Journal of AgriSearch 4(4):276-279
An Open Access International Peer Reviewed Quarterly
ISSN : 2348-8808 (Print), 2348-8867 (Online)
http://dx.doi.org/10.21921/jas.v4i04.10209

Metallothionein: Protein structure prediction and sequence analyses in pigeonpea (*Cajanus cajan*)

SAKSHI CHAUDHARY[1], ANIL KUMAR SINGH[2] AND JESHIMA KHAN YASIN[1*]

Division of Genomic Resources, ICAR- NBPGR, Pusa campus, New Delhi, India

ARTICLE INFO

Received on	:	18.07.2017
Accepted on	:	16.11.2017
Published online	:	30.11.2017

ABSTRACT

Metallothioneins are a special group of small proteins capable of detoxifying non-essential metal ions present in excess within a plant cell. Metallothioneins are cysteine-rich diverse classes of heavy metal binding protein molecules which are essential for plant growth. These proteins are present in all taxa, except eubacteria. The similarity in protein sequences provides a basis for the method which predicts structural features of a protein with that of a known protein structure. Structural similarity of entire sequence or large sequence fragment enables prediction and modeling of entire structural domain, while distribution of local features of known protein structure make it possible to predict such features in structure of unknown or uncharacterised proteins. In this study, from available genomic resources metallothionein of pigeonpea was identified, structure of metallothionein was predicted and validated. We have presented a step-wise methodology to model a given protein and to validate the structures.

Keywords: Bioinformatics, deficiency, legumes, pigonepea, proteins, toxicity

INTRODUCTION

Proteins constitute about 80% of the total biomolecules found in a biological system and are known as the biochemical compounds consisting of two or more polypeptides that facilitate a biological function. According to Mount (2004), for deriving the functionality involved with a certain protein, information about its three-dimensional structure is essential and for this the amino-acid (primary) sequence of the protein must be known.

The simple, linear arrangement of amino-acid residues is known to be a protein's primary structure. One of the major goals of bioinformatics is to understand the relationship between three-dimensional structure and the primary sequence of proteins; by using this, a reliable prediction of tertiary structure can be done. But, as the relationship between sequence and structure is not that simple, we are not much sure about the quality of the predicted structure. To study and perform protein modeling, we must know about the basics of protein structure. Much progress has been made in categorizing proteins on the basis of structure or sequence homology, and this type of information is very useful for protein modeling.

Plants, like other organisms, have adaptive mechanisms (Yasin and Chaudhary, 2017) whereby they are able to respond to both nutrient deficiencies and toxicities. The metallothionein super family syndicates a hugevariability of small cysteine rich proteins from nearly all phyla of life that have the ability to coordinate various transition metal ions, including Zn, Cd, and Cu. Metallothioneins are a group of low molecular mass and cysteine-rich metal-binding proteins. Small proteins are the interacting molecules activating the major signaling pathways and regulating molecules (Yasin and Seema, 2015). They play an important role in maintaining intracellular metal homeostasis, eliminating metal toxification and protecting against intracellular oxidative damages. They can bind metals via the thiol groups of their cysteine (Cys) residues (Coyle *et al.*, 2002). Since being first purified as Cd-binding protein from horse kidney (Margoshes andVallee, 1956), MTs have been widely found in diverse organisms including mammals, plants, and fungi as well as some prokaryotes (Vallee 1991; Coyle *et al.*, 2002; Cobbett and Goldsbrough, 2002).

Pigeonpea belongs to the sub-tribe Cajaninae of the agriculturally most important tribe Phaseoleae under sub-family Papilionoideae of the family Fabaceae (Leguminosae). Globally, pigeon pea is grown on 4.5 million hectares land in more than 20 countries with an annual production of 3.48 million tons (Singh, 2012). It is one of the major pulse crops of the tropics and sub tropics including India, America, Australia, Hawaii, Uganda, Italy, East and West Indies and South-East Africa. Legumes contribute from nitrogen fixation in soil to nutritional security of human (Singh *et al.*, 2016). It is an erect, branched, hairy hard shrub, 1-2 meters high. Leaves are oblong-lanceolate to oblanceolate with three leaflets (Pal *et al.*, 2011).

