Characterization of a novel gene involved in cadmium accumulation screened from sponge-associated bacterial metagenome

Tetsushi Mori[1*], Koji Iwamoto[2*], Satoshi Wakaoji[1], Hiroya Araie[2], Yotaro Kohara[1], Yoshiko Okamura[1 ‡], Yoshihiro Shiraiwa[2], and Haruko Takeyama[1 †]

[1] Faculty of Science and Engineering, Waseda University, Tokyo, Japan

[2] Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan

[*] These authors contributed equally to this work

[‡] Present address: Graduate School of Advanced Sciences of Matter, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8530, Japan.

[†] Corresponding author: Faculty of Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku, Tokyo 162-8480, Japan.

TEL: +81-3-5369-7326

FAX: +81-3-5369-7302,

E-mail address: haruko-takeyama@waseda.jp (H. Takeyama)

Abstract

Metagenome research has brought much attention for the identification of important and novel genes of industrial and pharmaceutical value. Here, using a metagenome library constructed from bacteria associated to the marine sponge, *Styllisa massa*, a high-throughput screening technique using radioisotope was implemented to screen for cadmium (Cd) binding or accumulation genes. From a total of 3,301 randomly selected clones, a clone 247-11C was identified to harbor an open reading frame (ORF) showing Cd accumulation characteristics. The ORF, termed as ORF5, was further analyzed by protein functional studies to reveal the presence of a protein, Cdae-1, comprised of a signal peptide and domain harboring an E(G/A)KCG pentapeptide motif, in which the later, enhanced Cd accumulation when expressed in *E. coli*. Although showing no direct binding to Cd *in vitro*, the presence of important amino acid residues related to Cd detoxification suggests that Cdae-1 may possess a different mechanism from known Cd binding proteins such as metallothioneins (MTs) and phytochelatins (PCs). In summary, using the advantage of bacterial metagenomes, our findings in this work suggest the first report on the identification of a unique protein involved in Cd accumulation from bacteria associated to a marine sponge.

Abbreviations

Cd, cadmium; MTs, metallothioneins; PCs, phytochelatins; Cdae, Cadmium accumulation element.

Keywords

Cd accumulation, sponge-associated bacterial metagenome, pentapeptide motif

Highlights

- We screened a sponge-associated bacterial metagenome library for novel cadmium accumulation genes.
- The discovered Cd accumulation protein, Cdae-1, although coupled with a signal peptide, promoted intracellular Cd accumulation.
- Cdae-1 showed different amino acid features and characteristics to metallothioneins and phytochelatins.
- Cdae-1 harbored a novel pentapeptide motif unique to a class of hypothetical or low-complexity proteins of unknown function.

1. Introduction

Cadmium (Cd) has been regarded as an important trace element due to its industrial applicability in nickel-cadmium batteries, Cd pigments, Cd coatings and as stabilizers in plastics and alloys (Morrow, 2010). However, long-term exposure to Cd or uptake at high levels has resulted in serious health and ecological problems (Jarup and Akesson, 2009; Boyd, 2010). Currently, the removal of Cd from the environment is conducted using chemical, membrane, ion exchange, solvent extraction and adsorption techniques (Rao et al., 2010). Alternatively, many organisms have adopted resistance mechanisms such as exclusion (Zhu et al., 2011), compartmentalization (Dehn et al., 2004), the formation of complexes (Inouhe et al., 1996) and the synthesis of metal binding proteins (Mejare and Bulow, 2001) to overcome Cd toxicity and heavy metal stress. Such mechanisms have brought much attention as these systems provide an alternative to conventional Cd removal techniques and can be utilized to further overcome current bioremediation challenges.

Among these resistance mechanisms, the introduction or overexpression of metal-binding proteins have been widely exploited to increase Cd binding capacity, tolerance or accumulation. Two of the most well characterized binding proteins are metallothioneins (MTs) and phytochelatins (PCs). MTs, characterized as low-molecular cytosolic gene-encoded polypeptides, bind to a range of heavy metals including Cd^{2+}, Pb^{2+}, Bi^{3+}, Ag^{2+}, Cu^{2+}, Zn^{2+} and Ni^{2+} while PCs, glutathione polymers enzymatically synthesized by phytochelatin synthases (PCS), are chelators important for heavy metal detoxification (Henkel and Krebs, 2004; Dar et al., 2013; Rigouin et al., 2013). Both

MTs and PCs have also been widely reported in various organisms including plants, yeasts, algae and fungi. In bacteria, the identification of Cd-binding proteins including MT and PC homologues has also been reported (Capasso et al., 1996; Harada et al., 2004; Tsuji et al., 2004; Blindauer, 2011). Bacteria also serve as an important expression system for the overexpression of Cd-binding proteins attained from plants (Kim et al., 2009), yeast (Preveral et al., 2009) or synthetically synthesized peptides (Bae et al., 2002), further suggesting the importance of bacteria in Cd bioremediation. However, since the majority of work related to Cd binding proteins have focused on currently identified proteins such as MTs or PCs, there is still a need for the discovery of novel Cd binding proteins or proteins enhancing Cd accumulation.

Thus, to attain such proteins, we focused our search on microbial metagenomes. In recent years, metagenomic research has supported the identification of novel and important genes from bacterial communities of both terrestrial and marine environments. Marine microbial metagenomes in particular, known for its unique and large genetic diversity, has served as a resource for genes such as lipases (Selvin et al., 2012), esterases (Okamura et al., 2010), fumarases (Jiang et al., 2010), beta-glucosidases (Fang et al., 2010), applicable to pharmaceutics, research and industry (Kennedy et al., 2010; Hentschel et al., 2012). Furthermore, metagenomic based research also provides an advantage to identify and discover genes that may harbor non-elucidated characteristics or undetermined phenotypic properties since screening and activity assays are frequently conducted in bacterial hosts such as *E. coli* and *Bacillus*. However, although metagenomic researches are currently widely conducted, proteins related to Cd binding or accumulation has not been reported thus far.

Therefore, in this research, to conduct a comprehensive search for genes related to Cd binding or accumulation, we conducted the screening of such genes from the metagenome library of bacteria associated to the marine sponge *Styllisa massa*. We focused on bacteria associated with marine sponges, since marine sponges are known to be one of the largest producer in secondary metabolites (Thomas et al., 2010) and holds high potential to harbor unique functional genes including those related to heavy metal accumulation (Selvin et al., 2009; Nelson and Slinger-Cohen, 2014). Subsequently, functional analysis and preliminary sequence comparison studies to determine the novelty of the discovered protein were conducted. Here, we report one of the first reports on the identification of a unique protein involved in the accumulation of Cd from a sponge-associated bacterial metagenome.

2. Materials and Methods

2.1 *Sponge, bacterial strains and plasmids*

The marine sponge, *Stylissa massa*, was collected from the offshore of Ishigaki island, Okinawa, Japan. The *E. coli* strains, EPI300^{TM} (Epicentre Biotechnologies) were used in metagenome library construction, DH5α (TOYOBO) and EC100^{TM} (Epicentre Biotechnologies) in cloning and BL21 (DE3) (Novagen) in recombinant protein expression, respectively. The plasmid pCC1FOS (Epicentre Biotechnologies) was used for metagenome library construction and in standard cloning procedures. The Zero Blunt TOPO PCR Cloning Kit for Sequencing (Life Technologies) was used for the cloning of PCR amplicons, and the pET25b (Novagen) vector was used for protein expression.

