions. The antigenic region of gliadin (18mer) is underlined here.

3.3 Antigen identification in Metalloproteases

This novel motif of wheat protein (HNVVHAIILHQQ) is an antigen [51] and also a toxin as per the BIOPEP database [52]. As per the antigenicity software analysis, the zinc binding site of metalloproteases (HExxHxxGxxH) is antigenic in Astacin and Predicted metal-dependent hydrolase except for MMP2 (Table 2). This study confirms with the report of antigenic nature of zinc binding site of Leishmania gp63 and mammalian endopeptidase-24.11. [53]

Protein	Predicted Antigenic Peptide	Antibody Epitope Prediction	Epitope Prediction
Astacin	SGAQQVSLQANGCVYH GTIIHELMHAIGFYH (121-151)	GAQQVSLQANGCVYHGT IIHELMHAIGFYHE(122-152)	I H E L M H A I G F Y (score 24)
MMP2	NA	NA	A H E F G H A M G L E (score 9)
Predicted metal-dependent hydrolase	VKRLIVHGIVHLLGY (109-123)	KRLIVHGIVHLLGYD(110-124)	L I V H G I V H L L G (score 4)

Table 2. The prediction of antigenic region from the zinc binding site of metalloproteases from *in silico* based methods for Astacin, MMP2 and Predicted metal-dependent hydrolase. The zinc binding site is underlined.

3.4 Homology modelling

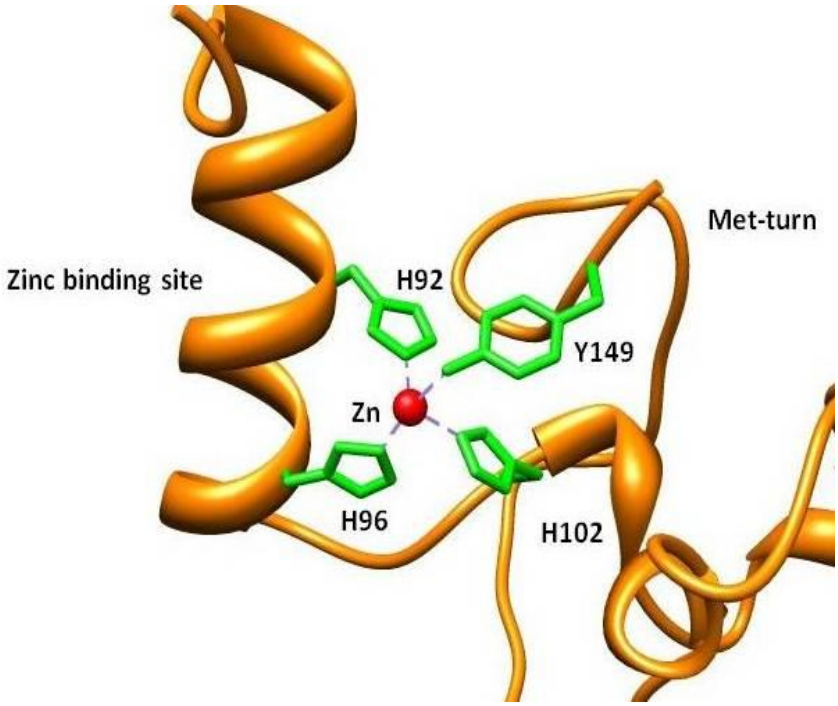
Template 2LVF was selected based on fold based method which was considered for modelling gliadin protein using Modeller software. Here regions like 119-217 and 250-285 showed better alignment with maximum query coverage of 39.86%. Thus only these regions could be modelled.

