ISSN: 2997-2728

Research Article

Journal of Biotechnology and Bioinformatics

Correlation between Protein Disorder and Post Translational Modifications in Resistance Protein of Plants

Samsed Ahmed Urmee[1]*, Rubel Hasan[2] and Jahangir Alam[1]

[1]Shahjalal University of Science & Technology, Faculty of Life Science, Department of Genetic Engineering & Biotechnology, Sylhet, 3114

[2]University of Dhaka, Department of Management Information systems, Dhaka

[*]Corresponding Author
Samsed Ahmed Urmee, Shahjalal University of Science & Technology, Faculty of Life Science, Department of Genetic Engineering & Biotechnology, Bangladesh.

Submitted: 2024 , Jun 06; **Accepted**: 2024, Aug 28; **Published**: 2024, Sep 04

Citation: Urmee, S. A., Hasan, R., Alam, J. (2024). Correlation between Protein Disorder and Post Translational Modifications in Resistance Protein of Plants. *J biotech bioinform,* 1(4), 01-18.

Abstract

Structural analysis and characterization of individual proteins in plants is less studied than animals and microbes. Intrinsic disordered region is highly abundant in plant proteins which play critical role in plant biology & act as an integrator of signals from multiple plant regulatory inputs. In this work, we calculated the degree of intrinsic disorder in the Resistance protein (R protein) of seven typical monocotyledonous (monocot) and dicotyledonous (dicot) plant species. R protein-based resistance of plants to various pathogens is a main area of interest in plant innate immunity. We further predicted multiple sites for phosphorylation, glycosylation, acetylation and methylation and systematically investigated the correlation between protein disorder and post-translational modifications (PTMs). It was found that phosphorylation, and acetylation displayed a clear preference for occurrence in disordered regions of plant proteins. In contrast, methylation prefers to be ordered, whereas N-glycosylation did not show a universal structural preference in monocotyledonous and dicotyledonous plants. In addition, this study revealed significant differences between some of the characteristics of R proteins in monocot and in dicot plant. They included disorder degree, increased rate of R-methylation, decreased rate of N-glycosylation and K-methylation in monocotyledonous plant species, as compared with dicotyledonous species. Altogether, this study sheds light on the connection between protein disorder and multiple PTMs in R proteins of plants.

1. Introduction

The activation of plant defense to restrict pathogen invasion is often conferred by Resistance (R) proteins. The most prevalent class of R proteins contains leucine rich repeats (LRRs), a central nucleotide binding site and a variable amino terminal domain. Other classes possess an extracellular LRR domain, a transmembrane domain and sometimes an intracellular serine/threonine kinase domain. R proteins function in pathogen perception and/or the activation of conserved defense signaling networks.

The functions of various R proteins require phosphorylation, protein degradation, or specific localization within the host cell. Some signaling components are shared by many R gene pathways whereas others appear to be pathway specific. Recent results provide new insights into how the signaling potential of R proteins might be created, managed and held in check until specific stimulation following infection [1].

Intrinsically disordered proteins (IDPs) are highly abundant in eukaryotic proteomes including R proteins. More than one third of eukaryotic proteins contain intrinsic disordered region (IDR) of more than 30 residues long [2]. Plant IDPs play critical roles in plant biology and often act as integrators of signals from multiple plant regulatory and environmental inputs. Binding plasticity allow IDPs to interact with multiple partners in protein interaction networks and provide important functional advantages in molecular recognition through transient protein-protein interactions. Short interaction-prone segments within IDPs, termed molecular recognition features, represent potential binding sites that can undergo disorder-to-order transition upon binding [3].

IDPs lack a unique 3D structure, either entirely or in parts, when alone in solution. This lack of structure in a protein can have several advantages. For instance (i) IDPs provide a larger interaction surface area than globular proteins of a similar length, (ii) conformational flexibility, and the exposure of short linear peptide motifs and interaction-prone structural motifs (Molecular Recognition Features, MoRFs) allow IDPs to scaffold and interact with numerous other proteins, and (iii) diverse post-translational modifications, such as phosphorylation and acetylation, facilitate

regulation of their function and stability in a cell. Due to their unusual structural features and important functional properties, the presence of IDPs in a cell needs to be carefully monitored [4].

Disorder-to-order transitions might be introduced by modifications of phospho-serine/-threonine mono/di/tri- methyllysine, sulfotyrosine, and 4-carboxyglutamate. Almost all proteins undergo certain chemical modifications on their side chains, called post-translational modifications (PTM) at some cellular state. Many PTMs sites have been shown to occur in disordered regions. For example, it has been reported that phosphorylation was overrepresented in disordered regions [5].

So, the concept of 'Intrinsic Disorder' in proteins has rapidly gained attention as the preponderance and functional roles of IDPs are increasingly being identified in eukaryotic proteomes. The structural plasticity allows IDPs to operate within numerous functional pathways, conferring multiple regulatory functions. Indeed, mutations in and deregulation of IDPs are associated with many diseases signifying that IDPs play vital roles in functional pathways [3]. This intrinsic disorder (ID) is highly abundant in eukaryotes, which reflect the greater need for disorder-associated signaling and transcriptional regulation in nucleated cells. It plays critical roles in plant biology and often acts as integrators of signals from multiple plant regulatory and environmental inputs.

There are several studies addressing correlations between intrinsic disorder and individual PTMs in eukaryotic proteins. Most of them include protein phosphorylation. Some researchers reported that protein phosphorylation on Ser, Thr and Tyr occurs predominantly in the disordered regions of animal proteins [5-8]. Recently, it has been confirmed for plant species also by using a predictive tool for analysis of protein phosphorylation in plants [8]. In addition, investigation on correlations between intrinsic disorder and N-glycosylation has also been done. It has been reported that the N-linked protein glycosylation prefer dissimilar structural, reflecting differences in the catalytic mechanisms of these PTMs [5,9]. Beside these, several studies addressed correlations between intrinsic disorder and the universal Lys modifications, such as acetylation and methylation. It has been reported that acetylation and methylation showed a preference for occurrence in the disordered regions of animal proteins [10]. In the present study, we analyzed the correlations between protein disorders with the main types of eukaryotic PTMs of plant Resistance (R) protein. In our study, we analyzed four major types of eukryotic protein PTMs such as Ser/Thr/Tyr- phosphorylation, Asn N-glycosylation, Lys- acetylation and Lys/Arg- methylation in several typical monocotyledonous and dicotyledonous plant species. The aim of this study was to analyze the disorder segment and post-translational modifications (PTMs) sites in R protein of plant to investigate correlation between them.

2. Materials & Methods

2.1 Data Sets

In this work, we analyzed the R protein or Resistance protein of following dicotyledonous plant species such as *Arabidopsis thaliana* (thale cress), *Citrus sinensis* (*orange*), *Solanum tuberosum* (potato), *Carica papaya* (papaya), as well as the proteins of monocotyledonous plants, such as *Oryza sativa* (rice), *Zea mays* (maize), and *Sorghum bicolor* (sorghum). The data sets of R protein were collected from the universal protein database Uniprot (http://www.uniprot.org). The final data sets were constructed using 21 proteins from each of the following: *Oryza sativa, Zea mays, Sorghum bicolor, and Carica papaya* and 20 proteins from *Arabidopsis thaliana, Citrus sinensis, Solanum tuberosum.* The total number of protein was 144.

2.2 Protein Disorder and Post-Translational Modifications (PTM) Sites Prediction

To calculate intrinsic disorder in plant proteins, the following predictive algorithms have been used in this study— RONN version 3, and MFDP2 [11,12]. The two predictive algorithms were reported to have comparable accuracy of 85–90%. RONN was used as the main predictor in this study because its scores were closer to the experimentally derived scores of intrinsic disorder for PDBdeposited protein structures [13]. However, most of the results observed using RONN have also been confirmed with the other tool.

In our study, 4 major types of eukaryotic protein PTMs which are methylation, phosphorylation, glycosylation, and acetylation have been analyzed. In specific, Ser, Thr and Tyr (S, T, Y) phosphorylation, N-linked Asn (N) glycosylation, Lys/Arg (K/R) methylation and Lys (K) acetylation have been investigated. They have been predicted using the bioinformatics tools freely available on the web. These are following

Phosphorylation sites were predicted with the Musite tool [5]. The sites of N-glycosylation were predicted with the NetNGlyc1.0 tool [14]. The sites of acetylation were predicted using another online tool PAIL, and the sites of methylation were identified with the web tool PMeS [15].

2.3 Correlation Analysis and Statistical Significance

The protein disorder degree has been correlated with the specific content of analyzed PTMs in plant proteins. The degree of correlation between the contents of various analyzed PTMs and protein disorder was evaluated by calculating Pearson correlation coefficients. The statistical significance of the Pearson correlation coefficients was determined by calculating one-tailed probability values, given the correlation value (r) and the sample size (n), with the significance level set to 0.05.

Calculations of correlation coefficients and P-values were performed using the statistics calculator IBM SPSS statistics (2014 version). Original SPSS manual which is used for statistical analysis in social science, now widely used by market researchers, health researchers, survey companies, government, education researchers, marketing organizations, data miners, and others.

We calculated the correlation coefficients and p-value between protein disorder and six different types of PTMs for every single plant. Disorder percentage was calculated separately for 144 proteins. We also determined the six types of PTMs sites for the proteins mentioned above.

