In Silico Characterization of Heavy Metal Binding Protein Predicted In the Genome Sequence of *Lactobacillus pentosus* KCA1

*Kingsley C Anukam [1, 2], Humphery B Osadolor [2], Ikemefuna C Uzochukwu [3]

1. TWAS Genomics Research Unit, Department of Medical Laboratory Science, School of Basic Medical Sciences, University of Benin, Nigeria.
2. Department of Medical Laboratory Science, School of Basic Medical Sciences, College of Medical Sciences, University of Benin, Nigeria.
3. Department of Pharmaceutical and Medicinal Chemistry, Faculty of Pharmaceutical Sciences, Nnamdi Azikiwe University, Awka 420281, Anambra State, Nigeria.

ABSTRACT

Heavy metals, such as lead, cadmium, arsenic and mercury can contaminate the food chain and pose serious health problems. We characterized the heavy metal binding protein predicted in the genome sequence of *Lactobacillus pentosus* KCA1. Bioinformatic tools such as BLASTp, ClustalW and iterative threading assembly refinement (I-TASSER) server were used for protein sequence similarity, secondary structure and 3-D model prediction for potential active binding sites. Two cysteine residues were identified at position 12 and 15 (Cys12 and Cys15), separated by proline (P) and serine (S). The secondary structure prediction revealed 2 alpha helices and 5 beta-strands. The 3-D model structure of KCA1_2798 has a confidence score of 1.26 that reflects a model of good quality with binding site residues predicted to occur at five positions (Thr11, Cys**12**, Ser14, Cys**15**, Gly63) of the KCA1_2798 sequence based on a protein (3iwlA) template. The heavy metal binding protein encoded in the genome sequence of *L. pentosus* KCA1 has laid a foundation for further investigation into its potential health applications.

KEY WORDS: *Lactobacillus pentosus* KCA1, metal-binding proteins, genome sequence.

INTRODUCTION

Heavy metals such as lead, cadmium, arsenic and mercury are widely distributed contaminants with varying concentrations, which accumulate through the food chain. The provisional tolerable weekly intake set by the World Health organization (WHO) stands at 25 and 7 μg/kg body weight for lead and cadmium respectively (JECFA, 2004). High concentrations of heavy metals are of particular concern to occupational workers such as painters (Sussell *et al.*, 1999), polluted areas (Dauwe *et al.*, 2004) and people with nutritional deficiencies (Andersen *et al*., 2004). Lead and cadmium are widely spread in nature and their various detrimental health effects have been well documented (Rosen, 1995; Satarug and Moore 2004). The toxicological properties of cadmium are associated with lack of biochemical elimination methods and its long half-life of about 10 years in the body (Satarug and Moore 2004). Cadmium has also been classified as a group 1 human carcinogen by the International Agency for Research on Cancer (IARC, 2012). Lead has various acute and chronic adverse effects on the health of humans and animals, including neurobehavioral-cognitive problems (Rosen 1995), heavy-metal-induced anaemia, generation of reactive oxygen species and alteration of antioxidant defense systems (Hsu and Guo, 2002), reduced fertility and blood-brain barrier dysfunction (Struzynska *et al.*, 1997). Even modern civilization has contributed to increased exposure as it has been reported that lead and cadmium migrate from ceramic materials into foods (Gonzalez-Soto *et al.*, 2000). In agricultural farmlands, cadmium levels in plants cultivated in polluted areas were found to be 6-9 times higher than in plants cultivated in non-polluted areas (Muller and Anke, 1994). Hand-to-mouth behaviour among children may increase ingestion of lead and cadmium from soil sources, including occupational painters that work with materials containing heavy metals. Poor and malnourished populations may be more vulnerable to the impacts of modern environmental health hazards, given that malnutrition increases susceptibility to toxicologic challenge (Tillet, 2009). Recently in 2012, hundreds of children lost their lives as a result of lead poisoning in Zamfara state in the Northern part of Nigeria, due to illegal mining of heavy metals. Heavy metal hazards abound in old leaded paints in Nigerian homes, schools, and workplaces and from old pipes that still carry municipal drinking water (Adebamowo *et al.*, 2006).

Heavy metal binding capability of some strains of lactobacilli has been mostly directed at waste management (Halttunen *et al.*, 2003). A pilot study has demonstrated the capacity of food-grade probiotic strains to bind heavy metals *in vitro* by identifying factors that affect the binding capacity (Ibrahim *et al.*, 2006). A recent study evaluated the protective effects of *Lactobacillus plantarum* CCFM8610, a selected probiotic with good cadmium binding capacity, against acute cadmium toxicity in mice (Zhai *et al*., 2013). A review of the literature has suggested the potential applications of *Lactobacilli* in bioremediation and detoxification of environmental contaminants (Monachese *et al.*, 2012). In an effort to explore novel ways of reducing the uptake of lead and cadmium that is ingested in our environment, we examined food grade bacteria for their metal binding capacity in a pilot study by ingestion of yogurt containing strain of *Lactobacillus pentosus* KCA1, isolated from a Nigerian subject. The result showed an apparent reduction in the blood concentrations of lead in subjects occupationally exposed to lead (Osadolor *et al.*, 2013). The genome of *L. pentosus* KCA1 has open reading frame coding for heavy metal binding protein and clusters of genes for exopolysaccharides (EPS) biosynthesis, which might aid in heavy metal binding (Anukam *et al.*, 2013).

The objectives of the present study are to use bioinformatic tools to analyze the heavy metal binding protein encoded in the genome sequence of *Lactobacillus pentosus* KCA1 and second to determine the 3-D model structure of the protein and potential active binding site residues.