3.5 3D Structural analysis of metalloprotease and gliadin protein

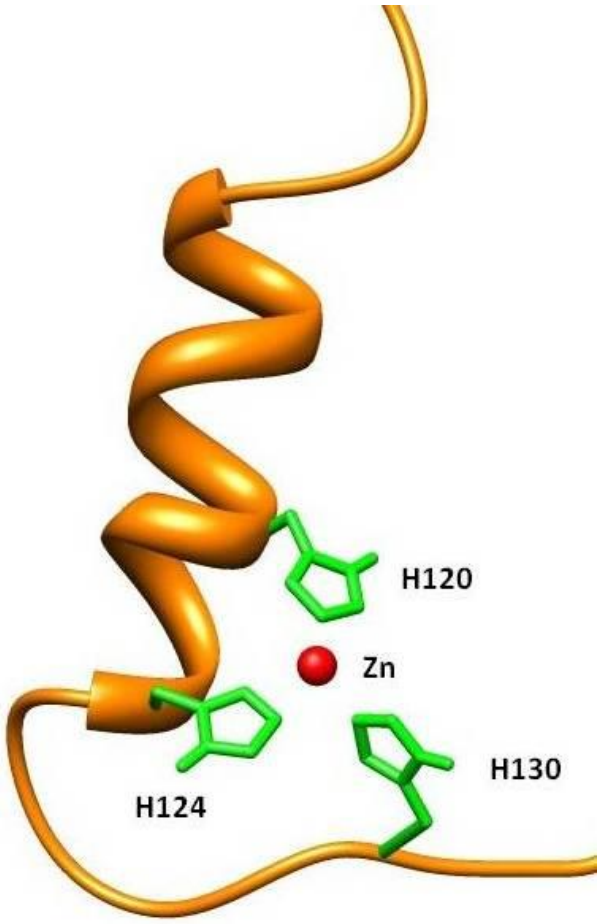
Modelled gliadin was superimposed with Astacin, MMP2 and predicted metal-dependent hydrolase active site region. The active site of Astacin 'HExxHxxGxxH' [54] has helix followed by a beta bridge. First two histidines (His92, His96) are present in the helix region and the third histidine (His102) in the loop region. Additionally, Y149 from the 'Met-turn' completes the active site (Figure 5a). In MMP2, His120 and His124 fall in helix region, while the His130 is observed in the loop region forming the zinc ion binding active site (Figure 5b). Again in Predicted metal-dependent hydrolase,

His1114 and His1118 are in the helix region, while His1124 is observed in the loop region (Figure 5c). However, in gliadin, His210, His214 and Gln220 (the substitute for third histidine) are observed in the extended helix region (Figure 5d). Thus metalloprotease has a 'helix-loop' active site which is found to be an extended helix region in gladin. Main reason behind this is the presence of glycine in 8^{th} position of metalloprotease which is reported to be a strong helix breaker. [55] This is substituted by isoleucine in gliadin which leads to extension of the helix secondary structure. As a result, it ultimately disorients the histidine residues present in the active site of the gliadin protein.

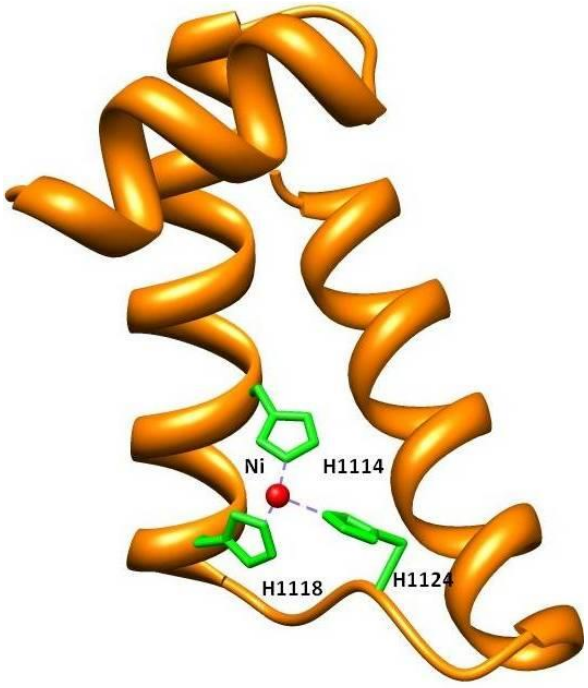
5(a)