2.4 PTM Content in Order and Disorder Regions

Total numbers of specific PTM sites in ordered and disordered segments of seven plant proteins have been calculated using the employed predictive algorithms. These numbers were divided by the total numbers of amino acids in the ordered or disordered segment of each proteome, providing the values of normalized PTM contents. The relative abundance of a specific PTM in the disordered and ordered segments of plant proteomes was analyzed using the following ratio: Rd/o = Nd/Ld:No/Lo, where No is the total number of PTM sites in the ordered segment of a proteome, Lo is the length of the ordered proteome segment, Nd is the total number of PTM sites in the disordered segment of a proteome and Ld is the length of the disordered proteome segment [13]. From this definition, the Rd/o value equals 1 if the relative abundances of a PTM in ordered and disordered regions are the same. It assumes a value of >1 when a PTM has a preference for occurrence in disordered regions, the more the value the more the preference. Value around 1 shows a neutral characteristics of PTM occurring. If the value is near 1 then it assumes to be in both order and disorder regions and it becomes <1 if a PTM tends to occur in ordered regions.

2.5 3D Homology Modeling of Protein

Three-dimensional structures of two proteins were built by homology modeling. First one is Acetyl-coenzyme A carboxylase carboxyl transferase from a dicotyledonous plant, S.tuberosum. The Uniprot ID of this protein is Q2VEG8 & its length is 490 amino acids. 3D structure was built by homology modeling based on the crystal structure of carboxyl transferase subunit of ACC from Staphylococcus aureus resolved at a resolution of 1.98 Å. The coordinate file of the template was retrieved from the Protein Data Bank (PDB: 2F9I chain B). The modeled range covered residues 221–471, sequence identity between the model and its template was 52% and query cover was 59%.

The next one is Cationic peroxidase from a monocotyledonous plant S.bicolor. It is 362 amino acids long & its Uniprot ID is P84516. Here, the template was crystal Structure of barley grain peroxidase from Hordeum vulgare resolved at a resolution of 1.90 Å. It was retrieved from the Protein Data Bank (PDB: 1BGP chain A). The modeled range covered residues 34–336, sequence identity between the model and its template was 78% and query cover was 85%. These sequence identities are considered to be high enough to make a reliable homology model. Homology modeling was carried out using the protein structure homology-modeling server SWISS-MODEL [16]. Structure visualization and mapping of predicted PTM sites in this model was done with PYMOL.

The generated model was validated using QMEAN analysis [17]. The QMEAN score, which ranges between 0 and 1, with higher values indicating better quality, for the first model was calculated to be 0.729, and 0.680 for the next indicating good overall quality of the generated structure.

3. Results

3.1 Comparison of Datasets and Intrinsic Disorder in Plant Proteins

A total of 144 proteins were taken from seven different plants. Then the average length of amino acid sequences in the plant proteins was determined, which varied from 324 amino acids in *S.bicolor* to 422 amino acids in *S.tuberosum*. In the following analysis, the number of predicted PTMs sites in proteins was normalized to the uniform length of 400 amino acids, rather than per sequence, considering the difference in the average protein lengths in the datasets. The following table shows the number of proteins, their total length, average length, and total length of disorder segment and the percentage of disorder region.

Plant's name	Total no. of Protein	Total length	Average length	Disorder length	Avg. disorder percentage
Monocot					
Rice	21	8473	403	2872	34%
Maize	21	8323	396	1945	24%
Sorghum	21	6795	324	958	16%
Dicot					
Arabidopsis	20	7317	366	1920	26.25%
Papaya	21	8438	402	1282	15.3%
Potato	20	8438	422	1248	15%
Orange	20	7246	363	1053	14.5%

Table 3.1: Total Number of Proteins, their total Length, Average Length, and Total Length of Disorder Segment and the Percentage of Disorder Region

A significant variation in the degree of protein disorder has been observed using RONN among the protein of the analyzed plants, ranging from 14.5% in *C.sinensis* to 34% in *O.sativa* species (Fig 3.1A). Notably, the disorder degree was significantly higher in monocots than in dicots. Considering the importance of this finding, its statistical significance was independently confirmed using the alternative disorder prediction tools MFDP2. The elevated disorder content in the rice. O.sativa, as compared with other plant species, has also been reported by other studies [6,18]. In accordance with the previous report, that the content of intrinsic disorder is generally independent of the proteome size, we calculate the pair wise correlation analysis between the disorder content and total protein length (Fig 3.1B) [18].

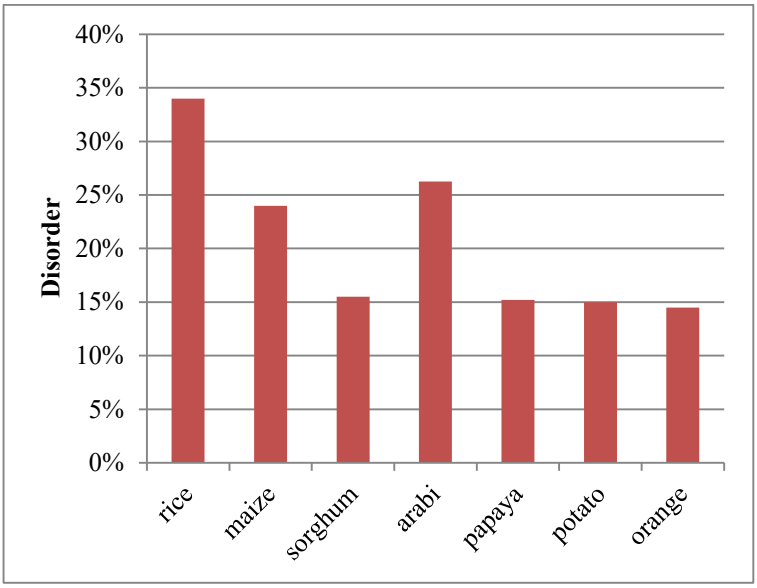


Figure 3.1A: Plant Protein Disorder Content was calculated with RONN

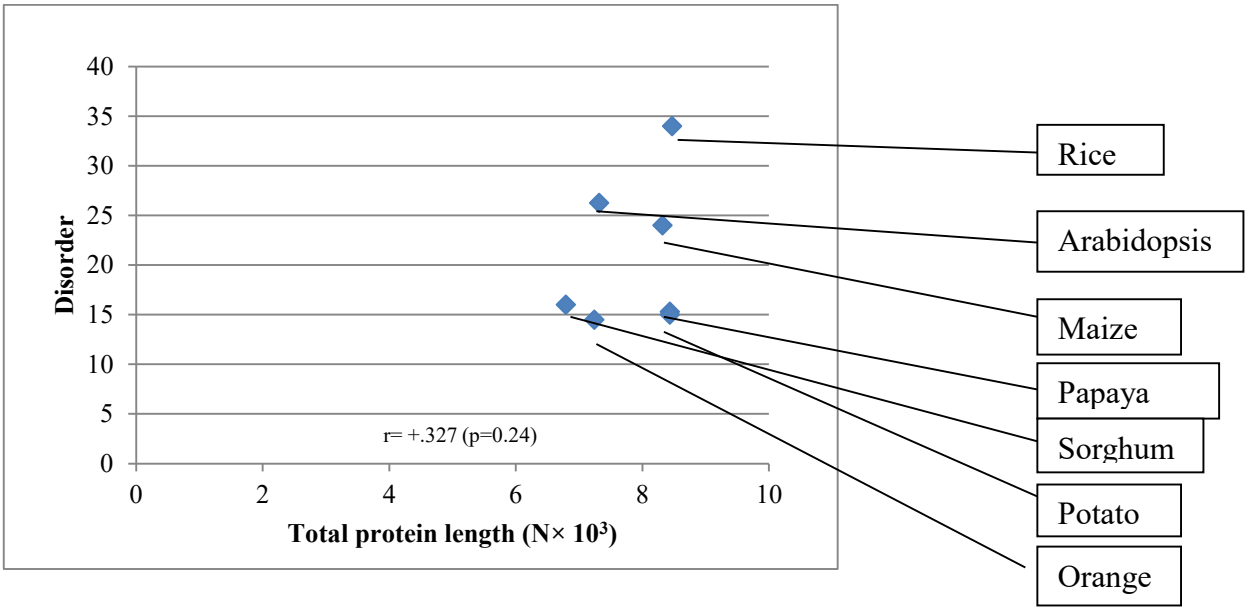


Figure 3.1B: Pair Wise Correlation Analysis between the Disorder Content and Protein Size

3.2 Correlation analysis and statistical significance

Calculations of correlation coefficients and P-values were performed using the statistics calculator SPSS which are presented in following table.

	pS,T,Y	N-gly	K-ace	R-met	K-met
Rice	**0.679** **<0.001**	**0.107** **0.322**	**0.800** **<0.001**	**0.096** **0.339**	**0.018** **0.482**
Maize	**0.667** **0.001**	**0.446** **0.021**	**0.715** **<0.001**	**0.167** **0.189**	**0.045** **0.418**
Sorghum	**0.623** **0.001**	**0.078** **0.372**	**0.737** **<0.001**	**0.202** **0.191**	**0.059** **0.400**
Arabidopsis	**0.556** **0.004**	**0.578** **0.003**	**0.585** **0.003**	**0.031** **0.449**	**0.044** **0.420**
Orange	**0.484** **0.015**	**0.599** **0.002**	**0.520** **0.009**	**0.624** **0.002**	**0.033** **0.475**
Potato	**0.380** **0.049**	**0.603** **0.002**	**0.506** **0.011**	**0.043** **0.425**	**0.009** **0.486**
Papaya	**0.525** **0.007**	**0.321** **0.084**	**0.665** **0.001**	**0.044** **0.422**	**0.037** **0.468**

Table 3.2: Statistical Significance of Correlations between Protein Disorder and Predicted Presence of PTMs. Pearson Correlation Coefficients and their Statistical Significance are Presented in the Large and Small Fonts, Accordingly for all Analyzed Correlation between Protein Disorder and PTMs

The column names stand for serine/threonine/tyrosine phosphorylation, N-linked glycosylation, lysine acetylation, arginine methylation and lysine methylation in order. The 'p' value <0.05 indicates a result that is statistically significant.