2.2 *Library construction*

The preparation of marine sponge bacterial fraction and DNA extraction were performed as described by Okamura et al. (Okamura et al., 2010). Fosmid library construction was conducted using the CopyControl Fosmid Library Production Kit (Epicentre Biotechnologies) based on the manufacturer's protocol. Briefly, blunt-ended and 5'-phosphorylated DNA was separated by pulsed-field gel electrophoresis (1% LMP agarose/1 x TBE gel, 0.5 x TBE buffer, 0.5 s pulse, 9 V/cm, 14 °C, 120 °, 3 h) and approximately 40 kbp of DNA was recovered with GELase. The attained DNA fragments were ligated with the pCC1FOS vector, packaged, titered and were infected into *E. coli* EPI300^{TM} cells. Upon plating on LB agar plates containing 12.5 μg/mL chloramphenicol, colonies were selected manually or using the BioPick automated colony picking system (Genomic Solutions). The metagenome library was stored at -80˚C in a 96-well plate format. Fosmid DNA were extracted from randomly selected clones following standard alkaline lysis procedure and digested by *Bam*HI and *Eco*RI/*Hin*dIII to estimate the average size of the DNA inserts.

2.3 ***Screening for Cd accumulation clones from metagenome library***

Screening of Cd accumulation from the metagenome library clones was conducted using the microplate-BAS method. Prior to the screening of Cd accumulation genes, 40 randomly selected 96-well plates; comprised of 3,301 metagenome library clones were cultured in LB medium containing 12.5 μg/mL of chloramphenicol overnight at 37˚C with agitation. The overnight cultures were diluted with 4 folds of culture medium described above with addition of 2.5 μM Cd including 37 kBq/mL of radioactive Cd (^{109}Cd, PerkinElmer Life and Analytical Science) and induction solution (Epicentre Biotechnologies) and were cultured at 37˚C for a subsequent 5 hours. After 5 hours, 50 μL of the bacterial culture was transferred to Multiscreen-GV filter plates (0.22 μm) and washed 3 times with 200 μL of 3% NaCl using a MultiScreen HTS Vacuum Manifold system (EMD Millipore). Upon drying, the plates were placed on imaging plates, overnight in the dark and the accumulation of Cd within the clones was detected using the Bio-imaging Analyzer BAS-1800 II (FUJIFILM). Positive clones showing high Cd accumulation were selected upon 2 rounds of screening. The selection criteria for Cd accumulation clones were determined based on the accumulation of Cd above the total average detection signal value. As such, the cut off point for positive clones was set to 2 folds and 3 folds for the 1st and 2nd round screening, respectively.

The clones attained from the 2nd screen were reanalyzed to determine Cd accumulation using the silicone oil centrifugation method. 0.4 mL sampling tubes were prepared containing a dense bottom layer, comprised of 50 μL silicone oil (Toray Dow Corning Silicone; SH550 : SH556 = 2 : 1). 200 μL of cells incubated with 2.5 μM Cd including ^{109}Cd radioisotope were pipetted over the silicone oil layer, centrifuged at 10,000 g for 1 min and the sample tubes were frozen in liquid nitrogen. The bottom layer with the cells was clipped into measurement tubes and the radioactivity was measured using the COBRAII γ-counter (Packard Instrument). For the plasmid reintroduction assay, 200 mL of positive cultures were grown in 2-YT medium containing 12.5 μg/mL chloramphenicol and induction solution, autoinduced and cultured overnight. Plasmids were extracted using the Qiagen Plasmid Buffer Set and the Qiagen-tip 100 (Qiagen) based on the manufacturer's protocol. The extracted plasmid was subsequently transformed by electroporation using the Gene Pulser II (Biorad) into *E. coli* EC100™ electrocompetent cells. The clone showing the highest Cd accumulation was cloned and sequenced to determine the regions harboring the Cd accumulation gene.

2.4 *Identification of the Cd accumulation gene*

Based on the sequencing results, plasmid from the Cd accumulating clone was enzyme digested with the restriction enzyme *Xho*I to determine the region in which Cd accumulation occurs. The enzyme digested fragments were cloned into the pCC1FOS vector and transformed into *E. coli* EC100^{TM} cells. Cd accumulation was measured from the attained clones using the silicone oil centrifugation method as described above. *E. coli* cells harboring only the pCC1FOS vector was used as a negative control. Subsequently, the attained fragment harboring the Cd accumulation region, was further enzyme digested with the restriction enzymes *Sph*I or *SanD*I, self-ligated and transformed into *E. coli* EPI300^{TM} cells. The clones attained from *Sph*I and *SanD*I digestion, were analyzed to locate the position of the Cd accumulation gene using the silicone oil centrifugation and microplate-BAS methods.

Specific ORFs with the predicted promoter regions were amplified by PCR using the PrimeSTAR Max DNA Polymerase（Takara Bio）based on the manufacturer's instructions. PCR conditions were as follows: 33 cycles of 98 °C for 10 s, 55 °C for 10 s, and 72 °C for 11 s, and the attained amplicons were cloned accordingly. Pre-culture of each transformant was inoculated at OD_{660}=0.03 in 5 mL of LB medium with proper antibiotics and Cd standard solution (1,000 ppm; Wako Chemicals) was added at 100 μM final concentration. After culturing for 15 hours at 37˚C with agitation, cells were harvested by centrifugation at 8,000 g, for 5 min at 4˚C and washed twice with 1 x PBS buffer supplemented with 25 mM EDTA in order to remove adsorbed Cd from the cell surface. Washed pellets were heated at 180°C with 9.6 N HNO_3 to degrade organic matter. Dried residues including Cd were dissolved in 5 mL of 0.5 N HNO_3 and Cd concentration was determined using an AA-6600G atomic adsorption spectrometer (Shimadzu). Cd standard solution (1,000 ppm) was used for the calibration curve.

2.6 *Analysis of the target clone*

The fragments, truncated at the designated regions were amplified by PCR and were cloned into the pCR4Blunt TOPO cloning vector provided with the Zero Blunt TOPO cloning kit for sequencing. The fragments for protein expression were similarly amplified by PCR and were cloned in the pET25b expression vector. The primers used in the amplification of the respective fragments are shown in Table 1. Protein expression was conducted using the Overnight Express^{TM} Autoinduction Systems 1 (EMD Millipore) according to the manufacturer's protocol. Preparation of the designated samples for intracellular Cd concentration measurement is as described in the section above and samples were measured using an ICPE-9000 ICP atomic emission spectrometer (Shimadzu).

For protein expression analysis,, the designated clones induced using the Overnight Express^{TM} Autoinduction Systems 1 were harvested by centrifugation at 13,000 rpm, for 2 min at room temperature. The pellets were treated using the BugBuster Master Mix (Novagen) for protein extraction and were regarded as the intracellular fraction. The supernatants on the other hand, were concentrated using the Amicon® Ultra-4 (EMD Millipore) filters and were regarded as the extracellular fraction. Both the intracellular and extracellular fractions were purified by affinity chromatography and the purified proteins were analyzed by standard SDS-PAGE.

