

An in silico approach for structural and functional analysis of Heavy Metal Associated (HMA) proteins in *Brassica oleracea*

Jasmin Šutković

International University of Sarajevo,
Hrasnicka cesta 15, Sarajevo,
Bosnia and Herzegovina
jsutkovic@ius.edu.ba

Mujo Kekić

International University of Sarajevo,
Hrasnicka cesta 15, Ilidza,
Bosnia and Herzegovina
mujo_kekic@hotmail.com

Maida Ljubijankić

International University of Sarajevo,
Hrasnicka cesta 15, Ilidza,
Bosnia and Herzegovina
maidaaa-k@hotmail.com

Peter Glamočlija

International University of Sarajevo,
Hrasnicka cesta 15, Ilidza,
Bosnia and Herzegovina
ammar_glamoclija@live.com

Abstract

Heavy metal ATPases (HMAs) are the most important proteins involved in heavy metal accumulation process. *Brassica oleracea* has 5 HMA (1-5) homologues whose 3D structure has been predicted and validated in this study by different bioinformatics tools. Phylogenetic and multiple sequence alignment analyses showed high relationship between HMA2 and HMA4, while two same domains were identified in all five HMA proteins: E1-E2 ATPase and haloacid dehydrogenase (HAD) domain. Four HMA (2-5) proteins were identified to be localized in the plasma membrane, while HMA1 localization is predicted to be in plastid. Interactome analysis revealed high interaction of all HMA (1-5) proteins with many metal ion binding proteins and chaperones. Among these, interesting and strong interaction is observed between all HMA (1-5) proteins and ATX1, while HMA1, HMA2 and HMA4 have been found to strongly interact with FP3 (farnesylated protein 3) and FP6 (farnesylated protein 6) proteins. Docking site predictions and electrostatic potentials between HMA2/HMA4 and the interactome proteins were explained and discussed in this study.

Keywords: Protein structure prediction, Heavy metals; accumulation; transport; interactome; docking site

1. Introduction

1.1. Brassicaceae family

Brassicaceae is the name of a medium-sized and economically important family of flowering plants, informally known as the mustards, mustard flowers, the crucifers, or the cabbage family. The name is derived from the included genus *Brassica*. *Brassica* is a one of the major crop worldwide with *Brassica oleracea* as a main consumed species in Europe and USA. *Brassica* is a genus with many beneficial characteristics for our health, such as reducing risk for age related chronic illness, degenerative diseases and it reduces risk of several types of cancer. *Brassica* contain many vitamins which are essential for our health, such as vitamin A, C, E, K and B-6, carotenoids (such as c- and b-carotene and zeaxanthin), anthocyanins, folate, soluble sugars and phenolic compounds which are known to be the major antioxidants of *Brassica* crops [1].

Interesting fact is that all parts of *Brassica* is used as a food, including root, stems, leaves, flowers, buds and seeds. *Brassica* has many species, thanks to difference

in phenotype within themselves. Like all species in *Brassica* family, *Brassica oleracea* is very rich with vitamins and other nutrients. *Brassica oleracea* has been bred into a wide range of cultivars, including cabbage, broccoli, cauliflower, brussels sprouts, collards, and kale, some of which are hardly recognizable as being members of the same genus, let alone species [2]. *Brassica* vegetables are highly regarded for their nutritional value. With high amounts of vitamin C and soluble fiber they are excellent candidates to fight cancer, including molecules known of anticancer properties such as cellsproperties:3,3'-diindolylmethane, sulforaphane and selenium [3]. Furthermore, *Brassica* vegetables are rich in indole-3-carbinol, a chemical which boosts DNA repair in cells in vitro and appears to block the growth of cancer cells in vitro [4]. They are also a good source of carotenoids, with broccoli having especially high levels [5] and goitrogens, some of which suppress thyroid function [6].

1.2. *Brassica oleracea* genetic characterization

A recent study done with AFLP markers, evaluated the genetic diversity in kale landraces through Europe and compared the diversity to that in the wild populations of *Brassica oleracea*. In total 17 accessions were collected from all around Europe, including Bosnia and Herzegovina, Croatia and Turkey. In Bosnia and Herzegovina 47 individuals were analyzed, in Rivine, Dubrave and city of Stolac and its interesting to say that among a total of 93 polymorphic markers which were scored, a unique allele was found in only one accession, and it is the one in Bosnia and Herzegovina. In addition, the AFLP analyses of genetic diversity in leafy kale (*Brassica oleracea* L. *convar. acephala*) landraces, showed that Herzegovina has a 58% of polymorphic loci, while Croatia had 69% and Turkey 76%. Accessions from Bosnia and Herzegovina, Croatia, Portugal and Turkey contain many individuals with mixed genotype, sharing parts of their genome with other accessions due to common ancestry or gene flow [7].

1.3. Brassicaceae family as heavy metal accumulators

Through various ways, as for example, gas exhausts, energy and fuel production, intensive agriculture, and sludge dumping activities, humans contaminate soils and aqueous streams with large quantities of toxic metals. A number of studies from developing countries have reported heavy metals contamination in wastewater and wastewater irrigated soils [8]. In this regards, heavy metals are harmful to humans and other life forms, as they can cause cancer, blindness, loss organ function, severe illness, and death. The fact that some of these Brassicaceae family plants can accumulate high amounts of toxic metals, without visible symptoms, and in the same time being important food crops as well, leads to potential contamination of our food chain and this has to be taken into account in any phytoremediation process [9]. In general, plants require at least 14 mineral elements for their nutrition. These include the macronutrients nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg) and sulphur (S) and the micronutrients boron (B), chlorine (Cl), iron (Fe), manganese (Mn), copper (Cu), zinc (Zn), nickel (Ni) and molybdenum (Mo). Crop production is often limited by low bioavailability of essential mineral elements and/or the presence of excessive concentrations of potentially toxic heavy metals, such as Fe, Mn, Cu, Cr, Cd, Pb, Zn and Al in the soil solution. High concentrations of heavy metals in the soil can inhibit plant growth and reduce crop yields, which can affect sustainable development severely [10]. Therefore, some known Brassicaceae family species are already proven to be effective heavy metal

accumulators. For example, *A.halleri* is one of the best model organism for the study of plant adaptation to extreme metallic conditions since it is considered as Zn and Cd hyper-accumulator [11]. *A.halleri* is found to cope with excessive metal ions and toxicity in a way that it uses effective metal uptake, increased xylem loading and increased detoxification in shoot tissues. In recent years, several types of transporters involved in these processes have been identified in Zn and Cd hyper-accumulators, specifically in *A.halleri* [12]. The most investigated proteins that transfer the toxic metals are named as Heavy metal ATPases. These are located within membrane complexes of plant cell. As their name implies they produce or utilize energy in form of ATP. There are three different types of heavy metal transporters: P-type, V type and F0-F1 type. Most common type that is found in plant organisms is P-type. The proteins of this type usually transport essential metal ions which are Cu^{2+} , Zn^{2+} , Mn^{2+} , Fe^{2+} and Co^{2+} . This type of transporters does not produce energy, but actually uses it in order to pump these metals. Another type is V type which also utilizes energy. The third type is F0-F1 ATPases which produce energy instead of using it. The function of these proteins is to regulate the concentration of these metals in all tissues found in plants [10]. Their proper functioning is highly important for plant, where high levels of essential and some non-essential metals can be very toxic for the plant [9]. For example, expression of the HMA1 gene from *Atriplex canescens* significantly increased the ability of yeast cells to adapt to and recover from exposure to excess iron. AChMA1 expression also provided salt, alkaline, osmotic and oxidant stress tolerance in yeast cells. In this regards these results suggest that HMA1 gene encodes a membrane-localized metal tolerance protein that mediates the detoxification of iron in eukaryotes and may be involved in the response to abiotic stress [13]. HMA2 is known for maintaining plant metal homeostasis by transporting Zn and Cd metal ions [14]. It is shown that HMA2 and HMA4 drive metal efflux out of the cell in *A. thaliana* [15] and promote xylem loading of metal in *N. caerulescens* [16]. HMA4 is responsible for zinc hyper-accumulation in *A. halleri* as it shown by a RNA interference approach for down regulation of its expression. Additionally, transfer of the HMA4 gene to *A. thaliana* enables zinc partitioning into xylem vessels and up-regulated specific genes characteristic for zinc hyper accumulators [17]. This example shows impressively the importance of regulatory gene expression and gene copy number expansions for the special trait of metal hyper-accumulation. Furthermore, AtHMA4 is shown to be responsible for the reduction of Cd uptake/accumulation [18]. In contrast, HMA3 is localized at the tonoplast enabling vacuolar metal influx and therefore cellular sequestration [19]. The

quantitative trait locus (QTL) analysis on chromosome 1 in *Arabidopsis thaliana* revealed that this QTL regulates Cu translocation capacity and involves Cu-transporting via HMA5 [20].

Furthermore, in *Arabidopsis*, the heavy metal P-type ATPase HMA5 is shown to interact with metallo-chaperones and function in copper detoxification of roots [21]. It is found that some HMA genes are highly expressed in *A.halleri*, suggesting their importance in hyperaccumulation process. Among HMA genes, HMA2 and HMA4 are discovered to be among the most important ones with HMA4 playing role ineffective root-to-shoot Zn/Cd translocation [12] and HMA3 playing role in Zn detoxification [22]. Additionally, HMA2 and HMA4 have both been demonstrated to be plasma membrane proteins. Finally, both of these proteins appear to function in Cd transport

within the plant. Analysis of whole plant demonstrates that HMA2 accumulate more Zn and Cd than wild type plants although they do not appear to have an increased sensitivity to either metal. HMA4 mutant plants accumulate more Zn and Cd in the roots but they accumulate less Zn and Cd in leaves [22].

