

12. Chan, F. K. *et al.*, Preventing recurrent upper gastrointestinal bleeding in patients with *Helicobacter pylori* infection who are taking low-dose aspirin or naproxen. *N. Engl. J. Med.*, 2001, **344**, 967–973.
13. Lai, K. C. *et al.*, Lansoprazole for the prevention of recurrences of ulcer complications from long-term low-dose aspirin use. *N. Engl. J. Med.*, 2002, **346**, 2033–2038.
14. Noda, S., Kitazono, E., Tanaka, T., Ogina, C. and Kondo, A., Benzoic acid fermentation from starch and cellulose via a plant-like  $\beta$ -oxidation pathway in *Streptomyces maritimus*. *Microb. Cell Fact.*, 2012, **11**, 49–58.
15. *Ayurvedic Pharmacopoeia of India*, Part II, Vol II, Monographs, pp. 70–73.
16. Anthon, G. E. and Barrett, D. M., Colorimetric method for the determination of lipoxygenase activity. *J. Agric. Food Chem.*, 2003, **49**, 32–37.
17. Baraniak, B. and Krzepiko, A., Inhibition of broccoli lipoxygenase by some phenolic compounds – a short report. *Pol. J. Food Nutr. Sci.*, 2004, **13**, 339–342.
18. Delfini, C., Gala, P., Bardi, L., Mariscalco, G., Contiero, M. and Pagliara, A., Production of benzaldehyde, benzyl alcohol and benzoic acid by yeasts and *Botrytis cinerea* from grape musts and wines. *Vitis*, 1991, **30**, 253–263.
19. Gonthier, M. P., Rios, L. Y., Verny, M., Remesy, C. and Scalbert, A., Novel liquid chromatography-electrospray ionization mass spectrometry method for the quantification in human urine of microbial aromatic acid metabolites derived from dietary polyphenols. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2003, **789**, 247–255.
20. Tomy, M. J., Dileep, K. V., Preethidan, D. S., Prasanth, S., Sabu, A., Sadasivan, C. and Haridas, M., Cuminaldehyde as a lipoxygenase inhibitor: *in vitro* and *in silico* validation. *Appl. Biochem Biotechnol.*, 2014, **174**(1), 388–397.
21. Tomy, M. J., Sharanya, C. S., Prasanth, S., Sabu, A., Sadasivan, C. and Haridas, M., Derivatives form better lipoxygenase inhibitors than piperine: *in vitro* and *in silico* study. *Chem. Biol. Drug Design.*, 2015, **85**(6), 715–721.

**ACKNOWLEDGEMENTS.** We gratefully acknowledge the Institute for Intensive Research in Basic Sciences (IIRBS) for NMR facility and Inter University Center for Instrumentation for LC-MS at Mahatma Gandhi University.

Received 8 September 2015; revised accepted 21 July 2016

doi: 10.18520/cs/v111/i11/1836-1838

## ***In silico* characterization of ferritin-1 chloroplast targeting peptide encoding sequence in chickpea and pigeonpea**

**Poornima K. Narayana\*, Alok Das and Narendra P. Singh**

Division of Plant Biotechnology, Indian Institute of Pulses Research, Kanpur 208 024, India

**Chloroplast is an important organelle and a hub of photosynthetic reactions that occur in all plants. Majority of the chloroplast proteins are encoded by the nucleus and then transported into the chloroplast. The chloroplast transport is aided by leader sequence, chloroplast targeting peptide (cTP) that directs the proteins into the chloroplast. We identified the genes encoding chloroplast targeting proteins from chickpea (*Cicer arietinum*) and pigeonpea (*Cajanus cajan* and *Cajanus cajanifolius*) genomes. All the sequences were aligned, conserved motifs were identified and their phylogeny was deciphered. In this study, structural and functional features of chloroplast transit peptides of Ferritin-1 and its putative role in chloroplast targeting have been examined. The study would impart better understanding of targeting heterologous proteins to chloroplasts in pulses.**

