

**STRUCTURE FUNCTION RELATIONSHIP OF PHOSPHATASES FROM  
*Vibrio cholerae*O395**

By

**SEEMA NATH**

**Enrolment no.: LIFE05200804003**

**Saha Institute of Nuclear Physics, Kolkata**

*A thesis submitted to the*

*Board of Studies in Life Sciences*

*In partial fulfillment of requirements*

*for the Degree of*

**DOCTOR OF PHILOSOPHY**

*of*

**HOMI BHABHA NATIONAL INSTITUTE**



**December, 2014**

# HOMI BHABHA NATIONAL INSTITUTE<sup>1</sup>

## RECOMMENDATIONS OF THE VIVA VOCE COMMITTEE

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Ms. Seema Nath** entitled "**Structure-Function Relationship of Phosphatases from *Vibrio cholerae*0395**" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman – Prof. Dipak Dasgupta

Date:

*D. Dasgupta*

28.05.15

Guide / Convener – Prof. Udayaditya Sen

Date:

*Udayaditya Sen*

28/5/15

Co-guide - <Name> (if any)

Date:

Examiner – Prof. Dinakar M Salunke

Date:

*D. Salunke*

May 28, 2015

Member 1- Prof. Sampa Biswas

Date:

*Sampa Biswas*

28/5/2015

Member 2- Prof. Abhijit Chakrabarti

Date:

*Abhijit Chakrabarti*

28/5/2015

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: Kolkata

Place: 28.05.2015

<Signature>

Co-guide (if applicable)

*Udayaditya Sen*

<Signature>

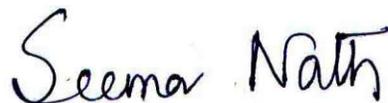
Prof. Udayaditya Sen  
Crystallography & Molecular Biology Division  
SAHA INSTITUTE OF NUCLEAR PHYSICS  
1/AF, Bidhannagar  
Kolkata - 700064

<sup>1</sup> This page is to be included only for final submission after successful completion of viva voce.

## STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

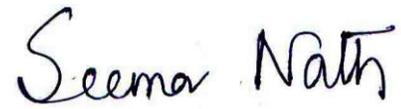
Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

A handwritten signature in black ink that reads "Seema Nath". The script is cursive and fluid, with the first letters of "Seema" and "Nath" being capitalized and prominent.

Seema Nath

## DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

A handwritten signature in black ink that reads "Seema Nath". The letters are cursive and fluid, with a distinct loop in the 'S' and a sharp peak in the 'N'.

Seema Nath

## List of Publications from the thesis

### Journal

1. “Atomic resolution crystal structure of VcLMWPTP-1 from *Vibrio cholerae* O395: insights into a novel mode of dimerization in the low molecular weight protein tyrosine phosphatase family”, **Nath S**, Banerjee R, Sen U. *Biochemical and Biophysical Research Communications*. 2014; 450(1):390-395.
2. “A novel 8-nm protein cage formed by *Vibrio cholerae* acylphosphatase”, **Nath S**, Banerjee R, Sen U. *Journal of Molecular Biology*. 2014; 426(1):36-38.
3. “Cloning, purification, crystallization and preliminary X-ray analysis of two low-molecular-weight protein tyrosine phosphatases from *Vibrio cholerae*”. **Nath S**, Banerjee R, Khamrui S, Sen U. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*. 2012; 68(10):1204-1208.

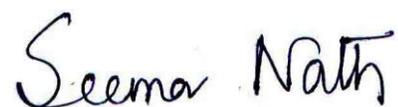
### Conferences

1. Poster presentation on “‘KNOT’ and ‘CAGE’ – two new dimensions in protein structures” in the 12th Conference of Asian Crystallographic Association (AsCA13) held at Hong Kong University of Science and Technology, Hong Kong, China.
2. Oral presentation on “Structure of a novel protein cage from *Vibrio cholerae* O395” in 42nd National Seminar on Crystallography and International Workshop on Application of X-ray Diffraction for Drug Discovery (NSC42) held at Jawaharlal Nehru University, New Delhi, India.

3. Oral presentation on “‘KNOT’ and ‘CAGE’ – two new dimensions in protein structures” in Frontiers in Modern Biology (FIMB) 2013 held at Indian Institute of Science Education and Research (IISER), Kolkata, India.
4. Workshop on “Cryoelectron Microscopy in LifeSciences” in International Conference on Electron Microscopy and XXXIV Annual Meeting of the Electron Microscope Society of India (EMSI) held at National Institute of Cholera and Enteric Diseases (NICED), Kolkata, India.
5. Poster presentation on “High resolution (1.6Å) crystal structure of Low Molecular Weight Phosphotyrosine Phosphatase from *Vibrio cholerae* 0395” in 5th International Symposium on Recent Trends in Macromolecular Structure and Function (ISRTMSF-2012) held at University of Madras, Chennai, India.

#### **Other Publications**

1. “The first structure of polarity suppression protein, Psu from enterobacteria phage P4, reveals a novel fold and a knotted dimer”. Banerjee R, **Nath S**, Ranjan A, Khamrui S, Pani B, Sen R, Sen U. *Journal of Biological Chemistry*. 2012; 287(53):44667-44675.
2. “Cloning, expression, purification, crystallization and preliminary X-ray analysis of a fructokinase from *Vibrio cholerae* O395”. Paul R, **Nath S**, Sen U. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* 2012; 68(12):1564-1567.



Seema Nath

*Dedicated to my parents, family members*

*and*

*all my well-wishers*

## **Acknowledgements**

I would like to take this opportunity to express my gratitude to the help and cooperation extended by my friends and well-wishers during the course of my work. First and foremost, I would like to convey heartfelt regards to my mentor Prof. Udayaditya Sen for his untiring cooperation. I have immensely benefited from his flair of innovative thinking, positive attitude and infinite optimism throughout my tenure in this institute. His worthy suggestions, symbolic optimism and unremitting inspirations have compassed my journey to the right direction. Whenever problems engulfed me, his encouragement and guidance have always helped me to find the proper way out. I am deeply obliged to Dr. Jhimli DasGupta for her kind support during the course of this study. I am also thankful to Prof. Jiban Kanti Dattagupta, Prof. Sampa Biswas and Prof. Abhijit Chakrabarti for their assistance. Financial assistance from the Department of Atomic Energy (DAE) and other laboratory facilities from Saha Institute of Nuclear Physics are acknowledged. I do thankfully acknowledge Ramanuj without whose active support and cooperation this thesis would not have been completed. I would like to thank Abhijit da and Bikram da for their excellent technical support. I also thank Bipin da for looking after daily chores in the lab. My sincere thanks are due to all the staff members of the Crystallography and Molecular Biology Division for their generous cooperation. I would specially like to mention my friends and lab mates for encouraging me through the highs and lows of my stay in this lab from the very beginning. Their friendship like attitude shall always be a treasure to me. Lastly, I want to heartily thank my family members whose constant support led my path towards this degree.

Seema Nath

## Contents

	Page No.
Synopsis.....	I-VIII
List of Figures	
List of Tables	
Chapter 1- Introduction	
1.A1. Phosphorylation event in bacteria.....	3
1.A2. Hydrolysis of carboxyl phosphate bond – A tale about another Phosphatase.....	4
1.A3. Structural aspects of Acylphosphatase in light of its catalytic action.....	5
1.A4. Structural and functional aspects of <i>Vibrio cholera</i> 0395- Acylphosphatase.....	5
1.A5. Role of Tyrosine phosphorylation in bacteria.....	6
1.A6. Bacterial tyrosine kinase (BYK) and Bacterial tyrosine phosphatases (BYP).....	6
1.A7. Low Molecular Weight PhosphoTyrosine Phosphatase.....	8
1.A8. Structural and functional aspects of BYP with respect to <i>Vibrio cholera</i> O395- Low molecular weight phosphotyrosine phosphatases ( <i>Vc</i> -LMWPTP).....	11
1.B1. Protein X-ray Crystallography to study the structure-function relationship of the aforementioned phosphatases.....	11
1.B2. Steps involved in protein crystallography study.....	12
1.B3. Protein preparation.....	12
1.B4. Crystallization technique and data collection strategy.....	13
1.B5. Data processing step.....	14
1.B6. ‘Phase’ problem in crystallography.....	14
1.B7. Refinement and Phase improvement.....	16
1.B8. Validation and deposition of the structure.....	17
Chapter 2- Materials & Methods	
2. Cloning Over-Expression, Purification, Crystallization, Data Collection, Processing and Refinement.....	18
2.1. Cloning and Over-expression of the proteins as 6× His-tagged construct.....	18

2.1A. Cloning and over-expression of the wild type Vc-AcP and its mutant.....	19
2.1B. Cloning and over-expression of the wild type VcLMWPTP-1 and its mutant.....	19
2.1C. Cloning and over-expression of the wild type VcLMWPTP-2 and its mutants.....	20
2.1D. Polymerase Chain Reaction (PCR).....	21
2.1E. Site directed mutagenesis (SDM).....	22
2.1F. Digestion and Ligation.....	23
2.1G. Transformation and Analysis.....	23
2.1H. Protein over-expression.....	24
2.2. Purification of recombinant proteins in <i>E. coli</i> using affinity chromatography and size exclusion chromatography.....	25
2.2A. Purification of 6×His-tagged proteins using affinity chromatography.....	25
2.2B. Removal of His-tag with thrombin.....	27
2.2C. Fast Protein Liquid Chromatography (FPLC) using Sephacryl S-100.....	27
2.3. Characterization of protein.....	29
2.3A. Dynamic Light Scattering (DLS).....	29
2.3B. Characterization of protein by Transmission Electron Microscopy (TEM).....	29
2.3C. Chemical Cross Linking.....	30
2.4. Crystallization of target proteins.....	30
2.5. Data Collection and Processing.....	33
2.6. Structure determination of the target proteins.....	34
2.6A. Molecular Replacement (MR) method.....	34
2.6B. Refinement and Model Building step.....	36
2.7. Structural analysis.....	37
2.8. Enzyme assay.....	38
2.8A. Enzyme kinetics of Vc-AcP and its mutant.....	38
2.8B. Enzyme kinetics of VcLMWPTPs and their mutants.....	38

## Chapter 3- Results & Discussion

<b>3.1. Acylphosphatase (Vc-AcP)</b> .....	<b>39</b>
<b>3.1A. Cloning and Overexpression of Vc-AcP and its mutant</b> .....	<b>39</b>
<b>3.1B. Ni-NTA affinity chromatographic purification of the target proteins</b> .....	<b>39</b>
<b>3.1C. Thrombin cleavage standardization and Fast Protein Liquid Chromatography (FPLC)</b> .....	<b>40</b>
<b>3.1D. Dynamic Light Scattering (DLS) profile of Vc-AcP and Vc-AcP-C20R</b> .....	<b>40</b>
<b>3.1E. Crystallization of Vc-AcP and Vc-AcP-C20R</b> .....	<b>42</b>
<b>3.1F. Data collection and processing of Vc-AcP and Vc-AcP-C20R</b> .....	<b>43</b>
<b>3.1G. Structure solution of Vc-AcP-C20R</b> .....	<b>45</b>
<b>3.1H. Structure solution of Vc-AcP</b> .....	<b>46</b>
<b>3.1I. Structural analysis of Vc-AcP-C20R: Overall structure of Vc-AcP-C20R monomer</b> .....	<b>48</b>
<b>3.1J. Crystal structure of Vc-AcP-C20R reveals a trimeric assembly involving a set of unique residues</b> .....	<b>48</b>
<b>3.1K. Active site structure of Vc-AcP-C20R</b> .....	<b>50</b>
<b>3.1L. Nature of electrostatic surface of Vc-AcP-C20R</b> .....	<b>51</b>
<b>3.1M. Structural analysis of Vc-AcP: Overall crystal structure of Vc-AcP</b> .....	<b>52</b>
<b>3.1N. Arrangement and symmetry of the dodecameric protein cage</b> .....	<b>54</b>
<b>3.1O. Nature of electrostatic surface of Vc-AcP</b> .....	<b>57</b>
<b>3.1P. Role of unique residues in the formation of ordered multimeric structures: An outcome of Multiple Sequence Alignment result for Vc-AcP</b> .....	<b>58</b>
<b>3.1Q. Kinetic Studies of Vc-AcP-C20R</b> .....	<b>59</b>
<b>3.1R. Transmission Electron Microscopy (TEM)</b> .....	<b>60</b>
<b>3.2. <u>L</u>ow <u>M</u>olecular <u>W</u>eight <u>P</u>rotein <u>T</u>yrosine <u>P</u>hosphatase-1 (<i>VcLMWPTP-1</i>)</b> .....	<b>61</b>
<b>3.2A. Cloning and Overexpression of <i>VcLMWPTP-1</i> and its mutant</b> .....	<b>61</b>
<b>3.2B. Ni-NTA affinity chromatographic purification of the target proteins</b> .....	<b>61</b>
<b>3.2C. Thrombin cleavage standardization and Fast Protein Liquid Chromatography</b> .....	<b>62</b>
<b>3.2D. Determination of oligomers using FPLC</b> .....	<b>63</b>
<b>3.2E. Enzyme kinetics</b> .....	<b>65</b>
<b>3.2F. Crystallization of <i>VcLMWPTP-1</i></b> .....	<b>65</b>
<b>3.2G. Data collection and Processing of <i>VcLMWPTP-1</i></b> .....	<b>66</b>
<b>3.2H. Structure solution of <i>VcLMWPTP-1</i></b> .....	<b>67</b>
<b>3.2I. Structural analysis of <i>VcLMWPTP-1</i>: Overall structure of <i>VcLMWPTP-1</i> monomer</b> .....	<b>69</b>
<b>3.2J. Active site architecture of <i>VcLMWPTP-1</i></b> .....	<b>70</b>

<b>3.2K. Detection of different conserved ‘Motif’s in the LMWPTP family - A comparison based on the Multiple Sequence Alignment result.....</b>	<b>72</b>
<b>3.2L. Dimerization of <i>Vc</i>LMWPTP-1.....</b>	<b>75</b>
<b>3.2M. Insight into differential substrate specificity: Diverged surface property of the active site.....</b>	<b>77</b>
<b>3.2N. Glutaraldehyde crosslinking to capture the dimeric state in solution.....</b>	<b>79</b>
<b>3.3. Low Molecular Weight Protein Tyrosine Phosphatase-2 (<i>Vc</i>LMWPTP-2).....</b>	<b>80</b>
<b>3.3A. Cloning and Overexpression of <i>Vc</i>LMWPTP-2 and its mutants.....</b>	<b>80</b>
<b>3.3B. Ni-NTA affinity chromatographic purification of the target proteins.....</b>	<b>80</b>
<b>3.3C. Thrombin cleavage standardization and Fast Protein Liquid Chromatography.....</b>	<b>81</b>
<b>3.3D. Enzyme kinetics.....</b>	<b>81</b>
<b>3.3E. Crystallization of <i>Vc</i>LMWPTP-2.....</b>	<b>82</b>
<b>3.3F. Data collection and Processing of <i>Vc</i>LMWPTP-2.....</b>	<b>82</b>
<b>Chapter 4- Discussion and Summary</b>	
<b>4. <i>Vc</i>-AcP and <i>Vc</i>-AcP-C20R.....</b>	<b>85</b>
<b>4.1A. Unique set of residues involved in the tetrahedral cage formation of <i>Vc</i>-AcP and triangular <i>Vc</i>-AcP-C20R.....</b>	<b>85</b>
<b>4.1B. Single mutation C20R on <i>Vc</i>-AcP affect the assembly formation.....</b>	<b>86</b>
<b>4.1C. Designer cage- new horizon to carrier molecule.....</b>	<b>89</b>
<b>4.2. <i>Vc</i>LMWPTP-1.....</b>	<b>91</b>
<b>4.2A. <i>Vc</i>LMWPTP-1 forms enzymatically active dimer in solution.....</b>	<b>91</b>
<b>4.2B. Molecular details of the dimeric crystal structure of <i>Vc</i>LMWPTP-1 reveal the altered mode of dimerization.....</b>	<b>91</b>
<b>4.2C. Electrostatics surface indicates similarity of <i>Vc</i>LMWPTP-1 with eukaryotic LMWPTPs.....</b>	<b>92</b>
<b>4.3. <i>Vc</i>LMWPTP-2.....</b>	<b>93</b>

## References

## Publications

## List of Figures

**Figure 1: Proposed catalytic mechanism for Acylphosphatase**

**Figure 2: The catalytic mechanism followed by LMWPTPs**

**Figure 3: Genomic organization of BY-kinases and their cognate phosphotyrosine-phosphatase genes in proteobacteria**

**Figure 4: The Phase Diagram in Protein Crystallization**

**Figure 5: Comparison of 'hanging drop' and 'sitting drop' Vapor Diffusion method**

**Figure 6a-6b. Cloning and overexpression check of Vc-AcP**

**Figure 7. 15% SDS-PAGE profile showing the Ni-NTA purification profile of Vc-AcP**

**Figure 8a-8d. Dynamic light scattering profiles of Vc-AcP and Vc-AcP-C20R**

**Figure 9a-9d. Dynamic light scattering profiles of Vc-AcP with different oxyanions**

**Figure 10a-10b. Crystals of Vc-AcP & Vc-AcP-C20R**

**Figure 11a-11b. Diffraction pattern of Vc-AcP & Vc-AcP-C20R crystals upon exposure to X-rays**

**Figure 12a-12b. Structure of monomeric AcP and Vc-AcP-C20R**

**Figure 13a-13b. Molecular details of Vc-AcP-C20R and its active site**

**Figure 14. Molecular details of the active site Vc-AcP-C20R**

**Figure 15a-15b. Electrostatic surface view of Vc-AcP-C20R**

**Figure 16a-16b. Dodecameric assembly of Vc-AcP**

**Figure 17a-17c. Different faces of Vc-AcP**

**Figure 18a-18c. Molecular details of the ST-face and sulfate assembly**

**Figure 19. Symmetry elements in Vc-AcP protein cage with tetrahedral symmetry.**

**Figure 20a-20c. Electrostatic surface view of different faces of Vc-AcP**

**Figure 21a-21b. Electrostatic view of exterior and interior of Vc-AcP cage**

**Figure 22. Multiple sequence alignment of Vc-AcP**

**Figure 23. Reaction kinetics of acetylphosphate by Vc-AcP-C20R and Vc-AcP**

**Figure 24. TEM image**

**Figure 25a-25b. Cloning and overexpression check of VcLMWPTP-1**

**Figure 26. 15 % SDS-PAGE profile showing the Ni-NTA purification profile of *Vc*LMWPTP-1**

**Figure 27. Thrombin cleavage standardization of *Vc*LMWPTP-1**

**Figure 28. Existence of oligomeric *Vc*LMWPTP-1 in solution by FPLC**

**Figure 29. FPLC analysis of *Vc*LMWPTP-1 with different protein and salt concentrations**

**Figure 30a-30b. Enzyme kinetics of *Vc*LMWPTP-1 and C8S**

**Figure 31a-31b. Crystals and diffraction pattern of *Vc*LMWPTP-1**

**Figure 32a-32b. Overall structure and ligand (MOPS) binding at the active site of *Vc*LMWPTP-1**

**Figure 33. Stereo representation of the active site of *Vc*LMWPTP-1**

**Figure 34a-34b. Zoomed view of the active site of *Vc*LMWPTP-1**

**Figure 35. Similarity of binding of MOPS with that of phosphotyrosine binding**

**Figure 36. Structure based multiple sequence alignment of *Vc*LMWPTP-1 and other LMWPTPs from different organisms**

**Figure 37a-37d. Comparison of dimeric interface of *Vc*LMWPTP-1 with that of bovine LMWPTP**

**Figure 38. Comparison of the electrostatic surface around the active site cleft of LMWPTPs from different organisms**

**Figure 39a-39b. Glutaraldehyde crosslinking of *Vc*LMWPTP-1.**

**Figure 40a-40b. Cloning and overexpression check of *Vc*LMWPTP-2**

**Figure 41. 15% SDS-PAGE profile showing the Ni-NTA purification profile of *Vc*LMWPTP-2**

**Figure 42. Enzyme kinetics of *Vc*LMWPTP-2 and its mutants**

**Figure 43a-43b. Crystals and diffraction pattern of *Vc*LMWPTP-2**

**Figure 44. Superposition of 12 monomers of *Vc*-AcP and monomer of *Vc*-AcP-C20R**

**Figure 45. Superposition on *Vc*-AcP-C20R on *Vc*-AcP cage**

**Figure 46. *Vc*-AcP-C20R superposed on *Vc*-AcP showing possible clashes which prevent the mutant cage formation**

**Figure 47. Construction of disulfide-bridged cage.**

**Figure 48. Schematic representation of programmable designer protein cage**

## List of Tables

- Table 1. Functions of several BYPs from different organisms**
- Table 2. Function of paired BYP and BYK from several gram-positive and gram-negative bacteria**
- Table 3: Primer Sequences of wild type proteins**
- Table 4: Primer Sequences of mutant proteins for the 1st step PCR of SDM**
- Table 5: Composition of the PCR reaction mixture**
- Table 6: Calculation of the two-step-PCR reaction mixture**
- Table 7: Composition of the restriction enzyme double digestion and ligation**
- Table 8: Buffer compositions for the purification of recombinant proteins**
- Table 9: Crystallization conditions of recombinant proteins**
- Table 10: Data collection statistics of all crystals**
- Table 11: Data collection and processing statistics of Vc-AcP-C20R and Vc-AcP**
- Table 12: Improvement of Phaser statistics of Vc-AcP-C20R and Vc-AcP**
- Table 13: Data collection, phasing and refinement statistics of Vc-AcP and Vc-AcP-C20R**
- Table 14: PISA server data for Vc-AcP-C20R and Vc-AcP**
- Table 15: Data Collection and processing statistics of VcLMWPTP-1**
- Table 16: Improvement of Phaser statistics of VcLMWPTP-1**
- Table 17: Statistics for data collection and refinement of crystal structure of VcLMWPTP-1**
- Table 18: Motif table**
- Table 19: Results of VcLMWPTP-1 (one dimer) from the PISA server**
- Table 20: Data Collection and processing statistics of VcLMWPTP-2**

# **Synopsis**

The phenomenon of reversible phosphorylation in proteins plays an important role in all living organisms. The enzymes responsible for this action are of two types, kinase which transfers phosphoryl group to a specific residue and phosphatase which removes that group. Innumerable kinases-phosphatases are known to be present in eukaryotic systems, while only a handful of them are known and investigated in case of prokaryotes (1). Substrates with hydroxyl group in Ser/Thr belong to protein phosphatase superfamily which is subdivided into PPP and PPM family (2) and substrates with phenolic group in Tyr belongs to protein tyrosine phosphatase (PTP) family which is again subdivided into classical PTP, low molecular weight PTP (LMWPTP) and dual specificity PTP (DSP) (3). Substrate containing amine/imine residues are hydrolyzed by Protein Histidine Phosphatase (PHP) which acts as regulatory checkpoints in the multistep His-Asp phosphorelay system (4). Another group contains substrates with phosphorylation at acidic amino acids and is unstable at physiological condition and are dephosphorylated by Acylphosphatase (AcP) (5).

#### Acylphosphatase (AcP):

It is one of the smallest (~10kDa) enzymes ubiquitously found in all organisms. It acts on the substrates having carbamoyl-phosphate bond and are intermediate products of various metabolic and biosynthetic pathways. The active site contains an arginine (arg) which captures and binds to the substrate and an asparagine (asn) which polarizes a water molecule to cleave the substrate (5).

In case of *Vibrio cholerae*0395, the arg is replaced by a naturally occurring cys in the wild type acylphosphatase. So it becomes nearly catalytically inactive and it may have an influence on its structure and subsequent functions.

## Low Molecular Weight Protein Tyrosine Phosphatase (LMWPTP):

LMWPTP is a class of large PTP superfamily sharing the common structural motif “CX<sub>5</sub>GN(S/T)” and follows the same catalytic pathway. The difference with other members of the PTP superfamily is that it has comparatively smaller structure and apart from the catalytic site motif the overall structure is completely different. It contains two loops- P-loop containing a cysteine initiates the catalysis by a nucleophilic attack on the substrate forms a cysteinyl-phosphate intermediate and an aspartic acid from DPY-loop hydrolyzes the intermediate releasing the product (6). Apart from their involvement in various physiological processes, they are also shown to be involved in pathogenicity in some organism (7).

*Vibrio cholerae*0395 (*Vc*) possesses two LMWPTPs that have been designated as *Vc*LMWPTP-1 (155 aminoacids) and *Vc*LMWPTP-2 (166aminoacids). They share only ~37% sequence identity between themselves while the closest structural homologues of *Vc*LMWPTP-1 and *Vc*LMWPTP-2 present in PDB bear a maximum identity of ~43% (3IDO – *Entamoeba histolytica*) and ~37% (2CWD – *Thermus thermophilus*) respectively. Although they have the same functional motifs, low sequence identity among them and also with other LMWPTPs suggest that three dimensional structures of these two PTPs may provide distinct features adding to the existing knowledge.

**Purpose of the present work** : Broadly the objective of the present work is to investigate the structure(s) and function(s) of the phosphatases (LMWPTP and AcP) from the gram-negative bacteria *Vibrio cholerae*0395.

## **Chapter 2: Materials & Methods**

### **Cloning, expression and large culture purification of wild type (WT) proteins:**

The respective DNA fragments of *Vc*-genomic DNA encoding the protein-of-interest (*VcAcP*, *VcLMWPTP-1*, *VcLMWPTP-b* and respective mutants) were cloned in the pET-28a(+) vector and overexpressed as N-terminal 6X-his-tagged recombinant proteins. The respective His-tagged proteins were purified from the cell lysate by Ni-NTA affinity column with increasing concentration of imidazole, were cleaved by thrombin and were further purified by size-exclusion chromatography, using Sephacryl S-100 column (78 × 1.4 cm) at room temperature. The homogeneity of the purified proteins was checked by 15% SDS-PAGE and was taken for several rounds of crystallization trials by hanging drop vapor diffusion method at 277K and 293K.

### **Dynamic Light Scattering:**

Homogeneity of the protein samples were checked prior to crystallization. Samples with multimodal population were further screened with different additives.

### **Protein crystallization and data collection:**

Crystallization trials were performed in 'hanging drop vapour diffusion' method where equal volume of protein and precipitant solutions was mixed and were incubated in Hampton crystal trays at 277K and 293K. The crystals were fished out from the crystallization drops using a 10-20 μm nylon loop, briefly soaked in a suitable cryoprotectant solution and flash-frozen in a stream of nitrogen at 100 K. X-ray data were collected in-house at 100K using a MAR Research image-plate detector of diameter 345 mm and Cu K alpha radiation (1.54 Å) generated by a Bruker-Nonius FR591 rotating-anode generator equipped with Osmic MaxFlux confocal optics and running at 50 kV and 90 mA.

### **Structure solution and analysis:**

The data was integrated and scaled. For phasing, the coordinates of models having higher identity with respective proteins were used for molecular replacement(MR) CCP4 (8). Model building was done with Coot (9) and refinement was carried out with Phenix refine (10).

### **Kinetic assay:**

Kinetic parameters were calculated using acetyl phosphate for *VcAcP* and *VcAcP-C20R* and *p*-nitrophenylphosphate for *VcLMWPTP-1*.

### **Fast protein liquid chromatography:**

Only *VcLMWPTP-1* at different concentration was run in FPLC with buffer containing increasing ionic strength at room temperature.

### **Transmission Electron Microscopy:**

TEM was performed only with AcP incubated with 0.8M ammonium sulfate, applied to a carbon coated TEM grid and stained with 0.5% uranyl acetate.

## **Chapter 3: Results**

### **Oligomeric state in solution:**

1. *VcAcP* showed three different oligomeric states in solution. The hydrodynamic radius ( $R_H$ ) of the three oligomers with expected molecular weight (MW) varies from ( $2.5 \pm 0.3$ nm;  $32 \pm 8$ kDa) to ( $3.7 \pm 0.3$ nm;  $68 \pm 12$ kDa) to ( $4.6 \pm 0.3$ nm;  $119 \pm 13$ kDa). Addition of different oxy-anions at an optimum concentration in *VcAcP* removed heterogeneity. *VcAcP-C20R* mutant showed to exist as a trimer in solution.

2. *VcLMWPTP-1* was shown to exist as a dimer in solution.

### **Enzymatic properties:**

1.  $K_m$  values obtained for *VcAcP* and its C20R mutant were  $8.53 \pm 0.7$  mM and  $2.36 \pm 0.6$  mM respectively.
2. The determined  $K_m$  value for *VcLMWPTP-1* was  $2.07 \pm 0.2$  mM (pH 4.8) and  $2.03 \pm 0.4$  mM (pH 7.6) respectively performed at 25°C using *p*-nitrophenyl phosphate as a substrate. The cys-mutant (C8S) resulted in complete loss of enzymatic activity.

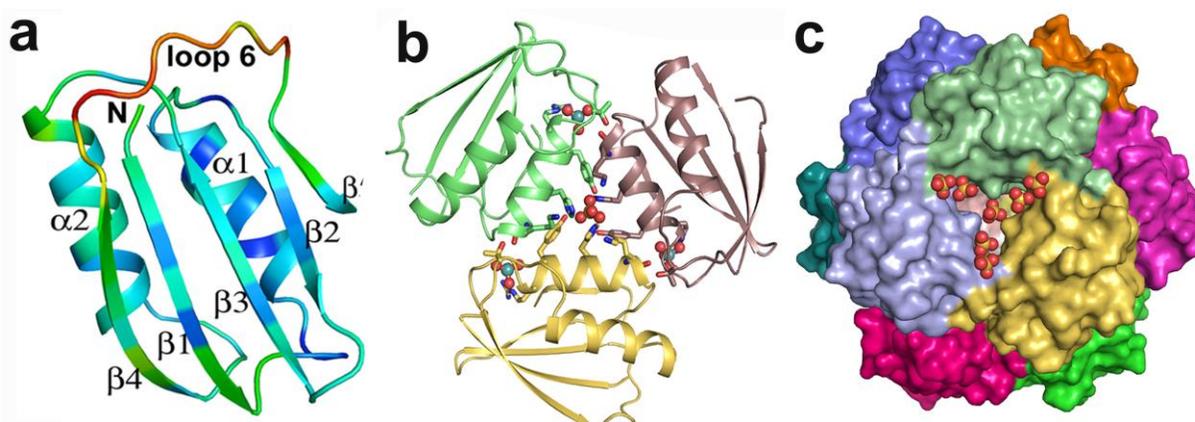
### **Crystallization & data collection:**

The diffraction patterns of the crystals were collected in-house and were solved by MR.

### **Structural details:**

1. The asymmetric unit of *VcAcP-C20R*, *VcAcP* and *VcLMWPTP-1*, and contained three, twelve and four monomers respectively. Coordinates and structure factor files have been deposited with the accession code 4HI1, 4HI2 and 4LRQ respectively.
2. The monomeric structures of *VcAcP* and its mutant contained ferredoxin like fold (Fig.1a) and have similar structures like other AcPs. These two structures differ from each other at the biological assembly. The wild type (WT) forms a dodecameric cage (Fig.1b) and the mutant remained in a trimeric assembly (Fig.1c) which is supposed to be the building block of the WT.
3. *VcAcP-C20R* contained a tightly bound molybdate ion at its active site. *VcAcP* had a total of 43 sulfates and they stabilized the cage-like structure.
4. The hollow dodecameric cage-like structure had an outer and inner diameter of ~8nm and ~4.5nm respectively. TEM images showed circular dots of average diameter 7-9nm.

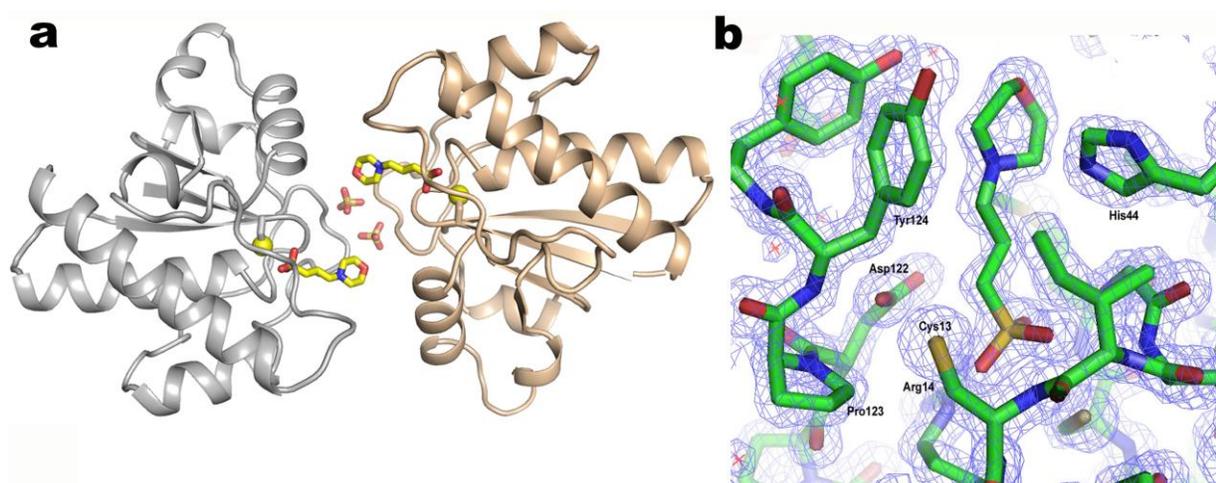
**Fig 1.**



**Fig.1 a)** Monomeric *VcAcP* and its mutant with ferredoxin fold of two  $\alpha$ -helices and a four stranded antiparallel  $\beta$ -sheet. **b)** Trimeric assembly of its mutant, C20R and **c)** Dodecameric assembly of *VcAcP* mediated by sulfate ions (shown in sphere) at their respective asymmetric unit.

5. *VcLMWPTP-1* existed as a dimer with an extended surface area excluding the active site and possessed 3-(N-morpholino)propanesulfonic acid (MOPS), a phosphotyrosine mimicking ligand at its active site (Fig.2a-2b).

**Fig. 2**



**Fig.2 a)** Dimeric *VcLMWPTP-1* having MOPS (shown in stick) at its active site, **b)** Zoomed view of the active site at the  $2f_o-f_c$  electron density map with a clear density of MOPS surrounded by aromatic residues.

## Chapter 4: Discussion & Summary

*Vibrio cholerae* 0395 acylphosphatase (*VcAcP*) was almost enzymatically inactive and resulted in the ordered oligomerisation to form a symmetric cage stabilized by sulfates which was confirmed through independent lines of evidence. Although a point mutation at the active site (*VcAcP*-C20R) converted the enzyme to its active form, it resulted in the disruption of the cage. Although oligomerization of AcP in solution had been reported earlier, *VcAcP* oligomerization differentiated itself from the other oligomeric AcPs by forming an ordered symmetric cage. Despite having similar fold, only *VcAcP* with cysteine at the active site, unique set of aromatic and basic residues in strategic location (at  $\alpha 1$ ) alongwith a flexible loop-6 formed such protein cage.

The monomeric structure of *VcLMWPTP*-1 having a four-stranded parallel  $\beta$ -sheet sandwiched between two sets of  $\alpha$ -helices, was similar to other LMWPTPs solved from other organism. It contained the prerequisite cysteine and aspartic acid residue at its catalytically important P-loop and DPY-loop respectively. The active site was captured by a MOPS molecule mimicking a phosphorylated tyrosine residue. It existed as a dimer in solution and was catalytically active. The crystal structure indicated exclusion of active site pocket from the extended dimeric interface unlike other reported structures. This observation accounted for the catalytic efficiency of the dimer.

The structure of *VcLMWPTP*-2 could not be solved due to low homology with others and the crystals were also not reproducible eliminating the chance of experimental phasing.

## Chapter 5: References

1. Kennelly P.J. (2002) *FEMS Microbiol. Letters*, 206; 1-8.
2. Kennelly, P, and Potts, M. (1999) *Frontiers in BioScience*, 4: 372-385.
3. Kennelly, P. et al.(1998) *FEMS Microbiol. Rev.*22: 229-253.
4. Stephenson, K. and Hoch, J. (2002) *Mol. Microbiol.*2: 297-304.
5. Ramponi, G. et al. (1997) *Cell Mol. Life Sci.* 53: 141-151.
6. Ramponi, G. and Stefani, M. (1997) *Int J Biochem Cell Biol.* 29:279-292.
7. Fallman, M. et al. (1997) *EMBO J*, 16: 2307–2318.
8. Dodson, E. et al. (2003) *Acta Crystallogr D Biol Crystallogr.* 59: 1131–1137.
9. Emsley, P. and Cowtan, K. (2004) *Acta. Crystallogr. D Biol. Crystallogr.* 60: 2126–2132.
10. Zwart, P. H. et al. (2010) *Acta. Crystallogr. D Biol. Crystallogr.* 66: 213–221

# **Chapter-1**

## **Introduction**

The event of protein phosphorylation-dephosphorylation plays an important role in the livelihood of cells. Though it was considered simply a mere post-translational modification once, it comes out as one of the necessary components in controlling the cell fate later on. This phenomenon in eukaryotes seeks the attention initially and then spreads in the prokaryotic level also. The addition of phosphoryl group in a specific residue is as much important as its removal. The synchronized way of addition and removal of phosphoryl group is achieved generally by kinase-phosphatase pair. The kinase adds the phosphoryl group using any nucleotide triphosphate to initiate the downstream processes and phosphatase hydrolyzes the phosphoryl group upon completion of the downstream cascade. Disturbance in any of the two processes may be fatal to the survival of the cell. In short, the maintenance of the phosphorylated pool is important to achieve proper functioning of the cell. Since the role of phosphatases being very specific, nonredundant and highly regulated in many cellular processes, they are considered as better drug targets than kinases [1].

Phosphorylation is most frequently observed on the side chain of serine, threonine, tyrosine, histidine and aspartate residues of a protein. Based on the phosphorylated substrates, phosphatases are categorized broadly in three different classes, first where the phosphatase acts on phosphorylated serine and/threonine is called Ser/Thr phosphatase, second where the substrate is phosphorylated tyrosine is called Phospho/Protein Tyrosine Phosphatase (PTP) and lastly a dual specificity phosphatase (DSP) which is capable of removing the phosphoryl group from all of the three residues. Apart from that, several other phosphatases are also available acting on other than these residues. Each class of phosphatases is again subdivided into many other groups based on other factors. The Ser/Thr phosphatases have been classified into three structurally distinct

families: (1) the PPM family of  $Mg^{2+}$ -dependent phosphatases, (2) the  $Mg^{2+}$ -dependent FCP family and (3) the PPP family.

PTPs are mainly of two types –(1) Cys-based PTPs and (2) Asp-based PTPs. The first type is again categorized in three classes- I, II and III and the second type i.e. the Asp-based PTPs are exemplified by the Eya (eyes absent) tyrosine phosphatases [2]. The latter one is also a part of the HAD family, which is now emerging as a very large protein family with representatives in plants [3], prokaryotes [4], mammals and includes numerous enzymes with other than Tyr-specificity [5-8]. Class I Cys-based PTPs are structurally related to the first PTP - PTP1B. There are at least 99 members of this family in the human genome [9] and they can be further subclassified into the classical PTPs (receptor-like and non receptor), and the VH1-like phosphatases, which contain the mitogen-activated protein (MAP) kinase phosphatase (MKPs), the atypical DSPs, the slingshots, the PRLs, the CDC14s, the PTEN group and the myotubularins. The class II Cys-based protein phosphatases comprise a small group of cell cycle regulators known as the CDC25 phosphatases. Although their catalytic machinery is very similar to the class I enzymes, they are structurally unrelated and, instead, bear considerable resemblance to bacterial rhodanese enzymes [10], and are thought to have evolved relatively late in eukaryote evolution. The class III Cys-based protein phosphatases are widely distributed in all kingdoms of life and most bacteria have the genes for one or two such enzymes in their genomes. In *Escherichia coli*, one such phosphatase regulates the tyrosine phosphorylation of a transmembrane tyrosine autokinase, which regulates synthesis of polysaccharides of the bacterial capsule [11]. In the Gram-negative *Bacillus subtilis*, the two class III phosphatases YfkJ and YwlE have clearly distinct properties and bacterial knockout strains have distinct phenotypes [12]. The human genome contains a single gene for a class III PTP, the low-*Mr* PTP

(LMWPTP), which undergoes alternative splicing to yield two active and one inactive isoforms. Although a polymorphism in this gene correlates with numerous common human diseases [13], the exact function of LMPTP has remained obscure.

### **1.A1. Phosphorylation event in bacteria**

For a long time, it was believed that the phosphorylation systems in bacteria and eukaryotes were mutually exclusive and the latter one possess only serine-threonine and tyrosine kinases, while bacteria were thought to possess only histidine and aspartate kinases. The search for genes encoding bacterial serine-threonine kinases were stemmed as genomic sequencing became readily available in the 1990s. As researchers realized that these proteins can play a crucial role in pathogenicity, they came under intense scrutiny. Many of these bacterial kinases and phosphatases have close eukaryotic counterparts, indicating that they may interfere with the host defense mechanisms. Phosphorylating bacterial proteins on serine-threonine and tyrosine residues also controls transcription of some bacterial virulence genes, including those of *M. tuberculosis* and various *Streptococcus* genes. The first bacterial protein kinase for which the structure was solved, was the catalytic domain of HPr kinase/phosphorylase from *Lactobacillus casei* [14].

Protein phosphorylation in eukaryotes is not only limited to regulate enzyme activities but also to control protein-protein interactions. Some eukaryotic proteins containing SH2 domain and other characteristic domains specifically interact with phospho-tyrosine-containing proteins [15]. These so-called SH2 domains have not been described in bacteria. However, FHA domains, which recognise phospho-threonine-containing proteins are present not only in eukaryotes [16] but also in clostridia, mycobacteria, *Streptomyces* and several gram-negative bacteria [17, 18].

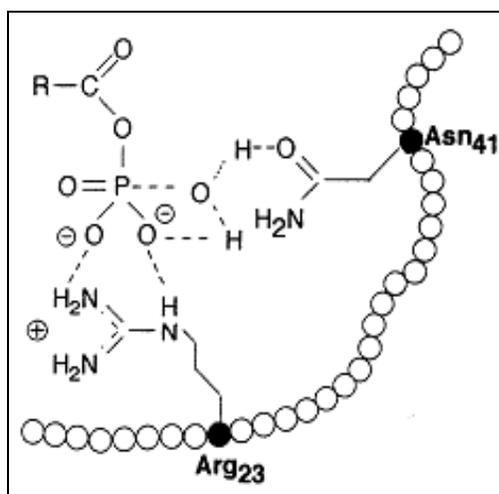
There are also other types of protein-protein interactions in bacteria, not requiring a specific domain, that involve recognition of a certain type of phosphorylated amino acid residue in one of the two proteins by the other protein. The catabolite control protein A (CcpA) is an example where it needs to interact with its co-repressor P-Ser-HPr in order to bind to its DNA target sites specifically [19, 20]. The structure of this repressor/co-repressor complex gives an insight into phosphorylation-dependent protein/protein interactions [21].

### **1.A2. Hydrolysis of carboxyl phosphate bond – A tale about another Phosphatase**

Phosphatases discussed so far are functionally defined to hydrolyze phosphor-ester bonds of aryl or alkyl group generally. If the hydrolysis reaction occurs on a carboxyl phosphate bond, then the acting enzyme is termed as Acylphosphatase. Acylphosphatase is one of the smallest (~10kDa) enzymes ubiquitously found in all organisms. It acts on the substrates having carbamoyl-phosphate bond and gives phosphate as a product along with carboxylate which are intermediate products of various metabolic and biosynthetic pathways [22]. The substrates include acetyl phosphate, 1,3-bisphosphoglycerate, succinoyl phosphate, and carbamoyl phosphate, which are intermediates in glycolysis, tricarboxylic acid cycle, pyrimidine, and urea biosynthesis [23]. It is also shown to cleave the phosphodiester bond of nucleoside diphosphate and nucleoside triphosphate, though at a slower rate [24]. The enzyme is shown to increase the rate of glycolysis which is further supported by the fact that the enzyme is overexpressed in tissues that are very active in glycolysis (e.g., brain, heart, and skeletal muscle) [23]. AcP can also hydrolyze the  $\beta$ -aspartyl phosphate intermediate of the  $\text{Ca}^{2+}$ - and  $\text{Na}^+/\text{K}^+$ -ATPases and may play a role in regulating ion-transport across the membrane [25-29].

### 1.A3. Structural aspects of Acylphosphatase in light of its catalytic action

Acylphosphatase is a closely packed  $\alpha/\beta$  protein composed of two  $\beta\alpha\beta$  motifs intercalated into each other [30], forming an antiparallel  $\beta$ -sheet. The residues important for its catalytic action are Arginine and Asparagine. Arg binds the hydrolysable moiety of the substrate through its guanidinium group while the Asn residue, which lie in close spatial proximity to the former residue [31]. The role of Asn residue is to polarizes an water molecule that initiate the hydrolysis of O–P bond during catalysis as previously suggested [32] (Figure 1).



**Figure 1: Proposed catalytic mechanism for Acylphosphatase.** The P-atom of the incoming substrate is stabilized by the side chain of Arginine (Arg23) at the catalytic site and a water molecule polarized by Asparagine (Asn41) hydrolyzes the substrate. (Figure courtesy: Ramponi et al. 1997; *Cell Mol Life Sci.*)

### 1.A4. Structural and functional aspects of *Vibrio cholera*0395- Acylphosphatase

In case of *Vibrio cholerae* 0395 acylphosphatase (Vc-AcP) the catalytically important Arginine residue is naturally mutated to Cysteine residue. Structural information related to other AcPs shows the Arginine to be positioned at the surface to bind the substrate efficiently. Since Vc-AcP lacks this residue at the surface, it may influence its structure and catalytically it may become inactive.

### **1.A5. Role of Tyrosine phosphorylation in bacteria**

The importance of tyrosine phosphorylation on bacterial physiology at the level of gene regulation and control of enzyme activity is now well-accepted. Higher stability of phosphorylated tyrosine than phosphorylated histidine/aspartic acid moiety drew much attention for the experimentalists. Also the fact that tyrosine residues are less abundant than both serine and threonine makes it more valuable. Therefore, tyrosine phosphorylation could represent a means to produce signals with the aim to regulate specific functions including the virulence mechanisms of many pathogens. Hence they can be the ideal targets for drug development.

### **1.A6. Bacterial tyrosine kinase (BYK) and Bacterial tyrosine phosphatases (BYP)**

Bacterial protein-tyrosine kinases (BYKs) are unique enzymes which exploits the ATP/GTP-binding Walker motif to catalyze phosphorylation of specific tyrosine residues of other proteins [33]. Generally a typical BYK consists of a variable-sized N-terminal transmembrane loop, connected to its C-terminal tail of by a cytosolic catalytic domain containing the ATP-binding sites [34].

