University of Florence

International Doctorate in Structural Biology

Cycle XXI (2006-2008)

Metal-Based Drug Screening and Design in Metallophosphatases

Ph. D. thesis of

Eleonora Talluri

Tutor

Prof. Ivano Bertini

Coordinator

Prof. Claudio Luchinat

S.S.D. CHIM/03

This thesis has been approved by the University of Florence,
the University of Frankfurt and the Utrecht University

Contents

1

INTRODUCTION

1.1 Role of Protein Phosphatases in biological systems

The reversible phosphorylation of proteins catalyzed by protein kinases and protein phosphatases (Fig.1) determines the biological activities of many proteins and it is recognized as a major mechanism controlling cellular processes in both prokaryotes and eukaryotes[1,2].

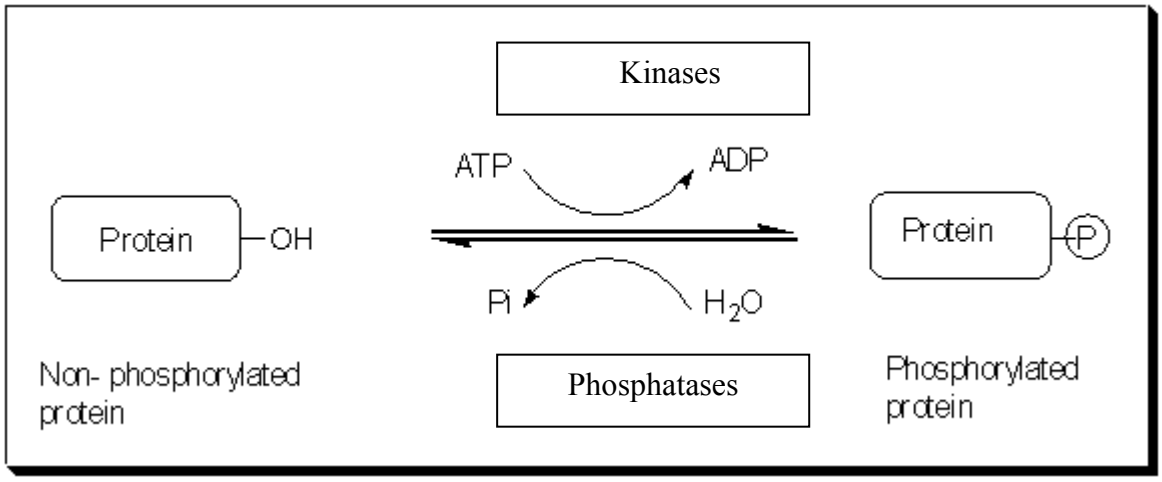

Fig.1: Reversible phosphorylation of proteins.

In eukaryotes, during protein phosphorylation, phosphate moieties are transferred to the serine, threonine or tyrosine residues of proteins from adenosine triphosphate molecules by protein kinases, and are hydrolysed by protein phosphatases[2]. In addition, phosphorylation occurs on the basic amino acid residues histidine or arginine or lysine in prokaryotic proteins [3].

Many dephosphorylation processes can result in changes in polypeptide folding, enzymatic activity and protein–protein interactions and they are proved to occur very frequently in both intra- and extracellular processes. Phopshatases are involved in:

- signal transduction cascades and the general effect of dephosphorylation, in many cases, is to provide an 'on-off switch' for signalling pathways [4-6]**;**
- signalling in the nucleus. The importance of phosphorylation in this highly dynamic cellular compartment is related to its dominant regulatory role in processes as diverse as DNA replication and repair, chromosome condensation, ribosome biogenesis and chromatin remodelling[7-9];
- exocytosis [10,11];
- cell-matrix adhesion and migration [12].

Many other specific biological roles of phosphatases are reported in Tab.1 where the most recent classification of these enzyme is shown [8].

1.2 Protein Phosphatases Gene Families: a focus on Human Protein Serine/Threonine Phopshatase PP5 , PP2A and PP6

According to the last evolutionary studies, human protein phosphatases can be classified into three groups on the basis of sequence, structure and catalytic mechanism[13]. A recent classification of human phosphatases is reported in Tab.1.

Tab.1: Main features and classification of Human Protein Phosphatases (Reprinted from Kerk *et al.*[13])

Family	*Class*	*Regulatory Subunits*	*Example of function and/or (substrate)*
Ser/Thr phosphatases			
PPP family	PP1	Yes	Chromosome condensation
	PP2A	Yes	Chromatid cohesion
	PP4	Yes	DNA repair
	PP5	None	Cellular stress
	PP6	Yes	NF_kB pathway
	PP2B	Yes	Immune response
	PP7	Unknown	
PPM family	PPC	None	TGFß signalling
PTP superfamily			
Class I PTPs	Receptor PTP		Cell adhesion/cytoskeletal
	Non-receptor PTP		Insulin signalling (insulin receptor)
Class I PTPs (DSPs)	MAPKP		MAPK (MAPK)
	Slingshots		Actin dynamics (cofilin)
	PRLs		Unknown
	Atypical DSP		Mostly unknown (mRNA)
	CDC14		Cytokinesis, mitotic exit
	PTEN		PIP phosphatase
	Myotubularins		Ptdlns3P, Ptdlns(3,5)p phosphatase
Class II PTPs	CDC25s		Promotes mitosis
Class III PTPs	LMWPTP		Unknown
Asp-based catalysis			
FCP/SCP family	FCP1	Yes	Transcription (pol II)
	SCP		TGFß signalling (SMADs)
	FCP/SCP-like		Unknown
HAD family			Actin dynamics (cofilin)

The first group comprises the classic Ser/Thr phosphatases: the large phosphoprotein phosphatase (PPP) family (PP1, PP2A, PP2B, PP4, PP5, PP6 and PP7) and the protein phosphatase, Mg^{2+} or Mn^{2+} dependent (PPM) family (PP2C). They catalyze the dephosphorylation of serine and threonine residues. The protein Tyr phosphatase (PTP) superfamily forms the second group and they dephosphorylate the tyrosine residues of their substrates. The third group consists of the Asp-based protein phosphatases with a DXDXT/V catalytic signature. In general, genomics studies showed that there are ~147

human protein phosphatases. In human phosphoproteome, proteins are phosphorylated predominantly on Ser, Thr and Tyr residues, with each accounting for approximately 86.4, 11.8 and 1.8%, respectively [7].

Although over 98% of protein-phosphorylation occurs on Ser and Thr side chains, only ~ 40 of the 147 known protein phosphatases are specific for these residues [8]. The fact that also the total number of phosphatases is significantly lower than the number of kinases implies that some phosphatases have more than one function[12]. Dephosphorylation brought about by protein Ser/Thr phosphatases represents a relevant cellular process and investigations aimed at understanding phosphatase catalysis are needed.

In mammals, PPP family consists of multiple high homologies isoforms with peculiar specificities. In particular, there are three isoforms of PP1 (identity >90%), two isoforms of PP2A (identity >97%), and three isoforms of PP2B (identity >80%). All but one, PP7, are ubiquitously expressed in the organism [14].

PP2A, PP5, and PP6 are largely expressed in humans but their specific biological functions are still poorly characterized although different studies show their involvement in cell cycle regulation, apoptosis and stress response [15-18].

The members of the PPP family show common structural features and distinctive characteristics. In particular, all PPP proteins share a catalytic domain containing two metal ions. At the same time, regulatory subunits that restrict and control the catalytic activity and regulate the substrate specificity have been identified in PP1, PP2A, PP4, PP2B and PP6[19 ,20, 21-23]. A short N-terminal section TPR domain with autoinhibitory function distinguish PP5 from the other members. Moreover, a C-terminal region that includes an unusual nuclear localization sequence posses an autoinhibitory function, too[24]. The catalytic domain of PP5 (PP5c) shares 35-45% sequence identity with the catalytic domain of other PPP phosphatases[25].

For all serine/threonine phosphatase, a common catalytic mechanism of metal ion-mediated hydrolysis has been proposed[27].

1.3 Catalytic mechanism

As all the phosphatases belonging to PPP family, PP5 is a metalloprotein in which highly conserved three motifs (GDXHG, GDXXDGR and GNH(E/D) in the N-terminal subdomain of catalytic domain are present[26]. These motifs contain most residues which coordinate metal ions in the active centre and are considered as the signature of the phosphatase family[27].

X-ray crystal structure of other related phosphatases (Protein phosphatase 1 and Protein phosphatase 2B) revealed in their catalytic subunit a metal-bound water molecule acting as a nucleophile to attack the phosphorus atom of a phosphate group in a S_N2 mechanism. Metals assist in this catalysis as Lewis acids. Protein phosphatase 2C too contains binuclear metal ions (2 Mn^{2+}) in its catalytic site. It promotes protein dephosphorylation by activation of water molecules to catalyse hydrolysis in a single-step reaction [27,28]. The catalytic mechanism of PP5 has been proposed by Swingle *et al.* (Fig.3)[25] . Precise positioning of two metal ions within a conserved Asp^{271}-$Metal_1$:$Metal_2$-Water-His^{427}-His^{274}-Asp^{274} catalytic motif is requested [25,27].

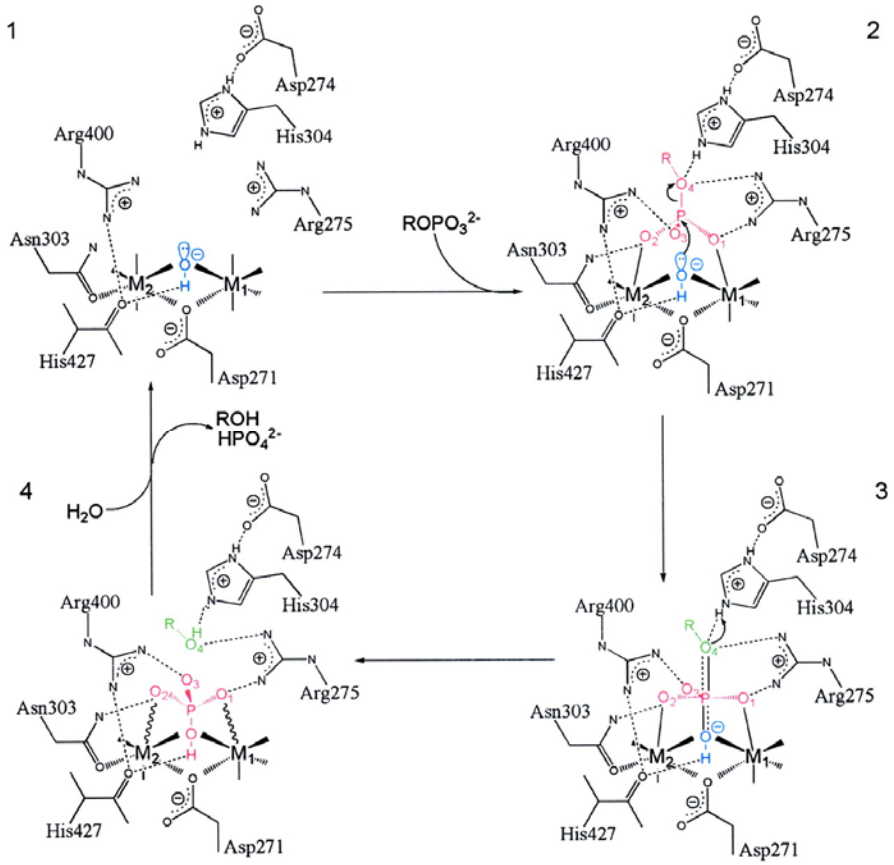

Fig.3: Catalytic activity metal ion-mediated hydrolysis of PO3 derived from the crystal structure of PP5c. The attacking hydroxide W1 is shown in blue, and the leaving group of the substrate is in green. The substrate, the planar PO3 moiety of the transition state, and the phosphate product are all shown in red. Solid lines to the metal ions denote metal-ligand bonds, and solid or dashed wedges indicate metal-ligand bonds directed above or below the plane of the page, respectively. Wavy lines to the metal ions indicate strained contacts with poor coordination geometry. Dotted lines indicate hydrogen bonds, and the nearly dissociated axial bonds in the transition state are shown by half-dotted double line (Reprinted from Swingle *et al.*[25]).

Despite these similarities in sequence, each phosphatase in the PPP family appears to have different metal ion requirements[29]. The dinuclear metal site is situated in a unique protein fold, a β-α-β-α-β motif which provides the majority of ligands to the metal ions. A similar

fold is also seen in plant purple acid phosphatases where the two metal ions are bridged by a solvent molecule and a carboxylate group from an aspartic acid residue, juxtaposing the two metal ions to within 3.0–4.0Å of each other. A similar motif has been identified in a number of other enzymes which exhibit phosphoesterase activity, implicating that several of them are metalloenzymes which contain dinuclear metal ion cofactors[30]. However, the *in vivo* identities of relevant metal ions in most PPP family phosphatases have not been completely resolved except for PP2B, PP1 and PP2A for which a dinuclear center containing respectively Fe^{2+} /Zn^{2+}, Mn^{2+}/Fe^{2+} and Mn^{2+}/Mn^{2+} was identified[31-33] . A bimetallic center cotaining Mn^{2+}/Mn^{2+} has been identified also in the bacteriophage λ protein phosphatase (PP λ)[31].

Manganese is required in the formation of an active metalloenzyme form of PP1 but its treatment with a combination of iron and zinc lead to a more active PP1[34]. Because the catalytic domains of PPP members are so highly conserved, PP2B and PP1 has been recently used as a model system for studying the PPP family [35] as well as the PP λ that was found to represent the minimal catalytic core of PP5 and PP1[36]. In this PhD project further structural and biochemical investigations have been focused on the catalytic domain of PP5.

1.4 PPPs as pharmaceutical targets

The catalytic activity of protein phosphatases is crucial for cellular life. Dysregulated phosphorylation/dephosphorylation in humans underlies numerous diseases [10,37]. Embryogenesis, angiogenesis, nerve growth, tissue repair, inflammation, atherosclerosis, arthritis, osteoporosis, invasion and metastasis are all biological processes that require cell attachment and migration in which many phosphatases are involved. Abnormal phosphorylation has been associated with excitotoxicity, in response to perturbed calcium homeostasis. Neuronal loss due to excitotoxicity may contribute to the pathophysiology of chronic neurodegenerative illnesses such as Alzheimer's disease, amyotrophic lateral sclerosis, Parkinson's disease [10].

Therefore, the targeting of phosphatase activity might represent a new approach to treat specific diseases and to study PPP function. In the case of holoenzyme, the association with multiple regulatory subunits is considered the basis for achieving diversity of biological functions and modulation of enzymatic activity. Therefore, both the catalytic domain and the regulatory subunits represent a possible pharmacological targets.

During this PhD project, among the members of PPP family, PP2A, PP6 and PP5 have been cloned and expressed. While for PP5 soluble constructs of the catalytic domain with and without the autoinhibitory domain have been successfully expressed, purified and characterized, for PP2A and PP6 the research activity is still in progress, due to the difficulties in obtaining soluble forms of the enzyme in *E.coli* and in *cell-free* system.

PP5

Protein phosphatise 5 has been investigated and structurally characterized by Yang and Swingle [25,28] that have revealed important aspects of the catalytic mechanism and enzyme regulation. PP5 is constituted by an N- terminal TPR domain, by a catalytic domain and by a C-terminal subdomain. TPR domain acts as an interface for protein-protein interactions in several proteins included PP5 [38,40]. The best-known binding partners of PP5 is the Heat Shock Protein 90 (HSP90) that binds PP5 throught the EEVD sequence[41,42].

In PP5, the TPR-domain has an additional function, e.g. to keep the enzyme inactive in concert with the C-terminal domain. Actually, the interaction of TPR with the catalytic domain restricts the access to the catalytic site. Due to this, PP5, unlike other phosphatases, displays low-phosphatase activity under normal conditions, but it can be activated when the TPR domain is displaced or truncated from PP5 [43]. A model for the *in vivo* activation and regulation of PP5 has been proposed by Zeke in 2006 [44] and is shown in Fig. 2.

PP5 protein can be activated by glucocorticoid receptor–Hsp90 complexes, by other Hsp90 complexes or by the interaction of TPR domain with other ligands and proteins[41,43,45,46]. It is expressed in all mammalian tissues, with high levels in brain and neurons. Within the cell, this enzyme is also broadly distributed. PP5 is involved in several phosphorylation cascades including MAPK-mediated growth and differentiation, cell cycle arrest and DNA damage repair via the p53 and ATM/ATR pathways, regulation of ion channels via the membrane receptor for atrial natriuretic peptide, the cellular heat shock response as mediated by heat shock transcription factor, and steroid receptor signaling, especially glucocorticoid receptor (GR) [47].

It has been recently suggested a potentially important regulator role of PP5 in human breast cancer [48-50] and in neurodegenerative disease like Alzheimer's disease, where the dephosphorylation /hyperphosphorylation of Tau protein play a crucial role in intracellular neurofibrillary tangles[51,52]. Therefore, PP5 is a promising therapeutic target in the treatment of these diseases. In order to interfere with the biological activity of the enzyme, both the catalytic site and the regulation mechanism can be targeted by designing specific molecules.

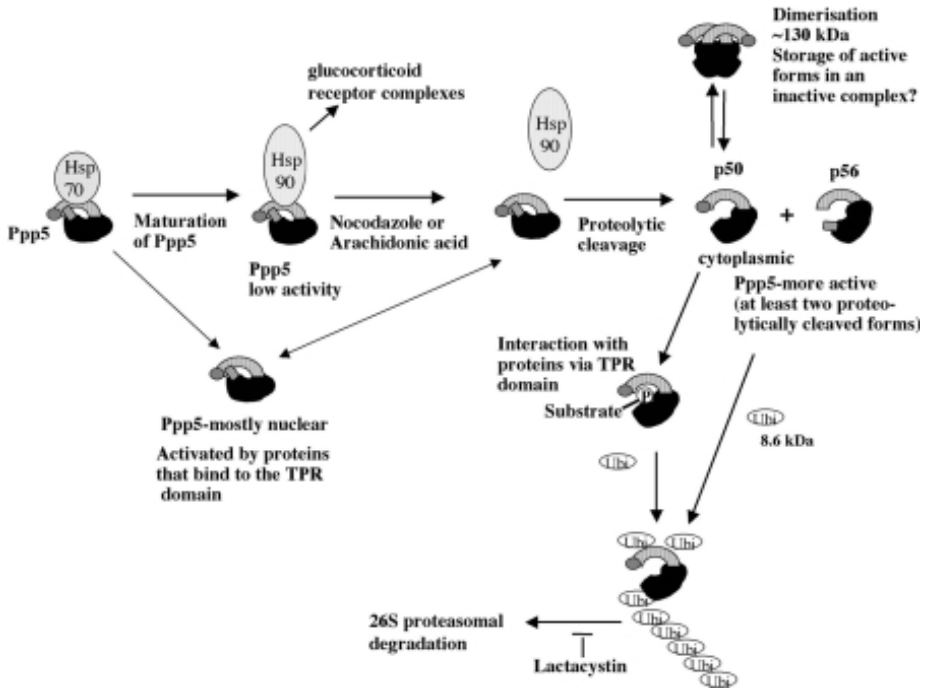

Fig.2: PP5 *in vivo* regulation: a model. Proteolytic cleavage may remove the C-terminal hatched region or the N-terminal grey section. Newly synthesized PP5 forms a complex with Hsp70. Via multichaperone complexes, Hsp90 may displace Hsp70, and PP5 bound to Hsp90 may move to the nucleus in a low-activity form. In response to stress stimuli, PP5 (58 kDa) dissociates from the Hsps and is proteolytically cleaved at either the C-terminus (50kDa isoform) or the N-terminus (56kDa isoform). Cleavage at the C-terminus allows PP5 to translocate to the cytoplasm. A small fraction of proteolytically cleaved PP5 may form dimers (or multimers) by interaction of the TPR domains, which may serve as a store for the production of active monomers. Proteolytic cleaved PP5 monomers dephosphorylate phosphoprotein substrate. The phosphate moiety of the substrate may initially bind to the TPR domain, allowing the phosphatase domain to catalyse the cleavage of phosphate from the substrate. Active PP5 is then polyubiquitinated and destroyed via the proteasome [44].

PP5 activation

Activation of protein phosphatase 5 can occurs by limited proteolysis or the binding of ligands to the TPR domain that seems to acts by inducing a conformational change in the enzyme making the catalytic site accessible to substrates. Two potential physiological activators (long chain fatty acid-CoA esters and the C-terminal domain of the heat shock protein 90) have been identified [53] looking at which some new molecules can be designed to modulate PP5 activity.

PP5 inhibition

In general, the physiological importance of protein phosphatase 5 is demonstrate by the fact that it is the target of many naturally occurring toxins [54].

PP5 is reported to be sensitive to inhibition by okadaic acid, microcystin, cantharidin, tautomycin and calyculin A [55].

Structural analysis of human PP5 showed that the mechanism of TPR inhibition of the phosphatase domain is similar to the mechanism of inhibition of PP2B by its C-terminal regulatory segment and to that of PP1 by toxins [28].

Apparently, inhibitor molecules show their inhibitory function by interacting with the catalytic site of the enzyme.

Phosphatase inhibition has become one of the major area for therapeutic intervention [56-58] and drug design in pharmaceutical companies becomes increasingly important in an ageing society. Anyway, efforts in the field of structure-based design of protein phosphatase activators/inhibitors and ligand screening methodology are strongly appreciated.

1.5 Aims and topics of the research

The present PhD project was focused on the development of new tools that can support a structure-based drug discovery research on metallophosphatases and on a detailed characterization of protein phosphatase 5 that is a representative member of PPP family.

Attention was mainly devoted on the structural and functional characterization of the catalytic domain of serine/threonine phosphatase PP5. This protein is known to participate in signaling networks that either initiate or regulate a variety of cellular events.

Another goal of the research was to clarify the biological role of different metal ions in the regulation of PP5 activity.

There have been several studies that directly implicate PP5 in the development of many diseases like cancer and Alzheimer's syndrome. To better define the biological roles of PP5 and to develop novel antitumor drugs, specific inhibitors of PP5 are desired. PP5 is sensitive to inhibition by several natural toxins commonly employed to study the actions of PP1 and PP2A. In this respect the interaction of PP5 with a group of structurally related inhibitors (cantharidin, cantharidic acid and endothal) has been investigated by X-ray. The three-dimensional structures obtained during the present research have been compared with the few adducts of protein phosphatases already available in the protein databank. The informations obtained have provided new hints to elucidate the structural determinants relevant for the inhibition of protein phosphatases. During the present research a new ligand screening approach based on the analisys of the relaxation rate of water proton has been developed. Relaxometry, that is not so popular as high resolution NMR, is able to provide structural and dynamical informations on diamagnetic and paramagnetic proteins. The relaxometric approach, here decribed, provides informations on the affinity constant

and on the binding mode of the ligand by using very low ammount of protein and inhexpensive instrumentation. Moreover, relaxometric approach can be used in a high-throughput screening strategies for large libraries of compounds or already identified ligands.

A part of the work has been focused on the expression of other members of PPP family using an alternative protein expression technique. Starting from the observation that the human recombinant proteins are usually difficult to be expressed in *E. coli* system, the *in vitro cell-free* system has been taken into consideration even if further optimization in the protocol steps are still in progress.

1.6 Reference List

1. Cortay JC, Negre D, Cozzone AJ. Analyzing Protein-Phosphorylation in Prokaryotes. Methods In Enzymology 1991; 200:214-227.

2. Barford D, Das AK, Egloff MP. The structure and mechanism of protein phosphatases: Insights into catalysis and regulation. Annual Review of Biophysics and Biomolecular Structure 1998; 27:133-164.

3. Stock JB, Stock AM, Mottonen JM. Signal Transduction in Bacteria. Nature 1990; 344(6265):395-400.

4. Koss DJ, Hindley KP, Riedel G, Platt B. Modulation of hippocampal calcium signalling and plasticity by serine/threonine protein phosphatases. Journal of Neurochemistry 2007; 102(4):1009-1023.

5. Neel BG, Tonks NK. Protein tyrosine phosphatases in signal transduction. Current Opinion in Cell Biology 1997; 9(2):193-204.

6. Hunter T. Protein-Kinases and Phosphatases - the Yin and Yang of Protein-Phosphorylation and Signaling. Cell 1995; 80(2):225-236.

7. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 2006; 127(3):635-648.

8. Moorhead GBG, Trinkle-Mulcahy L, Ulke-Lemee A. Emerging roles of nuclear protein phosphatases. Nature Reviews Molecular Cell Biology 2007; 8(3):234-244.

9. Bollen M, Beullens M. Signaling by protein phosphatases in the nucleus. Trends in Cell Biology 2002; 12(3):138-145.

10. Gee CE, Mansuy IM. Protein phosphatases and their potential implications in neuroprotective processes. Cellular and Molecular Life Sciences 2005; 62(10):1120-1130.

11. Sim ATR, Baldwin ML, Rostas JAP, Holst J, Ludowyke RI. The role of serine/threonine protein phosphatases in exocytosis. Biochemical Journal 2003; 373:641-659.

