Bioremediation: Discovery of a Novel Alkaline Phosphatase through X-Ray Crystallography

Subhash C. Bihani, Shree Kumar Apte, Jean-Luc Ferrer and Madhusoodan V. Hosur*

Solid State Physics Division and Molecular Biology Division

Abstract

Alkaline phosphatase (AP) is a bi-metalloenzyme with potential applications in biotechnology and bioremediation. The enzyme hydrolyses nonspecifically phosphate monoesters under alkaline conditions to yield inorganic phosphate. The AP isolated from the bacterium Sphingomonas sp. Strain BSAR-1 (SPAP) was shown to be particularly useful for the recovery of uranium from alkaline radioactive waste. We report here the first X-ray crystal structure of SPAP determined by using the Multi-wavelength Anomalous Diffraction method to solve the phase problem. The crystal structure shows that the enzyme active site is similar to those in other APs with two Zn^{2+} ions in the core, even when there is a large difference in the rest of the tertiary structure. Structural differences observed are: 1) threonine as the catalytic residue instead of serine, 2) absence of third metal ion binding pocket, 3) deletion of the arginine residue forming bidentate hydrogen bonding with the substrate phosphoryl group, and 4) recruitment into the active site of lysine171 and asparagine110 residues to bind the substrate phosphoryl group in a manner not observed before in any other AP. These structural differences suggest that SPAP represents a new class of APs and also provides important insights into evolutionary relationships between members of AP superfamily. Because of its direct contact with the substrate phosphoryl group, lysine171 is proposed to play a significant role in catalysis.

Introduction

(a) Bioremediation

Nuclear materials are being used for a variety of purposes ranging from electrical power generation, to nuclear weapons development to disease diagnosis and treatment. Various steps in the nuclear fuel cycle generate, as intermediates, large amounts of radioactive materials [1]. Environmental contamination due to release of radionuclides during storage is a major worry for the nuclear industry. According to U.S. Environmental Protection Agency (EPA), over 1,000 United States locations are contaminated with highly energetic radionuclides including 235U [2]. Currently available chemical methods for cleanup are costly and often lack the specificity required to properly decontaminate. Bioremediation of heavy metal contamination using radiation-resistant microbes provides an effective and more economical alternative. Microbial precipitation of uranium in the form of uranyl phosphate also allows recovery of precious metal from very low concentrations [3]. Recently, Kayzad et al. [4] have cloned into Escherichia coli, the alkaline phosphatase from Sphingomonas sp strain BSAR-1, labeled here as SPAP, and the recombinant E. coli strain over-expressing SPAP was found to be very efficient in bioprecipitation of uranium as uranyl phosphate. SPAP exhibits several unique features, such as constitutive expression, thermo-lability, extracellular release and high specific activity [5], which make SPAP an ideal reagent also for many biotechnological applications. Atomic-level structure in three dimensions, that can be determined using crystallography, would provide critical input for further improving the utility of SPAP in bioremediation and biotechnology [6].

(b) The Crystallographic method

The basic principle of the crystallographic method is the Fourier relationship between scattering density distribution ρ(xyz) and its diffraction pattern F(hkl). When the sample is in the form of a single crystal of the molecule under study, the Fourier transform is sampled at discrete points determined by the unit cell parameters. Then the Fourier relationship can be written as a series summation given below:

*Corresponding author
The oscillation angle and the total number of frames to be recorded depend on the crystal symmetry, unit cell parameters, and the redundancy required. As already mentioned above, in a diffraction experiment, one obtains only the magnitude of the structure factor. Not knowing the phases of reflections is described as the 'phase problem' in crystallography, and a variety of methods to obtain reflection phase have been developed over the last century. One of the more recently developed methods is the Multi-wavelength Anomalous Diffraction method (MAD) [7, 8]. This method exploits the physical fact that atomic X-ray form factors become complex and change rapidly with wavelength near the atomic absorption edges. The exact atomic absorption edge is sensitive to the environment within the crystal, and therefore, has to be derived by using in-situ X-ray absorption/fluorescence measurements. In a MAD experiment, the wave lengths used for data collection differ in energy by only a few electron volts, and therefore synchrotron beamlines with energy resolution of the order of 10000 are required. Further, since the anomalous differences are rather small, the reflection intensities have to be measured very accurately. The

\[ \rho(xyz) = \frac{1}{V} \sum_{\mathbf{hkl}} |F_{\mathbf{hkl}}| \exp(1\alpha_{\mathbf{hkl}}) \exp(-2\pi i (hx + ky + lz)) \]