BYK autophosphorylation has been linked to their functional role in exopolysaccharide production in several bacteria. They are usually encoded by genes in the large operons involved in biosynthesis and export of sugar polymers [35]. One of the most studied systems related to the importance of BYKs in exopolysaccharide synthesis is the Proteobacteria *Escherichia coli* which contains two BYKs Wzc and Etk. The first one is encoded by a gene in the operon which participates in the biosynthesis of polysaccharide polymers [36] and the latter one, encoded by a gene in the G4C operon which is required for formation of group 4 capsules (G4C) polysaccharide [37]. It has been demonstrated that the presence of both kinases is essential for

synthesis of corresponding extracellular polysaccharides. More to the point, the auto-phosphorylation of tyrosine residues in the C-termini of these BY-kinases is the key feature in the assembly of capsular polysaccharides. These BYKs have been shown to influence the amount of polysaccharides as well as the length of the produced polymer probably via an interaction with a polysaccharide polymerase Wzy [38-40]. In particular, BYKs control the exopolysaccharide content of the cell envelope, thus affecting the attachment of pathogens to host cells and their recognition by lymphocytes. Apart from the virulence gene expression, this pair of kinase-phosphatases are also found in some nonpathogenic bacteria where they are thought to play some other roles like metabolite synthesis, cell wall material production etc.

Protein tyrosine phosphatase (PTP) superfamily enzymes have been discovered in a range of prokaryotes, and most of them appear to play a roles that mimic their better-known eukaryotic counterparts as regulators of cellular function [41,42]. The catalysis reaction in turn can result in either the propagation or inhibition of phospho-dependent signaling. Protein tyrosine phosphatases are capable of possessing dual functions, in some cases, they can stimulate actions of cognate protein tyrosine kinases, and in other cases, they may antagonize their actions [43]. Depending on the substrate specificity bacterial tyrosine phosphatases can be categorized into three different families: (1) The eukaryotic-like phosphatases (PTPs) and dual-specific phosphatases (dsPTPs) that also display activity against phosphoserine and phosphothreonine; (2) The low molecular weight protein-tyrosine phosphatases (LMW-PTPs), a family of small acidic enzymes also found in eukaryotes and (3) The polymerase-histidinol phosphatases (PHP), a family of phosphoesterases commonly found in grampositive bacteria.

Both class I and II enzymes utilize a common catalytic mechanism involving the conserved signature C(X)<sub>5</sub>R motif in the phosphate binding loop or P-loop where cysteine, acts

as a nucleophile and makes an attack on the phosphorus atom of the phosphotyrosine residue of the substrate. The arginine residue interacts with the phosphate moiety of the Phosphotyrosine [44]. The cysteinyl-phosphate intermediate thus formed is hydrolyzed by an aspartic acid residue, the location of which varies among the families. Though BYPs are involved in many cellular processes, their major function is limited to the production of polysaccharide including exopolysaccharide and capsular polysaccharide and other proteins involved in the disruption of host cell machinery. Since capsular polysaccharides are considered as one of the key virulence determinants in some organisms, the function of BYPs can be assigned as a central player in the information flow that controls pathogenic activity. The function of several BYPs is listed in **Table 1.**

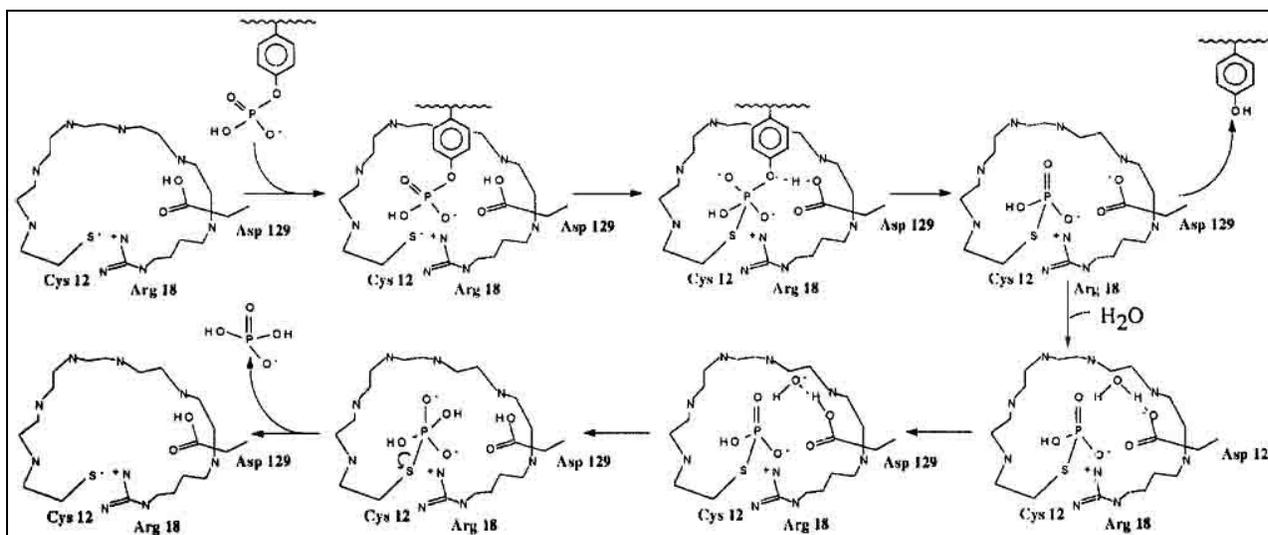
**Table 1. Functions of several BYPs from different organisms**

<b>Organism name</b>	<b>BYP-type</b>	<b>Function</b>	<b>Reference</b>
<i>Yersinia pseudotuberculosis</i>	conventional phosphatase	YopH- host infection	Bliska et al.1991
<i>Escherichia coli</i> K-30	Low Molecular weight phosphatase	Wzb- Group 1 capsule assembly	Wugeditsch et al.2001
<i>Bacillus subtilis</i>	Polymerase and histidinol phosphatase	YwqE- exopolysaccharide synthesis	Mijakovic et al.2003
<i>Anabaena</i> sp. strain PCC7120	dual specific phosphatase	PTP- function not reported	Kennelly and Potts 1999

### **1.A7. Low Molecular Weight PhosphoTyrosine Phosphatase**

LMWPTP is a class of large PTP family sharing the common structural motif “CX<sub>5</sub>GN (S/T)” and follows the same catalytic pathway [45]. The difference with other members of the PTP-superfamily is their molecular weight i.e. it has comparatively smaller size and apart from the catalytic site motif the overall 3-D structure is completely different. It contains two important loops at the active site, P-loop having the aforementioned motif where the cysteine initiates the

catalytic action by making a nucleophilic attack on the incoming phosphorylated substrate thereby forming a cysteinyl-phosphate intermediate which in turn gets cleaved by an activated water molecule. The water molecule responsible for the hydrolysis step is activated by a conserved Aspartic acid residue coming from the DPY-loop [46]. A schematic presentation of the catalytic mechanism is shown in **Figure 2**.



**Figure 2: The catalytic mechanism followed by LMWPTPs.** The cysteine (Cys12) forms a bond with the phosphorus atom of the phosphorylated tyrosine and the P-O bond linking the phosphate group to the tyrosine is protonated, either by Asp129 or a water molecule. The cysteinyl-phosphate intermediate is then hydrolysed by another water molecule, regenerating the active site for another reaction. The Arginine (Arg18) of the active site CX<sub>5</sub>R-motif along with the amide (-NH) environment of the P-loop stabilizes the reaction intermediate throughout the process. (Figure courtesy: Ramponi and Stefani; 1997; *Biochimica and Biophysica Acta*)

In case of gram-negative bacteria, the gene encoding the LMW-PTP generally is upstream of the tyrosine kinase in the same operon (**Figure 3**). They efficiently dephosphorylate these autokinases and thereby regulate their biological function [35].



**Figure 3: Genomic organization of BY-kinases and their cognate phosphotyrosine-phosphatase genes in proteobacteria.** In proteobacteria, the BY-kinase gene and the gene encoding a phosphatase of the LMW-PTP class are present in the same operon. (Figure Courtesy: Mijakovic et al. 2007; *Trends Biochem. Sci.*)

Generally one pair of LMWPTP is found in bacterial systems investigated so far. They differ from each other structurally and act on different substrates [47]. Some of them are also shown to be involved in pathogenicity of that organism [48] details of which are given in **Table 2**. Shaded rows contain information about gram-positive bacteria and the rest is gram-negative.

**Table 2. Function of BYP and BYK from several gram-positive and gram-negative bacteria**

Organism	BYK	BYP	Function	References
<i>Acinetobacter johnsonii</i>	Ptk	Ptp	-	Duclos et al.1996; Grangeasse et al.1998
<i>Acinetobacter lwoffii</i>	Wzc	Wzb	Emulsion production	Nakar and Gutnick 2003
<i>Erwinia amylovora</i>	AmsA	AmsI	Amylovoran production	Bugert and Geider 1997
<i>Escherichia coli</i> K-12	WzcCA	Wzb	Colanic acid synthesis	Vincent et al.1999
<i>Escherichia coli</i> K-12/K-30	Etk	Etp	Exopolysaccharide production	Ilan et al.1999
<i>Escherichia coli</i> K-30	WzcCPS	Wzb	Group 1 capsule assembly	Wugeditsch et al.2001
<i>Escherichia coli</i> L-form	UspA	Etp	Stress	Freestone et al. 1997
<i>Klebsiella pneumoniae</i>	Yco6	Yor5	Capsule synthesis	Arakawa et al. 1997; Preneta et al. 2002
<i>Pseudomonas solanacearum</i>	EpsB	EpsP	Exopolysaccharide transport	Huang and Schell 1995
<i>Bacillus subtilis</i>	-	PtpA	-	Li and Strohl 1996
<i>Mycobacterium tuberculosis</i>	-	MPtpA, MptpB	-	Koul et al.2000; Cowley et al.2002
<i>Staphylococcus aureus</i>	-	PtpA, PtpB	-	Soulat et al.2002
<i>Streptomyces coelicolor</i> A3	-	PtpA	-	Li and Strohl 1996

### **1.A8. Structural and functional aspects of BYP with respect to *Vibrio cholerae*0395-Low molecular weight phosphotyrosine phosphatases (Vc-LMWPTP)**

*Vibrio cholerae*0395 possesses two LMWPTPs that have been designated as VcLMWPTP-1 (155 aa) and VcLMWPTP-2 (166aa). They share only ~37% of identity between themselves while the closest structural homologues of VcLMWPTP-1 and VcLMWPTP-2 present in PDB (3IDO – *Entamoeba histolytica* and 2CWD – *Thermus thermophilus* respectively) bears a maximum identity of ~43% and ~37% respectively. Although VcLMWPTP-1 and VcLMWPTP-2 have the same functional motifs, low sequence identity amongst them indicates that the three dimensional structures may provide distinct features unlike others reported so far. It will also be useful to gain insights in to their catalytic functions at the atomic level. Moreover, knowledge about the structure and function of these phosphatases might be of use in drug design against this bacterial pathogen. As the molecular details of the function of a protein can be best predicted by its structure, the present study on LMWPTP from a gram-negative gammaproteobacteria *Vibrio cholerae* 0395 may shed some light on the prediction of their functions.

### **1.B1. Protein X-ray Crystallography to study the structure-function relationship of the aforementioned phosphatases**

Generally, different methods such as X-ray crystallography, Nuclear magnetic resonance spectroscopy, and Cryo-electron microscopy are employed to determine 3D structures of biological macromolecules. Of them, X-ray crystallography is the most convenient method to enable us to visualize protein structures at the atomic level and to enhance our understanding of protein function. The study involves how proteins interact with other molecules, how they

undergo conformational changes, and how they perform catalysis in the case of enzymes. Based on this information novel drugs can be designed to target a particular protein, or rationally engineer an enzyme for a specific industrial process. Hence the three dimensional structures of protein molecules and their respective functions are directly related to each other and understanding of this structure-function relationship has the potential applications in various fields ranging from basic science to pharmaceutical industries. Initially, X-ray diffraction was used for the determination of structures of small inorganic and organic compounds. Since proteins could also be crystallized, crystal structures of proteins could also be determined using X-ray crystallography. First protein structure was determined in 1958 when structure of myoglobin was solved by John C Kendrew. Subsequently, structure of hemoglobin was also determined by Max F Perutz. Since then, there is an exponential rise in the number of macromolecular structures solved and deposited in the Protein Data Bank (PDB) crossing 100000+ structures.

### **1.B2. Steps involved in protein crystallography study**

In order to obtain the crystal structure of a protein, several steps are to be performed starting with the preparation of the protein sample. Then the sample is used for successful crystallization followed by data collection and structure solution. Finally interpretation of the structure enables us to explain the detailed architecture and subsequent function of the protein.

### **1.B3. Protein preparation**

Firstly a pure sample of the target protein is required. This can be done either isolating it from its source, or by cloning its gene into a high expression system. The sample is then assessed for, stability, purity, homogeneity, proper folding, monodispersity, enzyme activity -all providing

information about the suitability for crystallization. If the sample fails one or more of the above criteria, it may be worthwhile returning to the expression and purification protocols and trying something different, such as the addition of ligands known to interact with the protein, or adding extra purification steps. In extreme cases it may be worthwhile switching to a different expression system altogether or working with a mutated or truncated construct. It may be possible to refold protein successfully using chaotropic reagents such as urea. Aggregated or polydisperse samples may be made monodisperse by simply changing pH or adding some salt. However, without DLS, this is very difficult to assess.

#### **1.B4. Crystallization technique and data collection strategy**

After the fulfillment of the most of the criteria, the protein sample is concentrated to screen a reasonable number of conditions containing various reagents. If however there is/are one or more "hits" in the screening step, then varying the concentrations of all components in the crystallization, changing the temperature and pH slightly, using additives, switching to similar buffers or precipitants, or even using different crystallization methods (e.g. dialysis, batch, in-gel) may give rise to diffraction-quality crystals. Occasionally good crystals will form overnight, but more typically they will take from several days to several weeks to grow.

Once crystals are obtained, they are shot with X-rays to obtain the diffraction pattern. For that, generally a single crystal needs to be mounted in some way either in a capillary at room temperature or flash-cooled to 100 K in a loop and then attached to a device called a goniometer head which enables the sample to be accurately positioned in the X-ray beam by means of a number of adjustment screws. For cryogenic data collection, a cold nitrogen gas stream keeps the crystal at 100 K throughout the experiment. Focused X-rays emerge from a narrow tube called a

collimator and strike the crystal to produce a diffraction pattern which is recorded on the X-ray detector. A good diffraction pattern should be devoid of any salt or ice crystals and spot streaking. Using the initial few images, it is needed to ascertain the crystal symmetry, the unit cell parameters, the crystal orientation and the resolution limit from the automated software. Depending on this information a data collection strategy is derived to maximize both the resolution and completeness of the data set. For that the crystal is rotated through a small angle called the oscillation, typically 1 degree, and record the X-ray diffraction pattern. The lower the symmetry, then more data is required [49].

### **1.B5. Data processing step**

Now the data coming from a 3D crystal is recorded in a 2D detector in the form of varying intensities of several spots. Each spot corresponds to a different type of variation in the electron density; determining which variation corresponds to which spot (indexing), the relative strengths of the spots in different images (merging and scaling) and how the variations should be combined to yield the total electron density (phasing) are important steps to obtain the structure. Interpretation of these intensities in terms of electron density maps and explaining the structure of the protein indeed requires a robust computing system. Prior to that sophistication, important mathematical tools such as, Fourier transforms are employed to answer these questions and later on integrated in computing system.

### **1.B6. 'Phase' problem in crystallography**

The data collected from a diffraction experiment is a reciprocal space representation of the crystal lattice. The position of each diffraction 'spot' is governed by the size and shape of the

unit cell, and the inherent symmetry within the crystal. The intensity (I) of each diffraction 'spot' is recorded, and this intensity is proportional to the square of the structure factor amplitude (|F|).

$$I(\mathbf{h},\mathbf{k},\mathbf{l}) \propto |F(\mathbf{h},\mathbf{k},\mathbf{l})|^2$$

where (h,k,l) is the co-ordinate of the reciprocal lattice of any point with the real space co-ordinate (x,y,z) .

The structure factor is a complex number containing information relating to both the amplitude and phase of a wave. In order to obtain an interpretable electron density map, both amplitude and phase must be known.

$$\rho(xyz) = \frac{1}{V} \sum_{hkl}^{+\infty} |F(hkl)| \cdot e^{-2\pi i[hx+ky+lz-\phi(hkl)]}$$

where,  $\rho$  is the electron density with a certain value at each point within the unit cell given by the coordinates (x, y, z), F(hkl) represents the resultant diffracted beams of all atoms contained in the unit cell in a given direction. And the magnitudes, one for each diffracted beam, are known as structure factors. (h, k, l) are the Miller indices of the diffracted beams (the reciprocal points) and  $\Phi(hkl)$  represent the phases of the structure factors.

The various methods that are applied to obtain the 'phase' are as follows:

1. **Ab initio phasing** or **direct methods** which is applicable for for small molecules (<1000 non-hydrogen atoms). If the resolution of the data is better than 1.4 Å (140 pm), direct methods can be used to obtain phase information, by exploiting known phase relationships between certain groups of reflections.

2. **Molecular replacement** method requiring a sequence identity >25% and an r.m.s. deviation of <2.0 Å between the C<sup>α</sup> atoms of the model and the final new structure,. Patterson methods are usually used to obtain first the orientation of the model in the new unit cell and then the translation of the correctly oriented model relative to the origin of the new unit cell.

3. **Isomorphous replacement (SIR, MIR)** method where the native data are compared with two or more heavy-atom (Au, Pt, Ag, Hg etc) derivatives collected at the same wavelength. The derivative contains a small and isomorphous modification that yet scatters strongly enough to detectably alter the diffraction pattern.

4. **Anomalous scattering (SAD/MAD)** technique where a crystal containing an anomalous scatterer (Se, S) are collected at different wavelengths. The positions of the anomalous scatterer can be determined from the differences and serve as first phase information.

### **1.B7. Refinement and Phase improvement**

After getting the initial phases, a model can be built and can be used to refine the phases, leading to an improved model, and so on. Solvent flattening, histogram matching and non-crystallographic averaging are the main techniques used to modify electron density and improve phases. Solvent flattening is a powerful technique that removes negative electron density and sets the value of electron density in the solvent regions to a typical value of  $0.33 \text{ e } \text{\AA}^{-3}$ , in contrast to a typical protein electron density of  $0.43 \text{ e } \text{\AA}^{-3}$ . Histogram matching alters the values of electron-density points to concur with an expected distribution of electron-density values. Non-crystallographic symmetry averaging imposes equivalence on electron-density values when more than one copy of a molecule is present in the asymmetric unit [50]. Given a model of some atomic positions, these positions and their respective B-factors, accounting for the thermal motion of the atom can be refined to fit the observed diffraction data, ideally yielding a better set of phases. A new model can then be fit to the new electron density map and a further round of refinement is carried out. This continues until the correlation between the diffraction data and the model is maximized. The agreement is measured by an R-factor or reliability factor defined as,

$$R = \frac{\sum_{\text{all reflections}} |F_o - F_c|}{\sum_{\text{all reflections}} |F_o|}$$

where,  $F_o$  and  $F_c$  are observed and calculated structure factor respectively.

Another quality criterion is  $R_{\text{free}}$ , calculated from a subset (~5-10%) of reflections that were never used in the structure refinement. Both R factors depend on the resolution of the data.

### 1.B8. Validation and deposition of the structure

Once the R-factors cease to decrease further, the reliability of a structural model has to be assessed in terms of several tests by the ‘model validation’ procedure. Thus, the structural model should be continuously checked and validated using consistent stereochemical criteria (for example, bond lengths and bond angles must be acceptable). Ramachandran plot is used to assess the geometrical restrictions followed by the peptide bonds along with the torsional angles of the usual conformations shown by the amino acid chains. Similarly, the values of the thermal factors associated with each atom should show physically acceptable values. These parameters account for the thermal vibrational mobility of the different structural parts. Thus, in the structure of a macromolecule, these values should be consistent with the internal or external location of the chain, being generally lower for the internal parts, and higher for external parts near the solvent. Therefore The structure which maintains the following the criteria –(1) a reasonable agreement between observed and calculated structure factors, (2) bond distances, bond angles and torsional angles that meet stereochemical criteria and (3) physically reasonable thermal vibration factors.

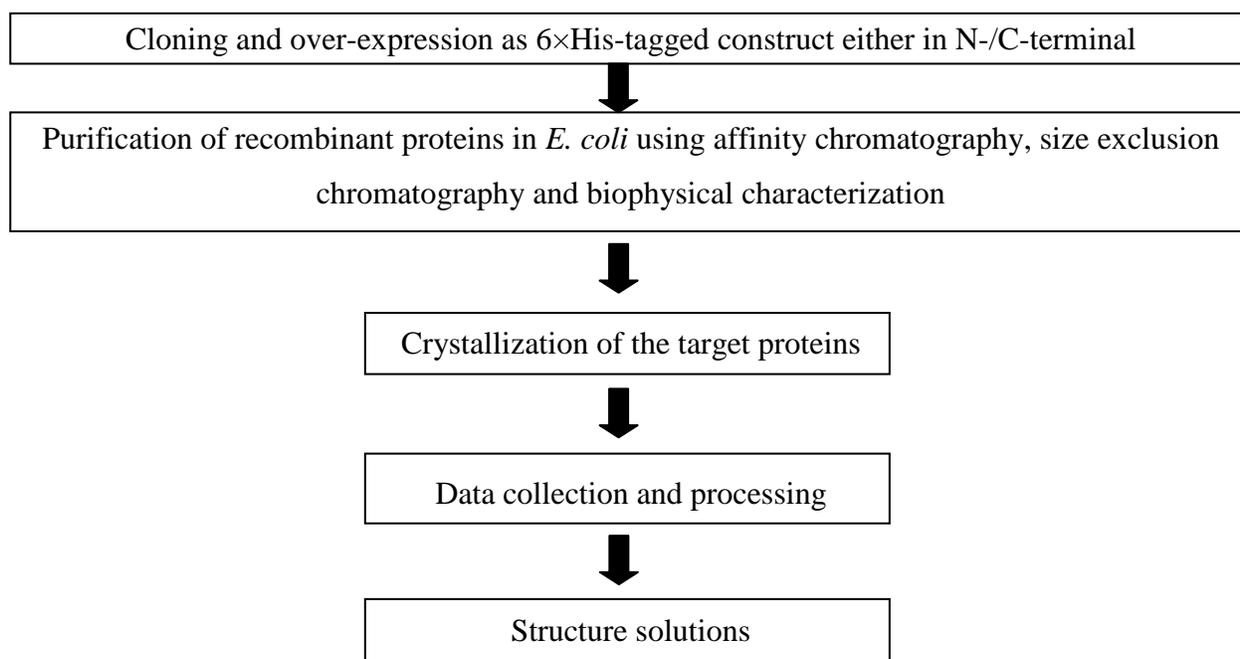
Finally the structural model along with the scaled reflection file (.mtz) is deposited to the Protein Data Bank (PDB).

# **Chapter-2**

## **Materials & Methods**

## 2. Cloning, Over-Expression, Purification, Crystallization, Data Collection, Processing and Refinement

To decipher the structural basis of their probable function grossly the same flow-diagram was adopted for Vc-AcP and Vc-PTPs proteins and their mutants. Detailed descriptions of each step of the flow-diagram are described thereafter.



### 2.1. Cloning and Over-expression of the proteins as 6x His-tagged construct

The genomic DNA of *Vibrio cholerae* O395 was used as a template to clone each wild type protein (Vc-AcP, VcLMWPTP-1 and VcLMWPTP-2) as N-terminal 6xHis tagged recombinant constructs. Mutants of each protein (Vc-AcP-C20R, VcLMWPTP-1-C8S and VcLMWPTP-2-C12S) were prepared by site directed mutagenesis technique to overexpress in the same way as that of the wild type protein. A truncated construct of VcLMWPTP-2 wild type protein was also cloned and overexpressed as N-terminal 6xhis tagged recombinant protein.

Apart from that, *Vc*LMWPTP-1-C8S and wild type *Vc*LMWPTP-2 was cloned to over-express as C-terminal 6×his tagged recombinant proteins.

Initially the genomic DNA was targeted for PCR amplification with appropriate pair of primers followed by purification of the PCR-product. The purified product was then digested with restriction enzymes followed by another step of purification. The vector was also isolated, digested and purified simultaneously. The vector and insert obtained thus was ligated properly and transformed to electrocompetent cells. Cells carrying the ligated product was then screened and again transformed to check over-expression of the protein of interest by the addition of inducer.

### **2.1A. Cloning and over-expression of the wild type *Vc*-AcP and its mutant**

*Vc*-AcP (1-94) was cloned into pET-28(a+) vector under the control of T7 promoter. The complete DNA sequence of this construct was confirmed by DNA sequencing. The N-terminal 6×His-tagged wild-type proteins were overexpressed into B121 (DE3) strain of *E. coli*. The same procedure was followed to clone and overexpressed the single mutant (C20R).

### **2.1B. Cloning and over-expression of the wild type *Vc*LMWPTP-1 and its mutants**

*Vc*LMWPTP-1 (1-155) was cloned into pET-28(a+) vector under the control of T7 promoter. The complete DNA sequence of the construct was confirmed using DNA sequencing by SciGenome. The N-terminal 6×His-tagged wild-type protein was overexpressed into B121(DE3) strain of *E. coli*. The same procedure was followed to clone and overexpressed its N-terminal 6×His-tagged single C8S mutant. The C8S mutant of *Vc*LMWPTP-1 was also cloned into pET-24b (+) vector and overexpressed into B121 (DE3) strain of *E. coli* as C-terminal 6×His-tagged protein due to unsuccessful purification of the N-terminal 6×His-tagged mutants.

### 2.1C. Cloning and over-expression of the wild type *VcLMWPTP-2* and its mutants

*VcLMWPTP-2* (1-166) and truncated *VcLMWPTP-2* or *VcLMWPTP-2<sup>tr</sup>*(1-155) were cloned into pET-28(a+) vector under the control of T7 promoter. The complete DNA sequences of these constructs were confirmed by DNA sequencing. The N-terminal 6×His-tagged wild-type proteins were overexpressed into B121 (DE3) strain of *E. coli*. The same procedure was followed to clone and overexpressed the single C12S mutant. Wild-type *VcLMWPTP-2* was also cloned into pET-24b(+) vector and overexpressed into B121 (DE3) strain of *E. coli* as C-terminal 6×His-tagged proteins, *VcLMWPTP-2<sup>Ctag</sup>*. Details of each step are given below.

The sequences of the forward and reverse oligos of the wild type and mutants along with the recognition sequences of the restriction enzymes are listed in **Table 3** and **Table 4** respectively.

**Table 3: Primer Sequences of wild type proteins**

Clone name	Primer Sequences
<i>Vc</i> -AcP	G <sup>f</sup> P : 5'-GGAATTC <u>CATATG</u> ATGAAAAAATGGAAAAACAGTGCAGTAAG-3'
	G <sup>r</sup> P : 5'-GGGGATCCTTACTACAGAATTTCAAACCCTTGATAG-3'
<i>VcLMWPTP-1</i>	G <sup>f</sup> P : 5'-CCAGC <u>CATATG</u> CAGAAGGTACTCGTGGTGTGC-3'
	G <sup>r</sup> P : 5'-CGGGGATCCTTAATGCTGGCCTTGCTGTTTTAG-3'
<i>VcLMWPTP-2</i>	G <sup>f</sup> P : 5'-GCCGCC <u>CATATG</u> AAGGTTAAAGGTTTATCAG -3'
	G <sup>r</sup> P : 5'- CGGGGATCCTTATTGAGATAAATTTTCGTTGCACGC -3'
<i>VcLMWPTP-2<sup>tr</sup></i>	Same as G <sup>f</sup> P of <i>VcLMWPTP-2</i>
	G <sup>r</sup> P : 5'-CGGGGATCCTTACTGCTCTTGCC-3'
<i>VcLMWPTP-2<sup>c</sup></i>	Same as G <sup>f</sup> P of <i>VcLMWPTP-2</i>
	G <sup>r</sup> P : 5'-CCGCTCGAGTTGAGATAAATTTTCGTT-3'

**Table 4: Primer Sequences of mutant proteins for the 1<sup>st</sup> step PCR of SDM**

Clone name	Primer Sequences for the 1st step PCR of Site Directed Mutagenesis
VcLMWPTP-1-C8S	M <sup>f</sup> P : 5'-CAGAAGGTACTCGTGGTG <u><b>A</b></u> GCATGGGCAATATTTGCCGTTTC-3'
	M <sup>r</sup> P : 5'-GAACGGCAAATATGCCCATG <u><b>C</b></u> TACCACGAGTACCTTCTG-3'
VcLMWPTP-2-C12S	M <sup>f</sup> P : 5'-GGTTTATCAGTATTAGTCGTG <u><b>A</b></u> GTACAGGTAATTTGTGTCGTTTCG-3'
	M <sup>r</sup> P : 5'-CGAACGACACAAATTACCTGTACT <u><b>T</b></u> CACGACTAATACTGATAAACC-3'
Vc-AcP-C20R	M <sup>f</sup> P : 5'-GTGCAAGGCGTGGGTTTT <u><b>A</b></u> GGTATCACACGTCTCATCAAGG-3'
	M <sup>r</sup> P : 5'-CCTTGATGAGACGTGTGATAC <u><b>C</b></u> TAAAACCCACGCCTTGCAC-3'

*G<sup>f</sup>P* -gene forwarding primer, *G<sup>r</sup>P* –gene reverse primer; *Restriction enzyme sites are underlined within the primer sequences. M<sup>f</sup>P* -mutation forwarding primer, *M<sup>r</sup>P* –mutation reverse primer; *The nucleotide written in bold and underlined mark the changed position of required mutation.*

### 2.1D. Polymerase Chain Reaction (PCR)

PCR technique [51] was used to amplify the gene *acyP* encoding Vc-AcP from the genomic DNA of *Vibrio cholera*O395 using as template. The same template was used to amplify the gene with the ordered locus names of VC0395\_A0559 and VC0395\_A0440 encoding VcLMWPTP-1 and VcLMWPTP-2 respectively. The annealing temperature  $T_m$  was chosen to be approximately 5K below the lowest (theoretical) melting temperature of both oligo and 30 PCR-cycles were carried out. Initially ‘gradient PCR’ with varying annealing temperature was carried out and depending on that second step PCR was done for the wild type proteins. The program used for both of the steps of PCR is given below and calculation of the reaction mixture for 100µl of total volume is listed in **Table 5**.

- Initial denaturation at 95°C for 5 minutes
- [Denaturation at 95°C for 30 seconds
- Annealing at  $X \pm 5^\circ\text{C}$  for 30 seconds
- Polymerization at 72°C for 45 seconds]<sup>30 cycles</sup>
- Final extension at 72°C for 10minutes. (where, X varies for different constructs)

**Table 5: Composition of the PCR reaction mixture**

Components	Final concentrations
Template- <i>Vc</i> -genomic DNA	~20ng
Forward primer	0.8pmol
Reverse primer	0.8pmol
dNTP mix-100X	1X
10X PCR buffer	1X
<i>Taq</i> polymerase	2U
Autoclaved water to make the final volume 100µl.	

**2.1E. Site directed mutagenesis (SDM):**

SDM by two-step PCR method as described in [52] was employed to prepare the single and double mutants. Briefly, the first step was equivalent to ‘gradient PCR’ to judge the proper annealing temperature using ‘gene forward’ and ‘mutation reverse’ primers in one set and ‘gene reverse’ and ‘mutation forward’ primers in another. The PCR products from both the sets were taken as templates for the second step which is similar to the second step of PCR of wild type fragments. The calculation is given in **Table 6**.

**Table 6: Calculation of the two-step-PCR reaction mixture**

Components	Final concentrations 1 <sup>st</sup> step PCR	Components	Final concentrations 2 <sup>nd</sup> step PCR
<i>Vc</i> -genomic DNA	~20ng	1 <sup>st</sup> step PCR products	~10ng each
‘Gene forward’ and ‘Mutation reverse’ primer – set I	0.8pmol each	Gene forward primer	0.8pmol
‘Gene reverse’ and ‘Mutation forward’ primer – set II	0.8pmol each	Gene reverse primer	0.8pmol
dNTP mix-100X	1X	dNTP mix-100X	1X
10X PCR buffer (Fermentas)	1X	10X PCR buffer	1X
<i>Taq</i> polymerase (Fermentas)	2U	<i>Taq</i> polymerase	2U
Autoclaved water to make the final volume 100µl.			

## 2.1F. Digestion and Ligation

The PCR amplified products were purified with the Qiagen “PCR purification kit”, digested with restriction enzymes (*NdeI* and *BamHI*) for N-terminal his-tagged or (*NdeI* and *XhoI*) for C-terminal his-tagged, purified using Qiagen “Gel extraction kit”. pET28a(+) and pET24b(+) vectors were previously digested with (*NdeI* and *BamHI*) or (*NdeI* and *XhoI*) enzymes respectively to create the ‘sticky ends’ and purified in the same way as that of the insert. Finally the insert was ligated into the respective target vector(s). The compositions are given in **Table 7**.

**Table 7: Composition of the restriction enzyme double digestion and ligation**

Restriction enzyme double digestion		Ligation	
Components	Final concentration	Components	Final concentration
10X Buffer <sup>3</sup> / Buffer <sup>4</sup> (NEB)	1X	10X Ligase Buffer (Promega)	1X
Purified PCR product	2-4µg	pET 28a(+) / pET 24b(+) Vector (Promega) : Insert	1:2
Enzyme I and II	1U each		
100X BSA	1X	T4 DNA Ligase (Promega)	1U
Autoclaved water to make the final volume 20µl.		Autoclaved water to make the final volume 10µl.	

## 2.1G. Transformation and Analysis

The ligation product was directly transformed into electrocompetent cells (XL1-blue, Stratagene) according to the manufacturer’s instructions and plated on agarose plates with 30µg/ml Kanamycin. Plasmids were purified from colonies by mini-preps using ‘alkaline lysis method’ [53] and analyzed on agarose gels after digestion with the restriction enzymes used during cloning. Finally the sequence of the clones had been confirmed by sequencing.

## 2.1H. Protein over-expression

Over-expression of the 6×His-tagged (both N-and C-terminal) proteins was carried out in *E. coli* strain BL21 (DE3) (Novagen) in LB medium at 37°C. The BL21 (DE3) competent cells were transformed with the recombinant pET28a (+) plasmids containing the insert. From the transformed colonies, four single colony were inoculated in individual falcons of 2ml fresh LB medium containing *Kanamycin* up to a final concentration of 30 µg/ml and grown overnight at 37°C at 150rpm. Next day, this overnight culture was inoculated to 3ml of fresh LB medium in 1:50 ratio and allowed to grow at 37°C at 175rpm till OD<sub>600</sub> reached to ~0.4-0.5. 1.5 ml aliquot was taken out at that point to preserve as ‘uninduced’ cell and rest of the culture was induced by IPTG with a final concentration of 1mM followed by growth for another 4 hours at the same temperature. Aliquots from both uninduced and induced cells were harvested in a 1.5 ml microcentrifuge tube by centrifugation at 12000 g for 2 min and the resulting pellet was suspended in 30 µl of H<sub>2</sub>O, vortexed and mixed with 10µl of 4X SDS-PAGE sample buffer (250 mM Tris.HCl of pH 6.8, 10 % SDS, 40 % glycerol, 5 % Bromophenol blue and 20 % β-mercaptoethanol). The cells were lysed by boiling in water bath for 5 min and after centrifuging at 12000 g for 5 min, 15µl of the supernatant was loaded in SDS-polyacrylamide gel and allowed to run at 200V;20 mA. Over-expression of the recombinant proteins were checked in the gel stained with Coomassie R-250 (Sigma). Appearance of a broader band of the respective proteins in the induced cells compared to the uninduced cells and comparing it to the protein molecular weight marker confirmed the over-expression of the recombinant proteins.

## **2.2. Purification of recombinant proteins in *E. coli* using affinity chromatography and size exclusion chromatography**

### **2.2A. Purification of 6×His-tagged proteins using affinity chromatography**

Purification of the His-tagged protein by Ni<sup>+2</sup>-NTA is based on the remarkable affinity between the six consecutive histidine residues – ‘the 6×His tag’, either at the N-terminus or at the C-terminus of the proteins, and Ni<sup>+2</sup> ion immobilized on Nitrilotriacetic acid (NTA) resin. This affinity is even greater than that of the antigen-antibody or the enzyme-substrate interactions. Untagged proteins present in the cell lysate having histidines in the surface/close proximity to the surface may also bind to the Ni<sup>+2</sup>-NTA along with the desired ‘His tagged’ protein but their interactions with Ni<sup>+2</sup>-NTA, in practice, will be much weaker compared to the specific interactions of the 6×His tagged protein with Ni<sup>+2</sup>-NTA. So, the unbound and/or nonspecifically bound proteins to the matrix can easily be washed away under relatively stringent conditions without affecting the target protein. The target protein can be extracted by elution with either higher concentration of imidazole or slight reduction in pH.

For the purification of the recombinant proteins two methods were adopted. In one method, overnight cultures of BL21 (DE3) cells harboring the recombinant pET28a (+) plasmid that carries the corresponding inserts were inoculated (1:100 dilution) to fresh LB medium supplemented with *Kanamycin* (final concentration of 30 µg/ml). The cultures were then grown at 37°C, 175rpm until the OD<sub>600</sub> reached to approximately 0.4-0.5. Expression of the desired proteins was induced at that point by addition of IPTG to a final concentration of 0.2mM followed by an incubation of another 4 hours at 37°C, 175rpm. In another method, the large cultures after inoculation were allowed to grow at 37°C, 175rpm for 1hr. and then at 20°C,

175rpm for another 1hr. Then they were induced by IPTG and allowed to grow at 20°C, 100rpm for 12-14hrs.

The bacterial cells, harvested from the liquid culture by centrifugation at 4500 rpm for 15 min at 4°C, were resuspended in ice-cold lysis buffer (LyB) with no imidazole. Suitable buffers for different proteins are listed in **Table 8**. Lysozyme (final concentration 1mg/ml) and PMSF (final concentration 1 mM) were added in the resuspended solution right then and incubated on ice for 30 minutes. The cell suspension was sonicated using 25-30 bursts of 15 secs at 200 W with an interval of 45 seconds. The supernatant was separated from the insoluble cell component that comes into pellet by centrifugation of the crude lysate at 12000g for ~45 min at 4°C. The supernatant was collected carefully without disturbing the pelleted debris and was subjected to Ni<sup>+2</sup>-NTA affinity chromatography.

The cleared lysate containing the soluble 6×His tag protein was applied onto a column packed with 1.5 ml slurry of Ni<sup>+2</sup>-NTA resin, pre-equilibrated with the lysis buffer. Weakly/non-specifically bound proteins were washed from the resin with the wash buffers having gradually increased imidazole concentrations of 5mM (WB1), 10mM (WB2) and 15 mM (WB3). The recombinant His-tagged proteins were stripped from the Ni<sup>+2</sup>-NTA columns with elution buffer (EB) containing 150-250 mM imidazole in LyB in the same buffer and the elution was monitored by measuring the absorbance at 280 nm. The purity of the eluted protein was checked in 15% SDS-PAGE. The elution fractions contain the target proteins with few minor contaminants. These fractions were pooled and concentrated in the lysis buffer using 3-10 kDa cut-off centrifugal unit of Millipore at 4500g at 4°C.

## **2.2B. Removal of His-tag with thrombin**

Prior to the cleavage of the fusion proteins, the protein was dialyzed against suitable gel filtration buffer (GB) having no imidazole. NaCl was added accordingly to avoid any sort of aggregation that may form in the highly concentrated protein and its concentration was standardized accordingly. This exchange of buffer was necessary as the high salt and imidazole present in the elution buffer may interfere with the proteolytic cleavage efficiency of thrombin. Moreover, it is reported that imidazole can influence crystallization trials and the presence of imidazole often results in protein aggregation. For the cleavage reaction, it is necessary to optimize the amount of thrombin, temperature and time of incubation which varies for each individual protein and even sometimes for their mutants. In general, about 6-10 mg of the fusion protein was digested with 2 units of thrombin (Novagen) at 4°C for 38-40 hr.

## **2.2C. Fast Protein Liquid Chromatography (FPLC) using Sephacryl S-100**

After the thrombin cleavage, the proteins were further purified from the thrombin, cleaved His tag and other few minor contaminants eluted from the Ni<sup>+2</sup>-NTA column by size exclusion chromatography using Sephacryl S-100 (GE Healthcare Biosciences). The reaction mixture set for the thrombin cleavage was loaded directly on a Sephacryl S-100 column of dimension 85×1.5 cm., pre-calibrated with protein mixture of different molecular mass and pre-equilibrated with respective GB. 0.02 % Na-Azide, was added to the GB to avoid any sort of microbial contamination. The elution profiles were recorded in a chart recorder based on the absorbance profile at 280nm. The eluted fractions were then analysed by SDS-PAGE to assess the purity of the samples and finally pooled using 3-10 kDa cut-off centrifugal unit of Millipore at 4500g at 4°C.

FPLC was used to purify and detect oligomeric states of the protein of interest. VcLMWPTP-1 (0.85mg/ml, 1.75mg/ml, 2.58mg/ml and 4.25mg/ml) in three different 50mM Tris buffer, each at pH 7.6 containing 150mM, 300mM and 500mM NaCl was fractionated by a Sephacryl S-100 (Amersham Biosciences) column (46 × 1.6 cm) at 0.9MPa, preequilibrated with respective buffers and precalibrated with a protein mixture containing Lysozyme (MW14.3 kDa) FluC, a monomeric protein from *V.cholerae* (MW 31.0 kDa) and Bovine serum albumin (MW 66.45kDa) at room temperature. Fractions were collected at a flow rate of 0.4ml per minute using an ÄKTAPrime chromatographic system. The elution profile was determined by monitoring the absorbance at 280 nm.

**Table 8: Buffer compositions for the purification of recombinant proteins**

<b>Protein name : Vc-AcP and its mutants</b>	
Lysis Buffer -LyB	50 mM Tris (pH 7.5), 300 mM NaCl
Wash Buffer I – WB1	50 mM Tris (pH 7.5), 300 mM NaCl, 10mM Imidazole
Wash Buffer II – WB2	50 mM Tris (pH 7.5), 300 mM NaCl, 50mM Imidazole
Elution buffer - EB	50 mM Tris (pH 7.5), 300 mM NaCl, 150mM Imidazole
Gel filtration buffer - GB	50 mM Tris (pH 7.5), 150 mM NaCl
<b>Protein name : VcLMWPTP-1</b>	
Lysis Buffer -LyB	50 mM HEPES (pH 7.0), 300 mM NaCl
Wash Buffer I – WB1	50 mM HEPES (pH 7.0), 300 mM NaCl, 10mM Imidazole
Wash Buffer II – WB2	50 mM HEPES (pH 7.0), 300 mM NaCl, 20mM Imidazole
Elution buffer - EB	50 mM HEPES (pH 7.0), 300 mM NaCl, 200mM Imidazole
Gel filtration buffer - GB	50 mM MOPS (pH 7.6), 300 mM NaCl, 0.5 mM DTT
<b>Protein name : VcLMWPTP-2 and VcLMWPTP-2<sup>tr</sup></b>	
Lysis Buffer -LyB	50 mM HEPES (pH 7.0), 300 mM NaCl
Wash Buffer I – WB1	50 mM HEPES (pH 7.0), 300 mM NaCl, 10mM Imidazole
Wash Buffer II – WB2	50 mM HEPES (pH 7.0), 300 mM NaCl, 20mM Imidazole
Elution buffer - EB	50 mM HEPES pH 7.0, 300 mM NaCl, 200mM Imidazole
Gel filtration buffer - GB	50 mM HEPES (pH 7.0), 300 mM NaCl, 0.5 mM DTT
<b>Protein name : N- and C-terminal Hig-tagged mutants of VcLMWPTP-1 and N-terminal mutants of VcLMWPTP-2</b>	
Lysis Buffer -LyB	50 mM Phosphate (pH 7.0), 300 mM NaCl
Wash Buffer I – WB1	50 mM Phosphate (pH 7.0), 300 mM NaCl, 10mM Imidazole
Wash Buffer II – WB2	50 mM Phosphate (pH 7.0), 300 mM NaCl, 20mM Imidazole
Elution buffer - EB	50 mM Phosphate (pH 7.0), 300 mM NaCl, 300mM Imidazole
Gel filtration buffer - GB	50 mM Tris (pH 7.5), 150 mM NaCl

## **2.3. Characterization of protein**

### **2.3A. Dynamic Light Scattering (DLS)**

The main purpose of using DLS in crystal screening is to help the investigator understand the size distribution, stability, and aggregation state of macromolecules in solution. It can also be used to understand how experimental variables influence aggregation. DLS is primarily used to assess the aggregation state of a sample and to measure polydispersity, which is predictive of crystallizability. A microcuvet of protein solution is illuminated by laser light. The molecules in solution are undergoing Brownian motion and cause fluctuations in the scattered light intensity. This change in light intensity is measured by a detector placed at a 90° angle to the incident laser light. The translational diffusion coefficient ( $D_T$ ) is derived from these data using an auto-correlation function. A hydrodynamic radius ( $R_H$ ) of the molecules in solution can be calculated from  $D_T$ .

Measurements were done for proteins with concentration range of ~0.9-1.2mg/ml and analyzed using Dynapro equipped with temperature control using a 12µl micro cuvette. Dynamics V6 software were used to calculate the hydrodynamic radius, % polydispersity and number of oligomeric state(s) of all samples.

### **2.3B. Characterization of protein by Transmission Electron Microscopy (TEM)**

In TEM, a beam of electrons is transmitted through a specimen stained by suitable reagents, interacting with the specimen as it passes through. An image thus formed from the interaction of the electrons transmitted through the specimen is then magnified. The micrograph obtained is then used to analyze the overall structure of the sample in solution.

