Ecotoxicology and Environmental Safety 163 (2018) 7–18

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

# *In silico* prediction, phylogenetic and bioinformatic analysis of *SoPCS* gene, survey of its protein characterization and gene expression in response to cadmium in *Saccharum officinarum*

Maryam Kolahi[a,*], Milad Yazdi[b], Andrea Goldson-Barnaby[c], Mohammad Reza Tabandeh[d]

[a] Department of Biology, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[b] Department of Genetic, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[c] Department of Chemistry, University of the West Indies, Mona, Jamaica
[d] Department of Biochemistry and Molecular Biology, Faculty of Veterinary Medicine, Shahid Chamran University of Ahvaz, Ahvaz, Iran

ARTICLE INFO

*Keywords:*
3D structure
Cadmium
*SoPCS*
Sugarcane

ABSTRACT

Phytochelatin synthase isolated from microorganisms, yeasts, algae and plant, serve a fundamental role in reducing heavy metals. In this research the *in silico* PCS gene structure (*SoPCS*) of sugarcane, its secondary and 3D protein structure, physicochemical properties, cell localization and phylogenetic tree were predicted utilizing bioinformatics tools. *SoPCS* expression in the leaves and roots of sugarcane in tissue culture treated with cadmium was also studied utilizing real time PCR. The predicted *SoPCS* gene contains 1524 nucleotides, a protein encoded with 508 amino acids of which the molecular weight is 55953.3 Da, 6 exons and 5 introns. The subcellular position of the enzyme is mitochondrion or cytoplasmic. Two domains belonging to the phytochelatin synthase family with similar features was found in Pfam having more than 97% similarity with the predicted *SoPCS* protein. Phylogeny analyses of plant species were well isolated from other organisms. Ten disulfide-bonded cysteines were excluded from the structure of *SoPCS*. The predicted 3D structure of *SoPCS* showed that it is able to bind to L-gamma-glutamylcysteine as substrate. The binding site sequence of PCS included amino acids 52(Q),55(P),56(A),57(F), 58(C),103(G),104(I),151(S),163(G),165(F),206(D), 213(R). The common amino acid with conserved sequence in the binding site of the plant was 103Gly. Gene expression indicated that *SoPCS* has an important role in the response of sugarcane to cadmium with potential use in genetic engineering to remove metal contaminants in the environment. This is the first characterization of a PCS from sugarcane.

## 1. Introduction

Phytochelatin synthase (PCS) proteins are also referred to as γ-glutamylcysteine dipeptidyl transpeptidase (EC 2.3.2.15). These enzymes are able to synthesize phytochelatins. Phytochelatin synthase is a member of the papain superfamily of cysteine proteases which mechanistically deglycinates glutathione (GSH) via a catalytic triad comprised of Cys-His-Asp. PCS are thiol-containing peptides formed in post-translational synthesis from glutathione, and are made from glutamine, cysteine and glycine. PCS enzymatically synthesizes peptides and are not gene-encoded (Merlos et al., 2014). Studies on the physiological, biochemical and genetic properties of this enzyme have revealed that the substrate for PCS biosynthesis is GSH (Mehra et al., 1994; Rigouin et al., 2013; Zayneb et al., 2017). Various studies on yeast and Arabidopsis have established that mutants which are GSH-deficient are also PC deficient with hypersensitivity to cadmium (Ha et al., 1999). A structural feature of PCS is that of progressively increasing repetitions of the γ-Glu-Cys dipeptide followed by a terminal Gly tail; $(\gamma\text{-Glu-Cys})_n$-Gly, with *n* ranging between 2 and 11, but usually not exceeding more than five. With the exception of glycine, Ser, Glu, Gln and Ala can also be found on the C-terminal of $(\gamma\text{-Glu-Cys})_n$ peptides in certain plant species with the same function from essentially similar biochemical pathways (Vatamaniuk et al., 1999).