**Keywords:** Chloroplast targeting peptide, domains, ferritin-1, motifs, transgenes.

THE chloroplast is the site of photosynthesis that supports the life of most of the living organisms. Chloroplast targeting has emerged as an important tool for heterologous expression of transgenes in plants. The targeting of transgenes into the chloroplast by using a suitable strategy with an efficient expression cassette containing all the required elements has been the need of the day. Majority (90%) of the 3000 different proteins necessary for fully functional chloroplasts are known to be encoded by nuclear DNA. These proteins are synthesized in their precursor forms with an amino-terminal signal peptide called the transit peptide<sup>1</sup>. A signature peptide sequence (chloroplast target peptide) drives protein into chloroplast and gets cleaved off after reaching target destination, delivering the protein to its destined cellular compartment<sup>2</sup>. Chloroplast targeted expression of genes such as anthranilate synthase for increasing tryptophan production in maize has been achieved using novel constructs encoding chloroplast transit peptide<sup>3</sup>. Looking into the importance of cTP sequences in localizing transgenes, this study is based on the isolation of three cTP candidate sequences from chickpea and pigeonpea genome and their structural characterization.

\*For correspondence. (e-mail: poornimakn04@gmail.com)

**Table 1.** Ferritin 1 transit peptide sequences of different crops retrieved from Uniprot database and the Ferritin 1 chloroplast transit peptide sequence isolated from chickpea and pigeonpea

Peptide name	Uniprot ID	Transit peptide sequence
Ferritin 1 – <i>Glycine max</i>	P19976	MALAPSKVSTFSGFSPKPSVGGGAQKNPTCSVSLSFLNEKLGSRNLRVCA
Ferritin 3 – <i>Vigna unguiculata</i>	O65100	MALSCSKVLTFSLSSVVGDDAKKKLSLCSSSSLASVNGGSRNMRVCAAASN
Ferritin 1 – <i>Pisum sativum</i>	P19975	MALSSSKFSSFGFSLSPVSGNGVQKPCFCDLRVGEKWGSRKFRVSA
Ferritin – <i>Malus baccata</i>	Q94FY2	MALAPSKVSTFSGFSPKPSVGGGAQKNPTCSVSLSFLNEKLGSRNLRVCA
Ferritin – <i>Phaseolus vulgaris</i>	P25699	MALAPSKVSPFSGFSLSDGVGAVRNPTCSVSLSFLNKKVGSRNLRVSA
Ferritin 1 – <i>Arabidopsis thaliana</i>	Q39101	MASNALSSFTAANPALSPKPLLPHGASPSVSLGFSRKGGRVAVVVA
Ferritin 1 – <i>Brassica napus</i>	Q96540	MASKALSSFTAAPAVSLLPHGVSSASPSVMSLSFSRHTGGRGVVAA
Ferritin 1 – <i>Nicotiana tobaccum</i>	Q8RX97	MLLKAAPAFALLNTQGENLSPLFSSSKSFSFKNGRNFVVSASKAT
Ferritin 1 – <i>Zea mays</i>	P29036	MMLRVSPSPAAA VPTQLSGAPATPAPVVRVAAPRGVVASPAGAACR
Ferritin 1 – <i>Cajanus cajan</i> (Asha)		MALFSSSKVSSFGFSLSPISGNLKKPTLSFCSCCLR VNEKWGSRKLSVCAATTVP LV GVIFEPFEEVKKDNLA VPIV
Ferritin 1 – <i>Cajanus cajanifolius</i> (ICP 15629)		MALFSSSKVSSFGFSLSPISGNLKKPTLSFCSCCLR VNEKWGSRKLSV CAAILSSFNWAIYLNHLRRIRRIILLFQLCX
Ferritin 1 – <i>Cicer arietinum</i> (DCP 92-3)		MALFSSSKVSSFGFSLSPISGNLKKPTLSFCSCCLR VNEKWGSRKLSV CAATTVP LTVGIVIFEPFEEVKKDNLA VPIV