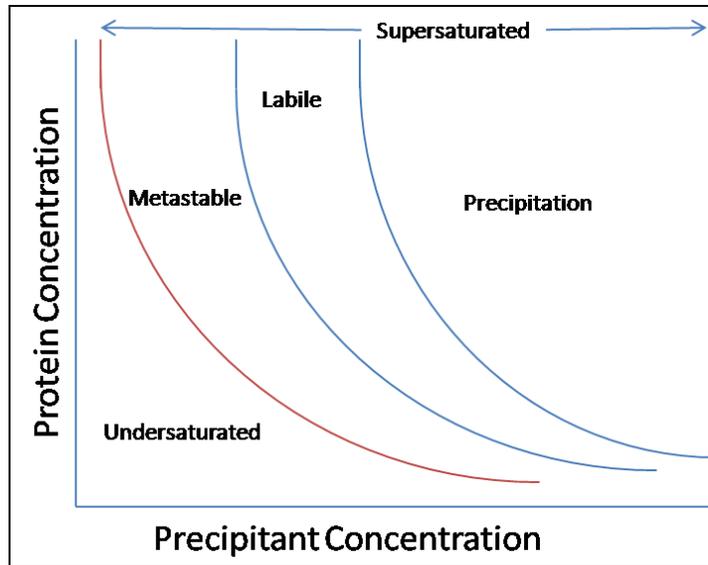
For TEM, the Vc-AcP protein was incubated with 0.8M AMS, applied to a carbon coated TEM grid and stained with 0.5% uranyl acetate. After negative staining, the grids were dried slowly before observation with transmission electron microscopy (FEI, Tecnai S-twin).

### **2.3C. Chemical Cross Linking**

Crosslinking by glutaraldehyde is often used to obtain preliminary information on the quaternary association of proteins. The method involves mixing of glutaraldehyde with the protein solution and direct detection of crosslinked products by SDS-PAGE [54]. For that a 2.3% freshly prepared solution of glutaraldehyde was added to a reaction mixture of 100  $\mu$ l containing about 50 $\mu$ g of the protein VcLMWPTP-1 and the reaction was carried out at room temperature (25°C). Samples were collected at regular intervals of 15sec., 30 sec., 1min., 2 min., 3 min., 4 min. and 5 min. and quenched by addition of 10  $\mu$ l of 1 M Tris-HCl, pH 8.0. Cross-linked proteins were analyzed through 15% SDS-PAGE.

### **2.4. Crystallization of target proteins**

The goal of crystallization is to produce a well-ordered crystal large enough to provide a diffraction pattern when exposed to X-rays. This diffraction pattern can then be analyzed to discern the protein's tertiary structure. The elaborate information that can be obtained from the three-dimensional structure of a protein is useful in a variety of ways. The interaction between protein molecules, however, is much weaker than for small (especially ionic) molecules. Finding appropriate conditions is one of the bottlenecks towards obtaining a crystal structure. The phase diagram in protein crystallization is a schematic representation of how protein and precipitate concentration are related. Protein crystals are formed in supersaturated solutions whereas low protein and/or precipitate concentrations will cause under saturation with no crystals (**Figure 4**).

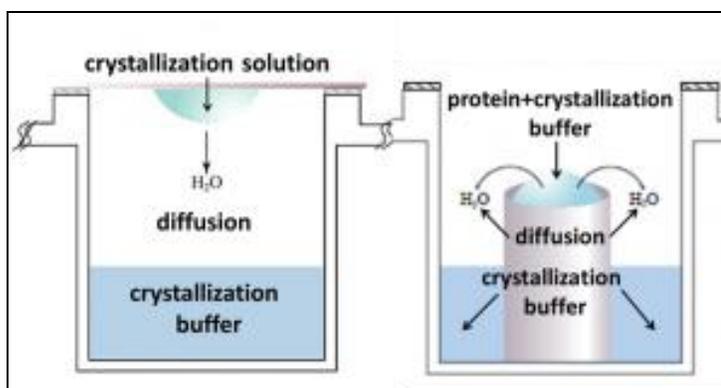


**Figure 4: The Phase Diagram in Protein Crystallization.** Schematic representation of a two-dimensional phase diagram, illustrating the change of protein molecules concentration against precipitating agent concentration. The concentration space is divided by the solubility curve into two areas corresponding to undersaturated and supersaturated state of a protein solution. The supersaturated area comprises of the metastable, nucleation and precipitation zones. The red line that separates undersaturated conditions from supersaturated is known as the solubility curve.

A benefit of determining the solubility curve is that it can help to analyze crystal growth conditions. A crystallization setup that is undersaturated or in the metastable zone will appear clear, however, the latter has the possibility of crystal growth if seeded. The phase diagram is often broken down into 4 distinct zones. Undersaturated zone is described previously. Precipitation zone is when the protein comes out of solution as an aggregate and therefore is not useful for crystallographic studies. The labile zone (or nucleation zone) is important since this is where crystal nucleation and initial growth occur. As the crystal forms the protein concentration will be depleted causing one to move from the labile to metastable zone (**Figure 4**).

Vapor diffusion is the most commonly employed method of protein crystallization. In this method, a droplet containing purified protein and precipitant are allowed to equilibrate with a larger reservoir containing similar buffers and precipitants in higher concentrations. Initially,

the droplet of protein solution contains comparatively low precipitant and protein concentrations, but as the drop and reservoir equilibrate, the precipitant and protein concentrations increase in the drop. If the appropriate crystallization solutions are used for a given protein, crystal growth will occur in the drop. Vapor diffusion can be performed in either hanging-drop or sitting-drop format (**Figure 5**). Hanging-drop apparatus involve a drop of protein solution placed on an inverted cover slip, which is then suspended above the reservoir. Sitting-drop crystallization apparatus place the drop on a pedestal that is separated from the reservoir. Both of these methods require sealing of the environment so that equilibration between the drop and reservoir can occur.



**Figure 5: Comparison of ‘hanging drop’ and ‘sitting drop’ Vapor Diffusion method**, where the green represents protein solution and the blue represents reservoir.

The proteins were concentrated to ~10-12 mg/mL in a suitable buffer and crystal growth was accomplished using hanging drop vapour diffusion method against 600-700 $\mu$ L of reservoir and incubated at either 4° or 20 °C. Hampton grid screen (PEG grid screen, Ammonium Sulfate Grid Screen, MPD grid screen), Hampton Crystal Screen, Hampton Crystal Screen 2 and Qiagen Nextal Suite were used as precipitants. The crystallization conditions are given in following **Table 9**.

**Table 9. Crystallization conditions of recombinant proteins**

Protein name	Composition of precipitant	Composition of reservoir solution	Incubating temperature
Vc-AcP	1.6M Ammonium sulfate, 0.1M Bicine (pH 9.0), 4.0% Glycerol.	2.4M Ammonium sulfate, 0.1M Bicine (pH 9.0), 4.0% Glycerol.	293K
Vc-AcP-C20R	1.6M Ammonium sulfate, 0.1M HEPES (pH 7.0)	2.0M Ammonium sulfate, 0.1M HEPES (pH 7.0)	277K
VcLMWPTP-1	2.4 M Ammonium sulfate, 0.1 M Citric acid (pH 5.0), 2.0% Glycerol	3.0 M Ammonium sulfate, 0.1 M Citric acid (pH 5.0), 4.0% Glycerol	277K
VcLMWPTP-2	0.2 M Ammonium sulfate, 30% (w/v) Polyethylene glycol 8000	0.17 M Ammonium sulfate, 25.5% (w/v) Polyethylene glycol 8000, 15.0% Glycerol	293K

## 2.5. Data Collection and Processing

Crystals of respective proteins were fished out from the crystallization drops using a 10-20 mm nylon loop (Hampton Research, Laguna Niguel, California, USA), briefly soaked in a suitable cryoprotectant solution and flash-frozen in a stream of nitrogen (Oxford Cryosystems) at 100 K. Then diffraction data set was collected using an in-house MAR Research image plate detector of diameter 345 mm and Cu-K $\alpha$  radiation generated by a Bruker–Nonius FR591 rotating-anode generator equipped with Osmic MaxFlux confocal optics and running at 50 kV and 90 mA. Different number of frames was collected with a varying crystal-to-detector distance. Detailed statistics of the no. of frames, crystal-to-detector distance (D), the exposure time for each image, the oscillation range and composition of cryo-protectants are listed in **Table 10**. Data were processed and scaled using *iMOSFLM* [55] and *SCALA* [56]. In *iMOSFLM*, diffraction images are accumulated for ‘Index’ing to determine the possible crystal lattice and estimation of mosaicity. Then ‘Cell Refinement’ task is done to allow the refinement of cell parameters, crystal orientation and mosaicity based on a post-refinement procedure [57-59] that provides more accurate values. This is applicable only when the resolution of the data is better

than  $\sim 3.5$  Å. Finally ‘Integration’ step is performed to refine various parameters. The output file is then used in *SCALA* to merge multiple observations of reflections and produces a file that contains averaged intensities for each reflection.

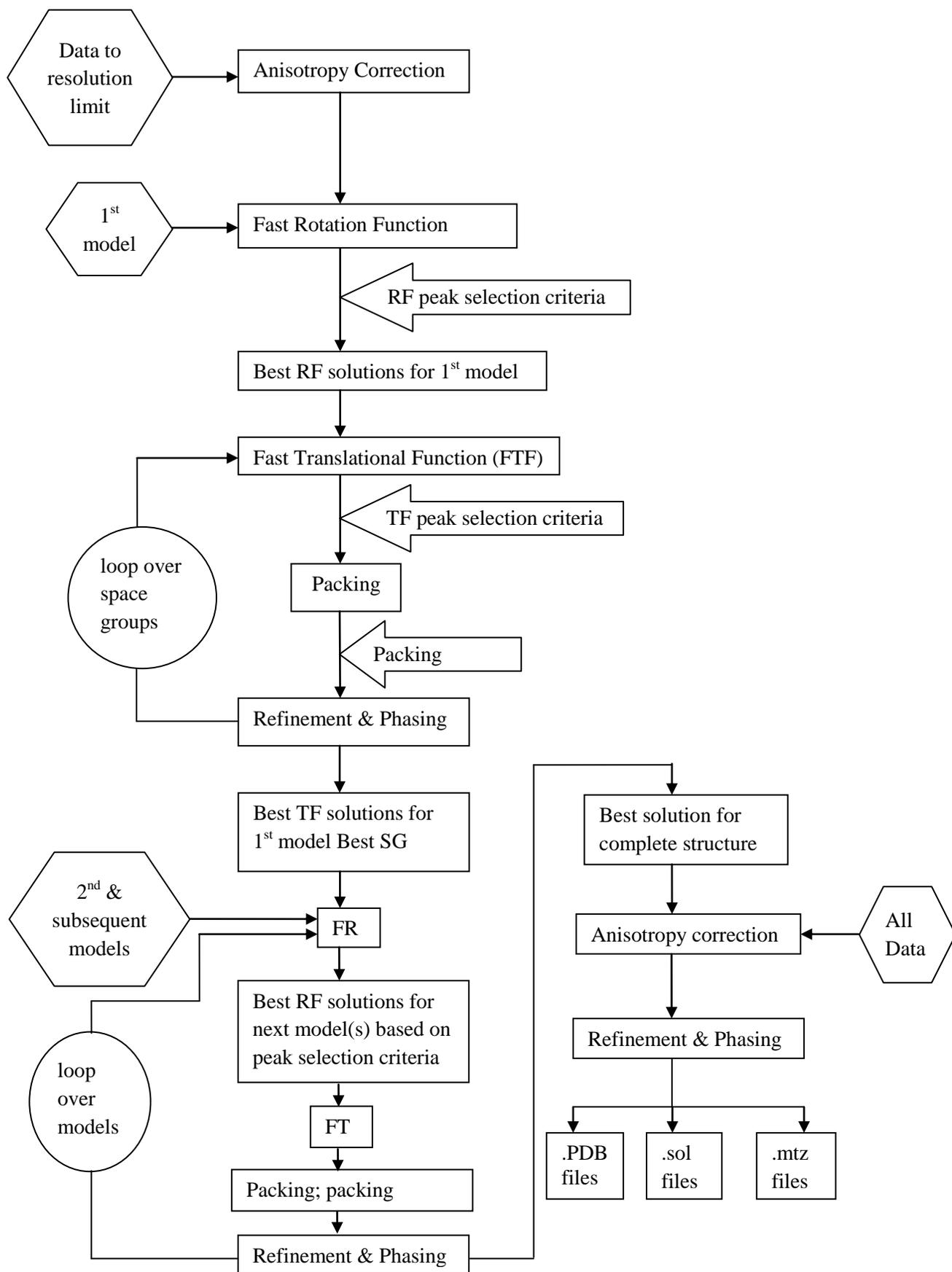
**Table 10. Data collection statistics of all crystals**

<b>Crystal name</b>	<b>Composition of cryo-protectant</b>	<b>No. of frames collected</b>	<b>D in mm.</b>
Vc-AcP	15% (v/v) glycerol with reservoir solution	150	250
Vc-AcP-C20R	1mM Molybdate and 10% (v/v) glycerol with reservoir solution	194	150
VcLMWPTP-1	10% (v/v) glycerol with reservoir solution	60	80
VcLMWPTP-2	Reservoir solution	401	260

## **2.6. Structure determination of the target proteins**

### **2.6A. Molecular Replacement (MR) method**

The structures of all of the proteins were solved using ‘Molecular Replacement’ (MR) method where the coordinates of the highest homologue to the target protein was taken as a ‘model’ for phasing. In this method the Patterson map, which is an interatomic vector map, is derived from data obtained from the unknown structure and from the structure of a previously solved homologue. Now using ‘rotation functions’ to determine the orientation of the molecule followed by a ‘translational function’ these two maps are orientated in correct position in the unit cell. Following this correctly oriented and translated phasing models, phases are calculated to derive an electron density map which can be used to build and refine in subsequent steps to get an atomic model of the unknown structure. The flow-diagram of the program *PHASER* is given below.



The model was used as an input to the 'Automated MR mode' of the program - *PHASER* [60] in CCP4 suite [61] to place the unknown structure correctly within the unit cell using 'likelihood' based method [62]. The 'Automated MR mode' combines i) anisotropy correction i.e. removing directional dependence of the diffraction data, ii) likelihood enhanced fast rotation function, iii) likelihood enhanced fast translational function, iv) packing and refinement modes for multiple search models and v) a set of possible spacegroups to automatically solve a structure by MR. The flow diagram of 'Automated MR mode' is given above.

## 2.6B. Refinement and Model Building step

The initial phases are generally quite poor and it is often difficult to start building the model. To improve the phases and also the interpretation of the electron density map refinement methods are a very important step. Refinement is used as a final step in structure solution, as well as an intermediate step to improve models and obtain improved electron density to facilitate further model rebuilding. Refinement is done through statistical adjustment of the atomic coordinates to fit the diffraction data better. As a measure of the fitting the "work R-factor" is used, that measures how far the calculated amplitudes differ from the observed amplitudes. The " $R_{\text{work}}$ " can get trapped in local minima giving the false impression of having a good model. For this reason  $R_{\text{free}}$  is mostly used to validate the refinement process. The factor is calculated from 5% of the reflections that are excluded from refinement and it therefore gives an independent measure of the refinement progress. Subsequent refinement cycles were then carried out with *Phenix refine* [63] in Phenix suite or with *REFMAC5* [64] in CCP4 suite to maximize the agreement between the model and the X-ray data. The progress in refining the model is measured by  $R_{\text{work}}$  and  $R_{\text{free}}$ . Two methods are widely used in refinement are i) maximum likelihood and ii) simulated annealing. Both methods use restraints to how an atomic model has to look like in

respect to bond distances, angles and torsions and temperature factors (B-factors). In maximum likelihood the phases are adjusted to minimize the R-factor. In simulated annealing the structure is "heated" to add randomness and slowly cooled and refined. The randomness reduces the probability of falling into a wrong local minimum. Additional methods can improve the maps. Noncrystallographic symmetry (NCS) restraints average the phases in symmetrical parts (different molecules) of the asymmetric unit (AU) and are applied in case of multiple copies of molecule in the AU. Density modification (DM) aims to adjust the density to the expectations of how it should generally look like. The solvent does not diffract normally and the electron density should therefore be zero in the solvent region. In protein often the temperature factors are averaged in all dimensions (isotropic) instead of individual (anisotropic). TLS refinement (for translation, libration (small movements) and screw-rotation of a group of atoms) can give a good approximation of anisotropy with much fewer parameters (ref.webs8). Model building was done with Coot [65].

## 2.7. Structural analysis

Average B-factors for each residue were calculated using Baverage in CCP4 [66]. The oligomeric state of the protein was analysed using the PISA webserver [67]. Structural analysis was done by PISA server [67] and PIC server [68]. Sequence alignment of was done using *ClustalO* [69,70] and Multalin [71]. Figures were prepared using Pymol [72]. The surface electrostatic potential of the structure was mapped using Chimera [73] [-10kT/e (red) to +10kT/e (blue)] and *GRASP* [74].

## **2.8. Enzyme assay**

### **2.8A. Enzyme kinetics of Vc-AcP and its mutants**

Kinetic parameters were calculated for Vc-AcP and its mutants using acetyl phosphate as the substrate as described [75]. Briefly, acetyl phosphate at a concentration range of 1-3.5mM was treated with 100 $\mu$ M Vc-AcP and its mutants and quenched with trichloroacetic acid after 15 minutes. Residual amount of acetyl phosphate was calculated from the standard curve of acetyl phosphate. For standard curve, acetyl phosphate was treated with equal volume of 2M hydroxylamine at room-temperature to form the product hydroxamic acid which was acidified by 1N hydrochloric acid and finally 5% ferric chloride solution was used to develop the color which absorbs at 510 nm.

### **2.8B. Enzyme kinetics of VcLMWPTPs and its mutants**

Kinetic parameters were calculated for VcLMWPTPs using *p*-nitrophenyl phosphate (pnpp) as the substrate as described [76]. Briefly, pnpp at a concentration range of 1-40mM was treated with 100 $\mu$ M VcLMWPTP-1, VcLMWPTP-2 and their mutants and quenched with 1N NaOH after 10 minutes. The absorbance of the product, *p*-nitrophenol, thus formed is measured at 405nm. The amount of *p*-nitrophenol was calculated from the standard curve of *p*-nitrophenol. For standard curve, stock solution of *p*-nitrophenol was diluted with 0.05N NaOH and the absorbance of the samples was measured at 405nm. To check the effect of temperature on the enzymatic activity samples were incubated at 5 $^{\circ}$ C interval in the water-bath prior to check the absorbance.

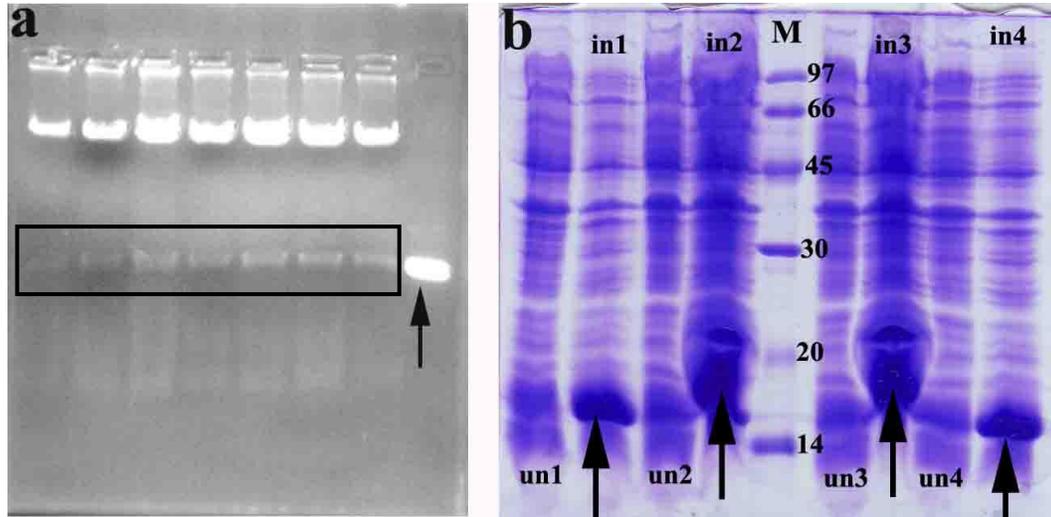
# **Chapter-3**

## **Results**

### 3.1. Acylphosphatase (Vc-AcP)

#### 3.1A. Cloning and Overexpression of Vc-AcP and its mutants

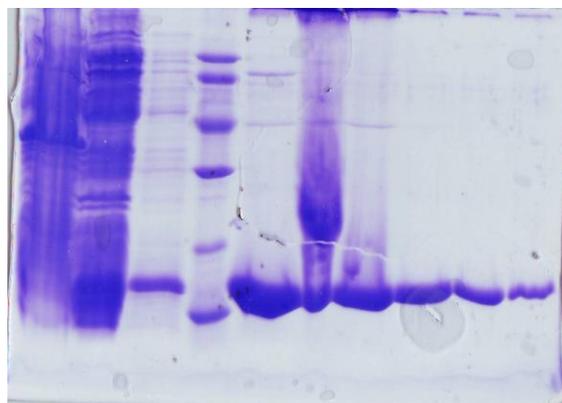
Vc-AcP and its mutant C20R was successfully cloned in pET28a(+) and overexpressed as N-terminal 6×His-tagged recombinant proteins. Cloning and overexpression profiles of the wild type are given in **Figure 6a-6b**.



**Figure 6:** (a) Clone check of Vc-AcP in 1% agarose gel after restriction digestion with *NdeI* and *BamHI*. lane 1-7: plasmid. All samples (marked with box) contained the insert, lane 8: PCR product of the insert from *V.cholerae* genomic DNA (b) 15% SDS-PAGE profile, showing the **overexpression of Vc-AcP**.lanes 1, 3, 6 and 8: whole cell lysate of the uninduced cells (marked as un-1, un-2, un-3 and un-4), lanes 2, 4, 7 and 9: whole cell lysate of the induced cells (marked as in-1, in-2, in-3 and in-4); induced band is marked with arrow, lane 5: Marker.

#### 3.1B. Ni-NTA affinity chromatographic purification of the target proteins

The proteins were purified by Ni-NTA affinity chromatography using buffers containing increasing concentration of imidazole. The buffers used for the purification and their respective elution profiles are given in **Table 8**. The elution profile of only the wild type protein is shown in **Figure 7**.



**Figure 7. 15% SDS-PAGE profile showing the Ni-NTA purification profile of Vc-AcP.** (from left to right) lane 1: Resuspended pellet after sonication and centrifugation, lane 2: flow through (FT) collected after passing the cell supernatant onto the buffer equilibrated Ni-NTA column, lane 3: sample collected after washing the column with WBI, lane 4: marker, lane 5: sample collected after washing the column with WBII, lane 6-9: Eluted fractions, lane 10: Ni-beads.

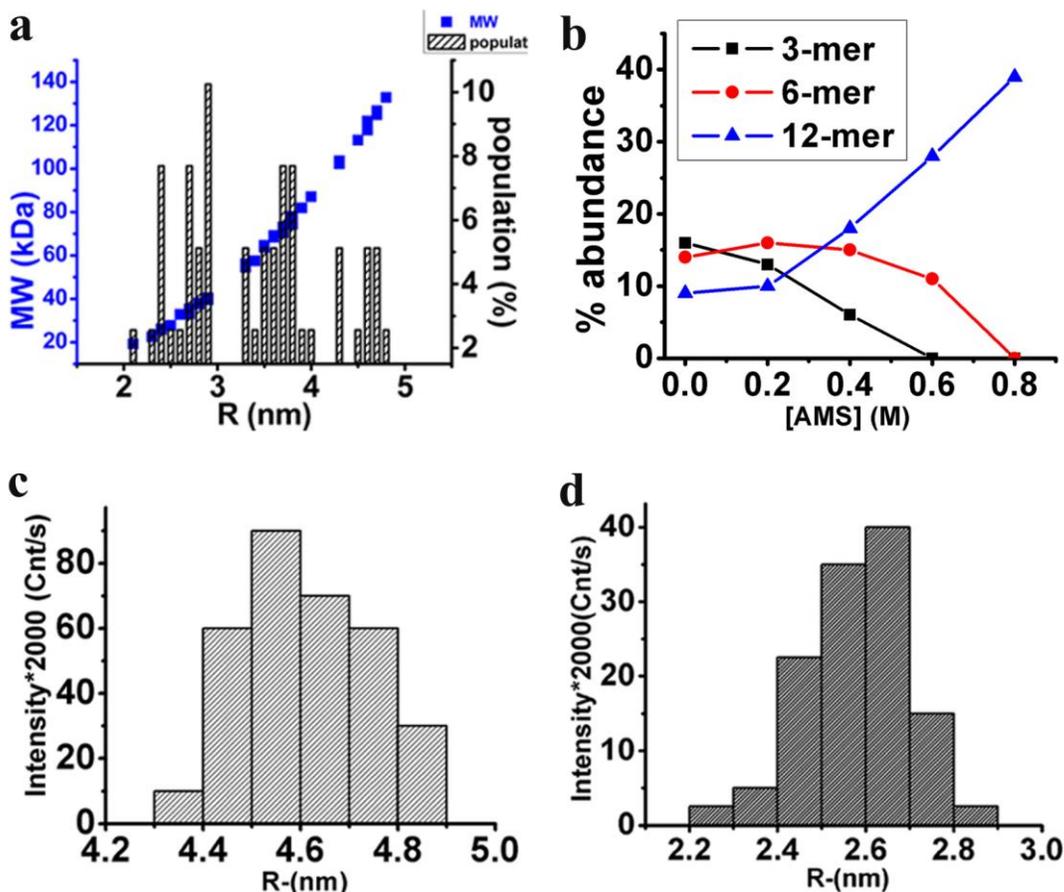
### **3.1C. Thrombin cleavage standardization and Fast Protein Liquid Chromatography (FPLC)**

Size exclusion chromatography was performed after the standardization of thrombin cleavage and the elution of the sample was checked by monitoring the OD at 280nm. Purity of the eluted samples was judged by 15% SDS-PAGE and fractions containing less contaminants were further concentrated for crystallization and other biophysical studies.

### **3.1D. Dynamic Light Scattering (DLS) profile of Vc-AcP and C20R**

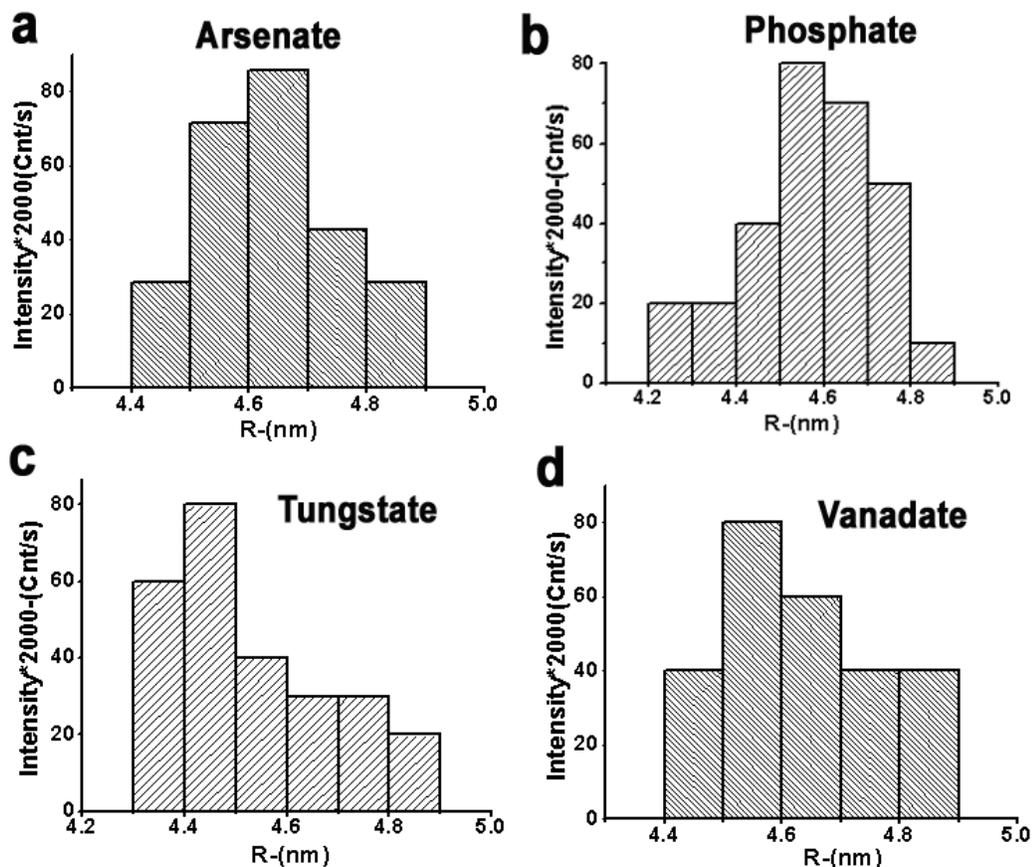
Previous reports suggested that AcP purified from other organisms forms oligomers and aggregates in solution [77]. Samples containing heterogeneous population of different types of oligomeric forms and/or aggregated protein are not suitable for successful crystallization. Therefore the oligomeric state of Vc-AcP was checked through DLS experiment prior to crystallization trials. The data indeed showed a multimodal heterogeneous population of Vc-AcP consisting of three different oligomeric species for Vc-AcP. This heterogeneous solution of Vc-AcP was not suitable for successful crystallization and to overcome this hurdle, extensive screening of pH and additives were performed. Interestingly with increasing concentration of

oxyanions (sulfate, vanadate, molybdate, tungstate, phosphate) gradual conversion of multimodal to monomodal population of Vc-AcP was observed (**Figure 8a-8b**). This oxyanion concentration dependent change in oligomerisation data of Vc-AcP, obtained from DLS experiments, was used for subsequent Vc-AcP crystallization.



**Figure 8. Dynamic light scattering profiles of Vc-AcP and Vc-AcP-C20R.** (a) DLS profile of Vc-AcP showing ‘%’ population of three different oligomeric states with increasing molecular weight (MW-kDa) and hydrodynamic radius (R-nm). The data were taken without any additives. (b) % abundance of three different oligomeric states (trimer, hexamer and dodecamers) of Vc-AcP in presence of increasing concentration of ammonium sulfate (AMS). All of the three populations exists upto 0.4M AMS, though with increasing concentration of AMS, relative population of trimer declines and that of the hexamer remains almost steady. At 0.8M AMS the solution turns into a monomodal population containing only dodecameric Vc-AcP.(c) DLS profile of Vc-AcP incubated with 0.8M AMS having a  $R_H$  value of  $(4.6 \pm 0.3 \text{ nm})$ .(d) DLS profile of Vc-AcP-C20R mutant showing a monomodal population with hydrodynamic radius  $R_H$   $(2.5 \pm 0.3 \text{ nm})$ .

In contrast to the DLS data of the wild type protein, Vc-AcP-C20R mutant showed the presence of a single species in solution. The  $R_H$  of the sample was similar to one of the oligomeric states of the wild type (Figure 8c-8d). Other tetrahedral anions like sulfate, vanadate, phosphate, tungstate, and arsenate, above a threshold concentration, were observed to be suitable to shift the multimodal population to a monomodal state that corresponds to the highest MW and  $R_H$  (Figure 9a-9d).



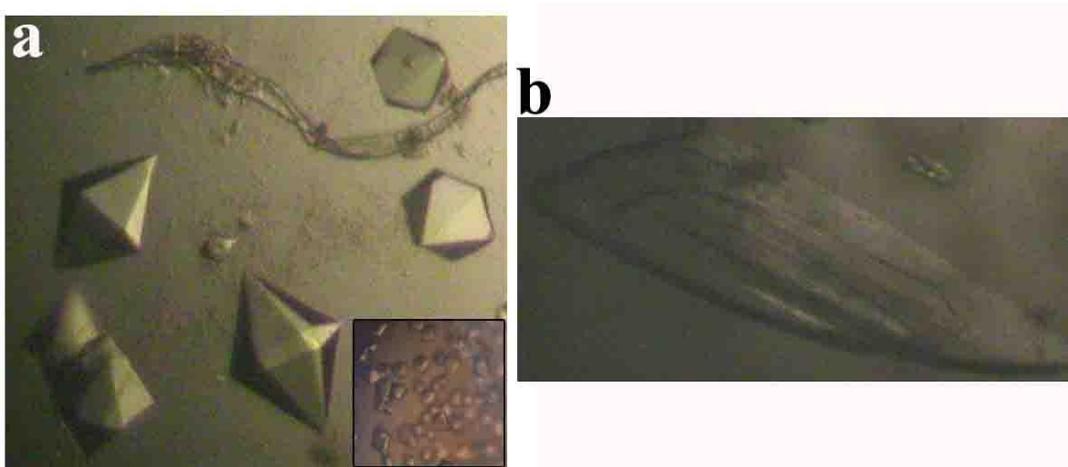
**Figure 9. Dynamic Light Scattering profiles of Vc-AcP with different oxyanions.** A monomodal population of Vc-AcP with the hydrodynamic radius ( $R_H$ ) of around  $(4.6 \pm 0.3 \text{ nm})$  at (a) 0.85M sodium arsenate, (b) 0.75M sodium phosphate, (c) 0.87M sodium tungstate and (d) 0.68M sodium vanadate is obtained

### 3.1E. Crystallization of Vc-AcP and Vc-AcP-C20R

Vc-AcP was crystallized using Ammonium Sulfate Grid Screen from Hampton as precipitant. Typically Vc-AcP protein ( $\sim 10 \text{ mg/ml}$ ) was crystallized using 1.6M ammonium sulfate, 0.1M BICINE, pH 9.0 as precipitant and a reservoir containing 2.4M Ammonium sulfate,

0.1M Bicine (pH 9.0) produced showers of small octahedral shaped crystals within a couple of days at 293K (**Figure 10a inset**). Quality of the crystals was improved by varying the protein and/or precipitant concentration, along with the variation of different additives e.g. glycerol, dioxan. Best crystals (**Figure 10a**) appeared when 2 $\mu$ l of the protein was mixed with 2  $\mu$ l of the precipitant containing glycerol (listed in **Table 9**) and incubated at 293K.

Vc-AcP-C20R was crystallized using the precipitant from Ammonium Sulfate Grid Screen from Hampton, condition containing 1.6M Ammonium sulfate, 0.1M HEPES (pH 7.0) against the reservoir solution of 2.0M Ammonium sulfate, 0.1M HEPES (pH 7.0). A single wing-shaped crystal (**Figure 10b**) appeared within a couple of days at 277K. The precipitant condition is (listed in **Table 9**).

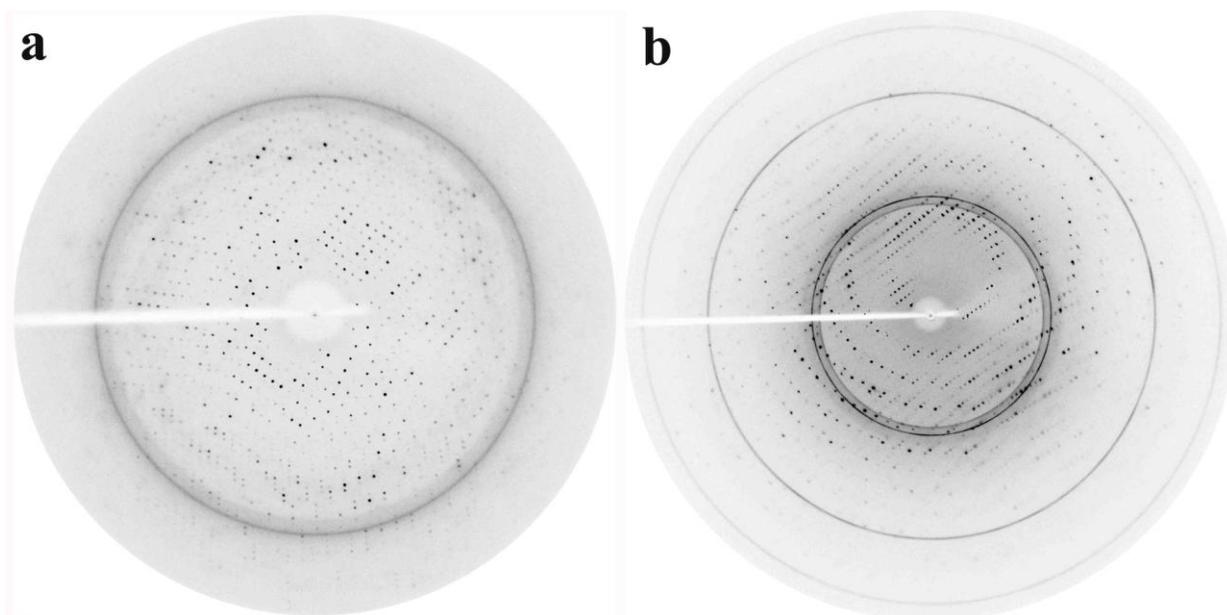


**Figure 10. Crystals of Vc-AcP and Vc-AcP-C20R.** (a) (Inset) Initial showers of Vc-AcP crystals appeared in conditions containing 1.6M Ammonium sulfate, 0.1M Bicine (pH 9.0) at 293 K. Crystals of Vc-AcP appeared after optimization of initial condition and addition of 4% glycerol which diffracted to higher resolution. (b) single crystal of Vc-AcP-C20R obtained using precipitant containing 1.6M Ammonium sulfate, 0.1M HEPES (pH 7.0) at 277K.

### 3.1F. Data collection and processing of Vc-AcP and Vc-AcP-C20R

For Vc-AcP-C20R crystal a diffraction data set with 194 frames was collected to 1.9 $\text{\AA}$  resolution (**Figure 11a**) keeping the crystal to detector distance 150mm.

The diffraction data set of Vc-AcP crystal with 150 frames was collected to 3.1 Å resolution (**Figure 11b**) with a crystal to detector distance 250mm. In both cases Cu-K $\alpha$  radiation generated by a Bruker–Nonius FR591 rotating-anode generator equipped with Osmic MaxFlux confocal optics running at 50 kV and 90 mA was used as X-ray source and oscillation range was set at 1° per image.



**Figure 11.** Diffraction pattern of (a)Vc-AcP and (b)Vc-AcP-C20R crystals upon exposure to X-rays.

Both the data was processed and scaled using *iMOSFLM* [55] excluding the ice ring followed by *SCALA* [56]. No translational NCS and twinning fractions were observed. Details of data collection and processing statistics for Vc-AcP-C20R and Vc-AcP are given in **Table 11**.

**Table 11: Data collection and processing statistics of Vc-AcP-C20R and Vc-AcP**

Protein name		Vc-AcP-C20R	Vc-AcP
Space group		P2 <sub>1</sub>	P3 <sub>1</sub>
Cell dimension	a, b, c	39.71, 81.3, 40.53	104.94, 104.94, 147.811
	$\alpha, \beta, \gamma$ (°)	90.0, 90.52, 90.0	90.0, 90.0, 120.0
Resolution (Å)*		39.71-1.964 (2.034-1.964)	34.23-3.10 (3.211-3.10)
R <sub>merge</sub>		0.062 (0.205)	0.089 (0.368)
I / $\sigma$ (I)		12.5 (5.5)	5.8 (1.8)
Completeness (%)		93.5 (100)	96.8 (96.9)
Redundancy		3.8 (3.7)	1.8 (1.7)
Total no. of observations		65296	59085
No. of unique reflections		17237	31972
Mosaicity		0.62	0.78
Matthew's coefficient (V <sub>m</sub> ) (ANGSTROMS**3/DA)		2.1	3.9
% solvent content		43.6	68.6

### 3.1G. Structure solution of Vc-AcP-C20R

For phasing, the coordinates of *Sulfolobus solfataricus* acylphosphatase (PDB ID: 2BJD) was modified by the program *CHAINSAW* [78] in CCP4 using a sequence alignment file and residues that differs with Vc-AcP were truncated to C <sup>$\beta$</sup> . This modified PDB file was used as search model for molecular replacement using Phaser [60] in CCP4 [61]. Phaser identified three copies of the search molecule with final TFZ= 11.4 and LLG=191 (**Table 12**) that corresponds to the Mathews coefficient (V<sub>m</sub>) of 2.1 (solvent content 43.6%). The model was improved with few cycle of rebuilding with Coot [65] and refined with Phenix refine [63]. TLS refinement was done during the final stages of refinement to improve the anisotropic displacements (e.g. motions) of the atoms of the model at medium to low resolution. [79].

**Table 12. Improvement of Phaser statistics of Vc-AcP-C20R and Vc-AcP**

Vc-AcP-C20R				
	RFZ	TFZ	LLG	PAK
Molecule 1	7.1	6.3	59.0	0
Molecule 2	4.8	8.9	52.0	0
Molecule 3	5.6	11.4	191	0
Vc-AcP				
Ensemble 1	3.0	5.4	55	0
Ensemble 2	3.9	8.4	226	0
Ensemble 3	5.1	11.5	457	2
Ensemble 4	7.9	34.8	1460	3

where, RFZ and TFZ denotes the Z-score of Rotation and Translational function respectively. Signal-to-noise is judged by Z-score which is computed by comparing the LLG values from the rotation/translational search with LLG for a set of random rotation/translation. PAK denotes no. of packing clashes. By default upto 5% of C<sup>α</sup> atoms for proteins are allowed to be involved in clashes. LLG is the log-likelihood gain which is an indication of how much better the solution is compared to a random solution.

### 3.1H. Structure solution of Vc-AcP

MR trials using the co-ordinates of Vc-AcP-C20R monomer did not produce any clear-cut solution. However, the coordinates of the trimeric Vc-AcP-C20R yielded a clear solution in molecular replacement using *PHASER* [60] in CCP4 [61]. Four Vc-AcP-C20R trimer or 12 monomers in the asymmetric unit produced final TFZ=34.2 and LLG=1460 (**Table 12**) with V<sub>m</sub> of 3.92 (solvent content 68.61%). Model building was done using the 12-fold NCS averaged map with Coot [65] and refinement was carried out with *Phenix refine* [63].

**Table 13: Data collection, phasing and refinement statistics of Vc-AcP and Vc-AcP-C20R**

Protein name		Vc-AcP	Vc-AcP-C20R
Space group		P3 <sub>1</sub>	P2 <sub>1</sub>
Cell dimension	a, b, c	104.94, 104.94, 147.81	39.71, 81.3, 40.53
	$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 120.0	90.0, 90.52, 90.0
Resolution (Å)*		34.23-3.10 (3.21-3.10)	39.71-1.96 (2.03-1.96)
R <sub>merge</sub>		0.089 (0.368)	0.062 (0.205)
I / $\sigma$ (I)		5.8 (1.8)	12.5 (5.5)
Completeness (%)		96.8 (96.9)	93.5 (100)
Redundancy		1.8 (1.7)	3.8 (3.7)
Refinement			
Resolution (Å)		34.23-3.1	39.71-1.964
R <sub>cryst</sub> / R <sub>free</sub>		22.09 / 28.91	18.74/24.87
No. of Atoms		8548	2340
Protein		8333	2061
Ligand		215	30
Water		0	249
B-factor (Å <sup>2</sup> )	Average	59.50	21.4
	Protein	58.9	20.5
	Ligand	79.64	37.48
	Water		26.98
R.M.S. deviations	Bond length (Å)	0.009	0.027
	Bond angles (°)	1.34	2.03
Ramachandran statistics (%)	Most favoured	92	97.0
	Additionally allowed	6.2	3
	Disallowed	1.8	0

\*Numbers in parentheses refer to the highest resolution shell with all data collected from a single crystal.

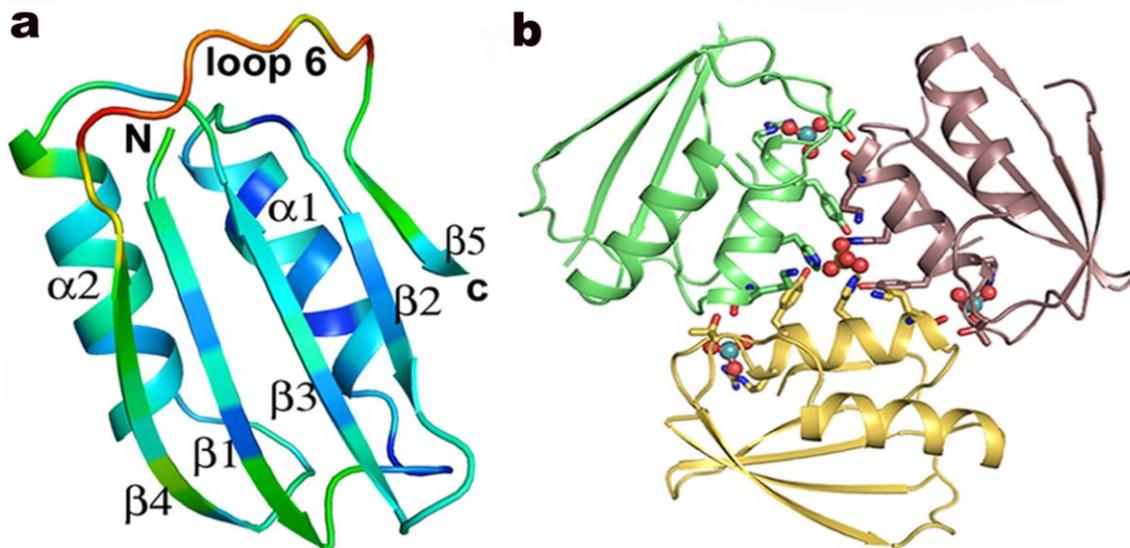
We could locate 43 sulfates in the electron density map which were placed in the final cycles of refinement. TLS refinement [79] was done towards the end of the refinement to approximate the observed atomic displacement parameters ("B values") obtained in crystallographic refinement. Refinement statistics of both the protein crystals, Vc-AcP and Vc-AcP-C20R are given in **Table 13**. Both the structures are deposited in the PDB with ID 4HI1 and 4HI2 respectively.

### **3.1I. Structural analysis of Vc-AcP-C20R: Overall structure of Vc-AcP-C20R monomer**

The crystals of Vc-AcP-C20R grew using AMS as precipitant and the structure has been solved using the 1.96Å diffraction data (**Table 13**). Vc-AcP-C20R crystals belong to space group  $P2_1$  and the asymmetric unit contains three monomers each having a molybdate ion at its active site. Vc-AcP-C20R monomer exhibits the characteristic ferredoxin-like fold having 2  $\alpha$ -helices on one side of a 4 stranded anti-parallel twisted  $\beta$ -sheet [31, 80-82]. Apart from the 4-stranded anti-parallel twisted  $\beta$ -sheet, one extra  $\beta$ -strand ( $\beta_6$ ) parallel to  $\beta_2$  is present. This C-terminal  $\beta$ -strand is the signature of acylphosphatase fold (**Figure 12a**).