2. Material and methods

2.1. Retrieving HMA sequences and Multiple Sequence Alignment

The sequences of Heavy metal ATP proteins were obtained from the National Center for Biotechnology Information (NCBI) Protein Database [23]. Sequences' accession numbers are listed in table 1.

Table 1: Accession numbers of HMA proteins

HMA proteins	<i>Arabidopsis thaliana</i>	<i>Brassica oleracea</i>	<i>Brassica napus</i>	<i>Brassica rapa</i>
HMA1	At4g37270	XP_013595810.1	XP_013722994.1	XP_009137459.1
HMA2	At4g30110	XP_013609995.1	XP_013748024.1	XP_009128090.1
HMA3	At4g30120	XP_013591300.1	XP_013704754.1	XP_009128077.1
HMA4	At2g19110	XP_013629797.1	XP_015765843.1	XP_009150707.1
HMA5	At1g63440	XP_013606061.1	XP_013684388.1	XP_009112946.1

Multiple sequence alignment (MSA) has been performed using the Clustal Omega software located on the website of the European Bioinformatics Institute (EBI), using default options [24]. Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. MSA is an invaluable bioinformatics tool used to measure the similarity between sequences, examine patterns of conservation and variability and derive evolutionary relationships [25].

2.2. Phylogenetic tree construction

In order to infer the evolutionary relationship between the HMA proteins, a phylogenetic tree was constructed using ClustalW2 phylogeny, a web service for phylogenetic analysis of molecular sequences [26]. The service was run on default settings, and the steps that it performed to construct the phylogenetic tree involved MSA, alignment organization and construction, and visualization of the phylogenetic tree using different integrated tools.

2.3. 3D structure prediction and validation

The structures of HMA proteins were hereby predicted with the help of the Phyre2 protein homology modeling server [27]. Phyre2 is a web-based service

for protein structure prediction that is free for non commercial use, and being one of the most popular methods for protein structure prediction. Cited over 1000 times, it is able to generate reliable protein models. Phyre2 has been designed and funded by the Biotechnology and Biological Sciences Research Council (BBSRC) from United Kingdom. A practical and widely cited molecular visualization tool PyMOL was used for structure visualization and representation. PyMOL vs 1.31 edu (The PyMOL Molecular Graphics System), is a molecular visualization tool that provides viewing, customizing and exporting of the visualized molecules.

The validation and stereo-chemical analysis of the predicted structures was performed using several tools. The first one was QMEAN6, available as the structure assessment tool at ExpASy server. QMEAN6 is a scoring function that is actually a linear combination of six terms: torsion angle potential over three consecutive amino acids, two distance-dependent interaction potentials, solvation potential and two terms describing the agreement of the predicted structure and the solvent accessibility of the model [28]. Also provided is the Z-score of the QMEAN6, which compares the estimated score to the score from a high-resolution reference structure solved experimentally by X-ray crystallography, with strongly negative Z-scores expected from low quality models.

In QMEAN6 score better predictions have higher scores (between 0 and 1) and in Z-score lower quality predictions have more negative scores. Next, we used Verify3D, which assesses protein structures using three-dimensional profiles, analyzing the compatibility of a 3D model with its own amino acid sequence (1D). The scores range from -1 (bad score) to +1 (good score) [29]. Stereo-chemical quality of the protein models was assessed with the PROCHECK software [30]. PROCHECK compares the geometry of the residues in the predicted model with the known stereo-chemical values from well-known structures. It results in Ramachandran plots, providing information about the dihedral angles ϕ and ψ of amino acid residues in the protein structure. What PROCHECK basically does is comparison of the geometry of the residues in the predicted model with the known stereo-chemical values from well-known structures [30].

2.4. Localization of proteins

For the subcellular localization, we used the recently developed tool PSI (Plant Subcellular localization integrative predictor) which uses the group voting strategy and machine learning to combine the results of 11 independent subcellular localization tools: cello, mPloc, Predotar, mitoProt, MultiLoc, TargetP, WolfPSORT, subcellPredict, iPSort, Yloc and PTS1 [31].

2.5. Domain search and interaction prediction

The identification of domains in the five HMA proteins was performed using the online tool SMART (Simple Modular Architecture Research Tool) located on the website of the European Molecular Biology Laboratory (EMBL). The tool is able to detect more than 500 domain families from chromatin-associated, extracellular and signaling proteins [32].

Interactome of the HMA proteins was determined by using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins). It was used for searching interlogs of the five proteins. The STRING database consists of known and predicted protein interactions of currently 9 643 763 proteins and 2031 organisms. The predicted protein interactions are classified into physical or functional associations. What the program does basically is to determine binary interactions of each individual protein with predicted proteins.

Further, each interaction is assigned to a confidence score which depicts the quality and number of experimental technique used for the detection of these protein interactions [33].

2.6. Docking sites prediction

The docking site prediction was undertaken by the ClusPro 2.0 online software, an automated docking tool from the Structural Bioinformatics Lab of Boston University. ClusPro works on the basis of providing 70 000 rotations for the ligand protein from which 1000 rotations with the lowest score are chosen [34].

3. Results

3.1. MSA and Phylogenetic tree construction

Multiple sequence alignment was performed by the Clustal Omega online tool. The results of the sequence alignment are in table 2.

Table 2. Multiple sequence alignment of *Brassica oleracea* HMA proteins

Seq. 1	Length in aa	Seq.e 2	Length in aa	Clustal Omega score in %
HMA 1	817	HMA 5	999	26.81
HMA 1	817	HMA 3	758	25.60
HMA 1	817	HMA 2	883	26.24
HMA 1	817	HMA 5	999	26.41
HMA 5	999	HMA 3	758	27.82
HMA 5	999	HMA 2	883	29.94
HMA 5	999	HMA 4	1195	29.15
HMA 3	758	HMA 2	883	69.14
HMA 3	758	HMA 4	1195	67.60
HMA 2	883	HMA 4	1195	71.71

aa: Amino acids

Using the same tool we have constructed the phylogenetic tree that shows the evolutionary relationship between the aligned HMA proteins from *Brassica oleracea*.



Figure 1. Phylogenetic tree (cladogram) of HMA proteins in *Brassica oleracea*

In figure 1 we observe that we have two sister groups, one group being HMA2 and HMA4 proteins and on another side HMA1 and HMA5 proteins, each group having a common ancestor. HMA3, as a lone taxon, shares common ancestor with HMA1 and HMA5 but is more distant. However, it shows more homology with HMA2 and HMA4, as confirmed by Table 2.

In addition, phylogenetic tree was constructed combining the five *B.oleracea* HMA proteins with

HMA proteins from *A. thaliana* and *B.rapa* (see Figure 2). In this figure, it's clearly visible that *A.thaliana* HMA proteins share the common ancestors with all other taxa analyzed in the phylogenetic tree. Furthermore, this indicated the possibility of an evolutionary change of *A.thaliana* HMAs into *B.rapa* and *B.oleracea* HMA proteins. The HMA proteins from

B.rapa and *B.oleracea* share the common ancestor, being sister taxa with all HMA proteins, except HMA3. HMA3 protein from *B.rapa* and *A.thaliana* shares the common ancestor whereas *B.Oleracea* has evolved separately.

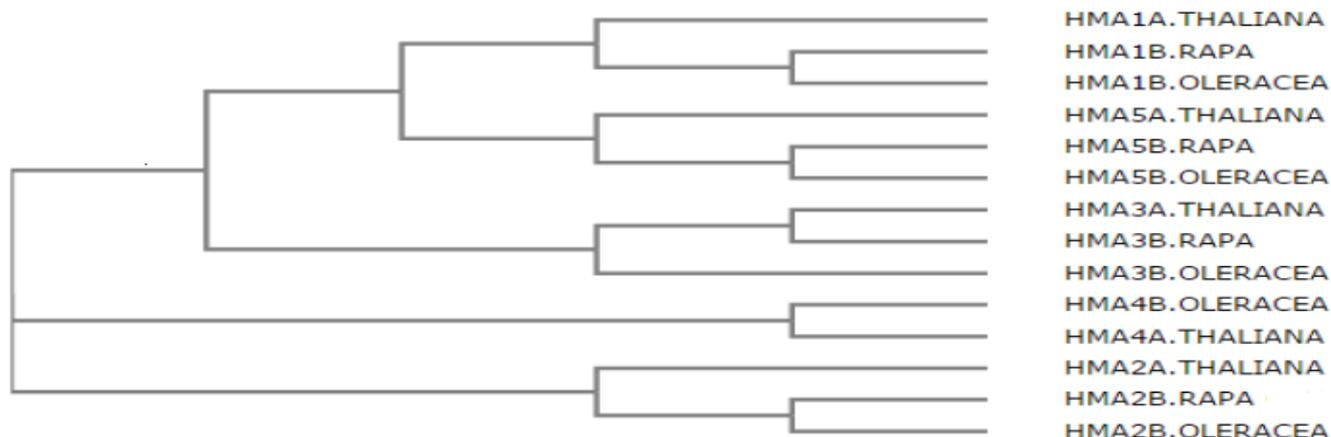


Figure2. Phylogenetic tree (cladogram) of HMA proteins with other Brassicaceae family members

3.2. Protein localization

By PSI, the proteins are localized to 10 possible locations, with a score from 0 to 1 and higher implying higher confidence in the presence of the protein in a particular subcellular compartment. Results are shown for protein localization prediction in table below:

Table3. Results of protein localization by PSI

Protein name	Predicted Subcellular Localization	Score
HMA 1	Plastid	0.689
HMA 2	Membrane	0.676
HMA 3	Membrane	0.694
HMA 4	Membrane	0.513
HMA 5	Membrane	0.393

3.3. Predicted and varified 3D structure models

The determination of the structure of proteins is vital for total understanding of the function, interactions and possible ligands, conserved domains and their homologues and many other purposes. However, experimental determination of the 3D structure is a demanding and time consuming process, so bioinformatics tools are used to predict the structures of proteins of interest. The 3D structures of HMA proteins predicted by Phyre2 tool are seen in figure 3.