12. Larsen M, Tremblay ML, Yamada KM. Phosphatases in cell-matrix adhesion and migration. Nature Reviews Molecular Cell Biology 2003; 4(9):700-711.

13. Kerk D, Templeton G, Moorhead GBG. Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. Plant Physiology 2008; 146(2):351-367.

14. Huang XZ, Honkanen RE. Molecular cloning, expression, and characterization of a novel human serine/threonine protein phosphatase, PP7, that is homologous to Drosophila retinal degeneration C gene product (rdgC). Journal of Biological Chemistry 1998; 273(3):1462-1468.

15. Goshima G, Iwasaki O, Obuse C, Yanagida M. The role of Ppe1/PP6 phosphatase for equal chromosome segregation in fission yeast kinetochore. Embo Journal 2003; 22(11):2752-2763.

16. Bastians H, Ponstingl H. The novel human protein serine/threonine phosphatase 6 is a functional homologue of budding yeast Sit4p and fission yeast ppe1, which are involved in cell cycle regulation. Journal of Cell Science 1996; 109:2865-2874.

17. Janssens V, Goris J. Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. Biochemical Journal 2001; 353:417-439.

18. Hinds TD, Sanchez ER. Protein phosphatase 5. International Journal of Biochemistry & Cell Biology 2008; 40(11):2358-2362.

19. Cho US, Xu WQ. Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. Nature 2007; 445(7123):53-57.

20. Stefansson B, Ohama T, Daugherty AE, Brautigan DL. Protein phosphatase 6 regulatory subunits composed of ankyrin repeat domains. Biochemistry 2008; 47(5):1442-1451.

21. Janssens V, Goris J. Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. Biochemical Journal 2001; 353:417-439.

22. Janssens V, Longin S, Goris J. PP2A holoenzyme assembly: in cauda venenum (the sting is in the tail). Trends in Biochemical Sciences 2008; 33(3):113-121.

23. Bollen M. Combinatorial control of protein phosphatase-1. Trends in Biochemical Sciences 2001; 26(7):426-431.

24. Borthwick EB, Zeke T, Prescott AR, Cohen PTW. Nuclear localization of protein phosphatase 5 is dependent on the carboxy-terminal region. FEBS Letters 2001; 491(3):279-284.

25. Swingle MR, Honkanen RE, Ciszak EM. Structural basis for the catalytic activity of human serine/threonine protein phosphatase-5. Journal of Biological Chemistry 2004; 279(32):33992-33999.

26. Barton GJ, Cohen PTW, Barford D. Conservation Analysis and Structure Prediction of the Protein Serine/Threonine Phosphatases - Sequence Similarity with Diadenosine Tetraphosphatase from Escherichia-Coli Suggests Homology to the Protein Phosphatases. Eur J Biochem 1994; 220(1):225-237.

27. Barford D. Molecular mechanisms of the protein serine threonine phosphatases. Trends in Biochemical Sciences 1996; 21(11):407-412.

28. Yang J, Roe SM, Cliff MJ, Williams MA, Ladbury JE, Cohen PTW et al. Molecular basis for TPR domain-mediated regulation of protein phosphatase. Embo Journal 2005; 24(1):1-10.

29. Rusnak F, Yu LA, Mertz P. Metalloenzymes and signal transduction: The protein serine/threonine phosphatases, a novel class of binuclear metal-containing enzymes. JBIC 1996; 1(5):388-396.

30. Rusnak F, Yu LA, Mertz P. Metalloenzymes and signal transduction: The protein serine/threonine phosphatases, a novel class of binuclear metal-containing enzymes. JBIC 1996; 1(5):388-396.

31. Reiter TA, Reiter NJ, Rusnak F. Mn2+ is a native metal ion activator for bacteriophage lambda protein phosphatase. Biochemistry 2002; 41(51):15404-15409.

32. Namgaladze D, Hofer HW, Ullrich V. Redox control of calcineurin by targeting the binuclear Fe2+-Zn2+ center at the enzyme active site. Journal of Biological Chemistry 2002; 277(8):5962-5969.

33. Nishito Y, Usui H, Shinzawa-Itoh K, Inoue R, Tanabe O, Nagase T et al. Direct metal analyses of Mn2+-dependent and -independent protein phosphatase 2A from human erythrocytes detect zinc and iron only in the Mn2+-independent one. FEBS Letters 1999; 447(1):29-33.

34. Chu YF, Lee EYC, Schlender KK. Activation of protein phosphatase 1 - Formation of a metalloenzyme. Journal of Biological Chemistry 1996; 271(5):2574-2577.

35. Jackson MD, Denu JM. Molecular reactions of protein phosphatases - Insights from structure and chemistry. Chemical Reviews 2001; 101(8):2313-2340.

36. Ansai T, Dupuy LC, Barik S. Interactions between a minimal protein serine/threonine phosphatase and its phosphopeptide substrate sequence. Journal of Biological Chemistry 1996; 271(40):24401-24407.

37. den Hertog J. Regulation of protein phosphatases in disease and behaviour - Meeting on protein phosphatases. Embo Reports 2003; 4(11):1027-1032.

38. Cortajarena AL, Regan L. Ligand binding by TPR domains. Protein Sci 2006; 15(5):1193-1198.

39. Cliff MJ, Harris R, Barford D, Ladbury JE, Williams MA. Conformational diversity in the TPR domain-mediated interaction of protein phosphatase 5 with Hsp90. Structure 2006; 14(3):415-426.

40. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends in Biochemical Sciences 2003; 28(12):655-662.

41. Chen MS, Silverstein AM, Pratt WB, Chinkers M. The tetratricopeptide repeat domain of protein phosphatase 5 mediates binding to glucocorticoid receptor heterocomplexes and acts as a dominant negative mutant. Journal of Biological Chemistry 1996; 271(50):32315-32320.

42. Russell LC, Whitt SR, Chen MS, Chinkers M. Identification of conserved residues required for the binding of a tetratricopeptide repeat domain to heat shock protein 90. Journal of Biological Chemistry 1999; 274(29):20060-20063.

43. Chen MX, Cohen PTW. Activation of protein phosphatase 5 by limited proteolysis or the binding of polyunsaturated fatty acids to the TPR domain. FEBS Letters 1997; 400(1):136-140.

44. Zeke T, Morrice N, Vazquez-Martin C, Cohen PTW. Human protein phosphatase 5 dissociates from heat-shock proteins and is proteolytically activated in response to arachidonic acid and the microtubule-depolymerizing drug nocodazole. Biochemical Journal 2005; 385:45-56.

45. Morita K, Saitoh M, Tobiume K, Matsuura H, Enomoto S, Nishitoh H et al. Negative feedback regulation of ASK1 by protein phosphatase 5 (PP5) in response to oxidative stress. Embo Journal 2001; 20(21):6028-6036.

46. Urban G, Golden T, Aragon IV, Cowsert L, Cooper SR, Dean NM et al. Identification of a functional link for the p53 tumor suppressor protein in dexamethasone-induced growth suppression. Journal of Biological Chemistry 2003; 278(11):9747-9753.

47. Hinds TD, Sanchez ER. Protein phosphatase 5. International Journal of Biochemistry & Cell Biology 2008; 40(11):2358-2362.

48. Golden T, Swingle M, Honkanen RE. The role of serine/threonine protein phosphatase type 5 (PP5) in the regulation of stress-induced signaling networks and cancer. Cancer and Metastasis Reviews 2008; 27(2):169-178.

49. Golden T, Aragon LV, Rutland B, Tucker JA, Shevde LA, Samant RS et al. Elevated levels of Ser/Thr protein phosphatase 5 (PP5) in human breast cancer. Biochimica et Biophysica Acta-Molecular Basis of Disease 2008; 1782(4):259-270.

50. Golden T, Aragon IV, Zhou GF, Cooper SR, Dean NM, Honkanen RE. Constitutive over expression of serine/threonine protein phosphatase 5 (PP5) augments estrogen-dependent tumor growth in mice. Cancer Letters 2004; 215(1):95-100.

51. Liu F, Grundke-Iqbal I, Iqbal K, Gong CX. Contributions of protein phosphatases PP1, PP2A, PP2B and PP5 to the regulation of tau phosphorylation. European Journal of Neuroscience 2005; 22(8):1942-1950.

52. Liu F, Iqbal K, Grundke-Iqbal I, Rossie S, Gong CX. Dephosphorylation of tau by protein phosphatase 5 - Impairment in Alzheimer's disease. Journal of Biological Chemistry 2005; 280(3):1790-1796.

53. Ramsey AJ, Chinkers M. Identification of potential physiological activators of protein phosphatase 5. Biochemistry 2002; 41(17):5625-5632.

54. Mackintosh C, Mackintosh RW. Inhibitors of Protein-Kinases and Phosphatases. Trends in Biochemical Sciences 1994; 19(11):444-448.

55. Honkanen RE, Golden T. Regulators of serine/threonine protein phosphatases at the dawn of a clinical era? Current Medicinal Chemistry 2002; 9(22):2055-2075.

56. Bialy L, Waldmann H. Inhibitors of protein tyrosine phoshatases: next genereation drugs? Angew Chem Int Ed Engl. 2005; 44(25):3814-39

57. Nören-Müller A, Reis-Corrêa I Jr, Prinz H, Rosenbaum C, Saxena K, Schwalbe HJ, Vestweber D, Cagna G, Schunk S, Schwarz O, Schiewe H, Waldmann H. Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. Proc Natl Acad Sci U S A. 2006; 103(28):10606-11

58. Nören-Müller A, Wilk W, Saxena K, Schwalbe H, Kaiser M, Waldmann H. Discovery of a new class of inhibitors of Mycobacterium tuberculosis protein tyrosine phosphatase B by biology-oriented synthesis. Angew Chem Int Ed Engl. 2008;47(32):5973-7

2

MATERIALS AND METHODS

2.1 Structural and Functional Genomic

Impressive progress in genome sequencing, protein expression and high-throughput crystallography and NMR has radically transformed the opportunities to use protein three-dimensional structures to accelerate drug discovery, but the quantity and complexity of the data have ensured a central place for informatics. Structural biology and bioinformatics have assisted in lead optimization and target identification where they have well established roles; they can now contribute to lead discovery, exploiting high-throughput methods of structure determination that provide powerful approaches to screening of fragment binding[1]. With the advent of modern high throughput technologies in both genomics and biological screening, and at the same time the enormous advances in computer technology, it is now feasible to use these tools in rational approaches in the search for new medicines.

The genome sequencing projects developed in the last few years have made available the primary sequence of an enormous number of new genes. The draft sequence of the human genome [2,3] is now available together with the complete genomes of many other organisms (Web site: www.ncbi.nlm.nih.gov). This wealth of genomic data can be fully exploited only through elucidation of the three-dimensional structures of the corresponding proteins. The availability of the human genome provides access to a broad variety of phosphatases. Thus, the use of sophisticated techniques to identify potential inhibitors allows systematic structure-based drug design, that is the final objective of the "classic" Structural Genomics.

The study of the protein function can be achieved by the identification, production and characterization of partner proteins and by protein-protein interaction studies. The various Structural Genomics initiatives commencing worldwide differ in the choice of protein targets in an effort to provide a complete coverage of protein structures and relative associated function. So, there are actually two different possible philosophies that distinguish the various research consortia.

The Structural Genomics projects are characterised by a High-Throughput Production (HTP) of proteins and structures, and the goal of this kind of projects is to provide a complete repertoire of protein folding by determining a representative structure. In this case the effort for the target selection is limited. Entire genomes or very large protein families (e.g.: all soluble human proteins under 100 residues) are selected, cloned and expressed in parallel using automatic systems. The great advantage of this approach is that a huge number of genes are screened for protein expression and solubility. However, a common

cloning and expression protocol has to be used for all the genes, limiting the rate of success (usually 20-30%) for each target.

The Functional Genomics projects use a Medium Throughput Production approach (MTP) and focus the effort on a limited number of proteins that are linked to diseases or are involved in some specific metabolic pathways. Large bibliographic and bioinformatics research is at the basis of the target selection. A limited number of targets are cloned but, in order to maximize the rate of success, many simultaneous cloning and expression strategies are performed.

For both strategies a multidisciplinary approach is necessary; different scientific fields like bioinformatics, biochemistry and molecular biology have to cooperate in order to proceed in the target selection, protein expression, and finally structural and functional characterization.

In this work, the multidisciplinary approach was applied to study Protein Phosphatases. In particular, NMR spectrospcopy, X-ray and relaxometry have been used to get informations on physiological metals of the targeted metalloenzyme, on its catalytic mechanism and on inhibition operated my small molecules.

2.2 Genome browsing and domain definition

One of the main challenges facing the molecular biology community today is to make sense of the wealth of data that has been produced by the genome sequencing projects. In the past decade, bioinformatics has become an integral part of research and development in the biomedical sciences. Bioinformatics now has an essential role both in deciphering genomic, transcriptomic and proteomic data, generated by high-throughput experimental technologies, and in organizing information gathered from traditional biology. It is an interdisciplinary research area at the interface between the biological and computational science. The ultimate goal of bioinformatics is to uncover the biological information hidden in the mass of data and obtain a clearer insight into the fundamental biology of organisms.

A number of data banks with databases and programs have been created with the aim of providing the scientific community with tools for searching gene banks, for the analysis of protein sequences, for the prediction of a variety of protein properties. Primary databases contain information and annotations of DNA and protein sequences, DNA and protein structures and protein expression profiles.

Some available databases for genome browsing are:

- **NCBI** (www.ncbi.nlm.nih.gov/Entrez/) - This web site integrates information from several databases (Swissprot, EMBL, all geneBank, etc...)

- **PDB** (www.rcsb.org/pdb) - A 3-D biological macromolecular structure database
- **Pfam** (http://pfam.wustl.edu) - A collection of different protein families organized in terms of different domains as obtained from multiple alignment.

Secondary or derived databases are so called because they contain the results of analysis of the primary resource including information on sequence patterns or motifs, variants and mutations and evolutionary relationships.

Once all biological data is stored, the requirement is to provide bioinformatics tools for extracting the meaningful information.

The most used programs for genome browsing are:

- **BLAST** (www.ncbi.nlm.nih.gov/BLAST/): Standard BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. PHIBLAST is designed to search for proteins that contain a pattern specified by the user, and simultaneously are similar to the query sequence.
- **CLUSTALW** (www.ebi.ac.uk/clustalw/) is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms
- **PROSITE** (www.expasy.org/prosite/) SCANPROSITE allows to scan a protein sequence, provided by the user, for the occurrence of patterns and profiles stored in the PROSITE database, or to search in protein databases all sequences with a user entered pattern.
- **STRING** (dag.embl-heidelberg.de) STRING is a database of predicted functional associations among genes/proteins. Genes of similar function tend to be maintained in close neighbourhood and tend to be present or absent together.

Once individuated a protein of particular interest, this is subjected to further bioinformatics investigations in order to predict important features like stability, solubility, hydrophobicity, secondary and tertiary structures.

For each target protein several domains are selected for expression, and of these domains various construct are cloned in order to increase the probability of obtaining a soluble and well-folded protein. Domain selection and construct design are performed considering several aspects.

The first important step is to perform the transmembrane regions prediction using dedicated bioinformatics programs. (e.g. TMHMM: www.cbs.dtu.dk/services/TMHMM-2.0) The result of this analysis make us know if we are dealing with a completely soluble protein, with an integral transmembrane protein or with a protein mainly soluble with a small portion inserted into the phospholipidic bilayer. In the latter case it is possible to try to express a recombinant protein without the transmembrane tail. Using the program SIGNALP (www.cbs.dtu.dk/services/SignalP) it is also possible to predict the presence of a signal-peptide sequence that enables the expression of the protein in the periplasmic area.
Another important analysis is the secondary structure prediction, exploiting the web server http://npsa-pbil.ibcp.fr/, in order to have an indication of the regions with tendency to form helices, beta-sheet, coil or turns. Combining these predictions with ClustalW alignments of the protein family, we can evaluate which is the best residue where to cut the domain, still maintaining the correct folding of the protein. It is also useful to search for the presence of intrinsically unstructured/disordered region using IUPred (http://iupred.enzim.hu/).
Once individuated a protein of particular interest, this is subjected to further bioinformatics investigations in order to predict important features like stability, solubility, hydrophobicity, secondary and tertiary structures.
On the basis of bioinformatic analysis and solubility predictions the following Protein Serine/Threonine Phosphatases were taken into consideration and the following constructs have been selected:

PP2A catalytic domain: PP2Ac (residues 1-290) and $PP2Ac_1$ (residues 23-290)

PP6 catalytic domain: PP6c (residues 1-298) and $PP6c_1$ (residues 1-305)

PP5: 16-499

PP5 catalytic domain: PP5c (residues 169-499)

2.3 DNA preparation techniques

2.3.1 Gene cloning

In the Functional genomic approach, the cloning strategy has to be carefully designed, since it will influence the behaviour of the target protein, e.g. yield, solubility, folding, etc.
The best strategy to maximize the probability of obtaining a soluble and correctly folded target protein is to select, following preliminary bioinformatics analyses, several different domains and clone each domain into several different expression vectors, which allow the expression of the target not only in the native state, but also using several different fusion partners. They are reported to assist protein solubility and folding [4].

Even if the number of fusion partners is increasing progressively during the years it has been experienced that none of these tags work universally with each partner protein.

Classical cloning uses restriction enzymes that cut the DNA of interes at specific recognition sites. To be able to clone a DNA insert into a cloning or expression vector, both have to be treated with two restriction enzymes that create compatible ends. The next step is the ligation of the insert into the linearized vector. This involves the formation of phosphodiester bonds between adjacent 5'-phosphate and 3'-hydroxyl residues, a reaction catalyzed by the bacteriophage T4 DNA ligase. Finally an aliquot of this reaction is transformed in DH5α *E. coli* competent cells and positives clones are screened both by PCR and DNA sequencing. The yield of all this procedure is particularly influenced by the ligation step that is dependent from several factors, like DNA concentration, vector/insert ratio and transformation efficiency of the cells. Moreover the parallelization of this approach is not so easy because it is a very time-consuming technique and not all the host vectors and the target genes can be cut with the same restriction enzymes.

When recombinant proteins are expressed *in vivo*, genes are typically subcloned and expressed from medium- or low-copy plasmids to avoid residual gene expression or "leakiness" that occurs in high-copy plasmids. Low-copy plasmids provide a lower gene dosage, which minimizes background levels and maintains tight regulation of expression [5].
In a *in vitro* protein synthesis (also called *cell-free,* see paragraph 2.4.2) high copy plasmids are used beacuse high quantity of available DNA template is required.
A more recent cloning technique, called Gateway technology, was described by Landy *et al.* [6] (Fig. 5). This technique allows a rapid cloning of one or more genes into virtually any expression vector using site-specific and conservative recombination (LR reaction), thus eliminating the requirement to work with restriction enzymes and ligase. Moreover, the Gateway system resulted faster due to the higher reaction efficiency and less sequencing demand since only the starting pENTR should be sequence verified.

PP5c, PP2Ac/c_1 and PP6c/c_1 constructs were cloned according to the Gateway technology. Different fusion tags have been used (His, GST, MBP, NusA, Trx, GB1). PP6 c/c_1 and PP2A c/c_1 have been also cloned inside an expression vector that does not contain any fusion tag in order to get native proteins.

High copy expression vectors are usually used in *cell*-free system. The MCSI vector, a pET derivative plasmid, has been used in this work and cloning has been performed according to the classical ligase reaction.

Vector deriving from Gateway cloning have been used to test the expression of PP6 c/c_1 and PP2A c/c_1.

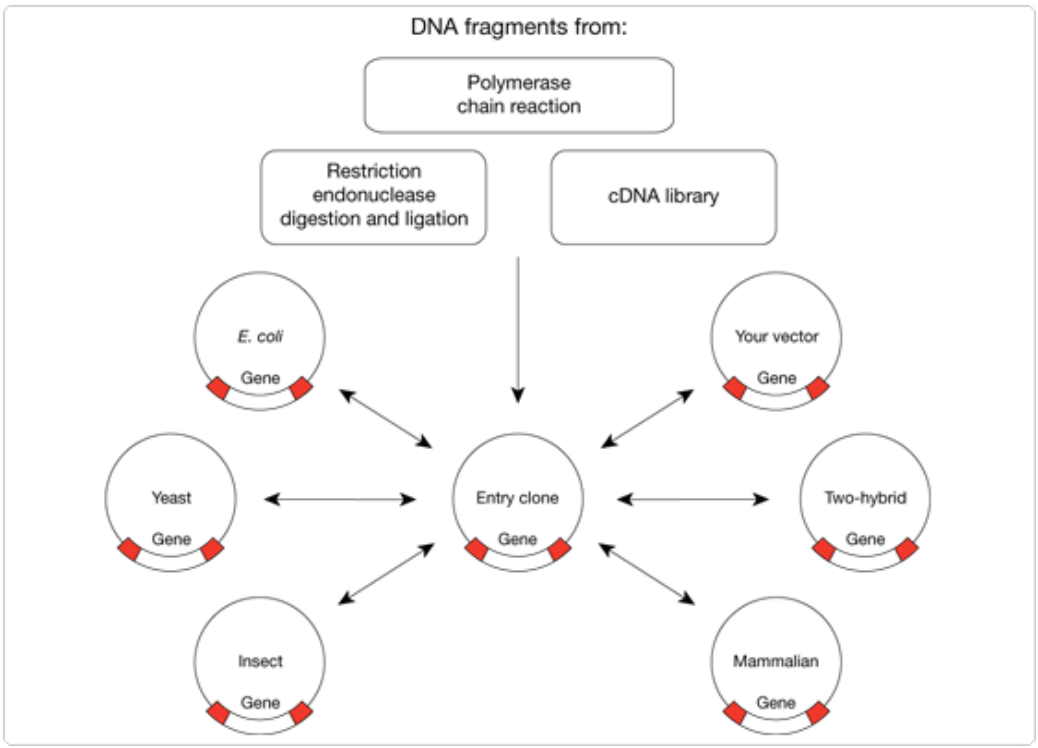

Fig.5:**Gateway® technology principle.** The startegy facilitates cloning of genes, into and back out of, multiple vectors via site-specific recombination. Once a gene is cloned into an Entry clone, the DNA fragment can be moved into one or more destination vectors simultaneously (Reprinted from Invitrogen website)

2.3.2 PCR amplification for *in vitro* expression

Large-scale plasmid preparation in *in vitro* expression system is not always required [7].

In the last years, a new technique has been developed. Starting from a DNA as template, some PCRs steps are carried out, generating stable PCR templates ready to be used in *in vitro* transcription/translation system. This startegy it is more expensive but faster than the classic cloning technique and it also offers the advantage to have more pure DNA preparations. The obtained template contains all the elements that are needed to start the transcription of the target gene (Fig.6).

This technique still remain to be experienced in our laboratories since we have been carring out the *in vitro* synthesis by using expression vectors deriving from Gateway cloning.

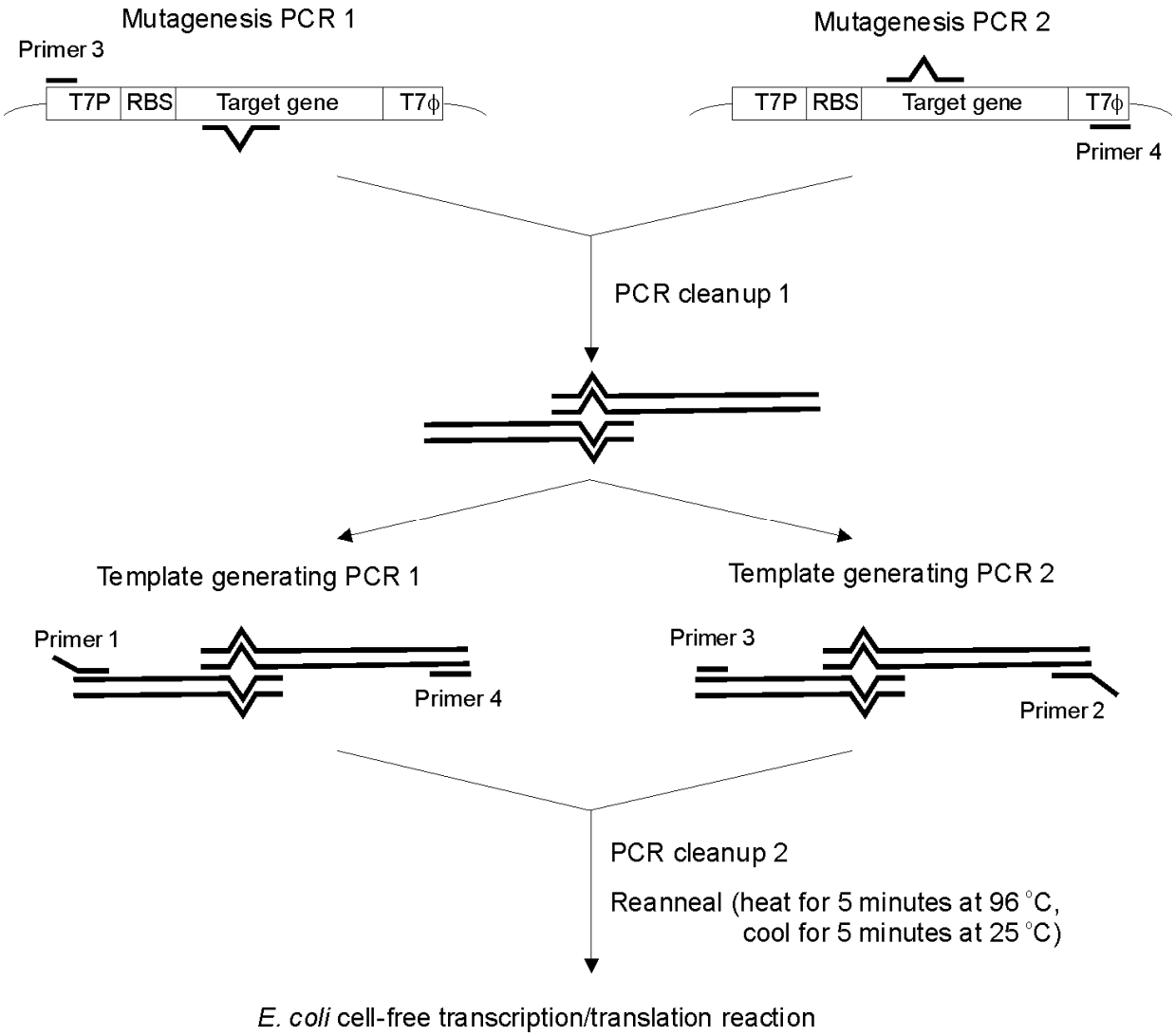

Fig.6 Generation of stable PCR templates from a construct with aT7 f10 promoter, a ribosome binding site, and a T7 terminator (Reprinted from Wu *et al.*[7]).