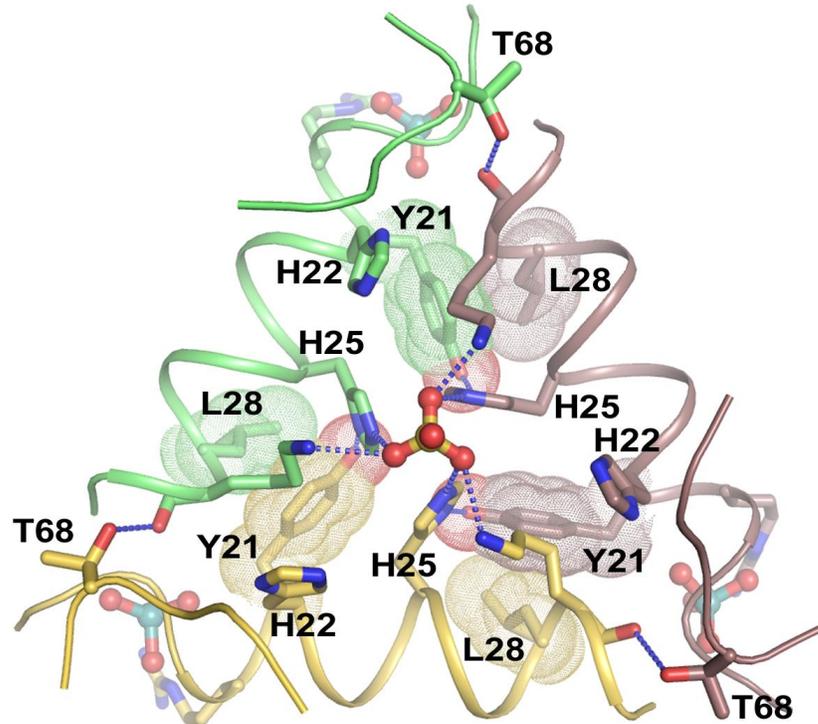
### **3.1J. Crystal structure of Vc-AcP-C20R reveals a trimeric assembly involving a set of unique residues**

Crystal structure of Vc-AcP-C20R exhibits a trimeric assembly where three monomers are arranged in a triangular fashion (**Figure 12b**). Interestingly, the size and mass of this trimeric assembly matches well with that of the oligomeric state obtained in DLS experiments (**Figure 8d**). The trimeric assembly is stabilized by a set of residues which are unique for Vc-AcP. The residues (Y21, H22, H25, L28, and K29) involved in the trimerisation are contributed exclusively by helix  $\alpha_1$  of each monomer and the trimerisation can be termed as helix mediated (**Figure 13**).



**Figure 12. Structure of monomeric AcP and Vc-AcP-C20R.** (a) Cartoon representation of AcP monomer exhibiting ferredoxin like fold with two helices ( $\alpha 1$ ,  $\alpha 2$ ) and four stranded antiparallel  $\beta$ -sheet. One extra  $\beta$ -strand  $\beta 5$  at the C-terminal, connected by loop6, is the characteristic of AcP family of proteins. The monomer is coloured based on B-factor (from blue to red) where blue being the least flexible and red being most flexible region. (b) The trimeric assembly of Vc-AcP-C20R (PDB ID : 4HI1) as seen in the crystal structure; three chains (yellow, green and brown) of Vc-AcP-C20R are shown in cartoon, central trimerisation region is shown in sticks and central sulphate and active site molybdate ions are shown in ball-and-stick.

At the interior of the triangle, these residues are arranged in three layers and pack themselves neatly with each other. The top layer is constituted by H22 and K29, the middle layer by H25 and the bottom layer by Y21 and L28. A sulfate ion is present at the center and three oxygen atoms of the tetrahedral sulfate ion interact with K29 and H25 side chains and favor the formation of the trimeric assembly (**Figure 13**). H22 of the top layer, H25 of the middle layer and Y21 of the bottom layer form an aromatic cage providing the base of the sulfate ion. Y21 plays a crucial role in the stabilization of the assembly as the aromatic ring of Y21 makes hydrophobic interactions



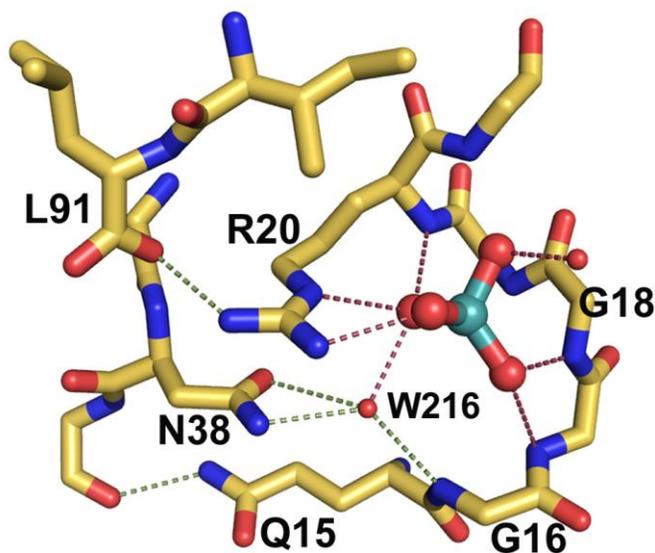
**Figure 13. Molecular details of Vc-AcP-C20R.** Zoomed view of the triangular helix mediated trimeric assembly. The corners of the triangle are occupied by three molybdate ions (molybdate in green and oxygen in red color) and the corners are clamped by the H-bonding between main chain carbonyl of K29 and side chain of T68. The residues H22, K29 (top layer; shown in sticks); H25 (middle layer; shown in sticks) and Y21, L28 (bottom layer; shown in dot) from  $\alpha 1$  play a vital role to stabilize the trimeric assembly. A sulfate ion (sulfur in yellow and oxygen in red) is captured at the centre by positively charged K29 and H25 on top of an aromatic core formed by Y21 and H25 of the three monomers. The sulfate and molybdate ions form an imaginary pyramid where the sulfate ion occupies the vertex.

with L28, K29 and H25 of the adjacent monomer while its polar head makes H-bond with H25 of the adjacent monomer. The side-chain of T68 interacts with the adjacent monomer through H-bond involving the main-chain carbonyl of K29.

### 3.1K. Active site structure of Vc-AcP-C20R

Each monomer binds a molybdate ion on the opposite face of the assembly where sulphate ion binds. These three molybdate ions are captured at the positively charged amide environment of  $\alpha 1$  and are further stabilized through strong interaction with the catalytic R20 side chain (**Figure**

**14).** A network of water molecules is present at the active site of Vc-AcP-C20R and one of them is closely placed between the molybdate ion and N38 residue. The water molecule (W216, W204 and W215 corresponding to chain A, B and C respectively) is also stabilized by H-bonding to the main chain amide of G16. Conserved nature, close proximity with N38 (2.9 Å for Chain A and B and 2.8 Å for chain C) and molybdate ion along with the strategic location probably qualifies this water molecule to act as the polarisable moiety required to initiate the nucleophilic attack.

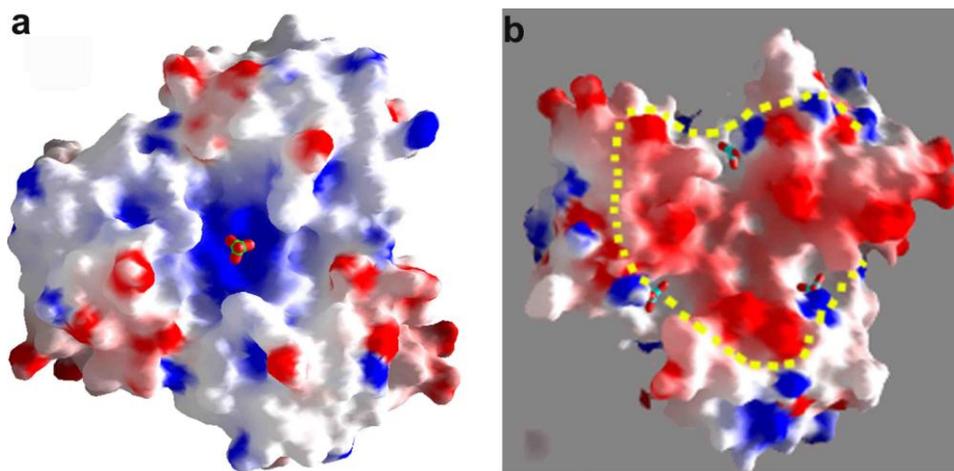


**Figure 14. Molecular details of the active site Vc-AcP-C20R.** Stabilization of molybdate ion through interaction with the R20 side chain and also with the main chain amides of G17, G18, R20 and Y21. Close proximity of the catalytic water (W216 shown in red dot) with N38, molybdate ion and the main chain of G16 is indicated.

### 3.1L. Nature of electrostatic surface of Vc-AcP-C20R

Surface presentation of the face of Vc-AcP-C20R trimer that contains the deeply buried sulfate ion is shown in **Figure 15a**. Sulfate ion occupies the vertex of a positively charged center created by H25 and K29 and makes strong interactions with them. The molybdate ions occupy the active site at the other face of the trimer (**Figure 15b**). One sulfate ion and three molybdate ions form an imaginary pyramid where the sulfate is placed at the positively charged vertex. Each

monomer of Vc-AcP-C20R trimer is related through a three-fold rotation axis that passes through the vertex sulfate. The molybdate ions at the base of the pyramid are  $\sim 23\text{\AA}$  apart from each other while the slant height of the pyramid is  $\sim 16\text{\AA}$ .

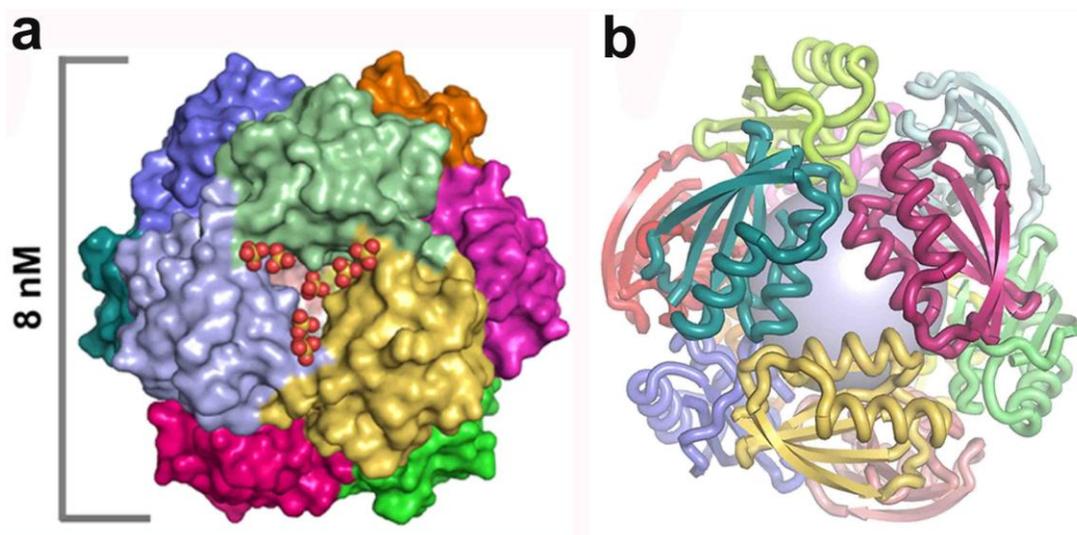


**Figure 15: Electrostatic surface of Vc-AcP-C20R** (a) Electrostatic surface of the trimeric assembly showing a strong positively charged (blue) patch where the sulfate ion (shown in ball and stick) binds to. (b) Opposite face of the trimer has a predominant negative charge. Three molybdate ions (shown in ball and stick) occupying the positively charged amide pockets near the catalytic R20.

### 3.1M. Structural analysis of Vc-AcP: Overall crystal structure of Vc-AcP

Crystals of Vc-AcP grew using AMS as precipitant which belong to space group  $P3_1$  and diffracted to a resolution of  $3.1\text{\AA}$ . The structure was solved using the trimeric coordinates of Vc-AcP-C20R where the Arg20 was truncated to Ala. Four trimers or twelve Vc-AcP monomers are arranged within the asymmetric unit to form a hollow octahedral cage-like structure having an outer and inner diameter of  $\sim 8\text{nm}$  and  $\sim 4.5\text{nm}$  respectively (**Figure 16a-16b**). The cage-like self assembly of Vc-AcP is also supported by PISA [67] where, upon cage formation, Vc-AcP buries about  $\sim 26040\text{\AA}^2$  surface area which is about 40% of the total surface area, with a  $\Delta G_{\text{diss}}$  value of  $105.5\text{kcal/mol}$  depicting the high stability of the cage. Detailed numerical values are given in

**Table 14.** Altogether 43 sulfate ions are located at several junction points of the monomers including the nonfunctional active site of each monomer. The sulfate ions play key roles to stabilize the self-assembled dodecameric cage.



**Figure 16.** Dodecameric assembly of Vc-AcP (PDB ID: 4HI2) (a) Solid surface view with sulfate ions (shown in sphere) showing nine monomers each in different colors. (b) Cartoon representation showing the cage structure with a hollow interior which is demonstrated as a grey sphere inside the cage.

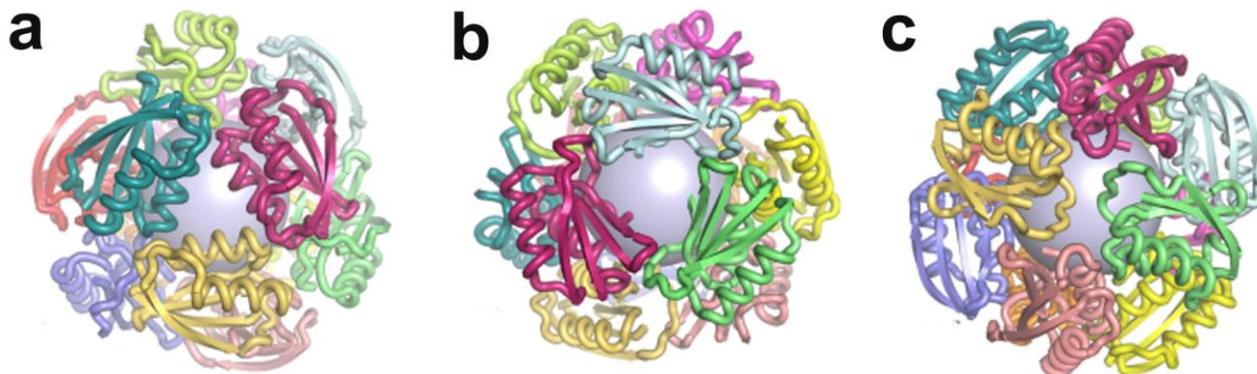
**Table 14: Details of Vc-AcP and Vc-AcP-C20R surface areas obtained from the PISA server**

Protein name	Total Accessible Surface area [SA] (sq.Å)	Buried Surface area [BSA] (sq.Å)	$\Delta G^{\text{int}}$ , kcal/mol	$\Delta G^{\text{diss}}$ , kcal/mol	% surface area buried upon oligomerisation/cage formation
Vc-AcP	43050	26050	-685.4	105.6	60
Vc-AcP-C20R	12796.5	3389.2	-46.9	1.5	26

Where,  $\Delta G_{\text{int}}$  indicates the solvation free energy gain upon formation of the assembly, in kCal/M. The value is calculated as difference in total solvation energies of isolated and assembled structures. This value does not include the effect of satisfied hydrogen bonds and salt bridges across the assembly's interfaces and  $\Delta G_{\text{diss}}$  indicates the free energy of assembly dissociation, in kCal/M. The free energy of dissociation corresponds to the free energy difference between dissociated and associated states. Positive values of  $\Delta G_{\text{diss}}$  indicate that an external driving force should be applied in order to dissociate the assembly, therefore assemblies with  $\Delta G_{\text{diss}} > 0$  are thermodynamically stable.

### 3.1N. Arrangement and symmetry of the dodecameric protein cage

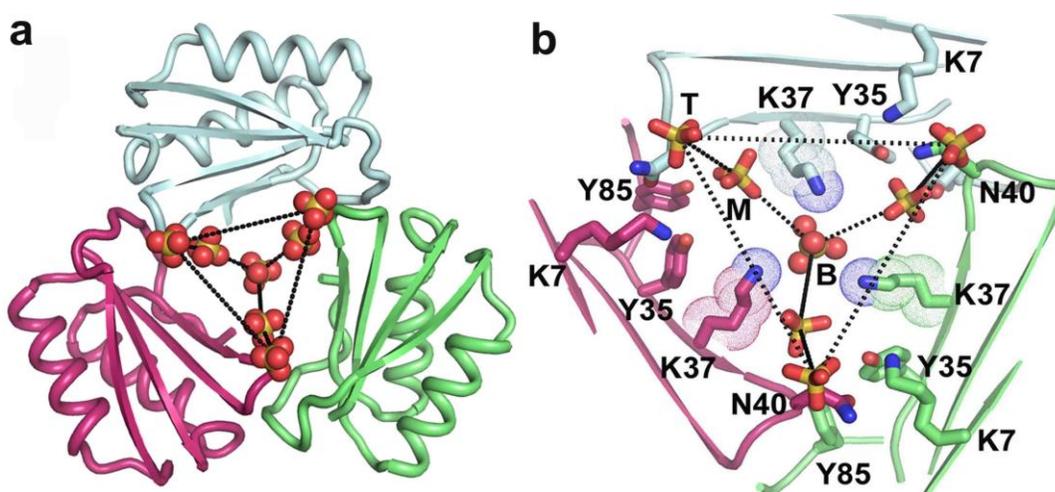
In this dodecameric cage structure, two types of trimeric building blocks namely the ‘Helix-mediated trimer’ (HT) and the ‘Sheet-mediated trimer’ (ST) can be envisaged. The HT closely resembles the Vc-AcP-C20R assembly, the only difference being the presence of sulfate ions instead of molybdate ions at the active site. The other trimer, ST, as the name bears, is formed and stabilized by the residues coming from the twisted  $\beta$ -sheet portion of the monomer. A closer look at the cage structure reveals that each monomer is utilized twice in making the HT and ST assembly. The  $\alpha$ -helices take part in HT formation while the twisted  $\beta$ -sheet of the same monomer takes part in ST formation (**Figure 17a-17b**).



**Figure 17. Different faces of Vc-AcP.** (a) HT face, (b) ST face and (c) LT face of the spherically octahedral hollow protein cage where the Vc-AcP-C20R like helix mediated trimeric assembly corresponds to the HT face, each monomer of the three different HT associates through their  $\beta$ -sheets to forms the ST face whereas LT represents the junction point of two HT and two ST faces. A grey sphere is drawn inside to depict the hollow interior.

Stabilization of the ST is dependent on the packing of the C-terminal flexible loop-6 of one chain with the residues from loop-1 and loop-3 of another. Polar side chains of H13 of  $\alpha$ 1, N38 of  $\beta$ 2, N40 of loop-3 of one chain interact with the charged residues of loop-6 (E82, H83, S79) of the other at each corner of the ST. The main chain of N40 is also involved in the direct stabilization of

the ST and interacts with the main chain of loop-6 carrying residues E80 and L81. Again, the main chain and side chain of the residues of loop-6 of one chain interacts with residues of loop-1 and  $\beta 2$  of another. Apart from the aforementioned interactions, K7 of  $\beta 1$  of one chain interacts with N40 of another. The sulfate ions of 7-sulfate-inverted pyramid play a crucial role in stabilizing the ST through several polar/ionic interactions at the junction of the monomers forming the HT (**Figure 18a**). The top (T) and middle (M) layers of the inverted 7-sulfate pyramid at the ST are held by the interactions with the main chain and side chain of Asn (N38, N40, N41) residues of  $\beta 2$  and with Y35, and Y85. The vertex sulfate of 7-sulfate-inverted pyramid is held only by K37 of the  $\beta 2$  of each chain (**Figure 18b**).

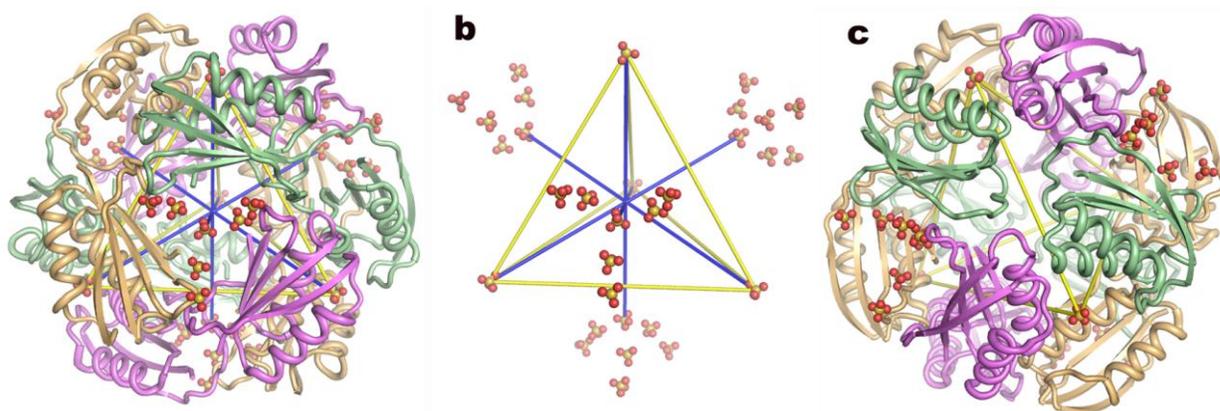


**Figure 18. Molecular details of the ST-face and sulfate assembly.** (a) A 7-sulfate-inverted pyramid formed at the center of ST face (b) Zoomed view of the funnel like ST face having a 7-sulfate-inverted pyramid at its center. Seven sulfates are arranged in three layers top (T) middle (M) and bottom (B). The vertex sulfate (B) is stabilized by K37 (shown in stick and dot) from each chain whereas the sulfates of the middle (M) and top (T) layers are held by N40 and N41 (shown in stick) of one monomer and K7, Y35 and Y85 of the adjacent monomer (shown in stick) and thereby stabilizing the ST face.

The sulfate ions of HT, occupying the active site of Vc-AcP, position themselves at the interface of the cage. The ST assembly assumes a 3-blade, 4-stranded  $\beta$ -propeller like fold where the hollow funnel-like center is occupied by a 7-sulfate-stitched inverted pyramid (**Figure 18a-**

**18b).** Unlike 4-sulfate pyramid of HT, one extra layer of three sulfate moieties exists in case of 7-sulfate-stitched inverted pyramid at the ST. If the 4-sulfate assembly of each HT is joined by an imaginary line, it reveals a tetrahedral shape where the HT is situated at the vertices. The occurrence of such flat vertices makes it a ‘truncated’ tetrahedron. A three-fold rotation axis, like HT of Vc-AcP or the trimeric assembly of Vc-AcP-C20R, passes through the flat vertices. All sulfates of the ST are located at the exterior surface of the cage. The vertex sulfate is situated at a depth of  $\sim 11\text{\AA}$  from the top layer surface of the base and slant height of the pyramid is  $\sim 14\text{\AA}$ . The side of the base is  $\sim 16.7\text{\AA}$ , slightly shorter than that of the HT-sulfate-pyramid. The side of the base of second layer is  $\sim 10.5\text{\AA}$  and slant height from this layer being  $\sim 7.5\text{\AA}$ , depth of core sulfate from this layer is  $\sim 5.5\text{\AA}$ . The first two layers of sulfates are closer ( $\sim 6.5\text{\AA}$ ) to each other than the distance of the middle layer to the core sulfate ( $\sim 7.5\text{\AA}$ ).

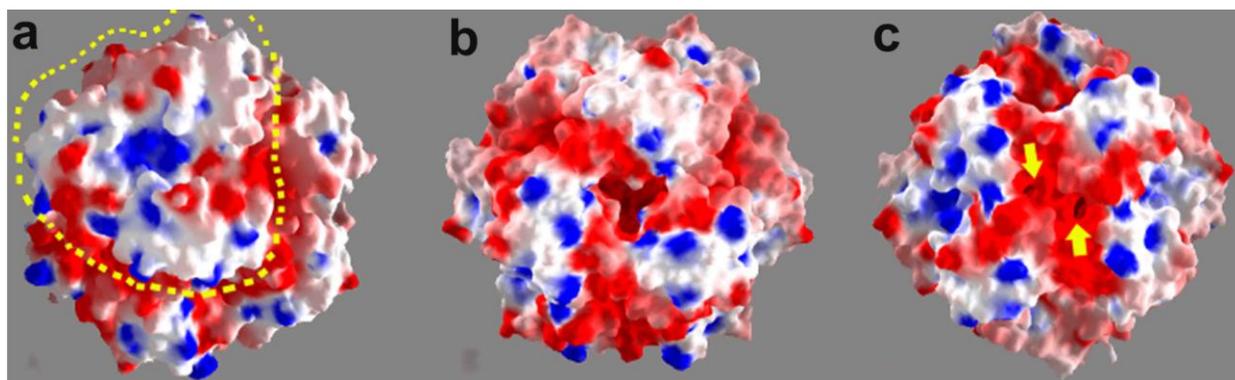
The hollow protein cage has tetrahedral symmetry (T<sub>32</sub>) where the four 3-fold rotational axes pass through the center of two opposite faces and three 2-fold axes passing through the opposite edges of the tetrahedron connecting the vertex sulfates of two oppositely positioned HT faces. The symmetry elements are shown in **Figure 19**.



**Figure 19. Symmetry elements in Vc-AcP protein cage with tetrahedral symmetry.** (a)-(b) Tetrahedral (shown in red) arrangement of Vc-AcP monomers with four 3-fold rotation axes (blue line) (c) three 2-fold axes passing through the middle of two opposite edges (yellow line).

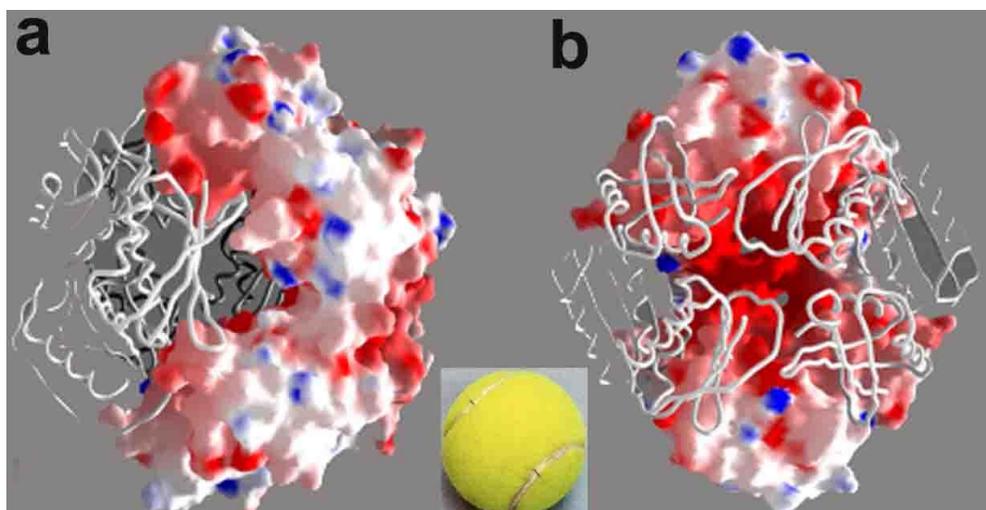
### 3.10. Nature of electrostatic surface of Vc-AcP

The center of HT of Vc-AcP cage is positively charged similar to that observed in Vc-AcP-C20R (**Figure 20a**). In contrast, mosaic distribution of charged patches can be seen at the ST surface along with a hole. This hole is connected by three negatively charged canyons, suitable enough to allow diffusion of small molecules (**Figure 20b**). The junction point of the two HT, at the LT face (**Figure 17c**), is also dominated by a negatively charged ‘X’ shaped deep canyon having smaller sized holes (**Figure 20c**).



**Figure 20. Electrostatic surface view of different faces of Vc-AcP.** (a) Mosaic distribution of charged (positive, negative and neutral) patches is present all over the exterior surface. HT face (marked with yellow dotted line) is similar to the trimeric assembly of Vc-AcP-C20R with a positively charged (blue) patch at the centre. (b) ST face contains a shrunken triangular hole at its center connected by three negative charged canyons and (c) LT face is dominated by a ‘X’ shaped negatively charged (red) canyon containing two small holes (marked with yellow arrow).

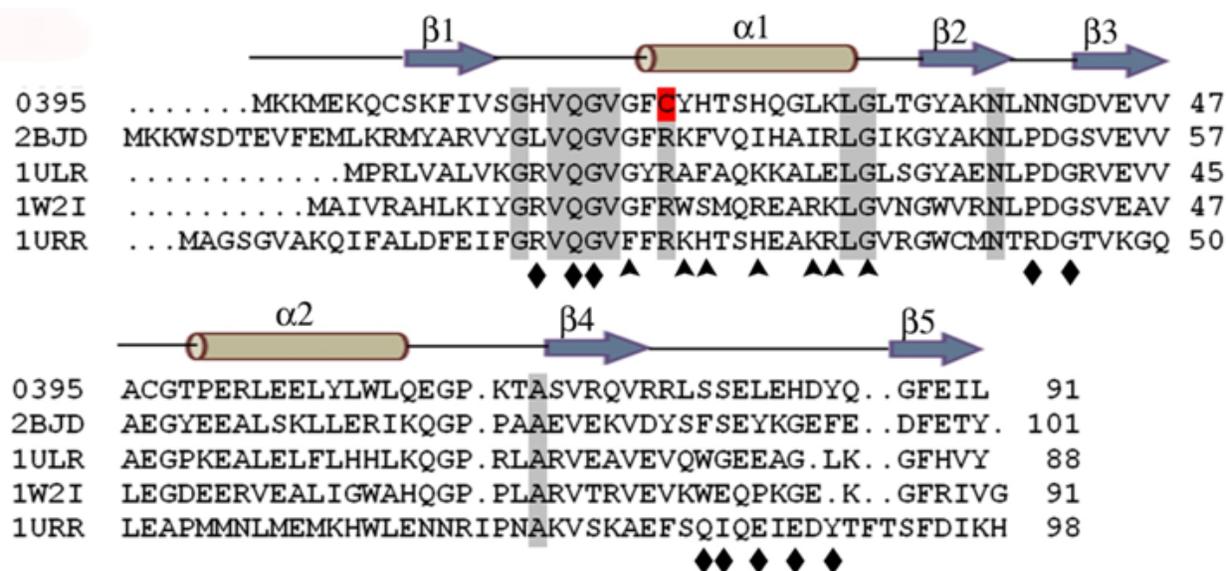
The opposite face of HT contains negative charge patches surrounded by a hydrophobic layer (**Figure 21a**). Association of two HT generates a tilted hexameric plane which further associates to form the dodecamer. The negatively charged patches inside the cage are contributed by E82, E89 and the terminal carboxylate ion from each monomer and these residues cluster together to provide a highly negatively charged environment which is capable of holding positively charged small molecules (**Figure 21b**).



**Figure 21. Electrostatic view of exterior and interior of Vc-AcP cage.** (a) Side view of the cage (half of the cage is shown in worm and the other half in surface) showing a hydrophobic lining across the periphery through which two adjoining HT faces tether together. Each hexameric plane takes up a tilted arc-like shape, analogous to tennis ball, suitable to acquire a spherical architecture upon their association. (b) The interior of the cage is highly negatively charged (red) and capable of binding oppositely charged small molecules.

### **3.1P. Role of unique residues in the formation of ordered multimeric structures: An outcome of Multiple Sequence Alignment result for Vc-AcP**

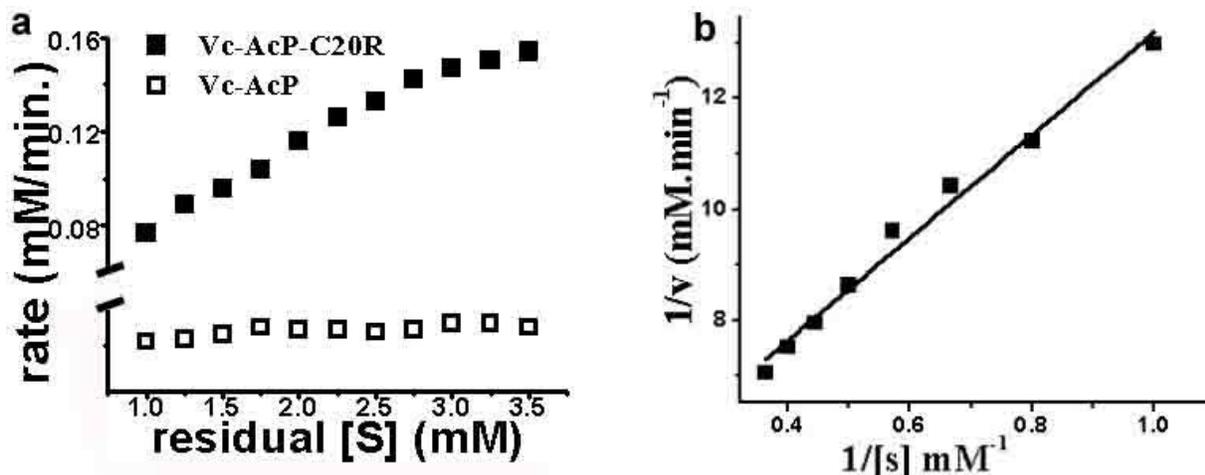
The trimer was the building block to attain the higher order structure and association of two such trimers (HT-face) formed a hexamer. When two hexamer joined together through the hydrophobic lining as shown in **Figure 21a**, the dodecamer was formed. The sequence of Vc-AcP when aligned with other sequences (**Figure 22**), the residues important in maintaining the stabilization of the trimer and hexamer were shown to be unique for *V.cholerae* O395 strain.



**Figure 22. Multiple sequence alignment of Vc-AcP with AcPs from *Sulfolobus solfataricus* (2BJD), *Thermus thermophilus* (1ULR), *Pyrococcus horikoshii* (1W2I), *Drosophila melanogaster* (1URR) showing the conserved region in grey and the natural variant Cys-residue at 20th position in red. Location of  $\alpha$ -helices and  $\beta$ -strands are indicated on top of the sequence. The residues marked with triangle and rhombus shapes are playing crucial role for trimerization and dodecamerisation of Vc-AcP.**

### 3.1Q. Kinetic Studies of Vc-ACP-C20R

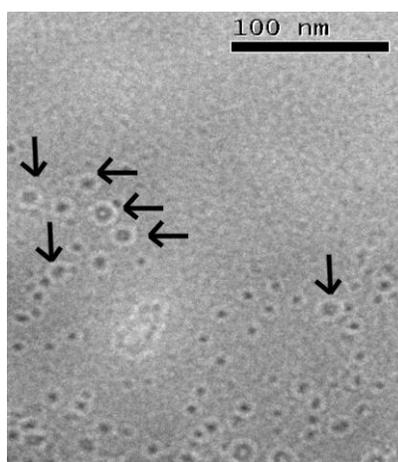
Kinetic parameters calculated for Vc-ACP-C20R using acetyl phosphate as the substrate gives  $K_m$  and  $K_{cat}$  value of  $2.36 \pm 0.6$  mM and  $2.55 \pm 0.5$   $\text{min}^{-1}$  respectively. The activity of the wild type protein is almost abolished due to the natural mutation at the active site and  $K_m$  and  $K_{cat}$  values are  $8.53 \pm 0.7$  mM and  $9.29 \pm 0.3 \times 10^{-5}$   $\text{min}^{-1}$  respectively. The reaction kinetics graph is shown in **Figure 22**.



**Figure 23. Reaction kinetics of acetylphosphate by Vc-AcP-C20R and Vc-AcP** (a) rate of the reaction (in mM/min) of both the wild type and mutant was plotted against residual substrate [S] concentration (in mM) (b) Double reciprocal plot of catalytically active (inset showing rate curve of hydrolysis of acetylphosphate) Vc-AcP-C20R mutant indicating a  $K_m$  and  $K_{cat}$  value of 2.36 mM and 2.55 min<sup>-1</sup> respectively.

### 3.1R. Transmission Electron Microscopy (TEM)

TEM images of AMS incubated Vc-AcP showed presence of circular dots on the carbon coated grid (**Figure 24**). The average diameter of those dots are around 8-10nm which was similar to the diameter of the crystal structure and half of the hydration radius ( $R_H$ ) obtained in DLS obtained for AMS (and other oxy-anions) treated Vc-AcP.

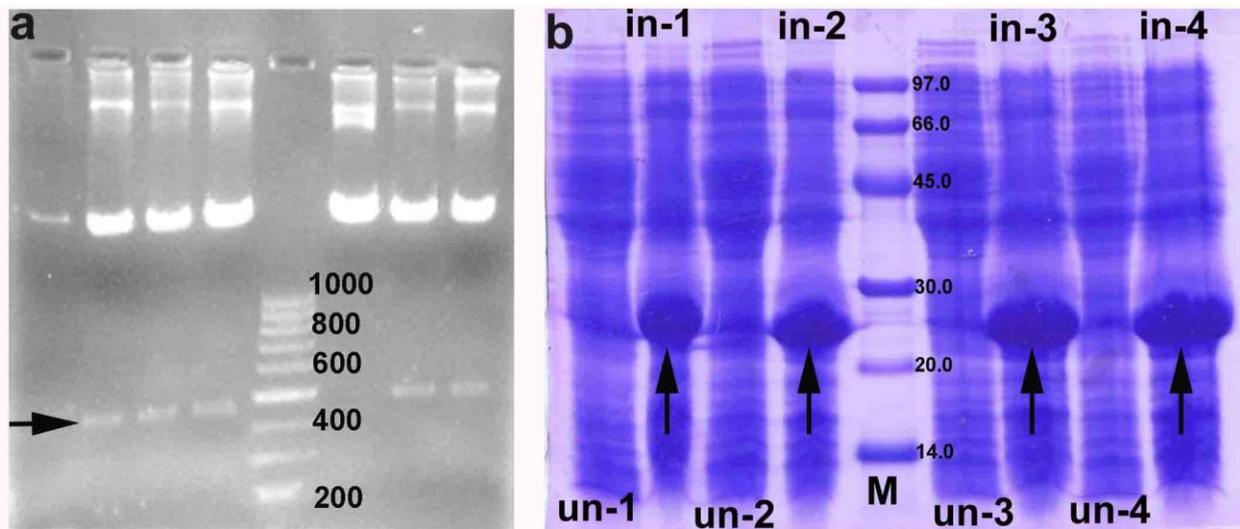


**Figure 24. TEM image** showing the formation of circular dots (marked with arrows) in solution of AMS incubated Vc-AcP.

### 3.2. Low Molecular Weight Protein Tyrosine Phosphatase-1 (*VcLMWPTP-1*)

#### 3.2A. Cloning and Overexpression of *VcLMWPTP-1* and its mutants

All the constructs of *VcLMWPTP-1* and its N-terminal 6×His-tagged mutant C8S was successfully cloned in pET28a(+) and overexpressed as N-terminal 6×His-tagged recombinant proteins. C8S mutant of *VcLMWPTP-1* bearing C-terminal 6×His-tag was also cloned in pET 24b(+) and overexpressed as C-terminal 6×His-tagged recombinant protein. Cloning and overexpression profiles of the wild type are given in **Figure 25a-25b**.

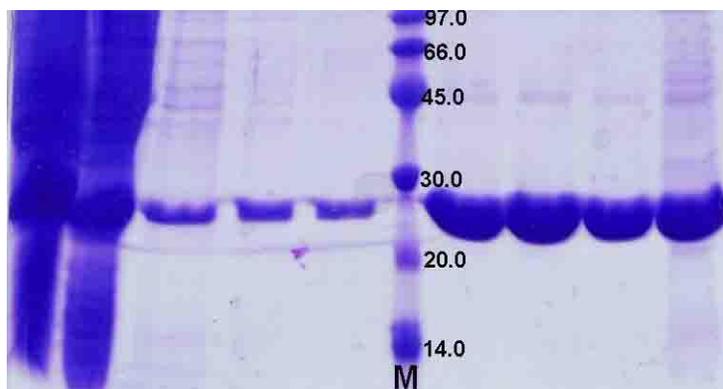


**Figure 25: (a) Clone check of *VcLMWPTP-1* in 1% agarose gel after restriction digestion with *NdeI* and *BamHI*. Lane 1-4, 6-8: plasmids, all samples (marked with arrow) except 1 and 5 contained the insert, lane 5: DNA ladder (b) 15% SDS-PAGE profile, showing the overexpression of *VcLMWPTP-1*. Lane 1, 3, 6 and 8: whole cell lysate of the uninduced cells (marked as un-1, un-2, un-3 and un-4), lane 2, 4, 7 and 9: whole cell lysate of the induced cells (marked as in-1, in-2, in-3 and in-4); induced band is marked with arrow, lane 5: Protein molecular weight marker.**

#### 3.2B. Ni-NTA affinity chromatographic purification of the target proteins

The proteins were eluted in buffers containing increasing concentration of imidazole in Ni-NTA affinity chromatography. The buffers used for the purification and their respective elution

profiles are given in **Table 8** and **Figure 26** respectively. The elution profile of only the wild type protein is shown.

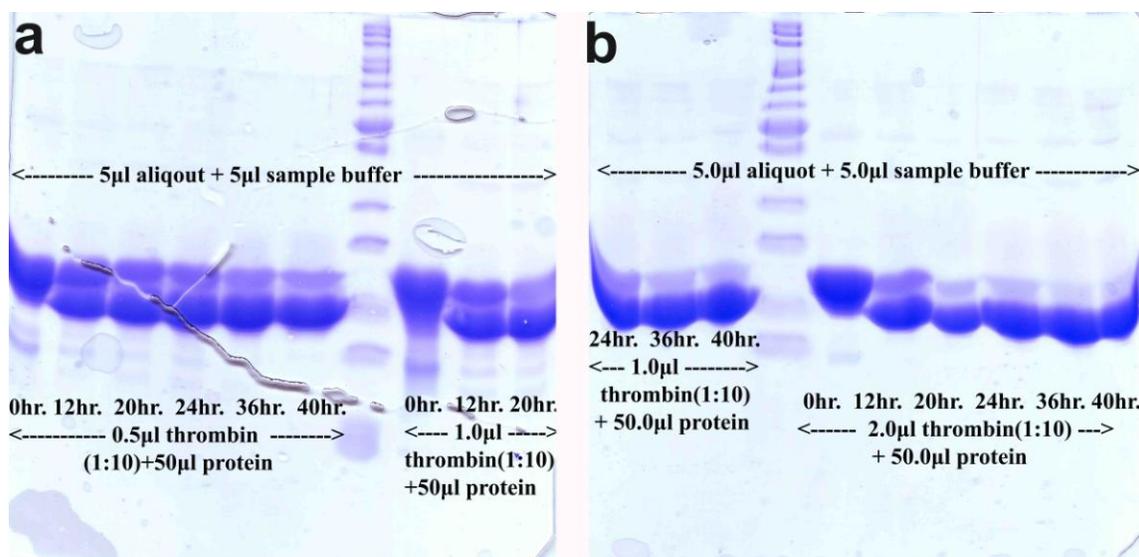


**Figure 26: 15% SDS-PAGE showing the Ni-NTA purification profile of *VcLMWPTP-1*:** Lane 1: Resuspended pellet after sonication and centrifugation, lane 2: flow through (FT) collected after passing the cell supernatant onto the buffer equilibrated Ni-NTA column, lane 3, 4,5: samples collected after washing the column with LB, WBI and WBII respectively, lane 6: Marker, lane 7-9: Eluted fractions, lane 10: Ni-beads.

The N-terminal 6×His-tagged C8S mutant of *VcLMWPTP-1* could not be purified using Ni-NTA as it eluted in the flow-through. Different trials with variation in buffer composition, ionic strength, additives did not help. The mutant was subsequently cloned as C-terminal 6×His-tagged and successfully purified.

### **3.2C. Thrombin cleavage standardization and Fast Protein Liquid Chromatography**

Prior to size exclusion chromatography using S100 column, the cleavage of the His tag by thrombin was standardized for wild type protein (*VcLMWPTP-1*). The SDS-PAGE profile of the thrombin cleavage standardization is given below (**Figure 27a-27b**).

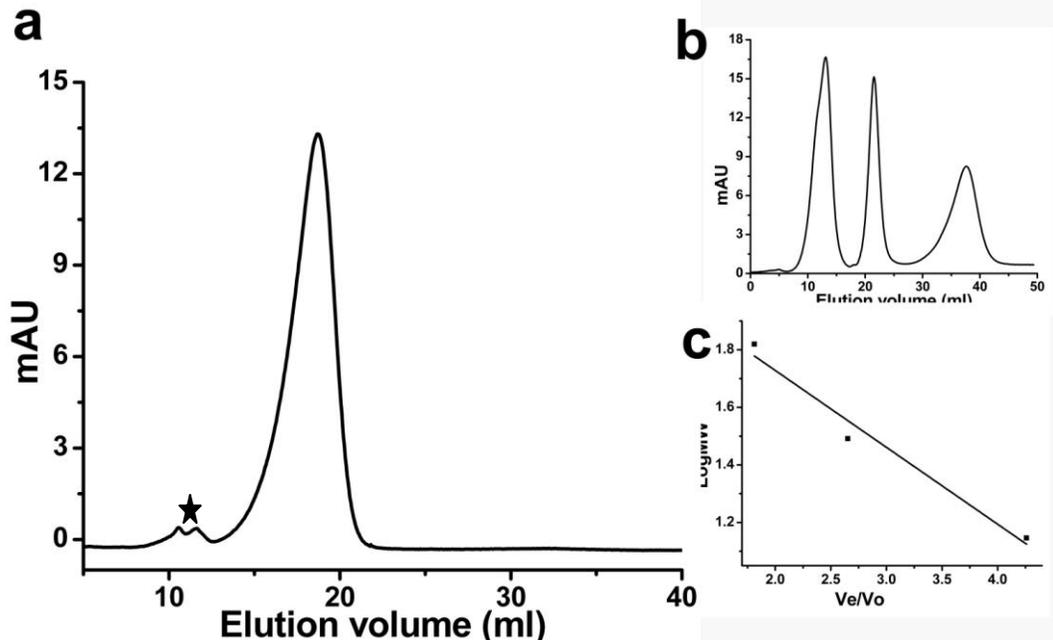


**Figure 27: Thrombin cleavage standardization of VcLMWPTP-1.** (a) Lane 1-6: thrombin cleavage profiles with 0.5 $\mu$ l of 5U/ $\lambda$  thrombin at 0hr., 12hrs, 20hrs, 24hrs, 36hrs and 40hrs respectively, lane 7-Marker and lane 8-10: thrombin cleavage profiles with 1.0 $\mu$ l of 5U/ $\lambda$  thrombin at 0hr, 12hrs and 20hrs, (b) Lanes 1-3: thrombin cleavage profiles with 1.0 $\mu$ l of 5U/ $\lambda$  thrombin at 24hrs, 36hrs and 40hrs respectively, lane 4: Marker, lane 5-10: thrombin cleavage profiles with 2.0 $\mu$ l of 5U/ $\lambda$  thrombin at 0hr., 12hrs, 20hrs, 24hrs, 36hrs and 40hrs respectively.

