

Predicted metal binding sites for phytoremediation

Ashok Sharma*, Sudeep Roy, Kumar Parijat Tripathi, Pratibha Roy, Manoj Mishra, Feroz Khan, Abha Meena

Bioinformatics and *In-Silico* Biology, Central Institute of Medicinal and Aromatic Plants (Council of Scientific and Industrial Research), P.O Lucknow (226015), India.; Ashok Sharma - Email: ashoksharma@cimap.res.in; * Corresponding author

Received March 05, 2009; Revised April 15, 2009; Accepted July 18, 2009; Published September 05, 2009

Abstract:

Metal ion binding domains are found in proteins that mediate transport, buffering or detoxification of metal ions. The objective of the study is to design and analyze metal binding motifs against the genes involved in phytoremediation. This is being done on the basis of certain pre-requisite amino-acid residues known to bind metal ions/metal complexes in medicinal and aromatic plants (MAP's). Earlier work on MAP's have shown that heavy metals accumulated by aromatic and medicinal plants do not appear in the essential oil and that some of these species are able to grow in metal contaminated sites. A pattern search against the UniProtKB/Swiss-Prot and UniProtKB/TrEMBL databases yielded true positives in each case showing the high specificity of the motifs designed for the ions of nickel, lead, molybdenum, manganese, cadmium, zinc, iron, cobalt and xenobiotic compounds. Motifs were also studied against PDB structures. Results of the study suggested the presence of binding sites on the surface of protein molecules involved. PDB structures of proteins were finally predicted for the binding sites functionality in their respective phytoremediation usage. This was further validated through CASTp server to study its physico-chemical properties. Bioinformatics implications would help in designing strategy for developing transgenic plants with increased metal binding capacity. These metal binding factors can be used to restrict metal uptake by plants. This helps in reducing the possibility of metal movement into the food chain.

Keywords: Phytoremediation; medicinal and aromatic plants (MAPs); putative metal binding sites

Background:

Phytoremediation [1] can be defined as “the efficient use of plants to remove, detoxify or immobilize environmental contaminants in a growth matrix (soil, water or sediments) through the natural biological, chemical or physical activities and processes of the plants”. The plants can be subsequently harvested, processed and disposed. Heavy metal ions such as Cu²⁺, Zn²⁺, Mn²⁺, Fe²⁺, Ni²⁺ and Co²⁺ [2] are essential micronutrients for plant metabolism. These ions when present in excess, along with non-essential metals such as Cd²⁺, Hg²⁺ and Pb²⁺, can become extremely toxic. Thus mechanisms must exist to satisfy the requirements of cellular metabolism but also to protect cells from toxic effects. High levels of metals in the soil interfere with the uptake of essential ions, biosynthesis of chlorophyll and nucleic acids and lipid metabolism, thus profoundly affecting nutrition, respiration and photosynthesis. Wide ranges of gene families are involved in transition/heavy metal transport. These includes the heavy metal ATPases (HMAs), the natural resistance associated macrophase proteins (Nramps), the cation diffusion facilitator (CDF) family, the ZIPs (ZRT, IRT- like proteins) family, and the Cation anti-porters. Complexation of toxic metal ions by peptides or proteins also helps in detoxification process. One recurrent general mechanism for heavy metal detoxification in plants and other organisms is the chelation of metal by a ligand. Plants make two types of peptide metal binding ligands: Metallothioneins (MT's) and Phytochelatins (PC's) [3]. PCs are enzymatically synthesized peptides, whereas MTs are gene-encoded polypeptides.