2.3.3 Site-directed mutagenesis

The functional analysis of a protein often requires the generation of a number of single or multiple amino acid variants of the wild type gene. Site-directed mutagenesis is a technique for carrying out vector modification and characterizing the dynamic, complex relationships between protein structure and function. The basic procedure utilizes a supercoiled double-stranded DNA vector with insert the gene of interest and two synthetic oligonucleotide primers, both containing the desired mutation. The primers, each complementary to opposite strands of the vector, are extended during PCR reaction by a

high fidelity DNA polymerase (PfuUltra, Stratagene). Extension of the oligonucleotide primers generates a mutated plasmid. After PCR reaction, the product is treated with an endonuclease specific for methylated and hemimethylated DNA, DpnI. This enzyme is used to digest the parental DNA template and thus select only the mutation-containing synthesized DNA; this happens because DNA isolated from almost all *E. coli* strains is dam methylated and therefore susceptible to endonuclease digestion. The vector containing the desired mutations is then transformed into XL1-Blue supercompetent cells and subsequently subjected to sequencing analysis.

Insertion of a PCR fragment into a recipient plasmid DNA can be performed expanding the Site-directed mutagenesis technique without using restriction enzimes. This methods was recently optimized by Novartis [8].

Originally, the expression vector containing PP5 full-lenght gene was generate to express a GST- PP5- His-tagged protein. While it was possible to remove the GST, the absence of a tag-removal enzyme site in the protein sequence made not possible the His-tag removal. The presence of an His-tag has been supposed to influence the binding of metals in the catalytic pocket. The Novartis mutagenesis strategy was used to engineerize a new expression vector containing a TEV recognition site at the PP5 C-terminus.

2.4 Protein expression

2.4.1 *In vivo* expression

E.coli expression system is, nowadays, the most widely used, since it is the easiest, quickest and cheapest one. Soluble protein expression has continued to be a major bottleneck in the field though by the moment that many proteins belonging to interesting families, especially the human proteins, are often accumulated as inclusion bodies. *E. coli* is a prokaryote and lacks intracellular organelles, such as the endoplasmic reticulum and the Golgi apparatus that are present in eukaryotes, which are responsible for post-translation modifications of the proteins being produced. Many eukaryotic proteins can be produced in *E. coli,* but sometimes they are produced like a non-functional inclusion bodies protein, since glycosylation or post-translational modifications such as proteolytic processing, folding and disulfide bond formation do not occur.

Because of this, considerable efforts are still in progress in order to optimize both the *in vitro* refolded processes and alternative strategies (mammalian, yeast and insect host) for overcoming some of the problems inherent in soluble and membrane protein expression[9]. As general strategy, a preliminary expression test is performed in a small-volume scale

using different strains (e.g.:, Bl21 (DE3) Gold, Bl21 (DE3) pLysS a protease deficient strain, Rosetta (DE3) and Bl21 (DE3) RIPL for genes containing rare codons and Origami (DE3) for proteins containing disulphide bounds), three expression temperatures (37-25-17°C), three inducer (arabinose or isopropyl-beta-D-thiogalactopyranoside) concentrations, different induction times (4h and 16h) and three culture media of different composition. Expression results has been checked on SDS polyacrilamide gel (SDS-PAGE). This kind of approach allows us to explore a large set of expression conditions and to evaluate which one gives the best yield of soluble protein.

Factors such as reduced temperature, changes in the *E.coli* expression strain or induction conditions and co-expression of molecular chaperones have all been deeply examined at this step and refinement of the expression conditions for the soluble proteins made before starting massive culture that is performed in flasks or fermentors to increase protein yield and have a better parameter control (e.g. pH, temperature and pO2).

If all the proteins are expressed in the insoluble fraction as inclusion bodies, the choice falls in redefining the domain, in prorocceeding with an " in vitro" refolding screening or in turning into a different expression system.

2.4.2 *In vitro* expression

Cell-free protein synthesis is an alternative strategy for overcoming some of the problems linked to the *in vivo* expression system . Many reviews can be found in literature [10-14]. The principle of the *in vitro* protein synthesis is based on the production of proteins without the use of living cells. The necessary catalytic machinery required for metabolism, transcription, translation, and proper folding *in vitro* is extracted from a living source. To date, a number of lysates have been successfully prepared, including *E. coli,* yeast, rabbit reticulocyte, wheatgerm, a hyperthermophile, Drosophila embryo, hybridoma, *Xenopus* oocyte or egg, insect, mammalian and human cells.

The cellular extract preparation represent one of the critical step in *cell-free* technology. Soft cell-lysis must be performed by the moment that sonication may completely or partially destroy ribosome apparatus, thus resulting in a not-working translation.

The most used source is *E.coli* because it is easy to menaged with . This cellular extract is named S30 because it is obtained by centrifuging the lysated cell at 30 000*g*, that is the right speed to get a functional catalytic machinery.

The system lacks of a cell wall or membrane barrier, and of requirements to satisfy biochemical needs associated with growth and maintenance of a living cell. This aspect is hightly advantegeous because energy and other resources can be more focused and efficiently utilized for processes directed toward synthesis and folding of the target and

purification steps are not required. The cellular extract is washed through centrifugation steps. Membranes, some of proteases and RNA and others unuseful molecules are removed. All the purified components such as ribosomes, T7RNA polymerase and the most of the components of the cytoplasm of the living source (Fig.7) are transferred in an *in vitro* apparatus simulating living cells.

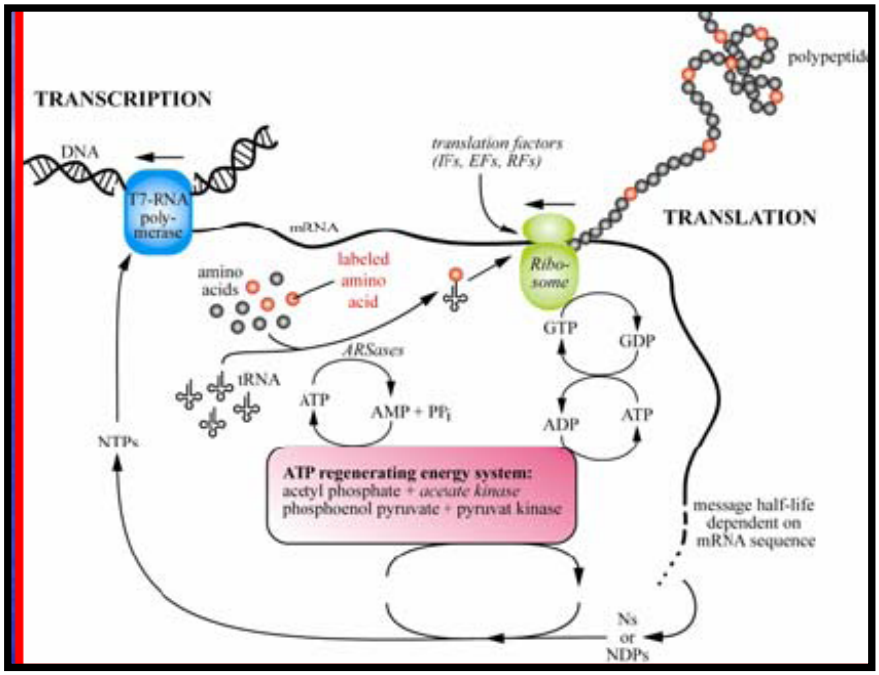

Fig.7: Essential components of transcription/translation machinery (Reprinted from Bernhard F.[42]).

An example of apparatus is represented by a dyalisis sistem in which the dialysis membrane may represent the cellular membrane. (Fig.8)

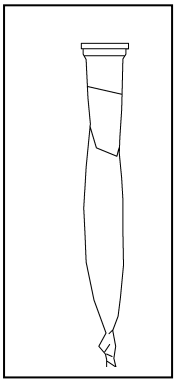

Fig.8: Reaction chamber assembly. Soak one 10 cm length of Spectrapor #2 dialysis tubing per reaction in MQ water for 5 min, then tie a knot as close as possible to one end. Trim the knot to have minimal overhang. Cut off the lid and hinde section of 500 µL Eppendorf. The lid is retained and the hinge is discarded. Removal of the hinge allows the tube to fit inside a 10 mL tube. The bottom 0.5 cm of the tube is then sliced off using a scalpel. Insert each cut Eppendorf tube into the top of a pre-tied length of dialysis tubing, The Eppendorf tube ensures the contents of the dialysis tube are readily accessible. Remove the dialysis tubing from the MQ water and insert the Eppendorf tube immediately prior to use, so that the tubing does not dry out before assembly of the complete reactionPipette the reaction mixture into the dialysis tubing and seal the Eppendorf tube with the pre-cut lid. Place the dialysis tubing inside the 10 mL tube containing the outside buffer. The entyre assembly is then collocated at 37°C with shaking for 6-8 hours, then the inner solution centrifugated and protein expression checked by SDS-acrylamide electrophoresis. Reprinted from Appnonyi *et al.* [15]

Dialysis membrane separates the external solution that is in exchange with the inner solution where the protein synthesis occurs. The outer solution contains a buffered solution of pH 7.5, aminoacid mix, salts, energy - regenerating factors and alanine for T7RNAP; the inner solution has the same composition of the external buffer, but it is enriched with cellular extract, DNA plasmid, T7RNA polymerase plasmid (or purified T7RNA enzyme), RNasin inhibitors. The continuous exchange of the reagents it was proved to increase the yield of protein expression when compared with old methods based on not-continuous exchange system [16,17].

The reaction is allowed to proceed in a autoinducted manner[18]. Starting input is given by the bacterial T7 RNA polymerase that it can be recovered in the cellular extract. In this case, T7RNA polymerase is freshly generate making the procedure cheeper and faster. Previously expressed and purified T7RNA polymerase is added to the reaction when labelled protein are expressed. In this way labelled aminoacid incorporation in T7RNA polymerase does not occur.

By the moment that the system is not a living organism, many parameters and conditions can be directly controlled. In fact, the completely open nature of *cell-free* systems allows external molecules and components to be added to create favourable environments for protein folding and activities. Consequently, candidate are proteins which are toxic or sensitive to proteolysis *in vivo* [19] and proteins requiring binding partners for expression in soluble form[20] including membrane proteins[21], metals [22] or inhibitors. A cellular extract from *E. coli* containing rare codons may help the synthesis of human proteins that are rich in. Addition or coexpression of trigger factors to the reaction is also possible in *vitro.* Some other parameters can be controlled and efforts to prolong the protein synthesis and to increase the final yield are continuously in progress [23, 24, 25].

Cell-free protein synthesis has a great potential in structural biology, since Kigawa *et al*. demonstrated the possibility of obtaining milligram-per-mL yields in a coupled cell-free transcription-translation system based on *E. coli* cell extract[26]. These concentrations are sufficient for analysis by nuclear magnetic resonance (NMR) spectroscopy, providing a convenient and rapid way to analyze proteins, which would be difficult or costly to make .

One of the most useful application is the specific labeling of proteins with isotopes , unnatural or fluorescence-enhanced amino acid. [27,28]. *Cell-free* protein synthesis uses isotope-labelled amino acids much more economically than *in vivo* expression systems and is much less affected by isotope-scrambling by metabolic enzymes[29]. Moreover, the target protein is the only protein synthesized during the reaction. Consequently, purification of

the protein by chromatography or concentration of the sample is not necessary[30] and proteins can be made from DNA and structurally characterized within one day [31].

Many commercial kits for high-yield *cell-free* protein expression have become available (*e.g.* the RTS system by Roche and the Expressway system by Invitrogen), but the daily cost tends to be prohibitive. An "hand-made" *E. coli* based cell-free protein synthesis protocol has been developed in the Otting's laboratory at the Australian National University [15] and applied in our laboratories.

However, exspecially for NMR purpose, which requires the production of high yield of labelled 15N and 13C samples, *E.coli* still remain the best choice.

PP2Ac/c_1 and PP6 c/c_1 have been expressed as insoluble in *E.coli*. *Cell-free* protein sinthesys has been considered as a new approach to test the expresion and solubility of the target proteins.

2.5 Protein purification follows the *in vivo* expression

Depending on the subcellular localization of the expressed protein (e.g. periplasmic space or within the cytoplasm as soluble or insoluble) different purification ways can be performed.

After disrupting the cells by sonication, next steps must be compatible with different physical, chemical and biological characteristics of the proteins, the amount of material to be processed and the intended downstream applications. All the purifications involve several chromatographic steps. Ion exchange and size exclusion chromatography are commonly used to purify proteins expressed as native. On the other hand, fusion tags increase the efficiency of protein purification by affinity chromatography. In particular, gel filtration, immobilized metal ion affinity chromatography (IMAC) and agarose matrix having GST (glutathione-S-transferase) affinity have been used in PPP sample preparation. Generally the GST affinity for its matrix is not so strong so further purification steps are required.

IMAC exploits the interaction between chelated transition metal ions like Zn^{2+} or Ni^{2+} and side-chains of specific amino acids (mainly histidine) on the protein . Methods used to elute the protein of interest include changing the pH, or adding a competitive molecule, such as imidazole. Excess of reduced glutathione is used to displace the GST tagged protein.

Enzymatic digestion using a specific protease is then necessary to remove the fusion partner from the target protein and a second IMAC purification is generally performed in order to separate them. TEV, Fatt.Xa, Thrombin, Prescission Protease, recombinant

Enterokinase are some example of proteases that are normally used for the cleavage of fusion proteins.

PP5 (residues 16-499) protein has been purified by GST affinity chromatography and GST cutted by Prescission Protease enzyme (Amersham). A second GST affinity chromatography has been performed in order to separate GST-tag and the not cutted protein from the cutted one.
and). A further purification step was done using IMAC by His-trap column charged with Ni^{2+} considering the presence of a C-terminal His-tag. Finally size exclusion cromatography has been performed.
The protein sample (in Tris 50 mM, NaCl 150 mM, DTT 10 mM, ZnCl2 0.1 mM, pH 7.5) concentrated up to 0.25 mM, was analyzed by 1H,15N-TROSY NMR spectrum recorded at 900 MHz.

In case of PP5c (residues 169-499), the first purification step has been performed using GST-affinity cromatography. His-GST tag has been cleaved with 2ul of Ac-TEV protease/1mg of fusion protein (Invitrogen, Carlsbad, CA) under over night incubation at room temperature.The IMAC by His-trap column charged with Ni^{2+} purification has been made. The unbounded protein coming from this step represents the protein lacking the His-GST tag. This sample has been further purified by gel filtration. Samples of the purified PP5 protein were dialyzed against either solutions containing MnCl2 or ZnCl2 in order to produce the paramagnetic protein and the diamagnetic analogue. A fast dyalisis has been done and unbound metals removed.

PP2A c/c_1 -PP6 c/c_1
The two His-tagged proteins have been recovered in the inclusion bodies. IMAC purification has been performed by His-trap column charged with Ni^{2+}. Denaturising conditions (urea) have been used in each purification step. Consequently the use of Ac-TEV protease for the removal of His-tag was not possible. In this respect different refolding protocols have been tested. Precipitation of more than 90% of the sample occurred so that further characterizations were not allowed.
The *in vitro* synthesis resulted in a insoluble expressed product for all the PPP constructs.

2.6 Biochemical and biophysical protein characterization

2.6.1 Metals content determination by ICP

Inductively Coupled Plasma (ICP) is an analytical technique used for the detection of trace metals and allows multi-elemental, simultaneous analysis of most elements in the periodic table, with excellent sensitivity (ng/L or ppt) and high sample throughput. The analytical principle used is optical emission spectroscopy, triggered by plasma ionization. The liquid sample is introduced in the prechamber, usually by mean of a peristaltic pump, then nebulised and entrained in the flow of plasma support gas, which is typically Argon. The analytes are excited into a state of radiated light emission by the plasma vapour and the emitted radiation is converted to an electrical signal that can be measured quantitatively by resolving the light into its component radiation (almost always by means of a diffraction grating). The light intensity is measured, at the specific wavelength for each element line, with a CCD (*charge coupled device*, basically an array of tiny, light-sensitive diodes) which converts it to electrical signals. The electron signal intensity is then compared to previously measured intensities of standard solutions of the target elements, and a concentration is computed. This spectroscopic technique is very suitable for the proteins metal content determination, since it requires very small amounts of sample, and is not perturbed by the polypeptidic matrix.

Both the non-metallated and metallated samples of PP5c (169-499) have been then analyzed by ICP mass for the metal contenent.

2.6.2 Enzymatic assay

Although serine/threonine phosphatases participate in multiple signaling transduction events the actual biological substrates for most phosphatase are not known. As such, activity assays rely on synthetic surrogate substrates.

Basically, activity assays for phosphatases consider the hydrolysis of a arylphosphate moiety from a synthetic substrate yielding a spectroscopically active product. Many different substrates and assays have been developed to detect the hydrolytic activity of protein phosphatases [32].

Hydrolysis products that can be quantified without interfering with the reaction (continuous enzyme assays) are advantageous over those that require the reaction to be stopped (discontinuous enzyme assays) beacuse they require decreased sample manipulation [33].

Para-nitrophenyl phosphate (*p*NPP) is a chromogenic substrate for most phosphatases such as alkaline phosphatases, acid phosphatases, protein tyrosine phosphatases and serine/threonine phosphatases.The reaction yields para-nitrophenol (Fig. 9).

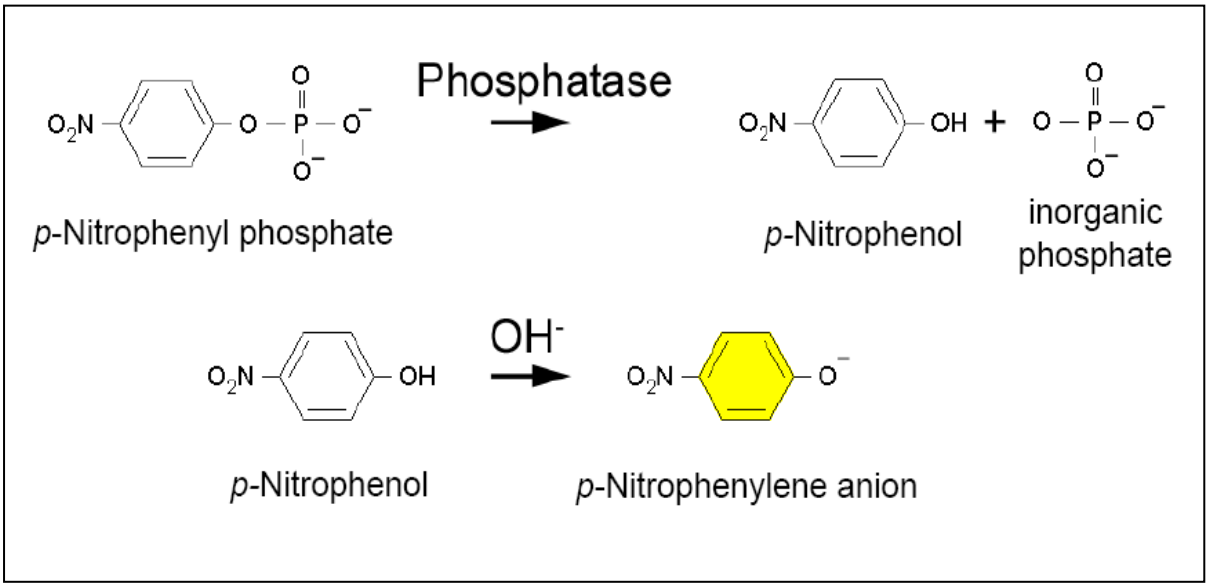

Fig. 9: pNPP Assay Principle: deprotonation of the phenolicgroup gives rise to the yellow colored p-nitrophenylene anion.

The phenolate ion becomes an intense yellow soluble product under alkaline conditions that are required by most serine/threonine phosphatases. Production of *p*NP can be followed by monitoring the increase in absorbance at 405nm on a spectrophotometer in a continuous way. An additional advantage is represented by the fact that serine/threonine phosphatses show high efficiency hydrolysis for *p*NPP[34].

Consequently, progressive suppression of *p*NPP phosphatase activity can be monitored by adding competitive phosphatase inhibitors for which the binding site is the catalytic centre, which is also accessible to *p*NPP [34]. The test can be applied also to check the metal requirement for phosphatase activity. False positive may represent one of the most disadvantage in colorimetric assay.Consequently, the assay can be utilized as a simple and preliminary approach for ligand screening studies.

The activity of catalytic PP5c has been tested at 25C° using *p*-nitrophenyl phosphate (*p*NPP) in Cary 50 Eclypse Spectrophotometer. Activity was expressed as Unit according to the *Jena Bioscience* manufacturing (Unit definition: amount of the phosphatase required to relase 1 mol of phosphate from pNPP in one minute at 25°C).

The larger enzymatic activity of catalytic manganese-loaded PP5 versus the catalytic zinc-loaded protein suggests a physiological role for the paramagnetic ion. This result further supports the use of the manganese-loaded PP5 in ligand selection and optimization that is

also based on the knowledge of the protein structure. Specific protein-ligand interaction can be investigated, increasing efforts in structure-based drug discovery.

2.7 Structural characterization

In order to understand completely the function of a protein, the determination of the 3D structure is of basic importance. X-ray crystallography and NMR spectroscopy are the two main techniques that can provide structures of macromolecules at atomic resolution. Both techniques are well established and play a key role in Structural Biology as a basis for a detailed understanding of molecular functions. Their respective different advantages and disadvantages in terms of sample preparation and data collection and analysis make these two approaches complementary in Structural Proteomics.
Additionally, relaxometry can provide structural and dynamical information on diamagnetic and paramagnetic biomolecules.

2.7.1 Structural determination by NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is unique among the methods available for 3D- structure determination of proteins at atomic resolution, since the NMR data can be recorded in solution, simulating a phisiological state of the protein. An highly purified protein sample is required and in homogeneous preparation and/or aggregation of the protein as well as low molecular weight impurities may be harmful in structural determination. Limiting factors in NMR spetroscopy still remain the long time request for the data analysis as well as the size limit of protein that is currently around 35KD. Recent advances in both hardware and experimental design promise to allow the study of much larger protein [35].

PP5 (16-499)
The protein sample (in Tris 50 mM, NaCl 150 mM, DTT 10 mM, ZnCl2 0.1 mM, pH 7.5) concentrated up to 0.25 mM, was analyzed and folding checked by 1H,15N-TROSY NMR spectrum recorded at 900 MHz
PP5c (169-499)
The protein sample (in Tris 50 mM, NaCl 150 mM, DTT 10 mM, MnCl2 0.1 mM, pH 7.5), concentrated up to 0.3 mM, was analyzed by 1H,15N-TROSY NMR spectrum recorded at 700 MHz.

2.7.2 Proton relaxometry

Proton relaxometry can be applied to provide structural and dynamical informations on biomolecules containing a paramagnetic metal in the active site [36] The paramagnetic relaxation enhancement experienced by proton nuclei is mainly influenced by a direct interaction of the solvent molecules with the metal ion [37]. In this case, the relaxation rate of solvent nuclei can act as a probe to monitor the solvent accessibility to the paramagnetic center. Several pharmaceutical targets are metalloproteins, where a diamagnetic or paramagnetic metal ion plays a catalytic function. Many commercial drugs inhibits these pathologically relevant metalloenzymes interacting directly with the catalytic metal ion or preventing the access of the physiologic substrate to the active site[38]. In most of the case, the interaction of the solvent molecules with the metal ion is also prevented by the ligand binding. Therefore relaxometry can be exploited in drug discovery to monitor the interaction of ligands at the metal binding site. Moreover, a relaxometric investigation performed on already identified ligands permits a rapid selection of the molecule for their ability to bind the catalytic pocket.