3.3 Intrinsic Disorder and Phosphorylation

Protein phosphorylation is one of the most pervasive post-translational modifications (PTM), controlling signaling pathways as Phosphorylation and dephosphorylation act as an important switch in signal transduction, metabolic, and cellular processes [16,20]. Previously, the bioinformatics tool Musite has been developed to predict phosphorylation sites based on the distinctive features of protein sequence [5]. It has been successfully applied to predict phosphorylation sites in plant proteins. An average accuracy of Musite was estimated to be 82.4% for serine, 78.6% for threonine and 89.0% for tyrosine models [4]. In our work, the superior algorithm Musite was used to predict phosphorylation sites in the analyzed plant species. It was found that the average abundance of phosphorylation sites in different species varied on a small scale except rice (O.sativa) (Fig 3.3A), suggesting common phosphorylation requirements in various plant species. Importantly, strong positive correlation was evident between protein disorder and phosphorylation. All of these correlations had a high statistical significance, as it could be judged from the one-tailed probability values of the calculated correlation coefficients (Table 3.2). Notably, the positive correlation between protein disorder and phosphorylation has also been confirmed when the alternative bioinformatics tool MFDP2 was used to predict the degree of intrinsic disorder. Normalized content of phospho-S/T/Y sites in the studied plant protein per 400 amino acids, Number of S/T/Y phosphorylation sites in ordered and disordered region, Presence and absence of S/T/ Y phosphorylation sites in disorder region have been observed here (Table 3.3; Fig 3.3 A, B, & C respectively). The number of phosphorylation sites is higher in disorder region than the order region (Fig 3.3B).

These data agree well with the previous reports [13]. Thus, they provide validation for the employed method of bioinformatics analysis.

Plant's name	Total no. of PTM sites	Order	Disorder
rice	90	6	84
maize	37	13	24
sorghum	14	4	10
arabidopsis	23	6	17
papaya	10	1	9
potato	21	4	17
orange	19	4	15

Table 3.3: Total Numbers of S/T/Y Phosphorylation Sites, Number of PTMs in Ordered and Disordered Region are given

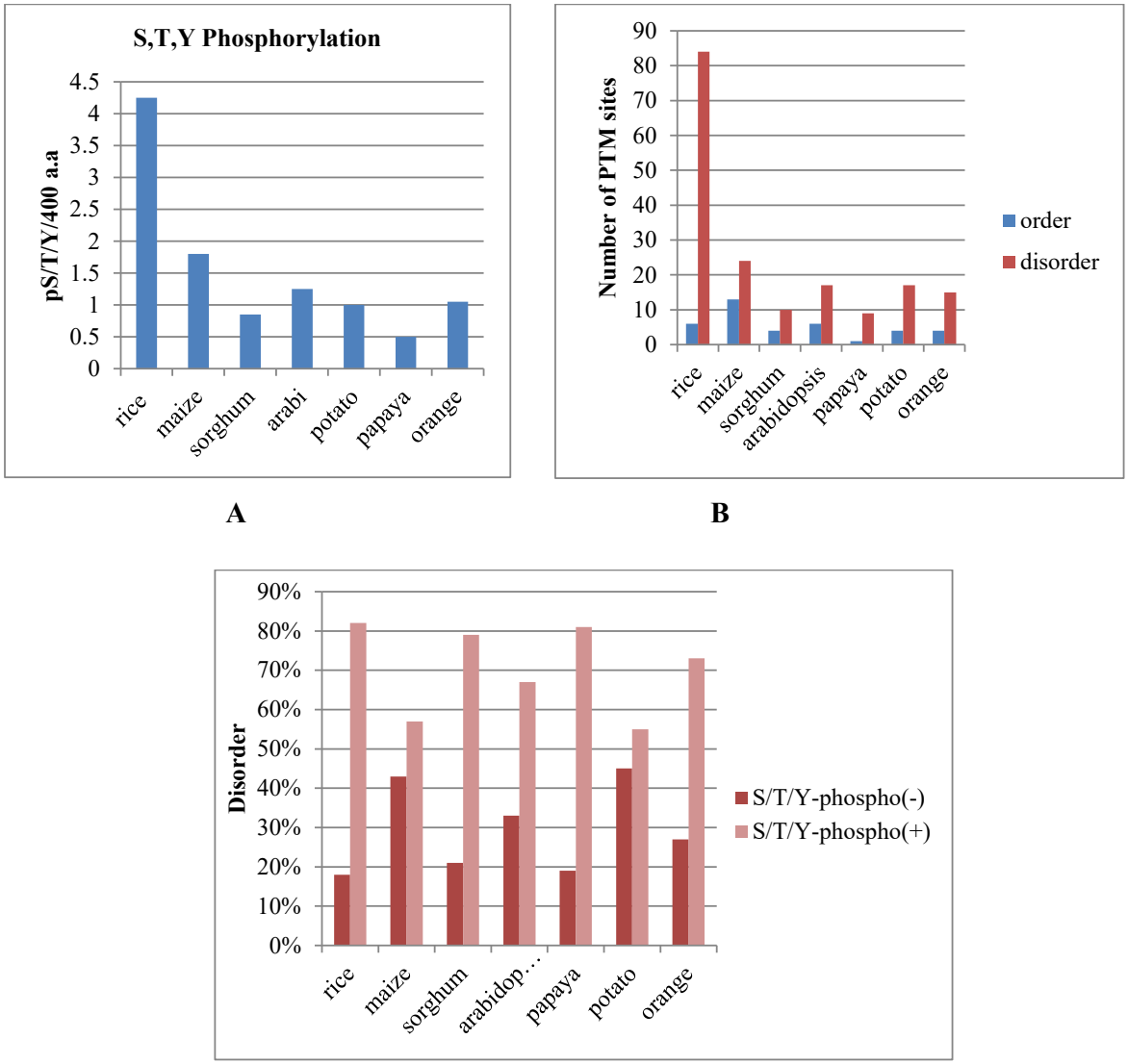


Figure 3.3: Normalized Content of Phospho-S/T/Y sites in the Studied Plant Protein Per 400 Amino Acids is Presented in (A). Number of S/T/Y Phosphorylation Sites in Ordered and Disordered Region (B) Presence (showed as '+') and Absence (showed as '-') of S/ T/ Y Phosphorylation Sites in Disorder Region (C)

3.4 Intrinsic Disorder and Acetylation

Acetylation of internal lysine is the highly reversible enzymatic reactions that changes electrostatic properties of a protein molecule by neutralizing the positive charge of lysine and arginine residues and it plays a crucial role in regulating protein function, chromatin structure, and gene expression. Here, we used the general acetylation prediction tool PAIL, freely available on the Internet, to predict the sites of lysine acetylation. The accuracies of PAIL were reported to be 85.13, 87.97 and 89.21% at low, medium and high thresholds, respectively [14]. This algorithm was trained on a set of experimentally verified acetylation sites from different eukaryotic proteins belonging to various biological species, including plants. We have found that the content of acetyllysine varied in the analyzed plant species from 6 to 17 sites per protein Fig 3.4A. A significant correlation has been observed between the predicted presence of lysine acetylation sites and protein disorder. Number of K-acetylation sites in ordered and disordered region has been observed (Table 3.4, Fig 3.4B) Percentage of K-acetylation sites presence in disordered region has showed that most of the disorder segment contains K-acetylation sites. (Fig 3.4C).

Plant's name	Total no. of acetylation sites	Order	Disorder
rice	277	171	106
maize	305	185	120
sorghum	152	116	36
Arabidopsis	303	185	118
papaya	284	227	57
potato	225	182	43
orange	171	116	55

Table 3.4: Total Numbers of K-Acetylation Sites, Number of PTMs in Ordered and Disordered Region are given