*** Corresponding author: E-mails: kanukam@gmail.com; Kingsley.anukam@uniben.edu. Tel: +234 7060921660**

MATERIALS AND METHODS

Location of the heavy metal binding protein

The Ensembl genome annotation system developed jointly by the European Buoinformatic Institute (EBI) and the Wellcome Trust Sanger Institute was used for the location, extraction of the nucleotide base sequence or open reading frame (ORF) and the amino acid translation of the heavy metal binding protein from L. pentosus KCA1 (KCA1_2798) (http://ensemblgenomes.org/id/EIW12399). Sequences similar to L. pentosis KCA1 were searched for in UniProt® database using BLASTp algorithm.

Multiple sequence alignments

The amino acid translations from the nucleotide bases of 17 bacterial organisms were selected from the BLASTp of UniProt® database (http://www.uniprot.org/) based on product annotation hit (heavy metal binding protein), gene name, % identity, matrix score and E-value. These 17 amino acids sequences along with the sequence of L. pentosus KCA1 (http://www.uniprot.org/uniprot/I9KWT3) were imported into the ClustalW algorithm for multiple sequence alignments.

Metal binding sites predictor

The amino acid sequence was submitted to the metal binding site predictor, (http://metaldetector.dsi.unifi.it/v2.0/), which looks for cysteine and histidine residues as ligands. Prediction is limited to transition metals (with the addition of heme and Fe/S clusters). The Metal Detector predicts the number of bound metal ions and, for each ion, and the number of CYS (Cysteine) and HIS (Histidine) ligands in the sequence (Passerini et al., 2011).

Prediction of the transmembrane region of the heavy metal binding protein

Transmembrane regions in peptides were deduced using the SOSUI program [Hirokawa et al., 1998] (http://bp.nuap.nagoya-u.ac.jp/sosui/) and PRED-TMBB, a web server for predicting the topology of beta-barrel outer membrane proteins [Cole et al., 2008].

Prediction of secondary structure, 3-D model, similarity structure in PDB, functional and binding sites predictions with I-TASSER.

The iterative threading assembly refinement (I-TASSER) server is a four stage integrated platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm (Roy et al., 2010). The amino acid sequence was submitted online (Yang, 2008) for the prediction of the 3D structure, similar structures in PDB, function and the binding site by this integrated algorithm. .

RESULTS

The heavy metal binding protein is located in contig AKAO01000083.1 from the DNA assembly and found between 3,165,352 and 3,165,570 along the L. pentosus KCA1 chromosome (figure 1). The amino acid composition of the gene coding for the heavy metal binding protein (72 amino acid residues) from L. pentosus KCA1 comprised of 216 nucleotide base sequence with a molecular mass of 7842 Dalton. The protein belonged to the protein family (Pfam) number Pf00403. The Basic Local Alignment Search Tool for proteins (BLASTp) from the UNIPROTEIN database yielded 250 hits with KCA1_2798 amino acid sequence. Seventeen organisms were selected based percentage identity (94-100 %) and e-value cutoff of 9.0×10-48. (Table1). Three Lactobacillus species (L. plantarum EGD-AQ4, L. rhamnosus ATCC 8530 and L. pentosus MP-10) had 99 % amino acid sequence identity to KCA1_2798 with the same e-value of 4.0X10-50.

ClustalW multiple sequence alignments showed that L. pentosus KCA1 metal binding protein had significant amino acid sequence identity with the metal binding proteins from the selected Lactobacillus strains (figure 2). A unique amino acid substitution was observed in KCA1 sequence at position 54 with Threonine (T-Thr), while all the other strains have Alanine (A-Ala) at the same position. Five genes including; I9KWT3 (L. pentosus_KCA1), T5JRS1 (L. plantarum_EGD-AQ4), G7UZN5 (L. rhamnosus_ATCC_8530), G0M4Y9 (L. pentosus_IG1) and F6IXZ0 (L. pentosus_MP-10) had serine substitution at position Ser26 and Ser51 and glutamic acid at Glu29. The remaining organisms had Alanine at position 26, (Ala26), Aspartic acid at position 29 (Asp29), and Asparagine at position 51 (Asn51). Two cysteine residues were identified in all the strains at position 12 and 15 (Cys12 and Cys15), separated by proline (P) and serine (S).

The secondary structure prediction with I-TASSER revealed 2 alpha helices and 5 beta-strands as shown in figure 3. The first □-helix started at position 13 (P-Proline) and ends at position 23 (V-valine) and the second □-helix has 10 amino acid residues at position Ala51 to Gly60. It appears the longest beta-strand occurred between position 28 (V-valine) and 34 (L-leucine) while the shortest □-strand has only 2 amino acid residues at position Leu10 and Thr11.

The 3-D model of KCA1_2798 has a C-score of 1.26 and an estimated accuracy of 0.89 ± 0.07 (TM-score-template modeling score) and 1.1 ± 1.1 Å (RMSD-root mean square deviation) based on the 10 templates used for alignments with 3dxsX-PDB having the top normalized Z-score of 4.02. The co-ordinate file of the L. pentosus KCA1 heavy metal binding protein model was downloaded in PDB format and Jmol molecular visualization program (Hanson, 2010) was used to view the predicted structure as shown in figure 4. Proteins with highly similar structure in PDB as identified by TM-align are shown in Table 2. The protein 3dxsX PDB-hit as the top rank has a TM-score of 0.975 and a coverage of 1.000.

Five enzyme homologues were identified in PDB as having similar functions to the predicted KCA1_2798 sequence (Table 3). Notably, 1mwyA PDB-hit has the top rank with confidence score of 0.541 for the Enzyme Classification (EC) number (3.6.3.3 and 3.6.3.5). The predicted active site residues were identified as residues at position 18, 35 and 40 in the KCA1_2798 amino acid sequence.