The first MT identified in plant was the wheat Ec-MT protein (Lane *et al.*, 1987), and since then more than 900 MT sequences have been recorded from various plant species. MTs typically contain two cysteine-rich domains in the terminal ends of the protein and have historically been divided into three classes according to the number and distribution of cysteines. Class I MTs, found in mammals and vertebrates, are characteristic for their highly conserved cysteine residues, whereas the distribution of cysteines in Class II MTs, present in plants, fungi and invertebrates are less conserved. Class III MTs include the phytochelatins, which are a family of enzymatically synthesized cysteine-rich peptides (Fowler *et al.*, 1987).

The latest and widely accepted classification based on

[2]*ICAR Research Complex for Eastern Region, Patna, India*
*Corresponding Author Email: yasinlab.icar@gmail.com

taxonomic parameters and the pattern of distribution of Cys residues in metallothioneins was given by BinzandKagi (1999), according to which all MTs were divided into 15 families and the 15th family belongs to the plant MTs. Plant MTs have further been divided into four sub-types based on their amino acid sequences. Unique structural feature of most plant MTs are long variable cysteine-free spacer region that separates the two cysteine-rich domains (Cobbett and Golds brough, 2002; Robinson *et al*., 1993). For the plant MT1, MT2, and MT3 proteins the presence of the aromatic amino acids Tyr and Phe caused their classification as MT-like.

The C-terminal region constitutes a conserved pattern of (CxCxxxCxCxxCxC), while the N-terminal region contains a characteristic arrangement of Cys residues that becomes the classification basis. The Ec proteins stick out as they contain three Cys-rich regions, all clearly separated in sequence by linker region. There are also two fully conserved His residues in the Ec subfamily, which are located close to and within the central Cys-rich region. Several biological roles have been ascribed to these MTs, including their involvement in zinc, cadmium and copper homeostasis, protection against environmental heavy metals and oxidative stress, the control of redox status of the cell, anti-apoptotic role, etc. (Yang *et al*., 2011). In many plants metallothioneins are not fully characterised. In the present investigation, we have predicted and validated the metallothionein structure of pigeon pea(*Cajanas cajan*).

The Ec proteins stick out as they contain three Cys-rich regions, all clearly separated in sequence by linker region. There are also two fully conserved His residues in the Ec subfamily, which are located close to and within the central Cys-rich region. Several biological roles have been ascribed to these MTs, including their involvement in zinc, cadmium and copper homeostasis, protection against environmental heavy metals and oxidative stress, the control of redox status of the cell, anti-apoptotic role, etc. (Yang *et al*., 2011). In many plants metallothioneins are not fully characterised. In the present investigation, we have predicted and validated the metallothionein structure of pigeonpea (*Cajanas cajan*).

MATERIALS AND METHODS

The databases and software used in carrying out this study are discussed below. Although data retrieval is the main purpose of all databases, biological databases often have a higher level of requirement, known as knowledge discovery, which refers to the identification of connections between pieces of information that were not known when the information was first entered. To have the basic information consensus sequences of plant metallothioneins were analysed.

Sequence retrieval

In order to model protein secondary structure, a pre-requisite is the primary sequence of that protein. Pigeonpea (*Cajanus cajan*) metallothionein sequence was retrieved through NCBI's life sciences search engine Entrez. The key-feature of Entrez is its ability to integrate information, which comes from cross-referencing between NCBI databases based on pre-existing and logical relationships between individual entries.

Structure modelling

The sequence was submitted in onlineprotein modeling software Modeller and confirmed with RaptorX. Further downstream sequence analyses were done using Pymol. The results obtained were downloaded in graphical form and as their atomic co-ordinate files as well. It is a new protein structure prediction program; excelling at the alignment of distantly-related proteins with sparse sequences profile and that of a single target to multiple templates. Currently, RaptorX consists of four major modules: single-template threading, alignment quality assessment, multiple-template threading and fragment-free approach to free modeling. In a structure prediction job, a protein structure is built for each of the 10 top-ranked alignments between the target sequence and the structures in the template library. The secondary structure for the sequence is also predicted. The PDB code of the template for the currently selected structure model is also provided which will take you to structure record at the Protein Data Bank http://www.pdb.org.