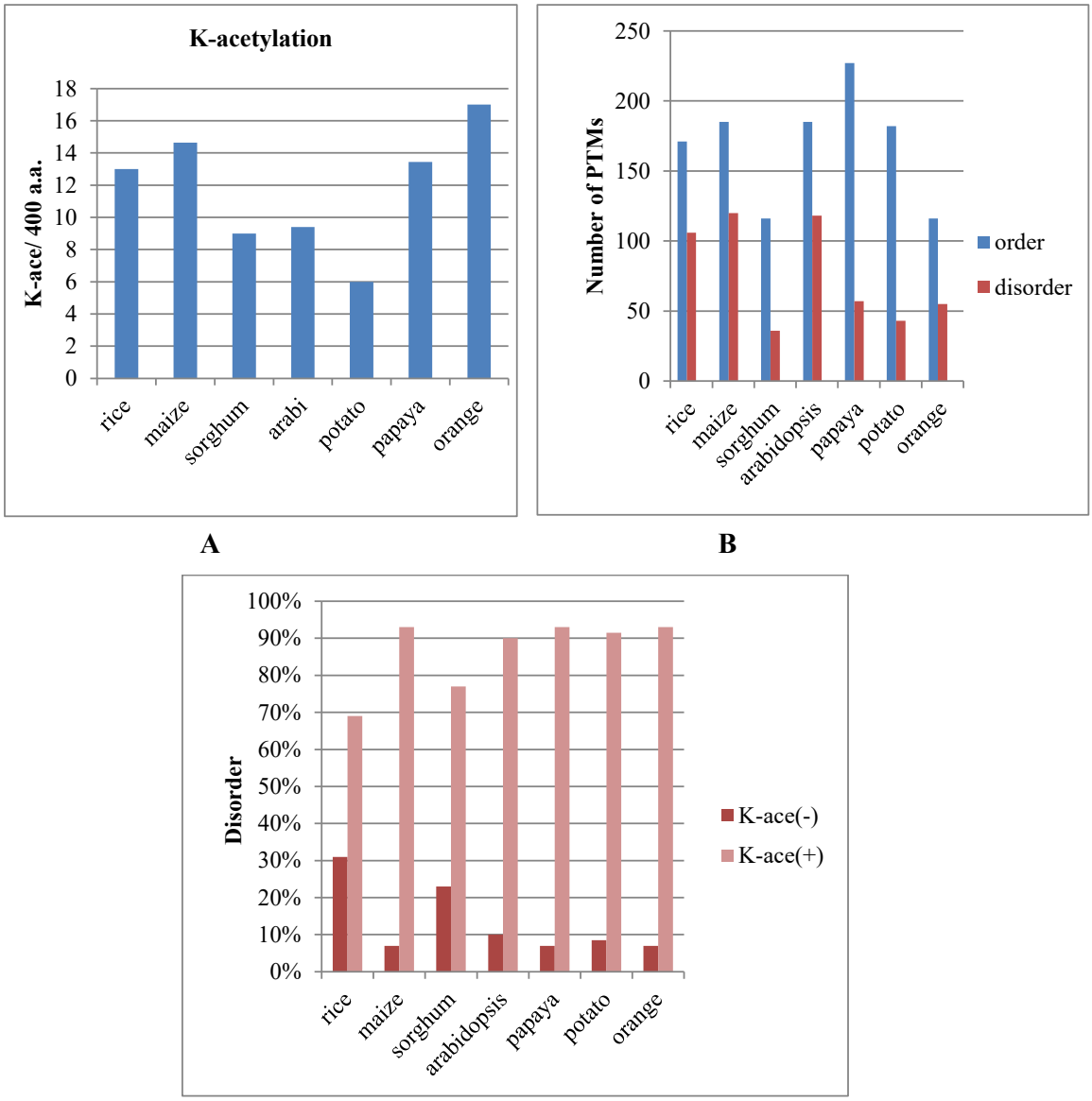


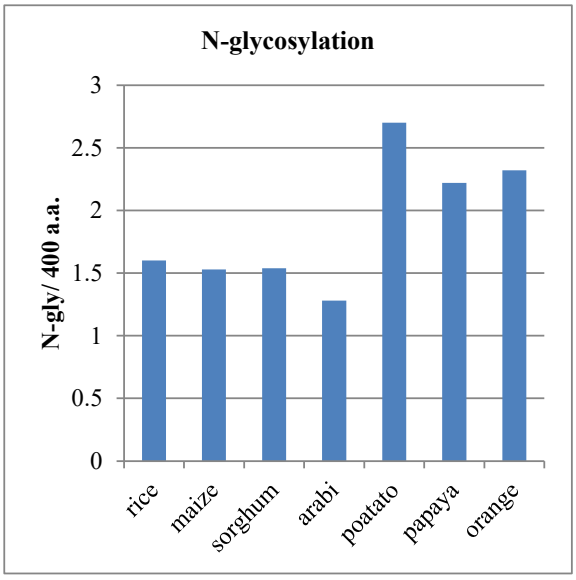
Figure 3.4: Normalized Content of K-Acetylation Sites in the Studied Plant Protein per 400 Amino Acids (A). Number of K-Acetylation Sites in Ordered and Disordered Region (B) & Presence of PTM Sites (showed as '+') in Disorder Region (C)

3.5 Intrinsic disorder and N-glycosylation

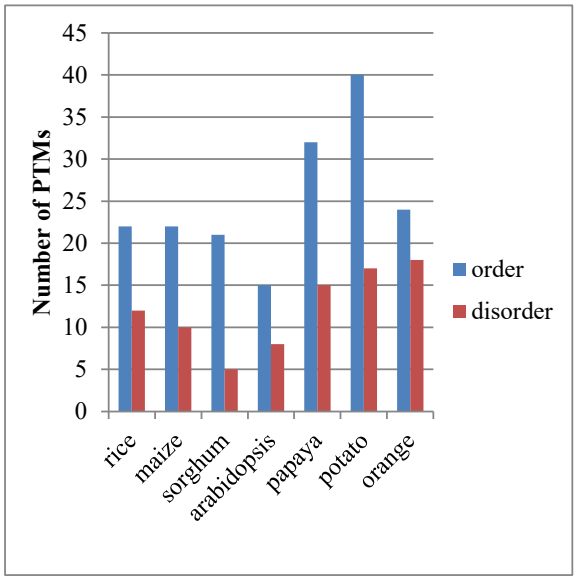
The most conserved form of protein glycosylation in eukaryotes is N-linked glycosylation, another major type of protein glycosylation in plants. Here, we used the general prediction algorithm NetNgly1.0 for predicting N-glycosylation in plant proteins. Some degree of variation in the average abundance of N-glycosylation in the studied plant species is observed here (Fig 3.6A). The content of PTM sites is higher in dicots than in monocots. Importantly, unlike Oglycosylation, no universal tendency has been observed between the predicted presence of Nglycosylation sites and protein disorder. N-glycosylation tended to correlate positively with disorder content in the dicotyledonous species, such as *A.thaliana, S.tuberosum, C.sinensis*; however, the weak correlation between protein disorder and N-glycosylation was observed in the monocotyledonous *O.sativa,* and *S.bicolor* (Table 3.2). Notably, many of the observed correlations had a low statistical significance (Table 3.2).

Plant's name	Total no. of N-glycosylation sites	Order	Disorder
rice	34	22	12
maize	32	22	10
sorghum	26	21	5
Arabidopsis	23	15	8
papaya	47	32	15
potato	57	40	17
orange	42	24	18

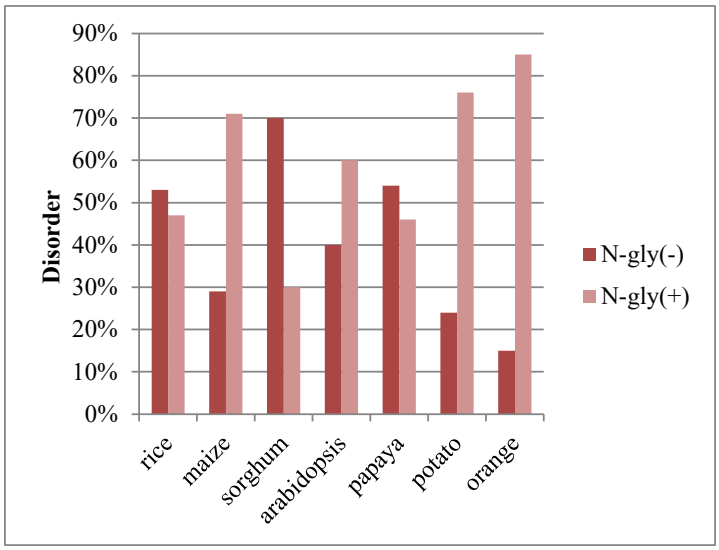
Table 3.5: Total Numbers of N-Glycosylation Sites, Number of PTMs in Ordered and Disordered Region are given



A



B



C

Figure 3.6: Normalized content of N-glycosylation Sites in the Studied Plant Protein Per 400 Amino Acids (A). Number of N-glycosylation Sites in Ordered and Disordered Region (B) Presence of Nglycosylation sites (showed as '+') in Disorder Region (C)

3.6 Intrinsic Disorder and Methylation

Protein methylation is predominantly found on lysine and arginine residues, and carries many important biological functions, including gene regulation and signal transduction. It offers great functional diversity to the primary sequence of a protein. In this study, the protein methylation tool PMeS has been used which achieved a promising performance with a sensitivity of 92.45%, a specificity of 93.18%, an accuracy of 92.82% and a Matthew's correlation coefficient of 85.69% for arginine as well as a sensitivity of 84.38%, a specificity of 93.94%, an accuracy of 89.16% and a Matthew's correlation coefficient of 78.68% for lysine in 10-fold cross validation. Compared with other existing methods, the PMeS provides better predictive performance and greater robustness.

We found that this modification showed a weak correlation with protein disorder in the analyzed plant species, suggesting that it has a preference for occurrence in ordered regions (Fig 3.6B, Fig 3.6E). Most of these correlations were statistically significant, as confirmed by the one-tailed probability values of the correlation coefficients (Table 3.2). Notably, in contrast to lysine methylation, the content of arginine methylation was found to be significantly higher in monocots than in dicots (Table 3.6A, Fig 3.6B). Most probably, this can be attributed to the relative abundance of lysine and arginine in the plant species analyzed [13].

