Available online at www.elixirpublishers.com (Elixir International Journal)

Bio Technology

Elixir Bio. Tech. 49 (2012) 10152-10159

ISSN: 2229-712X

In silico analysis of phytochelatin: a stress related protein of *Cynodon dactylon*

N.Vinod Kumar[1], E.G.Wesely[2] and Lizzy Mathew[3]

[1]Department of Biotechnology, Muthayammal College of Arts and Science, Rasipuram 637 408, Namakkal

[2]Department of Botany, Arinagar Anna Government Arts College, Namakkal.

[3]Department of Botany, St. Teresa's College, Ernakulam.

ARTICLE INFO

Article history:

Received: 22 June 2012;

Received in revised form:

16 August 2012;

Accepted: 27 August 2012;

Keywords

Phytochelatin,

Cadmium stress,

Phylogenetic analysis,

Structure prediction.

ABSTRACT

Phytochelatins (PCs), are enzymatically synthesized Cys-rich peptides, found widely distributed in the plant kingdom and was the functional equivalent of metallothioneins (MTs). Synthesis of PC was activated by the exposure to heavy metal ions like cadmium. PC may have a role in the metal ion detoxification and help in phytoremediation approach. Hence this study was focused on the structural aspects of phytochelatin protein isolated from *Cynodon dactylon*. Protein sequence was retrieved from NCBI protein database with gene bank id AAO138102. Various bioinformatics online tools like GOR 4, Geno3D, ProtParam, ProtScale, TmPred, EsyPred3D analysis and GlobPlot were used to predict the structure and properties of the protein. An unstable protein with 504 amino acid residues and of molecular weight of 55803.5 Daltons was identified by *In Silico* studies. Further analysis using a wide range of tools revealed that the protein is a structural variant of the phytochelatin protein.

Introduction

Phytochelatin (PC) is one of the important proteins involved in the heavy metal ion detoxification and plays a crucial role in phytoremediation. The expression of this protein occurs usually on exposure to toxic heavy metals by plants. A variety of metal-binding ligands have been described in plants, and their respective roles in heavy-metal detoxification have been reviewed (Rauser, 1999). Low molecular weight phytochelatin–metal complexes are transported into vacuoles by ATP binding cassette transporters, as shown for *S.pombe* (Ortiz *et al*., 1992, 1995).To store metals or to prevent toxic effects of essential and non-essential metals, metal ions can be directly sequestered into the vacuole or bound to chelators (metallothioneins, phytochelatins, glutathione, organic acids, amino acids), which can be also transported into vacuoles or excreted from the cell (Lee et al., 1996; Clemens, 2001). The activation of phytochelatin synthase occurs by the binding of metal ions and metal-glutathione complexes (Cobbett, 2000; Vatamaniuk et al., 2000).

PC play similar role to that of an intracellular chealtor in metal ion transport. Inactivation of toxic metal ions through the metal-induced synthesis of phytochelatins and the formation of metal–phytochelatin complexes is the general cytoplasmic mechanism for trace metal homeostasis in plants and provides tolerance against non-essential heavy metals owing to its relatively unspecific mode of action (Zenk, 1996). Though the proteins are expressed in many plant species, the exact properties and their role in mechanism of detoxification remains unclear. There are no studies which focused on the comparison of the PC in different plant species. Moreover, the stability and other biochemical properties are not well studied.

The role of phytochelatin in heavy metal ion detoxification has been studied by many but the interaction mechanism and site of interactions are yet to be traced out. The protein which is expressed only on exposure to heavy metals can be used as a bioindicator for heavy metal contamination. The principle can be utilized for designing a suitable biosensor. The detoxification effect of the protein is very crucial in remediation of the heavy metal polluted site. The tolerance and expression level of the phytochelatin as well as its detoxification efficiency need to be well studied. The degree of detoxification of heavy metals may influence the rate of bioaccumulation into the food web. The biomagnification of these contaminants to the final consumers like humans can be reduced to a greater extend.

In the present study, PC from *Cynodon dactylon* was studied based on the bioinformatics tools available. These tools enable us to predict certain properties of this protein based on the various matrices made out of the protein sequence. *C. dactylon* also called Bermuda grasses is a typically common species in many grazing fields. PC protein of this plant has a great importance in reduction of heavy metal accumulation into food web. This study remains as a preliminary effort to trace the exact properties of the protein but needs a confirmative scale of experiments to be performed regarding its prediction. The analysis resulted in elucidation of a different conformation property of the PC protein thereby a changed biochemical characteristics. The results of the work motivates one for further analysis of the protein deeper for understanding the prediction variations with existing structural and property facts.

Materials and Methods

Phytochelatin protein sequence of *Cynodon dactylon* [AAO13810.2] was retrieved from the NCBI data base (http://www.ncbi.nlm.nih.gov) and made as the query sequence for the structure and properties prediction and modeling. The sequence obtained was analyzed using various softwares available in the ExPASy server (Gasteiger *et al*., 2005). The GOR 4 analysis was performed to understand the presence of helices, beta turns and coils in the protein structure (Garnier *et al*., 1996). The Geno3D analysis (http://geno3d-pbil.ibcp.fr) was performed and the PSIBLAST results obtained were observed

Tele:

E-mail addresses: nvkibt@gmail.com

(Combet *et al.,* 2002). First similar sequence was choosen and full model analysis was performed. Energy of models was computed and represented in Kcal/mol. Ramachandran plots for five models were obtained. Ramachandran plots for different residues in a model were also studied. Main chain parameters and side chain parameters were analyzed using PROCHECK (Laskowski *et al*., 1993). Local deviation by residues, restrain repartition along sequence and its violation repartition was also plotted.

