

RESEARCH

Open Access



# In silico structural homology modelling of EST073 motif coding protein of tea *Camellia sinensis* (L)

K. H. T. Karunaratna<sup>1,2\*</sup> , N. H. K. S. Senathilake<sup>1</sup>, K. M. Mewan<sup>3</sup>, O. V. D. S. J. Weerasena<sup>1</sup> and S. A. C. N. Perera<sup>4</sup>

## Abstract

**Background:** Tea (*Camellia sinensis* (L.) O. Kuntze) is known as the oldest, mild stimulating caffeine containing non-alcoholic beverage. One of the major threats in south Asian tea industry is the blister blight leaf disease (BB), caused by the fungus *Exobasidium vexans* Masse. SSR DNA marker EST SSR 073 is used as a molecular marker to tag blister blight disease resistance trait of tea. The amino acid sequences were derived from cDNA sequences related to EST SSR 073 of BB susceptible (TRI 2023) and BB resistant (TRI 2043) cultivars. An attempt has been made to understand the structural characteristics and variations of EST SSR 073 locus that may reveal the factors influencing the BB resistance of tea with multiple bioinformatics tools such as ORF finder, ExPasy ProtParam tools, modeler V 9.17, Rampage server, UCSF-Chimera, and HADDOCK docking server.

**Results:** The primary, secondary, and tertiary structures of EST SSR 073 coding protein were analyzed using the amino acid sequences of both BB resistant TRI 2043 and BB susceptible TRI 2023 tea cultivars. The coding amino acid sequences of both the cultivars were homologous to photosystem I subunit protein (PsaD I) of *Pisum sativum*. The predicted 3D structures of proteins were validated and considered as an acceptable overall stereochemical quality. The BB resistant protein showed CT repeat extension and did not involve in topology of the PsaD I subunit. The C terminal truncation of BB resistance caused the formation of hydrogen bonds interacting with PsaD I and other subunits of photosystem I in the modeled three-dimensional protein structure.

**Conclusions:** *Camellia sinensis* EST 073 SSR motif coding protein was identified as the PsaD I subunit of photosystem I. The exact mechanism of PsaD I conferring the resistance for blister blight in tea needs to be further investigated.

**Keywords:** Blister blight disease, EST-SSR, Homology modeling, Molecular markers, Tea

## Background

Tea, *Camellia sinensis* (L.) O. Kuntze, is the second most popular, healthy non-alcoholic beverage in the world. It is an economically important tree crop, grown in several countries in Asia and Africa. Globally, Sri Lanka is the third-largest producer and the 2<sup>nd</sup> largest

exporter of tea [1] with its popular brand “Ceylon Tea”, playing a key role in the international tea trade.

Blister blight leaf disease (BB), caused by the obligatory fungal pathogen, *Exobasidium vexans* Masse (Basidiomycetes) is one of the most devastating biotic constraints, commonly found in a majority of tea plantations in south Asia including Sri Lanka, India, Indonesia, Bangladesh, Thailand, Nepal, Vietnam, Cambodia, and Japan [2]. The BB leaf disease causes approximately 25 to 30% crop loss annually depending on the agro-ecological region (AER) of Sri Lanka [3]. The disease infection also causes a reduction of the quality of black tea by changing the composition of leaf

\* Correspondence: [thissa1234@gmail.com](mailto:thissa1234@gmail.com)

<sup>1</sup>Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Colombo, Sri Lanka

<sup>2</sup>Current address: Department of biosystems Technology, Faculty of Technology, University of Ruhuna, Matara, Sri Lanka

Full list of author information is available at the end of the article



biochemical components such as polyphenols, catechins, and enzymes which highly influence the quality of black tea [4].

Presently, the control of this disease is solely based on chemical means, where spraying of Cu-based fungicides directly on to the foliage, before infection, being the recommended practice. The disease is very common in major tea growing areas of Sri Lanka throughout the year, and the repeated application of Cu-based fungicides may lead to chemical residues in the end product “Black tea.” Though tea, is a popular healthy beverage, exceeding maximum residual levels (MRLs) of pesticides, heavy metals, and other chemical impurities, leads to a non-tariff trade barrier in exporting and consumption of tea [5]. Therefore, to overcome the said constraints and also to maintain the quality of the symbol “Ceylon Tea”, the development of resistant cultivars to BB disease would be the most effective and sustainable approach to control the disease.

Tea is a perennial crop, which requires 20–25 years to develop a new improved cultivar and therefore, the application of marker-assisted selection (MAS) techniques would be highly desirable to increase the efficiency and effectiveness of the breeding program. Bulk segregant analysis (BSA) approach has successfully been applied to identify a SSR DNA marker EST SSR 073 to tag blister blight disease resistance trait using a segregating population derived from the two parents: TRI 2043 (resistant cultivar) × TRI 2023 (susceptible cultivar) [6]. EST SSR 073 motif correlates with the photosystem I subunit D (PsaD I) and identifying the structural model of a protein of the motif is one of the key points for understanding the underlying biological mechanism at a molecular level. The available knowledge on the structure and the role of PsaD I protein is scarce. The experimental elucidation of the tertiary structure of a protein is a huge and a difficult endeavor [7]. The X-ray crystallography or nuclear magnetic resonance techniques (NMR), which are applied to identify the tertiary structure of a protein, are time consuming and expensive [8, 9]. However, the “*In silico* homology modeling” provides an alternative application to predict the 3D structure of proteins with better validation. Homology modeling is known to be one of the best and extensively used computational methods to generate three-dimensional structures when there is more than 35% sequence identity between the known protein structure (template) and the unknown protein structure [10–13].

*In silico* homology modeling has been successfully applied to predict the structure of Matrix metalloproteinase 25 (MMP 25) and it can be used as a target for the inhibition of airway remodeling in asthma disease by using *in silico* drug designing methods [14]. Furthermore, an acceptable protein structure of nif A which is involved in nitrogen fixation of rhizobial strains, has

been identified and validated by using *in silico* structure homology modeling [15]. *In silico* characterization of ChiLCV coat proteins of *Begomovirus* in chilli aided in the development of strategies to control *Begomovirus* disease of crops [16]. Vascular wilt disease of tomato caused by *Fusarium oxysporum* f. sp. *lycopersici* is controlled by targeting a novel candidate protein FOXG\_04696 which has been developed by homology modeling [17].