Plant's name	Total no. of R-methylation sites	Order	Disorder
rice	110	72	38
maize	108	85	23
sorghum	95	81	14
arabidopsis	78	60	18
papaya	73	64	9
potato	87	79	8
orange	80	74	6

Table 3.6A: Total Numbers of R-Methylation Sites, Number of PTMs in Ordered and Disordered Region are given

Plant's name	Total no. of K-methylation sites	Order	Disorder
rice	11	9	2
maize	17	11	6
sorghum	11	10	1
Arabidopsis	12	10	2
papaya	21	19	2
potato	17	16	1
orange	9	8	1

Table 3.6B: Total numbers of K-methylation sites, number of PTMs in ordered and disordered region are given

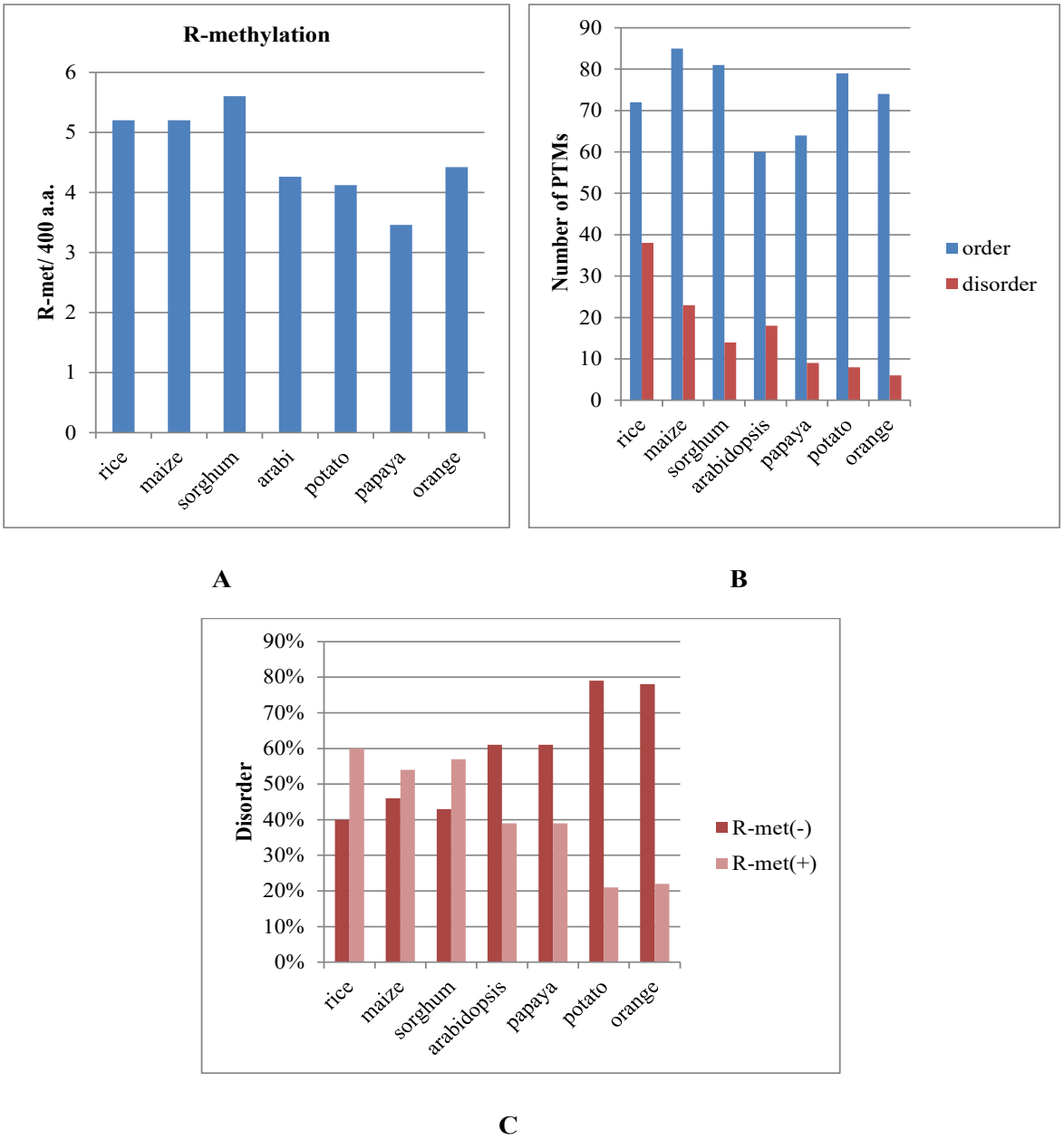


Figure 3.6: Normalized content of R-methylation sites in the studied plant protein per 400 amino acids (A). Number of R-methylation sites in ordered and disordered region (B) Presence of Rmethylation sites in disorder region (C)

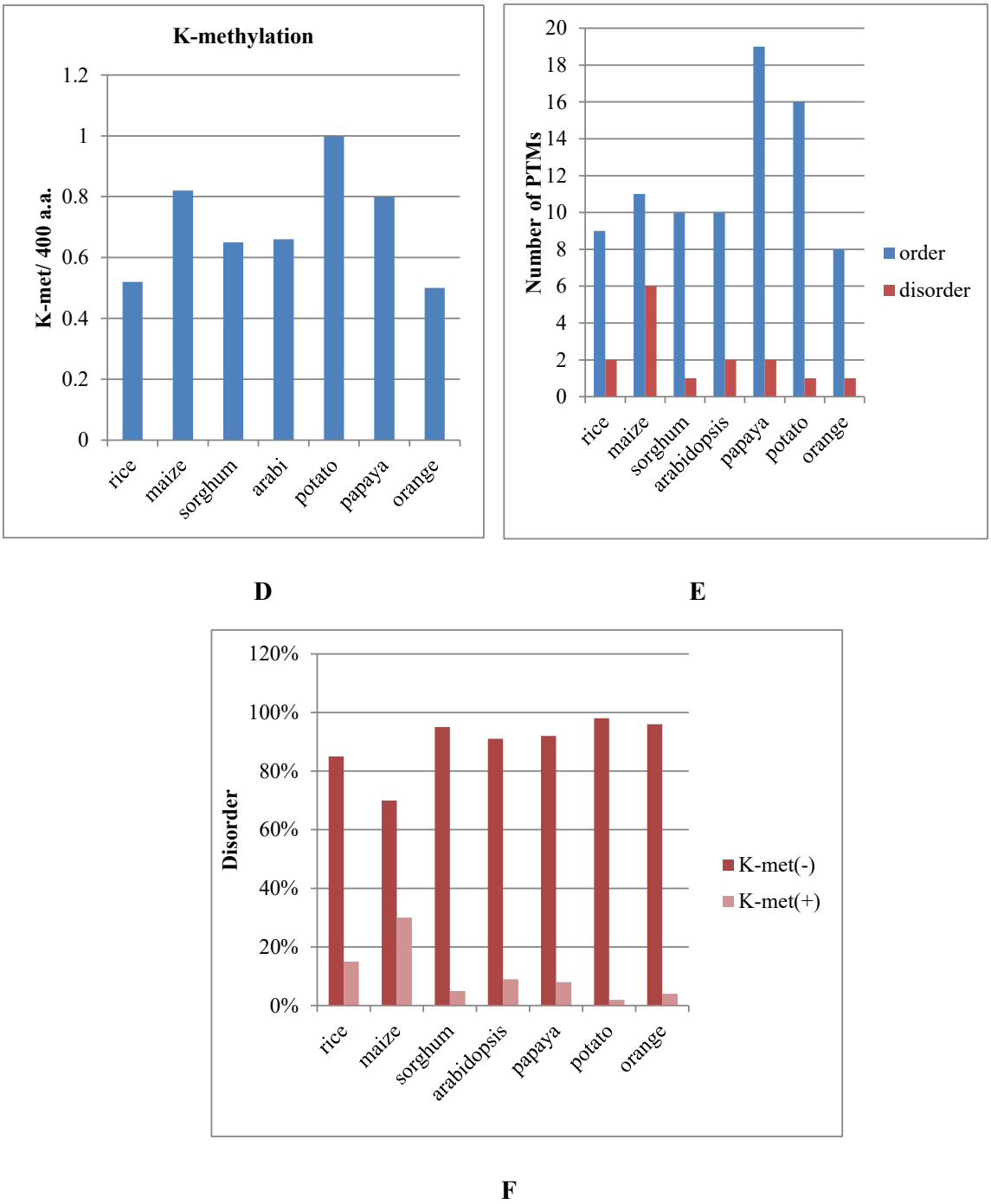


Figure 3.6: Normalized content of K-methylation sites in the studied plant protein per 400 amino acids (D). Number of K-methylation sites in ordered and disordered region (E) Presence of Kmethylation sites in disorder region (F)

3.7 Relative Contents of PTMs in Ordered and Disordered Regions of Plant Proteins

To independently verify the observed correlations, relative abundances of specific PTMs in ordered and disordered regions of plant proteins have been determined. The values of the relative abundances were presented as a ratio of normalized PTM contents (Rd/o) in disordered and ordered segments of plant proteomes. The results of these calculations are shown in Table 3.7. In general, they confirm the major findings of correlation analysis, such as the preference of phosphorylation, and acetylation for occurrence in disordered regions and the opposite tendency for methylation. Moreover, this approach allowed us to estimate the robustness of the observed correlations. For instance, high values of the Rd/o parameter obtained for phosphorylation (≥6.03) was indicative of robust relationships, whereas the correlation between protein disorder and N-glycosylation was less robust, as it could

be judged from the lower Rd/o value (≥1.06) calculated for this PTM (Table 3.7). Arginine methylation, which has a preference for occurrence in ordered protein regions, displayed more robust correlation in monocots than in dicots, as it is suggested by Rd/o values (Table 3.7). This finding is consistent with the results of correlation analysis, indicating a higher statistical significance of this correlation in monocots (Table 3.2). In, dicots, they displayed a weak correlation whereas in monocots they showed the opposite.