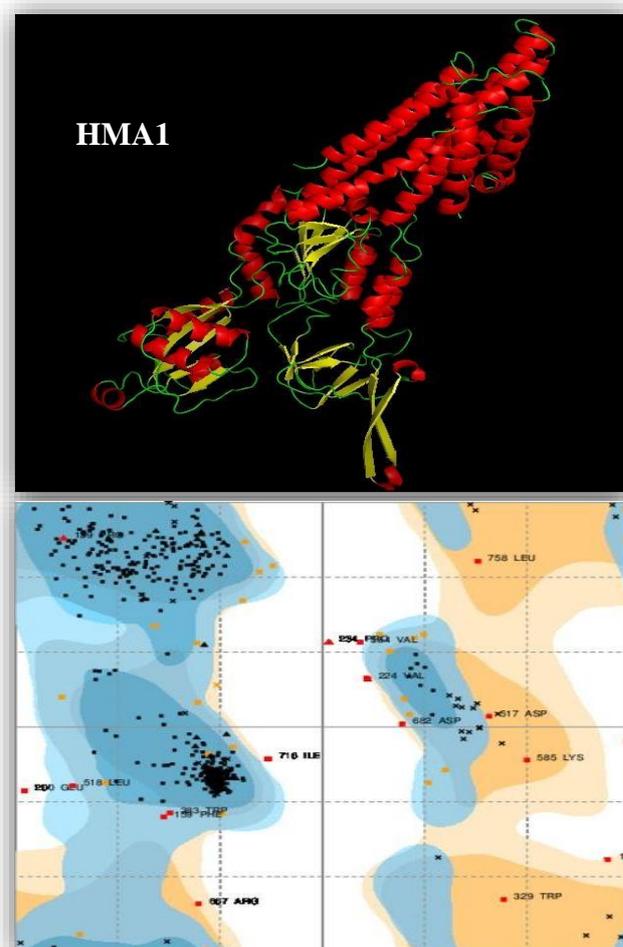


Figure 3: Predicted 3D structures and respective Ramachandran plots for HMA1 protein

After the prediction, the 3D structures underwent the process of validation by several structure assessment tools. The results are shown in table 4.

Table4. Structure assessment of predicted 3D structures

Name	QMEAN6 score	QMEAN Z-score	Verify3D	PROCHECK most favored region in %
HMA 1	0.546	-2.40	0.70	93.5
HMA 2	0.516	-2.73	0.68	90.4
HMA 3	0.523	-2.66	0.78	92.3
HMA 4	0.554	-2.31	0.72	90.3
HMA 5	0.440	-3.57	0.80	91.1

3.4. Domain identification

The domains in the five HMA proteins from *Brassica oleracea* were identified by the SMART software. The results are presented in the table 5:

Table5. The domains and their positions in the five HMA proteins

Protein Names	DOMAIN NAME			
	E1-E2 ATPase		HAD	
	Start	End	Start	End
HMA 1	202	438	446	687
HMA 2	164	383	391	603
HMA 3	167	386	394	606
HMA 4	175	394	403	614
HMA 5	414	654	662	879

The SMART analysis revealed that all HMA proteins have the same domains but on different locations. The E1-E2 ATPase domain is a trans-membrane domain, which is basically membrane-bound enzyme complex/ion transporter that uses ATP hydrolysis to drive the transport of protons across a membrane. Some trans-membrane ATPases also work in reverse, harnessing the energy from a proton gradient, using the flux of ions across the membrane via the ATPase proton channel to drive the synthesis of ATP [35]. There are many different classes of P-ATPases, which transport specific types of ion: H⁺, Na⁺, K⁺, Mg²⁺, Ca²⁺, Ag⁺ and Ag²⁺, Zn²⁺, Co²⁺, Pb²⁺, Ni²⁺, Cd²⁺, Cu⁺ and

Cu²⁺. P-ATPases can be composed of one or two polypeptides, and can usually assume two main conformations called E1 and E2 [36]. The HAD domain, haloacid dehydrogenase (HAD) superfamily domain, includes phosphatases, phosphonates, P-type ATPases, beta-phosphoglucosyltransferases, phosphomannosyltransferases, and dehalogenases, which are involved in a variety of cellular processes ranging from amino acid biosynthesis to detoxification [37].

3.5. Interactome of HMA proteins

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is used for searching interlogs of the five proteins. The STRING database consists of known and predicted protein interactions of currently 9.6 million proteins and 2031 organisms [38]. The predicted protein interaction is classified into physical or functional associations. Before entering FASTA format of sequences into STRING, we need to specify which organism to search for our sequence. STRING doesn't offer *Brassica oleracea* as a model organism so we used organism *Brassica rapa*, because it shares more than 90 % of homology with HMA proteins from *Brassica oleracea* species.

Table6. MSA of *Brassica oleracea* and *Brassica rapa* HMA protein family (1-5)

Proteins		Similarity %
BoHMA 1	BrHMA 1	97.1
BoHMA 2	BrHMA 2	95.9
BoHMA 3	BrHMA 3	97.2
BoHMA 4	BrHMA 4	90.0
BoHMA 5	BrHMA 5	97.4

Bo: *Brassica oleracea*

Br: *Brassica rapa*

The interactome analysis revealed strong interactions of HMA1, HMA2, HMA4 proteins with FP3 (farnesylated protein 3) and FP6 (farnesylated protein 6), whereas all HMA proteins show strong interactions with ATX1 (copper metallochaperone) protein and other related copper and ion binding proteins (see supplement Table 1 and Figure 7).

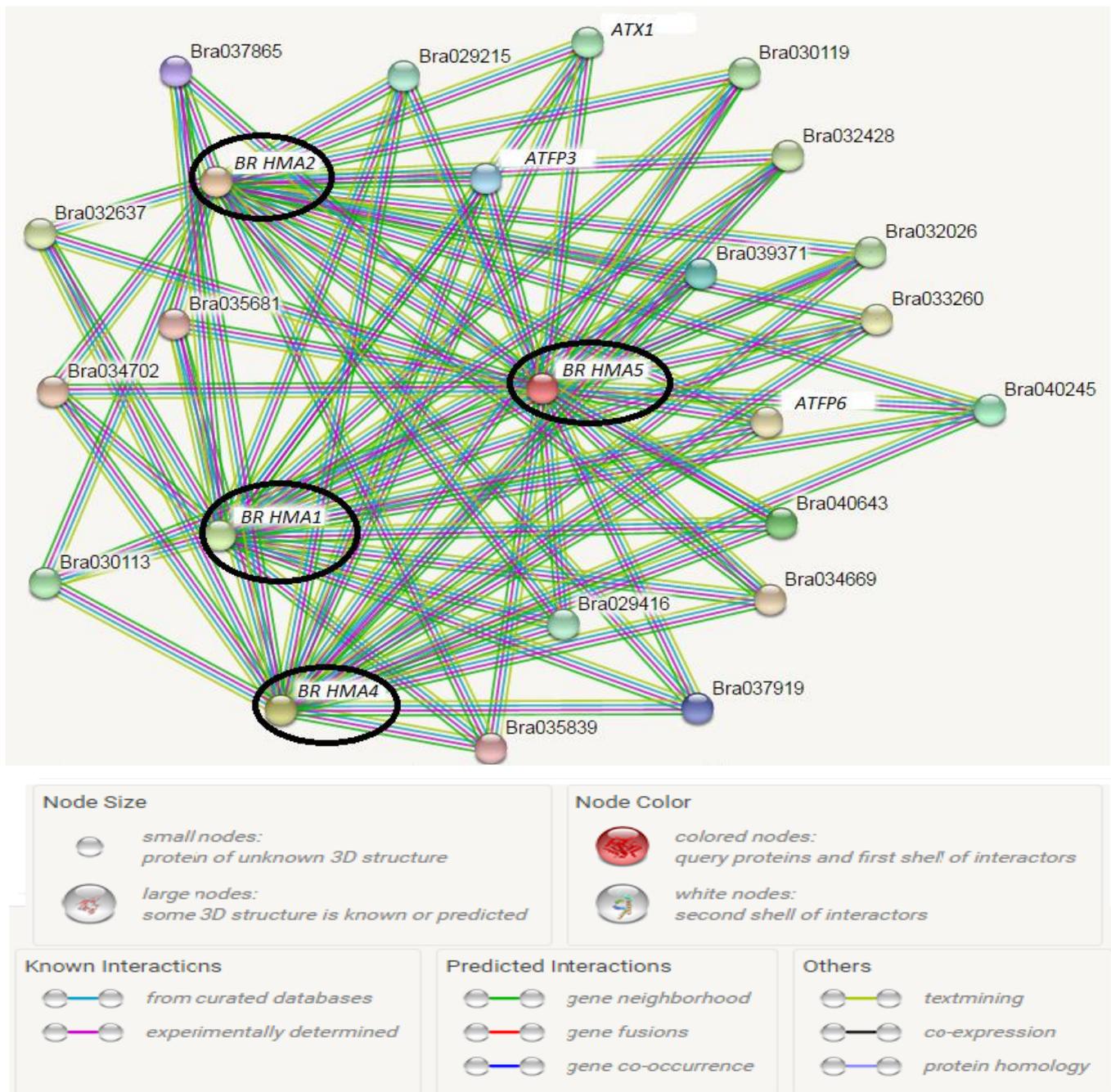


Figure 5: STRING interactome of *Brassica rapa* HMA1, HMA2, HMA4 and HMA5 protein family (original figure shown)