With the above background, molecular modeling of EST SSR 073 motif coding protein was the objective of the current study to provide a topology for revealing protein folding and functional structure which would help in understanding the blister blight fungal infection for combating the disease.

## Methods

### DNA sequence of EST SSR 073 motif

The EST SSR 073 motif containing cDNA sequence of blister blight disease resistant tea cultivar TRI 2043 (BBR) (GenBank accession no: MT303817) [18] and the DNA sequence of EST SSR 073 motif of blister blight disease susceptible tea cultivar TRI 2023 (BBS) (GenBank accession no: MT303818) [6] were retrieved. The sequences of BBR and BBS were aligned with BLASTn program [19].

### Amino acid sequence analysis and template retrieval

All possible open reading frames (ORFs) for both the nucleotide sequences were identified by ORF finder (NCBI) [20]. Amino acid sequences derived by conceptual translation of each of the ORFs were used as the query for searching homologous sequences using BLASTp [21] against uniprotKB/swissprot database to identify potential orthologs [22]. The search was repeated against Protein Data Bank (PDB) and the amino acid sequence which contained the putative conserved domain and showed the highest sequence similarity and the lowest *E* value, was selected for structure modeling.

### Homology modeling and energy minimization

PsaD subunits have been reported to possess N (1Met to GLY 90) and C (Asp171 to Gly 193) terminal unstructured domains which are involved in the assembly of photosystem I super complex [23, 24]. Accordingly, the three-dimensional (3D) structure of the identified proteins were built using modeller V 9.17 [25] using the crystal structure of PsaD subunit of *Pisum sativum* photosystem I super-complex (PDB ID: 5l8r\_D) [26] as the template and viewed by UCSF Chimera [27]. The generated model of *C. sinensis* PsaD-like protein was superimposed on the PsaD subunit of 5l8r in energy minimized state while keeping the rest of the complex fixed. Superimposition was carried out using Matchmaker function of UCSF-Chimera [28, 29]. Energy minimization

Score	Expect	Identities	Gaps	Strand
1149 bits(622)	0.0	675/696(97%)	21/696(3%)	Plus/Plus
Query 1	AGTCGACGGTACCGACATATGCCCGGAATTCGGCCATTACGGCCGGGGAAAAACCACC	60		
Sbjct 1	AGTCGACGGTACCGACATATGCCCGGAATTCGGCCATTACGGCCGGGGAAAAACCACC	60		
Query 61	CACCCATTATCCCTAACAACTCctctctctctc-----tctcCAT	100		
Sbjct 61	CACCCATTATCCCTAACAACTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT	120		
Query 101	GGCCATGGCAACCAAGCTTCCCTATTACCCCAACACTGTCCACCCCAAACTCCGGCGA	160		
Sbjct 121	GGCCATGGCAACCAAGCTTCCCTATTACCCCAACACTGTCCACCCCAAACTCCGGCGA	180		
Query 161	CCGCCTCATCTCATCGTGGAAAGCAACCGTCAACTTTGTCTTTACCAAGTGCAGTCTGC	220		
Sbjct 181	CCGCCTCATCTCATCGTGGAAAGCAACCGTCAACTTTGTCTTTACCAAGTGCAGTCTGC	240		
Query 221	CAGATCCAGCATGCCATCGCGGACTATTAAGGCCGCGGCGCTGAGGGAAAGGCAGAAAGC	280		
Sbjct 241	CAGATCCAGCATGCCATCGCGGACTATTAAGGCCGCGGCGCTGAGGGAAAGGCAGAAAGC	300		
Query 281	ACCGGTGAAGGAGGCTCCTGTGGGCTTACCCCAACGCAATTGGACCGGACCCCAATC	340		
Sbjct 301	ACCGGTGAAGGAGGCTCCTGTGGGCTTACCCCAACGCAATTGGACCGGACCCCAATC	360		
Query 341	ACCAATTTTGGTGGCAGCACTGGGGCTGCTTAGAAAAGCCCAAGTGAAGAGTTTTA	400		
Sbjct 361	ACCAATTTTGGTGGCAGCACTGGGGCTGCTTAGAAAAGCCCAAGTGAAGAGTTTTA	420		
Query 401	TGTGATCACTTGGGAGTCCCAGGAACGAATATTGAAATGCCACTGGTGGTGCAGC	460		
Sbjct 421	TGTGATCACTTGGGAGTCCCAGGAACGAATATTGAAATGCCACTGGTGGTGCAGC	480		
Query 461	CATAATGAGAGAAGGCCCTAACCTACTCAAGTTGGCTAGGAAAGAACAGTCTTGCACT	520		
Sbjct 481	CATAATGAGAGAAGGCCCTAACCTACTCAAGTTGGCTAGGAAAGAACAGTCTTGCACT	540		
Query 521	TGGGACTAGACTTAGGTCCAAATATAAGATTAAAGTACCAATTTTATAGGGTGTCCCGAA	580		
Sbjct 541	TGGGACTAGA-TTAGGTCCAAATATAAGATTAAAGTACCAATTTTATAGGGTGTCCCGAA	599		
Query 581	TGGGGAGGTACAATATTTGCACCCTAAGGATGGAGTGTACCCGGAGAGGTGAACCCGGG	640		
Sbjct 600	TGGGGAGGTACAATATTTGCACCCTAAGGATGGAGTGTACCCGGAGAGGTGAACCCGGG	659		
Query 641	TCGACAAGGTGTCTGGGCTGAATATGCGGTGCGATTGG	676		
Sbjct 660	TCGACAAGGTGTCTGGGCTGAATATGCGGTGCGATTGG	695		

**Fig. 1** Global pairwise alignment of EST 073 loci related to BB resistant and BB susceptible nucleotide sequences. Query denotes the BBS sequence and sbjct denotes BBR sequences.

was carried out using AMBER force field [30–32] in chimera with 100 steps of steepest descents followed by 10 steps of conjugate gradients.

**Protein model validation**

The quality of generated models was validated with respect to backbone and side chain geometry. To validate protein backbone quality, Ramachandran plot [33] was generated using Rampage server (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) and the backbone quality was validated by analyzing  $\phi$  and  $\psi$  angles using Ramachandran plot. Further; VERIFY3D, ERRAT,

PROVE, PROCHECK, AND WHATCHECK [34] servers were used to analyze the overall quality of the model.