2.7 *Gene accession number*

The nucleotide sequence of the Cd accumulation gene has been submitted to the DNA Data Bank of Japan (DDBJ) and has been assigned the accession number AB969736.

3. Results and Discussion

3.1 *Metagenome library construction and screening of cadmium accumulating clones*

In the past, we have reported on the identification of a novel halotolerant esterase, EstHE1, from the metagenome library of bacteria associated to the marine sponge, *Hyrtios erecta* (Okamura et al., 2010). Using this understanding, we subsequently focused on the marine sponge, *Styllisa massa*, which was collected off the coast of Ishigaki island, Okinawa, Japan. As a result, we were successful in establishing a bacterial metagenome library comprised of 65,043 clones, with DNA insert size averaging at 37 kbp totaling up to a sequence size of approximately 2.4 Gbp.

For the screening of the clones harboring Cd binding or accumulation genes, randomly selected 40 96-well plates comprised of 3,301 clones were selected and subjected to 2 rounds of selection using ^{109}Cd radioisotope. From the first screen, 52 clones showing approximately 2 folds accumulation of Cd above the average was attained (Fig. 1a). In the second screen, 6 out of the 52 clones significantly showed high Cd uptake (Fig. 1b). These 6 clones designated as 66-11E, 217-9E, 219-3D, 238-4B, 247-11C and 361-5A was reanalyzed and Cd accumulation was determined by using the silicone oil centrifugation method. The Cd accumulation of these clones ranged in between 4-9 folds in comparison to the control, with clones 238-4B and 247-11C showing relatively high Cd accumulation at approximately 8 - 9 folds (Fig. 1c). To further validate our results and to determine the clone showing the highest Cd accumulation, reintroduction of the plasmids into *E. coli* was conducted, in which the clone 247-11C showed the most stable Cd accumulation (Fig. 1d). Further analysis was conducted against this clone to identify the Cd accumulation gene.

3.2 *Identification of the gene involved in cadmium accumulation.*

To identify the gene involved in Cd accumulation, the plasmid from clone 247-11C was extracted and sequenced, revealing a metagenome fragment comprised of 33.7 kbp. Restriction enzyme digestion using *Xho*I resulted in 2 fragments in which the smaller fragment sized at 9.7 kbp showed high Cd accumulation similar to that of the non-truncated 247-11C clone (Fig. 2a). The newly identified clone, 247-11CX1 was further subjected to restriction enzyme digestion with *Sph*I or *SanD*I, in which 3 clones designated 247-11CX1S0, 247-11CX1Sa and 247-11CX1Sp was attained and analyzed for Cd accumulation. The 247-11CX1S0 and 247-11CX1Sa clones that had deleted regions as shown in Fig. 2b did not show any Cd accumulation, suggesting that the Cd accumulation gene may exist within the 3 kbp and 7 kbp region of the 247-11CX1 fragment. As predicted, when Cd accumulation was conducted with the 247-11CX1Sp clone, high Cd accumulation was observed (Fig. 2c). Prior amino acid sequence similarity search of the open reading frames (ORFs) within the 9.7 kbp fragment using ORF Finder showed the presence of 3 ORFs within this region. ORF4, encodes for an acetyl-CoA synthetase (amino acid similarity of 70%) and ORF5 and ORF6, each represents hypothetical proteins, respectively. To further identify the gene responsible for Cd accumulation, Cd accumulation assay was conducted, resulting in ORF5 showing the highest Cd accumulation to as high as 11 folds in comparison to ORF4 and ORF6 (Fig. 2d). Here we clearly indicate that ORF5 was involved in Cd accumulation.

3.3 *Domain conformation and functional analysis of Cdae-1*

ORF5, comprised of a 744 bp long fragment, was analyzed at the nucleotide and amino acid level to further identify possible functional domains. Based on prediction and similarity searches, we hypothesized the presence of 3 domains within ORF5, each comprised of a putative promoter sequence region (R1), a signal peptide region (R2) and an unknown protein region harboring a repetitive unique pentapeptide motif, E(G/A)KCG (R3) (Fig. 3). Hereon, subsequent analysis was narrowed down to R2 and R3 of ORF5, in which this region was designated as Cdae-1 (Cdae: Cadmium accumulation element). The functional prediction of R2 as a signal peptide was determined by using SignalP in which the cleavage point was identified at AHA-EG.

The function of R2 as a signal peptide and R3 as the protein region involved in Cd accumulation was further confirmed by protein expression experiments. Based on truncation experiments of ORF5 at R2 (ORF5 (T2)) and R3 (ORF5 (T1)), it was clearly shown that R3, the domain harboring the E(G/A)KCG repetitive motif played an important role in Cd accumulation. This was further clarified by the cloning of R2 and R3 (Cdae-1) or only R3 (Cdae-1R3) into the pET-25b expression vector and by performing a Cd accumulation assay (Fig. 4a). To confirm the function of R2 as a signal peptide, SDS-PAGE of the extracellular and intracellular fractions of the designated clones, Cdae-1 and Cdae-1R3, were conducted and only the clone expressing Cdae-1 showed the presence of a band of approximately 18 kDa at the extracellular fraction. This result clearly indicates that R2 does function as a signal peptide (Fig. 4b).

3.4 *Comparative study of Cdae-1 with other metal binding proteins and similar homologs.*

Cysteine (C) residues are known to be reactive and conjugates to nitric oxide or metal ions via its sulfhydryl group (Hynek et al., 2012) while histidine (H) residues are known to coordinate metal ion binding with its imidazole substituent (Blindauer, 2008). MTs for example contains C residues (about 10-30%) that are arranged in the form of clusters known as ‘‘metal binding motifs’’ in which they are usually present in different combinations of C–X–C, C–X–X–C and C–C form (Cobbett and Goldsbrough, 2002), while PCs, comprises of (γGlu–Cys)$_n$–Gly, n = 2–11, small C-rich peptides (Oven et al., 2002). To elucidate the role of the R3 region of Cdae-1 in Cd accumulation, we first identified the presence of C or H residues within the amino acid sequence. Our observations showed that in comparison to MTs and PCs, only 10.6% of the total amino acid residues of Cdae-1R3 comprised of C residues, where all were found within the E(G/A)KCG repetitive motif, while only 1.8% were H residues. This observation showed a significant difference between the composition of C and H residues to that of currently known proteins or peptides associated to Cd binding. Thus, in reference to the results attained from the *in vivo* accumulation of Cd (Fig. 4a) and to determine the direct binding of Cdae-1R3 to Cd, *in vitro* binding assays using ^{109}Cd radioisotopes were conducted. However, contradicting to our expectations, no direct binding of Cd to Cdae-1R3 was observed (data not shown).

The unexpected results from the *in vitro* experiments suggested that Cdae-1R3, although coupled with a signal peptide, promoted Cd accumulation intracellularly and not extracellularly. This observation suggests that Cdae-1R3 may instead possess a unique characteristic within the cell to enhance Cd accumulation and we speculate that the signal peptide may be regulated by intracellular mechanisms that serve as a control for the release of Cdae-1R3 when required. Looking at the total composition of amino acid residues within Cdae-1, it was found that Cdae-1R3 itself has a high composition of glycine (G; 30.1%) and lysine (K; 19.5%) residues. Lysine in particular has been reported to play an important role in the detoxification properties of MTs (Cody and Huang, 1993) while glycine acts as a crucial constituent in PTs (Cobbett, 2000). Nevertheless, these speculations are still preliminary, as further protein functional analyses need to be conducted. Attempts to further speculate the function of Cdae-1R3 by generation of a tertiary model using SWISS-MODEL (http://swissmodel.expasy.org) resulted in a 12.82% sequence identity to a soluble cytochrome b562 from *Salmonella enterica* which did not correspond to the E(G/A)KCG repetitive motif of Cdae-1R3. This low structural similarity further suggests that Cdae-1R3 is highly unique and crystallization studies may provide new insights to the mechanism and function of this protein.