In the current work, an attempt was made to look for putative metal binding motifs responsible for phytoremediation in plants especially MAP's [15]. Medicinal and aromatic plants appear to be a good choice for phytoremediation since these species are mainly grown for secondary metabolites thus eliminating the possibility of the contamination of the food chain with heavy metals. Aromatic and medicinal plants also have demonstrated ability to accumulate heavy metals. Research has shown that heavy metals accumulated by aromatic and medicinal plants do not appear in the essential oil and that some of these species are able to grow in metal contaminated sites without significant yield reduction [4]. Multiple sequence alignment of representative proteins from different plant species helps in identifying such conserved regions. These can further be characterized as sequence motifs or patterns. Generating a functional motif involves

identifying residues in a protein sequence that impart functional properties to the protein. Motifs when used to probe sequence databases help to find and annotate members belonging to a particular family of proteins. These family of proteins of motifs were subjected to binding site prediction software [5] that shows binding site capability of metal ion ligands. PROSITE is a database for motifs and patterns. PROSITE includes heavy metal associated domain signatures and profiles with PROSITE and ATPase signature bearing IDs. These are broad spectrum motifs. In this study, specific putative metal binding motifs have been designed for the ions nickel, lead, molybdenum, copper, manganese, cadmium, zinc, iron, cobalt, aluminium, magnesium and xenobiotic compounds.

Methodology:**SWISS PROT:**

Based on gene name search, metal ion binding, protein sequences belonging to bryophytes, pteridophytes, gymnosperms and angiosperms against the genes involved in phytoremediation were retrieved from the UniProtKB/Swiss-Prot (release 55.3) database.

CLUSTALX

Sequences of the proteins for the respective genes were subjected to multiple sequence alignment to generate conserved regions of evolutionary importance.

Bioedit

The alignments obtained were run on Bioedit for manipulation and editing purpose in order to generate consensus sequences. Motifs were designed (Table 1 in supplementary material) keeping in mind that carboxylic acids and amino acids such as citric acid, malic acid, histidine, glutamic acid, aspartic acid and cysteine are potential ligands for heavy metals. Therefore, they could play a role in tolerance and detoxification [6]

Prosite Scan Tool

The designed motifs were run on Scan Prosit tool at Prosit database to detect functional and structural intra-domain residues. The hits obtained were recorded for each class of proteins. Representative structures were obtained from the Protein Data Bank and the motifs were correlated with the metal binding residues.