**Structural comparison of modeled proteins**

Optimized energy minimized protein models generated for the sequence derived from BB susceptible TRI 2023 and the sequence derived from BB resistant TRI 2043 were superimposed using Matchmaker function of UCSF Chimera and RMSD (root mean square deviation) value was obtained. Further, structural comparison was carried out by superimposing and RMSD evaluation against the template protein that was used to generate the protein models.

Identity:	95.2 %
Query: 69	GFTPPQLDPSTPSPIFGGSTGGLLRKAQVEEFYVITWESPKERIFEMPTGGAAIMREGPN 128
	GFTPP+LDP+TPSPIFGGSTGGLLRKAQVEEFYVITWESPE+IFEMPTGGAAIMREGPN
Templ: 1	GFTPPELDPNTPSPIFGGSTGGLLRKAQVEEFYVITWESPKQIFEMPTGGAAIMREGPN 60
Query: 129	LLKLARKEQCLALGTRLRSKYKIKYQFYRVFPNGEVQYLHPKDGVPYKVNPNPGRQGVGLN 188
	LLKLARKEQCLALGTRLRSKYKIKYQFYRVFP+GEVQYLHPKDGVPYKVNPNPGRQGVG+N
Templ: 61	LLKLARKEQCLALGTRLRSKYKIKYQFYRVFPNGEVQYLHPKDGVPYKVNPNPGRQGVGN 120
Query: 189	MRSIG 193
	RSIG
Templ: 121	FRSIG 125

**Fig. 2** Protein sequence alignment of BBS and *P. sativum*

```

Query: 69  GFTPPQLDPSTPSPIFGGSTGGLLRKAQVEEFYVITWESPKERIFEMPTGGAAIMREGPN 128
          GFTPP+LDP+TPSPIFGGSTGGLLRKAQVEEFYVITWESPK+IFEMPTGGAAIMREGPN
Templ: 1   GFTPPELDPNTPSPIFGGSTGGLLRKAQVEEFYVITWESPKQIFEMPTGGAAIMREGPN 60

Query: 129 LLKLLARKEQCLALGTRLGPNIRLSTNF 155
          LLKLLARKEQCLALGTRL  ++  F
Templ: 61 LLKLLARKEQCLALGTRLRSKYKIKYQF 87
  
```

**Fig. 3** Protein sequence alignment of BBR and *P. sativum*

### Validation of physiological parameters

Structure-function relationship of the derived protein models was further validated using ProtParam tool of ExPASy Proteomics Server [35] for various parameters such as estimated half-life, theoretical pI, instability index, aliphatic index, and grand average of hydropathicity (GRAVY). The values were compared with the protein sequence used as the template as well as the BB susceptible and resistant genotypes.

### Molecular docking

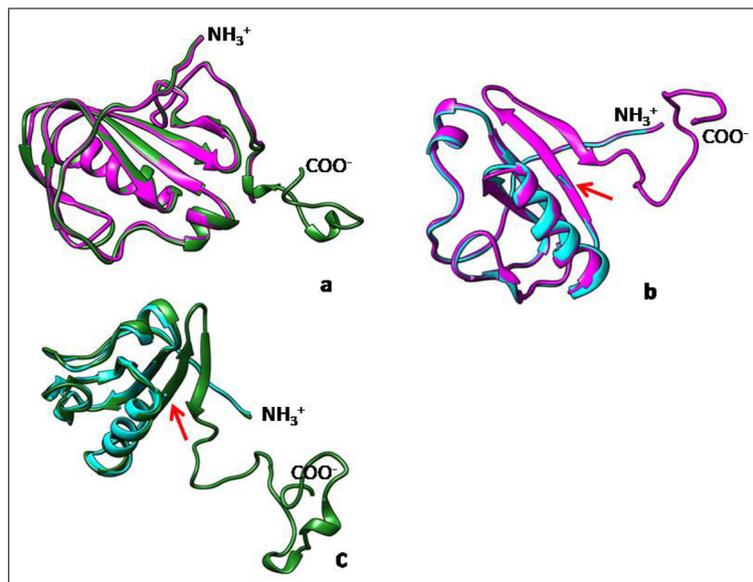
Crystal structure of PSI-I complex of *Pisum sativum* (PDB ID: 5l8r) was retrieved from RCSB-PDB. Subunits that interact with PsaD were retained and other subunits were removed using UCSF-Chimera. Modeled PsaD was docked against the binding site of the complex by HADDOCK docking server [36]. Docking results were viewed using UCSF Chimera. Default parameters were used for docking process and energy (E) values of each docking event were obtained. For comparative analysis, docked complexes were compared with the interactions of the PSI complex of *Pisum sativum*.

