U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Training Period Report

Integrated Masters in Medicine

HUMAN MOLECULAR EVOLUTION

THE SERINE/THREONINE-SPECIFIC PHOSPHATASE PROTEINS FAMILY

Rui Tiago Aguiar Cardoso

Supervisor

Agostinho Antunes Pereira, Ph.D (Biology)

Co-supervisor

Susana Rodrigues Pereira, Ph.D (Chemistry)

Porto, 1 March 2010

Summary

This training period occurred at **Laboratório de Ecotoxicologia, Genómica e Evolução, Centro Intersdisciplinar de Investigação Marinha e Ambiental (CIIMAR), Universidade do Porto,** from November 2008 to November 2009, with a six hours a week schedule (288 hours).

Objectives:

1. To introduce the student in research activity
2. To study the molecular aspects of gene expression
3. To study the role of protein phosphatases in vertebrates
4. To understand the biomolecular variations, in the context of species evolution, as a response to environmental changes
5. To make an approach to bioinformatics and molecular analysis of developmental genes in metazoan organisms
6. To introduce the student to the laboratorial work, namely DNA extraction, PCR and DNA sequencing
7. To interpret data using phylogenetic and molecular evolutionary tools
8. To collaborate in manuscript writing

Developed activities:

1. Bibliographic research about molecular genetics, protein phosphatases and gene evolution
2. Laboratorial work: DNA extraction, PCR and DNA sequencing
3. Bioinformatic analysis of gene variations among different protein phosphatases in metazoans
4. Analysis of the three-dimensional structure of protein phosphatases
5. Collaboration in data analysis using phylogenetic and molecular evolutionary tools
6. Collaboration in manuscript writing

Results:

The Phosphoprotein Phosphatase (PPP) enzymes are regarded as very ancient, with most members distributed in all studied metazoans. The catalytic subunits of PPP enzymes, especially in those enzymes that need regulatory subunits (Ppp1, 2, 3, 4 and 6) are highly conserved, being the monomeric PPP enzymes (Ppp5 and 7) the more genetically divergent. This fact can be explained by the structural and functional plasticity conferred by the regulatory subunits.

PP enzymes evolve in two different ways: some (like PTP) acquire additional domains by a gene level fusion and thus consist of multidomain monomeric enzymes with a strict substrate specificity and a tight regulation; other (like most PPP) consist of highly conserved small catalytic subunits that diversified their biological function by binding novel regulatory subunits encoded by separate genes.

List of Abbreviations

3D: Three-dimensional

DNA: Deoxyribonucleic Acid

PCR: Polymerase Chain Reaction

HIV: Human Immunodeficiency Virus

ATP: Adenosine Triphosphate

RNA: Ribonucleic Acid

PPP: Phosphoprotein Phosphatase

PTP: Protein Tyrosine Phosphatase

PPM: Metallo-Dependent Protein Phosphatase

FCP: Fyn Carboxyl-Terminal Peptide

EST: Expressed Sequence Tags

MEGA 4: Molecular Evolutionary Genetics Analysis (Version 4)

NCBI: National Center for Biotechnology Information

BLAST: Basic Local Alignment Search Tool

TPR: Tetratricopeptide Repeat

MHC: Major Histocompatibility Complex

NMDA: N- Metil-D-Aspartate

Acknowledgements

I want to express my thanks to Dr. Agostinho Antunes and Dr. Susana Pereira supervisors of this training period.

I´m also grateful to all my colleagues of the Laboratório de Ecotoxicologia Genómica e Evolução.

A thanks to CIIMAR (Centro Interdisciplinar de Investigação Marinha e Ambiental) and to FCT (Fundação para a Ciência e Tecnologia), by the initiative of this grant, and for making this training period possible.

A special thanks to my parents and my sister. Thank you all for always believe in me.

Thanks to Cecilia Reis.

Thanks to Cátia Costa.

Thanks to my teachers at Liceu Carolina Michaelis, Dr. Helena Leite, Dr. Maria do Céu Vieira and Dr. Adriana Moreira, for having stimulated me in sciences study!

Table of Contents

	Summary	i
	Abbreviations	iii
	Acknowledgements	iv
	Table of Contents	v
1.	Introduction	1
1.1	The Evolution of Biological Systems	3
1.2	Proteins Evolution and Disease	5
1.3	The Biological Importance of Phosphorylation	6
1.4	General Aspects of Protein Phosphatases Classification	7
1.5	The Serine/Threonine-Specific Protein Phosphatases	7
1.6	The Three-Dimensional Structure of PPP Catalytic Subunits	14
2.	Discussion	16
3.	Conclusion	29

Introduction

The present report concerns a training period that took place at **Laboratório de Ecotoxicologia, Genómica e Evolução** of **Universidade do Porto,** from November 2008 to November 2009, with a six hours per week schedule (in a total of 288 hours).

The main purpose of this training period was to study the genomics, the molecular structure and the physiology of serine/threonine protein phosphatases, from a molecular evolution perspective.

Molecular evolution is the process by which the molecules of living forms diverge from one another over time. Its study embraces the evolution of molecules themselves, their origins, the effects of the changes on function, and the integration of the molecules on the biochemical pathways of an organism (Cooper DN *et al*, 2008).

Molecular evolution – the use of molecular data to study every aspect of evolution – is now standard. The discoveries of molecular biology can only be understood in an evolutionary context. Its purposes are to describe the structure of a protein, and explain its function in terms of physicochemical principles, but also to understand the mechanisms for the origin of the protein. The major contributions include (Cooper DN *et al*, 2008):

Measuring the nature and amount of genetic variation in and between populations;

Studying quantitatively the relationships between all forms of life;

Giving a more-centered view of evolution;

Modeling evolution of macromolecules;

Discovering mechanisms for the increase of genetic information with time;

Identifying important functional (conserved) parts of sequences;

Allowing realistic and testable theories for the origin of life;

Understanding human origins and evolution.

The age of comparative genomics will have a major impact on evolutionary biology and biological research in general. Already this field has played a role for the resolution of key issues in the genetic relationships between organisms and the elucidation of the emergence of basic cellular processes (Merchant SS *et al*, 2007) and the evolution of species-specific traits (Hanikenne M *et al*, 2008), thus constituting the basis to new medical and pharmaceutical research.

The comparative genomics is an emergent area, very promising in clarifying the evolution of biological systems and its implications for the interaction of different species with the environmental factors that can lead to their survival or death.

Proteins are important markers of species evolution and are used in molecular taxonomy to reconstruct the phylogenetic trees; they are “molecular clocks”. Proteins may evolve at different rates; for some proteins such as cytochrome c, evolution rates seem to be very constant along time; other proteins, like those related to antibiotic activity in bacteria, have higher rates of evolution.

Protein phosphatases are a large group of enzymes that catalyze the dephosphorylation of phosphoamino acids, mainly phosphotyrosine, phosphoserine and phosphothreonine, previously phosphorylated by the corresponding protein kinases.

The reversible phosphorylation of proteins determines their activity and, being the main intracellular language for signal transduction, it assumes a prominent role in the regulation of multiple physiological processes. In fact, genetic and biochemical studies showed that these proteins up-regulate and down-regulate signaling pathways and physiological processes in many tissues (Zhang *et al*, 2002. Zhang *et al*, 2001). Investigation of these phosphatases has revealed their role in processes such as glycogen metabolism, cell cycle, RNA splicing, T-cell activation, memory formation, and regulation of numerous protein kinases (Andreeva AV *et al*, 2001).

Protein phosphatases are also important in immune system function, cell migration and adhesion (Janssens V *et al*, 2009), as well as in immune protection against lethal pathogens (Panchal RG *et al*, 2009).

Given the importance of protein phosphorylation for the regulation of such a wide number of physiological processes, it is only natural to find that some of the more important human pathologies can be related to an abnormal degree of phosphorylation in some key-proteins. As examples we can cite: Alzheimer’s Disease (Liu R *et al*, 2009), heart failure (Champion HC, 2005), diabetes (Parameswara VK *et al*, 2005), HIV infection (Ammosova T *et al*, 2005; Faulkner NE *et al*, 2002), carcinoma (Reeder JE *et al*, 2003), and endometrial cancer (Sugiyama M *et al*, 2003).

For these reasons protein phosphatases are important targets in new pharmaceutical research. Cyclosporin A and Tacrolimus (FK-506), two macrocyclic natural compounds and excellent inhibitors of the protein phosphatase calcineurin, are already used as immunosuppressors to prevent the rejection of transplants. Clinical research has also showed the usefulness of these

compounds in auto immune diseases such as rheumatoid arthritis, asthma, psoriasis, and Crohn's disease (Silva EFC, 1998).

These were some of the main reasons that lead to the choice of the investigation topic developed in this report.

1.1 The Evolution of Biological Systems

Biological systems are dynamic entities that evolve naturally in normal circumstances (especially by mutation and genetic recombination), with the natural selection being the cornerstone of this complex process.

Mutations generate new alleles that fall in three broad categories: neutral, selectively advantageous (when it increases the chance of survival and reproduction of its bearer – increases its fitness) and selectively disadvantageous (when it decreases the survival and reproduction of its bearer).

Spontaneous mutational events, although very low in frequency, become important when considering a geological scale of time. The process of evolution changes the distribution of genotypes and phenotypes in successive generations; spontaneous mutations (estimated as 10^{-12}-10^{-10} per base pair per generation), recombination, gene duplication and gene flow are some of the most important aspects implicated in gene evolution (Lesk AM, 2005).

Since there are many sites at which mutation can occur in the DNA of each organism, the probability of two species fixing the same mutation is low; the fixation of different mutations leads to the molecular divergence of species – their molecular evolution. The essence of evolution at the molecular level is, therefore, the differential fixation of mutations in distinct groups of organisms (Cooper DN *et al*, 2008).

Random genetic recombination during fecundation, genetic drift, migrations and absence of panmixy are also important factors that influence the evolution of biological systems.

Natural selection affects the frequency of beneficial and deleterious mutations and also the frequency of nearby neutral mutations, especially in regions of low recombination. Thus, natural selection links the fates of beneficial, deleterious and neutral mutations (Cooper DN *et al*, 2008).

Selection can act to maintain alleles within a population through many mechanisms, which, taken together, constitute balancing selection. Balancing selection has been a force in shaping human evolution since the beginning. While its effects are isolated and transitory, they can be

seen especially in those regions of the genome where selective forces are traditionally stronger, namely the genes related to the immune system and, in particular, the MHC – Major Histocompatibility Complex. While trans-species polymorphisms, which are nearly always a certain indicator of balancing selection, are rare outside the MHC, the patterns of variation that it creates within a species can be found strewn across the human genome.

Other forms of selective forces are positive and negative selection. While the implications of the former in human health are unclear, negative selection is ubiquitous and its effects are certainly a major cause of the genetic diseases observed today (Cooper DN *et al*, 2008).

Natural selection rejects with variable strength, mutations that reduce the individual's capability to survive and reproduce. Thus, evolutionary theory predicts that mutations producing disease will be under strong selective constraints. Selective strengths at the codon level will determine if a mutation frequency will increase, decrease or change randomly during evolution. This strength finally serves in the prediction of nonsynonymous single nucleotide polymorphisms (nsSNP) that produce disease in humans. By using comparative genomics data and maximum likelihood phylogenetic approaches it was demonstrated that mutations on residues showing low rates of evolution are significantly associated to disease but not to human genetic polymorphisms (Cooper DN *et al*, 2008).

The greater the selective pressure, the greater the adaptive mechanisms that living beings have to make available for their survival. In limit conditions, genetic variations can make the difference in individual and species survival.

Nowadays, humans and the remaining living beings are exposed to a high diversity of new toxicants: as examples we can cite the food additives (preservatives and coloring), the pesticides used in vegetable and fruit cultures, the antibiotics and steroid hormones used in animal production, the pharmaceutical drugs (whose total effects in cellular physiology can never be completely understood), and the environmental pollutants (soil, water and atmosphere). All species are directly or indirectly affected by new toxicants due to the cyclic nature of ecosystems' dynamics. Organisms need to make an effort to deal with this huge diversity of new compounds, reinforcing and recruiting new adaptation mechanisms. All these toxicants constitute forms of direct (giving rise to DNA mutations) and indirect (favoring certain phenotypes) selective pressure. Immune and metabolizing systems must recognize and eliminate these new compounds, with the less possible disrupting effects in homeostasis. When this purpose is not well achieved disease can emerge.