Subsequently, we conducted a conserved domain search and a blastp search against the NCBI and Uniprot databases to determine the presence of similar proteins to Cdae-1. The conserved domain search conducted using the NCBI Conserved Domain Search tool resulted in Cdae-1 harboring a domain classified to an uncharacterized low-complexity protein superfamily, COG3767, while the blastp search resulted in the identification of a group of hypothetical or low-complexity proteins of unknown or uncharacterized function that similarly harbors a signal peptide and the E(G/A)KCG repetitive motif (Table 2). Amino acid similarity alignment of Cdae-1 to these proteins is also shown in Fig. 5. Interestingly, all of these proteins we identified were only found to be highly conserved within the class γ-proteobacteria. The presence of other similar proteins and together with our functional characterization of Cdae-1 could provide evidence of γ-proteobacteria strains with unique Cd accumulation properties.

In summary, although further analytical studies including crystal structural conformation and detailed analysis on the mechanism of Cdae-1 to Cd accumulation within *E. coli* needs to be conducted, we showed that by using bacterial metagenome as a genetic resource, we were successful in the discovery of a novel candidate Cd accumulation gene showing different features to that of MTs and PCs. We hope that this research will trigger for the search and identification of novel and unique genes that may not only further promote Cd bioremediation, but also for other heavy metal pollutants.

4. Acknowledgements

This work was funded in part by the New Energy and Industrial Technology Development Organization (NEDO) as part of a research and development project of the Industrial Science and Technology Frontier Program and Scientific Research (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

The authors have no conflict of interest directly relevant to the contents of this research.

Figure Legends

Fig. 1 Screening of cadmium (Cd) binding or accumulation genes from the marine sponge, *Stylissa massa*, associated bacterial metagenome library. a. First screen of Cd accumulation clones using the microplate-BAS method. Clones showing Cd accumulation of approximately 2 folds were selected for the second screen. b. Second screen of Cd accumulation. Clones showing Cd accumulation of >3 folds were selected. c. Third screen and determination of Cd accumulation clones using the silicone oil centrifugation method. d. Reintroduction and reconfirmation of Cd accumulation ability from clones attained from the 3^{rd} screen.

Fig. 2 Identification of the cadmium (Cd) accumulation gene. a. Fragmentation of 247-11C using the *Xho*I restriction enzyme and Cd uptake of the fragment harboring the Cd accumulation gene. b. Fragmentation of 247-11CX1 using the *Sph*I or *SanD*I restriction enzymes and Cd accumulation analysis. The dotted arrows indicate the deleted regions within each of the 247-11CX1S0 and 247-11CX1Sa clones c. Cd accumulation using the 247-11CX1Sp clone. d. Analysis of the gene involved in the accumulation of Cd. Open reading frames (ORFs) were predicted using ORF Finder. For all samples, fold increase of Cd accumulation was determined by referring to the fold increase of pCC1FOS plasmid (negative control) as 1.

Fig. 3 The nucleotide and amino acid sequence of ORF5

Fig. 4 Functional analysis of ORF5 and Cdae-1. a. Illustration of the regions truncated or analyzed within ORF5 (Top); cadmium accumulation analysis to determine the regions involved in accumulation (Bottom). b. Functional analysis of the signal peptide and R3 region (Cdae-1R3) by SDS-PAGE.

Fig. 5 Comparative analysis of the amino acid sequences between Cdae-1 and similar proteins extracted from online databases.

Table 1 Primer sequences used in the amplification of the target fragments for the functional analysis of ORF5

Table 2 Comparative study of Cdae-1 with related proteins

References

Bae, W., Mulchandani, A. and Chen, W., 2002. Cell surface display of synthetic phytochelatins using ice nucleation protein for enhanced heavy metal bioaccumulation. J Inorg Biochem 88, 223-7.

Blindauer, C.A., 2008. Metallothioneins with unusual residues: histidines as modulators of zinc affinity and reactivity. J Inorg Biochem 102, 507-21.

Blindauer, C.A., 2011. Bacterial metallothioneins: past, present, and questions for the future. Journal of Biological Inorganic Chemistry 16, 1011-24.

Boyd, R.S., 2010. Heavy metal pollutants and chemical ecology: exploring new frontiers. J Chem Ecol 36, 46-58.

Capasso, C., Nazzaro, F., Marulli, F., Capasso, A., La Cara, F. and Parisi, E., 1996. Identification of a high-molecular-weight cadmium-binding protein in copper-resistant Bacillus acidocaldarius cells. Res Microbiol 147, 287-96.

Cobbett, C. and Goldsbrough, P., 2002. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. Annu Rev Plant Biol 53, 159-82.

Cobbett, C.S., 2000. Phytochelatins and their roles in heavy metal detoxification. Plant Physiol 123, 825-32.

Cody, C.W. and Huang, P.C., 1993. Metallothionein detoxification function is impaired by replacement of both conserved lysines with glutamines in the hinge between the two domains. Biochemistry 32, 5127-31.

Dar, S., Shuja, R.N. and Shakoori, A.R., 2013. A synthetic cadmium metallothionein gene (PMCd1syn) of Paramecium species: expression, purification and characteristics of metallothionein protein. Mol Biol Rep 40, 983-97.

Dehn, P.F., White, C.M., Conners, D.E., Shipkey, G. and Cumbo, T.A., 2004. Characterization of the human hepatocellular carcinoma (hepg2) cell line as an in vitro model for cadmium toxicity studies. In Vitro Cell Dev Biol Anim 40, 172-82.

Fang, Z., Fang, W., Liu, J., Hong, Y., Peng, H., Zhang, X., Sun, B. and Xiao, Y., 2010. Cloning and characterization of a beta-glucosidase from marine microbial metagenome with excellent glucose tolerance. J Microbiol Biotechnol 20, 1351-8.

Harada, E., von Roepenack-Lahaye, E. and Clemens, S., 2004. A cyanobacterial protein with similarity to phytochelatin synthases catalyzes the conversion of glutathione to gamma-glutamylcysteine and lacks phytochelatin synthase activity. Phytochemistry 65, 3179-85.

Henkel, G. and Krebs, B., 2004. Metallothioneins: zinc, cadmium, mercury, and copper thiolates and selenolates mimicking protein active site features--structural aspects and biological implications. Chem Rev 104, 801-24.

Hentschel, U., Piel, J., Degnan, S.M. and Taylor, M.W., 2012. Genomic insights into the marine sponge microbiome. Nat Rev Microbiol 10, 641-54.

Hynek, D., Krejcova, L., Sochor, J., Cernei, N., Kynicky, J., Adam, V., Trnkova, L., Hubalek, J., Vrba, R. and Kizek, R., 2012. Study of Interactions between Cysteine and Cadmium(II) Ions using Automatic Pipetting System off-line Coupled with Electrochemical Analyser. International Journal of Electrochemical Science 7, 1802-1819.