The *Brassica rapa* HMA3 protein interactome is not shown in figure 7 due to the absence of this particular protein in STRING interactome server. In order to analyze the interactome of HMA3 protein we have used the homolog from *Arabidopsis thaliana* as a template (AT4G30120), due to the high similarity of 90% with the *B.oleracea* HMA3 protein. In this analysis we have confirmed that the HMA3 from *Arabidopsis thaliana* is also predicted to interact with FP3 and FP6 as the homolog heavy metal accumulator's protein from *Brassica rapa*. Furthermore, the results show strong

interactions with other metal transporting proteins such as Copper chaperone (CCH) related, heavy metal associated isoprenylated plant protein 27 (HIP27) and several heavy metal transport/detoxification domain-containing proteins (see supplement table 2).

3.6. Docking sites prediction results

In figure 6, 7 and 8, the results of docking site prediction of HMA2 and HMA4 are shown with AtX1, FP3 and FP6, respectively.

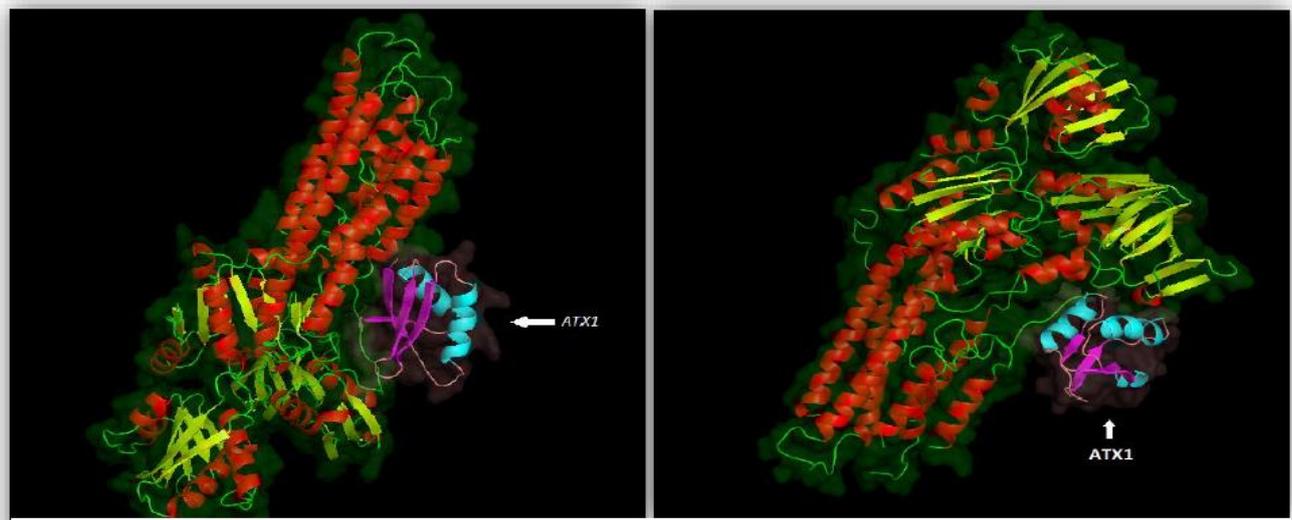


Figure 6: Docking site prediction of ATX1 with HMA2 and HMA4 from *Brassica rapa*

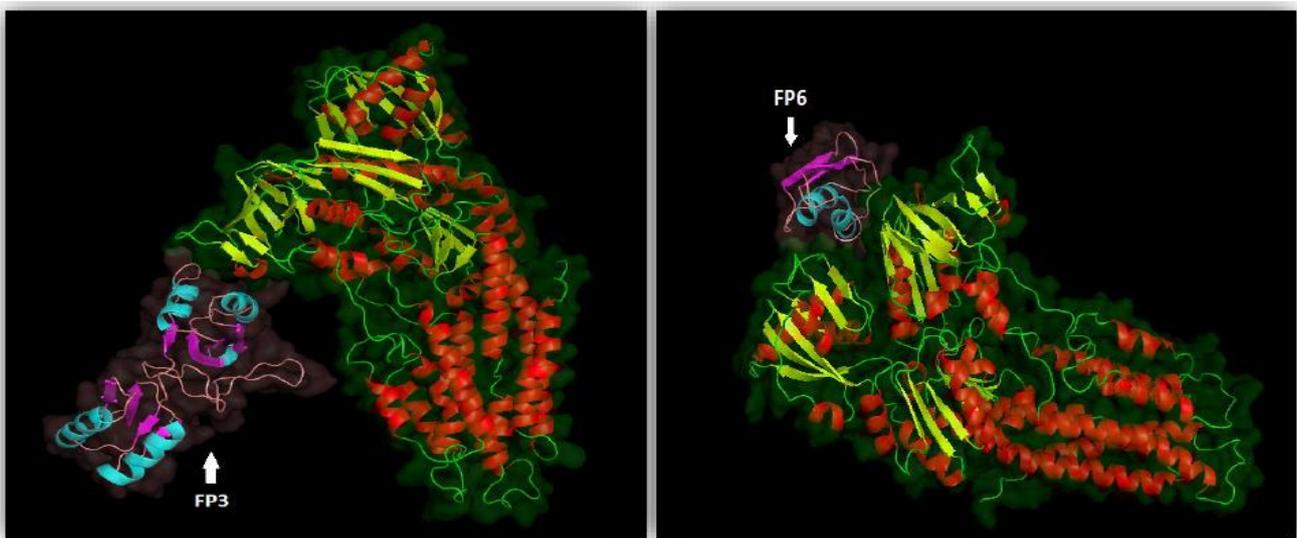


Figure 7: Docking site prediction of FP3 and FP6 with HMA2 from *Brassica rapa*

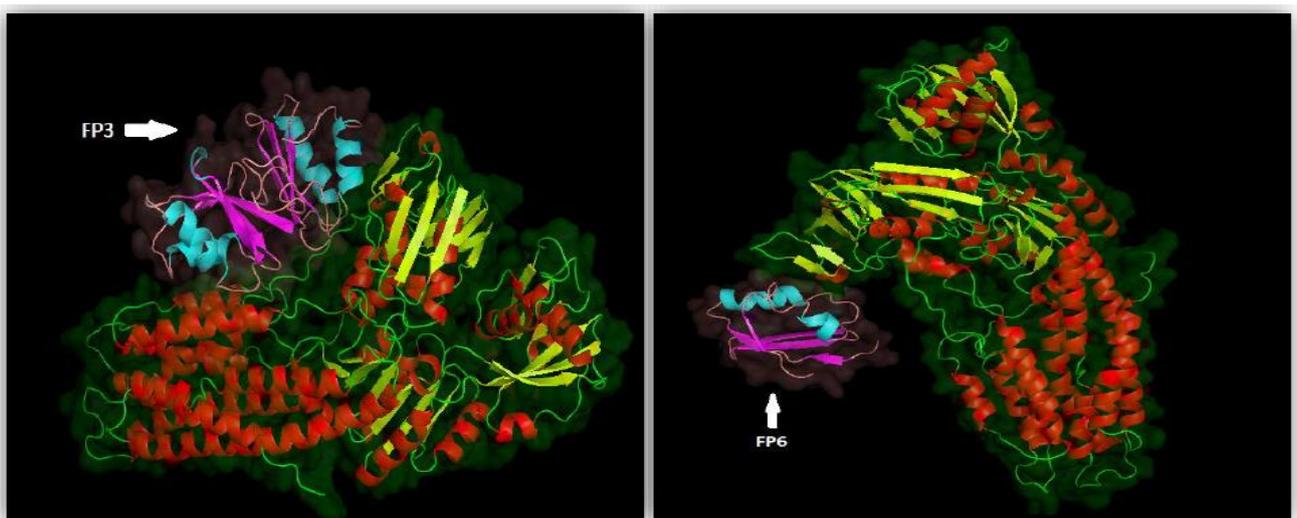


Figure 8: Docking site prediction of FP3 and FP6 with HMA4 from *Brassica rapa*

The protein structures modeled with ClusPro were checked by the same verification tools as with the HMA (1-5) proteins (Table 10). In addition, we have introduced an additional verification tool, DFire. This tool estimates the non-bonded atomic interactions in a model, thus providing the energy estimation which is considered closer to the native conformation if the DFire energy score is lower (supplement table 3) [39].