5(b)



5(c)



5(d)

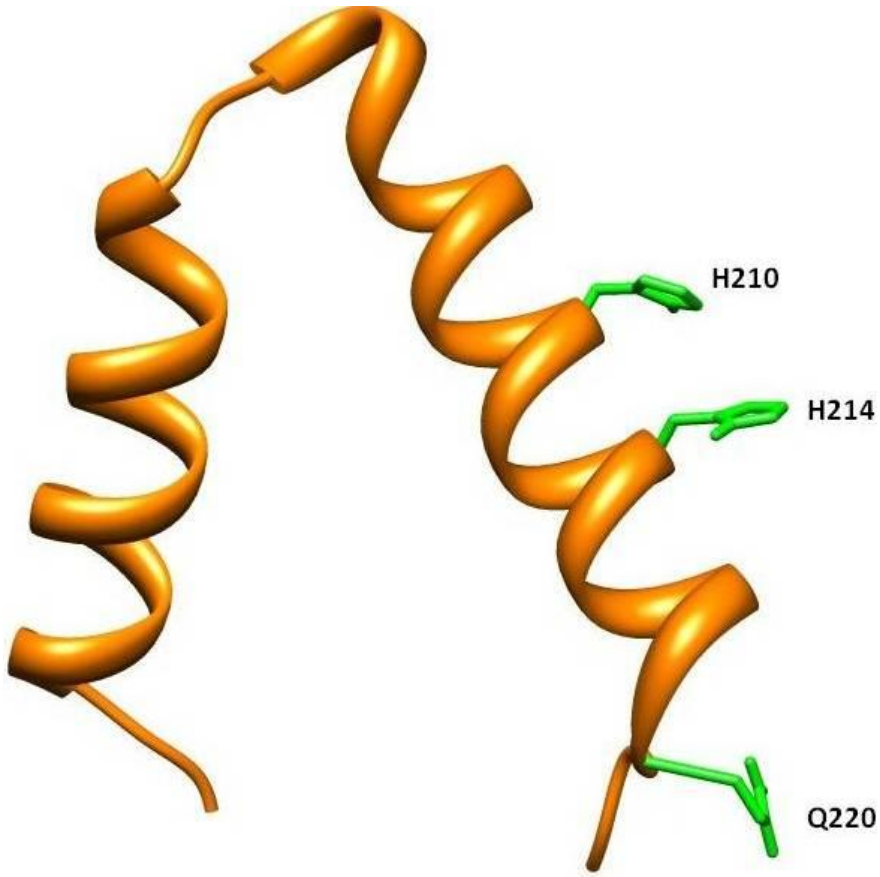


Figure 5. The crystal structure of metalloprotease and gliadin generated using CHIMERA software (a) Astacin protein with zinc binding site and Met-turn site interacting with zinc ion through 'helix–loop' conformation (b) MMP2 protein with a 'helix-loop' conformation interacting with zinc ion (c) Predicted metal-dependent hydrolase with a 'helix-loop' conformation interacting with nickel ion (d) Gliadin protein with two histidine and a glutamine residue in helical conformation.

4. CONCLUSION

This *in silico* based study report a broken zinc binding like motif in gliadin with highly conserved histidine residues which resembles like a distant homolog of metalloproteinase from sequential and structural perspective with conserved helix region. They have partially substituted residues in their predicted zinc binding site. Interestingly the aligned subsection portion of gliadin protein (HNVVHAIILHQQ) against the metalloproteases is reported to be an antigen and a toxin peptide which gets partially digested during the enzyme activity as reported earlier. Furthermore, as per the antigenicity report, the zinc binding site of metalloproteases is an antigen with a maximum score for Astacin protein sequence. To identify the antigenicity, this short active region of metalloproteases needs to be sequenced and their antigenicity needs to confirm through immunological studies. With the approval of this hypothesis we can confirm that the gliadin and the selected metalloproteases are similar from antigenic (immunological), sequential and structural perspective.