Predicted gene ontology (GO) terms (Table 4) associated with the KCA1_2798 query sequence identified 1fe0B from the PDB-hit with 12 GO terms. This protein has the top GO confidence score of 0.69 and TM-score of 0.8063. One template protein (3iwlA) with similar binding site residues was predicted to occur at positions 11,12,14,15,63 of the KCA1_2798 sequence (Table 5).

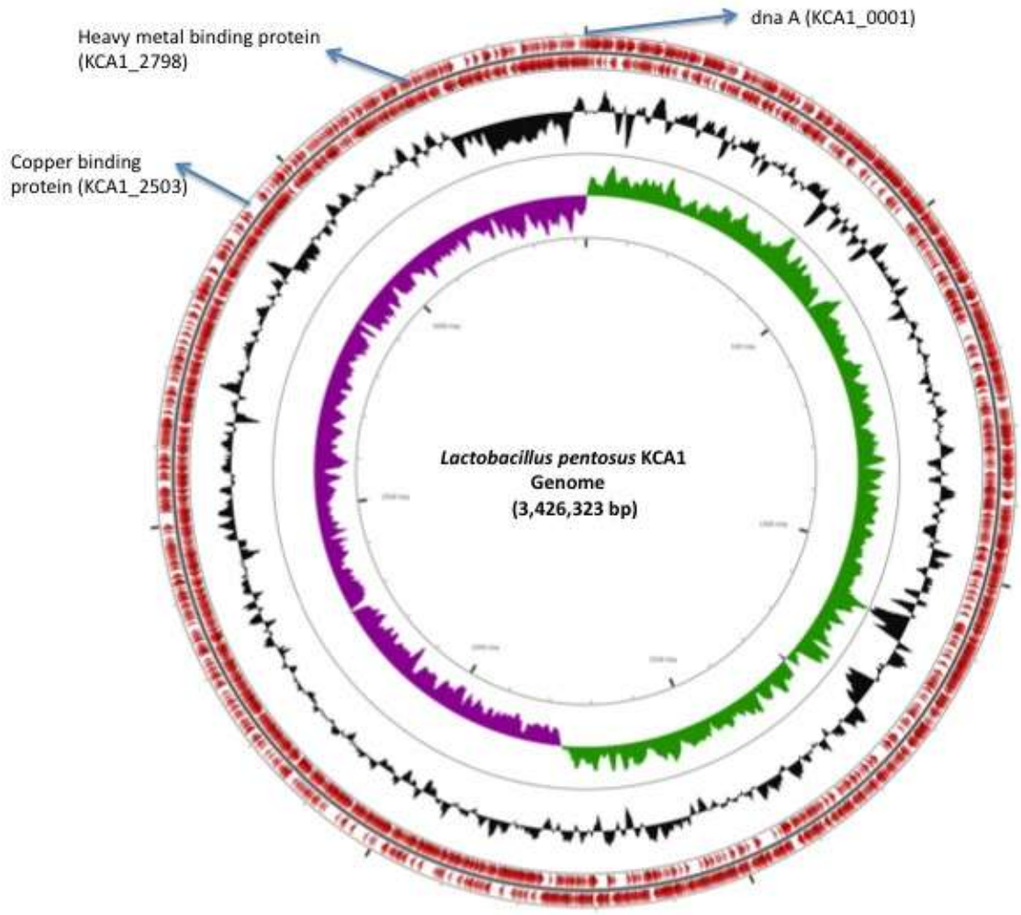

Figure 1: Genome atlas view of L. pentosus KCA1
showing the location of the heavy metal binding protein (KCA1_2798).
The red arrows shows the open reading frames,
Black (GC content), Green (GC skew +, Purple (GC skew-)

```
R9X7H5_L.plantarum_16           MIMKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 50
H3NZ86_L.plantarum_NC8          MIMKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 50
U2H776_L.plantarum_2165         MIMKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 50
C6VMM1_L.plantarum_JDM1         MIMKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 50
E1TTC9_L.plantarum_ST-III       MIMKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 50
F9UUB8_L.plantarum_WCFS1        MIMKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 50
D7VEZ3_L.plantarum_ATCC_14917   --MKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 48
M4KEH0_L.plantarum_ZJ316        --MKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 48
M7C4Y4_L.plantarum_UCMA_3037    --MKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 48
R4Q7J2_L.plantarum_P-8          --MKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 48
S2W4I7_L.plantarum_IPLA88       --MKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 48
U2WJF8_L.plantarum_AY01         --MKIIMQLGTLTCPSCMTKIEKAVANHDGVENVKVLFNASKVKANFDPE 48
I9KWT3_L.pentosus_KCA1          --MKIIMQLGTLTCPSCMTKIEKAVSNHEGVENVKVLFNASKVKANFDPE 48
T5JRS1_L.plantarum_EGD-AQ4      --MKIIMQLGTLTCPSCMTKIEKAVSNHEGVENVKVLFNASKVKANFDPE 48
G7UZN5_L.rhamnosus_ATCC_8530    --MKIIMQLGTLTCPSCMTKIEKAVSNHEGVENVKVLFNASKVKANFDPE 48
G0M4Y9_L.pentosus_IG1           --MKIIMQLGTLTCPSCMTKIEKAVSNHEGVENVKVLFNASKVKANFDPE 48
F6IXZ0_L.pentosus_MP-10         --MKIIMQLGTLTCPSCMTKIEKAVSNHEGVENVKVLFNASKVKANFDPE 48
                                  **********************:**:**********************

R9X7H5_L.plantarum_16           VTNADDLAQVVTGLGYEVENVKVK 74
H3NZ86_L.plantarum_NC8          VTNADDLAQVVTGLGYEVENVKVK 74
U2H776_L.plantarum_2165         VTNADDLAQVVTGLGYEVENVKVK 74
C6VMM1_L.plantarum_JDM1         VTNADDLAQVVTGLGYEVENVKVK 74
E1TTC9_L.plantarum_ST-III       VTNADDLAQVVTGLGYEVENVKVK 74
F9UUB8_L.plantarum_WCFS1        VTNADDLAQVVTGLGYEVENVKVK 74
D7VEZ3_L.plantarum_ATCC_14917   VTNADDLAQVVTGLGYEVENVKVK 72
M4KEH0_L.plantarum_ZJ316        VTNADDLAQVVTGLGYEVENVKVK 72
M7C4Y4_L.plantarum_UCMA_3037    VTNADDLAQVVTGLGYEVENVKVK 72
R4Q7J2_L.plantarum_P-8          VTNADDLAQVVTGLGYEVENVKVK 72
S2W4I7_L.plantarum_IPLA88       VTNADDLAQVVTGLGYEVENVKVK 72
U2WJF8_L.plantarum_AY01         VTNADDLAQVVTGLGYEVENVKVK 72
I9KWT3_L.pentosus_KCA1          VTSADDLTQVVTGLGYEVENVKVK 72
T5JRS1_L.plantarum_EGD-AQ4      VTSADDLAQVVTGLGYEVENVKVK 72
G7UZN5_L.rhamnosus_ATCC_8530    VTSADDLAQVVTGLGYEVENVKVK 72
G0M4Y9_L.pentosus_IG1           VTSADDLAQIVTGLGYEVENVKVK 72
F6IXZ0_L.pentosus_MP-10         VTSADDLAQVVTGLGYEVENVKVK 72
                                **.****:*:**************
```