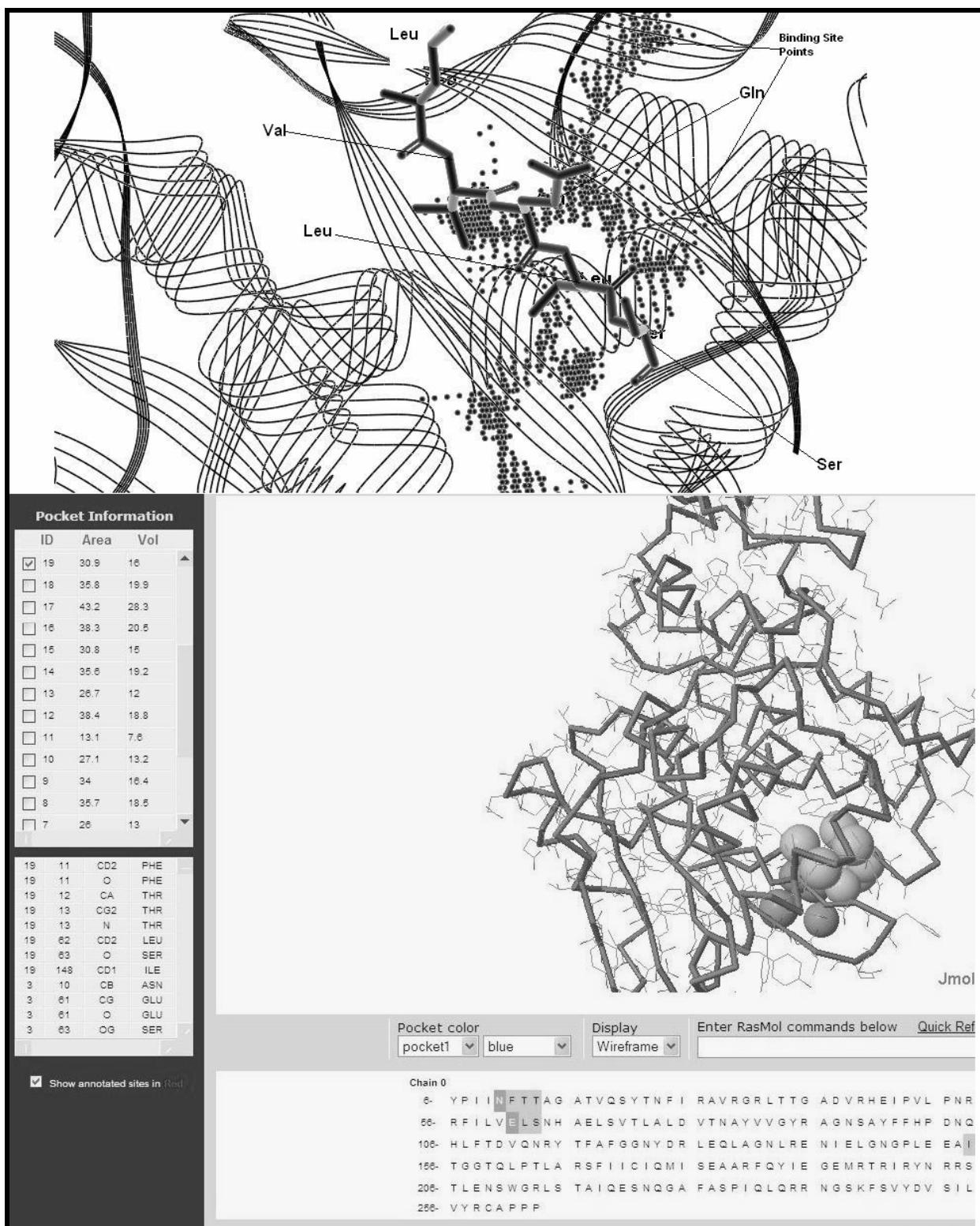


Figure 1: (a) Predicted motif for Nramp1 (LVELS, PDB Id 1IFS-A). The above (a) depicts the presence of binding cavities in the vicinity of metal binding motifs LVELS in pdb 1IFSa (binding site module in Discovery studio). Here binding crevices are depicted as binding point ranges in large proportions that can be seen preferentially around valine, Glutamine and leucine amino acid. (b) Validation of the motif for Nramp1 protein through CASTp server.

Discovery Studio

Binding site prediction for the PDB ID'S obtained for the motifs designed for metal binding site was done by using the Discovery Studio 1.6 full version binding site module [8].

CASTp Server

CASTp server was used to validate probable binding site visualized through Discovery Studio.

Discussion:

Motifs were designed and validated for each protein family. Conserved motifs/putative metal ion binding motifs corresponding to heavy metals with respective genes are given in **Table 1 (see supplementary material)**. Site directed mutation studies which confirm the involvement of the residues in certain motifs were also recorded [7]. The 3D structure of these motifs were generated and analyzed. The patterns generated out of motifs gives an indication of probable binding sites as visualized through Discovery Studio (**Figure 1a**) for phytoremediation. These were further validated out through CASTp server [14]. CASTp web server was used to study surface features, functional regions and specific roles of key residues of proteins out of the designed motifs. The motifs designed give an insight into the physicochemical properties needed for a protein to perform its function. Pocket 3 and Pocket 19 are involved in Nramp1 protein which imparts its binding properties. The results reveal its binding as hydrophilic in nature. Similarly Pocket 25, 51, 85 and 93 are involved for Laccase protein and Pocket 46, 50, 87 and 119 are involved for MGT protein.

ATP driven P-type heavy metal pumps represent a class of proteins that translocate toxic and essential metal ions across biological membranes. They are also called as the CPx type ATPases and form the ion translocation subclass of P-type ATPases. They contain a conserved CPC or CPX motif and varying number of C-Xaa-Xaa-C motifs in the N-terminal domain of the proteins [12]. They contain a conserved aspartate as phosphorylation site. One signature has been designed for the ion translocating P-type ATPases in this study. The motif [ND]L[QLRF][HY][DP][PS] designed for MGT protein is part of a hinge motif in P-type ATPases and site directed mutagenesis studies have shown that the conserved aspartate and proline residues are important for catalytic activity [13]. Some of the motifs could not be validated due to the lack of structural data. However, since most of these motifs have residues considered to be important for metal binding, that could have a role in metal binding or structural integrity.

