

Crystal structure of a unique LMWPTP from *Vibrio cholerae* and its functional studies.

By

SHRAMANA CHATTERJEE

Enrollment No: CHEM05201404001

SAHA INSTITUTE OF NUCLEAR PHYSICS, KOLKATA

*A thesis submitted to the
Board of Studies in Chemical Sciences*

*In partial fulfillment of requirements
for the Degree of*

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE




December, 2019

Homi Bhabha National Institute

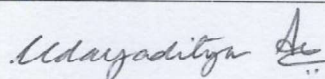
Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by SHRAMANA CHATTERJEE entitled "Crystal structure of a unique LMWPTP from *Vibrio cholerae* and its functional studies" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman - Prof. Abhijit Chakrabarti

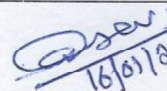
 16/1/2020

Guide / Convener - Prof. Udayaditya Sen

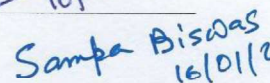
 16/1/20

Co-guide - None

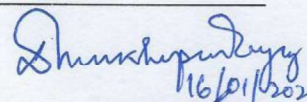
Examiner - Prof. Amit Kumar Das, IIT-Kharagpur

 16/01/2020

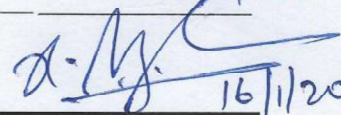
Member 1- Prof. Sampa Biswas

 16/01/2020

Member 2- Prof. Debashis Mukhopadhyay

 16/01/2020

Member 3- Prof. H. Raghuraman

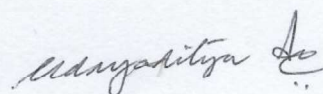
 16/1/2020

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

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

Date: 16/1/20

Place: Kolkata



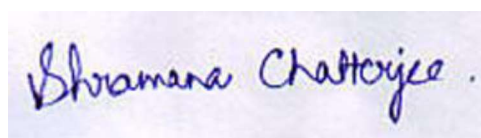
Signature

Guide

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.

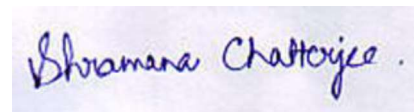
Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when 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 blue ink that reads "Shramana Chatterjee .". The signature is written in a cursive style with a period at the end.

Shramana Chatterjee

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.



Shramana Chatterjee

List of Publications arising from the thesis

Journal

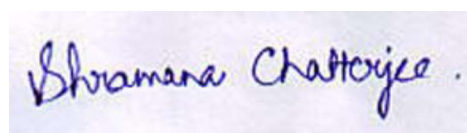
1. “*Vibrio cholerae* LMWPTP-2 display unique surface charge and grooves around the active site: Indicative of distinctive substrate specificity and scope to design specific inhibitor”, S. Chatterjee, S. Nath, B. Ghosh, U. Sen, *BBA Proteins Proteom*, **2019**, 1867:2, 114-124

Conferences

1. **National Symposium on Frontiers of Biology: The DAE Spectra -2015**
21-22 January, 2015
Saha Institute of Nuclear Physics, Kolkata
2. **44th National Seminar on Crystallography (NSC44) -2016**
10-13 July, 2016
Indian Institute of Science Education and Research (IISER), Pune.
Presented poster titled “Structural and functional studies on LMWPTP from *Vibrio cholera* O395”
3. **24th Congress and General Assembly of the International Union of Crystallography (IUCR) -2017**
21-28 August, 2017
Hyderabad International Convention Centre, Hyderabad
Presented Poster titled “Structural and enzymatic aspects of LMWPTP from *Vibrio cholerae* 0395”
4. **CCP4 Crystallography School and Workshop: From data processing to structure refinement and beyond -2018**
22-26 October, 2018
CSIR-IMTech, Chandigarh

Others

1. “Large-scale conformational changes and redistribution of surface negative charge upon sugar binding dictate the fidelity of phosphorylation in *Vibrio cholerae* Fructokinase”, R. Paul, S. Chatterjee, S. Nath, U. Sen, *Scientific Reports*, **2018**, 8 , 16925
2. “High resolution structure of *Vibrio cholerae* acylphosphatase (VcAcP) cage: identification of drugs, location of its binding site and engineering to facilitate cage formation”, S. Chatterjee, S. Nath, U. Sen, *BBRC* (Just accepted).



Shramana Chatterjee

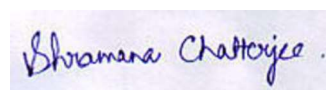
Dedicated to my family.....

ACKNOWLEDGEMENTS

It is a formidable task to construct a proper acknowledgement list within one page, as there are innumerable persons, who have been involved behind these PhD days.

First of all I would like to express my sincere gratitude to my mentor Prof. Udayaditya Sen. I have immensely benefited from his daring and positive attitude, flair of innovative thinking and symbolic optimism throughout my tenure in this institute. I am thankful to all the members of our division for their assistance. I would like to thank my seniors and lab mates Seemadi, Sanghatidi, Kamalenduda, Maltidi and Tulika for introducing me to various techniques of molecular biology and protein chemistry and for encouraging me through the highs and lows. I do thankfully acknowledge my Post MSc batch mates and members of MSA-1 hostel specifically our cooks & sweepers, without whose active support and cooperation this thesis would not have been completed.

I owe special thanks to my husband, Arindam, who continuously encouraged me to be optimistic and also for his continuous and unfailing support and indulgence during my pursuit of Ph.D. My heart felt regard goes to my father in law and mother in law, for their love and moral support. Lastly, I would like to express my respect, love and gratitude to the persons from whom a child gets her first inspiration, my family (baba, ma, kaka and dada), who has always been the chief architect behind my every success, who has always been last shelter during my each turmoil.



Shramana Chatterjee

Contents

	Page No.
SUMMARY.....	1
List of Figures.....	3
List of Tables.....	6
Chapter 1- Introduction.....	7
Part I.....	8
1.I.1. ‘PTP-ome’	9
1.I.2. Classification	10
1.I.3. Bacterial phosphorylation/de-phosphorylation-Historical outlook.....	12
1.I.4. Bacterial tyrosine kinases (BYK).....	13
1.I.5. Bacterial tyrosine phosphatases (BYP).....	14
1.I.6. Low molecular weight protein tyrosine phosphatase (LMWPTP).....	16
1.I.6a. Regulation of LMWPTP.....	17
1.I.6b. Catalytic mechanism of LMWPTP.....	20
1.I.7. LMWPTP: potential target of therapeutics.....	21
1.I.8. LMWPTP : In search of inhibitors.....	23

1.I.9. Scope of the thesis.....	26
Part II.....	27
1.II. Structure-function relationship of <i>Vc</i> LMWPTP-2: Study by protein X-ray crystallography.....	28
1.II.1 X-ray crystallographic study: steps involved.....	29
1.II.2 Preparation of protein sample.....	29
1.II.3. Technique of crystallization.....	29
1.II.4. Solubility Curve.....	30
1.II.5. Phase diagram.....	30
1.II.6. Vapor diffusion.....	32
1.II.7. Screening.....	33
1.II.8. Strategy of data collection.....	34
1.II.9. Data processing step.....	34
1.II.10. Phase problem.....	35
1.II.11. Patterson methods.....	37
1.II.12. Direct methods.....	37
1.II.13. Multiple Isomorphous Replacements.....	38
1.II.14 Multi wavelength Anomalous Dispersion.....	39
1.II.15 Molecular Replacement.....	39
1.II.16. Phase improvement and refinement.....	43
1.II.17. Deposition of structure after validation.....	44
Chapter 2- Materials & Methods.....	45

2.1 Cloning, Over-Expression, Purification, Crystallization, Data-collection, Data processing, Refinement.....	46
2.1.1. Cloning and over-expression of <i>VcLMWPTP-2</i> and its mutants.....	47
2.1.1.1a. <u>P</u>olymerase <u>C</u>hain <u>R</u>eaction (PCR).....	49
2.1.1.1b. <u>S</u>ite <u>d</u>irected <u>m</u>utagenesis (SDM).....	50
2.1.1.2. Digestion and Ligation.....	51
2.1.1.3. Transformation.....	52
2.1.1.4. Screening and Analysis.....	52
2.1.1.5. Protein over-expression.....	53
2.1.2. Protein purification by affinity and size exclusion chromatography.....	54
2.1.2.1. Affinity chromatography: purification of 6×His Tagged protein.....	54
2.1.2.2. 6×His tag removal by thrombin.....	56
2.1.2.3. Purification by Size exclusion chromatography.....	57
2.1.3. Crystallization of protein of interest.....	57
2.1.4. Collection of Data and further processing.....	58
2.1.5. Structure determination of protein of interest.....	58
2.1.6. Structural analysis.....	60
2.2. Phylogenetic Analysis.....	60

2.3. FPLC (Fast Protein Liquid Chromatography) analysis of <i>Vc</i>LMWPTP-2	61
2.4. Dynamic Light Scattering	61
2.5. Molecular Dynamics Simulation	62
2.6. Enzyme Kinetics	63
2.6.1 Phosphatase activity	63
2.6.2 Reactivation kinetics of hydrogen peroxide-inactivated <i>Vc</i>LMWPTP-2 using β-ME as reducing agent	64
Chapter 3- Results	65
Part-I	66
3.I.1. Phylogenetic analysis	67
3.I.2. Cloning, Overexpression, Purification, Crystallization, Data Collection and Structure solution of <i>Vc</i>LMWPTP2	69
3.I.2.1. Cloning and Over-expression of <i>Vc</i>LMWPTP-2	69
3.I.2.2. Purification of target protein by Ni-NTA affinity chromatography	70
3.I.2.3. Thrombin cleavage and size exclusion chromatography	71
3.I.2.4. Crystallization	72
3.I.2.5. Data Collection	73
3.I.2.6. Structure solution of <i>Vc</i>LMWPTP-2	74
3.I.3. Structure based sequence alignment	76

3.I.4. Structural Analysis of <i>Vc</i>LMWPTP-2.....	78
3.I.4.1 Overall structure of <i>Vc</i>LMWPTP-2 monomer.....	78
3.I.4.2 Structural superposition.....	79
3.I.4.3 MOPS binds like a substrate mimetic at the active site.....	80
3.I.4.4 Architecture of the active site.....	81
3.I.4.5 Comparison with ligand bound/unbound structures.....	82
3.I.5 Dimerization of <i>Vc</i>LMWPTP-2.....	83
3.I.6 Multiple sequence alignments.....	86
3.I.7 Kinetic studies of <i>Vc</i>LMWPTP-2.....	90
3.I.7.1. Phosphatase activity of <i>Vc</i>LMWPTP-2.....	90
3.I.7.2. Redox regulation of <i>Vc</i>LMWPTP-2.....	92
3.I.8. Molecular dynamics simulation.....	93
3.I.8.1. Mechanical stability and dynamic nature of loops.....	93
3.I.8.2. Dynamic nature of residues around the active site.....	94
3.I.9 Unique surface charge and groves around the P-loop.....	96
Part-II.....	98
3.II. Role of Cys17 in <i>Vc</i>LMWPTP-2: Is it protective or catalytic?	99
3.II.1 Cloning and Over-expression of <i>Vc</i>LMWPTP-2-C12S.....	99

3.II.2 Purification of target protein by Ni-NTA affinity chromatography....	100
3.II.3 Thrombin cleavage and size exclusion chromatography.....	101
3.II.4 Crystallization.....	103
Chapter 4- Discussion.....	104
References.....	109

Summary

Reversible phosphorylation of tyrosine residues is the basis for controlling many diverse cellular processes. The extent of tyrosine phosphorylation is dynamically controlled by two groups of enzymes, phospho tyrosine kinases (PTKs) and phospho tyrosine phosphatases (PTPs). Although protein tyrosine phosphorylation and de-phosphorylation was considered as mere post translational modification recently it has been emerged as an important factor for controlling the fate of cell. PTPs are family of enzymes that can be divided into four subgroups: classical pTyr specific, dual specificity phosphatases, Cdc25 and low molecular weight protein tyrosine phosphatases (LMWPTP). LMWPTPs, an enigmatic member of PTP family, are small cytosolic enzymes (~18kDa) ubiquitously found across a spectrum of genera from prokaryotes to higher eukaryotes, whereas other PTPs (PTP1B, CDC45, LAR, Shp1, Shp2, PTP α , PTP β , PTP γ) are expressed exclusively in eukaryotes. The gram-negative bacteria, *Vibrio cholerae*, is the causative organism of the severe diarrheal disease cholera that continues to be a significant cause of death in developing countries claiming approximately 95,000 lives annually. The genome of *Vibrio cholerae* O395 contains two LMWPTPs (*Vc*LMWTP-1 and *Vc*LMWTP-2). Upon vigorous sequence alignment and analysis we have found that *Vc*LMWPTP-2 is unique among bacterial LMWPTPs. Occurrence of two LMWPTP in the same organism and their biological role, especially the role of *Vc*LMWPTP-2, remains to be established. Structural as well as functional information of this unique LMWPTP from *Vibrio cholerae* is therefore useful for designing of inhibitor which will lead to drug design against this pathogenic bacterium. Starting with this knowledge, the work of the present thesis represents detailed structural as well as functional aspects of *Vc*LMWTP-2. The X-ray

structural work alongwith in detail comparative studies with other LMWPTPs has been presented here. With various orthogonal techniques it has been demonstrated that LMWPTP differed significantly from other homologous proteins (structures bearing closest identity 30-37%, similarity ~55%). Oligomerization, considered as one of the regulatory mechanisms in this family, proved to be quite different in this LMWPTP that led to an active dimer. The result is unique and is in sharp contrast with the inactive dimers observed in other cases. Moreover, mutagenesis experiments coupled with kinetic data, demonstrated that the role of P-loop Cys17 is involved in the protection of catalytic cysteine Cys12 along with its possible role in substrate turnover. The roles of several conserved motifs within this family that were key in maintaining the architecture of the active site have been highlighted. Structural fluctuations observed in MD simulations, in combination with crystal structure, identified several functional loops and hydrogen bonds that were key in defining the architecture of the substrate binding cavity either in apo or substrate bound forms. Collectively, the work presented here, describes in depth the structural description defining the uniqueness of this LMWPTP while the MD simulation and functional studies led to the modulation of substrate binding cavity and P-loop mediated regulatory role. As distribution of surface charge and grooves around the 'P-loop' is believed to govern the specificity of LMWPTPs, we performed detailed comparison between two LMWPTPs from the same organism (*Vc*LMWPTP-1 and *Vc*LMWPTP-2) and also with other LMWPTPs. This comparative study from various LMWPTPs along with substrate analogue bound LMWPTPs, delineates unique features of *Vc*LMWPTP-2 and also provides an idea about the plausible nature of its true substrate would likely be which could be utilized to design specific inhibitors.

List of Figures

Figure 1: PTP classification of Human

Figure 2: Event of bacterial tyrosine phosphorylation has been revealed after many years than eukaryotic system

Figure 3: Oxidation schemes for the PTPs by Hydrogen peroxide (H_2O_2)

Figure 4: Schematic diagram for regulation through dimerization

Figure 5: Schematic diagram of LMWPTP catalytic mechanism

Figure 6: Physiological or pathological signaling pathways by LMWPTP

Figure 7: Inhibitor of human LMWPTP

Figure 8: Inhibitors of bovine LMWPTP

Figure 9: The Phase diagram in protein crystallization experiment

Figure 10: Vapour diffusion method

Figure 11: Multiple Sequence alignment between *Vc*LMWPTP-1 and *Vc*LMWPTP-2

Figure 12: Comparative analysis between PTP family proteins and *Vc*LMWPTP-2

Figure 13a-13b: Clone check and overexpression of *Vc*LMWPTP-2

Figure 14: Ni-NTA purification profile represented by 15% SDS-PAGE

Figure 15a-15b: Elution profile by SEC and concentration checked for crystallization by SDS-PAGE.

Figure 16: Diffraction pattern of *Vc*LMWPTP-2 crystals upon X-ray exposure

Figure 17: Structure based Sequence alignment of *Vc*LMWPTP-2 with *Vc*LMWPTP-1 and other organisms

Figure 18: Structure of *Vc*LMWPTP-2 monomer

Figure 19a-19b: Structural superposition of *Vc*LMWPTP-2

Figure 20: $2F_{\text{obs}} - F_{\text{calc}}$ electron density map around the active site of *Vc*LMWPTP-2

Figure 21: 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 R18.

Figure 22: Surface representation of MOPS (ball and stick) binding pocket; residues lining the active site pocket are shown in surface and sticks

Figure 23a-23b: Comparison with 'open'/'closed' structure

Figure 24a-24f: Dimerization of *Vc*LMWPTP-2

Fig. 25a-25b: Sequence and structural alignment of LMWPTP

Figure 26a-26b: Enzyme kinetics of *Vc*LMWPTP-2 using p-NPP as a substrate

Figure 27a-27b: Redox regulation of *Vc*LMWPTP-2 and its mutants

Figure 28a-28b: Structural flexibility of *Vc*LMWPTP-2

Figure 29: Superposition of the snapshots of MD simulation trajectories

Figure 30: Phylogenic tree and comparison of the electrostatic surface charge distribution

Figure 31a-31b: Clone check and overexpression of *Vc*LMWPTP-2 C12S

Figure 32: Ni-NTA purification profile of *Vc*LMWPTP-2 C12S represented by 15% SDS-PAGE

List of Tables

Table1. Biochemically characterized tyrosine kinases and tyrosine phosphatases and their cellular function from various organisms

Table 2: Sequences of Primer

Table 3: Calculation of the PCR reaction mixture

Table 4: Calculation for the two-step-PCR reaction mixture

Table 5: Calculation for restriction enzyme digestion

Table 6: Components of Ligation

Table 7: Data collection statistics and refinement statistics

Table 8: Results found from Dali Server using *Vc*LMWPTP-2 (*Vc2*) sequence as input

Table 9: Area and Volume of active site of LMWPTPs

Table 10: Kinetic parameters of wild-type *Vc*LMWPTP-2 and its mutants

Chapter-1

Introduction

PART-I

The human genome sequencing is followed by this ‘omics’ research era where signaling molecules and their families are being defined. Regulation of signal transduction and its complexities are the state of the art. Cascades of activation and inactivation are involved in signal transduction which is governed by various cellular processes such as cell proliferation, differentiation etc. Highly conserved example is mitogen-activated protein-kinase cascades which are consists of three kinases formed linear cascades where each kinase phosphorylate the next one in-line [1]. Kinases are counteracted by phosphatases and they results in a concerted action to decide the active/inactive state of a downstream enzyme [2]. Signal transduction process by which cells receive signals from outside and they are transmitted into the nucleus of cell, is well scrutinized [3]. The importance of protein phosphorylation and de-phosphorylation by kinases and phosphatases respectively, has been established but phosphatases as merely housekeeping enzyme: this point of view is changing. Initially signaling pathways were viewed as simple linear arrangements of phosphorylation cascades which work in isolation but now it has been understood as multiple interacting signaling network which works consequently to result in a physiological outcome. Phosphatases and their cognate kinases are distinct but complementary enzymes where amplitude of signals is controlled by kinases but phosphates play an important role in the rate and duration of signal responses. [1,4].

1.1.1. PTP-ome

Protein kinases have been evolved from a common ancestor although protein phosphatases have been originated and evolved from structurally and functionally distinct families. Among them serine/threonine phosphatases exists *in vivo* as a range of holo-enzyme complexes which are consisted with multiple subunits and through

this catalytic and regulatory subunits a broad spectrum of signaling pathways are controlled [5]. Largest family of phosphatase genes encodes the protein tyrosine phosphatases (PTPs). PTPs are defined by signature motif CX₅R at the active site where cysteine residue of this motif acts as nucleophile and without this cysteine residue catalysis wouldn't take place. Now it is possible to catalogue the genes that comprise PTP superfamily due to completion of human genome sequencing. Similar complexity between kinases and phosphatases has been compared by the fact that there are ~100 human PTP superfamily genes compared to ~90 PTK genes. The number genes encoding PTP/PTKs only indicates marginal complexity; additional complexities are added with alternative mRNA splicing, promoters and post-translational modifications. Thereafter functional importance of negative as well as positive regulation of signal transduction by PTPs is relatable from this structural diversity. Moreover, *in vivo* PTPs display attractive substrate and functional specificity. Therefore, establishment of a detailed analysis on the structure, regulation, physiological function of this important enzyme can provide a definition of 'PTP-ome' [6].

1.1.2. Classification

Based on protein sequences and function, PTP superfamily can be subdivided into 4 groups with 107 members [Fig.1] [6,7]. Class I PTP is the major group consisted with 99 members and also subdivided into various subclasses as shown in figure1, whereas class II consists only low molecular weight (LMW) PTP. Class III has three members, viz., Cdc25A, Cdc25B, Cdc25C. For the three abovementioned classes Cys residue acts as catalytic residue although class IV PTPs are consisted with 4 members and for

all of them Asp acts as catalytic residue which directs towards unique evolutionary origin [8].

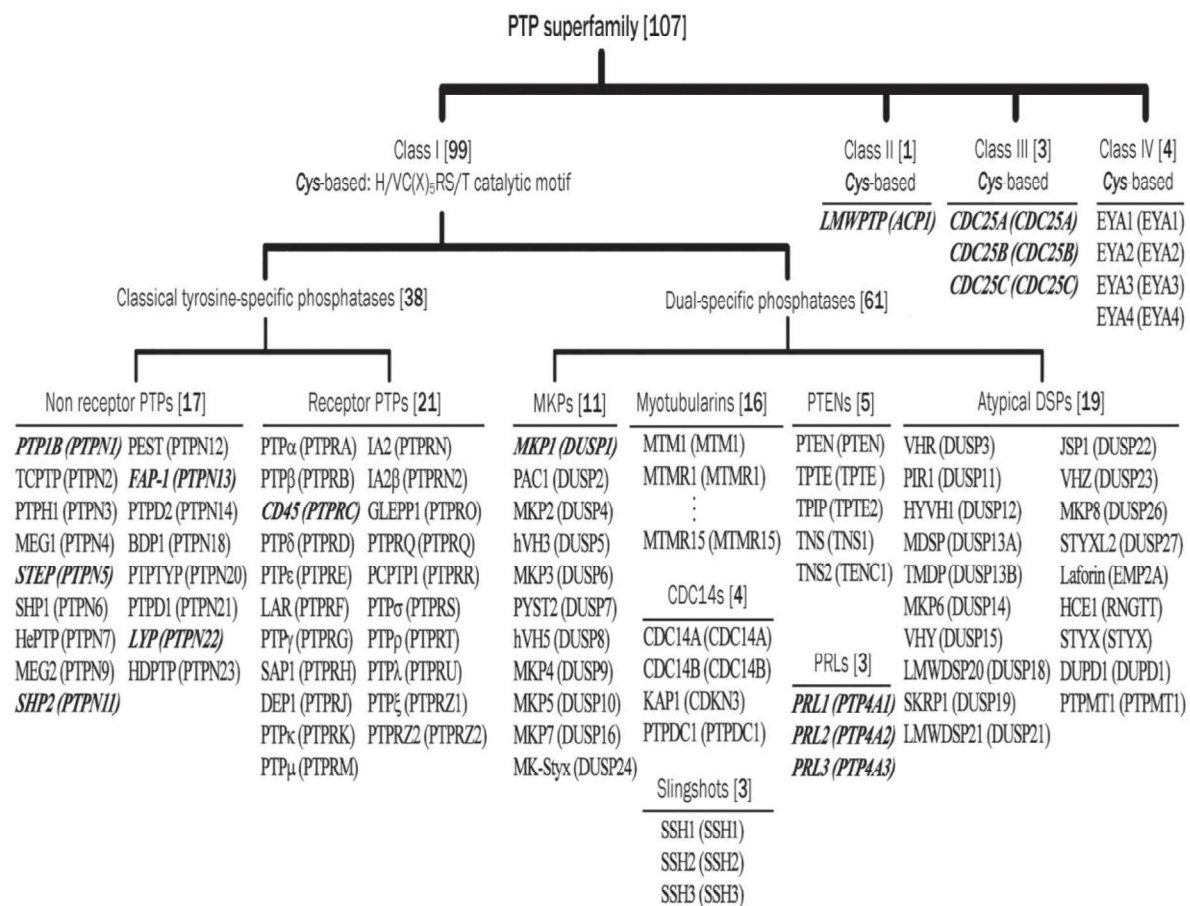


Figure 1. PTP classification of Human. Individual PTP encoding gene is shown in parenthesis (after the PTP name). (Figure courtesy: Rong-jun HE et al. 2014 Acta Pharmacologica Sinica)

1.1.3. Bacterial phosphorylation/de-phosphorylation- Historical outlook

The most widespread and important post-translation modification is known as protein phosphorylation, originally discovered in 1950's and known to regulate numerous cellular functions in eukaryotes. It was established that phosphorylation takes place exclusively on serine, threonine and tyrosine residues where ATP is used as phosphate donor. Initially protein kinases similar to eukaryotes were not found in prokaryotes and therefore it was thought that bacteria are devoid of any serine/threonine/tyrosine phosphorylation. The existence of this documentation was modified in late 1970's [9, 10] although early studies on bacterial system were focused on phosphorylation on histidine/aspartic acid residues.

A clear distinction was made on those days that bacteria uses histidine / aspartic acids and eukaryotic system uses serine/threonine/tyrosine residues for phosphorylation. In early 1990s first serine/threonine kinase which was known as eukaryotic like till date, was characterized on bacterium *Myxococcus xanthus* [11]. Bacterial phosphorylation on tyrosine residues has not been documented quickly and to date a very few kinases in bacterial system was identified to use tyrosine as protein substrate [12, 13]. The presumption that tyrosine kinases are restricted into eukaryotic system only started to erode by 1980's [14] and obviated by 1997 with characterization of bacterial gene which encodes an auto-phosphorylating protein tyrosine kinase [15]. Since then numerous bacterial tyrosine phosphorylation/de-phosphorylation by tyrosine kinases and phosphatases respectively have been characterized [Fig. 2].

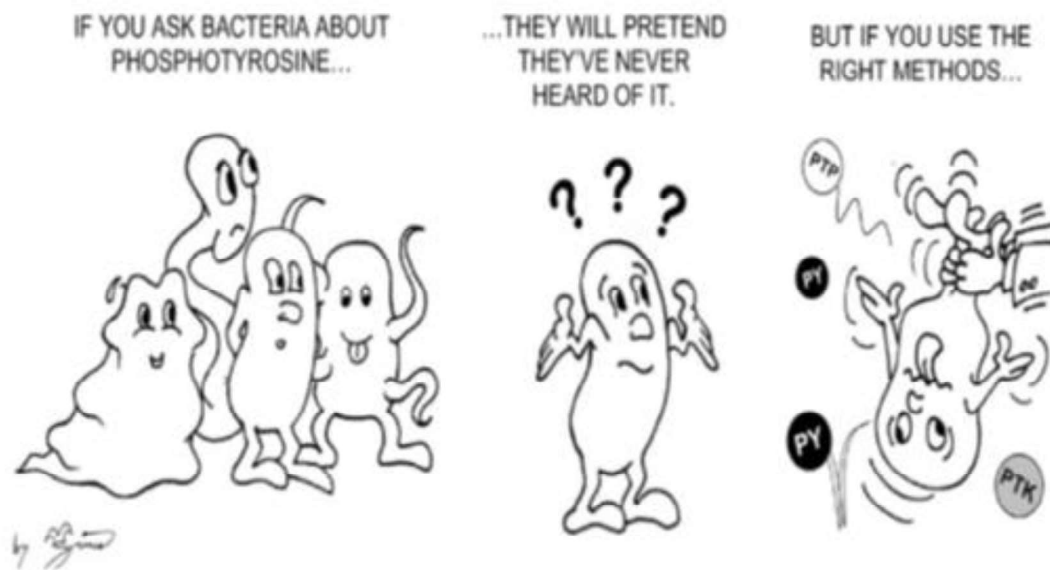


Figure 2. Event of bacterial tyrosine phosphorylation has been revealed after many years than eukaryotic system. (Cartoon by I. Mijakovic. Figure courtesy: Christophe Grangeasse et. al. 2006 TRENDS in Biochemical Sciences)

1.I.4. Bacterial tyrosine kinases (BYK)

A trans-membrane domain and an intracellular catalytic domain which can be linked either by single polypeptide or through a specific interaction of their alpha helices consists a BY-kinase [16-19]. The trans-membrane domain is considered as a kinase anchor into the cell-membrane as well a sensor domain which triggers the activity of BY-kinase. This domain also interacts with other proteins situated at outer membrane and influences the cellular function of BY-kinase [20]. Canonical Walker motifs, Walker A and Walker B define the active site of catalytic domain and this ATP/GTP binding Walker motifs phosphorylates tyrosine residues [21].

Exo-polysaccharide production in several bacterial systems has been linked with the functional role of BYK auto-phosphorylation. Genes in large operons which are generally involved in bio-synthesis and sugar polymer export, generally encode them

[22]. Relation between functional role of BYK and exopolysaccharide synthesis has been well studied in proteobacteria *Escherichia coli*. Wzc and Etk are two BYKs of this system where the first one is encoded by the gene which is involved in biosynthesis of polysaccharide polymers [23] and the second one is encoded from the gene which is involved in synthesis of group 4 capsules (G4C) polysaccharides [24]. Presence of both Wzc and Etk are important for the synthesis of extra cellular polysaccharide. The amount of polysaccharide produced and the length of the polymer is controlled by BYKs and this occurs probably with an interaction with Wzy, a polysaccharide polymerase. [25-27]. Exopolysaccharide content of the cell envelope is controlled by BYKs which affects the pathogen attachment to the host cell and also acknowledgement by lymphocytes. In metabolite synthesis, cell wall material production etc. also kinase-phosphatase pair play some roles in non-pathogenic bacteria.

1.1.5. Bacterial tyrosine phosphatases (BYP)

The extent of tyrosine phosphorylation is dynamically controlled by two groups of enzymes, phospho tyrosine kinases (PTKs) and phospho tyrosine phosphatases (PTPs). PTPs are family of enzymes that can be divided into four subgroups: classical pTyr specific, dual specificity phosphatases, Cdc25 and low molecular weight phospho tyrosine phosphatases (LMWPTP). The common feature of the four subgroups of PTPs is the occurrence of a consensus sequence motif C(X)₅RS/T, where X can be any amino acid, but are otherwise different in their sequence and topology [28]. Surprisingly, this motif adopts a highly similar loop conformation which represents an example of convergent evolution. Functionally, this motif efficiently

binds the phosphoryl group of the substrate phospho-tyrosine to be hydrolyzed and is known as P-loop. The invariant cysteine of P-loop acts as a nucleophile [29] to carry out the dephosphorylation reaction and mutation of this residue completely abrogate the enzymatic activity [30].

In a wide range of prokaryotes PTP have been discovered which plays a regulatory role in cellular function and in most of the cases they are found as a mimetic enzyme of their better-known eukaryotic counterpart [31,32]. Either transmission or inhibition of the phosphor-dependent signaling can be resulted by the catalysis reaction. Stimulation or alienation in the action of cognate protein kinases may also be controlled by protein tyrosine phosphatases [33].

BYPs are involved in various cellular processes among them involvement in exopolysaccharide synthesis and capsular polysaccharide production is the most important and other proteins are involved in to disrupt the host cell machinery. BYPs can be thought to play a central role in flow of information that controls the pathogenic activity because capsular polysaccharides are one of the key determinants in the virulence factors in some organisms. Biochemically characterized tyrosine kinases and tyrosine phosphatases and their cellular function from various organisms are listed in table 1.

Table 1. Biochemically characterized tyrosine kinases and tyrosine phosphatases and their cellular function from various organisms

Organism	BYK	BYP	Function
<i>Escherichia coli</i>	Wzc	LMWPTP* (Wzb)	Production of Capsular polysaccharide and colanic acid
<i>Escherichia coli</i>	Etk	LMWPTP (Etp)	Heat shock response
<i>Acinetobacter sp.</i>	Ptk	LMWPTP (Ptp)	Emulsion production
<i>Erwinia amylovora</i>	AmsA	LMWPTP (AmsI)	Succinoglycan production
<i>Klebsiella pneumoniae</i>	Yco6	LMWPTP (Yor5)	Capsular polysaccharide production
<i>Pseudomonas solanacearum</i>	EpsB	LMWPTP (EpsP)	Exopolysaccharide I production

*LMWPTP represents low molecular weight protein tyrosine phosphatase as mentioned in 1.2 classifications.

1.1.6. Low molecular weight protein tyrosine phosphatase (LMWPTP)

LMWPTPs, an enigmatic member of PTP family, are small cytosolic enzymes (~18 kDa) ubiquitously found across a spectrum of genera from prokaryotes to higher eukaryotes, whereas other PTPs (PTP1B, CDC45, LAR, Shp1, Shp2, PTP α , PTP β , PTP γ) are expressed exclusively in eukaryotes [34]. As mentioned in table1, some prokaryotic LMWPTPs are virulence factors that mimic eukaryotic phosphatases and dephosphorylate eukaryotic proteins, thereby interfering with the host defense

response. They are also involved in the biosynthesis and transport of virulence factors, such as exo- and capsular polysaccharides or in bacterial stress resistance.

From the prokaryotic counterpart, structures of LMWPTP from gram-positive eubacteria *Staphylococcus aureus*– PtpA (PDB: 3ROF) [35] gram-negative eubacteria *Thermus thermophilus*–TT1001 (PDB: 2CWD), gram-positive proteobacteria *Bacillus subtilis*– YwIE (PDB: 4ETI), gram-negative proteobacteria *Escherichia coli*– Wzb (PDB: 2WMY) [36] and pathogenic *Mycobacterium tuberculosis* MPtpA (PDB: 1U2Q) [37] have been reported.

1.I.6a. Regulation of LMWPTP

Among various reported regulatory mechanisms of PTPs, regulation through reversible oxidation of the active-site cysteine drew much attention [38, 39]. PTEN (phosphatase and tensin homolog), Cdc25 and LMWPTPs have a unique feature of hosting two cysteines in the P-loop, and for LMWPTP these are catalytic Cys12 and the vicinal Cys17, the later plays a major regulatory role by protecting the catalytic Cys12 from irreversible oxidation [40] although for classical PTP 1B sulfenic-amide intermediate is formed [Fig. 3]

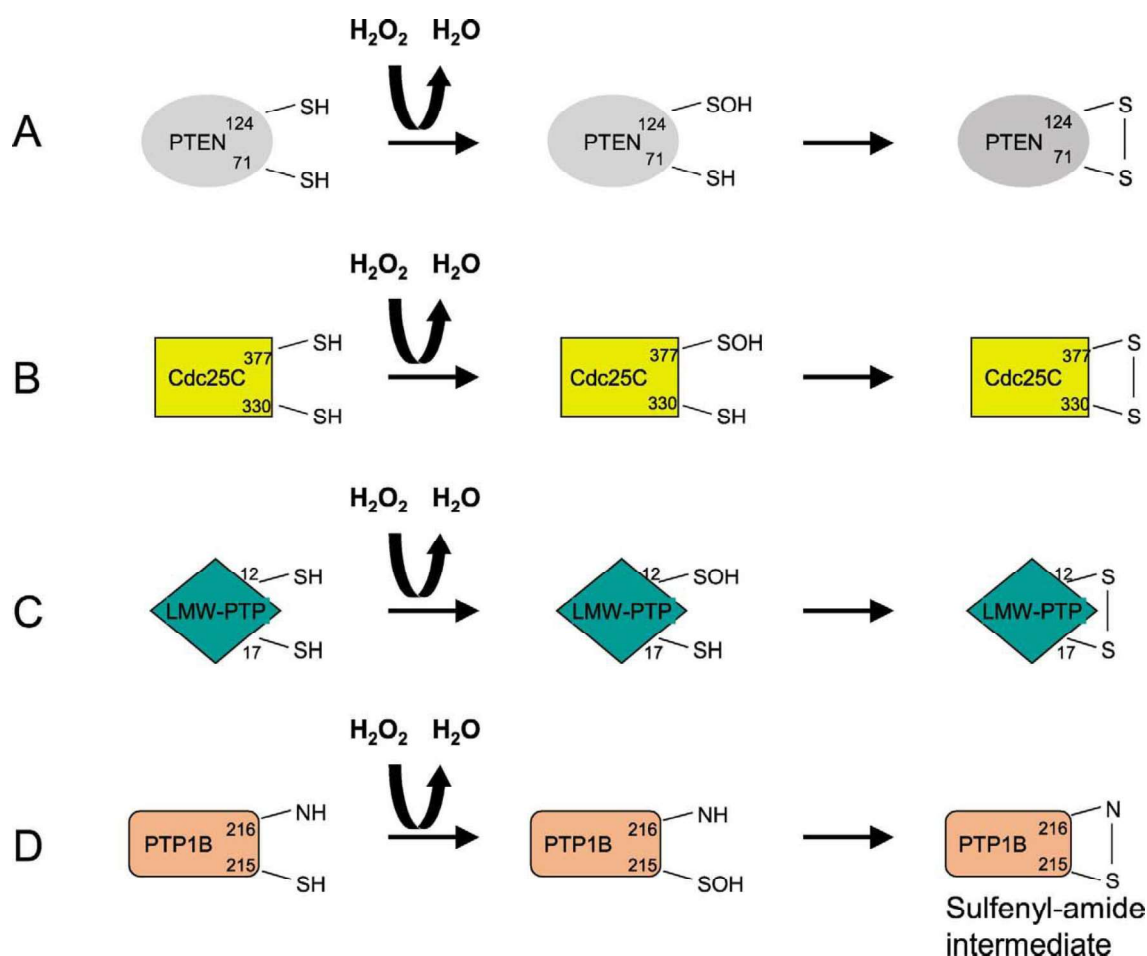


Figure 3. Oxidation schemes for the PTPs by Hydrogen peroxide (H_2O_2). PTEN (A), Cdc25C (B), and LMW-PTP (C) are oxidized to form a disulphide bond with Cys-SOH as an intermediate whereas reactive cysteine of PTP1B is oxidized to Cys-SOH and forms sulfenic-amide intermediate. (Figure Courtesy: S.H.Cho et.al. 2004 FEEBS Letter)

Another regulatory mechanism proceeds for LMWPTPs is through oligomerization. It has been reported that the self association of mammalian LMWPTP (viz. BPTP) produces inactive oligomers that are in equilibrium with its active monomers [41] [Fig. 4]. Inactivation occurs through direct involvements of the active site residues of

one monomer with the tyrosines of the DPY-loop of other. Similar but weak dimerization has also been reported in prokaryotic LMWPTPs like YwlE from *Bacillus subtilis*, PtpB from the Gram-negative bacterium *Salmonella aureus* and Wzb from *Escherichia coli* [42]. Higher order oligomerization of YwlE was also detected by NMR experiments [42]. Overall, the oligomerization processes were found to be weak, with in vitro dissociation constants in the millimolar range.