Structure validation

The stereo chemical model structures validation of proteins is an important part of the comparative molecular modeling process. The selection of high quality structures for inclusion in loop dictionaries is important for the simple reason that these coordinate sets will be used to build future models. The structural evaluation of comparative modeling output must be used to identify possible problematic regions. Macromolecular structure validation is the process of evaluating reliability for 3-dimensional atomic models of large biological molecules such as proteins and nucleic acids. The atomic co-ordinate file of structure was submitted in RAMPAGE: Assessment of the Ramachandran Plot for validating the structures. RAMPAGE is a web server designed to evaluate and validate protein structures.

RESULTS AND DISCUSSION

Sequence of identified metallothionein of pigeonpea

>ADD11816.1 metallothionein type 1 [*Cajanus cajan*]
MSSCGCGSSCNCGSNCGCNKYSFDLSYAEKTTTETLVLG
VGPVEAQLEGAEMGVAAENGGCNCGSSCPCDPCNCK

Protein Structure Prediction

We have focused on plant metallothioneinsin this study and derived the structure as shown in figure 1. Not much work has been done on these molecules especially when it comes to the 15th family of plant metallothioneins. Only three structures are present in RCSB PDB and are incomplete with a single domain present in the structure. The *Cajanus cajan* metallothionein sequence was submitted to protein prediction server RaptorX for predicting 3D structures of protein sequences without close homologs in the Protein Data Bank (PDB). Given an input sequence, RaptorX predicts its secondary and tertiary structures, contacts, solvent accessibility, disordered regions and binding sites. RaptorX also assigns some confidence scores to indicate the quality of a predicted 3D model: P-value for the relative global quality and modeling error at each residue. Homology modeling is useful when the target protein of known sequence is related to at least one other

protein of known structure and sequence. This relatedness must be in terms of at least 40-50% of identical residues present in an optimal alignment.

The similarity in protein sequences provides a basis for the method which predicts structural features of a protein with that of a known protein structure. Structural similarity over entire sequence or large sequence fragment enables prediction and modeling of entire structural domain, while statistics derived from distribution of local features of known protein structure enables in predicting such features in proteins with unknown structure. The accuracy of models of protein structure is important for many practical purposes.

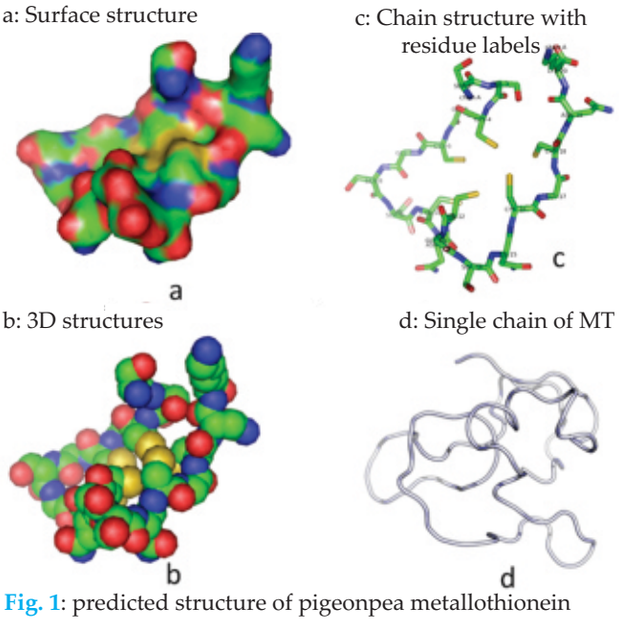

Fig. 1: predicted structure of pigeonpea metallothionein

Protein Structure Validation

We have validated the predicted structure using RAMPAGE. The aim of RAMPAGE is to assess how normal, or conversely how unusual, the geometry of the residues in a given protein structure is, as compared with stereo chemical parameters derived from well-refined, high-resolution structures.

It calculates the:

Torsion angles phi-psi combination, chi1-chi2 combination, chi1 torsion for those residues that do not have a chi-2, combined chi-3 and chi-4 torsion angles, omega torsion angles.

Covalent geometry

Main-chain bond lengths, main-chain bond angles. It is a log-odds score based on the observed distributions of these stereo chemical parameters. When applied to a given residue, a low *G*-factor indicates that the property corresponds to a low-probability conformation.