ProtParam software analysis was done to understand about the amino acid composition, total number of negatively and positively charged residues, atomic composition, formula, number of atoms, instability index, estimated half-life, aliphatic index and grand average of hydropathicity (GRAVY) (Gasteiger *et al*., 2005). ProtScale analysis was done to analyze the variation of 20 residues for different scales like HPLC / TFA retention, HPLC / HFBA retention, alpha-helix / Chou & Fasman, beta-turn / Chou & Fasman, beta-sheet / Deleage & Roux and Polarity / Grantham (Gasteiger *et al*., 2005). Disorder propensity sum and dy/dx propensity was computed using GLOBPLOT2 (Rune *et al*., 2003). ESyPred3D analysis was done for obtaining protein models similar to the query protein (Lambert *et al*., 2002).The model was visualized in Raswin software and measurements were obtained by pick distance option. The length, width of the protein structure and length of each coil etc. were measured. Absolute surface accessibility of amino acid residues were calculated based on Petersen method (Petersen *et al.,* 2009).The *C.dactylon* phytochelatin sequence was used as query sequence and pBLAST was performed. Sequences of fifteen different plant sources were retrieved with higher similarity.

Clustal W multiple sequence alignment was done for those sequences using BioEdit5.0. The aligned sequences were used to construct phylogentic tree using neighbour joining method (Saitou and Nei, 1987) in MEGA4.0 (Tamura *et al.,* 2007). Tajima' relative test was also performed by varying different sequences as out groups but *C.dactylon* sequences remain constant. Three sequences were alone used at a time. Drawhca analysis for hydrophobic clusters was performed for the reference sequence (Callebaut *et al.,* 1997; Gaboriaud *et al.,* 1987). RADAR analysis was done in EMBL server. GlobPlot Analysis of the query protein was also performed to understand the disorder propensity according to Linding and Russell definition (Linding *et al.,* 2003). TargetP analysis of the sequence was done for understanding sub cellular localization (Emanuelsson *et al.,* 2000).

Results

GOR 4 analysis shows that there is 33.53% of alpha helix, 16.87% extended strands and 49.60% random coils. The PSI BLAST analysis resulted in many sequences among which only the first model (pdb2bwA-0) alone was choosen for further study. 5 different models were predicted after full model analysis using Geno3D. The deviation on carbon alpha with respect to amino acid position was represented in the Fig 2.Restarin repartition and violation of repartition was plotted in fig 3a and fig 3b respectively. The models resulted after the Geno3D analysis is shown below. Their energy was represented in kcal/mol. Ramachandran plots of each models (Fig 4) showed variations in allowed, disallowed and additional regions which is depicted in table1.The model energies of model 1 to 5 in kcal/mol are -8228.62, -8406.69, -8082.24, -8455.48 and -8236.17 respectively. Secondary structure and the accessibility sites were graphical represented in fig 5.

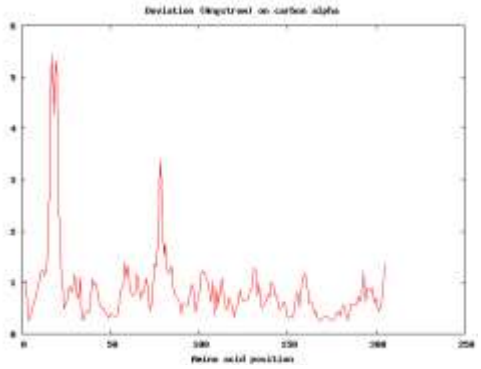


Figure 2.The deviation on carbon alpha with respect to amino acid position

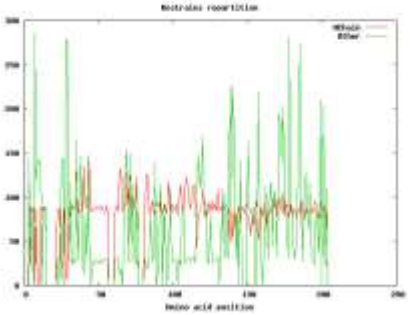


Fig.3a

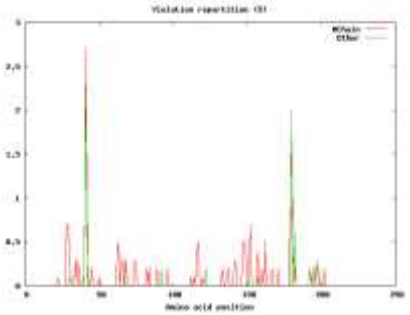
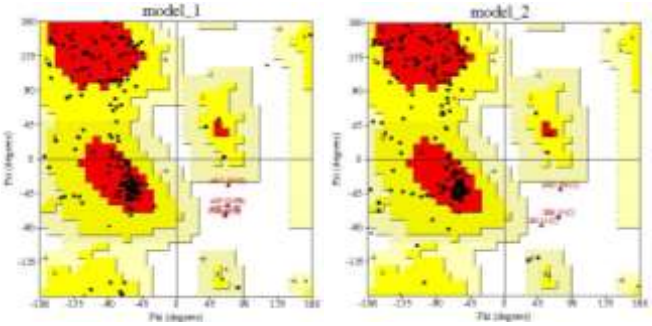


Fig.3b

Figure 3: Restrain repartition (a) and violation of repartition (b) of amino acid residues.

