Biotechnology Reports 33 (2022) e00709

Contents lists available at ScienceDirect

Biotechnology Reports

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

Research article

Isolation and characterization of a recombinant class C acid phosphatase from *Sphingobium* sp. RSMS strain

Shyam Sunder Rangu [a], Rahul Singh [b], Neeraj Kailash Gaur [b], Devashish Rath [a], Ravindra D. Makde [b,*], Rita Mukhopadhyaya [c,**]

[a] *Applied Genomics Section, Bhabha Atomic Research Centre, Mumbai, 400085, India*
[b] *Beamline Development and Application Section, Bhabha Atomic Research Centre, Mumbai, 400085, India*
[c] *Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai, 400085, India*

ARTICLE INFO

Keywords:
Class C acid phosphatase
Sphingobium sp. RSMS strain
Monobutyl phosphate

ABSTRACT

Tributyl phosphate (TBP) is extensively used in nuclear industry and is a major environmental pollutant. The mechanism for TBP degradation is not identified in any TBP-degrading bacteria. Here, we report identification of an acid phosphatase from *Sphingobium* sp. RSMS (Aps) that exhibits high specific activity towards monobutyl phosphate (MBP) and could be a terminal component of the TBP degradation process. A genomic DNA library of the bacteria was screened using a histochemical method which yielded 35 phosphatase clones. Among these, the clone that showed the highest MBP degradation was studied further. DNA sequence analysis showed that the genomic insert encodes a protein (Aps) which belongs to class C acid phosphatase. The recombinant Aps was found to be a dimer and hydrolysed MBP with a K_{cat} 68.1 ± 5.46 s^{-1} and K_m 2.5 mM ± 0.50. The protein was found to be nonspecific for phosphatase activity and hydrolyzed disparate organophosphates.

1. Introduction

Tributyl phosphate (TBP) is extensively used in the nuclear and automobile industry and has been a major environmental pollutant [9]. *Sphingobium* sp. RSMS strain has been identified for its ability to degrade TBP efficiently [16]. The final products of TBP degradation are n-butanol and inorganic phosphate. This degradation requires three different phosphoryl transfer enzymes or catalytic sites. These are phosphotriesterase, phosphodiesterase and phosphomonoesterase (popularly called as phosphatases), which sequentially catalyze the conversion of TBP to dibutyl phosphate (DBP), DBP to monobutyl phosphate (MBP) and MBP to n-butanol and inorganic phosphate, respectively. So far, the genetic pathway for TBP degradation has not been identified in any of the TBP-degrading bacteria [[16], [18]].

Enzymes with different structural architectures and catalytic sites have evolved to carry out phosphomonoesterase activity with different specificity towards the R-group attached to phosphate moiety. Bacterial non-specific acid phosphatases (bNSAP) are the subgroup of phosphatases known to hydrolyze the phosphomonoester bonds in disparate organophosphates at acidic to neutral pH [[8], [24]]. bNSAP are periplasmic or secretory phosphatases with polypeptide components of 25–30 kDa. They typically act on a broad range of organophosphates such as nucleotides, glucose phosphates, and glycerol phosphate but lack activity towards diesters and cyclic nucleotides. bNSAPs are grouped into three different families based on the sequence similarity, namely, class A, class B and class C [24]. Class A bNSAPs are histidine phosphatases with alpha-helical fold and belong to phosphatidyl acid phosphatase type-2 superfamily. Class B and class C bNSAPs are evolutionarily related and belong to haloacid dehalogenase (HAD) superfamily; they are magnesium-dependent metalloenzymes and share a Rossmannoid class of alpha-beta fold [3]. Class B, class C and some plant class C-like phosphatases have been grouped together as a DDDD family based on the conservation of the four aspartate residues at the four distinct positions of the polypeptides [29]. These aspartate residues are responsible for binding magnesium divalent cofactor into the active site. Class C enzymes contain a signal peptide typical of bacterial lipoprotein and anchor on the outer membrane of Gram-negative bacteria using cysteine residue [[7], [13], [19], [21], [22]]. OlpA enzyme of *Chryseobacterium meningosepticum* was the prototype enzyme of class C bNSAPs [15]. Subsequently, class C enzymes were characterized from

* Corresponding author.
** Corresponding author.
E-mail addresses: ravimakde@rrcat.gov.in (R.D. Makde), rita45@gmail.com (R. Mukhopadhyaya).

https://doi.org/10.1016/j.btre.2022.e00709
Received 24 September 2021; Received in revised form 28 January 2022; Accepted 7 February 2022
Available online 8 February 2022

Haemophilus influenzae (*e* (P4)) [[21], [22]], *Bacillus anthracis* [7], *Streptococcus equisimilis* (LppC) [13], *Staphylococcus aureus* (SapS) [6], *Clostridium perfringens* [20], *Mycoplasma bovis* [27] and *Helicobacter pylori* (HppA) [19]. OlpA and HppA have been shown to possess activity for both 5′ and 3′ nucleotides with preference of former over later at acidic to neutral pH. *e* (P4) enzyme has been shown to possess nicotinamide mononucleotide (NMN) hydrolase activity [26].

Here, we report isolation and characterization of an efficient phosphatase from *Sphingobium* sp. RSMS. The recombinant protein is shown to degrade MBP with high-specific activity. Sequence analysis suggests that this polypeptide belongs to class C acid phosphatase; however, it lacks the typical signature of signal sequence at the N-terminal region for periplasmic localization or secretion. The recombinant protein was found to be non-specific and showed catalytic activity towards a variety of organophosphate substrates such as adenosine triphosphate (ATP), adenosine diphosphate (ADP), adenosine monophosphate (AMP), phenolphthalein diphosphate (PDP), *para*-nitrophenyl phosphate (*p*NP) and β-glycerol phosphate (β-GP) besides monobutyl phosphate (MBP).

2. Materials and methods

2.1. Bacterial strains and growth conditions

Sphingobium sp. RSMS strain was grown in Luria Bertani (LB) medium and *E. coli* strains were grown in LB or 2xYT medium at 30 °C and 37 °C, respectively. Aeration for liquid medium was provided by continuous shaking at 180 rpm. The *E. coli* EP1300-T1^R genomic DNA library clones were grown under chloramphenicol (33 μg ml^{-1}) antibiotic selection.