In this context, microevolution refers to relatively small changes in few genes, which, in most cases, lead to relatively small changes in phenotypes; it affects the individual within the

population. Modern techniques allow scientists to follow microevolution at the molecular level, through the measurement of genome sequences and protein expression patterns. In contrast, macroevolution refers to large-scale changes in populations as a whole, including formation of new species. A major challenge in modern biology is to understand how large-scale events, such as the development of new species, can occur as a result of microevolutionary events (Lesk AM, 2005). It is important to recognize the biochemical aspects of species adaptation and evolution, and biological research assumes in this field a fundamental role.

1.2 Protein Evolution and Disease

Molecular evolution begins with allele-frequency changes in a population. To study these initial changes, researchers sample a population and test all the individuals in the sample for the presence of variants of a particular molecule. The tests reveal different frequencies of alleles at a given locus in the population – polymorphism (Cooper DN *et al*, 2008).

Scientists estimate that approximately 70% of disease-causing mutations occur at structurally and functionally important sites with well defined properties such as less than 5% of solvent accessibility, sometimes located in β-strands, active sites, disulfide bonds or evolutionary conserved sites. Moreover, it was found that most of the allelic variants map to the same structurally and functionally important regions of proteins, suggesting that many of them probably have negative effects on the phenotype (Sunyaev S *et al*, 2000).

Human mutations that result in disease are overabundant at amino acid positions highly conserved throughout the long-term history of metazoans. Conversely, human polymorphic replacement mutations and silent mutations are randomly distributed across sites with respect to their level of conservation.

The age of mutations can be estimated from the decay of linkage disequilibrium with flanking or intragenic polymorphisms; because of recombination the frequency of mutation itself is a consequence of genetic drift. Several methods have been proposed, and the results of their application can be combined with population data to provide a critical view of the origin and the natural history of the mutation (Cooper DN *et al*, 2008).

Natural variations in families of homologous proteins, which retain a common function, reveal how structures accommodate changes in amino acid sequence. Surface residues not involved in function are usually free to mutate. Loops on the surface can often accommodate changes by local refolding. Mutations that change the volume of buried residues usually do not change the conformation of individual helices or sheets, but produce distortions of their spatial assembly.

Families of related proteins tend to retain common folding patterns. However, although the general folding pattern is preserved, there are distortions that increase as the amino acid sequences progressively diverge. Until the fraction of identical residues in the sequence drops below about 40-50%, these effects are relatively modest. With increasing sequence divergence, some regions refold entirely, reducing the size of the core, and the distortions of the residues remaining within the core increase in magnitude (Lesk AM, 2005).

It is estimated that the number of nuclear human genes range from 25000 to 30000. Considering that coding sequences have in average 1000 base pairs, they represent only 1% of the human genome, and the remaining 99% are non-coding regions. It should be noted that the larger number of genes in humans (and eukaryotes in general) as compared with bacteria is mainly due to the fact that many eukaryotic genes exist as multigene families, which are the result of genome and gene duplications during evolution (Cooper DN *et al*, 2008).

Studies have consistently shown that genetic diversity in humans is low when compared with other species, including mice and apes, suggesting that human lineage underwent an evolutionary recent *population bottleneck* (Strachan T *et al*, 2004).

The sequencing of the human genome has not only helped us understand its structure and function, and explore the full range of human genotypic diversity, but has also provided the key to understand the evolutionary history of the human species and that of individual human populations (Cooper DN *et al*, 2008).

1.3 The Biological Importance of Phosphorylation

Why do protein phosphatases have such an important role in biochemical steps across all living beings? Reversible protein phosphorylation is a dynamic and ubiquitous mechanism, indispensable for the regulation of virtually any cellular function. Therefore, protein kinases and phosphatases are of paramount importance for the normal function of all metabolic and signaling pathways (Andreeva AV, 2004). Reversibility is the main characteristic of this regulatory mechanism.

Phosphorylation is a favorable thermodynamic process, which occurs efficiently in the presence of the adequate catalyzer (kinase or phosphatase). The protein phosphorylation system has three main constituents: a protein kinase, responsible for the transference of a phosphate group from ATP to the target-protein; a protein phosphatase, responsible for the hydrolysis of the phosphate group from the target-protein (dephosphorylation), and the target-protein, also named phosphoprotein (Silva EFC, 1998).

Early studies on protein phosphatases were focused on their role in modifying enzyme activity; more recently they have been found to play a role also in protein localization, in protein turnover and in providing a specific interaction site for other proteins (Moorhead *et al*, 2009).

1.4 General Aspects of Protein Phosphatases Classification

Scientists think that of all the proteins expressed by a single cell, 1/3 are phosphoproteins (Silva EFC, 1998). Until now there have been identified more than 500 kinase and 140 phosphatase coding genes (Gonzalez-Kristeller, 2007). Recently it has been postulated that 2-4% of the genes, in a typical eukaryotic genome, codify protein phosphatases and kinases (Moorhead *et al*, 2009).

Protein phosphatases are classified according to the amino acid residue of the protein substrate that they dephosphorylate. The amino acid residues that are known to be involved in reversible protein phosphorylation are serine, threonine, tyrosine, histidine, arginine, lysine and aspartate. In eukaryotes, the phosphoesterification of proteins (on serine, threonine and tyrosine) is the predominant – in humans, phosphorylation on serine, threonine and tyrosine is approximately 86.4%, 11.8% and 1.8% respectively (Olsen JV *et al*, 2006).

Thus eukaryotic protein phosphatases are mainly tyrosine specific phosphatases (PTP), serine/threonine specific phosphatases (PP) and dual-specificity phosphatases (DSP), which are structurally related to PTP, but can dephosphorylate the three phosphoesterified residues. This work concerns serine/threonine specific protein phosphatases.

1.5 The Serine/Threonine-Specific Protein Phosphatases

Concerning the current classification of PP (serine/threonine-specific protein phosphatases) it reflects the structural differences of the enzymes and is based on the amino acid sequences of the catalytic subunits. They are classified in PPM (Metallo-Dependent Protein Phosphatases), FCP (Fyn Carboxy-Terminal Peptide) and PPP (Phosphoprotein Phosphatases).

Table 1. Serine/Threonine Protein Phosphatases Categories

Serine/Threonine Phosphatases Cathegories (PP)	
PPM	Metallo-Dependent Protein Phosphatases
FCP	Fyn Carboxy-Terminal Peptide
PPP	Phosphoprotein Phosphatases

The PPM comprises Ppm1 (previously known as PP2C), a mitochondrial piruvate dehidrogenase-phosphatase and some prokaryotic enzymes (Gonzalez-Kristeller, 2007). PPM are monomeric enzymes (not modulated by inhibitors or regulatory subunits) that require Mg^{2+} or Mn^{2+} to its activity and are insensitive to okadaic acid (Barford, 1995; Wera *et al*, 1995). In eukaryotes, PPM enzymes are related with the cellular response to stress, the dephosphorylation in some signaling pathways and in cell cycle regulation (Gonzalez-Kristeller, 2007).

The FCP is a more recently identified family of PP enzymes, which dephosphorylate the carboxy-terminal domain of RNA polymerase II (Meinhart A *et al*, 2005). Their catalytic mechanism involves the metal-assisted phosphorylation of the first aspartate residue in the active site -DXDX(T/V)- conserved motif (Kamenski T *et al*, 2004).

The PPP family is defined by three signature motifs in the N-terminal subdomain (-GDXHG-, -GDXVDRG- and -GNHE-), within a ~280 amino acid catalytic domain. These motifs contain most residues that coordinate the metal ions in the active centre (Barford D, 1996).

Remarkably, PPM and PPP families are unrelated in sequence and have probably evolved from two unique ancestral genes, but converged to have a highly related structure at the catalytic centre (Das AK *et al*, 1996).

PPP enzymes are regarded as very ancient, with most members widely distributed in all eukaryotes, and most bacterial and archeal genomes having at least one PPP-like member (Moorhead *et al*, 2009).

The amino acid sequences of most PPP catalytic subunits have been extremely conserved throughout the evolution of multicellular eukaryotes, particularly Ppp1c and Ppp2c, which are among the most slowly evolving proteins known (Moorhead *et al*, 2009). This may be explained by their many roles in crucial cellular processes and the fact that they interact with different regulatory subunits.

In mammals there are seven PPP enzymes, in a total of 13 catalytic subunits, including three isoforms of Ppp1c, two of Ppp2c, three of Ppp3c, one Ppp4c, one Ppp5c, one Ppp6c, and two Ppp7c.

The seven human PPP enzymes will now be discussed in detail.

Ppp1

It is one of the main serine/threonine protein phosphatases found in eukaryotic cells. It is formed by the association of the catalytic subunit with one or more regulatory subunits, which determine

its subcellular localization, substrate specificity and catalytic activity (Hubbard MJ *et al*, 1993).

The three isoforms of the catalytic subunit - Ppp1cα, Ppp1cβ, and Ppp1cγ - are highly homologous and present similar substrate specificity, but exhibit different patterns of distribution. The subcellular localization and the specific substrate are determined by the regulatory proteins, which are integral subunits of the enzyme (example: the M subunit targets Ppp1 to myofibrils in smooth and skeletal muscle).

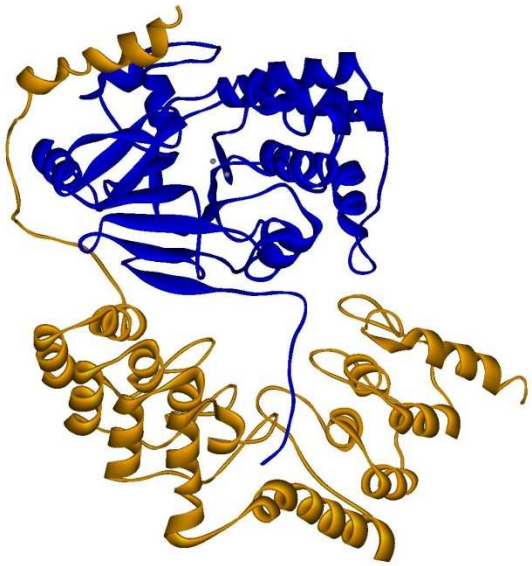


Fig. 1 Ppp1cα (in blue) interacting with the regulatory M subunit (brown).

The regulatory subunits are generally unrelated, but most possess common features, particularly the -(R/K)(V/I)XF- motif – a canonic sequence found to be critical for interaction with Ppp1c (Terrak M *et al*, 2004; Egloff MP *et al*, 1997).

The cell nucleus also contains high concentrations of Ppp1, where it is believed to play an important role in dephosphorylation of transcription factors critical to control gene expression (example: HIV-1 Tat protein is known to interact with Ppp1c and, as a consequence, activate HIV-1 genes transcription) (Ammosova T *et al*, 2005).

The physiological processes where Ppp1 plays a regulatory role include: muscle contraction/relaxation, glycogen metabolism, synaptic transmission, gene transcription, RNA splicing and cell cycle progression.

Ppp2

Ppp2 is a heterotrimeric enzyme constituted by a structural, a regulatory and a catalytic subunit, denominated A, B and C respectively. Subunit C (the catalytic subunit) interacts with subunit A (the scaffolding subunit) to form the so called core enzyme, which associates with the variable

subunit B to form the active Ppp2 holoenzyme (Xu YH *et al*, 2006).

In mammals there are two isoforms of the catalytic subunit, α and β, with 97% of genetic homology (Hemmings BA *et al*, 1990). While the precise functions of each B subunit remain incompletely understood, it appears that by being expressed differentially in mammalian tissues, they determine the cell type-specific functions of Ppp2.

Ppp2 is mainly a cytosolic enzyme but also exists in the cell nucleus. It plays a role in cell-cycle regulation, cell growth control, development, regulation of multiple signal transduction pathways, cytoskeleton dynamics, cell mobility, metabolism, transcription, translation, RNA splicing, DNA replication, apoptosis, inflammation, differentiation, among others (Janssens V *et al*, 2001). It is also considered an important tumor-suppressor protein (Eichhorn PJA *et al*, 2009).

Ppp3

Ppp3, also called calcineurin (Ca^{2+}/calmodulin-activated enzyme), has two major subunits: subunit A which has the catalytic activity and subunit B which contains two calmodulin-like domains that bind Ca^{2+}.

There are three isoforms of the catalytic subunit A in mammals: Ppp3cα, Ppp3cβ and Ppp3cγ. The expression of the first two is ubiquitous, although presenting higher levels in the brain, particularly in striatal regions, but Ppp3cγ is testis-specific.

The catalytic subunit possesses a globular catalytic domain, similar to the other Pppc, but with two small inserted regions, plus C- and N-terminal extensions. The long carboxy-terminal extension includes a subunit B-binding domain, a calmodulin-binding domain and an auto-inhibitory domain. The short N-terminal extension seems only to support the binding of subunit B. The auto-inhibitory domain interacts with the active site, blocking the entry of the substrate, thus inhibiting the catalytic activity. This inhibition is somewhat relieved by the interaction of Ca^{2+} bound subunit B, but only the subsequent binding of Ca^{2+}/calmodulin can completely displace the short α-helical auto-inhibitory peptide from the active site and enhance the catalytic activity (Kissinger CR *et al*, 1995).