4. Discussion

Brassica oleracea is a plant known as metal hyper-accumulator that, as such can have an important role in environmental aspects. Among these, phytoremediation technology is the most interesting one and is the one that brought high attention of researchers in the last decade. However, beside phytoremediation and positive effects they can produce, metals-accumulating plants are directly or indirectly responsible for much of the dietary uptake of toxic heavy metals by humans and animals. Vegetables such as cabbage (*Brassica juncea*, *Brassica oleracea*) cultivated in wastewater-irrigated soils take up heavy metals in large enough quantities to cause potential health risks to the consumers [8].

Metal accumulation and translocation potential varies from plant to plant and metal to metal [8], therefore it is important to investigate both potentials in plants which are considered as metal-accumulators, which, in our work was *B.oleracea*.

Analysis of proteins responsible for metal accumulation and transport is of great importance to understand how those plants perform their functions in hyper-accumulation of metals. The HMA proteins (1-5) analyzed in our work are already known to play important roles in heavy metal accumulation processes. Beside their primary function as metal-accumulators, it is important to investigate other processes in which HMAs can be involved. In order to investigate such processes, we were looking for potential interactions with other proteins that are currently unknown to the literature, not known to interact with the analyzed HMA proteins.

In this work we have confirmed that HMA2 and HMA4 proteins share the most homology among other HMA family proteins [40], with 71.7% similarity. The phylogenetic tree analysis between the HMA proteins, additionally confirmed the similarity among these two proteins, where HMA2 and HMA4 share same ancestor, separated by other groups in the tree. These results suggest that due to their close evolutionary relationship, they play important biochemical roles by performing same or similar functions within the cell.

As reviewed by Hussain and colleagues [41], HMA2 and HMA4 play an important role in Zn transport and

homeostasis in *A.thaliana*. By mutating the HMA4 and HMA2 genes they have observed a significant decrease in Zn accumulation. Furthermore, they observed that only the *hma2-hma4* double mutant and neither of the single mutants exhibited an obvious nutritional deficiency in soil, suggesting that HMA2 and HMA4 have a level of functional redundancy, which can be consistent with sequence comparisons that show that HMA4 is the most closely related to HMA2, as confirmed in this study.

For further phylogenetic investigation for HMA homologues from the Brassicaceae family, a ClustaOmega cladogram was constructed (Figure 5), revealing clear and strong phylogenetic relationship between the target species (*B.oleracea*) and other Brassicaceae family species, *Arabidopsis thaliana* and *Brassica rapa*.

Due to high sequence similarity (with more than 90%) and common conserved domains of all HMA proteins of *B.oleracea* with *B.rapa* and *A.thaliana* HMA proteins, we may conclude that all known functions observed in *A.thaliana* and *B.rapa* HMA proteins can be attributed to *B.oleracea* homologues HMA proteins. As it is seen in Figure 5, *A.thaliana* HMA proteins share the common ancestors with all other taxa, indicating the possibility of an evolutionary change of *A.thaliana* HMAs into *B.rapa* and *B.oleracea* HMA proteins. The HMA proteins from *B.rapa* and *B.oleracea* share a common ancestor, being sister taxa with all HMA proteins, except HMA3. HMA3 proteins from *B.rapa* and *A.thaliana* share the common ancestor, whereas *B.Oleracea* has evolved separately.

The 3D structures of proteins enables additional functional studies, domain analysis, molecular interaction studies, estimation of structural similarity between proteins etc. In this study, we used Phyre2 tool, a protein homology modeling server, used to create models of target proteins. These models contain information about the tendency for mutation of each amino acid in a sequence and are unique for each protein. They are created for a set of known 3D structures as well as for the user sequence, and then scanned to find a match [42]. Further confirmation and verification of the modeled structure was tested by three validation methods. QMEAN6, PROCHECK and Verify 3D.

The verification results of all five HMA protein in *B.oleracea*, showed sufficient quality, required for further analysis. According to our results, the Verify3D score for HMA1 is 0.70, for HMA2 it is 0.68, for HMA3 0.78, for HMA4 it is 0.72 and the highest score was observed with HMA5, being 0.80. The Ramachandran plots analysis revealed that all of the structural regions lie in the range of acceptance, with having more than 90% favored regions. QMEAN6

results revealed good Z scores. To be precise, QMEAN6 score for HMA1 is 0.546, for HMA2 score is 0.516. 0.523 is score for HMA3 while 0.554 is score for HMA4. The lowest score is 0.444 and it stand for HMA5 (see Table 5). For the Z-score analysis we observe that all models have negative Z-scores (in average of -2 Z-score), being a median score for structural validations. Models of low quality are expected to have strongly negative QMEAN Z-scores, less then -3.5 [43]. Obtained Z-scores are in line to scores obtained for high-resolution experimental structures of similar sizes solved by X-raycrystallography. Therefore, the Phyre2 generated models appeared acceptable for the protein and metal docking site prediction.

We have shown that all proteins share two domains, starting from different residues. The domains identified are P-ATPases (E1-E2 ATPases), membrane-bound enzyme complexes/ion transporters that use ATP hydrolysis to drive the transport of protons across a membrane [44] and the HAD domain, haloacid dehydrogenase (HAD) superfamily domains which are involved in a variety of cellular processes ranging from amino acid biosynthesis to detoxification [37]. The interactome analysis revealed strong interactions of HMA1, HMA2, HMA4 proteins with FP3 (farnesylated protein 3) and FP6 (farnesylated protein 6), whereas all HMA proteins show strong interactions with ATX1 (copper metallochaperone) protein and other related copper and ion binding proteins. In order to analyze the interactome of HMA3 protein we have used the homolog from *A.thaliana* as a template (AT4G30120), due to the high similarity of 90% with the *B.oleracea* HMA3 protein. The HMA3 protein from *A.thaliana* is also predicted to interact with FP3 and FP6 as the homolog heavy metal accumulator's protein from *B.rapa*. Furthermore, the results show strong interactions with other metal transporting proteins such as Copper chaperon (CCH) related, heavy metal associated isoprenylated plant protein 27 (HIPP27) and several heavy metal transport/detoxification domain-containing proteins (supplement table 1).

FP6 (also known as HIPP26) is characterized by a heavy metal binding domain (HMA) and an additional isoprenylation motif on C-terminus. This family of HIPPs embraces at least 44 proteins in *A.thaliana* with HMA domain being responsible for heavy metal binding, metal transport and metal homeostasis processes. Isoprenylation motif is added through the process of isoprenylation [45].

Isoprenylation, also known as farnesylation, is a post-translational protein modification that involves addition of a C-terminal hydrophobic anchor that is important for interaction of the protein with membranes or other proteins [46].

In a study conducted by Barth and colleagues [45], it is confirmed that HIPP26 exhibits a nuclear localization signal (NLS), thus being localized in the nucleus. In their work, they also concluded that for the exact spatial localization of HIPP26 within the nucleus, the isoprenylation seems to be important, which probably by its hydrophobic nature determines the correct spatial arrangement of this protein within the nucleus.

Furthermore, their study confirmed that HIPP26 strongly interacts with ATHB29, a zinc fingerhomeodomain transcription factor (ZF-HD proteins) which is found to be induced by drought, high salinity and abscisic acid, thus playing role in regulation of stress response of plants [45]. Furthermore, Gao and colleagues [47] showed that FP6 in *A.thaliana* (AtFP6) upon interaction with plasma membrane acyl-CoA-binding protein 2 (ACBP2) mediate cadmium Cd(II) tolerance [47].

Due to the strong interaction with FP6, the represented data confirms HMA protein family involvement in Cd(II) transport and tolerance, since all three HMAs are found to be cadmium/zinc transporting ATPases. In addition, we may suggest that these three HMA proteins may be important in stress-induced tolerance, since it was the case for FP6 protein [45]. FP3 from *A.thaliana*, if soluble and isoprenylated, is capable of reversibly binding a copper-chelate matrix in tobacco BY2 cell homogenates, suggesting a ubiquitous role for these proteins in diverse plants [48]

In this study, we confirm the interaction of all *HMA* proteins with CCH (Copper chaperone) or CCH-related proteins, which has been shown to functionally complement *atx1* mutants, but the *ATFP3* gene expression is not regulated in the same manner as CCH gene expression [49].

ATX1 (copper metallochaperone) protein shows strong interactions with *HMA1*, *HMA2*, *HMA4* and *HMA5* proteins. ATX1 is related with copper metallochaperones which assist copper in reaching vital destinations without inflicting damage or becoming trapped in adventitious binding sites [50]. ATX1 is shown to bind Cu(I) in the cytoplasm which delivers it to a copper transporter in the membrane of a post-Golgi vesicles. In the vesicle, the copper is inserted into a multicopper oxidase essential for high-affinity iron uptake, so ATX1 can be involved in both, copper transport and defense against oxidative stress [49] ATX1 is also proposed to be involved in Cu homeostasis by its Cu-binding activity and interaction with the Cu transporter heavy metal-transporting P-type ATPase5, suggesting a regulatory role for the plant-specific domain of the CCH Cu chaperone, therefore, a role for *HMA5* in Cu compartmentalization and detoxification [21].