Name of plants	S,T,Y-phosphorylation	N- gly	K- ace	R- met	K- met
Monocot					
Rice	27.1	1.064	2.82	1.029	0.43
Maize	6.03	1.49	2.12	0.89	1.79
Sorghum	15.5	1.48	1.92	1.07	0.62
Dicot					
Arabidopsis	8.00	1.29	2.79	0.84	0.56
Orange	22.06	4.4	2.78	0.48	0.74
Potato	24.45	2.45	1.96	0.58	0.36
Papaya	22.3	2.62	2.40	0.79	0.59

Table 3.7: Relative PTMs Contents in Ordered and Disordered Regions

3.8 Mapping PTM Sites on a Protein Structure

To highlight the correlations observed, the predicted sites of PTMs analyzed in this study were mapped onto the modeled 3D structure of the plant protein Acetyl-coenzyme A carboxylase carboxyl transferase from a dicotyledonous plant, *S.tuberosum* and the Cationic peroxidase from a monocotyledonous plant *S.bicolor* were built by homology modeling based on the crystal structure of carboxyltransferase subunit of ACC from *Staphylococcus aureus* and the crystal Structure of barley grain peroxidase from Hordeum vulgare respectively. Details about model information are given in Fig 3.7A and Fig 3.7B. The PTM sites in this protein have been predicted with the bioinformatics tools used in the present study. These sequence identity are considered to be high enough to make a reliable homology model. The summary and validation of the molecular modeling are presented in. All in all, 12 PTM sites have been mapped in protein Acetyl-coenzyme A carboxylase carboxyl transferase and 17 sites in protein Cationic peroxidase. It agrees well with the correlations revealed by this study. Most notably, the sites of phosphorylation are mapped preferentially in unstructured regions, the sites of N-glycosylation and acetylation display no clear preference for either ordered or disordered sequence, and the sites of methylation are mapped mainly in the structured regions of the protein molecule. The model illustrates well the major findings of this study. Modeled 3D structure with mapped PTM sites are presented in Fig 3.7C.

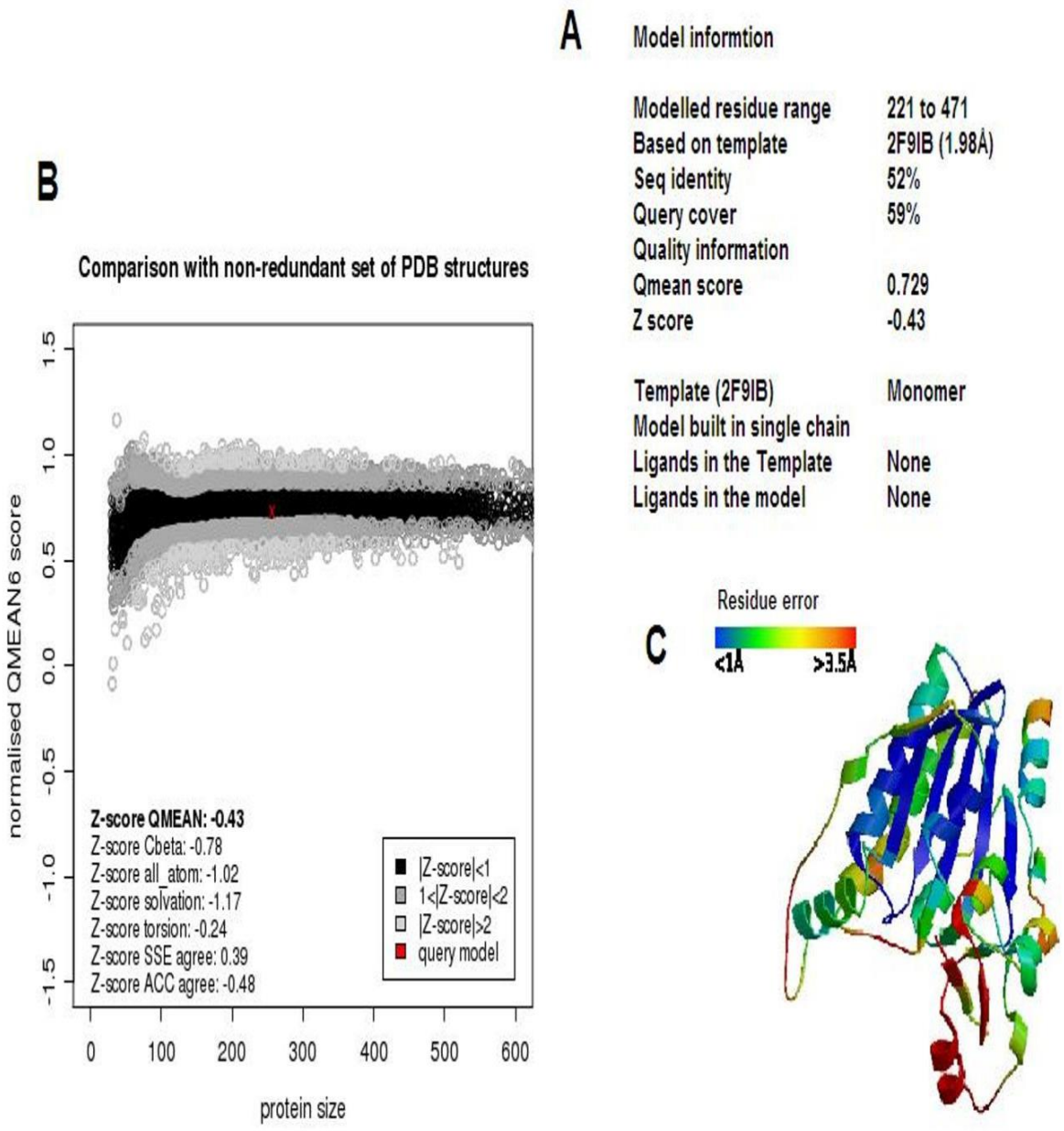


Fig 3.7A: Structural homology modeling of the plant protein Acetyl-coenzyme A carboxylase carboxyl transferase from S.tuberosum. Model summary is presented in (A) and estimation of the QMEAN score for the generated 3D model is shown in (B). Estimated residue error in the model is visualized using a color gradient in panel (C)