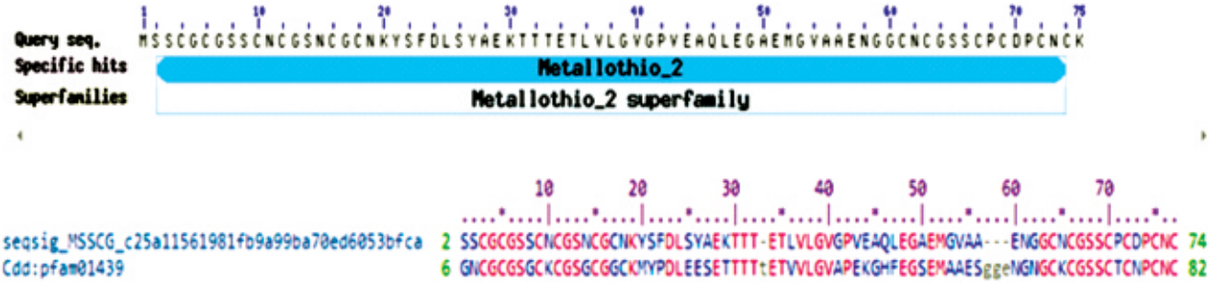

Fig. 2: Conserved domain details of pigeonpea metallothionein

Fig. 3: Predicted motif of pigeonpea metallothionein

Conserved domain analyses indicated the presence of metallothionein group of protein in the query sequence as depicted in figure 2. Motif prediction resulted in a single motif (Fig. 3). To validate the predicted protein structure of identified metallothionein, the pigeon pea predicted protein structure was analysed with Ramachandran plot. This plot revealed the structure validity as none of the residue was found to be outside the favoured region. In table 1, the plot statistics were explained with the number and type of residues.

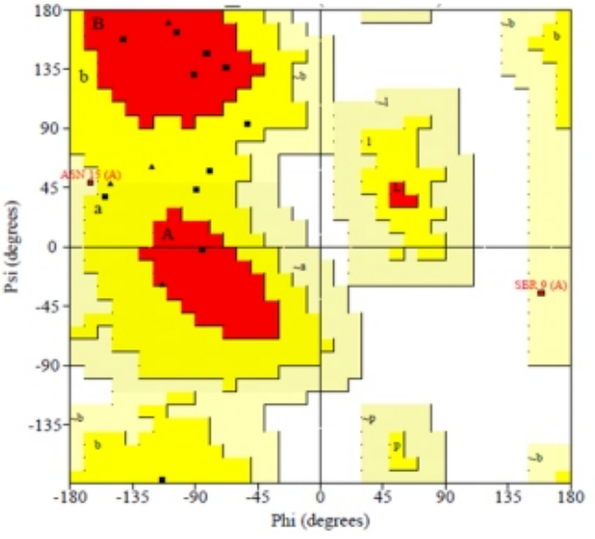

Fig. 4: Ramachandran plot analysis of predicted protein structure

Table 1:Plot statistics

Residue analysis	No. of residue	Percentage to the total
Residues in most favoured regions [A,B,L]	6	46.2%
Residues in additional allowed regions [a,b,l,p]	5	38.5%
Residues in generously allowed regions [~a,~b,~l,~p]	2	15.4%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non prolineresidues	13	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	0	
Total number of residues	19	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Here we presented the identified protein sequence, its structure and validation in this manuscript. The predicted protein structure is valid as per the results as none of the residue is found in disallowed region. The goal of structural proteomicswits is to offer template structures for utmost protein families. In many intracellular conditions presence of a DNA sequence of gene or open reading frame is insignificant as they are regulated by set of non-coding RNAs (Yasin and Chaudhary, 2016). The functional protein plays the actual role as an end product of gene action (Krzysztof *et al*., 2005). To know more about a protein its structural analyses are important. Structure prediction methodologies are meant to become limited to comparative modeling, since a close homolog of known structure would be available for most targets. Here we presented models from denovo as well as comparative modelling. This is the first report of pigeon pea metallothionein and can initiate a series of new hypotheses and findings. However, the limited success of the structural proteomics contributes to the booming interest in structure prediction methods as measured by the growing number of servers and groups taking part in community-wide prediction quality assessment experiments. But, targeted proteomics will play a major role in translational research.