Model	Residues Present In(%)			
	Most favoured regions	Additional Allowed Region	Generously Allowed Region	Disallowed Region
1	71.3	25.3	1.1	2.2
2	71.3	23	3.9	1.7
3	75.3	23	1.7	0
4	72.5	23	1.7	2.8
5	71.9	25.3	1.7	1.1

Table1: Plot Statistics of Ramachandran plot for each models



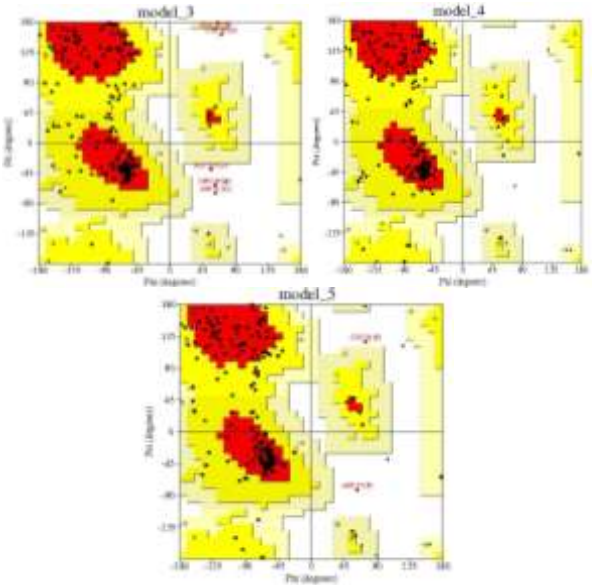
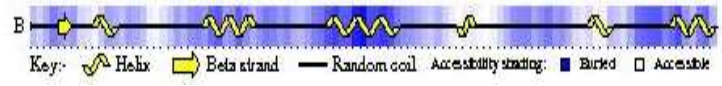


Figure 4: Ramachandran plots of similar predicted models using Geno3D full model analysis.

Model 1



Model 2

Model 3

Model 4

Model 5

Figure 5: Secondary structure and accessibility

The ProtParam analysis showed that the query protein sequence was composed of 504 amino acid residues and found to possess molecular weight of 55803.50 Da. Theoretical pI value was found to be 6.85. The amino acid composition of the protein was analyzed and found that it is a leucine rich (12.1%) protein with 61 residues. It was followed by amino acids like alanine (8.7%), Serine (8.9%), glutamine (6.9%) and valine (6.2%). Pyl and Sec residues were not found. Trp (8%) and Tyr (9%) were among the least.

Total number of negatively charged (Asp+Glu) and positively charged (Arg+Lys) residues were found to be 56 and 55 respectively. Atomic formula was found to be $C_{2459}H_{3925}N_{679}O_{726}S_{37}$. Total number of atoms was 7826. The estimated half life were 30 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*) and > 10 hours (*E.coli*, *in vivo*) when considering Methonine residue found at the N terminal. The protein was predicted to be unstable which has an instability index of 55.80. Aliphatic index of the protein was estimated to be 90.02 and grand average of hydropathicity found to be -0.057. ProtScale analysis based on six scales namely HPLC/TFA retention (fig 6a), Alpha helix/ Chou and Fasman(fig 6b), Beta turn/ Chou and Fasman(fig 6c), Beta sheet/ Deleage and Roux(fig 6d), Polarity/ Grantham(fig 6e) and HPLC/HFBA retention(fig 6f),. Scores of individual amino acids were given in the table 2.

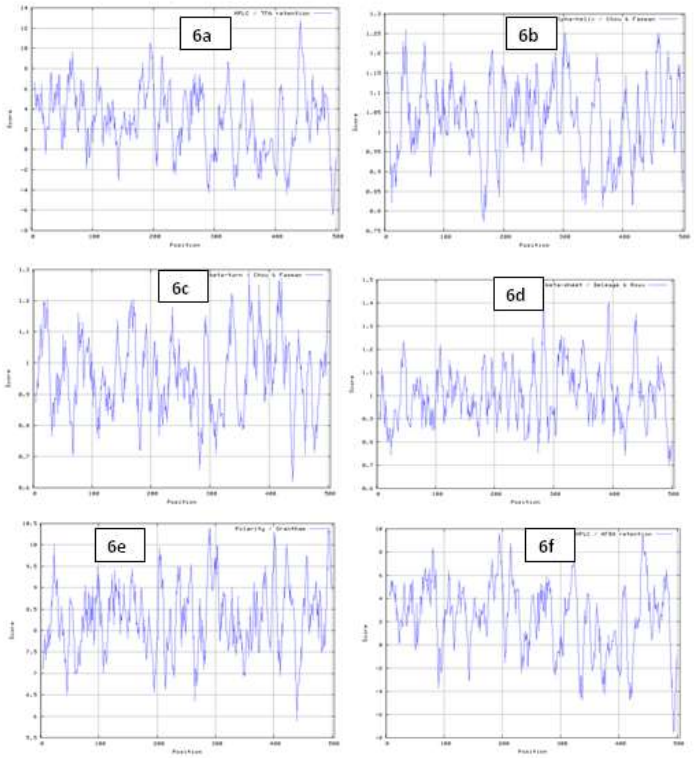


Figure 6- ProtScale analysis plot for various amino acid residues of different parameters

a) HPLC/TFA retention
b) Alpha helix/ Chou and Fasman
c) Beta turn/ Chou and Fasman
d) Beta sheet/ Deleage and Roux
e) Polarity/ Grantham
f) HPLC/HFBA retention

Amino acids	Analysis scores for each amino acids					
	HPLC/TFA retention	Alpha-helix	Beta turn	Beta sheet	Polarity	HPLC/HFBA retention
Ala	7.300	1.420	0.709	8.100	3.900	0.660
Asn	-5.700	0.670	0.604	11.600	-2.800	1.560
Cys	-9.200	0.700	1.191	5.500	-14.300	1.190
Glu	-7.100	1.510	0.567	12.300	-7.500	0.740
His	-2.100	1.000	0.863	10.400	2.000	0.950
Leu	20.000	1.210	1.261	4.900	15.000	0.590
Met	5.600	1.450	1.210	5.700	4.100	0.600
Pro	5.100	0.570	0.354	8.000	5.600	1.520
Thr	0.800	0.830	1.221	8.600	1.100	0.960
Tyr	5.900	0.690	1.266	6.200	3.800	1.140
Arg	-3.600	0.980	0.920	10.500	3.200	0.950
Asp	-2.900	1.010	0.541	13.000	-2.800	1.460
Gln	-0.300	1.110	0.840	10.500	1.800	0.980
Gly	-1.200	0.570	0.657	9.000	-2.300	1.560
Ile	6.600	1.080	1.799	5.200	11.000	0.470