The activity of Ppp3 is influenced by a series of proteins, also called immunophilins because they were first identified as the targets of potent immunosuppressive drugs. The immunophilins cyclophilin and FKBP12 are targeted by Cyclosporin A and Tacrolimus (FK-506), respectively (Liu *et al*, 1991).

It is now known that immunossupressive drugs promote the interaction of immunophilins with Ppp3, noncompetitively inhibiting its catalytic activity and preventing the activation of T-

lymphocytes.

Ppp4

Ppp4c interacts with one of two different regulatory subunits - Ppp4r1 and Ppp4r2 -, which interestingly results in a substantial decrease of its catalytic activity. This can happen because the regulatory subunits induce narrow substrate specificity or because they inhibit the catalytic subunit. In this latter case, Ppp4r1 and Ppp4r2 may be in fact an equivalent to the scaffolding subunit of Ppp2, which means that the complex of Ppp4c with either regulatory subunit will interact with a variable third subunit in order to form a catalytically active heterotrimeric holoenzyme. This scenario is supported by the finding that the amino acid sequence of Ppp4r1 has homology to the A subunit of Ppp2 (Kloeker S *et al*, 1999).

Ppp4 is present in most human tissues, with the highest levels of expression in brain, liver and testis (Kloeker S *et al*, 1997). It is predominantly nuclear but is also present in the cytoplasm and, more interestingly, in mitotic centrosomes, where it seems to play a role in microtubule organization (Helps NR *et al*, 1998). It is Ppp4r2 that targets Ppp4 to centrosomes (Hastie CJ *et al*, 2000). Ppp4 also seems to have a pro-apoptotic role in T-lymphocytes (Mourtada *et al*, 2003).

Ppp5

Contrarily to other PPP enzymes, Ppp5 consists of a single polypeptide chain that has catalytic, regulatory and subcellular targeting functions.

The catalytic domain of Ppp5 is closely related to the catalytic subunits of the other PPP enzymes, but the enzyme presents an additional amino-terminal extension that modulates the binding of the target proteins and the maintenance of a low basal activity (Swingle MR *et al*, 2004). This extension is called the TPR domain, because it presents several tetratricopeptide repeats (TPR); these consist of tandem repeats of a 34-amino acid consensus sequence, which usually mediate protein-protein interactions and are present in proteins having diverse functions and subcellular localizations (Blach GL *et al*, 1999).

Ppp5 was found to be activated *in vitro* by arachidonic acid and other polyunsaturated fatty acids, through binding to the TPR domain (Chen MX *et al,* 2004).

This enzyme is ubiquitously expressed, although its levels vary among different tissues, being higher in the brain. The enzyme is present both in the nucleus and in the cytoplasm, and appears to be involved in glucocorticoid signaling (Chen MS *et al* 1996; Zuo Z *et al*, 1999), cell-cycle regulation (Urban G *et al*, 2003; Zuo Z *et al*, 1998), atrial natriuretic peptide signaling

(Chinkers M, 1994), ion channels regulation (Yamaguchi Y *et al*, 2002), circadian rhythm regulation (Chinkers M, 2001), DNA repair (Wechsler T *et al*, 2004; Ali A *et al*, 2004) and cell response to stress (Huang SL *et al*, 2004; Zhou GF *et al*, 2004; Morita K *et al* 2002).

Ppp6

Ppp6c presents the highest levels of expression in testis, heart and skeletal muscle (Bastians H, 1996). A genome-wide screen for kinases and phosphatases that affect cell survival, identified Ppp6 as an important negative regulator of apoptosis (MacKeigan JP *et al*, 2005). Until now, three potential regulatory subunits of Ppp6 were identified (Stefansson B *et al*, 2006). This enzyme appears to be involved in the regulation of transcription, translation, morphogenesis and cell cycle (Cohen PTW, 1997).

Ppp7

Ppp7, like Ppp5, is a monomeric enzyme. Its polypeptide chain contains a phosphatase catalytic domain with all the invariant motifs of the PPP family, but presents unique N- and C-terminal regions plus an additional inserted 43 amino acids long domain in the phosphatase core region. The amino-terminal extension is a potential calmodulin-binding site (Kutuzov MA *et al*, 2002), whereas the carboxy-terminal extension includes multiple Ca^{2+} binding sites. Ca^{2+} greatly enhances the activity of Ppp7, which is only basal in the absence of the ion (Huang XZ *et al*, 1998).

Contrarily to the other PPP, the expression of Ppp7 is not ubiquitous. The enzyme was only detected in retina and in human fetal brain, which suggests that its functions are more limited and possibly confined to those tissues. Ppp7 seems to be involved in the control of the phosphorylation status of G-protein coupled receptors (Huang XZ *et al*, 1998).

As referred above, all catalytic subunits of the PPP family present a common globular domain, which is responsible for the catalytic activity, and corresponds to a highly conserved amino acid sequence that includes the characteristic signature motifs of the family. In addition to this common structural and functional part, Ppp3c, Ppp5c and Ppp7c possess fused amino- and/or carboxy-terminal domains, which confer them distinct properties, such as auto-inhibitory functionality or sensitivity to Ca^{2+} through a calmodulin-binding sequence. These N- and C-terminal extensions can be regarded as fused regulatory subunits, representing the evolutionary strategy adopted mainly by PTP and PPM that resulted in multi-domain monomeric phosphatases. In fact, Ppp5 and Ppp7 are the only enzymes within the PPP family, that are monomeric and do not need to interact with separate subunits to become catalytically active.

(There is also another group of phosphatase enzymes used in clinical medicine as important biochemical markers, the acidic and alkaline phosphatases, which are not specific and occur in particular intracellular or extracellular environments.

Table 2 summarizes some information regarding the seven human PPP enzymes:

Table 2. The mammalian serine/threonine specific protein phosphatases of the PPP family

Protein Phosphatase		Catalytic Subunit				
Name	**Main Function**	**Name**	**Active-Site Metals**	**Gene**	**Gene Location (Human)**	**Tissue Expression**
Ppp1	Muscle relaxation, synaptic transmission, gene expression, glycogen metabolism, RNA splicing, cell cycle progression.	Ppp1cα Ppp1cβ Ppp1cγ	Fe^{3+}/Zn^{2+}	PPP1CA PPP1CB PPP1CC	11q13 2p23 12q24.1-q24.2	Overall
Ppp2	Cell cycle regulation, cell growth control, cytoskeleton dynamics, cell mobility, metabolism, transcription, translation, RNA splicing, DNA replication, apoptosis, inflammation, differentiation.	Ppp2cα Ppp2cβ	Mn^{2+}	PPP2CA PPP2CB	5q31.1 8p12	Overall
Ppp3	Physiologic response to neurotransmitters and nerve impulses, involved in NMDA-receptor signaling pathway and T-lymphocyte activation.	Ppp3cα Ppp3cβ Ppp3cγ	Ca^{2+}	PPP3CA PPP3CB PPP3CC	4q21-q24 10q21.q22 8p21.3	Overall “ Testis
Ppp4	Centrosome maturation, microtubule organization, histone phosphorylation (regulation of chromatin activities), apoptosis.	Ppp4c		PPP4C	16p12-p11	Overall
Ppp5	Cell growth, ribosomal RNA transcription regulation, atrial natriuretic peptide signaling, steroid signaling, blue light signal transduction (circadian cycle regulation?), ion channels regulation.	Ppp5c		PPP5C	19q13.3	Overall
Ppp6	Cell cycle regulation, transcription, translation, morphogenesis.	Ppp6c		PPP6C	9q33.3	Overall
Ppp7	Control of phosphorylation status of G protein coupled receptors.	Ppp7cα Ppp7cβ		PPP7CA PPP7CB	Xp22.2-p22.1 4q21.1	Retina and Brain

Adaptado de Gonzalez-Kristeller, 2007

1.6 The Three-Dimensional Structure of PPP Catalytic Subunits

As an example we will now discuss the Ppp1c structure.

Until date PPP1c protein was crystallized in complex with some of its inhibitors and regulatory subunits (e.g.: calyculin A, microcystins, okadaic acid and MYPT1). These studies showed that PPP1c has an ellipsoid structure composed by 11 β-sheets, surrounded by seven α-helices at one side and a subdomain composed by three α-helices and three β-sheets in another. In the central protein region, three β-sheets and two α-helices are connected forming a β-α-β-α-β motif. This motif is probably related to the active site of the enzyme and to the metallic ions coordination by Asp64, His66 and Asp92 (Gonzalez-Kristeller, 2007).

Associating structural and kinetic studies, scientists concluded that the dephosphorylation of the substrates is catalyzed in one step without involvement of intermediates. This is not observed in acidic and tyrosine phosphatases (Gonzalez-Kristeller, 2007).

Some studies indicated that the connection between β-sheets 12 and 13 (β12-β13 loop) is essential to the inhibition of the enzyme by toxins. Mutations in this region increases or decreases PPP1c sensibility, suggesting that critical elements for enzyme inhibition are localized between residues 269-301 (Gonzalez-Kristeller, 2007).

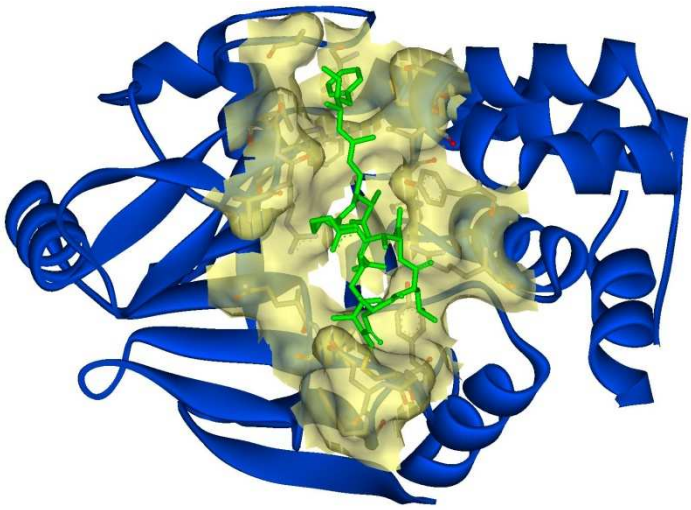


Fig 2. The Ppp1c structure interacting with microcystin-LR (inhibitor).

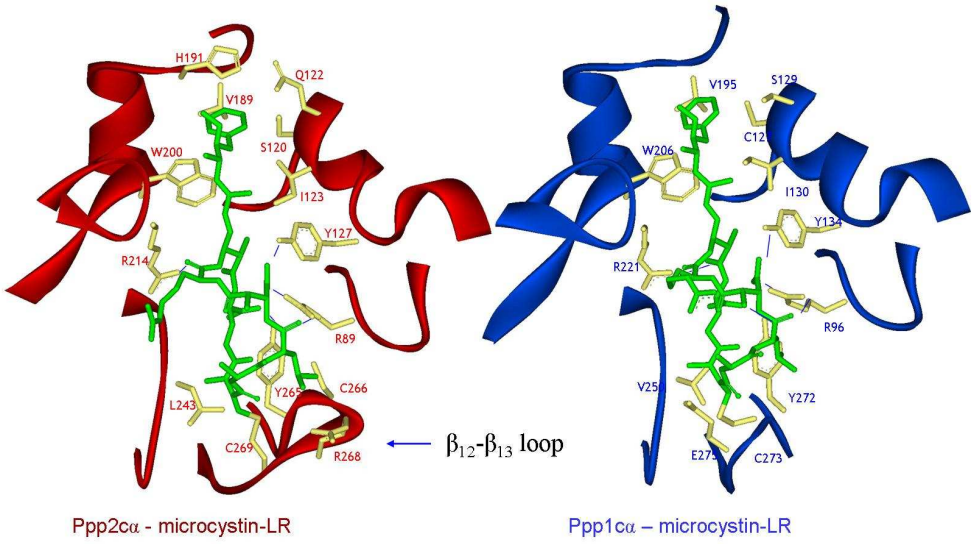


Fig 3. The Ppp2cα (red) and Ppp1Cα (blue) active sites interacting with microcystin-LR (inhibitor)

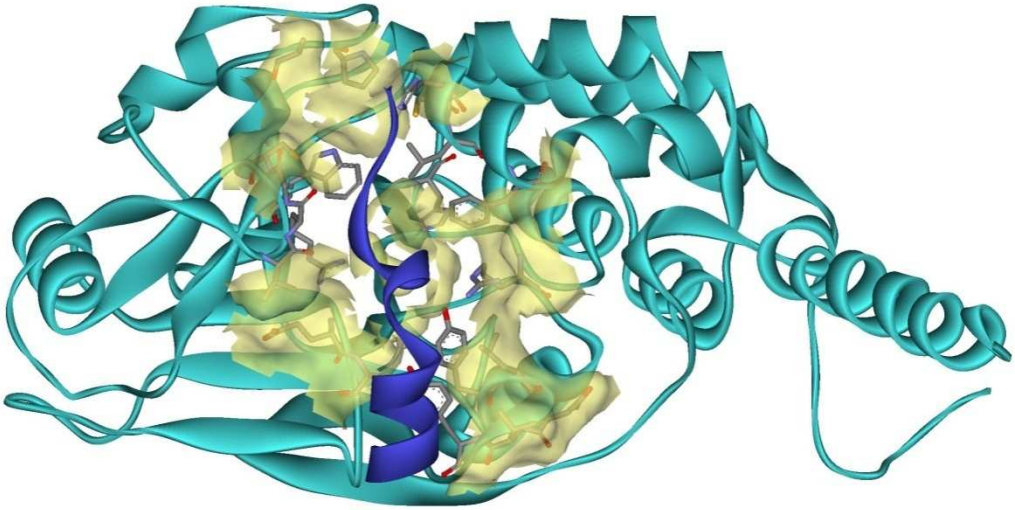


Fig 4. The Ppp3cα structure.

