ISAS- INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Pseudomonas aeruginosa PA3859: From Structure to Function

Thesis submitted for the degree of
"Doctor Philosophiae"

CANDIDATE
Alessandro Pesaresi

SUPERVISOR
Dr. Doriano Lamba

Trieste

December 2005

SISSA – Via Beirut 2-4 – TRIESTE – ITALY

Tabel of contents

Chapter 1

Introduction

1.1 *Pseudomonas aeruginosa*

1.1.1 *Pseudomonas aeruginosa* as an opportunistic pathogen

Pseudomonas aeruginosa (PA) is a Gram-negative aerobic bacterium that is commonly found in soils and waters. It occurs regularly on the surfaces of plants and occasionally on the skin of animals and is regarded as the epitome of an human opportunistic pathogen: this bacterium almost never infects uncompromised tissues, yet there is hardly any tissue that it cannot infect if the tissue defenses are compromised in some respect. It causes infections of the urinary tract, respiratory system, dermatitis, soft tissue, bacteremia, bone and joint, gastrointestinal and a variety of systemic infections, particularly in patients with severe burns and in cancer and AIDS patients who are immunosuppressed. PA infection is a serious problem in patients hospitalized with cancer, cystic fibrosis, and burns. The case fatality rate in these patients is 50 %.
PA is primarily a nosocomial pathogen: the overall incidence of its infections in hospitals averages to about 0.4 %, and the bacterium is the fourth most commonly-isolated nosocomial pathogen accounting for 10 percent of all hospital-acquired infections. It demands very simple nutritional requirements. In the laboratory, the simplest medium for its growth consists of acetate as carbon source and ammonium sulfate as nitrogen source. Moreover it is notorious for its resistance to antibiotics and is, therefore, a particularly dangerous and dreaded pathogen. The bacterium is naturally resistant to many antibiotics due to permeability barrier built up by its outer membrane.

Figure 1.1. *Pseudomonas aeruginosa* scanning electron micrograph: the cell is a rod measuring 0.5 to 0.8 µm by 1.5 to 3.0 µm. Almost all strains are motile by means of a single polar flagellunm.

Also, its tendency to colonize surfaces as a biofilm form makes the cells impervious to therapeutic antibiotics doses. Since its natural habitat is the soil, living in association with the bacilli, actinomycetes and molds, it has developed resistance to a variety of their naturally-occuring antibiotics.
Only a few antibiotics are effective against PA, including fluoroquinolones, gentamicin and imipenem, and even these antibiotics are not effective against all strains. The futility of treating PA infections with antibiotics is most dramatically represented in cystic fibrosis patients that typically become infected with strains so resistant that it cannot be treated.

For an opportunistic pathogen such as PA, the disease process begins with some alteration or circumvention of normal host defenses. The pathogenesis of PA infections is multifactorial, as suggested by the number and wide array of virulence determinants possessed by the bacterium. Multiple and diverse determinants of virulence are expected in the wide range of diseases caused, which include septicemia, urinary tract infections, pneumonia, chronic lung infections, endocarditis, dermatitis, and osteochondritis. A brief description of these infective diseases is reported in the following list.

Endocarditis. PA infects prosthetic heart valves. The organism establishes itself on the endocardium by direct invasion from the blood stream.
Bacteremia and Septicemia. PA causes bacteremia primarily in immunocompromised patients. Predisposing conditions include hematologic malignancies, immunodeficiency

relating to AIDS, neutropenia, diabetes mellitus, and severe burns. Most PA bacteremia is acquired in hospitals and nursing homes.

Central Nervous System infections. PA causes meningitis and brain abscesses. The organism invades the central nervous system from a contiguous structure such as the inner ear or paranasal sinus, or is inoculated directly by means of head trauma, surgery or invasive diagnostic procedures, or spreads from a distant site of infection such as the urinary tract.

Ear infections including external otitis. PA is the predominant bacterial pathogen in some cases of external otitis. The bacterium is infrequently found in the normal ear, but often inhabits the external auditory canal in association with injury, maceration, inflammation, or simply wet and humid conditions.

Eye infections. PA can cause devastating infections in the human eye. It is one of the most common causes of bacterial keratitis, and has been isolated as the etiologic agent of neonatal ophthalmia. It can colonize the ocular epithelium by means of a fimbrial attachment to sialic acid receptors. If the defenses of the environment are compromised in any way the bacterium can proliferate rapidly and, through the production of lytic enzymes cause a rapidly destructive infection that can lead to loss of the entire eye.

Bone and joint infections. PA infections of bones and joints result from direct inoculation of the bacteria or the hematogenous spread of the bacteria from other primary sites of infection. Blood-borne infections are most often seen in conjunction with urinary tract or pelvic infections. PA has a particular tropism for fibrocartilagenous joints of the axial skeleton. It causes chronic contiguous osteomyelitis, usually resulting from direct inoculation of bone, and is the most common pathogen implicated in osteochondritis after puncture wounds of the foot.

Urinary tract infections. Urinary tract infections (UTI) caused by PA are usually hospital-acquired and related to urinary tract catheterization, instrumentation or surgery. It is the third leading cause of hospital-acquired UTIs, accounting for about 12 % of all infections of this type. The bacterium appears to be among the most adherent of common urinary pathogens to the bladder uroepithelium. As in the case of *E. coli* urinary tract infection can occur via an ascending or descending route. In addition, PA can invade the bloodstream from the urinary tract, and this is the source of nearly 40 percent of PA bacteremias.

Gastrointestinal infections. PA can produce disease in any part of the gastrointestinal tract from the oropharynx to the rectum. As in other forms of PA disease, those involving the gastrointestinal tract occur primarily in immunocompromised individuals. The organism has been implicated in perirectal infections, pediatric diarrhea, typical gastroenteritis, and necrotizing enterocolitis. The gastrointestinal tract is also an important portal of entry in PA septicemia.

Skin and soft tissue infections, including wound infections, pyoderma and dermatitis. PA can cause a variety of skin infections, both localized and diffuse. The common predisposing factors are breakdown of the integument which may result from

burns, trauma or dermatitis; high moisture conditions such as those found in the ear of swimmers and the toe webs of athletes and combat troops, in the perineal region and under diapers of infants, and on the skin of whirlpool and hot tub users. Individuals with AIDS are easily infected. PA has also been implicated in folliculitis and unmanageable forms of acne vulgaris.

Respiratory infections. Respiratory infections caused by PA occur almost exclusively in individuals with a compromised lower respiratory tract or a compromised systemic defense mechanism. Primary pneumonia occurs in patients with chronic lung disease and congestive heart failure. Bacteremic pneumonia commonly occurs in cancer patients undergoing chemotherapy. Lower respiratory tract colonization of cystic fibrosis patients by mucoid strains of PA is common and difficult, if not impossible, to treat.

1.1.2 *Pseudomonas aeruginosa* modulation of epithelial function

The lung is exposed to a constant barrage of inhaled harmful agents and microorganisms. It has therefore evolved both innate and acquired defense mechanisms to combat infections. The innate host defense mechanisms include mucociliary clearance, surfactant proteins, antimicrobial peptides and complement, all of which have important roles in immune modulation. The airways are also protected by alveolar macrophages, neutrophils, lymphocytes and circulating antibodies. A loss of individual components frequently results in lung infection. Infection of human airways by PA leads to varied pathological responses and lung injury. One extreme is the acute nosocomial necrotizing pneumonia that is associated with a high incidence of mortality despite antibiotic therapy, due in part to epithelial destruction and bacterial invasion of the pulmonary vasculature with resultant bacteremia and its sequellae. At the opposite extreme is persistent airway infection in individuals with cystic fibrosis (CF), where organisms remain confined to the airway as biofilms, triggering a chronic neutrophilic inflammatory response.

PA infection of normal and CF airways

Numerous effects of PA components/products on various aspects of epithelial cell function, including direct cytotoxicity, activation of cellular signaling pathways and proinflammatory responses, have been examined extensively *in vitro* and *in vivo* [1–14]. DNA microarray technology has revealed that epithelial gene expression patterns in response to different PA-derived products and intact organisms vary considerably [15–18]. Abnormalities of the cystic fibrosis transmembrane regulator (CFTR), which result

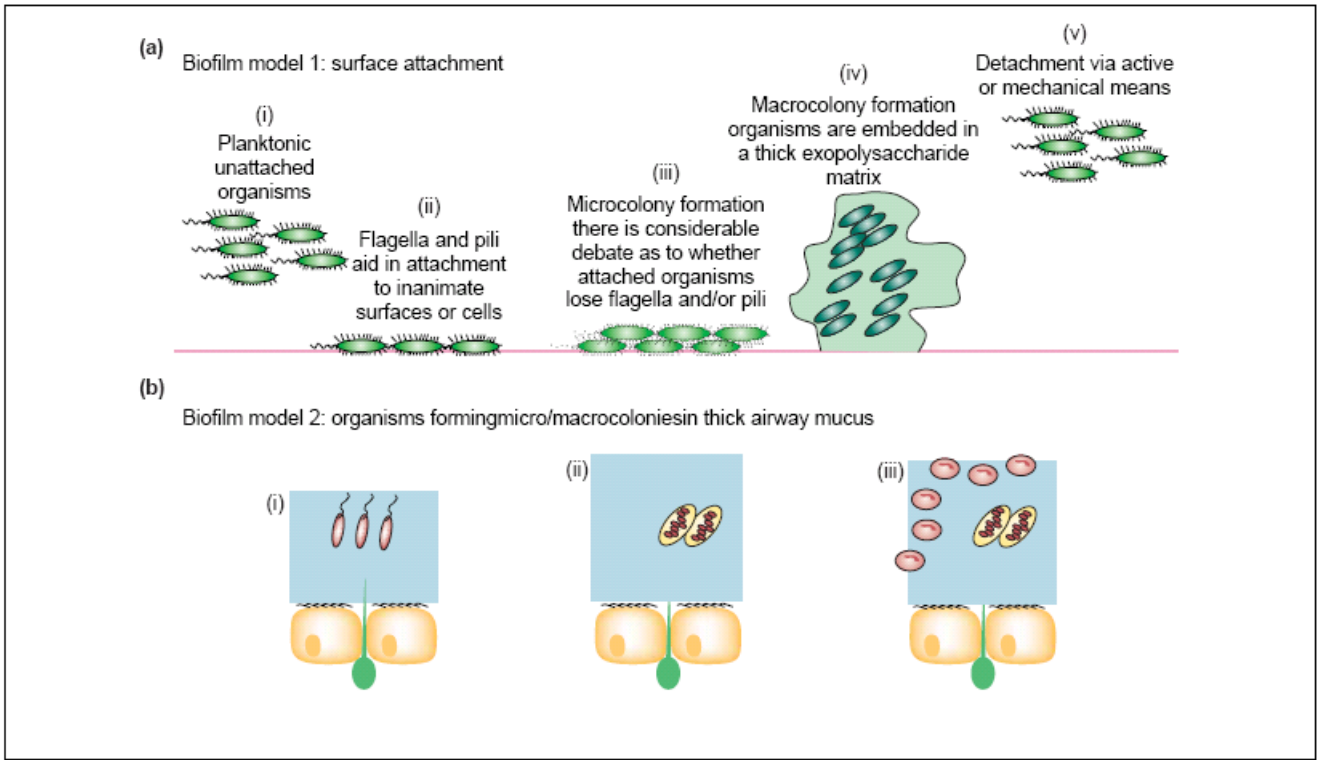


Figure 1.2. Models of biofilm formation. (**a**) In biofilm model 1, planktonic *Pseudomonas aeruginosa* (i) attach to a surface via flagella and/or type IV pili (ii). Structures known as 'microcolonies' form after significant cell division and biosynthesis of a multi-component exopolysaccharide matrix (iii). These microcolonies mature into 'macrocolonies' as the biofilm thickens (iv). Signaling molecules, nutrient and oxygen gradients or mechanical perturbation trigger release of bound organisms from the biofilm, often referred to as biofilm detachment or dispersion (v).
(**b**) In biofilm model 2, which occurs in the cystic fibrosis (CF) airways, *P. aeruginosa* are embedded in a layer of thickened mucus and rarely come into indirect contact with epithelial surface (i). Bacteria lose their surface appendages, mutate to a mucoidy phenotype with a heavy coat of alginate, and form biofilm macrocolonies (ii). Thick mucus layer and biofilm both protect *P. aeruginosa* from clearance by neutrophils and macrophages (iii), enabling them to establish chronic infections. (Taken from Hassett D.J., *et al.* (2003) *Expert Opinion Biol. Ther*. 3, 1201-1207).