ACKNOWLEDGEMENT:

The authors acknowledge the management of D.Y. Patil University, CBD Belapur for providing the facilities to carry out this work.

CONFLICT OF INTEREST:

The authors have no conflict of interest.

REFERENCES

1. Schuppan D. Current concepts of celiac disease pathogenesis. Gastroenterol. 2000; 119 (1):234-42.
2. McAdam SN, L.M.Sollid. Gut.2000; 47:743-745.
3. Sollid LM, Molecular basis of celiac disease. Annu Rev Immunol.2000; 18:53-81.
4. Shewry PR, Tatham A S. The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J.1990; 267: 1–12.
5. Maki M, Collin P. Coeliac disease. The Lancet. 1997; 349(9067): 1755-1759.
6. Corrao G, Corazza GR, Bagnardi, V., et al., Mortality in patients with coeliac disease and their relatives: a cohort study. Lancet.2001; 358:356–61.
7. Knut, EA. Lundin, Armin Alaedini. Non-celiac Gluten Sensitivity. Gastrointest Endoscopy Clin N Am. 2012; 22: 723–734.
8. Wild D, Robins GG, Burley VJ, Howdle PD. Evidence of high sugar intake, and low fibre and mineral intake, in the gluten-free diet. Aliment. Pharmacol Ther. 2010; 32(4):573-81.
9. Wieser H. The precipitating factor in coeliac disease. Baillieres Clin Gastroenterol.1995; 9(2):191-207.
10. Dicke WK, Weijers HA, Van De Kamer J H. Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. Acta Paediatr. 1953; 42(1):34–42.
11. Arentz-Hansen H, Körner R, Molberg Ø, Quarstena H, Vaderd W, YMC. Kooy , Lundina KEA, Koning F, Roepstorff P, Sollid LM, McAdam SN. The intestinal T cell response in adult coeliac disease to α-gliadin is focused on a single deamidated glutamine targeted by tissue transglutaminase. J Exp Med. 2000; 191:603–12.
12. Arentz-Hansen H, McAdam SN, Molberg Ø, Fleckenstein B, Lundin KE, Jørgensen TJ, Jung G, Roepstorff P, Sollid LM. Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues.Gastroenterology.2002;123, 803.
13. Qiao SW, Bergseng E, Molberg Ø, Xia J, Fleckenstein B, Khosla C,Sollid LM. Antigen Presentation to Celiac Lesion-Derived T Cells of a 33-Mer Gliadin Peptide Naturally Formed by Gastrointestinal Digestion. J Immunol.2004; 173:1757-1762.
14. Shan L, Molberg Ø, Parrot I, Hausch F , Filiz F , Gray G.M , Sollid L M, Khosla C, Structural basis for gluten intolerance in celiac sprue. Science. 2002; 297, 2275–2279.