Inouhe, M., Sumiyoshi, M., Tohoyama, H. and Joho, M., 1996. Resistance to cadmium ions and formation of a cadmium-binding complex in various wild-type yeasts. Plant Cell Physiol 37, 341-6.

Jarup, L. and Akesson, A., 2009. Current status of cadmium as an environmental health problem. Toxicol Appl Pharmacol 238, 201-208.

Jiang, C., Wu, L.L., Zhao, G.C., Shen, P.H., Jin, K., Hao, Z.Y., Li, S.X., Ma, G.F., Luo, F.F., Hu, G.Q., Kang, W.L., Qin, X.M., Bi, Y.L., Tang, X.L. and Wu, B., 2010. Identification and characterization of a novel fumarase gene by metagenome expression cloning from marine microorganisms. Microb Cell Fact 9, 91.

Kennedy, J., Flemer, B., Jackson, S.A., Lejon, D.P., Morrissey, J.P., O'Gara, F. and Dobson, A.D., 2010. Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. Mar Drugs 8, 608-28.

Kim, I.S., Shin, S.Y., Kim, Y.S., Kim, H.Y. and Yoon, H.S., 2009. Expression of a glutathione reductase from Brassica rapa subsp. pekinensis enhanced cellular redox homeostasis by modulating antioxidant proteins in Escherichia coli. Mol Cells 28, 479-87.

Mejare, M. and Bulow, L., 2001. Metal-binding proteins and peptides in bioremediation and phytoremediation of heavy metals. Trends Biotechnol 19, 67-73.

Morrow, H., 2010. Cadmium and Cadmium Alloys, Kirk-Othmer Encyclopedia of

Chemical Technology. pp. 1-36.

Nelson, W. and Slinger-Cohen, N., 2014. Trace metals in the sponge Ircinia felix and sediments from North-Western Trinidad, West Indies. J Environ Sci Health A Tox Hazard Subst Environ Eng 49, 967-72.

Okamura, Y., Kimura, T., Yokouchi, H., Meneses-Osorio, M., Katoh, M., Matsunaga, T. and Takeyama, H., 2010. Isolation and characterization of a GDSL esterase from the metagenome of a marine sponge-associated bacteria. Mar Biotechnol (NY) 12, 395-402.

Oven, M., Page, J.E., Zenk, M.H. and Kutchan, T.M., 2002. Molecular characterization of the homo-phytochelatin synthase of soybean Glycine max: relation to phytochelatin synthase. J Biol Chem 277, 4747-54.

Preveral, S., Gayet, L., Moldes, C., Hoffmann, J., Mounicou, S., Gruet, A., Reynaud, F., Lobinski, R., Verbavatz, J.M., Vavasseur, A. and Forestier, C., 2009. A common highly conserved cadmium detoxification mechanism from bacteria to humans: heavy metal tolerance conferred by the ATP-binding cassette (ABC) transporter SpHMT1 requires glutathione but not metal-chelating phytochelatin peptides. J Biol Chem 284, 4936-43.

Rao, K.S., Mohapatra, M., Anand, S. and Venkateswarlu, P., 2010. Review on cadmium removal from aqueous solutions. International Journal of Engineering, Science and Technology 2, 81-103.

Rigouin, C., Nylin, E., Cogswell, A.A., Schaumloffel, D., Dobritzsch, D. and Williams, D.L., 2013. Towards an understanding of the function of the phytochelatin synthase of Schistosoma mansoni. PLoS Negl Trop Dis 7, e2037.

Selvin, J., Kennedy, J., Lejon, D.P., Kiran, G.S. and Dobson, A.D., 2012. Isolation identification and biochemical characterization of a novel halo-tolerant lipase from the metagenome of the marine sponge Haliclona simulans. Microb Cell Fact 11, 72.

Selvin, J., Priya, S.S., Kiran, G.S., Thangavelu, T. and Bai, N.S., 2009. Sponge-associated marine bacteria as indicators of heavy metal pollution. Microbiol Res 164, 352-363.

Thomas, T.R., Kavlekar, D.P. and LokaBharathi, P.A., 2010. Marine drugs from sponge-microbe association--a review. Mar Drugs 8, 1417-68.

Tsuji, N., Nishikori, S., Iwabe, O., Shiraki, K., Miyasaka, H., Takagi, M., Hirata, K. and

Miyamoto, K., 2004. Characterization of phytochelatin synthase-like protein encoded by alr0975 from a prokaryote, Nostoc sp. PCC 7120. Biochem Biophys Res Commun 315, 751-5.

Zhu, X.F., Zheng, C., Hu, Y.T., Jiang, T., Liu, Y., Dong, N.Y., Yang, J.L. and Zheng, S.J., 2011. Cadmium-induced oxalate secretion from root apex is associated with cadmium exclusion and resistance in Lycopersicon esulentum. Plant Cell Environ 34, 1055-64.

Target region		**Primer sequences**
ORF5	Forward	ATGCAGGTTCGTGCGCGGACACG
	Reverse	TTAGTGCCCGCACTTGCCCTC
ORF5 (T1)	Forward	ATGCAGGTTCGTGCGCGGACA
	Reverse	CGCCCTTATGGGCATCGGCGA
ORF5 (T2)	Forward	ATGCAGGTTCGTGCGCGGACA
	Reverse	GGCGGTCACGGAAACGGCGAA
Cdae-1	Forward	AAAAA**CATATG**AGCGAAGAGAAGAAATCC
	Reverse	AAAAA**CTCGAG**GCTGCCGCGCGGCACCAGGTGCCCGCACTTGCCCTCGC
Cdae-1R3	Forward	AGC**CATATG**GTCGAGTATGGAGGCGGC
	Reverse	**GGATCC**TTAGTGCCCGCACTTGCCCTC

* Bold sequences show the restriction enzyme sites

** Underline sequence show the thrombin cleavage site

Table 1

Protein	NCBI Accession Number	Origin/Microorganism	Assigned abbreviations	Sequence length (aa)	Signal peptide cleavage sequences	No. of pentapeptide repeats (EX*KCG motif)
Cdae-1		Sponge-associated bacterial metagenome		145	AHA-EG	12
Related proteins						
Hypothetical protein	WP_016956805	*Catenovulum agarivorans*	CatA-HP	165	ANA-NP	16
Hypothetical protein	WP_007641548	*Cellvibrio* sp. BR	CelV-HP	120	ASA-NT	8
Hypothetical protein	WP_019025790	*Colwellia piezophila*	ColP-HP	139	AKA-ET	2
Hypothetical protein	WP_008284847	gamma proteobacterium HTCC5015	GamP-HP	129	AQA-DQ	8
Hypothetical protein	WP_017444670	*Gayadomonas joobiniege*	GayJ-HP	138	NA	12
Hypothetical protein	WP_008292033	*Methylophaga thiooxydans*	MetT-HP	114	ASA-EA	8
Hypothetical protein	WP_016900409	*Pseudoalteromonas* sp. PAMC 22718	PseuA-HP	139	ASA-DV	12
Hypothetical protein Q91 1866	YP_006838403	*Cycloclasticus* sp. P1	CycC-HP	117	VNA-DT	8
Hypothetical protein Sama 1304	YP_927181	*Shewanella amazonensis* SB2B	ShewA-HP	147	AFA-AE	10
Hypothetical protein Shal 1584	YP_001673809	*Shewanella halifaxensis* HAW-EB4	ShewH-HP	149	ALA-TS	12
Hypothetical protein Shew 2241	YP_001094366	*Shewanella loihica* PV-4	ShewL-HP	146	AFA-AQ	10
Hypothetical protein Spea 1502	YP_001501362	*Shewanella pealeana* ATCC 700345	ShewPE-HP	160	ALA-TS	16
Hypothetical protein Ssed 2858	YP_001474593	*Shewanella sediminis* HAW-EB3	ShewS-HP	181	AFA-AE	8
Low-complexity protein	YP_002311067	*Shewanella piezotolerans* WP3	ShewPI-LCP	145	VQA-SP	2
Putative periplasmic low complexity protein	NP_717613	*Shewanella oneidensis* MR-1	ShewO-PLCP	143	VNA-QT	8