2. Discussion

Concerning the developed activities during this training period, it was initially organized in four parts, each one comprising different activities:

Part I: Bibliographic research, scientific information reading and selection of the investigation theme
Part II: Introduction to laboratorial work
Part III: Introduction to bioinformatics
Part IV: Training period report writing

Bibliographic research, scientific information reading and selection of the investigation theme

This part had the duration of two months (48 hours) and was devoted to bibliographic research and selection of appropriate information sources.

During this period the student read some selected information related to scientific research, genetics, biochemistry, bioinformatics and human health. It was the time to select the scope of investigation and to program the study.

Introduction to laboratorial work

This part had the duration of two months (48 hours). The student had opportunity to read some laboratorial protocols related to DNA extraction, DNA purification, PCR and DNA sequencing (DNA extraction had been addressed in a practical lesson in the Biological Chemistry subject of the Medicine course and PCR techniques were taught in different subjects along the course). These activities allowed the student to be familiarized with the laboratorial equipment, to become aware of the length and sequence of the procedures, and to realize the rigor necessary to manipulate biological specimens. It is essential to refer that all the techniques, which the student observed and/or collaborated in, are used nowadays as diagnostic procedures in clinical medicine, especially in genetic and oncologic areas.

Concerning DNA extraction, it can be done on different biologic specimens. The protocol for DNA extraction from blood leucocytes is described in attachment 4.

PCR (Polymerase Chain Reaction) is a technique that allows the production of more than 10 million copies of a target DNA sequence from only a few molecules. The sensitivity of this technique requires that the sample should not be contaminated with any other DNA or

previously amplified products (amplicons), which may reside in the laboratory environment. PCR is a series of controlled heating and cooling steps performed in a thermal cycling equipment. The steps that are repeated cyclically are three: (1) heating at 94ºC to denature the double-stranded DNA; (2) cooling to the annealing temperature (dependent on the melting temperature of the expected duplex), which permits the chemical primers to bind to their specific target; (3) raising the temperature to the optimum value (usually 70-75ºC) for a suitable heat-stable DNA-polymerase to originate a new strand of DNA. In the presence of the four DNA precursors (deoxynucleoside triphosphates), the primers initiate the synthesis of new DNA strands, which are complementary to the individual DNA strands of the target DNA segment. The design of the primers is aided by various commercial software programs and by freeware programs such as those compiled at http://www.hgmp.mrp.ac.uk/GenomeWeb/nucprimer.html (Strachan T *et al*, 2004).

In the medicine field, the PCR technique is currently used in forensics, clinical genetics, clinical microbiology, oncology and other areas.

Regarding DNA sequencing (Protocol in attachment 5), it has applications in many fields of clinical medicine, some examples of which are genetics, oncology, microbiology, hematology, and forensics. The vast majority of current DNA sequencing techniques uses an enzymatic method: the DNA is provided in a single-stranded form and serves as a template for a new complementary DNA strand using a suitable DNA polymerase *in vitro*. The automated DNA sequencing uses fluorescent labeling. During electrophoresis, a monitor detects and records the fluorescence signal as the DNA passes through a single point in the gel. This allows an output in the form of intensity profiles for each of the differently colored fluoroscopes, while simultaneously storing the information electronically. If the DNA sequence is known to be a coding sequence, the output can immediately be translated in different reading frames to infer a polypeptide sequence (Strachan T *et al*, 2004).

Introduction to bioinformatics

This part – bioinformatics data search and analysis – had the duration of six months (144 hours) and, in the student's opinion, was the most difficult but also the most challenging. This phase comprised the following aspects:

Introduction to gene and protein databases;
Introduction to BLAST (Basic Local Alignment Search Tool) procedures;
Utilization of FASTA format for information saving;

Introduction to the MEGA 4™ program, to make comparative genetic analysis;

Interpretation of the data and reflection about the obtained results;

Construction of phylogenetic trees and their interpretation;

Introduction to Accelrys DS Visualizer 2.0™ program (to study the 3D structure of proteins obtained from X-ray crystallography).

Interestingly, these activities occurred at the same time as the Clinical Genetic subject of 5th year of the Medicine course, in which the student also learned the BLAST procedure.

Most of the organized activity of bioinformatics has been focused on the analysis of the following principles:

DNA sequence determines protein sequence;

Protein sequence determines protein structure;

Protein structure determines protein function;

Regulatory mechanisms, including but not limited to control of expression patterns, deliver the right amount of the right function to the right place at the right time (Lesk AM, 2005).

In this work, those principles were applied to the study of the PPP family of serine/threonine-specific protein phosphatases. First we studied their role in cellular physiology and then we evaluated the evolution of their catalytic subunits among metazoan species. To achieve the first purpose we researched and read scientific information, mainly articles and books. For the second purpose, we searched and compiled PPPc sequences of several metazoan species, from online databanks, which are archives of biochemical information available for research studies. Our sources of information were mainly NCBI and ENSEMBL databanks.

Regarding the choice of the species to include in the evolutionary study, it was conditioned, on the one hand, by the animals whose genomes were already sequenced and, on the other, by the effort to generate a broad and representative sample, including the main *Phyla* of animal taxonomy.

To retrieve sequences from the databanks, it was necessary to use BLAST tools. The BLAST program and its variants (Table 3.) check each entry in the databank independently, against a query sequence. These programs compare amino acid sequences, using by default the BLOSUM62 matrix. Searches involving nucleotide sequences, either as query sequence or as database sequences to be searched, are carried out by translating nucleotide sequences to amino acid sequence in all six possible reading frames. Another program in this family,

BLASTN, compares nucleotide query sequences with nucleotide entries of the databanks directly (Lesk AM, 2005).

Here, two concepts must be clarified: sensitivity and selectivity. As an example, if we identify a gene in the human genome, which is responsible for some disease, and we wish to determine whether related genes appear in other species, the ideal method is both sensitive – it picks up even distant relationships – and selective – all the relationships it reports are true. For this purpose, a powerful tool for searching sequence databases with a probe sequence is PSI (Position Sensitive Iterated)-BLAST from NCBI (Lesk AM, 2005).

Table 3. Type of BLAST programs

Program	Type of query sequence	Search in database of
BLASTP	Amino acid sequence	Protein sequences
BLASTX	Translated nucleotide sequence	Protein sequences
TBLASTN	Amino acid sequence	Translated nucleotide sequences
TBALSTX	Translated nucleotide sequence	Translated nucleotide sequences
PSI-BLAST	Amino acid sequence	Protein sequence database

In our study, TBLASTN was the main tool used: given a protein sequence, or a fragment of a protein sequence (for example, the conserved motifs of PPP catalytic subunits), it searched the databases (in a first approach the genes database of each organism) for translated sequences that were similar to it.

In fact, the first PPPc coding sequences were retrieved from the database without using a BLAST procedure. Regarding the genomes already well studied and annotated (e.g. the human genome), the genes were simply searched by their name. These sequences were then used to BLAST all the genomes, in order to get the homolog sequences from genomes not yet annotated and to confirm that we had all the PPPc sequences of the genomes already annotated.

Table 4. Number of Sequences for Each PPP Catalytic Subunit Per Species

SPECIES	COMMON NAME	Ppp1c	Ppp1c-related	Ppp2c	Ppp4c	Ppp6c	Ppp2/4/6c-related	Ppp3c	Ppp5c	Ppp7c (rdgC)
Homo sapiens	man	3		2	1	1		3	1	2
Mus musculus	mouse	3		2	1	1		3	1	2
Bos taurus	cow	3		2	1	1		3	1	2
Sus scrofa	pig	3		2	1	1		3	1	2
Tursiops truncatus	dolphin	3		2	1	1		3	1	2
Canis familiaris	dog	3		2	1	1		3	1	2
Pteropus vampyrus	megabat	3		2	1	1		3	1	2
Monodelphis domestica	opossum	3		2	1	1		3	1	2
Ornithorhynchus anatinus	platypus	3		2	1	1		3	1	1
Gallus gallus	chicken	3		2	-	1		2	-	2
Taeniopygia guttata	darwin's finch	3		2	-	1		3	-	2
Anolis carolinensis	green anole	3		2	1	1		3	1	2
Xenopus laevis	african clawed frog	4		2	1	-		1	1	-
Xenopus tropicalis	frog	3		2	1	1		3	1	1
Danio rerio	zebra-fish	5		3	2	1		4	1	3
Gasterosteus aculeatus	stickleback	6		3	1	2		2	1	3
Oryzias latipes	japanese medaka	5		-	1	2		2	1	3
Tetraodon nigroviridis		5		-	1	1		2	1	3
Takifugu rubripes	fugu	5		2	1	1		2	1	3
*Petromyzon marinus**		5								
Branchiostoma floridae		2	1	1	2	1	1	1	1	1
Ciona intestinallis		3		1	1	2		2	1	-
Strongylocentrotus purpuratus	purple	3		1	1	2		1	1	1
Daphnia pulex	zooplancton			1						
Drosophila melanogaster	fruit fly	4	5	1	1	1	1	3	1	1
Lottia gigantea		2		1	1	1		2	-	1
Nematostella vectensis		2		1	1	1		1	-	1
Trichoplax adhaerens		3		1	1	1		1		

When the sequences obtained were incomplete, had some obvious errors or when there was some PPP gene missing in a certain genome, a cross search was performed in a second databank and, if not sufficient, in databases of more raw data like nucleotide sequences or EST (Expressed Sequence Tags – short sequences of cDNA, synthesized from mRNA). Some initial sequences had missing exons that were completed from one or several EST.

Table 4 summarizes the data retrieved from the online databanks, indicating the number of sequences of catalytic subunits that were found for each PPP enzyme per species.

The FASTA format (Pearson WR) was used for sequence data reading, alignment and reporting results.

One important issue, when analyzing the results, is the distinction between similarity and homology. Similarity is the observation or measurement of resemblance or difference, independently of the source of resemblance. Homology means specifically, that the sequences and the organisms in which they occur are descendents of a common ancestor, which means that the similarities constitute shared ancestral characteristics. Homology is thus an inference from observations of similarity (Lesk AM, 2005).

But how can we define a quantitative measure of sequence similarity? To compare the nucleotides or amino acids that appear at corresponding positions in two or more sequences, we must first assign those correspondences. Sequence alignment is the identification of residue-residue correspondences. Any assignment of correspondences that preserves the order of the residues within the sequences is an alignment. It is the basic tool of bioinformatics (Lesk AM, 2005). The quality of the alignment is measured by a score, a percentage number of the sequences' similarity. The alignments can be pairwise or multiple, in this study we used multiple alignments. The alignments can also be classified as global or local alignments: the first case refers to the alignment of one entire sequence with another entire sequence; the second refers to the alignment of one sequence against a segment of another sequence (Lesk AM, 2005).

The obtained sequences were aligned and the alignment was used to construct phylogenetic trees, using the Mega 4™ software.

Phylogeny is the description of biological relationships, usually expressed as a tree. In computer science, a tree is a graph equivalent, containing nodes (abstract points) connected by edges (represented as lines between the points). Theoretically we can define a tree as a connected graph in which there is exactly one path between every two points. The basic principle of phylogeny is that the origin of similarity is common ancestry. The field of phylogeny has the goal

of working out the relationships among species, populations, individuals or genes. Evolutionary trees determined from genetic data are often based on inferences from the patterns of similarity (Lesk AM, 2005).