Figure 2: ClustalW multiple amino acid sequence alignments of the metal binding protein from the selected Lactobacillus strains.

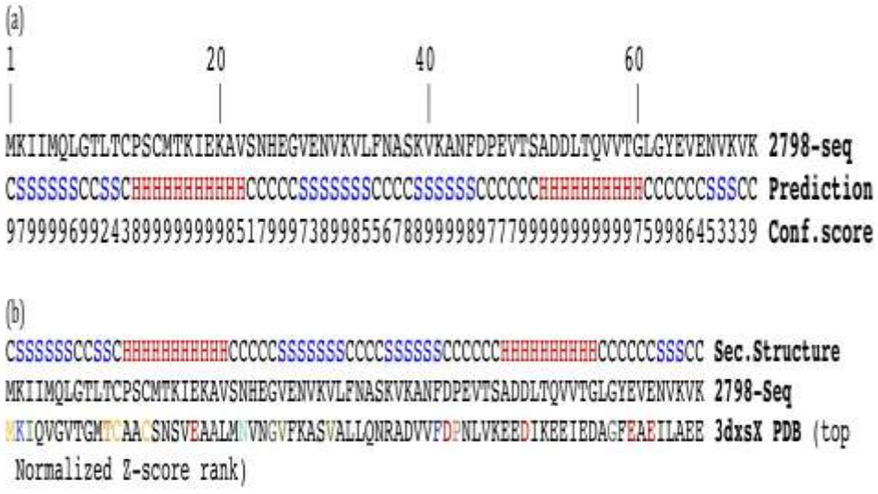

Figure 3:The secondary structure of KCA1_2798 prediction with I-TASSER (a) revealing 2 alpha helices (Red H) and 5 beta-strands (Blue S), (b) alignment with 3dxsX protein template with top rank Z-score.

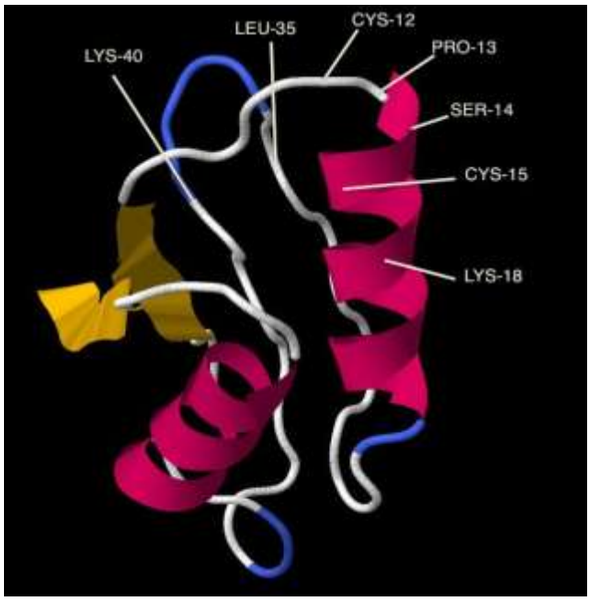

Figure 4: The 3-D model structure of KCA1_2798 as predicted by I-TASSER

(visualized with Jmol program) showing the location of the active binding site residues.

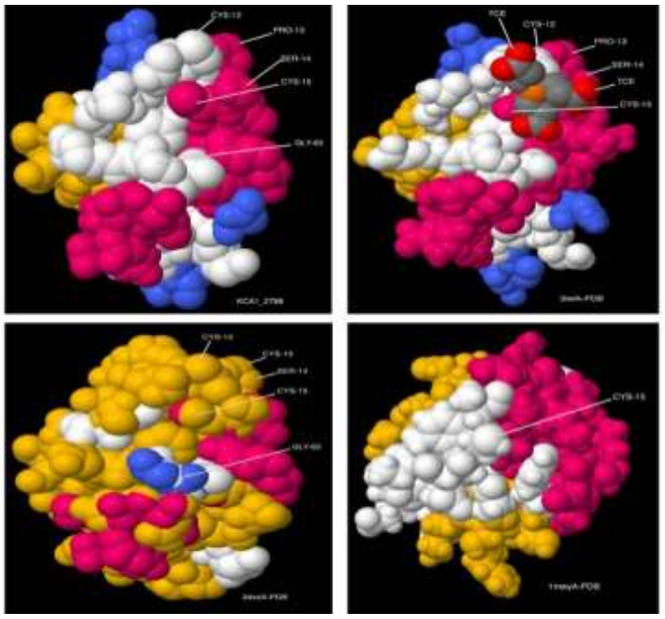

Figure 5: Comparative structure of 100% van der waal forces (visualized with Jmol program) between KCA1_2798 heavy metal binding protein and the top rank proteins in the PDB for similar binding site ligands (3iwlA), functional enzyme