Research has found that the distribution of the *PCS* gene between organisms is much more than what was initially thought (Clemens, 2006). The common structure of PCS has been found in a variety of microorganisms, yeasts, algae and plants. Some researchers have recognized similar genes in several invertebrates and many animals, but vertebrates do not have *PCS* genes. In the nematode, *Caenorhabditis elegans*, the expected amino acid sequence of the N-terminal region displayed approximately 50% identical amino acids with the conforming regions of plant and yeast gene products (Cobbett, 2000). PC

* Corresponding author.
*E-mail address:* m.kolahi@scu.ac.ir (M. Kolahi).

https://doi.org/10.1016/j.ecoenv.2018.07.032
Received 6 February 2018; Received in revised form 2 July 2018; Accepted 8 July 2018

synthase activities have been identified in pea (Klapheck et al., 1995), tomato (Chen et al., 1997), Arabidopsis (Howden et al., 1995), date palm (Zayneb et al., 2017), algae (Gekeler et al., 1989), as well as a variety of fungal species (Mehra et al., 1994) and marine diatoms (Morelli and Pratesi, 1997).

Over 200 individual plants have been studied and all of them without exception possess PCS. In plants PC–Cadmium complexes are confined to the vacuole (Ha et al., 1999). Increased production of heavy metals following industrialization and a rise in industrial pollutants has led to environmental problems. Production of metal-binding ligands such as metallothioneins and phytochelatins is vital to regulate cellular uptake as well as responding to metal accumulation. Research has shown that phytochelatins introduced as proteins result in metal detoxification. Following exposure to various metal ions such as cadmium, phytochelatins were immediately produced resulting in adaptation of the metal-sensitive organism which became tolerant. Organisms with a deficiency in PCS are hypersensitive to cadmium and are unable to tolerate cadmium toxicity.

Other than their role in metal detoxification, PCS proteins also play a role in xenobiotic metabolism and the detoxification of plants and fungi. Recent studies showed that phytochelatin synthase isolated from rice, played a fundamental role in decreasing the levels of arsenic in rice grains which is a major challenge in agriculture (Uraguchi et al., 2017). Some research has indicated that PCS also plays a role in the metabolism of glutathione conjugates which is plant specific (GS-conjugates) (Blum et al., 2007). Phytochelatins have recently received attention as a potential drug target which could be utilized in the treatment of Schistosomiasis disease, as *Schistosoma mansoni* the agent of the disease has PCS, which have been able to defend against the toxic effects of metal in this particular parasite (Rigouin et al., 2013).

*Saccharum officinarum* belongs to the Poaceae or Gramineae family providing approximately 70% of the world's sugar. Commercially, it is cultivated as a source of sucrose and is also the best source of biofuels (Hamdi, 2016). Extensive and industrial cultivation of sugarcane in several countries on the basis of high performance and incredible biomass encourages the use of this plant as a phytoremediation plant. Presently there is no information on the gene structure of PCS in sugarcane. In this paper we have predicted the *in silico* PCS gene structure (*SoPCS*) in sugarcane which has not been previously reported, as well as investigated the secondary and 3D protein structure, enzyme features such as physicochemical properties, and its localization in the cell and position in the phylogenic tree.

The study of PCS in sugarcane is highly valuable because sugarcane is widely cultivated in some countries and it has excellent characteristics for phytoremediation including high biomass, high yield, rapid growth and growth potential in contaminated areas. Also sugarcane as a model of a C4 plant was the most energy efficient C4 perennial plant. Proper cultivation, extensive cultivation, vegetative characteristics and the uptake of heavy metals by sugarcane suggests that it can potentially be a good candidate for the removal of heavy metals from contaminated environments. This plant is high yielding and has an extensive root system that is highly branched. Root depths in excess of 6 m have been reported. Sugarcane production continues to increase in Iran.