in CF, enable enhanced Na^+, Cl^- and water reabsorption. CF goblet cells hypersecrete mucus to the apical surface. CF epithelial cilia are matted down by thick mucus and either do not beat or beat erratically. This mucus becomes a haven for PA [1]. In contrast to single cell (planktonic) growth *in vitro*, the mode of PA growth in CF airways is termed 'biofilm' [19, 20]. Biofilm is a highly ordered surface-attached mode of growth whose formation involves attachment, cell division, differentiation and detachment. Biofilm formation requires flagella, pili and other gene products (Fig. 1.2). PA in biofilms are associated with high resistance to antibiotics and phagocytic killing [20]. Two models of biofilms are known. The first involves a series of steps that lead to bacterial attachment to surfaces (Fig. 1.2, **a**) and the other is characterized by bacterial 'rafts' (Fig. 1.2, **b**). Model 1 could represent the initial phase of CF airway colonization

Virulence determinants		Impact on airway epithelial cells
LPS		Ligand for TLR4, TLR2. CFTR?
Alginate		Stimulates mucin and lysozyme secretion
Flagella and pili		Adhesin
		TLR5 and TLR2 signaling
		NF-κB induction
		IL-8 induction
		Induce matrilysin transcription
QS molecules	PAI-1	Induces IL-8 release
T3SS	ExoU	Cytotoxic phospholipase
		Activates proapoptotic pathway
		Changes phospholipids profiles
	ExoS and ExoT	Modulates small GTPases
		Invasion of basolateral epithelium layer
	ExoY	Increases intracellular cAMP
		Causes cell rounding and detachment
		Disrupts intercellular tight junctions
Phenazines	PCN	Inhibits respiration (ATP depletion)
		Causes ciliary dyskinesis
		Inhibits epidermal cell growth
		Inhibits α1-protease inhibitor
		Disrupts Ca^{2+} homeostasis
		Inhibits catalase
		Inhibits V-ATPase
		Disrupts glutathione flux
		Modulates the levels of IL-8 and RANTES
	I-HP	Ciliary dysfunction
	PCA	Increases IL-8 release
		Increases ICAM-1 expression
		Decreases RANTES and MCP-1 release
Proteases	LasA	Syndecan shedding
	LasB, alkaline protease and Protease IV	Degrades innate immunity proteins, chemokines and cytokines
		Increase IL-8 release
		Disrupt intercellular tight junctions
Phospholipases	PLCs, PldA	Digest phosphatidylcholine, lung infections
	PlcB	Phospholipids chemotaxis
Exotoxin A		Inhibits protein synthesis
		Causes apoptotic and non-apoptotic cell death
		Disrupts intercellular tight junctions
Rhamnolipids		Causes ciliostasis
		Releases glycoconjugates from trachea and bronchi
		Disrupt intercellular junctions
Nitrite reductase		Induces IL-8 release

Table 1.1. Virulence determinants of *Pseudomonas aeruginosa* and their impact on lung epithelia. (taken from Lau W. G., *et al*. (2005) *TREND Microbiol*. 8, 389-397.)

by environmental isolates of PA, or acute pneumonias in non-CF airways. As the chronic stage of infection ensues, biofilm model 2 predominates. It is associated with over production of alginate, loss of flagella, mutations in antigenic B-band O-antigen genes, and reduced expression of extracellular virulence determinants [1]. Thus, PA is subject to selection towards reduced cytotoxicity and attenuated virulence during chronic CF lung infection. Microscopic examination of airways removed from lung

transplantation patients strongly argues that PA rarely localizes to the airway epithelial surface in end-stage CF [7]. Rather, the bacteria are embedded in thick mucus (biofilm model 2). Levels of oxygen within the mucus are low, creating 'steep hypoxic gradients' [7]. However, sufficient levels of the alternative electron acceptors NO_3^- and NO_2^- are present in CF sputa, suggesting that some bacteria might undergo anaerobic biofilm formation and respiration [21].

PA products and their effects on the epithelium

Numerous PA components have been reported to modulate lung epithelial functions (Tab. 1.1). Key virulence determinants and areas of controversy are highlighted in the subsequent section.

Membrane-bound virulence determinants Lipopolysaccharide.

The lipid A component of bacterial lipopolysaccharide (LPS) is a potent modulator of epithelial cell functions via its interaction with the toll-like receptor 4 (TLR4) signaling complex [22] (Fig. 1.2, **a**). LPS is not important in PA-mediated modulation of airway cell functions when compared with other Gram-negative pathogens [23, 24]. However, exceptions have been described [13, 25].

Flagella and pili.

Flagella and pili are appendages required for PA motility and lung infection [26, 27]. They enable attachment of PA to cells by binding to respiratory mucin [12] and the glycolipid, asialoGM1 [11], inducing release of interleukin (IL)-8 through a Src-Ras-ERK1/ 2-NF-kB pathway [10]. Interestingly, mucoid isolates of PA from CF patients lack flagella, which might enhance their survival by decreasing neutrophil recruitment to the airway [15].

Alginate.

Alginate is not a significant component of biofilms in vitro in PA strains that possess an intact mucA allele [28]. Although purified alginate stimulates mucin and lysozyme secretion in airway epithelia [29], the primary pathophysiologic role of alginate might be to dampen the local immune response towards PA [15].

Secreted virulence determinants

Pathogenic interactions between PA and hosts are often guided by exoproducts secreted by the bacteria that interact with specific host targets (Tab. 1.1).

Quorum sensing molecules: role in pathogenesis and biofilm formation.

PA communicates intercellularly via small, diffusible acylated homoserine lactones: Pseudomonas autoinducers (PAIs). Local PAI concentrations increase at high cell densities, enter bacteria and bind to transcriptional activators to activate or repress the

expression of genes. The precise interaction among these quorum sensing systems remains to be defined but they clearly control production of numerous virulence factors, including exoproteases, exotoxin A, rhamnolipids, phenazines and hydrogen cyanide [30]. Interestingly, PA in CF airway and in vitro biofilms produce higher levels and identical ratios of PAI than planktonic bacteria [19].

Type III secretion system (T3SS).
PA injects four type III secretion system (T3SS) effector proteins, ExoU, ExoS, ExoT and ExoY into host cells [31]. ExoU, a lysophospholipase, which inhibits ERK1/2-p38 survival pathways and activates pro-apoptotic pathways through JNK1/2 [32], probably has the greatest role in virulence [33]. N-terminal domains of ExoS and ExoT encode a GTPase activating protein activity (GAP) and C-terminal domains have ADP ribosyltransferase (ADPRT) activity [31]. ExoY, an adenylyl cyclase, increases lung epithelial intracellular cAMP 800-fold, causes cell rounding/detachment, might inhibit phagocytosis of PA, and disrupts pulmonary endothelial barrier function [44].

Phenazines.
Pseudomonas species produce phenazines as secondary metabolites that function in microbial competitiveness and virulence [38]. The most studied phenazines are: pyocyanin, 1-hydroxyphenazine and phenazine-1-carboxylic acid. Bioactivities of these compounds are linked to their ability to affect the redox-cycle, which leads to generation of reactive oxygen species and host cell damage [38]. In mice pyocyanin-deficient PA mutants are attenuated in both acute and chronic lung infections in mice, providing strong evidence that pyocyanin contributes to PA virulence [38].

Proteases.
PA secretes elastases (LasA and LasB), alkaline protease and protease IV. These exoproteases degrade host defense and immunoregulatory proteins and damage epithelia [39–41]. Elastase activates the mitogen-activated protein kinase (MAPK) pathway, increasing IL-8 expression and augments respiratory epithelium permeability by attacking tight junctions [5].

Phospholipases.
PA produces several phospholipases: hemolytic PlcH and non-hemolytic PlcN, PlcB [42] and PldA [43]. PlcB is responsible for PA chemotaxis toward phospholipid gradients [42] and PlcH exhibits sphingomyelin synthase activity [44]. The substrate for these enzymes, phosphatidylcholine is abundant in lung surfactant, therefore, phospholipases probably have a role in virulence. Anti-PLC antibody is readily detectable in CF patients, and the isolation of PA strains capable of secreting PLCs is

usually associated with a poor clinical status. A tatC mutant, that fails to secrete multiple exoproducts including PlcH and PlcN [45], and the pldA mutant [43], are both attenuated in mouse chronic lung infection.

Exotoxin-A (ETA).
Most clinical PA isolates secrete ETA, a potent inhibitor of protein synthesis via ADP-ribosylation of elongation factor 2. ETA expression is regulated by QS and by the availability of iron, which is chelated by two siderophores pyochelin and pyoverdin [46]. ETA levels in bronchial secretions correlate with exacerbation periods in CF. ETA binds to 2 macroglobulin receptors located exclusively on the basolateral side of polarized epithelia [47]. It increases the permeability of the airway epithelia by inhibiting the repletion of tight junction proteins [5], which partially contributes to apoptotic cell death [48]. ETA also kills airway epithelia non-apoptotically through mitochondrial dysfunction, superoxide production and DNA degradation [49].

1.2 Bacterial phospholipase/lysophospholipase and pathogenesis

Phospholipases are a heterologous group of proteins produced by bacteria and their eukaryotic hosts. In eukaryotic cells, phospholipases perform the ordinary task of phospholipid turnover in addition to their role in signal transduction; similar enzymatic activities involved in phospholipid turnover for membrane maintenance also release precursors of lipid second messengers. Some mammalian cells also produce secreted phospholipases which are present in extracellular fluids including serum and synovial fluid; as part of the inflammatory response, secreted phospholipase A2 (PLA2, see below) are produced and secreted in response to cytokines during inflammatory events [50]. The secreted PLA2 have also been implicated in the progression of damaging inflammatory responses, septic shock, acute lung injury, and other inflammatory disorders [50]. Bacteria produce membrane-associated phospholipases that are thought to function primarily in membrane maintenance, but their function has not been definitively established. *Escherichia coli* mutants lacking a phospholipase have been isolated, suggesting these enzymes are not essential for viability [51]. Various bacterial genera also produce secreted phospholipases and lipases. In general, secreted phospholipases are thought to function in phosphate acquisition, carbon source acquisition, and in some cases as virulence factors for pathogenic species.
The amphipathic nature of phospholipids creates obstacles for the enzymes as the substates are assembled into bilayers or even micelles, and are not present in significant amounts as single soluble substrates. It should be noted that the amphipathic

phospholipid substrates are defined by their polar head group, of which the most common in mammalian cells are phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylinositol (PI), and phosphatidylethanolamine (PE) (Fig. 1.3). These designations actually define a group of molecules with differing degrees of saturation and length of the two fatty acyl groups; only a limited range of degrees of saturation and potential acyl chain lengths are found in biological systems. Both the acyl chains and the phospho-head groups influence the substrate specificity of phospholipases, some enzymes being much more restricted than others. Because the phospholipid composition varies with the source of the membrane, i.e., bacterial versus eukaryotes, the substrate specificity of a particular phospholipase can suggest or eliminate potential target membranes. In contrast to mammalian cells, *E. coli* membranes are primarily composed of PE with some phosphatidylglycerol, cardiolipin (double phospholipid joined at the phosphate by glycerol) and trace amounts of PS. Furthermore, only a subset of the products of phospholipase activity are direct precursors for second messengers in mammalian cells, i.e., the fatty acid arachidonic acid is the precursor for eicosanoid biosynthesis. Therefore, given some knowledge of the substrate specificity of a particular phospholipase, some inference regarding its function can be made.
Two major divisions of phospholipase activities can be defined by the site of cleavage, whether the cleavage is in the hydrophobic diacylglycerol moiety (PLA) or in the polar head group of the amphipathic phospholipid (PLC and PLD; Fig. 1.3). PLAs hydrolyze a fatty acid from the glycerol backbone (leaving a lysophospholipid) in a reaction similar to lipases hydrolyzing fatty acids from triacylglycerol lipids. In fact, a few enzymes have significant PLA activity as well as triacylglycerol lipase activity; therefore these enzymes hydrolyze either hydrophobic or amphipathic substrates with reasonable efficiency [52]. PLAs can be further defined by their positional specificity, preference for the acyl group attached to position 1 or 2 of the glycerol backbone, PLA1 and PLA2, respectively; PLBs have both PLA1 and PLA2 activity, i.e., little or no positional specificity. Lysophospholipases release free fatty acid from lysophospholids. The other types of phospholipases cleave on either side of the head group phosphate. Cleavage by PLC releases the phospho-head group (e.g., choline phosphate, inositol phosphate, and inositol triphosphate) and diacylglycerol, whereas PLD cleaves on the other side of the phosphate producing phosphatidic acid and the head group (e.g., choline and inositol). Phospholipases of all types have various substrate specificities with respect to the polar head group and also with respect to the length and saturation of the fatty acyl groups. Some phospholipases are very specific, hydrolyzing a particular phospholipids with greater efficiency.