Table 1: Heavy metal binding protein amino acid sequence identity as determined from the UniProt® database

Organism	Gene entry	Protein names	Length (AA)	Identity	Score	E-value	Gene name
L. pentosus KCA1	I9KWT3	Heavy metal binding protein	72	100.00%	388	6.0×10-51	KCA1_2798
L. plantarum EGD-AQ4	T5JRS1	Heavy metal binding protein	72	99.00%	383	4.0×10-50	N692_07270
L. rhamnosus ATCC 8530	G7UZN5	Heavy-metal-associated domain	72	99.00%	383	4.0×10-50	LRHK_2625
L. pentosus MP-10	F6IXZ0	Uncharacterized protein	72	99.00%	383	4.0×10-50	LPE_02587
L. pentosus IG1	G0M4Y9	Uncharacterized protein	72	97.00%	382	6.0×10-50	LPENT_01987
L. plantarum AY01	U2WJF8	Uncharacterized protein	72	94.00%	369	8.0×10-48	N644_2552
L. plantarum IPLA88	S2W4I7	Heavy metal binding protein	72	94.00%	369	8.0×10-48	L103_16023
L. plantarum P-8	R4Q7J2	Copper chaperon	72	94.00%	369	8.0×10-48	LBP_cg2745
L. plantarum UCMA 3037	M7C4Y4	Heavy metal binding protein	72	94.00%	369	8.0×10-48	H073_06708
L. plantarum ZJ316	M4KEH0	MerTP mercury (Hg2+) permease	72	94.00%	369	8.0×10-48	zj316_0076
L. plantarum ATCC 14917 HMPREF0531_12705	D7VEZ3	Heavy metal-associated protein	72	94.00%	369	8.0×10-48	
L. plantarum WCFS1	F9UUB8	Heavy metal binding protein	74	94.00%	369	9.0×10-48	lp_3442
L. plantarum ST-III	E1TTC9	Putative uncharacterized protein	74	94.00%	369	9.0×10-48	LPST_C2819
L. plantarum JDM1	C6VMM1	Putative uncharacterized protein	74	94.00%	369	9.0×10-48	JDM1_2743
L. plantarum 2165	U2H776	Heavy metal binding protein	74	94.00%	369	9.0×10-48	N574_01535
L. plantarum 16	R9X7H5	Heavy metal binding protein	74	94.00%	369	9.0×10-48	Lp16_2696
L. plantarum NC8	H3NZ86	Heavy metal binding protein	74	94.00%	369	9.0×10-48	nc8_2920

Table 2: Proteins with highly similar structure in PDB as identified by ™-align (Top 10 identified structural analogs in PDB)

Rank	PDB-Hit	TM-score	RMSDa	IDENa	Cov.
1	3dxsX	0.975	0.49	0.250	1.000
2	3cjkB	0.916	0.93	0.194	1.000
3	1aw0_	0.902	0.86	0.243	0.972
4	1fvsA	0.899	0.98	0.169	0.986
5	1yjvA	0.876	1.29	0.278	0.986
6	2l3mA	0.863	0.86	0.299	0.931
7	1cpzA	0.862	1.00	0.294	0.944
8	2ropA2	0.848	1.16	0.261	0.958
9	4a48A	0.844	0.84	0.277	0.903
10	1kvjA	0.844	1.38	0.194	0.986

(a) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(b) RMSDa is the RMSD between residues that are structurally aligned by TM-align.
(c) IDENa is the percentage sequence identity in the structurally aligned region.
(d) Cov. represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Table 3: Function prediction using CO-FACTOR (Predicted EC numbers)-Top 5 enzyme homologs in PDB.

Rank	$Cscore_{EC}$	PDB-Hit	TM-score	$RMSD_a$	$IDEN_a$	Cov	EC NumberPredicted	Active Site Residues
1	0.541	1mwyA	0.830	1.38	0.229	0.972	3.6.3.3 3.6.3.5	18,35,40
2	0.525	1p6tA	0.823	1.51	0.306	0.986	3.6.3.-	20,31,47,54
3	0.518	1y3jA	0.778	1.87	0.211	0.986	3.6.3.4	6,20,40,63
4	0.514	2kkhA	0.823	1.56	0.167	0.972	3.6.3.3 3.6.3.5	20,65
5	0.512	3cjkA	0.815	1.21	0.194	0.931	3.6.3.4	29,40

a) CscoreEC is the confidence score for the Enzyme Classification (EC) number prediction. CscoreEC values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.
(b) TM-score is a measure of global structural similarity between query and template protein.
(c) RMSDa is the RMSD between residues that are structurally aligned by TM-align.
(d) IDENa is the percentage sequence identity in the structurally aligned region.
(e) Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