In a more recent study conducted by Lung et al. [51], it is confirmed that overexpression of ATX1 enhancing Cu tolerance implies the potential use of ATX1 for phytoremediation in Cu-contaminated soil. In same study, they connected *HMA5* with ATX1 on the way that ATX1 was proposed to deliver Cu to *HMA5* for Cu detoxification in roots and translocation to shoots.

In the docking analysis, the HMA proteins were considered to be a ligand (according to ClusPro default settings, the ligand is the structure that gets rotated to fit into the receptor). Specific docking sites presented are ATX1 with *HMA2* and *HMA4*. In this research we have verified our structures where all predicted models of docking show sufficient quality (see supplement table 3).

For further analyses of predicted docking structures, the electrostatic potential between the HMA proteins and docking partners was calculated via DeepView. This tool is showing clouds of negative and positive electrostatic potential in the docking site predicted, from which we could conclude that at least part of all docking sites is due to electrostatic forces. In supplement figure 1 we can see the clear separation of charges between *HMA2* /*HMA4* and ATX1 on similar docking regions.

The predicted docking region lies in the of N terminus as shown in literature, where *HMA2* and *HMA4* N-terminal domain are essential for function in planta while the C-terminal domain, although not essential for function, may contain a signal important for the subcellular localization of the protein (supplement figure 2)[52].

These predicted docking sites of FP3 and FP6 to *HMA2* and *HMA4* lie in similar region, usually on N terminus, which confirms the good modelling of 3D structures by ClusPro. The visualization of the resulting docking site models for *HMA4* with FP3 and PP6 (supplement figure 3).

All the structures, verified by the electrostatic potential, given by DeepView, confirm the docking sites supported by the electrostatic forces. It has been shown that the electrostatic potentials at the interfaces of interacting molecules are anti-correlated. This means that at the interface, there is a good chance to find a patch of positive electrostatic potential on the surface of one molecule positioned next to a negative patch on the surface of the adjacent molecule and vice versa [53]. Furthermore, a big DFire score results for all models are indicating good models of docking, which estimates the non-bonded atomic interactions in a model, thus providing the energy estimation that is closer to the native conformation the lower it gets (lower quality predictions have more negative scores) [39].

5. Conclusion

Brassicaceae family plants are known to accumulate high amounts of toxic metals, such are: (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg) and sulphur (S) and the micronutrients boron (B), chlorine (Cl), iron (Fe), manganese (Mn), copper (Cu), zinc (Zn), nickel (Ni) and molybdenum (Mo). The subgroup, *Brassica oleracea* shows great potential to be hyper-accumulator of Zn and Cd heavy metals. In this research we have focused to further investigate the roles of HMA (1-5) proteins in *Brassica oleracea*, predicting their structures and interactomes. To achieve the most accurate result, the stated aim was enhanced and supported through the use of several common bioinformatics techniques. The *B.oleracea* HMA proteins (1-5) were subjected to multiple sequence alignment analysis with HMA proteins from *B.rapa*, *B.napus* and *A.thaliana* in order to obtain information about conserved regions among these proteins and to assess the phylogenetic relationship of the proteins. This further enabled further analysis of their 3D structures as well as their interactome analysis, in order to confirm current functional roles of each protein and possibly discover new interactome partners unknown to the literature, for new annotations of functional roles of *B.oleracea* HMA proteins (1-5).

It is through bioinformatics analysis that we identified and structurally predicted 5 homologues of HMA proteins in *Brassica rapa*, mostly similar to *Brassica oleracea*. Since they are similar, but not identical in structure and differentiate in two groups in phylogenetic analysis, further inference about the functions and localization of the homologues was required. For that purpose, localization tools were used to predict subcellular locations and the trend of differences between the homologues continued. Lastly, the interactome analysis showed similar functions and associations with many crucial processes of metal ion transportation, required for cellular integrity and stability maintenance. The results obtained in this study lead us to the conclusion that cellular functions of the 5 homologues are very similar, where *HMA2* and *HMA4* are directly involved in Zn/Cd transport, whereas *HMA5* functions as metallochaperones and functions in copper detoxification, as confirmed within this study. Furthermore, *HMA2* and *HMA4* are shown to have strong interaction with ATX1 protein, by now only know to interact with *HMA5*, which may indicate a specific involvement of *HMA2* and *HMA4* proteins in Cu(I) binding and the delivery to the post- Golgi vesicle, with strong possibility of Cu compartmentalization and detoxification, as shown for *HMA5*.

In addition, the interactome analysis revealed strong interaction of *HMA2* and *HMA4* proteins with FP3 (farnesylated protein 3) and FP6 (farnesylated protein 6). A study conducted by Dykema et al. [48], showed that FP3 from *A.thaliana* has function as ubiquitous protein in diverse plants. FP6 (also known as HIPP26) is characterized by a heavy metal binding domain (HMA) and an additional isoprenylation motif on C-terminus. It is shown that HIPP26 strongly interacts with ATHB29, a zinc finger homeodomain transcription factor (ZF-HD proteins) which is found to be induced by drought, high salinity and abscisic acid, thus playing role in regulation of stress response of plants. Furthermore it is shown that FP6 interacts with acyl-CoA-binding protein 2 (ACBP2) mediate cadmium Cd(II) tolerance protein, indicating the possibility the *HMA2* and *HMA4* proteins may share the above mentioned cellular functions.

References

- [1] A.M. Gardner, A. F. Brown, J. A. Juvik, "QTL analysis for the identification of candidate genes controlling phenolic compound accumulation in broccoli (*Brassica oleracea* L. var. *italica*)," *Mol Breeding*, vol. 36:81, 2014.
- [2] M. P. Mourato, I. N. Moreira, L. Leitão et al., "Effect of Heavy Metals in Plants of the Genus Brassica," *International Journal of Molecular Sciences*, ISSN 1422-0067, pp 1-24, 2015.
- [3] H. J. Korry, J. W. Finley, A. Sigrid-Keck, R. J. Robbins, "The Journal of Nutrition", vol. 135, no. 5, 2005.
- [4] L. Fengjuan, W. Yongsheng, F. Xiaoling, J. Yucui, W. Zhaojia, H. William, P. Carolyn, L. Lei et al., "A Novel Mechanism of Indole-3-Carbinol Effects on Breast Carcinogenesis Involves Induction of Cdc25A Degradation," *Cancer Prevention Research*, vol. 3, no. 7, 2010.
- [5] M. W. Farnham and D.A. Kopsell, "Importance of Genotype on Carotenoid and Chlorophyll Levels in Broccoli Heads," *HortScience*, vol. 44, no. 5, 2009.
- [6] B. Srilakshmi, *Nutrition Science*. New Age International. ISBN 978-81-224-1633-6, pp. 186-7, 2006.
- [7] S. Christensen, R. von Bothmer, P. Gert et al., "AFLP analyses of genetic diversity in leafy kale (*Brassica oleracea* L. convar. *acephala* (DC.) *alef*) landraces, cultivars and wild population in Europe," *Genet Reseour Crop Evol*, vol. 58, 2010.
- [8] G.A. Boamponsem, M. Kumi and I. Debrah, "Heavy Metals Accumulation In Cabbage, Lettuce And Carrot Irrigated With Wastewater From Nagodi Mining Site In Ghana," *International Journal of Scientific and Technology Research*, vol. 1, no. 11, pp. 124-129, 2012.
- [9] S. Mohammad, A. Khan, Z.R. Goel and J. Musarrat, "Biomangement of Metal-Contaminated Soils," In: Environmental Pollution. Springer Netherlands (online), 2011.
- [10] L. Subasic, H. Gavranovic, I. Muhovic and A. Memon, "Heavy metal induced gene expression in Brassicaceae," Department of Genetics and Bioengineering, Faculty of Engineering and Information Technologies, International Burch University, BiH, 2012.
- [11] M. Cho, C.N. Agnes and D. Karl-Josef, "Differential heavy metal tolerance of *Arabidopsis halleri* and *Arabidopsis thaliana*: a leaf slice test," *The New Phytologist*, vol. 158, no. 2, pp. 287-293, 2002.
- [12] V. Shanmugam, Lo, Jing-Chi and Kuo-Chen. Yeh, "Control of Zn uptake in *Arabidopsis halleri*: a balance between Zn and Fe," *Frontiers in Plant science*, vol. 4, pp. 281, 2013.
- [13] X. Sun, G. Yu, Jing-Tao. Li et al., "A Heavy Metal-Associated Protein (AcHMA1) from the Halophyte, *Atriplex canescens* (Pursh) Nutt., Confers Tolerance to Iron and Other Abiotic Stresses When Expressed in *Saccharomyces cerevisiae*," *Int J Mol Sci.*, vol. 15, no. 8, pp. 14891-14906, 2014.