The presence of two metal in the catalytic pocket of PP5 in the is revealed by X-ray analysis[39,40]. The possibility to achieve a metallated form of PP5 with the paramagnetic Mn[II] make PP5c suitable for this kind of validation. Moreover inhibition of its phosphatase activity can be efficiently achieved by blocking the catalytic channel with small ligands [41]. Cantharidin, a biologically active natural compound, is one of the prototypes of this class of inhibitors. The binding of cantharidine and of two structurally-related compounds to PP5 has been here investigated with relaxometry.

Both autoinhibitory and catalyic domains have a strong potential for modulating phosphatases activity by the interaction with targeted molecules.

This methodology is can be also applied on authoinhibithed form of PP5 with the aim to test the bind of both already known and hypotized activators.

All the protein samples (50 μM) in a buffer containing Tris 50 mM, NaCl 150 mM, DTT 1mM, at pH 7.5 were analysed by relaxometer at 298 K. The NMRD profiles were collected with a Stelar fast field cycling relaxometer in the 0.01–40 MHz proton Larmor frequency range.

2.7.3 X-ray and protein crystallization

X-ray crystallography enables to visualize protein structures at the atomic level and enhances our understanding of protein function. Specifically it is possible to study how proteins interact with other molecules, how they undergo conformational changes, and how they perform catalysis in the case of enzymes. In all forms of microscopy, the amount of detail or the resolution is limited by the wavelength of the electro-magnetic radiation used. In order to see proteins in atomic detail we need to work with electro-magnetic radiation with a wavelength of around 0.1 nm or 1 Å.

The diffraction from a single molecule would be too weak to be measurable. So it is necessary to use an ordered three-dimensional array of molecules, in other words a crystal, to magnify the signal. Even a small protein crystal might contain a billion molecules. If the crystal is well ordered, then diffraction will be measurable at high angles or high resolution and a detailed structure should result. The X-rays are diffracted by the electrons in the structure and consequently the result of an X-ray experiment is a 3-dimensional map showing the distribution of electrons in the structure.

X-ray studies usually require substantial investment of time in order to optimize the crystallization conditions and obtain a crystal with good diffraction properties. This process can take weeks or months, but once a well-diffracting (<2.5Å) crystal is obtained, the structure determination can proceed quite quickly. In order to crystallize a protein, the purified protein undergoes slow precipitation from an aqueous solution. As result, individual protein molecules align themselves in a repeating series of "unit cells" by adopting a consistent orientation. The crystalline “lattice” that forms is held together by non-covalent interactions. The non-covalent bonds that hold together the lattice are often formed through several layers of solvent molecules. The production of good crystals is dependent upon a number of environmental factors because so much variation exists among proteins, with each individual requiring unique conditions for successful crystallization. Especially protein purity and pH conditions are very important; for example different pH values can result in different packing orientations.

The two most commonly used methods for protein crystallization are both vapor diffusion techniques. These are known as the “hanging drop” and **“**sitting drop” methods. Both entail a droplet containing purified protein, buffer and precipitant being allowed to equilibrate with a larger reservoir containing similar buffers and precipitants in higher concentrations. Initially, the droplet of protein solution contains an insufficient concentration of precipitant for crystallization, but as water vaporizes from the drop and transfers to the reservoir, the

precipitant concentration increases to a level optimal for crystallization. Since the system is in equilibrium, these optimum conditions are maintained until the crystallization is complete.

In the past few years macromolecular crystallography has become a standard technique used by many pharmaceutical and biotechnology companies. This methodology offers details of protein-ligand interactions at levels of resolution virtually unmatched by any other technique, and this approach holds the promise of novel, more effective, safer and cheaper drugs. Although crystallography remains a laborious and rather expensive technique, remarkable advances in structure determination and structure based drug design (SBDD) have been made in recent years.

Crystals of PP5c were obtained using the vapour diffusion technique at 289 k from solutions containing MPD 40% and PEG MME 5000 20%. Crystals of the protein dialyzed against Mn^{2+} (PP5-2Mn) and from the protein dialyzed against Zn^{2+} (PP5-2Zn) have been obtained. The final protein concentration was in all cases 1.1 mM. Complexes of PP5c with cantharidin and structural analogues have been obtained by socking tecnique. The experimantal structures have been provided insight into the structural determinats responsible of the ligand binding.

2.8 Reference List

1. Blundell TL, Sibanda BL, Montalvao RW, Brewerton S, Chelliah V, Worth CL et al. Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philosophical Transactions of the Royal Society B-Biological Sciences 2006; 361(1467):413-423.

2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J et al. Initial sequencing and analysis of the human genome. Nature 2001; 409(6822):860-921.

3. Venter JC. The sequence of the human genome (vol 292, pg 1304, 2001). Science 2001; 292(5523):1838.

4. Kapust RB, Waugh DS. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 1999; 8(8):1668-1674.

5. Gruber DF, Pieribone VA, Porton B, Kao HT. Strict regulation of gene expression from a high-copy plasmid utilizing a dual vector system. Protein Expression and Purification 2008; 60(1):53-57.

6. Landy A. Dynamic, Structural, and Regulatory Aspects of Lambda Site-specific Recombination. Ann.Rev.Biochem. 58, 913-949. 1989.

7. Wu PSC, Ozawa K, Lim SP, Vasudevan SG, Dixon NE, Otting G. Cell-free transcription/translation from PCR-amplified DNA for high-throughput NMR studies. Angewandte Chemie-International Edition 2007; 46(18):3356-3358.

8. Geiser M, Cebe R, Drewello D, Schmitz R. Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. Biotechniques 2001; 31(1):88-+.

9. Boettner M, Prinz B, Holz C, Stahl U, Lang C. High-throughput screening for expression of heterologous proteins in the yeast Pichia pastoris. J Biotechnol 2002; 99(1):51-62.

10. Shimizu Y, Kuruma Y, Ying BW, Umekage S, Ueda T. Cell-free translation systems for protein engineering. Febs Journal 2006; 273(18):4133-4140.

11. Klammt C, Schwarz D, Lohr F, Schneider B, Dotsch V, Bernhard F. Cell-free expression as an emerging technique for the large scale production of integral membrane protein. Febs Journal 2006; 273(18):4141-4153.

12. Ozawa K, Wu PSC, Dixon NE, Otting G. N-15-Labelled proteins by cell-free protein synthesis - Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. Febs Journal 2006; 273(18):4154-4159.

13. Staunton D, Schlinkert R, Zanetti G, Colebrook SA, Campbell LD. Cell-free expression and selective isotope labelling in protein NMR. Magn Reson Chem 2006; 44:S2-S9.

14. Vinarov DA, Newman CLL, Markley JL. Wheat germ cell-free platform for eukaryotic protein production. Febs Journal 2006; 273(18):4160-4169.

15. Apponyi M, Ozawa K, Dixon N, Otting G. Cell-free protein synthesis for analysis by NMR spectroscopy. in Kobe B, Guss M, Huber T editors.Methods in Molecular Biology,Structural Proteomics: High-Throughput Methods.Totowa 426, 257-268. 2008.

16. Shirokov V.A. KAKVAaAS. Continuous-Exchange Protein-Synthesizing Systems. Methods Mol.Biol. 375, 19-55. 2007.

17. Jun S.Y.Continuous-exchange cell-free protein synthesis using PCR-generated DNA and an RNase E-deficient extract). Biotechniques 44, 387-391. 2008.

18. Ozawa K. Cell-free in vitro protein synthesis in an autoinduction system for NMR studies of protein-protein interactions. J Biomol NMR 32, 235-241. 2005.

19. Renesto P, Raoult D. From genes to proteins - In vitro expression of rickettsial proteins. Rickettsiology: Present and Future Directions 2003; 990:642-652.

20. Ozawa K, Dixon NE, Otting G. Cell-free synthesis of N-15-labeled proteins for NMR studies. IUBMB Life 2005; 57(9):615-622.

21. Klammt C, Lohr F, Schafer B, Haase W, Dotsch V, Ruterjans H et al. High level cell-free expression and specific labeling of integral membrane proteins. Eur J Biochem 2004; 271(3):568-580.

22. Matsuda T.. Cell-free synthesis of zinc-binding proteins. Struc Funct Genomics 7, 93-100. 2006.

23. Yamane T, Ikeda Y, Nagasaka T, Nakano H. Enhanced cell-free protein synthesis using a S30 extract from Escherichia coli grown rapidly at 42 degrees C in an amino acid enriched medium. Biotechnology Progress 2005; 21(2):608-613.

24. Calhoun KA SJ. Energy systems for ATP regeneration in cell-free protein synthesis reactions. Methods Mol.Biol. 375, 3-17. 2007.

25. Kim DM, Kigawa T, Choi CY, Yokoyama S. A highly efficient cell-free protein synthesis system from Escherichia coli. Eur J Biochem 1996; 239(3):881-886.

26. Kigawa T, Yabuki T, Yoshida Y, Tsutsui M, Ito Y, Shibata T et al. Cell-free production and stable-isotope labeling of milligram quantities of proteins. FEBS Letters 1999; 442(1):15-19.

27. Kigawa T, Yamaguchi-Nunokawa E, Kodama K, Matsuda T, Yabuki T, Matsuda N et al. Energy systems for ATP regeneration in cell-free protein synthesis reactionsv. J.Struct.Funct.Genomics 2, 29-35. 2001.

28. Ozawa K, Headlam MJ, Mouradov D, Watt SJ, Beck JL, Rodgers KJ et al. Translational incorporation of L-3,4-dihydroxyphenylalanine into proteins. Febs Journal 2005; 272(12):3162-3171.

29. Ozawa K , Headlam MJ., Schaffer P M., Henderson B R., Dixon N E., Otting G. Optimization of an Escherichia coli system for cell-free synthesis of selectively 15N-labelled proteins for rapid analysis by NMR spectroscopy. Eur J Biochem 271, 4084-4093. 2004.

30. Guignard L, Ozawa K, Pursglove SE, Otting G, Dixon NE. NMR analysis of in vitro-synthesized proteins without purification: a high-throughput approach. FEBS Letters 2002; 524(1-3):159-162.

31. Keppetipola S, Kudlicki W, Nguyen BD, Meng X, Donovan KJ, Shaka AJ. From gene to HSQC in under five hours: High-throughput NMR proteomics. J Am Chem Soc 2006; 128(14):4508-4509.

32. Zhang ZY, Dixon JE. Protein-Tyrosine Phosphatases - Mechanism of Catalysis and Substrate-Specificity. Advances in Enzymology and Related Areas of Molecular Biology, Vol 68 1994; 68:1-36.

33. McCain DF, Zhang ZY. Protein tyrosine phosphatase: enzymatic assays. Methods Enzymol 345, 507-518. 2002.

34. Takai A, Mieskes G. Inhibitory Effect of Okadaic Acid on the Para-Nitrophenyl Phosphate Phosphatase-Activity of Protein Phosphatases. Biochemical Journal 1991; 275:233-239.

35. Wüthrich K. The second decade--into the third millenium. Nat Struct Biol 1998; 5 Suppl:492-495.

36. Halle B, Denisov VP. Magnetic relaxation dispersion studies of biomolecular solutions. Nuclear Magnetic Resonance of Biologica Macromolecules, Pt A 2001; 338:178-201.

37. Bertini I, Luchinat C, Parigi G. Solution NMR of paramagnetic molecules. Elsevier, Amsterdam 2001 . 2008.

38. Fisher JF, Mobashery S. Recent advances in MMP inhibitor design. Cancer and Metastasis Reviews 2006; 25(1):115-136.

39. Yang J, Roe SM, Cliff MJ, Williams MA, Ladbury JE, Cohen PTW et al. Molecular basis for TPR domain-mediated regulation of protein phosphatase. Embo Journal 2005; 24(1):1-10.

40. Swingle MR, Honkanen RE, Ciszak EM. Structural basis for the catalytic activity of human serine/threonine protein phosphatase-5. Journal of Biological Chemistry 2004; 279(32):33992-33999.

41. Schmid AC, Woscholski R. Phosphatases as small-molecule targets: inhibiting the endogenous inhibitors of kinases. Biochemical Society Transactions 2004; 32:348-349.

42. Bernhard F. Cell-free expression of integral membrane proteins. EMBL Workshop, Heidelberg, 13-19. 9. 2006.

3

RESULTS

3.1

Biochemical and structural characterization of the binuclear metal site in Human Protein Phosphatase PP5[1]

[1]Ivano Bertini, Claudio Luchinat and Marco Fragai designed the research, Eleonora Talluri done all the Molecular Biology work, sample preparation and protein crystallization, Vito Calderone done the X-ray crystallography experiments.

Abstract

Three forms of the catalytic domain of Human Protein Serine/Threonine Phosphatase PP5 have been characterized. The form containing a monometallic center is catalytically inactive and anomalous dispersion data suggest that a manganese ion is present at the catalytic site. The other two forms contain one a Mn-Mn bimetallic center and the other a Zn-Zn center. The former is fully active while the Zn-Zn form is virtually inactive, thereby suggesting a possible biological role of the two metal ions in the regulation of PP5 activity. The three-dimensional structure of the enzyme has provided insight into the characterization of the nuclear metal site.

Introduction

Protein phosphatases reverse the function of their most known counterpart. These enzymes are involved in the hydrolysis of various phosphate esters and include acid and alkaline phosphatases, exonucleases, nucleotidases, diadenosine tetraphosphatases, cyclic nucleotide phosphodiesterases, sphingomyelin phosphodiesterases, the eukaryotic serine/threonine protein phosphatases, and a protein phosphatase from bacteriophage λ. They can be classified into three groups on the basis of sequence, structure and catalytic mechanism. The first group comprises the classic Ser/Thr phosphatases: the large phosphoprotein phosphatase (PPP) family (PP1, PP2A, PP2B, PP4, PP5, PP6 and PP7) and the protein phosphatase, Mg^{2+} or Mn^{2+} dependent (PPM) family (PP2C). The protein Tyr phosphatase (PTP) superfamily forms the second group, and the third group consists of the Asp-based protein phosphatases with a DXDXT/V catalytic signature[1]. Human protein phosphatase 5 (PP5) is a member of Serine/Threonine Phosphatases that dephosphorylates their substrate in serine and threonine residues. PP5 posses a series of three N-terminal TPR domains that function to keep the enzyme inactive and that make PP5 unique in its structural features[2,3]. Structural analysis of human PP5 showed that the mechanism of TPR inhibition of the phosphatase domain is similar to the mechanism of inhibition of PP2B by its C-terminal regulatory segment and to that of PP1 by toxins. The enzyme is cytoplasmic and nuclear in both humans and *S. cerevisiae*, and the C-terminus contains a conserved sequence that is required for nuclear localization [4,5].

PP5 is involved in regulation of rRNA transcription; it promotes cellular proliferation by binding to Hsp90; it promotes progression into S-phase and regulates progression through mitosis by binding to CDC16 and CDC27 of the anaphase-promoting complex, thus it is

assumed it has an effect on the development of cancer cells and Alzheimer's diseases [5-7]. PP5 can be considered as a promising therapeutic target for the treatment of many diseases.

The catalytic domain of PP5 (PP5c) shares 35–45% sequence identity with the catalytic domains of other PPP phosphatases, including protein phosphatase 1 (PP1), 2A (PP2A), 2B/calcineurin (PP2B), 4 (PP4), 6 (PP6), and 7 (PP7). As all the serine/threonine phosphatases, PP5 is a metalloprotein. Three motifs (DXH(X)nGDXXD(X)mGNHD/E) highly conserved in the catalytic domain of all phosphatases[8] contain most residues which coordinate metal ions in the active centre and are considered as the signature of the phosphatase family [5,9]. The X-ray structures of others related phosphatases (Protein phosphatase 1 and Protein phosphatase 2B) indicate that the phosphoesterase motif is represented as a β-α-β-α-β secondary element that provides the framework for an active site dinuclear metal cofactor [10]. The high degree of sequence conservation within the phosphoesterase motif in this family of enzymes suggests that other members will share a common active site architecture and catalytic mechanism. The bacteriophage λ protein phosphatase is a metal requiring enzyme by demonstrating that Mn^{2+} and Ni^{2+} are necessary for activation of the enzyme [10]. λ protein phosphatase can be reconstituted to contain a dinuclear Fe^{3+}-Fe^{2+} center, analogous to the Fe^{3+}-Zn^{2+} and Fe^{3+}-Fe^{2+} centers of purple acid phosphatase further advancing the hypothesis that enzymes containing the phosphoesterase motif utilize an active site dinuclear metal center [10]. The identity of the physiological metal ion cofactors for most phosphoesterases is still unknown, but a Fe^{3+}-Zn^{2+}, Mn^{2+}/Fe^{2+} and Mn^{2+}/Mn^{2} metal center was identified in PP2B[11], PP1 and PP2A respectively[10]. Nevertheless, calcineurin phosphatase activity can be enhanced by the presence of a divalent metal ion, with Mn^{2+} and Ni^{2+} providing the highest level of activation [10]. The mechanism for activation is not known, but in PP2A the addition of a divalent metal such as Mn^{2+} appears to prevent a time-dependent inactivation process that may involve metal substitution[12].

X-ray crystal structure of PP1 and PP2B revealed in their catalytic subunit a metal-bound water molecule acting as a nucleophile to attack the phosphorus atom of a phosphate group in a S_N2 mechanism. Metals assist in this catalysis as Lewis acids. Protein phosphatase 2C too contains binuclear metal ions (2 Mn^{2+}) in its catalytic site that promote protein dephosphorylation by activation of water molecules to catalyse hydrolysis in a single-step reaction. [3]. The presence of two metal in the catalytic pocket of PP5 is revealed by X-ray analysis[3,13] elucidating the catalytic mechanism of action also for PP5 .From this structure

the authors propose a mechanism for PP5-mediated hydrolysis of phosphoprotein substrates, which requires the precise positioning of two metal ions within a conserved Asp^{271}-M_1:M_2-W^1-His^{427}-His^{304}-Asp^{274} catalytic motif (where M_1 and M_2 are metals and W^1 is a water molecule) [3,13]. The importance to investigate the nature of the metal center in PP5 is also due to the fact that metal-substitution studies on this enzyme have shown that manganese ion can be replaced by zinc with loss of activity [3,14]. This observation suggest a physiological role of manganese for PP5. The presence of manganese in the catalytic pocket of related phosphatases is also reported [15].

Structural analysis of both the Mn-Mn and Zn-Zn enzyme have been provided insight into the structure of this metal center. The catalytic domain of human PP5 (residues 169-499) has been here investigated.

Materials and Methods

The catalytic domain of PP5 (PP5c) was expressed by using Gateway tecnology as His-GST-binding fusion protein. The Gateway expression vector pDEST30 was used to transform *E. Coli* BL21 (DE3) RIPL strain. PP5c was purified by Glutathione Sepharose affinity chromatography (Amersham). The His-GST tag was removed by the TEV protease digestion according to the Invitrogen manifactures. The cutting leaves four vector-derived amino acid residues (Gly-Ser-Phe-Thr) at the N-terminus of PP5c. Ni^{2+}-saturated HiTrap Chelating column (Amersham) and Superdex 75 26/60 size exclusion chromatography (Amersham) were used to further purify the sample. Protein sample was concentrated to 0,1mM as determined by Bradford assay and UV spectrophotometer. Then, a fraction of a sample was dialyzed against 50mM Tris-Hcl,pH 7.5, 150mM NaCl, 1mM DTT and 0.2mM $MnCl_2$; an other fraction was dialyzed in the same buffer containing $ZnCl_2$ instead of manganese. The rest of the sample was in 50mM Tris-Hcl,pH 7.5, 150mM NaCl, 1mM DTT. Dialysis have been made in the ratio 1mL of protein per liter of solution and leaved to proceed for 6 hours. A second dialysis was performed in the same way. The last dialysis has been done in 4 hours in a metal-free solution to remove excess of the metal and samples have been analyzed by ICP-mass.

Activity has been tested against p-nitropenyl phosphate (pNPP) according to the Jena Bioscience protocols and expressed as Unit/µg enzyme were one unit is the amount of the phosphatase required to release 1 nmol of phosphate from pNPP in one minute at 298 K

under the tested assay conditions. A pNPP molar extinction coefficient of 17.8 μL·nmol-1·cm-1 was considered.

All the protein samples have been then concentrated until 1.1mM . Crystals of PP5 were obtained using the vapour diffusion technique at 289 k from solutions containing 10mM Tris-HCl pH 8.0, MPD 40% and PEG MME 5000 20%.

The PP5-2Mn dataset and the PP5-2Zn were measured in-house, using a PX-Ultra copper sealed tube source (Oxford Diffraction) equipped with an Onyx CCD detector, whereas the non-dialyzed PP5 was measured using synchrotron radiation at BM16 beamline (ESRF, Grenoble, France). All datasets were collected at 100 K and the crystals used for data collection were cryo-cooled without any cryo-protectant treatment.
The data were processed in all cases using the program MOSFLM[16] and scaled using the program SCALA[17]with the TAILS and SECONDARY corrections on (the latter restrained with a TIE SURFACE command) to achieve an empirical absorption correction. **Table 1** shows the data collection and processing statistics for all datasets.

Table 1. Data collection and refinement statistics.

	PP5	PP5-2Mn	PP5-2Zn
Spacegroup	$P4_12_12$	C2	C2
Cell dimensions (Å, °)	a=b= 67.99 c= 200.62	a= 154.34 b= 41.65 c= 105.67 β= 96.85	a= 154.25 b= 41.68 c= 106.09 β= 96.46
Resolution (Å)	34.5 – 2.0	38.2 – 2.0	38.3-2.6
Unique reflections	32495 (4530)	45387 (6419)	19742 (2660)
Overall completeness (%)	99.0 (97.7)	99.3 (97.3)	93.3 (87.2)
R_{sym} (%)	10.0 (69.4)	17.3 (31.4)	21.1 (45.3)
Multiplicity	11.7 (4.0)	4.4 (2.8)	3.9 (3.2)
I/(σI)	19.8 (1.6)	8.4 (4.1)	5.8 (2.7)
Wilson plot B-factor ($Å^2$)	27.94	0.9	13.2
R_{cryst} / R_{free} (%)	19.8 / 25.9	21.6 / 27.7	19.9 / 29.1
Protein atoms	2596	5058	5058
Water molecules	194	306	56
Ions per monomer	1	2	2
RMSD bond lengths (Å)	0.031	0.030	0.052
RMSD bond angles (°)	2.57	2.35	3.70
Mean B-factor ($Å^2$)	39.03	2.93	3.08

The structures were solved using the molecular replacement technique; the model used for all datasets was 1S95, where water molecules and ions were omitted from the models. The correct orientation and translation of the molecule within the crystallographic unit cell was determined with standard Patterson search techniques [18,19,19] as implemented in the program MOLREP[20,21].

The isotropic refinement was carried out using REFMAC5 [22] on all datasets. REFMAC5 default weights for the crystallographic and the geometrical term have been used in all cases.

In between the refinement cycles the models were subjected to manual rebuilding by using XtalView [23]. Water molecules have been added by using the standard procedures within the ARP/WARP suite [24]. The stereochemical quality of the refined models was assessed using the program Procheck.[25] The Ramachandran plot is in all cases of good quality.

Results and Discussion

PP5 is involved in signaling pathways that control cellular responses to stress, glucocorticoids and DNA damage. The catalytic domain of PP5 (PP5c) shares 35-45% sequence identity with the catalytic domain of other PPP phosphatases. This enzyme is sensitive to inhibition by okadaic acid, microcystin, cantharidin, tautomycin and calyculin A. PP5 is a two domains protein since it is formed by a catalytic domain and by an extended N-terminal domain containing three tetratricopeptide repeat (TPR) motifs. The TPR domain acts as an interface for protein-protein interactions and provides a means of targeting PP5 to particular regions within the cell and facilitating interactions with substrates. The TPR domain engages with the catalytic channel of the phosphatase domain, restricting access to the catalytic site.

All the members of the phosphatase family share a common phosphoesterase motif and an active site dinuclear metal center. Despite these similarities in sequence, each phosphatase in the family appears to have different metal ion requirements. The hallmark of metallophosphatases is a dinuclear center with separation between metal ions from 3.1 to 3.4 Å. The catalytic site of PP5, as obtained in a recent work [13] (PDB code: 1S95) and by us, can be identified from the position of these two metal ions. It is located at the base of a shallow depression on the surface formed by the residues of four loops (Fig.1). Anomalous dispersion from non-dialyzed PP5 crystals revealed the presence of a Mn^{2+} in a mononuclear binding site (see below) but the presence of Fe^{2+} or Zn^{2+} cannot be ruled out.

The presence of manganese, iron and zinc in the protein sample has been also revealed by ICP-mass (data not shown). X-ray fluorescence and anomalous dispersion from PP5 crystals obtained from the protein dialyzed against Mn^{2+} (PP5-2Mn hereafter) revealed the presence of two Mn^{2+} ions in the metal site. On the other hand, X-ray fluorescence and anomalous dispersion from PP5 crystals obtained from the protein dialyzed against Zn^{2+} (PP5-2Zn hereafter) revealed the presence of two Zn^{2+} ions in the metal site. The structures of PP5-2Mn and PP5-2Zn have an RMSD of 0.21 Å, and the catalytic sites are perfectly superimposable (Fig. 2A).