Table 4: Predicted GO Terms

Rank	$Cscore_{GO}$	TMscore	$RMSD_a$	$IDEN_a$	Cov.	PDB-Hit	Associated GO Terms
1	0.69	0.8063	1.25	0.19	0.93	1fe0B	GO:0005507 GO:0006811 GO:0030001 GO:0016530 GO:0006825 GO:0016531 GO:0005829 GO:0046872 GO:0006878 GO:0006810 GO:0032767 GO:0006979
2	0.65	0.7637	1.45	0.33	0.92	1yg0A	GO:0035434 GO:0046872 GO:0005507 GO:0030001 GO:0005375 GO:0006825
3	0.61	0.8629	0.86	0.30	0.93	2l3mA	GO:0005375 GO:0005507 GO:0006825 GO:0030001 GO:0035434 GO:0046872
4	0.60	0.8231	1.51	0.31	0.99	1p6tA	GO:0005507 GO:0006825 GO:0015097 GO:0015694 GO:0030001 GO:0046872
5	0.59	0.8619	1.00	0.29	0.94	1cpzA	GO:0005737 GO:0046872 GO:0004008 GO:0005507 GO:0005524 GO:0006754 GO:0006825 GO:0016020 GO:0030001 GO:0060003
6	0.56	0.8253	1.13	0.24	0.92	2roeA	GO:0046872 GO:0030001
7	0.56	0.7415	1.61	0.19	0.89	2k2pA	GO:0030001 GO:0046872
8	0.56	0.7795	1.87	0.28	0.96	2kt2A	GO:0030001 GO:0046872
9	0.55	0.9750	0.49	0.25	1.00	3dxsX	GO:0004008 GO:0005524 GO:0006754 GO:0016020 GO:0030001 GO:0046872 GO:0060003
10	0.54	0.7876	1.43	0.22	0.93	2kyzA	GO:0046872 GO:0030001

(a) CscoreGO, which is a combined measure for evaluating global and local similarity between query and template protein. CscoreGO values range in between [0-1]; where a higher value indicates a better confidence in predicting the function using the template.
(b) TM-score is a measure of global structural similarity between query and template protein.
(c) RMSDa is the RMSD between residues that are structurally aligned by TM-align.
(d) IDENa is the percentage sequence identity in the structurally aligned region.
(e) Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

Table 5: Template proteins with similar binding site:

Rank	$Cscore_{LB}$	PDB-Hit	TM-score	RMSDa	IDENa	Cov.	BS-score	Lig. Name	Predicted binding site residues
1	0.14	3iwlA	0.792	1.29	0.167	0.917	1.31	TCE	11,12,14,15,63

(a) $Cscore_{LB}$ is the confidence score of predicted binding site. CscoreLB values range in between [0-1]; where a higher score indicates a more reliable ligand-binding site prediction.
(b) BS-score is a measure of local similarity (sequence & structure) between template binding site and predicted binding site in the query structure. Based on large scale benchmarking analysis, we have observed that a BS-score >1 reflects a significant local match between the predicted and template binding site.
(c) TM-score is a measure of global structural similarity between query and template protein.
(d) $RMSD_a$ the RMSD between residues that are structurally aligned by TM-align.
(e) $IDEN_a$ is the percentage sequence identity in the structurally aligned region.
(f) Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

The template protein 3iwlA from PDB-hit has TCE (3, 3', 3''-phosphanetriyltripropanoic acid) as the ligand name and ligand-binding site prediction confidence score of 0.14. Figure 5 shows the comparative structure of 100 % van der waal forces between KCA1_2798 heavy metal binding protein and the top rank proteins in the PDB for similar binding site ligands (3iwlA), functional enzyme homologues (1mwyA) and highly similar structure (3dxsX)

DISCUSSION

The heavy metal binding protein from L. pentosus KCA1 was found to be associated with cluster of orthologous group (COG) of proteins in class P responsible for inorganic ion transport and metabolism. The protein family identity number of PF00403.19 indicates that the protein is strongly linked with heavy metal associated domain (HMA) with e-value of 3.7e-13. HMA is a conserved protein domain found in a number of heavy metal transport or detoxification proteins (Bull and Cox, 1994). It has been suggested that proteins that transport heavy metals in microorganisms, including bacterial heavy-metal-resistance proteins and mammals share similar identity in their sequences and structures (Gitschier et al., 1998).

The heavy metal binding protein of L. pentosus KCA1 was found to be closely related to the protein sequence of one pentosus strain notably, L. pentosus MP-10 with 99 % amino acid sequence identity (Abriouel et al., 2011). Two Lactobacillus species, L. plantarum EGD-AQ4 and L. rhamnosus ATCC 8530 has the same 99 % amino acid identity. The 1 percent difference is a unique amino acid substitution observed in KCA1_2798 sequence at position 54 with Threonine (T-Thr), while other selected Lactobacillus species have Alanine (A-Ala) at the same position. The functional differences in this substitution remains to be determined.

The HMA domain contains two conserved cysteine residues that are previously believed to be involved in metal binding. Interestingly, two cysteine residues were identified in all the strains thus suggesting that the protein is universal in nature and has the ability to interact with heavy metals. The conserved cysteine domain appears to occur at position 12 and 15 (Cys12 and Cys15) in the L. pentosus KCA1

sequence. It is still debatable that these genes containing HMA have any function in metal binding or detoxification. Though they may serve a biological purpose, it is most likely inconclusive to assume that they have a function in detoxification. This assertion is based on the fact that the heavy metal binding protein is not a membrane protein but a soluble protein as predicted by SOSUI (Hirokawa et al., 1998). Over the past decade, several authors have studied interactions of Lactobacilli with lead and cadmium (Halttunen et al., 2008 ; Mrvcic et al., 2009). The data shows that binding of metals by lactobacilli is facilitated by the cell wall exopolysaccharides and not cytosolic proteins. Previous reports have revealed that some Lactobacilli, including L. rhamnosus, L. plantarum, and L. brevis, can bind and remove heavy metals such as cadmium, lead, and copper in vitro (Halttunen et al., 2008 ; Mrvcic et al., 2009). A recent study has shown that Glutamic acid and Aspartic acid carboxylate side chain groups were found to be the main interacting molecular fragments with the metal ions in both Lactobacillus kefir CIDCA 8348 and JCM 5818 S-layers (Gerbino et al., 2011). This result is in agreement with previously obtained data for the interaction of Bacillus sphaericus JG-A12 with Pd2+ (Fahmy, 2006). The amino acid sequence of KCA1_2798 contains six Glutamic acid residues and three Aspartic acids. It will be interesting to determine if these molecular fragments have any effect on the functional activities of KCA1_2798 gene.