## Results

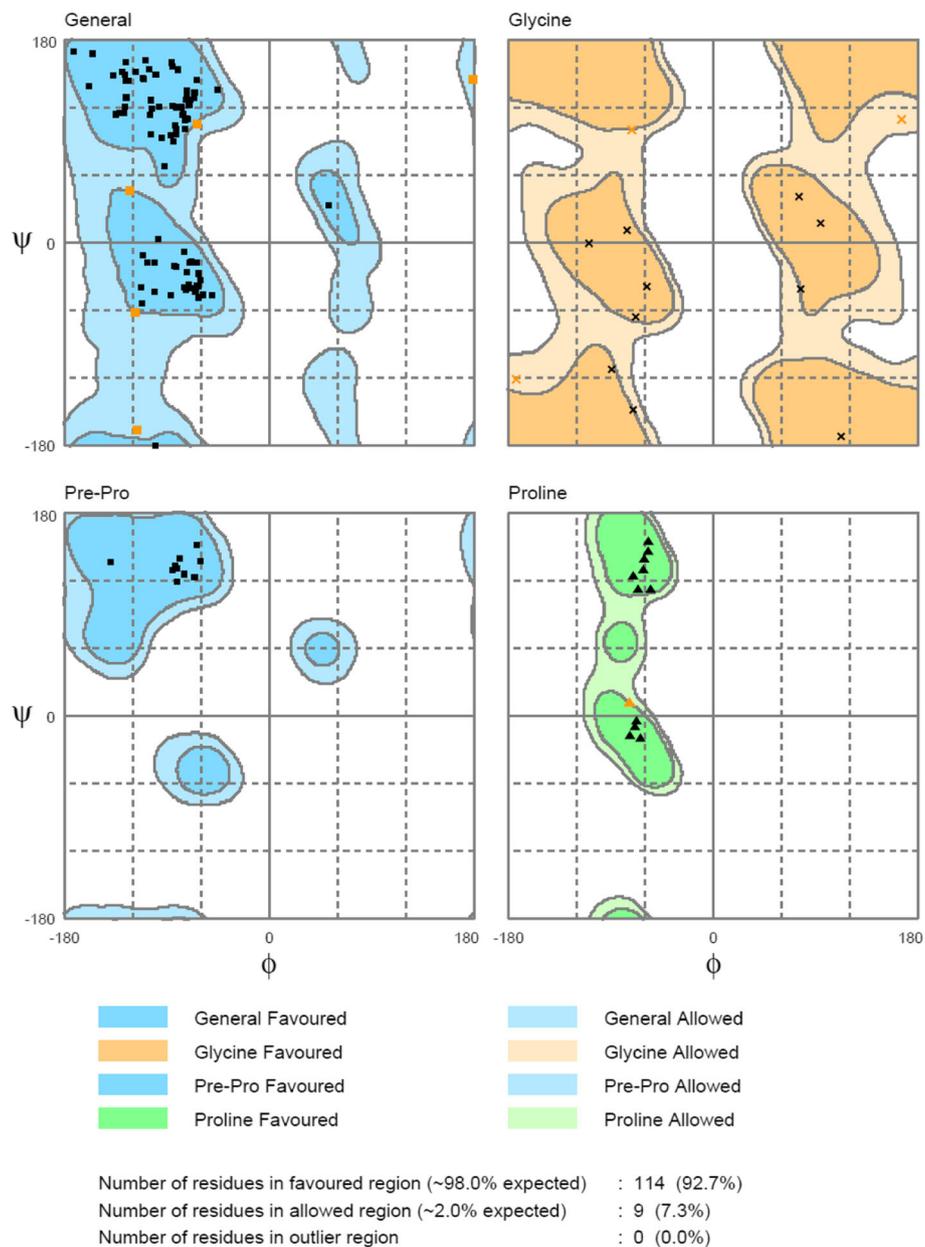
### Nucleotide sequence analysis and comparison

The best aligning nucleotide sequence of the Genbank database for both sequences was *Camellia sinensis* photosystem I reaction center subunit II, chloroplast-like (LOC114287061), and mRNA (sequence ID: [XM\\_028230326.1](#)) derived from shuchazao tea cultivar of China. The EST SSR 073 motif containing cDNA sequence of blister blight disease resistant tea cultivar TRI 2043 displayed 99.07% identity with 0.0% *E* value by covering 92% query coverage. The cDNA sequence of blister blight disease susceptible tea cultivar TRI 2023 showed 99.33% identity with 0.0% *E* value by covering 92% query coverage. Further, nucleotide sequences obtained for both BBS and BBR had over 80% identity with the coding sequence of *Diospyros kaki* photosystem I subunit D-I (PsaD-1) mRNA (ID: KX871204.1) with *E* value of  $5e-117$ .

Pairwise alignment of BBS and BBR sequences indicated 20 microsatellite CT repeat extension in 5'UTR of BBR and a single nucleotide deletion at 552 bp (deletion of C) (Fig. 1).



**Fig. 4** Superposition of the generated models with the PsaD subunit of *P. sativum* (red colored arrows indicate the truncation of PsaD subunit of BBR). **a-** PsaD subunits from BBS (magenta) and *P. sativum* (green). **b-** PsaD subunits from BBR (cyan) and BBS (magenta). **c-** PsaD subunits from BBR (cyan) *P. sativum* (green)



**Fig. 5** Ramachandran plot statistics of BBS model

**Amino acid sequence analysis**

The amino acid sequence obtained from the longest ORF of BBS sequence was 193 amino acid in length and had 97% similarity and 95.2% identity over the full length of PsaD subunit of *Pisum sativum* photosynthesis complex I (Fig. 2).

The BBR protein sequence had a 86.2% identity and 89.1% similarity with the PsaD subunit of *P. sativum*. Both the nucleotide and protein sequence comparison clearly showed a frame shift mutation (single nucleotide deletion) leading to truncated C

terminal caused by the single nucleotide deletion of BBR at 552bp as shown in the sequence alignments in Fig. 3. The search was repeated against PDB database and the crystal structure of *P. sativum* photosynthesis complex I was retrieved for homology modeling.

**Homology modeling and structure comparison**

After homology modeling and energy minimization superimposed model of the BBS sequence retained the same fold and domain structures as the PsaD subunit of