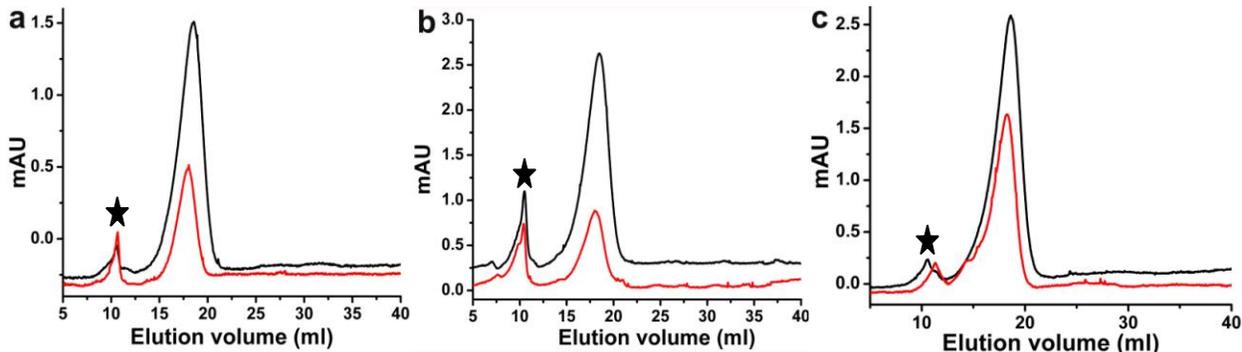
Size exclusion chromatography was performed after the standardization of the thrombin cleavage and the elution of the sample was monitored at OD<sub>280</sub>. Purity of the eluted samples was checked by 15% SDS-PAGE and concentrated for crystallization and other biophysical studies.

### 3.2D. Determination of oligomers using FPLC

FPLC analysis of VcLMWPTP-1 indicated that the protein existed as a dimer in solution as calculated from the standard curve (**Figure 28a-28c**). The dimeric species exists even at different concentrations of the protein (0.85mg/ml, 1.75mg/ml, 2.58mg/ml and 4.25mg/ml) and in buffers having varying ionic strength (150mM, 300mM and 500mM NaCl) (**Figure 29a-29c**).



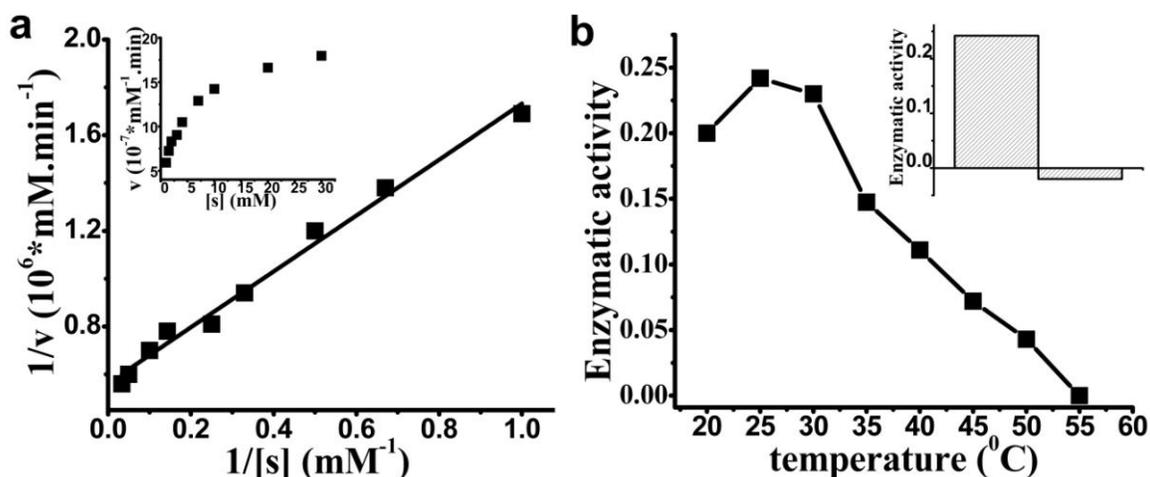
**Figure 28. Existence of oligomeric *VcLMWPTP-1* in solution by FPLC** (a) The elution volume of peak-2 corresponds to a dimeric *VcLMWPTP-1* species, peak-1 (marked with star) is impurity. (b) Calibration of the column using a mixture of proteins (from left to right: peak 1-Lysozyme, MW14.3 kDa; peak 2-ovalbumin, MW 36.0 kDa and peak 3-Bovine serum albumin, MW 66.45kDa) and (c) the standard curve is drawn based on the elution profile of protein mixtures of known molecular weights.



**Figure 29. FPLC analysis of *VcLMWPTP-1* with different protein and salt concentrations.** The chromatogram in each figure containing two curves corresponds to (a) 0.85mg/ml (red) and 1.75mg/ml (black) protein in 50mM Tris, pH 7.6, 150mM NaCl, (b) 0.85mg/ml (red) and 2.58mg/ml (black) protein in 50mM Tris, pH 7.6, 300mM NaCl and (c) 0.85mg/ml (red) and 1.75mg/ml (black) protein in 50mM Tris, pH 7.6, 500mM NaCl. Peak 1 (marked with star) in each curve of all the figures is impurity and peak 2 corresponds to the *VcLMWPTP-1*. The elution volume is same for all the buffer conditions containing varying amount of NaCl.

### 3.2E. Enzyme kinetics

Enzyme kinetic assays for VcLMWPTP-1 were performed at pH 4.8 and pH 7.6 at 25°C using p-nitrophenyl phosphate as a substrate. The determined  $K_m$  value under these conditions was  $2.07 \pm 0.2 \text{ mM}$  and  $2.03 \pm 0.4 \text{ mM}$  respectively **Figure 30a**, which is consistent with the  $K_m$  values reported for other LMWPTPs like PtpA and PtpB (1.2 and 1.5 mM) [83]. Kinetic data indicate that pH variation and dimerization has no effect on its catalytic activity. The enzyme shows optimum activity at 25°C and the activity decreases with increasing temperature leading to complete loss of activity at 55°C **Figure 30b**. Mutating the active site Cys8 to Ser (C8S) results in complete loss of enzymatic activity **Figure 30b (inset)** as reported for other LMWPTPS [84].



**Figure 30. Enzyme kinetics of VcLMWPTP-1 and C8S.** (a)  $V/[S]$  graph of VcLMWPTP-1 (inset), Double Reciprocal plot of VcLMWPTP-1 at pH 4.8 (b) Enzymatic activity ( $\mu\text{moles min}^{-1} \mu\text{g}^{-1}$ ) of VcLMWPTP-1 with temperature ( $^{\circ}\text{C}$ ); optimum enzymatic activity is observed at 25°C while complete loss of activity occurs at 55°C. Comparative enzymatic activity of VcLMWPTP-1 and C8S mutant (inset) at 25°C demonstrating that Cys to Ser mutation at the active site results in the complete loss of catalytic activity.

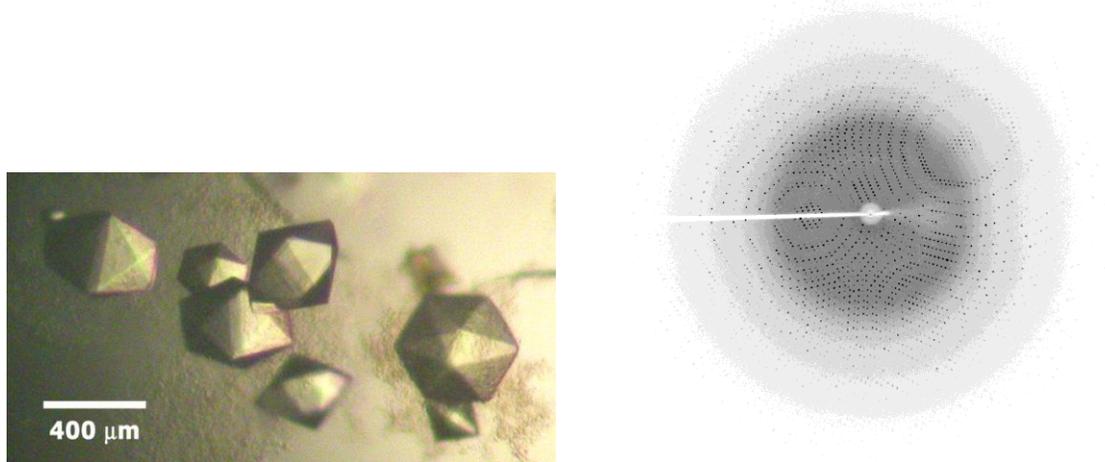
### 3.2F. Crystallization of VcLMWPTP-1

VcLMWPTP-1 was crystallized using the Ammonium Sulfate Grid Screen from Hampton, containing 2.4 M Ammonium sulfate, 0.1 M Citric acid (pH 5.0) as precipitant equilibrated against

a reservoir solution containing 3.0 M Ammonium sulfate, 0.1 M Citric acid (pH 5.0) at 277K. The crystals were of suitable size (**Figure 31a**) and were flash frozen for data collection.

### 3.2G. Data collection and Processing of VcLMWPTP-1

Initially a diffraction data set of VcLMWPTP-1 crystal with 58 frames was collected to 1.66 Å resolution using an in-house MAR Research image plate detector of diameter 345 mm kept at a distance of 120mm. and Cu K $\alpha$  radiation generated by a Bruker–Nonius FR591 rotating-anode generator equipped with Osmic MaxFlux confocal optics and running at 50 kV and 90 mA. Then the same crystal was used to collect another dataset of 63 frames with a resolution upto 1.54Å keeping the detector at 90mm. Finally, a second crystal was used to collect another dataset of 60 frames with a resolution upto 1.45 Å (**Figure 31b**) keeping the detector at 80mm. The oscillation range was kept at 1° in all the cases. The data was processed and scaled using *iMOSFLM* [55] followed by *SCALA* [56]. No translational NCS was detected. Details of data collection and processing statistics are given in **Table 15**.



**Figure 31. Crystals and diffraction pattern of VcLMWPTP-1. (a)** Diffraction quality crystals appeared in ammonium sulfate condition. **(b)** Diffraction pattern of VcLMWPTP-1 crystal upon exposure to X-rays.

**Table 15: Data collection and processing statistics of VcLMWPTP-1**

Protein name		VcLMWPTP-1
Space group		P3 <sub>1</sub>
Cell dimension	a, b, c	87.47,87.47,73.85
	$\alpha, \beta, \gamma$ (°)	90.00,90.00,120.00
Resolution (Å)*		23.89-1.45 (1.53-1.45)
R <sub>merge</sub>		0.057 (0.350)
I / $\sigma$ (I)		7.8 (2.1)
Completeness (%)		98.3 (95.5)
Redundancy		1.8 (1.7)
Total no. of observations		198204
No. of unique reflections		110187
Mosaicity		0.54
Matthew's coefficient (V <sub>m</sub> ) (ANGSTROMS**3/DA)		2.26
% solvent content		46.0

**3.2H. Structure solution of VcLMWPTP-1**

For phasing, the coordinates of *Entamoeba histolytica* (PDB code: 3IDO) was modified using *CHAINSAW* [78] in CCP4 using a sequence alignment file and residues that differ with VcLMWPTP-1 were truncated to C <sup>$\beta$</sup> . This modified PDB file was used for molecular replacement with *PHASER* [60] in CCP4 [61]. *PHASER* identified four molecules with Mathews coefficient (V<sub>m</sub>) of 2.26 (solvent content 46.0%), final TFZ= 10.3 and LLG=247 (**Table 16**). The model was improved with few cycle of rebuilding with Coot [65] and refined with Phenix refine [63] with twin law (-k,-h,-l). TLS refinement was performed during the final stages of refinement [79]. Refinement statistics are given in **Table 17**. The structure is deposited in the PDB with ID: 4LRQ.

**Table 16.Improvement of Phaser statistics of VcLMWPTP-1**

	RFZ	TFZ	LLG	PAK
Molecule 1	3.1	5.0	27	0
Molecule 2	3.5	8.0	85	1
Molecule 3	3.6	10.9	159	3
Molecule 4	3.8	10.3	247	8

where, RFZ and TFZ denotes the Z-score of Rotation and Translational function respectively. Signal-to-noise is judged by Z-score which is computed by comparing the LLG values from the

rotation/translational search with LLG for a set of random rotation/translation. PAK denotes no. of packing clashes. By default upto 5% of C<sup>α</sup> atoms for proteins are allowed to be involved in clashes. LLG is the log-likelihood gain which is an indication of how much better the solution is compared to a random solution.

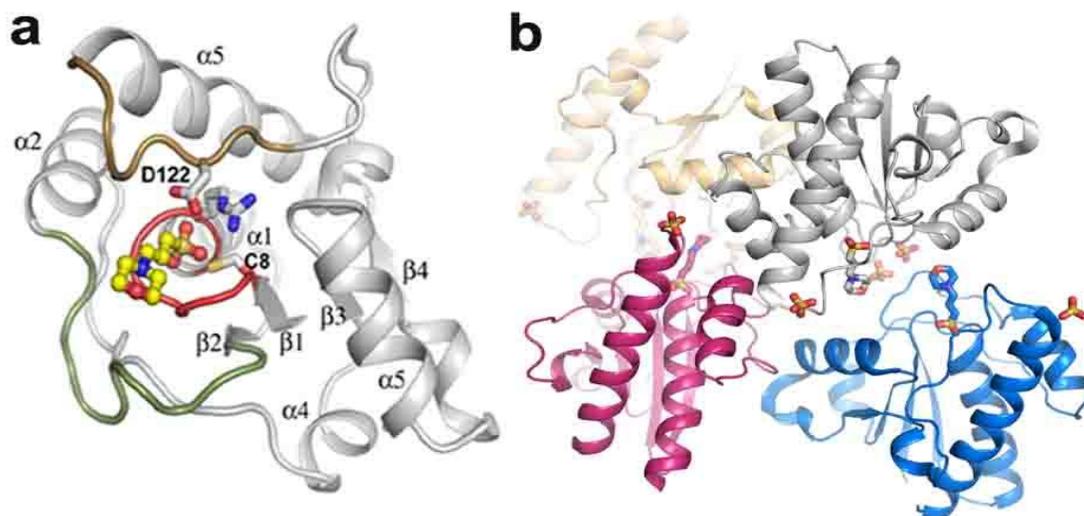
**Table 17. Statistics for data collection and refinement of crystal structure of VcLMWPTP-1**

Protein name		VcLMWPTP-1
Space group		P3 <sub>1</sub>
Cell dimension	a, b, c	87.47,87.47,73.85
	α, β, γ (°)	90.00,90.00,120.00
Resolution (Å)*		23.89-1.45 (1.53-1.45)
Rmerge		0.057 (0.350)
I / σ(I)		7.8 (2.1)
Completeness (%)		98.3 (95.5)
Redundancy		1.8 (1.7)
<b>Refinement</b>		
Resolution (Å)		23.89-1.45
Total no. of reflections/reflections used in Rfree calculation		104633/5518
Rcryst / Rfree		15.2/16.1
No. Atoms		5129
Protein		4645
Ligand		92
Water		392
B-factor (Å <sup>2</sup> )	Average	14.8
	Protein	13.6
	Ligand	14.6
	Water	16.1
R.M.S. deviations	Bond length (Å)	0.004
	Bond angles (°)	0.9
Ramachandran statistics (%)	Most favoured	97.9
	Additionally allowed	2.4
	Disallowed	0

*\*Numbers in parentheses refer to the highest resolution shell with all data collected from a single crystal.*

### 3.2I. Structural analysis of *Vc*LMWPTP-1: Overall structure of *Vc*LMWPTP-1 monomer

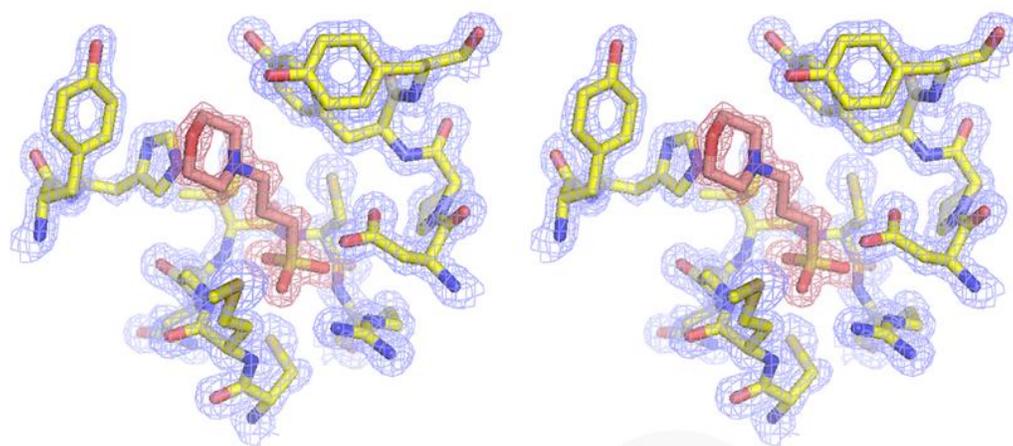
*Vc*LMWPTP-1 crystallized in spacegroup  $P3_1$  using ammonium sulfate (AMS) as precipitant and the structure has been solved at 1.45Å (Table 17). The asymmetric unit contains four molecules of *Vc*LMWPTP-1 and their arrangement suggests that molecules A and C and molecules B and D assembled together to form two dimers (Figure 32b). The high resolution electron density map allowed us to locate 149 residues out of 155 residues of each *Vc*LMWPTP-1 molecule. The terminal six residues could not be located as the electron density was not evident, probably due to their high flexibility. Each of the four molecules in the asymmetric unit binds a ligand, MOPS in its active site and one sulfate at the C-terminal loop. The fold and overall secondary structure of *Vc*LMWPTP-1 is similar to other LMWPTP structures solved till date with slight differences in some loops. Briefly, the structure is comprised of three layer  $\beta\alpha\beta$  sandwiched architecture with a topology of Rossmann fold and the four stranded parallel  $\beta$ -sheet is sandwiched between three long ( $\alpha 1$ ,  $\alpha 2$  and  $\alpha 6$ ) and two short ( $\alpha 3$ , distorted  $\alpha 4$ ) helices (Figure 32a).



**Figure 32. Overall structure and ligand (MOPS) binding at the active site of *Vc*LMWPTP-1 (PDB ID: 4LRQ)** (a) Secondary structure of *Vc*LMWPTP-1 monomer. The ligand, MOPS (ball and sticks) binds at the active site where residues C8, R14 (red P-loop), D122 (brown DPY loop) and the loop between  $\alpha 2$ - $\beta 2$  (olive) responsible for ligand binding and subsequent catalytic action. (b) Cartoon representation of the asymmetric unit content of *Vc*LMWPTP-1 showing four molecules (chain A-grey, chain B- magenta, chain C- blue and chain D-light orange). Sulfates are at the interface and C-terminal loop of each molecule.

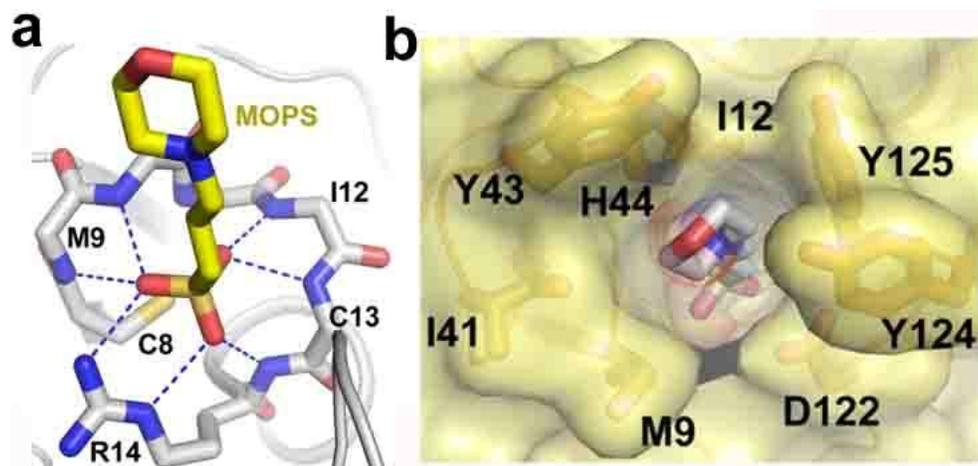
### 3.2J. Active site architecture of *Vc*LMWPTP-1

The active site is located at the crevice formed by loop-1, also called P-loop, located between  $\beta 1$  and  $\alpha 1$ . The active site is surrounded by long, flexible loop-3 and short loop-6 and the DPY-loop, which is a portion of loop-8 preceding the last helix ( $\alpha 6$ ), that contribute residue (D122) necessary for the catalytic action. From the well defined electron density in **Figure 33**, it was evident that each *Vc*LMWPTP-1 molecule binds tightly binds a MOPS molecule at the active site.



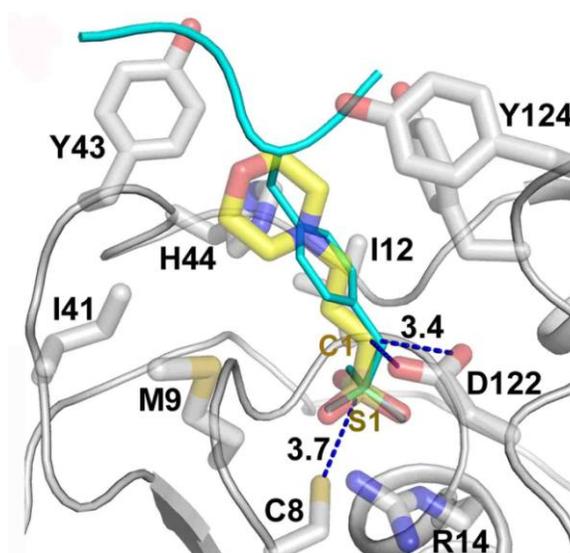
**Figure 33.** Stereo representation of the active site of *Vc*LMWPTP-1 with  $2F_{\text{obs}} - F_{\text{calc}}$  electron density map (blue mesh; contoured at  $1.8\sigma$ ) overlaid on it. The MOPS molecule is overlaid with red mesh contoured at  $2\sigma$ .

Main-chain amide nitrogens of the residues constituting the P-loop and the side-chain amide nitrogens of Arg14 are engaged in the binding and stabilization of the phosphate group of the ligand (**Figure 34a**). The piperazine ring of MOPS molecule is in close contact with several polar and aromatic residues (T40, Y43, H44, N47, D87, E89, N90, D122, Y124, and Y125) surrounding it. These polar and aromatic residues define the architecture of the wall of the deep phosphotyrosine binding pocket of *Vc*LMWPTP-1 with an opening at one side (**Figure 34b**).



**Figure 34. Zoomed view of the active site of VcLMWPTP-1.** (a) Extensive hydrogen bonds made by the sulfonyl group of the MOPS molecule with the amide environment of the P-loop and the side chain of R14.(b) Surface representation of VcLMWPTP-1 depicting the residues defining the active site pocket (surface and stick) that tightly bind the MOPS molecule (dots and stick).

The MOPS molecule closely mimics the substrate phosphotyrosine as seen in PTP1B (PDB code: 1G1F) and is also analogous to the histidine and a phosphate, together mimicking a phosphotyrosine substrate in PtpA (PDB code: 1ROF) or phenylalanine and a phosphate together mimicking a phosphotyrosine (PDB code: 1YWF) **Figure 35**. The S-atom is within 3.7Å from the catalytic S<sup>γ</sup> of Cys-8 and D122 of DPY loop comes close to after substrate binding.



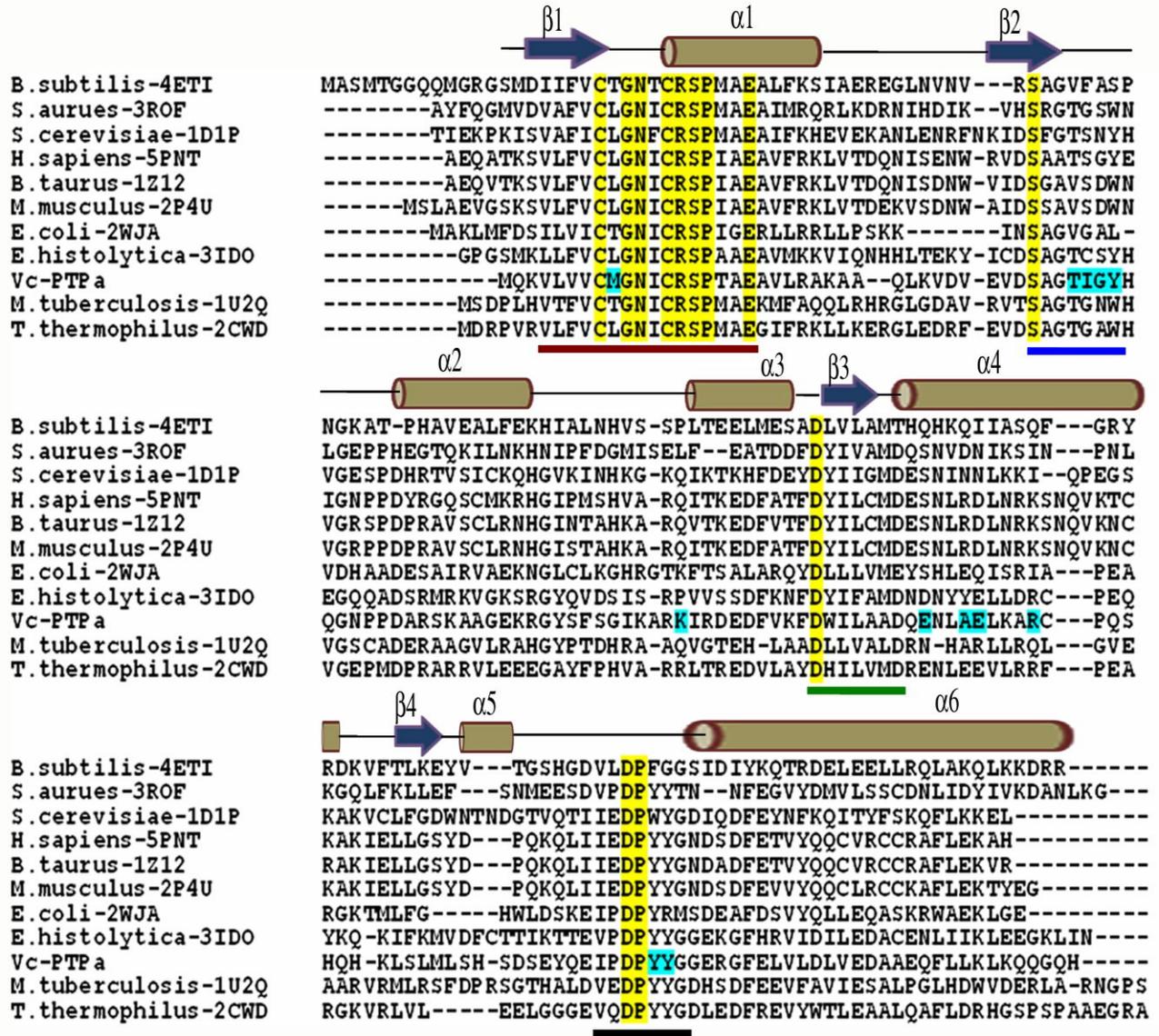
*Figure legend-next page*

**Figure 35. Similarity of binding of MOPS with that of phosphotyrosine (cyan) (PDB code: 1XXV) binding.** The residues of VcLMWPTP-1 that define the active site along with the MOPS molecule are shown as transparent sticks while the side chain of nonhydrolyzable phosphotyrosine in thick lines (cyan). The position of sulfonyl group of the MOPS molecule (stick) is almost indistinguishable with the phosphate group of phosphotyrosine. Distance between S $\gamma$  atom of Cys8 and the S1 atom of MOPS (that corresponds the P atom of phosphotyrosine) is 3.7Å, capable to accept the phosphoryl group. Distance between OD1 (or OD2) of D122 and C1 of MOPS (that corresponds the OH atom of phosphotyrosine) is 3.4Å and in a position to serve as an acid catalyst.

### **3.2K. Detection of different conserved ‘Motif’s in the LMWPTP family - A comparison based on the Multiple Sequence Alignment result**

MSA and 3D superposition of LMWPTP structures determined from other organisms show that their active site is highly conserved (**Figure 36**). Among these structures VcLMWPTP-1 shares highest sequence identity (43%) with that of *Entamoeba histolytica* (PDB code: 3JVI) with an average RMS of superposition 0.480Å for 122 residues. Structural superposition indicates highest amount of variability at the end of the distorted  $\alpha$ 4 and region preceding the catalytically important DPY-loop. Significant structural difference has also been observed around the loop-3, positioned in semicircle orientation around the P-loop, compared to PtpA, LMWPTP from gram-positive eubacteria *Staphylococcus aureus* (PDB code: 3ROF); YwIE, LMWPTP from gram-positive proteobacteria *Bacillus subtilis* (PDB code: 4ETI) and Wzb, LMWPTP from gram-negative proteobacteria *Escherichia coli* (PDB code: 2WJA). Besides the active site P-loop and DPY-loop, two more motifs are evident from the sequence alignment result. Details of the motifs and possible role of the conserved residues are listed in **Table 18**. The active site structure of PTPase follow the active site motif of CX<sub>5</sub>R, which is the only common feature between the large PTPase family and LMWPTP [85]. Sequence alignment helps to dissect the ‘X<sub>5</sub>’ portion of the P-loop sequence revealing ‘(T/L/M) GN (I/L) C’ as a consensus motif in this family, contrary to the

notion of being tagged as ‘any amino acids’. It is also to be noted that the flanking residues(s) centering the P-loop is hydrophobic in nature.



**Figure 36. Structure based multiple sequence alignment of VcLMWPTP-1 and other LMWPTPs from different organisms.** Strictly conserved residues are highlighted in yellow and residues engaged at the dimeric interface of VcLMWPTP-1 are highlighted in cyan color. The conserved motifs (in Table S2) are marked with solid lines (motif-1 in brown; motif-2 in blue; motif-3 in green and motif-4 in black) at the bottom of the alignments. The secondary structures of VcLMWPTP-1 are indicated at the top of the alignment. PDB IDs of each sequence is mentioned beside the sequence name.

Motif 1 and Motif 4 are parts of P-loop and DPY-loop respectively. Motif 2 containing loop 2 is in close proximity to the active site whereas Motif 3 containing  $\beta 3$  is remotely distant from the active site. Motif 2 region superposes well in 3D structures between *VcLMWPTP-1* with eukaryotic LMWPTPs but a marked distortion with the prokaryotic counterpart is evident, especially with Wzb, YwIE and PtpA. PDBemotif [86] shows that ‘ststaple’ and ‘asxturniil’ motifs are present in portions of motif 1 and 4 respectively. In the ‘CRST’ sequence of motif 1, Thr17(i) is H-bonded to main chain of Cys13(i-4) which is one of the characteristics of ‘ststaple’ motif. Similarly in the ‘DPY’ region of motif 4, Asp122 (i) is H-bonded to main chain of Tyr124 (i+2) which is one of the features of ‘asxturniil’ motif.

**Table 18. Motif table**

Motif	Motif sequence	Proposed role of residues
Motif 1 (residue4-19); $\beta 1$ -loop- $\alpha 1$	( <u>V</u> / <u>I</u> / <u>L</u> )X( <u>F</u> / <u>V</u> )( <u>V</u> / <u>I</u> )C <sup>8</sup> ( <u>L</u> <sup>9</sup> / <u>T</u> / <u>M</u> )GN <sup>11</sup> ( <u>I</u> <sup>12</sup> / <u>L</u> / <u>M</u> / <u>F</u> )C <sup>13</sup> RS <sup>19</sup> PX( <u>A</u> / <u>G</u> )E	C8S mutation in <i>VcLMWPTP-1</i> results in complete loss of activity. N15A (N11 of <i>VcLMWPTP-1</i> ) mutation in 5PNT results in disruption of H-bonding network [87]. L12T (L9 in <i>VcLMWPTP-1</i> ) mutation in MPtpa may play an important role in the stabilization of the p-loop [88]. I12T (I12 in <i>VcLMWPTP-1</i> ) mutation in PtpA decrease pTyr activity and increase pArg activity [89]. C17 in HCPTP (C13 of <i>VcLMWPTP-1</i> ) is shown to be involved in the formation of an S–S intramolecular bond, protecting the catalytic C12 (C8 of <i>VcLMWPTP-1</i> ) from further and irreversible oxidation [90]. S19A (S14 of <i>VcLMWPTP-1</i> ) mutant of BPTP results in thiolate destabilization [91].
Motif 2 (residue37-43); $\beta 2$ -loop2	SX( <u>G</u> / <u>A</u> )( <u>T</u> / <u>V</u> )X <sub>2</sub> ( <u>W</u> / <u>Y</u> / <u>S</u> / <u>L</u> )	No such information is available till date.
Motif 3 (residue81-87); $\beta 3$	DX( <u>I</u> / <u>L</u> / <u>V</u> )X <sub>2</sub> ( <u>M</u> / <u>A</u> / <u>L</u> )( <u>D</u> <sup>87</sup> / <u>T</u> / <u>E</u> )	D92A (D87 in <i>VcLMWPTP-1</i> ) mutation on BPTP disrupts the salt bridge with R18 and hampers the affinity of binding the substrate [92].
Motif 4 (residue120-125); loop- $\alpha 5$	( <u>V</u> / <u>I</u> )XD <sup>122</sup> P( <u>Y</u> / <u>W</u> / <u>F</u> )( <u>Y</u> / <u>R</u> / <u>G</u> )	D129A (D122 of <i>VcLMWPTP-1</i> ) mutation on BPTP shows >2000 fold reduction in $V_{max}$ [92].

### 3.2L. Dimerization of VcLMWPTP-1

Crystal structure of VcLMWPTP-1 shows that each dimer (A: C and B: D) is stitched by two sulfate ions at the interfacial region with an extensive dimerization surface (**Figure 37a**). Calculation of buried surface area (BSA) by PISA [67] shows 2910Å<sup>2</sup> BSA for the B: D dimer and 2820 Å<sup>2</sup> BSA for the A: C dimer. The BSA of both the dimers is ~20% of their respective accessible surface area (ASA) and ~10% of their total ASA. PISA also indicates that the average  $\Delta G_{\text{diss}}$  value of both the dimers is ~14kcal.mol<sup>-1</sup> indicating they are stable in solution. Detailed numerical values are given in **Table 19**. At the interfacial region, twelve residues (M9, T40, I41, G42, Y43, K71, E89, A92, E93, R97, Y124, and Y125) play pivotal role to stabilize the dimers through numerous hydrophobic and H-bonding interactions (**Figure 37b**). Y43 and K71 from each chain are involved in cation- $\pi$  interaction with its dimeric counterpart and at the dimerization interface two such interactions are present.

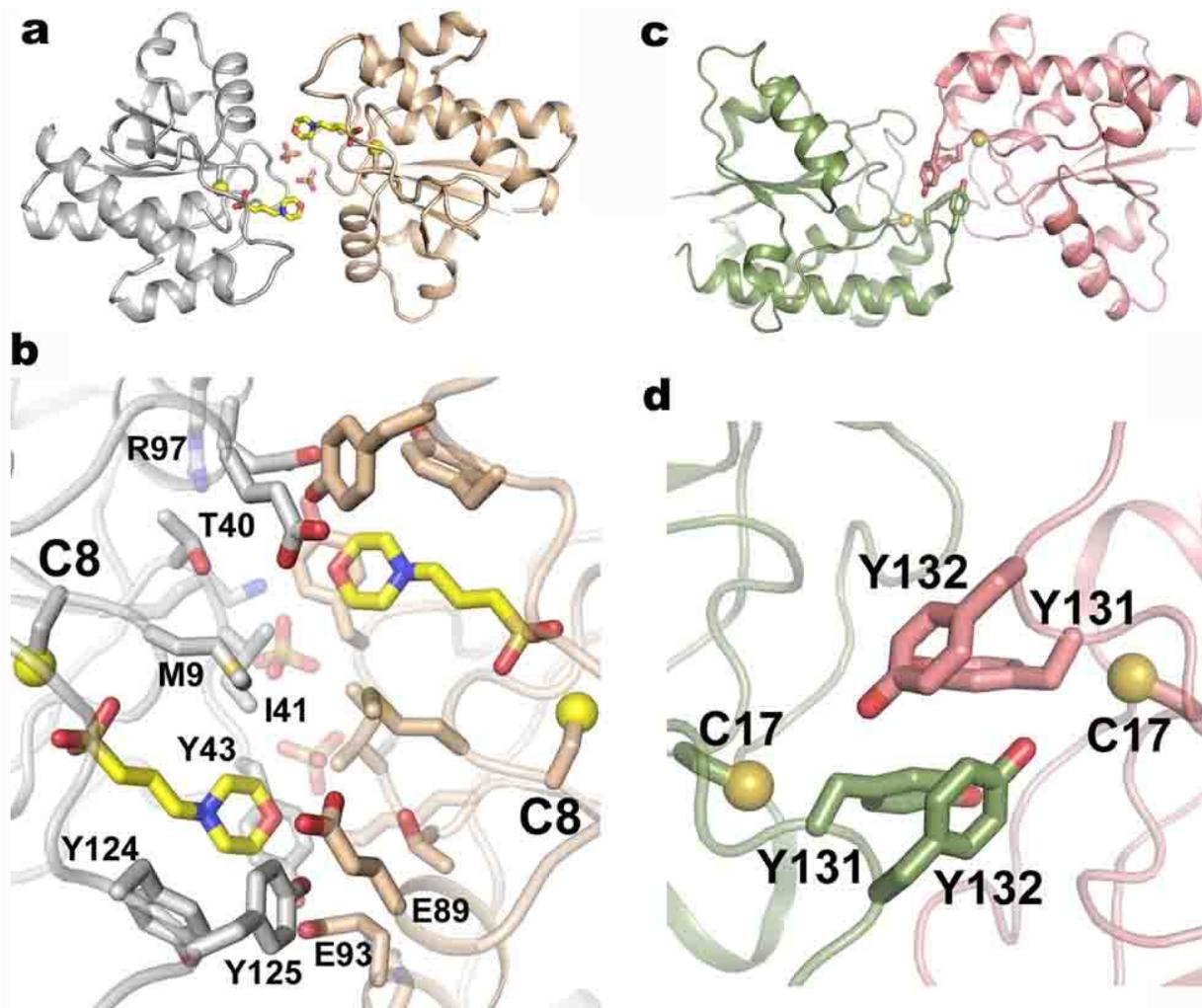
**Table 19: Results of VcLMWPTP-1 (one dimer) from the PISA server**

Protein name	Total Surface area [SA] (sq.Å)	Buried Surface area [BSA] (sq.Å)	$\Delta G^{\text{int}}$ , kcal/mol	$\Delta G^{\text{diss}}$ , kcal/mol
VcLMWPTP-1	14030	2930	-57.7	14.3

Where,  $\Delta G_{\text{int}}$  indicates the solvation free energy gain upon formation of the assembly, in kCal/M. The value is calculated as difference in total solvation energies of isolated and assembled structures. This value does not include the effect of satisfied hydrogen bonds and salt bridges across the assembly's interfaces and  $\Delta G_{\text{diss}}$  indicates the free energy of assembly dissociation, in kCal/M. The free energy of dissociation corresponds to the free energy difference between dissociated and associated states. Positive values of  $\Delta G_{\text{diss}}$  indicate that an external driving force should be applied in order to dissociate the assembly, therefore assemblies with  $\Delta G_{\text{diss}} > 0$  are thermodynamically stable.

Only R97 of chain-D is involved in this interaction with Y43 of chain-B. Sulfates that stitch two monomers connect main chain G42 and side chain K71. The dimeric zone of

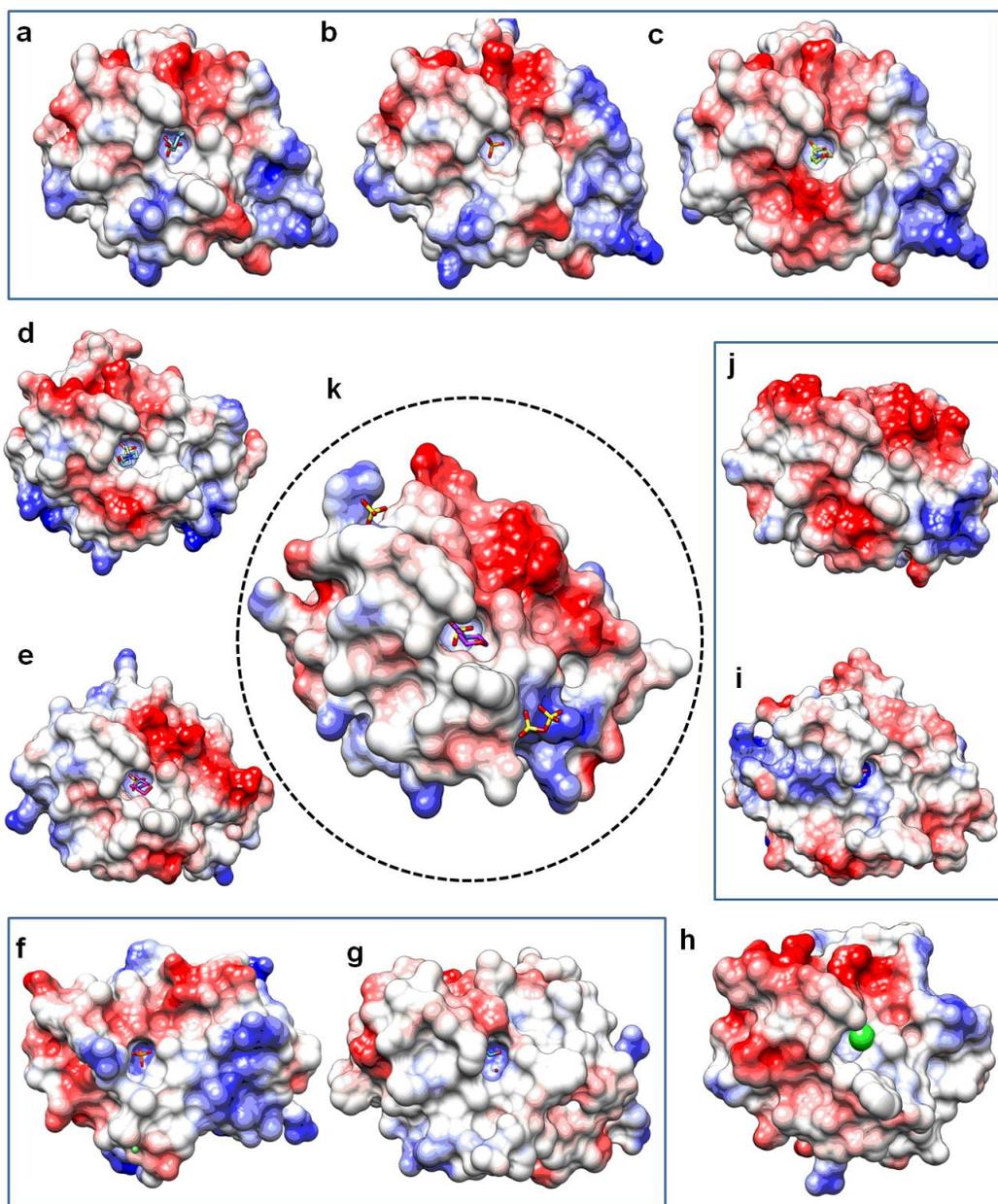
*Vc*LMWPTP-1 encompasses a larger surface area than that of the BPTP which contains a narrow region formed by the residues from the active site (**Figure 37c-37d**).



**Figure 37. Comparison of dimeric interface of *Vc*LMWPTP-1 (left) with that of bovine LMWPTPs (BPTP) (right).** (a) Cartoon representation of the *Vc*LMWPTP-1 dimer (grey and wheat) with the bound MOPS molecule (yellow) and sulfate ions (in sticks). (b) Close-up view of the dimeric interface of *Vc*LMWPTP-1 highlighting the residues involved in dimerization. Y124 and Y125 of the DPY loop are engaged in dimerization but keep the catalytic pocket accessible. (c) Cartoon representation of the BPTP dimer (green and salmon). (d) Close-up view showing Y131, Y132 of the DPY loop of one molecule comes in close proximity to the active site of the other molecule resulting in the inactivation of the enzymes.

### 3.2M. Insight into differential substrate specificity: Diverged surface property of the active site

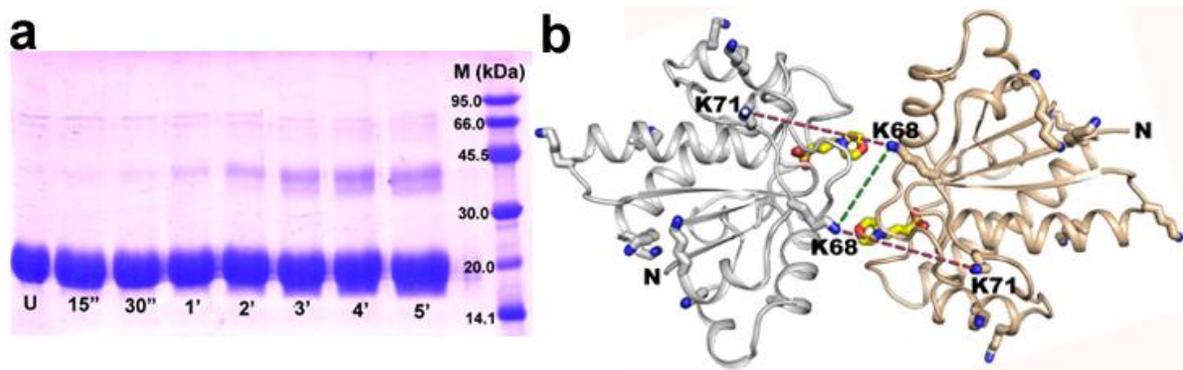
The active site of two monomers faces each other at the dimerization zone. CASTp server [93] shows  $256.5\text{\AA}^3$  as the volume of the active site cavity. Comparison of the structure and electrostatic surface around the active site of *VcLMWPTP-1* with others reveals close resemblance with EhPtp and other mammalian LMWPTPs than with LMWPTPs of bacterial origin (**Figure 38**). *VcLMWPTP-1* differs with prokaryotic LMWPTPs at two regions which are implicated in substrate recognition and binding. The W-loop, harboring a Trp-residue almost in every LMWPTP, is proposed to play an important role in substrate recognition [94]. However, corresponding residue is Tyr43 in *VcLMWPTP-1* and Tyr49 in HCPTPA, important for substrate recognition for the latter. But in bacteria, Tyr43 is substituted by Leu44 in Wzb and Ser42 in YwIE. Asn53 in HCPTPA plays a crucial role in determining the substrate specificity [95], corresponding residue Asn47 in *VcLMWPTP-1* may play a similar role. Again this is in sharp contrast with bacterial LMWPTPs like Wzb, Ptp where the corresponding residue K43 and Y44 respectively play indispensable role in substrate recognition [94]. Overall, these differences from prokaryotic LMWPTPs make *VcLMWPTP-1* active site more like eukaryotic than of prokaryotic origin which may dictate its substrate specificity.