*** X: Alanine (A) or Glycine (G)**

Table 2

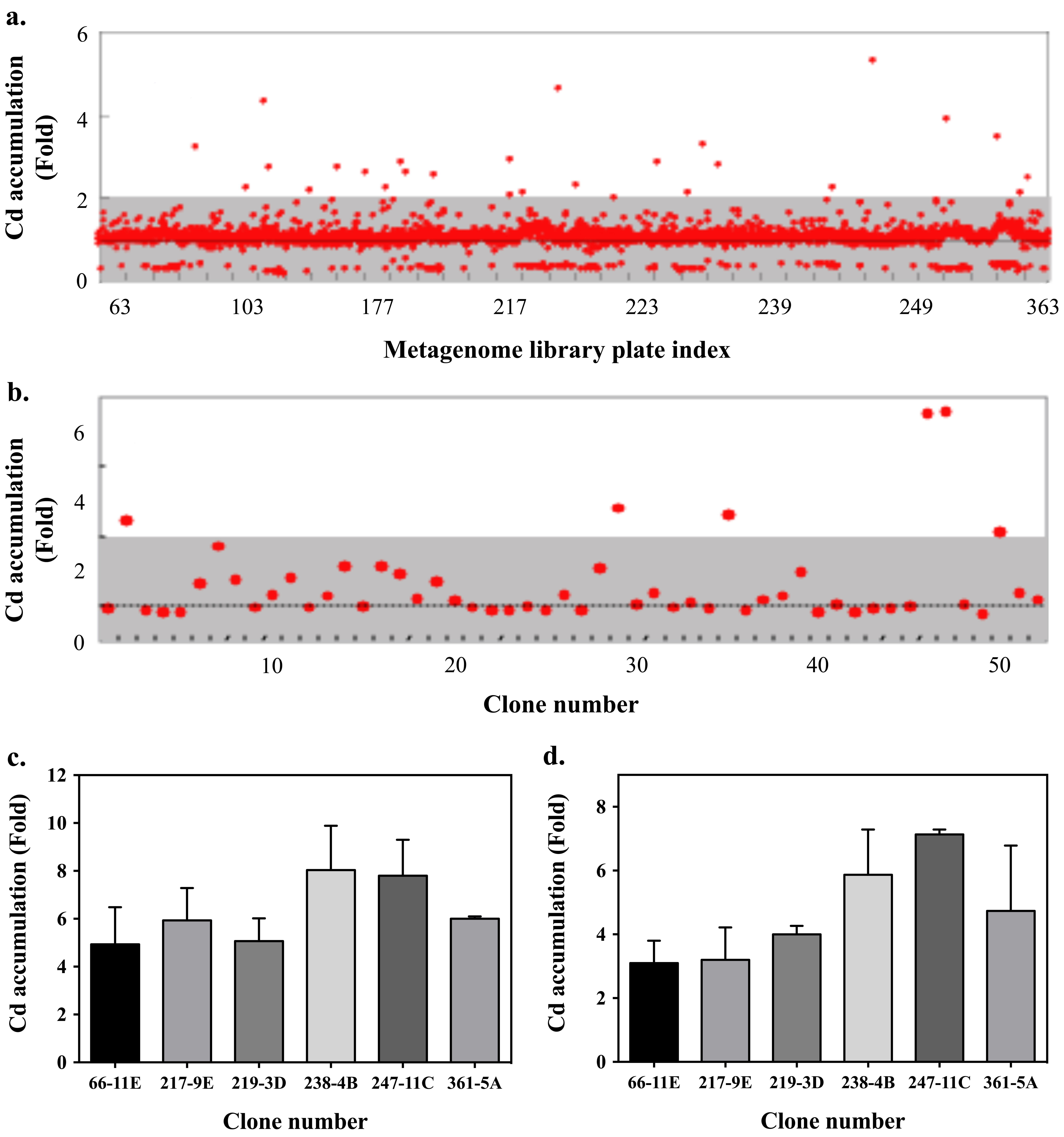

a.
Cd accumulation (Fold)
6
4
2
0
63
103
177
217
223
239
249
363
Metagenome library plate index
b.
Cd accumulation (Fold)
6
4
2
0
10
20
30
40
50
Clone number
c.
Cd accumulation (Fold)
12
10
8
6
4
2
0
66-11E
217-9E
219-3D
238-4B
247-11C
361-5A
Clone number
d.
Cd accumulation (Fold)
8
6
4
2
0
66-11E
217-9E
219-3D
238-4B
247-11C
361-5A
Clone number