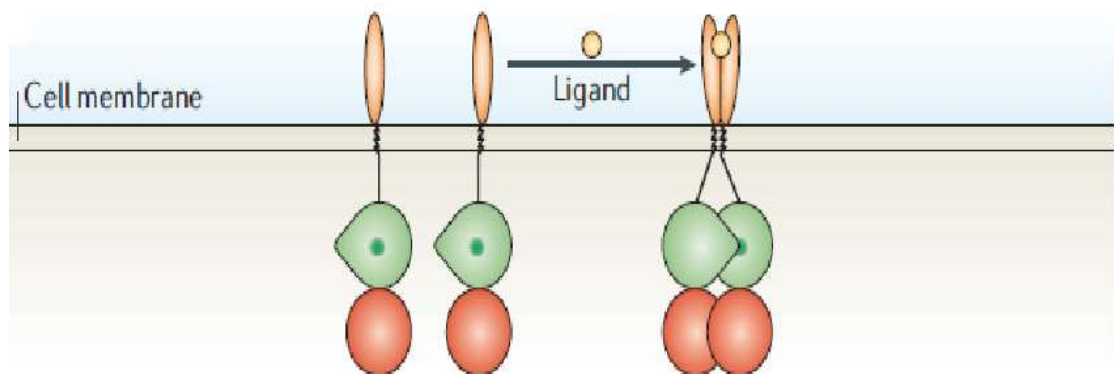


Figure 4. Schematic diagram for regulation through dimerization. Ligand binding has been shown as inhibition of LMWPTP activity by induction of dimerization. (Figure Courtesy: N.K. Tonks 2006 Nature review)

1.I.6b. Catalytic mechanism of LMWPTP

Schematic diagram of catalytic mechanism for LMWPTP is shown in figure 5. In the first step a Michaelis complex by enzyme-substrate is formed and then nucleophilic attack takes place by the catalytic Cys 12 at the substrate phosphorous which produces a thio-phosphate covalent bond. Asp129 then facilitates to depart phenolate moiety. An activated and geometrically oriented water molecule attacks to the phosphorous atom so that inorganic phosphate (Pi) can be produced. In the final step proton dissociation occurs by Cys12 which recovers the initial thiolate ion structure.

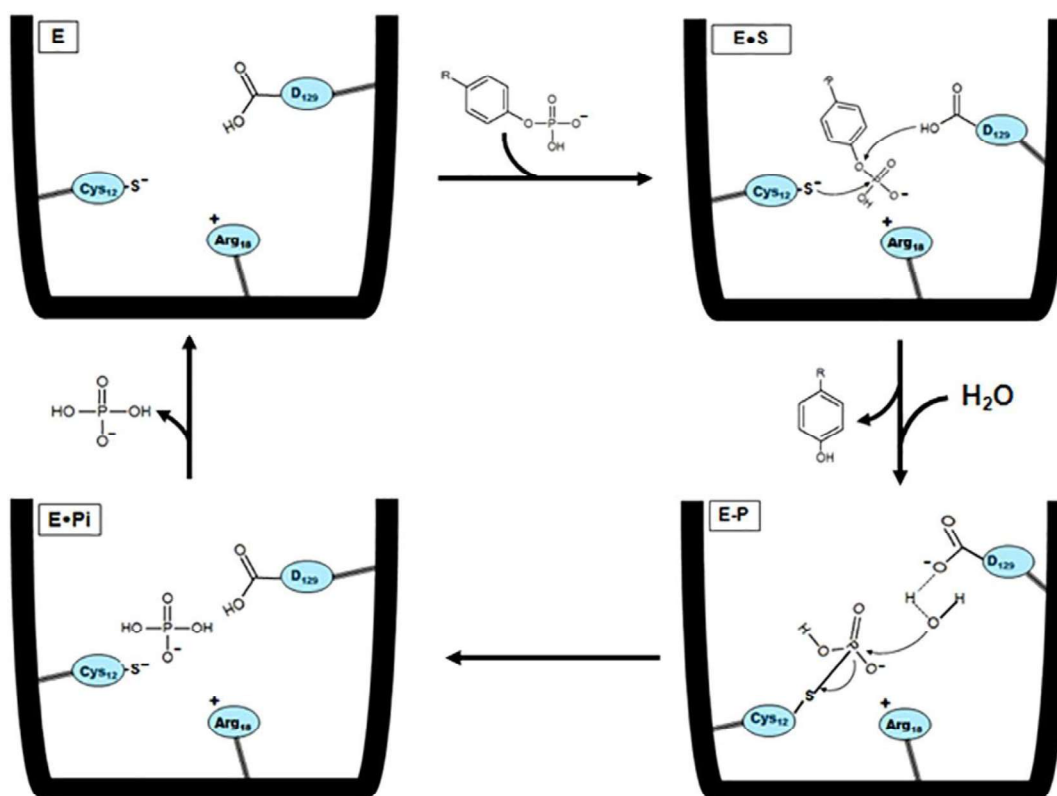


Figure 5. Schematic diagram of LMWPTP catalytic mechanism.

(Figure Courtesy: A. Caselli et al. 2016 BBA)

1.1.7. LMWPTP: potential target of therapeutics

Cell growth, differentiation and adhesion are controlled by growth-factor induced signaling pathways. Receptor tyrosine kinases and growth-factor induced signaling pathways are regulated by LMWPTPs [Fig. 6] e.g. PDGF receptor which is phosphorylated is a well-known substrate of LMWPTP, dephosphorylated in its β subunit and this is the regulatory residue for the kinase activity on PDGF receptor [43-45]. De-phosphorylation of Src kinase and STAT family transcription factors are done by LMWPTP which results in blockage of their PDGF receptor-induced activation [46-48]. FAK and p190RhoGAP are involved in matrix adhesion of cell-extracellular and they are also inactivated and dephosphorylated by LMWPTP [49, 50]. EphA2 is a receptor tyrosine kinase which is overexpressed in aggressive and metastatic types of human cancers and in tumor cells LMWPTP prefers EphA2 as substrate [51-53]. An oncogenic axis may be formed for the promotion of tumor initiation, progression and metastasis by LMWPTP and EphA2 because in many tumor types LMWPTP is overexpressed.

Negative regulation of insulin signaling takes place through de-phosphorylation of tyrosine residue in the β subunits of insulin receptor by LMWPTP [54]. Insulin signaling in liver and skeletal muscle is regulated by LMWPTP as inhibition of LMWPTP initially affects liver and adipose tissues and this is in contrast with PTP1B which suggests that PTP1B and LMWPTP has different tissue specificity [55]. On the other hand highest level of LMWPTP has been detected in diabetic and aged patients [56]. Clinical studies demonstrated that diabetic retinopathy, a diabetic complication is highest with elevated LMWPTP (specifically with IF1 isoform) [57].

53 primary human mammary carcinoma samples were analyzed by Kidd et al in 1993. They have found high level of PTP activity among ~68% of all the samples [58]. High levels of LMWPTP or LMWPTP mRNA expression have also been with many types of cancer, specifically colon cancer where LMWPTP mRNA level increased expression resulted in unfavorable outcome [59, 60]. Increase in LMWPTP level follows a step-wise increase in different levels of dysplasia [61]. Abovementioned analysis suggests that, LMWPTP can be a predictive biomarker as well as a rational novel target for the treatment of various diseases.

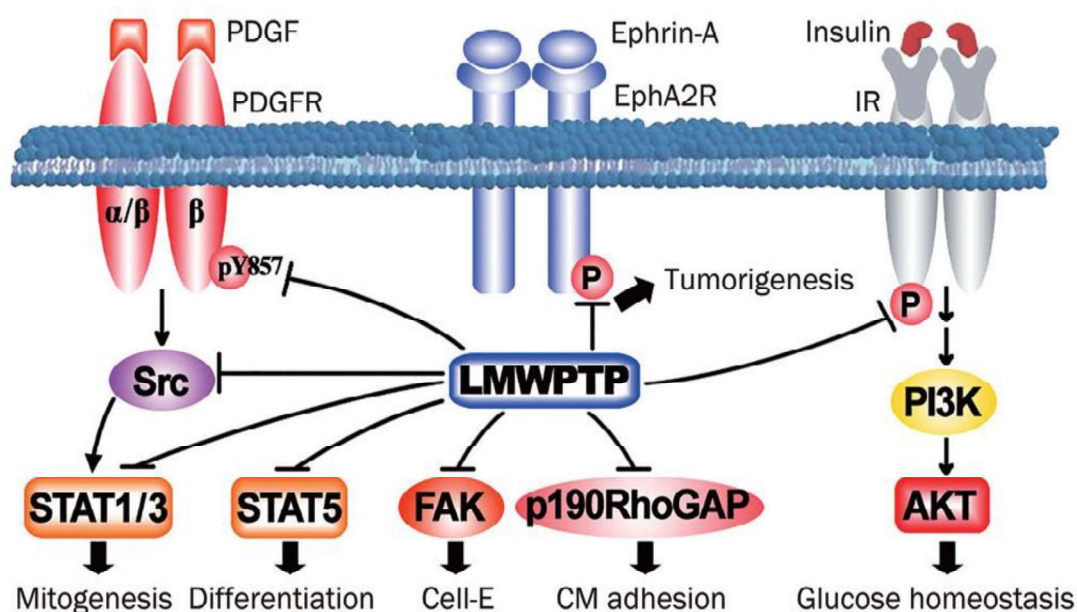


Figure 6. Physiological or pathological signaling pathways by LMWPTP where arrow represents positive regulation and T-bar represent either negative regulation or de-phosphorylation if the T-bar ends to a phosphate group or specific pY. (Figure Courtesy: Rong-jun HE et al. 2014 *Acta Pharmacologica Sinica*)

1.I.8. LMWPTP: In search of inhibitors

All the evidences discussed in 1.I.7, suggests the fact that LMWPTP to be considered as a novel therapeutic agent. A number of LMWPTP inhibitors have been identified till date as LMWPTP has been recognized as potential drug target. Until a few years ago, early transition metal anions including vanadate, molybdate, tungstate were known for their inhibitory action towards LMWPTP. Despite of being potential inhibitor and their ability to function as insulin mimetic and antitumor agents, they are not potential inhibitor for clinical usage because of their non-selectivity towards all members of PTP-family and also towards phosphodiesterase and phosphomonoesterase [62, 63]. Ionic surfactants viz., n-alkyl sulphates, n-alkyl trimethyl ammonium sulphates were able to act as inhibitors probably due to interaction with active sites and neighboring regions.[64].

First rationally designed inhibitor of Human-LMWPTP, a phosphonated-5-azaindole [Fig. 7], was synthesized only after crystal structure solution, although with weak mM affinity towards the enzyme [65]. It was clear from this inhibitor that two portions of the inhibitor is required to bind successfully with the enzyme, the first part should contain an ionic portion which can anchor into the positively charged catalytic pocket and the second part should contain aromatic moiety with suitable functional groups so that interaction with neighboring portions of the enzyme can take place.

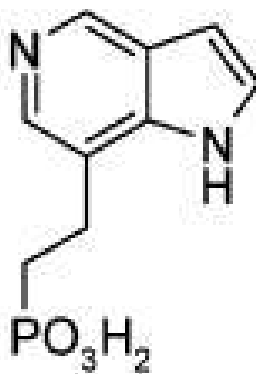


Figure 7. Inhibitor of human LMWPTP which contains phosphonic acid group

In view of this bioisosteric replacement of pTyr-phosphonate group with carboxylic acid or the isothiazolidinone system, has been considered as effective inhibitor with appreciable in vitro or cellular activity [62, 66-68]. On the other hand some compounds were selected and assayed with respect to bovine LMWPTP since it shares 94% sequence and a high degree of structural similarity with isoform 2 (IF2) of human LMWPTP. The first subset were selected with K_i ranging from 50-900 μM , all being derivatives of carboxylic acid [17-20, Fig. 8] and the second subset leads to identification of compounds 21 and 22 with K_i values 9 and 11 μM respectively and weaker inhibitor identification with K_i values 60 and 170 μM leads to 23 and 24 respectively [Fig. 8] [69, 70]. Interestingly all these inhibitors act as competitive inhibitor.

Preference in inhibition of IF1 over IF2 for human LMWPTP represents a feature which is promising towards pharmaceutical point of view that is IF1 appears to be more implicated in cancer development as well as in diabetics. Molecular docking study has shown that phenoxy- and benzyloxybenzylidene substituted derivatives are the most active one which can act as bidentate inhibitor [71-73]. Depending of the

inhibitory activity of flavonoids, a series of chromones which are active as human LMWPTP inhibitor has been reported recently [74] which are the most promising agent for further development. Several inhibitors of LMWPTP from *Mycobacterium tuberculosis* (both MPTPA and MPTPB) have also been designed and developed as anti-mycobacterial agents [75-79].

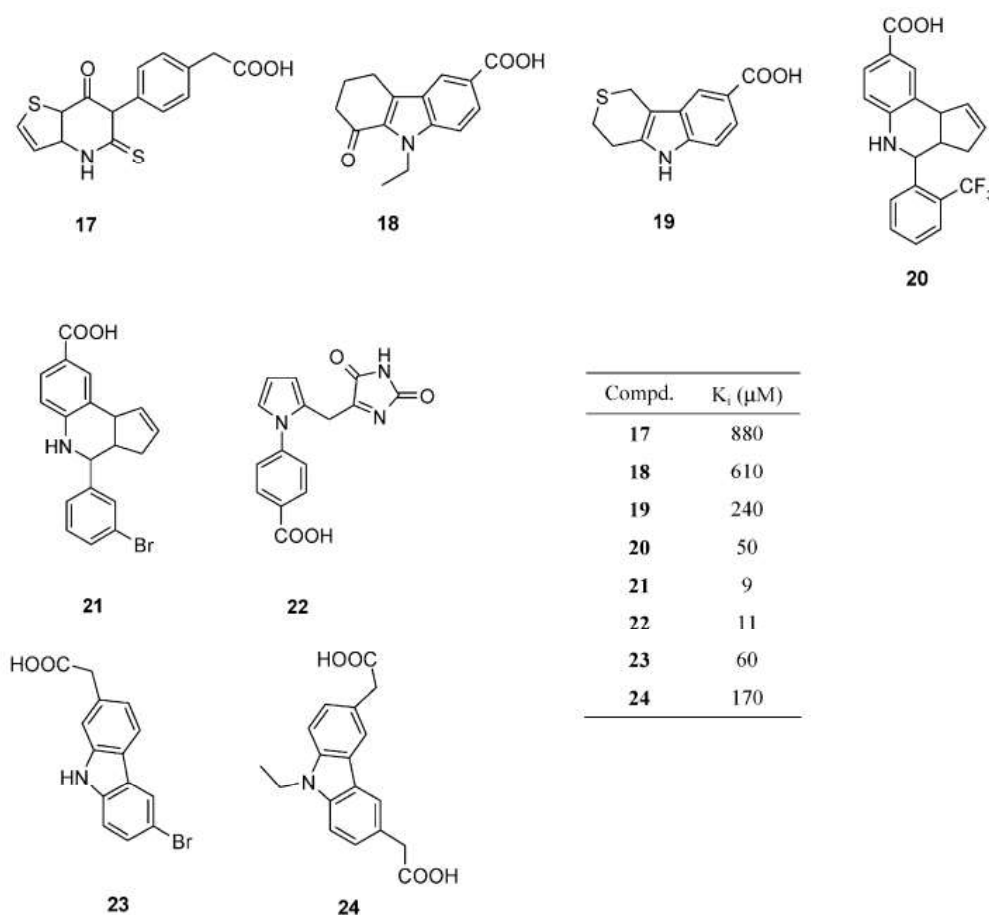


Figure 8. Inhibitors of bovine LMWPTP which contains carboxylic acid group

1.1.9. Scope of the thesis

The gram-negative bacteria, *Vibrio cholerae*, is the causative organism of the severe diarrheal disease cholera that continues to be a significant cause of death in developing countries claims approximately 95,000 lives out of 2.9 million annually [80]. The genome of *Vibrio cholerae* O395 contains two LMWPTPs (*Vc*LMWTP-1 and *Vc*LMWTP-2). These two LMWPTPs share only ~30% sequence identity among themselves and upon vigorous sequence alignment and analysis we have found that *Vc*LMWTP-2 is unique among bacterial LMWPTPs. Low sequence identity of this unique LMWTP indicates that three dimensional structure may deliver some distinct feature which has not been reported so far. It can also be useful to study the catalytic features at atomic level.

Occurrence of two LMWTP in the same organism and their biological role, especially the role of *Vc*LMWTP-2, remains to be established. Knowledge of structure as well as function of this idiosyncratic LMWTP from *Vibrio cholera* may become useful for designing of inhibitor which will lead to drug design against this pathogenic bacterium. Starting with this knowledge, the work of the present thesis represents structural as well as functional aspects of *Vc*LMWTP-2.

PART-II

1.II. Structure-function relationship of *Vc*LMWPTP-2: Study by protein X-ray crystallography

Three dimensional structures of biological macromolecules are determined by various methods such as nuclear magnetic resonance (NMR), X-ray crystallography and also cryo-electron microscopy. To enhance understanding of protein function as well as to visualize 3D structure of protein, X-ray crystallography is the most convenient among all the above-mentioned methods. Structural and functional study of proteins by X-ray crystallography involves the mode of interaction of the protein of interest with other proteins, conformational changes of the particular protein depending upon various parameters and moreover if the protein of interest is enzyme the X-ray crystallographic study also helps to decipher the catalytic mechanism. Depending upon these information rational designing of drugs (if the protein is a therapeutic target) or engineering of the protein can be performed for specific purposes. Therefore it can be said that determination of three dimensional structures directly helps to understand the functional role of the protein and relationship between structure and function has potential application among different fields of study as well as for commercial purposes. During initial days X-ray crystallography was used to study inorganic and organic compounds until the first structure of protein was solved in 1958 by John C Kendrew which was the structure of myoglobin and after that structure of hemoglobin was solved by Max F Perutz. Since then solution of macromolecular crystal structure has coped up its momentum and till date more than 10 million structures have been solved and deposited in PDB.

1.II.1 X-ray crystallographic study: steps involved

To study three dimensional structure of biological macromolecules by X-ray crystallographic technique various steps are involved. The first step is to prepare protein sample followed by successful crystallization of the purified protein. Upon crystallization data collection from this crystal is performed which leads to structure solution. Solution of structure requires proper interpretation to understand the architecture as well as function of the protein of interest.

1.II.2 Preparation of protein sample

One of the most important criteria to crystallize protein sample is to prepare pure target protein which can be achieved by isolation of the protein from source or by cloning the gene of interest into a high expression system. To get information about suitability for crystallization, stability and purity along with homogeneity of the protein sample is assed. In some cases study on enzymatic activity, proper folding and also monodispersity helps to find whether the purified protein sample is suitable for crystallization and if it is found that the target protein is not suitable the expression as well purification protocols can be modified and some extra purification steps can also be added or in some cases addition of some additives viz., ligands, metals, glycerol, reducing agents etc., mutation of particular residue, different expression system, etc. also helps to stabilize the protein which results to an ease of crystallization.

1.II.3. Technique of crystallization

When most of the criteria are filled up, protein samples are screened with a reasonable number of conditions which contains various conditions with various reagents. If ‘hit’ of crystals appeared in screening step then variation in all components viz., protein as

well as precipitant and also with reservoir is performed. Trials with different temperature, pH, additives, even with different crystallization methods are also performed so that diffraction quality crystals can appear. In most of the cases typically it takes from several days to several weeks to grow crystals.

1.II.4. Solubility Curve

Solubility curve is a curve of protein solution in the phase diagram where protein concentration vs precipitant concentration is plotted at constant temperature and this curve divides the phase diagram into various conditions where separation of phase from liquid, amorphous, solid and crystal nuclei takes place from the specific phase where the protein solution remains for indefinite time in a condition which is above the saturation point and the protein solution remains in this phase without any separation. It can be said that, the specific region in the phase diagram where kinetic phenomenon is dominant, can be identified by solubility curve. [81].

1.II.5. Phase diagram

Variation of protein solubility with precipitant concentration (e.g. poly ethylene glycol (PEG) or salt viz., ammonium sulphate (AMS) etc.) is schematically presented with phase diagram which is represented in figure 9. If the concentration is below protein solubility the crystals get dissolved and they grow in a super-saturated condition and this supersaturated condition is subdivided into three conditions, metastable, labile, and precipitation zones.

The concentration of protein is kept almost three times than the solubility in most of the cases because crystals are generated only when the solution becomes

supersaturated. With this favorable condition minimum energy barrier of crystallization activation energy can be overcome and nucleus (which is small ordered array of clusters) formation occurs at this condition. From nucleus fully formed crystals are grown although nucleation is time taking due to higher energy barrier. With supersaturated solution nucleus formation may not take place if the solution lies at the metastable zone of phase diagram although if the solution goes beyond the dotted line as shown in the phase diagram, small clusters or nucleus can be formed and this is called labile zone [Fig. 9]. This labile zone is the starting point for the formation of bigger crystals from nucleus by slow although consecutive precipitation. Phase diagram also contains a precipitation zone where no crystal formation occurs because of oversaturation of the protein solution [82].

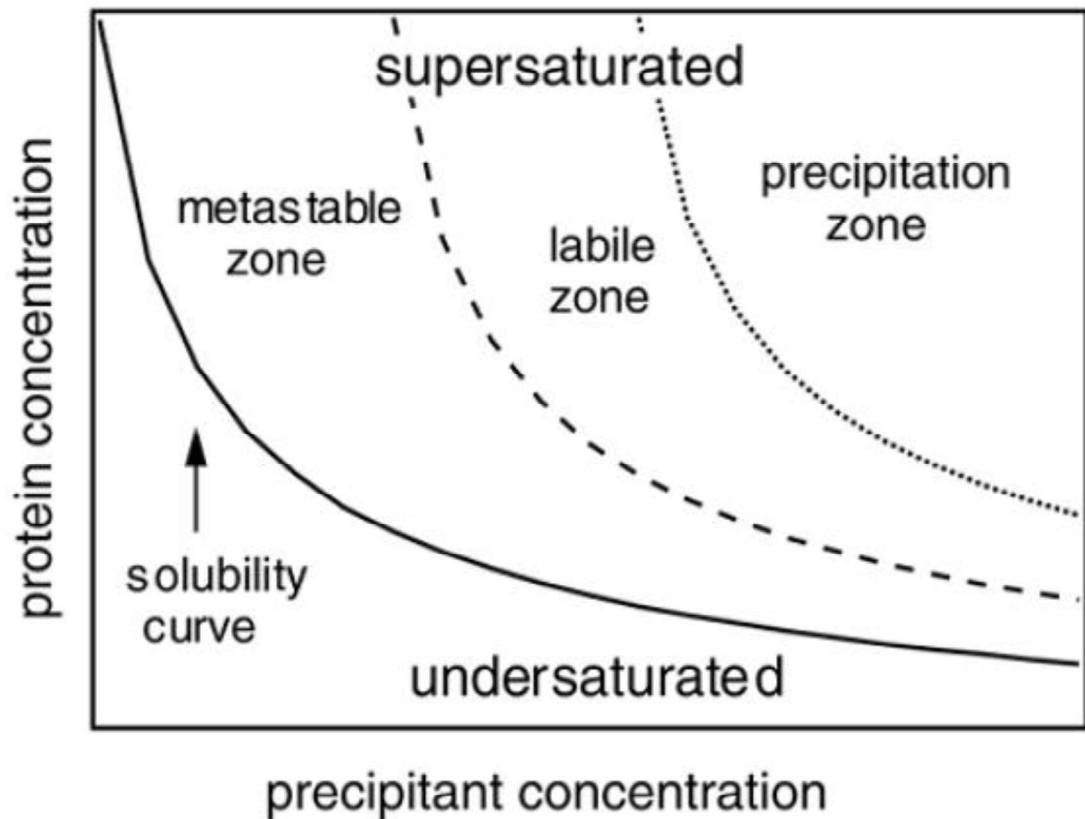


Figure 9.The Phase diagram in protein crystallization experiment which represents solubility curve as precipitant concentration function at constant temperature. (Figure Courtesy: N. Asherie 2004 Methods)

1.II.6. Vapor diffusion

The most commonly employed method in crystallization is vapor diffusion method where a droplet which contains supersaturated protein and precipitant are allowed to equilibrate against a large reservoir which contains same precipitant with the droplet but at higher concentration. This system is kept as closed system. The droplet contains lower concentration of protein as well precipitant initially but with time the drop starts to equilibrate with reservoir and with equilibration concentration of protein as well precipitant increases and when it reaches to the appropriate crystallization condition

(i.e. protein and precipitant concentration at particular temperature) growth of crystals take place in the drop. Two most popular formats for vapor diffusion method are hanging drop and sitting drop [Fig. 10]. In hanging drop vapor diffusion method a drop of protein and precipitant is suspended over the reservoir from an inverted cover-slip whereas in sitting drop method, the drop is placed on a pedestal which is separated from the reservoir. In both these methods the connecting point between cover-slip and reservoir-well is sealed by grease so that the closed system can be maintained.

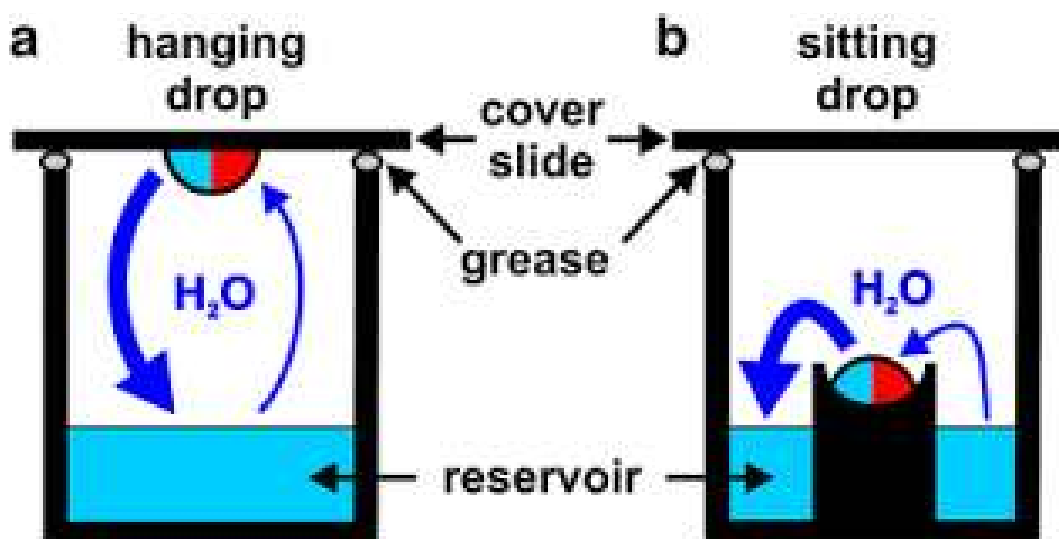


Figure 10. Vapour diffusion method: comparison between (a) hanging drop & (b) sitting drop.

1.II.7. Screening

The phase diagram is multidimensional as it depends of various functions such as temperature, pH, concentration of protein and precipitant, ionic strength (should be low ideally), and presence or absence of monovalent or divalent anions or cations etc.

Therefore screening for proper condition for crystallization of target protein is actually the search for either global or local maximum of ‘crystallization’ function.

1.II.8. Strategy of data collection

After getting crystals, they are shot to get diffraction pattern by X-ray beam and this is generally performed with single crystal of the target protein. Single crystal is either mounted in a capillary (at room temperature) or it is flash cooled with liquid nitrogen in a nylon loop and placed on a goniometer head where constant flow of liquid nitrogen is maintained so that temperature can be kept at 100K. The goniometer head is aligned with X-ray beam by adjusting screws present there. Converged single wavelength of X-ray beam is emerged from collimator and this beam strikes on the protein crystal so that diffraction data can be collected at the detector. Position of the detector is also adjusted to collect best diffraction pattern and a good diffraction pattern doesn’t consist any diffraction from salt or ice-crystal and the spots should not be streaked. With few initial images crystal symmetry, unit cell parameters, highest resolution limit and also orientation of the crystal is ascertained. With this information strategy to collect data with highest completeness and resolution is calculated from automated software. To get a complete set of diffraction pattern crystal is rotated through a small angle which is called oscillation. With higher symmetry, lesser data is required [83].

1.II.9. Data processing step

Data is recorded in a two dimensional detector although it is generated from a three dimensional crystal with a variation in intensity in each spot which corresponds to the

variation in electron density. Two separate pieces of information are achieved from this diffracted spots, one is the geometrical arrangement of this reflection from which information about the crystal lattice can be acquired and the other is intensity of the reflection from which part of information about lattice can be acquired. With indexing the spots are found, integer numbers (hkl) are given and crystal geometry is assigned properly so that this spots can be integrated. The spots are also merged and scaled and finally the phases are calculated to obtain the electron density map which is important to solve the structure of the target protein.

1.II.10. Phase problem

X-rays have suitable wavelength to interact with electrons of atoms. Therefore upon exposure of crystals into x-rays, from every point of crystal lattice diffraction occurs in proportion with the electron density of that point. Depending on the arrangement of atoms a particular diffraction pattern is composed and therefore position of atoms can be deduced from the analysis of the particular pattern of diffraction. On the other hand the intensity of diffraction depends on the fact that whether these X-rays interfere with each other in a particular direction in constructive way or destructive way which actually depends on the particular position and spacing of electron density within the crystal. Constructive interference leads to strong intensity of diffracted spots in a particular direction whereas destructive interference leads to weak one. As the crystal is consisted with a repetitive pattern of identical molecules the pattern of electron density also shows a similar feature upon regular interval and this repetitive unit is called the unit cell of crystal lattice which is defined by three dimension of crystal axes, a, b, and c. As scattering from unit cell is repetitive, it shows strong intensity

towards a particular direction and scattering from other directions become weak. Therefore the overall diffraction patterns formed from a three dimensional lattice on reciprocal spacing and directions with real lattice and each diffraction images are two dimensional section from through this pattern and obtained from a particular orientation of the single crystal and is known as reciprocal lattice. The diffraction spots are identified with integer values (hkl) known as Miller indices. The intensity of each diffracted spots is dependent with the feature of spacing of electron density within each unit cell and this electron density and intensity is related with a mathematical relationship which is famously known as ‘Fourier Transform’. So diffraction pattern can also be predicted if the unit cell is known or from the scattering unit cell can also be determined. Therefore determination of electron density from a single unit cell is required through Fourier transformation and therefore only the diffraction spots need to be considered not the space. Every diffraction spots may then be considered as a single wavelength with amplitude and phase related to X-rays which are scattered at a particular direction and the wave is mathematically represented as ‘structure factor’ although from experiments only amplitude of this structure factor can be determined and since the phase information is absent, it is called as ‘phase problem’ in X-ray crystallography. The relation between intensity of each diffraction spots measured with the amplitude of the wave or structure factor is expressed as equation (1)

$$I(\mathbf{h},\mathbf{k},\mathbf{l}) \propto |F(\mathbf{h},\mathbf{k},\mathbf{l})|^2 \text{ -----(1)}$$

Where (h,k,l) is the co-ordinate of any point at reciprocal lattice relates to real lattice (x,y,z).

It should be noted that without any information about phase reconstruction of electron density in the unit cell is impossible and in order to get an interpretable electron density map it is necessary to obtain both amplitude and phase of structure factor.

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

In the above mentioned equation (2) ρ is the electron density of a point in unit cell with coordinates (xyz), $F(hkl)$ the magnitude of structure factor, (hkl) are Miller indices of reciprocal lattice points and $\Phi(hkl)$ represent phases of structure factors.

There are several methods available to solve phase problem.

1.II.11. Patterson methods

To solve small molecule structure Patterson method is very effective although with increasing number of atoms, effectiveness of this method falls. Small molecule structures can be solved without phase information because for small molecules magnitude of structure factor alone contains information about the spacing of the atoms in the structure to be solved.

1.II.12. Direct methods

Well-ordered small as well intermediate sized molecules from a very high diffraction angle structure factor can be calculated because information about fine features of unit

cell can be obtained from high angled diffraction spots. Missing information about phase can be calculated directly from mathematical formulations between structure factors and for this reason it is called 'direct methods'.

1.II.13. Multiple Isomorphous Replacements

Larger protein molecules are less rigid than small inorganic molecules and due to this fact imperfections are included in an unit cell of crystal structure which limits the diffraction upto $\sim 2\text{-}3\text{ \AA}$ and this problem is solved by a common approach namely multiple isomorphous replacements (MIR) [84]. A known change is made in the unit cell and then diffraction pattern is measured to get information about unknown phase and this includes inclusion of heavy atoms without any change in the structure of the target protein so that both the structures become 'isomorphous'. By use of direct method or Patterson method it is generally possible to locate the heavy atoms since they diffract more strongly than other atoms and locating the heavy atoms scattering from those atoms give the information about magnitude as well phase. From diffraction spots structure factor of native crystals are also calculated in parallel and when structure factor upon inclusion of heavy atom becomes stronger than native one it is concluded that diffraction from heavy atom included structure is interfering constructively and in this case the unknown phase of native structure becomes similar to that of the known phase of heavy atom included structure and from a comparison between these two, unknown phase can be calculated.

1.II.14 Multi wavelength Anomalous Dispersion

Upon introduction of a tunable wavelength at the synchrotron Multi wavelength Anomalous Dispersion has become a popular alternative of MIR [85]. Though scattering from atoms are independent with wavelength, there are ‘atomic edges’ where variation in scattering takes place with altering wavelength and this variation occurs in both amplitude and phase. Therefore by varying wavelength contribution of a particular atom (in most of the cases heavy atoms) can be calculated which helps to calculate the position of that atom and then phase information can be calculated following the same procedure as MIR.

1.II.15 Molecular Replacement

To estimate phase of a new protein sometimes phases from the structure factors of an already known protein is used. A single native data set is used in this process to determine structure of the new protein and the known structure is referred as ‘phasing model’ and this method which involves placement of model in the unit cell of target protein is known as ‘molecular replacement’ [86]. Molecular replacement method will be discussed in detail as we have solved the structure of our protein of interest using this method.

To illustrate about this method, serine proteases like trypsin, chymotrypsin, elastase are similar in structure and conformation. If target protein is serine protease (specifically mammalian) and if sequence homology suggests similarity in structure then any of the abovementioned serine protease can be used as phasing model for the target protein. On the other hand upon binding of ligand in most of the cases conformational changes occur. If structure of protein ligand complex has to be solved

the ligand free structure of the same protein (if available) can also be used to solve protein-ligand complex structure and there is no need to start from the scratch.

When target protein and phasing model are isomorphous, then to compute $\rho(x, y, z)$ from native intensities of the target protein phases from the known protein can be used directly following the equation,

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}^{\text{target}}| e^{-2\pi i(hx + ky + lz - \alpha_{hkl}^{\text{model}})} \quad \text{-----}(3)$$

$|F_{hkl}^{\text{target}}|$ can be obtained directly from native intensities of target protein and $\alpha_{hkl}^{\text{model}}$ are phase of the phasing model. In most of the cases iterative process of phase improvement changes the phase of the model into the phase of the target protein which reveals the desired structure.

When the phases of model are not isomorphous with the phases of target protein the problem becomes difficult. Phases of atomic as well as molecular structure factor changes with location of atoms in the unit cell. To use the known protein as a phasing model, superimposition of the known protein with the target protein in its unit cell is performed and then phase of the properly oriented model is calculated and then only calculation of the structure factor from properly oriented can be done. The phase from these computed structure factor can be used as initial estimate of the desired phase.

From native data unit cell-dimension and symmetry of new protein can be determined. Therefore the model is placed according to the symmetry of the target protein which gives some constrains on the model although not enough for phase

estimation. Theoretically it is possible to search all orientations and positions of the model in the new unit cell and structure factor of the model F_{calc} is calculated accordingly and their amplitudes are compared with the measured amplitudes which is obtained from the intensities of the diffraction of target protein. The orientation and position with best match of computed phase is used as starting phase.

Innumerable possibilities of the search for probable orientation and position has been simplified by choosing to search for best orientation over best position since it is possible to search orientation without depending on the location by using the Patterson function as Patterson map changes with change in orientation of the molecule. From this it can be said that Patterson map provides an idea about the orientation of the model in the unit cell of the target protein. Therefore model and target of same orientation in the similar unit cell dimension with same symmetry will give similar Patterson map.

To compare results and also to determine optimum values of parameters the search system goes through all trial values (including angles and coordinates) for the model. Comparison between orientations is performed by rotation function computation. Rotation function evaluates correlation of Patterson map between target protein and phasing model in various orientations. In this orientation search or rotation search, large values of model Patterson function ($P^{\text{model}}(u, v, w)$) is searched at the locations of peak for the Patterson map of protein of interest. One of the method is to integrate the product of model Patterson and target Patterson over Patterson unit cell for each orientation and where either of the Patterson has a peak but if the other doesn't the product is zero but when the peaks of each Patterson coincides product becomes large. Therefore if there are many coincident peaks the integral value

of the product will be very large. For this type of search the rotation function is expressed as

$$R(\phi, \varphi, \chi) = \int_{u,v,w} P^{\text{target}}(u, v, w) P^{\text{model}}\{(u, v, w) \times [\phi, \varphi, \chi]\} du dv dw \quad \text{-----}(4)$$

This equation can be expressed as for each rotation angles ϕ , φ , and χ the integral of two Patterson function, target molecule $[P^{\text{target}}(u, v, w)]$ and model $[P^{\text{model}}(u, v, w)]$ produces rotation function R and (u, v, w) represents co-ordinates which are operated on by rotation matrix² $[\phi, \varphi, \chi]$ which produces specific orientation which is related to the target Patterson. This abovementioned function exhibit maxima where the two Pattersons have many peaks which are coincident and this maxima suggests the placement of phasing model with best orientation in the unit cell of the desired protein.

The criteria for location or translational search is that there should have a correspondence between the expected structure-factor amplitude from the phasing model in a trial location and the actual amplitude which is derived from the native data set of the protein of interest and this criteria is expressed as R-factor, which is also a criteria for phase improvement for final structure determination. R-factor has been defined as

$$R = \frac{\sum ||\mathbf{F}_{\text{obs}}| - |\mathbf{F}_{\text{calc}}||}{\sum |\mathbf{F}_{\text{obs}}|} \quad \text{-----}(5)$$

This equation can be expressed in words as, for each reflection, difference is computed between the observed structure-factor amplitude which is derived from native data set, $|F_{obs}|$, and the structure-factor calculated in the current trial location of the model, $|F_{calc}|$, and their absolute value is taken, which gives the magnitude of the difference. These magnitudes for all reflections are added and then it is divided by the sum of the observed structure-factor amplitudes. Therefore if the observed and calculated structure-factors agree with each other, R becomes small and for perfect agreement R is zero. For proteins R values with 0.3-0.4 gives the best placement of the phasing model which often provides adequate initial estimates for determination of phases.

1.II.16. Phase improvement and refinement

Upon generation of initial model phases are refined to improve the model using solvent flattening, histogram matching, noncrystallographic averaging etc. A typical protein electron density is $0.43 \text{ e } \text{\AA}^{-3}$ and through solvent flattening negative electron densities are removed which sets the values at solvent region $0.33 \text{ e } \text{\AA}^{-3}$. Electron density values are altered by histogram matching to coincide with an expected electron density distribution. If more than one copy of molecules are present in the asymmetric unit, equivalence on electron density values are imposed by non crystallographic averaging [87]. The atomic positions of a model and respective B-factors are also taken care to yield a better phase from which a new model is generated and this goes on with a cycle until correlation between diffraction data and generated model is maximized which is measured by R-factor or reliability factor defined in

equation 5 previously. Method for cross verification is calculation of R_{free} where ~5-10% of diffraction data set is set aside and never used for refinement.