2.2. Genomic DNA library preparation and screening of phosphatase clones

Genomic DNA library (with a mean insert of 40 kb) of *Sphingobium* sp. RSMS strain was constructed in *E. coli* EP1300-T1^R strain (EP1300) using Copy-Control Fosmid Library production kit (Epicenter Biotechnologies, Madison, WI). Briefly, genomic DNA of *Sphingobium* sp. RSMS was fragmented to 40–50 kb by repeated pipetting. The fragmented DNA was cloned into pCC1FOS as per the manufacturer instructions. Using the formula, $N = \ln(1\text{-}P)/\ln(1\text{-}f)$, we determined the number of genomic DNA library clones required to ensure that any given DNA sequence is represented within the library. Here, P is the probability (expressed as a fraction), f is the proportion of the genome contained in a single clone and N is the required number of genomic DNA library clones. The number of clones required to ensure representation of entire *Sphingobium* sp. RSMS genome (5.12 Mb), at 99% probability with a mean insert size of 40 kb was found to be 590 ($N = \ln(1 - 0.99)/\ln(1 - [4 \times 10^4$ bases/5.12×10^6 bases$])$).

We have used PDP substrate (a phosphatase substrate) for selecting the phosphatase positive clones. PDP carries two phosphate moieties linked through two separate phosphomonoester bonds. Entire genomic DNA library (30,000 clones) was screened for phosphatase (phosphomonoesterase) clones by spreading them as isolated colonies on LB agar medium supplemented with PDP, at 1 mg ml^{-1} and methyl green at 50 μg ml^{-1} [14]. The plates were incubated for one week at 37 °C for the development of green coloration on the colonies, which indicated presence of phosphatase activity.

2.3. Isolation of an efficient MBP-degrading clone

A single colony of EP1300-T1^R-pho (35 clones) was inoculated into 5 ml of LB medium supplemented with chloramphenicol (33 μg ml^{-1}) and incubated overnight at 37 °C/180 rpm. The overnight grown cells were harvested, washed and resuspended in 1 ml of 10 mM Tris-Cl (pH 8) and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cell density was equalized using OD_{600nm} and the cells were lysed on ice by sonication (1 s on, 1 s off for 1 min at 30% amplitude). The lysate was subjected to centrifugation at 5000 rpm (4700 x *g*) for 10 min at 4 °C to remove unbroken cells. The protein concentration in the cellular extract was measured using Bradford method (Sigma). Equal amount of protein from each clone was incubated with 10 mM Tris-Cl (pH 7) supplemented with 5 mM MBP. The reaction mixtures were incubated overnight at 30 °C and the inorganic phosphate released due to MBP hydrolysis was estimated using phosphomolybdate method [2]. The clones (designated as EP1300-T1^R-MBPase clones) that showed highest end product (indicating MBP degradation) were studied further.

2.4. Isolation of acid phosphatase gene from MBP-degrading clones

Both ends of the genomic insert from three isolated MBP-degrading clones (EP1300-T1^R-MBPase) were sequenced using vector specific pCC1FOS-F and pCC1FOS-R primers (Table 1). The obtained partial sequences of genomic insert were aligned against the *Sphingobium* sp. RSMS genome (**GenBank accession: CP053222.1)** using NCBI-BLAST similarity search to identify the genomic region inserted into these clones. From amongst the list of annotated genes present in this reconstructed region of the *Sphingobium* sp. RSMS genome, an annotated *acid phosphatase* (*aps*) gene was chosen for further investigation.

2.5. Similarity search and signal peptide analysis

Similarity search for Aps was carried out using NCBI-BLAST analysis with default settings. Further, Sphingomonad strains were excluded from search criteria in order to identify similar proteins from other organisms. Multiple sequence alignment for these sequences was carried out using *multAlin* program (http://multalin.toulouse.inra.fr/multalin/) [4]. Signal peptide prediction for the proteins was carried out using SignalP-5.0 (http://www.cbs.dtu.dk/services/SignalP/) [1].

2.6. Cloning, expression and purification of Aps

Coding DNA for full-length Aps (**NCBI accession: WP_137709535.1**) and truncated Aps (Aps-t; deletion of 19 residues at N-terminus, which could be a periplasmic tag) were PCR amplified (using the primers mentioned in Table 1) from *Sphingobium* sp. RSMS strain. The amplified PCR products were purified, digested with *Bam*HI and *Hind*III and cloned into the same restriction sites of pST50Tr. This formed an in-frame fusion construct of the *aps/aps-t* with N-terminal Streptavidin-His6x-TEVtag (3.8 kDa). The identity and integrity of the Aps encoding region in the constructs were verified by DNA sequencing. The cloned constructs were transformed into an expression host, *E. coli* Rosetta (DE3) pLysS strain. The expressing clone for both the constructs (Rosetta-*aps* and Rosetta-*aps-t*) were grown in 2 L of 2xTY broth containing ampicillin (50 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}) at 37 °C till OD_{600nm} reached to ~0.6 and the culture was induced for four hours by addition of 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37 °C. The cells were harvested at 4 °C by centrifugation at 9500 x *g* for 5 min. The cell pellet was suspended in T200-$MgCl_2$ buffer (20 mM Tris-Cl pH 8.0, 200 mM NaCl, 2 mM $MgCl_2$). Suspended cells were flash frozen in liquid nitrogen, thawed and then subjected to probe sonication on ice (pulse 3 s on/off; 40% amplitude; frequency-20 kHz; 750 W) for cell lysis. The lysate was subjected to centrifugation at 4 °C on 28,000 x *g* for 30 min to separate the soluble fraction from the pellet. SDS-PAGE analysis concluded the presence of Aps and Aps-t in pellet fractions as inclusion bodies. The pellet was suspended and washed twice using T200-$MgCl_2$ containing 0.1% Triton-X100 (v/v); followed by washing using T200-$MgCl_2$ to remove traces of Triton-X100. The pellet was then solubilized in T200-$MgCl_2$ containing 8 M urea. SDS-PAGE showed that the pellet was largely composed of Aps. The urea soluble fractions were subjected to Ni-affinity chromatography by batch method at 37 °C. The solutions were passed several times through 3 ml manually packed Ni-Sepharose resin column (GE Healthcare)

Table 1
Strains, plasmid vectors and oligonucleotide sequences used in this study.