Fig. 1

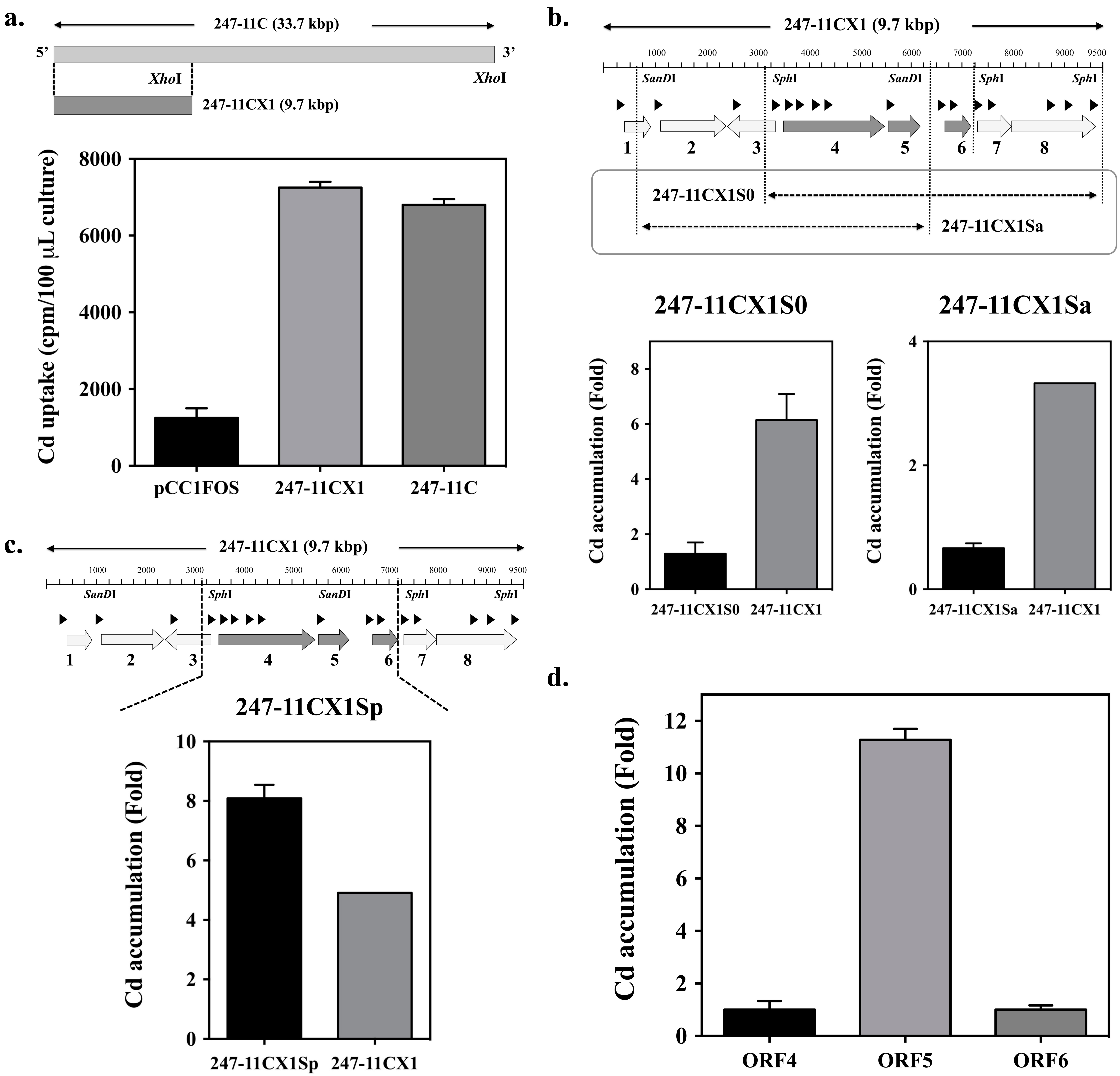
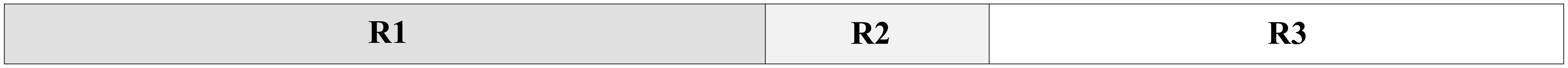

a.
247-11C (33.7 kbp)
5'
3'
XhoI
XhoI
247-11CX1 (9.7 kbp)
Cd uptake (cpm/100 μL culture)
8000
6000
4000
2000
0
pCC1FOS
247-11CX1
247-11C
b.
247-11CX1 (9.7 kbp)
1000
2000
3000
4000
5000
6000
7000
8000
9000
9500
SanDI
SphI
SanDI
SphI
SphI
1
2
3
4
5
6
7
8
247-11CX1S0
247-11CX1Sa
247-11CX1S0
247-11CX1Sa
Cd accumulation (Fold)
8
6
4
2
0
247-11CX1S0
247-11CX1
Cd accumulation (Fold)
4
2
0
247-11CX1Sa
247-11CX1
c.
247-11CX1 (9.7 kbp)
1000
2000
3000
4000
5000
6000
7000
8000
9000
9500
SanDI
SphI
SanDI
SphI
SphI
1
2
3
4
5
6
7
8
247-11CX1Sp
Cd accumulation (Fold)
10
8
6
4
2
0
247-11CX1Sp
247-11CX1
d.
Cd accumulation (Fold)
12
10
8
6
4
2
0
ORF4
ORF5
ORF6

Fig. 2

ORF5



```
ATGCAGGTTCGTGCGCGGACACGGGGACACGGGCGAAATCTTTGGTCGCATGGCGTCGCAAGC

GTATTTCCGAAACCTCGCAATAGGACATTTCTGCTTTGACACCCCCTCCGCATGGGGTGGCAT

TGGCCCGGGGATTGTGGCAGACTCTCCCCCCAGTTGCCCTGTTGCCGAGGAAGAGGCGCAGGG

ACGCAGAACCCGAAAGCCGAAGTGGGCCCAAAGAGCCCCGGCAATAGGGGTGGCGATCGTGAA

GGCGGGCTTTTTGGCCAGTCTGTTACGTTTAGGAAATCGAGAGGAGAACCGAAGATGAGCGAA
                                                       M  S  E
GAGAAGAAATCCAGTTTGAAGCCTATCGCGGTGGCCCTTGGGGCTACCTTCGCCGTTTCCGTG
 E  K  K  S  S  L  K  P  I  A  V  A  L  G  A  T  F  A  V  S  V
ACCGCCGCGGGCCCGGCCCATGCGGAGGGCAATCCCTTCGGCATGGTCGAGTATGGAGGCGGC
 T  A  A  G  P  A  H  A  E  G  N  P  F  G  M  V  E  Y  G  G  G
TACATGGTCGCCGATGCCCATAAGGGCGAGGGCAAATGCGGCGAAGGCAAATGCGGCGAGGGC
 Y  M  V  A  D  A  H  K  G  E  G  K  C  G  E  G  K  C  G  E  G
AGGAAAGCCGAAGGCAAGTGCGGCGAAGGCAAGTGCGGCGAGGGCAAGAAAGCCGAAGGCAAG
 R  K  A  E  G  K  C  G  E  G  K  C  G  E  G  K  K  A  E  G  K
TGCGGCGAAGGCAAGTGCGGCGAGGGCAAGAAAGCCGAGGGCAAGTGCGGCGAAGGCAAATGC
 C  G  E  G  K  C  G  E  G  K  K  A  E  G  K  C  G  E  G  K  C
GGCGAGGACAAGAGAGCCGAAGGCAAGTGCGGCGAAGGCAAGTGCGGCGAGGGCAAGGAAGCC
 G  E  D  K  R  A  E  G  K  C  G  E  G  K  C  G  E  G  K  E  A
AAGTCCAAGGACGCCGAGGGCAAATGCGGCGAGGGCAAGTGCGGGCACTAA
 K  S  K  D  A  E  G  K  C  G  E  G  K  C  G  H
```