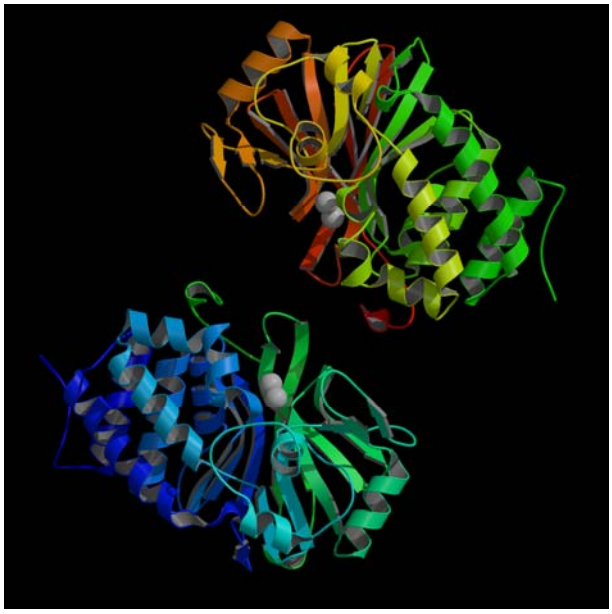

Fig.1 Ribbon representation of the crystallization unit of the catalytic domain of PP5 in complex with manganese ions (Green spheres). The figure was generated with PyMol..

Nevertheless, PP5-2Zn shows no catalytic activity against the *p*NPP substrate, while PP5-2Mn retains full enzymatic activity. The affinity constant of manganese for PP5 was found to be around 10μM. This value has been obtained by *p*NPP enzimatic assay. Interestingly, the incubation of the zinc-loaded PP5 with different concentration of Mn(II) did not increase the activity of the protein, thereby suggesting a larger affinity of PP5 for zinc ion.
In this work we refer to the metal ions as M1 and M2, based upon the coordinating residues. Each metal ion is coordinated by five ligands (Fig. 2B), with the bridging carboxylate oxygen of Asp271 and a water molecule (W1) coordinating both M1 and M2. Asp242, His244 and W2 also coordinate M1, and Asn303, His352 and His427 coordinate M2. The metals are 3.3 Å apart. This metal binding site corresponds to the one already reported in literature[13] where a phosphate ion provides an additional oxygen atom as a ligand for both metals. Interestingly, the structure of the non-dialyzed PP5 obtained in the

present work is quite different from the one already present in literature and from both PP5-2Zn and PP5-2Mn.

The non-dialyzed protein crystallizes in the space group $P4_12_12$ with one molecule in the asymmetric unit, at variance with 1S95, PP5-2Zn and PP5-2Mn, which belong to space group C2 with two molecules in the asymmetric unit.

The catalytic site of non-dialyzed PP5 has only one metal, namely M2, which shows a similar coordination to the one in PP5-2Zn and PP5-2Mn, where M1 is replaced by a water molecule (see Fig.3).

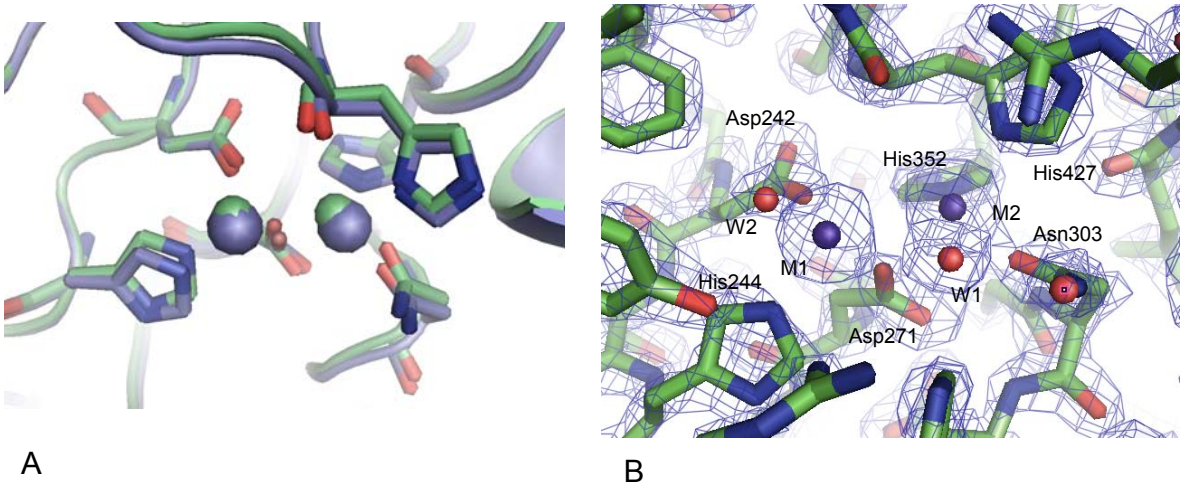

Fig. 2 Crystal structures of PP5-2Mn (Blue) and PP5-2Zn (Green). Superimposition of the two structures shows that no structural differences are appreciated (A). The ligands that coordinates the metal ions are shown in B.

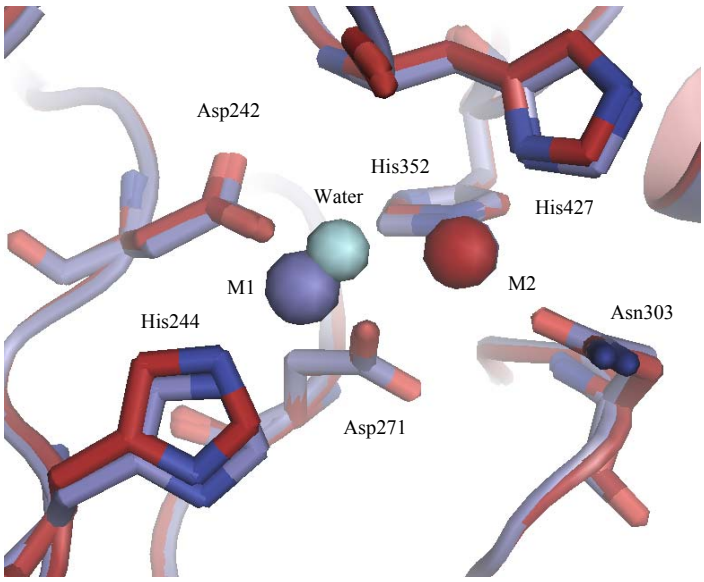

Fig. 3 Superimposition of crystal structure of not-dialyzed PP5 (Red) and PP5-2Zn (Blue). A monometallic center is present at the catalytic site of not-dialyzed PP5. Differently, the metal-loaded PP5 contains a binuclear metal center.

The structures of PP5-2Mn and non-dialyzed PP5 have an RMSD of 0.29 Å, and the same RMSD is found between the structures of PP5-2Zn and non-dialyzed PP5. The different crystal packing and the lack of a metal in non-dialyzed PP5 is certainly due to the presence of the C-terminal of another molecule belonging to another asymmetric unit in the catalytic site. The C-terminal of PP5 is actually present in the active site in the auto inhibited form of the protein, as described by Yang *et al.*[3]. In the dialyzed protein the presence of the two ions in the active site hinders the interaction with the C-terminal of another adjacent molecule in such a way that the last ten residues are not stabilized by any interaction: they are therefore extremely mobile and do not show any visible electron density.

It has been proposed a catalytic mechanism for PP5 in which the nucleophile is a bridging hydroxide ion ligated to both metal ions in the active site and the leaving group oxygen does not coordinate an active site metal. In the proposed metal-ion mediated hydrolysis of phosphorylated substrates, one bridged hydroxide serves as the nucleophile that attacks the phosphorus atom of the substrate.It has been proposed[13] that is is W1 that provides the bridging hydroxide which serves as the nucleophile attacking the phosphorous atom of the subtrate phosphoryl moiety.

The interaction of W1 with His427 and the dinuclear metal center helps fix the orientation of a lone pair of electrons from the hydroxide W1 toward the phosphorous atom, thereby helping to reduce the entropic barrier for nucleophilic attack.

Conclusions

The biochemical and structural characterization of the binuclear active site of PP5 has provided relevant information on the nature and on the role of the metal ions responsible for the catalytic activity of the enzyme. The ICP analysis and partially the X-ray investigation revel that manganese, iron, zinc and can be present at the catalytic site. A full-loaded forms of the protein with manganese or zinc have been prepared by exhaustive dialysis in presence of submillimolar concentration of the metal ion in solution. However, only the protein containing manganese exhibits full enzymatic activity suggesting a possible roles for these two metal ions in the regulation of PP5 function. In this respect also the different affinity for PP5, observed for the two metal ions, could play a crucial role.

Reference List

1. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 2006; 127(3):635-648.

2. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends in Biochemical Sciences 2003; 28(12):655-662.

3. Yang J, Roe SM, Cliff MJ, Williams MA, Ladbury JE, Cohen PTW et al. Molecular basis for TPR domain-mediated regulation of protein phosphatase. Embo Journal 2005; 24(1):1-10.

4. Chen MX, Mcpartlin AE, Brown L, Chen YH, Barker HM, Cohen PTW. A Novel Human Protein Serine/Threonine Phosphatase, Which Possesses 4 Tetratricopeptide Repeat Motifs and Localizes to the Nucleus. Embo Journal 1994; 13(18):4278-4290.

5. Chen MS, Silverstein AM, Pratt WB, Chinkers M. The tetratricopeptide repeat domain of protein phosphatase 5 mediates binding to glucocorticoid receptor heterocomplexes and acts as a dominant negative mutant. Journal of Biological Chemistry 1996; 271(50):32315-32320.

6. Ollendorf V, Donoghue DJ. The serine/threonine phosphatase PP5 interacts with CDC16 and CDC27, two tetratricopeptide repeat-containing subunits of the anaphase-promoting complex. Journal of Biological Chemistry 1997; 272(51):32011-32018.

7. Liu F, Iqbal K, Grundke-Iqbal I, Rossie S, Gong CX. Dephosphorylation of tau by protein phosphatase 5 - Impairment in Alzheimer's disease. Journal of Biological Chemistry 2005; 280(3):1790-1796.

8. Barton GJ, Cohen PTW, Barford D. Conservation Analysis and Structure Prediction of the Protein Serine/Threonine Phosphatases - Sequence Similarity with Diadenosine Tetraphosphatase from Escherichia-Coli Suggests Homology to the Protein Phosphatases. Eur J Biochem 1994; 220(1):225-237.

9. Barford D. Molecular mechanisms of the protein serine threonine phosphatases. Trends in Biochemical Sciences 1996; 21(11):407-412.

10. Rusnak F, Yu LA, Mertz P. Metalloenzymes and signal transduction: The protein serine/threonine phosphatases, a novel class of binuclear metal-containing enzymes. JBIC 1996; 1(5):388-396.

11. Namgaladze D, Hofer HW, Ullrich V. Redox control of calcineurin by targeting the binuclear Fe2+-Zn2+ center at the enzyme active site. Journal of Biological Chemistry 2002; 277(8):5962-5969.

12. Nishito Y, Usui H, Shinzawa-Itoh K, Inoue R, Tanabe O, Nagase T et al. Direct metal analyses of Mn2+-dependent and -independent protein phosphatase

2A from human erythrocytes detect zinc and iron only in the Mn2+-independent one. FEBS Letters 1999; 447(1):29-33.

13. Swingle MR, Honkanen RE, Ciszak EM. Structural basis for the catalytic activity of human serine/threonine protein phosphatase-5. Journal of Biological Chemistry 2004; 279(32):33992-33999.

14. Bertini I, Fragai M, Luchinat C, Talluri E. Water-based ligand screening for paramagnetic metalloproteins. Angewandte Chemie-International Edition 2008; 47(24):4533-4537.

15. Reiter TA, Reiter NJ, Rusnak F. Mn2+ is a native metal ion activator for bacteriophage lambda protein phosphatase. Biochemistry 2002; 41(51):15404-15409.

16. Leslie AGW. In Crystallographic Computing V. In: Moras D, Podjarny AD, Thierry J-C, editors. Molecular data processing. Oxford: Oxford University Press, 1991: 50-61.

17. "Data Reduction", Proceedings of CCP4 Study Weekend.: 1993.

18. Rossmann MG, Blow DM. The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst 1962; D15:24-31.

19. Crowther RA. The Molecular Replacement Method. In: Rossmann MG, editor. New York: Gordon & Breach, 1972.

20. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Crystallogr 1997; 30:1022-1025.

21. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr 2000; 56:1622-1624.

22. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Cryst 1997; D53:240-255.

23. McRee DE. "XtalView: A Visual Protein Crystallographic Software System for XII/XView". J Mol Graphics 1992; 10:44-47.

24. Lamzin VS, Wilson KS. Automated refinement of protein models. Acta Crystallogr D Biol Crystallogr 1993; 49(Pt 1):129-147.

25. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993; 26:283-291.

3.2

Crystal structure of Cantharidin with the catalytic domain of Human Serine/Threonine Phosphatase PP5[1]

[1]Ivano Bertini, Claudio Luchinat and Marco Fragai designed the research, Eleonora Talluri done all the Molecular Biology work, sample preparation and protein crystallization, Vito Calderone done the X-ray crystallography experiments.

Abstract

Human Protein Serine/Threonine Phosphatase PP5 is inhibited by a number of natural toxins[13] However, the binding modes of these inhibitors with PP5 are unknown. To elucidate the binding mode of selected inhibitors to PP5, it is necessary to determine the three-dimensional structure of their complex. Among well known PPP inhibitors, cantharidin, cantharidic acid and endothal have been selected in this study. Their low molecular weight may facilitate a further structure –based drug design. The crystal structures of the catalytic domain of PP5 in complex with cantharidin, cantharidic acid and endothal have been solved, providing interesting informations on the determinants responsible of the binding.

Introduction

The reversible phosphorylation of serine/threonine residues in proteins has been known to play a crucial role in cellular signal transduction. Protein-serine/threonine phosphatases (PPP) which catalyze the reverse reaction of kinases , comprise protein phosphatase 1, 2A, 2B, 4, 5, 6, 7 and 2C. PP1, PP2A ,PP2B, PP4, PP5, PP6 and PP7 have highly homologous catalytic domains but differ in their substrate specificities and interactions with regulatory molecules, whereas PP2C shares no sequence homology to the others[1].

The crystal structures of the catalytic subunits of PP1, PP2A, PP2B and PP5 were recently solved. These structures revealed a common motif containing a dinuclear metal ion center located at the active site and every active site residue involved in metal coordination or implicated in catalysis is strictly conserved[2].

PP5 differs from PP1-PP7 in that it contains an N-terminal domain characterized by the presence of three tetratricopeptide repeats which posses an authoinhibitory funciton and mediate protein protein interaction. Moreover, a flexible linker region connects the N-terminal domain to the catalytic domain that is strickingly similar to that of the catalytic subunit of other related phosphatases. A C-terminal subdomain helps to mantein a closed, inactive conformation by stabilizing the N-terminal domain over the catalytic site, thereby blocking the substrate access [3,4].

Recent findings suggest the interaction of PP5 with critical substrates is regulated by cellular stress[5]. PP5 is required for activation of the DNA-damage-activated protein kinase. Both knock down of PP5 by antisense and overexpression of catalytically inactive phosphatases attenuated activation of ATM by ionizing radiation and allowed DNA synthesis in the presence of radiation damage Hypoxia, oxidative stress and rapamycin

induce apoptosis in sensitive cell lines in a pathway that requires activation of the apoptosis signal regulated kinase 1 (ASK1). Several recent reports suggest that PP5 regulates ASK1-dependent apoptosis and may promote tumor growth. PP5 binds to and inhibits ASK1, blocking activation of MKK4 and JNK Over-expression of PP5 represses ASK1 activation, blocks apoptosis after hypoxia and promotes tumor growth in a MCF-7 mouse xenograph model [Rapamycin inhibits the activity of the protein kinase mTOR and induces a stress response that includes activation of ASK1. In sensitive cells this results in apoptosis. ASK1 activation correlated with PP5 inhibition or depletion, and over-expression of PP5 prevented ASK1

activation and blocked apoptosis in a rhabdomyosarcoma cell line Hence, one major role of PP5 seems to be regulating the activation of stress and DNA-damage-regulated protein kinases. PP5 is assumed it has also an effect on the development of cancer cells and Alzheimer's diseases[5-7]. Consequently, PP5 is a promising therapeutic target in the treatment of many diseases. Removal of the N-terminal domain leads to the activation of PP5 [8]. Successful development of drugs may require targeting of the phosphatase domain even though in some cases PP5 is needed [9].

The structural features of the PP5 were reveled by the high resolution X-ray analysis [4]. In particular, the structure of PP5 catalytic domain shows that the essential catalytic motif contains six conserved features: Asp274, His304, Asp271, the backbone carbonyl of His427, the two active site metal ions (M1 and M2), and W1 (water/hydroxide coordinated to the two active site metal ions). Sequence alignments and comparisons to other crystal structures reveal that this motif is common to all members of the PPP family phosphatases, as is the hydrogen bond network that ensures this catalytic activity. Structure-based sequence alignment of PP5 with the eukaryotic PPP family members reveals an active site motif consisting of ten conserved amino acids, Asp^{242}-Xaa^{243}-His^{244}-$Xaa_{26–27}$-Asp^{271}-Xaa^{272}-Xaa^{273}-Asp^{274}-Arg^{275}-Xaa_{28}-Asn^{303}-His^{304}-Xaa_{48}-His^{352}-$Xaa_{48–54}$-Arg^{400} Xaa_{27}-His^{427}. Six of these amino acids act as metal-coordinating residues (Asp^{242}, His^{244}, Asp^{271}, Asn^{303}, His^{352}, and His^{427}), and four (Arg^{275}, Asn^{303}, His^{304}, and Arg^{400}) help position the phosphate ion through hydrogen bonds to O^1, O^2, O^3, and O^4 [4]. The arginine residues along with His^{304} when protonated and the metal ions create the only large pocket of positive electrostatic potential on the surface of PP5c.

The dinuclear metal site is situated in a unique protein fold, a β-α-β-α-β motif which provides the majority of ligands to the metal ions[10]. A similar fold is also seen in plant purple acid phosphatases, which also contain a dinuclear iron–zinc cofactor. In these enzymes, the two metal ions are bridged by a solvent molecule and a carboxylate group

from an aspartic acid residue, juxtaposing the two metal ions. A similar motif has been identified in a number of other enzymes which exhibit phosphoesterase activity, implicating several of them as metalloenzymes which contain dinuclear metal ion cofactors[10]. In the catalytic site of PP5, each metal ion is coordinated by six ligands in a slightly distorted octahedral geometry, with the bridging carboxylate oxygen of Asp271 and W1 each coordinating both M1 and M2 Asp242, His244, W2, and O1 of the phosphate ion also coordinate M1, and Asn303, His352, His427, and O2 of the phosphate coordinate M2. The phosphate ion coordinates the metal ions in a bidentate mode, indicating that the close spacing of the metal ions is necessary for catalysis. The nature of the metal ions present at the active site has been longely debuted. The biochemical characterization performed on the purified protein showed that Mn, Zn as well as Fe can be present at the catalytic site.

The larger enzymatic activity of Mn(II)-PP5 versus the Zn(II)-loaded protein suggests a physiological role for the manganese ion [11].

The necessary alignment of substrate and nucleophile is facilitated by substrate contacts with M1, M2, Arg275, Asn303,His304, and Arg400 and interactions of W1 with M1, M2, and His427. Thus, there are many sites for inhibitor binding that should disrupt catalytic activity[5]. Despite this, the development of a specific inhibitor may be challenging, as the catalytic core of PP5 is highly homologous to other eukaryotic PPP family phosphatases. In addition, both structural and mutational studies indicate that PP1, PP2A, PP2B and PP5 share a common catalytic mechanism that is most likely common to the whole PPP family[5]. Superposition of the structure of PP5c onto the structures of other eukaryotic PPP phosphatases such as PP1, PP2A and PP2B gives root mean square deviations of <2.0Å within the highly homologous ~270-residue region used in the calculations. Nevertheless, comparisons of the loop regions of PP5, PP1, PP2A and PP2B reveal significant variation that may be relevant for drug discovery efforts, such as sequence/conformational differences in the β12/β13 loop that plays an important role in microcystin-LR and okadaic acid-mediated inhibition of catalytic activity[5]. Such structural differences closely apposed to the conserved catalytic site suggest the feasibility of developing type-specific inhibitors[5]. Many commercial drugs inhibits this pathologically relevant metalloenzyme interacting directly with the catalytic metal ion or preventing the access of the physiologic substrate to the active site [12,13]. To date several compounds that inhibit the activity of PP5 have been identified (i.e. okadaic acid, microcystins, cantharidin, calyculin A, tautomycin and fostriecin)[13]. Most of these compounds were originally derived from extract of natural products, and many were initially identified as eukaryotic cell toxins. The organisms that

make these inhibitors are biologically diverse, with potent inhibitors produced by dinoflagellates (i.e. okadaic acid), cyanobacteria (i.e. variants of microcystin and nodularin), Streptomyces (i.e. tautomycin and fostriecin) and insects (cantharidin). The ultimate natural source of calyculin A, a potent inhibitor identified from a marine sponge extract, is still unknown. PP5 is sensitive to inhibition of okadaic acid, microcystins, calyculin A ,tautomycin cantharidin at nanomolar concentrations [11,14].

Cantharidin (exo. exo-2. 3-dimethyl-7-oxobicyclo [2.2.1] heptane-2,3-dicarboxylic acid anhydride), is a major component of the Chinese blister beetles: Mylabris phalerata or Al. cichorii). The dried body of these beetles has been used by the Chinese as a natural remedy for the past 2000 years. Although Western medicine decreed cantharidin to be too toxic in the early 1900's. Its purported aphrodisiac qualities (the active ingredient of"Spanish Fly"), and its widespread occurrence in cattle feed still results in numerous human and livestock poisonings[15]. The cellular mechanisms underlying the toxicity of cantharidin have been under investigation for several years. Cantharidin interfers with mitochondrial respiration and was then shown to bind with high affinity to a specificand saturable binding site in a cytosolic fraction produced from mouse liver. This high-affinity cantharidin binding site was subsequently identified in several other tissues including heart. kidney, lung, pancreas, kin. spleen, brain, blood and stomach.

Due to its low molecular weight, cantharidin can be considered as a lead to indentify structural determinants of the binding mode for a selective design of drugs against PP5.

In recent years, there has been intense interest in the development of potent and selective inhibitors of the serine/threonine class of protein phosphatases, PP1 and PP2A. These attempts have included analogues of cantharidin [16] and the more complex toxins such as the microcystin which resulted in moderate PP1 selectivity [17].

Of all the known naturally occurring toxins, cantharidin represents by far the simplest synthetic target. The synthesis of numerous analogues of cantharidin, and their inhibition of PP1 and PP2A is reported [18].

Here the catalytic domain of PP5 in complex with the inhibitors cantharidin, cantharidic acid and endothal has been investigated. Crystal structure of PP5 complexed with these molecules has been solved revealing interesting feautures in the binding mode that support further structure-based drug-design.

Materials and methods

The catalytic domain of PP5 corresponding to the construct 169-449 has been cloned in the expression vector pDEST30 by Gateway technology and expressed in soluble form in *E.*

Coli BL21(DE3) RIPL. The growing conditions were optimized in fermenters and the GST-fused protein has been purified by means of both affinity and size exclusion chromatography after the removal of the tag.

Samples of the purified PP5 protein were dialyzed against a solution containing MnCl2 in order to produce a Mn-Mn enzyme. ICP was performed to verify the metal content. The enzymatic activity of the dialyzed sample has been assayed by monitoring the dephosphorylation of the p-nitro-phenol-phosphate (*p*NPP) substrate. The inhibition of dephosphorylation of *p*NPP was monitored after the addition of selected compounds to the reaction misture.

Crystals of the 2Mn-PP5 (1.1mM) were obtained using vapor diffusion technique at 16°C from a solution containing 10mM Tris-HCl pH 8.0, 40% MPD and 20% PEG MME 5000.

Crystals of PP5-2Mn in complex with the inhibitors endothal, cantharidin and cantharidic acid were obtained through soaking PP5-2Mn crystals with the mother liquor containing the inhibitor itself.

The PP5-2Mn complexed with cantharidin, cantharidic acid and endothal were acquired at beamline ID23-1 (ESRF, Grenoble, France).

All datasets were collected at 100 K and the crystals used for data collection were cryo-cooled without any cryo-protectant treatment.

The data were processed in all cases using the program MOSFLM[19] and scaled using the program SCALA[20] with the TAILS and SECONDARY corrections on (the latter restrained with a TIE SURFACE command) to achieve an empirical absorption correction.