Lys	-3.700	1.160	0.721	11.300	-2.500	1.010
Phe	19.200	1.130	1.393	5.200	14.700	0.600
Ser	-4.100	0.770	0.928	9.200	-3.500	1.430
Trp	16.300	1.080	1.306	5.400	17.800	0.960
Val	3.500	1.060	1.965	5.900	2.100	0.500

Table 2: Score of Individual Amino acids in the ProtScale Analysis

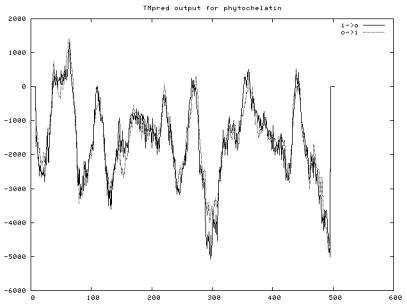


Figure 7: Transmembrane topology prediction for the phytochelatin sequence of *C.dactylon*.

From the TmPred analysis, totally eleven helices possibilities were found and among them five were from inside to outside and six from outside to inside. Two possible models for transmembrane topology have been suggested. The first model with a score of 2057 suggest an outside to inside region from residue number 29 to 49 and an inside to outside region from residue 56 to 75. The alternative model suggests a single segment from inside to outside from residue number 56 to 75 with a score of 1298.TmPred plot is depicted in fig 7. Among the models predicted only five predictions were found significant which are shown in Table 3.

```
Inside to outside helices
from        to     score center
56 (  56)  75 (  73)    1298     65
351 ( 351) 373 ( 369)     513    361
430 ( 430) 451 ( 451)     510    440

Outside to inside helices
from        to     score center
29 (  29)  49 (  49)     759     39
56 (  56)  72 (  72)    1398     64
```

Table 3: Possible transmembrane models suggested for the phytochelatin using TmPred

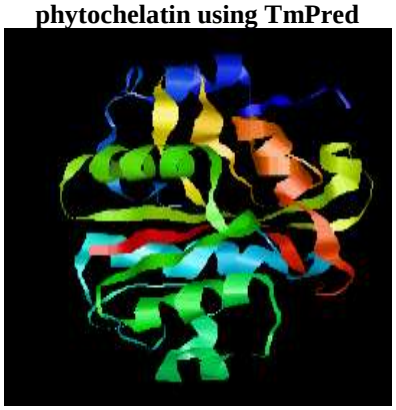


Figure 8: Modeling of Phytochelatin query protein using ESyPred3D analysis

During the ESyPred3D analysis, a 3D model of query protein has been built using the 3D structure 2BTW chain 'A' as template (Fig 8). This template shares 15.0% identities with your query sequence (using the ALIGN program).The measurement of the protein structure predicted by ESyPred3D was done. Length and width of the protein structure was measured by visualizing in Raswin software (Fig 9). The approximate length from distant amino acid residues were considered as its length and was measured as 43.08, 44.69 and 43.52 Å. Width was measured as 32.87 and 38.99 Å. Length of each coils were measured and were found to be 24.02, 21.34, 17.66, 25.04, 13.34 and16.81 Å . The distance between the free terminal end amino acid residues SER15 CA and SER219 CA was found to be 32.87 Å (Fig 10)

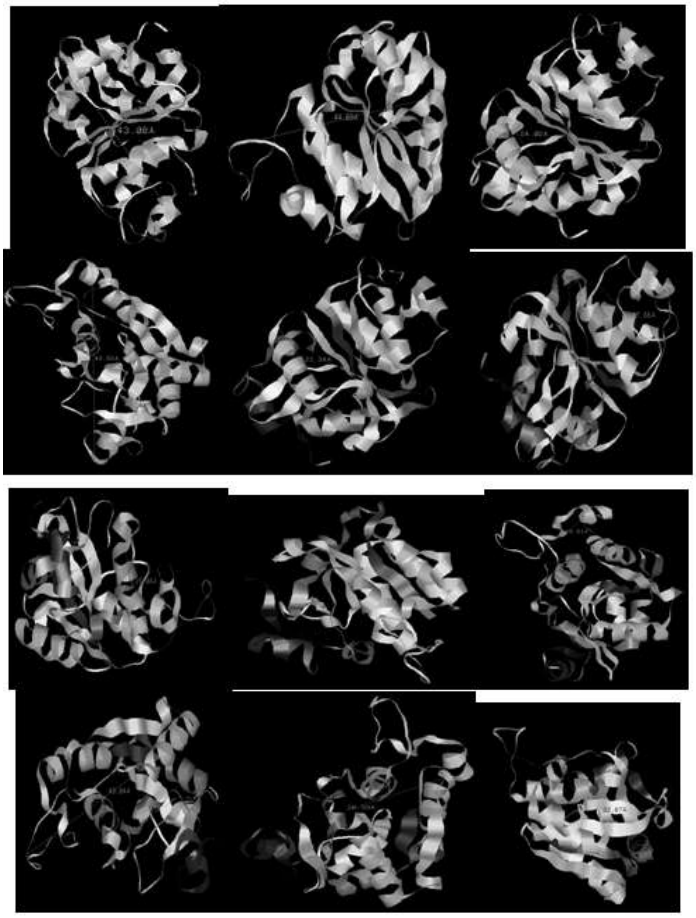


Figure 9: Raswin analysis and measurement of the length, width of protein and length of coils in the protein structure.

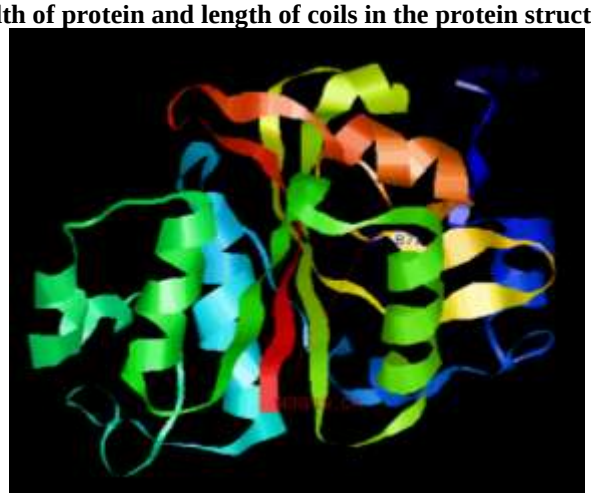


Figure 10: Measuring the distance between two terminal amino acid residues using Raswin.