Nevertheless, hydrolysis by any phospholipase type releases products that destabilize membranes; if sufficiently active, bacterial phospholipases may initiate further degradation of the phospholipids by other cellular enzymes including phospholipases.

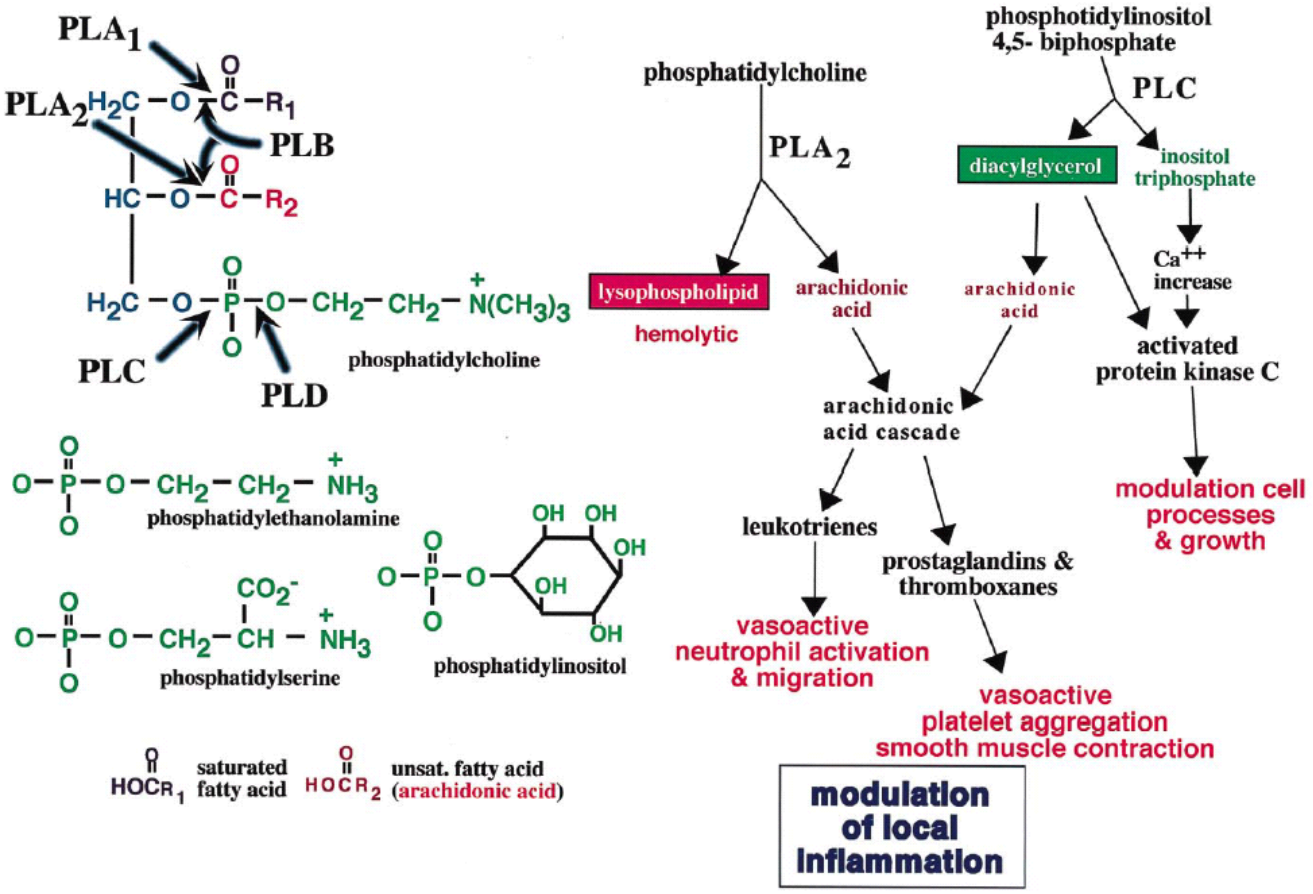


Figure 1.3. Structures of common mammalian phospholipids and the signaling pathways induced by the second messengers released by phospholipases. Portions of phospholipid molecules are color coded: polar head groups in green, glycerol backbone in blue and fatty acids in red and purple. Diacylglycerol encompasses all of the phospholipid except the green polar head group, and likewise the lysophospholipid encompasses everything but the fatty acid shown in red. Arrows do not necessarily indicate a direct induction. (Taken from Schmiel D. H., *et al*. (1999) *Microb. Infect.* 1, 1103-1112).

Phospholipases are considered virulence factors for bacterial species which cause disparate disease syndromes, from infections causing massive tissue destruction, such as gas gangrene and the skin and lung infections of *Pseudomonas aeruginosa* to food-borne listeriosis (Tab. 1.2) [53-55]. In the disease process phospholipases have been implicated in a number of roles, which are considered below (Fig. 1.4). Even though the importance of some bacterial phospholipases in pathogenesis is certain and the enzyme has been characterized, it is clear that the exact mechanism of the phospholipase action *in vivo* has not been definitively determined. Historically, the first indication that bacterial phospholipases were virulence factors was the realization that some bacterial

.

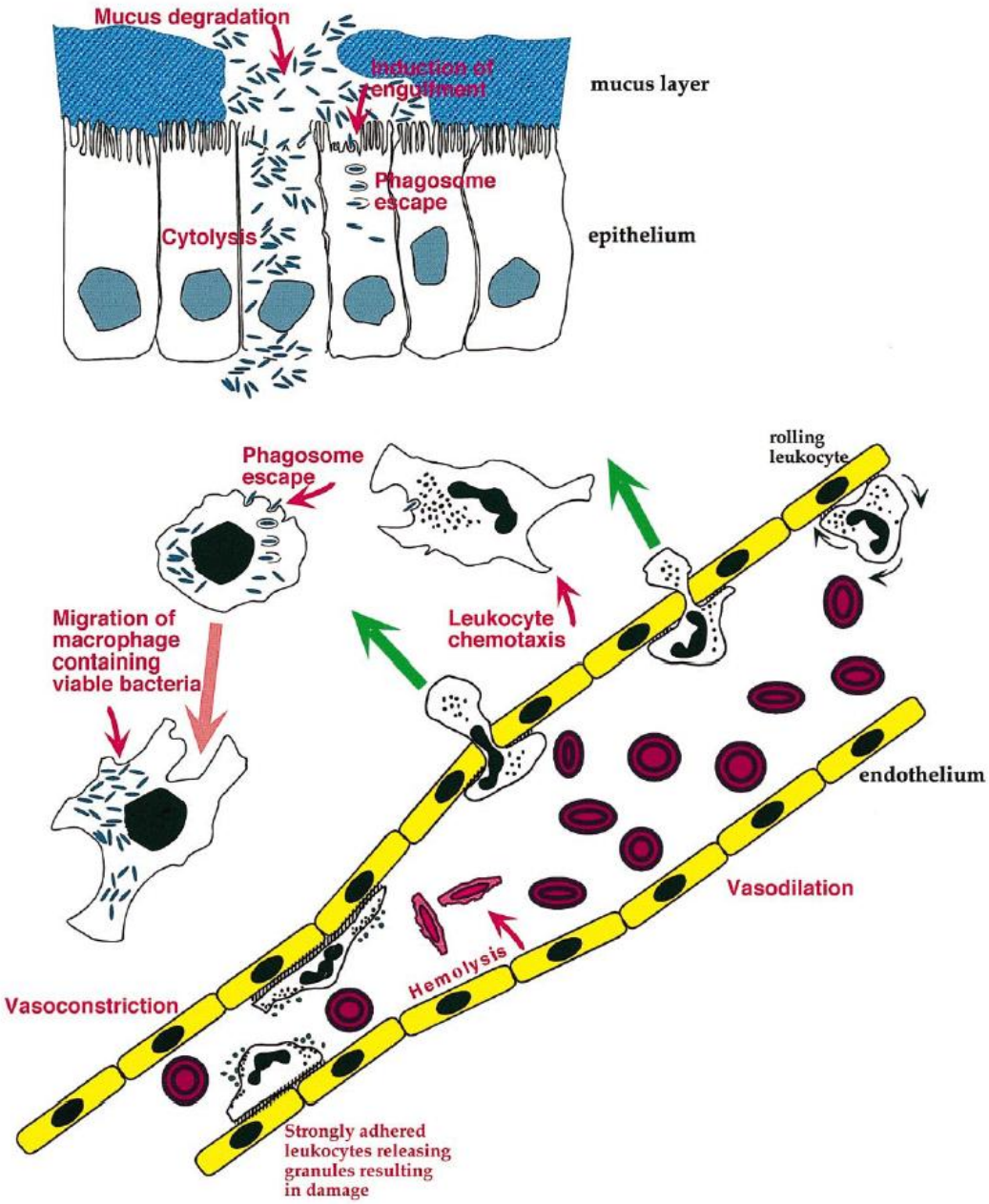


Figure 1.4. Drawing depicting the roles of bacterial phospholipases during infection. The imaginary bacterial phospholipase is shown performing most of the well-established as well as proposed functions ascribed to all the bacterial phospholipases discussed. From the upper left corner, the bacteria depicted as green rods secrete a phospholipase which aids degradation and penetration of the mucus layer (*H. pylori* and *P. aeruginosa*). Once the bacteria reach the epithelial surface phospholipase action triggers engulfment by an epithelial cell, a nonprofessional phagocyte (*Rickettsia*). Phospholipase activity then promotes escape from the vacuole or phagosome in the right-hand epithelial cell and the macrophage below (*L. monocytogenes*) and promotes cytolysis in the left-hand epithelial cell. Cytolysis can be triggered from within or without (*C. perfringens* and many others). Bacteria which have escaped the macrophage phagosome are being carried from the site of

infection by the migrating macrophage (*L. monocytogenes*). Effects associated with the blood vessel, bottom right, are depicted in the absence of bacteria because they are mostly indirect effects attributed to eicosanoid or cytokine induction (excepting hemolysis). The vessel is drawn partly dilated and partly constricted because different eicosanoids have opposite vasoactive properties. When blood leukocytes are induced to migrate into tissues, the vessel generally dilates, reducing blood velocity, and several different leukocyte adhesion receptors are sequentially induced in the endothelial cells.Weaker interactions promote transient binding, and the leukocyte rolls along the endothelial wall. Subsequent strong binding stops the rolling and allows the leukocyte to pass between endothelial cells. The migrating leukocytes follow the chemokine gradient to reach infected tissues (*C. perfringens* and *C. pseudotuberculosis*). The bottom left depicts an abnormal state where overexpression of adhesion integrins on the endothelium promotes sustained leukocyte adherence; degranulation of the static leukocytes could then cause damage to the vessel (*C. perfringens*). (Taken from Schmiel D. H., *et al*. (1999) *Microb. Infect*. 1, 1103-1112).