where \( h, k, \) and \( l \) are integers that specify spatial direction of the Bragg reflection \((hkl)\), and \( R_{hkl} \) is the complex structure factor characterized both by a magnitude, \(|R_{hkl}|\), and a phase, \( \alpha_{hkl} \). The set of values of the function \( \rho(x,y,z) \) at different \((x,y,z)\) positions constitutes the electron density map, which represents the three dimensional structure of the molecule. All atoms in the molecule will lie inside the cage defined by positive contours in the electron density map. However, in a diffraction experiment, only the magnitudes of the structure factors are derived from the measured intensities, and separate strategies have to be used to obtain the phases before Fourier summation can be carried out. Although there are many different methods for actual recording of Bragg reflection intensities, the fastest one is the oscillation method, which uses a single-axis goniometer and a stationary two-dimensional X-ray detector positioned perpendicular to the monochromatic X-ray beam incident on the crystal, as shown in Fig. 1a. The crystal is oscillated about an axis perpendicular to the X-ray beam direction, and the diffracted rays emanating in different directions are intercepted on the detector. The total diffraction data is measured as a series of non-overlapping oscillation frames recorded about contiguous positions of the crystal, and all these frames together constitute one data set. A typical oscillation diffraction photograph is shown in fig. 1b. Each dark spot visible in Fig. 1b is a Bragg reflection, whose characteristic indices \((h,k,l)\) and intensity \(I_{hkl}\) are established during ‘processing’ of the oscillation frame.
phases derived using MAD method are generally superior compared to other methods, and lead to cleaner and easily interpretable electron density maps. The structure of SPAP is one of the first structures to be determined in India using the MAD method.

(c) Alkaline phosphatase

Alkaline phosphatases (APs) are enzymes that hydrolyze phosphate monoesters under alkaline conditions to yield an alcohol and inorganic phosphate. APs are metallo-enzymes, and crystal structures show that their active sites contain the bi-metallo-zinc core and a Mg$^{2+}$ ion, which are involved in a two stage catalytic mechanism [9, 10]. Multiple sequence alignment shows that there is very little similarity in the amino acid sequences of SPAP and other APs. Interestingly, SPAP has highest sequence similarity with nucleotide pyrophosphatases (NPP), but does not have pyrophosphatase activity. The three dimensional structure shows that SPAP is a unique alkaline phosphatase representing a new class.

Experimental

Overexpression, Purification and Crystallization of SPAP protein

Gene for SPAP protein was cloned and protein was over-expressed in *E. coli* as described [4]. Crystallization was performed at 298 K by sitting-drop vapor diffusion method in 96-well crystallization plates (Greiner, 3cup) using Cy-Bio HTPC robot [5, 6]. Both native as well as Seleno-MET (SM-SPAP) crystallized under several conditions. Crystallization condition for SM-SPAP was optimized manually by hanging-drop vapor diffusion method [6].

X-ray Diffraction Data collection & Structure Determination

X-ray diffraction data for both native and SM-SPAP crystals were collected at 100 K by using the oscillation method and the FIP beamline on the European Synchrotron Radiation Facility (ESRF). A fluorescence scan on SM-SPAP crystals enabled
identification of the three wavelengths for a MAD data collection. At each wavelength, 180 diffraction images, each for an oscillation angle of 1° and an exposure time of 60 seconds, were recorded. For native crystals 180 diffraction images at a single wavelength were recorded, again for an oscillation angle of 1° and an exposure time of 60 seconds. Experimental phase determination and initial molecular model building were carried out using the AutoSol and AutoBuild tools in the software suite PHENIX [11]. The fit of the atomic model to the experimental electron density was improved manually by using software package Coot [12]. The software phenix-refine was used to execute standard protocols of simulated annealing (SA) and TLS refinement [13]. The native structure was later solved by using the protein part of SM-SPAP structure as a search model in Molecular Replacement calculations [11]. The procedures used for refinement of the native structure were same as described above.