CastP server (1b) is used to validate the results obtained from Discovery studio. The figure shows that in PDB 1IFS-A, large no. of binding cavities are present on the basis of topographical geometry. Binding cavities 3 and 19 depicts the involvement of similar residues E and L from the motif "LVELS", as observed within the sequence below the three dimensional figure. Similar residues are conserved in pdb structures involved in the binding of Iron (Fe) and Manganese (Mn) within the protein structures 1IFS-A corresponding to genes Nramp1. This gene is reported to be involved in Fe and Mn uptake (Maser P *et al.*, 2001, L. E. Williams *et al.*, 2000 [10]. Similiarly all those motifs are identified in plants which are involved in phytoremediation of heavy metals , with having knowledge of particular genes involved in the process even if the genes are not known in that plants.

On the left hand side bar of CASTp server, lists of all the atoms from these residues are tabulated which provides interaction or forms the structure of these metal ion binding cavities. It also gives an account of the hypothetical motifs that shows a binding cavity in its vicinity and suggests its important role in phytoremediation within plants especially in those plants where it is conserved in the sequence through evolutionary history.

Table 1 (see supplementary material) gives an account of all the proteins involved in phytoremediation along with the references. It also gives descriptive information of the metals that are involved in binding the proteins. Detailed information of the important amino acid residues required to bind the protein to form a metal binding complex is also quoted. The following table shows the entries obtained for the sequences obtained in UniProt/SwissProt which is preceded by the hits obtained through Scan Prosise for motifs search. In most cases, the motifs were able to retrieve all instances of the corresponding protein. In cases where the number of hits obtained by a motif search is higher than the instances of the protein in UniProtKB, the additional hits were either related proteins or hypothetical proteins. In some cases where the motifs missed some of the UniProtKB entries, some of the entries were fragments and it is plausible that the motif designed does not fall in the fragment. The last two columns gives a note of the PDB id's available, generated out of the Scan Prosise result which finally terminates into the CastP results giving account of the binding pockets involvement.

Parallel Approach

All the EST related to corresponding heavy metal uptake genes are retrieved and assembled through CAP3 assembler [16]. Orthologous EST matches were determined in plants related to MAPs family through blast. As a result large number of EST were determined in MAPs which share orthologous relationship with metal uptake specific EST. The results led to the involvement of *Artemisia annua* and *Allium cepa* plants for xenobiotic phytoremediating genes; *Helianthus annus* for manganese uptake and *Hordeum vulgare*, *Brassica napus* and *Helianthus annus* for zinc uptake. E-value adjusted ≤ 0.001 for all blast hits with percentage $\geq 85\%$. Similarly, with this approach large number of plants were studied with context to genes related to large number of heavy metal uptake viz Nickel, cadmium, Copper, Cobalt, Iron, Manganese and Xenobiotics (**Table 2 in supplementary material**). Datasets reveals the orthologous relations amongst contigs related to genes EST dataset for Xenobiotic uptake of genes with almost more than 85% similarity in case of *Artemisia annua*. Similar results were obtained for *Allium cepa*.

Conclusion

The motifs represent conserved regions that lie in protein structural core and are formed by three-dimensional structural topology of amino acids. All the motifs that have been constructed in this study retrieved the corresponding metal binding proteins from UniProtKB indicating that they are all specific to the protein families taken into consideration. The specificity of amino acids plays a pivotal role in the process of phytoremediation. A clear hypothesis can be drawn that the amino acids and its various associations are important for the process of phytoremediation. The efficiency of some of the possible amino acids those possessing with same physico-chemical properties useful as a ligand can be targeted for more uptake of toxic metals. The dual role of several motifs conserved cannot be ignored. Specific motifs can be effectively used for functional annotation of proteins. Probing non-redundant databases with such motifs will help catalog plants with a potential for metal tolerance, metal resistance, metal accumulation, metal transport & metal detoxification. This knowledge can be used for developing efficient bioremediation strategies in future.