1.II.17. Deposition of structure after validation

When R-factors are stopped to decrease a validation model in terms of bond angles, bond lengths etc. (stereochemical) are performed and usage of Ramachandran Plot is the most popular one to assess geometrical validation (specifically peptide bonds and torsional angles from usual conformations of the amino acids). Thermal factors associated with each atoms are also assessed and in general it is lower for the internal parts and higher for the external parts which are exposed towards solvent. Finally the structure solved which meets the abovementioned criteria are deposited into Protein Data Bank (PDB).

Chapter-2

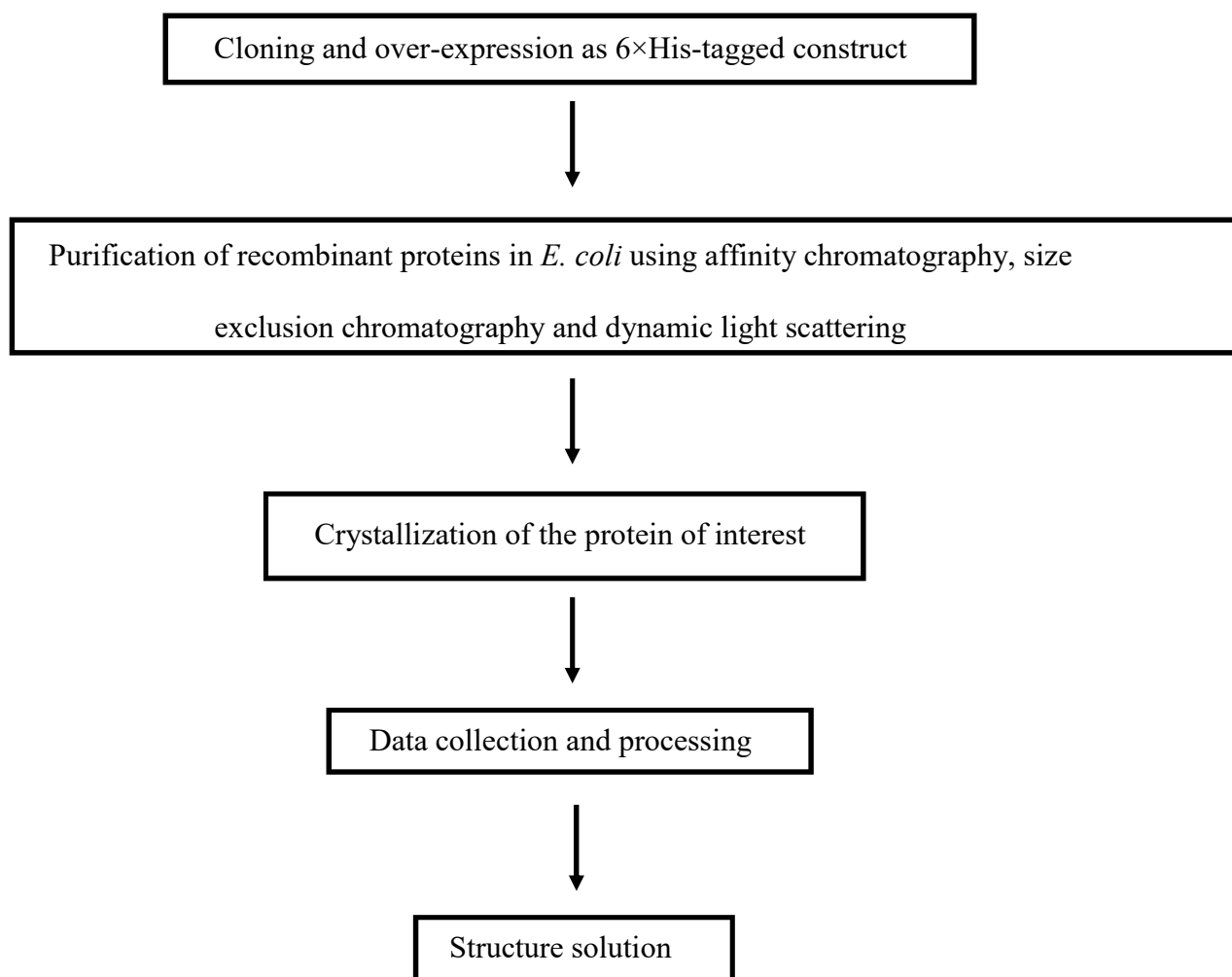
Materials

&

Methods

2.1 Cloning, Over-Expression, Purification, Crystallization, Data-collection, Data processing, Refinement

Structural and structure based probable functional aspects of *Vc*LMWPTP-2 and its mutants were subjected to follow the flow diagram given below. Thereafter detailed description of each procedure has been elaborated.



2.1.1. Cloning and over-expression of *Vc*LMWPTP-2 and its mutants:

For successful cloning first the gene of interest was amplified by polymeric chain reaction (PCR) with proper forward and reverse primer. PCR product was then purified by Qiagen Gel Extraction Kit (following the user's guide) and the purified product was then digested with proper restriction enzymes followed by purification of digested product. Simultaneously vector was also digested with the same restriction enzyme(s) and the digested vector was also purified. Digested vector and PCR product was then allowed for ligation. Ligated product was transformed into electro-competent cells and cells which carried ligated product were then screened. After plasmid purification from these cells, plasmids carrying insert was again transformed to check over-expression of the protein of interest by addition of inducer.

Wild type construct was cloned from the genomic DNA of *Vibrio cholerae* and mutants were prepared by site directed mutagenesis (by two step PCR method) following the same procedure as wild type. *Vc*LMWPTP-2 (166 amino acids) was cloned into pET-28(a+) vector under the control of T7 promoter and 6×His-tagged wild-type protein was overexpressed into B121 (DE3) strain of *E. coli*. The complete DNA sequence of this construct was confirmed by DNA sequencing. The same procedure was followed to clone and overexpress the single (C12S, C17S, T13L) and double (C12S/C17S) mutants.

The sequences of the forward and reverse primers along with the recognition sites of the restriction enzymes are listed in **Table 2**.

Table 2: Sequences of Primer

Clone name	
<i>Vc</i> LMWPTP-2	Forward Primer- with <i>NdeI</i> site 5'-GCCGCC <u>CATATG</u> AAGGTTAAAGGTTTATCAG -3'
	Reverse Primer – with <i>BamHI</i> site 5'-CGGGGATCCTTATTGAGATAAAATTTTCGTTGCACGC -3'
<i>Vc</i> LMWPTP-2 C12S	Mutation forward primer for the 1st step PCR of SDM 5'- GGTTTATCAGTATTAGTCGTG <u>A</u> GTACAGGTAATTTGTGTCGTTTCG -3'
	Mutation reverse primer for the 1st step PCR of SDM 5'- CGAACGACACAAATTACCTGT <u>A</u> CTCACGACTAATACTGATAAAC C-3'
<i>Vc</i> LMWPTP-2 C17S	Mutation forward primer for the 1st step PCR of SDM 5'-GTCGTGTGTACAGGTAATTTG <u>A</u> GTTCGTTCCCAATGGCAGAG- 3'
	Mutation reverse primer for the 1st step PCR of SDM 5'-CTCTGCCATTGGCGAACG <u>A</u> CTCAAATTACCTGTACACACGAC- 3'
<i>Vc</i> LMWPTP-2 T13L	Mutation forward primer for the 1st step PCR of SDM 5'-TTATCAGTATTAGTCGTGTGT <u>C</u> TAGGTAATTTGTGTCGTTTCG-3'
	Mutation reverse primer for the 1st step PCR of SDM 5'- CGAACGACACAAATTACCT <u>A</u> GACACACGACTAATACTGATAA-3'

For the mutants the nucleotide written in bold and underlined mark indicates the changed position of required mutation.

2.1.1.1a. Polymerase Chain Reaction (PCR):

PCR technique [88] was used to amplify the gene *VC0395_A0440* which encodes *VcLMWPTP-2* from the genomic DNA of *Vibrio cholerae*. Annealing temperature T_m was chosen to be around below 5°C than the lowest (theoretical) melting temperature of both forward primer and reverse primer. 32 PCR-cycles were carried out. To optimize T_m , initially ‘gradient PCR’ with varying temperature was carried out and depending on that large scale PCR was performed for the wild type protein. The program used for PCR is given below:

- (i) Initial denaturation at 95°C for 5 minutes.
- (ii) Denaturation at 95°C for 30 seconds
- (iii) Annealing at $X \pm 5^\circ\text{C}$ for 30 seconds (X varies for different constructs)
- (iv) Polymerization at 72°C for 45 seconds (32 cycles)
- (v) Final extension at 72°C for 10minutes.

Calculation of the reaction mixture for 100µl of total volume is listed in Table 3.

Table 3: Calculation 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.	

2.1.1.1b. Site directed mutagenesis (SDM):

Single and double mutants were prepared by Site Directed Mutagenesis (SDM) by two-step-method as described in [89] where proper annealing temperature was judged by the first step which is equivalent to the first step of ‘gradient PCR’ using forward primer of wild type and mutation reverse primer in one set whereas reverse primer of wild type and mutation forward primer in another set. PCR products from the first step of both the sets were taken as template for the second step. Second step for SDM is equivalent with the second step of wild type fragment. The calculation is tabulated in Table 4.

Table 4: Calculation for the two-step-PCR reaction mixture

Components	Final concentrations 1st step PCR	Components	Final concentrations 2nd step PCR
Template- <i>Vc</i> -genomic DNA	~20ng	Template- PCR products from set—I & set – II	~10ng each
‘Forward’ and ‘Mutation reverse’ primer – set I	0.8pmol each	Forward primer	0.8pmol
‘Reverse’ and ‘Mutation forward’ primer – set II	0.8pmol each	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.			

2.1.1.2. Digestion and Ligation:

Qiagen Gel Extraction Kit was used to purify the PCR product and digested overnight at 37°C water-bath with *NdeI* and *BamHI* from NEB. pET28a(+) vector was also digested (two times) overnight following the same procedure with the same restriction enzymes. The calculation for the restriction enzyme digestion has given below in Table 5.

Table 5: Calculation for restriction enzyme digestion

Restriction enzyme double digestion	
Components	Final concentration
10X Buffer3 (NEB)	1X
Purified PCR product/Vector	2-4µg
Enzyme NdeI and BamHI	1U each
100X BSA	1X
Autoclaved water to make the final volume 20µl.	

The purified digested vector and insert were ligated at 4°C for overnight following the protocol mentioned in Table 6.

Table 6: Components of Ligation

	Standard reaction	Positive control	Negative control
10X Ligase Buffer (Promega)	1 μ l	1 μ l	1 μ l
pET 28a+ Vector (Promega): PCR product	1:3	1:3	Only vector added
T4 DNA Ligase (Promega)	1 μ l	1 μ l	1 μ l
Autoclaved water to make the final volume 10μl.			
Total	10 μ l	10 μ l	10 μ l

2.1.1.3. Transformation:

Ligated product was directly transformed into electro-competent *E.coli* XL1Blue cells according to the manufacturer's instructions. At 37°C when the transformed cells were grown upto log phase, they plated on agar plates containing both tetracycline and kanamycine antibiotics upto 12.5 μ g/ml and 30 μ g/ml final concentrations respectively.

2.1.1.4. Screening and Analysis:

Few handpicked colonies were subjected for overnight cultures and for further screening. Plasmid isolation was performed from overnight cultures by Qiagen miniprep kit following the manufacturer's protocol. Purified plasmids were subjected for digestion with the corresponding restriction enzymes for digestion product or insert. The insert was checked by running agarose gel although final conformation was performed by DNA sequencing from Amnion Sequencing.

2.1.1.5. Protein over-expression:

In *Escherichia coli* strain BL21 (DE3) (Novagen) over-expression of 6×His-tagged proteins were carried out in Luria broth medium. Transformation of BL21 (DE3) competent cells were performed with the recombinant pET28a(+) plasmids which contains the insert. At 37°C when the transformed cells were grown upto log phase they were further spread onto agar plates containing 30 µg/ml kanamycin. Next day some of the colonies were picked from the plate and grown overnight in 2mL of LB medium in presence of kanamycin antibiotic at 37°C at 175 rpm. The overnight grown culture was inoculated next day in a fresh 3mL of LB medium at 1:50 ratio which was allowed to grow till OD reached to ~0.4-0.5 at 37°C at 175rpm. Then 1.5 ml from this culture was taken out to preserve as ‘uninduced’ cell with the induction of the rest cells with IPTG at a final concentration of 0.5mM. Both ‘induced’ and ‘uninduced’ cells were allowed to grow for another 4-5 hours maintaining the same condition. The cells were then harvested in a 1.5mL micro-centrifuge tube by centrifugation at 13000g for 2 minutes. The supernatant was thrown and pellet was dissolved in 30µl of autoclaved H₂O. It was then vortexed and mixed with 20 µl 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). Cell lysis was then performed by boiling the sample prepared in water for 10 minutes followed by centrifugation at 12000g for 7 minutes and finally 15µl of the sample was loaded in SDS-PAGE which was allowed to run at 200V and 30mA at room temperature and then overexpression of recombinant proteins were checked by staining the gel with Coomassie R-250 (Sigma) from uninduced vs. induced gel profile. Positions of the protein of interest were confirmed by protein marker.

2.1.2. Protein purification by affinity and size exclusion chromatography

2.1.2.1. Affinity chromatography: purification of 6×His Tagged protein

Histidine tagged proteins (6×His tagged) show remarkable affinity towards Ni^{2+} ion immobilized Nitrilotriacetic acid (NTA) resin. Histidine residues in the His tag binds with the vacant co-ordinate of immobilized co-ordinate sphere of Ni^{2+} and this interaction is highly specific. Due to this specificity the proteins which are untagged by histidine although present in cell lysate may also bind if the Histidine residue of these unspecific proteins remains in the close proximity to the surface, although untagged protein will bind with lesser affinity than that of tagged proteins where consecutively six histidine residues are present. The non-specific proteins can be washed out under relatively stringent conditions without affecting target protein. Elution of target protein is possible under higher concentration of imidazole or by reducing pH of the buffer.

To obtain maximum protein two methods were followed. In the first method, overnight grown culture of BL21 (DE3) cells containing recombinant plasmid of pET28(a+) vector which contains insert of the protein of interest was inoculated (1:100 dilution) into freshly prepared 1litre LB medium supplemented with Kanamycin antibiotic (with a final concentration upto 30 $\mu\text{g/ml}$). It was then allowed to grow at 37°C at 175rpm until OD_{600} reached upto 0.4-0.5. Then 0.2-0.5mM IPTG induction was performed maintaining the same condition for another 4-5 hours. In the second method, same procedure was followed but after inoculation it was allowed to grow for ~2hours at 37°C 190rpm. Then the temperature was reduced at 20°C (which took ~30-45 minutes) and at this reduced temperature IPTG was added and allowed to

grow for another 12-14 hours at 100rpm. Addition of 0.4mM IPTG in the second method produced best result.

Maintaining the optimized condition, i.e. low temperature (20°C) and slow shaking (100 rpm) after IPTG induction (0.4mM), the bacterial cells were allowed to grow for overnight and then harvested in 50ml oakridge centrifuge tubes at 7000g for 5 minutes. Resuspension of cell pellets of wild type (WT) protein was performed in lysis buffer containing 50mM MOPS, pH7.7, 300mM NaCl, 5% glycerol, 2mM PMSF (Phenylmethylsulfonylfluoride which is a serine protease inhibitor) along with pinch of (~0.5–1mg/ml) lysozyme and allowed to incubate for ~1hour in ice. The cells were then sonicated with ~45-50 bursts at 20 seconds at 200W with an interval of 45seconds after each bursts. Cell lysate after sonication was centrifuged for 45 minutes at 12000g and after centrifugation supernatant was loaded in a Ni-NTA column which was pre-equilibrated with lysis buffer. To elute weakly bound non-specific proteins maximum wash was performed. Starting from lysis buffer wash, wash with 5mM imidazole, 20mM imidazole and 50mM imidazole along with lysis buffer was performed several times. Then the 6×His tagged target protein was eluted by elution buffer which contains 200mM imidazole along with lysis buffer. The purity of the eluted protein as then checked with 15% SDS-PAGE. Eluted protein contains few contaminants. It was then pooled and concentrated by centrifugation at 5000 rpm at 4°C and finally buffer exchange was performed in presence of lysis buffer and concentrated upto ~1ml for thrombin cleavage.

Mutants were also purified with same buffer maintaining same conditions except for C12S and C12/17S mutants, they were purified at 10°C maintaining high percentage of glycerol as they were very prone towards precipitation. For mutants

lysis buffer 50mM MOPS, pH7.7, 300mM NaCl, 10-15% glycerol was used as lysis buffer and then the abovementioned procedure was followed for crystallization trials. For functional studies, WT and all the mutants were purified in 50mM Tris, pH7.7, 300mM NaCl, 10% glycerol lysis buffer. All other procedures and conditions remained similar.

2.1.2.2. 6×His tag removal by thrombin

Hexa histidine tag is deliberately introduced into recombinant protein to ease the purification by metal ion chelation in affinity chromatography and this hexa-histidine tag is cleaved by the use of thrombin. The consensus sequence Leu-Val-Pro-Arg-Gly-Ser, deliberately introduced between His tag and the desired protein sequence, is recognized by thrombin and cleavage occurs between Arg-Glysite. In presence of Ca^{2+} ion at pH 8.0 it works most efficiently. It is reported that presence of imidazole interferes in thrombin cleavage. Therefore we have performed buffer-exchange by lysis buffer to decrease concentration of imidazole upto ~5-10mM. Optimization of the amount of thrombin, time and temperature for cleavage has been standardized. After optimization, 2 μL thrombin (1.6 units/ μL , Novagen) was added into ~1ml of protein solution containing ~4mM of Ca^{2+} ion and kept at 4°C for ~40-45 hours for complete cleavage of His tag.

2.1.2.3. Purification by Size exclusion chromatography

Upon completion of thrombin cleavage reaction the protein was further purified from cleaved histidine tag, enzyme thrombin and other contaminants from Ni-NTA affinity chromatography by size exclusion chromatography. After complete removal of His tag, the sample was directly loaded into the column of ÄKTA Prime chromatographic system using Superdex 200 increase (GE-Healthcare) column (10×300mm; bed volume ~24ml). Before loading the sample mixture the column was pre-equilibrated with gel-filtration buffer containing 50mM MOPS, pH7.7, 300mM NaCl, 3% glycerol, 0.02% sodium azide (sodium azide used to avoid any microbial contamination). The elution profile was monitored by absorbance at 280nm and throughout this procedure temperature was maintained at 10°C with flow rate 0.4ml.min⁻¹. The fractions containing protein was then collected and concentrated upto ~4-5mg.ml⁻¹ by centrifugation at 5000 rpm using 10kDa cut-off centrifugal unit of Millipore.

2.1.3. Crystallization of protein of interest

Before starting crystallization trials, solubility of the target protein was checked by using 3M ammonium sulphate and 50% PEG6K as stock of precipitant and upon dilution of precipitant as well as protein concentration the optimized start point of crystallization was determined. 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 [90] were used to explore the initial crystallization conditions. Typically, 2µl protein solution was mixed

with 2µl precipitant solution, inverted over a reservoir containing 600µl precipitant solution and maintained at both 277 and 293K. Plate shaped crystals of *VcLMWPTP-2* appeared in 2.4M ammonium sulfate, 0.1M HEPES pH7.0, 5% glycerol as precipitant against reservoir 3 M ammonium sulphate, 0.1 M Tris-HCl, 0.2 M NaCl at 277K.

VcLMWPTP-2 C12S mutant crystallization trials were performed using various home-made and commercial grids both at 277 and 293K. Out of which ammonium sulphate produced small sized crystals at 277K which need further improvement for diffraction.

2.1.4. Collection of Data and further processing:

Crystals of *VcLMWPTP-2* were looped out from the crystallization drops using a 20µm nylon loop and flash-cooled in liquid nitrogen (Oxford Cryo systems) at 100K. Cryo-protectant used during crystal freezing composed of 30% glycerol, 2M AMS, 0.1M MOPS. X-ray diffraction data was collected on PX BL-21 [91] beam line of Indus-2 synchrotron at RRCAT, INDORE. Collected data were further processed and scaled using *iMOSFLM* [92].

2.1.5. Structure determination of protein of interest

At first number of molecules in the asymmetric unit were determined by calculating Matthews coefficient and solvent content. A ‘model’ for phasing was then generated from the nearest homologue of *VcLMWPTP-2* using the program Chainsaw of CCP4 package [93]. As chainsaw identifies several file formats, we used PIR (.pir) as an input to build a model by trimming non-conserved and (or) all residues. We started structure solution by molecular replacement by Phaser [94]. The model was used as

input for ‘Molecular Replacement’ (MR) of the program PHASER so that the unknown structure can be placed correctly in the unit cell using ‘likelihood’ based method. The MR method combines anisotropy correction, fast rotation and translational function which are likelihood enhanced, packing and refinement modes which are applicable for multiple search models and also a set of all possible space groups.

Upon indexing, refining parameters and integration the .mtz file generated from iMOSFLM [92] was provided as experimental data into Phaser. The model generated from chainsaw i.e. the co-ordinate file (.pdb) was used as model for molecular replacement. Whether Phaser has solved data, was judged initially by LLG and TFZ (Translation Function Z-score) values and the initial model (.pdb file) was modified accordingly. Single PDB file produced with the top solutions (LLG, TFZ, RFZ, PAK etc.) was used for further refinement.

Refinement was carried out by Phenix refine [95]. Just after molecular replacement refinement started with refining the co-ordinates (specifically rigid body refinement and real space coordinate refinement) along with atomic displacement parameters (isotropic B-factors). As R-factors resulted <50% by refining more than 5 cycles on each macro-cycle, the co-ordinates generated was used for model-building into maps from X-ray data using Coot [96] and the output was then used for another cycle of refinement. With improvement of X-ray data confident building of model was performed keeping in mind about Ramachandran as well as rotamer restrains. With further improvement, inclusion of ligands in all the molecules present was performed and these cycles were gone on with refinement of co-ordinates and atomic displacement parameters. In the mean time we also performed simulated annealing

(both Cartesian and torsional) as well as non-crystallographic symmetry (NCS) restraints. Until then we refined isotropic ADPs only and upon adequate decrease of R-factors we started TLS refinement. We picked water molecules after that and performed an update with every macro-cycle with refinement of individual coordinate and B-factors. Water molecules with $B < 1.0$ or $> 50 \text{ \AA}^2$, occupancy other than 1.0, peak height ($mF_o - DF_c$ map) lesser than 2 sigma and also distance between water-water or water-macromolecule $< 1.8 \text{ \AA}$ or $> 6 \text{ \AA}$ were eliminated. We also scrutinized electron density map to evaluate presence of solvents other than water or ions which were present either in crystallization (like sulphate ion from AMS) or in purification (like glycerol in this case) condition. In each cycles 5% of the reflections were kept aside as a cross-validation to calculate R_{free} values.

2.1.6. Structural analysis

Average B-factors for each residue were calculated using B average in CCP4 [97]. PISA webserver [98] was used for the analysis of the interfaces and assemblies of the structure. Sequence alignment of *Vc*LMWPTP-2 with other LMWPTPs was done using Multalin [99] and PBIL [100]. Figures were prepared using Pymol (<http://www.pymol.org>). Mapping of surface electrostatic potential was performed using Chimera [101] and the area as well as volume of active site pocket was calculated using CASTp server [102].

2.2. Phylogenetic Analysis

Phylogenetic trees were generated using the highly conserved LMWPTP catalytic domain sequences as indicated [103]. The sequence of *Vc*LMWPTP-2 was taken as seed with which to search the NR (UniprotKB 60% identity maximum+PDB) at the

PRABI- GERLAND: RHONE –ALPES BIOINFORMATIC POLE GERLAND SITE at NPS@server. E threshold lesser or equal to $8e-15$ were used in database searches. Uncharacterized proteins and proteins other than PTP were excluded. The remaining 124 sequences were analyzed and aligned by ClustalW and MULTALIN which prepares the multiple sequence alignment with hierarchical clustering [99,104] with blosum62 matrix, gap weight 12, gap length weight 2 and with minimal distance between sequences (in PAM) 20. The phylogenic tree was generated with maximal number of clustals 12. MULTALIN follows pairwise alignment whereas for ClustalW calculation is performed by using distance matrix and finally neighbor joining (NJ) method is used for building of rooted as well as unrooted trees.

2.3. FPLC (Fast Protein Liquid Chromatography) analysis of *Vc*LMWPTP-2

Oligomeric states of *Vc*LMWPTP-2 were analyzed by size exclusion chromatography on ÄKTA Prime chromatographic system using Superdex 200 increase (GE-Healthcare) column (10×300mm; bed volume ~24ml). The column was pre-equilibrated with lysis buffer 50mM MOPS, pH7.7, 300mM NaCl, 3% Glycerol. The analysis was performed at 283K with flow rate of 0.4ml min^{-1} . The elution profile was determined by monitoring the absorbance at 280 nm. Albumin (66.5 kDa), Ovalbumin (45 kDa), Chymotrypsin (25 kDa) and Ribonuclease A (13.7 kDa) were used as molecular mass standards.

2.4. Dynamic Light Scattering

DLS measurements, to determine the oligomeric states in solution, were performed on a Malvern Nano Series DLS spectrometer, equipped with thermostat chamber and 633 nm laser. Fractions eluted from the Superdex200 gel filtration column were directly

used for this experiment. Before carrying out the DLS experiments all solutions were passed through 0.22 μm membrane filter (Millipore) and degassed. All measurements were conducted at least three times.

2.5. Molecular Dynamics Simulation

Molecular Dynamics (MD) simulation was carried out on a monomer (Residue 5–157, Chain A) of *VcLMWPTP-2* protein where the MOPS molecule and the sulfate ions were removed from the crystal structure. The simulation was performed for 1000 ns using Gromacs-5.0.4 simulation package [105] with CHARMM27 [106] all atom force field and the water molecules were modeled explicitly using the TIP3P model. The protonation state of ionizable groups was appropriately chosen for pH 7.0. Suitable number of counter ions were added at random locations to neutralize the net charge of the system. The simulation box was a dodecahedron with minimum protein-edge distance of 12 \AA and periodic boundary condition was applied on all the three (xyz) directions. After the energy minimization of the whole system using the steepest descent algorithm, the system was gradually heated to 300 K using NVT ensemble and later equilibrated using NPT ensemble. During the equilibrations, the protein backbone was restraint with a harmonic potential of force constant 1000 kJ/mol. The leap-frog integrator with a time-step of 2 fs was used. The Parrinello-Rahman algorithm [107] was employed to control the pressure at 1 bar with a coupling constant of 2 ps and the modified Berendsen (V-rescale) [108] thermostat was used to control the temperature of the system at 300 K with a time constant of 0.1 ps. The Particle Mesh Ewald (PME) [109] method was used to compute the electrostatic interactions with a real space cut-off distance of 12 \AA . The same cut-off value was used for calculations of the van der Waals interactions. After 5 ns of equilibration

using position restraints on the protein, the production MD simulation run was carried out for 1000 ns. The position and the velocity of all the atoms were recorded in the trajectory file at every 20 ps for analysis of the dynamics.

2.6. Enzyme Kinetics

2.6.1 Phosphatase activity

To determine dephosphorylating activity or phosphatase activity of *Vc*LMWPTP-2 wild type protein and all the mutants prepared were purified in 50mM Tris, pH 7.7, 300mM NaCl, 10% glycerol buffer. Phosphatase activity of *Vc*LMWPTP-2 WT and mutants were calculated at room temperature (298 K) using p-Nitrophenylphosphate (p-NPP) as the substrate as described previously [110]. p-NPP (1–40 mM) was treated with 10 μ M *Vc*LMWPTP-2 (WT or its mutants C12S, T13L, C17S, C12/17S) and the reaction was quenched by the addition of 0.1N NaOH after 10 min. The absorbance of the liberated product, p-nitrophenol, thus formed is measured at 405 nm. The amount of p-nitrophenol liberated was calculated from the standard curve prepared under identical conditions. For standard curve, stock solution of p-nitrophenol was diluted with 0.05N NaOH and the absorbance of the samples was measured at 405 nm. For baseline correction, the non-enzymatic hydrolysis of substrates was measured using the control sample that did not contain *Vc*LMWPTP-2. The hydrolysis rate was measured in triplicate for all substrates. Kinetic constants were calculated using initial velocity data, which were fitted to the Michaelis-Menten curves and the Lineweaver-burk plots.

2.6.2 Reactivation kinetics of hydrogen peroxide-inactivated *Vc*LMWPTP-2 using β -ME as reducing agent

After several trials a minimum of 50 μ M H₂O₂ was found to be sufficient to inactivate *Vc*LMWPTP-2 (26 μ M) completely. The H₂O₂ inactivation reaction was carried out in Tris buffer at pH7.0 at 298 K and allowed to proceed for 7 min at room temperature. Substrate p-NPP was added to the resulting H₂O₂ treated enzyme and the mixtures were allowed to incubate for 10 min before measuring the Phosphatase activity. For reactivation, 2-Mercaptoethanol (β -ME) was added (final concentration 30mM) to the inactivated mixture and phosphatase activity was assayed after 40 min of β -ME treatment.

Chapter-3

Results

PART-I

3.I.1. Phylogenetic analysis

The gram-negative bacteria, *Vibrio cholerae*, is the causative organism of the severe diarrheal disease cholera that continues to be a significant cause of death in developing countries claims approximately 95,000 lives out of 2.9 million annually [80, 111]. The genome of *Vibrio cholerae* O395 contains two LMWPTPs (*VcLMWTP-1* and *VcLMWTP-2*). Multiple sequence alignment, as shown in figure 11, shows these two LMWPTPs share only~37% sequence identity but their active site residues are conserved. Detailed analysis with more than hundred non-redundant (NR) LMWPTPs sequences indicate that *VcLMWTP-2* distinguishes itself as an idiosyncratic type of bacterial LMWPTP which has been performed by phylogenetic trees generation using the highly conserved LMWPTP catalytic domain sequences as indicated [103]. To analyze the phylogenetic relationships between PTP family proteins and *VcLMWTP-2*, we analyzed and aligned more than 100 sequences by ClustalW and MULTALIN and finally generated rooted as well as un-rooted tree by neighbor joining method, based on the full length *VcLMWTP-2* sequence as seed. Uncharacterized proteins and proteins other than PTP were excluded manually from this analysis. Remaining non- redundant 124 sequences were analyzed to generate phylogenetic tree where *VcLMWTP-2* emerges as a single taxon as shown in figure 12. To study this unique bacterial LMWPTP we started further analysis with this protein.

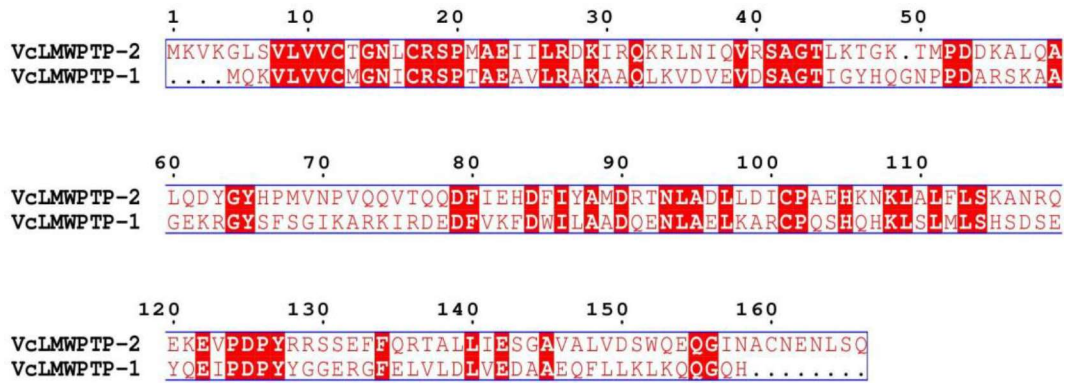


Figure 11. Multiple Sequence alignment between *VcLMWPTP*-1 and *VcLMWPTP*-2 shows catalytically important residues are conserved

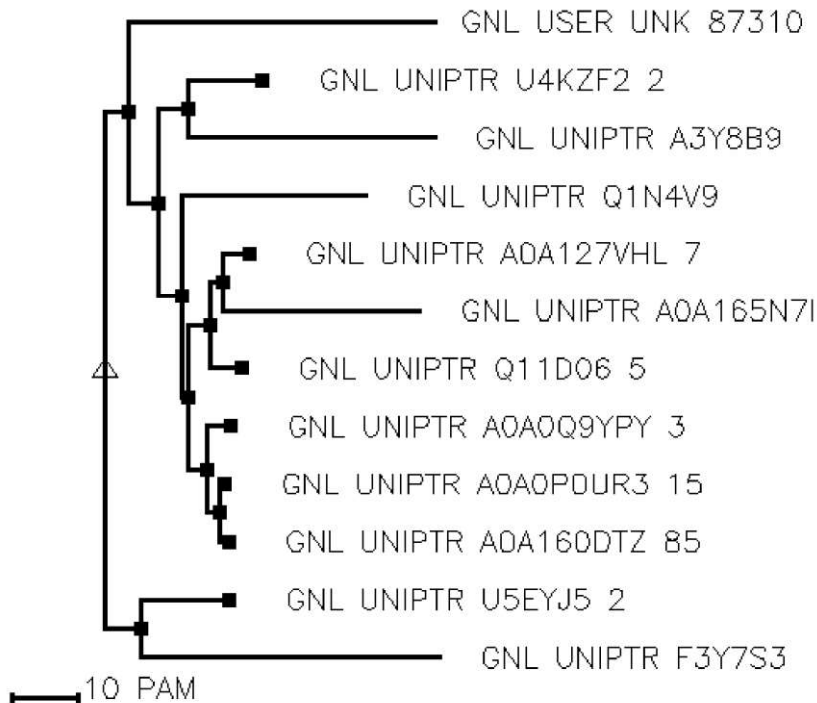


Figure 12. Comparative analysis between PTP family proteins and *VcLMWPTP*-2. The phylogenetic tree was generated using CLUSTLW and MULTALIN using *VcLMWPTP*-2 sequence as seed. Branch length has been displayed from the unit for the scale bar. Single taxon ‘GNL USER UNK 87310’ represents seed sequence (*VcLMWPTP*-2). Δ sign represents ‘tree root’; ■ sign represents ‘root of a subfamily which goes across domains of life’; after the ID the number of sequences of each subfamily has been indicated.

3.I.2. Cloning, Overexpression, Purification, Crystallization, Data Collection and Structure solution of *VcLMWPTP-2*

3.I.2.1. Cloning and Over-expression of *VcLMWPTP-2*

Cloning of *VcLMWPTP-2* was performed in pET28a (+) vector and the protein of interest was successfully overexpressed as 6×-histidine tagged recombinant protein. Different mutants were generated to study the role of conserved residues. These mutants were C12S, T13L, C17S and a double mutant C12/17S. Profiles of cloning and overexpressed protein of interest are shown in figure 13a and 13b. Condition of overexpression was kept constant as mentioned earlier.

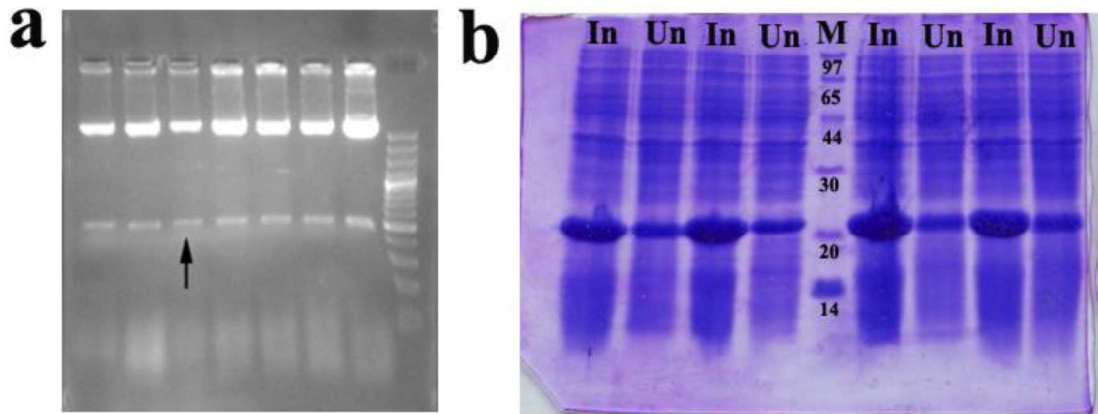


Figure 13. Clone check and over-expression of *VcLMWPTP-2* in (a) 1% agarose gel after digestion with restriction enzyme *NdeI* and *BamHI*. Lane 1-7: all samples contain insert into the plasmid, lane 8: DNA ladder (b) 15% SDS-PAGE profile, showing the overexpression of *VcLMWPTP-2*. Lanes 1, 3, 6 and 8: whole cell lysate of the induced cells (marked as ‘In’), lanes 2, 4, 7 and 9: whole cell lysate of the uninduced cells marked as ‘Un’), lane 5: Protein Marker (marked as ‘M’)

3.I.2.2. Purification of target protein by Ni-NTA affinity chromatography

Purification of target protein with 6×-Histidine tag was performed by Ni-NTA affinity chromatography. Elution of Histidine tagged bound proteins were performed by step gradient method. In this method we used lysis buffer with increasing concentration of imidazole which was used as an eluent to elute target protein from Ni-NTA column.

The elution pattern of wild type protein is shown in figure 14.

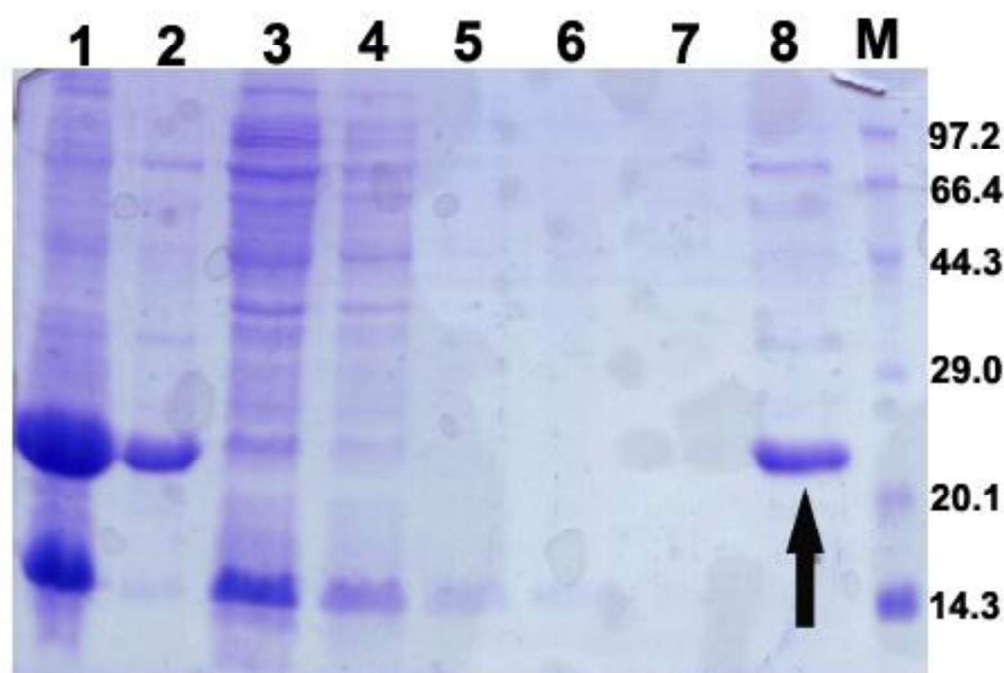


Figure 14. Ni-NTA purification profile represented by 15% SDS-PAGE. Lane 1 represents resuspended pellet after sonication and centrifugation; Lane 2 represents Ni-NTA bead after final elution; Lane 3 represents flow through after passing the cell supernatant through Ni-NTA beads equilibrated with lysis buffer. Lane 4 represents sample collected after washing the supernatant equilibrated column with lysis buffer. Lane 5 represents sample collected after washing the column with lysis buffer containing 5mM imidazole. Lane 6 represents sample collected after washing the column with lysis buffer containing 20mM imidazole. Lane 7 represents sample collected after washing the column with lysis buffer containing 30mM imidazole. Lane 8 represents sample collected after washing the column with lysis buffer containing 200mM imidazole (presence of purified protein marked with black arrow). Lane 9 represents protein marker.