- [14] J. M. Argüello, E. Eren, M. González-Guerrero, "The structure and function of heavy metal transport PIB-ATPases," *M. Biometals*, vol. 20, 2007.
- [15] E. Eren and José M. Argüello, "Arabidopsis HMA2, a Divalent Heavy Metal-Transporting PIB-Type ATPase, Is Involved in Cytoplasmic Zn²⁺ Homeostasis," *Plant Physiol.*, vol. 136, no. 3, pp. 3712–3723, 2004.
- [16] P. Ashton and K. Leon, "Identification of *Thlaspi caerulescens* Genes That May Be Involved in Heavy Metal Hyperaccumulation and Tolerance. Characterization of a Novel Heavy Metal Transporting ATPase," *Plant Physiology*, vol. 136, no. 3, pp. 3814–3823, 2004.
- [17] K. Ute, I. N. Talke and M. Hanikenne, "Transition metal transport," *FEBS Letters*, vol. 581, no. 12, pp. 2263–2272, 2007.
- [18] O. Siemianowski, A. Barabasz, M. Kendziorek, E. Bulska, L. E Williams and D.M. Antosiewicz, "HMA4 expression in tobacco reduces Cd accumulation due to the induction of the apoplastic barrier," *J Exp Bot.*, vol. 65, no. 4, 2014.
- [19] A. Gravot, A. Lieutaud, F. Verret, P. Auroy, A. Vavasseur and P. Richaud, "AtHMA3, a plant PIB-ATPase, functions as a Cd/Pb transporter in yeast," *FEBS Lett.*, vol. 561, 2004.
- [20] Y. Kobayashi, K. Kuroda, K. Kimura et al., "Amino acid polymorphisms in strictly conserved domains of a P-type ATPase HMA5 are involved in the mechanism of copper tolerance variation in *Arabidopsis*," *Plant Physiol.*, vol. 148, no. 2, 2008.
- [21] A. S. Colás, V. S. Rodríguez-Navarro et al., "The *Arabidopsis* heavy metal P-type ATPase HMA5 interacts with metallochaperones and functions in copper detoxification of roots," *The Plant Journal*, vol. 45, no. 2, pp. 225–236, 2005.
- [22] N. Grotz and M. L. Gueriot, "Molecular aspects of Cu, Fe and Zn homeostasis in plants," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1763, no. 7, pp. 595–608, 2006.
- [23] E.W. Sayers, T. Barrett, S.H. Benson, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Res.*, vol. 37, D5-15, 2009.
- [24] M. Goujon, H. McWilliam, W. Li et al., "A new bioinformatics analysis tools framework at EMBL-EBI," *Nucleic acids research*, vol. 38, Suppl: W695-9, 2010.
- [25] A. M. Lesk, *Introduction to Bioinformatics*. Oxford University Press Inc., New York, USA, 2002.
- [26] M.A. Larkin, B. G. Blackshields, N.P. Brown et al., "ClustalW and ClustalX version 2," *Bioinformatics*, vol. 23, no. 21, pp. 2947-2948, 2007.
- [27] L.A. Kelley and M.J.E. Sternberg, "Protein structure prediction on the web: a case study using the Phyre server," *Nature Protocols*, vol. 4, pp. 363 – 371, 2009.
- [28] P. Benkert, M. Biasini and T. Schwede, "Toward the estimation of the absolute quality of individual protein structure models," *Bioinformatics*, vol. 27, no. 3, 2011.
- [29] D. Eisenberg, R. Lüthy and J.U. Bowie, "Assessment of protein models with three-dimensional profiles," *Nature*, vol. 356, no. 6364, 1992.
- [30] R.A. Laskowski, M.W. MacArthur, D.S. Moss and J.M. Thornton, "PROCHECK - a program to check the stereochemical quality of protein structures," *J. App. Cryst.*, vol. 2, pp. 283-291, 1993.
- [31] L. Liu, Z. Zhang, Q. Mei and M. Chen, "PSI: A Comprehensive and Integrative Approach for Accurate Plant Subcellular Localization Prediction," *PLoS ONE*, vol. 8, no.10, 2013.
- [32] J. Schultz, F. Milpetz, P. Bork and C.P. Ponting, "SMART, a simple modular architecture research tool: Identification of signaling domains," *Proc. Natl. Acad. Sci.*, vol. 95, no. 11, pp. 5857-5864, 1998.
- [33] D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: proteinprotein interaction networks, integrated over the tree of life," *Nucleic Acids Res.*, vol. 43, 2015.
- [34] D. Kozakov, D. Beglov, T. Bohnuud et al., "How good is automated protein docking?" *Proteins*, vol. 81, no. 12, pp. 2159-2166, 2013.
- [35] R.L. Cross and V. Muller, "The evolution of A-, F-, and V-type ATP synthases and ATPases: reversals in function and changes in the H⁺/ATP coupling ratio," *FEBS Lett.*, vol. 576, pp. 1-4, 2004.
- [36] K. B. Axelsen and M.G. Palmgren, "Evolution of substrate specificities in the Ptype ATPase superfamily," *J Mol Evol.*, vol. 46, no. 1, pp. 84-101, 1998.
- [37] E.V. Koonin and R.L. Tatusov, "Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity.

Application of an iterative approach to database search,” *J Mol Biol*, vol. 244, no. 1, 1994.

[38] I. Letunic, T. Doerks and P. Bork, “SMART: recent updates, new developments and status in 2015,” *Nucleic Acids Res.*, vol. 43, 2015.

[39] H. Zhou and Y. Zhou, “Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction,” *Protein Sci.*, vol. 11, pp. 2714–2726, 2002.

[40] K. Viehweger, “How plants cope with heavy metals,” *Botanical Studies*, vol. 55, no. 35, 2014.

[41] H. D. Haydon, M. J. Wang et al., “P-Type ATPase Heavy Metal Transporters with Roles in Essential Zinc Homeostasis in Arabidopsis,” *Plant Cell*, vol. 16, no. 5, pp. 1327–1339, 2004.

[42] L. A. Kelley et al., “Phyre2 Protein Fold Recognition Server,” *Nature Protocols*, vol. 10, pp. 845–858, 2015.

[43] P. Benkert, M. Biasini and T. Schwede, “Toward the estimation of the absolute quality of individual protein structure models,” *Bioinformatics*, vol. 27, no. 3, 2011.

[44] K. B. Axelsen and M.G. Palmgren, “Evolution of substrate specificities in the Ptype ATPase superfamily,” *J Mol Evol.*, vol. 46, no. 1, pp. 84–101, 1998.

[45] B. Olaf, S. Vogt, R. Uhlemann, Z. Wiebke and H. Klaus, “Stress induced and nuclear localized HIPP26 from Arabidopsis thaliana interacts via its heavy metal associated domain with the drought stress related zinc finger transcription factor ATHB29,” *Plant Mol Biology*, vol. 69, pp. 213–226, 2009.

[46] Braga de Abreu-Neto, A. C. Joao, L.F.V. Turchetto-Zolet et al., “Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants,” *The FEBS Journal*, vol. 280, no. 7, 2013.

[47] G. Wei, Y.L. Hong, X. Shi and Mee-Len Chye, “Protein interactors of acyl-CoA-binding protein ACBP2 mediate cadmium tolerance in Arabidopsis,” *Plant Signal Behav.*, vol. 5, no. 8, pp. 1025–1027, 2010.

[48] E. Dykema, R. Sipes, A. Marie et al., “A new class of proteins capable of binding transition metals,” *Plant Mol Biol.*, vol. 41, no. 1, 1999.

[49] E. Himelblau, H. Mira, S. Lin et al., “Identification of a Functional Homolog of the Yeast Copper Homeostasis Gene ATX1 from Arabidopsis,” *Plant Physiol.*, vol. 117, no. 4, pp. 1227–1234, 1998.

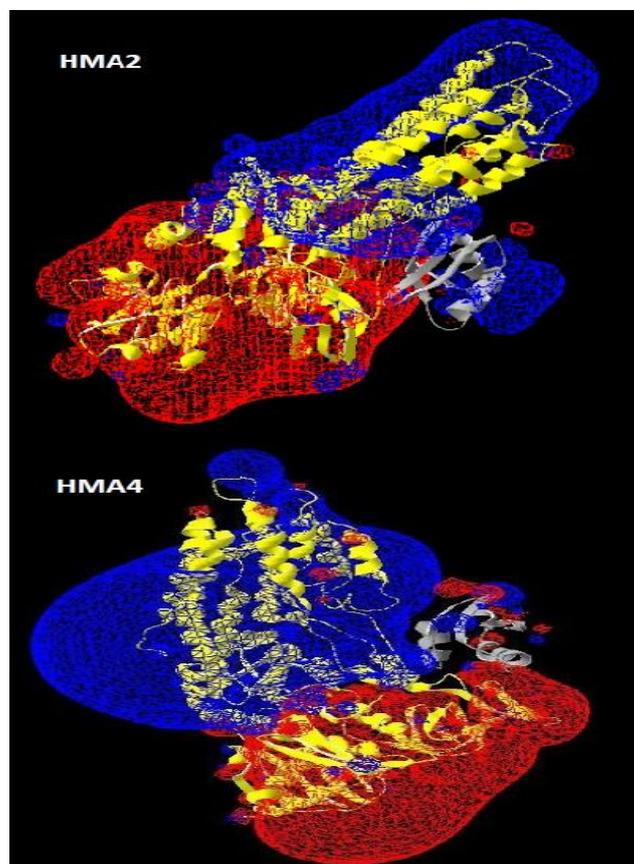
[50] N. J. Robinson and Dennis R. Winge, “Copper Metallochaperones,” *Annu Rev Biochem.*, vol. 79, pp. 537–562, 2010.

[51] S. Lung-Jiun, L. Jing-Chi and Y. Kuo-Chen, “Copper Chaperone Antioxidant Protein1 Is Essential for Copper Homeostasis,” *Plant Physiol.*, vol. 159, no. 3, pp. 1099–1110, 2012.