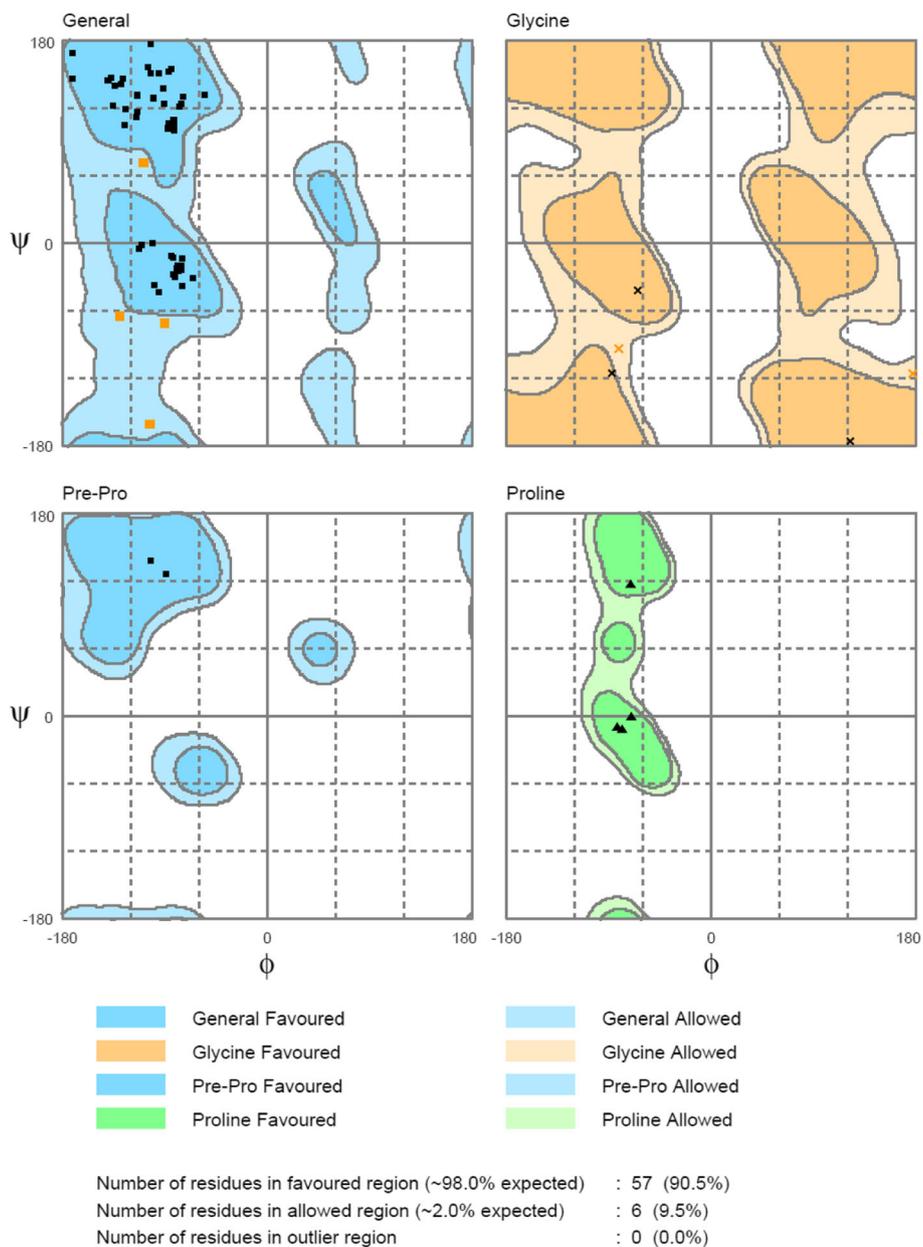


Fig. 6 Ramachandran plot statistics of BBR model

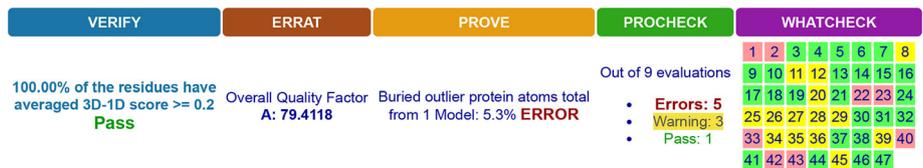
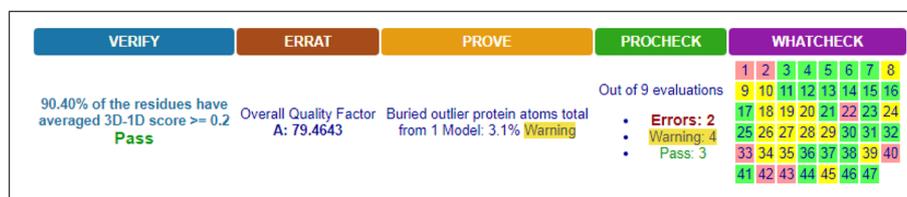


Fig. 7 Summary of BBS model quality analysis by VERIFY, ERRAT, PROVE, PROCHECK, and WHATCHECK servers



**Fig. 8** Summary of BBR model quality analysis by VERIFY, ERRAT, PROVE, PROCHECK, and WHATCHECK servers

*P. sativum* with a RMSD value of 0.530 Å. More importantly, all amino acids that form H bonds with other subunits of the complex were conserved in *P. sativum* and *C. sinensis* (Fig. 4a). The model generated from BBR sequence clearly showed truncation of the C terminus, completely eliminating the anti-parallel beta strands and the C terminal unstructured domain (Fig. 4b, c). Rest of the structure was the same as the BBS and had an RMSD value of 0.616 Å when compared to aligning region of PsaD of *P. sativum*.

**Structure validation**

After energy minimization, the models generated for BBS and BBR were showed good overall stereochemical quality as expected for modeled with high sequence identity with the template [37–39]. The BBS model had no residues in the outlier region, while 92.7% of residues lied in the favored region (Fig. 5), and BBR had 90% of its residues in the favored region, while no residues were found in the outlier region (Fig. 6).

Further quality analysis using VERIFY3D, ERRAT, PROVE, PROCHECK, AND WHATCHECK servers indicated a good overall quality of both BBS (Fig. 7) and BBR (Fig. 8) homology models.

Generated structures and the PsaD subunit of *P. sativum* were further compared with physicochemical parameters such as theoretical isoelectric point (pi), estimated half-life, instability index, aliphatic index, and the grand average hydropathicity (GRAVY) of BBR, BBS, and PsaD subunit of *P. sativum* as given in Table 1.

**Molecular docking and interaction analysis**

For comparative analysis, docked complexes were compared with the interactions of the PSI complex of *Pisum*

**Table 1** Physiological parameters of PsaD subunits

Parameter	<i>P. sativum</i>	BBR	BBS
Estimated half-life	30 H	30 H	30 H
Theoretical pi	9.46	10.10	9.91
Instability index	47.74	60.65	45.55
Aliphatic index	72.87	69.04	69.17
GRAVY	−0.513	−0.310	−0.419

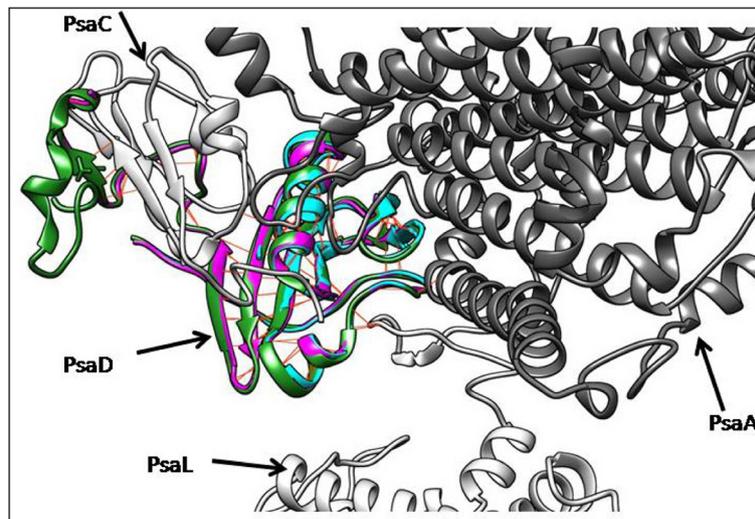
*sativum* (Fig. 9). Interaction analysis showed that in both *P. sativum* and BBS PsaD s, all residues that involve in H bonding with PsaA, Psa C, and Psa L to be conserved and showing similar interaction patterns. However, BBR proteins possessed a C terminal truncation which prevents PsaD from interacting with PsaC.