toxins were in fact secreted phospholipases, for example the *Clostridium perfringens* toxin [56] and *Staphylococcus aureus* toxin [57]. In general, phospholipase toxicity has been linked to cytolytic activity and is presumed to be directly due to phospholipase activity on membrane phospholipids and membrane destruction (Fig. 1.4). Cytolysis is one of the more common characteristics attributed to bacterial phospholipase virulence factors (Tab. 1.2). Whether cytolysis results from the accumulation of membrane destabilizing products or by the whole sale destruction of membrane phospholipids, it can be caused directly by a very active bacterial phospholipase with broad specificity or in concert with host degradative enzymes induced by the bacterial phospholipase. Naturally, the cytolytic activity varies greatly amongst bacterial phospholipases. The cytolytic properties of the phospholipase itself are dependent on the ability to interact directly with and hydrolyze phospholipids in the membrane. Thus the enzymatic activity determined using solublized substrates is not indicative of the cytolytic potential of a particular phospholipase if it interacts poorly with membranes. Frequently, the lytic activity is measured as hemolysis for convenience, regardless of the likelihood for a particular pathogen to cause systemic infection or bacterial phospholipase to enter the bloodstream. Furthermore, different eukaryotic membranes are composed of different phospholipids, even the two leaflets (inner and outer) of a particular membrane may have very distinct phospholipid composition. Therefore, an extracellular phospholipase would not be expected to lyse a target cell unless the outer leaflet phospholipids were efficiently hydrolyzed substrates of the enzyme. Of course, this does not exclude possible cooperativity between several phospholipases or between a phospholipase and another type of ‘lysin’ (hemolysin or cytolysin) molecule as has been proposed for Listeria monocytogenes [57]. Specificity for phospholipids in a particular membrane type or leaflet has had some interesting implications. In addition to eukaryotic cell cytolysis, bacterial phospholipases have been implicated in the very

Species	*Phospholipase type*	*Cytokines/ eicosanoids*[a]	*Properties*
Bacillus cereus	PLC		Hemolytic with sphingomyelinase
Campylobacter coli	PLA_2		Hemolysis, null mutant nonhemolytic
Clostridium perfringens	PLC (α toxin)	IL-8, TNF-α PAF	Cytolysis and tissue destruction, leukocyte killing; purified α toxin induces many features of septic shock and tissue necrosis in animals, induces arachidonic acid cascade
Corynebacterium pseudotuberculosis	PLD		Purified PLD is dermonecrotic and lethal; null mutant no or limited abscess formation at inoculum causing 100% abscesses with wild type; interferes with neutrophil chemotaxis and increases vascular permeability
Helicobacter pylori	PLA_2		With other exoenzymes destroys protective gastric mucus; lysophospholipid product cytolytic and acts as mucin surfactant
Listeria monocytogenes	PlcA PlcB		Specific digestion of endosomal membranes Synergy with listeriolysin O; double mutant attenuated in animal model different phenotypes, yet overlap of function
Mycobacterium tuberculosis	PLC (PLD)		Genes identified, hemolytic when expressed in *E. coli*
Pseudomonas aeruginosa	PLC	IL-8, thromoboxanes, leukotrienes	Tissue destruction, hemolysis; purified PLC reproduces many of the pathologies seen in infected burn patients and fleecerot in sheep
Rickettsia prowazekii	PLA		Hemolysis, production of lysophospholipids; proposed to induce phagocytosis, endosomal escape, and host cell lysis
Salmonella newport	PLA		Injection purified PLA induces same histopathologic changes in ligated ileal loops as bacteria
Staphylococcus aureus	PLC (β toxin)	IL-1β	Hemolytic, cytotoxic for monocytes and epithelial cells; null mutant causes less damage in animal models; pure PLC causes inflammation in keratitis model
Vibrio parahaemolyticus	PLA		Hemolysis
Vibrio cholerae	PLA?		Cytotoxic for intestinal cell line; mutant demonstrated no difference in fluid influx into ligated ileal loops compared to wild type
Yersinia enterocolitica	PLA		Mutant reduced colonization of tissues and caused less inflammation and necrosis than wild type in infected mouse tissues

Table 1.2. Bacterial phospholipases implicated in pathogenesis. (Taken from Schmiel D. H., *et al*. (1999) *Microb. Infect*. 1, 1103-1112).

specific destruction of endosomal or phagosomal membranes [58-61]. This particular membrane may have a very distinct phospholipid composition. For instance, the plasma membrane of most eukaryotic cells contains predominantly PC and sphingomyelin in the outer leaflet and PI, PE, and PS in the inner leaflet. Therefore, an extracelluler phospholipase would not be expected to lyse a target cell unless the outer leaflet phospholipids were efficiently hydrolysable substrates of the enzyme. In addition to eukaryotic cell cytolysis, bacteria phosphoipases have been implicated in the very specific destruction of endosomal or phagosomal membranes. This releases the bacterium into the nutrient-rich host cell cytoplasm and is one method of escape from possible annihilation by the lysosome (Fig. 1.4). As some of the more active cytolytic phospholipases are linked to diseases characterized by tissue destruction, one could speculate that destruction is solely due to lysis of individual cells comprising the tissues. However, some of the best-studied examples of phospholipase virulence factors implicated in destructive infections, for example the two *P. aeruginosa* PLC and the *C. perfringens* toxin (see below), have also demonstrated phospholipase dependent induction of immunomodulatory cytokines that stimulate the inflammatory response [62-64]. Consequently, in these examples tissue destruction has been proposed to be partially due to the overstimulation of the inflammatory response indirectly caused by the phospholipase. In addition, *C. perfringens* toxin [65], and the hemolytic *P. aeruginosa* PLC [66] were shown to stimulate the arachidonic acid cascade probably by further hydrolysis of diacylglycerol by endogenous cytoplasmic enzymes. Hydrolysis of diacylglycerol could release arachidonic acid, the precursor for the eicosanoids, lipid second messengers which include prostaglandins, thromboxanes, and leukotrienes [67]. Prostaglandins and thromboxanes are vasoactive and modulate platelet activation locally [68]. Leukotrienes are important mediators of acute inflammation affecting neutrophil migration, superoxide generation, and degranulation. Interestingly, hydrolysis of phospholipids by PLA2 releases arachidonic acid directly because it is most commonly found in position 2 of mammalian phospholipids. Though bacterial PLAs have been implicated as virulence factors, the potential for stimulation of the arachidonic acid cascade has not been examined.

1.3 Carboxylesterases

Esterases (EC 3.1.1.x) represent a diverse group of hydrolases catalyzing the cleavage and formation of ester bonds. They are widely distributed in animals, plants and

microorganisms. Many of them show a wide substrate tolerance which led to the assumption that they have evolved to enable access to carbon sources or to be involved in catabolic pathways. Moreover, esterases also show high regio- and stereospecificity, which makes them attractive biocatalysts for the production of optically pure compounds in fine-chemicals synthesis [69-74]. The interest in these enzymes also resides in the fact that they do not require cofactors, are usually rather stable and are even active in organic solvents.

Two major classes of hydrolases are of utmost importance: lipases (EC 3.1.1.1, triacylglycerol hydrolases) and 'true' esterases (EC 3.1.1.3, carboxyl ester hydrolases). The three-dimensional (3D) structures of both enzymes show the characteristic α/β hydrolase fold [75] which is also found in haloperoxidases and epoxide hydrolases (Fig. 1). The catalytic triad is composed of Ser-Asp-His (Glu instead of Asp for some lipases) and usually also a consensus sequence (Gly-X-Ser-X-Gly) is found around the active site serine. More recently, esterases have been identified containing a Gly-X-X-Leu motif [76] as well as enzymes showing high homology to class C β-lactamases [77].

Lipases can be distinguished from esterases by the phenomenon of interfacial activation, which was only observed for lipases. Whereas esterases obey classical Michaelis-Menten kinetics, lipases need a minimum substrate concentration before high activity is observed. Structure elucidation revealed that this interfacial activation is due to a hydrophobic domain (lid) covering the active site of lipases. Only in the presence of a minimum substrate concentration, i.e. a triglyceride phase or a hydrophobic organic solvent, the lid moves apart, making the active site accessible. Furthermore, lipases prefer water-insoluble substrates, typically triglycerides composed of long-chain fatty

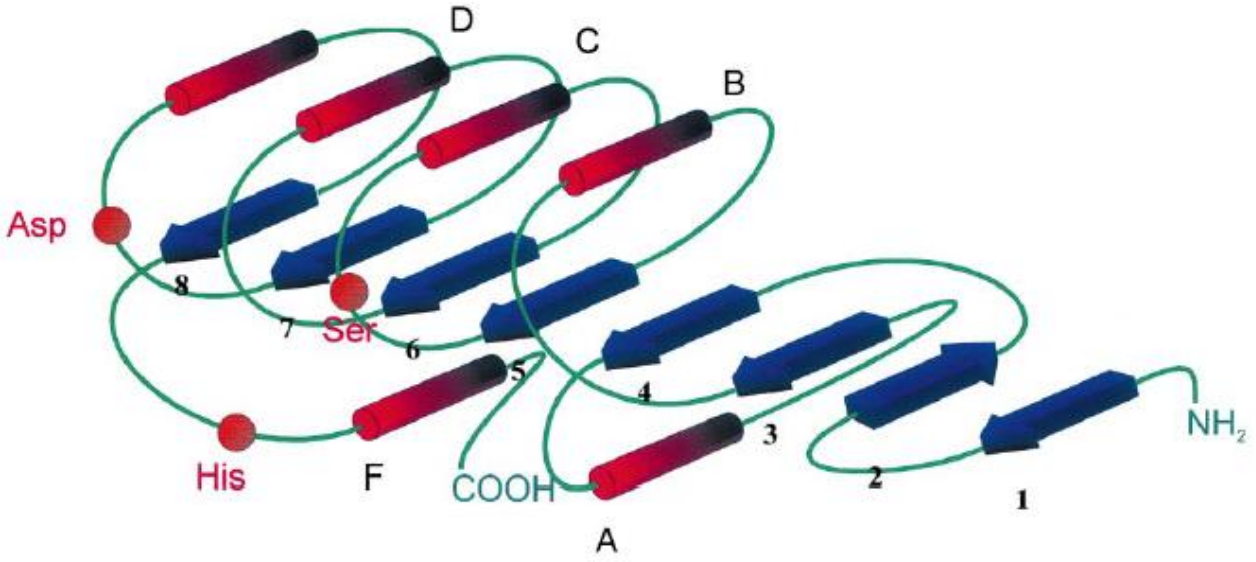


Figure 1.1. Schematic presentation of the prototypic α/β hydrolase fold. β-strands (1-8) are shown as blue arrows, α-helices (A-F) as red cylinders. The relative positions of the amino acids of the catalytic triad are indicated as red circles.

acids, whereas esterases preferentially hydrolyze ‘simple’ esters (e.g. ethyl acetate) and usually only triglycerides bearing fatty acids shorter than C6. Both enzymes have been shown to be stable and active in organic solvents, but this feature is more pronounced with lipases. A comparison of amino acid sequences and 3D-structures of lipases and esterases suggested that they can be differentiated by a pH-dependent electrostatic `signature' [78]; the active site of lipases displays a negative potential in the pH-range associated with their maximum activity (typically at pH 8), whereas esterases show a similar pattern, but at pH values around 6, which correlates with their usually lower pH-activity optimum [78]. Screening of esterases is usually performed either by employing chromophoric substances (e.g. p nitrophenyl esters) or by using tributyrin-supplemented agar plates. Organisms producing active esterases are then identified by halo formation on plates containing dispersed water-insoluble substrates such as triglycerides. Hydrolysis of these dispersed lipid droplets leads to a clearing zone around the colonies. Then, lipases can be distinguished from carboxyl esterases by their substrate spectra, using p-nitrophenyl palmitate (cleaved by lipases) vs. *p*-nitrophenyl butyrate (cleaved by esterases and sometimes also by lipases). In contrast to this traditional approach, modern molecular biology techniques allow for DNA library expression screening.

