Global Journal of Science Frontier Research: I
Interdisciplinary
Volume 16 Issue 3 Version 1.0 Year 2016
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Physiochemical and Functional Characterization of a Dominant Grain Endosperm Protein Called Glutelin in Rice (*Oryza Sativa* L.) using *in Silico* Methods

By E. Ramprasad, MNV Prasad Gajula, Ch. V. Durga Rani, G. Padmavathi & S. Vanisri

Professor Jayashankar Telangana State Agricultural University

Abstract- Glutelin protein is the most well-known abundant seed storage protein in rice seed endosperm. A total of 9 glutelin and glutelin type protein sequences from *Oryza* species available in uniport were evaluated by using bioinformatics tools to investigate physico-chemical properties, secondary structure prediction, putative phosphorylation sites and conserved motif search. Physicochemical analysis offers data such as pI, EC, AI, GRAVY and II about these sequences and the results showed that all glutelin protein sequences are basic, hydrophilic, thermo stable, having some extracellular portion. The secondary structure of the protein sequences were also predicted using SOPMA server. It was observed that alpha helix was predominant, followed by random coil, extended strand and least beta turn was found. Putative phosphorylation sites were also identified which are found to be conserved in plant species and the results showed that the most abundant phosphorylation site is serine residues in glutelin protein sequences.

Keywords: *glutelin protein, cupin family proteins, in silico, and homology modeling.*

GJSFR-I Classification: *FOR Code: 060799*

Strictly as per the compliance and regulations of:

Physiochemical and Functional Characterization of A Dominant Grain Endosperm Protein Called Glutelin in Rice (*Oryza Sativa* L.) using *in Silico* Methods

E. Ramprasad [α], MNV Prasad Gajula [σ], Ch. V. Durga Rani [ρ], G. Padmavathi [ω] & S. Vanisri [¥]

Abstract- Glutelin protein is the most well-known abundant seed storage protein in rice seed endosperm. A total of 9 glutelin and glutelin type protein sequences from *Oryza* species available in uniport were evaluated by using bioinformatics tools to investigate physico-chemical properties, secondary structure prediction, putative phosphorylation sites and conserved motif search. Physicochemical analysis offers data such as pI, EC, AI, GRAVY and II about these sequences and the results showed that all glutelin protein sequences are basic, hydrophilic, thermo stable, having some extracellular portion. The secondary structure of the protein sequences were also predicted using SOPMA server. It was observed that alpha helix was predominant, followed by random coil, extended strand and least beta turn was found. Putative phosphorylation sites were also identified which are found to be conserved in plant species and the results showed that the most abundant phosphorylation site is serine residues in glutelin protein sequences. Conserved protein motifs subjected to MEME to obtain the best possible matches. Other protein motifs found in the glutelin proteins are most of them belongs to cupin family proteins. The obtained results could be used for further *in silico* analysis and homology modeling studies of these glutelin proteins.

Keywords: glutelin protein, cupin family proteins, in silico, and homology modeling.

Abbreviations

pI	:	Isoelectric point
EC	:	Extinction coefficient
AI	:	Aliphatic index
GRAVY	:	Grand average of hydropathy
II	:	Instability index

I. Introduction

Rice (*Oryza sativa* L.), is one of the staple food crop for millions of people worldwide, provides 27 per cent of dietary energy supply and 20 per cent of dietary protein intake. Rice protein is superior in lysine content to wheat, corn and sorghum (Hegsted, 1969) and has a more balanced amino-acid profile. High-protein rice has the potential to enhance human nutrition in poor rural families where rice serves as the staple food (Li *et al.,* 2004). Therefore, in the improvement of rice storage protein, the main target has been to improve the quantity and nutritional quality of the protein in rice.