**Figure 38. Comparison of the electrostatic surface around the active site cleft of LMWPTPs from different organisms.** (a-e) Eukaryotic LMWPTPs. (a) Bovine LMWPTP or BTP (PDB code: 1DG9), (b) Mouse LMWPTP or RTP (PDB code: 2P4U), (c) Human LMWPTP or HCPTPa (PDB code: 5PNT), (d) Yeast LMWPTP or LTP1 (PDB code: 1D1P) and (e) Protozoan parasitic LMWPTP or EPtp (PDB code: 3IDO). (f-j) Prokaryotic LMWPTPs. (f-g) Proteobacterial source, (f) LMWPTP from gram-negative gamma proteobacteria *E.coli* or Wzb (PDB code: 2WJA), (g) LMWPTP from gram-positive proteobacteria *B.subtilis* or YwIE (PDB code: 4ETI), (h) LMWPTP from *M.tuberculosis* or MPtpA (PDB code:1U2P), (i-j) Eubacterial source, (i) LMWPTP from gram-positive eubacteria *S.aureus* or PtpA (PDB code:3ROF) and (j) LMWPTP from gram-negative eubacteria *T.thermophilus* or TT1001(PDB code: 2CWD). (k) VcLMWPTP-1. Bound ligand at their active site cleft is shown in stick to orient the reader.

### 3.2N. Glutaraldehyde crosslinking to capture the dimeric state in solution

The crosslinking assay conducted on *VcLMWPTP-1* is given in **Figure 39a**. A band corresponding to molecular weight of between 30.0-45.0kDa which was double of the monomeric protein was appeared at around 30 seconds of quenching and became prominent at the end of 5minutes. The gradual appearance of this band can be explained by the dimeric crystal structure where the lysine residues from each monomer (K68) at the dimeric interface are within the distance between 8-10Å preferable to crosslink (**Figure 39b**). Other residue K71 or the terminal amine groups are beyond the condensable distance ( $>10\text{Å}$ ).

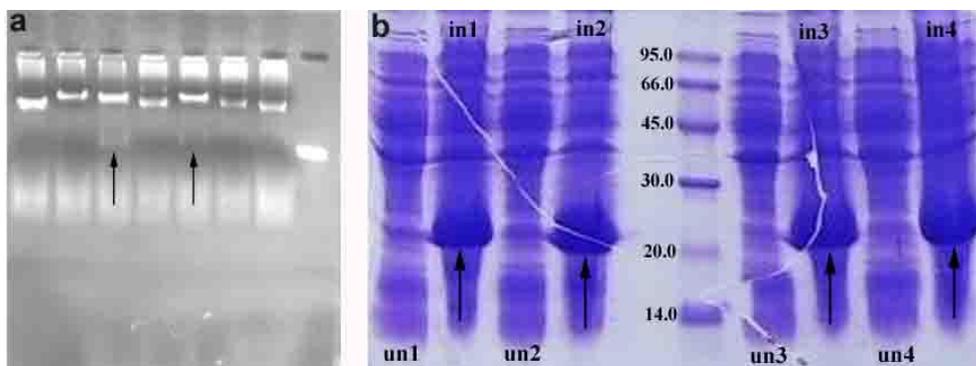


**Figure 39. Glutaraldehyde crosslinking of *VcLMWPTP-1*.** (a) 15% SDS-PAGE profile of Glutaraldehyde crosslinked *VcLMWPTP-1* showing a gradual increase in the dimeric species (~40kDa) with time. Lane 1: untreated protein, lane 2-8: glutaraldehyde treated protein, lane 10: protein marker. The band corresponding to the dimeric *VcLMWPTP-1* (between 45.0 and 30.0kDa) becomes prominent from 1 minute. (b) Dimeric structure of *VcLMWPTP-1* and Lysine residues that may react with glutaraldehyde. The shortest distance is 10Å (shown in green dashed line) between K68 from each monomer that can cross-link with each other while the next larger distance is 17Å (red dashed line) which is beyond the cross-linking range of glutaraldehyde.

### 3.3. Low Molecular Weight Protein Tyrosine Phosphatase-2 (*VcLMWPTP-2*)

#### 3.3A. Cloning and Overexpression of *VcLMWPTP-2* and its mutants

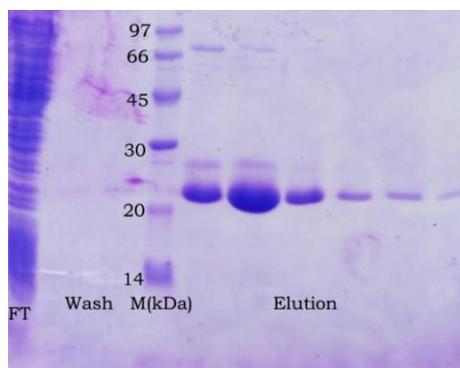
The wild type construct of *VcLMWPTP-2* was successfully cloned in pET 28a(+) and overexpressed as N-terminal 6×His-tagged recombinant proteins. Another truncated construct was also prepared in the same way deleting last 11 residues at the C-terminal end and named as *VcLMWPTP-2<sup>tr</sup>*. Mutation only at 12<sup>th</sup> position of wild type *VcLMWPTP-2* (C12S mutant) was successfully introduced and overexpressed as N-terminal 6×His-tagged protein. Mutations at other positions (17,125) were not successful. Cloning and overexpressed profiles of the wild type are given in **Figure 40a-b**.



**Figure 40. (a) Clone check of *VcLMWPTP-2* in 1% agarose gel after restriction digestion with *NdeI* and *BamHI*. lane 1-7: plasmid. Samples (marked with arrow) contained the insert, lane 8: PCR product of the insert from *V.cholerae* genomic DNA (b) 15% SDS-PAGE profile showing the overexpression of *VcLMWPTP-2*. lane 1, 3, 7 and 9: whole cell lysate of the uninduced cells (marked as un1, un2, un3 and un4), lane: 2, 4, 8 and 10: whole cell lysate of the induced cells (marked as in1, in2, in3 and in4); induced band is marked with arrow, lane 6: Marker.**

#### 3.3B. Ni-NTA affinity chromatographic purification of the target proteins

The proteins were purified in buffers containing increasing concentration of imidazole in Ni-NTA affinity chromatography. The buffers used for the purification and their respective elution profiles are given (**Table 8**). The elution profile of the wild type protein is shown in **Figure 41**.



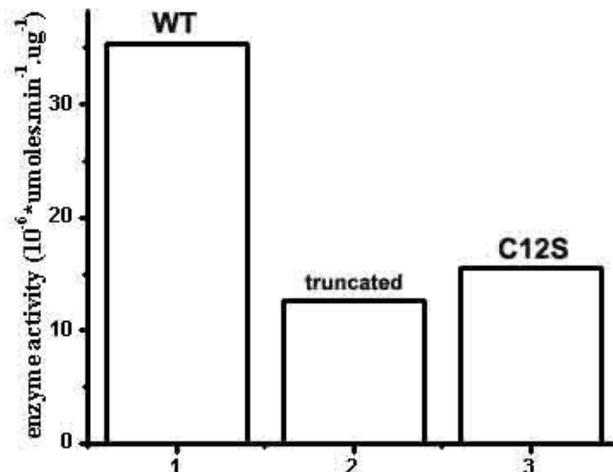
**Figure 41. 15% SDS-PAGE profile showing the Ni-NTA purification profile of *VcLMWPTP-2*:** lane 1: flow through (FT) collected after passing the cell supernatant onto the buffer equilibrated Ni-NTA column, lane 2, 3: samples collected after washing the column with WBI and WBII respectively, lane 4: marker, lane 5-9: Eluted fractions, lane 10: Ni-beads.

### 3.3C. Thrombin cleavage standardization and Fast Protein Liquid Chromatography

Size exclusion chromatography was performed after the standardization of thrombin cleavage and the elution of the sample was monitored at OD 280nm.

### 3.3D. Enzyme kinetics

Relative enzyme activity of the wild type *VcLMWPTP-2*, truncated construct *VcLMWPTP-2<sup>tr</sup>* and C12S mutant was determined. Enzyme kinetic assay was performed at pH 4.8 at RT using p-nitrophenyl phosphate as a substrate. The determined values under these conditions were  $35.41 \times 10^{-6}$ ,  $12.6 \times 10^{-6}$  and  $15.57 \times 10^{-6}$   $\mu\text{moles} \cdot \text{min}^{-1} \cdot \mu\text{g}^{-1}$  respectively (**Figure 42**). This leads to suggest two inferences, firstly unlike *VcLMWPTP-1* mutant C8S, the C12S mutant does not lose its activity fully and secondly the C-terminal tail may be involved in the catalytic activity.



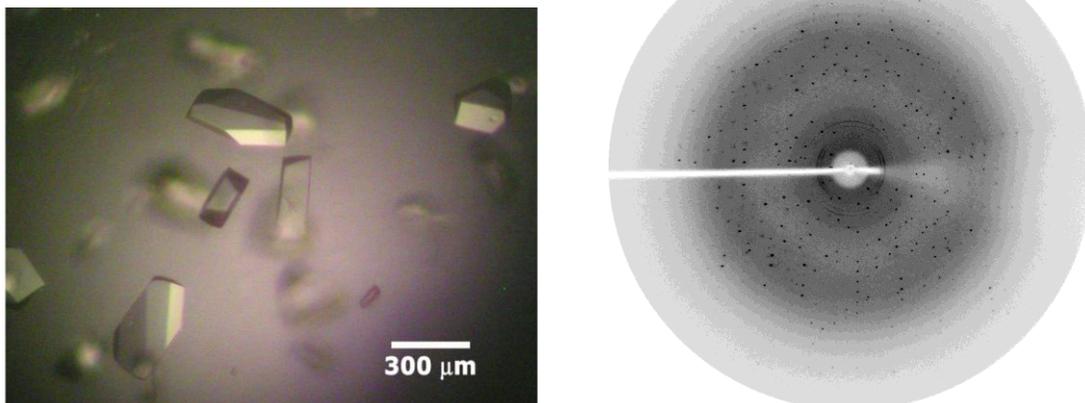
**Figure 42. Enzyme kinetics of *VcLMWPTP-2* showing reduced activity of truncated and C12S mutant of *VcLMWPTP-2* compared to the wild type protein.**

### 3.3E. Crystallization of *VcLMWPTP-2*

*VcLMWPTP-2* was crystallized using the precipitant from Crystal Screen from Hampton Research, containing 0.2 M Ammonium sulfate, 30% (w/v) Polyethylene glycol 8000 against a reservoir solution of the same. Pyramid shaped crystal appeared within a couple of days at 293K. The crystallization conditions were further modified using Crystal Screen Cryo solution containing 0.17 M Ammonium sulfate, 25.5% (w/v) Polyethylene glycol 8000, 15.0% Glycerol from Hampton Research. Best crystals appeared when 4  $\mu\text{l}$  of the protein was mixed with 1  $\mu\text{l}$  of the precipitant (**Figure 43a**) and incubated at 293K. But the crystals were not reproducible with different batches of purified protein.

### 3.3F. Data collection and Processing of *VcLMWPTP-2*

Diffraction data of *VcLMWPTP-2* crystal were collected to 2.8 Å resolution (**Figure 43b**) using an in-house MAR Research image plate detector of diameter 345 mm kept at a distance of 260mm. and Cu K $\alpha$  radiation generated by a Bruker–Nonius FR591 rotating-anode generator equipped with Osmic MaxFlux confocal optics and running at 50 kV and 90 mA.



**Figure 43. Crystals and diffraction pattern of VcLMWPTP-2.** (a) Crystals of VcLMWPTP-2 grown using 7.5% (w/v) PEG 6000, 5% (v/v) glycerol, 0.5 mM DTT and 300 mM NaCl in 0.1 M MES pH 6.0 at 277 K (b) Diffraction pattern of VcLMWPTP-2 crystals upon exposure to X-rays.

**Table 20: Data Collection and processing statistics of VcLMWPTP-2**

Crystal Name		VcLMWPTP-2
Space group		C2
Cell dimensions	a, b, c	121.38, 45.25, 88.56,
	$\alpha, \beta, \gamma$ (°)	90, 121.08, 90
Resolution (Å) *		2.8
$R_{\text{merge}}^{\dagger}$		0.043 (0.183)
I / $\sigma(I)$		19.6 (6.1)
Completeness (%)		92.1 (74.9)
Total no. of observations		36313
No. of unique reflections		10945
Mosaicity		1.33
Matthew's coefficient ( $V_m$ ) (ANGSTROMS**3/DA)		2.77
% solvent content		56.0

\*Numbers in parentheses refer to the highest resolution shell with all data collected from a single crystal.

$\dagger R_{\text{merge}} = \frac{\sum hkl \sum i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum i I_i(hkl)}$ , where  $I_i(hkl)$  is the observed intensity of the measurement of reflection  $hkl$  and  $\langle I(hkl) \rangle$  is the mean intensity of reflection  $hkl$  calculated after scaling.

Due to lower completeness and high mosaicity of the collected data, it did not give any successful solution. Also the sequence identity of *VcLMWPTP-2* with other homologues being >35%, proper search model was also unavailable. Again the crystals being not-reproducible with other batches of purified protein, the chance of experimental phasing was also ruled out.

# **Chapter-4**

## **Discussion**

## 4. Vc-Acp and Vc-AcP-C20R

### 4.1A. Unique set of residues involved in the tetrahedral cage formation of Vc-AcP and triangular Vc-AcP-C20R

Oligomerization of AcP is a common phenomenon reported earlier. Other than that, human muscle acylphosphatase is also reported to form aggregates in presence of heparan sulfate which is a component of the extracellular matrix and universally associated with amyloid deposition followed by plaque formation [96]. When oligomeric structures of Vc-AcP and its active site mutant Vc-AcP-C20R were obtained, they were compared with the structures available till date. Firstly, the oligomers formed in these cases are of definite shapes unlike other reported structures. The mutant forms a trimer which is the building block of the dodecameric wild type protein.

Multiple sequence alignment and molecular details of Vc-AcP-C20R structure reveals that the residues coming from  $\alpha 1$  are arranged in three layers at the trimeric interface and these residues are solely present in *Vibrio cholerae*0395 (**Figure 14**). Looking from the vertex sulfate, in the first layer the side chain of the K29 residue binds to the oxygen atoms of the sulfate. H22 is also present in that layer. In the middle layer, H25 is placed below the sulfate and at the bottom layer H21 and L28 are aligned. Exclusive presence of these residues (especially Y21, H22, H25 and T68) helps in the formation and stabilization of the trimeric assembly. Though the two histidine residues are available in case of *Drosophila melanogaster* AcP (PDB ID: 1URR), the tyrosine and threonine residue implicated in trimer formation are lysine and asparagine here. So the interactions found between the side chains of Y21 /H25 pair and between main chains of Y21 and side chain of T68 are missing, thereby losing important interactions responsible for the

stabilization of the trimer. Absence of these crucial residues probably hinders the formation and stabilization of the ordered trimeric or higher oligomeric assembly in case of other AcPs.

Analysis of the cage structure reveals that the helix mediated trimeric structure is the basic building block of the cage and four such trimer assemble tetrahedrally to build up the higher order dodecameric cage formed by the wild type Vc-AcP. The association of two trimers produces a tilted plane and two such planes join together to form the dodecamer (**Figure 21a**). While the interactions between  $\alpha 1$  are necessary for the stabilization of the trimer, interactions among four trimers are crucial in the formation and the stability of the cage. Structure of the dodecameric protein cage emphasizes the exclusive positional advantages of specific residues from three different loops viz., loop-1 (H13, Q15 and G16), loop-3 (N40 and G42) and loop-6 (S78, S79, L81, H83 and Y85). Out of these residues only Q15, G16 and G42 are conserved in AcPs. Q15 maintains the active site architecture and main chain amide of G16 stabilizes the ligand at the active site pocket. G42(i) rigidifies the position of loop-3, hosting the catalytic N38, by interacting with the main chain of L39(i-3). Therefore, conservation of the residues is justified due to their importance in maintaining the active site structural integrity. Except Q15, G16 and G42 are specific for Vc-AcP and they are exclusively taking part in cage formation.

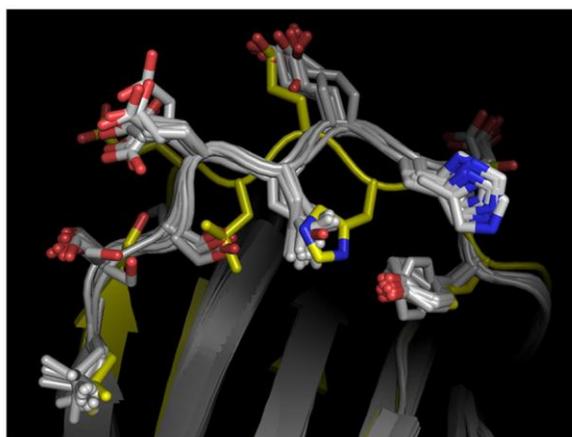
#### **4.1B. Single mutation C20R on Vc-AcP affects the assembly formation**

The presence of a natural mutation at the active site of Vc-AcP enquires about its significance in the context of its structure and functional property. Kinetic study reveals that the catalytic property of the wild type protein is almost impaired due to the natural active site mutation Arg20  $\rightarrow$  Cys. The normal catalytic property is, however, regained by a single Cys20  $\rightarrow$  Arg mutation. In solution, the mutant exhibits a single population but the wild type shows

three oligomeric populations having different hydrodynamic radius ( $R_H$ ) and predicted MW. Interestingly, the population having lowest  $R_H$  and MW of wild type Vc-AcP matches with that of the C20R mutant. However, the C20R mutant never exhibited the higher  $R_H$  and MW population, irrespective of salt and pH variation.

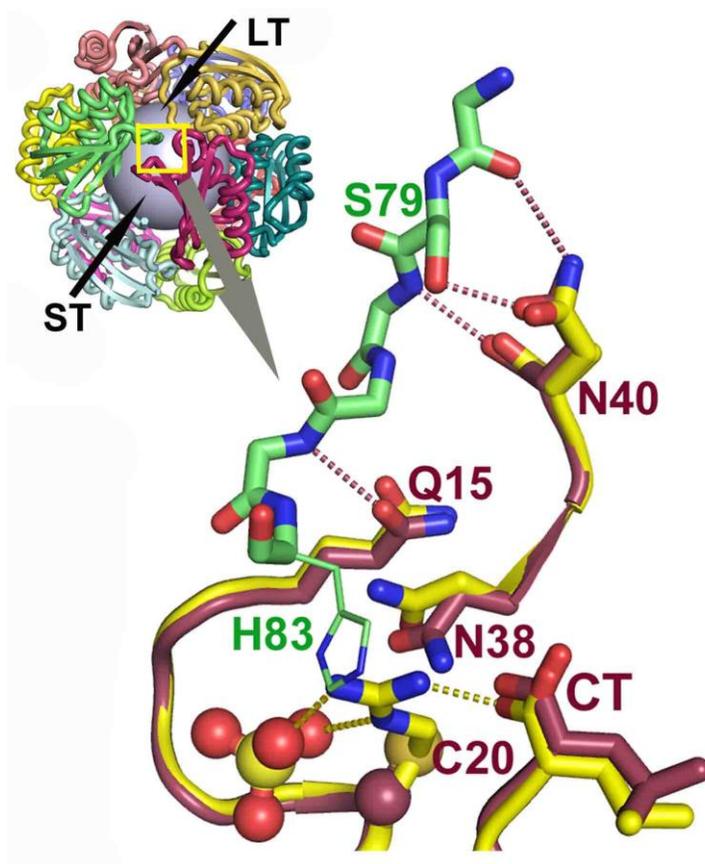
Crystal structure of the C20R mutant gives a trimer the size and MW of this trimer matches with the  $R_H$  and MW obtained from DLS experiments. The wild type protein, on the other hand, exhibits a dodecamer protein cage structure with tetrahedral arrangement of four trimeric assemblies found in case of the mutant. This finding enquires how a point mutation at the active site can act as a guiding factor to dictate the formation of different types of assembly.

Reconstruction of the dodecameric cage by superposition of the trimeric C20R structure on it shows several steric clashes. To avoid steric clash C-terminal loop-6 adjust itself and freezes to a conformation during cage formation in wild type Vc-AcP. In C20R mutant residues from this loop may clash with the locked side chain conformation of R20. This clash is easily avoided by shorter C20 thereby favoring the formation of the higher order structure (**Figure 44**).

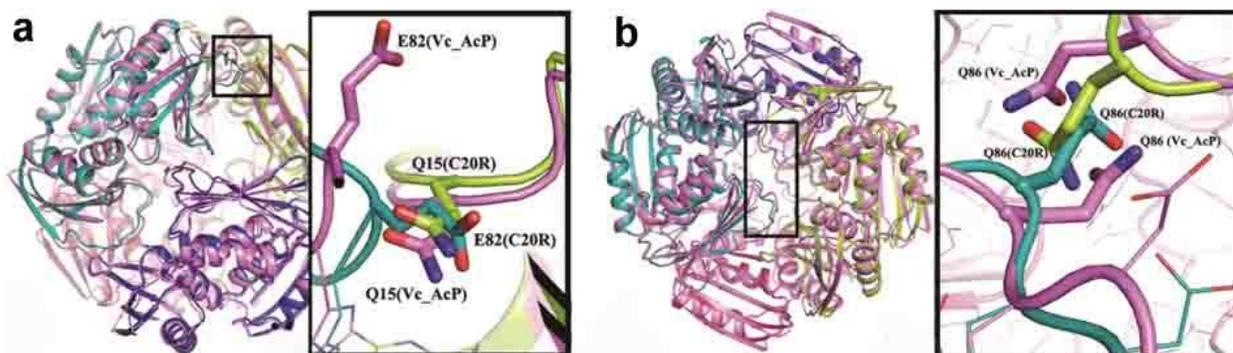


**Figure 44: Superposition of 12 monomers of Vc-AcP (grey) and one monomer of Vc-AcP-C20R (yellow) showing the structural change at loop-6.**

Moreover, strong salt bridge interactions between R20 and C-terminal carboxylate (**Figure 45**) reduce the flexibility of Loop-6 in case of Vc-AcP-C20R. Due to this loss of flexibility favourable interactions made by loop-6 with the main chain and side chains of N40, Q15 would be difficult to be form at the ST and LT interfaces. Loss of flexibility of Loop-6 also causes severe clashes between H83 with R20 and N38, thereby hindering the cage formation (**Figure 46a-46b**).



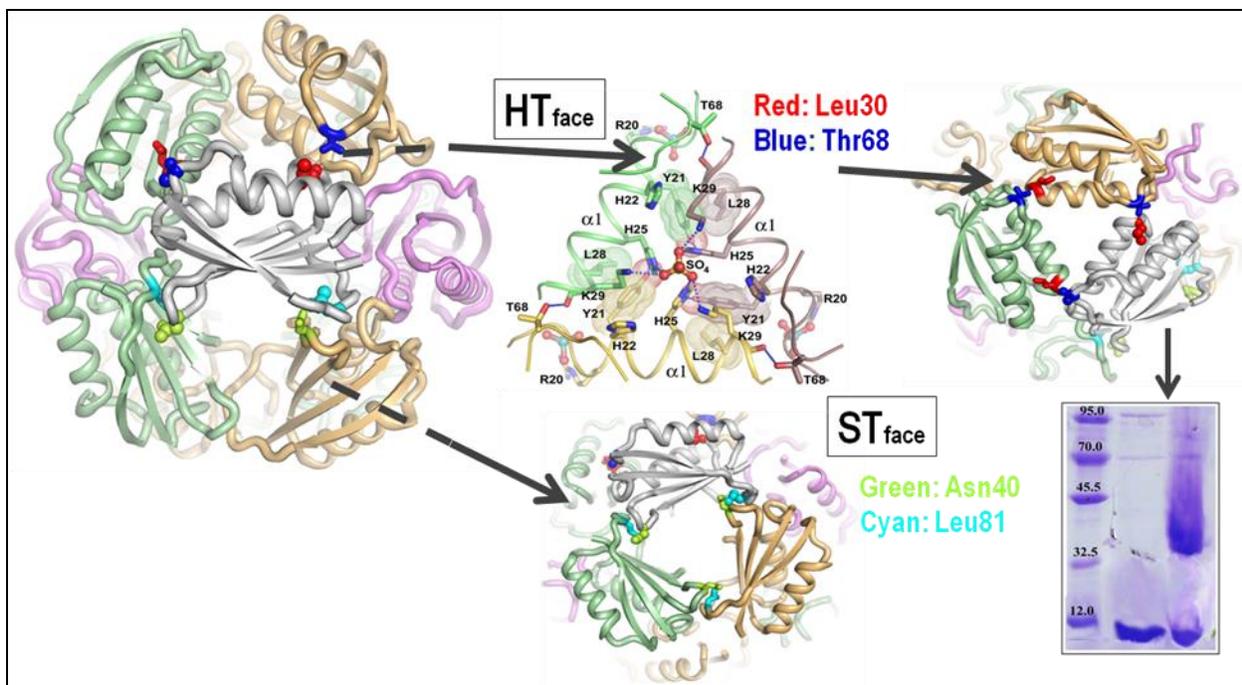
**Figure 45: Superposition on Vc-AcP-C20R on Vc-AcP cage** showing the ST and LT face. Zoomed view of one of the LT corner (yellow box) where loop-6 of one monomer (green) is sandwiched with loop-2 and loop-3 of the other monomer (magenta). Vc-AcP-C20R (yellow) is superposed on one of the monomer (magenta) of the Vc-AcP cage to show how R20 (shown in stick) clashes with H83 (shown in line) of other monomer and hinders the cage formation (sulfate is shown in stick). In case of Vc-AcP the clash does not occur due to smaller size of C20 (shown in ball and stick).



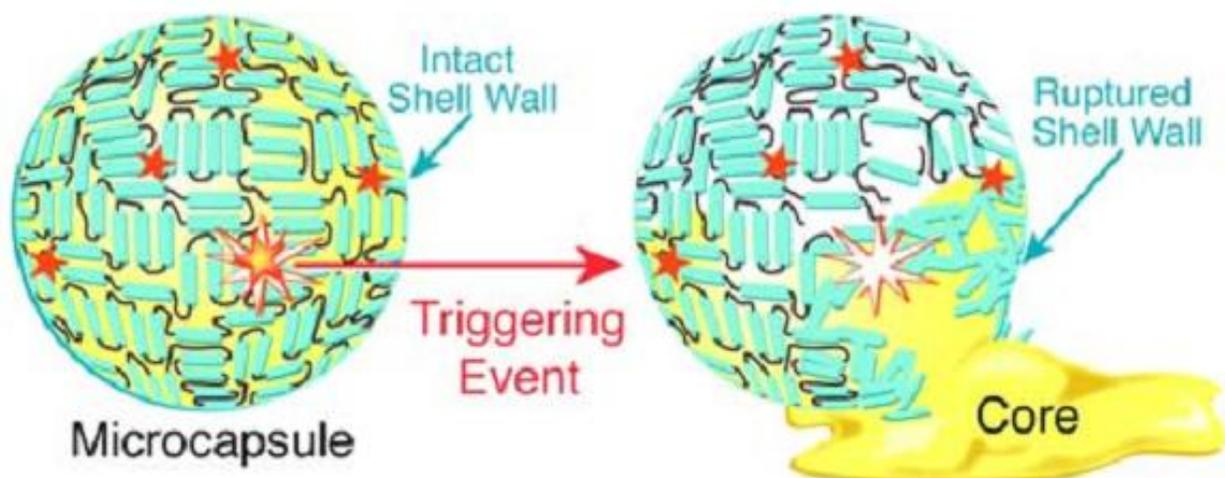
**Figure 46: Vc-AcP-C20R (green, teal) superposed on Vc-AcP (violet) showing possible clashes which prevent the mutant cage formation (a) Probable clash at the ST interface. Zoomed view of the position showing Q15 (teal) from loop1 of one chain collides with E82 (green) of loop6 of another in case of Vc-AcP-C20R and distortion in loop6 of Vc-AcP (violet) escapes this clash, (b) Possible clash at the LT face. Zoomed view showing Q86 of loop6 clashes in case of the mutant (green and teal) instead Q86 of Vc-AcP (violet) interacts to stabilize the junction.**

The structure of the wild type protein can be suitably modified by means of several mutations (L30, T68, N40 and L81) at proper junction points so that a designer cage can be planned (**Figure 47**). If the chosen residues are mutated to cysteine in such a way that whenever one monomer comes in close contact of the other a disulfide bridge can be formed. The formation of this disulfide bridged cage can be controlled using reducing agents accordingly.

The cage thus formed when allowed to encapsulate small molecule according to the size and charge of the hollow interior, it can carry the molecule inside until collapsed by a reducing agent. The presentation of the small-molecule carrier designer cage can be viewed in **Figure 48**.



**Figure 47. Construction of disulfide-bridged cage.** One pair of residues at HT- and ST-face mutated to Cysteine to form artificial bridge to join the trimers.



**Figure 48: Schematic representation of programmable designer protein cage.** The cage wall is a dodecameric assembly like the Vc-AcP structure (green) with a disulfide bridge (star) between mutated cysteine residues at different faces. The cage releases its content upon activation from a triggering event of addition of reducing agent in solution. This event dissolves the disulfide linkages (star) initiating a collapse of the cage and release of core contents. (Figure courtesy: Moore et al.2010 JACS)

## **4.2. VcLMWPTP-1**

### **4.2A. VcLMWPTP-1 forms enzymatically active dimer in solution**

In solution VcLMWPTP-1 forms exclusively dimeric species which is evident from the FPLC experiment. The dimer thus formed is independent of protein concentration and ionic strength (fig). Kinetic data, on the other hand, suggests that the wild type protein is enzymatically active. Together it can be proposed that the dimeric form of VcLMWPTP-1 is enzymatically active. This observation is in sharp contrast with the previous reports of LMWPTP dimers from both prokaryotes and eukaryotes are enzymatically inactive. Therefore the mode of dimer formation in VcLMWPTP-1 has to be different from the previous ones. Crystal structure of VcLMWPTP-1 shows that the dimer forms in such a way that the active site remains accessible to the substrate and hence it shows appreciable activity. Glutaraldehyde crosslinking further strengthen the fact of dimer-formation.

The kinetic parameters obtained for the wild type protein did not alter with the pH of the system. The active site C8S mutant lost its full activity in comparison to the wild type protein suggesting this residue as the primary attacking residue for the catalytic action.

### **4.2B. Molecular details of the dimeric crystal structure of VcLMWPTP-1 reveal the altered mode of dimerization**

The phenomena of ‘Dimerization/weak oligomerization’ of LMWPTP are already reported for both eukaryotic and prokaryotic LMWPTPs. But the orientation of the dimer reported so far from X-ray crystallographic structure and NMR studies is markedly different from that obtained in case of VcLMWPTP-1. Generally a head-to-head orientation of the active site of two monomers is observed. The crystal structure of BPTP shows the side chain of tyrosine

residue just after the DPY-loop gets inserted inside the catalytic pocket leaving the space unavailable to the incoming substrate thereby reporting it to be catalytically inactive. Comparison of the dimer formation mode with that of the *VcLMWPTP-1* dimer indicates similar head-to-head orientation that differs in the positioning of the residues at the catalytic site. As the dimeric zone is comparatively more extended in case of *VcLMWPTP-1*, the tyrosine residue positioned after the DPY-loop is unable to protrude inside the catalytic pocket. Instead it remains closer to the ligand at the active site. Accessibility of the cleft easily houses the incoming ligand thereby rendering the protein enzymatically active.

Analysis of the position of Lys residues at the dimeric interface of the *VcLMWPTP-1* shows only a pair of amine (Lys68 from each monomer at a distance of  $\sim 10\text{\AA}$ ) which could crosslink with each other in presence of glutaraldehyde (**Figure 39a-39b**). Distance between all other inter-monomeric Lys residues and the N terminus amines are beyond the condensable distance. As this mode of dimerization is stabilized by hydrophobic and polar interactions, it would show a monomeric band in SDS-PAGE. However, if the dimer is crosslinked by glutaraldehyde a band near its dimer appears with time. The dimeric structure of *VcLMWPTP-1* correlates well with glutaraldehyde crosslinking. Since the distance between Lys68 in the dimer is little larger than the condensable distance ( $6-8\text{\AA}$ ), more dimers become crosslinked with gradual increase in time.

#### **4.2C. Electrostatics surface indicates similarity of *VcLMWPTP-1* with eukaryotic LMWPTPs**

The residues around the active site cleft of *VcLMWPTP-1* differ from other bacterial LMWPTPs and hence the electrostatic charge distribution around the active site is also different.

Among the eukaryotic LMWPTPs, EhPTP bears maximum similarity with *Vc*LMWPTP-1 at the active site residues suggesting close similarity in substrate(s) preference to eukaryotic LMWPTPs, than its bacterial homologue. Since the dimeric surface is closer to the active site of *Vc*LMWPTP-1 it may further modulate the substrate recognition and specificity.

### **4.3. *Vc*LMWPTP-2**

The data collected for the *Vc*LMWPTP-2 crystal was not fruitful to give a structure. The protein was then tried to crystallize with freshly prepared batches of protein to try experimental phasing. The fresh batches failed to reproduce any crystals in previous conditions. Hence no further work was done on this protein. Kinetics data suggested that the truncated construct with a deletion of C-terminal 11 residues had a declining impact on the enzyme activity data w.r.t. the wild type protein. Again the C12S mutant also showed comparable activity w.r.t the wild type protein. Interestingly the truncated protein and the mutant shared approximately similar kinetic parameter. This observation suggested that the first one might play an important role in the catalytic mechanism of the protein. In the latter case, C12 was not the sole residue for maintaining the catalytic power unlike *Vc*LMWPTP-1 where C8S mutant sequentially analogous to C12S of *Vc*LMWPTP-2 held the same and mutation resulted in complete loss of activity. However the relation of these residue(s) or region(s) with the catalytic mechanism of the protein *Vc*LMWPTP-2 will be revealed when the atomic details are available.

# References

## References:

1. T. Mustelin (2007) A brief introduction to protein phosphatase families. *Protein Phosphatase Protocols*; Chapter-2: 9-22.
2. I.Rebay, S.J.Silver and T. L.Tootle (2005) New vision from Eyes absent: transcription factors as enzymes. *Trends Genet.*; 21: 163–171.
3. J. P.Rayapureddi, C.Kattamuri, F.H.Chan and R.S.Hegde (2005) Characterization of a plant, tyrosine-specific phosphatase of the aspartyl class. *Biochemistry*; 44: 751–758.
4. A.Roberts, S.Y.Lee, E.McCullagh, R.E.Silversmith and D. E. Wemmer, (2005)Ybiv from Escherichia coli K12 is a HAD phosphatase. *Proteins*; 58: 790–801.
5. S. J.Roberts, A. J.Stewart, P. J.Sadler and C.Farquharson (2004) Human PHOSPHO1 exhibits high specific phosphoethanolamine and phosphocholine phosphatase activities. *Biochem. J.*; 382: 59–65.
6. S.Allegri, A.Scaloni, M.G.Careddu, G.Cuccu, C.D'Ambrosio, R.Pesi, M.Camici, L.Ferrara and M.G.Tozzi (2004) Mechanistic studies on bovine cytosolic 5'-nucleotidase II, an enzyme belonging to the HAD superfamily. *Eur. J. Biochem.*; 271: 4881–4891.
7. J. E.Lunn, A. R. Ashton, M. D. Hatch and H. W. Heldt (2000) Purification, molecular cloning, and sequence analysis of sucrose-6F-phosphate phosphohydrolase from plants. *Proc. Natl. Acad. Sci. USA*; 97: 12,914–12,919.
8. A.Cronin, S.Mowbray, H.Dürk, S.Homburg, I.Fleming, B.Fisslthaler, F.Oesch, and M.Arand (2003) The N-terminal domain of mammalian soluble epoxide hydrolase is a phosphatase. *Proc. Natl. Acad. Sci. USA*; 100; 1552–1557.
9. A.Alonso, J.Sasin, N.Bottini, I.Friedberg, I.Friedberg, A.Osterman, A.Godzik, T.Hunter, J. Dixon and T.Mustelin (2004) The PTPs in the human genome. *Cell*; 117: 699–711.
10. D. Bordo and P.Bork (2002) The rhodanese/Cdc25 phosphatase superfamily. Sequence–structure–function relations. *EMBO Rep.*; 3: 741–746.
11. C.Vincent, B.Duclos, C.Grangeasse, E. Vaganay, M. Riberty, A.J. Cozzone and P.Doublet (2000) Relationship between exopolysaccharide production and protein-tyrosine phosphorylation in Gramnegative bacteria. *J. Mol. Biol.*; 304: 311–321.
12. L. Musumeci, C. Bongiorno, L. Tautz, R.A. Edwards, A. Osterman, M. Perego, T.Mustelin and N.Bottini, (2005) Low-molecular-weight protein tyrosine phosphatases of *Bacillus subtilis*. *J. Bacteriol.*; 187(14):4945-4956.

13. N. Bottini, E. Bottini, F. Gloria-Bottini and T. Mustelin (2002) LMPTP and human disease: in search of biochemical mechanisms. *Arch. Immunol. Ther. Exp.*; 50: 95–104.
14. S. Fieulaine, S. Morera, S. Poncet, V. Monedero, V. Gueguen-Chaignon, A. Galinier, J. Janin, J. Deutscher, S. Nessler (2001) X-ray structure of HPr kinase: a bacterial protein kinase with a P-loop nucleotide-binding domain. *EMBO J.*; 20(15):3917-3927.
15. T. Pawson, G. D. Gish (1992) SH2 and SH3 domains: From structure to function. *Cell*; 71(3):359–362.
16. K. Hofmann, P. Bucher (1995). The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem. Sci.*; 20: 347–349.
17. D. Durocher, S.J. Smerdon, M.B. Yaffe, S.P. Jackson (2000) The FHA Domain in DNA Repair and Checkpoint Signaling. *Cold Spring Harb. Symp. Quant. Biol.*; 65: 423-432.
18. M. Pallen, R. Chaudhuri, A. Khan (2002) Bacterial FHA domains: neglected players in the phospho-threonine signalling game? *Trends Microbiol.*; 10: 556–563.
19. J. Deutscher, E. Küster, U. Bergstedt, V. Charrier, W. Hillen, (1995) Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in Gram-positive bacteria. *Mol Microbiol* ;15: 1049–1053.
20. A. Fujita, T. Takeuchi, H. Nakajima, H. Nishio, F. Hata, (1995) Involvement of heterotrimeric GTP-binding protein and rho protein, but not protein kinase C, in agonist-induced Ca<sup>2+</sup> sensitization of skinned muscle of guinea pig vas deferens. *J. Pharmacol. Exp. Ther.*; 274: 555–561.
21. S. Schumacher, T. Gryzik, S. Tannebaum, H.A.J. Mueller (2004). The RhoGEF Pebble is required for cell shape changes during cell migration triggered by the Drosophila FGF receptor Heartless. *Development*; 131: 2631-2640.
22. Y.Y. Cheung, S.Y. Lam, W.K. Chu, M.D. Allen, M. Bycroft, K.B. Wong (2005) Crystal Structure of a Hyperthermophilic Archaeal Acylphosphatase from Pyrococcus horikoshii Structural Insights into Enzymatic Catalysis, Thermostability, and Dimerization. *Biochemistry.*; 44 (12):4601–4611.
23. M. Stefani, N. Taddei, G. Ramponi (1997) Insights into acylphosphatase structure and catalytic mechanism. *Cell Mol. Life Sci.*; 53:141–151.
24. P. Paoli, G. Camici, G. Manao, E. Giannoni, G. Ramponi (2000) Acylphosphatase possesses nucleoside triphosphatase and nucleoside diphosphatase activities. *Biochem. J.*; 349: 43–49.

25. P. Nassi, C. Nediani, G. Liguri, N. Taddei, G. Ramponi (1991) Effects of acylphosphatase on the activity of erythrocyte membrane  $\text{Ca}^{2+}$  pump. *J. Biol. Chem.*; 266: 10867–10871.
26. C. Nediani, C. Fiorillo, E. Marchetti, R. Bandinelli, D. Degl'Innocenti, P. Nassi (1995) Acylphosphatase: A potential modulator of heart sarcolemma  $\text{Na}^+, \text{K}^+$  pump. *Biochemistry.* ; 34:6668–6674.
27. C. Nediani, C. Fiorillo, E. Marchetti, A. Pacini, G. Liguri, P. Nassi (1996) Stimulation of cardiac sarcoplasmic reticulum calcium pump by acylphosphatase. Relationship to phospholamban phosphorylation. *J. Biol. Chem.*; 271:19066–19073.
28. C. Nediani, C. Fiorillo, S. Rigacci, F. Magherini, M. Francalanci, G. Liguri, A. Pacini, P. Nassi (1999) A novel interaction mechanism accounting for different acylphosphatase effects on cardiac and fast twitch skeletal muscle sarcoplasmic reticulum calcium pumps, *FEBS Lett.*; 443: 308–312.
29. C. Nediani, A. Celli, C. Fiorillo, V. Ponziani, L. Giannini, P. Nassi (2003) Acylphosphatase interferes with SERCA2a-PLN association, *Biochem. Biophys. Res. Commun.*; 301: 948–951.
30. N. Taddei, M. Stefani, F. Magherini, F. Chiti, A. Modesti, G. Raugei, G. Ramponi (1996) Looking for Residues Involved in the Muscle Acylphosphatase Catalytic Mechanism and Structural Stabilization: Role of Asn41, Thr42, and Thr46. *Biochemistry.*; 35: 7077–7083.
31. A. Pastore, V. Saudek, G. Ramponi, R.J. Williams (1992) Three-dimensional structure of acylphosphatase. Refinement and structure analysis. *J Mol Biol.*; 224:427-440.
32. D.P. Satchell, N. Spencer, G.F. White (1972) Kinetic studies with muscle acylphosphatase. *Biochim Biophys Acta.*; 268:233-248.
33. L. Shi, A. Kobir, C. Jers, I. Mijakovic (2010) Bacterial Protein-Tyrosine Kinases. *Current Proteomics*; 7: 1-7.
34. C. Grangeasse, A.J. Cozzone, J. Deutscher, I. Mijakovic (2007) Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem. Sci.*; 32:6-94.
35. C. Whitfield (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.*; 75: 39-68.
36. C. Vincent, P. Doublet, C. Grangeasse, E. Vaganay, A.J. Cozzone, B. Duclos (1999) Cells of *Escherichia coli* contain a protein tyrosine kinase, Wzc, and a phosphotyrosine-protein phosphatase, Wzb. *J. Bacteriol.*; 181: 3472-3477.

37. A.Peleg, Y.Shifrin, O.Ilan, C.Nadler-Yona, S.Nov, S.Koby, K.Baruch, S.Altuvia, M.Elgrably-Weiss, C.M.Abe, S.Knutton, M.A. Saper, I.Rosenshine (2005) Identification of an Escherichia coli operon required for formation of the O-antigen capsule. *J. Bacteriol.*; 187: 5259-5266.
38. B.Obadia, S.Lacour, P.Doublet, H.Baubichon-Cortay, A.J.Cozzone, C.Grangéasse, (2007) Influence of tyrosine-kinase Wzc activity on colanic acid production in Escherichia coli K12 cells. *J. Mol. Biol.*; 367: 42-53.
39. A.Tocilj, C.Munger, A.Proteau, R.Morona, L.Purins, E.Ajamian, J.Wagner, M.Papadopoulos, L.Van Den Bosch, J.L.Rubinstein, J.Féthière, A.Matte, M.Cygler (2008) Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. *Nat. Struct. Mol. Biol.*; 15: 130-138.
40. C.Whitfield, K.Larue (2008) Stop and go: regulation of chain length in the biosynthesis of bacterial polysaccharides. *Nat. Struct. Mol. Biol.*; 15: 121-123.
41. S. E. Whitmore, R. J. Lamont (2012) Tyrosine phosphorylation and bacterial virulence. *Int J Oral Sci.*; 4(1): 1-6.
42. P.J. Kennelly, M. Potts (1999) Life among the primitives: protein O-phosphatases in prokaryotes. *Front Biosci.*; 4:D372-85.
43. Z.Y.Zhang (2005) Functional studies of protein tyrosine phosphatases with chemical approaches. *Biochim Biophys Acta.*; 1754:100-107.
44. T.Tiganis (2002) Protein tyrosine phosphatases: dephosphorylating the epidermal growth factor receptor. *IUBMB Life.*; 53: 3-14.
45. Z.Y.Zhang, J.E.Dixon (1994). Protein tyrosine phosphatases: Mechanism of catalysis and substrate specificity. *Adv. Enzymol.*; 68:1-36.
46. G.Ramponi, M.Stefani (1997) Structural, catalytic, and functional properties of low M(r), phosphotyrosine protein phosphatases. Evidence of a long evolutionary history. *Int J Biochem Cell Biol.*; 29(2):279-292.
47. Z.Jia, D.Barford, A.J. Flint, N.K.Tonks (1995) Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science.*; 268: 1754-1758.
48. C. Persson, N. Carballeira, H. Wolf-Watz, M. Fallman (1997) The PTPase YopH inhibits uptake of Yersinia, tyrosine phosphorylation of p130Cas and FAK, and the associated accumulation of these proteins in peripheral focal adhesions. *EMBO J.*; 16: 2307-2318.
49. <https://www.jic.ac.uk/>

50. G. Taylor (2003) The Phase Problem. *Acta D*; 59(11): 1881-1890.
51. K. Kleppe, E. Ohtsuka, R. Kleppe, I. Molineux, H.G. Khorana (1971). Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.* ;56 (2): 341-361.
52. S.H.Ke, E.L.Madison (1997) Rapid and efficient site-directed mutagenesis by single-tube 'megaprimer' PCR method. *Nucleic Acids Res.*; 25 (16): 3371-3372.
53. H.C. Birnboim and J. Doly. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* ;7 (6): 1513-1523.
54. V. E. Fadoulgloua, M. Kokkinidis and N. M. Glykos (2008) Determination of protein oligomerization state: Two approaches based on glutaraldehyde crosslinking. *Analytical Biochemistry*; 373(2): 404-406.
55. T.G.G. Batty, L. Kontogiannis, O. Johnson, H.R. Powell, A.G.W. Leslie. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr.*; 67 (4): 271-281.
56. P.R.Evans (2006) Scaling and assessment of data quality. *Acta Cryst. D*; 62:72-82.
57. M.G.Rossmann, A.G.W.Leslie, S. S.Abdel-Meguid and T.Tsukihara (1979). Processing and post-refinement of oscillation camera data. *J. Appl. Cryst.*; 12: 570-581.
58. F.K.Winkler, C.E.Schutt and S.C.Harrison, (1979). The oscillation method for crystals with very large unit cells. *Acta Cryst. A*; 35: 901-911.
59. A.G.W. Leslie (2006). The integration of macromolecular diffraction data. *Acta Cryst. D*; 62: 48-57.
60. A.J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D.Winn, L. C. Storoni, R. J. Read (2007) Phaser crystallographic software. *J. Appl. Cryst.*; 40: 658-674.
61. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson (2011) Overview of the CCP4 suite and current developments. *Acta. Crystallogr. D Biol. Crystallogr.*; D67: 235-242.
62. R. J. Read, J. L. Sussman and J. Sussman (2007) Likelihood-based molecular replacement in phaser. *Evolving Methods for Macromolecular Crystallography*; 245: 91-100.
63. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R.

- Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta. Crystallogr. D Biol. Crystallogr.*; 66: 213–221.
64. G.N. Murshudov, A.A.Vagin, E.J.Dodson (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood method. *Acta. Crystallogr. D Biol. Crystallogr.*; 53: 240-255.
65. P. Emsley, K. Cowtan (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr.*; 60: 2126–2132.
66. CCP4 : The CCP4 suite: programs for protein crystallography. (1994) *Acta Crystallogr D Biol Crystallogr.*; 50: 760–763.
67. E.Krissinel, K. Hendrick (2007) Inference of macromolecular assemblies from crystalline state. *J.Mol.Biol.*;372:774–797 and PISA web server: [http://www.ebi.ac.uk/msd-srv/prot\\_int/cgi-bin/piserver](http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver)
68. K. G. Tina, R. Bhadra, N.Srinivasan (2007) PIC: Protein Interactions Calculator. *Nucleic Acids Research.*; 35:W473–W476.
69. M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson , D.G. Higgins (2007) ClustalW and ClustalX version 2. *Bioinformatics.*; 23(21): 2947-2948.
70. H. McWilliam, W. Li, M. Uludag, S.Squizzato, Y.M.Park, N.Busso, A.P.Cowley, R. Lopez (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research.*; 41:W597-600.
71. F.Corpet (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research.*; 16 (22):10881-10890.
72. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
73. E.F.Pettersen, T.D.Goddard, C.C.Huang, G.S.Couch, D.M.Greenblatt, E.C.Meng, T.E.Ferrin (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput. Chem.*; 25: 1605-1612.
74. D.Petrey, B.Honig (2003) GRASP2: Visualization, Surface Properties, and Electrostatics of Macromolecular Structures and Sequences. *Methods in Enzymology.*; 374: 492-509.
75. F.Lipmann, L. C.Tuttle (1945) A specific micromethod for the determination of acyl phosphates. *Journal of Biological Chemistry.*; 159: 21-28.

76. Z.Y.Zhang, R.L.Van Etten (1990) Purification and characterization of a low-molecular-weight acid phosphatase--a phosphotyrosyl-protein phosphatase from bovine heart. *Arch Biochem Biophys.*; 282: 39-49.
77. F. Bemporad and F.Chiti (2009) "Native-like aggregation" of the acylphosphatase from *Sulfolobus solfataricus* and its biological implications. *FEBS Lett.*; 20;583(16):2630-8.
78. N. Stein (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.*; 41: 641 - 643.
79. J.Painter, E.A.Merritt (2006) TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.*; 39: 109-111.
80. Y.Y.Cheung, S.Y.Lam, W.K.Chu, M.D.Allen, M.Bycroft and K.B.Wong, (2005) Crystal structure of a hyperthermophilic archaeal acylphosphatase from *Pyrococcus horikoshii*- structural insights into enzymatic catalysis, thermostability, and dimerization. *Biochemistry* ;44: 4601–4611.
81. Thunnissen, M.M., Taddei, N., Liguri, G., Ramponi, G. and Nordlund, P. (1997) Crystal structure of common type acylphosphatase from bovine testis. *Structure* 5, 69–79.
82. Zuccotti, S., Rosano, C., Ramazzotti, M., Degl’Innocenti, D., Stefani, M., Manao, G. and Bolognesi, M. (2004) Three-dimensional structural characterization of a novel *Drosophila melanogaster* acylphosphatase. *Acta Crystallogr. D Biol. Crystallogr.* 60, 1177–1179.
83. D.Soulat, E.Vaganay, B.Duclos, A.L.Genestier, J.Etienne, A.J. Cozzone (2002) *Staphylococcus aureus* contains two low-molecular-mass phosphotyrosine protein phosphatases. *J. Bacteriol.*; 184: 5194–5199.
84. J.P.Davis, M.M.Zhou, R.L.Van Etten (1994) Kinetic and site-directed mutagenesis studies of the cysteine residues of bovine low molecular weight phosphotyrosyl protein phosphatase. *J Biol Chem.*; 269: 8734-8740.
85. E.B.Fauman, M.A.Saper (1996) Structure and function of the protein tyrosine phosphatases. *Trends Biochem. Sci.*; 21: 413–417.
86. Golovin, K, Henrick (2008) MSDmotif: exploring protein sites and motifs. *BMC Bioinformatics.*; 9:312.
87. B.Evans, P.A.Tishmack, C.Pokalsky, M.Zhang, R.L.Van Etten (1996) Site-directed mutagenesis, kinetic, and spectroscopic studies of the P-loop residues in a low molecular weight protein tyrosine phosphatase. *Biochemistry*; 35: 13609-13617.

88. C.Madhurantakam, E.Rajakumara, P.A.Mazumdar, B.Saha, D.Mitra, H.G.Wiker, R.Sankaranarayanan, A.K.Das (2005) Crystal structure of low-molecular-weight protein tyrosine phosphatase from *Mycobacterium tuberculosis* at 1.9-Å resolution. *J. Bacteriol.*; 187:2175-2181.
89. J.Fuhrmann, B.Mierzwa, D.B.Trentini, S.Spiess, A.Lehner, E.Charpentier, T.Clausen, (2013) Structural basis for recognizing phosphoarginine and evolving residue-specific protein phosphatases in gram-positive bacteria. *Cell Rep.*; 27:1832-1839.
90. P.Chiarugi, T.Fiaschi, M.L.Taddei, D.Talini, E.Giannoni, G.Raugei, G.Ramponi (2001) Two vicinal cysteines confer a peculiar redox regulation to low molecular weight protein tyrosine phosphatase in response to platelet-derived growth factor receptor stimulation. *J Biol Chem.*; 276:33478-33487.
91. L.Tabernerero, B.N.Evans, P.A.Tishmack, R.L.Van Etten, C.V.Stauffacher (1999) The structure of the bovine protein tyrosine phosphatase dimer reveals a potential self-regulation mechanism. *Biochemistry* ; 38:11651-11658.
92. Z.Zhang, E.Harms, R.L. Van Etten (1994) Asp129 of low molecular weight protein tyrosine phosphatase is involved in leaving group protonation. *J Biol Chem.*; 269:25947-25950.
93. J.Dundas, Z.Ouyang, J.Tseng, A.Binkowski, Y.Turpaz, J.Liang (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.*; 1: 34.
94. G.Hagelueken, H.Huang , I.L.Mainprize, C.Whitfield, J.H.Naismith (2009)Crystal structures of Wzb of *Escherichia coli* and CpsB of *Streptococcus pneumoniae*, representatives of two families of tyrosine phosphatases that regulate capsule assembly. *J. Mol. Biol.*; 392: 678-688.
95. Y.Y.Wo, A.L.McCormack, J.Shabanowitz, D.F.Hunt, J.P.Davis, G.L.Mitchell, R.L.Van Etten (1992) Sequencing, cloning, and expression of human red cell-type acid phosphatase, a cytoplasmic phosphotyrosyl protein phosphatase. *J Biol. Chem.*; 267: 10856-10865.
96. N.Motamedi-Shad, T.Garfagnini, A.Penco, A.Relini, F.Fogolari, A.Corazza, G.Esposito, F.Bemporad and F.Chiti (2012) Rapid oligomer formation of human muscle acylphosphatase induced by heparan sulfate. *Nat Struct Mol Biol.*;19(5):547-554.

# **Publications**



# Atomic resolution crystal structure of VcLMWPTP-1 from *Vibrio cholerae* O395: Insights into a novel mode of dimerization in the low molecular weight protein tyrosine phosphatase family <sup>☆</sup>



Seema Nath, Ramanuj Banerjee, Udayaditya Sen <sup>\*</sup>

Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF, Bidhannagar, Kolkata 700064, India

## ARTICLE INFO

### Article history:

Received 19 May 2014

Available online 6 June 2014

### Keywords:

Bacterial protein phosphatases  
Low molecular weight protein tyrosine phosphatases  
X-ray crystallography  
Active dimer

## ABSTRACT

Low molecular weight protein tyrosine phosphatase (LMWPTP) is a group of phosphotyrosine phosphatase ubiquitously found in a wide range of organisms ranging from bacteria to mammals. Dimerization in the LMWPTP family has been reported earlier which follows a common mechanism involving active site residues leading to an enzymatically inactive species. Here we report a novel form of dimerization in a LMWPTP from *Vibrio cholerae* O395 (VcLMWPTP-1). Studies in solution reveal the existence of the dimer in solution while kinetic study depicts the active form of the enzyme. This indicates that the mode of dimerization in VcLMWPTP-1 is different from others where active site residues are not involved in the process. A high resolution (1.45 Å) crystal structure of VcLMWPTP-1 confirms a different mode of dimerization where the active site is catalytically accessible as evident by a tightly bound substrate mimicking ligand, MOPS at the active site pocket. Although being a member of a prokaryotic protein family, VcLMWPTP-1 structure resembles very closely to LMWPTP from a eukaryote, *Entamoeba histolytica*. It also delineates the diverse surface properties around the active site of the enzyme.

© 2014 Elsevier Inc. All rights reserved.

## 1. Introduction

Reversible protein phosphorylation is an important event involved in intracellular signal transduction pathways in response to triggering factors which regulate crucial metabolic activities in cell. The net result of incorporation of phosphate groups in proteins is diverse and this probably acts as a key step in cellular regulation [1]. The overall content of phosphotyrosine in cells is reciprocally controlled by protein tyrosine kinases (PTKs) which specifically phosphorylate tyrosines in proteins [2]. An abundant member of this class of proteins is the low molecular weight protein tyrosine phosphatases (LMWPTPs) ubiquitously found in all organisms ranging from prokaryotes to higher eukaryotes. LMWPTPs belong to class II PTPs which act on tyrosine phosphorylated proteins, low molecular weight aryl phosphates and natural and synthetic acyl phosphates [3].

**Abbreviations:** LMWPTP, low molecular weight protein tyrosine phosphatases; FPLC, fast protein liquid chromatography; MOPS, 3-(N-morpholino) propanesulfonic acid.

<sup>☆</sup> Protein Data Bank: Coordinates and structure factor files have been deposited with the accession code 4LRQ.

<sup>\*</sup> Corresponding author. Fax: +91 33 2337 4637.

E-mail address: [udayaditya.sen@saha.ac.in](mailto:udayaditya.sen@saha.ac.in) (U. Sen).

<http://dx.doi.org/10.1016/j.bbrc.2014.05.129>

0006-291X/© 2014 Elsevier Inc. All rights reserved.

Several structures of LMWPTP from eukaryotic organism such as human – HCPTPA (PDB: 5PNT) [4], bovine – BPTPA (PDB: 1Z12) [5], mouse – RPTPA (PDB: 2P4U) [6], yeast – LTP1 (PDB: 1D1P) [7] and protozoan parasite *Entamoeba histolytica* – EhPtp (PDB: 3IDO) [8] are available. From the prokaryotic counterpart, structures of LMWPTP from gram-positive eubacteria *Staphylococcus aureus* – PtpA (PDB: 3ROF) [9], gram-negative eubacteria *Thermus thermophilus* – TT1001 (PDB: 2CWD), gram-positive proteobacteria *Bacillus subtilis* – YwIE (PDB: 4ETI) [10], gram-negative proteobacteria *Escherichia coli* – Wzb (PDB: 2WJA) and pathogenic *Mycobacterium tuberculosis* MPtpA (PDB: 1U2Q) [11] are available. It has been reported that the self association of mammalian LMWPTP (viz. *Bos taurus* LMWPTP, BPTP) produces inactive oligomers that are in equilibrium with its active monomers [12]. Among the prokaryotic LMWPTPs weak oligomerization has been found to exist in YwIE from *B. subtilis*, PtpB from the Gram-negative bacterium *Salmonella aureus* and *E. coli* Wzb. However the mode of dimerization is similar in both the cases and takes place through the direct involvements of the active site residues and the tyrosines of the DPY-loop, leading to a catalytically inactive species [13].

Here we report a novel mode of dimerization of VcLMWPTP-1, a 17.9 kDa (155 amino acids) protein from *Vibrio cholerae* O395. Fast protein liquid chromatography and glutaraldehyde crosslinking

reveal the existence of the dimeric species of the protein in solution. Kinetic studies of VcLMWPTP-1 using para-nitro phenyl phosphate (pNPP) as the substrate depicts that the protein is active, suggesting that the mode of dimerization in VcLMWPTP-1 is different from other LMWPTPs. In an attempt to investigate the molecular mechanism of this form of dimerization, we solved a high resolution crystal structure (1.45 Å) of VcLMWPTP-1 which depicts a mode of dimerization markedly different from those reported for other dimeric LMWPTPs. The extensive dimerization interface area of VcLMWPTP-1 is about double than that of other mode of dimerization and active site residues are not seen to be involved in the oligomerization. The active site of each monomer is totally accessible to the substrate which is evident from the crystal structure where the active site of each monomer is occupied by a tightly bound MOPS molecule in a substrate like manner. Comparison of the VcLMWPTP-1 structure and surface properties with similar structures in the PDB illuminates its structural convergence with LMWPTP from a eukaryote, *E. histolytica*.

## 2. Materials and methods

### 2.1. Cloning, expression and purification

VcLMWPTP-1 protein was cloned, over expressed and purified as mentioned earlier [14]. Active site mutant of VcLMWPTP-1, C8S was cloned using two-step PCR method and inserted after the start codon of pET24b(+). The mutant was overexpressed as C-terminal His-tag and was purified in similar way of the wild type.

### 2.2. Determination of oligomers using FPLC

VcLMWPTP-1 (0.85 mg/ml, 1.75 mg/ml, 2.58 mg/ml and 4.25 mg/ml) in three different 50 mM Tris buffer, each at pH 7.6 containing 150 mM, 300 mM and 500 mM NaCl was fractionated by a Sephacryl S-100 (Amersham Biosciences) column (46 × 1.6 cm) at 0.9 MPa, pre-equilibrated with respective buffers and precalibrated with a protein mixture containing Lysozyme (MW14.3 kDa) Ovalbumin (MW 36.0 kDa) and Bovine serum albumin (MW 66.45 kDa) at room temperature. Fractions were collected at a flow rate of 0.4 ml per minute using an ÄKTAPrime chromatographic system. The elution profile was determined by monitoring the absorbance at 280 nm.

### 2.3. Crystallization, data collection and structure solution

Crystallization and data collection of VcLMWPTP-1 was reported earlier [14]. For phasing the 1.45 Å data of the crystal, the coordinates of *E. histolytica* (PDB: 3IDO) [8] was used for molecular replacement with Phaser [15] in CCP4 [16] with  $V_m$  of 2.18 (solvent content 43.63%), final TFZ = 13.3 and LLG = 157. Model building was done with Coot [17] and refinement was carried out with Phenix refine [18] with twin law -k,-h,-l. TLS refinement was performed during the final stages of refinement [19].

### 2.4. Structural analysis

Average B-factors for each residue were calculated using B average in CCP4 [20]. PISA webserver [21] and PIC server [22] were used for analysis of the structure and oligomeric state. The oligomeric state and other structural of the protein was analyzed using the. Sequence alignment of VcLMWPTP-1 with other lmwptps was done using ClustalO [23,24]. Figures were prepared using Pymol (<http://www.pymol.org>). The surface electrostatic potential was mapped using Chimera [25] [-10 kT/e (red) to +10 kT/e (blue)].

### 2.5. Enzyme kinetics

Kinetic parameters were calculated for VcLMWPTP-1 using *p*-nitrophenyl phosphate (pNPP) as the substrate as described [26]. Briefly, pNPP at a concentration range of 1–40 mM was treated with 100 μM VcLMWPTP-1 and C8S mutant and quenched with 1 N NaOH after 10 min. The absorbance of the product, *p*-nitrophenol, thus formed is measured at 405 nm. The amount of *p*-nitrophenol was calculated from the standard curve of *p*-nitrophenol. For standard curve, stock solution of *p*-nitrophenol was diluted with 0.05 N NaOH and the absorbance of the samples was measured at 405 nm. To check the effect of temperature on the enzymatic activity samples were incubated at 5 °C interval in the water-bath prior to check the absorbance.

### 2.6. Glutaraldehyde crosslinking to capture the dimeric state in solution

For crosslinking assays, 2.3% freshly prepared solution of glutaraldehyde was added to a reaction mixture of 100 μl containing about 50 μg of the protein in 50 mM MOPS, pH 7.6, 300 mM NaCl and the reaction carried out at room temperature (25 °C). Samples were collected at 15 s, 30 s, 1 min and then up to 5 min at an interval of 1 min and quenched by addition of 10 μl of 1 M Tris-HCl, pH 8.0. Cross-linked proteins were analyzed through 15% SDS-PAGE.

## 3. Results and discussion

### 3.1. VcLMWPTP-1 forms dimer in solution

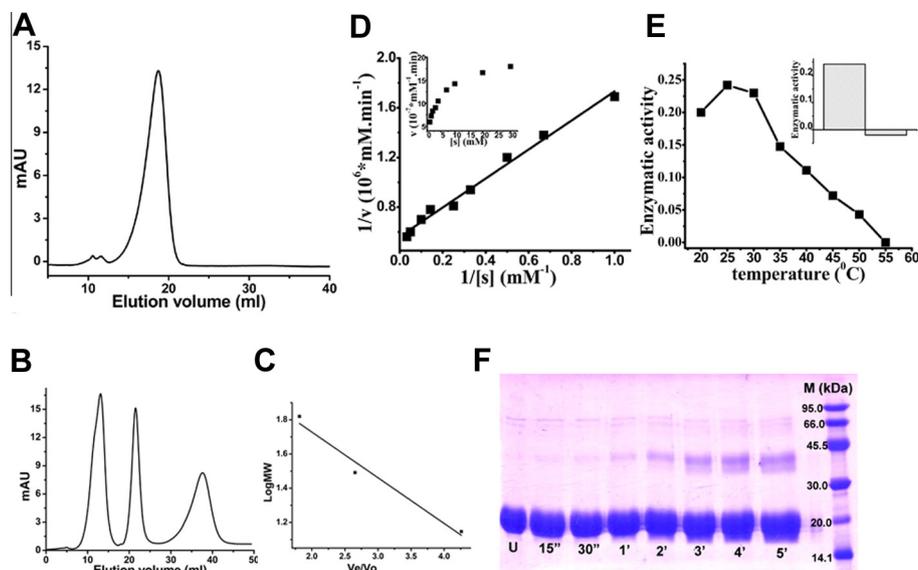
VcLMWPTP-1 elutes as a single peak in fast protein liquid chromatography (FPLC) (Fig. 1A) and calculations based upon standard curve shows that this species is the dimeric form of VcLMWPTP-1 (Fig. 1B and C). Moreover, presence of the dimeric form is observed irrespective of protein or salt (NaCl) concentration. Further confirmation of the dimeric form was performed through glutaraldehyde crosslinking where a gradual increase in the dimeric band with time could be seen in 15% SDS-PAGE (Fig. 1F).

### 3.2. Phosphatase activity of VcLMWPTP-1

Enzyme kinetic assays for VcLMWPTP-1 were performed at pH 4.8 and pH 7.6 at 25 °C using *p* nitrophenyl phosphate as a substrate (Materials and Methods). The determined  $K_m$  value under these conditions was  $2.07 \pm 0.2$  mM and  $2.03 \pm 0.4$  mM respectively (Fig. 1D), which is consistent with the  $K_m$  values reported for other LMWPTPs like PtpA and PtpB (1.2 and 1.5 mM) [27]. This leads us to propose that the respective active site of the dimer is accessible to the substrate. So the dimer formed in case of VcLMWPTP-1 does not involve active site residues as commonly found in previously reported inactive dimeric species of LMWPTPs and pH variation has no effect on its catalytic activity. The enzyme shows optimum activity at 25 °C and the activity decreases with increasing temperature leading to complete loss of activity at 55 °C (Fig. 1E). Mutating the active site Cys8 to Ser (C8S) results in complete loss of enzymatic activity (Fig. 1E inset) as reported for other LMWPTPS [28].

### 3.3. Overall structure of VcLMWPTP-1 monomer

VcLMWPTP-1 crystallized in space group P3<sub>1</sub> using ammonium sulfate (AMS) as precipitant and the structure has been solved at 1.45 Å (Supplementary Table S1). The asymmetric unit contains four molecules of VcLMWPTP-1 and their arrangement suggests that molecules A and C and molecules B and D assemble together



**Fig. 1.** Existence of oligomeric VcLMWPTP-1 in solution and its activity. (A) FPLC analysis of VcLMWPTP-1. The elution volume of the peak-2 corresponds to a dimeric VcLMWPTP-1 species, peak-1 is impurity. (B) Calibration of the column using a mixture of proteins (peak 1-Lysozyme, MW14.3 kDa; peak 2-ovalbumin, MW 36.0 kDa and peak 3-Bovine serum albumin, MW 66.45 kDa) and (C) the standard curve is drawn based on the elution profile of protein mixtures of known molecular weights. (D)  $V/[S]$  graph of VcLMWPTP-1 (inset), Double Reciprocal plot of VcLMWPTP-1 at pH 4.8. (E) Enzymatic activity ( $\mu\text{moles min}^{-1} \mu\text{g}^{-1}$ ) of VcLMWPTP-1 with temperature ( $^{\circ}\text{C}$ ); optimum enzymatic activity is observed at  $25^{\circ}\text{C}$  while complete loss of activity occurs at  $55^{\circ}\text{C}$ . Comparative enzymatic activity of VcLMWPTP-1 and C8S mutant (inset) at  $25^{\circ}\text{C}$  demonstrating that Cys to Ser mutation at the active site results in the complete loss of catalytic activity. (F) 15% SDS-PAGE profile of Glutaraldehyde crosslinked VcLMWPTP-1 showing a gradual increase in the dimeric species ( $\sim 40$  kDa) with time. Lane 1: untreated protein, lane 2–8: glutaraldehyde treated protein, lane 10: protein marker. The band corresponding to the dimeric VcLMWPTP-1 becomes prominent from 1 min between 45.0 and 30.0 kDa.

to form two dimers. The high resolution electron density map allowed us to locate 149 residues out of 155 residues of each VcLMWPTP-1 molecule. Each VcLMWPTP-1 binds a MOPS molecule in its active site and one sulfate at the C-terminal loop. The fold and overall secondary structure is comprised of three layer  $\beta\alpha\beta$  sandwiched architecture with a topology of Rossmann fold and the four stranded parallel  $\beta$ -sheet is sandwiched between three long ( $\alpha 1$ ,  $\alpha 4$  and  $\alpha 6$ ) and three short ( $\alpha 2$ ,  $\alpha 3$ ,  $\alpha 5$ ) helices (Fig. 2A).

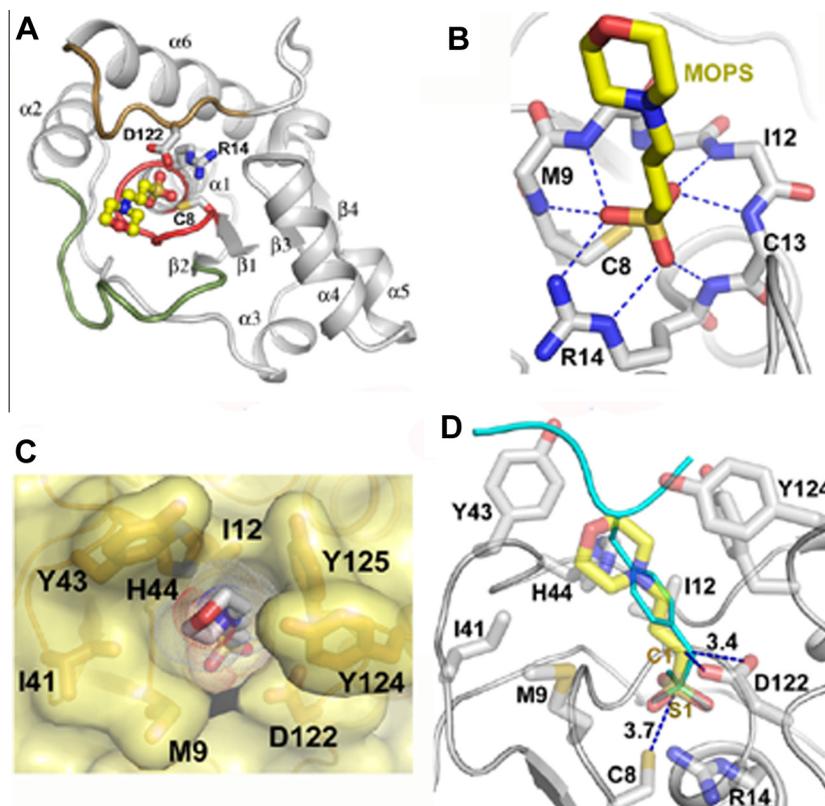
### 3.4. Ligand bound 'closed structure' conformation

The active site is located at the crevice formed by loop-1 between  $\beta 1$  and  $\alpha 1$ , also called P-loop. It is surrounded by long, flexible loop-3 and short loop-6 and the DPY-loop, which is a portion of loop-8 preceding the last helix ( $\alpha 6$ ) that contributes residue (D122) necessary for the catalytic action (Fig. 2A). From the well defined electron density it was evident that each VcLMWPTP-1 molecule tightly binds a ligand molecule, MOPS at the active site. Main-chain amide nitrogens of the residues constituting the P-loop and the side-chain amide nitrogens of Arg14 are engaged in the binding and stabilization of the sulfonate group of the ligand (Fig. 2B). A stereo representation of the electron density map around the active site of VcLMWPTP-1 molecule is shown in Supplementary Fig. S1. The piperazine ring of MOPS molecule is surrounded with several polar and aromatic residues (Thr40, Tyr43, His44, Asn47, Asp87, Glu89, Asn90, Asp122, Tyr124, and Tyr125) and they define the architecture of the wall of the deep phosphotyrosine binding pocket of VcLMWPTP-1 with an opening at one side (Fig. 2C). The ligand mimics the substrate phosphotyrosine as seen in PTP1B (PDB: 1G1F) [29] (Fig. 2D). The catalytic  $\text{S}^{\text{Y}}$  of Cys8 point towards the S1-atom of MOPS (that corresponds to the P atom of Phosphotyrosine) at a distance of  $3.8 \text{ \AA}$  and is in a position to accept the phosphoryl group. Asp122 of DPY-loop comes close to the P-loop after MOPS binding and its side-chain oxygen atoms are within  $3.4 \text{ \AA}$  with the C1 atom of the MOPS

molecule (that corresponds to the phenolate O atom of the Phosphotyrosine). At this distance Asp122 serves as an acid catalyst for the leaving alcoholic group. The covering of MOPS molecule with Thr40, Tyr43, His44, Asn47, Asp87, Glu89, Asn90, Asp122, Tyr124, Tyr125 and the disposition of catalytic Cys8 and Asp122 near the leaving group together resembles a substrate bound 'closed structure'. This closure at the P-loop pocket is also evident from the structural alignment with an 'open structure' of apo-MPtpA (PDB: 2LUO) [30] to the ligand-bound structures (PDB: 1U2P, 1U2Q) [11] where the position of DPY-loop is shown to be closer to P-loop in case of the ligand-bound structure than the apo-form [30].

### 3.5. Multiple sequence alignment showing conserved motifs

Multiple sequence alignment and 3D superposition of LMWPTP structures determined from other organisms show that their active site is highly conserved (Supplementary Fig. S2). Among these structures VcLMWPTP-1 shares highest sequence identity (43%) with that of *E. histolytica* (PDB: 3JVI) [8] with an average RMSD of superposition  $0.48 \text{ \AA}$  for 122 residues. Structural superposition indicates that the end of the distorted  $\alpha 4$  and region preceding the catalytically important DPY-loop exhibits highest variability. Comparison of the structure of VcLMWPTP-1 to PtpA, LMWPTP from gram-positive eubacteria *Staphylococcus aureus* (PDB: 3ROF) [9]; YwIE, LMWPTP from gram-positive proteobacteria *B. subtilis* (PDB: 4ETI) and Wzb, LMWPTP from gram-negative proteobacteria *E. coli* (PDB: 2WJA) [10] also indicates significant structural difference around loop-3 which is closer to the P-loop. Besides the active site P-loop and DPY-loop, two more conserved sequences (motif) are evident from the alignment result. Details of the four motifs and possible role of the conserved residues are listed in Supplementary Table S2. The active site structure of PTPs follow the active site motif of  $\text{CX}_5\text{R}$ , the only common feature between the large PTP family and LMWPTP [31]. Sequence alignment helps to dissect the



**Fig. 2.** Overall structure and MOPS binding at the active site of VcLMWPTP-1. (A) Cartoon representation of the secondary structure of VcLMWPTP-1 monomer. The MOPS molecule (ball and sticks) at active site where residues C8, R14 in red colored P-loop, D122 in brown colored DPY loop and loop  $\alpha 2$ - $\beta 2$  (olive) responsible for catalytic action. (B) Extensive hydrogen bonds made by the sulfonyl group of the MOPS molecule with the amide environment of the P-loop and the side chain of R14. (C) Surface representation of VcLMWPTP-1 depicting the residues defining the active site pocket (surface and stick) that tightly bind the MOPS molecule (dots and stick). (D) Similarity of binding of MOPS with that of phosphotyrosine (cyan) (PDB: 1XXV) binding. The residues of VcLMWPTP-1 that define the active site along with the MOPS molecule are shown as transparent sticks while the side chain of nonhydrolyzable phosphotyrosine in thick lines (cyan). The position of sulfonyl group of the MOPS molecule (stick) is almost indistinguishable with the phosphate group of phosphotyrosine. Distance between S $\gamma$  atom of Cys8 and the S1 atom of MOPS (that corresponds the P atom of phosphotyrosine) is 3.7 Å, capable to accept the phosphoryl group. Distance between OD1 (or OD2) of D122 and C1 of MOPS (that corresponds the OH atom of phosphotyrosine) is 3.4 Å and in a position to serve as an acid catalyst. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

'X<sub>5</sub>' portion of the P-loop sequence revealing '(T/L/M)GN(I/L)C' as a consensus motif in this family, contrary to the notion of being tagged as 'any amino acid'. It is also to be noted that the flanking residue(s) beside the P-loop is hydrophobic in nature. Motif-1 and Motif-4 are parts of P-loop and DPY-loop respectively. Motif-2 containing loop-2 is in close proximity to the active site whereas Motif-3 containing  $\beta 3$  is remotely distant from the active site. Motif-2 region superposes well in 3D structures between VcLMWPTP-1 with eukaryotic LMWPTPs but a marked distortion with the prokaryotic counterpart is evident, especially with Wzb, YwIE and PtpA.

### 3.6. Dimerization of VcLMWPTP-1: role of a set of unique residues

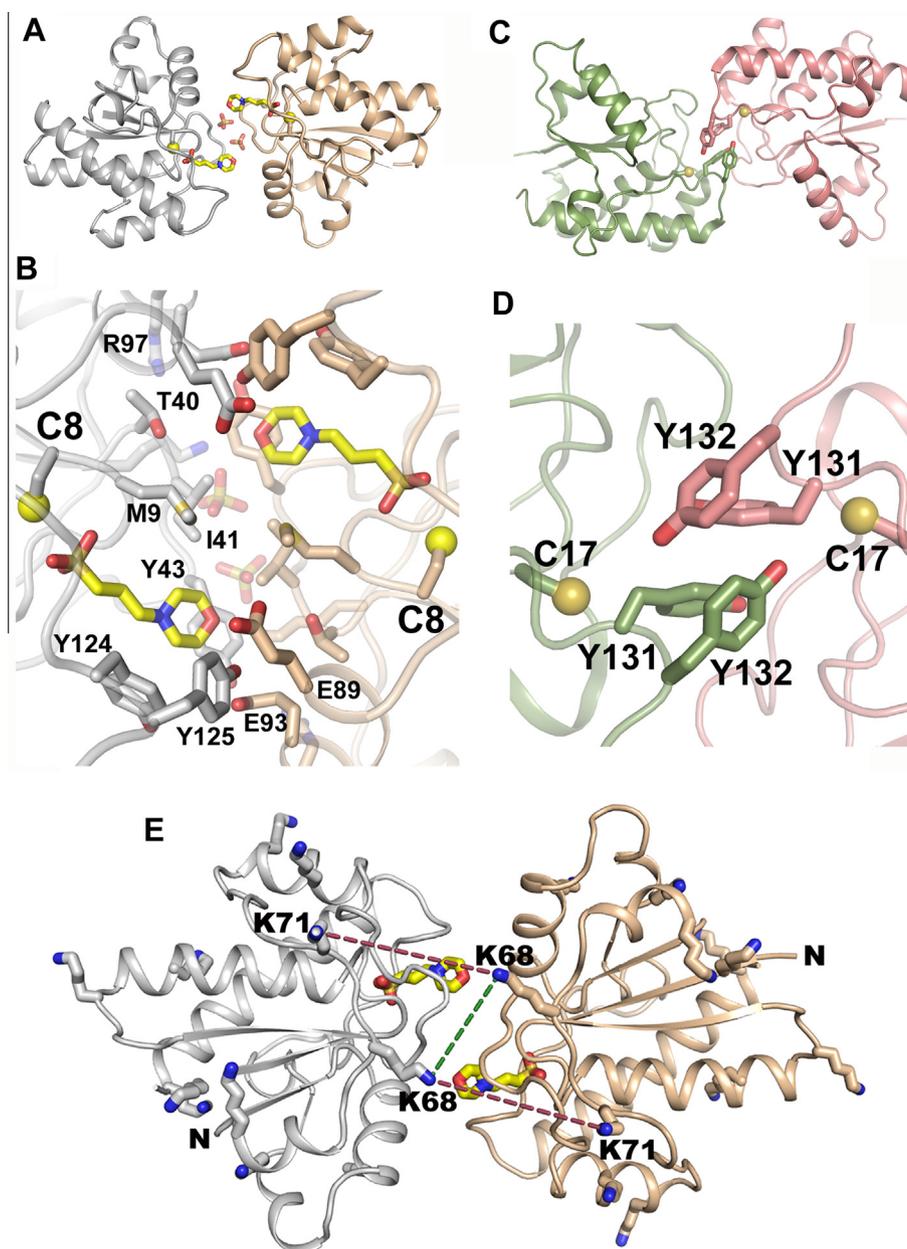
Crystal structure of VcLMWPTP-1 shows that each dimer (A:C and B:D) faces an extended surface area that is further stitched by two sulfate ions at the interfacial region (Fig. 3A). Calculation of buried surface area (BSA) by PISA [21] shows 2910 Å<sup>2</sup> BSA for the B:D dimer and 2820 Å<sup>2</sup> BSA for the A:C dimer. The BSA of both the dimers is ~20% of their respective accessible surface area (ASA) and ~10% of their total ASA. PISA server also indicates that the average  $\Delta G_{\text{diss}}$  value of both the dimers is ~14 kcal.mol<sup>-1</sup> indicating they are stable in solution. At the interfacial region, twelve residues (Met9, Thr40, Ile41, Gly42, Tyr43, Lys71, Glu89, Ala92, Glu93, Arg97, Tyr124, and Tyr125) play pivotal role to stabilize the dimers through numerous hydrophobic and H-bonding interactions (Fig. 3B). Tyr43 and Lys71 from each chain are involved in cat-

ion-TT interaction with its dimeric counterpart. Sulfate ions which stitch the two monomers interacts with the main chain NH of Gly42 and the side chain NZ of Lys71 of one monomer and the main chain NH of Tyr43 of the other monomer. In contrast to the buried surface area (BSA) of VcLMWPTP-1 (average BSA 2865 Å<sup>2</sup>) the dimerization region of BPTP encompasses a much smaller BSA (1589 Å<sup>2</sup>) that consist only the residues from the active site and DPY-loop (Fig. 3C and D).

Analysis of the position of Lys residues at the dimeric interface of the VcLMWPTP-1 (Fig. 3E) shows only a pair of amine (Lys68 from each monomer at a distance of ~10 Å) which could crosslink with each other in presence of glutaraldehyde (Fig. 3E). Distance between all other inter-monomeric Lys residues and the N terminus amines are beyond the condensable distance. As this mode of dimerization is stabilized by hydrophobic and polar interactions, it would show a monomeric band in SDS-PAGE. However, if the dimer is crosslinked by glutaraldehyde a band near its dimer appears with time. Since the distance between Lys68 in the dimer is little larger than the condensable distance (6–8 Å), more dimers become crosslinked with gradual increase in time (Fig. 1F).

### 3.7. Insight into substrate specificity – diverged surface property around the active site

CASTp server [32] shows 256 Å<sup>3</sup> as the volume of the active site cavity. Comparison of the structure and electrostatic surface around the active site of VcLMWPTP-1 with others reveals



**Fig. 3.** Comparison of dimeric interface of VcLMWPTP-1 (left) with that of bovine LMWPTPs (BPTP) (right). (A) Cartoon representation of the VcLMWPTP-1 dimer (gray and wheat) with the bound MOPS molecule (yellow) and sulfate ions (in sticks). (B) Close-up view of the dimeric interface of VcLMWPTP-1 highlighting the residues involved in dimerization. Y124 and Y125 of the DPY loop are engaged in dimerization but keep the catalytic pocket accessible. (C) Cartoon representation of the BPTP dimer (green and salmon). (D) Close-up view showing Y131, Y132 of the DPY loop of one molecule comes in close proximity to the active site of the other molecule resulting in the inactivation of the enzymes. (E) Dimeric structure of VcLMWPTP-1 and amine residues (Lysine and N-terminus residues) that may react with glutaraldehyde. The smallest distance is 10 Å (shown in green dashed line) between K68 from each monomer that can cross-link with each other while the next larger distance is 17 Å (red dashed line) which is beyond the cross-linking range of glutaraldehyde. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

close resemblance with EhPtp and other mammalian LMWPTPs than with LMWPTPs of bacterial origin (Supplementary Fig. S3). VcLMWPTP-1 differs with prokaryotic LMWPTPs at two regions which are implicated in substrate recognition and binding. The W-loop, harboring a Trp-residue almost in every LMWPTP, is proposed to play an important role in substrate recognition [10]. However, corresponding residue is Tyr43 in VcLMWPTP-1 and Tyr49 in HCPTPA, important for substrate recognition for the latter. But in bacteria, Tyr43 is substituted by Leu44 in Wzb and Ser42 in YwIE. Asn53 in HCPTPA plays a crucial role in determining the substrate specificity [3], corresponding residue Asn47 in VcLMWPTP-1 may play a similar role. Again this is in sharp contrast with bacterial

LMWPTPs like Wzb, Ptp where the corresponding residue K43 and Y44 respectively play indispensable role in substrate recognition [10]. Overall, these differences from prokaryotic LMWPTPs make VcLMWPTP-1 active site more like eukaryotic than of prokaryotic origin which may dictate its substrate specificity.

In conclusion, we report an exclusive dimeric species of VcLMWPTP-1 in solution through FPLC while atomic resolution X-ray structure shows that the dimeric species is 'novel' where the active site is not occluded. This dimer is catalytically active as demonstrated by kinetic studies while a MOPS molecule bound in a substrate like manner at the active site proves the active pocket is accessible for the substrate. VcLMWPTP-1 also distinguishes itself

from other bacterial LMWPTPs in terms of electrostatic charge distribution around the active site thereby suggesting close similarity in substrate(s) preference to eukaryotic LMWPTPs, than its bacterial homologue. Since the dimeric surface is closer to the active site it may further modulate the substrate recognition and specificity.

### Acknowledgments

The laboratory of US is partly supported by the MSACR project, DAE, Govt. of India, SINP. RB is a Senior Research Fellow of the Council of Scientific and Industrial Research, India. SN is a Senior Research Fellow of the Department of Atomic Energy, India. US thank Dr. J. Dasgupta for her help during FPLC experiments.

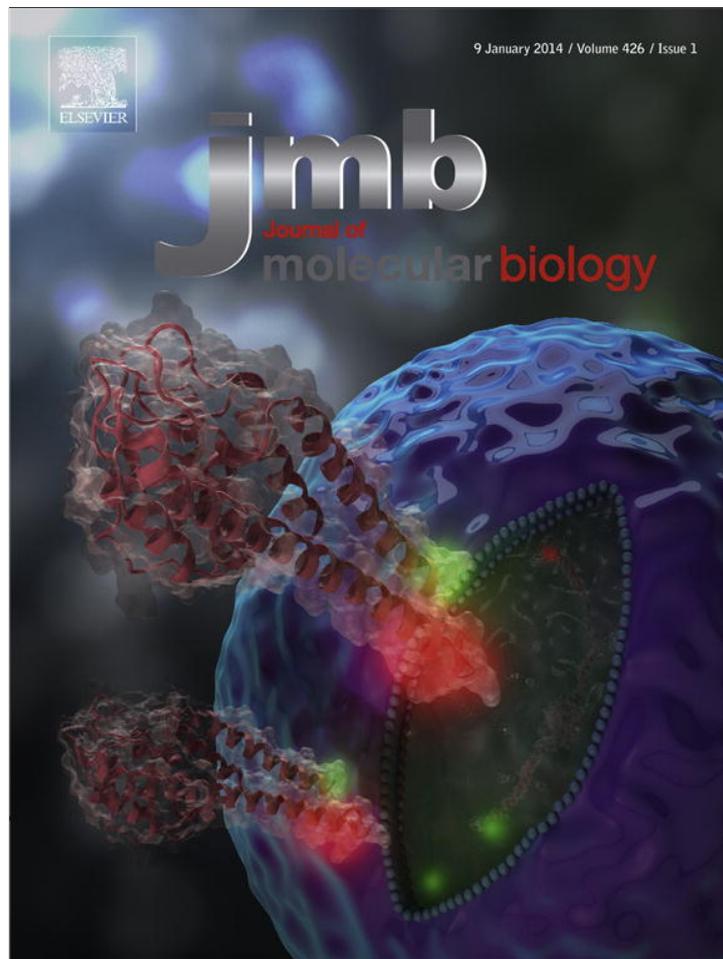
### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bbrc.2014.05.129>.

### References

- [1] L.N. Johnson, The regulation of protein phosphorylation, *Biochem. Soc. Trans.* 37 (2009) 627–641.
- [2] T. Mustelin, K. Taskén, Positive and negative regulation of T-cell activation through kinases and phosphatases, *Biochem. J.* 371 (2003) 15–27.
- [3] Y.Y. Wo, A.L. McCormack, J. Shabanowitz, D.F. Hunt, J.P. Davis, G.L. Mitchell, R.L. Van Etten, Sequencing, cloning, and expression of human red cell-type acid phosphatase, a cytoplasmic phosphotyrosyl protein phosphatase, *J. Biol. Chem.* 267 (1992) 10856–10865.
- [4] M. Zhang, C.V. Stauffacher, D. Lin, R.L. Van Etten, Crystal structure of a human low molecular weight phosphotyrosyl phosphatase. Implications for substrate specificity, *J. Biol. Chem.* 273 (1998) 21714–21720.
- [5] M. Zhang, M. Zhou, R.L. Van Etten, C.V. Stauffacher, Crystal structure of bovine low molecular weight phosphotyrosyl phosphatase complexed with the transition state analog vanadate, *Biochemistry* 36 (1997) 15–23.
- [6] S.C. Almo, J.B. Bonanno, J.M. Sauder, S. Emtege, T.P. Dilorenzo, V. Malashkevich, S.R. Wasserman, S. Swaminathan, S. Eswaramoorthy, R. Agarwal, D. Kumaran, M. Madegowda, S. Ragumani, Y. Patskovsky, J. Alvarado, U.A. Ramagopal, J. Faber-Barata, M.R. Chance, A. Sali, A. Fiser, Z.Y. Zhang, D.S. Lawrence, S.K. Burley, Structural genomics of protein phosphatases, *J. Struct. Funct. Genomics* 8 (2007) 121–140.
- [7] S. Wang, L. Taberero, M. Zhang, E. Harms, R.L. Van Etten, C.V. Stauffacher, Crystal structures of a low-molecular weight protein tyrosine phosphatase from *Saccharomyces cerevisiae* and its complex with the substrate *p*-nitrophenyl phosphate, *Biochemistry* 39 (2000) 1903–1914.
- [8] A.S. Linford, N.M. Jiang, T.E. Edwards, N.E. Sherman, W.C. Van Voorhis, L.J. Stewart, P.J. Myler, B.L. Staker, W.A. Petri Jr., Crystal structure and putative substrate identification for the *Entamoeba histolytica* low molecular weight tyrosine phosphatase, *Mol. Biochem. Parasitol.* 193 (2014) 33–44.
- [9] C. Vega, S. Chou, K. Engel, M.E. Harrell, L. Rajagopal, C. Grundner, Structure and substrate recognition of the *Staphylococcus aureus* protein tyrosine phosphatase PtpA, *J. Mol. Biol.* 413 (2011) 24–31.
- [10] G. Hagelueken, H. Huang, I.L. Mainprize, C. Whitfield, J.H. Naismith, Crystal structures of Wzb of *Escherichia coli* and CpsB of *Streptococcus pneumoniae*, representatives of two families of tyrosine phosphatases that regulate capsule assembly, *J. Mol. Biol.* 392 (2009) 678–688.
- [11] C. Madhurantakam, E. Rajakumara, P.A. Mazumdar, B. Saha, D. Mitra, H.G. Wiker, R. Sankaranarayanan, A.K. Das, Crystal structure of low-molecular-weight protein tyrosine phosphatase from *Mycobacterium tuberculosis* at 1.9-Å resolution, *J. Bacteriol.* 187 (2005) 2175–2181.
- [12] L. Taberero, B.N. Evans, P.A. Tishmack, R.L. Van Etten, C.V. Stauffacher, The structure of the bovine protein tyrosine phosphatase dimer reveals a potential self-regulation mechanism, *Biochemistry* 38 (1999) 11651–11658.
- [13] J. Blobel, P. Bernadó, H. Xu, C. Jin, M. Pons, Weak oligomerization of low-molecular-weight protein tyrosine phosphatase is conserved from mammals to bacteria, *FEBS J.* 276 (2009) 4346–4357.
- [14] S. Nath, R. Banerjee, S. Khamrui, U. Sen, Cloning, purification, crystallization and preliminary X-ray analysis of two low-molecular-weight protein tyrosine phosphatases from *Vibrio cholerae*, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 68 (2012) 1204–1208.
- [15] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, *J. Appl. Crystallogr.* 40 (2007) 658–674.
- [16] E. Potterton, P. Briggs, M. Turkenburg, E. Dodson, A graphical user interface to the CCP4 program suite, *Acta Crystallogr. D Biol. Crystallogr.* 59 (2003) 1131–1137.
- [17] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.* 60 (2004) 2126–2132.
- [18] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive python-based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr.* 66 (2010) 213–221.
- [19] J. Painter, E.A. Merritt, TLSMD web server for the generation of multi-group TLS models, *J. Appl. Crystallogr.* 39 (2006) 109–111.
- [20] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, *Acta Crystallogr. D* 67 (2011) 235–242.
- [21] E. Krissinel, K. Hendrick, Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol.* 372 (2007) 774–797 and PISA web server: <[http://www.ebi.ac.uk/msd-srv/prot\\_int/cgi-bin/piserver](http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver)>.
- [22] K.G. Tina, R. Bhadra, N. Srinivasan, PIC: protein interactions calculator, *Nucleic Acids Res.* 35 (2007) W473–W476.
- [23] F. Sievers, A. Wilm, D.G. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.* 11 (2011) 539.
- [24] M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, R. Lopez, A new bioinformatics analysis tools framework at EMBL-EBI, *Nucleic Acids Res.* 38 (2010) W695–W699.
- [25] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera – a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612.
- [26] Z.Y. Zhang, R.L. Van Etten, Purification and characterization of a low-molecular-weight acid phosphatase – a phosphotyrosyl-protein phosphatase from bovine heart, *Arch. Biochem. Biophys.* 282 (1990) 39–49.
- [27] D. Soulat, E. Vaganay, B. Ducloux, A.L. Genestier, J. Etienne, A.J. Cozzone, *Staphylococcus aureus* contains two low-molecular-mass phosphotyrosine protein phosphatases, *J. Bacteriol.* 184 (2002) 5194–5199.
- [28] J.P. Davis, M.M. Zhou, R.L. Van Etten, Kinetic and site-directed mutagenesis studies of the cysteine residues of bovine low molecular weight phosphotyrosyl protein phosphatase, *J. Biol. Chem.* 269 (1994) 8734–8740.
- [29] A. Salmeen, J.N. Andersen, M.P. Myers, N.K. Tonks, D. Barford, Molecular basis for the dephosphorylation of the activation segment of the insulin receptor by protein tyrosine phosphatase 1B, *Mol. Cell* 6 (2000) 1401–1412.
- [30] T. Stehle, S. Sreeramulu, F. Löhr, C. Richter, K. Saxena, H.R. Jonker, H. Schwalbe, The apo-structure of the low molecular weight protein-tyrosine phosphatase A (MptpA) from *Mycobacterium tuberculosis* allows for better target-specific drug development, *J. Biol. Chem.* 287 (2012) 34569–34582.
- [31] E.B. Fauman, M.A. Saper, Structure and function of the protein tyrosine phosphatases, *Trends Biochem. Sci.* 21 (1996) 413–417.
- [32] J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang, CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, *Nucleic Acids Res.* 1 (2006) 34.