Classification of enzymes can be done either based on their substrate specificity or by sequence alignments. The former requires that all enzymes which need to be compared have been assayed with the same or at least related substrates, preferentially under similar reaction conditions. Although this would allow the biochemist to directly identify a suitable enzyme for a given synthetic problem, this is not a common strategy. Due to the increasing availability of sequence information in public databases, the comparison of amino acid sequences can provide a clearer picture about the evolutionary relationship between enzymes of different origin. On the other hand, it often turns out that high sequence homology cannot be related to the enzyme’s properties (substrate specificity, stereoselectivity, pH and temperature optima, solvent stability), and in some cases completely different types of reactions are catalyzed. For instance, a bromoperoxidase from *Streptomyces aureofaciens* shares 55% sequence identity to an esterase from *Pseudomonas fuorescens* [79, 80] but they share little substrate specificity. Until a few years ago, it was believed that for all lipases and carboxyl esterases only the consensus sequence motif Gly-X-Ser-X-Gly (where X represents an arbitrary amino acid residue) occurs around the active site serine. More recently, a thorough comparison of 53 amino acid sequences of lipases and esterases revealed that other motifs also exist [81]. For instance, some lipases and an esterase from *Streptomyces scabies* contain a GDSL (Gly-Asp-Ser-Leu) consensus sequence. Moreover, structure elucidation of this esterase revealed that it contains a catalytic Ser-His dyad instead of the common Ser-Asp- His triad [76]. The acidic side chain, which usually stabilizes the positive charge of the active site histidine residue, is replaced by the backbone carbonyl of Trp315 located three positions upstream of the His itself. The

enzyme also has an α/β tertiary fold, which differs substantially from the α/β hydrolase fold. Other esterases in the GDSL group includes those from *Pseudomonas aeruginosa* (accession code: AF005091), *Salmonella typhimurium* (AF047014) and *Photorhabdus luminescens* (X66379), the first two being outer-membrane-bound esterases. Other enzymes show high homology to the mammalian hormone-sensitive lipase family. Here, conserved sequence motifs were found, which initially have been related to activity at low temperature. However, it was found that esterases from psychrophilic (e.g. *Moraxella* sp., X53869; *Psychrobacterimmobilis*, X67712) as well as mesophilic (*Escherichia coli*, AE000153) and thermophilic (Archeoglobus fulgidus, AAE000985) origins belong to this family. Members of family V, such as esterases from *Sulfolobus acidocaldarius* (AF071233) and *Acetobacter pasteurianus* (AB013096), share significant homology to non-lipolytic enzymes, e.g. epoxide hydrolase, dehalogenase and haloperoxidase. Rather small (23-26 kDa) enzymes are found in family VI, which includes an esterase from *P. fluorescens*, of which the structure is known [82]. The esterase is active as a dimer, has a typical Ser-Asp-His catalytic triad and hydrolyzes small substrates, but not long-chain triglycerides. Interestingly, 40% homology to eukaryotic lysophospholipases is found for members of this family. In contrast, esterases from family VII are rather large (55 kDa) and share significant homology to eukaryotic acetylcholine esterases and intestine or liver carboxyl esterases (e.g. pig liver esterase). A *p*-nitrobenzyl esterase from *Bacillus subtilis* [83, 84] and an esterase from *Arthrobacter oxydans* (Q01470) active against phenylcarbamate herbicides [85] belong to this group. In the last family, VIII, high homology to class C β-lactamases is observed. These enzymes contain a Gly-X-Ser-X-Gly motif and a Ser-X- X-Lys motif, but it has recently been demonstrated by site directed mutagenesis studies of an esterase (EstB) from *Burkholderia gladioli* that the Gly-X-Ser-X-Gly motif does not play a significant role in enzyme function [77]. The most prominent member is an esterase from *Arthrobacter globiformis* (AAA99492) [86].

Enzymes belonging to the α/β hydrolase fold family, all contain a catalytic triad consisting of a nucleophile, a base, and an orientating acid [2].

The existence of the catalytic triad was first established with acetylcholinesterase (AChE) a well-studied member of the α/β hydrolase fold family. It has been known for some time that both the serine and the histidine are members of the catalytic triad in AchE [3–5]. Further investigation via mutagenesis of these amino acids in AchE (*Torpedo californica*) confirmed that they were essential for catalysis [6]. The third member of the catalytic triad, the glutamic acid, was identifyed by crystal structure analysis of AChE [7]. This was again supported biochemically via the mutation of the Glu334 in human AChE in which all catalytic activity was lost [8]. This procedure has

been repeated for many lipases and esterases (e.g., [9–11]), including the highly specific juvenile hormone esterase [12].

Carboxylesterases cleave esters via a two-step process that involves the formation and degradation of an acyl-enzyme intermediate. The hydrolysis initially occurs through a nucleophilic attack via the serine onto the carbon of the carbonyl group (Fig. 1.2) [2]. The catalytic serine is able to accomplish this with the assistance of the histidine acting as a general base. In turn, the protonated histidine is stabilized via a hydrogen bond to the glutamic acid. These two catalytic amino acids, His–Glu, are often called a "charge relay" system because they work together to activate the serine. The initial nucleophilic attack produces what is thought to be the first of two tetrahedral intermediates (Fig. 1.2, **A**) that are stabilized by the presence of two main chain amide group in the active site (oxyanion hole).

This tetrahedral intermediate rapidly collapses, aided by the protonated-histidine acting as a general acid, to produce an acyl-enzyme complex (Fig. 1.2, **B**) while the alcohol component of the ester is released. The acyl-enzyme complex then undergoes an attack by a histidine-activated water molecule (Fig. 1.2, **C**) which produces the second tetrahedral intermediate (Fig. 1.2, **D**). After rapid rearrangement of this intermediate, the enzyme is regenerated and the acid component released. Much debate still surrounds the exact nature of the mechanism described above, which includes the possibility of a short, strong bond or low barrier hydrogen bond (LBHB)

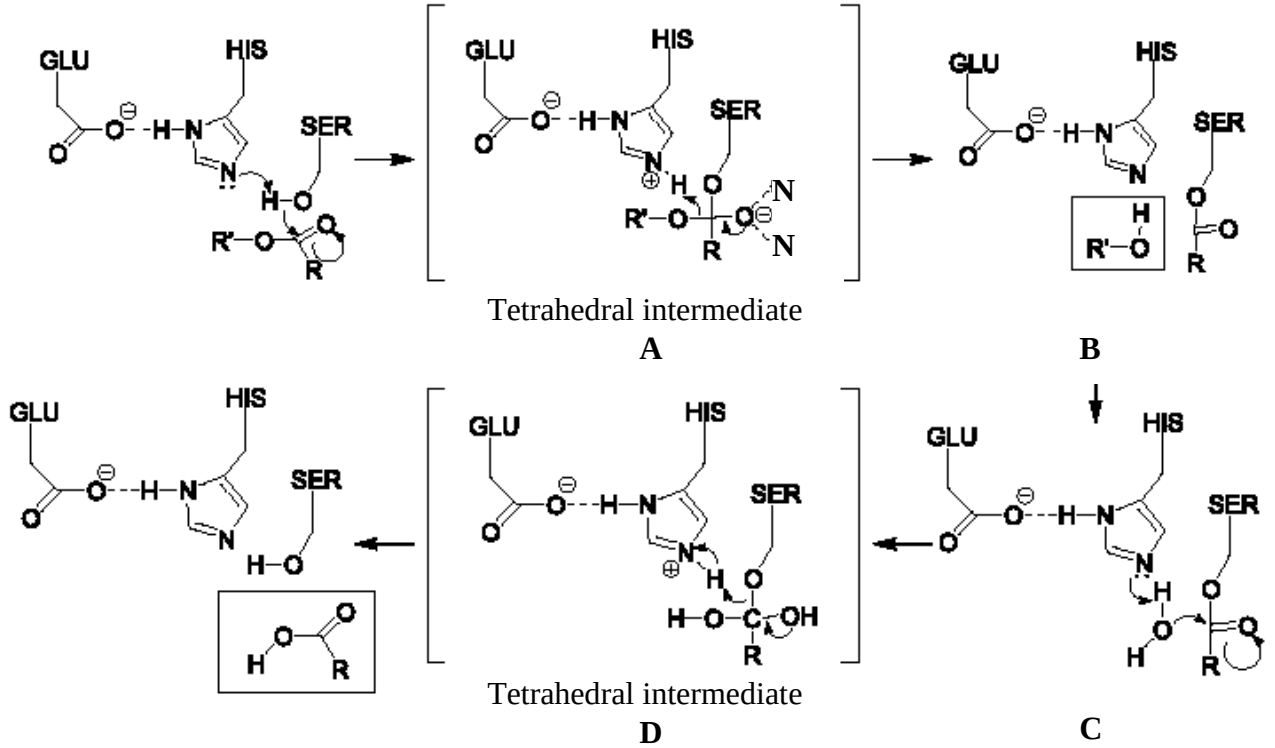


Figure 1.2. Detailed mechanism of the hydrolysis of esters by carboxylesterases. Compounds in boxes are the products of the hydrolysis.

between the glutamic acid and the histidine during catalysis [13–16]. By lowering the activation energy required for the formation of the first tetrahedral intermediate the LBHB is believed to assist in the activation of the catalytic serine.

1.4 Aim of the research project

The aim of this work has been the structural and functional study of a putative carboxylesterase purified from *P. aeruginosa*, namely PA3859. The protein has been purified from the wild type and a preliminary biochemical charcterization was carried out. The PA3859 gene was then cloned and the recombinant protein was expressed in *E. coli* (Chapter 2).
The recombinant PA3859 was successfully crystallized and its 3D crystal structure was determined (Chapter 3 and 4).
Starting from the enzyme 3D structure, an approach involving *in silico*, *in vitro* and *in vivo* assays lead to the reliable determination of the PA3859 physiological function (Chapter 5, 6, 7).

1.5 References

[1] Lyczak, J.B. *et al*. (2002) Lung infections associated with cystic fibrosis. *Clin. Microbiol. Rev.* **15**, 194–222.

[2] Leroy-Dudal, J. *et al*. (2004) Role of avb5 integrins and vitronectin in *Pseudomonas aeruginosa* PAK interaction with A549 respiratory cells. *Microbe Infect*. **6**, 875–881

[3] Lee, A. *et al*. (1999) Airway epithelial tight junctions and binding and cytotoxicity of *Pseudomonas aeruginosa*. *Am. J. Physiol*. **277**, 204–217.

[4] Evans, D.J. *et al*. (1998) *Pseudomonas aeruginosa* induces changes in fluid transport across airway surface epithelia. *Am. J. Physiol*. **275**, 1284–1290.

[5] Azghani, A.O. *et al*. (2000) Virulence factors from *Pseudomonas aeruginosa* increase lung epithelial permeability. *Lung* **178**, 261–269.

[6] Kanthakumar, K. *et al*. (1996) The effect of bacterial toxins on levels of intracellular adenosine nucleotides and human ciliary beat frequency. *Pulm. Pharmacol*. **9**, 223–230.