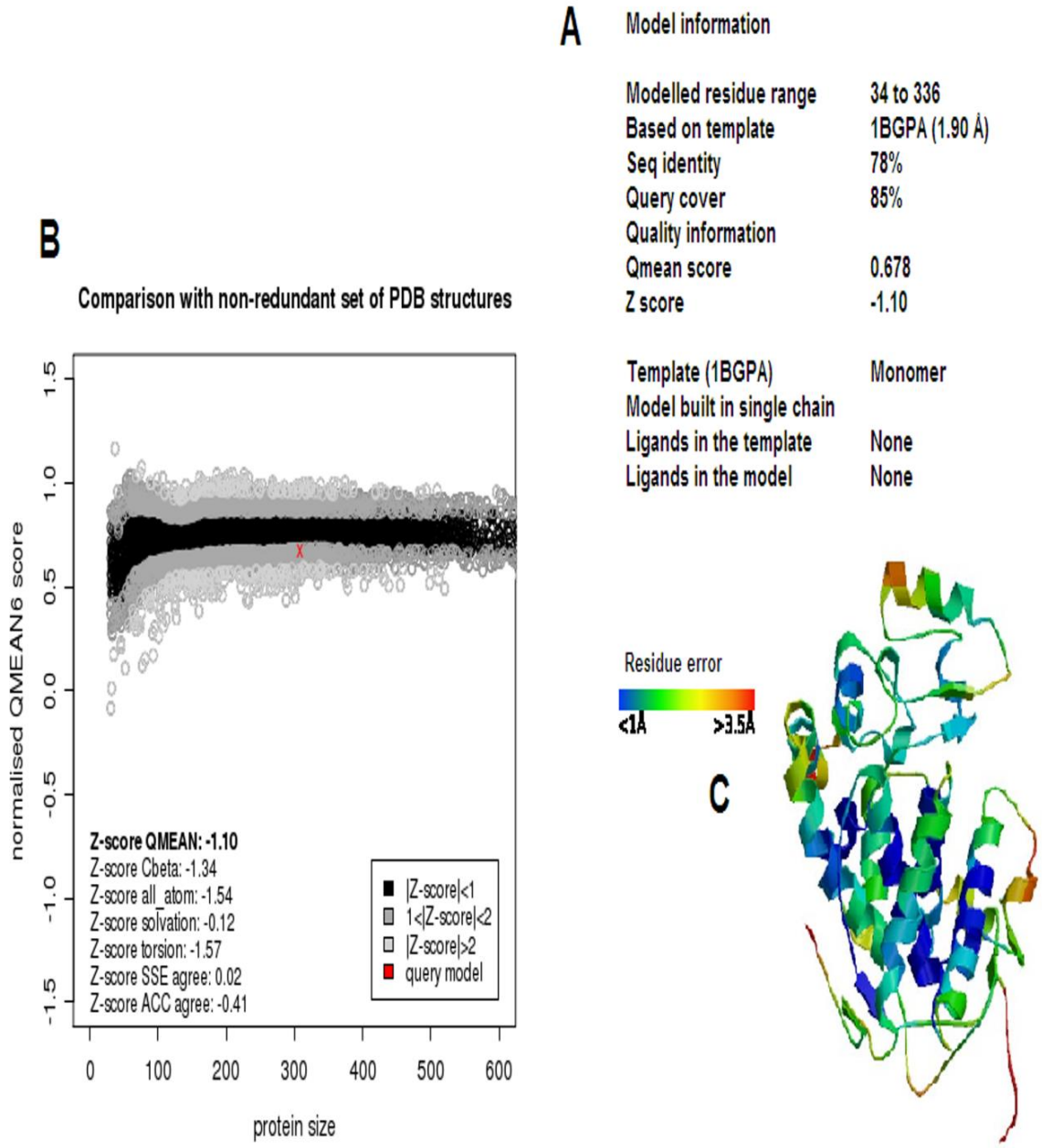
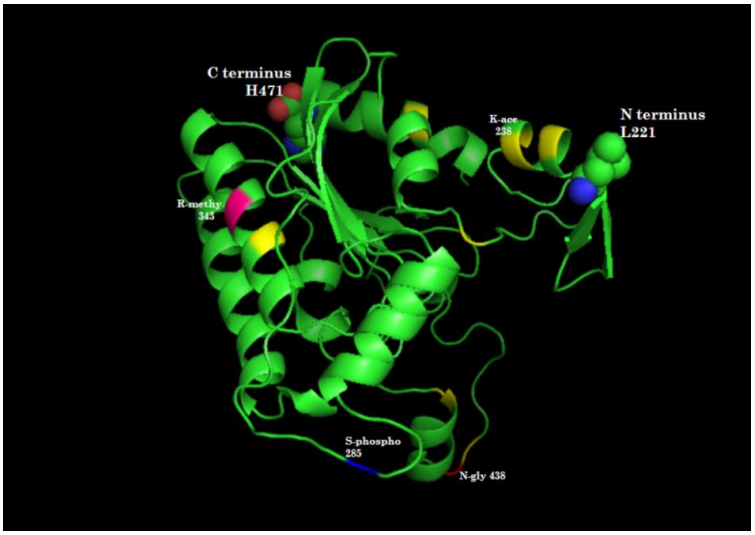
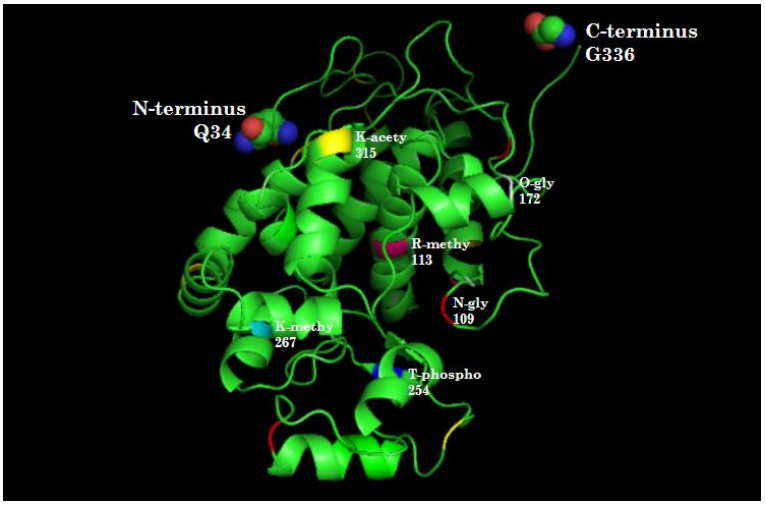


Fig 3.7B: Structural homology modeling of the plant protein Cationic peroxidase from S.bicolor. Model summary is presented in (A) and estimation of the QMEAN score for the generated 3D model is shown in (B). Estimated residue error in the model is visualized using a color gradient in panel (C).



A



B

Fig 3.7C: Modeled 3D structure of the plant protein Acetyl-coenzyme A carboxylase carboxyl transferase from S.tuberosum and the Cationic peroxidase from S.bicolor with mapped sites of analyzed PTMs are showed in Fig A & Fig B respectively. In above figures yellow color indicates the K-acetylation sites. Blue color indicates S/T/Y- phosphorylation sites & N-glycosylation sites is indicated by Red color. R-methylation, & K-methylation sites are indicated by magenta, & cyan respectively.

4. Discussion

In agreement with the previous studies [13,19,20], we observed that phosphorylation, acetylation displayed a preference for occurrence in the disordered regions of plant proteins. However, in contrast to the previous results obtained using combined datasets of eukaryotic proteins, we found, based on the set of tools used in this study, that methylation tended to occur in ordered protein regions. Also, we found that N-glycosylation did not show a clear preference for either ordered or disordered regions of plant proteins. These findings also resemble with previous studies for whole plant proteome [13]. This work examined the correlations between protein disorder and multiple PTMs in 144 R proteins from three monocotyledonous and four dicotyledonous plant species. R protein activates plant defence to restrict pathogen invasion by various defense signaling networks. These are activated via R protein post-translational modification, oligomerization, degradation, conformational changes and by the shuttling of R proteins between the plant cell cytoplasm and the nucleus. Enzyme-mediated reversible PTMs have a preference for surface accessible and disordered environments are suggested by previous studies. In accordance with this idea, we have found that disorder content correlates positively with the presence of

predicted PTM sites.

Protein phosphorylation represents an important regulatory mechanism in eukaryotic cells. At least one-third of all eukaryotic proteins are estimated to undergo reversible phosphorylation [21]. Reversible protein phosphorylation is the most extensively studied PTM because it provides a major regulatory mechanism in eukaryotic cells. Including plants, Phosphorylation was reported to be overrepresented in the disordered regions of eukaryotic proteins [5,7]. It has been experimentally showed that the regions of intrinsic disorder are associated with phosphorylation. Amino acid composition, sequence complexity, hydrophobicity, charge and other sequence attributes of regions adjacent to phosphorylation sites are similar to those of intrinsically disordered protein regions; supporting the findings that phosphorylation occurs predominantly within IDRs [7].

In our study, Ser, Thr and Tyr phosphorylation sites in these analyzed plant proteins showed positive correlations with protein disorder and had a high statistical significance. The high rate of phosphorylation occurring sites in disordered region can be easily noticed. These observations are in good agreement with the previous studies of eukaryotic proteins from animal species. For instance, data mining of human proteome revealed that phosphorylation occurs two to three times more often within disordered than ordered regions [6]. But in case of plant R protein we observed phosphorylation occurs six to twenty seven times more often within disordered than ordered using calculation of relative PTM abundance in ordered and disordered segments of plant protein. In addition, similar conclusion was obtained previously on the investigation of PDB-annotated structures of plant proteins with the disorder-assisted tool for prediction of phosphorylation sites Musite [8] that the phosphorylation frequencies among serine, threonine, and tyrosine correlate with their relative proportion in disordered regions. Thus, it can be concluded that the analytic approach applied in our study is adequate in general, considering consistency of the obtained results on phosphorylation that it tends to occur in disorder region.

On the other hand, N-glycosylation is known to occur co-translationally before a protein is fully folded. This should result in the lack of any structural preference for this modification. In accordance with this consideration, no clear structural preference has been reported for Nglycosylation in animal proteins [9]. Consistently, the correlation analysis performed in our study also failed to reveal a strong universal unidirectional relationship between N-glycosylation and protein disorder in plant proteins. N-glycosylation was found to correlate with disorder content strongly in dicotyledonous plant species than in monocotyledonous species. In both species some plant shows a low statistical significance, suggesting species-specific patterns of N-glycosylation in plants. For which, it is difficult to draw a line for correlation between Nglycosylation and disorder content. Previous studies which cover the whole proteome of plant also suggest this conclusion [13]. It should be also noted in this connection that previous studies pointed to the existence of species-, organ- and development-specific Nglycosylation patterns in plants. Thus, it can be hypothesized that the modifying enzyme, oligosaccharyltransferase, may display selective recognition for the specific modification sites in different plant species [13]. Still, at present, no experimental evidence for existence of these differences has been presented. So systematic studies on the glycosylation patterns in various plant species is an urgent need.

Notably, in contrast to correlation analysis, the alternative approach based on the calculation of relative PTM abundance in ordered and disordered segments of plant proteomes revealed that this PTM has a weak preference for disordered sequence in some plant species analyzed and good preference but not satisfactory in comparison with others value in some plant species. Presumably, this discrepancy may be attributed to the low accuracy of the NetNGlyc predictor. Its reported overall accuracy was the lowest among all the tools used, reaching only 76% in a cross-validated experiment. Considering that the reported predictive accuracy of RONN is ~85%, it can be approximated that the Rd/o values within 0.6–1.4 may not be statistically significant [11,13]. Thus, the correlations between N-glycosylation and disorder observed in monocots & in dicots is of low robustness, as it is suggested by the Rd/o values. It has been reported in previous studies that acetyl-lysine is more likely to be found within surfaceaccessible and disordered protein regions [13,22]. Among posttranslational modifications, there are some conceptual similarities between Lys-Nε-acetylation and Ser/Thr/Tyr O-phosphorylation. Alongside with protein phosphorylation, this PTM plays a major regulatory role in eukaryotic cells. The evidence has been presented that these PTMs may counteract phosphorylation, suggesting that the balance between phosphorylation and acetylation/methylation is important for physiologically relevant regulation [23]. The most recent study of correlations between PTMs and intrinsic disorder reported that acetylation did not show any significant preference for either disordered or ordered regions [8]. Now, our present work demonstrates that acetylation shows a preference for disordered regions but not as much as phosphorylation shows.