15. Green PH, Cellier C. Celiac disease. N Engl J Med. 2007; 357:1731–1743.
16. Rawlings ND, Barrett AJ. Evolutionary families of metallopeptidases. Meth. Enzymol. 1995; 248: 183–228.
17. Matthews BW, Jansonius JN, Colman MM, Schoeborn BB, Dupourge D. Three-dimensional structure of thermolysin. Nature New Biol.1972; 238: 37-41.
18. Matthews BW. Structural basis of the action of thermolysin and related Zinc Peptidases. Acct Chem Res. 1998; 21, 333-340.
19. Neurath H, Walsh KA. Role of proteolytic enzymes in biological regulation (a review). Proc Natl AcadSci.1976; 73: 3825–3832.
20. Page-McCaw A, Ewald AJ, Werb Z. Matrix metalloproteinases and the regulation of tissue remodelling. Nat Rev Mol Cell Biol.2007; 8: 221–233.
21. Turk B, Stoka V Protease signalling in cell death: caspases versus cysteine cathepsins. FEBS Lett. 2007; 581: 2761–2767.
22. Turk B, Turk du SA, Turk V. Protease signalling: the cutting edge. EMBO J. 2012;
31: 1630–1643.
23. Lenart A, Dudkiewicz M, Grynberg M, Pawłowski K. CLCAs - a family of metalloproteases of intriguing phylogenetic distribution and with cases of substituted catalytic sites. PLoS One. 2013; 8:e62272.
24. Banerjee B, Kurup VP, Phadnis S, Greenberger PA, Fink JN. Molecular cloning and expression of a recombinant Aspergillus fumigatus protein Asp f2 with significant immunoglobulin E reactivity in allergic bronchopulmonary aspergillosis. J Lab Clin Med. 1996; 127: 253-262
25. Nishimura H, Cho C, Branciforte DR, Myles DG, Primakoff P. Analysis of loss of adhesive function in sperm lacking cyritestin or fertilin beta. Dev Biol.2001; 233: 204–213.
26. Oh JS, Han C, Cho C. ADAM7 is associated with epididymosomes and integrated into sperm plasma membrane. Mol Cells. 2009; 28: 441–446.
27. Maki M, Collin P. Coeliac disease. Lancet. 1997; 349: 1755.
28. Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-Prot: Juggling between Evolution and Stability. Brief Bioinform.2004; 5: 39-55.
29. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, LipmanDJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acid Res. 1997; 25: 3389-3402.
30. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res.2011; 39: D225-D229.

31. Bernstein FC,Koetzle TF, Williams GJ, Meyer Jr EE, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures. J of MolBiol.1977; 112: 535-542.
32. Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, Genome Res. 2004; 14:1188-1190.
33. Hubbard T J. Ailey B, Brenner, S E, Murzin A G, Chothia C. SCOP: A Structural Classification of Proteins database". Nucleic Acids Research. 1999; 27 (1): 254–256.
34. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J MolBiol.2000; 302: 205–217.
35. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006; 2:2.
36. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990; 276:172-174.
37. Rammensee HG, Bachmann J, Emmerich NPN, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics.1999;50: 213-219
38. Bode W, Gomis-Ruth F X, Huber R, Zwilling R, Stocker W.Structure of astacin and implications for activation of astacins and zinc-ligation ofcollagenases. Nature. 1992; 358: 164-167.
39. Feng Y, Likos JJ, Zhu L, Woodward H, Munie G, McDonald JJ, Stevens AM, Howard C P, De Crescenzo G A , Welsch D, Shieh HS, Stallings W C.Solution structure and backbone dynamics of the catalytic domain of matrix metalloproteinase-2 complexed with a hydroxamic acid inhibitor. Biochim.Biophys.Acta.2002; 1598: 10-23.
40. Zhan C, Fedorov E V, Shi W, Ramagopal UA, Thirumuruhan R , Manjasetty B A, Almo SC, Fiser A , Chance M R , Fedorov A A. The ybeY protein from Escherichia coli is a metalloprotein. ActaCrystallogr Sect F Struct BiolCrystCommun. 2005; 61: 959-963.
41. Henderson, K.N., Tye-Din, J.A., Reid, H.H., Chen, Z., Borg, N.A., Beissbarth, T., Tatham, A., Mannering, S.I., Purcell, A.W., Dudek, N.L., van Heel, D.A., McCluskey, J., Rossjohn, J., Anderson, R.P.A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. Immunity. 2007 27: 23-34.
42. Lobley A, Sadowski MI, Jones DT. pGenTHREADER and pDomTHREADER: New Methods for Improved Protein Fold Recognition and Superfamily Discrimination. Bioinformatics. 2009; 25: 1761-1767.
43. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Meth. Enzymol.2003; 374: 461–491.
44. Laskowski R A, MacArthur M W, Moss D S, Thornton J M. PROCHECK - a program to check the stereochemical quality of protein structures. J App Cryst. 1993; 26, 283-291.

45. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature.1992; 356:83-85.
46. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE UCSF Chimera--a visualization system for exploratory research and analysis. J ComputChem.2004; 25:1605-12.
47. Bode W, Gomis-Ruth F X, Huber R, Zwilling R, Stocker W.Structure of astacin and implications for activation of astacins and zinc-ligation ofcollagenases. Nature. 1992; 358: 164-167.
48. Feng Y, Likos JJ, Zhu L, Woodward H, Munie G, McDonald JJ, Stevens AM, Howard C P, De Crescenzo G A , Welsch D, Shieh HS, Stallings W C.Solution structure and backbone dynamics of the catalytic domain of matrix metalloproteinase-2 complexed with a hydroxamic acid inhibitor. Biochim.Biophys.Acta. 2002; 1598: 10-23.
49. Rundqvist L, Tengel T, Zdunek J, Bjorn E, Schleucher J, Alcocer MJ, Larsson G. Solution structure, copper binding and backbone dynamics of recombinant Ber e 1-the major allergen from Brazil nut. PLOS ONE. 2012; 7: e46435-e46435.
50. Krupicková S, Tucková L, Flegelová Z, Michalak M, Walters JR, Whelan A, Harries J, Vencovský J, Tlaskalová-HogenováH.Identification of common epitopes on gliadin, enterocytes, and calreticulinrecognised by antigliadin antibodies of patients with coeliac disease Gut. 1999; 44 (2):168-73.
51. Minkiewicz P, Dziuba J, Iwaniak A, Dziuba M, Darewicz M. BIOPEP database and other programs for processing bioactive peptide sequences.J AOAC Int. 2008; 91: 965-980.
52. Dziuba M, Dziuba J, Iwaniak A. Bioinformatics-aided characteristics of the structural motifs of selected potentially celiac-toxic proteins of cereals and leguminous plants.Pol. J. Food NutrSci.2007; 57: 405-414.
53. SoteriadouKP,Tzinia AK, Panou-Pamonis E, Tsikaris V, Sakarellos-Daitsiotis M, Sakarellos C, Papapoulou Y, Matsas R. Antigenicity and conformational analysis of the Zn2+-binding sites of two Zn2+-metalloproteases : Leishmania gp63 and mammalianendopeptidase-24.11,Biochem J. 1996; 313: 455-46.
54. Bode W, Gomis-Ruth, F-X and Stocker W. Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metazincin'. FEBS Lett. 1993; 331: 134-140.
55. Chakrabartty A, KortemmeT, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. ProtSci.1994; 3:843-852.

SUPPLEMENTARY MATERIAL

(a)ASTACIN-GLIADIN.clustalw

```
ASTACIN         MQCAVLLVLLGVVAASPII------------------PEAARALYYNDGM
GLIADIN         MKTFLILVLLAIVATTATTAVRFPVPQLQPQNPSQQQPQEQVPL-VQQQQ
                *:  ::****.:**::.                    *:   .*  ::

ASTACIN         F-----------------------------------------------
GLIADIN         FLGQQQPFPPQQPYPQPQPFPSQLPYLQLQPFPQPQLPYSQPQPFRPQQP
                *

ASTACIN         --------------------EGDIKLRAGRQPARVGAAILGDEYLWSGGV
GLIADIN         YPQPQPQYSQPQQPISQQQQQQQQQQQQQQQQQQILQQILQQQLIPCMDV
                                    : : : :  :*  ::   ** :: : . .*

ASTACIN         IPYTFAGVSGADQSAILSGMQELEEKTCIRF--VPRTTESDYVEIFTSGS
GLIADIN         VLQQHNIAHGRSQVLQQSTYQLLQELCCQHLWQIPEQSQ-----------
                :   .  . * .*    *  * *:*  * ::  :*. ::

ASTACIN         GCWSYVGRISGAQQVSLQANGCVYHGTIIHELMHAIGFYHEHTRMDRDNY
GLIADIN         ------------------------CQAIHNVVHAIILHQQQ-KQQQQPS
                                         **:::*** ::::: : :::

ASTACIN         VTINYQNVDPSMTSNFDIDTYSRYVGEDYQYYSIMHYGKYSFSIQWGVLE
GLIADIN         SQVSFQQ--------------------PLQQYPL-GQGSFRPSQQNPQAQ
                  :.:*:                      * *.:   *.:  * *    :

ASTACIN         TIVPLQNGIDLTDPYDKAHMLQTDANQINNLYTNECSL-------RH
GLIADIN         GSVQPQQLPQFEEI--RNLALQTLPAMCNVYIPPYCTIAPFGIFGTN
                  *  *:  :: :   :   *** .   *   .  *::        :
```