## 2. Material and methods

### 2.1. Propagation and cadmium exposure

Commercial sugarcane samples (cultivar CP48–103) were collected from the Karoon culture and Industrial Co of Shushtar. The terminal buds were chosen as explants, after being washed under running water (1 h), 70% ethanol (1 min) and 20% NaClO for sterilization (20 min). Explants were allowed to grow for 12 weeks on modified MS media supplemented with kinetin (2 mg/l), 1-naphthaleneacetic acid (0.05 mg/L) and gibberellic acid (0.15 mg/L). The medium contained sucrose (30 g/L) and agar (7% (w/v) at a final pH of 5.7. Explants were incubated at a temperature of 25 ± 2 °C with a photoperiod of 16/8 h. The shoots were cultured in MS medium containing indole-3-acetic acid (0.1 mg/L) and 6-benzylaminopurine (0.07 mg/L) for 75 days. Subcultures were prepared every two weeks. After propagation, samples were cultured in triplicate for 14 days using modified MS medium containing 0 and 150 μmol $CdCl_2$ and then incubated at a temperature of 25 ± 2 °C with a photoperiod of 16/8 h. Plants were harvested in 3 replicates, with shoot and root samples being taken from each plant. For molecular studies, samples were diced into smaller segments, placed in polypropylene tubes (50 mL) and immediately frozen with liquid nitrogen.

### 2.2. Prediction of SoPCS gene

By using nucleotide blast and the CDS portion of "NM_001175170.1 *Zea mays* glutathione gamma-glutamylcysteinyltransferase 2 (LOC100382428), mRNA" as query and two database "whole genome shotgun containing (WGS)" and "transcriptome shotgun assembly (TSA)" limited by "Saccharum (taxid: 4546)" both putative transcript, the full gene sequence of *SoPCS* was deduced. The alignment of these sequences was acquired by utilizing Mega 7. Gene structure was drawn using "GSDS2" software (Mascher and Stein, 2014; Hu et al., 2014). Physicochemical characteristics of PCS were deduced from ProtParam (Gasteiger et al., 2003). The predicted protein sequence was accessed utilizing the translation option of mega 7 and by searching the Pfam database for similar domain (Kumar et al., 2008; Finn et al., 2013). The signal peptide and the cell compartment that PCS is expected to act was predicted by the use of several software including kSignalP 4.1 Server, Signal-3 L 2.0, PrediSi, Psort II, MitoProt and DeepLoc-1.0.3d (Claros, 1995; Hiller et al., 2004; Horton et al., 2007; Shen and Chou, 2007; Petersen et al., 2011; Almagro Armenteros et al., 2017; Zhang and Shen, 2017).

### 2.3. Phylogenetic analysis of the SoPCS genes and protein sequence collection

The phylogeny of PCS proteins was assessed from 98 sequences, representing all major phylogenetic lineages. For phylogenetic analysis, PCS enzyme sequences were extracted from the NCBI protein database using advanced search keywords "phytochelatin synthase", "PCS" and "Glutathione gamma-glutamylcysteinyltransferase" so as to locate them in bacteria, cyanobacteria, fungi and plants. All the sequences found for plants, some bacteria, fungi and algae were used to assemble a phylogenetic tree. Actinobacteria was used as an outgroup. The Jones-Taylor-Thornton model was selected to obtain the phylogenies using the Maximum Likelihood (mL) method. Multiple sequence alignments were performed by utilizing the muscle algorithm of mega 7 software to detect conserved residues (Kumar et al., 2016).

### 2.4. Secondary and 3D structure of SoPCS

Secondary and 3D structures, binding site and ligand binding site were predicted using COACH, I-TASSER and phyre 2 softwares (Yang et al., 2012; Kelley et al., 2015). Disulfide connectivity was predicted by utilizing the DiANNA software (Ferrè and Clote, 2005). By use of the profile and model of the secondary structure, a full length 3D model was generated. The resulting conformation was visualized using phyre 2 (Wass et al., 2010; Kelley et al., 2015). Conserved regions of plants were visualized by ccp4mg (McNicholas et al., 2011).

### 2.5. RNA extraction, Reverse transcription and Quantitative real time PCR

Total RNA was extracted from frozen samples utilizing YTzol (Pure RNA isolation reagent) (Yekta Tajhiz Azma Co., Iran) and stored at − 80 °C. First-strand cDNA was synthesized utilizing RNA (8 μg), Oligo-dt primer and M-MLV reverse transcriptase. The PCR reaction was set