The GlobPlot analysis of the query sequence showed the disorder propensity graph. The dy/dx propensity values plotted against the amino acid residues were also obtained (Fig11). The disorder regions were marked by uppercase letters in the sequence. 13-24, 160-172, 367-371, 411-420 and 497-502

amino acid residues were shown to have maximum disorder propensity. Its worth mentioning that none of the disordered region tends to form a transmembrane structure predicted earlier using TmPred.

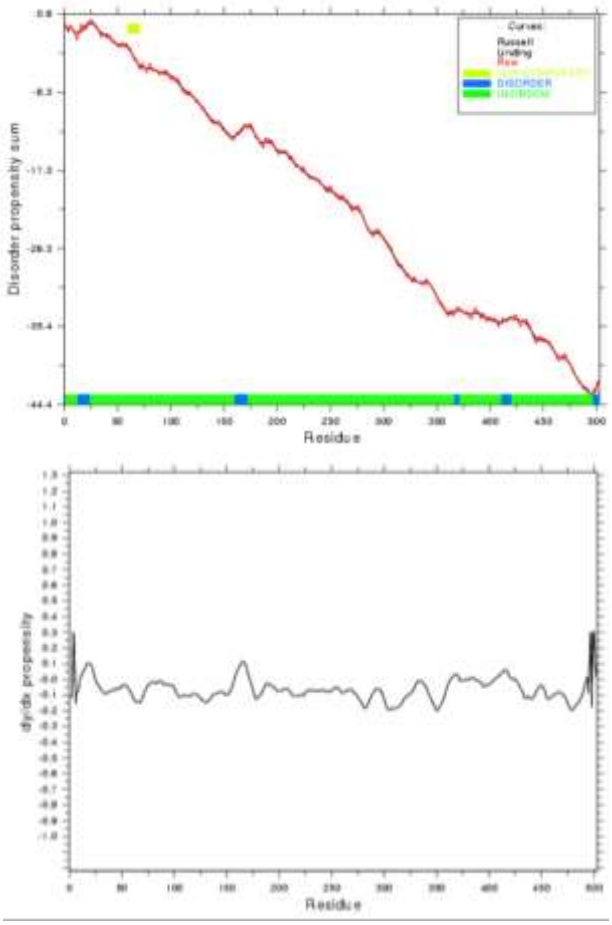


Figure 11: The disorder propensity and dy/dx propensity plot for the phytochelatin sequence.

Amino acid residues with absolute surface accessibility (ASA) higher than 100 were choosen for prediction of the suitable sites for accessing the particular protein. There were 42 residues with ASA values higher than 100. ASA values higher than 700 were grouped and compared. It showed that 6 residues resulted in maximum probability value (above 0.900) for coiled regions.4 residues with ASA values between 500-600 shown higher probability values to take alpha helical structures. It is notable that no residues were to take the beta conformations. This resulted in an understanding that coils are most accessible region in the particular protein but the accessible proteins are few in number.

The Phytochelatin protein sequence of *C.dactylon* was given as the query sequence and pBLAST was performed. Various similar sequences of different plant species were obtained. Among those, protein sequences from fifteen different plant species were retrieved and Clustal W multiple sequence alignment was performed using BioEdit. Phylogenetic tree was constructed among these plants using MEGA 4.0 by neighbour joining method (Fig 11). Tajima' relative test resulted in high divergence among the sequences (Table 3a-g).

In Table 3a, the equality of evolutionary rate between AAO13810.2 (*Cynodon dactylon)* and XP 002454970.1 (*Sorghum bicolor)* is tested using NP 001168641.1 (*Zea mays)* as an out group in Tajima' relative rate test in MEGA4. The χ^2 test statistics was 60.84 (P=0.00000 with 1 degree(s) of freedom). P value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages. In Table 3b, the equality of evolutionary rate between AAO13810.2 (*Cynodon dactylon)* and EEE64936.1 (*Oryza sativa)* is tested using XP 003560864.1 (*Brachypodium distachyon)* as an out group in Tajima' relative rate test in MEGA4. The χ^2 test statistics was 4.26 (P=0.03913 with 1 degree(s) of freedom

In Table 3c, the equality of evolutionary rate between AAO13810.2 (*Cynodon dactylon)* and BAJ85525.1 (*Hordeum vulgare)* is tested using Q9SWW5.1 (*Triticum sp.)* as an out group in Tajima' relative rate test in MEGA4. The χ^2 test statistics was 118.03 (P=0.00000 with 1 degree(s) of freedom). In Table 3d, the equality of evolutionary rate between AAO13810.2 (*Cynodon dactylon)* and AAO13809.1 (*Allium sativum)* is tested using ABX10958.1 (*Nicotiana glauca)* as an out group in Tajima' relative rate test in MEGA4. The χ^2 test statistics was 11.89 (P=0.00056 with 1 degree(s) of freedom). In Table 3e, the equality of evolutionary rate between AAO13810.2 (*Cynodon dactylon)* and XP 002529835.1 (*Ricinus communis)* is tested using ACT87974.1 (*Sesbania rostrata)* as an out group in Tajima' relative rate test in MEGA4. The χ^2 test statistics was 54.73 (P=0.00000 with 1 degree(s) of freedom).