[7] Worlitzsch, D. *et al*. (2002) Reduced oxygen concentrations in airway mucus contribute to the early and late pathogenesis of *Pseudomonas aeruginosa* cystic fibrosis airway infection. *J. Clin. Invest*. **109**, 317–325.

[8] Grassme´, H. *et al*. (2000) CD95/CD95 ligand interactions on epithelial cells in host defense to *Pseudomonas aeruginosa*. *Science* **290**, 527–530.

[9] Cannon, C.L. *et al*. (2003) *Pseudomonas aeruginosa*-induced apoptosis is defective in respiratory epithelial cells expressing mutant cystic fibrosis transmembrane conductance regulator. Am. J. Respir. Cell Mol. Biol. **29**, 188–197.

[10] Ratner, A.J. *et al*. (2001) Cystic fibrosis pathogens activate Ca^{2+}-dependent mitogen-activated protein kinase signaling pathways in airway epithelial cells. *J. Biol. Chem.* **276**, 19267–19275.

[11] Adamo, R. *et al*. (2004) *Pseudomonas aeruginosa* flagella activate airway epithelial cells through asialo GM1 and toll-like receptor 2 as well as toll-like receptor 5. *Am. J. Respir. Cell Mol. Biol*. **30**, 627–634.

[12] Lillehoj, E.P. *et al*. (2004) *Pseudomonas aeruginosa* stimulates phosphorylation of the airway epithelial membrane glycoprotein Muc1 and activates MAP kinase. *Am. J. Physiol. Lung Cell Mol. Physiol*. **287**, L809–L815.

[13] Li, J.D. *et al*. (1998) Activation of NF-kBvia a SRC-dependent Ras-MAPK-pp90rsk pathway is required for *Pseudomonas aeruginosa* induced mucin overproduction in epithelial cells. *Proc. Natl. Acad. Sci*. *U. S. A.* **95**, 5718–5723.

[14] Schroeder, T.H. *et al*. (2002) CFTR is a pattern recognition molecule that extracts *Pseudomonas aeruginosa* LPS from the outer membrane into epithelial cells and activates NF-kappaB translocation. *Proc. Natl. Acad. Sci. U. S. A*. **99**, 6907–6912.

[15] Cobb, L.M. *et al*. (2004) *Pseudomonas aeruginosa* flagellin and alginate elicit very distinct gene expression patterns in airway epithelial cells: implications for cystic fibrosis disease. *J. Immunol*. **173**, 5659–5670.

[16] Wolfgang, M.C. *et al*. (2004) *Pseudomonas aeruginosa* regulates flagellin expression as part of a global response to airway fluid from cystic fibrosis patients. *Proc. Natl. Acad. Sci. U. S. A*. **101**, 6664–6668.

[17] Ichikawa, J.K. *et al*. (2000) Interaction of *Pseudomonas aeruginosa* with epithelial cells: Identification of differentially regulated genes by expression microarray analysis of human cDNAs. *Proc. Natl. Acad. Sci. U. S. A*. 97, 9659–9664.

[18] Erridge, C. *et al*. (2004) Lipopolysaccharides of Bacteroides fragilis, *Chlamydia trachomatis* and *Pseudomonas aeruginosa* signal via tolllike receptor 2. *J. Med. Microbiol.* **53**, 735–740.

[19] Singh, P.K. *et al*. (2000) Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**, 762–764.

[20] Hall-Stoodley, L. *et al*. (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol*. **2**, 95–108.

[21] Yoon, S.S. *et al*. (2002) *Pseudomonas aeruginosa* anaerobic respiration in biofilms: relationships to cystic fibrosis pathogenesis. *Dev. Cell* **3**, 593–603.

[22] Kopp, E. and Medzhitov, R. (2003) Recognition of microbial infection by Toll-like receptors. *Curr. Opin. Immunol*. **15**, 396–401.

[23] Kulshin, V.A. *et al*. (1991) Structural characterization of lipid A component of *Pseudomonas aeruginosa* wild-type and rough mutant lipopolysaccharides. *Eur. J. Biochem.* **198**, 697–704.

[24] Lopez-Boado, Y.S. *et al*. (2001) Regulation of matrilysin expression in airway epithelial cells by *Pseudomonas aeruginosa* flagellin. *J. Biol. Chem*. **276**, 41417–41423.

[25] Koyama, S. *et al*. (1999) The potential of various lipopolysaccharides to release monocyte chemotactic activity from lung epithelial cells and fibroblasts. *Eur. Respir. J.* **14**, 545–552.

[26] Feldman, M. *et al*. (1998) Role of flagella in pathogenesis of *Pseudomonas aeruginosa* pulmonary infection. *Infect. Immun*. **66**, 43–51.

[27] Comolli, J.C. *et al*. (1999) *Pseudomonas aeruginosa* gene products PilT and PilU are required for cytotoxicity in vitro and virulence in a mouse model of acute pneumonia. *Infect. Immun*. **67**, 3625–3630.

[28] Wozniak, D.J. *et al*. (2003) Alginate is not a significant component of the extracellular polysaccharide matrix of PA14 and PAO1 *Pseudomonas aeruginosa* biofilms. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7907–7912.

[29] Kishioka, C. *et al*. (1999) *Pseudomonas aeruginosa* alginate is a potent secretagogue in the isolated ferret trachea. *Pediatr. Pulmonol*. **27**,174–179.

[30] Smith, R.S. and Iglewski, B.H. (2003) *Pseudomonas aeruginosa* quorum-sensing systems and virulence. *Curr. Opin. Microbiol*. **6**, 56–60.

[31] Frank, D.W. (1997) The exoenzyme S regulon of *Pseudomonas aeruginosa*. *Mol. Microbiol*. **26**, 621–629.

[32] Sato, H. *et al*. (2003) The mechanism of action of the *Pseudomonas aeruginosa*-encoded type III cytotoxin. *ExoU. EMBO J*. **22**, 2959–2969.

[33] Shaver, C.M. and Hauser, A.R. (2004) Relative contributions of *Pseudomonas aeruginosa* ExoU, ExoS, and ExoT to virulence in the *lung. Infect. Immun*. **72**, 6969–6977.

[34] Sun, J. *et al*. (2004) How bacterial ADP-ribosylating toxins recognize substrates. *Nat. Struct. Mol. Biol*. **11**, 868–876.

[35] Kazmierczak, B.I. *et al*. (2004) Epithelial cell polarity alters Rho-GTPase responses to *Pseudomonas aeruginosa*. *Mol. Biol. Cell* 15, 411–419.

[36] Sundin, C. *et al*. (2004) ADP-ribosylation by exoenzyme T of *Pseudomonas aeruginosa* induces an irreversible effect on the host cell cytoskeleton in vivo. *FEMS Microbiol. Lett*. **234**, 87–91.

[37] Banwart, B. *et al*. (2002) Children with cystic fibrosis produce an immune response against exoenzyme S, a type III cytotoxin of *Pseudomonas aeruginosa*. J. *Infect. Dis*. **185**, 269–270.

[38] Lau, G.W. *et al*. (2004) The role of pyocyanin in *Pseudomonas aeruginosa* infection. *Trends Mol. Med*. **10**, 599–606.

[39] Mariencheck, W.I. *et al*. (2003) *Pseudomonas aeruginosa* elastase degrades surfactant proteins A and D. *Am. J. Respir. Cell Mol. Biol*. **28**, 528–537.

[40] Schmidtchen, A. *et al*. (2002) Proteinases of common pathogenic bacteria degrade and inactivate the antibacterial peptide LL-37. *Mol. Microbiol*. **46**, 157–168.

[41] Leidal, K.G. *et al*. (2003) Metalloproteases from *Pseudomonas aeruginosa* degrade human RANTES, MCP-1, and ENA-78. J. Interferon Cytokine Res. **23**, 307–318.

[42] Barker, A.P. *et al*. (2004) A novel extracellular phospholipase C of *Pseudomonas aeruginosa* is required for phospholipid chemotaxis. *Mol. Microbiol*. **53**, 1089–1098.

[43] Wilderman, P.J. *et al*. (2001) Genetic and biochemical analyses of a eukaryotic-like phospholipase D of *Pseudomonas aeruginosa* suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol. Microbiol*. **39**, 291–303.

[44] Luberto, C. *et al*. (2003) Purification, characterization, and identification of a sphingomyelin synthase from *Pseudomonas aeruginosa*. PlcH is a multifunctional enzyme. *J. Biol. Chem*. **278**, 32733–32743.

[45] Ochsner, U.A. *et al*. (2002) Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis. Proc. *Natl. Acad. Sci. U. S. A*. **99**, 8312–8317.

[46] Lamont, I.L. *et al*. (2002) Siderophore-mediated signaling regulates virulence factor production in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7072–7077.

[47] Kounnas, M.Z. *et al*. (1992) The 2-macroglobulin receptor/low density lipoprotein receptor-related protein binds and internalizes *Pseudomonas* exotoxin A. *J. Biol. Chem*. **267**, 12420–12423.

[48] Rajan, S. *et al*. (2000) *Pseudomonas aeruginosa* induction of apoptosis in respiratory epithelial cells: analysis of the effects of cystic fibrosis transmembrane conductance regulator dysfunction and bacterial virulence factors. *Am. J. Respir. Cell Mol. Biol*. **23**, 304–312.

[49] Plotkowski, M.C. *et al*. (2002) Early mitochondrial dysfunction, superoxide anion production, and DNA degradation are associated with non-apoptotic death of human airway epithelial cells induced by *Pseudomonas aeruginosa* exotoxin *A. Am. J. Respir. Cell Mol. Biol*. **26**, 617–626.

[50] Vadas, P. *et al*. (1993) phospholipase A2 expression and inflammation: therelationship with associated disease states, J. Lipid Mediators **8**, 1–30.

[51] Ohki, M. *et al*. (1972) Mutant of *Escherichia coli* K-12 deficient for the detergent-resistant phospholipase A, *J. Bacteriol*. **110**, 864–869.

[52] Van Oort, G.M. *et al*. (1989) Purification and substrate specificity of *Staphylococcus hyicus* lipase, *Biochemistry* **28**, 9278–9285.

[53] Titball, R.W., (1993) Bacterial phospholipases C, *Microbiol. Rev*. **57**, 347–366.

[54] Songer, J.G., (1997) Bacterial phospholipases and their role in virulence, Trends *Microbiol.* **156**, 156–161.

[55] Titball, R.W., (1198) Bacterial phospholipases, *J. Appl. Microbiol*. **84**, 127S–137S.

[56] Macfarlene, R.G. *et al*. (1941) Hemolysis and the production of opalesence in serum and lecithovitellin by the alpha toxin of *Clostridium perfringens*, *J. Pathol. Bacteriol*. **52**, 99–103.

[57] Doery, H.M. *et al*. (1963) A phospholipase in staphylococcal toxin which hydrolyzessphingomyelin, *Nature* **198**, 1091–1092.

[58] Camilli, A. *et al*. (1993) Dual roles of plcA in *Listeria monocytogenes* pathogenesis, *Mol. Microbiol*. **8**, 143–157.

[59] Portnoy, D.A. *et al*. (1994) Cellular biology of *Listeria monocytogenes* infection, Molecular Genetics of Bacterial Pathogenesis, American Society for Microbiology, ASM Press, Washington, DC, pp. 279–293.

[60] Sheehan, B. *et al*. (1994) Molecular and genetic determinants of the *Listeria monocytogenes* infectious process, *Curr. Top. Microbiol. Immunol.* **192**, 187–216.

[61] Goldfine, H. *et al*. (1998) Bacterial phospholipases and intracellular growth: the two distinct phospholipases C of *Listeria monocytogenes*, *J. Appl. Microbiol.* **84**, 7–14.