Strain/primer/vector	Purpose	source
Oligonucleotide primers		
pCC1FOS-F (GGATGTGCTGCAAGGCGATTAAGTTGG)	forward primer for partial sequencing of insert from pCC1FOS-MBPase	epicenter
pCC1FOS-R (CTCGTATGTTGTGTGGAATTGTGAGC)	reverse primer for partial sequencing of insert from pCC1FOS-MBPase	epicenter
aps-F (cgtggatccATGTTGGC GGGCTGTTCGGCC)	Forward primer for amplification of *aps* gene	This study
aps-t-F (cgtggatccACGCCCC CGGCCGGGATG)	Forward primer for amplification of *aps-t* gene	This study
aps-R (cgtaagcttaTTTCTCC CCA TCGGCAGGCGC)	Reverse primer for amplification of *aps* and *aps-t* gene	This study
Plasmids		
pCC1FOS	plasmid vector used for genomic DNA library preparation	epicenter
pCC1FOS-MBPase	pCC1FOS plasmid carrying genomic insert of RSMS strain responsible for MBP degradation	This study
pST50Tr	Used for the expression of Aps	[28]
pST50Tr-*aps*	pST50Tr carrying *aps* gene	This study
pST50Tr-*aps*-t	pST50Tr carrying *aps*-t gene	This study
Bacterial strains		
Sphingobium sp. RSMS strain (RSMS strain)	TBP-degrading bacterial strain	Rangu et al. .,2014
E. coli EP1300-T1^R (EP1300)	an *E. coli* strain used for the construction of genomic library of RSMS strain	epicenter
EP1300-T1^R-pho	35 phosphatase clones isolated from genomic DNA library of RSMS strain.	This study
EP1300-T1^R-MBPase	*E. coli* EP100 carrying pCC1FOS-MBPase	This study
E. coli XL-1 blue	For cloning *aps* and *aps*-t genes	Lab collection
E. coli Rosetta (DE3) pLysS	For overexpression of Aps-t	Novagen
Rosetta-*aps*	*E. coli* Rosetta (DE3) pLysS carrying pST50Tr-*aps*	This study
Rosetta-*aps*-t	*E. coli* Rosetta (DE3) pLysS carrying pST50Tr-*aps*-t	This study

equilibrated with T200-$MgCl_2$ containing 8 M urea. The column was washed using T200-$MgCl_2$ containing 8 M urea and 50 mM imidazole for removing weakly bound proteins. The bound proteins were eluted using 20 ml of T200-$MgCl_2$ containing 8 M urea and 500 mM imidazole and immediately diluted using 20 ml of 20 mM Tris-Cl pH 8.0, 400 mM NaCl, 2 mM $MgCl_2$ to reduce the urea concentration to 4 M and to increase NaCl concentration to 300 mM. The 40 ml solution was subjected to overnight dialysis at 4 °C against 2 L refolding buffer (20 mM Tris-Cl pH 8.0, 200 mM NaCl, 2 mM $MgCl_2$, 5 mM 2-mercaptoethanol and 200 mM L-arginine). The refolded 40 ml protein solution was subjected to high-speed centrifugation 30,000 x *g* for 30 min at 4 °C. and the supernatant was concentrated using stirred cells containing ultrafiltration membrane of 10 kDa molecular weight cut-off (Amicon). The protein solubility (after refolding steps) was verified using analytical size-exclusion chromatography with T200-$MgCl_2$ buffer using Superdex 200, 10/300 GL column on AKTA purifier system (GE Healthcare). The refolded Aps-t was used for all further experiments. Glycerol (v/v concentration 20%) was added as a cryoprotectant to the purified protein before storing at −70 °C until further use.

2.7. Enzymatic analysis

The general phosphatase activity of refolded Aps-t was measured by *p*NPP assay. The *p*NPP assay was also used to find the other optimum parameters (temperature and pH). All enzymatic assays were performed at 37 °C in buffer containing 50 mM MES (pH 6), 100 mM NaCl and 1 mM $MgCl_2$. The Malachite green (MG) phosphate assay [[9], [11]] was used to evaluate the enzymatic activity for the substrates, AMP, ADP, ATP, 2′-deoxyadenosine 5′-triphosphate (dATP), monobutyl phosphate (MBP) and β-GP. Both *p*NPP and MG phosphate assays were performed as previously described, with some modifications. The substrate for the *p*NPP assay was prepared by dissolving *p*NPP (Sigma-Aldrich) in 0.1 M sodium acetate buffer at pH 4.5 and the assay was performed in a reaction volume of 500 µl with 5 mM *p*NPP substrate and 80 ng of purified Aps-t. The enzymatic reaction was stopped after 5 min of incubation by adding 1 ml of 0.2 M NaOH to the reaction mixture. The amount of product (paranitrophenylate, *p*NP) was estimated by colorimetric measurement at 405 nm (extinction coefficient of 18,000 M^{-1} cm^{-1}). For MG phosphate assay, the MG reagent (300 ml) was prepared by mixing 350 mg malachite green pigment along with 3.2 g of ammonium molybdate in 4.5 to 5 M HCl followed by filtration of the solution through 0.4 µm filter. The enzymatic reaction was performed in a reaction volume of 100 µl with 2 mM substrate and 40 to 120 ng of purified Aps-t. The enzymatic reaction was stopped after 5 min of incubation by adding 1 ml of MG reagent in the reaction mixture. The dark green color developed within 5 min and remained stable for 30 min. The intensity of the coloured product was measured at 660 nm by using the "POLARstar OMEGA microplate reader" (read volume 200 µl). The specific activity of the enzyme was expressed as µmoles of phosphate equivalent (sodium phosphate) mg^{-1} min^{-1}. Kinetic studies were performed by changing substrate concentration between 0.1 to 8 mM in buffer containing 50 mM MES (pH 6), 100 mM NaCl and 1 mM $MgCl_2$. The kinetic parameters were deduced by curve fitting to Michaelis Menten enzyme kinetics model through Graphpad Prism™ software.

2.8. Sequence conservation and structure prediction of Aps

Aps polypeptide sequence was used to retrieve its sequence homologs from Uniprot, NCBI and PDB databases. To explore the evolutionary aspect, a BLAST similarity search for Aps was conducted against the Uniref 50 database. This resulted in about 450 protein sequences, each representing a particular cluster of amino acid sequences that are not more than 50% identical thus avoiding any bias that could arise due to over or under representation of any particular organism (as model organisms tend to have more recurring sequences in the database).

Clustal Omega [25] was used to perform multiple sequence alignment (MSA) of Aps homologs that were either enzymatically or structurally characterized. Further, three-dimensional structures of Aps-t was predicted using the AlphaFold2 server [10]. AlphaFold2 is a deep learning algorithm that incorporates physical and biological information about protein structure using multi-sequence search/alignments (MMSeqs2). This accurately predicts the three-dimensional monomeric structure of a protein. The structure was analyzed using PyMol [5]. The Mg^{2+} was incorporated into active site of Aps-t model by structural overlap with *Haemophilus influenzae e* (P4) acid phosphatase (PDB: 3OCY). The superpositions of the Aps-t model with that of other class C phosphatases and root mean square deviation (RMSD) were computed using PyMol software.

3. Results

3.1. Genomic DNA library preparation

Genome size of the *Sphingobum* sp. RSMS strain is 5.12 MB [17]. About 30,000 genomic DNA library clones (with a mean insert size of 40 kb) were obtained using the Copy Control Fosmid Library production

kit. About 590 clones represent the entire genome of *Sphingobium* sp. RSMS strain with 0.99 probability. Hence, 30,000 genomic DNA library clones which represent a 50 X coverage of the *Sphingobium* sp. RSMS strain genome would cover the entire genome.