Amount of soluble protein was increased after several trials, with addition and (or) increase of glycerol into lysis as well as elution buffer, with increase in sonication time i.e. number of pulses and moreover with an increase in initial amount of bacterial culture (~2 litre) and also by keeping low temperature after inducing the system with IPTG.

3.I.2.3. Thrombin cleavage and size exclusion chromatography

Removal of Histidine tag was performed before crystallization trial experiments as protein of interest was cloned with 6×-His tag followed by thrombin cleavage site. Amount of thrombin added, incubation time and temperature was optimized by several trials. Finally 2μL thrombin (1.6 units/μl, Novagen) was added into ~1ml of protein solution which contained 4mM of Ca²⁺ ion at 4°C for 45 hours and complete removal of tag was checked by SDS-PAGE analysis.

Upon complete removal of Histidine tag, size exclusion chromatography was performed to further purify the protein from His-tag, thrombin and other impurities and elution of protein was monitored from absorbance at 280nm. The elution profile of protein from S-200, 10×300mm column has been shown in figure 15a. The fractions containing protein was collected and concentrated upto >5mg/ml. Before crystallization, concentration was monitored from absorbance at 280nm as well as from SDS-PAGE. Protein concentration checked by SDS-PAGE analysis before crystallization has been shown in figure 15b.

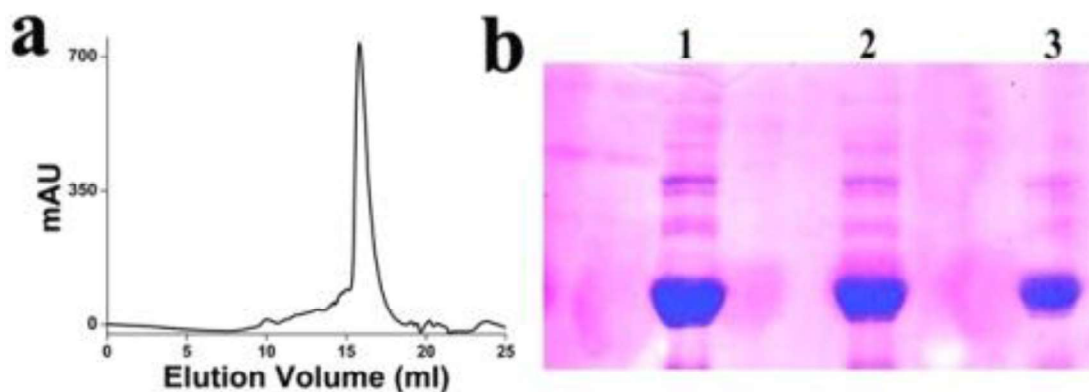


Figure 15. Elution profile by SEC and concentration checked for crystallization by SDS-PAGE. (a) The elution profile of SEC is shown where peak represents existence of target protein. Elution fractions of peak were concentrated for further use. (b) Upon concentrating elution fractions of peak concentration checked by SDS PAGE where Lane1 represents loading of 6 μ L protein, Lane2 represents loading of 4 μ L protein and Lane3 represents loading of 2 μ L protein.

3.I.2.4. Crystallization

*V_c*LMWPTP-2 crystallization trials were performed with various grid screens and crystal screens among them crystals were appeared at C4 of ammonium sulphate (AMS) grid screen from Hampton which is composed of 0.1M HEPES, pH7.0, 2.4M ammonium sulphate. Appearance of crystals took place with varying concentration of glycerol in the precipitant. Crystals of diffraction quality were appeared in presence of 5%glycerol with C4 of AMS grid screen as precipitant and upon mixing 2 μ L of protein with 2 μ L of precipitant at 277K temperature. Shape of the crystals was plate-shaped although all the crystals were very thin in texture. Looping out the crystals from the crystal drop with nylon loop was very difficult as crystals were very prone towards skin formation and they were coming along with skin formed into the drop.

3.1.2.5. Data Collection

Diffraction data were collected upto 2.6Å resolution at MARCCD 225 detector of PX BL-21 [91] beam line of Indus-2 synchrotron at RRCAT, INDORE. 180 frames were recorded with incident X-ray beam wavelength of 0.97Å. Typical image of a diffraction pattern is given in figure 16.

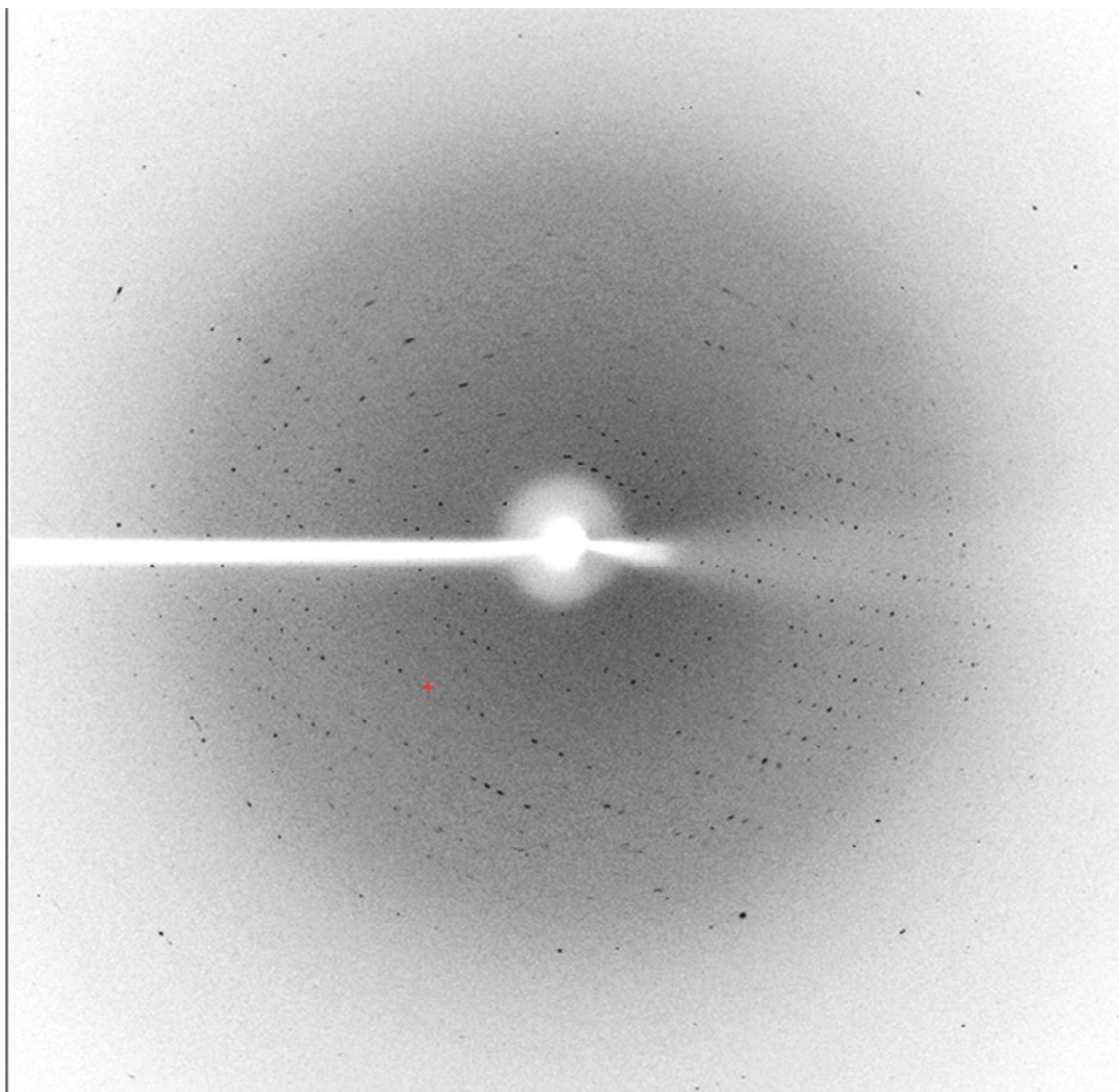


Figure 16. Diffraction pattern of *Vc*LMWPTP-2 crystals upon X-ray exposure

3.I.2.6. Structure solution of *VcLMWPTP-2*

Data were processed and scaled using iMOSFLM [92] and calculations of Matthews coefficient indicated the presence of two molecules of *VcLMWPTP-2* in the asymmetric unit with a solvent content 66%. Based upon sequence alignment between *VcLMWPTP-2* and *VcLMWPTP-1*, a model was generated from the coordinates of *VcLMWPTP-1* (PDB: 4LRQ) using the program Chainsaw of CCP4 package [93]. This model was used to generate the initial solution by molecular replacement using Phaser [94], which yielded a solution with final TFZ=10.5 and LLG=389. Just after molecular replacement refinement we started with refining the co-ordinates. During initial cycles rigid body refinement and real space coordinate refinement along with atomic displacement parameters (isotropic B-factors) were performed. Model building with Coot [96] was gone parallel with data refinement by Phenix refine [95]. These leads to improvement of data quality along with model improvement which allows gradual inclusion of the ligands, solvents, individual B-factors and TLS refinement produced a final R_{cryst} of 25.1% (R_{free} =28.8%). The coordinates have been submitted with PDB code: 5Z3M. Details of the data-collection and processing statistics are given in Table 7.

Table 7: Data collection statistics and refinement statistics

Protein name		<i>Vc</i> LMWPTP-2
Space group		<i>C</i> 2
Cell dimension	a, b, c (Å)	125.91,51.80,83.80
	α , β , γ (°)	90.0,103.60,90.0
Molecule(s)/ASU		2
Mathews coefficient, V _m (Å ³ Da ⁻¹)		3.7
Resolution (Å)*		19.4-2.6 (2.69-2.60)
Avg. Mosaicity (degrees)		0.3
Rmerge		0.136 (0.535)
I / σ (I)		8 (16.5)
Completeness (%)		98.9 (99.8)
Redundancy		3.7 (3.8)
CC1/2		99.1(88.4)
Refinement		
Resolution (Å)		19.4-2.6
No. of reflections/No. of reflections		31285/16209
Rcryst / Rfree		25.1/28.8
No. Atoms		2927
Protein		2433
Ligand		239
Water		255
Average B, all atoms (Å ²)		46.0
R.m.s.d.	Bond length (Å)	0.005
	Bond angles (°)	1.06
Ramachandran statistics (%)	Most favoured	90.03
	Additionally allowed	9.63
	Disallowed	0.33
PDB ID		5Z3M

* Values within parenthesis indicate highest resolution bin

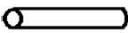
3.I.3. Structure based sequence alignment

Genome of *Vibrio cholera* O395 contains two LMWPTPs; *Vc*LMWPTP-1 & *Vc*LMWPTP-2. From Dali Server [112] we have found organisms which shares identity ~30% and Z-score >20 as shown in Table 8.

Table 8: Results found from Dali Server using *Vc*LMWPTP-2 (*Vc*2) sequence as input

Organism	PDB	Z	rmsd	%id
<i>V. cholerae</i> 1(<i>Vc</i> 1)	4LRQ	26	1.1	37
<i>E. histolitica</i> (e.his)	3IDO	25	1.5	31
<i>S. aureus</i> (s.aur)	3ROF	22	1.9	30
<i>B. subtilis</i> (b.sub)	4ETM	24	1.5	33
<i>T. thermophiles</i> (t.t)	2CWD	24	1.3	34
<i>H. sapiens</i> (human)	1XWW	23.7	1.5	28
<i>B. Taurus</i> (bovine)	1Z12	23.8	1.5	29
<i>E. coli</i> (e.coli)	2WJA	21.1	1.7	29

Sequence alignment of *Vc*LMWPTP-2 with *Vc*LMWPTP-1 and other organisms, show conserved patches at active site and near active site as shown in figure 17 where

➡ sign represents β -sheet and  sign represents α -helix.

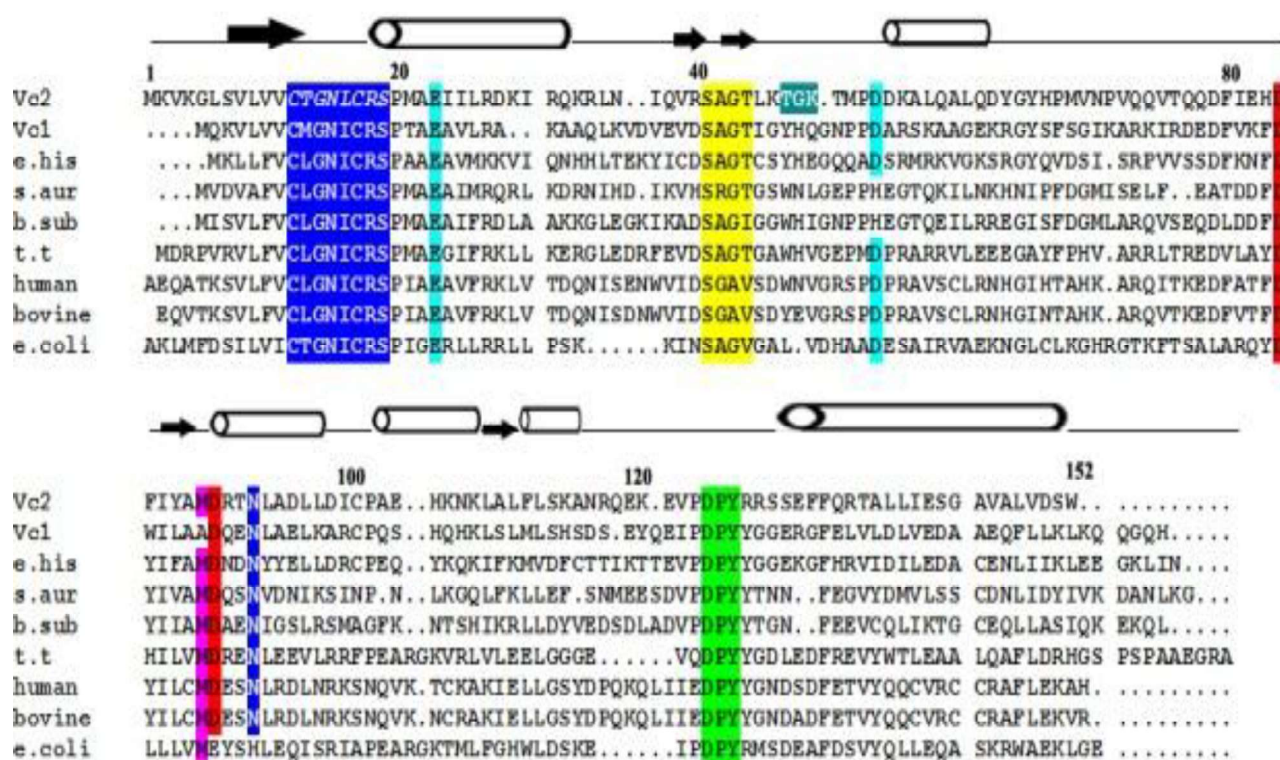


Figure 17. Structure based sequence allingment of *VcLMWPTP-2* with *VcLMWPTP-1* and other organisms where *Vc2* represents *VcLMWPTP-2*, *Vc1* represents *VcLMWPTP-1*, **e.his** represents *Eentamoeba histoltica*, **s.aur** represents *Staphylococcus aureus*, **b.sub** represents *Bacillius subtilis*, **t.t** represents *Thermos thermophilus*, **human** represents *Homo sapiens*, **bovine** represents *B. Tauruis*, **e.coli** represents *Escherichia coli*.

HIGHLIGHT represents P-loop residues

HIGHLIGHT represents conserved Glutamic acid (E) and Aspartic acid (D) residues

HIGHLIGHT represents residues of SAGT-loop

HIGHLIGHT represents conserved Aspartic acid (D) residues

HIGHLIGHT represents conserved hydrophobic Methionine (M) residue

HIGHLIGHT represents conserved Asparagine (N) residue

HIGHLIGHT represents residues of DPY-loop

HIGHLIGHT represents residues responsible for formation of active site cavity with smallest area and volume.

3.I.4. Structural Analysis of *Vc*LMWPTP-2

3.I.4.1 Overall structure of *Vc*LMWPTP-2 monomer

Crystal structure of *Vc*LMWPTP-2 has been solved in space group C2 at 2.6 Å. Each asymmetric unit of the crystal contains two molecules of *Vc*LMWPTP-2 together forming a dimeric unit. From the electron density map 157 residues out of 165 residues of each molecule could be located whereas last 9 residues have non interpretable density. Each *Vc*LMWPTP-2 molecule binds a 3-morpholinopropane-1-sulfonic acid (MOPS: which was present in buffer) at its active site and one sulfate ion near the DPY loop [Fig. 18].

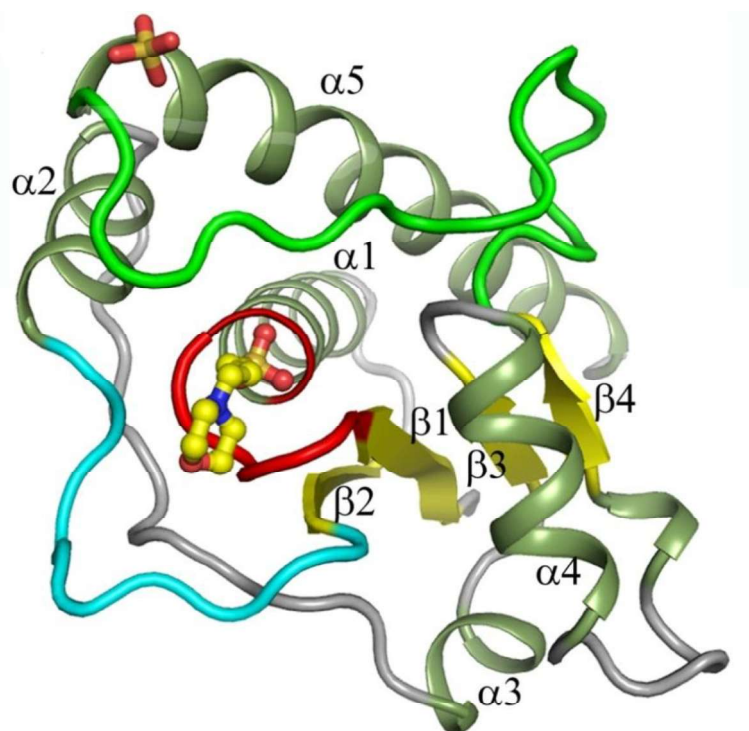


Figure 18. Structure of *Vc*LMWPTP-2 monomer. Structure has been shown in cartoon representation with MOPS at the active site (shown in ball and sticks). α -helices and β -sheets are represented in olive and yellow color respectively. The catalytic P-loop is colored in red while β 2- α 2 loop and DPY loop are shown in cyan and green respectively.

The three dimensional structure of each *Vc*LMWPTP-2 molecule is composed of four parallel β -strands [β 1 (Ser7 to Val11), β 2 (Gln38 to Glu43), β 3 (Phe85 to Ala88), and β 4 (Leu109 to Leu111)] sandwiched by five α -helices [α 1(Arg18 to Lys33), α 2(Asp54 to Gln62), α 3(Gln77 to Glu82), α 4(Arg91 to Leu98),and α 5(Ser131 to Gln155)], which are arranged to form a Rossmann like fold typical of LMWPTPs [37,63,113,114] [Fig. 18]. The catalytic cysteine residue (Cys12) of *Vc*LMWPTP-2 is part of the highly conserved motif C¹²TGNLCR¹⁸S which is located at the tip of β 1- α 1 loop. P-loop is surrounded by β 2- α 2 loop and β 4- α 5 loop. Amino acid sequence of these loops and their disposition define the depth and shape of the catalytic pocket. The β 4- α 5 loop (also called the DPY loop or simply ‘D’ loop) harbors an aspartic acid residue (Asp125 in *Vc*LMWPTP-2) that acts as a general acid/base necessary for the dephosphorylating activity [Fig. 18]. Interestingly, in *Vc*LMWPTP-2, the fourth residue of the motif is an arginine not a tyrosine, as in the case of *Escherichia coli* LMWPTP.

3.I.4.2 Structural superposition

Two *Vc*LMWPTP-2 molecules superpose with each other with an RMSD value of 0.45 Å whereas the same upon superposition with *Vc*LMWPTP-1 is 0.67 Å [Fig. 19]. As expected, *Vc*LMWPTP-1 and *Vc*LMWPTP-2 differ mainly at their surface exposed loops although their P-loop superposes very well.

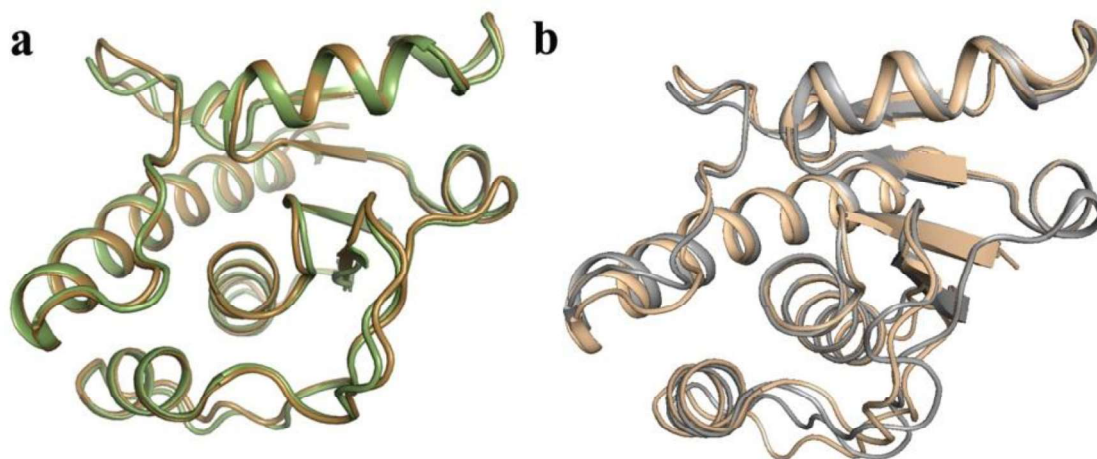


Figure 19. Structural superposition of *VcLMWPTP-2* where (a) Superposition between two molecules of *VcLMWPTP-2* and (b) Superposition between *VcLMWPTP-2* (wheat) and *VcLMWPTP-1* (Gray).

3.I.4.3 MOPS binds like a substrate mimetic at the active site

An electron density map around the active site of *VcLMWPTP-2* molecule is shown in figure 20. From the electron density map it is evident that each *VcLMWPTP-2* tightly binds a MOPS molecule, a phosphotyrosine mimetic, at the active site.

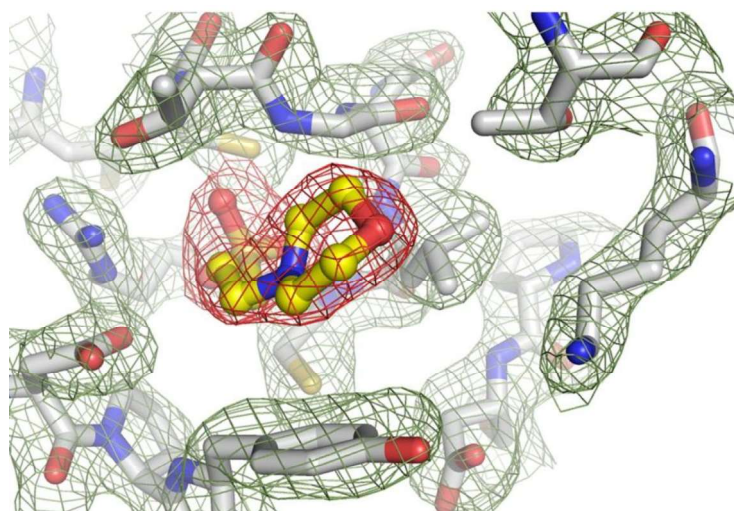


Figure 20. $2F_{\text{obs}} - F_{\text{calc}}$ electron density map around the active site of *VcLMWPTP-2* (green mesh; contoured at 1.5σ , red mesh around MOPS contoured at 1σ)

It forms several hydrogen bonds with the main chain amide nitrogens of the P-loop residues [Fig. 21] and the side chain nitrogen atoms of conserved Arg18. The MOPS molecule closely resembles the substrate phosphotyrosine as seen in PTP1B (PDB: 1G1F) [115] and is also analogous to the histidine and a phosphate, together mimicking a phosphotyrosine (PDB: 1YWF) [116]. The sulphonate group of MOPS mimics the phosphate group of the substrate phosphotyrosine as seen in PTP1B (PDB: 1G1F) [115].

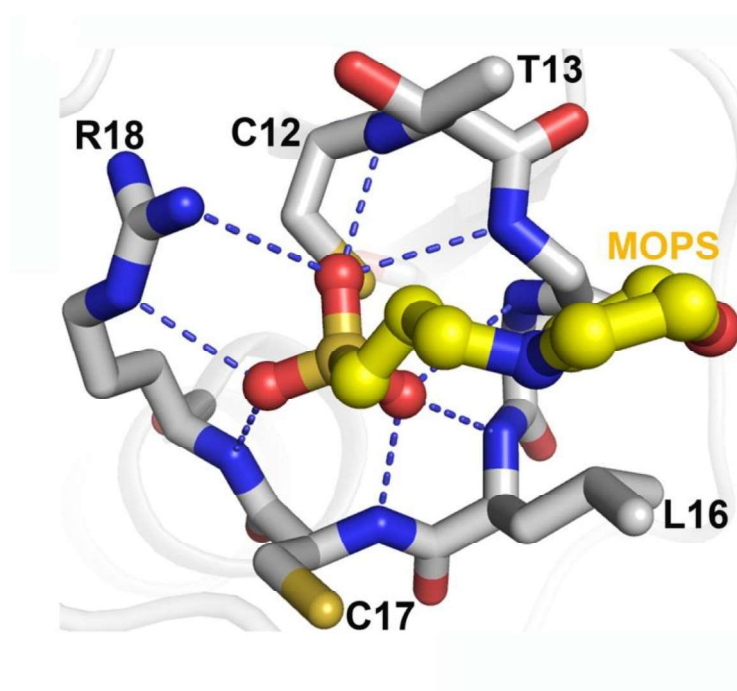


Figure 21. Extensive hydrogen bonds made by the sulfonyl group of the MOPS molecule with the amide environment of the P-loop and the side chain of R18.

3.I.4.4 Architecture of the active site

The piperazine ring of MOPS molecule is surrounded with several polar and aromatic residues (Thr44, Thr47, Thr50, Asp90, Asn93, Asp125, Tyr127, and Arg128). These residues define the architecture of the wall of the phospho-tyrosine binding pocket of *VcLMWPTP-2* with an opening at one side [Fig. 22]. It is to be noted that only one

large hydrophobic residue (Tyr127) is present around the active site in *Vc*LMWPTP-2 while in other LMWPTPs there are three or four Trp/Tyr/His residues. Lining of these residues around the active site define the architecture of the deep phosphotyrosine binding pocket.

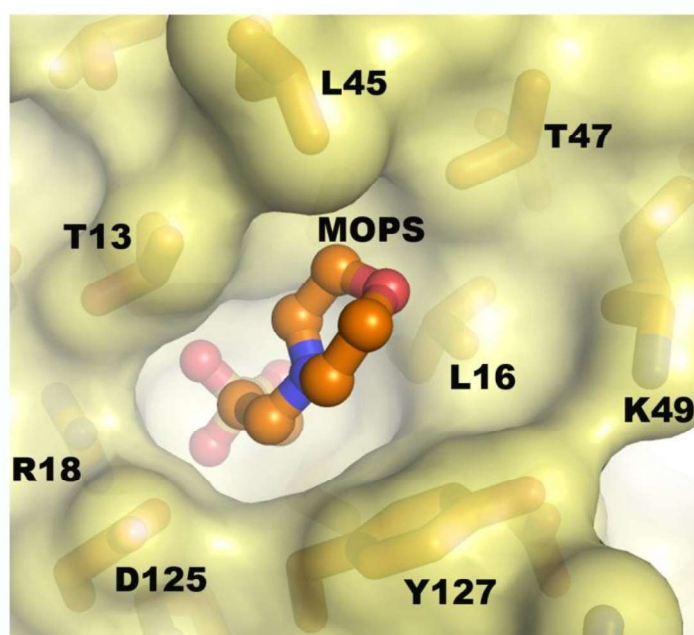


Figure 22. Surface representation of MOPS (ball and stick) binding pocket; residues lining the active site pocket are shown in surface and sticks.

3.I.4.5 Comparison with ligand bound/unbound structures

The orientation of S γ of Cys12 and its proximity (~ 3.8 Å) with the S1-atom of MOPS (that corresponds to the P $^-$ atom of Phosphotyrosine) indicates that it is in a position suitable for a S $_N$ 2 attack. Asp125 of DPY-loop is seen very close to the P-loop and its side chain oxygen atoms are within 3.7 Å from the C1 atom of the MOPS molecule (corresponding to the phenolate O $^-$ atom of phosphotyrosine). At this distance, Asp125 serves as an acid catalyst for the leaving alcoholic group. Therefore, MOPS bound state with the disposition of catalytic C12 and D125 resembles a substrate bound ‘closed structure’. This closure at the P-loop is also evident from the structural

alignment with an ‘open structure’ of apo-MPtpA (PDB: 2LUO) [117] to the ligand bound structures (PDB: 1U2P, 1U2Q) where the position of DPY-loop in the ligand-bound structure is shown to be closer to the P-loop [37] [Fig. 23].

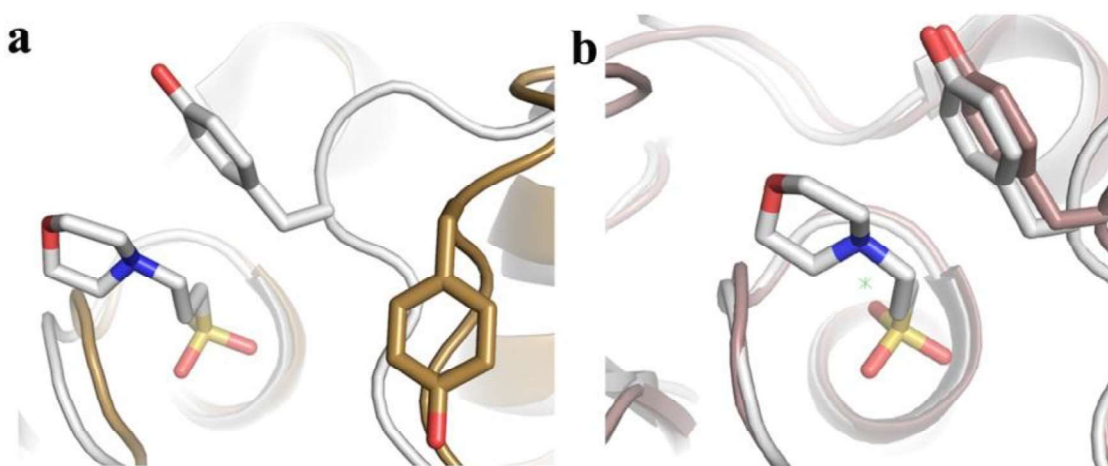


Figure 23. Comparison with ‘open’/‘closed’ structure. Upon structural alignment of *Vc*LMWPTP-2 (Gray) with (a) apo-MPtpA (PDB: 2LUO) (sand) and (b) Ligand bound MPtpA (PDB: 1U2Q) (Chocolate) shows, *Vc*LMWPTP-2 resembles a substrate bound ‘closed structure’.

3.1.5 Dimerization of *Vc*LMWPTP-2

Crystal structure of *Vc*LMWPTP-2 indicates a dimeric form where as the kinetic data exhibit strong phosphatase activity. Analysis of the mode of dimerization, seen in the crystal structure, provides an explanation of its catalytic activity. Figure 24a–c depicts the cartoon representation of the dimer of *Vc*LMWPTP-2, *Vc*LMWPTP-1 and BPTP respectively. In case of inactive BPTP dimer, the P-loop of one monomer is totally occupied by the residues of the YY loop of other [41]. While in *Vc*LMWPTP-1 P-loop is not occluded by dimerisation but remain close the dimeric surface rendering the dimer catalytically active. Sulphate ions near the interfacial region probably playing a

stitching role in dimer formation [113]. The mode of dimerization in *Vc*LMWPTP-2 is quite different from *Vc*LMWPTP-1 and BPTP. Here the active site is fully available to the substrate and remain far from the dimeric interface as reflected from its high phosphatase activity. A comparison of the buried surface area (BSA) indicates that *Vc*LMWPTP-2 has much higher BSA (3100 \AA^2) than that of *Vc*LMWPTP-1 (average BSA 2865 \AA^2) or BPTP (BSA 1589 \AA^2). At the interfacial region, eighteen residues from each monomer (Arg27, Gln38, Arg40, Lys49, Thr50, Met51, Pro52, Asp54, Leu57, Gln58, Gln61, His66, Pro67, Met68, Val69, Asn70, Pro71, Gln73) are engaged to stabilize the dimer through hydrophobic and H-bonding interactions. The set of residues involved in dimer formation in *Vc*LMWPTP-2 is quite different from that of *Vc*LMWPTP-1. The dimeric form is also well supported in solution. *Vc*LMWPTP-2 elutes exclusively as a single peak at ~ 15.8 ml in SEC irrespective of the salt concentration and the elution volume indicate a higher oligomeric form. Molecular mass calculation, based on a standard curve, confirms that this oligomeric form is a dimer [Fig. 24d-e]. To further confirm the oligomeric state, protein eluted from the FPLC column was used directly for DLS experiments. The intensity weighted fit shows presence of some aggregates which is probably due to the measurement with higher concentration of protein although volume weighted fit shows that maximum population exists as dimer in solution [Fig. 24f].

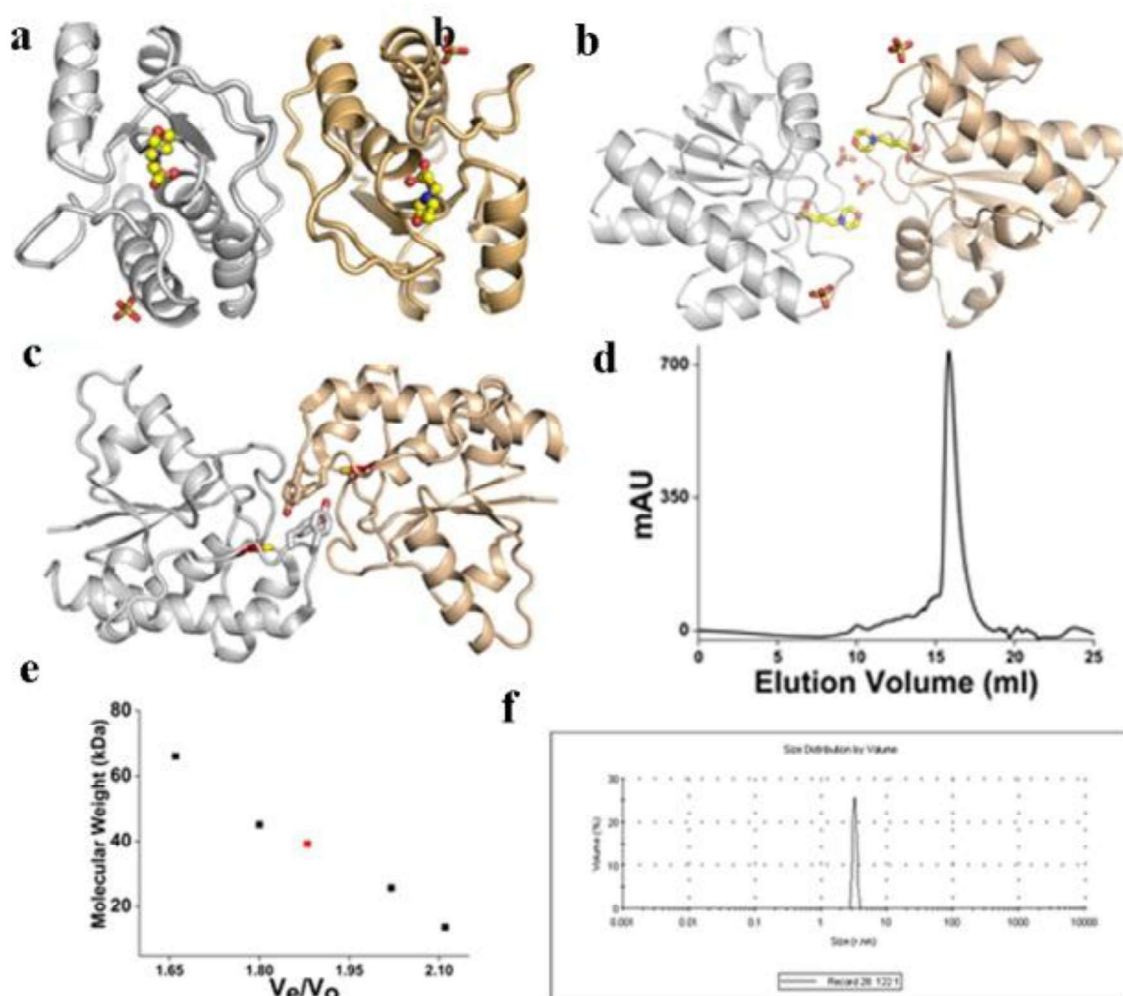


Figure 24. Dimerization of *VcLMWPTP-2*. Cartoon representation of the dimeric (shown in gray and wheat) form of (a) Novel catalytically active *VcLMWPTP-2* with bound MOPS in the active site (ball and stick) and sulfate ions (stick), (b) catalytically active *VcLMWPTP-1* with the bound MOPS molecule (yellow) and sulfate ions (in sticks) and (c) catalytically inactive BPTP where DPY loop of one molecule comes in close proximity to the active site of the other molecule resulting in the inactivation. (d) Elution profile of *VcLMWPTP-2* plotted as mAU (absorbance at 280 nm) vs. elution volume (ml) depicts the existence of exclusive dimeric form of *VcLMWPTP-2* in solution. (e) Standard curve [Molecular mass (in kDa) vs. V_e/V_o (elution volume/void volume)] is drawn based on the elution volumes of protein mixtures of known molecular mass (Albumin 66.5 kDa; Ovalbumin 45 kDa; Chymotrypsin 25 kDa and RibonucleaseA 13.7 kDa). Respective location of the protein (*VcLMWPTP-2*) in standard curve is shown in red. (f) *VcLMWPTP-2* eluted by gel filtration and the fractions analyzed by dynamic light scattering (DLS). Volume weighted fit shows existence of dimer in solution.