(b)MMP2-GLIADIN.clustalw

```
GLIADIN         MKTFL--------------ILVLLAIVATTATTAVRFPVPQLQPQ-----
MMP2            MEALMARGALTGPLRALCLLGCLLSHAAAAPSPIIKFPGD-VAPKTDKEL
                *::::              :  **: .*::.:. ::**   : *:

GLIADIN         -----NPSQQQPQEQVPLVQQQQFLGQQQPFPPQQPYPQPQPFPSQLPYL
MMP2            AVQYLNTFYGCPKESCNLFVLKDTLKKMQKFF-----GLPQTGDLDQNTI
```

```
                    *.    *:*.  *.  :: * : * *        **.   :   :

GLIADIN         QLQPFPQPQLPYSQPQPFRPQQPYPQPQPQYSQPQQPISQQQQQQQQQQQQ
MMP2            ETMRKPRCGNPDVANYNFFPRKPKW------------------------
                :    *:   *      * *::*

GLIADIN         QQQQQQQILQQILQQQLIPC------------------------------
MMP2            ---DKNQITYRIIGY--TPDLDPETVDDAFARAFQVWSDVTPLRFSRIHD
                   :::**  :*:     *

GLIADIN         --MDVVLQQHNIAH-------------------------------------
MMP2            GEADIMINFGRWEHGDGYPFDGKDGLLAHAFAPGTGVGGDSHFDDDELWT
                   *::::  .  *

GLIADIN         -GRSQVLQQSTYQLLQELC--------------------------------
MMP2            LGEGQVVRVKYGNADGEYCKFPFLFNGKEYNSCTDTGRSDGFLWCSTTYN
                 *..**:: .  :   * *

GLIADIN         ---------CQH---------------------------------------
MMP2            FEKDGKYGFCPHEALFTMGGNAEGQPCKFPFRFQGTSYDSCTTEGRTDGY
                         * *

GLIADIN         -------------------------------------------------
MMP2            RWCGTTEDYDRDKKYGFCPETAMSTVGGNSEGAPCVFPFTFLGNKYESCT


GLIADIN         ----------------------LWQ-IPEQSQ---CQAIHNVVHAIILHQQ
MMP2            SAGRSDGKMWCATTANYDDDRKWGFCPDQGYSLFLVAAHEFGHAMGLEHS
                                      *   *:*.      * *:. **: *.:.

GLIADIN         QKQQQQPSSQVSFQQPLQQYPLGQGSFRPSQQNPQA--------------
MMP2            QD----PG---ALMAPIYTYT---KNFRLSQDDIKGIQELYGASPDIDLG
                *.    *.   ::  *:  *.    .** **:: :.

GLIADIN         ------QG-------------------------------------------
MMP2            TGPTPTLGPVTPEICKQDIVFDGIAQIRGEIFFFKDRFIWRTVTPRDKPM
```

```
                         *

GLIADIN          --------------------------------------------------
MMP2             GPLLVATFWPELPEKIDAVYEAPQEEKAVFFAGNEYWIYSASTLERGYPK


GLIADIN          ---------------------------------------SVQPQQLPQFEE
MMP2             PLTSLGLPPDVQRVDAAFNWSKNKKTYIFAGDKFWRYNEVKKKMDPGFPK
                                                      .*: :  * * :

GLIADIN          IRNLALQTLPAMCNVYIPPY--------------------CTIAPFGIFG
MMP2             LIADAWNAIPDNLDAVVDLQGGGHSYFFKGAYYLKLENQSLKSVKFGSIK
                 :   * :::*   :. :                        . . ** :

GLIADIN          T----N
MMP2             SDWLGC
                 :
```