The major storage proteins found in rice are the glutelins, which according to previous studies, account for 80% or more of the total seed protein (Tecson *et al.,* 1971; Juliano, 1972; Villareal and Juliano, 1978). The remaining 20% is divided as follows: albumins, 1 to 5%; globulins, 4 to 15%; and prolamines, 2 to 8% (Houston *et al.,* 1968). Till date there are some little efforts have been made to characterize this rice glutelin protein. Earlier report on characterization of glutelin protein shows a remarkable similarity exists between the globulin storage protein fraction of oat (Brinegar and Peterson, 1982; Walburg and Larkins, 1983) and the 11S globulin or legumin fraction of pea (Derbyshire *et al.,* 1976) and soybean (Derbyshire *et al.,* 1976). Hence, the present study was undertaken and it was predicts some of the properties of rice glutelin protein such as physico-chemical properties, secondary structure prediction, putative phosphorylation sites, motifs searches etc. The study will be valuable to understand the structural features and molecular function of rice glutelin protein and will raise the prospects of its potential use in research. The obtained results could be used for further *in silico* analysis and homology modeling studies.

II. Material and Methods

a) Sequence retrieval

Expasy (uniprot KB) that provides protein sequences and annotation data (Jain *et al.,* 2009) was used to retrieve the chalcone synthase 1 protein sequences. These were downloaded in FASTA format to be used for further analysis (http://www.uniprot.org).

b) Physio-chemical characterization

For Physio-chemical characterization, theoretical Isoelectric Point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand

Author α σ ρ ¥: Institute of Biotechnology, Professor Jayashankar Telangana State Agricultural University (PJTSAU), Rajendranagar, Hyderabad-500030, India. e-mail: rambiotech100@gmail.com
Author ω: Plant Breeding, Crop improvement section, Indian Institute of Rice Research, Rajendra nagar, Hyderabad-500030, India.

average of hydropathy (GRAVY) were computed using the Expasy Protparm server (Gasteiger *et al.,* 2005) (http://us.expasy.org/tools/protparam.html).

c) Secondary structure prediction

SOPMA tool (Self-Optimized Prediction Method with Alignment) (Geourjon and Deleage, 1995) was applied to extract the information regarding the secondary structures that consist of Alpha helix, Extended strand, Beta turn and Random coil.

d) Putative phosphorylation sites and motif search

The amino acid sequences of the selected plants were analyzed for the putative phosphorylation sites at the NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/) (Blom *et al.,* 1999). Motif search (www.genome.jp/tools/motif) was used to find the number of motifs, motif ID, description and position of the motif found. Analysis of domain and conserved protein motifs was performed using MEME (http://meme.sdsc.edu/meme/meme.html) (Timothy *et al.,* 1994).

III. Results and Discussion

a) Physiochemical characterization

Glutelin protein sequences of *Oryza sativa* (Table 1) were analyzed in this study and corresponding protein sequences were collected from Uniport (http://www.uniprot.org/). Physiochemical properties of these protein sequences computed using Expasy Protparm server and the analyzed results were presented in table 2. The isoelectronic point is the pH at which the protein does not migrate in an electric field. It plays an important role in protein purification. The computed pI value that was less than 7 (pI<7) indicates that proteins were considered as acidic or greater than 7 (pI>7) reveals that proteins were basic in character. The pI value of all the sequences under study having more than 7 (pI>7) reveals that these proteins were basic in nature. The computed isoelectric point will be useful for separating the protein on a polyacrylamide gel by isoelectric focusing. Total numbers of negatively charged residues are lower than the total number of positively charged residues implies that these proteins are having extracellular portion. The extinction coefficient of a protein as calculated by the program depends on the molar extinction coefficient of Tyrosine, Tryptophan and Cysteine residues. Difference in the extinction coefficient values these glutelin proteins as evident from Table 2 was due to the difference in concentration of these three residues. The extinction coefficient can be used to calculate the concentration of a protein in solution. Instability index relies upon the occurrence of certain dipeptides along the length of the protein to distinguish between the unstable and stable protein. If the index is less than 40, it is probably stable in the test tube. If the value is greater than 40, it is probably not stable (Guruprasad *et al.,* 1990). The value for instability index for glutelin proteins are more than 40, hence these proteins are probably not stable (Guruprasad *et al.,* 1990). The aliphatic index refers to the relative volume of a protein that is occupied by aliphatic side chains and contributes to the increased thermo stability of protein. The aliphatic index of a protein is a measure of the relative volume occupied by aliphatic side chain of the following amino acids *viz.,* alanine, valine, leucine and isoleucine. The aliphatic index values of glutelin protein sequences ranging from 74.26 to 80.95. The very high aliphatic index of all glutelin protein sequences supports the view that these may be stable for a wide range of temperatures. Grand average of hydropathicity (GRAVY) index indicates the solubility of proteins: a positive GRAVY value indicates that proteins are hydrophobic in nature whereas a negative GRAVY value indicates more surface accessibility of the protein to interact with water (hydrophilic in nature). GRAVY values of glutelin protein sequences were ranged from -0.456 to -0.568. The very low GRAVY index of glutelin protein sequences implies that these protein sequences could result in a better interaction with water (hydrophilic in nature).