Results and Discussion

The tetragonal crystals of SPAP are shown in Figure 2. Crystals of SM-SPAP also have similar morphology, and both belong to the space group P4_12_2, and contain one SPAP monomer (559 residues) per asymmetric unit. High resolution diffraction data from both SPAP and SM-SPAP crystals were collected using the FIP beamline on ESRF at Grenoble, France. Crystal parameters and data collection statistics are given in Table 1. The crystal structure was solved by using the MAD method to determine reflection phases. The wavelengths for MAD data collection were determined by analysing the X-ray fluorescence spectrum (Fig. 3, Table 1). Fig. 4 is one of the four Harker sections of the anomalous difference Patterson map showing good quality of measured anomalous signal that led to Se-atom substructure. Refinement statistics for the anomalous scatterers are given in Table 1. Fig. 5 shows the experimentally obtained electron density cage contoured at 1.5σ level for a short segment of the protein chain. It is clear that all atoms of the atomic model are very well contained within the electron density cage, highlighting the accuracy of the derived structure. The accuracy of the structure is also evident from the low crystallographic R-factors (Table 2), which are a measure of the agreement between observed and model-derived intensities of Bragg reflections. The final contents of the asymmetric unit are: 526 protein residues (31–556) comprising of 4022 non-hydrogen atoms, two zinc ions of partial occupancy (0.75), 278 water molecules,
5 glycerol molecules, 1 Ca$^{2+}$ ion and one molecule of an organic phosphate in the active site. The protein model has good stereochemistry with 98% residues lying in the most favoured region of Ramachandran map.

**Molecular Conformation**

The overall shape of the polymeric protein chain is shown by a space filling diagram in which each atom is drawn as a sphere of radius proportional to the van der Waal radius of the atom (Fig. 6). Segments of the protein backbone fold into α-helices and β-strands, which are shown in fig. 6 as a cartoon overlaid on the space filling diagram. These secondary structure elements then associate to give the protein a specific three dimensional shape. The shape can be described as that of a shallow cup with a thick base. The protein has two domains, one large N-terminal domain forming the base and most of the wall region of the cup, and a smaller C-terminal domain (right extreme portion in Fig. 6) forming the handle. N-terminal domain consists of an eight-stranded mixed central beta sheet (shown in magenta) surrounded by α-helices (shown in cyan) while the smaller C-terminal domain primarily consists of a single two-stranded antiparallel beta hairpin covered by two α-helices on the surface. The physical size of the protein molecule can be estimated by noting that the alpha helix has a pitch of 5.4 Å. The active site (shown in brick red), located at the bottom of the wide and shallow opening of the cup is exposed to the environment.

**Active Site**

Active site of SPAP is geometrically identical to that in ECAP, and contains two metal ions (M1 and M2), which were identified as Zn$^{2+}$ through X-ray fluorescence measurements. Distance between the two metal ions is 4.1 Å. M1 coordinates directly to the hydroxyl group of Thr89, confirming Thr89 to be the catalytic residue in SPAP, as suggested by sequence alignment. The other amino acid residues coordinated by the zinc ions are identical to those in ECAP.

**Polar Lysine residue replacing the third metal**

The third metal ion, mostly Mg$^{2+}$, invoked in the catalytic mechanism is bound at a site shown to be conserved in crystal structures of ECAP-like APs. However, in SPAP, there is no such third metal ion coordination site because, structural superposition shows that non-polar amino acids Ala174 and

---

Fig. 7: Structural superposition of SPAP (magenta) and ECAP (blue) around the Mg$^{2+}$-binding site in ECAP. (A) Abolition of Mg$^{2+}$ ion binding site. Polar residues coordinating Mg$^{2+}$ ion in ECAP are missing in SPAP. (B) Introduction of Lysine in the active site of SPAP. Lys171 in SPAP interacts with the substrate in a way analogous to Mg$^{2+}$ ion in ECAP.
Gly295 now occupy positions of amino acids Thr155 and Glu322 that coordinate Mg\(^{2+}\) ion in ECAP (Fig. 7A). Interestingly, the structure reveals that a lysine residue, which has a positively charged N atom at the end of the side chain, is positioned exclusively in the active site of SPAP. This N atom is involved in chemical interactions similar to those of Mg\(^{2+}\) in ECAP-like enzymes (Fig. 7B). These observations suggest that Lys171 in SPAP is substituting for Mg\(^{2+}\) ion in ECAP.

**SPAP – A novel AP representing a New Class**

There are several features which make SPAP a novel AP. First, the key catalytic residue in SPAP is Thr89 instead of serine found in all other APs. Second, the conserved third metal ion binding pocket in the active site in ECAP-like APs, is absent in SPAP (Fig. 7A and 7B). Third, the active site arginine residue conserved in ECAP-like APs is absent in SPAP (Fig. 8A and 9). Fourth, the sequence identity between SPAP and ECAP-like APs ranges from 12% to 14.6%, which is very low compared to the range of 30% - 50% observed among ECAP-like APs (Fig. 9). Fifth, the combined presence in the SPAP active site of key amino acid residues Thr89, Asn110, Lys171 and Arg173 (Fig. 8B and 9), observed only separately in the active sites of members of the AP superfamily. Interestingly, the relative positions and interactions of these residues with the substrate phosphoryl group are maintained in the individual active sites (Fig. 8A).