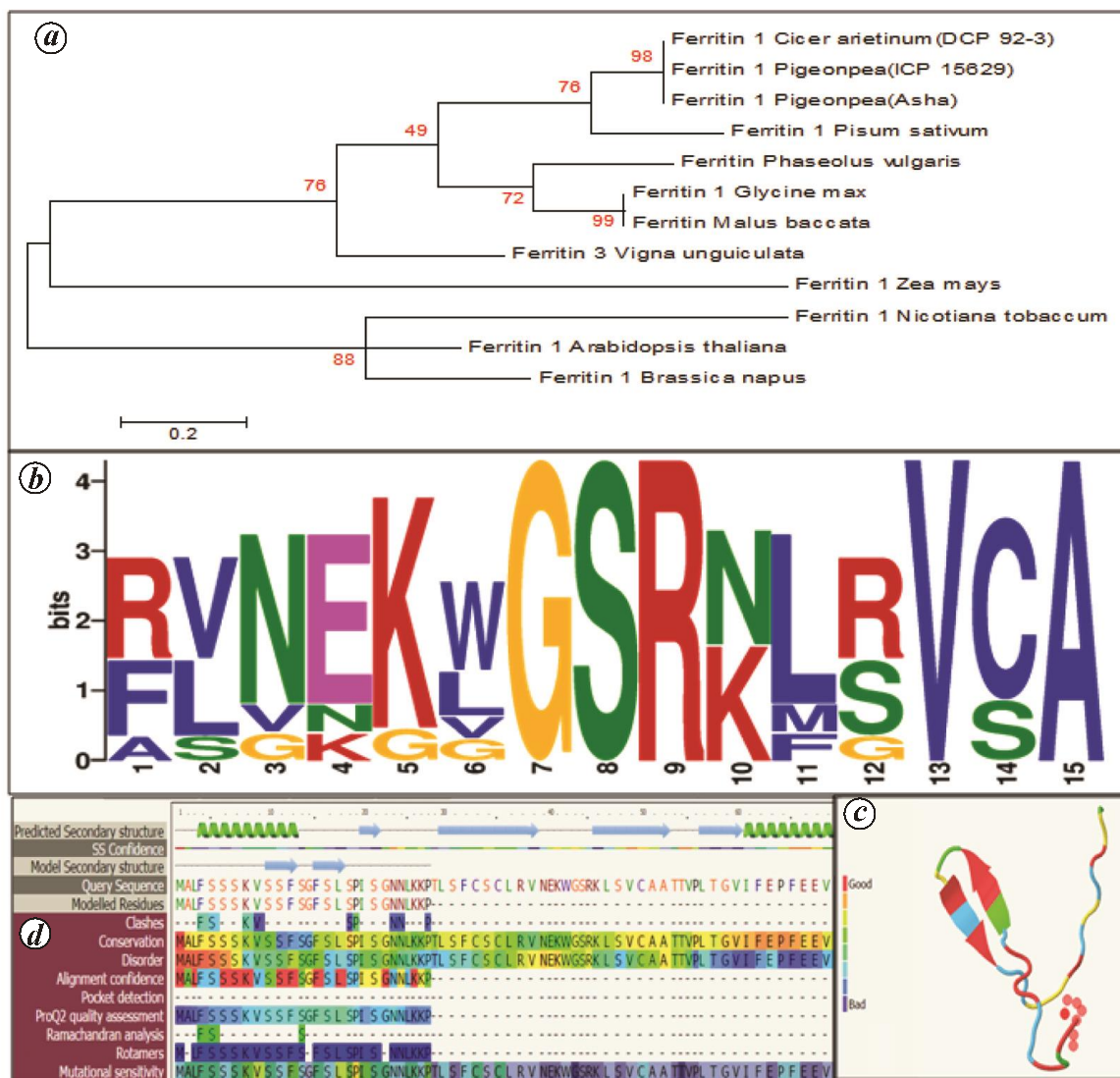
Chloroplast targeting proteins along with the predicted transit peptide sequences were retrieved from Uniprot (<http://www.uniprot.org/>). Heterologous transit peptide coding sequences of chloroplast targeted proteins from chickpea and pigeonpea genomes were retrieved from NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

The sequences obtained were translated and the peptides predicted for the presence of putative plastid targeting sequence using Predotar<sup>4</sup> (<https://urgi.versailles.inra.fr/predotar/test.seq>). The sequences were also subjected to cleavage site predictions using ChloroP<sup>5</sup> (<http://www.cbs.dtu.dk/services/ChloroP/>) and TargetP<sup>6</sup> (<http://www.cbs.dtu.dk/services/TargetP/>). Sequence alignment and phylogeny analysis was performed using Bioedit and MEGA6 (ref. 7) using Neighbour-Joining algorithm and a 1000 bootstrap value. Conserved motifs in cTPs were identified using MEME<sup>8</sup> (<http://meme.nbcr.net/meme/cgi-bin/meme.cgi>). The presence of secondary structures was predicted using the Jpred4 (ref. 9) (<http://www.compbio.dundee.ac.uk/jpred4>). The analysis of residues found within the peptide sequence was done using the Phyre2 (ref. 10) (<http://www.sbg.bio.ic.ac.uk/phyre2>).