b) Functional characterization

The secondary structure of the protein sequences were predicted using SOPMA server (Table 3). It was observed that alpha helix was predominant, followed by random coil, extended strand and least beta turn was found. The secondary structure was predicted by using default parameters (window width 17, similarity threshold: 8 and number of conformational states: 4). Using the NetPhos 2.0 Server the putative phosphorylation sites were identified for glutelin proteins (Table 4). The output score was given in 0.000-1.000 range and the score above the threshold (0.500) shows the confidence rate of true phosphorylation site by the server. Several putative phosphorylation sites are completely conserved in plant species and interestingly more phosphorylation sites were found in these protein sequences. Conserved protein motifs subjected to MEME to obtain the best possible matches (table 5). Other protein motifs found in the glutelin proteins are most of them belongs to cupin family proteins (fig 2).

c) Conclusion

The present *in silico* study describes some important physiochemical and functional properties of rice glutelin proteins. Physiochemical and functional analysis reveals that rice glutelins are a basic, hydrophilic, thermo stable, having some extracellular portion and which has many phosphorylation sites. Conserved protein motifs are observed in these proteins and other protein motifs found in the glutelin proteins are most of them belongs to cupin family proteins. The obtained results could be used for further *in silico* analysis and homology modeling studies of these glutelin proteins.

IV. Acknowledgements

We express our gratitude to the Department of Biotechnology, Government of India for providing fellowship and to the Indian Institute of Rice Research, Hyderabad for providing the facilities.

References Références Referencias

1. Tecson, E.M.S., B.V. Esmana., L.P. Lontok and B.O. Juliano. 1971. Studies on the extraction and composition of rice endosperm glutelin and prolamin. *Cereal Chemistry*. 48: 186-181.
2. Villareal R.M., B.O. Juliano. 1978. Properties of glutelin from mature and developing rice grain. *Phytochemistry*. 17: 177-182.
3. Mann C. Reseeding the green revolution. Science 1997; 277: 1038-43.
4. Brinegar, A.C., and D.M. Peterson. 1982. Separation and characterization of oat globulin polypeptides. *Arch Biochem Biophys*. 219: 71-79.
5. Walburg, G and B.A. Larkins. 1983. Oat seed globulin. *Plant Physiol*. 72: 161-165.
6. Derbyshire, E., D.J. wright and D. Boulter. 1976. Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry*. 15: 3-24.
7. Jain E, A Bairoch, S Duvaud, I Phan, N Redaschi, B E Suzek, M J Martin , P, McGarvey, E Gasteiger (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics, 10.
8. Gasteiger, C.Hoogland, A.Gattiker, S.Duvaud, M.R.Wilkins, R.D. Appel, A.Bairoch. Protein Identification and Analysis Tools on the ExPASy Server, (In) John M.Walker (ed): The Proteomics Protocols Handbook, Humana Press. (2005): 571-607.
9. Geourjon C, G Deleage (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci, 11, pp. 681-684.
10. Blom N., Gammeltoft S. and Brunak S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Biology, Vol: 294 (5) pp: 1351-1362.
11. Timothy, L., Bailey and Charles Elkan. (1994). "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California.