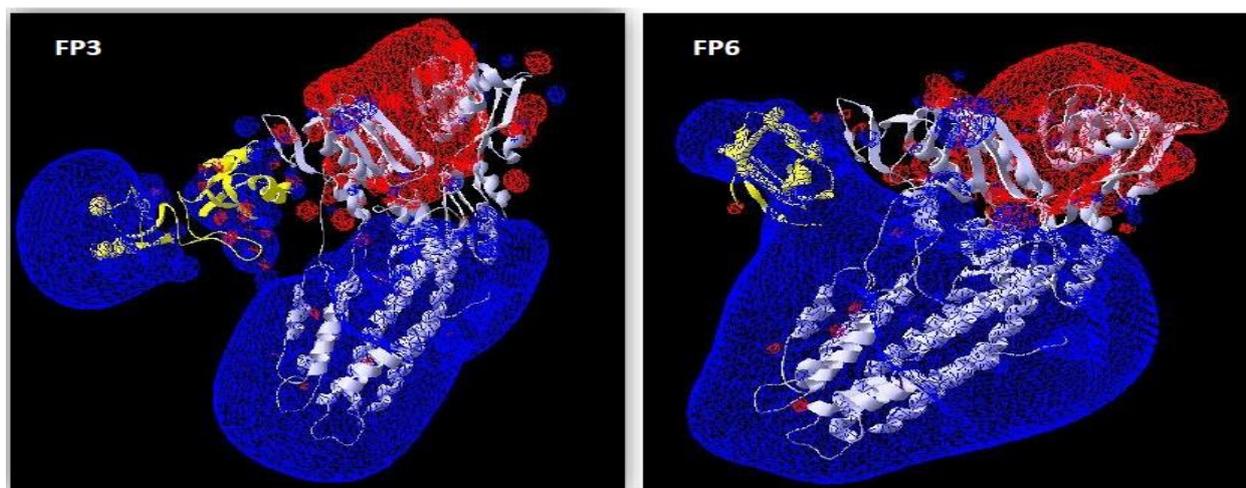
[52] K.C. Wong, E. Jarvis, S. Renée et al., “Functional analysis of the heavy metal binding domains of the Zn/Cd-transporting ATPase, HMA2, in Arabidopsis thaliana,” *New Phytologist Journal*, vol. 181, no. 1, pp. 79–88, 2009

[53] A. Heifetz, E. Katchalski-Katzir and M. Eisenstein, “Electrostatics in protein–protein docking,” *Protein Sci.*, vol. 11, no. 3, pp. 571–587, 2002.

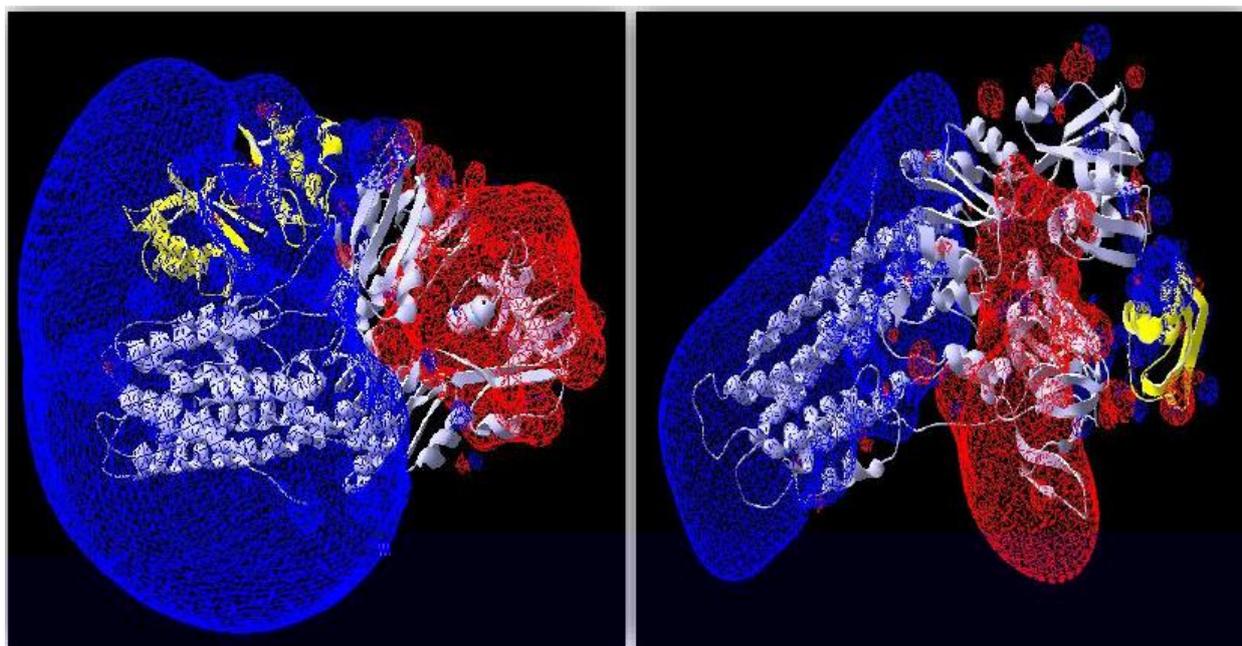
SUPPLEMENTARY DATA



Supplement Figure 1: Visualization of the electrostatic potential of the *HMA2* and *HMA4* with ATX1 docking site as modeled by ClusPro. Yellow=HMA ribbon, grey= ATX1 ribbon, red=negative potential, blue=positive potential



Supplement Figure 2: Visualization of the electrostatic potential of the FP3 and FP6 with *HMA2* docking site as modeled by ClusPro. Yellow= FP3/6 ribbon, grey= HMA2 ribbon, red=negative potential, blue=positive potential



Supplement Figure 3: Visualization of the electrostatic potential of the FP3 and FP6 with *HMA4* docking site as modelled by ClusPro. Yellow= FP3/6 ribbon, grey= HMA2 ribbon, red=negative potential, blue=positive potential

Supplement table 1: Interactome results of Brassica rapa HMA1, HMA2, HMA4 and HMA5 protein family

HMA s	Scores	Brassica rapa interactome ID	Brassica rapa STRING Annotation	Function(s)
HMA1	0.974	Bra035681	Copper-binding family protein	Copper ion binding
HMA4				
HMA1	0.974	Bra033260	Heavy-metal-associated domain-containing protein	Metal ion binding
HMA2				
HMA4				
HMA1	0.975	Bra029854	Arabidopsis homolog of anti-oxidant 1-ATX1	Chaperon protein. Predominant role in delivery of Cu to HMA and Copper homeostasis.
HMA2				
HMA4				
HMA5				
HMA1	0.974	Bra032026	Copper-binding protein-related	Copper ion binding
HMA2				
HMA4				
HMA1	0.975	Bra039371	Hydroxyproline-rich glycoprotein family protein	Involved in copper import and transfer through cells
HMA2				
HMA4				
HMA5				
HMA1	0.975	Bra037865	T15F16.6	Heavy-metal-associated domain-containing protein / copper chaperone (CCH)-related
HMA2				
HMA4				
HMA5				
HMA1	0.974	Bra037919	Metal ion binding	Metal ion binding

HMA2				
HMA4				
HMA5				
HMA1	0.974	Bra033570	HIPP26/FP6/ FARNESYLATED PROTEIN 6	Heavy-metal-binding protein, heat acclimation, Binds lead, cadmium and copper.
HMA2				
HMA4				
HMA1	0.974	Bra038642	FP3/ FARNESYLATED PROTEIN 3	Heavy-metal-binding protein. Binds lead, cadmium and copper. May be involved in heavy-metal transport
HMA2				
HMA4				

Supplement table 2: Interactome results of *Arabidopsis thaliana* HMA3 protein

HMA	Scores	Interactome TAIR ID	STRING Annotation	Functions
ATHMA3	0.961	AT3G56240.1	Copper chaperon (CCH) related.	Stress inducing. Metal ion binding
	0.961	AT5G02600.1	Sodium Potassium Root Defective -NAKKR1	Heavy metal transport/detoxification superfamily protein1
	0.961	AT5G66110.1	HIPP27	Heavy metal associated isoprenylated plant protein 27.Metal Ion Binding
	0.961	AT5G63530.1	Farnesylated protein 3- FP3/ MLE2_16	Heavy-metal-binding protein. Binds lead, cadmium and copper. May be involved in heavy-metal transport
	0.961	AT4G38580.1	Farnesylated protein 6 - FP6/HIPP26	Heavy-metal-binding protein, heat acclimation, Binds lead, cadmium and copper.
	0.961	AT5G60800.1	Heavy metal transport/detoxification domain-containing proteins	Heavy metal transport/detoxification domain-containing protein
	0.961	AT5G37860.1		
	0.961	AT5G27690.1		
	0.961	AT4G39700.1		
	0.961	AT5G03380.1		
	0.961	AT5G19090.1		
0.961	AT5G24580.1			

Supplement table 3: Verification analysis of all chosen docking candidates with HMA2 and HMA4

Name	QMEAN6 score	QMEAN6 score Z-score	DFire energy	PROCHECK most favored region in %
HMA2-ATX1	0.511	-2.795	-1007.50	85.0
HMA2-FP6	0.454	-3.413	-1006.70	84.2
HMA2-FP3	0.313	-4.979	-1124.00	82.08
HMA4-FP3	0.537	-4.484	-1111.77	83.0
HMA4-FP6	0.454	-3.413	-1006.70	84.2
HMA4-ATX1	0.484	-3.091	-996.22	84.9