3.I.6 Multiple sequence alignments

Multiple sequence alignment, using a non-redundant set of 124 LMWPTP sequences indicates several conserved patches [Fig. 25a]. Among them, the P-loop and the DPY loop are highly conserved having a consensus sequence of $^{12}\text{C}(\text{X})_4\text{CRS}^{19}$ (motif-1) and $\text{DPY}(\text{Y}/\text{F})$ (motif-4) [Fig. 25a]. This is quite expected since residues in these loops are directly involved in the phosphatase activity. P-loop harbors the catalytic residue Cys12 whereas the Asp residue of the DPY loop serves as general acid-base during catalysis. The microenvironment of Cys12 viz. H-bonding with Ser19 and the helix dipole of $\alpha 1$ make Cys12 a better nucleophile. C8S (corresponding to C12 in *Vc*LMWPTP-2) mutation in *Vc*LMWPTP-1 results in complete loss of activity implying that it is key catalytically residue [113]. Cys17 of the P-loop protects the catalytic Cys12 whereas Arg18 has a role in substrate binding. Besides these two highly conserved motifs, two less conserved motifs are evident from the multiple sequence alignment which are located around the P-loop like $\text{SAGT}(\text{S}^{41}\text{X}(\text{A}/\text{G}/\text{R})\text{X}(\text{G}/\text{A})\text{X}(\text{T}/\text{V}))$ (motif-2) and $\text{M}^{89}\text{D}(\text{X})_2\text{N}$ motif (motif-3) [Fig. 25]. Results of multiple sequence alignment when combined with the 3D superposition data of LMWPTP structures, determined from various organisms, illustrate the role of these four conserved motifs [Fig. 25b]. For example there is a network of hydrogen bonds involving conserved Asn15, Ser19, Cys12 and Ser41. This network not only provides proper conformation and required stability of the P-loop but along with the helix dipole of $\alpha 1$ also helps lowering down the pKa of Cys12. This has been shown in case of S19A mutant of BPTP which is reported to destabilize the thiolate [118]. Among the residues of $\text{M}^{89}\text{D}(\text{X})_2\text{N}$ of motif-3 D90 make a salt bridge with R18 which dictates the orientation of R18 for optimum substrate binding. For its strategic location

and negative charge it is also capable to neutralize the helix dipole of $\alpha 4$ helix. D→A mutation at this position in BPTP has been reported to hamper the affinity of binding the substrate [119]. The side chain of Met89 is projected towards a hydrophobic pocket formed by various residues. It consistently takes a typical conformation in all LMWPTPs and runs in parallel with the hydrophobic part of Arg18 keeping Arg18 in a particular orientation for substrate binding. The conserved Asn93 of this motif (motif-3) make H-bonds with the side chain of Arg18. In some LMWPTPs residues in 13th position is Thr and in those cases Asn93 could potentially form H-bond with them. In our case mutation of this Thr→Leu decreases the K_M value to 1.66 ± 0.1 mM. Structural superposition also indicates that $\beta 2$ - $\alpha 2$ loop, positioned semicircularly around the P-loop, exhibits highest structural variability. Sequence alignment also indicates a high degree of heterogeneity in this region. Strikingly, their P-loop superposes very well [Fig. 25]. Among the LMWPTP structures the $\beta 2$ - $\alpha 2$ loop of VcLMWPTP-2, *S.aureus* (3ROF) and *E.coli* (2WJA) differ significantly from others. Except for these three cases, it has been observed that $\beta 2$ - $\alpha 2$ loop contain two large hydrophobic residues W/Y and H. For 3ROF (*S.aureus*) conserved His has been replaced with Asn and for 2WJA (*E.coli*) with Leu residue. Moreover, orientation of their loop $\beta 2$ - $\alpha 2$ loop is quite different either due to insertion/deletion or sequence differences. Along with $\beta 2$ - $\alpha 2$ loop, the wall of the deep phosphotyrosine binding pocket is lined up by the two tyrosines of the DPY loop.

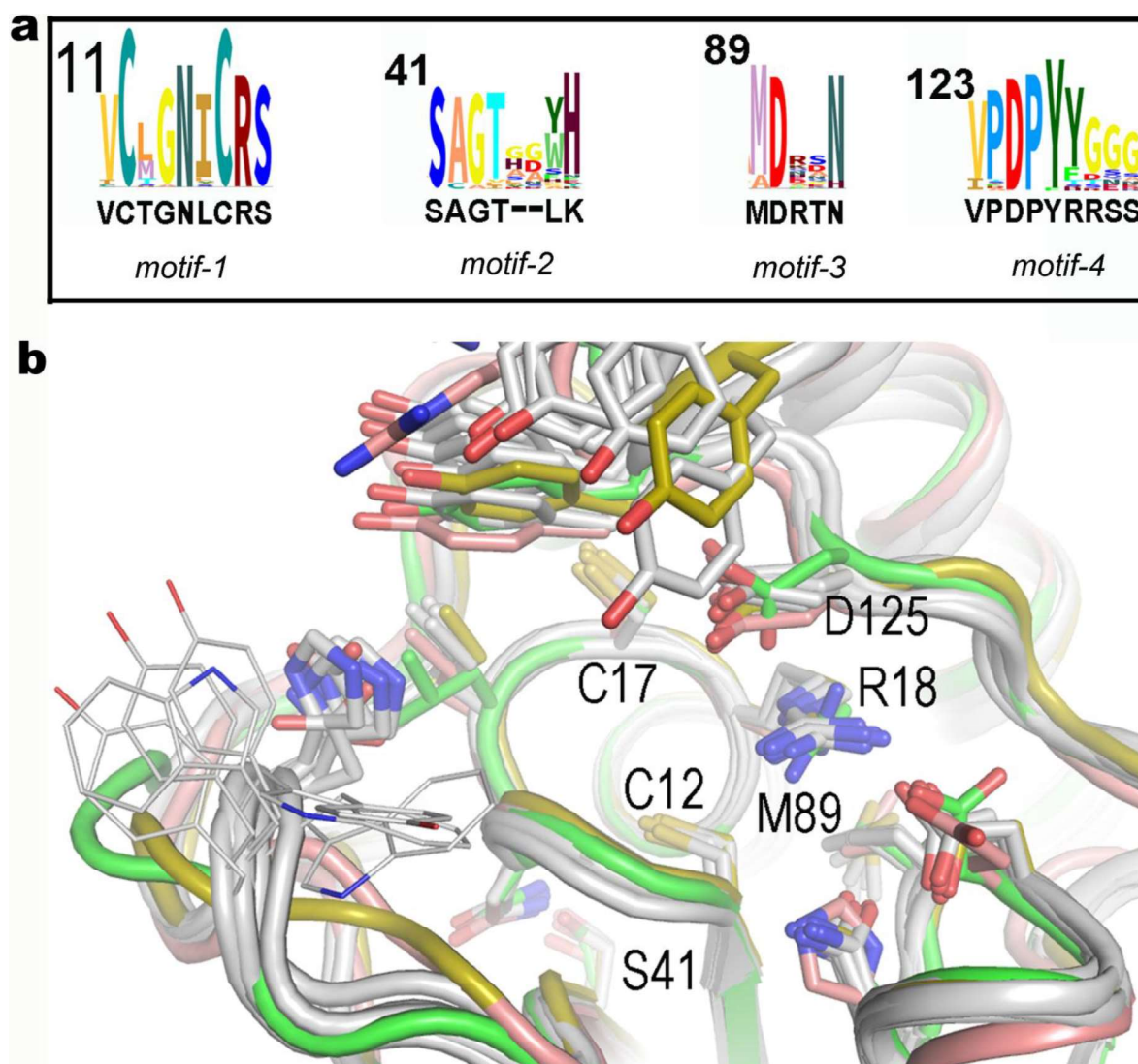


Figure 25. Sequence and structural alignment of LMWPTP. (a) Four conserved motifs viz., C(X)₅RS (motif-1), SAGT (motif-2), MD(X)₂N (motif-3), and DPYY (motif-4) identified through multiple sequence alignment of LMWPTP sequences where *Vc*LMWPTP-2 sequence has been shown in black color. Relative weightings of aligned sequence performed using Skylign [120]. (b) Structural alignment of *Vc*LMWPTP-2 (green; PDB ID: 5Z3M) with *Staphylococcus aureus* (olive; PDB ID: 3ROF), *Escherichia coli* (cherry; PDB ID: 2WJA) along with Human (PDB ID: 1XWW), *Thermos thermophilus* (PDB ID: 2CWD), *Vc*LMWPTP-1 (PDB ID: 4LRQ), *Entamoeba histoltica* (PDB ID: 3IDO), and *Bacillus subtilis* (PDB ID: 4ETM) shown in gray. Region near the P-loop is zoomed to show structural flexibility of DPY loop and β 2- α 2 loop.

The volume of the active site pocket was calculated for all these LMWPTP structures by CASTp server [102] and tabulated in Table 9. Table 9 indicates the volume of active site of *Vc*LMWPTP-2 is 140 Å³, whereas the volume of most other cases is ~250 Å³. No aromatic residue in the β2-α2 loop and presence of a single Tyr residue instead of two in the DPY loop for *Vc*LMWPTP-2 making the size of the cavity small. Small volume for 4ETM (*B.subtilis*) is probably due to different orientation of β2-α2 loop.

Table 9. Area and Volume of active site of LMWPTPs

PDB	Area(Å²)[‡]	Volume(Å³)[‡]
<i>Vc</i> LMWPTP-2	247.8	139.4
5jnv	280	196
4LRQ	545.4	284.9
4ETM	146.4	82.2
3ROF	334	232
3JVI	727	245
3IDO	421	202
2WJA	501	178
2CWD	784	270
1XWW	367	187

[‡]Calculation performed by CASTp server with probe radius 0.7 Å.

3.I.7 Kinetic studies of *Vc*LMWPTP-2

3.I.7.1. Phosphatase activity of *Vc*LMWPTP-2

Phosphatase activity of *Vc*LMWPTP-2 was determined by in vitro biochemical assays using p-NPP as a substrate. The K_M value determined for the WT protein is 1.41 ± 0.1 mM [Fig. 26a-b] which is comparable to that of other reported LMWPTPs like PtpA and PtpB [121] and *Vc*LMWPTP-1 [113]. Complete loss of phosphatase activity has been observed in presence of vanadate ions indicating that it acts as an irreversible inhibitor, which is in agreement with the data reported by Zhang et al. [63].

Role of active site residues has been studied by mutating those residues. Mutation of catalytic Cys12 to Ser reduces the phosphatase activity to a great extent but it is not abrogated. The K_M value increases to 8.06 ± 0.3 mM compared to that of WT protein (1.41 ± 0.1 mM). This is in contrast with other LMWPTPs along with *Vc*LMWPTP-1 [113, 121]. Interestingly, C17S mutant shows ~30% decreases of activity (K_M 2.49 ± 0.1 mM) with respect to WT. Strikingly, a double mutant (C12S/C17S), where both the active site cysteine was mutated to serine, exhibits complete loss of activity. Multiple sequence alignment [Fig. 17] shows the presence of threonine (13th position) instead of leucine in the P-loop unlike most of the LMWPTPs. Mutating threonine to leucine (T13L) reduces catalytic activity (K_M $\sim 1.66 \pm 0.1$ mM) [Fig. 26a-b]. All the kinetic parameters of wild-type *Vc*LMWPTP-2 and its mutants have been tabulated in Table 10.

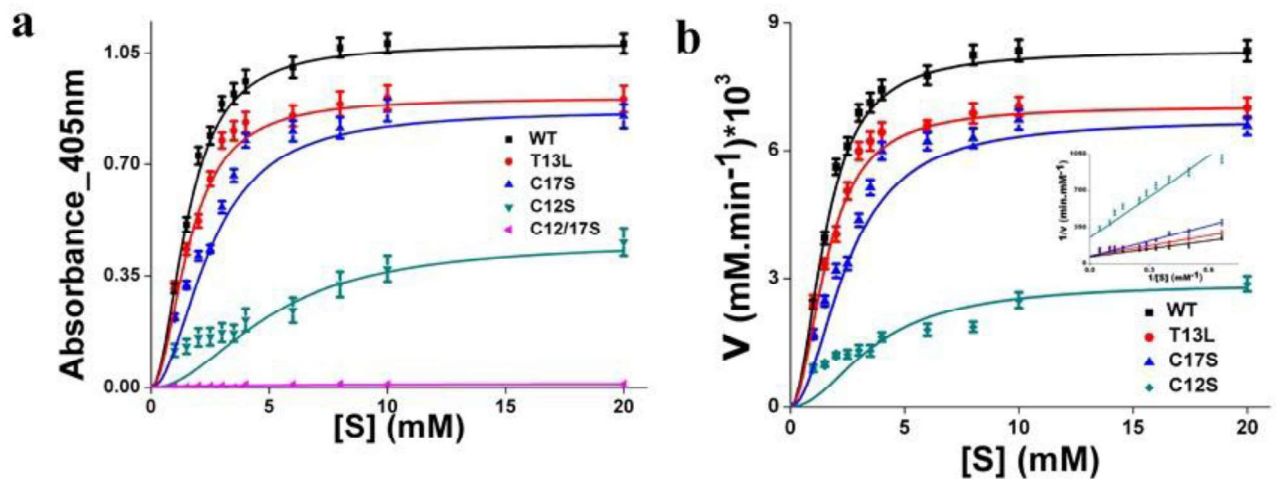


Figure 26. Enzyme kinetics of *VcLMWPTP-2* using p-NPP as a substrate. (a-b) Michaelis-Menten curves with Lineweaver-burk plots (inset) of WT *VcLMWPTP-2* and its various mutants (T13L, C17S, C12S).

Table 10. Kinetic parameters of wild-type *VcLMWPTP-2* and its mutants

Enzyme	K_M (mM)	V_{max} (mM.min ⁻¹)	V_{max}/K_M (min ⁻¹)
<i>VcLMWPTP-2</i> WT	1.41 ± 0.1	$(8.5 \pm 0.4) * 10^{-3}$	$(6.03 \pm 0.3) * 10^{-3}$
<i>VcLMWPTP-2</i> T13L	1.66 ± 0.1	$(7.1 \pm 0.3) * 10^{-3}$	$(4.28 \pm 0.1) * 10^{-3}$
<i>Vc-LMWPTP-2</i> C12S	8.06 ± 0.6	$(4.9 \pm 0.5) * 10^{-3}$	$(0.61 \pm 0.04) * 10^{-3}$
<i>Vc-LMWPTP-2</i> C17S	2.49 ± 0.1	$(6.5 \pm 0.3) * 10^{-3}$	$(2.61 \pm 0.1) * 10^{-3}$
<i>Vc-LMWPTP-2</i> C12/17S	ND	ND	ND

3.I.7.2. Redox regulation of *Vc*LMWPTP-2

To verify the role of two cysteine (Cys12 and adjacent Cys17) residing on P-loop of *Vc*LMWPTP-2, catalytic inactivation by oxidizing agent followed by reactivation with reducing agent has been performed. In presence of H_2O_2 , the catalytic activity of wild type *Vc*LMWPTP-2 diminished to basal level while addition of excess β -ME regains its activity almost completely upto $\sim 95\%$ [Fig. 27a], indicating that Cys17 protects the catalytic Cys12 under oxidative stress as seen in other LMWPTPs [40, 122, 123]. Recovery after addition of β -ME for H_2O_2 mediated inactive protein [Fig. 27b] is $\sim 10\%$ for C12S mutant, and for C17S it is $<2\%$ which further proves the participation of Cys17 to fortify catalytic cysteine. The rate of reactivation is however much slower than the rate of inactivation as excess β -ME and longer time of incubation is required to regain the activity.

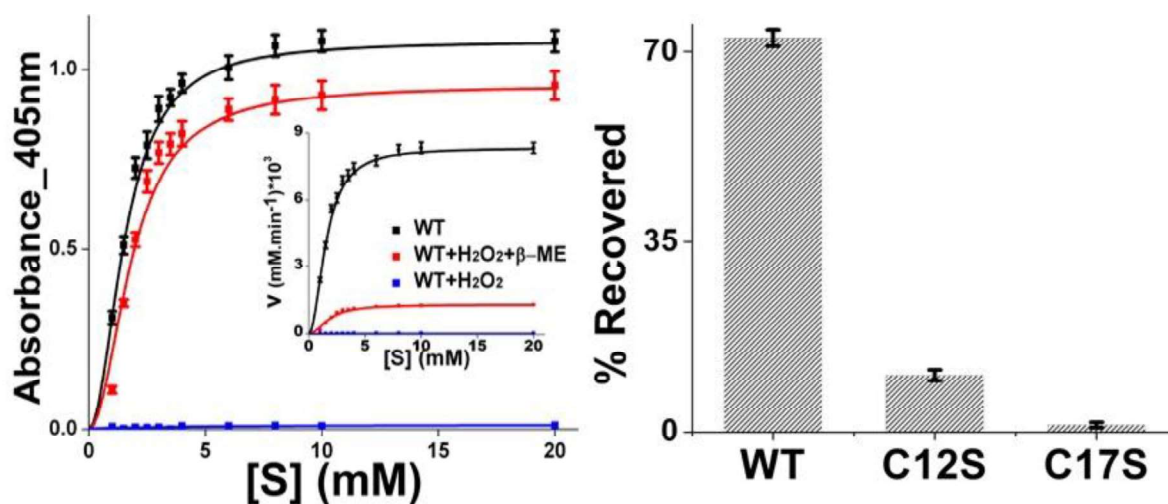


Figure 27. Redox regulation of *Vc*LMWPTP-2 and its mutants. (a) Reactivation kinetics of H_2O_2 inactivated *Vc*LMWPTP-2 using β -ME as reducing agent (Absorbance_405nm vs. [S] (mM)), along with $V/[S]$ Michaelis-Menten curves (inset) (b) Recovery by β -ME of H_2O_2 mediated inactive WT, C12S and C17S.

3.I.8. Molecular dynamics simulation

3.I.8.1. Mechanical stability and dynamic nature of loops

To understand the dynamic nature of *VcLMWPTP-2* active site we carried out MD simulations on ligand free *VcLMWPTP-2* for 1000 ns. The correlation between atomic positional fluctuations in *VcLMWPTP-2* has been investigated from (nanosecond) molecular dynamics (MD) simulations. It is evident from their C α RMSD values over the entire simulation time that *VcLMWPTP-2* is mechanically stable [Fig. 28a]. The RMSF values, averaged over 1000 ns of MD simulation trajectories, showed that some regions of *VcLMWPTP-2*, especially the $\beta 2$ - $\alpha 2$ loop and DPY loop, show significant dynamic behavior which might be required for ligand binding/release [Fig. 27b]. As expected, P-loop exhibits a rigid structure with minimum RMSF, due to intricate hydrogen bonding network, which is required to arrest the phosphotyrosine tightly [Fig. 27b].

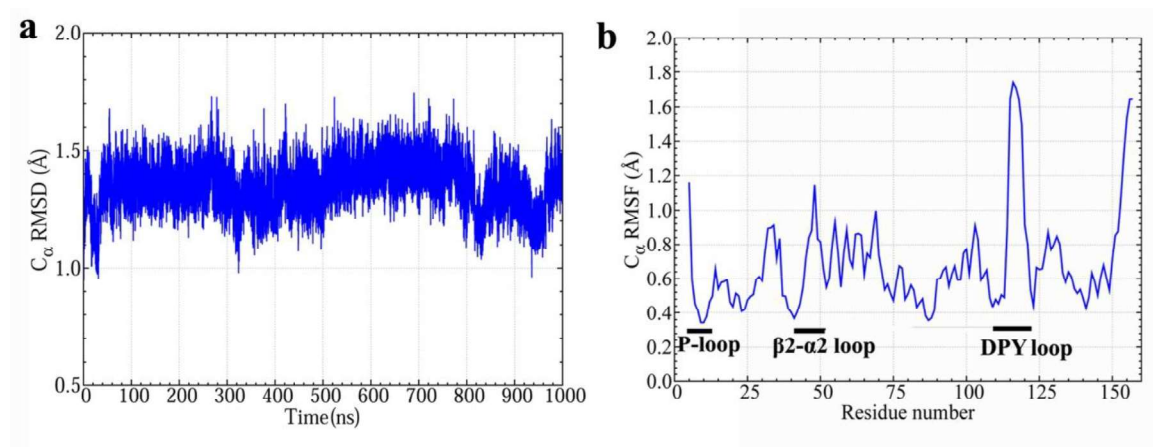


Figure 28. Structural flexibility of *VcLMWPTP-2*. (a) Root Mean Square Deviation (RMSD) for the C-alpha atoms for the full simulation range. (b) Root Mean Square Fluctuation (RMSF) for the C-alpha atoms is shown as a function of residue number for the entire 1000 ns MD trajectory. $\beta 2$ - $\alpha 2$ loop and DPY loop show significantly higher flexibility while the P-loop exhibits a rigid structure with minimum RMSF.

3.I.8.2. Dynamic nature of residues around the active site

Consideration of the dynamic nature of residues around the active site indicate that T13, N15, R18 (of motif 1) and D90, M91, N93 (of motif 4) exhibit minimum fluctuation [Fig. 29] which is in corroboration with figure 25a. Strong hydrogen bonding network amongst them persists in almost all the snapshots. These H-bonding interactions along with hydrophobic interaction with M91 are sufficient to maintain the proper conformation of R18 for ligand binding. Interestingly, side chain of D125 assumes multiple conformations in the absence of ligand while in figure 25b it assumes similar conformation in all ligand bound structures. Thus, the conformation of D125 is ligand induced and it does not influence the conformation of R18. In the MOPS bound state one side of the active site wall is defined by Y127, T47 and K49 with a weak hydrogen bond between K49 and Y127 [Fig. 20 & 22]. In our MD simulation we see that most of the conformations of K49 and Y127 are away from the active site and the hydrogen bonding interaction between them is abolished. Thus the resulting fluctuations in MD reflect the role of the conserved residues. While the residues of motif-1 and motif-3 together maintains the architecture of the active site. Large fluctuations of the residues of motif-2 and motif-4 and their movements away from active site, in apo form, reflect their role in substrate binding or release.

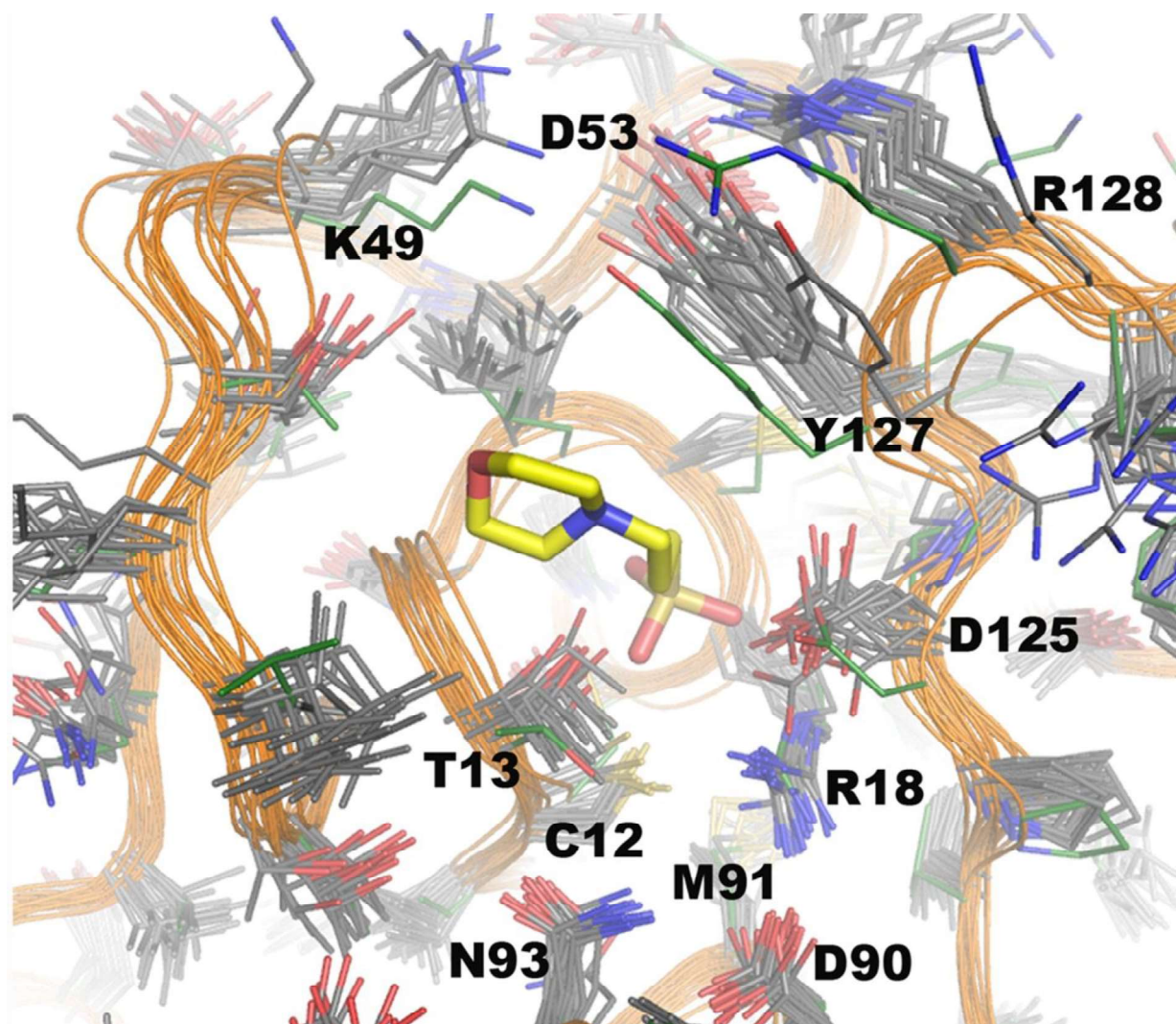


Figure 29. Superposition of the snapshots of MD simulation trajectories to show the movement of residues around P-loop. Green sticks indicate the starting conformation while its MOPS molecule is shown to indicate the location of active site.

3.1.9 Unique surface charge and grooves around the P-loop

*Vc*LMWPTP-2 exhibits a distinctive surface electrostatic charge and grooves distribution around the active site which outplays others. Multiple sequence alignment (Fig. 17) and phylogenetic analysis [Fig. 12] also indicated the uniqueness of amino acid sequence of *Vc*LMWPTP-2. Absence of several large hydrophobic residues around the active site makes the phosphotyrosine binding cavity shallow which is quite different from mammalian counterpart but somewhat similar to bacterial LMWPTPs. The amino acid composition around the active site of *Vc*LMWPTP-2 provides uniquely different surface charge around the active site [Fig. 30]. Interestingly, *Vc*LMWPTP-1(PDB: 4LRQ) resembles closely to LMWPTP from *E. histolytica* (PDB: 3IDO). But uniqueness of *Vc*LMWPTP-2 (PDB: 5Z3M) lies not only in the shallowness of the active site but also at the sub-sites that binds the residues either side of phosphotyrosine display quite different grooves and charges. The surface of 3ROF (*S.aureus*) is overlaid with a bound substrate mimicking peptide. The active site pTyr mimetic and other sub sites C-terminal to it are shown in space filling model [Fig. 30]. Comparison of surface charge and grooves of *Vc*LMWPTP-2, especially at these sub sites, indicate the idiosyncratic nature. Although definite information about the sub site location/charge at the N-terminus side of the pTyr is not available, presence of an exclusive positive patch and unmatched grooves at this region, definitely points towards a specific pTyr containing substrate for *Vc*LMWPTP-2.

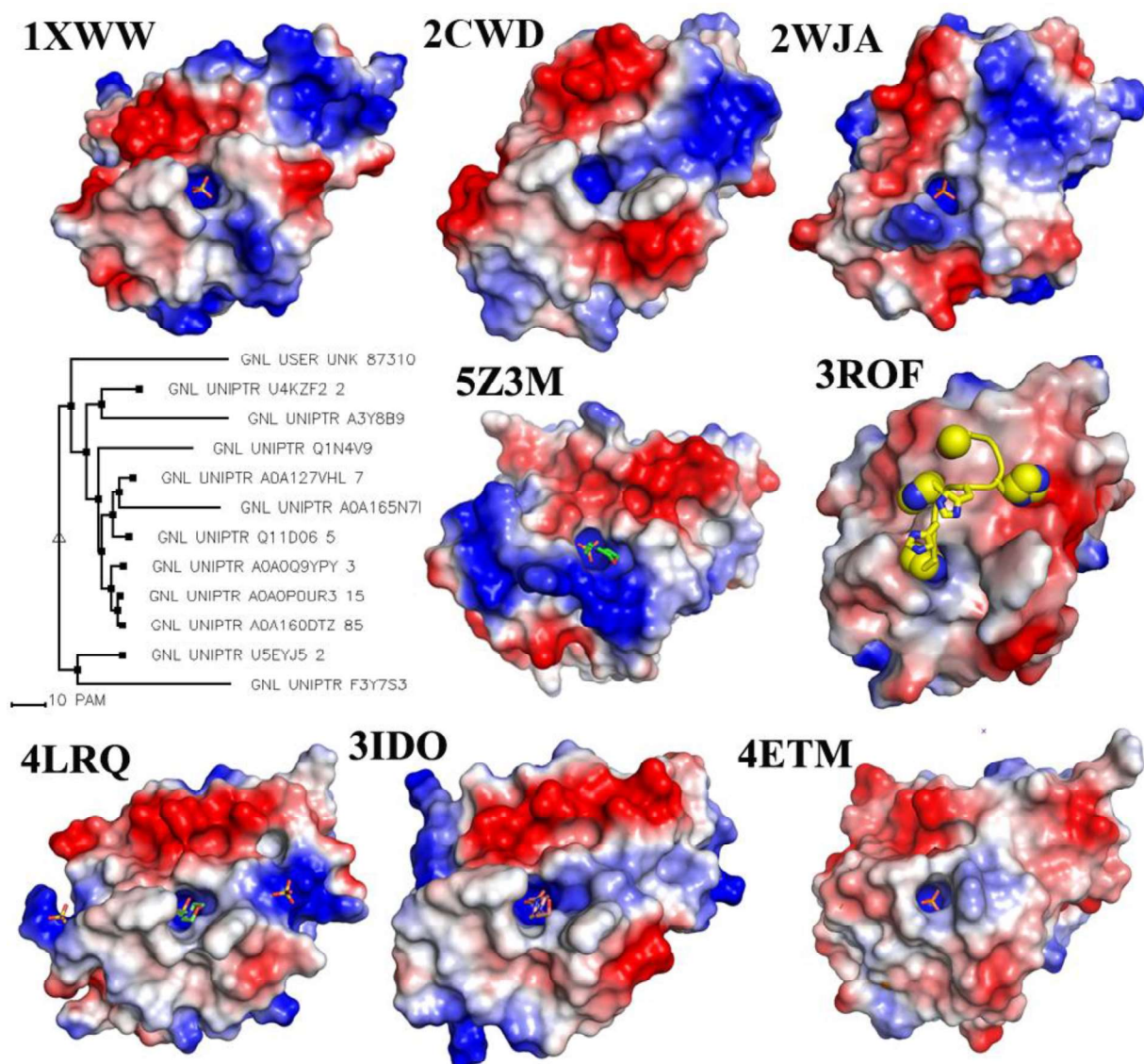


Figure 30. Phylogenetic tree and comparison of the electrostatic surface charge distribution of *VcLMWPTP-2* with homologous LMWPTPs (>30% sequence identity) around the active site cleft. Comparison of electrostatic surface charge distribution between *Human* (PDB ID: 1XWW), *Thermos thermophilus* (PDB ID: 2CWD), *Escherichia coli* (PDB ID: 2WJA), *VcLMWPTP-2* (PDB ID: 5Z3M), *Staphylococcus aureus* (PDB ID: 3ROF), *VcLMWPTP-1* (PDB ID: 4LRQ), *Eentamoeba histoltica* (PDB ID: 3IDO), and *Bacillius subtilis* (PDB ID: 4ETM) have been represented. Phylogenetic analysis of 124 non redundant LMWPTP sequences shows *VcLMWPTP-2* (Top; UNK 87310) lies as single taxon.

PART-II

3.II. Role of Cys17 in *Vc*LMWPTP-2: Is it protective or catalytic?

A detailed study on phosphatase activity including role of active site residues have been delineated in 3.I.7. Phosphatase activity revealed some interesting features of *Vc*LMWPTP-2 which is in contrast with all other LMWPTPs including *Vc*LMWPTP-1. Mutation of catalytic cysteine to serine (C12S) of the conserved motif C¹²TGNLCR¹⁸S, phosphatase activity has been decreased but not diminished unlike all other cases (including *Vc*LMWPTP-1). On the other hand C17S mutant shows ~30% decrease of activity (K_M 2.49± 0.1mM) with respect to WT but the double mutant (C12S/C17S), where both the cysteine were mutated to serine, exhibits complete loss of activity. This feature of *Vc*LMWPTP-2 tempted us to study further on C12S mutant.

3.II.1 Cloning and Over-expression of *Vc*LMWPTP-2-C12S

C12S mutant has been cloned by site directed mutagenesis by two-step PCR method in pET28a(+) vector and successfully this mutant was overexpressed as 6×histidine tagged recombinant protein in BL21(DE3) of *Escherichia coli*. Profiles of cloning and overexpression of mutant protein are given in figure 31.

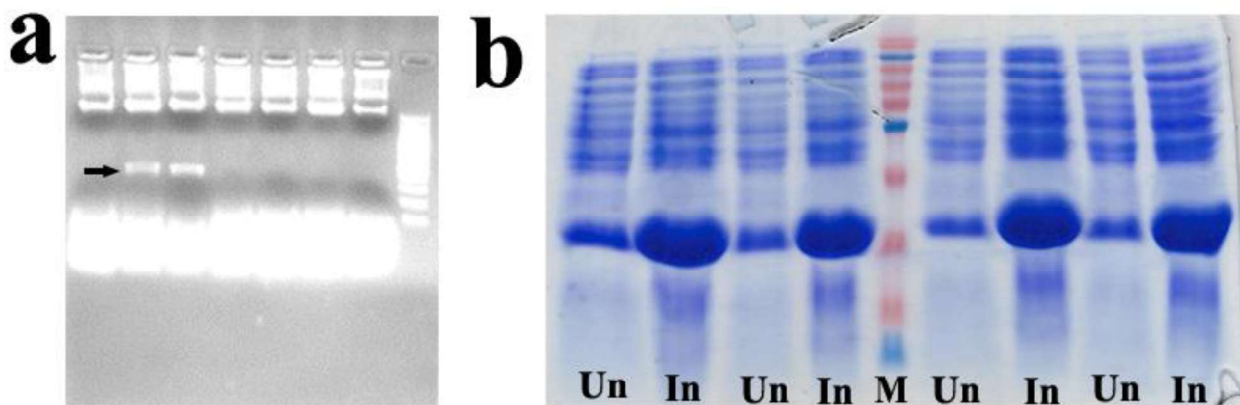


Figure 31. Clone check and overexpression of *VcLMWPTP-2 C12S* in (a) 1% agarose gel after digestion with restriction enzyme *NdeI* and *BamHI*. Lane 1-7: all samples contain the plasmid although lane 2 and 3 contains plasmid with insert as shown with black arrow, lane 8: DNA ladder (b) 15% SDS-PAGE profile, showing the overexpression of *VcLMWPTP-2 C12S*. Lanes 1, 3, 6 and 8: whole cell lysate of the uninduced cells (marked as 'Un'), lanes 2, 4, 7 and 9: whole cell lysate of the induced cells (marked as 'In'), lane 5: Protein Marker (marked as 'M') marking **10.5, 14, 22, 29, 42, 51, 62, 70, 95, 130, 175 kDa (from bottom to top).**

3.II.2 Purification of target protein by Ni-NTA affinity chromatography

Purification of target protein with 6×-Histidine tag was performed by Ni-NTA affinity chromatography. Elution of histidine tagged bound proteins were performed by step gradient method. In this method we used lysis buffer with increasing concentration of imidazole which was used as an eluent to elute protein from Ni-NTA column following the same procedure as wild type protein. The elution pattern of C12S mutant is shown in figure 32. C12S was purified at 10°C maintaining high percentage of glycerol as it was very prone towards precipitation. As shown in figure 32, protein was precipitated into the Ni-NTA bead also and the color of bead became bluish white after final elution with lysis buffer containing 300mM imidazole. Increase in

imidazole concentration (upto 500mM) results same as shown in the figure below. For purification 50mM MOPS, pH7.7, 300mM NaCl, 15% glycerol, 3mM β -ME was used as lysis buffer.

3.II.3 Thrombin cleavage and size exclusion chromatography

Histidine tag was not removed in case of *Vc*LMWPTP-2 C12S as the thrombin cleaved protein did not behave well and as it was very prone towards precipitation in size exclusion column, thereafter purification through SEC was also excluded after optimization.