In Table 3f, the equality of evolutionary rate between AAO13810.2 (*Cynodon dactylon)* and AAO74500.1 (*Nicotiana tabacum)* is tested using CAD68109.1 (S*olanum tuberosum)* as an out group in Tajima' relative rate test in MEGA4. The χ^2 test statistics was 149.76 (P=0.00000 with 1 degree(s) of freedom). In Table 3g, the equality of evolutionary rate between AAO13810.2 (*Cynodon dactylon)* and XP 002320627.1 (*Populus trichocarpa)* is tested using ACU44656.1 (*Sonchus arvensis)* as an out group in Tajima' relative rate test in MEGA4. The χ^2 test statistics was 5.20 (P=0.02259 with 1 degree(s) of freedom).

The phylogenetic tree was constructed using neighbour joining method. Two major groups were visible. Among the group the amino acid sequence of *C.dactylon* was grouped along with the phytochelatin sequences of *S.bicolor* and *Z. mays*. There was much divergence observed among the selected sequences of the particular protein. The stress related protein phytochelatin was only the basis for the comparison and hence the related of these plants based on stress physiological parameters are likely to be similar. Other predicted properties of the *C.dactylon* phytochelatin needs to be compared with the other plant phytochelatins for establishing a similarity of its functional and structural properties. The Tajima' relative test results for these plant species were high which strengthen the relational evidence between the phytochelatin protein of these three plants namely *C.dactylon, S. bicolor* and *Z. mays*.

The estimated value of the shape parameter for the discrete Gamma Distribution is 1.3060. Substitution pattern and rates were estimated under the Jones-Taylor-Thornton (1992) model (+G) (Jones *et al.*, 1992). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+*G*]). Mean evolutionary rates in these categories were 0.16, 0.44, 0.76, 1.23, 2.40 substitutions per site. The amino acid frequencies are 7.69% (A), 5.11% (R), 4.25% (N), 5.13% (D), 2.03% (C), 4.11% (Q), 6.18% (E), 7.47% (G), 2.30% (H), 5.26% (I), 9.11% (L), 5.95% (K), 2.34% (M), 4.05% (F), 5.05% (P), 6.82% (S), 5.85% (T), 1.43% (W), 3.23% (Y), and 6.64% (V). For estimating ML values, a user-specified topology was used. The maximum Log likelihood for this computation was -3464.473. The analysis involved 10 amino

acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 202 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al*., 2007).

Figure 11: Phylogenetic relationship among different plant species based on phytochelatin protein sequence using neighbour joining method.

Configuration	Count
Identical sites in all three sequences (m_{iii})	380
Divergent sites in all three sequences (m_{ijk})	11
Unique differences in Sequence A (m_{ijj})	82
Unique differences in Sequence B (m_{jji})	8
Unique differences in Sequence C (m_{iij})	19

Table 4a

Configuration	Count
Identical sites in all three sequences (m_{iii})	331
Divergent sites in all three sequences (m_{ijk})	43
Unique differences in Sequence A (m_{ijj})	57
Unique differences in Sequence B (m_{jji})	37
Unique differences in Sequence C (m_{iij})	31

Table 4b

Configuration	Count
Identical sites in all three sequences (m_{iii})	360
Divergent sites in all three sequences (m_{ijk})	11
Unique differences in Sequence A (m_{ijj})	121
Unique differences in Sequence B (m_{jji})	1
Unique differences in Sequence C (m_{iij})	5

Table 4c

Configuration	Count
Identical sites in all three sequences (m_{iii})	250
Divergent sites in all three sequences (m_{ijk})	93
Unique differences in Sequence A (m_{ijj})	69
Unique differences in Sequence B (m_{jji})	34
Unique differences in Sequence C (m_{iij})	52

Table 4d

Configuration	Count
Identical sites in all three sequences (m_{iii})	250
Divergent sites in all three sequences (m_{ijk})	65
Unique differences in Sequence A (m_{ijj})	119
Unique differences in Sequence B (m_{jji})	29
Unique differences in Sequence C (m_{iij})	35

Table 4e

Configuration	Count
Identical sites in all three sequences (m_{iii})	279
Divergent sites in all three sequences (m_{ijk})	29
Unique differences in Sequence A (m_{ijj})	175
Unique differences in Sequence B (m_{jji})	9
Unique differences in Sequence C (m_{iij})	7

Table 4f

Configuration	Count
Identical sites in all three sequences (m_{iii})	225
Divergent sites in all three sequences (m_{ijk})	83
Unique differences in Sequence A (m_{ijj})	78
Unique differences in Sequence B (m_{jji})	52
Unique differences in Sequence C (m_{iij})	45

Table 4g

Table 4: Tajima' relative test performed using phytochelatin protein sequences of different plants.

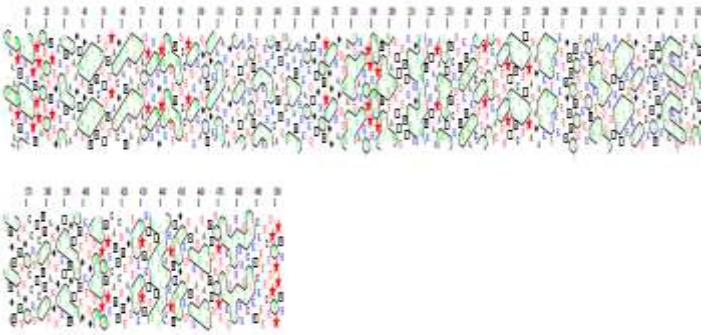


Figure 12: Hydrophobic cluster analysis of phytochelatin using Drawhca.

In the hydrophobic cluster analysis using Drawhca, it was found that there are many residues in the hydrophobic regions and their interaction with other molecules needs to be traced using a suitable methodology. Possible clusters were depicted in the graphic representation (Fig 12).The multicoil analysis showed zero probability for the possible dimer and trimer formation of this protein (Fig 13). In the RADAR analysis, totally four repeats were tracked with total score of 92.91 and 176.23. There are non aligned residues detected in the sequence during analysis. The identified repeats were significant when comparing to the total number of residues in the sequence. Table 4 depicts the target P predictions for the phytochelatin sequence of *C.dactylon*. It shows the probability of sub cellular localization site based on the presence of chloroplast peptides, mitochondrial peptides and other signal peptides. But the values are lower and moreover the reliability class value is 5 which interprets that the chance of the prediction possibility is too low. It is predicted that it could belong to any other location other than chloroplast and mitochondria.