3.2. Isolation of MBP-degrading gene from genomic DNA library

The genomic DNA library was screened to identify clones containing phosphatases from *Sphingobium* sp. RSMS. Our histochemical screening using PDP as a substrate identified 35 phosphatase positive clones (EP1300-T1^R-pho) from the genomic DNA library (Fig. 1a). The host strain, *E. coli* EP1300-T1^R (N1, N2 and N3) showed very less/nil green coloration on histochemical plates (Fig. 1a). *E. coli* strains can have periplasmic phosphatases (such as PhoA) which are usually expressed under phosphate limiting conditions [31]. The low background signal observed in the control strain (EP1300-T1^R) could be due to the use of LB medium (which have sufficient amount of inorganic phosphate) for the screening.

Nineteen of the 35 isolated phosphatase clones showed insignificant MBP degradation, equivalent to negative controls carrying empty vectors (N1 and N2) (data not shown). Sixteen out of 35 clones showed significant MBP degradation activity (Fig. 1b). MBP degradation was observed to be significantly higher for three clones (EP1300-T1^R-MBPase clones 1, 7 and 18) than the other MBP-degrading clones. Two clones (clones 8 and 14) displayed very high phosphatase activity for PDP substrate but exhibited much lower MBP hydrolysis activity than the best MBP active clones (Fig. 1a). Thus, clones 1, 7 and 18 were selected for further analysis. Partial sequencing (using vector specific primers, Table 1) of these three clones (pCC1FOS-MBPase) confirmed that they harboured identical genomic region.

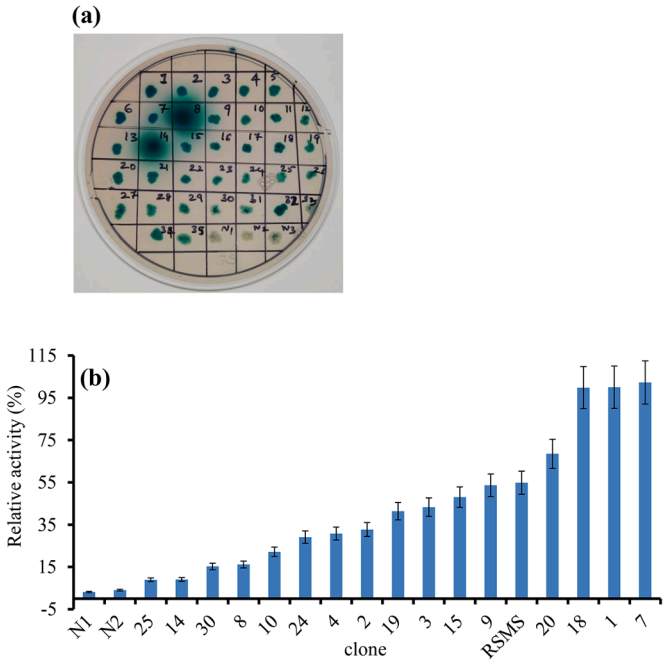

Fig. 1. Isolation of an efficient MBP-degrading clone. (a) For the isolation of phosphatase clones, the genomic DNA library clones were spread plated on to LB agar medium supplemented with PDP (1 mg ml^{-1}), methyl green (50 μg ml^{-1}) and incubated at 37 °C. The positive clones for phosphatase activity were identified by the green coloration of the colony after one week. Here, 1–35 represent isolated phosphatase clones and N1, N2 and N3 (negative controls) represent E. coli EP1300-T1R strain carrying empty pCC1FOS vector. (b) Equal amount of protein extracts from each clone was incubated with 10 mM Tris-Cl (pH 7) supplemented with 5 mM MBP, incubated overnight at 30 °C and the inorganic phosphate released due to MBP hydrolysis was estimated. Cellular extracts from *Sphingobium* sp. RSMS strain was used as control. Graph represents relative activity of each clone with respect to clone 1, which showed the maximum activity (100%). The values are average of three independent experiments. Fig. 1b shows data of the clones (a total of 16 clones) which showed more activity than the negative controls (N1 and N2).

Using similarity search (NCBI-BLAST) these partial sequences were mapped to chromosome 1 of *Sphingobium* sp. RSMS genome (**GenBank accession: CP053222.1**). Sequence analysis of the insert suggested presence of a 45,578 bp genomic insert (3,411,013 bp–3,456,591 bp of Chromosome 1) in these clones having 53 annotated ORFs (Table ST1), out of which one was annotated as *acid phosphatase* (**NCBI accession**: **WP_137709535.1**). This gene was considered as the putative MBP-degrading candidate (Aps). One of the EP1300-T1^R-MBPase clones (clone 1) was investigated further.

3.3. Similarity search and signal peptide analysis

Aps showed significant sequence similarity (67–100%) to several annotated acid phosphatases from Sphingomonad strains with query coverage of 90–100% (data not shown). Sphingomonads include organisms from four closely related genera, namely *Sphingobium, Sphingomonas, Sphingopyxis, Novosphingobium*. Notably, when Sphingomonads were excluded from the search criteria, an acid phosphatase from *Nostoc* sp. 3335mG (**NCBI accession: WP_110156888.1**) showed significant identity (59.7%) with 96% query coverage.

The SignalP-5.0 program predicts presence of conventional signal peptides (from Archaea, Gram-positive bacteria and Gram-negative bacteria) used by three known pathways, namely, 1) Sec/SPI, 2) Sec/SPII, 3) Tat/SPI. Signal peptide prediction by SignalP-5.0 revealed that Aps does not possess any conventional signal peptide. But, 86 out of 100 similar proteins (first 100 hits from BLAST search for Aps) from other Sphingomonad strains were predicted to have signal peptide by SignalP-5.0 (data not shown). Multiple sequence analysis of these 100 proteins suggested that the remaining 14 proteins (predicted to have no signal peptide) and Aps are shorter by a few amino acids at N-terminus, which corresponds to a part of the signal peptide for other 86 proteins (Fig. S1). All the characterized class C acid phosphatases reported till date exhibit signal peptide (Table ST2).