References

- [1] AD Peuke, H Rennenberg, *EMBO Rep.* **6**:497 (2005) [PMCID: PMC1369103].
- [2] S Eapen, SF D'Souza, *Biotechnology Advances* **23**:97 (2005) [PMID: 15694122].
- [3] C Cobbett, P Goldsborough, *Annual Review of Plant Biology* **53**:159 (2002) [PMID: 12221971].
- [4] <http://www.cprm.gov.br/pgagem/Manuscripts/jeliazkovaea.htm>
- [5] R Dayam *et al.*, *J. Med. Chem.* **48**:8009 (2005) [PMID: 16335925].

- [6] LE Williams *et al.*, *Biochimica et Biophysica Acta*, **1465**:104 (2000) [PMID: 14585824].
[7] J Lowe *et al.*, *J. Biol. Chem.* 279:25986–25994 (2004) [PMID: 15078884].
[8] www.accelrys.com/products/discovery-studio/
[9] D Hussain *et al.*, *Plant Cell* **16**:1327 (2004) [PMID: 15100400].
[10] LE Williams *et al.*, *Biochim. Biophys. Acta*. **1465**:104 (2000) [PMID: 10748249].
[11] L Li *et al.*, *Plant Cell* **13**:2761 (2001) [PMID: 11752386].
[12] M Solioz, C Vulpe, *Trends. Biochem. Sci.* **21**:237 (1996) [PMID: 8755241].
[13] J Okkeri *et al.*, *Eur. J. Biochem.* **269**:1579 (2002) [PMID: 11874474].
[14] J Dundas *et al.*, *Nucleic. Acids. Res.* **34**:W116-8 (2006) [PMID: 16844972].
[15] <http://www.isb.vt.edu/>
[16] X Huang and A Madan, *Genome. Res.* **9**:868 (1999) [PMID: 10508846].

Edited by P. Kanguane

Citation: Sharma *et al.*, Bioinformation 4(2): 66-70 (2009)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Motif assignment for phytoremediation genes

Gene	Metals	Functional residues	Pattern design	Motifs obtained	UniprotKB/Swiss Prot	Prosite Scan	PDB ID	Castp Results[Pockets involvement]
Cnx1	Mo	His[H],Cys[C]	[HL][AF][LM] K[QL][IL][KF] G[DG]	HALKQI KGD	4	1	1O8N-A 1O8N-B 1O8N-C	Pocket 41 and 45.
HMA3	Cd,Pb, Zn	Glu[D], Asp[E], Lys[K]	Ks[SA][HC]{}	KSSH	3	4	2AC1-A	Pocket 35
Nramp1	Fe,Mn	His[H],Glu[E],Ser[S]Glu[E]	[FL][VG]E[ML][SHG]	LVELS	8	1	1IFS-A	Pocket 3 and 19
MGT	Co	Glu[E],Cys[C],His[H]	DH][ED][VT]E	DEVE	5	6	2A3L-A	Pocket 46, 50, 87 and 119
Laccase	Xenobiotics	His[H]	HP][CR][HR][LFIMA[DE][AH	HCHIEP H	5	1	1ASP-A	Pocket 25, 51, 85 and 93
			PFSVIQR][HR					

Table 2: Plant EST Blast for Xenobiotic uptake genes *Artemisia annua*

query id	Subject ids (<i>Artemisia annua</i>)	% identity	E-value	bit score
Contig1	gi 159592739 gb EY052485.1	87.72	1.00E-26	121
Contig2	gi 159682384 gb EY051532.1	82.43	0.002	44.1
Contig3	gi 159682384 gb EY051532.1	100	1.00E-04	48.1