The secondary structure of heavy metal binding protein from KCA1-2798 indicates prediction with higher confidence as the confidence scores for both alpha helices and beta strands are high (figure 3). The confidence values are shown for each residue ranging between 0 and 9, in which a higher score indicates a prediction with strong confidence (Jones, 1999). The 3-D structure of KCA1_2798 has a confidence score (C-score) of 1.26 that reflects a model of good quality (Figure 3 and 4). The C-score was based on 3dxsX PDB. This protein contains three metal binding domains of which HMA-7 features a CxxC sequence motif, similar to CPSC motif of KCA1_2798, characteristic of Cu(I) binding sites. Those of HMA-2 and HMA-4 contain a CCxxE motif, unique for plant Zn(2+)-ATPases (Zimmermann et al., 2009). It remains to be determined if the CPSCMTKIE motif and DDLTQ of KCA1_2798 are possible metal binding domain specific to L. pentosus KCA1.

The enzyme homologue with top rank shows that 1mwyA PDB-hit produced a confidence score of 0.541 for the Enzyme Classification (EC) number 3.6.3.3 (Cadmium exporting ATPase) and 3.6.3.5 (Zinc exporting ATPase) relative to KCA1_2798. This protein contains a previously unknown protein coordination site for zinc that includes two cysteine residues, Cys59 and Cys62, and a carboxylate residue, Asp58 (Banci et al., 2002). It may probably be possible that any of the Asp residues in KCA1_2798 may play an important role in modulating the relative affinities and metal exchange rates for Zn (II)/Pb(II)/Cd(II) similar to ImwyA. However, the predicted active site residues in KCA1_2798 relative to the enzyme homologue 1mwyA PDB indicated Lys18, Lys40 and Leu35.

The predicted gene ontology (GO) terms for KCA1_2798 identified 1fe0B PDB with 12 GO terms for biological functions, of which the N-termini of its target proteins belong to a family of metal binding domains characterized by a conserved MT/HCXXC sequence motif specific for binding of copper and cadmium (Wernimont et al., 2000). The implication of high number of GO terms associated with KCA1_2798 suggests that the protein may have biological attributes interacting with various heavy metals. The template protein (3iwlA) with similar binding site residues was predicted to occur at five positions (Thr11, Cys12, Ser14, Cys15, Gly63) of the KCA1_2798 sequence. The crystal structures revealed conserved CXXC copper-binding motif associated with this protein, which may be similar to domain Cys12XXCys15 present in KCA1_2798. (Boal and Rosenzweig, 2009).

In conclusion, in silico analysis has shown that the heavy metal binding protein predicted in the L. pentosus KCA1 genome sequence has two cysteine residues identified in most of the microbial and mammalian cells. This suggests that the protein is universal in nature and has the ability to interact with heavy metals as it possesses the CXXC domain associated with heavy metal binding. Although recent findings appear to present other motifs that may be significant in binding. In this regard based on our findings, we propose that the CPSCMTKIE motif and DDLTQ motif of KCA1_2798, are possible metal binding domain specific to L. pentosus KCA1 as it has the two HMA cysteine residues, and presence of glutamic acid, and aspartic acid which may aid in binding of heavy metals. The characterization of the heavy metal binding protein encoded in the genome sequence of Lactobacillus pentosis KCA1, its 3-D structure and potential active site residues predictions have laid the foundation for further investigations into the use of L. pentosis KCA1 as probiotics for the mitigation of heavy metals uptake in susceptible or occupationally exposed persons.

ACKNOWLEDGMENT

Dr. Anukam KC research is partly supported by the Third World Academy of Sciences (TWAS), under the RESEARCH GRANT AGREEMENT (RGA) No.09-017RG/BIO/AF/AC_G-UNESCOFR:3240230312

REFERENCES

Abriouel H, Benomar N, Perez Pulido R, Canamero MM, Galvez A. (2011). Annotated genome sequence of Lactobacillus pentosus MP-10, which has probiotic potential, from naturally fermented Alorena green table olives. Journal of Bacteriology. 193: 4559–4560.

Adebamowo EO, Agbede OA, Sridhar MKC, Adebamowo CA. (2006).An evaluation of lead levels in residential paints sold in Nigeria markets. Indoor Built Environment.15:551–4.

Andersen O, Nielsen JB, Nordberg GF. (2004). Nutritional interactions in intestinal cadmium uptake-possibilities for risk reduction. Biometals. 17:543-547.

Anukam KC, Macklaim JM, Gloor GB, Reid G, Boekhorst J, Renckens B, van Hijum SA, Siezen RJ. (2013). Genome sequence of Lactobacillus pentosus KCA1: vaginal isolate from a healthy premenopausal woman. PLoS One. 8:e59239.

Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ. (2004). PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. Nucleic Acids Research. 32: W400-W404.

Banci L, Bertini I, Ciofi-Baffoni S, Finney LA, Outten CE, O'Halloran TV. (2002). A new zinc-protein coordination site in intracellular metal trafficking: solution structure of

the Apo and Zn(II) forms of ZntA(46-118). Journal of Molecular Biology. 8;323(5):883-97.

Boal AK, Rosenzweig AC. (2009). Crystal structures of cisplatin bound to a human copper chaperone. Journal of American Chemical Society.14;131:14196-7.