3.4. In-silico analysis of Aps

The Aps open reading frame codes for a 28.9 kDa (277 residues) protein with a theoretical pI of 5.67. Protein sequence analysis suggested that Aps belongs to class C acid phosphatases of haloacid dehalogenase (HAD) like hydrolases superfamily (Fig. 2). Analysis of the Aps showed the presence of VF(D)V(D)ETVMLNIGY and IAMGG(D)QLG(D)F sequence motifs respectively at the amino- and carboxy- terminals of the protein with highly conserved four aspartate residues (Fig. 2& 3), which is characteristic of class C acid phosphatases and DDDD family proteins. As indicated by previous studies the first aspartate of the conserved motif is known to act as a nucleophile to attack the phosphorus center of the substrate during reaction and others are essential for magnesium cofactor binding [12]. Aps appears to have a 20 amino acid long insertion towards the N-terminal of the protein (Fig. 2). The implications of such insertion remain elusive. Structural superpositions of predicted Aps-t structure with crystal structures of *Haemophilus influenzae e* (P4) acid phosphatase (PDB ID: 3OCY) and class C acid phosphatase from *B. anthracis* (PDB ID: 3I33) were carried out (Fig. S2). RMSD of about 145 C-alpha aligned atoms between the structures are 1.15 and 1.3 Å^2, respectively. In all the three structures the active site is open and highly solvated. Aps-t structure is rather unusual in two respects: a) C-terminal region of Aps-t does not show regular secondary structures and the region is involved in the dimer formation in other two structures. b) A unique insertion of a loop (magenta color in Fig. 3 and residues highlighted in blue in Fig. S2) is seen in Aps-t structure which is absent in other two structures.

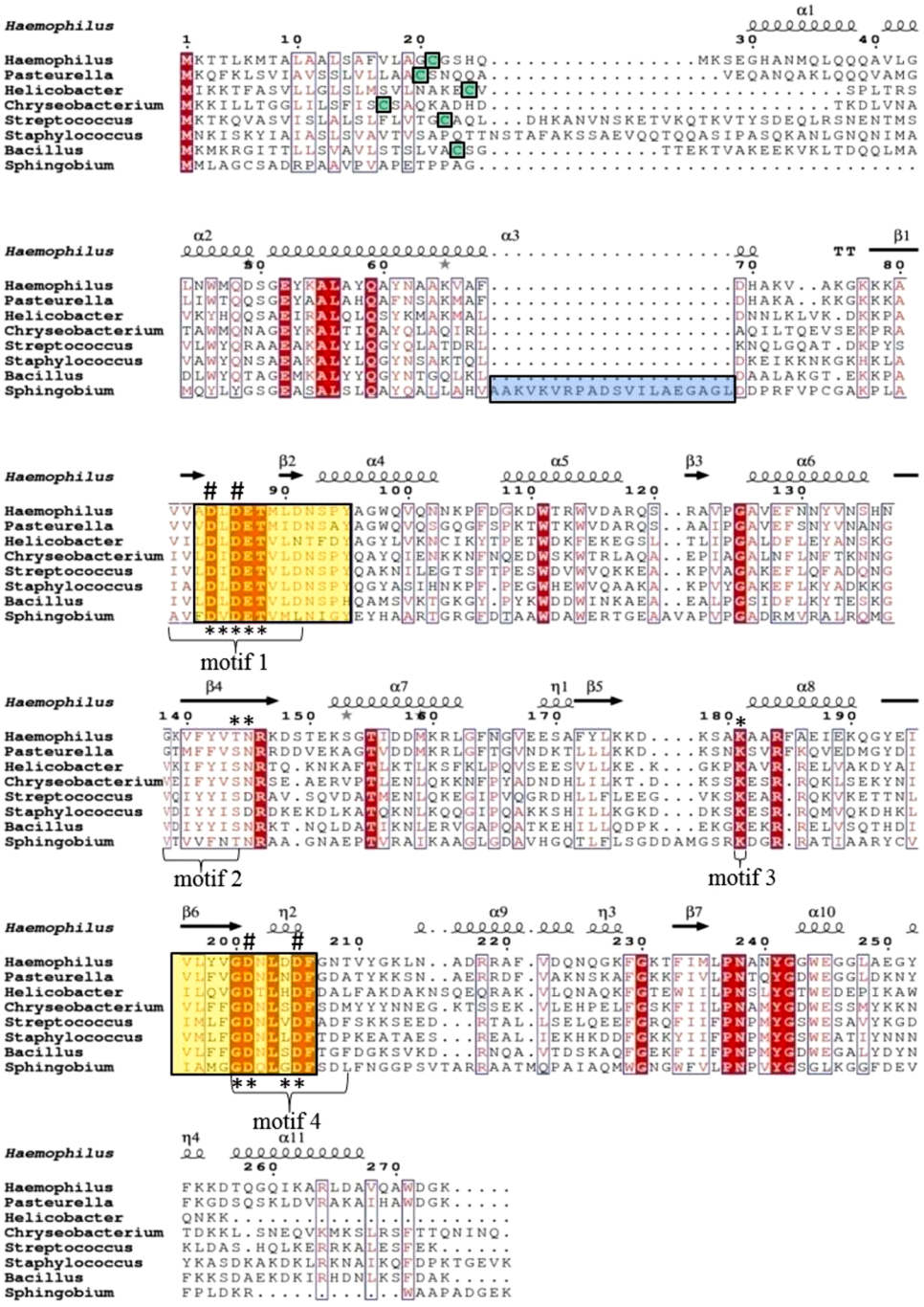

Fig. 2. Multiple sequence alignment of class C enzymes which are either structurally or functionally characterized. Figure representing multiple sequence alignment amongst characterized class C nonspecific acid phosphatases from *Haemophilus influenzae* (P26093), *Pasteurella multocida* (B9VWB2), *Helicobacter pylori* (AAQ88280.1), *Chryseobacterium meningosepticum* (O08351), *Staphylococcus aureus* (Q99WS0), *Streptococcus dysgalactiae* subsp. *equisimilis* (O05471), *Bacillus anthracis* (WP_000782788.1) and *Sphingobium* sp. RSMS strain (WP_137709535.1). * indicate residues at the active site of the class C acid phosphatases. The highly conserved bipartite aspartate residues (DDDD) are indicated with symbol #. The amino acid residues in the box (filled with yellow color) represent the consensus sequences of class C acid phosphatases. The cysteine residue at N terminal (first 15–35 residues) is highlighted in green color. Motif 1–4 represent the conserved sequences of HAD superfamily proteins. The residues highlighted in blue color represent insertion. The secondary structures are overlaid at the top of MSA. The analysis was carried out using CDD-NCBI and ESPript 3.0 [23] (https://espript.ibcp.fr/ESPript/ESPript/).

3.5. Purification and oligomeric state of Aps

Full-length protein was over expressed in different *E. coli* expression hosts including BL21(DE3), BL21(DE3) pLysS, Rosetta, Origami, but the protein was observed in inclusion bodies. Lowering the induction temperature to 28 °C and 18 °C, varying IPTG concentration and changing the expression host did not help to yield the protein in the soluble form in appreciable quantity. Aps belongs to class C acid phosphatase which usually have 20–30 residues N-terminal signal sequence. We decided to clone an alternative construct which excluded 19 N-terminal residues (Aps-t) which might hamper the solubility of the protein in *E. coli* cytoplasm. Aps-t too was observed in inclusion bodies. We therefore decided to purify the Aps and Aps-t from the inclusion bodies using refolding protocol. We could not obtain the full-length Aps in soluble form by refolding protocol. However, Aps-t was obtained in soluble form. A total of about 30 mg of Aps-t was purified from a 2-liter culture. The refolded protein was found to be pure as adjudged from SDS-PAGE. Aps-t showed a single band corresponding to molecular mass of 30.6 kDa on SDS-PAGE (Fig. S3). This matches very well to its calculated molecular mass, 31 kDa (including Streptavidin-His6x-TEVtag). Further, on the size-exclusion chromatography, the protein was eluted as a major peak corresponding to molecular mass of ~55 kDa (Fig. S3) which is reasonably close to its calculated dimeric mass (62 kDa). This indicates that the Aps-t exists as a dimeric assembly in solution. The protein is eluted slightly later than the exact dimeric mass; this may be due to some weak interactions of the protein with Superdex column matrix.