Methyl-lysine does not show any significant preference for surface accessibility or intrinsic disorder which has been reported previously [22]. On the contrary, it was found on building an SVM predictor for protein methylation, that both Arg and Lys methylation sites are likely to be intrinsically disordered But it is also reported that both Arg and Lys methylation are favored to be ordered [13,24]. The most recent study of correlations between PTMs and intrinsic disorder reported that methylation had a preference for occurrence in disordered regions [8]. But, our present work demonstrates that methylation tends to occur in ordered regions, especially K-methylation. A K-methylation site is in disorder region is hardly found. A very weak correlation has been observed between disorder and PTM sites in plants but some specific plant like orange (C.sinensis) shows a very strong correlation between Arg-methylation & disorder.

Most probably, little inconsistence can be attributed to the difference in the analyzed datasets. In the previous studies, the datasets included proteins from various eukaryotic species, or the whole proteome of a plant whereas our study was conducted only on the R protein of individual. The number of protein is not more than 144 which include 20 or 21 protein from each plant. On the other hand, those previous works covered a huge amount of protein. That is one of the important reasons of high magnitude relative PTM abundance in ordered and disordered segments of plant protein. Our study specially confined with Resistance protein of plant which are regulated by some forms of post-translational modification. Opposite correlations between the contents of protein disorder and specific PTM can be observed among the individual plant species. For instance, the weak correlation between protein disorder and N-glycosylation has been revealed in the monocotyledonous O.sativa, Z. mays and S.bicolor; however, the opposite tendency was evident in the dicotyledonous A.thaliana, S.tuberosum, C.sinensis, and C.papaya. In another case a very storng correlation was observed in C.sinensis for Arg methylation whereas all other plants showed the opposite.

Only a single study of protein disorder in the individual plant species has been investigated [13] and my findings are relevant compared with this study. Additional studies are required to confirm the observed tendencies in the particular plant species using various alternative bioinformatics and experimental approaches. Another factor that can potentially compromise the performed correlation analysis is applicability of the used prediction algorithms for analysis of plant proteins. In this connection, it should be noted that both the methylation predictor PMeS and the acetylation predictor PAIL, were trained on the databases that included plant proteins. The tools were annotated as the general PTM predictors for eukaryotic proteins. Admittedly, these algorithms have not been comprehensively applied to plants and their performance exclusively for plant proteins have not been evaluated [13]. Still, it doesn't mean that these tools systematically misrepresent the modification sites in certain plant proteins. Thus, one of the surprising finding of our study that methylation has a preference for occurrence in the ordered protein regions, reflects most probably a bona fide tendency in the investigated plant species. This finding was confirmed independently by both correlation analysis and by calculation of methylation content in the ordered and disordered segments of plant proteomes. Moreover, the observed tendency concerned both Arg and Lys methylation sites, in this connection, the PMeS prediction algorithm achieved the accuracies of 92.82 and 89.16% for arginine and lysine, respectively [15]. Notably, Arg methylation was found to be more abundant than Lys methylation; the two PTMs are known to be catalyzed by different enzymes. In light of our findings similar with the previous one it is tempting to surmise that both Arg- and Lys-specific plant methylases may recognize some structural determinants of modification sites, directing their specificity toward the substrate residues located in the ordered regions of plant proteins.

5. Conclusion

Our study reveals that some PTMs, display a clear preference for occurrence in disordered regions of plant proteins like phosphorylation, and acetylation. However, the opposite tendency is evident for methylation, and N-glycosylation does not display a universal preference for either ordered or disordered protein regions and preference to be ordered in case of lysine methylation. Many protein modifications can work together in regulation and signaling, so we expect that elaborate study concerning the relationships between R protein as well as other protein disorder and multiple posttranslational modifications in future. In our study, we only analyzed 4 types of PTMs. Besides these, the degree of intrinsic disorder needs to be correlated with other major types of PTMs like sulfation, ubiquitination, sumoylation, ribosylation etc. which will open a new era of plant structural biology. This investigation also revealed the marked differences between the integral characteristics of monocot and dicot proteomes including increased rate of disorder percentage, increased rate of and R-methylation, decreased rate of N-glycosylation and K-methylation in monocotyledonous plant species, as compared with dicotyledonous species. The evolutionary and environmental background behind this variation needs to be determined. However, these findings provided useful insight into the mechanisms of various PTMs occurring and may facilitate further investigations into the structural and functional implications of these PTMs. Further Studies on the other group of proteins are needed for better understanding the relation between proteins PTMs and IDR of plants.

References

1. Belkhadir, Y., Subramaniam, R., & Dangl, J. L. (2004). Plant disease resistance protein signaling: NBS–LRR proteins and their partners. *Current opinion in plant biology, 7*(4), 391-399.
2. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of molecular biology, 337*(3), 635-645.
3. Sun, X., Rikkerink, E. H., Jones, W. T., & Uversky, V. N. (2013). Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *The Plant Cell, 25*(1), 38-55.
4. Babu, M. M., van der Lee, R., de Groot, N. S., & Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. *Current opinion in structural biology, 21*(3), 432-440.
5. Gao, J., Thelen, J. J., Dunker, A. K., & Xu, D. (2010). Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. *Molecular & Cellular Proteomics, 9*(12), 2586-2600.
6. Fukuchi, S., Hosoda, K., Homma, K., Gojobori, T., & Nishikawa, K. (2011). Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC structural biology, 11*, 1-10.
7. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation.

Nucleic acids research, 32(3), 1037-1049.
8. Yao, Q., Gao, J., Bollinger, C., Thelen, J. J., & Xu, D. (2012). Predicting and analyzing protein phosphorylation sites in plants using musite. *Frontiers in plant science, 3*, 186.
9. Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A., & Wormald, M. R. (2004). Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology, 14*(2), 103-114.
10. Hansen, J. C., Lu, X., Ross, E. D., & Woody, R. W. (2006). Intrinsic protein disorder, amino acid composition, and histone terminal domains. *Journal of Biological Chemistry, 281*(4), 1853-1856.
11. Yang, Z. R., Thomson, R., McNeil, P., & Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics, 21*(16), 3369-3376.
12. Mizianty, M. J., Peng, Z., & Kurgan, L. (2013). MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically disordered proteins, 1*(1), e24428.
13. Kurotani, A., Tokmakov, A. A., Kuroda, Y., Fukami, Y., Shinozaki, K., & Sakurai, T. (2014). Correlations between predicted protein disorder and post-translational modifications in plants. *Bioinformatics, 30*(8), 1095-1103.
14. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry, 45*(22), 6873-6888.
15. Shi, S. P., Qiu, J. D., Sun, X. Y., Suo, S. B., Huang, S. Y., & Liang, R. P. (2012). PMeS: prediction of methylation sites based on enhanced feature encoding scheme. *PloS one, 7*(6), e38772.
16. Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics, 22*(2), 195-201.
17. Benkert, P., Künzli, M., & Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic acids research, 37*(suppl_2), W510-W514.
18. Gao, J., & Xu, D. (2012). Correlation between posttranslational modification and intrinsic disorder in protein. In *Biocomputing 2012* (pp. 94-103).
19. He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., & Dunker, A. K. (2009). Predicting intrinsic disorder in proteins: an overview. *Cell research, 19*(8), 929-949.
20. Uversky, V. N., & Dunker, A. K. (2008). Controlled chaos. *Science, 322*(5906), 1340-1341.
21. Marks, F. (1996). *Protein Phosphorylation*. VCH Weinhem, NewYork, Tokyo.
22. Pang, C. N. I., Hayen, A., & Wilkins, M. R. (2007). Surface accessibility of protein post-translational modifications. *Journal of proteome research, 6*(5), 1833-1845.
23. Krämer, O. H., & Heinzel, T. (2010). Phosphorylation–acetylation switch in the regulation of STAT1 signaling. *Molecular and cellular endocrinology, 315*(1-2), 40-48.
24. Daily, K. M., Radivojac, P., & Dunker, A. K. (2005, November). Intrinsic disorder and prote in modifications: building an SVM predictor for methylation. In *2005 IEEE symposium on computational intelligence in bioinformatics and computational biology* (pp. 1-7). IEEE.
25. Benkert, P., Tosatto, S. C., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics, 71*(1), 261-277.
26. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature protocols, 4*(1), 1-13.
27. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochemistry, 41*(21), 6573-6582.
28. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., ... & Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of molecular graphics and modelling, 19*(1), 26-59.
29. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of molecular biology, 323*(3), 573-584.
30. Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., & Dunker, A. K. (2005). Comparing and combining predictors of mostly disordered proteins. *Biochemistry, 44*(6), 1989-2000.
31. R gene, Wikipedia, modified in 2009, available at www.wikipedia.com
32. SPSS, Wikipedia, modified in 2014, available atwww.wikipedia.com
33. Uversky, V. N. (2011). Intrinsically disordered proteins from A to Z. *The international journal of biochemistry & cell biology, 43*(8), 1090-1103.
34. Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions?. *Proteins: structure, function, and bioinformatics, 41*(3), 415-427.
35. Xing, W., Zou, Y., Liu, Q., Liu, J., Luo, X., Huang, Q., ... & Chai, J. (2007). The structural basis for activation of plant immunity by bacterial effector protein AvrPto. *Nature, 449*(7159), 243-247.