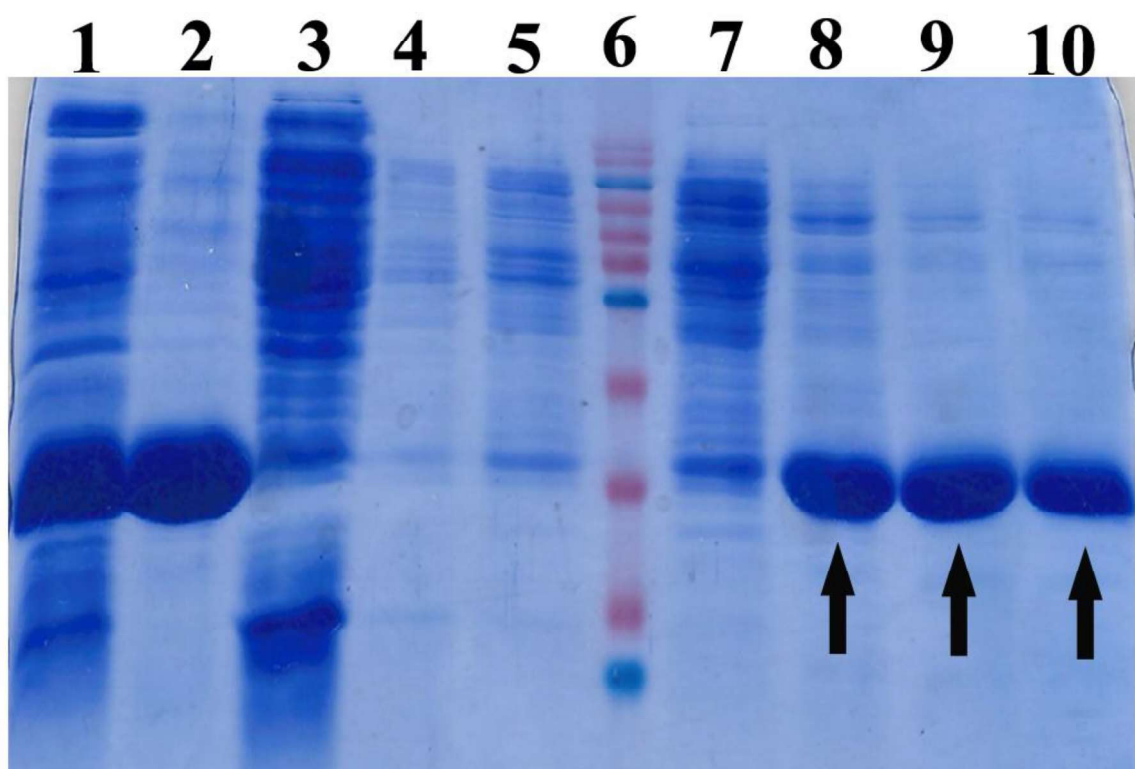


Figure 32. Ni-NTA purification profile of *VcLMWPTP-2 C12S* represented by 15% SDS-PAGE. Lane1 represents resuspended pellet after sonication and centrifugation; Lane 2 represents Ni-NTA bead after final elution; Lane 3 represents flow through after passing the cell supernatant through Ni-NTA beads equilibrated with lysis buffer. Lane 4 represents sample collected after washing the column with lysis buffer containing 5mM imidazole. Lane 5 represents sample collected after washing the column with lysis buffer containing 20mM imidazole. Lane 6 represents protein marker. Lane 7 represents sample collected after washing the column with lysis buffer containing 50mM imidazole. Lane 8 represents sample collected after washing the column with lysis buffer containing 100mM imidazole. Lane 9 represents sample collected after washing the column with lysis buffer containing 150mM imidazole. Lane 10 represents sample collected after washing the column with lysis buffer containing 300 mM imidazole (Presence of target protein represented with black arrow). **N:B: Protein Marker marking 10.5, 14, 22, 29, 42, 51, 62, 70, 95, 130, 175 kDa (from bottom to top).**

3.II.4 Crystallization

VcLMWPTP-2 C12S mutant was very susceptible for precipitation. Thereafter crystallization trials were performed at the same day of protein purification. After Ni-NTA chromatography it was subjected towards buffer exchange to remove imidazole up to <2mM imidazole concentration. Then it was concentrated until precipitation started and after ~1ml of volume addition of glycerol as well as β -ME was started with each spin to minimize precipitation. Finally we performed crystallization trials with ~20% glycerol and 5mM β -ME into the protein solution. On the other hand we added glycerol and β -ME into precipitant also during crystallization trials.

Crystallization trials were performed with various grid screens and crystal screens among them crystals were appeared at D4 of ammonium sulphate (AMS) grid screen from Hampton. Appearance of crystals took place with varying concentration of glycerol and β -ME at the precipitant. Crystals was appeared in presence of 15% glycerol and 2mM β -ME with D4 of AMS grid screen as precipitant and upon mixing 4 μ L of protein with 1 μ L of precipitant at 277K temperature. Needle-shaped crystals were obtained readily but the crystals were very thin in texture and the crystallization drop was very prone towards skin formation hindering harvesting the crystals through loop. Unfortunately we were unable to grow it big enough for further structural studies.

Chapter-4

Discussion

Low molecular weight protein-tyrosine phosphatases (LMWPTPs), an enigmatic member of PTP family ubiquitously found across a spectrum of genera from prokaryotes to higher eukaryotes. They play specific and dominant role in setting up the cellular tyrosine phosphorylation level [124]. The dephosphorylation reaction catalyzed by LMWPTPs is primarily governed by their active site cysteine in the P-loop. Because of its microenvironment the pK_a value of catalytic cysteine lies within 4-6. Therefore, at physiological pH catalytic cysteine exists as thiolate ($R-S^-$) ion and acts as nucleophile [125]. The catalytic mechanism of cysteine-dependent LMWPTPs, proceed in two distinct steps— S_N2 displacement of tyrosine by the thiol (Cys12) that is followed by protonation of the tyrosinate through either Asp125 or a second conserved cysteine (Cys17) [30, 126-127]. This creates an enzyme-bound thiophosphate ester intermediate and free tyrosine. In the second step, the thiophosphate ester is hydrolyzed by an incoming water molecule (activated by Asp125) in another S_N2 displacement, producing free phosphate and regenerating the thiol [119]. Switching between an “Open” and “Closed” conformation by ligand bound and unbound states respectively has been reported earlier [117]. Structural-superposition of *Vc*LMWPTP-2 with an ‘open structure’ of apo-MptpA (PDB: 2LUO) and ‘closed structure’ of chloride- and glycerol-bound MptpA (PDB: 1U2Q) shows that MOPS bound *Vc*LMWPTP-2 assumes a “closed” conformation although the surface of the substrate binding well is broadened by Thr47-Gly48-Lys49 residues forming an ‘open-mouthed closed’ substrate binding pocket. Our MD simulation data with apo *Vc*LMWPTP-2 structure, where the MOPS was removed, also indicate a movement of the DPY loop. Absence of a large hydrophobic residue in $\beta 2$ - $\alpha 2$ loop and a single such residue in DPY loop, compared to other LMWPTPs coupled with

the formation of hydrogen bond between K49 and Y127 may play a crucial role in substrate binding cavity formation.

It has been well characterized that LMWPTPs from both eukaryotic and prokaryotic sources produces inactive oligomers in solution that are in equilibrium with its active monomers [41-42]. The inactive oligomers thus formed are believed to act as supramolecular proenzymes. LMWPTP from *Bos taurus* (BPTP) becomes enzymatically inactive via dimerization in solution which is supported by crystal structure [41,128] [Fig. 24c]. It also forms higher oligomers in crowded condition and thus activity depends on the affinity towards the substrate in a competitive manner with oligomerization [129-130]. Dimerization of BPTP shows the involvement of two tyrosine residues occluding the active site thereby forming an enzymatically inactive dimer. The dimerization of both prokaryotes and eukaryotes involve the active site residues and the docking models are also in full agreement with the crystallographic inactive dimers [42]. However, the equivalence between dimerization and inactivation is not applicable for *Vc*LMWPTP-1 [113]. Crystal structure of *Vc*LMWPTP-2 indicates that dimerisation does not occlude the active site [Fig. 24a] rather bound substrate analog MOPS at the active site is an indicative of its accessibility towards phosphotyrosine containing substrates. Moreover, high PTPase activity determined by the kinetic data supports the availability of the active site for substrates.

Multiple sequence alignment shows the presence of threonine (13th position) instead of leucine in the P-loop unlike most of the LMWPTPs. Mutating threonine to leucine (T13L) reduces catalytic activity ($K_M \sim 1.66 \pm 0.1$ mM) as probably in this case Thr13 makes hydrogen bond with conserved Asn93 to keep the conformation of this residue in a particular orientation so that it can form H-bond with Arg18 which is a

critical residue for substrate binding. The reason for the residual activity of C12S mutant is difficult to explain. Credible explanation comes from the strong hydrogen bonding capability of serine which introduces some additional H bonding interaction that may slightly alter the conformation of the active site in such a way that the substrate orientation becomes optimal. Cirri et al. described about the ability of binding of Cys17 to PTP-inhibitor in the active site [131]. Presumably, in this case, Cys17 takes a major role in its catalytic pathway through a nucleophilic approach to the substrate although this prediction requires further structural conformation.

In cells downstream of many surface receptors H_2O_2 is produced which acts as an oxidant of PTPs [132]. While the low pK_a of Cys12 is beneficial for better nucleophilic attack to the phosphate group of phosphotyrosine containing substrates it renders Cys12 prone to oxidation. Even mild oxidation could convert the thiolate ion of active site cysteine to sulphenic acid (R-SOH) which is sufficient to inactivate the enzyme although reversible conversion of inactive R-SOH to R-SH is still possible. Upon further oxidation cysteine is converted to sulphinic (SO_2H) or sulphonic acid (SO_3H) derivatives which are probably irreversibly formed terminal products that do not permit cellular reduction systems to operate. Phosphatases, e.g., PTEN, CDC25, PTP-1B, LMWPTP also contains a second cysteine residue in close proximity to the catalytic cysteine, and its role is to prevent the irreversible oxidation of the thiolate to sulfinic and sulfonic acid by forming an intramolecular disulfide bond upon oxidation to sulfenic acid [122]. In presence of thiols, GSH, DTT, and β -ME cysteine reactivation of PTPs becomes fully reversible although the rate of reactivation by thiols is 10-100 folds slower than that of inactivation [123]. All members of the LMWPTP subfamily have two vicinal cysteines in catalytic pocket (for V_c LMWPTP-

2 Cys12 and Cys17), capable to form an intramolecular S–S bridge [39]. Under mild oxidative stressed condition redox regulation through the formation of intra-molecular S–S bridge with the Cys17. The disulphide bond is, however, reversibly broken down under reducing condition thereby indicating a definite role in protecting the catalytic cysteine.

Prokaryotic LMWPTPs are virulence factors that mimic eukaryotic phosphatases and dephosphorylate eukaryotic proteins, thereby interfering with the host defence response. An example is a phosphatase from *Mycobacterium tuberculosis* released into the extracellular medium and interfering with the host signaling pathways [133]. Sometime they work in tandem with the cognate kinase, like Wzb/Wzc pair, controlling the biosynthesis and transport of virulence factors, such as exo- and capsular polysaccharides [23, 134-135]. It has been suggested that LMWPTPs are essential in certain processes, such as bacterial stress resistance [136]. All these facts make LMWPTPs an attractive drug target. For facultative human pathogen *Vibrio cholerae*, dissimilarity in surface charge distribution around the active site between two LMWPTPs from the same organism (*Vc*LMWPTP-1 and *Vc*LMWPTP-2) indicates firmly about their diverse specificity on substrates. While *Vc*LMWPTP-1 resembles to a great extent with *E.histolytica* the distribution of surface charges and grooves around the active site of *Vc*LMWPTP-2 delineates unique features. Structure of *Vc*LMWPTP-2 provides an idea about the plausible nature of its true substrate would likely be which could be utilized to design specific inhibitors.

References

- [1] "Principles behind the multifarious control of signal transduction ERK phosphorylation and kinase/phosphatase control", J. J. Hornberg, F. J. Bruggeman, B. Binder, C. R. Geest, A. J. M. Bij de Vaate, J. Lankelma, R. Heinrich, H. V. Westerhoff, *FEBS Journal*, **2005**, 272, 244–258
- [2] "The coordinated action of protein tyrosine phosphatases and kinases in cell signaling.", H. Sun, N.K. Tonks, *Trends Biochem Sci*, **1994**, 19, 480–485
- [3] "Signal transduction: An eye on organ development", J. A. Epstein, B.G. Neel, *Nature*, **2003**, 426, 238–239
- [4] "Mathematical models of protein kinase signal transduction.", R. Heinrich, B. G. Neel, T. A. Rapoport, *Mol. Cell*, **2002**, 9, 957–970
- [5] "Overview of Protein Serine/Threonine Phosphatases", P. T. W. Cohen (Editors: J. Arino & Alexander D. R.) *Spring-Verlag*, Berlin, **2004**
- [6] "Protein tyrosine phosphatases: from genes, to function, to disease", N. K. Tonks, *Nat Rev Mol Cell Biol.*, **2006**, 7, 833–46
- [7] "Protein tyrosine phosphatases in the human genome.", Alonso A, Sasin J, Bottini N, Friedberg I, Osterman A, Godzik A, et al., *Cell*, **2004**, 6, 699–711
- [8] "Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis", Li X, Oghi KA, Zhang J, Krones A, Bush KT, Glass CK, et al., *Nature*, **2003**, 426, 247–54.
- [9] "Analysis of the protein-kinase activity of Escherichia coli cells", M. Manai, A.J. Cozzone, *Biochem. Biophys. Res. Commun.*, **1979**, 91, 19–826
- [10] "Evidence for protein kinase activities in the prokaryote Salmonella typhimurium", J.Y. Wang, Jr D.E. Koshland, *J. Biol. Chem.*, **1978**, 253, 7605–7608
- [11] "A gene encoding a protein serine/threonine kinase is required for normal development of M. xanthus, a gram-negative bacterium", J. Muñoz-Dorado et al., *Cell*, **1991**, 67, 995–1006
- [12] "No longer an exclusive club: eukaryotic signalling domains in bacteria", C.J. Bakal, J.E. Davies, *Trends Cell Biol.*, **2000**, 10, 32–38.
- [13] "WaaP of Pseudomonas aeruginosa is a novel eukaryotic type protein-tyrosine kinase as well as a sugar kinase essential for the biosynthesis of core lipopolysaccharide", X. Zhao, J.S. Lam, *J. Biol. Chem.* **2002**, 277, 4722–4730.
- [14] "Endogenous protein phosphorylation in Escherichia coli extracts", M. Manai, A.J. Cozzone, *Biochem. Biophys. Res. Commun.*, **1982**, 107, 981–988

- [15] "Characterization of a bacterial gene encoding an auto-phosphorylating protein tyrosine kinase", C. Grangeasse, et al., *Gene*, **1997**, 204, 259–265
- [16] "Transmembrane modulator dependent bacterial tyrosine kinase activates UDP-glucose dehydrogenases", I. Mijakovic, et al., *EMBO J.*, **2003**, 22, 4709–4718
- [17] "Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*", J.K. Morona, et al., *Mol. Micro biol.*, **2000**, 35, 1431–1442
- [18] "Staphylococcus aureus operates protein-tyrosine phosphorylation through a specific mechanism", D. Soulat, et al., *J. Biol. Chem.*, **2006**, 281, 14048–14056
- [19] "Structural organization of the protein-tyrosine autokinase Wzc within *Escherichia coli* cells", P. Doublet, et al., *J. Biol. Chem.*, **2002**, 277, 37339–37348
- [20] "Periplasmic protein-protein contacts in the inner membrane protein Wzc form a tetrameric complex required for the assembly of *Escherichia coli* group 1 capsules", R.F. Collins, et al., *J. Biol. Chem.*, **2006**, 281, 2144–2150
- [21] "Bacterial Protein-Tyrosine Kinases", L. Shi, A. Kobir, C. Jers, I. Mijakovic, *Current Proteomics*, **2010**, 7, 1-7.
- [22] "Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*", C. Whitfield, *Annu. Rev. Biochem.*, **2006**, 75, 39-68.
- [23] "Cells of *Escherichia coli* contain a protein tyrosine kinase, Wzc, and a phosphotyrosine protein phosphatase, Wzb.", C. Vincent, P. Doublet, C. Grangeasse, E. Vaganay, A.J. Cozzzone, B. Duclos, *J. Bacteriol.*, **1999**, 181, 3472-3477
- [24] "Identification of an *Escherichia coli* operon required for formation of the O-antigen capsule", M. Elgrably Weiss, C.M. Abe, S. Knutton, M.A. Saper, I. Rosenshine, A. Peleg, Y. Shifrin, O. Ilan, C. Nadler-Yona, S. Nov, S. Koby, K. Baruch, S. Altuvia, *J. Bacteriol.*, **2005**, 187, 5259-5266
- [25] "Influence of tyrosine-kinase Wzc activity on colanic acid production in *Escherichia coli* K12 cells.", B. Obadia, S. Lacour, P. Doublet, H. Baubichon-Cortay, A.J. Cozzzone, C. Grangeasse, *J. Mol. Biol.*, **2007**, 367, 42-53.
- [26] "Bacterial polysaccharide co-polymerases share a common framework for control of polymer length", M. Papadopoulos, L. Van Den Bosch, J.L. Rubinstein, J. Féthière, A. Matte, M. Cygler, A. Tocilj, C. Munger, A. Proteau, R. Morona, L. Purins, E. Ajamian, J. Wagner, *Nat. Struct. Mol. Biol.*, **2008**, 15, 130-138.
- [27] "Stop and go: regulation of chain length in the biosynthesis of bacterial polysaccharides.", C. Whitfield, K. Larue, *Nat. Struct. Mol. Biol.*, **2008**, 15, 121-123

- [28] "Low molecular weight protein tyrosine phosphatases: small, but smart", G. Raugei, G. Ramponi, P. Chiarugi, *Cell. Mol. Life Sci.*, **2002**,59, 941–949
- [29] "Evidence for Protein-tyrosine-phosphatase Catalysis Proceeding via a Cysteine-Phosphate Intermediate", K. Guan, J. E. Dixon, *J. Biol. Chem.*, **1991**, 266, 17026-17030
- [30] "Kinetic and Site-directed Mutagenesis Studies of the Cysteine Residues of Bovine Low Molecular Weight Phosphotyrosyl Protein Phosphatase", J. P. Davis, M. Zhou, R. L. Van Etten, *J. Biol. Chem.*, **1994**, 269, 8734-8740
- [31] "Tyrosine phosphorylation and bacterial virulence.",S. E. Whitmore, R. J. Lamont, *Int J Oral Sci.*, **2012**, 4 (1), 1–6.
- [32] "Life among the primitives: protein O-phosphatases in prokaryotes.", P.J. Kennelly, M. Potts, *Front Biosci.*, **1999** ,4 , D372-85
- [33] "Functional studies of protein tyrosine phosphatases with chemical approaches.", Z.Y.Zhang, *Biochim Biophys Acta.*, **2005** ,1754 , 100–107.
- [34] "Low molecular weight protein tyrosine phosphatase: Multifaceted functions of an evolutionarily conserved enzyme", A. Caselli, P. Paoli, A. Santi, C. Mugnaioni, A. Toti, G. Camici, P. Cirri, *BBA*, **2016** ,1864 , 1339–1355
- [35] "Structure and Substrate Recognition of the Staphylococcus aureus Protein Tyrosine Phosphatase PtpA", C. Vega, S. Chou, K. Engel, M. E. Harrell, L. Rajagopal, C. Grundner, *J Mol Biol.*, **2011** ,413 , 24–31
- [36] "Crystal Structures of Wzb of Escherichia coli and CpsB of Streptococcus pneumoniae, Representatives of Two Families of Tyrosine Phosphatases that Regulate Capsule Assembly", G. Hagelueken, H. Huang, I. L. Mainprize , C. Whitfield, J.H. Naismith, *J. Mol. Biol.*, **2009** ,392 , 678–688
- [37] "Crystal Structure of Low-Molecular-Weight Protein Tyrosine Phosphatase from Mycobacterium tuberculosis at 1.9-Å Resolution", C. Madhurantakam, E. Rajakumara, P. A. Mazumdar, B. Saha, D. Mitra, H. G. Wiker, R. Sankaranarayanan, A. K. Das, *J. Bacteriol.*, **2005** ,187 , 2175–2181
- [38] "Protein tyrosine phosphatases: from genes, to function, to disease", N. K. Tonks, *Nat Rev Mol Cell Biol.*, **2006** ,7 , 833-46
- [39] "Redox regulation of the protein tyrosine phosphatase PTP1B in cancer cells", Y.W. Lou, Y.Y. Chen, S.F. Hsu, R.K. Chen, C.L. Lee, K.H. Khoo, N.K. Tonks, T.C. Meng, *FEBS Journal*, **2008** ,275 , pp. 69–88
- [40] "The Inactivation Mechanism of Low Molecular Weight Phosphotyrosine-protein Phosphatase by H₂O₂", A. Caselli, R. Marzocchi, G. Camici, G. Manao, G. Moneti G. Pieraccini, G. Ramponi, *J. Biol. Chem.*, **1998** ,273 , 32554–32560

- [41] "The structure of the bovine protein tyrosine phosphatase dimer reveals a potential self-regulation mechanism", L. Tabernero, B.N. Evans, P.A. Tishmack, R.L. Van Etten, C.V. Stauffacher, *Biochemistry*, **1999**, *38*, 11651-11658
- [42] "Weak oligomerization of low-molecular-weight protein tyrosine phosphatase is conserved from mammals to bacteria", J. Blobel, P. Bernadó, H. Xu, C. Jin, M. Pons, *FEBS J.*, **2009**, *276*, 4346-4357
- [43] "PDGF receptor as a specific in vivo target for low m(r) phosphor tyrosine protein phosphatase", P. Chiarugi, P. Cirri, G. Raugei, G. Camici, F. Dolfi, A. Berti, et al., *FEBS Lett*, **1995**, *1*, 49-53.
- [44] "LowM(r) phosphotyrosine protein phosphatase interacts with the PDGF receptor direct via its catalytic site ", P. Chiarugi, P. Cirri, G. Raugei, G. Manao, M.L. Taddei, G. Ramponi, *Biochem Biophys Res Commun*, **1996**, *1*, 21-25.
- [45] "Insight into the role of low molecular weight phosphor tyrosine phosphatase (LAM-PTP) on platelet-derived growth factor receptor (PDGF-r) signaling — LMW-PTP controls PDGF-r kinase activity through TYR-de phosphorylation ", P. Chiarugi, P. Cirri, M.L. Taddei, E. Giannoni, T. Fiaschi, F. Buricchi, et al., *J Biol Chem*, **2002**, *40*, 37331-8.
- [46] "The Src and signal transducers and activators of transcription pathways as specific targets for low molecular weight phosphor tyrosine-protein phosphatase in platelet-derived growth factor signaling", P. Chiarugi, P. Cirri, F. Marra, G. Raugei, T. Fiaschi, G. Camici, et al., *J Biol Chem.*, **1998**, *12*, 6776-85
- [47] "Modulation of Src activity by Low molecular weight protein tyrosine phosphatase during osteoblast differentiation", W.F. Zambuzzi, J.M. Granjeiro, K. Parikh, S. Yuvaraj, M.P. Peppelenbosch, C.V. Ferreira., *Cell.Physiol.Biochem.*, **2008**, *5-6*, 497-506 0
- [48] "Modulation of STAT5 interaction with LMW-PTP during early megakaryocyte differentiation", S. Rigacci, V. Guidotti, M. Parri, A. Berti, *Biochemistry*, **2008**, *6*, 1482-9
- [49] "Low Mr phosphotyrosine protein phosphatase associates and dephosphorylates p125 focal adhesion kinase, interfering with cell motility and spreading.", S. Rigacci, E. Rovida, P. Dello Sbarba, A. Berti, *J Biol Chem.*, **2002**, *44*, 41631-6.
- [50] "The low M(r) protein-tyrosine phosphatase is involved in Rho mediated cytoskeleton rearrangement after integrin and platelet derived growth factor stimulation.", P. Chiarugi, P. Cirri, L. Taddei, E. Giannoni, G. Camici, G. Manao, et al., *J Biol. Chem.*, **2000**, *7*, 4640-6

- [51] "Regulation of the EphA2 kinase by the low molecular weight tyrosine phosphatase induces transformation.", K.D. Kikawa, D.R. Vidale, R.L. Van Etten, M.S. Kinch, *J Biol. Chem.*, **2002** ,42 , 39274–9.
- [52] "LMW-PTP is a positive regulator of tumor onset and growth.", P. Chiarugi, M.L. Taddei, N. Schiavone, L. Papucci, E. Giannoni, T. Fiaschi, et al., *Oncogene*, **2004** ,23 , 3905–14.
- [53] "EphrinA1 repulsive response is regulated by an EphA2 tyrosine phosphatase.", M. Parri, F. Buricchi, M.L. Taddei, E. Giannoni, G. Raugei, G. Ramponi, et.al., *J Biol. Chem.*, **2005** ,40 , 34008–18.
- [54] "LMWPTP is a negative regulator of insulin-mediated mitotic and metabolic signaling.", P. Chiarugi, P. Cirri, F. Marra, G. Raugei, G. Camici, G. Manao, et al., *Biochem Biophys Res Commun.*, **1997** ,2 ,pp.676–82.
- [55] "Reduction of low molecular weight protein-tyrosine phosphatase expression improves hyperglycemia and insulin sensitivity in obesemice.", S.K. Pandey, X.X. Yu, L.M. Watts, M.D. Michael, K.W. Sloop, A.R. Rivard, et.al., *J Biol. Chem*, **2007** ,19 , 14291–9.
- [56] "Association of the acid phosphatase(ACP1) gene with triglyceridelevels in obese women ", N. Bottini, J. MacMurray, W. Peters, M. Rostamkhani, D.E. Comings., *Mol Genet Metab.*, **2002** ,3 , 226–9.
- [57] "Low-molecular-weight protein tyrosine phosphatase and human disease: in search of biochemical mechanisms.", N. Bottini, E. Bottini, F. Gloria-Bottini, T. Mustelin, *Arch Immunol Ther Exp (Warsz)*, **2002** ,2 , 95–104
- [58] "Histochemical staining of protein tyrosine phosphatase activity in primary human mammary carcinoma: relationship with established prognostic indicators", K.R. Kidd, B.J. Kerns, R.K. Dodge, J.R. Wiener, *J. Histochem. Cytochem.*, **1992** ,40 ,729–735.
- [59] "Up-regulated expression of low molecular weight protein tyrosine phosphatases in different human cancers ", F. Malentacchi, R. Marzocchini, S. Gelmini, C. Orlando, M. Serio, G. Ramponi, G. Raugei, *Biochem. Biophys. Res. Commun.*, **2005** ,334 ,875–883.
- [60] "LMW-PTP is a positive regulator of tumor onset and growth", P. Chiarugi, M.L. Taddei, N. Schiavone, L. Papucci, E. Giannoni, T. Fiaschi, S. Capaccioli, G. Raugei, G. Ramponi, *Oncogene*, **2004** ,23 , 3905–3914.
- [61] "Low molecular weight protein tyrosine phosphatase (LMWPTP) up regulation mediates malignant potential in colorectal cancer" E. Hoekstra, L.L. Kodach, A.M. Das, R.R. Ruela-de-Sousa, C.V. Ferreira, J.C. Hardwick, C.J. van der Woude, M.P. Peppelenbosch, T.L. Ten Hagen, G.M. Fuhler, *Oncotarget*, **2015** ,6 , 8300–8312.

- [62] "Inhibitors of Protein Tyrosine Phosphatases: Next-Generation Drugs?", L. Bialy, H. Waldmann, *Angew. Chem., Int. Ed.*, **2005**, *44*, 3814–3839
- [63] "Crystal Structure of Bovine Low Molecular Weight Phosphotyrosyl Phosphatase Complexed with the Transition State Analog Vanadate", M. Zhang, M. Zhou, R. L. Van Etten, C. V. Stauffacher, *Biochemistry*, **1997**, *36*, 15–23.
- [64] "Effects of homologous series of n-Alkyl sulfates and n-alkyl trimethyl ammonium bromides on low molecular mass protein tyrosine phosphatase activity", J. M. Granjeiro, M. A. Miranda, M. G. S. T. Maia, C. V. Ferreira, E. M. Taga, H. Aoyama, P. L. O. Volpe, *Mol. Cell. Biochem.*, **2004**, *265*, 133–140.
- [65] "Inhibition Studies with Rationally Designed Inhibitors of the Human Low Molecular Weight Protein Tyrosine Phosphatase.", A. P. R. Zabell, S. Corden, P. Helquist, C. V. Stauffacher, O. Wiest, *Bioorg. Med. Chem.*, **2004**, *12*, 1867–1880
- [66] "PTP1B as a Drug Target: Recent Developments in PTP1B Inhibitor Discovery.", S. Zhang, Z. Y. Zhang, *Drug Discovery Today*, **2007**, *12*, 373–381.
- [67] "Structure-Based Design and Discovery of Protein Tyrosine Phosphatase Inhibitor Incorporating Novel Isothiazolidinone Heterocyclic Phosphotyrosine Mimetics.", A. P. Combs, E. W. Yue, M. Bower, et al., *J. Med. Chem.*, **2005**, *48*, 6544–6548.
- [68] "IsoxazoleCarboxylic Acids as Protein Tyrosine Phosphatase 1B(PTP1B) Inhibitors", H. Zhao, G. Liu, Z. Xin, et al., *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 5543–5546.
- [69] "Crystal Structure of Bovine Heart Phosphotyrosyl Phosphatase at 2.2-Å Resolution.", M. Zhang, R. L. Van Etten, C. V. Stauffacher, *Biochemistry*, **1994**, *33*, 11097–11105
- [70] "Structure-Based Discovery of New Small Molecule Inhibitors of Low Molecular Weight Protein Tyrosine Phosphatase.", D. Vidal, J. Blobel, Y. Pérez, M. Thormann, M. Pons, *Eur. J. Med. Chem.*, **2007**, *42*, 1102–1108
- [71] "5-Arylidene-2,4-thiazolidinediones as Inhibitors of Protein Tyrosine Phosphatases.", R. Maccari, P. Paoli, R. Ottanà, M. Jacomelli, R. Ciurleo, G. Manao, T. Steindl, T. Langer, M. G. Vigorita, G. Camici, *Bioorg. Med. Chem.*, **2007**, *15*, 5137–5149.
- [72] "5-Arylidene-2-phenylimino-4-thiazolidinones as PTP1B and LMW-PTP Inhibitors", R. Ottanà, R. Maccari, R. Ciurleo, P. Paoli, M. Jacomelli, G. Manao, G. Camici, C. Laggner, T. Langer, *Bioorg. Med. Chem.*, **2009**, *17*, 1928–1937.
- [73] "Structure-Based Optimization of Benzoic Acids as Inhibitors of Protein Tyrosine Phosphatase 1B and Low Molecular Weight Protein Tyrosine Phosphatase", R.

Maccari, R. Ottanà, R. Ciurleo, P. Paoli, G. Manao, G. Camici, C. Laggner, T. Langer, *Chem Med Chem*, **2009** ,4 , 957–962.

[74] "Synthesis, Activity and Molecular Modeling of a New Series of Chromones as Low Molecular Weight Protein Tyrosine Phosphatase Inhibitors.", M. Forghieri, C. Laggner, P. Paoli, T. Langer, G. Manao,; G. Camici, L. Bondioli, F. Prati, L. Costantino, *Bioorg.Med. Chem.*, **2009**, 17, 2658–2672.

[75] "Discovery of Mycobacterium tuberculosis Protein Tyrosine Phosphatase A (MptpA) Inhibitors Based on Natural Products and a Fragment-Based Approach. ", M. Manger, M. Scheck, H. Prinz, J. P. von Kries, T. Langer, K. Saxena, H. Schwalbe, A. Fürstner, J. Rademann. H. Waldmann, *Chem Bio Chem.*, **2005** ,6 ,1749–1753.

[76] "Design, Synthesis and Inhibition Activity of Novel Cyclic Peptides against Protein Tyrosine Phosphatase A from Mycobacterium tuberculosis.", K. Chandra, D. Dutta, A. K. Das, A. Basak, *Bioorg.Med.Chem.*, **2010** ,18 , 8365–8373

[77] "Fragment-Based Discovery of Selective Inhibitors of the Mycobacterium tuberculosis Protein Tyrosine Phosphatase PtpA.", K. A. Rawls, P. T. Lang, J. Takeuchi, S. Imamura, T. D. Baguley, C. Grundner, T. Alber, J. A. Ellman, *Bioorg. Med. Chem. Lett.*, **2009** ,19 , 6851–6854

[78] "Synthetic Chalcones as Efficient Inhibitors of Mycobacterium tuberculosis Protein Tyrosine Phosphatase PtpA.", L. D. Chiaradia, A. Mascarello, M. Purificação, J. Vernal, M. N. S. Cordeiro, M. E. Zenteno, A. Villarino, R. J. Nunes, R. A. Yunes, H. Terenzi, *Bioorg.Med. Chem. Lett.*, **2008** ,18 , 6727–6230

[79] "Inhibition of Mycobacterium tuberculosis Tyrosine Phosphatase PtpA by Synthetic Chalcones: Kinetics, Molecular Modeling, Toxicity and Effect on Growth.", A. Mascarello, L. D. Chiaradia, J. Vernal, et al., *Bioorg. Med. Chem.*, **2010** ,18 , 3783–3789

[80] "Current views and challenges on clinical cholera", C. Somboonwit, L.J. Menezes, D.A. Holt, J.T. Sinnott, P. Shapshak, *Bioinformation*, **2017** ,13 , 405-409

[81] "Crystallization of soluble proteins in vapor diffusion for X-ray crystallography.", M. Benvenuti, S. Mangani, *Nat Protoc.*, **2007** ,2 ,1633–51.

[82] "Protein crystallization and phase diagrams.", N. Asherie, *Methods.*, **2004** ,34 , 266–272

[83] <https://www.jic.ac.uk/>

[84] "HAD, a data bank of heavy-atom binding sites in protein crystals: A resource for use in multiple isomorphous replacement and anomalous scattering.", S.A. Islam, D. Carvin, M.J.E. Sternberg, T.L. Blundell., *Acta Crystallographica Section D: Biological Crystallography*. **1998**, 1199–1206.

- [85] "Anomalous diffraction in crystallographic phase evaluation.", W.A. Hendrickson, *Q Rev Biophys.*, **2014** ,47, 49–93.
- [86] "Molecular replacement - Historical background.", M.G. Rossmann, *Acta Crystallographica SectD Biol Crystallogr.*, **2001** ,57 , 1360–1366.
- [87] "The Phase Problem.", G. Taylor, *Acta D*, **2003** ,59 (11) , 1881-1890
- [88] "Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases.", K. Kleppe, E. Ohtsuka, R. Kleppe, I. Molineux, H.G. Khorana, *J. Mol. Biol.*, **1971** ,56 (2) , 341–361.
- [89] "Rapid and efficient site-directed mutagenesis by single-tube 'mega primer' PCR method.", S.H. Ke, E.L. Madison, *Nucleic Acids Res.*, **1997** ,25 (16) , 3371-3372
- [90] "Sparse matrix sampling: a screening method for crystallization of proteins", J. Jancarik, S.H. Kim, *J. Appl. Cryst.*, **1991** ,24 , 409-411
- [91] "Protein crystallography beamline (PX-BL21) at Indus-2 synchrotron ", A. Kumar, B. Ghosh, H. K. Poswal, K. K. Pandey, Jagannath, M. V. Hosur, A. Dwivedi, R. D. Makde, S. M. Sharma, *J. Synchrotron Rad.*, **2016** ,23 , 629-634
- [92] "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.", T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, A. G. W. Leslie, *Acta CrystallogrD Biol Crystallogr.*, **2011** ,67, 271–281
- [93]"A graphical user interface to the CCP4 program suite", E. Potterton, P. Briggs, M. Turkenburg, E. Dodson, *ActaCrystallogr. D Biol. Crystallogr.*, **2003** ,59 ,1131-1137
- [94]"Phaser crystallographic software", A.J. McCoy, R.W. Grosse Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, *J. Appl. Crystallogr.*, **2007** ,40 , 658-674
- [95] "PHENIX: a comprehensive python-based system for macromolecular structure solution",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. GrosseKunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson,J.S. Richardson, T.C. Terwilliger, P.H. Zwart, *ActaCrystallogr. D Biol. Crystallogr.*, **2010** ,66 , 213-221
- [96] "Coot: model-building tools for molecular graphics", P. Emsley, K. Cowtan, *ActaCrystallogr. D Biol. Crystallogr.*, **2004** ,60 , 2126-2132
- [97] "Overview of the CCP4 suite and current developments", 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, *Acta.Crystallogr. D*, **2011** ,67 , 235-242

- [98] "Inference of macromolecular assemblies from crystalline state", E. Krissinel, K. Hendrick, J. *Mol. Biol.*, **2007** ,372 , 774–797 and PISA web server: <http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver>.
- [99] "Multiple sequence alignment with hierarchical clustering.", F. Corpet, *Nucleic Acids Res.*, **1988** ,16 , 10881–10890
- [100] "NPS@: Network Protein Sequence Analysis", C. Combet, C. Blanchet, C. Geourjon, G. Deléage, *TIBS*, **2000** ,25 (3) , 147-150
- [101] "UCSF Chimera – a visualization system for exploratory research and analysis", E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. J. Meng, T.E. Ferrin, *Comput. Chem.*, **2004** ,25 , 1605-1612
- [102] " CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues", J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang, *Nucleic Acids Res.*, **2006** ,1 , 34
- [103] "Different Protein Tyrosine Phosphatase Super families Resulting from Different Gene Reading Frames", J.F. Huang, *Mol. Biol. Evol.*, **2003** ,20 , 815–820
- [104] "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.", J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.*, **1994** ,22 , 4673–4680
- [105] "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to super computers.", M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, *SoftwareX*, **2015** ,1-2 , 19–25
- [106] "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data.", N. Foloppe, A.D. MacKerell, A. D, Jr., *J. Comp. Chem.*, **2000** ,21 (2) , 86–104
- [107] "Polymorphic transitions in single crystals: A new molecular dynamics method", M. Parrinello, A. Rahman, *J. Appl. Physics*, **1981** ,52 (12) ,7182–7190
- [108] "Canonical sampling through velocity rescaling", G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.*, **2007** ,126 (1) , 014101
- [109] "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems ", T. Darden, D. York, L. Pedersen, *J. Chem. Phys.*, **1993** ,98 (12) , 10089–10092.
- [110] "Purification and characterization of a low-molecular-weight acid phosphatase – a phosphotyrosyl-protein phosphatase from bovine heart", Z.Y. Zhang, R.L. Van Etten, *Arch. Biochem. Biophys.*, **1990** ,282 , 39-49

- [111] "The Origin of the Haitian Cholera Outbreak Strain", C. Chin, J. Sorenson, J. B. Harris, W. P. Robins, R. C. Charles, R. R. Jean-Charles, J. Bullard, D.R. Webster, A. Kasarskis, P. Peluso, E. E. Paxinos, Y. Yamaichi, S. B. Calderwood, J. J. Mekalanos, E. E. Schadt, M.K. Waldor, *N Engl J Med.*, **2011** ,364 , 33–42
- [112] "Benchmarking fold detection by DaliLite v.5", Holm L, *Bioinformatics*, *btz536*, **2019** ,PMID: 31263867
- [113] "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", S. Nath, R. Banerjee, U. Sen, *BBRC*, **2014** ,450 , 390-395
- [114] "Crystal Structures of a Low-Molecular Weight Protein Tyrosine Phosphatase from *Saccharomyces cerevisiae* and Its Complex with the Substrate p-Nitrophenyl Phosphate", S. Wang, L. Tabernero, M. Zhang, E. Harms, R.L. Van Etten, C. V. Stauffacher, *Biochemistry*, **2000** ,39 , 1903-1914
- [115] "Molecular basis for the dephosphorylation of the activation segment of the insulin receptor by protein tyrosine phosphatase 1B", A. Salmeen, J.N. Andersen, M.P. Myers, N.K. Tonks, D. Barford, *Mol. Cell*, **2000** ,6 ,1401-1412
- [116] "Mycobacterium tuberculosis Protein Tyrosine Phosphatase PtpB Structure Reveals a Diverged Fold and a Buried Active Site", C. Grundner, H.L. Ng,T. Alber, *Structure*, **2005** ,13 , 1625–1634
- [117] "The Apo-structure of the Low Molecular Weight Protein-tyrosine Phosphatase A (MptpA) from *Mycobacterium tuberculosis* Allows for Better Target-specific Drug Development ", T. Stehle, S. Sreeramulu, F. Löhr, C. Richter, K. Saxena, H. R. A. Jonker, H. Schwalbe, *J. Biol. Chem.*, **2012** ,287 , 34569 –34582
- [118] "Energetics of Nucleophile Activation in a Protein Tyrosine Phosphatase", T. Hansson, P. Nordlund, J. Åqvist, *J. Mol. Biol.*, **1997** ,265 , 118–127
- [119] "Asp129 of Low Molecular Weight Protein Tyrosine Phosphatase Is Involved in Leaving Group Protonation", Z. Zhang, E. Harms, R. L. Van Etten, *J. Biol. Chem.*, **1994** ,269 , 25947-25950
- [120] "Licensed under a Creative Commons Attribution 3.0 Unported License and web server: <<http://skylign.org/>>", J. Clements, T. Wheeler, R. Finn
- [121] "Staphylococcus aureus contains two low-molecular-mass phosphotyrosine protein phosphatases", D. Soulat, E. Vaganay, B. Duclos, A.L. Genestier, J. Etienne, A.J. Cozzzone, *J. Bacteriol.*, **2002** ,184 , 5194-5199
- [122] "Redox regulation of PTEN and protein tyrosine phosphatases in H₂O₂-mediated cell signaling", S.H. Choa, C.H. Leea, Y. Ahna, H. Kima, H. Kimb, C.Y. Ahnb, K.S. Yanga, S.R. Lee, *FEBS Lett.*, **2004** ,560 , 7-13