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.

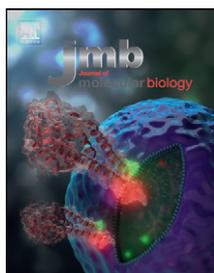


This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



## A Novel 8-nm Protein Cage Formed by *Vibrio cholerae* Acylphosphatase

Seema Nath, Ramanuj Banerjee and Udayaditya Sen

Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF, Bidhannagar, Kolkata 700064, India

Correspondence to :

<http://dx.doi.org/10.1016/j.jmb.2013.09.014>

Edited by T. Yeates

### Abstract

Here we show the formation of an ~8-nm cage formed by the self-assembly of acylphosphatase from *Vibrio cholerae* O395 (Vc-AcP). The 12-subunit cage structure forms spontaneously and is stabilized through binding of sulfate ions at its exterior face and interfacial regions. Crystal structure and studies in solutions illuminate the basis for the formation of the cage, while a single (Cys20 → Arg) mutation (Vc-AcP-C20R) transforms Vc-AcP to a potent enzyme but disrupts the assembly into a trimer.

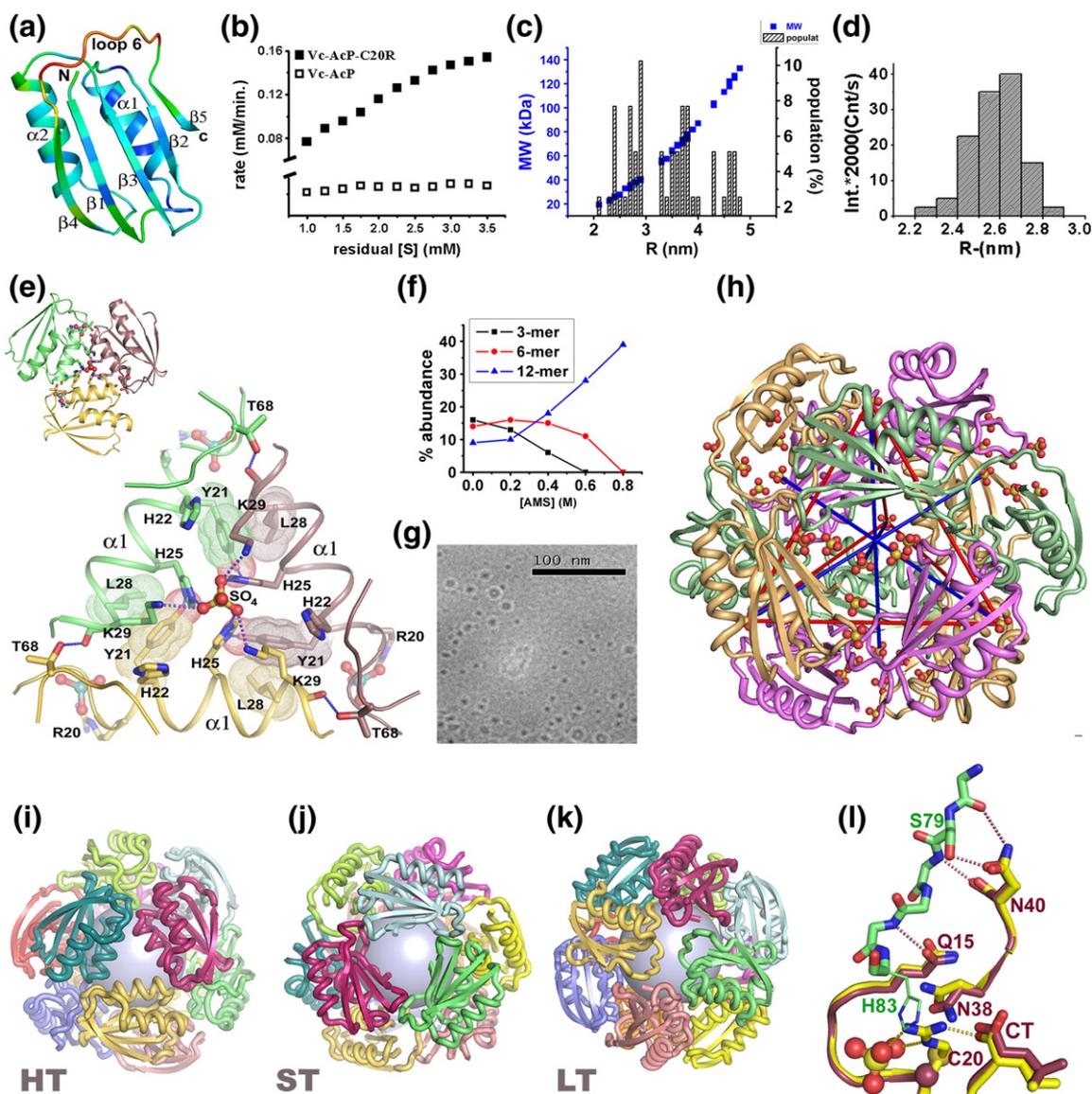
© 2013 Elsevier Ltd. All rights reserved.

Proteins can self-assemble in a highly specific and constrained way to form diverse architectures, from tubes to cages. Such architectures provide a suitable template for manipulation to create materials with a wide range of applications from biomedicine to material science [1–4]. Acylphosphatases (AcPs) are members of the hydrolase family, possessing ferredoxin-like folds (Fig. 1a), and are among the smallest enzymes known (molecular mass, ~10 kDa). Oligomerization of AcPs to form large aggregates is a common phenomenon, which has been reported for various sources [5,6]. Vc-AcP, a natural variant of *Vibrio cholerae* O395, shows nominal enzymatic activity due to the presence of a cysteine in place of catalytic arginine 20. A kinetic assay of Vc-AcP using acetylphosphate as the substrate has established  $K_m$  and  $K_{cat}$  values of  $8.53 \pm 0.7$  mM and  $9.29 \pm 0.3 \times 10^{-5} \text{ min}^{-1}$ , respectively (Fig. 1b). A point mutant, Vc-AcP-C20R, however, turns out to be a potent enzyme with  $K_m$  and  $K_{cat}$  values of  $2.36 \pm 0.6$  mM and  $2.55 \pm 0.5 \text{ min}^{-1}$ , respectively (Fig. 1b), indicating that the mutant is nearly  $10^5$  times more efficient in catalysis than the wild-type protein.

In our solution studies, Vc-AcP shows three different oligomeric states in solution, irrespective of the pH. The hydrodynamic radius (RH) and expected molecular mass (MM) of the oligomers vary from  $2.5 \pm 0.3$  nm and  $32 \pm 8$  kDa to  $3.7 \pm 0.3$  nm and  $68 \pm 12$  kDa, to  $4.6 \pm 0.3$  nm and  $119 \pm 13$  kDa (Fig. 1c), whereas Vc-AcP-C20R shows only one population with the RH and MM equivalent to that of the smallest oligomer of Vc-AcP (Fig. 1d). The crystal structure of the

Vc-AcP-C20R mutant (1.96 Å) shows a triangular trimeric assembly (Fig. 1e, Supplementary Table 1), the size and mass of which matches well with that of the oligomeric state obtained in DLS (dynamic light scattering) experiments (Fig. 1d). The helix-mediated trimeric assembly (HT) is stabilized by intricate hydrogen bonding network through a set of residues (Y21, H22, H25, L28, and K29 from helix  $\alpha_1$  of each monomer) that are specific for Vc-AcP compared to other AcPs.

Being evidently heterogeneous with respect to assembly state, wild-type Vc-AcP was not suitable for crystallization. Extensive pH and additive screening identified that tetrahedral anions such as sulfate, vanadate, phosphate, tungstate, and arsenate (~0.8 M) are capable of shifting the multimodal population to a monomodal state suitable for crystallization (Fig. 1f). This oligomeric state corresponds to the highest MM and RH. We determined the crystal structure of Vc-AcP (3.1 Å) revealing 12 monomers in the asymmetric unit, which are arranged to form a hollow cage-like structure obeying tetrahedral ( $T_{32}$ ) cubic symmetry and having outer and inner diameters of ~8 nm and ~4.5 nm, respectively (Fig. 1h, Supplementary Table 1). This cage-like self-assembly is also supported by the program PISA [7], which reports a buried surface area of  $26,040 \text{ \AA}^2$  (40% of the total surface area) and an estimated standard state free energy of dissociation  $\Delta G_{diss}$  of 105.5 kcal/mol. A total of 43 sulfate ions could be located at the junction points between monomers (Fig. 1h). Formation of cage-like



**Fig. 1.** (a) Cartoon representation of the AcP monomer showing its ferredoxin-like fold. The monomer is colored based on *B*-factor (from blue to red) with blue being the least flexible and red being the most flexible. (b) Hydrolysis of acetylphosphate by wild-type Vc-AcP and the single site mutant Vc-AcP-C20R. (c) DLS profile of Vc-AcP showing the population of three different oligomeric states with increasing molecular mass (MM, kDa) and hydrodynamic radius (*R*, nm). The data were taken without any additives. (d) DLS profile of the Vc-AcP-C20R mutant showing a monomodal population with hydrodynamic radius *R*<sub>H</sub> (2.5 ± 0.3 nm). (e) Crystal structure of the Vc-AcP-C20R trimeric assembly. Vc-AcP-C20R chains (yellow, green, and cadbury) are shown in cartoon, the central trimerization region is shown in sticks, and the central sulfate and active-site molybdate ions are shown in a ball-and-stick representation. Zoomed view of the triangular helix-mediated trimeric assembly depicts that the corners of the triangle are occupied by three molybdenum (molybdenum in green and oxygen in red color), and the corners are clamped by the hydrogen bonding between the main-chain carbonyls of K29 and the side chains of T68. (f) Percent abundance of three different oligomeric states (trimer, hexamer, and dodecamers) of Vc-AcP in the presence of increasing concentration of ammonium sulfate. At 0.8 M ammonium sulfate, the solution turns into a monomodal population containing only dodecameric Vc-AcP. (g) Transmission electron microscopy images showing the formation of the assembly in solution. (h) Tetrahedral (shown in red) arrangement of Vc-AcP monomers with four 3-fold rotation axes (blue lines); three 2-fold axes (not shown in the figure) pass through the middle of two opposite edges (red line). (i–k) Three different junction points “HT”, “ST”, and “LT” of in the Vc-AcP protein cage, where “HT” is the Vc-AcP-C20R-like helix-mediated trimeric assembly, “ST” is formed by three monomers that associate through their beta-sheets, and “LT” is the junction point of the monomers through their loop 6 and loop 1. A gray sphere is drawn inside to emphasize the hollow interior. (l) Vc-AcP-C20R (yellow) is superposed on one of the monomers (magenta) of the Vc-AcP cage to show how R20 (shown in stick) clashes with H83 (shown in line) of other monomer and hinders the cage formation (the sulfate ion is shown in sticks).

assembly was also confirmed through transmission electron microscopy, where circular dotted structures with an average diameter of 7–9 nm could be observed (Fig. 1g). In the cage-like structure, 12 monomers are assembled in the form of four trimers, where each trimer resembles the helix-mediated Vc-AcP-C20R assembly (HT). The trimers occupy the corners of a tetrahedron and are so arranged that, at each face of a tetrahedron, three monomers, from three different trimers, interact with each other to form a beta-sheet mediated trimer (ST), whereas along each axis of the tetrahedron, they form a loop-mediated tetrameric junction (LT) (Fig. 1i–k).

It is an intriguing observation that a single point mutation converts the marginally active enzyme to a potent form, while simultaneously disrupting the cage assembly. The crystal structure provides some insight. Loop 6 is the most flexible portion of Vc-AcP. It interacts with loop 1 at the LT interface. Superposition of the Vc-AcP-C20R trimer on the Vc-AcP cage shows that, upon sulfate binding, the arginine 20 side chain of the mutant comes in close contact with H83 of the neighboring monomer, thereby destroying the favorable interaction at the LT interface and hindering the cage formation (Fig. 1l). Moreover, Q15 in loop 1 of Vc-AcP-C20R clashes with E82 of loop 6 of another monomer, but loop 6 of Vc-AcP takes an alternative route to escape this clash (Fig. 1l). These interactions suggest why cage formation is incompatible with the catalytically active conformation.

It remains to be proven whether this cage represents a true biological form of the protein, and if so, what cellular function or evolutionary advantage it might provide. The very low enzymatic activity adds an interesting piece to the puzzle. Beyond its putative biological function, the Vc-AcP particle presents a novel architecture whose hollow interior might be exploited as a carrier of drugs or other small molecules. Its structure could also be modified to produce designer cages with altered properties.

#### Protein Data Bank accession codes

Coordinates and structure factor files have been deposited with the accession codes 4HI1 and 4HI2.

#### Acknowledgements

The laboratory of U.S. is partly supported by the SPGHGD project, Department of Atomic Energy, Government of India, Saha Institute of Nuclear Physics.

R.B. is a Senior Research Fellow of the Council of Scientific and Industrial Research, India. S.N. is a Senior Research Fellow of the Department of Atomic Energy, India. U.S. thanks Pulak Ray of Saha Institute of Nuclear Physics for help in electron microscopy.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2013.09.014>.

Received 21 August 2013;

Received in revised form 3 September 2013;

Accepted 7 September 2013

Available online 17 September 2013

#### Keywords:

bacterial protein phosphatases;  
dynamic light scattering;  
protein crystallization;  
protein self-assembly;  
X-ray crystallography

#### Abbreviation

LT, loop mediated tetramer.

#### References

- [1] Flenniken ML, Uchida M, Liepold LO, Kang S, Young MJ, Douglas T. A library of protein cage architectures as nanomaterials. *Curr Top Microbiol Immunol* 2009;327:71–93.
- [2] Lai YT, Cascio D, Yeates TO. Structure of a 16-nm cage designed by using protein oligomers. *Science* 2012;336:1129.
- [3] King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 2012;336:1171–4.
- [4] Uchida M, Flenniken ML, Allen M, Willits DA, Crowley BE, Brumfield S, et al. Targeting of cancer cells with ferrimagnetic ferritin cage nanoparticles. *J Am Chem Soc* 2006;128:16626–33.
- [5] Plakoutsi G, Taddei N, Stefani M, Chiti F. Aggregation of acylphosphatase from *Sulfolobus solfataricus*. *J Biol Chem* 2004;279:14111–9.
- [6] Motamedi-Shad N, Garfagnini T, Penco A, Relini A, Fogolari F, Corazza A, et al. Rapid oligomer formation of human muscle acylphosphatase induced by heparan sulfate. *Nat Struct Mol Biol* 2012;19:547–54.
- [7] Krissinel E, Hendrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007;372:774–97 [PISA web server:[http://www.ebi.ac.uk/msd-srv/prot\\_int/cgi-bin/piserver](http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver)].

## Cloning, purification, crystallization and preliminary X-ray analysis of two low-molecular-weight protein tyrosine phosphatases from *Vibrio cholerae*

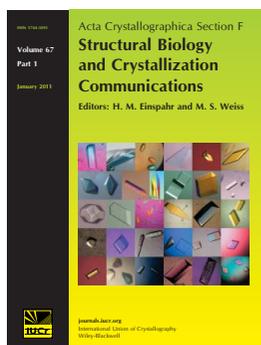
Seema Nath, Ramanuj Banerjee, Susmita Khamrui and Udayaditya Sen

*Acta Cryst.* (2012). **F68**, 1204–1208

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see <http://journals.iucr.org/services/authorrights.html>



*Acta Crystallographica Section F: Structural Biology and Crystallization Communications* is a rapid all-electronic journal, which provides a home for short communications on the crystallization and structure of biological macromolecules. Structures determined through structural genomics initiatives or from iterative studies such as those used in the pharmaceutical industry are particularly welcomed. Articles are available online when ready, making publication as fast as possible, and include unlimited free colour illustrations, movies and other enhancements. The editorial process is completely electronic with respect to deposition, submission, refereeing and publication.

Crystallography Journals **Online** is available from [journals.iucr.org](http://journals.iucr.org)

**Seema Nath, Ramanuj Banerjee,  
Susmita Khamrui and  
Udayaditya Sen\***

Crystallography and Molecular Biology  
Division, Saha Institute of Nuclear Physics,  
1/AF Bidhan Nagar, Kolkata 700 064, India

Correspondence e-mail:  
udayaditya.sen@saha.ac.in

Received 1 June 2012  
Accepted 8 August 2012

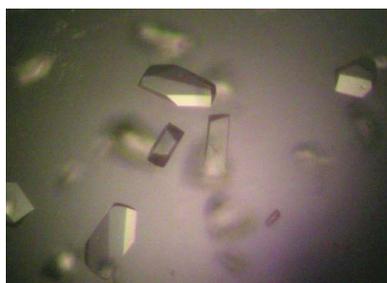
## Cloning, purification, crystallization and preliminary X-ray analysis of two low-molecular-weight protein tyrosine phosphatases from *Vibrio cholerae*

Low-molecular-weight protein tyrosine phosphatases (LMWPTPs) are small cytoplasmic enzymes of molecular weight  $\sim 18$  kDa that belong to the large family of protein tyrosine phosphatases (PTPs). Despite their wide distribution in both prokaryotes and eukaryotes, their exact biological role in bacterial systems is not yet clear. Two low-molecular-weight protein tyrosine phosphatases (*Vc*LMWPTP-1 and *Vc*LMWPTP-2) from the Gram-negative bacterium *Vibrio cholerae* have been cloned, overexpressed, purified by  $\text{Ni}^{2+}$ -NTA affinity chromatography followed by gel filtration and used for crystallization. Crystals of *Vc*LMWPTP-1 were grown in the presence of ammonium sulfate and glycerol and diffracted to a resolution of 1.6 Å. *Vc*LMWPTP-2 crystals were grown in PEG 4000 and diffracted to a resolution of 2.7 Å. Analysis of the diffraction data showed that the *Vc*LMWPTP-1 crystals had symmetry consistent with space group  $P3_1$  and that the *Vc*LMWPTP-2 crystals had the symmetry of space group  $C2$ . Assuming the presence of four molecules in the asymmetric unit, the Matthews coefficient for the *Vc*LMWPTP-1 crystals was estimated to be  $1.97 \text{ \AA}^3 \text{ Da}^{-1}$ , corresponding to a solvent content of 37.4%. The corresponding values for the *Vc*LMWPTP-2 crystals, assuming the presence of two molecules in the asymmetric unit, were  $2.77 \text{ \AA}^3 \text{ Da}^{-1}$  and 55.62%, respectively.

### 1. Introduction

Protein phosphorylation and dephosphorylation are involved in the regulation of many cellular processes such as cell growth, differentiation and metabolism (DeVinney *et al.*, 2000; Hunter, 1995; Mustelin *et al.*, 2005). The formation and hydrolysis of phosphate esters in proteins play key roles in signal transduction, through which external environmental stimuli are converted into internal cellular responses (Mustelin *et al.*, 2005; Neel & Tonks, 1997). Protein kinases and protein phosphatases are among many enzymes that catalyze such reversible reactions in a highly precise manner to control cellular activities. Defective or incorrect regulation of such systems result in loss of cell viability; thus, these proteins are ideal targets for drug design (Zhang, 2001; Hunter, 2000).

Protein tyrosine phosphatases (PTPs) belong to the protein phosphatase superfamily and catalyze the hydrolysis of phosphate esters on tyrosine residues in proteins. Based on their activity towards different phosphorylated amino acids, PTPs can be divided into two families: one class is exclusively active towards phosphorylated Tyr residues and the other acts on both phosphorylated Ser/Thr and Tyr residues in proteins. Low-molecular-weight PTPs fall into the first category (Ramponi & Stefani, 1997). The members of this family of proteins share very low sequence identities amongst themselves apart from the signature motif in the relatively flexible loop at the active site,  $\text{CX}_4\text{CR}$ , where  $X$  can be any amino acid. This loop is responsible for binding and hydrolyzing phosphorylated tyrosine residues and is thus known as the protein tyrosine phosphate-binding loop or P-loop (Zhang *et al.*, 1995).



© 2012 International Union of Crystallography  
All rights reserved

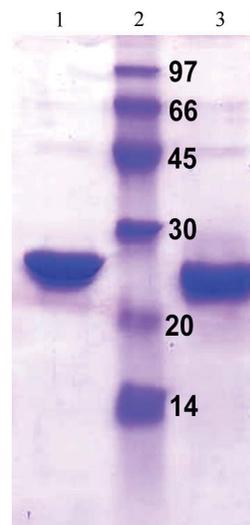
Low-molecular-weight protein tyrosine phosphatases (LMWPTPs) are widely distributed in prokaryotes and eukaryotes (Kennelly & Potts, 1999; Cozzone *et al.*, 2004) and play important roles in many biological processes. The reaction mechanism of eukaryotic LMWPTPs has been structurally, thermodynamically and kinetically characterized (Ramponi & Stefani, 1997). However, LMWPTPs from bacterial sources have been less explored in terms of their structure and function. *Vibrio cholerae* 0395 contains two LMWPTPs (accession Nos. A5F2Q3 and A5F3O7, hereafter termed *Vc*LMWPTP-1 and *Vc*LMWPTP-2, respectively). They have weak sequence identity (~30%), although their active-site signature motif CXGNXCR(S)P and the DPY loop, which play key roles in the hydrolysis of phosphorylated tyrosine, are conserved. However, the amino-acid residues that are located around the P-loop and the DPY loop differ significantly. It is believed that the residues around the P-loop are responsible for modulating substrate recognition, while the residues around the DPY loop are important in catalyzing the dephosphorylation reaction. Therefore, it seems that the physiological targets of these two LMWPTPs are quite different and that these two enzymes should have different dephosphorylation mechanisms. The primary sequences of these two LMWPTPs, especially *Vc*LMWPTP-2, do not produce significant matches with other LMWPTP structures reported in the PDB, suggesting that the three-dimensional structures of these two PTPs may show distinct features that have not been observed in other LMWPTP structures reported to date. Furthermore, no literature is available on the substrate recognition, catalytic mechanism and kinetic parameters of these two PTPs. Importantly, the physiological substrates of these two LMWPTPs are not known and identification of these will definitely shed light on the functional roles of these PTPs in *V. cholerae*. The three-dimensional structures of these two proteins will also be useful to obtain insights into their catalytic functions at the atomic level. Moreover, knowledge of the structure and function of these phosphatases might be of use in drug design against this bacterial pathogen. Here, we report the cloning, overexpression, purification, crystallization and preliminary structural analysis of *Vc*LMWPTP-1 and *Vc*LMWPTP-2 at resolutions of 1.6 and 2.67 Å, respectively.

## 2. Materials and methods

### 2.1. Cloning and expression

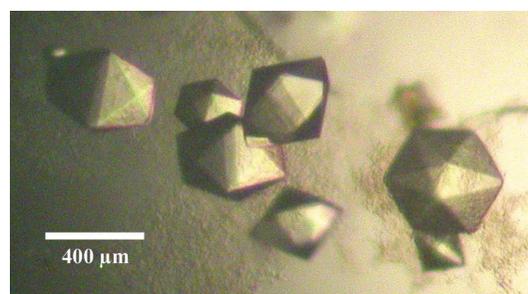
The genes encoding *Vc*LMWPTP-1 (155 amino acids) and *Vc*LMWPTP-2 (166 amino acids) (accession Nos. A5F2Q3 and A5F307, respectively) were amplified from *V. cholerae* 0395 genomic DNA using the polymerase chain reaction with the following primers: for *Vc*LMWPTP-1, forward primer 5'-*CCAGC**ATATG**CAGAAG-GTACTCGTGGTGTGC-3'* and reverse primer 5'-*CGGG**ATCC**-TTAATGCTGGCCTTGCTGTTTTAG-3'*; for *Vc*LMWPTP-2, forward primer 5'-*GCCG**CCATATG**AAGGTTAAAGGTTTATC-AG-3'* and reverse primer 5'-*CGGG**ATCC**TTATTGAGATAAA-TTTCGTTGCACGC-3'*. The primers were synthesized (NeuPro-Cell) with adaptor sites (shown in italics) and restriction-enzyme (*Nde*I and *Bam*HI) sites (shown in bold). Chromosomal DNA of *V. cholerae* strain O395 was used as a template to amplify the regions encoding *Vc*LMWPTP-1 and *Vc*LMWPTP-2. The purified PCR products were cloned into the *Bam*HI and *Nde*I sites of the expression vector pET-28a(+), which adds six consecutive histidines to the N-terminus of the desired protein followed by a thrombin cleavage site. Sequence-verified recombinant DNA was transformed into *Escherichia coli* strain BL21 (DE3) and subsequently selected on kanamycin plates for protein expression. The cells were grown in

Luria–Bertani medium containing 30 µg ml<sup>-1</sup> kanamycin at 310 K for 1 h at a shaker speed of 175 rev min<sup>-1</sup> followed by a further 1 h at 289 K and 175 rev min<sup>-1</sup>. The expression of recombinant protein was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and continued for 18 h at 289 K at a shaker speed of 100 rev min<sup>-1</sup>.

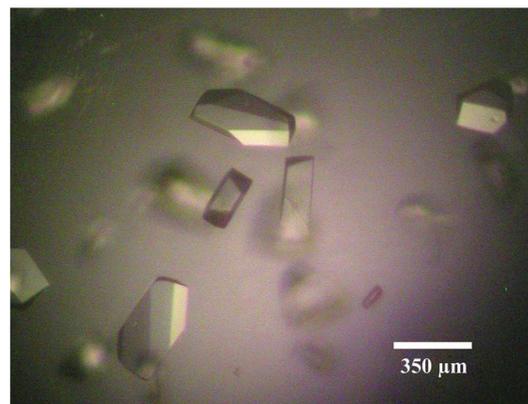


**Figure 1**

The homogeneity of the purified *Vc*LMWPTP-1 and *Vc*LMWPTP-2 proteins was checked by 12% SDS–PAGE. Lane 1, *Vc*LMWPTP-1; lane 2, molecular-mass markers (labelled in kDa); lane 3, *Vc*LMWPTP-2.



(a)



(b)

**Figure 2**

(a) Crystals of *Vc*LMWPTP-1 grown in the presence of ammonium sulfate pH 5 at 293 K. The maximum dimensions of the crystals were 0.4 × 0.4 × 0.3 mm. (b) Crystals of *Vc*LMWPTP-2 grown at 293 K (0.4 × 0.3 × 0.2 mm) appeared when 5% PEG 6000, 8% MPD pH 5.0 was used as a precipitant.

## 2.2. Purification

The cells were harvested by centrifugation at 4000g for 20 min at 277 K. The cell pellet was resuspended in lysis buffer (buffer A; 50 mM HEPES pH 7.0, 300 mM NaCl, 1 mM phenylmethylsulfonyl fluoride and lysozyme) and the cells were disrupted by sonication. The crude lysate was centrifuged at 12 000g for 40 min at 277 K. The supernatant was loaded onto an Ni<sup>2+</sup>-NTA column previously equilibrated with buffer A and subsequently washed with buffer A containing 5 and 10 mM imidazole. The protein was eluted using a gradient to 150 mM imidazole in buffer A. The 6×His tag was cleaved using restriction-grade thrombin (Novagen) and final purification of the protein from contaminating proteins, thrombin and cleaved 6×His tag was achieved by gel filtration using an S-100 (GE Healthcare) column pre-equilibrated with buffer B (50 mM HEPES pH 7.0, 300 mM NaCl, 0.5 mM DTT). The proteins thus purified were used for crystallization. The homogeneity of the purified protein was determined by SDS-PAGE using 15% (v/v) polyacrylamide gel (Fig. 1). The concentrations of both proteins were determined using the Bradford assay.

## 2.3. Crystallization of VcLMWPTP-1 and VcLMWPTP-2

For crystallization, thrombin-cleaved VcLMWPTP-1 (in a buffer consisting of 50 mM MOPS pH 7.6, 300 mM NaCl) and VcLMWPTP-2 (in a buffer consisting of 50 mM HEPES pH 7.0, 300 mM NaCl, 0.5 mM DTT) were concentrated to 6 mg ml<sup>-1</sup> using an Amicon ultracentrifugation unit (molecular-weight cutoff 10 000). Crystallization was performed by the hanging-drop vapour-diffusion method in 24-well crystallization trays (Hampton Research, Laguna Niguel, California, USA). Grid Screen Ammonium Sulfate, Grid Screen PEG 6000, Crystal Screen and Crystal Screen 2 from Hampton Research (Jancarik & Kim, 1991) were used to explore the initial crystallization conditions. 2 µl protein solution was mixed with 2 µl precipitant solution, inverted over a reservoir containing 600 µl precipitant

**Table 1**

Data-collection and processing parameters for VcLMWPTP-1 and VcLMWPTP-2 crystals.

Values in parentheses are for the outermost resolution shell.

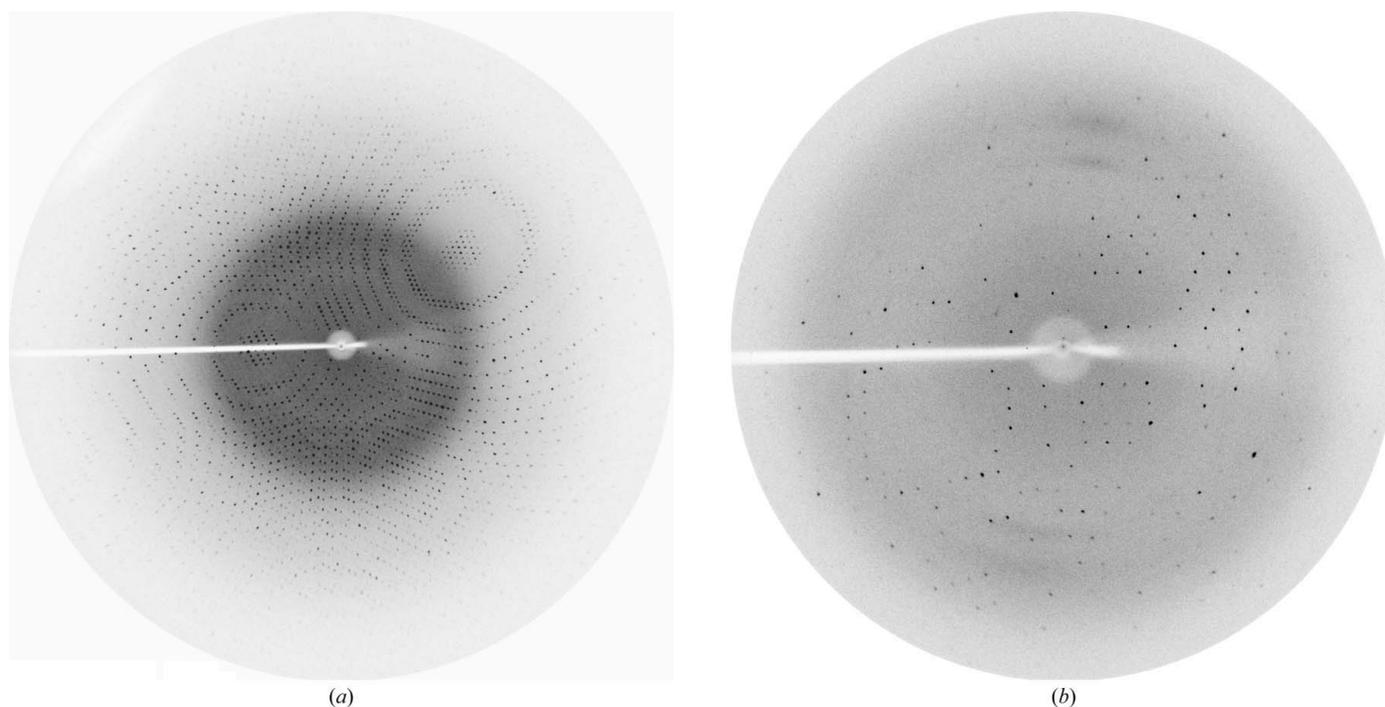
	VcLMWPTP-1	VcLMWPTP-2
Space group	<i>P</i> 3 <sub>1</sub>	<i>C</i> 2
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 87.47, <i>c</i> = 73.85, α = β = 90.0, γ = 120.0	<i>a</i> = 121.38, <i>b</i> = 45.25, <i>c</i> = 88.56, α = γ = 90, β = 121.08
Resolution (Å)	1.6	2.67
Molecules per asymmetric unit	4	2
Matthews coefficient <i>V</i> <sub>M</sub> (Å <sup>3</sup> Da <sup>-1</sup> )	2.26	2.77
Solvent content (%)	46	56
Total No. of reflections	158204 (22198)	36313 (4216)
No. of unique reflections	82757 (12178)	10945 (1286)
Mosaicity (°)	0.54	1.33
Completeness (%)	99.2 (99.3)	92.1 (74.9)
<i>R</i> <sub>merge</sub> <sup>†</sup> (%)	0.047 (0.268)	0.043 (0.183)
<i>I</i> /σ( <i>I</i> )	12.1 (3.0)	19.6 (6.1)

<sup>†</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the observed intensity of the  $i$ th measurement of reflection  $hkl$  and  $\langle I(hkl) \rangle$  is the mean intensity of reflection  $hkl$  calculated after scaling.

solution and maintained at both 277 and 293 K. VcLMWPTP-1 crystallized in 2.4 M ammonium sulfate, 0.1 M citric acid pH 5.0, 2% glycerol at 277 K (Fig. 2*a*) and VcLMWPTP-2 crystallized in 0.2 M ammonium sulfate, 30% (w/v) PEG 8K at 293 K (Fig. 2*b*).

## 2.4. Data collection and processing

Crystals of VcLMWPTP-1 and VcLMWPTP-2 were looped out from the crystallization drops using a 20 µm nylon loop and flash-cooled in a stream of nitrogen (Oxford Cryosystems) at 100 K. A diffraction data set was collected on an in-house MAR Research image-plate detector of diameter 345 mm using Cu *K*α radiation generated by a Bruker-Nonius FR591 rotating-anode generator equipped with Osmic MaxFlux confocal optics and operated at 50 kV



**Figure 3**

(*a*) X-ray diffraction image of a VcLMWPTP-1 crystal; the edge of the detector corresponds to a resolution of 1.6 Å. (*b*) X-ray diffraction image of a VcLMWPTP-2 crystal, which diffracted to a resolution of 2.67 Å.

and 65 mA. X-ray diffraction data were collected to a resolution of 1.6 Å from *VcLMWPTP-1* crystals (Fig. 3*a*) and to a resolution of 2.67 Å from *VcLMWPTP-2* crystals (Fig. 3*b*). Data were processed and scaled using *iMOSFLM* (Battye *et al.*, 2011). Data-collection and processing statistics are given in Table 1.

### 3. Results and discussion

Both *VcLMWPTPs* were successfully purified for crystallization and biochemical assays. They share very low amino-acid sequence identity with other bacterial LMWPTPs of known structure. The sequence identity between *VcLMWPTP-1* and *VcLMWPTP-2* is also low (30%) and their sequences are distinctly different around the P-loop and the DPY loop, implying that these two PTPs are not just two redundant versions of the same protein and that they probably target different physiological substrates for catalysis. Therefore, we crystallized both of these proteins and collected diffraction data. The *VcLMWPTP-1* crystals diffracted to 1.6 Å resolution and produced excellent-quality diffraction data (Table 1) with symmetry consistent with space group *P3<sub>1</sub>*. Packing considerations based on the molecular mass of 17 kDa indicated the presence of four molecules in the asymmetric unit, corresponding to a Matthews coefficient  $V_M$  (Matthews, 1968) of 2.26 Å<sup>3</sup> Da<sup>-1</sup> and a solvent content of 46%. The *VcLMWPTP-2* crystals only diffracted to a resolution of 2.67 Å with moderate data quality (Table 1) and with *C2* space-group symmetry. Packing considerations indicated a Matthews coefficient of 2.77 Å<sup>3</sup> Da<sup>-1</sup>, which corresponds to a solvent content of 56% considering two molecules of *VcLMWPTP-2* (molecular mass of ~18 kDa) in the asymmetric unit.

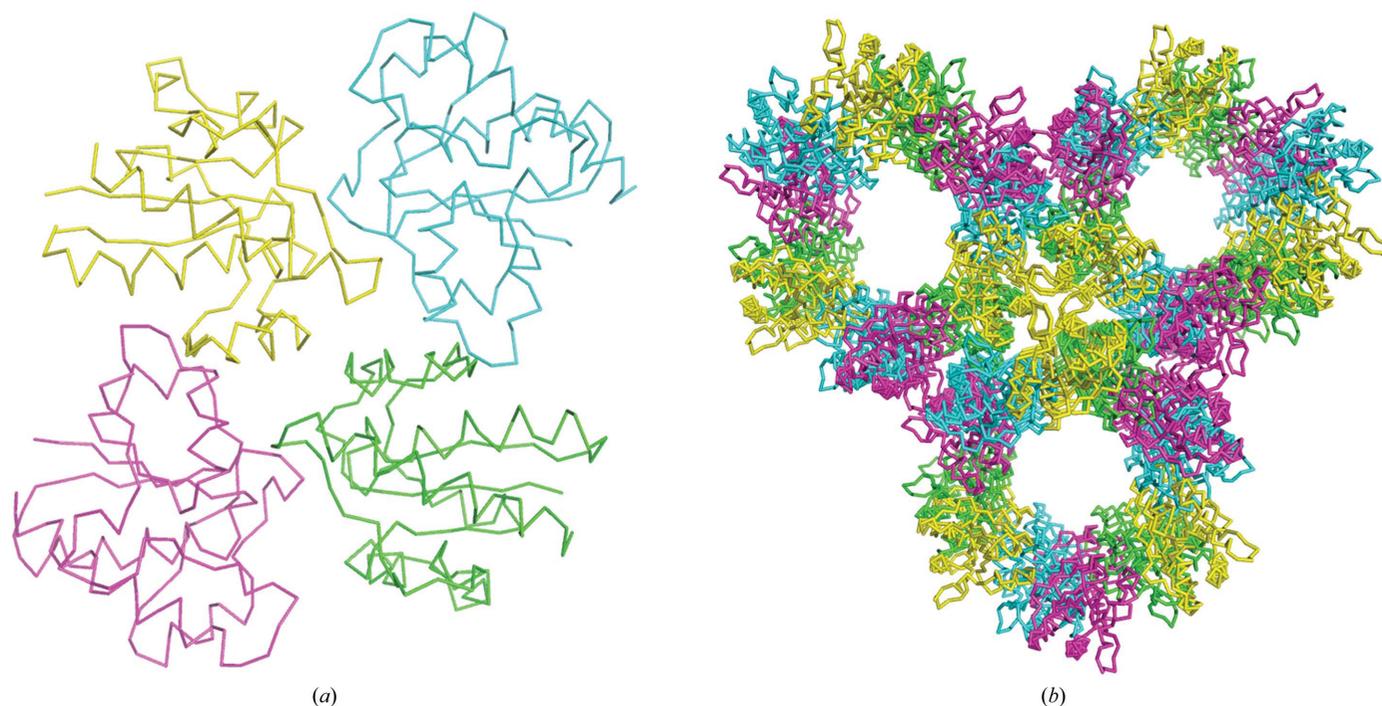
A *BLAST* (Altschul *et al.*, 1990) search for a homologous structure showed that the amino-acid sequence of *VcLMWPTP-1* possesses the highest identity (43%) to that of protein tyrosine phosphatase from *Entamoeba histolytica* (PDB entry 3ido; Seattle Structural Genomics

**Table 2**  
Improvement of *Phaser* statistics.

	RFZ	TFZ	PAK	LLG
Molecule 1	3.1	5.0	0	27
Molecule 2	3.5	8.4	1	85
Molecule 3	3.6	10.9	3	159
Molecule 4	3.8	10.3	8	247

Center for Infectious Disease, unpublished work) followed by human low-molecular-weight phosphotyrosyl phosphatase (40% identity; PDB entry 5pnt; Zhang *et al.*, 1998). *VcLMWPTP-2* has the highest identity (29%) to protein tyrosine phosphatase from *E. histolytica* (PDB entry 3ido). Although the structure of *E. histolytica* protein tyrosine phosphatase (PDB entry 3ido) showed marginally better identity than human LMWPTP (PDB entry 5pnt), human LMWPTP gave slightly better results during molecular-replacement calculations for *VcLMWPTP-1*. Before proceeding with molecular-replacement calculations, waters and a long loop preceding the DPY loop were deleted from the coordinates and mismatched residues were truncated to Ala. Using this truncated model, *Phaser* (McCoy *et al.*, 2007) placed four molecules in the asymmetric unit with RFZ = 3.8, TFZ = 10.3 and LLG = 247 (Table 2). This model was then subjected to several cycles of rigid-body refinement, which removed the clashes between the molecules; the orientations of the four molecules in the asymmetric unit after rigid-body refinement are shown in Fig. 4(*a*). An electron-density map calculated using the molecular-replacement solution thus obtained showed continuous electron density, and refinement of the structure is in progress to obtain the correct structure. The packing of the *VcLMWPTP-1* molecules in the crystal clearly indicates threefold symmetry and large solvent channels (Fig. 4*b*).

Initial molecular-replacement trials to solve the structure of *VcLMWPTP-2* did not produce any clear-cut solution. This might be because of the low sequence identity (29%) of the search models used



**Figure 4**

(*a*) Arrangement of four *VcLMWPTP-1* molecules in the asymmetric unit showing no intermolecular clashes. The four molecules are shown in four different colours. (*b*) Arrangement of the *VcLMWPTP-1* molecules in the crystal showing the threefold symmetry and the large solvent channels.

for molecular replacement. The moderate quality of the diffraction data of VcLMWPTP-2 (Table 1) might also be a reason why a clear solution was not obtained, especially when using models with low sequence identity. At present, we are trying to obtain the phases of VcLMWPTP-2 experimentally.

The laboratory of US is partly supported by the SPGHGD project, DAE, Government of India, SINP. RB is a senior research fellow of Council of Scientific and Industrial Research. US thanks Abhijit Bhattacharya of SINP for help in data collection.

### References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Cozzone, A. J., Grangeasse, C., Doublet, P. & Duclos, B. (2004). *Arch. Microbiol.* **181**, 171–181.
- DeVinney, R., Steele-Mortimer, O. & Finlay, B. B. (2000). *Trends Microbiol.* **8**, 29–33.
- Hunter, T. (1995). *Cell*, **80**, 225–236.
- Hunter, T. (2000). *Cell*, **100**, 113–127.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Kennelly, P. J. & Potts, M. (1999). *Front. Biosci.* **4**, D372–D385.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mustelin, T., Vang, T. & Bottini, N. (2005). *Nature Rev. Immunol.* **5**, 43–57.
- Neel, B. G. & Tonks, N. K. (1997). *Curr. Opin. Cell Biol.* **9**, 193–204.
- Ramponi, G. & Stefani, N. (1997). *Biophys. Biochim. Acta*, **134**, 137–156.
- Zhang, M., Stauffacher, C. V. & Van Etten, R. L. (1995). In *Advances in Protein Phosphatases*, Vol. 9, edited by W. Merlevede. Leuven University Press.
- Zhang, M., Stauffacher, C. V., Lin, D. & Van Etten, R. L. (1998). *J. Biol. Chem.* **273**, 21714–21720.
- Zhang, Z.-Y. (2001). *Curr. Opin. Chem. Biol.* **5**, 416–423.