Table 1: Details of glutelin protein sequences from *Oryza sativa*

Entry	Entry name	Protein names	Gene names	Organism	Length (No. of A.A)
Q09151	GLUA3_ORYSJ	Glutelin type-A 3	GLUA3 GLUA-3 GT22 GT3 Os03g0427300 LOC_Os03g31360 OSJNBa0083F15.19	*Oryza sativa* subsp. *japonica* (Rice)	496
P07728	GLUA1_ORYSJ	Glutelin type-A 1	GLUA1 GLUA-1 Os01g0762500 LOC_Os01g55690 P0460E08.38 P0512C01.36	*Oryza sativa* subsp. *japonica* (Rice)	499
P07730	GLUA2_ORYSJ	Glutelin type-A 2	GLUA2 GLUA-2 GT1 Os10g0400200 LOC_Os10g26060 OSJNBa0050N08.16	*Oryza sativa* subsp. *japonica* (Rice)	499
P14323	GLUB1_ORYSJ	Glutelin type-B 1	GluB1-A GluB-1 Os02g0249800 LOC_Os02g15169 OJ1113_G05.6 OSJNBa0011N12.36; GLUB1-B GLUB-1 Os02g0249900 LOC_Os02g15178 OJ1113_G05.4 OSJNBa0011N12.34	*Oryza sativa* subsp. *japonica* (Rice)	499
Q6ERU3	GLUB5_ORYSJ	Glutelin type-B 5	GLUB5 GLUB-5 Os02g0268100 LOC_Os02g16820 P0693E08.14	*Oryza sativa* subsp. *japonica* (Rice)	500
P14614	GLUB4_ORYSJ	Glutelin type-B 4	GLUB4 GLUB-4 Os02g0268300 LOC_Os02g16830 P0693E08.16	*Oryza sativa* subsp. *japonica* (Rice)	500
Q02897	GLUB2_ORYSJ	Glutelin type-B 2	GLUB2 GLUB-2 GluB-7 GLUB7 Os02g0249600 LOC_Os02g15150 OSJNBa0011N12.30	*Oryza sativa* subsp. *japonica* (Rice)	495
Q0E261	Q0E261_ORYSJ	Glutelin	Os02g0268300 OsJ_06189	*Oryza sativa* subsp. *japonica* (Rice)	500
Q40689	Q40689_ORYSA	Glutelin	Gt2	*Oryza sativa* (Rice)	499

Table 2: Details of Physiochemical Properties of chalcone synthase 1 protein sequences from different species

Entry	Entry name	Length (No. of A.A)	M.wt	pI	-R	+R	EC	II	AI	GRAVY
Q09151	GLUA3_ORYSJ	496	56015.0	8.81	44	51	45435	46.23	80.95	-0.456
P07728	GLUA1_ORYSJ	499	56246.9	9.09	42	52	45435	51.36	76.77	-0.539
P07730	GLUA2_ORYSJ	499	56306.1	8.93	42	50	50935	50.18	76.37	-0.509
P14323	GLUB1_ORYSJ	499	56550.5	9.26	39	50	50685	52.11	76.19	-0.495
Q6ERU3	GLUB5_ORYSJ	500	56808.0	9.00	42	50	43820	47.94	78.22	-0.508
P14614	GLUB4_ORYSJ	500	56818.0	9.00	42	50	43820	47.81	78.22	-0.510
Q02897	GLUB2_ORYSJ	495	56046.8	9.11	39	48	50685	51.22	74.26	-0.498
Q0E261	Q0E261_ORYSJ	500	56818.0	9.00	42	50	43820	47.81	78.22	-0.510
Q40689	Q40689_ORYSA	499	56239.9	9.09	42	52	45435	50.14	75.19	-0.568

Table 3: Details of secondary structures of chalcone synthase 1 protein sequences from different species