Bull PC, Cox DW. (1994). Wilson disease and Menkes disease: new handles on heavy-metal transport. Trends in Genetics. 10 (7): 246–52.

Cole C, Barber JD, Barton GJ. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Research. 36: W197-W201.

Dauwe T, Janssens E, Bervoets L, Blust R, Eens M. (2004). Relationships between metal concentration in great tit nestling and their environment and food. Environmental Pollution. 131:373-380.

Fahmy K, Merroun M, Pollmann K, Raff J, Savchuk O, Hennig C, Selenska- Pobell S. (2006). Secondary Structure and Pd(II) Coordination in S-Layer Proteins from Bacillus sphaericus Studied by Infrared and X-Ray Absorption Spectroscopy. Biophysics Journal. 91: 996.

Gerbino E, Mobili P, Tymczyszyn E, Fausto R, Gómez-Zavaglia A. (2011). FTIR spectroscopy structural analysis of the interaction between Lactobacillus kefir S-layers and metal ions. Journal of Molecular Structure. 987: 186–192.

Gitschier J, Moffat B, Reilly D, Wood WI, Fairbrother WJ. (1998). Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase. Nature Structural Biology. 5 (1): 47–54

Gonzalez-Soto E, Gonzalez-Rodriguez V, Lopez-Suarez C, Castro-Romero JM, Perez-Iglesias J, Fernandez-Solis JM. (2000). Migration of lead and cadmium from ceramic materials used in food preparation. Bulletin of Environmental Contamination and Toxicology. 65:598-603.

Halttunen T, Kankaanpaa P, Ouwehand A, Tahvonen R, Salminen S. (2003). Cadmium decontamination by lactic acid bacteria. Bioscience and Microflora. 22: 93–97.

Halttunen T, Collado MC, El-Nezami H, Meriluoto J, Salminen S. (2008). Combining strains of lactic acid bacteria may reduce their toxin and heavy metal removal efficiency from aqueous solution. Letter of Applied Microbiology. 46:160–165.

Hanson RM (2010). Jmol – a paradigm shift in crystallographic visualization. Journal of Applied Crystallography. 43(5): 1250-1260.

Hirokawa T, Boon-Chieng S, Mitaku S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics. 14: 378–379.

Hsu PC, Guo YL (2002). Antioxidant nutrients and lead toxicity. Toxicology. 180:33-44.

Ibrahim F, Halttunen T, Tahvonen R, Salminen S (2006). Probiotic bacteria as potential detoxification tools: assessing their heavy metal binding isotherms. Canadian Journal of Microbiology. 52:877-885.

IARC (2012). International Agency for Research on Cancer-IARC Monographs on the evaluation of carcinogenic risks to humans, vol. 100C.

JECFA (2004). Evaluation of certain food additives and contaminants. Sixty-first report of the Joint FAO/WHO Expert Committee on food Additives. WHO Tech. Rep. Ser. 922: 139-144.

Jones DT. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology. 292: 195-202.

Monachese M, Burton JP, Reid G. (2012). Bioremediation and Tolerance of Humans to Heavy Metals through Microbial Processes: a Potential Role for Probiotics? Applied and Environmental Microbiology. 78;18: 6397-6404.

Mrvcic J, Stanzer D, Bacun-Druzina V, Stehlik-Tomas V. (2009). Copper binding by lactic acid bacteria (LAB). Bioscience and Microflora. 28:1–6.

Muller M, Anke M (1994). Distribution of cadmium in the food chain (soil-plant-human) of a cadmium exposed area and the health risk of the general population. Science of Total Environment. 156:151-158.

Passerini A, Lippi M, Frasconi P (2011). MetalDetector v2.0: Predicting the Geometry of Metal Binding Sites from Protein Sequence. Nucleic Acids Research. 39; 2:W288-W292.

Rosen JF (1995). Adverse health effects of lead at low exposure levels. Trends in the management of childhood lead poisoning. Toxicology, 97:11-17.

Roy A, Kucukural A, Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols. 5: 725-738.

Satarug S, Moore MR (2004). Adverse health effects of chronic exposure to low-level cadmium in foodstuffs and cigarette smoke. Envrionmental Health Perspective. 112:1099-1103.

Struzynska L, Walski M, Gadamski R, Dabrowska-Bouta B, Rafalowska U (1997). Lead-induced abnormalities in blood-brain permeability in experimental chronic toxicity. Molecular Chemistry and Neuropathology. 31:207-224.

Sussell A, Hart C, Wild D, Ashley K (1999). An evaluation of worker lead exposures and cleaning effectiveness during removal of deteriorated lead-based paint. Applied Occupational and Environmental Hygiene. 14(3):177-85.

Tillet T (2009). The price of progress: modern environmental health hazards in Africa. Environmental Health Perspective. 117:A257.

Wernimont AK, Huffman DL, Lamb AL, O'Halloran TV, Rosenzweig AC. (2000). Structural basis for copper transfer by the metallochaperone for the Menkes/Wilson disease proteins. Nature Structural Biology. 7(9):766-71.

Yang Zhang. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 9:40

Zimmermann M, Clarke O, Gulbis JM, Keizer DW, Jarvis RS, Cobbett CS, Hinds MG, Xiao Z, Wedd AG (2009). Metal binding affinities of Arabidopsis zinc and copper transporters: selectivities match the relative, but not the absolute, affinities of their amino-terminal domains. Biochemistry. 48(49):11640-11654.

Zhai Q, Wang G, Zhao J, Liu X, Tian F, Zhang H, Chen W (2013). Protective Effects of Lactobacillus plantarum CCFM8610 against Acute
Cadmium Toxicity in Mice. Applied and Environmental Microbiology. 79(5):1508-1515.