The structures were solved using the molecular replacement technique; the model used for all datasets was 1S95, where water molecules and ions were omitted from the models. The correct orientation and translation of the molecule within the crystallographic unit cell was determined with standard Patterson search techniques [21,22,22] as implemented in the program MOLREP[23,24].

The isotropic refinement was carried out using REFMAC5 [25] on all datasets. REFMAC5 default weights for the crystallographic and the geometrical term have been used in all cases.In between the refinement cycles the models were subjected to manual rebuilding by using XtalView [26]. Water molecules have been added by using the standard procedures within the ARP/WARP suite [27]. The stereochemical quality of the refined models was assessed using the program Procheck[28]. The Ramachandran plot is in all cases of good quality.

Results and discussion

Similarly to other PPP family members, PP5 is potently inhibited by the tumor promoters okadaic acid, microcystin, cantharidin, calyculin A and tautomycin (Chen et al. 1994, Borthwick et al. 2001).

Here we solved the crystal structure of PP5 in complex with the inhibitors cantharidin, cantharidic acid and endothal. The three inhibitors share a common scaffold with three condensed ring. Cantharidin has an anhydride moiety while cantharidic acid is directly derived from cantharidin after addition of a water molecule, with disruption of the anhydride and formation of a dicarboxylic moiety. Endothal differs from the cantharid acid only for the presence of two methyls moieties. The three structure of PP5 in complex with the inhibitors are of very good quality and the RMSD with PP5-2Mn is 0.15, 0.37 and 0.19 Å for cantharidin, cantharidic acid and endothal, respectively. Data collection and refinement statistics are shown in **Tab 1**. The two metal ions are coordinated by one oxygen for each carbolxylate.

Table 1. Data collection and refinement statistics for PP5 complexed with the inhibitors.

	PP5-cntharidin	PP5-cantharidic acid	PP5-endothal
Spacegroup	C2	C2	C2
Cell dimensions (Å, °)	a= 153.97 b= 41.61 c= 105.90 β= 96.84	a= 158.87 b= 41.77 c= 104.99 β= 96.96	a= 154.54 b= 41.78 c= 105.48 β= 97.28
Resolution (Å)	40.0 – 2.6	39.4 – 1.7	38.3-2.1
Unique reflections	20601 (2872)	79891(9610)	39575 (5696)
Overall completeness (%)	98.7 (96.4)	96.5 (81.1)	99.9 (99.8)
R_{sym} (%)	19.2 (38.3)	9.1 (39.0)	11.4 (35.8)
Multiplicity	3.7 (1.9)	5.5 (3.5)	6.5 (4.9)
I/(σI)	5.8 (2.3)	14.7 (2.3)	16.3 (3.3)
Wilson plot B-factor ($Å^2$)	17.3	18.9	21.0
R_{cryst} / R_{free} (%)	18.4 / 20.8	19.4 / 23.4	16.2 / 22.7
Protein atoms per monomer	2528	2528	2528
Water molecules	846	547	343
Ions per monomer	2	2	2
RMSD bond lengths (Å)	0.009	0.020	0.026
RMSD bond angles (º)	1.58	1.78	1.96

The catalytic site of PP5 complexed with endothal and cantharidic acid is shown in Fig.1.
In the case of cantharidic acid the carboxylic moieties, which are very close to each other in the adduct, replace the water molecules which completed metal coordination in the dinuclear site. Surprisingly, endothal appears to be rotated of 180° with respect to the cantharidic acid, where the oxygen of the furanic ring is only 2.53 Å away from M1.
On the other hand, cantharidin sits in the catalytic in its open form (i.e. as a cantharidic acid molecule), where a double conformation is present.

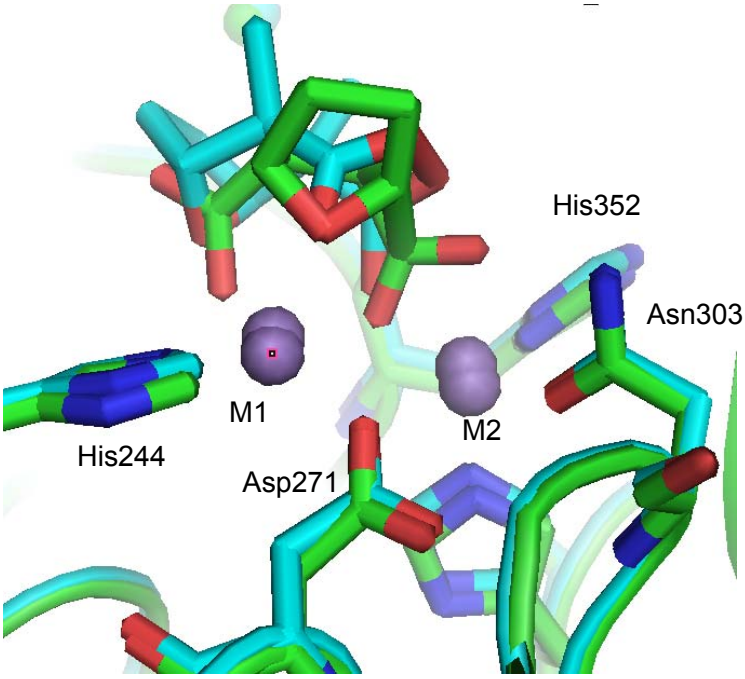

Fig.1 Crystal structure of catalytic domain of PP5 complexed with endothal (Green) and cantharidic acid (Cyan).

While one of the oxygen of each carboxylic acid moiety is still coordinating a metal ion, the other oxygen bascules in concert with the other carboxylic moiety in order to maximise the distance between the negatively charged groups (Fig. 2). Metal coordination is not affected by this conformational equilibrium.
Arg275 seems to exist in two different conformations which can help in stabilizing the inhibitor position in the active site with favorable H-bond interactions with one oxygen of one carbolxylate of the ligand.

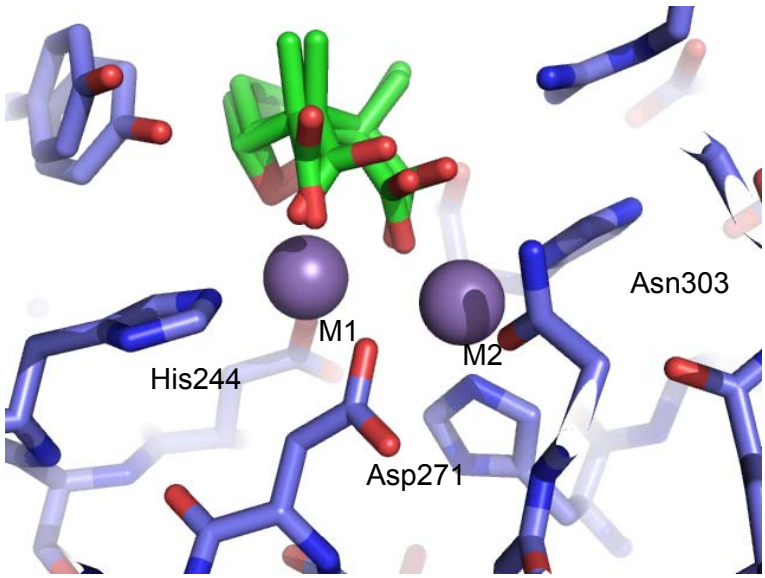

Fig.2. Crystal structure of catalytic domain of PP5 complexed with cantharidin.

The mobility of Arg275 seems to be a common feature of the catalytic site of many phosphatases.

Superimposition of crystal structure of related PPPs show that the aminoacid forming the active site are highly conserved. At the same time, the relative mobility of the side chain of the Arg275 is manteined. The variation of the position of these residues may have a critical role in both the interaction with the substrate and the catalytic mechanism.

The crystal structure of the adduct of cantharidin with PP5 reveals the capability of Arg275 to move in order to establish an interaction with an oxygen of the inhibitor. This plasticity of the active site makes the design of the inhibitor much more difficult even it can provide an additional site of interaction.

The availability of high resolution strcutures of related compounds complexed with PP5 made possible a detailed characterization of the catalytic site providing more detailed information on structure-activity relationship.

The distribution of the electronic density revels that the inhibitory capability canthaaridine is due to the hydrolysis of the anidridic ring and to the interaction of the free carboxylic groups with the metal ions present at the active site.

In particular the comparison of the binding pose of cantharidin with that of endothal reveals a role of the two methyls present only in the former molecule and apparently not involved in interactions with PP5.

Actually, the cantharidic acid is more stronlgy bounded to the catalytic pocket of PP5 than endothall thanks to a better coordination of the metal ions.

Conclusions

The complex of the cantharidin and of two structurally related molecules with a representative member of PPP family, have been solved for the first time. The analysis of the protein ligand complexes reveals the structural determinats of the interaction. In particular, has been clarified that the cantharidic acid and not cantharidine is the real inhibitor of PP5 since the electronic density is not consistent with the presence of the flat anydridic structure of the latter. An other relevant feature revealed by the experimental structures is the plasticity of catalytic site that appear to be a common characteristic of all the members of the family. In particular, the capability of Arg275 to move inside the catalytic site to establish new interactions can provide a further anchoring site to increase the affinity and the selectivity of new candidate drugs.

Reference List

1. Kerk D, Templeton G, Moorhead GBG. Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. Plant Physiology 2008; 146(2):351-367.

2. Rusnak F, Yu LA, Mertz P. Metalloenzymes and signal transduction: The protein serine/threonine phosphatases, a novel class of binuclear metal-containing enzymes. JBIC 1996; 1(5):388-396.

3. Yang J, Roe SM, Cliff MJ, Williams MA, Ladbury JE, Cohen PTW et al. Molecular basis for TPR domain-mediated regulation of protein phosphatase. Embo Journal 2005; 24(1):1-10.

4. Swingle MR, Honkanen RE, Ciszak EM. Structural basis for the catalytic activity of human serine/threonine protein phosphatase-5. Journal of Biological Chemistry 2004; 279(32):33992-33999.

5. Golden T, Swingle M, Honkanen RE. The role of serine/threonine protein phosphatase type 5 (PP5) in the regulation of stress-induced signaling networks and cancer. Cancer and Metastasis Reviews 2008; 27(2):169-178.

6. Ollendorf V, Donoghue DJ. The serine/threonine phosphatase PP5 interacts with CDC16 and CDC27, two tetratricopeptide repeat-containing subunits of the anaphase-promoting complex. Journal of Biological Chemistry 1997; 272(51):32011-32018.

7. Liu F, Iqbal K, Grundke-Iqbal I, Rossie S, Gong CX. Dephosphorylation of tau by protein phosphatase 5 - Impairment in Alzheimer's disease. Journal of Biological Chemistry 2005; 280(3):1790-1796.

8. Zeke T, Morrice N, Vazquez-Martin C, Cohen PTW. Human protein phosphatase 5 dissociates from heat-shock proteins and is proteolytically activated in response to arachidonic acid and the microtubule-depolymerizing drug nocodazole. Biochemical Journal 2005; 385:45-56.

9. Ali A, Zhang J, Bao SD, Liu I, Otterness D, Dean NM et al. Requirement of protein phosphatase 5 in DNA-damage-induced ATM activation. Genes & Development 2004; 18(3):249-254.

10. Rusnak F, Yu LA, Mertz P. Metalloenzymes and signal transduction: The protein serine/threonine phosphatases, a novel class of binuclear metal-containing enzymes. JBIC 1996; 1(5):388-396.

11. Bertini I, Fragai M, Luchinat C, Talluri E. Water-based ligand screening for paramagnetic metalloproteins. Angewandte Chemie-International Edition 2008; 47(24):4533-4537.

12. Fisher JF, Mobashery S. Recent advances in MMP inhibitor design. Cancer and Metastasis Reviews 2006; 25(1):115-136.

13. Honkanen RE, Golden T. Regulators of serine/threonine protein phosphatases at the dawn of a clinical era? Current Medicinal Chemistry 2002; 9(22):2055-2075.

14. Borthwick EB, Zeke T, Prescott AR, Cohen PTW. Nuclear localization of protein phosphatase 5 is dependent on the carboxy-terminal region. FEBS Letters 2001; 491(3):279-284.

15. Walter LD. Simon Bolivar and the cantharides. Revista Chilena de Infectologia 2007; 24(5):409-413.

16. McCluskey A, Keane MA, Mudgee LM, Sim ATR, Sakoff J, Quinn RJ. Anhydride modified cantharidin analogues. Is ring opening important in the inhibition of protein phosphatase 2A? Eur J Med Chem 2000; 35(10):957-964.

17. Aggen JB, Humphrey JM, Gauss CM, Huang HB, Nairn AC, Chamberlain AR. The design, synthesis, and biological evaluation of analogues of the serine-threonine protein phosphatase 1 and 2A selective inhibitor microcystin LA: Rational modifications imparting PP1 selectivity. Bioorganic & Medicinal Chemistry 1999; 7(3):543-564.

18. Hart ME, Chamberlin AR, Walkom C, Sakoff JA, McCluskey A. Modified norcantharidins: synthesis, protein phosphatases 1 and 2A inhibition, and anticancer activity. Bioorganic & Medicinal Chemistry Letters 2004; 14(8):1969-1973.

19. Leslie AGW. In Crystallographic Computing V. In: Moras D, Podjarny AD, Thierry J-C, editors. Molecular data processing. Oxford: Oxford University Press, 1991: 50-61.

20. "Data Reduction", Proceedings of CCP4 Study Weekend.: 1993.

21. Rossmann MG, Blow DM. The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst 1962; D15:24-31.

22. Crowther RA. The Molecular Replacement Method. In: Rossmann MG, editor. New York: Gordon & Breach, 1972.

23. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Crystallogr 1997; 30:1022-1025.

24. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr 2000; 56:1622-1624.

25. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Cryst 1997; D53:240-255.

26. McRee DE. "XtalView: A Visual Protein Crystallographic Software System for XII/XView". J Mol Graphics 1992; 10:44-47.

27. Lamzin VS, Wilson KS. Automated refinement of protein models. Acta Crystallogr D Biol Crystallogr 1993; 49(Pt 1):129-147.

28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993; 26:283-291.

3.3

Water-Based Ligand Screening for Paramagnetic Metalloproteins[1]

Ivano Bertini, Marco Fragai, Claudio Luchinat, and Eleonora Talluri

Angew. Chem. Int. Ed. 2008, 47, 4533 –4537

[1]Ivano Bertini, Claudio Luchinat and Marco Fragai designed the research, Eleonora Talluri done all the Molecular Biology work and sample preparation, Marco Fragai and Eleonora Talluri done relaxometric experimets.

((Catch Phrase))

DOI: 10.1002/anie.200((will be filled in by the editorial staff))

Water-based ligand screening for paramagnetic metalloproteins

Ivano Bertini, Marco Fragai, Claudio Luchinat and Eleonora Talluri*

((Dedication----optional))

NMR-based strategies for ligand screening and drug design are widely applied in pharmaceutical research.[1-11] Given the versatility of the NMR technique, many different methods have been developed to answer questions such as i) does the ligand bind, ii) where does it bind, and iii) what is the structure of the complex. Some or all of these questions are answered using two broad categories of experiments, *i.e.* the so-called ligand-based (Figure 1A) or protein-based (Figure 1B) experiments.[12-19]

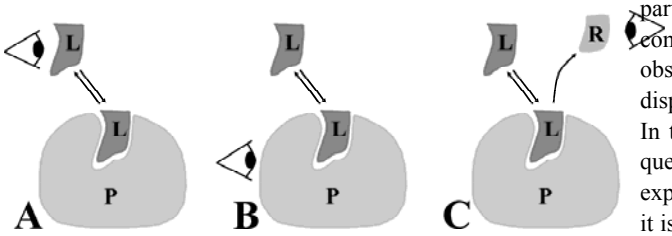

Figure 1. NMR-based strategies for ligand screening are either based on the observation of the bulk ligand, L, in exchange with the protein-bound ligand (A) or on the observation of the protein, P, in the presence of excess ligand (B). A subclass of ligand-based experiments is based on the observation of a reporter molecule, R, competing with the ligand for the same protein binding site (C).

Ligand-based experiments address question i), while protein-based experiments typically address questions ii) and iii). Ligand-based experiments require much smaller amounts of protein, as their sensitivity depends on the ligand concentration (which can be in the high millimolar range) rather than on the protein concentration (usually in the low micromolar range); therefore, they are faster and easily adaptable to high-throughput screening protocols. A particular subclass of ligand-based experiments, called competition experiments (Figure 1C), are based on the observation of one reporter ligand which may or may not be displaced from its binding site by the screened molecules.[20] In this case, if the binding site of the ligand is known, also question ii) can be answered. The sensitivity of competition experiments is similar to that of ligand-based experiments, as it is related to the concentration of the reporter ligand and not to that of the protein.

It occurred to us that any ligand-protein binding experiment often implies competition with a ubiquitarious probe-ligand, *i.e* water. Indeed, water molecules are almost always displaced upon binding of the protein ligand. Observation of the water signals instead of signals of a reporter ligand would provide a tremendous increase in sensitivity, due to the very high concentration of water protons (110 M versus submolar concentration of reporter ligand). Unfortunately, this very same feature prevents the direct use of water in competition experiments. In fact, the very large molar ratio between water and protein and the fact that only very few of the many protein-bound water molecules are displaced by the ligand make the differential effect negligibly small.[1]

However there is one case where the NMR parameters of a protein-bound water are so strongly altered that the effect is easily seen even when it is propagated from the bound to the bulk water molecules. This is the case of a water molecule coordinated to a paramagnetic metal ion, provided that it exchanges with bulk water molecules.[2] The presence of unpaired electrons, whose magnetic moment is 658 times larger than that of the protons, and the close proximity of the

[*] Prof. I. Bertini, Dr. M. Fragai, Prof. C. Luchinat, Dr. E. Talluri
Magnetic Resonance Center (CERM)
University of Florence
Via L. Sacconi 6, 50019 Sesto Fiorentino (Italy)
Fax: +390554574271
E-mail: bertini@cerm.unifi.it
Homepage: www.cerm.unifi.it

Prof. I. Bertini
Department of Chemistry
University of Florence
Via della Lastruccia 3, 50019 Sesto Fiorentino (Italy)

Dr. M. Fragai, Prof. C. Luchinat
Department of Agricultural Biotechnology
University of Florence
Via Maragliano 75-77, 50144 Florence (Italy)

[**] ((Acknowledgements)) This work was supported by the European projects NDDP n° LSHG-CT-2004-512077, Ortho and Para water n° 00503, and Nano4Drugs n° LSHB-CT-2005-019102; MIUR - RBLA032ZM7, and PRIN Cofin 2005 n° 2005039878; Ente Cassa di Risparmio di Firenze; GenExpress.

[1] Protein-bound water is indeed used in ligand screening, but only as a means to transfer magnetization to ligand molecules bound nearby in the so called WaterLogsy experiments.

[2] The maximal effect is observed when the residence time of the metal-bound water molecule is shorter or of the same order of the longitudinal relaxation time.

bound water protons to the paramagnetic center, increase the longitudinal (and transverse) relaxation rate of the latter by many orders of magnitude.[21] We propose here that water bound to a paramagnetic metal in the active site of a metalloenzyme which is a pharmaceutically relevant target can be an efficient reporter ligand to be used in competition experiments, answering questions i) and ii). We show that the efficiency of this method compares very favorably with the others commonly used. Given the fact that metalloenzymes are quite widespread, and that the metal can either be natively paramagnetic or be substituted by a paramagnetic one, this method can be of fairly general applicability. We have taken manganese(II)-activated protein phosphatase 5 (PP5) just to give an example. Manganese(II)-activated enzymes alone constitute around 15% of all metalloenzymes in the Protein Data Bank (PDB),[3] and several members of the two most studied protein targets in drug-discovery, kinases and phosphatases, are manganese-activated.[22-33] Manganese(II) is an ideal probe to demonstrate the feasibility of the approach, due to its strong paramagnetism and its favorable electronic relaxation properties.[21]

Protein Phosphatase 5 is constituted by two domains, a phosphatase domain and a regulatory domain.[34;35] The former is responsible for the enzymatic activity. The x-ray structures reveal that the phosphatase active site is formed by two metal ions placed at the bottom of a shallow pocket. The nature and the stoichiometry of the metal ions natively present in the catalytic domain of PP5 is still controversial, given that Fe^{2+}, Zn^{2+}, and Mn^{2+} compete for the two metal-binding sites and are all found present in variable amounts in enzyme preparations.[35] The active site structure of PP5 is shown in Figure 2. The two metal ions (reported as Mn^{2+} in the original publication) are bridged by a phosphate ion (P_i) and by a solvent-donated ligand (W_3, presumably a hydroxide). One of the two metals is coordinated by a water molecule (W_2), and other crystallographic molecules are present in the active site.[36]

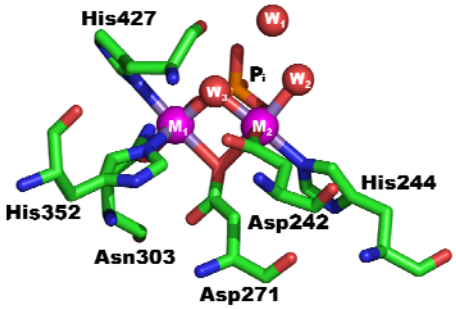

Figure 2. Active site structure of the catalytic domain of PP5. One metal ion is coordinated by Asp271, Asn303, His352, His427 and the other by Asp242, His244, Asp271, and a water molecule (W_2). A phosphate ion (P_i) and a further solvent-donated ligand (W_3), presumably a hydroxide ion, are bridging the two metal ions.

Four compounds known to be generic phosphatases inhibitors, with molecular weight in the 166 to 373 Da range, have been selected to test the method. These compounds can be considered as typical members of a compound library to be used in a high-throughput screening program to identify active site-directed PP5 ligands. The binding of the inhibitors at the catalytic site is believed to prevent the access of the substrate to the catalytic pocket and the subsequent interaction with the metal ions.[26] In doing so, the ligands may either displace the active site water or block its exchange with bulk water molecules: in either case, the propagation of relaxation enhancement to the bulk water molecules should be abolished. Therefore, PP5 may be a good test case for competition experiments with water as the reporter molecule.

The selected ligands were: i) cantharidin (inset of Figure 3A and Figure S1), a biologically active natural compound known to inhibit several phosphatases at nanomolar concentration;[37;38] ii) endothall, a nanomolar inhibitor of protein phosphatase 2A (PP2A);[39] iii) exo-3,6-epoxy-1,2,3,6-tetrahydrophthalic anhydride, a compound structurally related to norcantharidin, a micromolar inhibitor of PP2A;[40] iv) (-)-p-bromotetramisole, selected due to its ability to mimic the action of orthovanadate, a micromolar inhibitor of calcineurin (PP2B).[41] The structures of the last three compounds are reported in the Supporting Information (Figure S1).

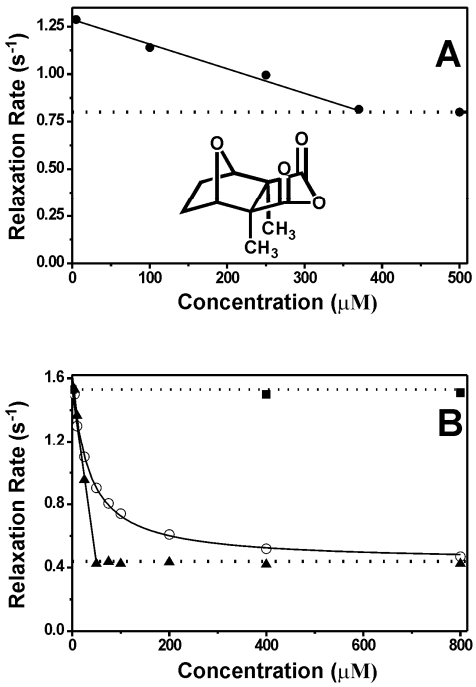

Figure 3. (**A**) Decrease of the water proton longitudinal relaxation rate at 400 MHz and 295 K, in a 370 μM solutions of Fe,Mn-PP5 upon addition of cantharidin (inset). (**B**) Titration curves of solutions of Fe,Mn-PP5 (50 μM) with endothall (open circles), with (-)-p-bromotetramisole oxalate (squares), and with exo-3,6-epoxy-1,2,3,6-tetrahydrophthalic anhydride (triangles) at 0.02 MHz and 295 K.

[3] Unpublished results from our laboratory.

Water proton longitudinal relaxation rates (R_1) were measured at 400 MHz, on 2 µL samples of Fe,Mn-PP5 370 µM, in the presence of increasing amounts of cantharidin. A linear decrease in R_1 could be observed until addition of one equivalent of cantharidin, after which the effect leveled off (Figure 3A). The residual relaxation rate is mostly due to the sum of the relaxation rate of pure water (ca. 0.35 s^{-1}) and the diamagnetic contribution from other water molecules bound to PP5 far from the active site, which can be measured on solutions of Fe,Zn-PP5 (not shown). Therefore, the effect of cantharidin is consistent with the binding of the ligand to the catalytic pocket preventing the access of the solvent to the paramagnetic center.

From Figure 3A it is apparent that the effect is clearly detectable, although not strong enough to allow to reduce the protein concentration much below the high micromolar range. On the other hand, given the strong intensity of the water signal, the sample volume can be much smaller than the 2 µL volume of the present experiment (see below), so that the quantity of enzyme needed is modest (much less than one nanomole).