Swiss Model analysis resulted in a structural prediction of the protein structures and a superimposed structure of various models predicted (Fig 14). Z score Q mean value for the prediction was calculated as -2.72 (Fig 15). Residue error plot was obtained for the model predicted and it shows a higher error rate between residues 50 and 100 (Fig 16). The error rate could be described more or less rhythmic from the plot.

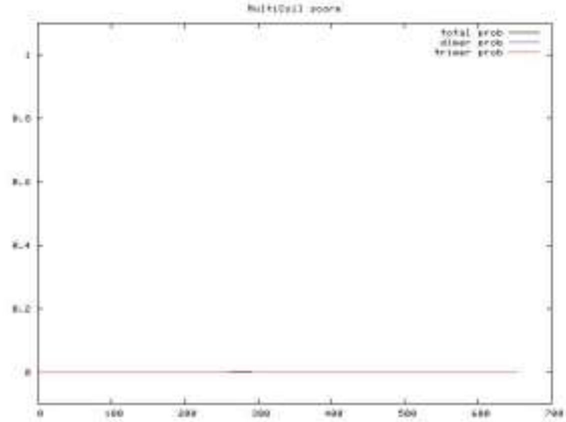


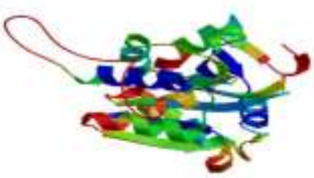
Figure 13: Multicoil score plot for dimer and trimer conformation probability.

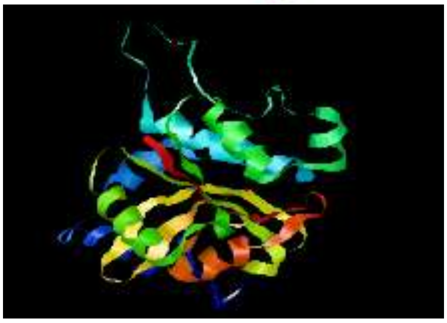
```
------------------------------------------------------------------------
No. of Repeats|Total Score|Length  |Diagonal| BW-From|   BW-To|   Level
             2|      92.91|      28|     257|     205|     236|       3
------------------------------------------------------------------------
  205-  236 |43.64/34.55)       TDESTPLLPGFMLISPHtaepAMLYTVSCPDE
  467-  494 |49.27/29.26)       TDNLPEVLKQEILHLPR....QLCYLESCREE
------------------------------------------------------------------------

------------------------------------------------------------------------
No. of Repeats|Total Score|Length  |Diagonal| BW-From|   BW-To|   Level
             2|     176.23|      57|     277|      92|     155|       4
------------------------------------------------------------------------
   92-  155 |86.02/83.16)       CCEP..LNKVKAQGTTPGHValLAhcaGADVQTPHANRITLedLPPHLVRCAS...SQDCHLIASTHGK
  372-  433 |90.21/63.75)       CCREtcFSSVQANGNGVKTV..IS...GTVVSEGHEQGVDM..LLPMSPRCASacmSEAHDEIIKYPSK
------------------------------------------------------------------------
```

Table 5: Radar analysis for the repeated templates in the sequence

Name	Len	cTP	mTP	SP	other	Loc	RC
Sequence	504	0.192	0.259	0.040	0.309	_	5
cutoff		0.000	0.000	0.000	0.000		

Table 6: TargetP prediction on protein localization





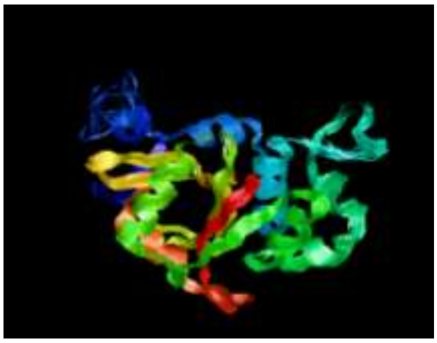


Figure 14: Swiss model prediction and superimposed model with other predicted models.

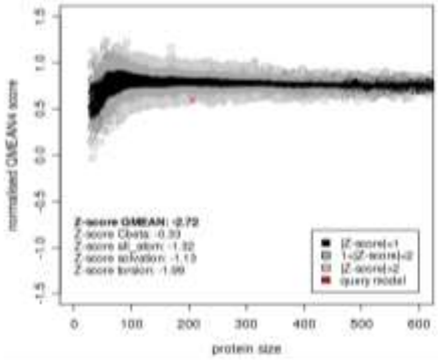


Figure 15: Comparison with non-redundant set of PDB structures and Z Score Q mean value.

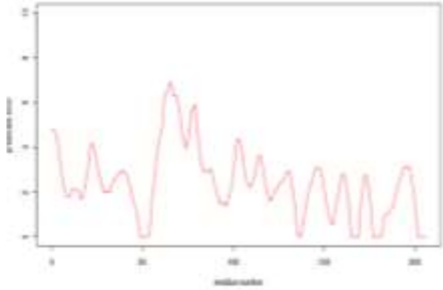


Figure 16: Predicted local errors for the protein model

Discussion

Phytochelatins are thiolate peptides with the primary structure (g-Glu-Cys)*n*-Gly, which are non-translationally synthesized from glutathione (Grill *et al*., 1989). These are expressed in plants, algae and yeast on exposure to heavy metals like cadmium. There are many structural variants found for phytochelatin but the functional variations are not much studied. Structural variations are also not so well understood though the differences in the pattern were traced to claim the variations. The cytosol and vacuoles are the prime cellular compartments where Cd-binding complexes would be located (Rauser, 1995). It has been reviewed of the various structural variants or PC-like peptides produced in plants and yeast (Inouhe, 2005). The study actual focused on the phytochelatin of C. dactylon and it was found a variant of the actual phytochelatin protein might have be expressed in the plant.