Most proteins involved in biosynthetic pathways of chloroplast are encoded by nuclear genes, made in the cytoplasm and localized in the chloroplast. Precise targeting of these proteins is essential for rendering their biosynthetic function. An N-terminal extension called the chloroplast targeting peptide (cTP) helps in this precise targeting of the proteins into the chloroplast. Bacterial genes when expressed in plants require to be attached to these cTP encoding oligos at the 5' region. On average, the length of plastid transit peptides is 50 amino-acids but it varies between 13 and 146 amino acids<sup>11</sup>. The transit peptide encoding sequences from chickpea and pigeonpea translate to about 61 amino acids lying within the expected range (Table 1).

The chloroplast targeting peptide encoding sequences from chickpea and pigeonpea ferritin 1 gene were found to be highly conserved and closely associated with cTP from *Pisum sativum*. No sequence variation was found among the cTPs from wild and cultivated pigeonpea genotypes (Figure 1a). The Predotar software has predicted the presence of chloroplast targeting sequence in our submitted sequence with a probability of 0.63–0.69 showing a 70% likelihood of the isolated sequences to be targeted into chloroplast (see [Supplementary Information Table S1 online](#)). ChloroP 1.1 predicted presence of cTP in the sequences with a significant score of 2.75 for peptide cleavage site (CS) (Table 2). A score above 2 for CS represents the presence of authenticated CS in the test sequence. TargetP 1.1 predicted chloroplast localization of the isolated sequences with a high reliability class score of 2 (see [Supplementary Information Table S2 online](#)). The prediction of cTPs with above software individually was found to be of low accuracy than using a combination of cTP predictors as demonstrated in *Arabidopsis*<sup>12</sup> with 2,450 proteins.

Initially, the transit peptide was thought to be enriched by serine and threonine<sup>13</sup>, but later they found the most abundant residue to be serine in *Arabidopsis* and alanine in rice. Serine has been found abundantly in the chickpea and pigeonpea ferritin 1 gene transit peptide and there are no threonine residues. The ferritin gene cTPs contain a conserved 'homology block' Gly-X-Arg-XXX-Val close to the cleavage or processing site (Figure 1b). This study has found many residues of serine and few of phenylalanine with no arginine in the N-terminal region of the transit peptide sequence. A low content of arginines, together with a high abundance of proline and serine in the N-terminal portion, would suggest a chloroplast targeting domain<sup>14</sup>. The three motifs identified in the study correlate with the common structure and modular organization of cTPs as described earlier<sup>15</sup>. The first motif



**Figure 1.** Sequence characterization of chloroplast targeting peptide. *a*, Phylogenetic tree representing the relatedness of cTP from different crops including legumes constructed using neighbour-joining method; *b*, Conserved 'homology block' found in the motif predicted at the cleavage site of cTP; *c*, Protein structure representing the alignment confidence (AC) of cTP with modelled protein (red colour represents a good AC and blue colour represent bad AC); *d*, Protein folding characteristics defined based on predicted secondary structures.

constitutes N-terminal residues beginning with M/A- and terminating with G/P, a second central motif lacking acidic residues but enriched in S/T and a C-terminal motif enriched in arginine have been observed. One  $\alpha$ -helix and three  $\beta$ -sheet structures were predicted in the C-terminal region of the sequence beyond the position of 45 amino acids. The analysis of secondary structures depicted that out of 79 residues, 45 (57%) are part of repeating secondary structures (helices and beta-sheets) and the remaining residues are involved in making H-bonds, to turns, bends or irregular structures. Absence of membrane interacting structures with a minimal content of regular secondary or higher order structures suggest cTPs

to be flexible peptides with relative tolerance to deletions and insertions in their central region. The ferridoxin 1 transit peptides contain two discontinuous alpha helical domains and are amphipathic in nature determined by the presence of hydroxylated amino acids (Ser, Thr) in accordance with the cTP properties reported. The Phyre2 analysis of the transit peptide sequence showed a N-terminal similarity hit by a transferase protein in the PDB database (Figure 1 *c*). The modelled residues were found to be present in the allowed regions of Ramachandran plot with a fairly good alignment confidence representing the reliability of pair-wise query-template alignment. A high degree of disordered regions were found in the

**Table 2.** ChloroP prediction of Ferritin 1 transit peptide sequences: Y represents presence of cTP (chloroplast transit peptide) of 61 amino-acids in the peptide sequence with a significant cleavage site (CS) score

Peptide name	ChloroP prediction				
	Length	Score	cTP	CS-score	cTP-length (amino acid)
Ferritin 1_Cp* (DCP 92-3)	236	0.55	Y	2.75	61
Ferritin 1_Pp* (Asha)	236	0.55	Y	2.75	61
Ferritin 1_Pp* (ICP 15629)	238	0.55	Y	2.75	61

\*Cp – Chickpea, Pp – Pigeonpea.