**Discussion**

The EST SSR 073 motif containing cDNA sequences of both TRI 2043 and TRI 2023 display significant similarity with PsaD I subunit nucleotide sequence of Chinese tea cultivar shuchazao by confirming the reliability of DNA sequence data were used in the study. Furthermore, the DNA sequences displayed high similarity with the *Diospyros kaki* photosystem I subunit D-I. Accordingly, both of the sequences were identified as putative PsaD subunit of photosystem I of *Camellia sinensis*. In almost all plants, the PsaD I gene exists as a single-copy gene [40].

The CT extension of BBR does not change the sequence and the structure of protein. However, it may be involved in the post-transcriptional regulation of PsaD I expression. In fact, 5'UTR of most of the mRNA sequences contains regulatory structures such as hairpins [41]. Furthermore, the BBR sequence possesses a single nucleotide deletion at 552 bp (deletion of C) which leads to a truncation of ORF and the formation of a shorter protein product.

Homology modeling of the resulted amino acid sequences from BBS sequences using PsaD I of *P. sativum* as a template produced a three-dimensional structure of the protein which is very similar to the template used. PsaD I subunit is reported to be expressed as unfolded protein having a leader sequence and later undergoing proper folding when complexed with the rest of the subunit of photosystem I super complex. Also, PsaD subunits possess N terminal and C terminal domains having no rigid structure. However, these unstructured domains of PsaD I, particularly the C terminal unstructured domain has reported to be involved in forming H bonds with PsaC subunit of the complex. Therefore, energy minimization of the models generated for PsaD subunits of BBS and BBR sequences were carried out within the



**Fig. 9** Superposition of docking poses of PsalD subunits from BBS, (magenta), BBR (cyan), and *P. sativum* (green). Brown color lines show H bonds.

binding surface of PsalD I complex with the PsalD subunit.

After energy minimization, the 3D model obtained for BBS was almost identical to the secondary and tertiary structure of the PsalD I of *P. sativum*. Both of the generated models for BBS protein sequence and the PsalD I of *P. sativum* showed identical hydrogen bonding pattern with the other subunits of photosystem I super complex. When both models of BBS were docked against PsalD binding site of the photosynthesis complex, interestingly both protein models retained the same H bonding pattern within the complex by further confirming the adoption of the PsalD I topology by them. Photosystem subunit D of photosystem I complex is hydrophobic and is exposed on stromal face of the thylakoid. The subunit interacts with ferredoxin in both cyanobacteria and eukaryotes [42].

The model generated for BBR had a C terminal truncation that eliminates the entire unstructured C terminal domain along with the C terminal anti-parallel beta sheets. It is unlikely that the BBR sequence would produce the functional protein because the interactions with PsalC is critical to maintain the stability of photosystem complex I. However, the N terminal extension of PsalD I in higher plants stabilizes the interactions with PsalC and rest of the photosystem I complex. Cross-linking study of barely, suggested that PsalD is stabilized with interaction between photosystem I H subunit [43] and PsalD is not tightly bound with photosystem I core [44]. Therefore, the C terminal truncation of the BBR may not involve in changing the stability of photosystem I complex and its main functions. Further, stability study of PsalD in *Synechocystis* has shown the reduced flavodoxin in, photosystem complex I without the PsalD

subunit [45]. The mutation of BBR may be involved in so far unreported function and the predicted model may lead to discover the functions of mutated PsalD subunit.

The half-life of protein is the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell. In this study, the half-life of all the proteins was 30 h. The instability index provides an estimate of the stability of the protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, while a value above 40 predicts that the protein will be unstable [46]. The results from this study recorded instability index higher than 40 in all the three proteins (*P. sativum*, BBR, and BBS) indicating unstable properties. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). If the aliphatic index is higher, the thermostability increases; therefore, the predicted proteins are thermostable. Isoelectric point is the condition of a solution where the amino acid produces the same amount of positive and negative charges and the ultimate charge will be zero. Isoelectric point (pI) of the three proteins was 9.4 to 10.1 and it seemed to be basic protein. The value of GRAVY spread between  $-0.310$  and  $-0.514$  and lower values are suggested to have good interactions between water and protein [47, 48].

In silico computational approach has been applied to predict protein structures of leaf rust disease resistance, and Lr 10 coding protein was identified as more resistant against the leaf rust disease of wheat [49]. However, the PsalD subunit has not been reported as associated with disease resistance up to now. Epicatechin (EC) and epigallocatechingallate (EGCG) are involved in BB disease resistance in tea [50]. Flavonoids biosynthesis

pathway which synthesizes EC and EGCG is light sensitive and therefore, the allele may indirectly involve in the BB disease resistance.

## Conclusions

The EST SSR 073 motif flanking sequences of *Camellia sinensis* is conserved in the Psad I subunit of photosystem I complex, and the developed in silico structures of homology proteins are reliable with their physico-chemical parameters. When compared with BBS, CT repeat extension of BBR did not change the topology of Psad I subunit but the single nucleotide deletion leads to C terminal truncation of BBR coding Psad I subunit by preventing hydrogen bond interaction with other complexes of photosystem I. It can be recommended that more sequence data of EST SSR 073 motif flanking sequences in different tea cultivars and analyzing the protein model would lead to unravel the mechanism of BB resistance.