Entry	Entry name	Alpha helix	Extended strand	Beta turn	Random coil
Q09151	GLUA3_ORYSJ	32.86%	21.98%	13.91%	31.25%
P07728	GLUA1_ORYSJ	32.67%	20.64%	14.83%	31.86%
P07730	GLUA2_ORYSJ	32.26%	21.84%	12.42%	33.47%
P14323	GLUB1_ORYSJ	33.87%	21.04%	12.63%	32.46%
Q6ERU3	GLUB5_ORYSJ	35.20%	22.00%	11.00%	31.80%
P14614	GLUB4_ORYSJ	35.20%	20.80%	11.00%	33.00%
Q02897	GLUB2_ORYSJ	36.36%	18.99%	13.74%	30.91%
Q0E261	Q0E261_ORYSJ	35.20%	20.80%	11.00%	33.00%
Q40689	Q40689_ORYSA	29.86%	21.44%	14.63%	34.07%

Table 4: Putative phosphorylation residues in chalcone synthase 1 protein sequences from different species

Entry	Entry name	Putative phosphorylation residues		
		Serine	Threonine	Tyrosine
Q09151	GLUA3_ORYSJ	15	2	3
P07728	GLUA1_ORYSJ	15	2	3
P07730	GLUA2_ORYSJ	21	3	4
P14323	GLUB1_ORYSJ	14	2	3
Q6ERU3	GLUB5_ORYSJ	20	3	4
P14614	GLUB4_ORYSJ	11	4	2
Q02897	GLUB2_ORYSJ	14	2	3
Q0E261	Q0E261_ORYSJ	14	2	3
Q40689	Q40689_ORYSA	15	2	3

Table 5: Different motifs commonly conserved in glutelin protein sequences with best possible match amino acid sequences

Motif	Width	Best Possible Match
1	50	SQSQKFRDEHQKIHRFRQGDIVALPAGVAHWCYNDGDAPVVAIYVTDLNN
2	50	HYVVLKKAEHEGCQYIAFKTNPNSMVSHMAGKNSIFRAMPVDVIANAYRI
3	50	ADTYNPRAGRITNLNSQKFPILNLVQMSATKVNLYQNAILSPFWNINAHS

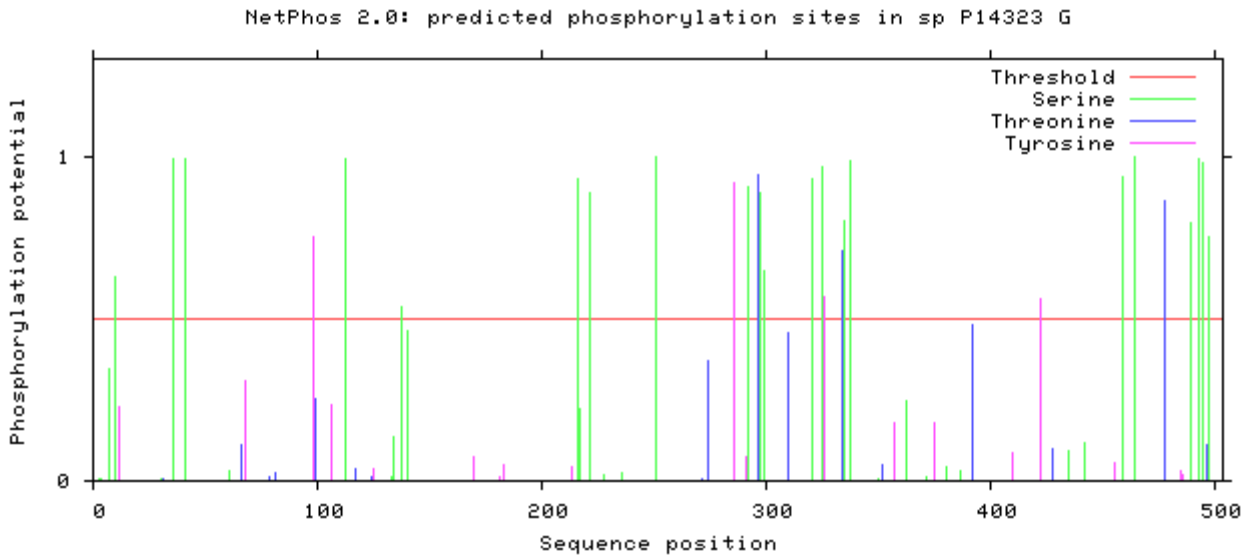

Fig.1: Putative phosphorylation residues in Glutelin type-B 1 protein sequences from *Oryza sativa* subsp. *japonica* (Rice)