Construction of evolutionary trees has many applications in science. As examples we can cite:

-understanding human origins and dispersal;

-understanding the rate of a gene sequence evolution;

-understanding mechanisms for the increase in genetic information (the discovery of gene duplication and the divergence of the two resulting copies);

-molecular epidemiology, including forensics: identifying the original host species of a new pathogen (e.g. HIV), estimating rates of transmission between hosts, identifying an individual as a source of a new infection, establishing paternity, identifying miscreants, are some frequent applications;

-as a sufficient model for genetic evolution (e.g. Markov model is a standard statistical process that can model sequence evolution, considering the continuity through time of both sequences and processes) (Cooper DN *et al*, 2008).

Although phylogeny and genealogy are two methods for modeling the processes by which population genetic variation has evolved, it is important to make a distinction between them: phylogeny is the compressed history of how genetic variants accumulated in lineages of successive copies of DNA sequences; genealogy refers to the actual family tree in which genetic variants arose and were passed from parents to children. They are two ways of seeing the same evolutionary process (Cooper DN *et al*, 2008).

Once an evolutionary tree has been derived, statistical methods can be used to gain a measure of its reliability. A popular method is *bootstrapping*, a form of Monte-Carlo simulation. Typically, a subsample of data is removed and replaced by a randomly generated equivalent data set, and the resulting *pseudosequenc*e is analyzed to see if the suggested evolutionary pattern is still favored. The resampling randomizes the data, but if there is a clear relationship between two sequences, randomization will not erase it (Strachan T *et al*, 2004).

The construction of phylogenetic trees was divided in three phases. First we did partial experimental trees, to certificate that the new genetic sequences, which were found by BLAST in the selected genomes were orthologs, relatively to sequences that we had previously classified as PPPc sequences. Second we constructed a final phylogenetic tree with the

sequences that we had previously classified as PPP sequences. Finally we interpreted the resulting tree based in taxonomic, genetic and biochemical parameters.

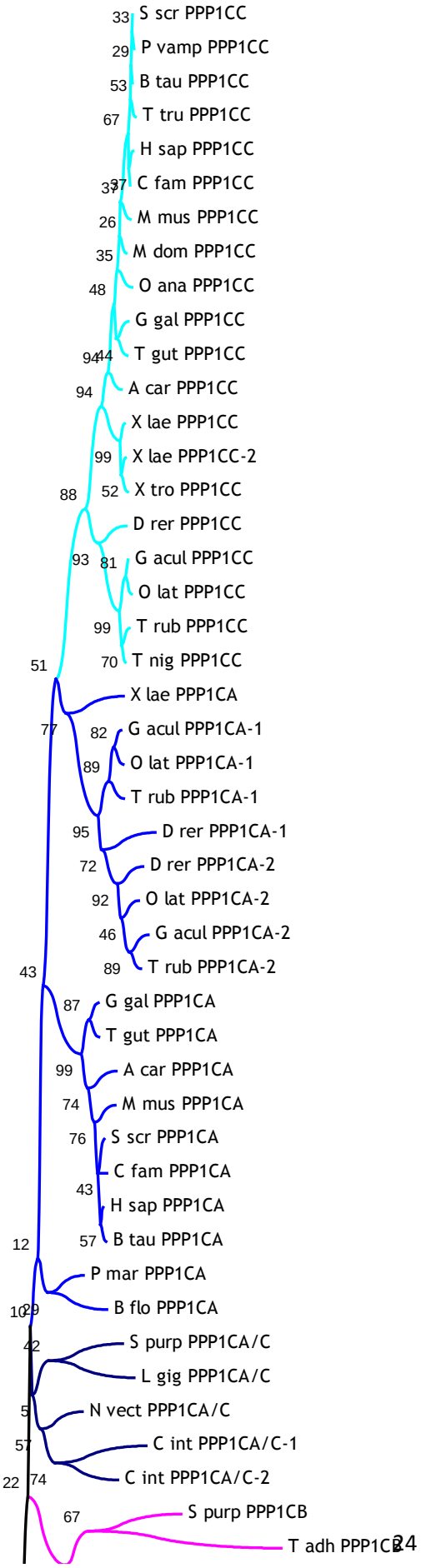
With the obtained data we constructed two final phylogenetic trees.

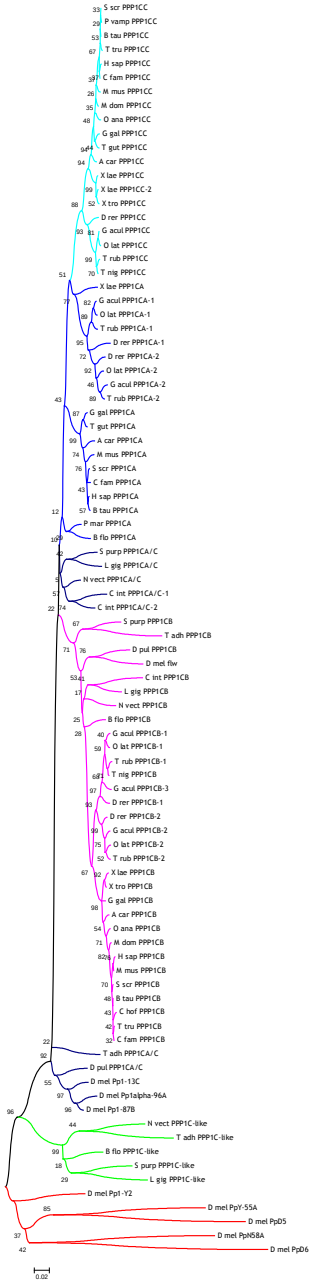
The first tree includes all the PPPc isofoms, of a representative sample of studied animals (*Homo sapiens*, *Canis familiaris*, *Danio rerio*, *Lottia gigantia*, *Nematostella vectensis* and *Trichoplax adhaerens*).

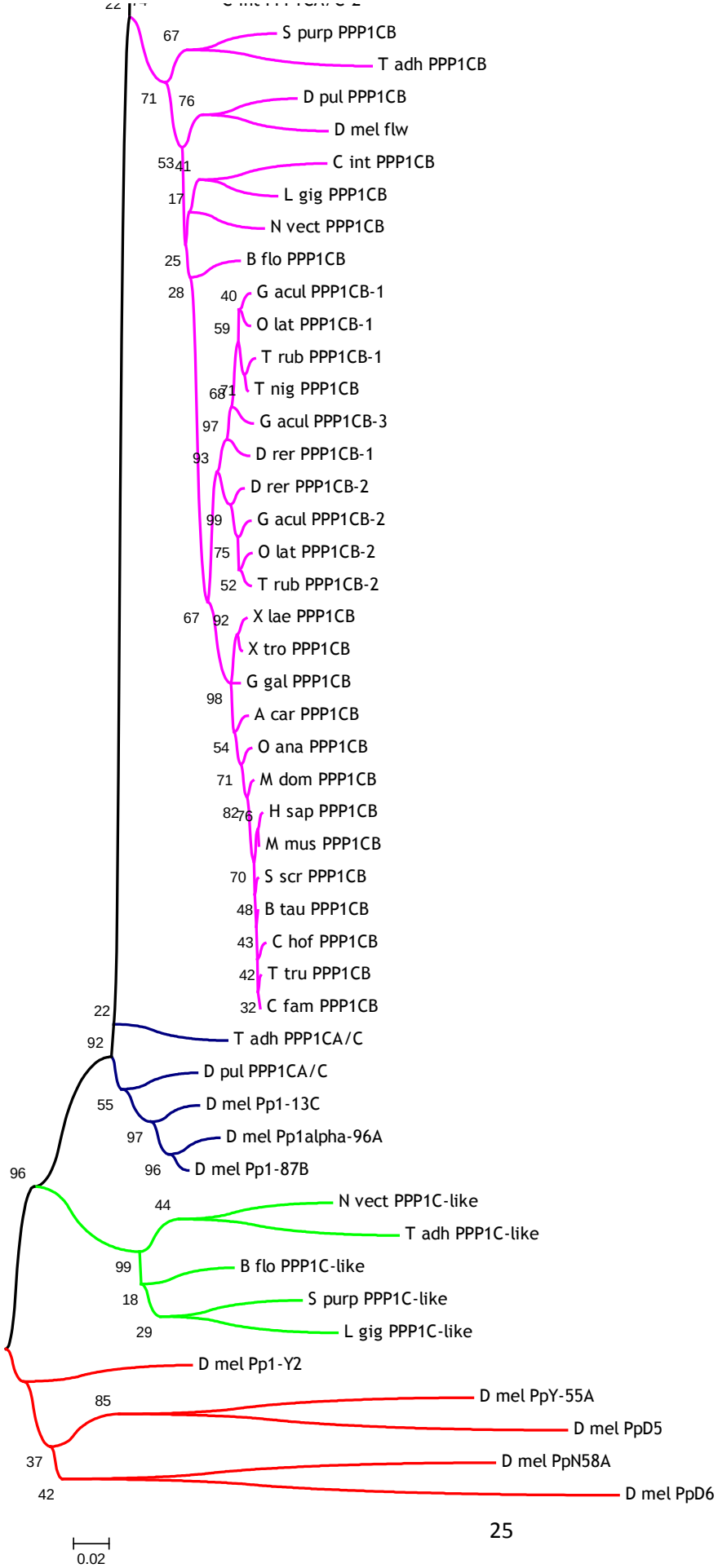
From the analysis of this tree some aspects became relevant, all the metazoans seems to have the same 7 enzymes also present in humans. The different PPP sub-families were well individualized among the metazoan studied species, suggesting that their divergence occurred before metazoan species divergence itself. The Ppp5 and Ppp7 sub-families are more divergent than the remaining PPP sub-families, which can be explained by the fact that Ppp5 and Ppp7 are monomeric enzymes, lacking regulatory subunits and thus more susceptible to evolution strengths in order to function in different biochemical environments. Interestingly the Ppp1c, 2c, 4c and 6c, which interact with regulatory subunits, are more conserved proteins.

It is also important to refer that in *Drosophila melanogaster*, we found some PPP isoforms not present in the other studied animals.

Next we show the tree of PPPc isofoms, of a representative sample of studied animals (*Homo sapiens*, *Canis familiaris*, *Danio rerio*, *Lottia gigantia*, *Nematostella vectensis* and *Trichoplax adhaerens*):


S scr PPP1CC
P vamp PPP1CC
B tau PPP1CC
T tru PPP1CC
H sap PPP1CC
C fam PPP1CC
M mus PPP1CC
M dom PPP1CC
O ana PPP1CC
G gal PPP1CC
T gut PPP1CC
A car PPP1CC
X lae PPP1CC
X lae PPP1CC-2
X tro PPP1CC
D rer PPP1CC
G acul PPP1CC
O lat PPP1CC
T rub PPP1CC
T nig PPP1CC
X lae PPP1CA
G acul PPP1CA-1
O lat PPP1CA-1
T rub PPP1CA-1
D rer PPP1CA-1
D rer PPP1CA-2
O lat PPP1CA-2
G acul PPP1CA-2
T rub PPP1CA-2
G gal PPP1CA
T gut PPP1CA
A car PPP1CA
M mus PPP1CA
S scr PPP1CA
C fam PPP1CA
H sap PPP1CA
B tau PPP1CA
P mar PPP1CA
B flo PPP1CA
S purp PPP1CA/C
L gig PPP1CA/C
N vect PPP1CA/C
C int PPP1CA/C-1
C int PPP1CA/C-2
S purp PPP1CB
T adh PPP1CB