- [123] "Specific and Reversible Inactivation of Protein Tyrosine Phosphatases by Hydrogen Peroxide: Evidence for a Sulfenic Acid Intermediate and Implications for Redox Regulation", J.M. Denu, K.G. Tanner, *Biochemistry*, **1998**, *37*, 5633-5642
- [124] "Protein tyrosine phosphatases as potential therapeutic targets", R. He, Z. Yu, R. Zhang, Z. Zhang, *Acta Pharmacol Sin.*, **2014**, *35*, 1227-46
- [125] "Electrostatic Evaluation of the Signature Motif (H/V)CX5R (S/T) in Protein-Tyrosine Phosphatases", G. H. Peters, T. M. Frimurer, O. H. Olsen, *Biochemistry*, **1998**, *37*, 5383-5393
- [126] "Structure and function of the low Mr phosphotyrosine protein phosphatases", G. Ramponi, M. Stefani, *BBA*, **1997**, *1341*, 137-156
- [127] "Aspartic-129 is an essential residue in the catalytic mechanism of the low Mr phosphotyrosine protein phosphatase", N. Taddei, P. Chiarugi, P. Cirri, T. Fiaschi, M. Stefani, G. Camici, G. Rauegi, G. Ramponi, *FEBS Letters*, **1994**, *350*, 328-332
- [128] "Intramolecular Dynamics of Low Molecular Weight Protein Tyrosine Phosphatase in Monomer-Dimer Equilibrium Studied by NMR: A Model for Changes in Dynamics upon Target Binding", T. Akerud, E. Thulin, R. L. Van Etten, M. Akke, *J. Mol. Biol.*, **2002**, *322*, 137-152
- [129] "Influence of excluded volume upon macromolecular structure and associations in 'crowded' media", A.P. Minton, *Curr Opin Biotechnol.*, **1997**, *8*, 65-69
- [130] "Combined Use of NMR Relaxation Measurements and Hydrodynamic Calculations To Study Protein Association. Evidence for Tetramers of Low Molecular Weight Protein Tyrosine Phosphatase in Solution", P. Bernado', T. Åkerud, J. G. Torre, M. Akke, M. Pons, *J Am Chem Soc.*, **2003**, *125*, 916-923
- [131] "The role of Cys-17 in the pyridoxal 5'-phosphate inhibition of the bovine liver low M(r) phosphotyrosine protein phosphatase", P. Cirri, P. Chiarugi, G. Camici, G. Manao, L. Pazzagli, A. Caselli, I. Barghini, G. Cappugi, G. Rauegi, G. Ramponi, *Biochim Biophys Acta.*, **1993**, *1161*, 216-22
- [132] "Hydrogen Peroxide: A Key Messenger That Modulates Protein Phosphorylation Through Cysteine Oxidation", S.G. Rhee, Y. S. Bae, S.R. Lee, J. Kwon, *Sci STKE.*, **2000**, *2000*, pe1
- [133] "Cloning and characterization of secretory tyrosine phosphatases of Mycobacterium tuberculosis." A. Koul, A. Choidas, M. Treder, A.K. Tyagi, K. Drlica, Y. Singh, A. Ullrich, *J Bacteriol*, **2000**, *182*, 5425-5432
- [134] "Characterization of the amsI gene product as a low molecular weight acid phosphatase controlling exopolysaccharide synthesis of Erwinia amylovora.", P. Bugert, K. Geider, *FEBS Lett.*, **1997**, *400*, 252-256

[135] "Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide.", O. Ilan, Y. Bloch, G. Frankel, H. Ullrich, K. Geider, I. Rosenshine, *EMBO J.*, **1999**, 18, 3241–3248

[136] "A new tyrosine phosphorylation mechanism involved in signal transduction in *Bacillus subtilis*.", J. Kirstein, K. Turgay, *J Mol Microbiol Biotechnol.*, **2005**, 9, 182–188

Reprint of thesis
related publication



Vibrio cholerae LMWPTP-2 display unique surface charge and grooves around the active site: Indicative of distinctive substrate specificity and scope to design specific inhibitor

Shramana Chatterjee^a, Seema Nath^a, Biplab Ghosh^b, Udayaditya Sen^{a,*}

^a Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, HBNI, 1/AF Bidhan Nagar, Kolkata 700064, India

^b High Pressure & Synchrotron Radiation Physics Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400085, India

ARTICLE INFO

Keywords:

Low molecular weight protein tyrosine phosphatases
X-ray crystallography
Active dimer
Catalytic activity
Molecular dynamics

ABSTRACT

Low molecular weight protein tyrosine phosphatases (LMWPTPs) are ubiquitously found as small cytoplasmic enzymes which act on phospho-tyrosine containing proteins that are engaged in various cellular functions. *Vibrio cholerae* O395 contains two LMWPTPs having widely different sequence. Phylogenetic analysis based on a non redundant set of 124 LMWPTP sequences, designate that LMWPTP-2 from *Vibrio cholerae* O395 (VcLMWPTP-2) is a single taxon. We have determined the crystal structure of VcLMWPTP-2 at 2.6 Å with MOPS bound in the active site. Tertiary structure analysis indicates that VcLMWPTP-2 forms dimer. Studies in solution state also confirm exclusive presence of a dimeric form. Kinetic studies demonstrate that VcLMWPTP-2 dimer is catalytically active while inactivation through oligomerisation was reported as one of the regulatory mechanism in case of mammalian LMWPTP viz., *Bos taurus* LMWPTP, BPTP. Kinetic studies using *p*-nitrophenyl phosphate (*p*-NPP) as a substrate demonstrate active participation of both the P-loop cysteine in catalysis. Vicinal Cys17, in addition plays a role of protecting the catalytic Cys12 under oxidative stress. Structural analysis and MD simulations allowed us to propose the role of several conserved residues around the active site. Distribution of surface charges and grooves around the active site delineates unique features of VcLMWPTP-2 which could be utilized to design specific inhibitor.

1. Introduction

Reversible phosphorylation of tyrosine residues is the basis for controlling many diverse cellular processes [1]. The extent of tyrosine phosphorylation is dynamically controlled by two groups of enzymes, phospho tyrosine kinases (PTKs) and phospho tyrosine phosphatases (PTPs). PTPs are family of enzymes that can be divided into four subgroups: classical pTyr specific, dual specificity phosphatases, Cdc25 and low molecular weight protein tyrosine phosphatases (LMWPTP). The common feature of the four subgroups of PTPs is the occurrence of a consensus sequence motif C(X)₅RS/T, where X can be any amino acid, but are otherwise different in their sequence and topology [2]. Surprisingly this motif adopts a highly similar loop conformation which represents an example of convergent evolution. Functionally, this motif efficiently binds the phosphoryl group of the substrate phosphotyrosine to be hydrolyzed and is known as P-loop. The invariant cysteine of P-loop acts as a nucleophile [3] to carry out the dephosphorylation reaction and mutation of this residue completely abrogate the enzymatic

activity [4].

LMWPTPs, an enigmatic member of PTP family, are small cytosolic enzymes (~18 kDa) ubiquitously found across a spectrum of genera from prokaryotes to higher eukaryotes, whereas other PTPs (PTP1B, CDC45, LAR, Shp1, Shp2, PTPα, PTPβ, PTPγ) are expressed exclusively in eukaryotes [5]. Some prokaryotic LMWPTPs are virulence factors that mimic eukaryotic phosphatases and dephosphorylate eukaryotic proteins, thereby interfering with the host defense response. They are also involved in the biosynthesis and transport of virulence factors, such as *exo*- and capsular polysaccharides or in bacterial stress resistance. From the prokaryotic counterpart, structures of LMWPTP from gram-positive eubacteria *Staphylococcus aureus*–PtpA (PDB: 3ROF) [6] gram-negative eubacteria *Thermus thermophilus*–TT1001 (PDB: 2CWD), gram-positive proteobacteria *Bacillus subtilis*–YwlE (PDB: 4ETI), gram-negative proteobacteria *Escherichia coli*–Wzb (PDB: 2WMY) [7] and pathogenic *Mycobacterium tuberculosis* MptpA (PDB: 1U2Q) [8] have been reported. Among various reported regulatory mechanisms of PTPs, regulation through reversible oxidation of the active-site cysteine drew

* Corresponding author.

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

much attention [9,10]. LMWPTPs have a unique feature of hosting two cysteines in the P-loop, catalytic Cys12 and the vicinal Cys17, the later plays a major regulatory role by protecting the catalytic Cys12 from irreversible oxidation [11]. Another regulatory mechanism proceeds for LMWPTPs is through oligomerization. It has been reported that the self association of mammalian LMWPTP (viz. BPTP) produces inactive oligomers that are in equilibrium with its active monomers [12]. Inactivation occurs through direct involvements of the active site residues of one monomer with the tyrosines of the DPY-loop of other. Similar but weak dimerization has also been reported in prokaryotic LMWPTPs like YwIE from *Bacillus subtilis*, PtpB from the Gram-negative bacterium *Salmonella aureus* and Wzb from *Escherichia coli* [13]. Higher order oligomerisation of YwIE was also detected by NMR experiments [13]. Overall, the oligomerization processes were found to be weak, with in vitro dissociation constants in the millimolar range. In spite of the proposed regulatory role of the oligomerization, the large dissociation constants of the oligomerization process shed some doubt on its physiological relevance.

The gram-negative bacteria, *Vibrio cholerae*, is the causative organism of the severe diarrheal disease cholera that continues to be a significant cause of death in developing countries claims approximately 95,000 lives out of 2.9 million annually [14,15]. The genome of *Vibrio cholerae* O395 contains two LMWPTPs (VcLMWTP-1 and VcLMWTP-2). Detailed analysis with more than hundred non-redundant (NR) LMWPTPs sequences indicate that VcLMWTP-2 distinguishes itself as an idiosyncratic type of bacterial LMWPTP. Occurrence of two LMWPTP in the same organism and their biological role, especially the role of VcLMWTP-2, remains to be established. Structural data of these two VcLMWPTPs would certainly be helpful to design specific and potential inhibitor. Through high resolution crystal structure as well as studies in solution we have previously demonstrated that VcLMWTP-1 forms a catalytically active dimer which is quite different from the inactive dimers of mammalian LMWPTP [16]. The surface properties of VcLMWTP-1, although have some distinct features, resembles closely to that of *E. histolytica* LMWPTP (PDB: 3IDO). Here we describe the 2.6 Å crystal structure of *Vibrio cholerae* LMWPTP-2 (VcLMWTP-2) with substrate analogue 3-(N-morpholino) propane sulfonic acid (MOPS) bound at the active site. The crystal structure of VcLMWTP-2 represents a dimeric species which is well supported by the results of size exclusion chromatography (SEC) and Dynamic Light Scattering (DLS) experiments. The mode of dimerization of VcLMWTP-2 is, however, quite different from the active dimer of VcLMWTP-1 or inactive dimer of BPTP. In concomitant with the crystal structure, kinetic studies of VcLMWTP-2 using chromogenic substrate *p*-nitrophenyl phosphate (p-NPP) exhibit high catalytic activity. Single Cys → Ser mutant (C12S or C17S) at the catalytic loop has diminished activity but the double mutant C12/17S exhibit nominal catalytic activity, suggesting the active participation of both the cysteines in the catalytic mechanism. Distribution of surface charge and grooves around the 'P-loop', believed to govern the specificity of LMWPTPs, is quite distinctive in VcLMWTP-2 which could be exploited to design specific inhibitors.

2. Materials and methods

2.1. Cloning, expression and purification

VcLMWTP-2 has been cloned, over-expressed and purified as mentioned earlier [17]. Active site mutants C12S, T13L, C17S, C12/17S were prepared using two step PCR method in pET28a(+) and over-expressed as C-terminal His Tag, to nullify any possible influence of N-terminal His tag on the active site. Resuspension of cell pellets of WT and all the mutants were performed in lysis buffer (50 mM MOPS, pH 7.7, 300 mM NaCl, 5% glycerol, 2 mM PMSF containing ~0.5–1 mg/ml lysozyme and 50 mM Tris, pH 7.7, 300 mM NaCl, 10% glycerol during kinetic studies). Among the mutants C12S and C12/17S are prone to precipitation and thus purified by maintaining the temperature

at 283 K. The 6 × His tag was cleaved using restriction-grade thrombin (Novagen) and final purification of the protein excluding trace amount of contaminants, thrombin and cleaved 6 × His tag was achieved by gel filtration.

2.2. Size exclusion chromatography (SEC) and dynamic light scattering (DLS)

Oligomeric states of VcLMWTP-2 were analyzed by size exclusion chromatography on a ÄKTAPrime chromatographic system using Superdex 200 increase (GE-Healthcare) column (10 × 300 mm; bed volume ~24 ml) pre-equilibrated with lysis buffer (50 mM MOPS, pH 7.7, 300 mM NaCl, 3% Glycerol) running at 283 K with a flow rate of 0.4 ml min⁻¹. The elution profile was determined by monitoring the absorbance at 280 nm. Albumin (66.5 kDa), Ovalbumin (45 kDa), Chymotrypsin (25 kDa) and Ribonuclease A (13.7 kDa) were used as molecular mass standards.

DLS measurements, to determine the oligomeric states in solution, were performed on a Malvern Nano Series DLS spectrometer, equipped with thermostat chamber and 633 nm laser. Fractions eluted from the Superdex200 gel filtration column were directly used for this experiment. Before carrying out the DLS experiments all solutions were passed through 0.22 µm membrane filter (Millipore) and degassed. All measurements were conducted at least three times.

2.3. Crystallization, data collection and structure determination

For crystallization, thrombin-cleaved VcLMWTP-2 (in a buffer consisting of 50 mM MOPS pH 7.7, 300 mM NaCl, 5% glycerol) were concentrated upto 4.5 mg/ml using an Amicon ultracentrifugation unit (molecular-weight cutoff 10,000 Da). 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 [18] were used to explore the initial crystallization conditions. Typically, 2 µl protein solution was mixed with 2 µl precipitant solution, inverted over a reservoir containing 600 µl precipitant solution and maintained at both 277 and 293 K. Plate shaped crystals of VcLMWTP-2 appeared in 2.4 M ammonium sulfate, 0.1 M HEPES pH 7.0, 5% glycerol at 277 K, which is different from what we reported earlier [17].

Crystals of VcLMWTP-2 were looped out from the crystallization drops using a 20 µm nylon loop and flash-cooled in liquid nitrogen (Oxford Cryosystems) at 100 K. X-ray diffraction data was collected on PX BL-21 [19] beamline of Indus-2 synchrotron at RRCAT, INDORE to a resolution of 2.6 Å. Data were processed and scaled using *iMOSFLM* [20] and details of the data-collection and processing statistics are given in Table 1. Calculations of Matthews coefficient indicated the presence of two molecules of VcLMWTP-2 in the asymmetric unit with a solvent content 66%. Based upon sequence alignment between VcLMWTP-2 and VcLMWTP-1, a model was generated from the coordinates of VcLMWTP-1 (PDB: 4LRQ) using the program *Chainsaw* of CCP4 package [21]. This model was used to generate the initial solution by molecular replacement using *Phaser* [22], which yielded a solution with final TFZ = 10.5 and LLG = 389. Few cycles of model building with *Coot* [23] and refinement by Phenix refine [24] with gradual inclusion of the solvents, individual B-factors and TLS refinement produced a final R_{cryst} of 25.1% (R_{free} = 28.8%) (Table 1). The coordinates have been submitted with PDB code: 5Z3M.

2.4. Phylogenetic analysis

Phylogenetic trees were generated using the highly conserved LMWPTP catalytic domain sequences as indicated [25]. The sequence of VcLMWTP-2 was taken as seed with which to search the NR (UniProtKB 60% identity maximum + PDB) at the PRABI- GERLAND:

Table 1
Data collection statistics and refinement statistics.

Protein name	VcLMWPTP-2
Space group	C2
Cell dimension	
a, b, c	125.91, 51.80, 83.80
α, β, γ (°)	90.0, 103.60, 90.0
Molecule(s)/ASU	2
Mathews coefficient, V_m (Å ³ Da ⁻¹)	3.7
Resolution (Å) _h	19.4–2.6 (2.69–2.60)
Avg. Mosaicity	0.3
Rmerge	13.6 (53.5)
I / σ (I)	8 (3.0)
Completeness (%)	98.9 (99.8)
Redundancy	3.7 (3.8)
Refinement	
Resolution (Å)	19.4–2.6
No. of reflections/No. of unique reflections	31285/16209
Rcryst/Rfree	25.1/28.8
No. Atoms	2927
Protein	2433
Ligand	239
Water	255
Average B, all atoms (Å ²)	46.0
R.m.s.d. deviations	
Bond length (Å)	0.005
Bond angles (°)	1.061
Ramachandran statistics (%)	
Most favoured	90.03
Additionally allowed	9.63
Disallowed	0.33
PDB ID	5Z3M

* Values within parenthesis indicate highest resolution bin.

RHONE –ALPES BIOINFORMATIC POLE GERLAND SITE at NPS@ server. E threshold lesser or equal to $8e^{-15}$ were used in database searches. Uncharacterized proteins and proteins other than PTP were excluded. The remaining 124 sequences were analyzed and aligned by ClustalW and MULTALIN which prepares the multiple sequence alignment with hierarchical clustering [26,27] with blosum62 matrix, gap weight 12, gap length weight 2 and with minimal distance between sequences (in PAM) 20. The phylogenetic tree was generated with maximal number of clustals 12.

2.5. Structural analysis

Average B-factors for each residue were calculated using B average in CCP4 [28]. PISA webserver [29] was used for the analysis of the interfaces and assemblies of the structure. Sequence alignment of VcLMWPTP-2 with other LMWPTPs was done using Multalin [27] and PBIL [30]. Figures were prepared using Pymol (<http://www.pymol.org>).

2.6. Enzyme kinetics

Phosphatase activity of VcLMWPTP-2 (purified in 50 mM Tris, pH 7.7, 300 mM NaCl, 10% glycerol) were calculated at room temperature (298 K) using p-NPP as the substrate as described previously [31]. Briefly, p-NPP (1–40 mM) was treated with 10 μ M VcLMWPTP-2 (WT or its mutants C12S, T13L, C17S, C12/17S) and the reaction was quenched by the addition of 0.1 N NaOH after 10 min. The absorbance of the liberated product, p-nitrophenol, thus formed is measured at 405 nm. The amount of p-nitrophenol liberated was calculated from the standard curve prepared under identical conditions. 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. For baseline correction, the non enzymatic hydrolysis of substrates was measured using the control sample that did not contain VcLMWPTP-2. The hydrolysis rate was measured in triplicate for all substrates. Kinetic constants were

calculated using initial velocity data, which were fitted to the Michaelis-Menten curves and the Lineweaver-Burk plots.

2.7. Reactivation kinetics of hydrogen peroxide-inactivated VcLMWPTP-2 using β -ME as reducing agent

After several trials a minimum of 50 μ M H₂O₂ was found to be sufficient to inactivate VcLMWPTP-2 (26 μ M) completely. The H₂O₂ inactivation reaction was carried out in Tris buffer pH 7.0 at 298 K and allowed to proceed for 7 min at room temperature. Substrate p-NPP was added to the resulting H₂O₂ treated enzyme and the mixtures were allowed to incubate for 10 min before measuring the Phosphatase activity. For reactivation, 2-Mercaptoethanol (β -ME) was added (final concentration 30 mM) to the inactivated mixture and phosphatase activity was assayed after 40 min of β -ME treatment.

2.8. Molecular dynamics simulation

Molecular Dynamics (MD) simulation was carried out on a monomer (Residue 5–157, Chain A) of VcLMWPTP-2 protein where the MOPS molecule and the sulfate ions were removed from the crystal structure. The simulation was performed for 1000 ns using Gromacs-5.0.4 simulation package [32] with CHARMM27 [33] all atom force field and the water molecules were modeled explicitly using the TIP3P model. The protonation state of ionizable groups was appropriately chosen for pH 7.0. Suitable number of counter ions were added at random locations to neutralize the net charge of the system. The simulation box was a dodecahedron with minimum protein-edge distance of 12 Å and periodic boundary condition was applied on all the three (xyz) directions. After the energy minimization of the whole system using the steepest descent algorithm, the system was gradually heated to 300 K using NVT ensemble and later equilibrated using NPT ensemble. During the equilibrations, the protein backbone was restraint with a harmonic potential of force constant 1000 kJ/mol. The leap-frog integrator with a time-step of 2 fs was used. The Parrinello-Rahman algorithm [34] was employed to control the pressure at 1 bar with a coupling constant of 2 ps and the modified Berendsen (V-rescale) [35] thermostat was used to control the temperature of the system at 300 K with a time constant of 0.1 ps. The Particle Mesh Ewald (PME) [36] method was used to compute the electrostatic interactions with a real space cut-off distance of 12 Å. The same cut-off value was used for calculations of the van der Waals interactions. After 5 ns of equilibration using position restraints on the protein, the production MD simulation run was carried out for 1000 ns. The position and the velocity of all the atoms were recorded in the trajectory file at every 20 ps for analysis of the dynamics.

3. Results

3.1. Overall structure of VcLMWPTP-2 monomer

Crystal structure of VcLMWPTP-2 has been solved in space group C2 at 2.6 Å. Each asymmetric unit of the crystal contains two molecules of VcLMWPTP-2 together forming a dimeric unit. From the electron density map 157 residues out of 165 residues of each molecule could be located whereas last 9 residues have non interpretable density. Each VcLMWPTP-2 molecule binds a MOPS at its active site and one sulfate ion near the DPY loop (Fig. 1a). Two VcLMWPTP-2 molecules superpose with each other with an RMSD value of 0.45 Å whereas the same upon superposition with VcLMWPTP-1 is 0.67 Å. As expected, VcLMWPTP-1 and VcLMWPTP-2 differ mainly at their surface exposed loops although their P-loop superposes very well. The three dimensional structure of each VcLMWPTP-2 molecule is composed of four parallel β -strands [β 1 (Ser7 to Val11), β 2 (Gln38 to Glu43), β 3 (Phe85 to Ala88), and β 4 (Leu109 to Leu111)] sandwiched by five α -helices [α 1(Arg18 to Lys33), α 2(Asp54 to Gln62), α 3(Gln77 to Glu82), α 4(Arg91 to Leu98),

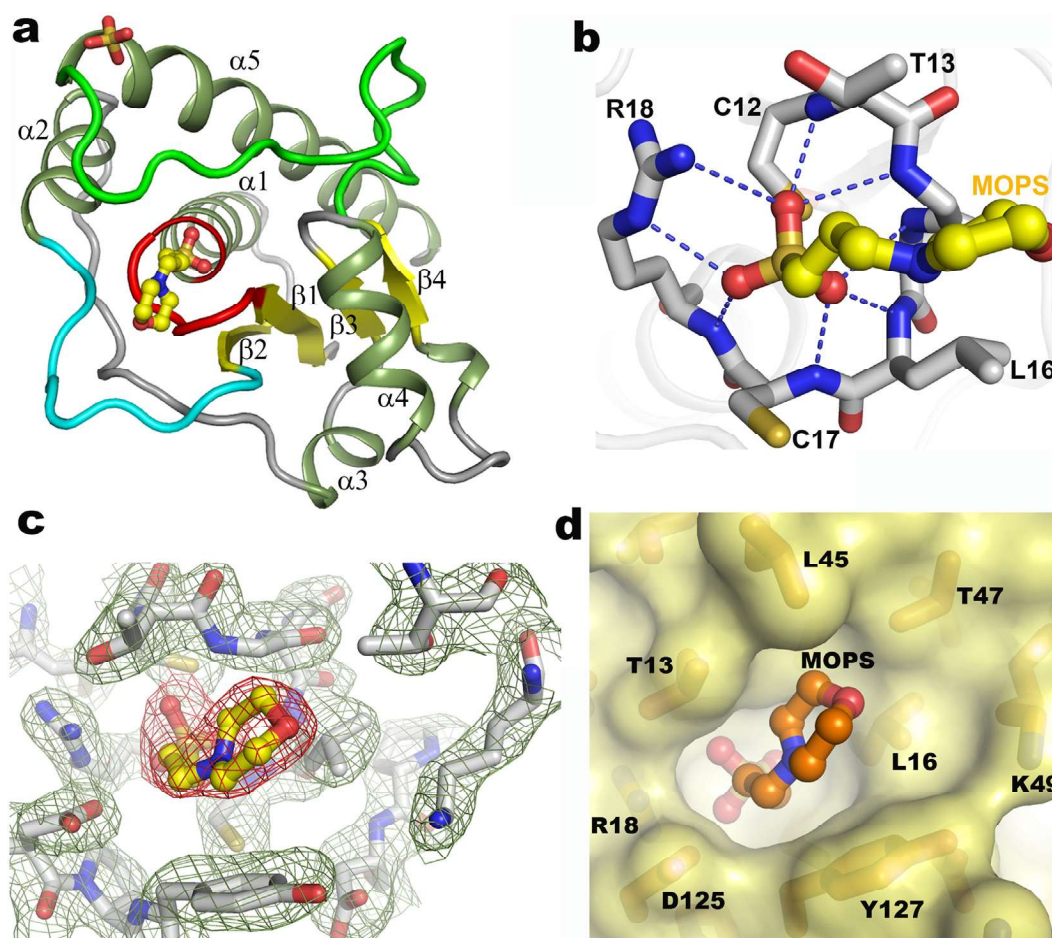


Fig. 1. Crystal structure of VcLMWPTP-2 and binding of MOPS at the active site. (a) Structure of VcLMWPTP-2 monomer shown in cartoon representation with MOPS at the active site shown in ball and sticks. α -helices and β -sheets are represented in olive and yellow color respectively. The catalytic P-loop is coloured in red while β 2- α 2 loop and DPY loop are shown in cyan and green respectively. (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 R18. (c) $2F_{\text{obs}} - F_{\text{calc}}$ electron density map around the active site of VcLMWPTP-2 (green mesh; contoured at 1.5σ , red mesh around MOPS contoured at 1σ). (d) Surface representation of MOPS (ball and stick) binding pocket; residues lining the active site pocket are shown in surface and sticks.

and α 5(Ser131 to Gln155)], which are arranged to form a Rossmann like fold typical of LMWPTPs [8,16,37,38] (Fig. 1a). The catalytic cysteine residue (Cys12) of VcLMWPTP-2 is part of the highly conserved motif C¹²TGNLCR¹⁸S which is located at the tip of β 1- α 1 loop. P-loop is surrounded by β 2- α 2 loop and β 4- α 5 loop. Amino acid sequence of these loops and their disposition define the depth and shape of the catalytic pocket. The β 4- α 5 loop (also called the DPY loop or simply D-loop) harbors an aspartic acid residue (Asp125 in VcLMWPTP-2) that acts as a general acid/base necessary for the dephosphorylating activity (Fig. 1a). Interestingly, in VcLMWPTP-2, the fourth residue of the motif is an arginine not a tyrosine, as in the case of *Escherichia coli* LMWPTP.

3.2. MOPS binds like a substrate mimetic at the active site

From the electron density map it is evident that each VcLMWPTP-2 tightly binds a MOPS molecule, a phosphotyrosine mimetic, at the active site. It forms several hydrogen bonds with the main chain amide nitrogens of the P-loop residues (Fig. 1b) and the side chain nitrogens of conserved Arg18. An electron density map around the active site of VcLMWPTP-2 molecule is shown in Fig. 1c. The MOPS molecule closely resembles the substrate phosphotyrosine as seen in PTP1B (PDB: 1G1F) [39] and is also analogous to the histidine and a phosphate, together mimicking a phosphotyrosine (PDB: 1YWF) [40]. The sulphonate group of MOPS mimics the phosphate group of the substrate phosphotyrosine as seen in PTP1B (PDB: 1G1F) [39]. The piperazine ring of MOPS

molecule is surrounded with several polar and aromatic residues (Thr44, Thr47, Thr50, Asp90, Asn93, Asp125, Tyr127, and Arg128). These residues define the architecture of the wall of the phosphotyrosine binding pocket of VcLMWPTP-2 with an opening at one side (Fig. 1d). It is to be noted that only one large hydrophobic residue (Tyr127) is present around the active site in VcLMWPTP-2 while in other LMWPTPs there are three or four Trp/Tyr/His residues. Lining of these residues define the architecture of the deep phosphotyrosine binding pocket. The orientation of S' of Cys12 and its proximity (~ 3.8 Å) with the S1-atom of MOPS (that corresponds to the P⁻ atom of Phosphotyrosine) indicates that it is in a feasible position for a S_N2 attack. Asp125 of DPY-loop is seen very close to the P-loop and its side-chain oxygen atoms are within 3.7 Å from the C1 atom of the MOPS molecule (corresponding to the phenolate O⁻ atom of phosphotyrosine). At this distance, Asp125 serves as an acid catalyst for the leaving alcoholic group. Therefore, MOPS bound state with the disposition of catalytic C12 and D125 resembles a substrate bound 'closed structure'. This closure at the P-loop is also evident from the structural alignment with an 'open structure' of apo-MptpA (PDB: 2LUO) [41] to the ligand-bound structures (PDB: 1U2P, 1U2Q) where the position of DPY-loop in the ligand-bound structure is shown to be closer to the P-loop [8].

3.3. VcLMWPTP-2 forms novel active dimer

Crystal structure of VcLMWPTP-2 indicates a dimeric form whereas

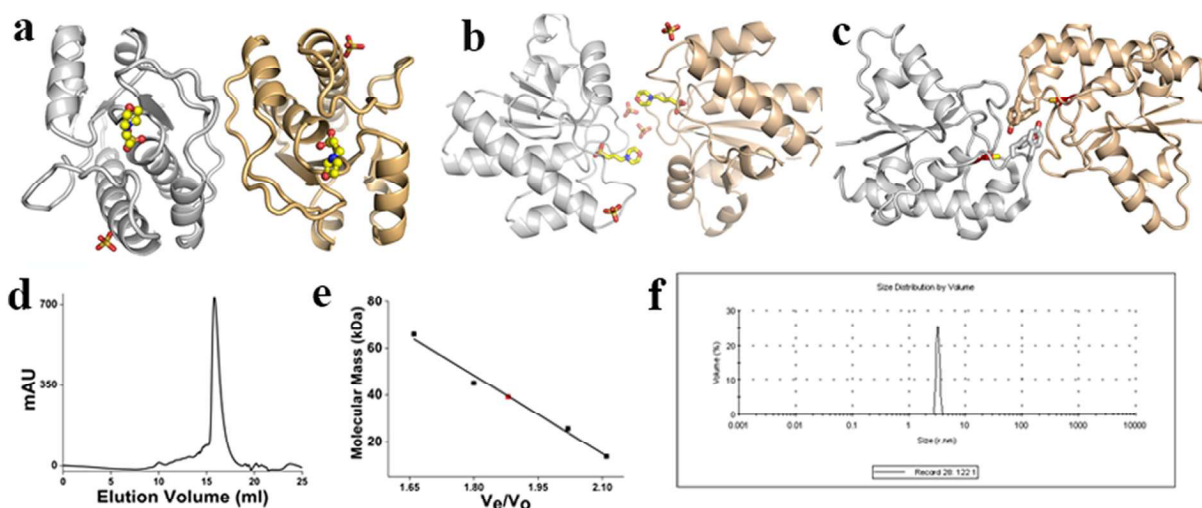


Fig. 2. Dimerization of VcLMWPTP-2. Cartoon representation of the dimeric (shown in gray and wheat) form of (a) Novel catalytically active VcLMWPTP-2 with bound MOPS in the active site (ball and stick) and sulfate ions (sticks), (b) catalytically active VcLMWPTP-1 with the bound MOPS molecule (yellow) and sulfate ions (in sticks) and (c) catalytically inactive BPTP where DPY loop of one molecule comes in close proximity to the active site of the other molecule resulting in the inactivation. (d) Elution profile of VcLMWPTP-2 plotted as mAU (absorbance at 280 nm) vs. elution volume (ml) depicts the existence of exclusive dimeric form of VcLMWPTP-2 in solution. (e) Standard curve [Molecular mass (in kDa) vs. V_e/V_o (elution volume/void volume)] is drawn based on the elution volumes of protein mixtures of known molecular mass (Albumin 66.5 kDa; Ovalbumin 45 kDa; Chymotrypsin 25 kDa and Ribonuclease A 13.7 kDa). Respective location of the protein (VcLMWPTP-2) in standard curve is shown in red. (f) VcLMWPTP-2 eluted by gel filtration and the fractions analyzed by dynamic light scattering (DLS). Volume weighted fit shows existence of dimer in solution.

the kinetic data exhibit strong phosphatase activity. The mode of dimerization seen in the crystal structure provide an explanation of its catalytic activity. Fig. 2a–c depicts the cartoon representation of the dimer of VcLMWPTP-2, VcLMWPTP-1 and BPTP respectively. In case of inactive BPTP dimer, the P-loop of one monomer is totally occupied by the residues of the YY loop of other [12]. While in VcLMWPTP-1 P-loop is not occluded by dimerisation but remain close the dimeric surface rendering the dimer catalytically active. Sulphate ions near the interfacial region probably playing a stitching role in dimer formation [16]. The mode of dimerization in VcLMWPTP-2 is quite different from VcLMWPTP-1 and BPTP. Here the active site is fully available to the substrate and remain far from the dimeric interface as reflected from its high phosphatase activity. A comparison of the buried surface area (BSA) indicates that VcLMWPTP-2 has much higher BSA (3100 \AA^2) than that of VcLMWPTP-1 (average BSA 2865 \AA^2) or BPTP (BSA 1589 \AA^2). At the interfacial region, eighteen residues from each monomer (Arg27, Gln38, Arg40, Lys49, Thr50, Met51, Pro52, Asp54, Leu57, Gln58, Gln61, His66, Pro67, Met68, Val69, Asn70, Pro71, Gln73) are engaged to stabilize the dimer through hydrophobic and H-bonding interactions. The set of residues involved in dimer formation in VcLMWPTP-2 is quite different from that of VcLMWPTP-1. The dimeric form is also well supported in solution. VcLMWPTP-2 elutes exclusively as a single peak at ~ 15.8 ml in SEC irrespective of the salt concentration and the elution volume indicate a higher oligomeric form. Molecular mass calculation, based on a standard curve, confirms that this oligomeric form is a dimer (Fig. 2d & e). To further confirm the oligomeric state, protein eluted from the FPLC column was used directly for DLS experiments. The intensity weighted fit shows presence of some aggregates which is probably due to the measurement with higher concentration of protein although volume weighted fit shows that maximum population exists as dimer in solution (Fig. 2f).

3.4. Multiple sequence alignments

Multiple sequence alignment, using a non redundant set of 124 LMWPTP sequences indicates several conserved patches (Fig. 3a). Among them, the P-loop and the DPY loop are highly conserved having a consensus sequence of $^{12}\text{C}(\text{X})_4\text{CRS}^{19}$ (motif-1) and $\text{DPY}(\text{Y}/\text{F})$ (motif-

4) [Fig. 3]. This is quite expected since residues in these loops are directly involved in the phosphatase activity. P-loop harbors the catalytic Cys12 residue whereas the Asp residue of the DPY loop serves as general acid-base during catalysis. The microenvironment of Cys12 viz. H-bonding with Ser19 and the helix dipole of $\alpha 1$ make Cys12 a better nucleophile. C8S (C12 in VcLMWPTP-2) mutation in VcLMWPTP-1 results in complete loss of activity implying that it is key catalytically residue [16]. Cys17 of the P-loop protects the catalytic Cys12 whereas Arg18 has a role in substrate binding. Besides these two highly conserved motifs, two less conserved motifs are evident from the multiple sequence alignment which are located around the P-loop like $\text{S}^{41}\text{X}(\text{A}/\text{G}/\text{R})\text{X}(\text{G}/\text{A})\text{X}(\text{T}/\text{V})$ (motif-2) and $\text{M}^{89}\text{D}(\text{X})_2\text{N}$ (motif-3) [Fig. 3]. Results of multiple sequence alignment when combined with the 3D superposition data of LMWPTP structures, determined from various organisms, illustrate the role of these four conserved motifs [Fig. 3b]. For example there is a network of hydrogen bonds involving conserved Asn15, Ser19, Cys12 and Ser41. This network not only provides proper conformation and required stability of the P-loop but along with the helix dipole of $\alpha 1$ also helps lowering down the pKa of Cys12. This has been shown in case of S19A mutant of BPTP which is reported to destabilize the thiolate [43]. Among the residues of $\text{M}^{89}\text{D}(\text{X})_2\text{N}$ of motif-3 D90 make a salt bridge with R18 which dictates the orientation of R18 for optimum substrate binding. For its strategic location and negative charge it is also capable to neutralize the helix dipole of $\alpha 4$ helix. $\text{D} \rightarrow \text{A}$ mutation at this position in BPTP has been reported to hamper the affinity of binding the substrate [44]. The side chain of Met89 is projected towards a hydrophobic pocket formed by various residues. It consistently takes a typical conformation in all LMWPTPs and runs in parallel with the hydrophobic part of Arg18 keeping Arg18 in a particular orientation for substrate binding. The conserved Asn93 of this motif (motif-3) make H-bonds with the side chain of Arg18. In some LMWPTPs residues in 13th position is Thr and in those cases Asn93 could potentially form H-bond with them. In our case mutation of this Thr \rightarrow Leu decreases the K_M value to 1.66 ± 0.1 mM. Structural superposition also indicates that $\beta 2$ - $\alpha 2$ loop, positioned semicircularly around the P-loop, exhibits highest structural variability. Sequence alignment also indicates a high degree of heterogeneity in this region. Strikingly, their P-loop superposes very well (Fig. 3b). Among the

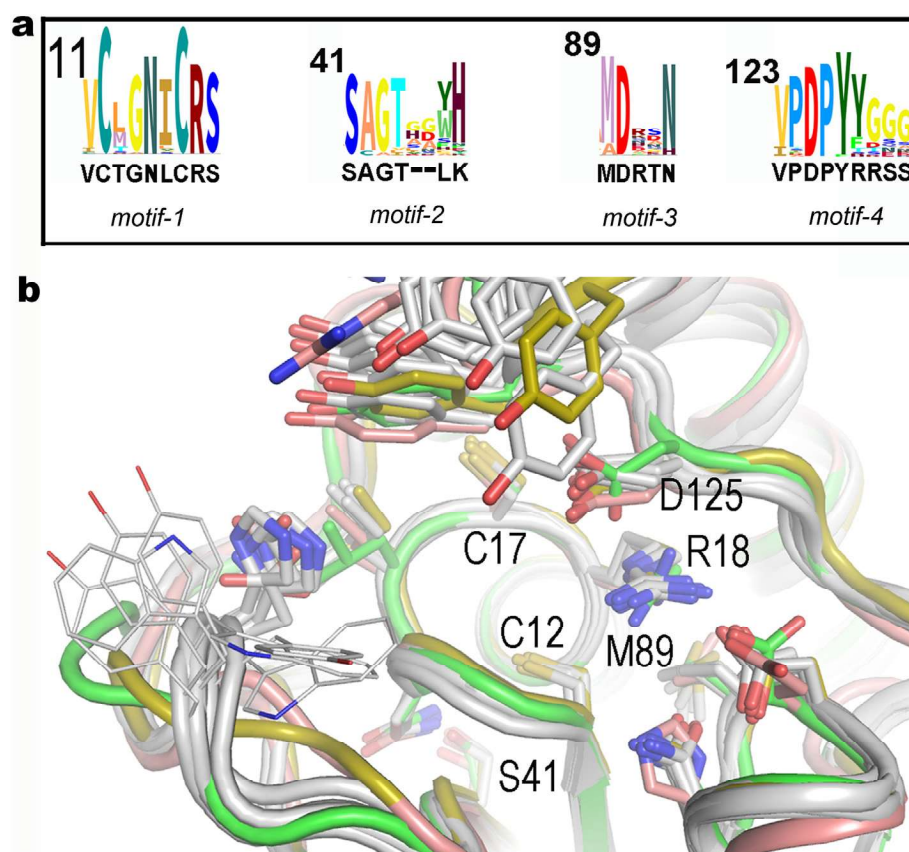


Fig. 3. Sequence and structural alignment of LMWPTP. (a) Four conserved motifs viz., C(X)₅RS (motif-1), SAGT (motif-2), MD(X)₂N (motif-3), and DPYY (motif-4) identified through multiple sequence alignment of LMWPTP sequences where VcLMWPTP-2 sequence has been shown in black color. Relative weightings of aligned sequence performed using Skyalign [45]. (b) Structural alignment of VcLMWPTP-2 (green; PDB ID: 5Z3M) with *Staphylococcus aureus* (olive; PDB ID: 3ROF), *Escherichia coli* (cherry; PDB ID: 2WJA) along with Human (PDB ID: 1XWW), *Thermos thermophilus* (PDB ID: 2CWD), VcLMWPTP-1 (PDB ID: 4LRQ), *Eentamoeba histolytica* (PDB ID: 3IDO), and *Bacillus subtilis* (PDB ID: 4ETM) shown in gray. Region near the P-loop is zoomed to show structural flexibility of DPY loop and β 2- α 2 loop.