[62] Konig, B. *et al*. (1997) Role of haemolytic and non-haemolytic phospholipase C from *Pseudomonas aeruginosa* in interleukin-8 release from human monocytes, *J. Med. Microbiol.* **46**, 471–478.

[63] Bryant, A.E. *et al*. (1996) Phospholipase C and perfringolysin O from *Clostridium perfringens* up regulate endothelial cell-leukocyte adherence molecule 1 and intercellular leukocyte adherence molecule 1 expression and induce interleukin-8 synthesis in cultured human umbilical vein endothelial cells, *Infect. Immun*. **64**, 358–362.

[64] Stevens D.L. *et al*. (1997) Clostridial gas gangrene: evidence that alpha and theta toxins differentially moduate the immune response and induce acute tissue necrosis, *J. Infect. Dis.* **176,** 189–195.

[65] Stevens, D.L. *et al*. (1997) Clostridial gas gangrene: evidence that alpha and theta toxins differentially moduate the immune response and induce acute tissue necrosis, *J. Infect. Dis*. **176**, 189–195.

[66] Fujii, Y. *et al*. (1989) Contraction of the rat isolated aorta caused by *Clostridium perfringens* alpha-toxin (phospholipase C): evidence for the involvement of arachidonic acid metabolism, *Br. J. Pharmacol*. **97**, 119–124.

[67] Meyers, D.J. *et al*. (1990) Characterization of phospholipase C from *Pseudomonas aeruginosa* as a potent inflammatory agent, *Infect. Immun*. **58**, 659–666.

[68] Williams K.I. *et al*. (1998) Eicosanoids and inflammation, *J. Pathol*. **156**, 101–110.

[69] Bornscheuer, U.T. and Kazlauskas, R.J. (1999) Hydrolases in organic synthesis – regio- and stereoselective biotransformations. Wiley-VCH, Weinheim.

[70] Phytian, S.J. (1998) Esterases. In: Biotechnology-Series, Vol. 8a (Rehm, H.J., Reed, G., Pu´ hler, A., Stadler, P.J.W. and Kelly, D.R., Eds.), pp. 193-241. Wiley-VCH, Weinheim.

[71] Patel, R.N. (2000) Stereoselective Biocatalysis. Marcel Dekker, New York.

[72] Faber, K. (2000) Biotransformations in Organic Chemistry. Springer, Berlin.

[73] Drauz, K. and Waldmann, H. (1995) Enzyme Catalysis in Organic Synthesis, Vol. 1 and 2. Wiley-VCH, Weinheim.

[74] Wong, C.-H. and Whitesides, G.M. (1994) Enzymes in Synthetic Organic Chemistry. Pergamon Press, Oxford.

[75] Ollis, D.L. *et al*. (1992) The α/β hydrolase fold. *Protein Eng*. **5**, 197- 211.

[76] Wei, Y. *et al*. (1995) A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nat. Struct. Biol*. **2**, 218-223.

[77] Petersen, E.I. *et al*. (2001) A novel esterase from *Burkholderia gladioli* shows high deacetylation activity on cephalosporins is related to L-lactamases and DD-peptidases. J. Biotechnol. **89**, 11-25.

[78] Fojan, P. *et al*. (2000) What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie* **82**, 1033-1041.

[79] Pelletier, I. and Altenbuchner, J. (1995) A bacterial esterase is homologous with non-haem haloperoxidases and displays brominating activity. *Microbiology* **141**, 459-468.

[80] Hecht, H.J. *et al*. (1994) The metal-ion-free oxidoreductase from *Streptomyces aureofaciens* has an α/β hydrolase fold. *Nat. Struct. Biol.* **1**, 532-537.

[81] Arpigny, J.L. and Jaeger, K.E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J*. **343**, 177-183.

[82] Kim, K.K. *et al*. (1997) Crystal structure of carboxylesterase from *Pseudomonas fuorescens*, an α/β-hydrolase with broad substrate specificity. *Structure* **5**, 1571-1584.

[83] Zock, J. *et al*. (1994) The Bacillus subtilis pnbA gene encoding pnitrobenzyl esterase: cloning, sequence and high-level expression in *Escherichia coli*. *Gene* **151**, 37-43.

[84] Moore, J.C. *et al*. (1996) Directed evolution of a paranitrobenzyl esterase for aqueous-organic solvents. *Nat. Biotechnol*. **14**, 458-467.

[85] Pohlenz, H.D. *et al*. (1992) Purification and properties of an *Arthrobacter oxydans* P52 carbamate hydrolase specific for the herbicide phenmedipham and nucleotide sequence of the corresponding gene. *J. Bacteriol*. **174**, 6600-6607.

[86] Nishizawa, M. *et al*. (1995) Stereoselective production of (+)-*trans*-chrysanthemic acid by a microbial esterase: cloning, nucleotide sequence, and overexpression of the esterase gene of *Arthrobacter globiformis* in *Escherichia coli*. *Appl. Environ. Microbiol*. **61**, 3208-3215.

[87] Satoh, T. *et al*. (1989) The mammalian carboxylesterases: from molecules to functions *Ann. Rev. Pharm. Toxicol*. **38**, 257–288.

[88] Krupka, R.M (1966) Hydrolysis of neutral substrates by acetylcholinesterase *Biochemistry* **5**, 1983–1988.

[89] Roskoski, R. (1974) Choline acetyltransferase and acetylcholinesterase: evidence for essential histidine residues *Biochemistry* **13**, 5141–5144.

[90] MacPhee-Quigley, K. *et al*. (1985) Primary structures of the catalytic subunits from two molecular forms of acetylcholinesterase. A comparison of NH2-terminal and active center sequences *J. Biol. Chem*. **260**, 12185–12189.

[91] Gibney, G. *et al*. Mutagenesis of essential functional residues in acetylcholinesterase *Proc. Natl. Acad. Sci. USA* 87 (1990) 7546–7550.

[92] Sussman, J.L. (1991) Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein *Science* **253**, 872–879.

[93] Shafferman, *et al*. (1992) Mutagenesis of human acetylcholinesterase. Identification of residues involved in catalytic activity and in polypeptide folding *J. Biol. Chem*. **267**, 17640–17648.

[94] Emmerich, J. *et al*. (1992) Human lipoprotein lipase. Analysis of the catalytic triad by site-directed mutagenesis of Ser-132, Asp-156, and His-241 *J. Biol. Chem*. **267**, 4161–4165.

[95] Haruki, M. *et al*. (1999) Identification of catalytically essential residues in Escherichia coli esterase by site-directed mutagenesis *FEBS Lett*. **454**, 262–266.

[96] Lohse, P. *et al*. (1997) Human lysosomal acid lipase/cholesteryl ester hydrolase and human gastric lipase: identification of the catalytically active serine, aspartic acid, and histidine residues *J. Lipid Res*. **38**, 892–903.

[97] Ward, V.K. *et al*. (1992) Analysis of the catalytic mechanism of juvenile hormone esterase by site-directed mutagenesis *Int. J. Biochem.* **24**, 1933–1941.

[98] Frey, P.A. *et al*. (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases *Science* **264**, 1927–1930.

[99] Warshel, A. *et al*. (1995) On low-barrier hydrogen bonds and enzyme catalysis *Science* **269**, 102–106.

[100] Viragh, C. *et al*. (2000) NMR evidence for a short, strong hydrogen bond at the active site of a cholinesterase *Biochemistry* **39**, 16200–16205.

[101] Massiah, M.A. *et al*. (2001) Short, strong hydrogen bonds at the active site of human acetylcholinesterase: proton NMR studies *Biochemistry* **40**, 5682–5690.

Chapter 2

Isolation, Characterization, and Heterologous Expression of Carboxylesterase PA3859 from *Pseudomonas aeruginosa* PAO1

Abstract. We purified to homogeneity an intracellular esterase from the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. The enzyme hydrolyzes *p*-nitrophenyl acetate and other acetylated substrates. The N-terminal amino acid sequence was analyzed and 11 residues, EPLILDAPNA, were determined. The corresponding gene PA3859 was identified in the *P. aeruginosa* PAO1 genome as the only gene encoding for a protein with this N-terminus. The encoding gene was cloned in *Escherichia coli,* and the recombinant protein expressed and purified to homogeneity. According to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis and analytical gel filtration chromatography, the esterase was found to be a monomer of approximately 24 kDa. The experimentally determined isoelectric point was 5.2 and the optimal enzyme activity was at 55 ºC and at pH 9.0. The esterase preferentially hydrolyzed short-chain fatty acids. It is inhibited by phenylmethylsulfonyl fluoride (PMSF) but not by ethylendiaminotetraacetic acid (EDTA). Native enzyme preparations typically showed a Michaelis constant (K_m) and V_{max} of 0.43 mM and 12,500 U mg^{-1} respectively, using *p*-nitrophenyl acetate as substrate. Homology-based database searches clearly revealed the presence of the consensus GXSXG signature motif that is present in the serine-dependent acylhydrolase protein family.

This chapter was adapted from:
Pesaresi A, Devescovi G, Lamba D, Venturi V, Degrassi G.
Isolation, characterization, and heterologous expression of a carboxylesterase of Pseudomonas aeruginosa PAO1.
Curr Microbiol. 2005 Feb; 50(2):102-9.

2.1 Introduction

Esterases catalyzing the hydrolysis of either aliphatic or aromatic ester bonds are widely distributed in nature and occur in most living organisms. They have been the subject of several studies due to their potential use in applied and medical biotechnology and also for the elucidation of their metabolic function, which is mostly unknown. The carboxylesterase family EC (3.1.1.1.) comprises a group of esterases hydrolyzing carboxylic ester bonds with relatively broad substrate specificity. They show a high degree of sequence similarity and are believed to be involved in the detoxification of xenobiotics [10], particularly in the detoxification of pesticides [11]. Most of the carboxylesterases have been studied from mammalian tissues, confirming their important role in the hydrolytic biotransformation of a vast number of drugs [22].
Since a significant number of drugs are metabolised by carboxylesterases, the alteration of this enzymatic activity can have important clinical implications. Indeed, the structural basis of heroin, cocaine, and organophosphate chemical weapon metabolism by a promiscuous human carboxylesterase has recently been elucidated [15].
Microbial esterases are also widely studied because of their potential industrial importance. A novel esterase from *Burkholderia gladioli* has been reported showing a high deacetylation activity on cephalosporin derivatives [19], and a thermostable carboxylesterase from *Sulfolobus solfataricus* has been characterized and the gene cloned [17]. Another carboxylesterase gene from *Acinetobacter lwoffii* has also been cloned and sequenced [14].
Pseudomonas spp. seems to be an important source of these enzymes since the presence of carboxylesterases has been reported in strain KWI-56 [24], in *Pseudomonas fluorescens* [9, 12], and in *Pseudomonas citronellolis* [3].
Pseudomonas aeruginosa is an ubiquitous environmental bacterium and is one of the three major causes of opportunistic human infections. The importance of *P. aeruginosa* as a pathogen is due to its intrinsic resistance to antibiotics and disinfectants. The *P. aeruginosa* PAO1 genome is the largest bacterial genome sequenced so far [23] and the sequence provides insights into its versatility and intrinsic drug resistance. It has been proposed that the size and complexity of the *P. aeruginosa* genome reflect an evolutionary adaptation to diverse environments, with a large number of genes involved in the catabolism of organic compounds as well as in the modification of anti-microbial substances.
Here is reported the purification and characterization of a carboxylesterase from *P. aeruginosa* PAO1. Most of the presently available biochemical data have proved that these hydrolytic enzymes have several features in common, such as a similar mechanism of hydrolysis and a catalytic triad consisting of serine, histidine, and aspartic acid residues, with the serine residue located in the highly conserved motif

GXSXG [1, 2]. A large percentage of esterases show sequence similarity to proteins with identified biochemical activity, but with uncertain physiological function. The integration of structural, bioinformatic, and enzymology data is expected to greatly enhance function determination, as has been recently reported for a new carboxylesterase from *E. coli* [21].
The identification of the correspondent gene from the genome sequence, cloning in *E. coli*, and the purification protocol of the recombinant protein are aimed at developing a system for heterologous expression and purification of large quantities of this enzyme, not only for its potential biotechnological application but also for its 3D crystal structure determination. The enzyme was also characterized from a biochemical point of view and compared to similar and previously reported proteins with the same activity.