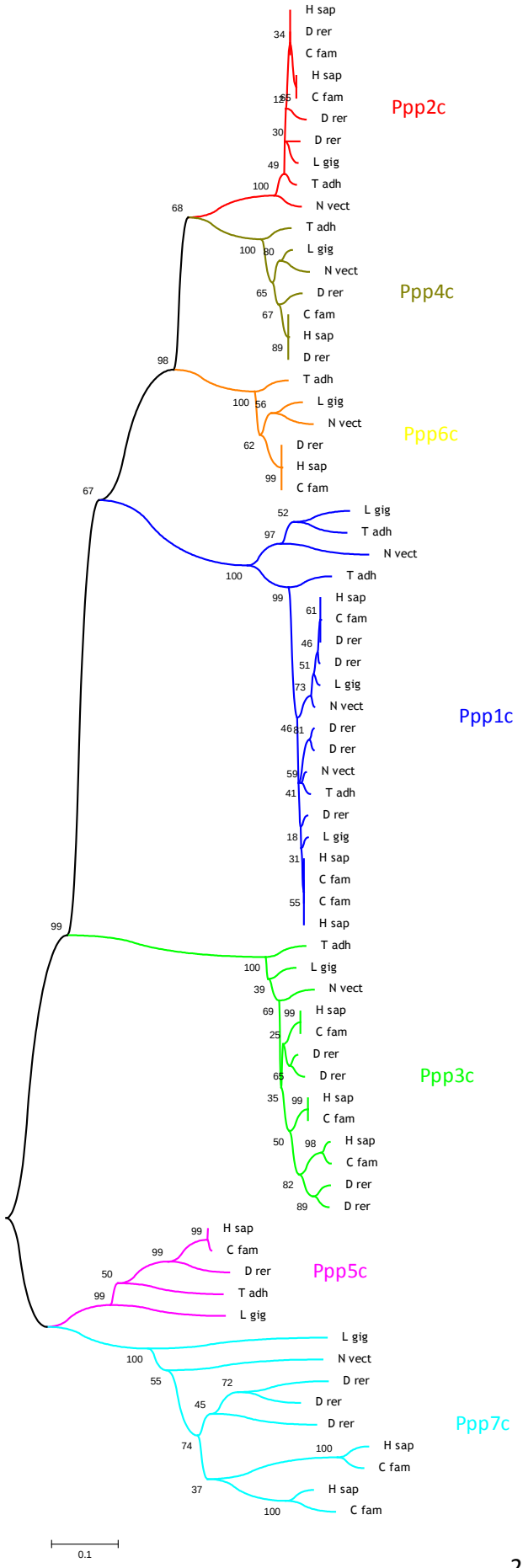



S purp PPP1CB
T adh PPP1CB
D pul PPP1CB
D mel flw
C int PPP1CB
L gig PPP1CB
N vect PPP1CB
B flo PPP1CB
G acul PPP1CB-1
O lat PPP1CB-1
T rub PPP1CB-1
T nig PPP1CB
G acul PPP1CB-3
D rer PPP1CB-1
D rer PPP1CB-2
G acul PPP1CB-2
O lat PPP1CB-2
T rub PPP1CB-2
X lae PPP1CB
X tro PPP1CB
G gal PPP1CB
A car PPP1CB
O ana PPP1CB
M dom PPP1CB
H sap PPP1CB
M mus PPP1CB
S scr PPP1CB
B tau PPP1CB
C hof PPP1CB
T tru PPP1CB
C fam PPP1CB
T adh PPP1CA/C
D pul PPP1CA/C
D mel Pp1-13C
D mel Pp1alpha-96A
D mel Pp1-87B
N vect PPP1C-like
T adh PPP1C-like
B flo PPP1C-like
S purp PPP1C-like
L gig PPP1C-like
D mel Pp1-Y2
D mel PpY-55A
D mel PpD5
D mel PpN58A
D mel PpD6
0.02

The second tree was constructed with all the Ppp1c sequences of the 28 listed in table 5, which are representative of the main animal *Phyla*.

From the analysis of the Ppp1c phylogenetic tree, we can see that, of the tree human isoforms, Ppp1cγ appears to have emerged more recently because it was only identified in vertebrates. Regarding Ppp1cα and Ppp1cβ, they have orthologues in all studied metazoans.

In the next page we show the Ppp1c phylogenetic tree constructed with all the Ppp1c sequences of the 28 listed in table 4:


H sap
D rer
C fam
H sap
C fam
D rer
D rer
L gig
T adh
N vect
T adh
L gig
N vect
D rer
C fam
H sap
D rer
T adh
L gig
N vect
D rer
H sap
C fam
L gig
T adh
N vect
T adh
H sap
C fam
D rer
D rer
L gig
N vect
D rer
D rer
N vect
T adh
D rer
L gig
H sap
C fam
C fam
H sap
T adh
L gig
N vect
H sap
C fam
D rer
D rer
H sap
C fam
H sap
C fam
D rer
D rer
H sap
C fam
D rer
T adh
L gig
L gig
N vect
D rer
D rer
D rer
H sap
C fam
H sap
C fam
Ppp2c
Ppp4c
Ppp6c
Ppp1c
Ppp3c
Ppp5c
Ppp7c
0.1

During this training period I have also the great opportunity to listen a lecture by the investigator Dr. Rita Colwell, about the relationship between *Vibrio cholerae* genetics, environment and public health; which had much importance in this text writing.

Training period report writing

This final part had the duration of two months (48 hours), during which the student analyzed the results obtained and wrote the training period report.

3. Conclusion

PP enzymes evolve in two different manners: some (like PTP) acquire additional domains by fusing at gene level and thus consist of multidomain monomeric enzymes with a strict substrate specificity and a thight regulation; other (like most PPP) consist of highly conserved small catalytic subunits that diversified their biological function by binding novel regulatory subunits encoded by separate genes.

The combinatorial subunit principle of PPP enzymes generates much more diversity and flexibility, allowing the same catalytic subunit to dephosphorylate a plethora of substrates.

The regulatory subunits in addition to substrate specificity also determine the tissue distribution and the subcellular localization of the active holoenzyme.

The PPP enzymes are regarded as very ancient, with most members distributed in all studied metazoans. The catalytic subunits of PPP enzymes, especially in those enzymes that need regulatory subunits (Ppp1, 2, 3, 4 and 6) are highly conserved, being the monomeric PPP enzymes (Ppp5 and 7) more divergent when orthologues from different species are compared. In fact, the conservation of the amino acid sequence of PPP catalytic subunits – roughly indicated in the tree by the length of the arm corresponding to each PPP orthologue – is intimately related to their interaction with a variety of regulatory subunits, which forced the catalytic subunits to the highest degree of conservation known among proteins. On the other hand, the increasing diversity of PPP function observed along eukaryotic evolution resides precisely on the structural and functional plasticity conferred by those regulatory subunits.

New pharmaceutical research must take this in account, targeting, for example, the regulatory and not the catalytic subunits, in order to act in specific pathways, allowing simultaneously other potentially vital roles of the same PPP enzyme. To achieve this purpose, it is necessary to identify the pathway involved in the pathology and the regulatory subunits that target the PPP enzyme to that particular pathway.

Gene duplication seems to be the way by which some PPPc isoforms of *Drosophila melanogaster* evolved. We also observed that one isoform of Ppp1c present in various invertebrates and also in *Branchiostoma floridae* (a *Chordate*) is absent in other superior vertebrates. This fact reflects the dynamics of gene's evolution even in a family of highly conserved enzymes.

Attachements

1. The 20 Amino Acids and Their Official Codes

NUMBER	CODE	NAME
1	A	Alanine
2	R	Arginine
3	N	Asparagine
4	D	Aspartic Acid
5	C	Cysteine
6	Q	Glutamine
7	E	Glutamic Acid
8	G	Glycine
9	H	*Histidine
10	I	*Isoleucine
11	L	*Leucine
12	K	*Lysine
13	M	*Methionine
14	F	*Phenylalanine
15	P	Proline
16	S	Serine
17	T	*Threonine
18	W	*Tryptophane
19	Y	Tyrosine
20	V	*Valine

Adapted from: Silverthorn, Fisiologia Humana. 2ª Ed. Manole 2003 P:32.

*Essential amino acids

2. The DNA Bases and Their Official Codes

CODE	BASE NAME	CHEMICAL GROUP
A	**Adenine**	**Purine**
T	**Thymine**	**Pyrimidine**
C	**Cytosine**	**Pyrimidine**
G	**Guanine**	**Purine**
N	**Any Nucleotide**	**--**
R	**Adenine or Guanine**	**Purine**
Y	**Thymine or Cytosine**	**Pyrimidine**

3. The PCR procedure

(in: http://www.fermentas.com/techinfo/PCR/dnaamplprotocol.htm**)**

A. Composition of the Reaction Mixture

1. **Template DNA.**

   Usually the template DNA amount is in the range of 0.01-1ng for plasmid or phage DNA and 0.1-1µg for genomic DNA, for a total reaction mixture of 50µl. Higher template DNA amounts usually increase the yield of nonspecific PCR products, but if the fidelity of synthesis is crucial, maximal allowable template DNA quantities in conjunction with limiting number of PCR cycles should be used to increase the percentage of "correct" PCR products. Nearly all routine methods are suitable for template DNA purification. Although even trace amounts of agents used in DNA purification procedures (phenol, EDTA, Proteinase K, etc.) strongly inhibit *Taq* DNA Polymerase, ethanol precipitation of DNA and repetitive treatments of DNA pellets with 70% ethanol is usually effective in removing traces of contaminants from the DNA sample.

2. **Primers.**

   Guidelines for primer selection:

   - PCR primers are usually 15-30 nucleotides in length. Longer primers provide sufficient specificity.
   - The GC content should be 40-60%. The C and G nucleotides should be distributed uniformly within the full length of the primer. More than three G or C nucleotides at the 3'-end of the primer should be avoided, as nonspecific priming may occur.
   - The primer should not be self-complementary or complementary to any other primer in the reaction mixture, in order to avoid primer-dimer and hairpin formation.
   - The melting temperature of flanking primers should not differ by more than 5℃, so the GC content and length must be chosen ac cordingly.
   - All possible sites of complementarity between primers and the template DNA should be noted.
   - If primers are degenerate, at least 3 conservative nucleotides must be located at the primer's 3'-end.
   - Estimation of the melting and annealing temperatures of primer: If the primer is shorter than 25 nucleotides, the approx. melting

temperature (T_m) is calculated using the following formula: **T_m= 4 (G + C) + 2 (A + T)**

G, C, A, T - number of respective nucleotides in the primer. Annealing temperature should be approx. 5℃ lower than the melting temperature.

If the primer is longer than 25 nucleotides, the melting temperature should be calculated using specialized computer programs where the interactions of adjacent bases, the influence of salt concentration, etc. are evaluated.

3. **$MgCl_2$ concentration.**

Since Mg^{2+} ions form complexes with dNTPs, primers and DNA templates, the optimal concentration of $MgCl_2$ has to be selected for each experiment. Too few Mg^{2+} ions result in a low yield of PCR product, and too many increase the yield of non-specific products and promote misincorporation. Lower Mg^{2+} concentrations are desirable when fidelity of DNA synthesis is critical. The recommended range of $MgCl_2$ concentration is 1-4mM, under the standard reaction conditions specified. In our experiments, at a final dNTP concentration of 0.2mM, a $MgCl_2$ concentration ranges of 1.5±0.25mM (in traditional PCR buffer) and of 2.0±0.5mM (in PCR buffer with $(NH_4)_2SO_4$) are suitable in most cases. If the DNA samples contain EDTA or other chelators, the $MgCl_2$ concentration in the reaction mixture should be raised proportionally.

4. **dNTPs.**

- The concentration of each dNTP in the reaction mixture is usually 200µM. It is very important to have equal concentrations of each dNTP (dATP, dCTP, dGTP, dTTP), as inaccuracy in the concentration of even single dNTP dramatically increases the misincorporation level.
- When maximum fidelity of the PCR process is crucial, the final dNTP concentration should be 10-50µM, since the fidelity of DNA synthesis is maximal in this concentration range. In addition, the concentration of $MgCl_2$ should be selected empirically, starting from 1mM and increasing in 0.1mM steps, until a sufficient yield of PCR product is obtained.

5. ***Taq* DNA Polymerase.**

Usually 1-1.5u of *Taq* DNA Polymerase are used in 50µl of reaction mix. Higher *Taq* DNA Polymerase concentrations may cause synthesis of nonspecific products. However, if inhibitors are present in the reaction mix (e.g., if the template DNA used is not highly purified), higher amounts of *Taq* DNA

Polymerase (2-3u) are helpful in obtaining a better yield of amplification products.

6. **Reaction overlay.**
   If necessary, the reaction mixture can be overlaid with mineral oil or paraffin (melting temperature 50-60℃) of special PCR grade. One-half of the total reaction volume is usually sufficient.

B. Temperature Cycling

1. **Initial Denaturation Step.**
   - The complete denaturation of the DNA template at the start of the PCR reaction is of key importance. Incomplete denaturation of DNA results in the inefficient utilization of template in the first amplification cycle and in a poor yield of PCR product. The initial denaturation should be performed over an interval of 1-3min at 95℃ if the GC content is 50% or less. This interval should be extended up to 10min for GC-rich templates.
   - If the initial denaturation is no longer than 3min at 95℃, *Taq* DNA Polymerase can be added into the initial reaction mixture. If longer initial denaturation or a higher temperature is necessary, *Taq* DNA Polymerase should be added only after the initial denaturation, as the stability of the enzyme dramatically decreases at temperatures over 95℃.
2. **Denaturation Step.**
   - Usually 0.5-2min denaturation at 94-95℃ is sufficient, since the PCR product synthesized in the first amplification cycle is significantly shorter than the template DNA and is completely denatured under these conditions. If the amplified DNA has a very high GC content, denaturation time may be increased up to 3-4min. Alternatively, additives facilitating DNA denaturation - glycerol (up to 10-15 vol.%), DMSO (up to 10%) or formamide (up to 5%) - should be used. In the presence of such additives, the annealing temperature should be adjusted experimentally, since the melting temperature of the primer-template DNA duplex decreases significantly. If additives are used, the amount of *Taq* DNA Polymerase in the reaction mix should be increased, because DMSO and formamide, at the suggested concentrations, inhibit the enzyme approx. 50%. Alternatively, a common way to decrease the melting temperature of the PCR product is to substitute dGTP with 7-deaza-dGTP in the reaction mix.
3. **Primer Annealing Step.**

- Usually the optimal annealing temperature is 5℃ lower than the melting temperature of primer-template DNA duplex. Incubation for 0.5-2min is usually sufficient. However, if nonspecific PCR products are obtained in addition to the expected product, the annealing temperature should be optimized by increasing it stepwise by 1-2℃.