(c)THERMOTO

GA-GLIADIN.clustalw

```
AQUIFEX          MSSTKRQKNRVLVKLKKRKVRKDKIEKWAELALSALGLNN------VELS
ECOLI            MSQVILDLQ-LACEDNSGLPEESQFQTWLNAVIPQF-QEE------SEVT
GLIADIN          MKTFLILVL-LAIVATTAT-------TAVRFPVPQL-QPQNPSQQQPQEQ
THERMOTOGAM      MIR-------ILGEGKGSKLLENLKEKLEEIVKKEI--GD------VHVN
                 *         :    .          .  .     :   :       .
AQUIFEX          VYITDDQEIRELNKTYRKKDKPTDV-------LSFPMGEE----------
ECOLI            IRVVDTAESHSLNLTYRGKDKPTNV-------LSFPFEVP----------
GLIADIN          VPLVQ-------QQQFLGQQQPFPPQQPYPQPQPFPSQLPYLQLQPFPQP
THERMOTOGAM      VILVSEDEIKELNQQFRGQDRPTDV-------LTFPLMEE----------
                 : :..       :  :  :::*           .**
AQUIFEX          ------------------FG-GYKILGDVVISQDTAERQARELGHSLE--
ECOLI            ------------------PGMEMSLLGDLVICRQVVEKEAQEQGKPLE--
GLIADIN          QLPYSQPQPFRPQQPYPQPQPQYSQ-PQQPISQQQQQQQQQQQQQQQQQ
THERMOTOGAM      ----------------------DVYGEIYVCPLIVEENAREFNNTFE--
                                      .   :  :.    :.: ::  :  :
AQUIFEX          --------------------------------------------------
ECOLI            --------------------------------------------------
```

```
GLIADIN          ILQQILQQQLIPCMDVVLQQHNIAHGRSQVLQQSTYQLLQELCCQHLWQI

THERMOTOGAM      --------------------------------------------------

AQUIFEX          -EEVKRLIVHGIVHLLGYDHEKGGEE-E----------------------

ECOLI            -AHWAHMVVHGSLHLLGYDHIE-DDE-A----------------------

GLIADIN          PEQSQCQAIHNVVHAIILHQQQKQQQQPSSQVSFQQPLQQYPLGQGSFRP

THERMOTOGAM      -KELLEVVIHGILHLAGYDHEFEDKN-S----------------------

                   .     :*. :*    .:    .:

AQUIFEX          ----------------KKFRE-LENYVLSKL------------------

ECOLI            ----------------EEMEA-LETEIMLAL-----GYED-----P--YI

GLIADIN          SQQNPQAQGSVQPQQLPQFEE-IRNLALQTLPAMCNVYIPPYCTIAPFGI

THERMOTOGAM      ----------------KEMFEKQKKYVEEVW-----GEWRSNPSED--SD

                                  ::    ..

AQUIFEX          SKAL

ECOLI            AEKE

GLIADIN          FGTN

THERMOTOGAM      PGKR
```

S1.Pairwise alignment of gliadin protein and metalloproteases showing complete sequences (a) Astacin and gliadin (b) MMP2 and gliadin (c) Predicted metal-dependent hydrolase comprising of sequences from *Aquifex aeolicus*, *Escherichia coli* and *Thermotoga maritima* against gliadin. The global alignment of gliadin against Astacin shows partial residue alignment in the N-terminal region and better alignment in the C-terminal region. MMP2 protein shows better alignment against gliadin both in the N-terminal and in the intermediate regions. Predicted metal-dependent hydrolase shows weak residue identity with gliadin. However, in all three cases, gliadin is found to be well aligned with the zinc binding site. The '*' mark represents residue identity and the ':' represents conservative substitutions and the '.' signifies the non-conservative substitutions.