The concentration of the enzyme could be further decreased if the paramagnetic effect were stronger. This is predicted to happen at low magnetic field, where the dipole-dipole coupling between the unpaired electrons and water protons is maximal. Indeed, it is well known that solutions of paramagnetic metalloproteins exhibit large proton longitudinal relaxation rates at values of magnetic field much lower than 400 MHz. To explore the field-dependence of the longitudinal relaxation rate of water protons in solutions of Fe,Mn-PP5, with and without cantharidin, relaxation rate profiles at magnetic fields ranging between 0.01 and 40 MHz were acquired on 50 µM Fe,Mn-PP5 using a Fast-Field Cycling relaxometer.[42] The relaxometric profile of Fe,Mn-PP5 (Figure 4) is indeed much higher all over the low field range with respect to 400 MHz and is typical of a Mn(II) profile.[21] The addition of an equimolar concentration of cantharidin to a solution of the paramagnetic protein decreases the relaxivity down to very low values, i.e. the values of Fe,Zn-PP5 (Figure 4). The decrease is much more striking than at 400 MHz, despite the enzyme concentration is about eight times less.

The acquired relaxometric profiles clearly indicate that a large gain in sensitivity can be achieved by performing the relaxation rate measurements at low magnetic field values, where the difference in relaxivity between the free and the inhibited form of the protein is more than one order of magnitude larger than at 400 MHz.

Such a strong effect also permits a safe estimate of the ligand dissociation constants K_D, provided they are of the order or higher than the concentration of the enzyme. Only an upper limit of about 1 µM can be estimated for cantharidin with this method, while K_D for cantharidin is in the nanomolar range.[38] For endothall, similarly to cantharidin, the relaxation rate decreases linearly with the concentration, while for exo-3,6-epoxy-1,2,3,6-tetrahydrophthalic anhydride the dependence on the concentration was non-linear so indicating a weaker binding constant (Figure 3B). The fit of the experimental relaxation rates, measured at the same magnetic field, as a function of the concentration of the latter inhibitor provided a K_D value of 19 ± 2.5 µM. This value compares well with the value of 13 ± 5 µM determined by an enzymatic assay (see Supporting Information). Finally, no effect was detected upon the addition of the (-)-p-bromotetramisole (Figure 3B), showing that this molecule does not bind PP5 at the metal binding site.

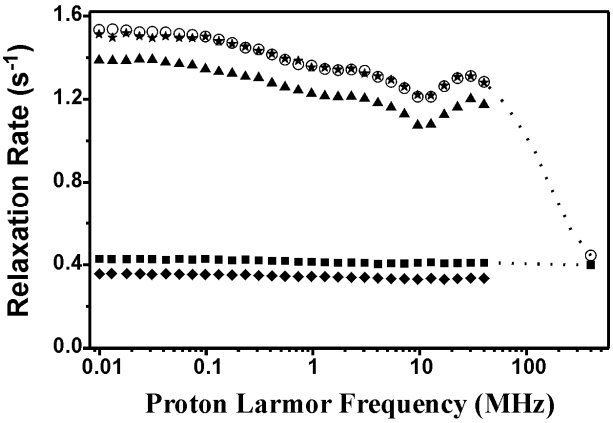

Figure 4. ^{1}H-NMRD profile of: a 50 µM Fe,Mn-PP5 solution at 288 K (stars), 295 K (open circles), and 303 K (triangles); a 50 µM Fe,Zn-PP5 solution at 295 K (diamonds); a 50 µM Fe,Mn-PP5 solution at 295 K in the presence of an equimolar concentration of cantharidin (squares). Dotted lines are only to better visualize the field dependence of the water relaxation rate.

Although in the present work the R_1 values have been acquired in a wide range of magnetic fields using a fast field-cycling relaxometer, the same information on the protein-ligand interaction could be achieved by performing the measurements at a specific value of the proton Larmor frequency. In particular the large difference in relaxivity present in the region 15-30 MHz in this case can be exploited by using commercial fixed-field relaxometers operating e.g. at 20 MHz. These instruments are relatively inexpensive and widely used in food analysis, biology and material science, and their sensitivity is not much smaller than a field-cycling relaxometer.

The amount of protein needed for the analysis is, together with speed, the main determinant of the success of a high-throughput screening methodology. The abundance of water protons makes the application of the present approach possible using very small sample volumes. From the present results, it can be concluded that at low field one can analyse samples with enzyme concentrations as low as 15 µM. With a conservative estimate of the minimal solution volume of 15 µL, the amount of protein needed would be as low as 60 picomoles, much lower than any other NMR-based method. At high field the minimal enzyme concentration would be much higher, i.e. around 400 µM, the paramagnetic effect being much less pronounced at high field. On the other hand, the sensitivity is so high that the sample volume could be in principle decreased by orders of magnitude, down to a few nanoliters.[4]

In a recent review,[14] Meyer et al published a comparative table (Table 2) of the merits and drawbacks of the most popular NMR-based methods in drug discovery.

[4] As an additional bonus, such a small volume ensures that the radiation damping phenomenon, that occurs for strong NMR signals and potentially makes relaxation measurements inaccurate, cannot be operative

TABLE 1. Comparison of the present water-based approach with the most common ligand screening strategies (from Meyer *et a.l*[14], Table 2). The most favorable features for each strategy are bold and underlined.

	SAR by NMR[1]	STD NMR[2]	Spin labelling[4]	Diffusion editing[43]	Inverse NOE pumping[44]	Water-Logsy[45]	Water-based
Large Protein (>30 kDa)	limited	**yes**	**yes**	no	**yes**	**yes**	**yes**
Small Protein (<10 kDa)	**yes**	no	**yes**	**yes**	no	no	**yes**
Isotope-labeled Protein required	yes	**no**	**no**	**no**	**no**	**no**	**no**
Binding epitope on protein	**yes**	no	no	no	no	no	**yes**
Binding epitope on ligand	no	**yes**	no	no	**yes**	**yes**	no
Amount of protein [nmol] at 500 MHz	25	**0.1**	~1	~100	~25	~25	Low field: 5 µL × 15 µM = **0.06** High field: 100 nL*× 400 µM = **0.04**
K_D tight binding	**no limit**	100 pM	100 pM	~100 pM	1 nM	100 pM	**no limit**
K_D weak binding	~1 mM	**~10 mM**	**~10 mM**	~1 mM	~1 mM	**~10 mM**	**no limit**
Identification of ligand	no	**yes**	**yes**	**yes**	**yes**	**yes**	no

*The amount of protein could be further reduced by appropriate nanodesign of a delivery system

Table 1 is based on Table 2 by Meyer et al., with the addition of a column for the present water-based method. In summary, the water-based strategy can be applied to both large and small proteins, does not require labelling, provides the binding epitope on the protein[5] and does not have lower or upper limits for K_D (ligand solubility being the upper limit), although it does not provide the binding epitope on the ligand and does not allow to identify the ligand in a mixture. Above all, water-based screening permits the use of unprecedentedly small amounts of protein per single assay.

The data here presented prove that the water-based approach can be successfully integrated in a drug discovery strategy aimed to design ligands for metalloenzymes where a paramagnetic metal ion is already present or can be inserted in place of the diamagnetic native metal. The analysis can be applied to large libraries of compounds or to a portfolio of ligands already identified by different high-throughput screening strategies, allowing us to select the molecules interacting at the metal binding site and providing, at the same time, information on the affinity constant and on the binding mode.

Experimental Section

The catalytic domain of PP5 corresponding to the construct 169-499 was expressed in soluble form in *E. Coli* and prepared as either Fe,Zn or Fe,Mn derivative (see Supporting Information). The Fe,Mn-enzyme was found to be fully active (3.9 ± 0.8 U/µg), while the Fe,Zn derivative was virtually inactive (0.1 ± 0.03 U/µg). This finding suggests a physiological role for the manganese ion, and further supports the use of the Fe,Mn derivative in PP5-directed ligand screening, selection and optimisation experiments.

[5] A reviewer pointed out that the ligand could bind away from the metal center and yet water accessibility be altered by allosteric effects. This is relatively uncommon, and in any case detecting such active-site-relevant allosteric effects could actually be considered an advantage.

Received: ((will be filled in by the editorial staff))
Published online on ((will be filled in by the editorial staff))

Keywords: Metalloenzyme · drug design · NMR spectroscopy

[1.] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996,** *274* 1531-1534.
[2.] M. Mayer, B. Meyer, *Angew.Chem.Int.Ed.* **1999,** *38* 1784-1788.
[3.] M. Coles, M. Heller, H. Kessler, *Drug Discov.Today* **2003,** *8* 803-810.
[4.] W. Jahnke, S. Rudisser, M. Zurini, *J.Am.Chem.Soc.* **2001,** *123* 3149-3150.
[5.] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. J. Stockman, *J.Biomol.NMR* **2001,** *21,* 349-359.
[6.] C. Dalvit, P. E. Fagerness, D. T. A. Hadden, R. W. Sarver, B. J. Stockman, *J.Am.Chem.Soc.* **2003,** *125,* 7696-7703.
[7.] B. Becattini, M. Pellecchia, *Chemistry* **2006,** *12,* 2658-2662.
[8.] K. E. Kover, P. Groves, J. Jimenez-Barbero, G. Batta, *J.Am.Chem.Soc.* **2007,** *129,* 11579-11582.
[9.] I. Bertini, M. Fragai, Y.-M. Lee, C. Luchinat, B. Terni, *Angew.Chem.Int.Ed.* **2004,** *43,* 2254-2256.
[10.] A. Ciulli, G. Williams, A. G. Smith, T. L. Blundell, C. Abell, *J.Med.Chem.* **2006,** *49,* 4992-5000.
[11.] S. Vanwetswinkel, R. J. Heetebrij, J. van Duynhoven, J. G. Hollander, D. V. Filippov, P. J. Hajduk, G. Siegal, *Chem.Biol.* **2005,** *12,* 207-216.
[12.] M. Pellecchia, B. Becattini, K. J. Crowell, R. Fattorusso, M. Forino, M. Fragai, D. Jung, T. Mustelin, Tautz L., *Exp.Opin.Ther.Targets* **2004,** *6,* 597-611.
[13.] M. Betz, K. Saxena, H. Schwalbe, *Curr.Opin.Cell Biol.* **2006,** *10,* 219-225.
[14.] B. Meyer, T. Peters, *Angew.Chem.Int.Ed.* **2003,** *42,* 864-890.
[15.] W. Jahnke, *J.Biomol.NMR* **2007,** *39* 87-90.

[16.] H. Schwalbe, H. U. Stilz, H. Kessler, *ChemBioChem* **2005,** *6,* 1475-1478.
[17.] C. A. Lepre, J. M. Moore, J. W. Peng, *Chem.Rev.* **2004,** *104* 3641-3675.
[18.] H. Kogelberg, D. Solis, J. Jimenez-Barbero, *Curr.Opin.Struct.Biol.* **2003,** *13,* 646-653.
[19.] X. Salvatella, E. Giralt, *Chem.Soc.Rev.* **2003,** *32*,365-372.
[20.] C. Dalvit, M. Flocco, S. Knapp, M. Mostardini, R. Perego, B. J. Stockman, M. Veronesi, M. Varasi, *J.Am.Chem.Soc.* **2002,** *124,* 7702-7709.
[21.] I. Bertini, C. Luchinat, G. Parigi, *Solution NMR of paramagnetic molecules*, Elsevier, Amsterdam **2001**.
[22.] T. Hunter, *Cell* **1995,** *80,* 225-236.
[23.] M. Gallego, D. M. Virshup, *Curr.Opin.Cell Biol.* **2005,** *17,* 197-202.
[24.] L. Tautz, M. Pellecchia, T. Mustelin, *Exp.Opin.Ther.Targets* **2006,** *10,* 157-177.
[25.] A. Noren-Muller, I. Reis-Correa, H. Prinz, C. Rosenbaum, K. Saxena, H. J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe, H. Waldmann, *Proc.Natl.Acad.Sci.USA* **2006,** *103*, 10606-10611.
[26.] A. C. Schmid, R. Woscholski, *Biochem.Soc.Trans.* **2004,** *32*, 348-349.
[27.] I. R. Correa, A. Noren-Muller, H. D. Ambrosi, S. Jakupovic, K. Saxena, H. Schwalbe, M. Kaiser, H. Waldmann, *Chemistry* **2007,** *2*, 1109-1126.
[28.] J. Klages, M. Coles, H. Kessler, *Mol.Biosys.* **2006,** *2*, 319-331.
[29.] T. Golden, I. V. Aragon, G. F. Zhou, S. R. Cooper, N. M. Dean, R. E.Honkanen, *Cancer Lett.* **2004,** *215*, 95-100.
[30.] W. D. Yong, S. D. Bao, H. Y. Chen, D. P. Li, E. R. Sanchez, W. Shou, *J.Biol.Chem.* **2007,** *282,* 14690-14694.
[31.] T. Zeke, N. Morrice, C. Vazquez-Martin, P. T. W. Cohen, *Biochem J.* **2005,** *385,* 45-56.
[32.] T. H. Davies, Y. M. Ning, E. R. Sanchez, *Biochemistry* **2005,** *44,* 2030-2038.
[33.] NDDP - NMR Tools for Drug Design Validated on Phosphatases (FP6-LIFESCIHEALTH 5012077) **2003.**
[34.] A.K Das, P.W. Cohen, D. Barford, *EMBO J.* **1998,** *17, 1192-1199*.
[35.] J. Yang, S. M. Roe, M. J. Cliff, M. A. Williams, J. E. Ladbury, P. T. W. Cohen, D. Barford, *EMBO J.* **2005,** *24,* 1-10.
[36.] M. R. Swingle, R. E. Honkanen, E. M. Ciszak, *J.Biol.Chem.* **2004,** *279,* 33992-33999.
[37.] R. E. Honkanen, *FEBS Lett.* **1993,** *330,* 283-286.
[38.] E. B. Borthwick, T. Zeke, A. R. Prescott, P. T. W. Cohen, *FEBS Lett.* **2001,** *491,* 279-284.
[39.] Y. M. Li, J. E. Casida, *Proc.Natl.Acad.Sci.USA* **1992,** *89,* 11867-11870.
[40.] A. McCluskey, C. Walkom, M. C. Bowyer, S. P. Ackland, E. Gardiner, J. A. Sakoff, *Bioorg.Med.Chem.Lett.* **2001,** *11,* 2941-2946.
[41.] M. Morioka, K. Fukunaga, T. Kawano, S. Haswgawa, K. Korematsu, Y. Kai, J. Hamada, E. Miyamoto, Y. Ushio, *Biochem.Biophys.Res.Commun.* **1998,** *253,* 342-345.
[42.] F. Noack, *Progr.NMR Spectrosc.* **1986,** *18,* 171-276.
[43.] M. F. Lin, M. J. Shapiro, J. R. Wareing, *J.Am.Chem.Soc.* **1997,** *119,* 5249-5250.
[44.] A. Chen, M. J. Shapiro, *J.Am.Chem.Soc.* **1998,** *120,* 10258-10259.
[45.] C. Dalvit, P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti, M. Sundstrom, *J.Biomol.NMR*, **2000,** *18,* 65-68.

Angewandte Chemie

Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Supporting Information

Water-based ligand screening for paramagnetic metalloproteins

Ivano Bertini[*,1,2], *Marco Fragai*[1,3], *Claudio Luchinat*[1,3] *and Eleonora Talluri*[1]

[1] Magnetic Resonance Center (CERM) – University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

[2] Department of Chemistry – University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

[3] Department of Agricultural Biotechnology, University of Florence, Via Maragliano, 75-77, 50144 Florence, Italy.

[*]**Prof. Ivano Bertini**
Magnetic Resonance Center (CERM)
University of Florence
Via L. Sacconi 6
50019 Sesto Fiorentino, Italy
e-mail: ivanobertini@cerm.unifi.it
Tel.: +390554574272
Fax: +390554574271

Plasmid DNA containing the gene of the human full length PP5 enzyme was kindly provided by Prof. D. Barford. The PP5 catalytic domain (PP5c, residues 169-499) was isolated fromthe template and then amplified by PCR. GATEWAY technology was used to clone the construct into the pENTR/TOPO/TEV vector and a LR reaction was performed to sub-clone the gene into the GATEWAY acceptor vector (pDEST30). The PP5c was expressed as His-GST-binding fusion protein in *E. Coli* BL21 (DE3) RIPL strain. PP5c was purified by Glutathione Sepharose affinity chromatography (Amersham) and then digested with TEV protease (Invitrogen). After the digestion

four vector-derived amino acid residues (Gly-Ser-Phe-Thr) remain at the N-terminus of PP5c. Further purification steps were carried out by using a combination of Ni^{2+}-saturated HiTrap Chelating column (Amersham) and Superdex 75 26/60 size exclusion chromatography (Amersham). The purified protein was checked by ICP mass spectrometry and found to contain one

equivalent of iron and variable amounts of zinc and manganese, accounting for another equivalent. Different aliquots were dialyzed against solutions containing either $MnCl_2$ or $ZnCl_2$. Apparently, the iron ion was not displaced by dialysis, while the other ion could be fully replaced by the selected ion, either zinc(II) or manganese(II). Protein samples were concentrated to 50 μM as determined by Bradford assay and activity tested against p-nitropenyl phosphate (*p*NPP, Jena Bioscience). For phosphatase activity assays *p*NPP was dissolved in the assay buffer solution (10mM pNPP in 20 mM Tris-HCl pH 7.4, 0.2 μg/mL BSA) and stored at 253 K until requested. The reaction was started by the addition of PP5 (50 nM) to 500 μL of the assay buffer containing 10mM *p*NPP. Production of *p*NP was followed by monitoring the increased absorbance at 405 nm using a Cary 50 Eclypse Spectrophotometer. Activity was expressed as Unit/μg enzyme were one unit is the amount of the phosphatase required to release 1 nmol of phosphate from *p*NPP in one minute at 295 K under the tested assay conditions. A *p*NPP molar extinction coefficient of 17.8 μL·nmol-1·cm-1 was considered. Assay was performed in triplicate and a standard error value was obtained. Endothall, exo-3,6-epoxy-1,2,3,6-tetrahydrophthalic anhydride and cantharidin were purchased from Sigma. Compounds were dissolved in DMSO at a stock concentration of 200 mM and added to solutions of PP5. The compounds were evaluated for their ability to inhibit the hydrolysis of *p*NPP. The enzyme was incubated at 298 K with increasing concentration of inhibitor and the absorbance at 405 nm monitored for 3 min after the

addition of the substrate. Fitting of rates as a function of inhibitor concentration provided Ki values.

The protein samples (50 μM and 370 μM) in a buffer containing Tris-HCl 50 mM, NaCl 150mM, DTT 1 mM, at pH 7.5 were analysed by 400 MHz NMR at 295 K and by relaxometry at 288, 295, and 303 K. The NMRD profiles were collected with a Stelar fast field cycling relaxometer in the 0.01–40 MHz proton Larmor frequency range. The instrument provides longitudinal relaxation rate measurements that are affected by an error of about ±1%. The analysis of the temperature dependence of the longitudinal water proton relaxation rate at low field reveals that, for this paramagnetic protein, τ_M and T_{1M} are of the same order. Actually the Fe,Mn-PP5 relaxivity was not affected by the temperature in the range 288-295 K while decreases from 295 to 303 K as shown in

Figure 4. T1 experiments at high field where recorded with a simple inversion recovery scheme at

295 K on a Bruker AV400 operating at 400.1 MHz and equipped with a BBO probe.

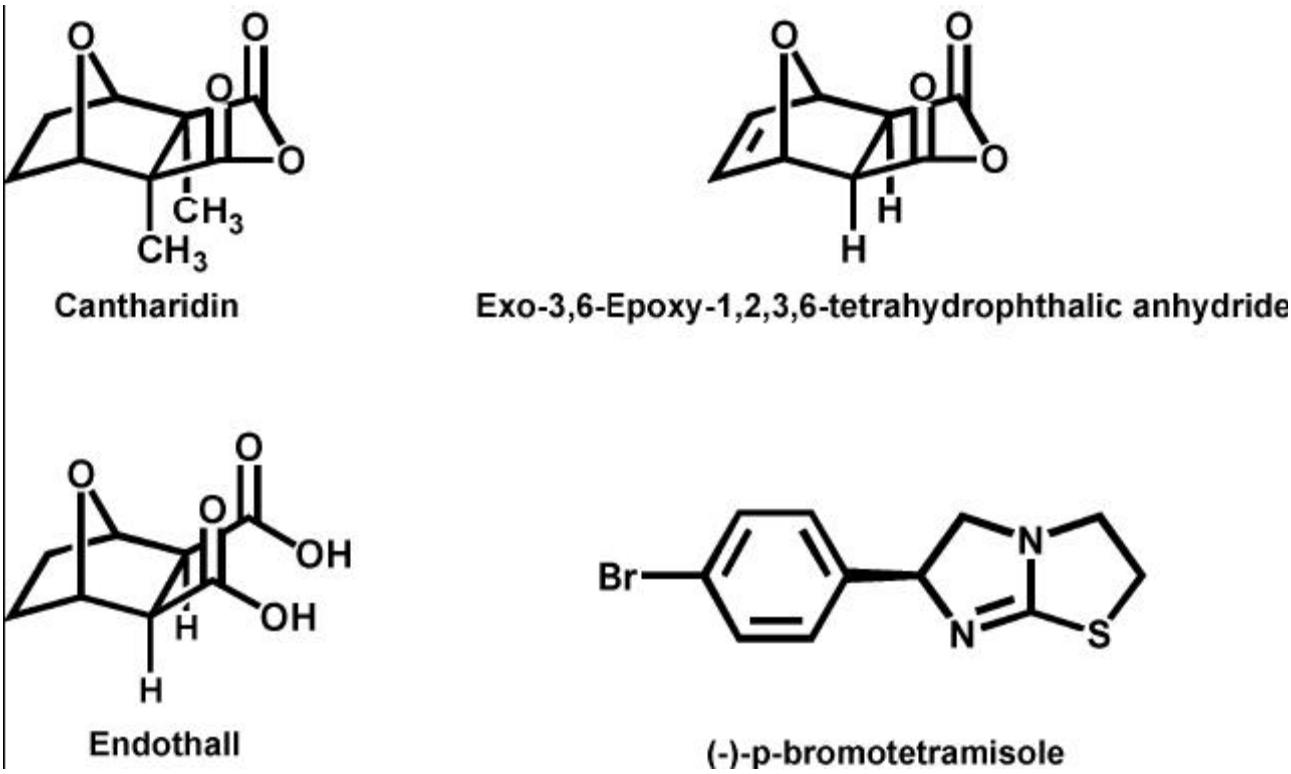

Figure S1. Chemical formula of the investigated ligands.

3.4

In vitro protein synthesis[1]

[1]Eleonora Talluri done all the Molecular Biology work.

Introduction

Several soluble proteins for NMR analysis have been expressed through the *in vitro* protein expression system, also called *cell-free* system, using S30 cellular extract from *E. coli (*Tab.1) as well as the same technique has been used for large scale production of integral membrane proteins[1,2].

Cell-free protein system technique has been taken into consideration to overcome the difficulties experienced in the *in vivo* expression of interesting metallophosphatases such as PP2A, PP6. The main advantage of this technique consists in the open and flexible nature of the *cell-free* systems. This peculiar feature allows the addition of components to facilitate protein folding and introduce modifications, producing proteins in controllable environments that may not be accessible *in vivo* systems. It is also capable of functionally producing both disulphide-bonded and membrane proteins, providing a platform for generation of 'difficult-to-express' proteins. The unique advantages of *cell-free* protein synthesis has been exploited. Expression of PP2A and PP6 have been tested. In addition, the expression of a set of targets has been tested to evaluate the capability of the *vitro* strategy to express proteins that differ in molecular weight, origin (human or bacterial protein), toxicity, solubility in water enviroment, presence of disulfide bounds. Features of the screening are reported in Tab.2. The expression and solubility of the most of the targets was already tested *in vivo.*

Tab.1 Soluble Proteins expressed in *cell-free* for NMR studies. S30 cellular extract from *E.coli* has been used.

Name	Origin	Function	NMR studies	Other studies	Ref.	Center/ University
PPiB	***E.coli***	**isomerase**	**X**	**Selective labeling, DOPA incorporation**	[3]	**ANU**
***h*CypA**	**Human**	**isomerase**	**X**	**Selective labeling, DOPA incorporation**	[3]	**ANU**
HMP	***E.coli***	**flavohaemoglobin**	**X**	**DOPA incorporation**	[3]	**ANU**
Dna-polymeraseIII(χ,γ ,ψ subunit)	***E.coli***	**Dna-polymerase**	**X**	**Protein-protein interactions**	[4]	**ANU**
Dna-polymeraseIII(τ subunit)	***E.coli***	**Dna-polymerase**	**X**	**Combinatorial isotope labeling**	[5]	**ANU**
Ras	**Human**	**oncogene**	**X**	**13C/15N labeling; D-glutamate replacement**	[6]	**RIKEN**
CAT	***E.coli***	**Cloramphenicol acetyltransferase**	**X**	**13C/15N labeling**	[6]	**RIKEN**
Transcription factors	***A. thaliana***	**Zinc-binding protein**	**X**	**Optimization of ligand's concentration**	[7]	**RIKEN**

Tab.2 Target selection for cell-free high trough-put screening. Proteins having different charatceristics have been chosen and class of protein has been create. For some of them, a soluble domain screening, a selective labelling and a coexpression with their partner have been planned. Green Fluorescent Protein and PpiB isomerase are not reported in the table even though they were used as positive controll in the expression and solubility screening.