4. **Extending Step.**
    - Usually the extending step is performed at 70-75℃. The rate of DNA synthesis by *Taq* DNA Polymerase is highest at this temperature (2-4 kb/min), and a 1min extending time is sufficient for the synthesis of PCR fragments up to 2 kb. When larger DNA fragments are amplified, the extending time is usually increased by 1min for each 1000 bp.
5. **Number of Cycles.**
    - The number of PCR cycles depends on the amount of template DNA in the reaction mix and on the expected yield of the PCR product. For less than 10 copies of template DNA, 40 cycles should be performed. If the initial quantity of template DNA is higher, 25-35 cycles are usually sufficient.
6. **Final Extending Step.**
    - After the last cycle, the samples are usually incubated at 72℃ for 5-15min to fill-in the protruding ends of newly synthesized PCR products. Also, during this step, the terminal transferase activity of *Taq* DNA Polymerase adds extra A nucleotides to the 3'-ends of PCR products. Therefore, if PCR fragments are to be cloned into T/A vectors, this step can be extended to up to 30min.

4. DNA Extraction From Periferal Blood Leucocytes Procedure

(in: http://geneticslab.topcities.com/DNAextraction.htm)

1. Whole blood was collected in disodium EDTA containers and stored at -20ºC and a number of samples were also stored at -80ºC. To facilitate haemolysis of RBCs it is recommended to store a fresh sample for a few hours in a freezer as freezing destroys the red cells.
2. After thawing, 3mls of whole blood are transferred to a sterile conical centrifuge tube (15ml volume) to which 9mls of 1 x erythrocyte lysing buffer (0.155M NH_4Cl; 10mM $KHCO_3$; 0.1mM Na_2 EDTA; pH 7.4) must be added.
3. The solution is left for 10 minutes at RT with occasional mixing by inversion followed by centrifugation for 5 minutes at 4000 rpm.
4. After centrifugation the supernatant is discarded and a white pellet will be observed at the bottom of the tube. This pellet must be washed for at least 3 times by adding 3mls of erythrocyte lysing buffer and repeating steps 3 and 4. It is important to breakdown the pellet and rinse it well in erythrocyte lysing buffer in order to clean the white blood cells from remaining heme.
5. 1.5mls of SE buffer (75mM NaCl; 25mM Na_2 EDTA; pH 8.0) containing 100mg/ml of Proteinase K and 1% sodium dodecyl sulphate (SDS w/v), are added to the pellet. The tubes are then incubated at a temperature of 37 - 55ºC (optimal temp for Proteinase K activity) overnight in a water bath or incubator. During this step the white blood cells' membranes are denatured and DNA goes out in solution.
6. After the incubation, 1.5mls of SE buffer together with 750ml of 6M (saturated) NaCl are added to each tube, followed by the addition of 3.75mls chloroform.
7. The tubes are mixed vigorously (on a vortex) for about 20 sec with occasional mixing for at least 30min. Alternatively you can leave the tubes on a rotator for 1 hour.
8. The emulsion will then be centrifuged for 10 minutes at 2000 rpm with minimal breaking force. After centrifugation 2 phases are observed and care must be taken not to disturb the interphase. During this step DNA is extracted into the supernatant and proteins separated into the lower phase.
9. The upper aqueous phase (containing the DNA) is transferred into a clean and sterile conical centrifuge tube using a sterile Pasteur pipette, followed by the addition of an equal volume of isopropanol.
10. DNA will be precipitated by gentle swirling of the tube and is observed visually as a white thread like strand.

11. Using a sterile spatula or loop transfer the DNA strand into a sterile microcentrifuge tube containing 1ml of 75% ethanol.
12. The DNA is then washed by inversion to clean it from any remaining salts and the tube centrifuged at 11000g for 4 minutes. The supernatant is discarded taking care not to discard the pellet. Repeat this step once more.
13. After discarding the supernatant the pellet is dried from excess ethanol either by using a vacuum centrifuge or by leaving the tubes open and inverted in an oven at around 50 - 65ºC for an hour.
14. The dried pellet is resuspended in TE buffer (1M Tris-HCl; 0.5M EDTA; pH 8.0) and left overnight on a rotator.
15. DNA concentration is determined either by agarose gel electrophoresis or spectrophotometry and adjusted to the desired concentration by adding more TE buffer. It is important to note that before adjusting and reading DNA concentrations one must obtain a homogeneous sample of DNA which is not quite easy acquired since DNA is very viscous. To adjust to lower concentration one must used other quantitation methods such as Picogreen (Molecular Probes) using spectrofluorometry as spectrophotometry is not accurate and sensitive at very low concentrations.

5. DNA Sequencing Method

(in: http://cb.bio.uminho.pt/tikiwiki/tiki-index.php?page=Sequencia%C3%A7%C3%A3o)

A sequenciação de DNA consiste em determinar a sequência exacta de unidades estruturais que compõem o DNA -os nucleótidos. Os primeiros métodos de sequenciação foram publicados quase em simultâneo por Sanger *et al*., e Maxam-Gilbert? em 1977. Desde então, a tecnologia de sequenciação de DNA progrediu rapidamente e actualmente vários genomas de microorganismos estão completamente sequenciados e muitos outros encontram-se em fase de sequenciação, tal como o genoma humano.

A sequenciação de DNA baseia-se desde os primeiros métodos descritos numa electroforese de alta resolução que permita resolver fragmentos de DNA que diferam de um nucleótido. Muitos dos equipamentos de sequenciação utilizam ainda geis de electroforese, no entanto a electroforese capilar predomina nos modelos mais recentes.

Para sequenciar um fragmento de DNA é necessário pelos menos 2 requisitos: marcação selectiva das quatro bases (A-adenina, T-Timina?, G-guanina, C-citosina); fragmentação do DNA marcado de forma a poder ser estabelecida uma sequência nucleotídica com base na electroforese dos fragmentos marcados. Existem dois métodos de sequenciação de DNA: **1-** utiliza primers marcados selectivamente para cada base. Este processo implica que para cada base é necessário fazer uma reacção independente, sendo misturadas uma porção de cada reacção antes da injecção no sequenciador. **2-** utiliza didesoxirribonucleótidos marcados-terminadores- (desoxirribonucleótidos truncados no grupo 3´OH o que impede a adição de outros desoxirribonucleótidos), sendo utilizados em simultâneo na reacção de sequenciação. A metodologia de sequenciação aplicada no LAG é o processo de sequenciação cíclica que recorre ao uso de terminadores. A sequenciação cíclica foi desenvolvida com base no método de Sanger *et al*. e consiste num método simples no qual sucessivos ciclos de desnaturação, emparelhamento e extensão num termociclador resultam na amplificação linear dos produtos da extensão. A sequenciação cíclica utiliza protocolos robustos, fáceis de realizar, requer uma quantidade de DNA molde inferior aos métodos de extensão baseados em apenas uma temperatura, usa temperaturas superiores que reduzem estruturas secondárias permitindo uma extensão mais completa, utiliza o mesmo protocolo para DNA simples e de cadeia, e funciona bem para a sequenciação directa de produtos de PCR e para amostras de DNA difíceis tais como os BACs (bacterial artificial chromosomes).

Preparação das amostras para sequenciação

O sucesso de uma reaccção de sequenciação é largamente dependente da qualidade e pureza do DNA molde. Deste modo, todo o DNA submetido para sequenciação tem de estar em bom estado e purificado, ou seja livre de contaminantes tais como proteínas, RNA, DNA cromossómico, fenol, clorofórmio, etanol, sais, detergentes, PEG, EDTA, etc.,. O método de extracção que recomendámos é a lise alcalina seguida de purificação com kits específicos. Após a extracção recomenda-se a ressuspensão do DNA em água ultra pura estéril. No caso de plasmídeos as mini-prep são aceitaveis desde que de boa qualidade, em geral, Qiagen, Promega, etc funcionam bem. Se as mini-prep não forem utilizadas deverá proceder a uma extracção fenol-clorofórmio seguida de uma precipitação com etanol. Se este procedimento não for efectuado ou se forem utilizadas preps de má qualidade poderão obter-se más sequências devido à interferência de contaminantes ou poderá mesmo ocorrer a inibição total da reacção de sequenciação. O LAG é capaz de processar qualquer produto amplificado mediante uma reacção de PCR, contudo os produtos de PCR têm restos de nucleótidos e primers e enzimas que impedem a sequencição de forma directa e produzem ruído de fundo dando origem a uma sequênciação de má qualidade. Por este motivo é necessário incorporar um passo de purificação prévio à respectiva sequenciação. A purificação de produtos de PCR pode realizar-me mediante a utilização de agarose ou de kits específicos. A purificação por qualquer um dos métodos depende do número dos fragmentos obtidos durante o processo de amplificação e poderá ser realizada no LAG caso o deseje. A quantidade do DNA-amostra é também muito importante. O DNA deverá ser quantificado (por espectrofotometria ou através de gel, utilize este último no caso de a concentração de DNA ser muito pequena) e enviado na quantidade adequada de acordo com a seguinte tabela.

Tabela 1. Quantidade de DNA a fornecer por reacção. Atenção: A tabela refere-se a quantidade e não a concentração.

Amostra de DNA	Quantidade
Produto de PCR	
100-200 bp	1-3 ng
200-500 bp|3-10 ng	

500-1000 bp	5-20 ng
1000-2000 bp|10-40 ng	

>2000 bp	40-100 ng
Single-stranded|50-100 ng	

Double-stranded	200-500 ng
Cosmid, BAC|0.5-1.0 µg	

Bacterial genomic DNA	2-3 µg

Deverá providenciar 10 µl de amostra por reacção tendo em conta os valores da concentração de DNA publicados na tabela anterior. No caso de utilizar primers especificos estes devem ser submetidos na concentração mínima de 20 µM e nunca num volume inferior a 10 µl.
Os primers e as amostras contendo DNA são submetidas em tubos separados e no caso do DNA este deve ser ressuspendido em água ultra pura e nunca em tampão TE. Especifique correctamente a designação do vector quando se tratar de plasmídeos.
Podem ser utilizados para a reacção de sequenciação dois tipos de primers: primers específicos para o fragmento a sequenciar (podem ser os mesmos que se utilizaram durante a amplificaçãoe por PCR) que são fornecidos pelo requisitante, ou primers universais. O LAG disponibiliza sem quaisquers custos os seguintes primers universais:

SP6 5´-ATTTAGGTGACACTATAG-3
T3 5´-ATTAACCCTCACTAAAG-3´
T7 5´-TAATACGACTCACTATAGGG-3
M13 Forward 5´- TGTAAAACGACGGCCAGT-3
M13 Reverse 5´- CAGGAAACAGCTATGAC-3

Todos os restantes primers universais que o utente necessite utilizar terão de ser providenciados em conjunto com a amostra.

Folha de requisição de serviços de sequenciação

Poderá solicitar o serviço de sequenciação, mediante o preenchimento da seguinte folha de requisição. Preeencha cuidadosamente todos os campos especialmente os que se referem às informações da amostra que submete para sequenciação. Toda a informação é relevante para planificação adequada e para o sucesso da sequenciação. Respeite as quantidades de amostra e primers sugeridas e quantifique adequadamente o DNA de cada amostra. O resultado da sequenciação é enviado por E-mail num período inferior a três dias. No caso de não ser possível sequenciar a amostra serão imputados os custos mínimos de laboração (reacções sem sucesso) descritos na tabela do preçário e será emitido um parecer com as dificuldades encontradas e as possíveis causas.