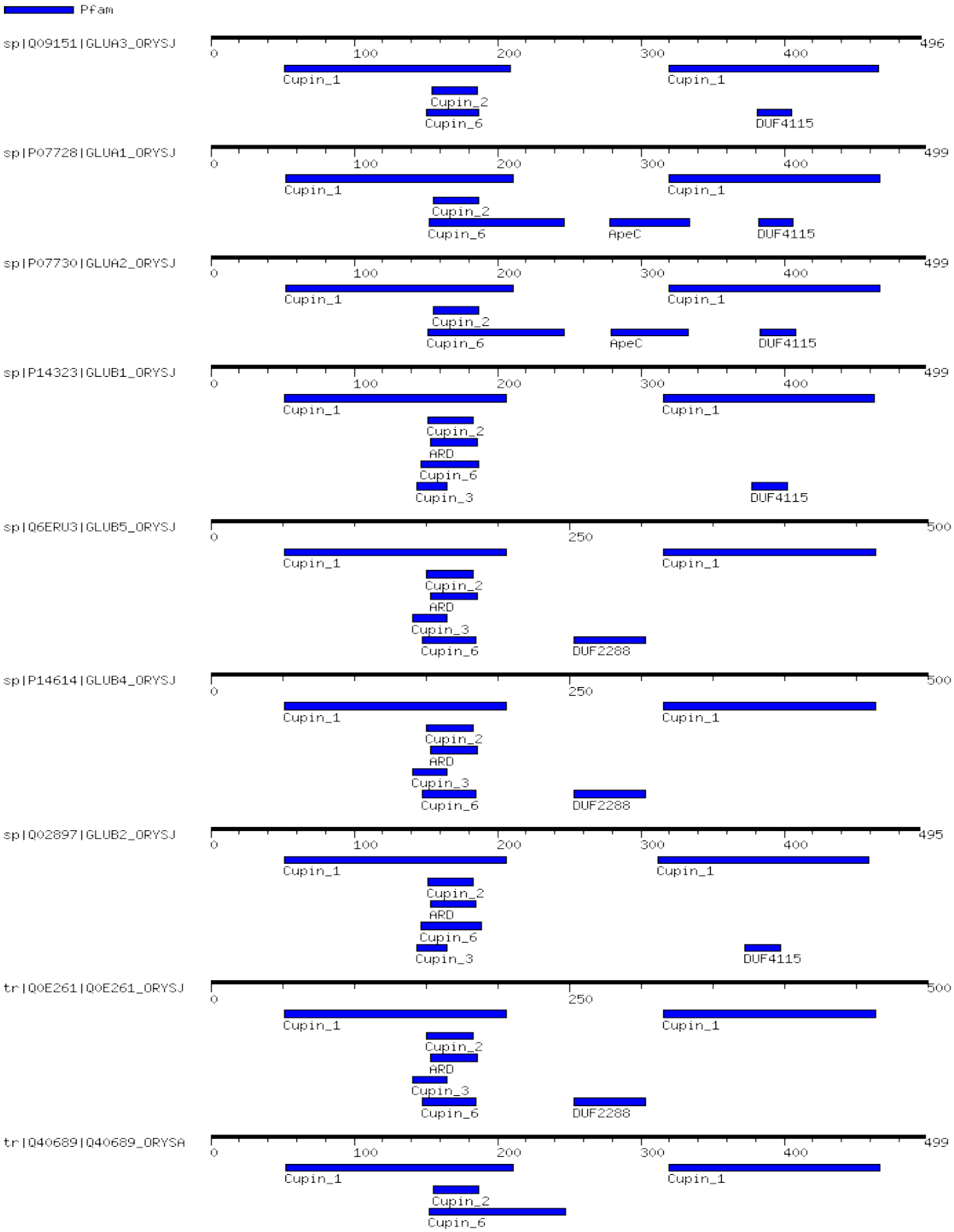

Fig. 2: Different other protein motifs found in glutelin proteins of *Oryza sativa*