Summary

Hence this study was focused on the structural aspects of phytochelatin protein isolated from *Cynodon dactylon*. Protein sequence was retrieved from NCBI protein database with gene bank id AAO138102. Various bioinformatics online tools like GOR 4, Geno3D, ProtParam, ProtScale, TmPred, EsyPred3D analysis and GlobPlot were used to predict the structure and properties of the protein. The sequence contained 504 amino acid residues and had molecular weight of 55803.5 Daltons and

a theoretical pI value of 6.8. The formula of the protein was found to be $C_{2459}H_{3925}N_{679}O_{726}S_{37}$. The sequence was found to have high leucine residues (12.1%). Total number of negatively charged residues (Asp+Glu) and positively charged residues (Arg+Lys) was found to be 56 and 55 respectively. The protein was classified to be unstable since it has an instability index of 55.80. Aliphatic index was estimated to be 90.02 and GOR 4 analysis showed that 49.60% of residues were belonging to random coils followed by 33.53% belonging to alpha helices and 16.87% in the extended strands. Transmembrane topology studies by TmPred resulted in two suggested models. First sequence model from the Geno3D analysis by PSIBLAST was choosen and 5 models were launched. Ramachandran plots for the models and for each residue were obtained. ProtScale analysis showed variation in amino acids when analyzed using different parameters. EsyPred3D analysis resulted in a protein model which was visualized by Raswin software. Measurements were picked like length, width, length of each coils, etc. Distance between the free terminal ends residues of the protein ie, SER15 CA and SER219 CA was measured as 32.87 Å. Surface accessibility of the sequence was studied and probable amino acid residues were choosen for further interaction studies. The PC protein was shown to exhibit a different structure compared to existing ones. Since it is *In Silico* study, extensive study of the proteins must be done in better understanding of the structural variance.

References

- Bent Petersen, Thomas Nordahl Petersen, Pernille Andersen, Morten Nielsen and Claus Lundegaard. 2009. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biol.* 9(51):6807-9-51.
- Clemens S. 2001. Molecular mechanisms of plant metal tolerance and homeostasis. *Planta*. 212: 475-486.
- Cobbett C.S. 2000. Phytochelatins and their roles in heavy metal detoxification. *Plant Physiol.* 123: 825-832.
- Combet C, Jambon M, Deléage G and Geourjon C. 2002. Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics*. 18: 213-214.
- Garnier J, Gibrat J.F and Robson B.1996. GOR secondary structure prediction method version IV. *Methods in Enzymol.* R.F. Doolittle Ed. 266: 540-553.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M.R, Appel R.D and Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. 571-607.
- Grill E., Lo¨ffler S., Winnacker E.L. and Zenk M.H. 1989. Phytochelatins, the heavy-metal-binding peptides of plants, are synthesized from glutathione by a specific ά-glutamylcysteine dipeptidyl transpeptidase (phytochelatin synthase). *Proc. Natl. Acad. Sci. USA*. 86: 6838–6842.
- Grill,E., Lo¨ffler,S., Winnacker,E.-L. and Zenk,M.H. 1989. Phytochelatins, the heavy metal-binding peptides of plants, are synthesized from glutathione by a specific g-glutamylcysteine dipeptidyl transpeptidase (phytochelatin synthase). *Proc. Natl Acad. Sci. USA*.86: 6838–6842.
- Heger A. and Holm L. 2000. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins.* 41(2): 224-237.
- Jones D.T., Taylor W.R., and Thornton J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 8: 275-282.
- Lambert C, Leonard N, De Bolle X and Depiereux E. 2002. ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics.* 18(9):1250-1256.
- Lee J.G., Ahner B.A and Morel F.M.M. 1996. Export of cadmium and phytochelatin by the marine diatom *Thalassiosira weissflogii*. *Environ. Sci. Technol*. 30: 1814-1821.
- Linding Rune, Robert B. Russell, Victor Neduva and Toby J. Gibson. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31(13): 3701-3708.
- Masahiro Inouhe. 2005. Phytochelatin. *Braz. J. Plant Physiol. 17(1): 65-78.*
- Olof Emanuelsson, Henrik Nielsen, Søren Brunak and Gunnar von Heijne. 2000.
- Ortiz D.F., Kreppel L., Speiser D.M., Scheel G., McDonald G. and Ow D.W.1992. Heavy metal tolerance in the fission yeast requires an ATP-binding cassette-type vacuolar membrane transporter. *EMBO J*. 11: 3491–3499.
- Ortiz D.F., Ruscitti T., McCue K.F and Ow D.W. 1995. Transport of metal-binding peptides by HMT1, a fission yeast ABC-type vacuolar membrane protein. *J. Biol. Chem*. 270: 4721–4728.

Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.*, 300: 1005-1016,

- Rauser E. WIlfried. 1995. Phytochelatin and related peptides. *Plant Physiol.* 109: 1141-1149.
- Rauser W.E. 1999. Structure and function of metal chelators produced by plants; the case for organic acids, amino acids, phytin and metallothioneins. *Cell Biochem. Biophys*. 31:19-48.
- Tajima F.1993. Simple methods for testing molecular clock hypothesis. *Genetics.* 135: 599-607.
- Tamura K., Dudley J., Nei M and Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol.Biol.& evol.* 24: 1596-1599.
- Vatamaniuk O.K., Mari S., Lu Y.P and Rea P.A. 2000. Mechanism of heavy metal ion activation of phytochelatin (PC) synthase. *J. Biol. Chem.* 275: 31451- 31459.
- Zenk M. H.1996. Heavy metal detoxification in higher plants – a review. *Gene*. 179: 21–30.