(requisição em format PDF)
SEQ Order Form Biology
(requisição em formato Word)

Modalidade de sequenciação e Interpretação dos resultados

A modalidade de sequenciação praticada no LAG é a sequenciação de cadeia simples. Esta modalidade consiste na sequenciação de um fragmento de DNA mediante uma reacção única. O fragmento de DNA pode ser: DNA clonado em um vector plasmídico, DNA inserido num cosmídeo ou um produto de PCR. Esta modalidade de sequenciação é recomendada quando se tratar dos seguintes objectivos: para verificar a identidade de um fragmento de DNA clonado ou de um produto de PCR; para comprovar a sequência de inserções génicas; para efectuar a comparação entre sequências com a finalidade de identificar genes homólogos ou detectar mutações pontuais, identificar genes novos mediante uma sequênciação parcial e outras aplicações que não necessitem uma sequenciação de maior fiabilidade.

O LAG utiliza o programa de análise de sequências ABI PRISM DNA Sequencing Analysis para efectuar a análise da amostra e a edição da sequência de bases nucleotídicas. A qualidade de uma corrida de sequenciação é dependente, tal como foi dito anteriormente, da qualidade do DNA mas também da intensidade do sinal. Uma corrida normal gera normalmente cerca de 650 bases com uma exactidão de 98-99% para as primeiras 300 bases. À medida que a leitura

ultrapassa as 300 bases a exactidão diminuí, no entanto em amostras de alta qualidade a exactidão pode chegar às 400-450 bases. Em geral a sequência é legível apenas a partir de 50 bases de distância do primer.

O serviço de sequenciação em cadeia simples incluí:

- Comprovação da quantidade e qualidade do DNA da amostra recebida mediante a realização de uma electroforese em gel de agarose.
- Primers universais gratuitos (ver na secção acima os primers disponíveis)
- Sequênciação automática da amostra
- Comprovação da sequência obtida e resolução de indeterminações
- Envio por E-mail da sequência em formato txt, assim como o cromatograma electrónico correspondente à reacção (para poder visualizar o electrofluorograma necessita de um dos programas indicados abaixo)
- Envio dos cromatogramas originais impressos por correio normal sem custos adicionais.

Bibliography

Ali A et al. Requirement of protein phosphatase 5 in DNA-damage-induced ATM activation. Genes & Development 2004;18(3):249-254.

Altschul SF *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Res 1997; 25: 3389-3402.

Ammosova T et al. Dephosphorylation of CDK9 by protein phosphatase 2A and protein phosphatase-I in Tat-activated HIV-I transcription. Retrovirology 2005; 2-47.

Ammosova T et al. Nuclear targeting protein phosphatase-1 by HIV-1 Tat protein. Journal of Biological Chemistry 2005;280(43):36364-36371.

Andreeva AV et al. Widespread presence of “bacterial like” PPP phosphatases in eukaryotes. BMC Evolutionary Biology 2004; 4: 47-61.

Andreeva AV et al. PPP Family of Protein Ser/Thr Phosphatases: Two Distinct Branches?. Mol. Evol. Biol. 2001; 18(02):448-452.

Barford D. Protein phosphatases. Curr Opin Struct Biol 1995; 5:728-734.

Barford D. Molecular mechanisms of the protein serine/threonine phosphatases. Trends Biochem Sci 1996; 21:407-412.

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

Blatch GL et al. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays 1999; 21(11):932-939.

Champion HC. Targeting Protein Phosphatase 1 in Heart Failure. Circ. Res. 2005; 96; 708-710.

Chen MS et al. The tetratricopeptide repeat domain of protein phosphatase 5 mediates binding to glucocorticoid receptor heterocomplexes and acts as a dominant negative mutant. J Biological Chemistry 1996; 271(50):32315-32320.

Chen MX et al. 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.

Chinkers M. Protein phosphatase 5 in signal transduction. Trends in Endocrinology and Metabolism, 2001; 12(1):28-32.

Cohen PTW. The structure and regulation of protein phosphatases. Annu Rev Biochem 1989; 58:453-508

Cohen PTW. Novel protein serine/threonine phosphatases: Variety in spice of life. Trends in Biochemical Sciences 1997; 22(7):245-251.

Cooper DN, Sawatzki HK, Handbook of Human Molecular Evlution, 2008. Vol 1,2; Wiley.

Das AK et al. Crystal structure of the protein phosphatase 2C at 2 Å resolution. EMBO J. 1996; 15: 6798-6809.

Egloff MP et al. Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. EMBO J 1997; 16(8):1876-1887.

Eichhorn PJA et al. Protein phosphatase 2A regulatory subunits and cancer. Biochimica et Biophysica Acta – Reviews on Cancer 2009. 1795(1):1-15.

Faulkner NE et al. ProteiN Phosphatase 2A Enhances Activation of Human Immunodeficiency Virus Type 1 by Phorbol Myristate Acetate. Journal of Virology 2003; 2276-2281.

Felsenstein J Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 1985; 39:783-791.

Gonzalez-Kristteler DC. Identificação e caracterização de subunidades catalíticas e reguladoras de proteínas fosfatases de *Dictyostelim discoideum*. Tese de Doutoramento 2007. Universidade de São Paulo.

Hanikenne M et al. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. Nature 2008; 453:391-395.

Hastie CJ et al. A novel 50 kDa protein forms complexes with protein phosphatase 4 and is located at the centrosomal microtubule organizing centres. Biochem J 2000; 347:845-855.

Helps, NR et al. Protein phosphatase 4 is an essential enzyme required for organization of microtubules at cebntromeres in Drosophila embryos. J Cell Science 1998; 111:1331-1340.

Hemmings BA et al. Alpha and Beta forms of 65 kDa subunit of protein phosphatase 2A have a similar 39 amino acid repeating structure. Biochemistry 1990; 29:3166-3173.

Huang XZ et al. Molecular cloning, expression, and characterization of a novel human serine/threonine protein phosphatase 7, PP7, that is homologous to Drosophila retinal degeneration C gene product (rdgC). J Biological Chemistry 1998;273(3):1462-1468.

Huang SL et al. Inhibition of mammalian target of rapamycin activates apoptosis signal-regulating kinase 1 by suppressing protein phosphatase 5 activity. J Biological Chemistry 2004;279(45):36490-36496.

Hubbard MJ et al. On target with a new mechanism for the regulation of protein phosphorylation. Trends Biochem Sci 1993; 18:172-177.

Hunter T et al. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. Proc Natl Acad Sci USA 1980. 77:1311-1315.

Janssens V et al. A Novel Biological Function of the PP2A PR130/B”α1 Complex in Cell Migration and Adhesion. EUROPHOSPHATASES 2009, EMBO second conference series; p.33.

Janssens V et al. Protein phosphatase 2A: A highly regulated family of serine/threonine phosphatases implicated in cell growth and signaling. Biochem J 2001; 353(3):417-439.

Kamenski T et al. Structure and mechanism of RNA polymerase IICTD phosphatases. Mollecular Cell 2004; 15(3):399-407.

Kloeker S et al. Purification and identification of a novel subunit of protein serine/threonine phosphatase 4. J Biological Chemistry 1999; 274(9):5339-5347.

Kloeker S et al. Carboxymethylation of nuclear protein serine/threonine phosphatase X. Biochem J 1997; 327:481-486.

Kennely PJ. Archaeal protein kinases and protein phosphatases insights from genomics and biochemistry. Biochem. J 2003; 370:373-389.

Kennely PJ. Protein phosphatases – a phylogenetic perspective. Chem Rev 2001; 101:2291-2312.

Kissinger CR et al. Cristal structures of human calcineurine and the human FKBP12-FK506-calcineurin complex. Nature 1995; 378(6557): 641-644.

Kutuzov MA et al. Protein serine/threonine phosphatases PPEF interact with calmoduline. Biochemical and Biophysical Communications 2002. 293(3):1047-1052.

Lesk AM. INTRODUCTION TO Bioinformatics, 2th Edition. Oxford 2005.

Liu R et al. Protein phosphatase 2A a key player in Alzheimer’s disease. Front. Med. China 2009; 3 (1): 8-12.

Liu J et al. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. Cell 1991; 66:807-815.

MacKeigan JP et al. Sensitizing RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. Nature Cell Biology 2005; 7(6):591-619.

Meinhart A et al. A structural perspective of CDT function. Genes & Development 2005, 19(12):1401-1415.

Merchant SS et al. Chlamydomonas genome reveals the key animal and plant functions. Science 2007; 318:245-250.

Miller MP et al. Understanding human disease mutations through the use of interspecific genetic variation. Human Molecular Genetics 2001; 10:2319-2328.

Moorhead GBG et al. Evolution of protein phosphatases in plants and animals. Biochem. J. 2009; 417: 401-409.

Morita K et al. Negative feedback regulation of ASK1 by protein phosphatase 5 (PP5) in response to oxidative stress. EMBO J 2001;20(21):6028-6036.

Morrison R et al. Química Orgânica. 13ª Ed. Fundação Calouste Gulbenkian 1996. 1370-1372.

Mourtada MM et al. Functional expression cloning reveals pró-apoptotic role for protein phosphatase 4. Cell Death and Differentiation 2003; 10(9):1016-1024.

Nei M & Kumar S *Molecular Evolution and Phylogenetics*. Oxford University Press 2000, New York.

Olsen JV et al. Global, in vivo and site specific phosphorilation dynamics in signaling networks. Cell 2006,127: 635-648.

Panchal RG et al, Immune Protection Against Multiple Lethal Phatogens by Modulation of a Protein Tyrosine Phosphatase. EUROPHOSPHATASES 2009, EMBO second conference series; p.40.

Parameswara VK et al. Have We Overlooked the Importance of Serine/ThreonineProtein Phosphatases In Pancreatic Beta-Cells? Role Played by Protein Phosphatase 2A in Insuline Secretion. JOP J Pancreas (Online) 2005; 6(4): 303-315.

Reeder JE et al. Inducible expression of catalytically active type I serine/threonine protein phosphatase in human carcinoma cell line. Cancer cell international 2003; 3-12.

Rusnak F et al. Calcineurine: form and function. Physiol Rev 2000; 80:1483-1521.

Rzhetsky A & Nei M A simple method for estimating and testing minimum evolution trees. *Molecular Biology and Evolution 1992;* 9:945-967.

Saitou N & Nei M The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution 1987;* 4:406-425.

Shenolikar S et al. Proteins phosphatases: recent progress. Adv Second Messenger Phosphoprotein Res 1991, 23:1-121.

Shenolikar S. Protein serine/threonine phosphatases-new avenues for cells regulation. Annu Rev Cell Biol 1994; 10:55-86.

Silva EFC. AS FOSFATASES E A LINGUAGEM INTRACELULAR. Fundação Calouste Gulbenkian 1998; 43-56.

Silverthorn, Fisiologia Humana, 2ª Edição, 2003. Manole. p.32

Stefansson B et al. Protein phosphatase 6 subunit with conserved sit4-associated protein domain targets I kappa B epsilon. J Biological Chemistry 2006. 281(32):22624-22634.

Strachan T & Read AP, HUMAN MOLECULAR GENETICS 3, 2004,Garland Science.

Sugiyama M et al. Membrane-associated serine/threonine protein phosphatase in endometrial cancer. Am J Obstet Gynecol 2003; 189: 1666-1669.

Sunyaev S et al. SNP frequencies in human genes: an excess of rare alleles and differing models of selection. Trends in genetics 2000; 16(8):335-337.

Swingle MR et al. Structural basis for the catalytic activity of human phosphatase-5. J Biological Chemistry 2004; 279(32):33992-33999.

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolu*

Tamura K, Nei, M & Kumar S Prospects for inferring very large phylogenies by using the neighbor-joining method. *PNAS 2004;* 101:11030-11035.

Terrak M et al. Structure basis of protein phosphatase 1 regulation. Nature 2004; 429(6993):780-784.

Wechsler T et al. DNA-PKcs function regulated specifically by protein phosphatase 5. Proceeding of the National Academy of Sciences of the United States of America

2004;101(5):1247-1252.

Wera S et al. Serine/threonine protein phosphatases. Biochem. J 1995; 311:17-29.

Xu, YH et al. Structure of the protein phosphatase 2A holoenzyme. Cell 2006; 127(6):1239-1251.

Zhang ZY. Protein tyrosine phosphatase: prospects for therapeutics. Chem. Biol. 2001; 5:416-423.

Zhang ZY et al. Modulation of protein kinase signaling by protein phosphatases and inhibitors. Pharm.Therap . 2002; 1:1-11.

Zhou GF et al. Ser/Thr protein phosphatase 5 inactivates hypoxia-induced activation of an apoptosis signal-regulating kinase 1/MKK-4/JNK signaling cascade. J Biological Chemistry 2004;279(45):46595-46605.

Zuo Z et al. Ser/Thr protein phosphatase 5 (PP5) is a negative regulator of glucocorticoid receptor-mediated growth arrest. Biochemistry 1999; 38(28):8849-8857.