Table 2
Area and volume of active site of LMWPTPs.

PDB	Area(Å ²) [‡]	Volume(Å ³) [‡]
VcLMWPTP-2	247.8	139.4
5jnv	280	196
4LRQ	545.4	284.9
4ETM	146.4	82.2
3ROF	334	232
3JVI	727	245
3IDO	421	202
2WJA	501	178
2CWD	784	270
1XWW	367	187

[‡] Calculation performed by CASTp server with probe radius 0.7 Å.

Table 3
Kinetic parameters of wild-type VcLMWPTP-2 and its mutants.

Enzyme	K _M (mM)	V _{max} (mM.min ⁻¹)	V _{max} /K _M (min ⁻¹)
VcLMWPTP-2 WT	1.41 ± 0.1	(8.5 ± 0.4) × 10 ⁻³	(6.03 ± 0.3) × 10 ⁻³
VcLMWPTP-2T13L	1.66 ± 0.1	(7.1 ± 0.3) × 10 ⁻³	(4.28 ± 0.1) × 10 ⁻³
VcLMWPTP-2 C12S	8.06 ± 0.6	(4.9 ± 0.5) × 10 ⁻³	(0.61 ± 0.04) × 10 ⁻³
VcLMWPTP-2 C17S	2.49 ± 0.1	(6.5 ± 0.3) × 10 ⁻³	(2.61 ± 0.1) × 10 ⁻³
VcLMWPTP-2 C12/17S	ND	ND	ND

LMWPTP structures the β 2- α 2 loop of VcLMWPTP-2, *S. aureus* (PDB: 3ROF) and *E. coli* (PDB: 2WJA) differ significantly from others. Except for these three cases, it has been observed that β 2- α 2 loop contain two large hydrophobic residues W/Y and H. For 3ROF (*S. aureus*) conserved His has been replaced with Asn and for 2WJA (*E. coli*) with Leu residue. Moreover, orientation of their loop β 2- α 2 loop is quite different either due to insertion/deletion or sequence differences. Along with β 2- α 2 loop, the wall of the deep phosphotyrosine binding pocket is lined up by

the two tyrosines of the DPY loop. The volume of the active site pocket was calculated for all these LMWPTP structures by CASTp server [46] and tabulated in Table 2. Table 2 indicates the volume of active site of VcLMWPTP-2 is 140 Å³, whereas the volume of most other cases is ~250 Å³. No aromatic residue in the β 2- α 2 loop and presence of a single Tyr residue instead of two in the DPY loop for VcLMWPTP-2 making the size of the cavity small. Small volume for 4ETM (*B. subtilis*) is probably due to different orientation of β 2- α 2 loop.

3.5. VcLMWPTP-2 display unique surface charge and grooves around the P-loop

VcLMWPTP-2 exhibits a distinctive surface electrostatic charge and grooves distribution around the active site which outplays others. Multiple sequence alignment and phylogenetic analysis also indicated the uniqueness of amino acid sequence of VcLMWPTP-2. Absence of several large hydrophobic residues around the active site makes the phosphotyrosine binding cavity shallow which is quite different from mammalian counterpart but somewhat similar to bacterial LMWPTPs. The amino acid composition around the active site of VcLMWPTP-2 provides uniquely different surface charge around the active site [Fig. 4]. Interestingly, VcLMWPTP-1(PDB: 4LRQ) resembles closely to LMWPTP from *E. histolytica* (PDB: 3IDO). But uniqueness of VcLMWPTP-2 (PDB: 5Z3M) lies not only in the shallowness of the active site but also at the sub sites that binds the residues either side of phosphotyrosine display quite different grooves and charges. The surface of 3ROF (*S. aureus*) is overlaid with a bound substrate mimicking peptide. The active site pTyr mimetic and other sub sites C-terminal to it are shown in space filling model [Fig. 4]. Comparison of surface charge and grooves of VcLMWPTP-2, especially at these sub sites, indicate the idiosyncratic nature. Although definite information about the sub site location/charge at the N-terminus side of the pTyr is not available, presence of an exclusive positive patch and unmatched

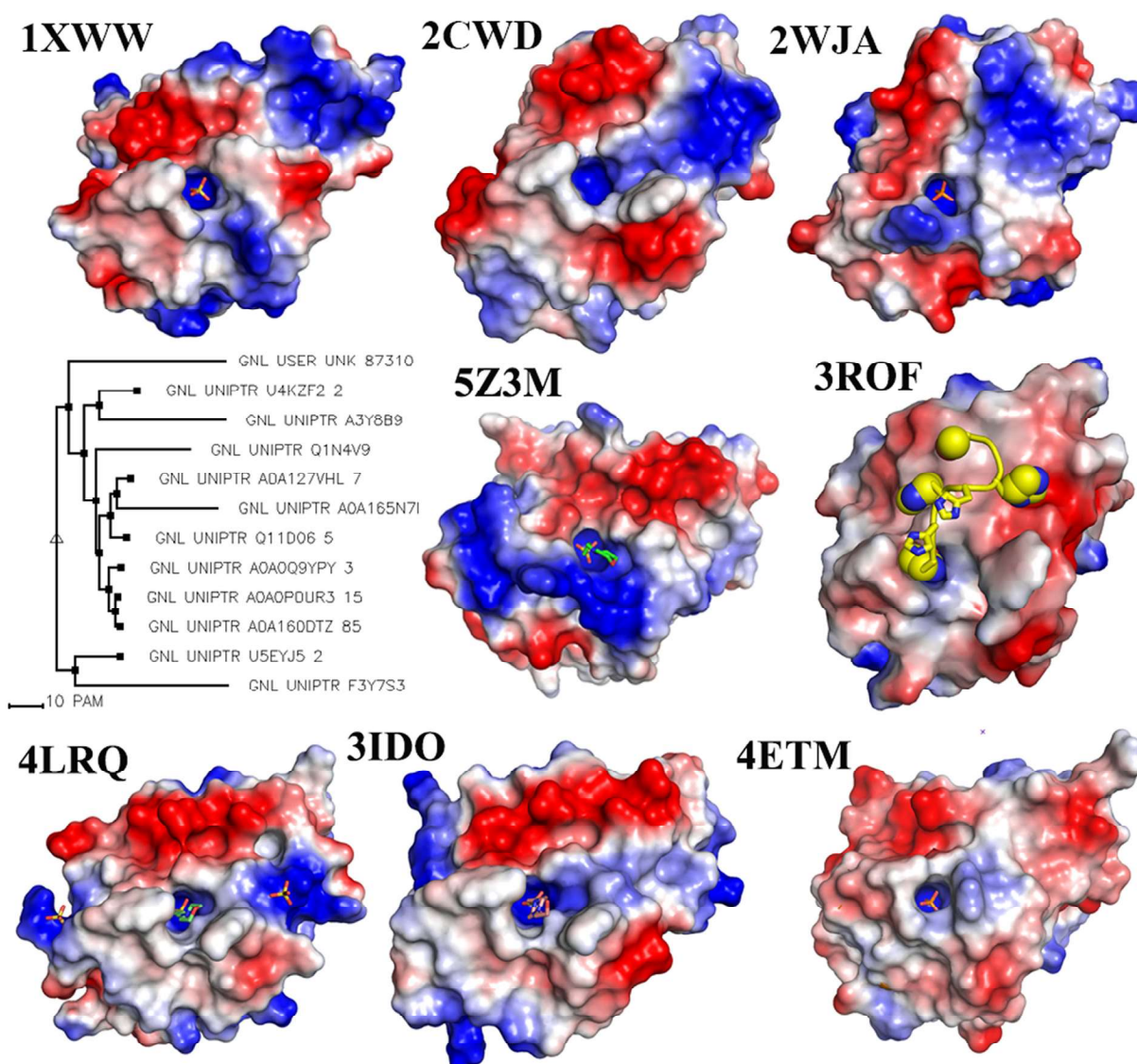


Fig. 4. Phylogenetic tree and comparison of the electrostatic surface charge distribution of VcLMWPTP-2 with homologous LMWPTPs (> 30% sequence identity) around the active site cleft. Comparison of electrostatic surface charge distribution between Human (PDB ID: 1XWW), *Thermophilus thermophilus* (PDB ID: 2CWD), *Escherichia coli* (PDB ID: 2WJA), VcLMWPTP-2 (PDB ID: 5Z3M), *Staphylococcus aureus* (PDB ID: 3ROF), VcLMWPTP-1 (PDB ID: 4LRQ), *Eentamoeba histolitica* (PDB ID: 3IDO), and *Bacillus subtilis* (PDB ID: 4ETM) have been represented. Phylogenetic analysis of 124 non redundant LMWPTP sequences shows VcLMWPTP-2 (Top; UNK 87310) lies as single taxon.

grooves at this region, definitely points towards a specific pTyr containing substrate for VcLMWPTP-2.

3.6. Role of C17 in VcLMWPTP-2: Is it protective or catalytic?

Phosphatase activity of VcLMWPTP-2 was determined by in vitro biochemical assays using p-NPP as a substrate. The K_M value determined for the WT protein is 1.41 ± 0.1 mM (Fig. 5a & b) which is comparable to that of other reported LMWPTPs like PtpA and PtpB [47] and VcLMWPTP-1 [16]. Complete loss of phosphatase activity has been observed in presence of vanadate ions indicating that it acts as an irreversible inhibitor, which is in agreement with the data reported by Zhang et al. [38].

To verify the role of two cysteine (Cys12 and adjacent Cys17) residing on P-loop of VcLMWPTP-2, catalytic inactivation by oxidizing agent followed by reactivation with reducing agent have been performed. In presence of H_2O_2 , the catalytic activity of wild type VcLMWPTP-2 diminished to basal level while addition of excess β -ME regains its activity almost completely upto $\sim 95\%$ (Fig. 5c), indicating

that Cys17 protects the catalytic Cys12 under oxidative stress as seen for other LMWPTPs [11,48,49]. Recovery after addition of β -ME for H_2O_2 mediated inactive protein (Fig. 5d) is $\sim 10\%$ for C12S mutant, and for C17S it is $< 2\%$ (Fig. 5d.) which further proves the participation of Cys17 to fortify catalytic cysteine. The rate of reactivation is however much slower than the rate of inactivation as excess β -ME and longer time of incubation is required to regain the activity. Mutation of catalytic Cys12 to Ser reduces the phosphatase activity to a great extent but it is not abrogated. The K_M value increases to 8.06 ± 0.6 mM compared to that of WT protein (1.41 ± 0.1 mM). This is in contrast with other LMWPTPs along with VcLMWPTP-1 [16,47]. Interestingly, C17S mutant shows $\sim 30\%$ decreases of activity (K_M 2.49 ± 0.1 mM) with respect to WT. Strikingly, a double mutant (C12/17S), where both the active site cysteine were mutated to serine, exhibits complete loss of activity.

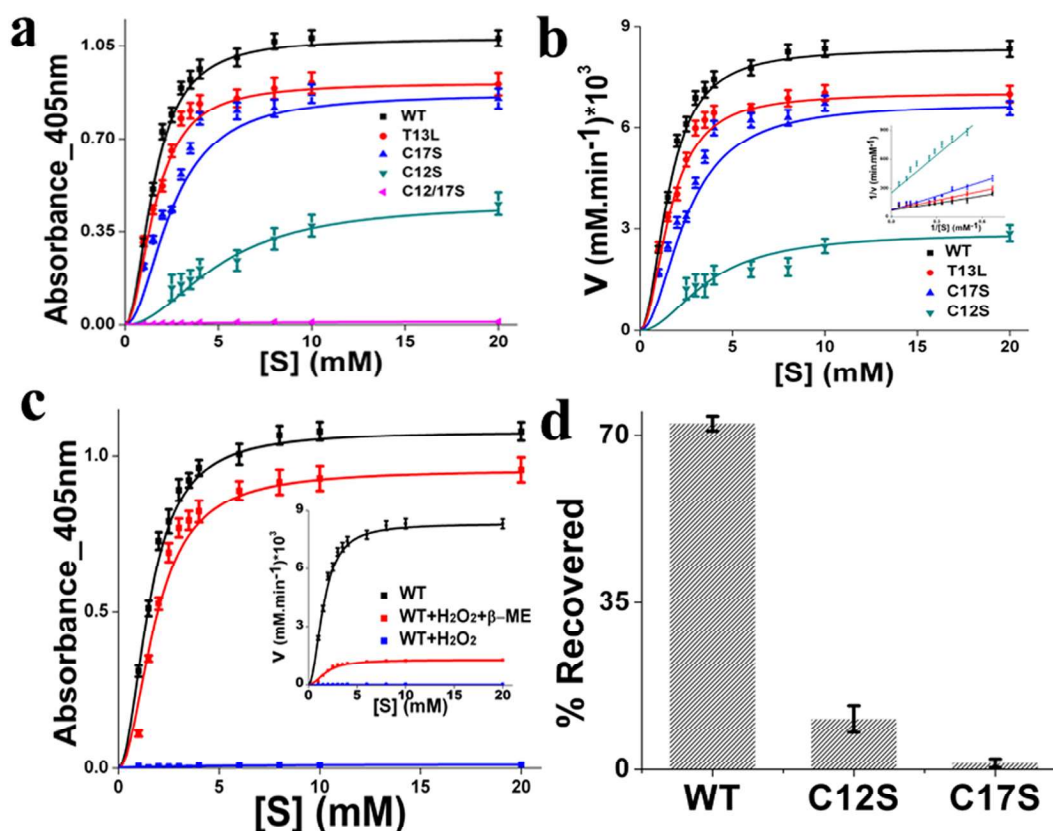


Fig. 5. Enzyme kinetics and redox regulation of VcLMWPTP-2 using p-NPP as a substrate. (a–b) Michaelis-Menten curves with Lineweaver-Burk plots (inset) of WT VcLMWPTP-2 and its various mutants (T13L, C17S, C12S). K_M , V_{max} and V_{max}/K_M values of WT and mutants are listed in Table 3. (c) Reactivation kinetics of H_2O_2 inactivated VcLMWPTP-2 using β -ME as reducing agent (Absorbance_{405nm} vs. [S] (mM)), along with $V/[S]$ Michaelis-Menten curves (inset) (d) Recovery by β -ME of H_2O_2 mediated inactive WT, C12S and C17S.

3.7. Molecular dynamics simulation reveal the role of residues around active site

To understand the dynamic nature of VcLMWPTP-2 active site we carried out MD simulations on ligand free VcLMWPTP-2 for 1000 ns. The correlation between atomic positional fluctuations in VcLMWPTP-2 has been investigated from (nanosecond) molecular dynamics (MD) simulations. VcLMWPTP-2 is mechanically stable as seen from their C_α RMSD values over the entire simulation time (Fig. 6a). The RMSF values, averaged over 1000 ns of MD simulation trajectories, showed that some regions of VcLMWPTP-2, especially the $\beta 2$ - $\alpha 2$ loop and DPY loop, show significant dynamics. Dynamic nature of these loops might be required for ligand binding/release (Fig. 6b). As expected, P-loop exhibits a rigid structure with minimum RMSF, due to intricate hydrogen bonding network, which is required to arrest the phosphotyrosine tightly (Fig. 6b). Consideration of dynamic nature of residues around the active site indicate that T13, N15, R18 (of motif 1) and D90, M89, N93 (of motif 4) exhibit minimum fluctuation (Fig. 6c) which is in corroboration with Fig. 3a. Strong hydrogen bonding network among them persists in almost all the snapshots. These H-bonding interactions along with hydrophobic interaction with M89 are sufficient to maintain the proper conformation of R18 for ligand binding. Interestingly, side chain of D125 assumes multiple conformations in the absence of ligand while in Fig. 3b it assumes similar conformation in all ligand bound structures. Thus, the conformation of D125 is ligand induced and it does not influence the conformation of R18. In the MOPS bound state one side of the active site wall is defined by Y127, T47 and K49 with a weak hydrogen bond between K49 and Y127 (Fig. 1c, d). In our MD simulation we see that most of the conformations of K49 and Y127 are away from the active site and the hydrogen bonding interaction between

them is abolished. Thus the resulting fluctuations in MD reflect the role of the conserved residues. While the residues of motif-1 and motif-3 together maintains the architecture of the active site. Large fluctuations of the residues of motif-2 and motif-4 and their movements away from active site, in apo form, reflect their role in substrate binding or release.

4. Discussion

Low molecular weight protein-tyrosine phosphatases (LMWPTPs), an enigmatic member of PTP family ubiquitously found across a spectrum of genera from prokaryotes to higher eukaryotes. They play specific and dominant role in setting up the cellular tyrosine phosphorylation level [50]. The dephosphorylation reaction catalyzed by LMWPTPs is primarily governed by their active site cysteine in the P-loop. Because of its microenvironment the pK_a value of catalytic cysteine lies within 4–6. Therefore, at physiological pH catalytic cysteine exists as thiolate ($R-S^-$) ion and acts as nucleophile [51]. The catalytic mechanism of cysteine-dependent LMWPTPs, proceed in two distinct steps— S_N2 displacement of tyrosine by the thiol (Cys12) that is followed by protonation of the tyrosinate through either Asp125 or a second conserved cysteine (Cys17) [4,52,53]. This creates an enzyme-bound thiophosphate ester intermediate and free tyrosine. In the second step, the thiophosphate ester is hydrolyzed by an incoming water molecule (activated by Asp125) in another S_N2 displacement, producing free phosphate and regenerating the thiol [44]. Switching between an “Open” and “Closed” conformation by ligand bound and unbound states respectively has been reported earlier [41]. Structural-superposition of VcLMWPTP-2 with an ‘open structure’ of apo-MtpA (PDB: 2LUO) and ‘closed structure’ of chloride- and glycerol-bound MtpA (PDB: 1U2Q) shows that MOPS bound VcLMWPTP-2 assumes a “closed”

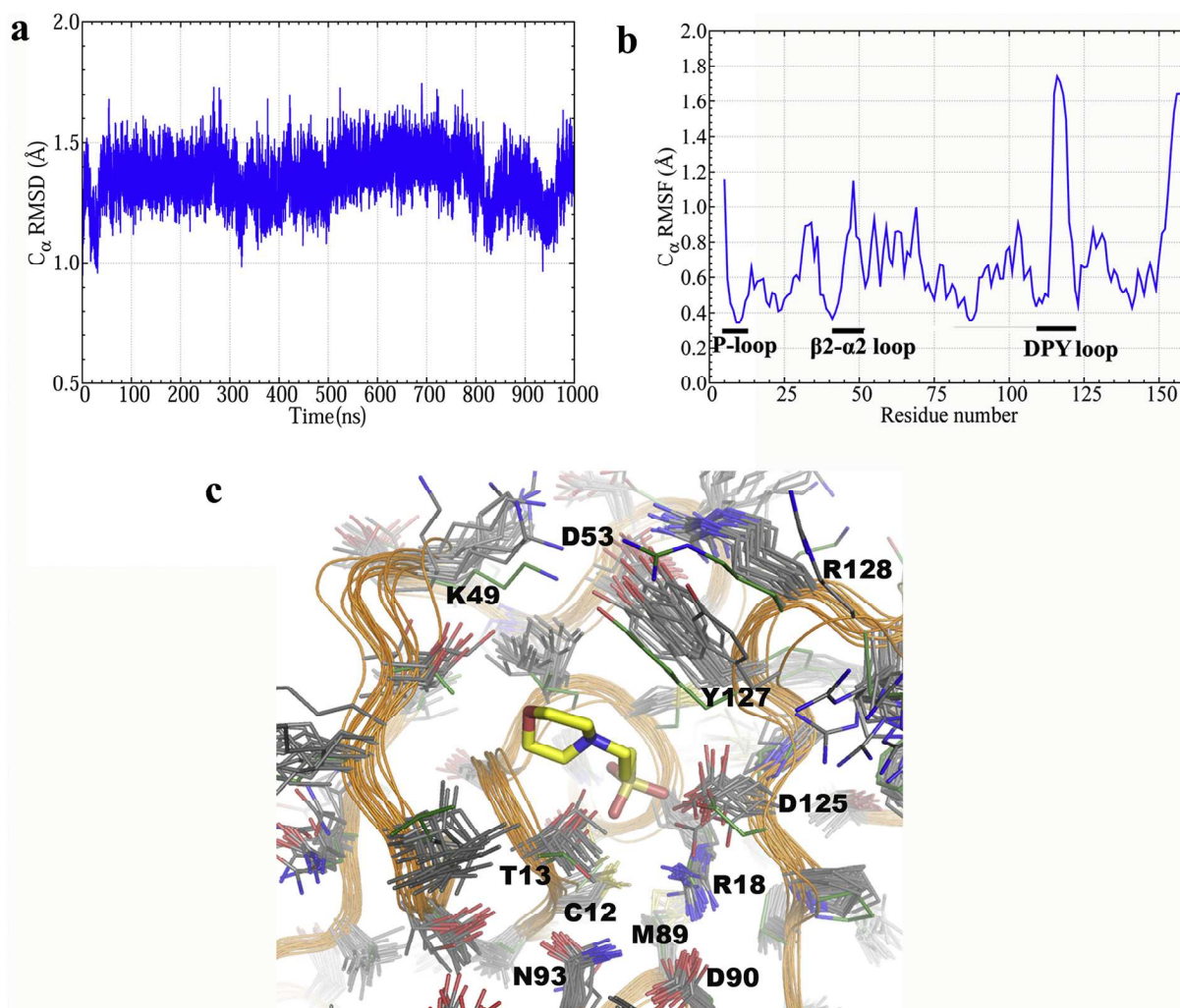


Fig. 6. Structural flexibility of VcLMWPTP-2. (a) Root Mean Square Deviation (RMSD) for the C-alpha atoms for the full simulation range. (b) Root Mean Square Fluctuation (RMSF) for the C-alpha atoms is shown as a function of residue number for the entire 1000 ns MD trajectory. β 2- α 2 loop and DPY loop show significantly higher flexibility while the P-loop exhibits a rigid structure with minimum RMSF. (c) Superposition of the snapshots of MD simulation trajectories to show the movement of residues around P-loop. Green sticks indicate the starting conformation while its MOPS molecule is shown to indicate the location of active site.

conformation although the surface of the substrate binding well is broadened by Thr47-Gly48-Lys49 residues forming an ‘open-mouthed closed’ substrate binding pocket. Our MD simulation data with apo VcLMWPTP-2 also indicate a movement of the DPY loop. In addition, we see that in the absence of a large hydrophobic residue in β 2- α 2 loop and a single such residue in DPY loop, compared to other LMWPTPs, hydrogen bond between K49 and Y127 may play a crucial role in substrate binding cavity formation.

It has been well characterized that LMWPTPs from both eukaryotic or prokaryotic sources produces inactive oligomers in solution that are in equilibrium with its active monomers [12,13]. The inactive oligomers thus formed are believed to act as supramolecular proenzymes. LMWPTP from *Bos taurus* (BPTP) becomes enzymatically inactive via dimerization in solution which is supported by crystal structure [12,54] (Fig. 2c). It also forms higher oligomers in crowded condition and thus activity depends on the affinity towards the substrate in a competitive manner with oligomerization [55,56]. Dimerization of BPTP shows the involvement of two tyrosine residues occluding the active site thereby forming an enzymatically inactive dimer. The dimerization of both prokaryotes and eukaryotes involve the active site residues and the docking models are also in full agreement with the crystallographic inactive dimers [13]. However, the equivalence between dimerization and inactivation is not applicable for VcLMWPTP-1 [16]. Crystal

structure of VcLMWPTP-2 indicates that dimerisation does not occlude the active site (Fig. 2a) rather bound substrate analog MOPS at the active site is an indicative of its accessibility towards phosphotyrosine containing substrates. Moreover, high PTPase activity determined by the kinetic data supports the availability of the active site for substrates.

Multiple sequence alignment shows the presence of threonine (13th position) instead of leucine in the P-loop unlike most of the LMWPTPs. Mutating threonine to leucine (T13L) reduces catalytic activity ($K_M \sim 1.66 \pm 0.1$ mM). The reason for the residual activity of C12S mutant is difficult to explain. Credible explanation comes from the strong hydrogen bonding capability of serine which introduces some additional H bonding interaction that may slightly alter the conformation of the active site in such a way that the substrate orientation becomes optimal. Cirri et al. described about the ability of binding of Cys17 to PTP-inhibitor which binds in the active site [57]. Probably in this case Cys17 takes a major role in its catalytic pathway through a nucleophilic approach to the substrate although this prediction requires further structural conformation.

In cells downstream of many surface receptors H_2O_2 is produced which acts as an oxidant of PTPs [58]. While the low pKa of Cys12 is beneficial for better nucleophilic attack to the phosphate group of phosphotyrosine containing substrates it renders Cys12 prone to oxidation. Even mild oxidation could convert the thiolate ion of active site

cysteine to sulphenic acid (R-SOH) which is sufficient to inactivate the enzyme although reversible conversion of inactive R-SOH to R-SH is still possible. Upon further oxidation cysteine is converted to sulphinic (SO_2H) or sulphonic acid (SO_3H) derivatives which are probably irreversibly formed terminal products that do not permit cellular reduction systems to operate. Phosphatases, e.g., PTEN, CDC25, PTP-1B, LMWPTP also contains a second cysteine residue in close proximity to the catalytic cysteine, and its role is to prevent the irreversible oxidation of the thiolate to sulfinic and sulfonic acid by forming an intramolecular disulfide bond upon oxidation to sulfinic acid [48]. In presence of thiols, GSH, DTT, and β -ME cysteine reactivation of PTPs becomes fully reversible although the rate of reactivation by thiols is 10–100 folds slower than that of inactivation [49]. All members of the LMWPTP subfamily have two vicinal cysteines in catalytic pocket (for VcLMWPTP-2 Cys12 and Cys17), capable to form an intramolecular S–S bridge [10]. Under mild oxidative stressed condition redox regulation through the formation of intra-molecular S–S bridge with the Cys17 is possible that protects the catalytic cysteine. The disulphide bond is, however, reversibly broken down under reducing condition.

Prokaryotic LMWPTPs are virulence factors that mimic eukaryotic phosphatases and dephosphorylate eukaryotic proteins, thereby interfering with the host defense response. An example is a phosphatase from *Mycobacterium tuberculosis* released into the extracellular medium and interfering with the host signaling pathways [59]. Sometime they work in tandem with the cognate kinase, like Wzb/Wzc pair, controlling the biosynthesis and transport of virulence factors, such as *exo*- and capsular polysaccharides [60–62]. It has been suggested that LMWPTPs are essential in certain processes, such as bacterial stress resistance [63]. All these facts make LMWPTPs an attractive drug target. For facultative human pathogen *Vibrio cholerae*, dissimilarity in surface charge distribution around the active site between two LMWPTPs from the same organism (VcLMWPTP-1 and VcLMWPTP-2) indicates firmly about their diverse specificity on substrates. While VcLMWPTP-1 resembles to a great extent with *E. histolytica* the distribution of surface charges and grooves around the active site of VcLMWPTP-2 delineates unique features. Structure of VcLMWPTP-2 provides an idea about the plausible nature of its true substrate would likely be which could be utilized to design specific inhibitor.

Acknowledgments

The laboratory of US is partly supported by the MSACR project, DAE, Govt. of India, SINP.

SC is a Senior Research Fellow of the Department of Atomic Energy, India.

Conflict of interest

There are no competing financial interests among the authors.

References

- [1] R. Maccari, R. Ottanà, Low molecular weight phosphotyrosine protein phosphatases as emerging targets for the design of novel therapeutic agents, *J. Med. Chem.* 55 (2012) 2–22.
- [2] G. Raugi, G. Ramponi, P. Chiarugi, Low molecular weight protein tyrosine phosphatases: small, but smart, *Cell. Mol. Life Sci.* 59 (2002) 941–949.
- [3] K. Guan, J.E. Dixon, Evidence for protein-tyrosine-phosphatase catalysis proceeding via a cysteine-phosphate intermediate, *J. Biol. Chem.* 266 (1991) 17026–17030.
- [4] J.P. Davis, 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.
- [5] A. Caselli, P. Paoli, A. Santi, C. Mugnaioni, A. Toti, G. Camici, P. Cirri, Low molecular weight protein tyrosine phosphatase: multifaceted functions of an evolutionarily conserved enzyme, *BBA* 1864 (2016) 1339–1355.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] N.K. Tonks, Protein tyrosine phosphatases: from genes, to function, to disease, *Nat. Rev. Mol. Cell Biol.* 7 (2006) 833–846.
- [10] Y.W. Lou, Y.Y. Chen, S.F. Hsu, R.K. Chen, C.L. Lee, K.H. Khoo, N.K. Tonks, T.C. Meng, Redox regulation of the protein tyrosine phosphatase PTP1B in cancer cells, *FEBS J.* 275 (2008) 69–88.
- [11] A. Caselli, R. Marzocchini, G. Camici, G. Manao, G. Moneti, G. Pieraccini, G. Ramponi, The inactivation mechanism of low molecular weight phosphotyrosine-protein phosphatase by H_2O_2 , *J. Biol. Chem.* 273 (1998) 32554–32560.
- [12] L. Tabernero, 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] C. Chin, J. Sorenson, J.B. Harris, W.P. Robins, R.C. Charles, R.R. Jean-Charles, J. Bullard, D.R. Webster, A. Kasarskis, P. Peluso, E.E. Paxinos, Y. Yamaichi, S.B. Calderwood, J.J. Mekalanos, E.E. Schadt, M.K. Waldor, The origin of the haitian cholera outbreak strain, *N. Engl. J. Med.* 364 (2011) 33–42.
- [15] C. Sombonwit, L.J. Menezes, D.A. Holt, J.T. Sinnott, P. Shapshak, Current views and challenges on clinical cholera, *Bioinformation* 13 (2017) 405–409.
- [16] S. Nath, R. Banerjee, U. Sen, 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, *BBRC* 450 (2014) 390–395.
- [17] 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.
- [18] J. Jancarik, S.H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, *J. Appl. Crystallogr.* 24 (1991) 409–411.
- [19] A. Kumar, B. Ghosh, H.K. Poswal, K.K. Pandey, M.V. Hosur Jagannath, A. Dwivedi, R.D. Makde, S.M. Sharma, Protein crystallography beamline (PX-BL21) at Indus-2 synchrotron, *J. Synchrotron Rad.* 23 (2016) 629–634.
- [20] T.G.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G.W. Leslie, iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr. D Biol. Crystallogr.* 67 (2011) 271–281.
- [21] E. Pottert, P. Briggs, M. Turkenburg, E. Dodson, A graphical user interface to the CCP4 program suite, *Acta Crystallogr. D Biol. Crystallogr.* 59 (2003) 1131–1137.
- [22] 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.
- [23] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.* 60 (2004) 2126–2132.
- [24] 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.
- [25] J.F. Huang, Different protein tyrosine phosphatase superfamilies resulting from different gene reading frames, *Mol. Biol. Evol.* 20 (2003) 815–820.
- [26] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [27] F. Corpet, Multiple sequence alignment with hierarchical clustering, *Nucleic Acids Res.* 16 (1988) 10881–10890.
- [28] 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. Pottert, 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.
- [29] 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.
- [30] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, NPS@: network protein sequence analysis, *TIBS* 25 (3 [291]) (2000) 147–150.
- [31] 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.
- [32] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1–2 (2015) 19–25.
- [33] N. Foloppe, A.D. MacKerell Jr., All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data, *J. Comp. Chem.* 21 (2) (2000) 86–104.
- [34] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.* 52 (12) (1981) 7182–7190.
- [35] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (1) (2007) 014101.
- [36] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an $N\log(N)$ method for Ewald sums in large systems, *J. Chem. Phys.* 98 (12) (1993) 10089–10092.

- [37] S. Wang, L. Tabernero, 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.
- [38] 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.
- [39] A. Salmee, 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.
- [40] C. Grundner, H.L. Ng, T. Alber, Mycobacterium tuberculosis protein tyrosine phosphatase PtpB Structure reveals a diverged fold and a buried active site, *Structure* 13 (2005) 1625–1634.
- [41] T. Stehle, S. Sreeramulu, F. Löhr, C. Richter, K. Saxena, H. R. A. 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), pp. 34569–34582.
- [42] T. Hansson, P. Nordlund, J. Åqvist, Energetics of nucleophile activation in a protein tyrosine phosphatase, *J. Mol. Biol.* 265 (1997) 118–127.
- [43] Z. Zhang, E. Harms, R.L. Van Etten, Asp129 of low molecular weight protein tyrosine phosphatase is involved in leaving group protonation, *J. Biol. Chem.* 269 (1994) 25947–25950.
- [44] J. Clements, T. Wheeler, R. Finn, Licensed under a Creative Commons Attribution 3.0 Unported License and Web Server, <http://skyline.org>.
- [45] 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), p. 34.
- [46] D. Soulat, E. Vaganay, B. Duclos, A.L. Genestier, J. Etienne, A.J. Cozzone, *Staphylococcus aureus* contains two low-molecular-mass phosphotyrosine protein phosphatases, *J. Bacteriol.* 184 (2002) 5194–5199.
- [47] S.H. Choa, C.H. Leea, Y. Ahna, H. Kima, H. Kimb, C.Y. Ahnb, K.S. Yanga, S.R. Lee, Redox regulation of PTEN and protein tyrosine phosphatases in H₂O₂-mediated cell signaling, *FEBS Lett.* 560 (2004) 7–13.
- [48] J.M. Denu, K.G. Tanner, Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation, *Biochemistry* Vol. 37 (1998) 5633–5642.
- [49] R. He, Z. Yu, R. Zhang, Z. Zhang, Protein tyrosine phosphatases as potential therapeutic targets, *Acta Pharmacol. Sin.* 35 (2014) 1227–1246.
- [50] G.H. Peters, T.M. Frimurer, O.H. Olsen, Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases, *Biochemistry* 37 (1998) 5383–5393.
- [51] G. Ramponi, M. Stefani, Structure and function of the low Mr phosphotyrosine protein phosphatases, *BBA* 1341 (1997) 137–156.
- [52] N. Taddei, P. Chiarugi, P. Cirri, T. Fiaschi, M. Stefani, G. Camici, G. Rauegi, G. Ramponi, Aspartic-129 is an essential residue in the catalytic mechanism of the low Mr phosphotyrosine protein phosphatase, *FEBS Lett.* 350 (1994) 328–332.
- [53] T. Akerud, E. Thulin, R.L. Van Etten, M. Akke, Intramolecular dynamics of low molecular weight protein tyrosine phosphatase in monomer–dimer equilibrium studied by NMR: a model for changes in dynamics upon target binding, *J. Mol. Biol.* 322 (2002) 137–152.
- [54] A.P. Minton, Influence of excluded volume upon macromolecular structure and associations in ‘crowded’ media, *Curr. Opin. Biotechnol.* 8 (1997) 65–69.
- [55] P. Bernado, T. Åkerud, J. G. Torre, M. Akke, M. Pons, Combined use of NMR relaxation measurements and hydrodynamic calculations to study protein association: evidence for tetramers of low molecular weight protein tyrosine phosphatase in solution, *J. Am. Chem. Soc.* 125 (2003) 916–923.
- [56] P. Cirri, P. Chiarugi, G. Camici, G. Manao, L. Pazzagli, A. Caselli, I. Barghini, G. Cappugi, G. Rauegi, G. Ramponi, The role of Cys-17 in the pyridoxal 5'-phosphate inhibition of the bovine liver low M(r) phosphotyrosine protein phosphatase, *Biochim. Biophys. Acta* 1161 (1993) 216–222.
- [57] S.G. Rhee, Y.S. Bae, S.R. Lee, J. Kwon, Hydrogen peroxide: a key messenger that modulates protein phosphorylation through cysteine oxidation, *Sci. STKE* 2000 (2000) pe1.
- [58] A. Koul, A. Choidas, M. Treder, A.K. Tyagi, K. Drlica, Y. Singh, A. Ullrich, Cloning and characterization of secretory tyrosine phosphatases of *Mycobacterium tuberculosis*, *J. Bacteriol.* 182 (2000) 5425–5432.
- [59] C. Vincent, P. Doublet, P. Grangeasse, E. Vaganay, A.J. Cozzone, B. Duclos, Cells of *Escherichia coli* contain a protein-tyrosine kinase, Wzc, and a phosphotyrosine-protein phosphatase, Wzb, *J. Bacteriol.* 181 (1999) 3472–3477.
- [60] P. Bugert, K. Geider, Characterization of the amsl gene product as a low molecular weight acid phosphatase controlling exopolysaccharide synthesis of *Erwinia amylovora*, *FEBS Lett.* 400 (1997) 252–256.
- [61] O. Ilan, Y. Bloch, G. Frankel, H. Ullrich, K. Geider, I. Rosenshine, Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide, *EMBO J.* 18 (1999) 3241–3248.
- [62] J. Kirstein, K. Turgay, A new tyrosine phosphorylation mechanism involved in signal transduction in *Bacillus subtilis*, *J. Mol. Microbiol. Biotechnol.* 9 (2005) 182–188.