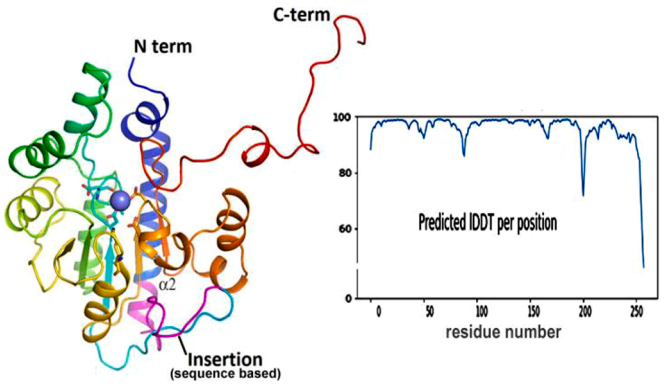

Fig. 3. Three-dimensional monomeric structure of Aps-t predicted by AlphaFold2 server. The structure of Aps-t was predicted using AlphaFold2 server (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipyn). The Local Distance Difference Test (LDDT), a confidence score, is shown in the right panel. Score >90 are expected to be modelled with very high accuracy and score 90–70 is predicted with reasonably good accuracy. The structural cartoon is shown in spectrum color (blue to red). The Mg^{2+} (sphere) was incorporated into active site by structural overlap with *Haemophilus influenzae e* (P4) acid phosphatase (PDB ID: 3OCY). It is showing strictly conserved aspartate residues (shown in sticks) near to Mg^{2+} (sphere) binding site. Multiple sequence alignment with the structurally or enzymatically characterized class C acid phosphatase shows presence of an insertion in Aps (magenta coloured).

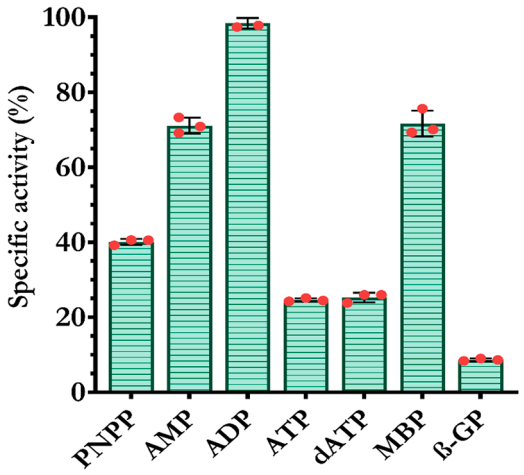

Fig. 5. Specific activities of Aps-t for various substrates. Average values are represented in bar graph with individual measurements in red dots. The error bars depict standard deviation between the replicates. 100% corresponds to specific activity of 69.2 μmol min^{-1} mg^{-1}. The inorganic phosphate released due to hydrolysis of AMP, ADP, ATP and MBP was estimated using Malachite green assay and the *para*-nitrophenol (*p*NP) released due to *p*NPP hydrolysis was estimated by *p*NPP assay.

3.6. Enzymatic activity

Aps-t showed enzymatic activity in a broad range of pH from 4.5 to 7 with maximum activity at pH 6 (Fig. 4b). As expected from an acid phosphatase, the activity of the Aps-t drops rapidly at the basic pH. The enzyme also showed activity in a broad range of temperatures (Fig. 4a). The enzymatic activity of Aps-t increased with an increase in reaction temperature with maximum activity at 55 °C, followed by a rapid decline in enzymatic activity with further increase in temperature. Since the optimum growth condition of the *Sphingobium* (alpha proteobacteria) is in the range of 30 to 37 °C, we performed rest of the enzymatic assays at 37 °C. The Aps-t showed no inhibition of activity by either EDTA (up to 20 mM) or inroganic phosphate (up to 500 mM) in the assay buffer (Fig. S4).

Bioinformatic studies suggested that the Aps belongs to class C acid phosphatase of haloacid dehalogenase (HAD) superfamily. Type C acid phosphatase represents a sub-group of phosphatases with a very broad range of substrate specificity known as nonspecific acid phosphatases (NSAPs). As expected from the representative member of NSAPs, Aps-t showed a broad range of substrate specificity. The enzyme had the ability to hydrolyze mono-phosphate, diphosphate, as well as triphosphate salts of adenosine nucleoside (Fig. 5). Among the tested nucleotides ADP was the most preferred substrate whereas ATP was the least preferred (Specificityconstant; Table 2). Surprisingly, the active site of the enzyme was capable of binding with a highly charged molecule like ATP. This signified either a highly solvated active site or the presence of a positive charge center like Mg^{2+}at the active site. This supported the bioinformatics finding of Aps being a metalloenzyme with Mg^{2+} as cofactor. The enzyme had shown significant activity for MBP. However, no enzymatic activity for di- and tri- butyl phosphate was observed. The enzyme showed the highest specific activity for substrate ADP (67.7 ± 0.79 μmole min^{-1} mg^{-1}), followed by AMP and MBP with almost similar substrate specificity of 50 μmole min^{-1} mg^{-1} (Fig. 5). Kinetic studies showed that Aps-t exhibits highest turnover number 68.1 ± 5.46 s^{-1} for MBP but has relatively higher K_m value 2.503 ± 0.5007 mM. Among all the tested substrates, Aps-t had the least enzymatic activity for β-GP. Due to poor enzymatic activity for β-GP, we could not perform kinetic studies with this substrate.