N-terminal with ordered peptides in the C-terminal region of the cTP. This shows that the N-terminal regions are highly flexible and are distinct from irregular loop structures that are static in solution. Such flexible regions hinder protein crystallization supporting the role of cTPs in protein transport and not in the formation of functional domain of proteins. The modelled peptides were found to be devoid of active sites represented by none of the pockets detected in the functional analysis (Figure 1d). The cTP sequence also rendered a moderate to low mutational sensitivity showing that the cTPs are mostly conserved for their function in protein transport into chloroplast<sup>10</sup>.

Previous studies have shown that ribulose biphosphate carboxylase small subunit promoter-transit peptide sequence (*rbcS-tp*) system facilitates the accumulation of the cry1A protein in chloroplast at a level equivalent to 2% of the total soluble proteins in transgenic crops<sup>16</sup>. The cTP sequences isolated from chickpea and pigeonpea genomes in this study could be used for targeting the genes with novel traits into the chloroplast representing a high-level expression system. Further, foreign proteins expressed at very low levels or toxic to the cytoplasm can be localized to intracellular compartments of chloroplast by fusing with cTP. It can also facilitate the genome editing of chloroplast genome and targeting of different cellular proteins into the chloroplast with higher and stable expression.

- Emanuelsson, O., Nielsen, H., Brunak, S. and Heijne, G. V., Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.*, 1997, **10**, 1–6.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution*, 2013, Mst 197.
- Bailey, T. *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucl. Acids Res.*, 2009, **37**, W202–W208.
- Drozdetskiy, A., Cole, C., Procter, J. and Barton, G. J., JPred4: a protein secondary structure prediction server. *Nucl. Acids Res.*, 2015, 10.1093/nar/gkv332.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Michael, J. E., The Phyre2 web portal for protein modelling, prediction and analysis. *Nat. Protocols*, 2015, **10**, 845–858.
- Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O. and Sun, Q., Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE*, 2008, **3**, e1994.
- Richly, E. and Leister, D., An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice. *Gene*, 329, 11–16.
- Zhang, X. P. and Glaser, E., Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. *Trends Plant Sci.*, 2002, **7**, 14–21.
- Bhushan, S., Kuhn, C., Berglund, A. K., Roth, C. and Glaser, E., The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. *FEBS Lett.*, 2006, **580**, 3966–3972.
- Bruce, B. D., Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol.*, 2000, **10**, 440–447.
- Kim, E. H. *et al.*, Chloroplast-targeted expression of synthetic cry1Ac in transgenic rice as an alternative strategy for increased pest protection. *Planta*, 2009, **230**, 397–405.

- Sakamoto, W., Miyagishima, S. and Jarvis, P., Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. *Arabidopsis Book*, 2008, **6**, e0110.
- Manjunath, S., Navarro, S. X., Rapp, W. D., Shi, X., Varagona, M. J., Winson, J. L. and Ye, G., Production of high tryptophan maize by chloroplast targeted expression of anthranilate synthase. US Patent and Trademark Office, Washington, DC, 2015, US Patent No. 8981188.
- Mace, E. S., Buhariwalla, H. K. and Crouch, J. H., A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol. Biol. Rep.*, 2003, **21**, 459a–459h.
- Small, I., Peeters, N., Legeai, F. and Lurin, C., Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, 2004, **6**, 1581–1590.
- Emanuelsson, O., Nielsen, H. and Von, H. G., ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.*, 1999, **5**, 978–984.

**ACKNOWLEDGEMENTS.** We thank our colleague Aravind K. Konda for his valuable inputs in preparation of the manuscript. This research was supported by Indian Council of Agricultural Research, Indian Institute of Pulses Research, Kanpur, India.

Received 30 November 2015; revised accepted 5 July 2016

doi: 10.18520/cs/v111/i11/1838-1841