**Implications toward Molecular Evolution**

Proteins with related functions and similar structure are combined into a larger group called as a superfamily. APs are part of a superfamily, which includes nucleotide pyro-phosphatases (NPP), cofactor independent phosphoglycerate mutases (IPGM), phosphonate monoester hydrolases (PMH) and aryl sulfatases (AS) [14]. Amino acid sequence similarity between different members of the superfamily is often too low to reveal relatedness. In such conditions, structural alignment method can provide better understanding of molecular evolution. Structural alignment, using the DALI server [15], shows that SPAP has highest similarity with NPP (Z-score 28.2) rather than with ECAP (Z-score 20.1) or other ECAP like APs. Structural superposition also shows that the substrate-interaction pattern observed in SPAP is different from that seen in ECAP and other APs. The two coplanar hydrogen bonds to the substrate phosphoryl group from the conserved active site arginine residue in all ECAP-like APs are absent in SPAP. Instead, there are two non-coplanar hydrogen bonds to the substrate phosphoryl group from two separate amino acids in the active site.
Fig. 9: Sequence alignment derived from structural superposition of SPAP, ECAP and NPP. The amino acid residues are described in single-letter code. Absolutely conserved residues are shown in red background. Following four novel features of SPAP are highlighted by arrows: 1) presence of Thr89 in place of catalytic serine - blue arrow, 2) presence of unique Lys171 and Arg173 - green arrows, 3) presence of Asn110 - blue filled arrow and 4) deletion, in SPAP, of conserved arginine – yellow arrow.

Fig. 10: Phylogenetic tree for alkaline phosphatases.
Engineering SPAP for higher activity

The alkaline phosphatase enzyme is widely used in the fields of molecular biology and biotechnology. In fact there are three commercial preparations of APs from different sources, and none of these has totally satisfactory properties of thermolability and...

Fig. 11: Hydrogen bonding interactions, near the active site, which may be modulated to increase flexibility of ARG-173 and ASN-110: (A) positioning of ARG-173 and (B) positioning of ASN-110.

Table 1: Crystal data, intensity data and anomalous phasing statistics for SM-SPAP.

<table>
<thead>
<tr>
<th></th>
<th>MAD dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>peak</td>
</tr>
<tr>
<td>Unit Cell (Å)</td>
<td>a = 87.20</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.980159</td>
</tr>
<tr>
<td>Upper Resolution limit (Å)</td>
<td>1.85 (1.9-1.85)*</td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁2₁2</td>
</tr>
<tr>
<td>Number of measured reflections</td>
<td>771391 (56396)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>7.44 (7.08)</td>
</tr>
<tr>
<td>R-sym (%)</td>
<td>7.3 (49.8)</td>
</tr>
<tr>
<td>Average I/σ(I)</td>
<td>20.2 (4.13)</td>
</tr>
</tbody>
</table>

**Anomalous Phasing**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Refined f' (e-)</td>
<td>-7.1</td>
<td>-9.5</td>
<td>-2.1</td>
</tr>
<tr>
<td>Refined f'' (e-)</td>
<td>10.6</td>
<td>6.6</td>
<td>6.2</td>
</tr>
<tr>
<td>Mean figure of merit</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The numbers between parentheses indicate the value in the outer resolution shell.
high activity. Therefore, there is a demand for an AP which has high activity, can be purified in large amounts, and can be easily and completely inactivated by heating to milder temperatures. SPAP has favourable properties in these respects. Since greater flexibility in and around the active site is correlated with higher activity, we have identified Glu88, Tyr125 and Asp172 as residues that can be mutated to impart flexibility to the active site residues Arg173 and Asn110, which are binding the phosphoryl moiety (Fig. 11A and 11B).

Acknowledgments

This work was carried out as part of an MOU between BARC, India and CEA, France, for cooperation in the field of life sciences. MVH is thankful to Drs. R. Chidambaram and K. K. Kannan for scientific discussions. We are thankful to the National Facility for Macromolecular Crystallography, SSPD, BARC for providing access to the equipment.

References