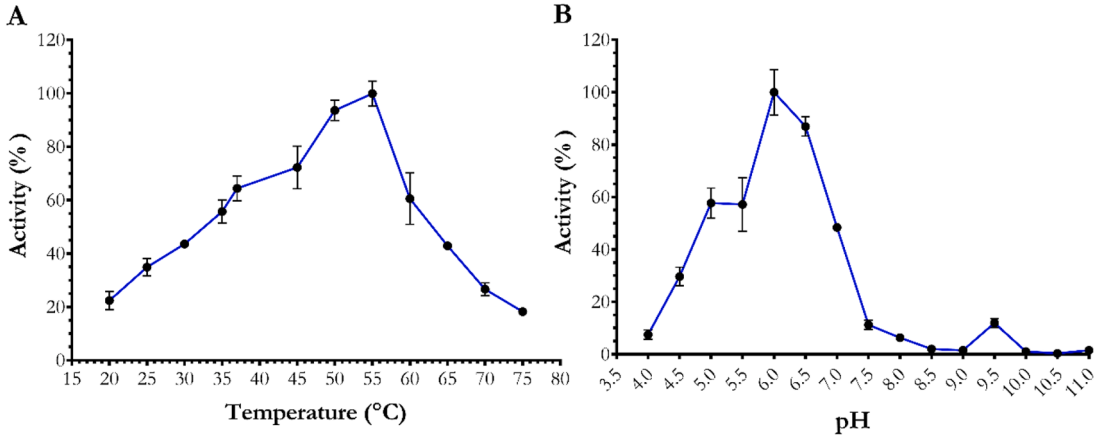

Fig. 4. Optimization of pH and temperature for Aps activity. The phosphatase activity was measured using the *p*NPP substrate at different temperature (A) and pH values (B). The activity was measured various pH values using buffers (each at 50 mM), sodium acetate (pH 4.0 to 5.5); MES (pH 6.0 and 6.5); Tris-Cl (pH 7.0 to 9.0); CAPS (pH 9.5 to 11).

Table 2
Enzyme kinetics of Aps-t with various substrates.

Sr. No	Substrate	V_{max} (μmol min^{-1})	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m (s^{-1} mM^{-1})
1.	AMP	55.7 ± 1.41	0.27 ± 0.02	29.3 ± 0.74	108.4 ± 10.84
2.	ADP	65.0 ± 1.36	0.19 ± 0.01	34.2 ± 0.71	178.5 ± 16.83
3.	ATP	33.1 ± 1.25	1.0 ± 0.12	17.4 ± 0.65	17.4 ± 2.78
4.	*p*NPP	29.7 ± 1.26	0.094 ± 0.014	15.6 ± 0.66	166.0 ± 32.70
5	MBP	129.5 ± 10.39	2.503 ± 0.5007	68.1 ± 5.46	27.2 ± 7.62

Purified Aps-t (40–120 ng) was incubated with increasing concentration of AMP, *p*NPP, ADP, ATP and MBP and incubated at 37 °C for 5 min. The inorganic phosphate released due to hydrolysis of AMP, ADP, ATP and MBP was estimated using Malachite green assay and the *para*-nitrophenol (*p*NP) released due to *p*NPP hydrolysis was estimated by *p*NPP assay. The rate of the reaction was plotted against substrate concentration and K_m, K_{cat} and K_{cat}/K_m were calculated.

4. Discussion

In this study we identified a novel non-specific acid phosphatase (Aps) from the alpha-proteobacterium, *Sphingobium* sp. RSMS strain. Screening of *Sphingobium* sp. RSMS strain genomic clones revealed three clones showing high specific activity for MBP hydrolysis. Sequence analysis of the clones showed that these three clones have identical insert from *Sphingobium* sp. RSMS genome and the insert harbors an *in silico* annotated phosphatase gene. This gene encodes polypeptide, Aps, which belongs to class C acid phosphatases of haloacid dehalogenase-like hydrolase (HAD) superfamily [29]. Aps possesses all the essential residues of class C motif including four aspartate residues which are required for binding to Mg^{2+} cofactor (Fig. 2 and Fig. 3). However, Aps lacks typical N-terminal signal sequence for periplasmic localization or lipoprotein as predicted by SignalP-5.0 server [1]. Aps orthologs are present in most of the known species of the genus *Sphingobium* and some related genera of alpha-proteobacteria. Majority of these proteins (86 out of 100) were predicted to have signal peptide, and those which do not (along with Aps), have short of few amino acids at N-terminal. Further, all the characterized class C acid phosphatases were predicted to have signal peptide by SignalP-5.0 server (Table ST2). Aps could have a non-canonical signal sequence, not recognized by the SignalP-5.0 server.

Purified recombinant Aps-t showed high specific activity for MBP hydrolysis and could be a candidate for MBP hydrolysis during TBP degradation process. Aps-t hydrolyses MBP with a turnover number of about 68 s^{-1}, highest among all the tested organophosphates. However, K_m of 2.5 mM for MBP is relatively higher. This could be due to structural simplicity of n-butyl group of MBP or non-physiological nature of MBP substrate. Aps was found to be a non-specific phosphatase enzyme since it hydrolyses terminal phosphate moieties of MBP, AMP, ADP, ATP, *p*NPP and β-GP. R-group attached to terminal phosphate moiety is structurally quite distinct from each other in all the organophosphates. It appears that Aps largely recognizes the terminal phosphate moiety of organophosphates in its active site with little attention to R-group attached to this phosphate moiety, suggesting that this protein could have potential applications in industry.

Aps shows significant sequence homology with two class C acid phosphatases in protein data bank (PDB) (PDB entries: 3PCT, 2I33) with about 28% and 24% sequence identity, respectively. Aps is dimeric in nature as suggested by size-exclusion chromatography result which is also consistent with the crystal structures of the published two proteins. Four highly conserved aspartate residues (Fig. 2) which defines DDDD family to which class C acid phosphatases belong, requirement for the magnesium cofactor binding are close enough to the predicted structural model of Aps (Fig. 3). In spite of having magnesium ion as cofactor for class C acid phosphatases, Aps activity is not inhibited by presence of up to 20 mM EDTA. Such kind of resistance to the inhibition by EDTA was earlier observed for at least two other class C acid phosphatases [[13], [30]]. The crystal structures of these two enzymes are not available. Structural superposition of Aps-t model with that of *Haemophilus influenzae e* (P4) acid phosphatase which is inhibited by EDTA (PDB: 3OCY; [26]) does not provide any obvious clue for resistance to EDTA inhibition (Fig. S2). Multiple sequence alignment with other class C acid phosphatases which are either enzymatically or structurally characterized reveals the presence of a 20 amino acid insertion in Aps (Ala 50 to Leu 69). In the predicted model, this insertion lies near α3 helice (Fig. 2 and 3; Fig. S2). Crystal structure of Aps will shed light on as to how Aps active site could accommodate structurally unrelated organophosphates including MBP and catalyze their hydrolysis. The crystal structure of any EDTA-resistant class C acid phosphatase structure is not yet available. Hence, the crystal structure of Aps may provide structural clue for its EDTA-resistance. Only such study could reveal exact monomeric and quaternary structures of this novel class C acid phosphatase.