▒ **Putative promoter region (R1)**

░ **Putative signal peptide region (R2)**

□ **Unknown region harboring pentapeptide repeats (R3)**

Fig. 3

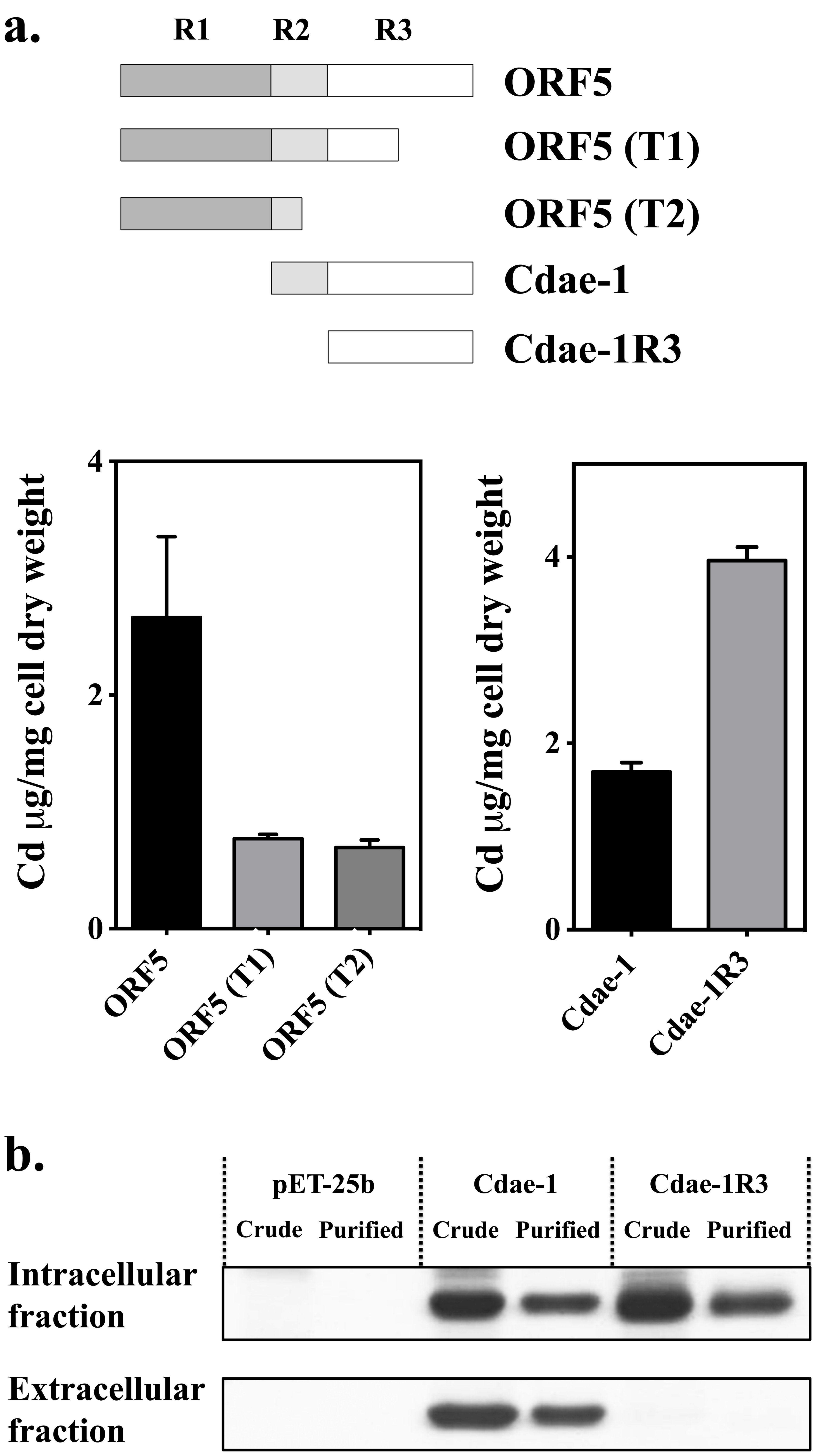
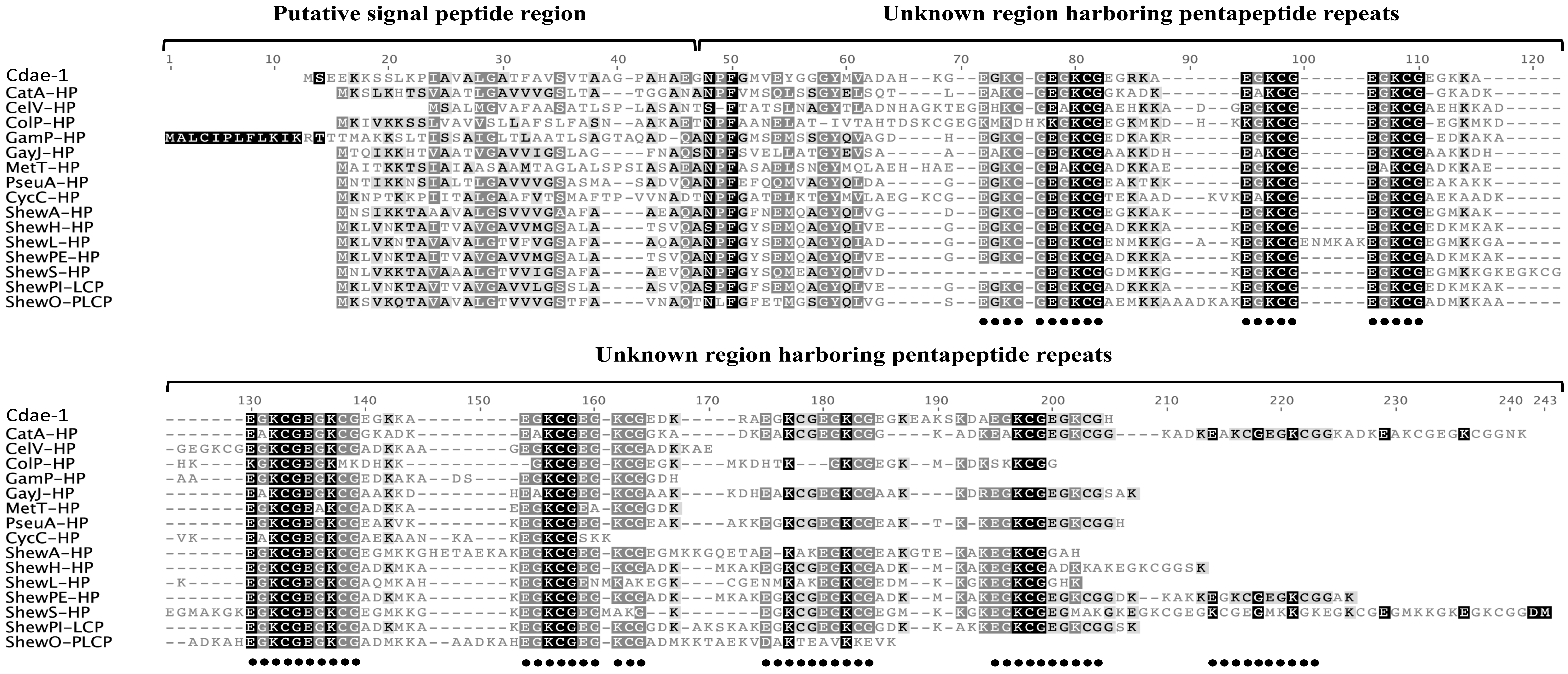

a.
R1
R2
R3
ORF5
ORF5 (T1)
ORF5 (T2)
Cdae-1
Cdae-1R3
Cd μg/mg cell dry weight
4
2
0
ORF5
ORF5 (T1)
ORF5 (T2)
Cd μg/mg cell dry weight
4
2
0
Cdae-1
Cdae-1R3
b.
pET-25b
Cdae-1
Cdae-1R3
Crude
Purified
Crude
Purified
Crude
Purified
Intracellular fraction
Extracellular fraction

Fig. 4



●: Amino acids identified to be highly conserved between sequences representing a unique pentapeptide repeat motif EXKCG (X: A/G)

Fig. 5