CONCLUSION

It is clear that prediction methods are not expected to replace experimental determination of protein structures in the nearest future, but are likely to complement such efforts and meanwhile fill the growing gap between the number of sequences and structures. Identification of protein, predicting its structure and validation of the predicted structure are major findings of the present investigation which will be very much applicable for in silico and translational research in pigeonpea and its cross validation in related crops (Parmar *et al*., 2017). This could revolutionise crop improvement.

REFERENCES

Binz PA and Kagi JHR. 1999. Metallothionein: molecular evolution and classification In: Klaassen C, editor. Metallothionein IV. Basel: BirkhäuserVerlag, 7–13.

Cobbett C and Goldsbrough P. 2002.Phytochelatins and Metallothioneins: Roles in Heavy Metal Detoxification and Homeostasis. *Annual Review of Plant Biology* **53**: 159-182.

Coyle P, Philcox JC, Carey LC and Rofe AM. 2002. Metallothionein: the multipurpose protein. *Cell Mol Life Sci*. **59**: 627–647.

Fowler BA, Hildebrand CE, Kojima Y and Webb M. 1987. Nomenclature of metallothionein.*Experientia Suppl*. **52**:21.

Krzysztof Ginalski, Nick V Grishin, Adam Godzik and LeszekRychlewski : 2005 : Practical lessons from protein structure prediction, *Nucleic Acids Research* **33** (6):1874–1891.

Lane B, Kajoika R and Kennedy T.1987. The wheat-germ E_c protein is a zinc-containing metallothionein. *Biochem Cell Biol* **65**:1001–1005.

Margoshes M and Vallee BL.1957.A cadmium protein from equine kidney cortex.*J. Amer. Chem. Soc* **79**: 4813–4814.

Mount D.2004. Bioinformatics: Sequence and Genome Analysis. CSHL Press (2nd edition), pp. 382.

Pal D, Mishra P, Sachan N and Ghosh AK. 2011. Biological activities and medicinal properties of *Cajanuscajan* (L) Millsp. *Journal of Advanced Pharmaceutical Technology and Research* **2**(4): 207–214.

Parmar N, Singh KH, Sharma D, Singh L, Kumar P, Nanjundan J, Yasin JK, Chauhan DK and Thakur AK. 2017. Genetic engineering strategies for biotic and abiotic stress tolerance and quality enhancement in horticultural crops: a comprehensive review. *3Biotech* 7(4):239 · DOI: 10.1007/s 13205-017-0870-y

Robinson N, Tommey A, Kuske C and Jackson P.1993. Plant metallothioneins (review). *Biochem J*. 295: 1-10.

Singh AK, Meena M, Yadav R, Khan YJ. 2016Assessment of soil fertility augmenting potential of faba bean (*Vicia faba* L.) germplasm and their performance. *Journal of AgriSearch*.**3** (3):136-41.

Singh AK**.** 2012. Vegetable type pigeonpea germplasm identified and explored from Vaishali district of Bihar. *HortFlora Research Spectrum***1**(4): 312-317.

Vallee BL.1991. Introduction to metallothionein.*Methods Enzymol*.**205**: 3-7.

Yang JL, Wang YC, Liu GF, Yang CP, Li CH.2011. *Tamarixhispida* metallothionein-like ThMT3, a reactive oxygen species scavenger, increases tolerance against Cd2+, Zn2+, Cu2+, and NaCl in transgenic yeast. *MolBiol Rep***38**: 1567–1574.

Yasin JK and Chaudhary S. 2017.Giving no quarter: Habitat, biotic and abiotic factors smash plant diversity and density.Science.http://science.sciencemag.org/content/356/6345/1389/tab-e-letters

Yasin JKand Seema KM. 2015. Role of rabgtpases and its interacting molecules in stress tolerance. *International Journal of Science and Research***4**(6): 1228-1232.

Yasin JK, Chaudhary S. 2016. Non-coding RNAs regulate quiescence: plays major role in stress tolerance of plants. *Science*.http://science. sciencemag.org/content/early/2016/10/12/science.aah5651/tab-e-letters

Citation:

Chaudhary S, Singh AK and Yasin JK. 2017. Metallothionein: Protein structure prediction and sequence analyses in pigeon pea (*Cajanus cajan*). *Journal of AgriSearch* **4** (4): 276-279