## Abbreviations

BBS: Blister blight susceptible DNA sequence of TRI 2023; BBR: Blister blight resistance DNA sequence of TRI 2043; BB: Blister blight; SSR: Simple sequence repeat; DNA: Deoxyribonucleic acid; EST: Expressed sequence tag; PsadI: Photosystem I subunit D protein; AER: Agro-ecological region; MRL: Maximum residual level; MAS: Marker-assisted selection; BSA: Bulk segregation analysis; NMR: Nuclear magnetic resonance; MMP: Matrix metalloproteinase; ORF: Open reading frame; NCBI: National Center for Biotechnology Information; PDB: Protein Data Bank; GRAVY: Grand average of hydropathicity; mRNA: Messenger ribonucleic acid; UTR: Untranslated region; EC: Epicatechin; EGCG: Epigallocatechingallate

## Acknowledgements

Authors appreciate Tea Research Institute of Sri Lanka (TRISL) for their supporting to receive the grant. Authors thank the emeritus Professor EH Karunanayake and Professor Kamani H Tennekoon for their guidance.

## Authors' contributions

KHTK and KMM conceived and designed the experiments. NHKS and KHTK analyzed the data. KHTK and KMM performed the experiment. KHTK and NHKS wrote the manuscript. SACN and KMM reviewed and revised the manuscript. OVDSJ and KMM supervised and validated the data. All authors have read and approved the manuscript.

## Funding

This work was financially supported by the National Research Council, Sri Lanka (Grant No: NRC 09-066). The funding body did not participate in the design of the study, nor in the collection, analysis, and interpretation of data or in writing the manuscript.

## Availability of data and materials

Authors declare that all generated and analyzed data are included in the article.

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Colombo, Sri Lanka. <sup>2</sup>Current address: Department of biosystems

Technology, Faculty of Technology, University of Ruhuna, Matara, Sri Lanka.

<sup>3</sup>Department of Biotechnology, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila, Sri Lanka.

<sup>4</sup>Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Peradeniya 20400, Sri Lanka.

Received: 16 March 2020 Accepted: 4 June 2020

Published online: 19 July 2020

## References

1. Anon (2017) Annual ITC Bulletin of Statistics. International Tea Committee. pp 51-61
2. Agnihothrudu V, Moulli BC (1991) Blister blight of tea, its control and future lines of research. In: Proceedings of International Symposium on Tea Science, Shizuoka, Japan. 26-29 Aug 1991. p. 655-59
3. TRISL (2002) TRI advisory Circular, no DM: 1 – protection of tea from blister blight. Talawakelle; TRISL-1-5
4. Baby UI, Balasubramanian S, Ajay D, Premkumar R (2004) Effect of ergosterol biosynthesis inhibitors on blister blight disease, the tea plant and quality of made tea. *Crop Protection*. 23:795–800
5. Barooahi AK, Kalita N, Borthakur MN, Barooah A, Barman T (2002) Non-tariff trade barriers - pesticide residues and heavy metals in tea and approaches to mitigate them. *Two and a Bud*. 59(2):1–8
6. Karunarathna KHT (2019) Use of SSR markers for genetic characterization and finger-printing of recommended tea (*Camellia sinensis* L) cultivars and identification of potential marker/s for blister blight disease resistance. Dissertation. Institute of Biochemistry, Molecular Biology and Biotechnology (IBMBB), University of Colombo
7. Canduri F, Jr WDA (2008) Protein Crystallography in Drug Discovery. *Current Drug Targets* 9(12):1048–1053. <https://doi.org/10.2174/138945008786949423>
8. Gupta CL, Akhtar S, Bajpai P (2014) In silico protein modeling: possibilities and limitations. *EXCLI J*. 13:513–5. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4467082/>
9. Fadel V, Bettendorff P, Herrmann T, Jr WFDA, Oliveira EB, Yamane T, Wüthrich K (2005) Automated NMR structure determination and disulfide bond identification of the myotoxin crostamine from *Crotalus durissus terrificus*. *Toxicol* 46(7):759–767. <https://doi.org/10.1016/j.toxicol.2005.07.018>
10. Fiser A (2010) Template-based protein structure modeling. *Methods Mol Biol* 673:73–94. [https://doi.org/10.1007/978-1-60761-842-3\\_6](https://doi.org/10.1007/978-1-60761-842-3_6)
11. Bitencourt-Ferreira G, Azevedo WFD (2019) Homology Modeling of Protein Targets with MODELLER. *Methods in Molecular Biology Docking Screens for Drug Discovery*. pp. 231–249. [https://doi.org/10.1007/978-1-4939-9752-7\\_15](https://doi.org/10.1007/978-1-4939-9752-7_15)
12. Azevedo WFD, Canduri F, Oliveira JSD, Basso LA, Palma MS, Pereira JH, Santos DS (2002) Molecular model of shikimate kinase from *Mycobacterium tuberculosis*. *Biochemical and Biophysical Research Communications* 295(1): 142–148. [https://doi.org/10.1016/s0006-291x\(02\)00632-0](https://doi.org/10.1016/s0006-291x(02)00632-0)
13. Pereira JH, Canduri F, Oliveira JSD, Silveira NJFD, Basso LA, Palma MS, Santos DS (2003) Structural bioinformatics study of EPSP synthase from *Mycobacterium tuberculosis*. *Biochemical and Biophysical Research Communications* 312(3):608–614. <https://doi.org/10.1016/j.bbrc.2003.10.175>
14. Joshi YN, Gajul SG (2019) In-silico Homology Modeling of MMP25 involved in Asthma. *IJSRST*. 4:9: 202-208
15. Satyanarayana SDV, Krishna MSR, Kumar PP, Jeeredy S (2018) In silico structural homology modeling of nif A protein of rhizobial strains in selective legume. *plants Journal of Genetic Engineering and Biotechnology* 16:731–737
16. Mistry P, Kalaria RK, Patel AI (2019) Molecular identification and in silico characterization of coat protein in chilli leaf curl virus associated in chilli from south Gujarat region of India, *International Journal of Chemical Studies* 7(4):711-718
17. Aamir M, Singh VK, Dubey MK, Meena M, Kashyap SP, Katari SK, Upadhyay RS, Umamaheswari A, Singh S (2018) In silico Prediction, Characterization, Molecular Docking, and Dynamic Studies on Fungal SDRs as Novel Targets for Searching Potential Fungicides Against Fusarium Wilt in Tomato. *Front. Pharmacol* 9:1038. <https://doi.org/10.3389/fphar.2018.01038>
18. Mewan KM (2011) Application of DNA Markers for Genetic Conservation and Breeding of Tea [dissertation]. Department of Plant Sciences, Faculty of Science, University of Colombo

19. Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403–410
20. Rombel IT, Sykes KF, Rayner S, Johnston SA (2002) "ORF-FINDER: a Vector for High-Throughput Gene Identification". *Gene* 282(1-2):33–41. [https://doi.org/10.1016/s0378-1119\(01\)00819-8](https://doi.org/10.1016/s0378-1119(01)00819-8)
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
22. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al (2016) UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Methods Mol Biol* 1374:23–54. [https://doi.org/10.1007/978-1-4939-3167-5\\_2](https://doi.org/10.1007/978-1-4939-3167-5_2)
23. Minai L, Cohen Y, Chitnis PR, Nechushtai R (1996) Photosystem I complex in two steps. *PNAS* 93(13):6338–6342. <https://doi.org/10.1073/pnas.93.13.6338>
24. Lagoutte B, Hanley J, Bottin H (2001) Multiple Functions for the C Terminus of the Psad subunit in the Cyanobacterial Photosystem I Complex. *Plant Physiology* 126:307–316
25. Fiser A, Sali A (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics application notes* 19(18):2500–2501. <https://doi.org/10.1093/bioinformatics/btg362>
26. Mazor Y, Borovikova A, Caspy I, et al (2017) Structure of the plant photosystem I super complex at 2.6 Å resolution. *Nature Plants* 3:17014. <https://doi.org/10.1038/nplants.2017.14>
27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of computational chemistry* 25(13):1605–12. <https://doi.org/10.1002/jcc.20084>
28. Bitencourt-Ferreira G, Azevedo WFD (2019) Docking with SwissDock. *Methods in Molecular Biology Docking Screens for Drug Discovery* 189–202. [https://doi.org/10.1007/978-1-4939-9752-7\\_12](https://doi.org/10.1007/978-1-4939-9752-7_12)
29. Croll TI, Sammito MD, Kryshtafovych A, Read RJ (2019) Evaluation of template-based modeling in CASP13. *Proteins* 1–15
30. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
31. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712–725
32. Bitencourt-Ferreira G, Veit-Acosta M, Azevedo WFD (2019) Van der Waals Potential in Protein Complexes. *Methods in Molecular Biology Docking Screens for Drug Discovery* 79–91. [https://doi.org/10.1007/978-1-4939-9752-7\\_6](https://doi.org/10.1007/978-1-4939-9752-7_6)
33. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95–99
34. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8:52–56
35. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al (2005) Protein identification and analysis tools on the ExPASy server. In Walker JM (ed) *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, pp. 571–607
36. Zundert VGCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastrius PL, Karaca E, et al (2016) The HADDOCK 2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* 428(4):720–725. <https://doi.org/10.1016/j.jmb.2015.09.014>
37. Vilar S, Sobarzo-Sánchez E, Uriarte E (2019) In Silico Prediction of P-glycoprotein Binding: Insights from Molecular Docking Studies. *Curr Med Chem* 26(10):1746–1760. <https://doi.org/10.2174/0929867325666171129121924>
38. Silveira NJFD, Bonalumi C E, Uchôa, HB, Pereira, JH, Candurí F, Azevedo WFD (2006) DBMODELING: A Database Applied to the Study of Protein Targets From Genome Projects. *Cell Biochem Biophys* 44(3):366–374. <https://doi.org/10.1385/cbb.44.3.366>
39. Azevedo WFD (2011) Molecular Dynamics Simulations of Protein Targets Identified in Mycobacterium tuberculosis. *Curr Med Chem* 18(9):1353–1366. <https://doi.org/10.2174/092986711795029519>
40. Minai L, Fish A, Yahana MD, Verchovsky L, nechustai R (2001) The Assembly Of The Psad Subunit Into The Membranal Photosystem I Complex Occurs via An exchange Mechanism. *Biochemistry* 40(43):12754–12760. <https://doi.org/10.1021/bi015694i>
41. Leppék K, Das R, Barna M (2017) Functional 5' UTR Mrna Structures In Eukaryotic Translation Regulation And How To Find Them. *Nature Reviews Molecular Cell Biology* 19(3):158–174. <https://doi.org/10.1038/nrm.2017.103>
42. Andersen B, Koch B, Scheller HV (1992) Structural and functional analysis of the reducing side of photosystem I. *Physiol Plant* 84:154–161
43. Naver H, Schott MP, Andersen B, Moller BL, Scheller HV (1995) Reconstitution of barley photosystem I reveals that the N-terminus of the PSI-D subunit is essential for tight binding of PSI-C. *Physiologia Plantarum* 95(1):19–26
44. Naver H, Haldrup A, Scheller HV (1999) Cosuppression of photosystem I subunit PSI-H in *Arabidopsis thaliana*. Efficient electron transfer and stability of photosystem I is dependent upon the PSI-H subunit. *J Biol Chem* 274: 10784–10789
45. Xu Q, Jung YS, Chitnis VP, Guikema JA, Golbeck JH, Chitnis PR (1995) Mutational analysis of photosystem I polypeptides in *Synechocystis* sp. PCC 6803. Subunit requirements for reduction of NADP+ mediated by ferredoxin and flavodoxin. *J Biol Chem* 269(34):21512–21518
46. Guruprasad K, Reddy BV, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng* 4(2): 155–161. <https://doi.org/10.1093/protein/4.2.155>
47. Verma A, Singh VK, Gaur S (2016) Computational based functional analysis of Bacillus phytases. *Comp Biol Chem* 60:53–58. <https://doi.org/10.1016/j.cmpbiolchem.2015.11.001>
48. Pramanik K, Soren T, Mitra S, Maiti TK (2017) In silico structural and functional analysis of Mesorhizobium ACC deaminase. *Comp Biol Chem* 68: 12–21. <https://doi.org/10.1016/j.cmpbiolchem.2017.02.005>
49. Kumar A, Mishra DC, Rai A, Gajula MNVP (2013) In Silico analysis of protein - protein interaction between resistance and virulence protein during leaf rust disease in wheat (*Triticum aestivum* L.). *World Research Journal of Peptide and Protein*. 2. 2278–4586
50. Punyasiri PAN, Abeyasinghe ISB, Kumar V (2005) Preformed and induced chemical resistance of tea leaf against *Exobasidium vexans* infection. *J Chem Ecol* 13:15–24

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)