Sphingobium sp. RSMS has been identified for its ability to efficiently degrade TBP [18]. It is likely that this process is completed by three different enzymes, phosphotriesterase, phosphodiesterase and phosphomonoesterase, which sequentially catalyze the conversion of TBP to DBP, DBP to MBP and MBP to n-butanol and inorganic phosphate, respectively. Here, we report a novel phosphatase form *Sphingobium* sp. RSMS strain which exhibits high specific activity for monobutyl phosphate (MBP) and could be a terminal component of the TBP degradation process.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Declaration of Competing Interest

Author 1, Shyam Sunder R, declares that he has no conflict of interest, Author 2, Rahul Singh, declares that he has no conflict of interest, Author 3, Neeraj K. Gaur, declares that he has no conflict of interest, Author 4, Devashish Rath, declares that he has no conflict of interest, Author 5, Ravindra D. Makde, declares that he has no conflict of interest, Author 6, Rita Mukhopadhyaya, declares that she has no conflict of interest

Acknowledgements

Authors acknowledge Dr. AVSSN Rao for his guidance and support during manuscript preparation. Authors acknowlegde the use of PX-BL21 Indus-2 biochemical facility.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.btre.2022.e00709.

References

[1] J.J. Almagro Armenteros, et al., SignalP 5.0 improves signal peptide predictions using deep neural networks, Nat. Biotechnol. 37 (4) (2019) 420–423.
[2] B.N. Ames, Assay of inorganic phosphate, total phosphate and phosphatases, in: F. Elizabeth, VG Neufeld (Eds.), Assay of inorganic phosphate, total phosphate and phosphatases, Methods in Enzymology (1966) 115–118.
[3] A.M. Burroughs, et al., Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes, J. Mol. Biol. 361 (5) (2006) 1003–1034.
[4] F. Corpet, Multiple sequence alignment with hierarchical clustering, Nucleic. Acids. Res. 16 (22) (1988) 10881–10890.
[5] W.L. DeLano, The PyMOL molecular graphics system, DeLano Scientific LLC, San Carlos, CA. PyMOLMol Graph. Syst. (2002). World Wide Web URL http//www pymol org.
[6] E.M. du Plessis, et al., Characterization of a phosphatase secreted by Staphylococcus aureus strain 154, a new member of the bacterial class C family of nonspecific acid phosphatases, Syst. Appl. Microbiol. 25 (1) (2002) 21–30.
[7] R.L. Felts, et al., Cloning, purification and crystallization of Bacillus anthracis class C acid phosphatase, Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 62(Pt 7) (2006) 705–708.
[8] N.U. Gandhi, B.S. Chandra, A comparative analysis of three classes of bacterial non-specific Acid phosphatases and archaeal phosphoesterases: evolutionary perspective, Acta Inform. Med 20 (3) (2012) 167–173. Sepdoi: 10.5455/aim.2012.20.167-173.
[9] K. Itaya, M. Ui, A new micromethod for the colorimetric determination of inorganic phosphate, Clin. Chim. Acta 14 (3) (1966) 361–366.
[10] J. Jumper, R. Evans, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.
[11] U. Lorenz, Protein tyrosine phosphatase assays, Curr. Protoc. Immunol. Chapter 11 (2011). Unit 11.17.
[12] R.D. Makde, et al., Structural and mutational analyses reveal the functional role of active-site Lys-154 and Asp-173 of Salmonella typhimurium AphA protein, Arch. Biochem. Biophys. 464 (2007) 70–79.
[13] H. Malke, Cytoplasmic membrane lipoprotein LppC of Streptococcus equisimilis functions as an acid phosphatase, Appl. Environ. Microbiol. 64 (7) (1998) 2439–2442.
[14] K.S. Nilgiriwala, et al., Cloning and overexpression of alkaline phosphatase PhoK from Sphingomonas sp. strain BSAR-1 for bioprecipitation of uranium from alkaline solutions, Appl. Environ. Microbiol. 74 (17) (2008) 5516–5523.
[15] C. Passariello, et al., The molecular class C acid phosphatase of Chryseobacterium meningosepticum (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5′-nucleotides, Biochim. Biophys. Acta 1648 (1–2) (2003) 203–209.
[16] S.S. Rangu, et al., Involvement of phosphoesterases in tributyl phosphate degradation in Sphingobium sp. strain RSMS, Appl. Microbiol. Biotechnol. 100 (1) (2016) 461–468.
[17] S.S. Rangu, et al., Whole-Genome Sequencing of sphingobium sp. strain RSMS, a highly efficient Tributyl phosphate-degrading bacterium, Microbiol. Resour. Announc. 9 (42) (2020).
[18] S.S. Rangu, et al., Tributyl phosphate biodegradation to butanol and phosphate and utilization by a novel bacterial isolate, Sphingobium sp. strain RSMS, Appl. Microbiol. Biotechnol. 98 (5) (2014) 2289–2296.
[19] T.J. Reilly, M.J. Calcutt, The class C acid phosphatase of helicobacter pylori is a 5′ nucleotidase, Protein Expr Purif 33 (1) (2004) 48–56.
[20] T.J. Reilly, et al., Characterization of a unique class C acid phosphatase from Clostridium perfringens, Appl. Environ. Microbiol. 75 (11) (2009) 3745–3754.
[21] T.J. Reilly, et al., Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase, J. Bacteriol. 181 (21) (1999) 6797–6805.
[22] T.J. Reilly, A.L. Smith, Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4), Protein Expr. Purif. 17 (3) (1999) 401–409.
[23] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic. Acids. Res. 42 (W1) (2014). W320-W324.
[24] G.M. Rossolini, S. Schippa, M.L. Riccio, F. Berlutti, L.E. Macaskie, M.C. Thaller, Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology, Cell. Mol. Life Sci. 54 (8) (1998) 833–850. Augdoi: 10.1007/s000180050212. PMID: 9760992.
[25] F. Sievers, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539.
[26] H et.al Singh, Recognition of nucleoside monophosphate substrates by Haemophilus influenzae class C acid phosphatase, J. Mol. Biol. 404 (4) (2010) 639–649.
[27] H. Singh, et al., Expression, purification and crystallization of an atypical class C acid phosphatase from Mycoplasma bovis, Acta Crystallogr. Sect. F Struct. Biol. Cryst .Commun. 67 (Pt 10) (2011) 1296–1299.
[28] S. Tan, et al., The pST44 polycistronic expression system for producing protein complexes in Escherichia coli, Protein Expr. Purif. 40 (2005) 385–395.
[29] M.C. Thaller, et al., Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily, Protein Sci. 7 (7) (1998) 1647–1652.
[30] R. Wang, et al., Genetic and biochemical analysis of a class C non-specific acid phosphatase (NSAP) of Clostridium perfringens, Microbiology 56 (Pt 1) (2010) 167–173.
[31] B.L. Wanner, B.D. Chang, The phoBR operon in Escherichia coli K-12″, J. Bacteriol. 169 (12) (1987) 5569–5574.