Protein name	Origin	Class of protein					Selective Labelling planned	Coexpression planned	Screening Domain planned
		Membrane protein	Protein containing S-S	Degradation *In vivo*	Difficult to be expessed *in vivo*	Proteases			
Phosphatases (PP6 and PP2A)	**Human**				**X**				**X**
MURR1	**Human**				**X**				
Bcl2	**Human**				**X**				
CCS	**Human**		**X**		**X**				
SCOI	**Human**						**X**		
Cox11	**Human**				**X**				
Cox17	**Human**		**X**						
Proteases (MMP12 – MMP1)	**Human**					**X**			
Calmodulin targets	**Human**			**X**				**X**	
YVGZ	**Bacterial** (*E. coli B. subtilis*)								
EDF1-α isophorm	**Human**			**X**					
Dynactin	**Human**				**X**				
ATP-Binding domain	**Human**				**X**				
Menkes	**Human**	**X**							
Surf	**Human**	**X**			**X**				
CusS	**Bacterial** (*E.coli*)	**X**			**X**				**X**

Materials and Methods

The expression of PP2A was tested using the following constructs: 1-290 (PP2Ac), 23-287 ($PP2Ac_1$). The expression of PP6 was tested using the construct 1-298 (PP6c) and 1-305 ($PP6c_1$). Gateway tecnology was applied to clone each gene into different plasmid expression vectors. Test expression was performed in small scale. The genes cloned through Gateway system into pDEST17 were used in c*ell-free* expression.

The following genes were cloned using Gateway technology: ccs, cuss (full and soluble domain), murr1 .

The following genes were cloned using traditional cloning : Matrix metalloproteases (MMP1 and MMP12).

The following genes were already cloned in our laboratories: edf1-α isophorm (both native and TRXA-fused), cox11,cox17, surf, menkes, ATP binding domain, dynactin, scoI , calmodulin target (tuberin, hrpap20, max), yvgz from *B. Subtilis* and its *E. Coli* omologue and gfp.

PPiB isomerase from *E. Coli* and T7RNAP clones were a kindly concession of Prof. Otting. PPiB has been used as positive control to test the efficiency of cellular extract in protein expression.

All the DNA plasmid samples were amplified and than purified through MIDI-plasmid preparation (Invitrogen)

All the reagents were prepared and stocked according to Ozawa *et al.* [4]

The prepration of S30 *E.coli* extract from *E.coli* Bl21 (DE3) star and *E.coli* Bl21 Rosetta (DE3) pRARE used the Ozawa's protocol [5]

Results and discussion

Firs step was the production and optimization of the S30 extract concentration (Fig. 1), made from both *E. coli* BL21 (DE3) star and Rosetta pRARE strain. The reaction volume was of 200 μL. The protein overexpressed was the isomerase PpiB (MW 18KDa).

The S30 activity was optimized in terms of both S30 and magnesium acetate concentration.

S30 *E.coli* extract from *E.coli* Bl21 (DE3) star enriched with different chaperone sets was also tested to get expression of soluble proteins.

Cell-free reactions was carried out as described by Ozawa *et al.*[5] except for the dialysis system. Membranes from pierce ready to be used were tested (Fig.2) instead of the

traditional membrane. The dilaysis reaction were allowed to proceed in a 24-well microtiter plate and expression checked after 1 hour , 3hours, 6 hours and over night.

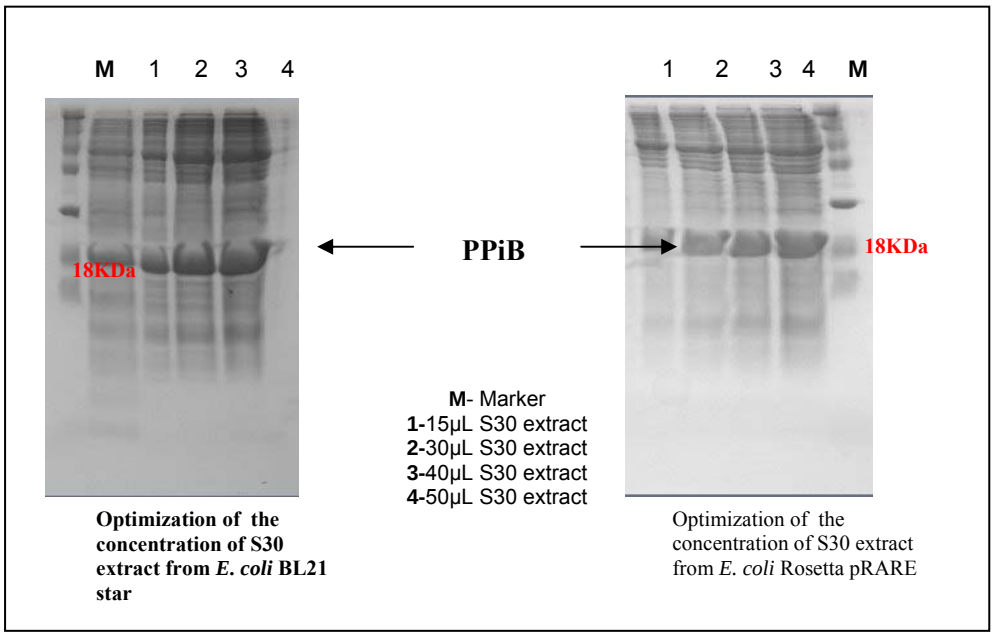

Fig.1 Titration of S30 extract from *E. coli* BL21 (DE3) star (A) and *E. coli* Rosetta pRARE (B) 10µL soluble protein+5µ FSB 3X have been mixed and 7µL of sample loaded on 17% acrilamide gel. PpiB isomerase has been used as positive control in test expression.

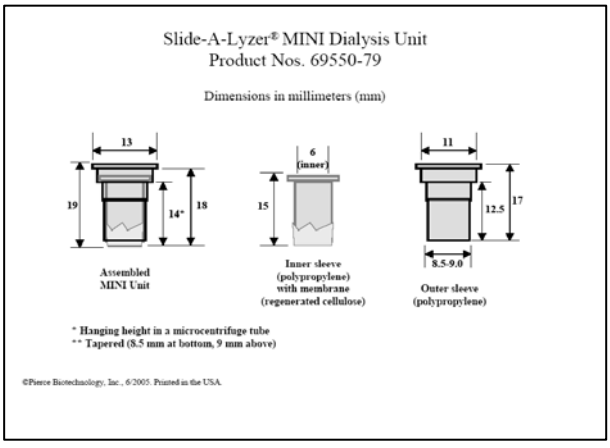

Fig. 2 Small scale dyalisis from PIERCE

Firstly, small scale 50µL test reactions were carried out on PpiB. The new system was as productive as the one got with the traditional dialysis system, allowing a scale-down of the reaction volume from 200µL to 50µL and reducing the hand-preparation time of the dialysis system.

The expression was verified by SDS-page and then confirmed by western –blotting for all the protein containing the His6-tag fused.

The expression and solubility of the most of the targets (not for all) was already tested *in vivo*. The not tested target were cloned into a pDEST 17 plasmid vector through Gateway technology.

Among these targets, the metalloproteases MMP1 and MMP12 showed the maximum expression yield *in vivo*, even if they were recovered in inclusion bodies. They were cloned into the MCSI vector but no expression occurred *in vitro.*

Most of the expressed targets were membrane proteins (SCOI, CusS full, SURF). Samples were expressed as insoluble according to the fact that the reaction was performed in an acqueous environment. The expression of CusS full, but not of the soluble domain, suggests that full-length constructs of memrane proteins that are composed of a soluble part anchorated to the membrain through two intramembrane fragments, are suitable for cell-free expression system.

The GFP, naturally produced by a jellyfish Aequorea, was easly overexpressed as it was for the bacterial YVGZ (both f from *B. Subtilis* and *E. Coli* omologue) and PpiB.

Among the soluble human targets, the expression was achieved for ATP binding domain, even if as insoluble, and EDF1-α isophorm with thioredoxin fusion-tag (Fig.3). The scale-up of the reaction for EDF1-α isophorm sample showed the same yield of the small scale test In. n Fig. 3 the scale- up result of EDF1-α isophorm expression is reported. After a column affinity purification and TRXA cutting, a final yield of less than 1mg/mL of reaction was recovered (Fig.4)

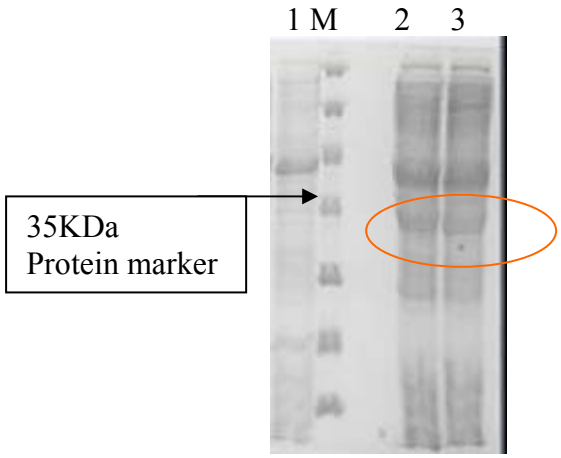

Fig.3 Cell-free expression of TRXA-EDF1-α in 1mL reaction. Before starting the reaction only a background deriving from S30 cellular extract is detectable by SDS- acrilamide gel (1); After 6 hours TRXA-EDF1-α is expressed. Total fraction (2) and soluble fraction (3) are indicated in the red circle.

M: Protein marker. 10μL protein+5μ FSB 3X have been mixed and 7μL of sample loaded on 17% acrilamide gel.

The calmodulin target HRPAP20 His6-fused was overexpressed as insoluble, while native protein was not espressed. The difficulty of expression of human-recombinant proteins is a real challenge for the *in-vitro* synthesis, even if one of the main advantages in this system is that the protein synthesis conditions can be adjusted and controlled, providing a defined environment for the expression and correct folding of individual proteins. For this purpose, the addition of detergents into the membrane protein's reaction may be useful and it still remain to be tested.

The *in vitro* attempt to express soluble form of PP2A and PP6 have failed and the two proteins can be recovered only as precipitated samples.

The addition of structural ions is not widely reported in the literature except for the zinc [7].

The cell-free reaction in presence of zinc helped the expression of CCS but not the solubility even tough this structural ion is naturally bounded to the protein[8].

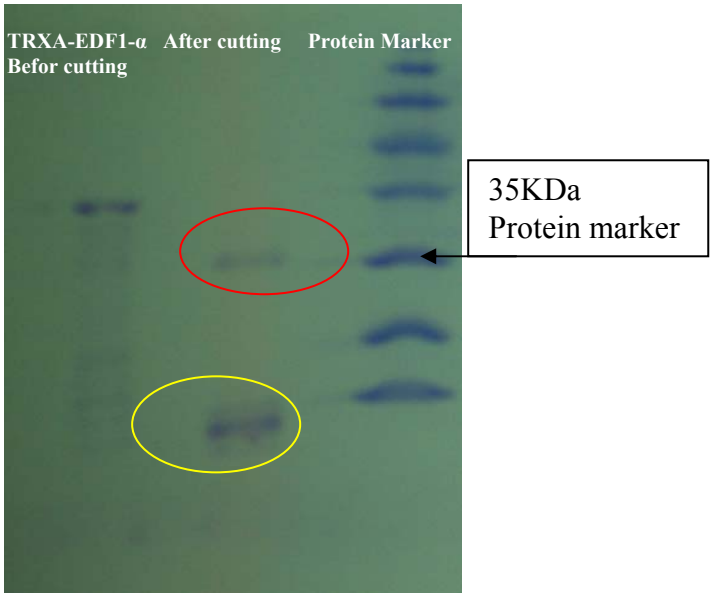

Fig.4 TRXA removal by TEV enzyme. TRXA in red circle and EDF1-α in yellow circle are shown**.**

MURR1, SD11 and CCS contain rare codons in their gene but *E.coli* Bl21 Rosetta (DE3) pRARE S30 extract was proved to be not more suitable than the *E.coli* Bl21 (DE3) star S30 extract for their expression and the same result was obtained. The expression for all the reported sample was verified both by SDS-page and western –blotting.

The expression and solubility of the targets were already tested *in vivo* except for CCS and CusS. All the proteins were overexpressed in incusion bodies except for GFP, PpiB and EDF1 α .

S30 *E.coli* extract from *E.coli* Bl21 (DE3) star enriched with different chaperone sets did not help the expression of the proteins in a soluble form.

A summary of the high through-put screening is reported in Tab.3.
It was hypothized that the small-scale dialysis system was allowing a faster cell-free reaction, thus resulting in a non-correct folding and insoluble sample. The traditional dialysis system was tested for all the expressed proteins with the same result. Different Temperature, DTT concentration, RPM were tested without success.

Further optimization of cell-free reaction conditions must be planned to help the yield of protein expression and production of labelled samples to be used both structural and biochemical studies.

Tab.3 Screening assay : main features and results

Protein name	description	Molecular weigth [Kda]	Cloning vector	Expression and solubilty *in vivo (E. coli)*	Expression and solubility *in vitro*
CCS	copper chaperone for SOD1	30	pDEST17 (His-tagged)	not tested	**insoluble**
MURR1	factor recently implicated in copper homeostasis	21	pDEST17 (His-tagged)	no expression	**insoluble**
MMP1	Metalloprotease	42	MCSI and pET21	Insoluble	No expression
MMP12	Metalloprotease	42	MCSI and pET21	Insoluble	No expression
Cox11	cytochrome c oxidase assembly protein	22	pDEST17(His-tagged)	Insoluble	No expression
Cox17	cytochrome c oxidase assembly protein	11	pDEST17(His-tagged)	Soluble	No expression
CusS	Membrane protein	53.7	pDEST17(His-tagged)	not tested	**insoluble**
SURF	membrane protein	28	pDEST17(His-tagged)	Insoluble	**Insoluble**
MENKES	Membrane protein	18	pDEST17(His-tagged)	Insoluble	N expression
ATP binding domain	ATP binding domain	40	pDEST17(His-tagged)	Insoluble	**Insoluble**
DYNACTIN	Dynain activator	53	pDEST17(His-tagged)	Soluble only if MBP-tagged	No expression
SCOI	protein essential for cytochrome c oxidase function is a Cu(I) - binding protein	30	pDEST17(His-tagged)	Soluble	**Insoluble**
PP6/PP2A	Serine/threonine phosphatase [PP6 residues 1-298; 1-305,PP2A residues: 1-290; 23-287]	34.3-35	pDEST17(His-tagged)	Insoluble	**Insoluble**
TUBERIN	Calmodulin target	26-30	pDEST42(native)-pDEST17(His-tagged)	Insoluble(native) Soluble (His-tagged)	No expression
HRPAP20	Calmodulin target	20	pDEST17(Hist tagged)	Insoluble	**Insoluble**
MAX	Calmodulin target	9.8	pDEST42 (native)	soluble	No expression
EDF1-α isophorm (native and TRXA-fused)	Calmodulin target	16.3 and 32.3	pET21 andpET20A	Insoluble	No expression (native) **Soluble TRXA-fused**
YVGZ from *B. Subtilis* and its *E. Coli* omologue	copper-sensing repressor that regulates the copZA operon	15	pDEST17 (His-tagged)	Soluble only with MBP-tagged	**Insoluble**
GFP	Green-fluorescent protein	27		Soluble	**Soluble**

Reference List

1. Klammt C, Schwarz D, Lohr F, Schneider B, Dotsch V, Bernhard F. Cell-free expression as an emerging technique for the large scale production of integral membrane protein. Febs Journal 2006; 273(18):4141-4153.

2. Savage D.F.Anderson C.L.Colmenares Y.R. NZEaSRM. Cell-free complements in vivo expression of the E. coli membrane proteome. Protein Sci 16, 966-976. 2007.

3. Ozawa K, Headlam MJ, Mouradov D, Watt SJ, Beck JL, Rodgers KJ et al. Translational incorporation of L-3,4-dihydroxyphenylalanine into proteins. Febs Journal 2005; 272(12):3162-3171.

4. Ozawa K.. Cell-free *in vitro* protein synthesis in an autoinduction system for NMR studies of protein-protein interactions. J Biomol NMR 32, 235-241. 2005.

5. Ozawa K, Wu PSC, Dixon NE, Otting G. N-15-Labelled proteins by cell-free protein synthesis - Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. Febs Journal 2006; 273(18):4154-4159.

6. Matsuda T, Koshiba S, Tochio N, Seki E, Iwasaki N, Yabuki T et al. Improving cell-free protein synthesis for stable-isotope labeling. Journal of Biomolecular NMR 2007; 37(3):225-229.

7. Matsuda T. Cell-free synthesis of zinc-binding proteins. Struc Funct Genomics 7, 93-100. 2006.

8. Lamb AL, Wernimont AK, Pufahl RA, Culotta VC, O'Halloran TV, Rosenzweig AC. Crystal structure of the copper chaperone for superoxide dismutase. Nature Struct Biol 1999; 6:724-729.

4

GENERAL DISCUSSION AND PERSPECTIVES

Structural biology provides detailed informations on the structure of biological macromolecules, with the ultimate goal of investigating the relationship between structure and function. The high resolution three-dimensional structure of a protein can be solved in the solid state, through X-ray crystallography, or in solution, through NMR spectroscopy.
With the development of the genome sequencing project, relating functional information to each of the proteins encoded in various genomes has become a crucial problem. The possibility of inferring from experimental three-dimensional structural data functional information more reliable than those derived from sequence comparison methods, has prompted worldwide efforts aimed at solving protein structures on a genome-wide scale.

Our research was aimed not only to elucidate the structure of metal-binding proteins and to understand the interactions of the selected enzyme with its targets, but also to develop new tools able to support the structure-based drug design for potential pharmacological targets.

The main steps of a structure-based drug discovery project are :

1. target selection
2. Clonign, expression and purification of the target protein
3. Structural and functional characterization of the target
4. Development of screening tools
5. Screening of library of chemicals and identification of lead compounds
6. Lead optimaization

Human Serine/Threonine Phosphatase PP5 was selected as potential pharmaceutical target beacuse is involved in may cellular precesses and the alteration of the catalytic activity may lead to many human disease. Protein sample has been obtained through recombinant expression in *E. coli* system. Investigations on the catalytic domain have been performed to refine the biochemical and structural characterization by the use of ICP-mass, enzymatic assay and X-ray spectrospcopy.
A part of the purified protein was dialyzed against either solutions of Mn(II) or Zn(II) to obtain different metallated protein samples. The manganese-dialyzed-enzyme was found to be fully active, while the zinc derivative was virtually inactive as it was for the not dyalized sample. The crystal structure of both metal-loaded protein (Fig.1) and not dyalized PP5 have been solved. The first two structures have been showed to be superimposable. A binuclear metal-center is present in the catalytic domain of these proteins.

PP5 crystals obtained from the protein dialyzed against Mn^{2+} revealed the presence of two Mn^{2+} ions in the metal site. On the other hand, X-ray fluorescence and anomalous dispersion from PP5 crystals obtained from the protein dialyzed against Zn^{2+} revealed the presence of two Zn^{2+} ions in the metal site.

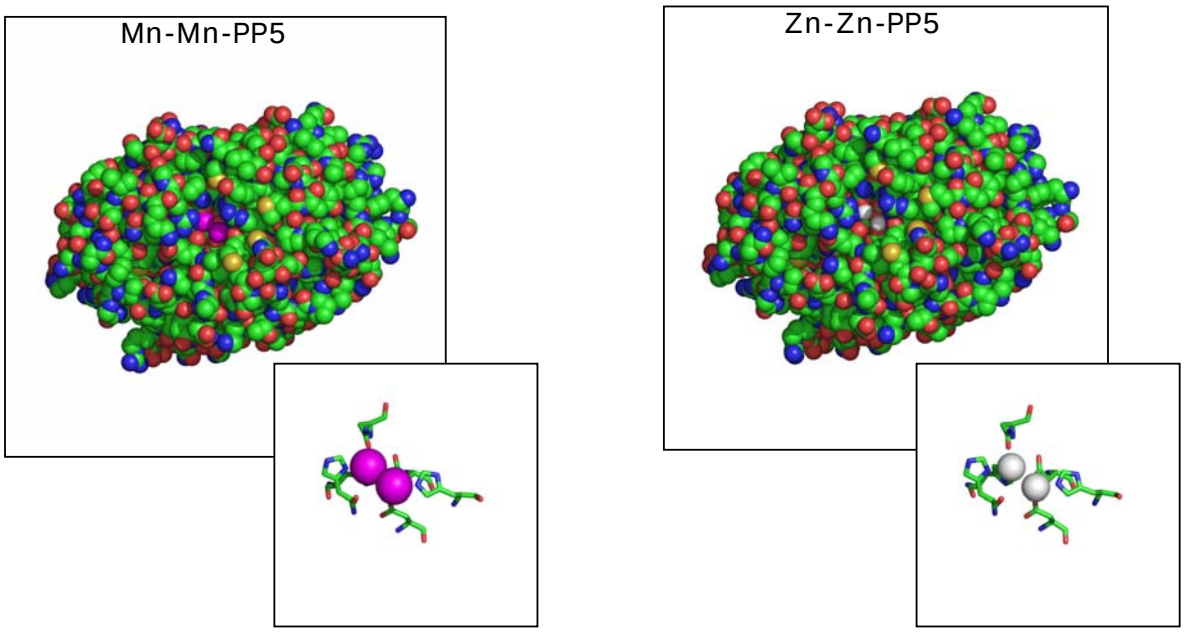

Fig.1 Crystal structure of Mn-Mn and Zn-Zn-PP5. Residues involved in metal cooridnation (Asp242,His244,Asp271;Asp27,Asn303,HEis352,His427) are zoomed in.

In the light of these results, the different roles for the two metal ions in the regulation of PP5 function have to be further investigated.

The paramagnetic properties of the fully active Mn-PP5 (catalytic domain) has been taken into account for the development of small ligand screening tools.

Taking advantage from the presence of a paramagnetic metal in the active site of the pharmaceutically relevant human protein serine threonine phosphatase PP5, water bound to the metal can be used as an efficient reporter ligand to perform competition experiments with a tremendous increase insensitivity due to the very high concentration of water protons.

Given the fact that metalloenzymes are quite wide spread, and that the metal can either be natively paramagnetic or be substituted by a paramagnetic one, this method should be of fairly general applicability. The data presented here show that this water-based approach can be successfully integrated into a drug discovery strategy aimed at designing ligands for

metalloenzymes where a paramagnetic metal ion is already present or can be inserted in place of the diamagnetic native metal.
A point we are going to address is the possible use of the relaxometric approach to investigate the binding of activators of PP5. In fact, PP5 differs from most PPP family phosphatases in that the principle substrate targeting, regulatory, and catalytic domains are contained in a single polypeptide chain. The displacement of this domain by the binding of the TPR domain with activators leads to the activation of PP5 making the paramagnetic site accessible to water molecules. Therefore, the activation of PP5 by small ligands can be easly monitored observing the increase of proton water relaxation rate.

Results obtained from relaxometric analysis increase the interest in chantaridin and endothal as PP5 inhibitors. Due to their low molecular weight, cantharidin and its analogues can be considered as leads to investigate the structural features of the binding site of PP5 useful to design new candidate drugs.
The crystal structure of the catalytic domain of PP5 in complex with the inhibitors cantharidin, cantharidic acid and endothal has been obtained by socking the lignads on crystals of free protein and all the structures solved by X-ray.
Endothal binds the catalityc pocket rotating in 180° with respect to the cantharidic acid. Moreover, cantharidin sits in the catalytic site as a mixture of open and closed conformations, where the closed one seems to be the less populated. Moreover, the Arg275 mobility gives plasticity to the catalytic site stabilizing the inhibitor position in the active site.
The comparison of the binding pose of cantharidin with that of endothal reveals a role of the two methyls present only in the former molecule and apparently not involved in interactions with PP5.
Actually, the cantharidic acid is more stronlgy bounded to the catalytic pocket of PP5 than endothall thanks to a better coordination of the metal ions. The distribution of the electronic density revels that the inhibitory capability canthaaridine is due to the hydrolysis of the anidridic ring and to the interaction of the free carboxylic groups with the metal ions present at the active site. The last aspect must be further elucidated.

The expression of PP2A and PP6 has been attempted by *cell-free* approach. Unfortunately, both the proteins have been recovered as insoluble samples. *Cell-free* systems will be further optimised in order to improve the expression yield and protein folding. Further development of the system would significantly reduce the cost and enhance the capability for high

throughput generation of proteins and screening of their function and interactions. It will be an important tool for the future analysis of proteomes.