2.2 Materials and Methods

2.2.1 Growth conditions, enzyme assay and purification of *P. aeruginosa* PA3859

P. aeruginosa PAO1 was grown aerobically overnight at 37 ºC in Luria–Bertani (LB) medium [20]. Acetyl esterase activity was measured using *p*-nitrophenyl acetate (Sigma Chemical Co., St. Louis, MO) as substrate at a concentration of 0.5 mM. Release of *p*-nitrophenol was detected spectrophotometrically at 400 nm at 25 ºC.
One unit of esterase activity was defined as the release of 1 μmole of *p*-nitrophenol per minute under these conditions. The enzyme was purified from the cell-free crude extract prepared from 1.6 L of culture. Cells were lysed as follows: the pellet was resuspended in 40 ml lysis buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 0.1 % Triton X-100, 1 mg ml^{-1} lysozyme), kept on ice for 1 h, and sonicated (three times at maximum power for 1 min each) using a Soniprep 150 sonicator (Sanyo). Cell walls were removed by centrifugation (12,000g, 15 min). Clear lysate was fractionated by $(NH_4)_2SO_4$ and the fraction precipitating from 30 to 80% saturation was recovered. This pellet was resuspended in 20 mM Bis-Tris, pH 6.5, and further dialyzed against the same buffer to eliminate residual ammonium sulfate. The sample was applied to a Q Sepharose FF column (Amersham Biosciences); the column was washed with 40 ml of 20 mM Bis-Tris, pH 6.5, and a 200 ml linear gradient from 0 to 0.4 M NaCl at a flow rate of 3 ml min^{-1} was applied. The collected fractions were tested for esterase enzymatic activity. Active fractions were pooled and the concentration of NaCl reduced to 0.2 M by dialysis against 20 mM Bis-Tris, pH 6.5, 0.2 M NaCl. The protein was loaded again onto the same column and eluted using a narrower gradient from 0.2 to 0.4 M NaCl in 200 ml. Active fractions were pooled, concentrated, dialyzed, and resuspended in 100 mM Na phosphate, pH 7.0, 1.7 M $(NH_4)_2SO_4$ before loading onto

phenyl Sepharose HP column (Amersham Biosciences). The column was washed with 40 ml of 100 mM Na phosphate, pH 7.0, 1.7 M ammonium sulfate before eluting the protein with a 240 ml linear gradient from 1.7 to 0 M ammonium suflate. The flow rate was 3 ml min^{-1} and fractions were 5 ml each. Active fractions were pooled, dialyzed against 100 mM Na phosphate, pH 7.0, and concentrated to 1 ml. The sample was loaded onto a gel filtration column (Sephacryl HR200, column XK16; Amersham Biosciences) previously equilibrated with 100 mM Na phosphate pH 7.0, 150 mM NaCl. Proteins were eluted at a flow rate of 0.5 ml min^{-1} and fractions of 2.5 ml were collected. The column was calibrated with an MW-GF-200 kit (Sigma).

2.2.2 Determination of molecular weight and pI

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli [16] using 12% acrylamide. Protein bands were stained with Coomassie Brilliant Blue R250 (Sigma). The broad range molecular weight marker kit (New England Biolabs) was used to estimate the relative molecular weight (Mr) of the purified protein under denaturing conditions. Gel filtration chromatography using a Superdex 75 column (Amersham Biosciences) was carried out to determine the molecular mass of the purified enzyme, using MWGF- 200 molecular weight markers for gel filtration (Sigma Chemicals). Analytical isoelectric focusing of the purified enzyme was performed with an Ampholine PAGplate precast polyacrylamide gel (Amersham Biosciences), with pH values ranging from 3 to 10, and the broad pI calibration kit (Amersham Biosciences) as pI marker.

2.2.3 Kinetics

The kinetic parameters K_m and V_{max} were determined in 30 mM Tris-HCl, pH 9.0, at 25 ºC over the substrate concentration range from 0.01 to 5 mM *p*-nitrophenyl acetate.

2.2.4 Temperature and pH effects

The optimal pH and temperature were determined in the range from pH 3 to pH 11 and from 15 to 85 ºC. For the pH stability determination, samples were incubated in buffers from pH 2.0 to pH 13 at 4 ºC for 24 h. For the determination of thermal stability, protein samples were kept at temperatures of 30, 40, 50, 60, 70, and 80 ºC and at pH 9 for 2 h, respectively, and the residual activity was detected as described above.

2.2.5 Other assays

Inhibition by PMSF, dithiothreitol (DTT), and EDTA at various concentrations was investigated by incubating 10 μg of the purified enzyme at 25 ºC in 1 ml of 30 mM Tris-HCl pH 9, for 5 min. The reaction was stopped by chilling on ice and aliquots were assayed with *p*-nitrophenyl acetate in order to determined the residual activity.
For the determination of substrate specificity, the following substrates were tested: α-naphthyl acetate, α-naphthyl butyrate, α-naphthyl caprylate, α-naphthyl laurate, and α-naphthyl oleate (Sigma Chemicals). Other substrates used were methylumbelliferyl acetate, fluorescein diacetate, and carboxyfluorescein diacetate (Sigma Chemicals). Esterase activity was assayed spectrophotometrically at 25 ºC. The release of α-naphthol was measured at 560 nm after the addition of 1 mg ml^{-1} Fast Garnet GBC (Sigma Chemicals) (stock solution, 5 mg ml^{-1} in 10% SDS). Methylumbelliferone was detected at 354 nm, fluorescein at 470 nm, and carboxyfluorescein at 490 nm. The reaction mixtures contained 30 mM Tris-HCl, pH 9.0, 0.5 mM substrate, and 2 μg of enzyme. One unit of esterase is defined as the amount of enzyme required to liberate 1 μmol of *p*-nitrophenol min^{-1}.
The effect of various metallic ions on the purified enzyme was tested. The activity was assayed at 25 ºC in 30 mM Tris-HCl, pH 9.0, 5 mM *p*-nitrophenyl acetate, and 1 mM cation.

2.2.6 Amino acid analysis and gene identification

The purified protein was resolved by 12% acrylamide SDS-PAGE, blotted onto a polyvinylidene difluoride (PVDF) membrane, and subjected to N-terminal amino acid sequence analysis by automated Edman degradation on a pulsed liquid-phase protein sequencer (Model 470A; Applied Biosystems) equipped with an on-line phenylthiohydantoin amino acid analyzer (Model 120A; Applied Biosystems).
(The protein sequencing was performed by J. Keen, University of Leeds, UK).

2.2.7 Cloning

The carboxylesterase gene from *P. aeruginosa* used in this study was amplified by PCR using two oligonucleotides (BamHI-start, 5′-AATGGATCCAAAGAGGAGAAATCC GTCATGAGCGAACCC; and HindIII-stop, 5′-GTCTGAGAGAAGCTTCGGGCGTC AGAGG).

The PCR product was cloned as a BamHI–HindIII fragment into the corresponding sites of pBlueScriptSK(+). The resulting plasmid was transformed into *E. coli* DH5α and the PCR product sequenced.
To obtain the His-tagged protein, the carboxylesterase gene was amplified by PCR using the oligonucleotides (Bam EST, 5′-CGCGGATCCAGCGAACCCCTGATCCTC GA; and EcoHind EST, 5′-CGGAATTCAAGCTTCAGAGGCGCTTGCGCAG). Once the sequence was confirmed, the PCR product was cloned as a BamHI–HindIII fragment into the corresponding sites of pQE30, and *Escherichia coli* M15 (pREP4) (Qiagen) transformed with the resulting plasmid. The recombinant protein carries an 11 amino acid residue extension at the N-terminus (MRGSHHHHHHG–SSEPLI) containing a histidine tag.

2.2.8 Purification of the *P. aeruginosa* PA3859 from *E. coli*

Affinity chromatography purification of the 6xHis-tagged protein was achieved from the crude extract of 1 L of *E. coli* pQE-31 culture grown in LB to OD_{600} 0.6 and then induced with 0.2 mM IPTG for 6 h at 25 ºC. For preparation of cell-free crude extract, bacterial cells were resuspended in lysis buffer (50 mM Na phosphate, pH 8.0, 500 mM NaCl, 10 mM imidazole, 1 mg ml^{-1} lysozyme), incubated on ice for 30 min, and sonicated. The lysate was centrifuged at 10,000g for 30 min at 4 ºC to remove the debris, and was loaded onto a 5-ml pre-packed HiTrap Chelating HP affinity column (Amersham Biosciences) previously equilibrated with 30 mM Na phosphate, pH 7.5, 500 mM NaCl, 20 mM imidazole. Nickel was used as chelating agent and immobilized onto the column following the manufacturer's instructions. After washing with 14 column vol of equilibration buffer, the protein was eluted by 8 column vol of a linear gradient, increasing the concentration of imidazole from 20 to 500 mM.
The protein fraction showing activity was further resolved by ion-exchange Q Sepharose FF16/10 chromatography (Amersham Biosciences). The resulting peaks were analyzed by electron ion spray–mass spectrometry (EIS-MS) using an API 150EX mass spectrometer (Applied Biosystem).
(Mass spectrometry analysis was performed by C. Guarnaccia, ICGEB, Trieste).

2.3 Results

2.3.1 Enzyme purification and characterization

The *P. aeruginosa* carboxylesterase is expressed in a rich medium such as LB. The purification of this enzyme is summarized in Table 2.1. After Q Sepharose FF column chromatography, a major protein peak showing esterase activity on *p*-nitrophenyl acetate as the substrate was eluted during the gradient at 300 mM NaCl. Approximately 20% of the total activity was found in another peak eluting at 180 mM NaCl. After gel filtration chromatography, only a single band of approximately 28 kDa was detected on a SDS–12% polyacrylamide gel (Fig. 2.1). The molecular mass of the purified protein estimated by Superdex 75 10/30 (Amersham Biosciences) gel filtration chromatography was 19.3 kDa. These results suggest the purified protein to be a monomer. The pI of the protein was estimated by isoelectric focusing to be approximately 5.2 (Fig. 2.1), in agreement with the theoretically predicted pI value of 5.0.

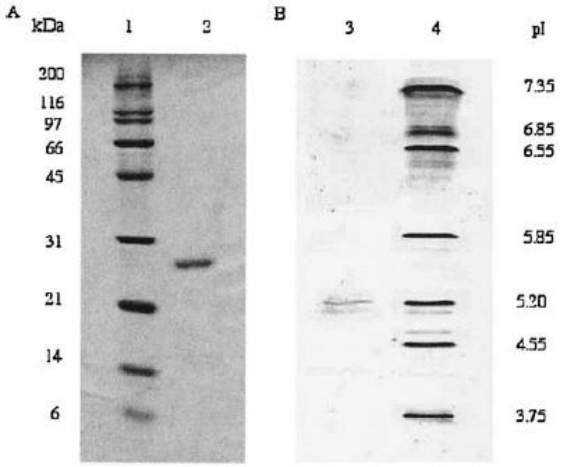


Figure 2.1. SDS-PAGE (**A**) and analytical isoelectric focusing (**B**) of the purified carboxylesterase. Lanes: 1, molecular mass standard; 2, 2,5 μg of enzyme; 3, 2.5 μg of enzyme; 4, pI markers.

The optimal pH and temperature were found to be 9.0 and 55 ºC, respectively (Fig. 2.2). The enzyme was found stable in the pH range between 4.0 and 8.0, and when incubated at different temperatures for 2 h, at pH 9.0, it showed to be stable up to 50 ºC (Fig. 2.2).

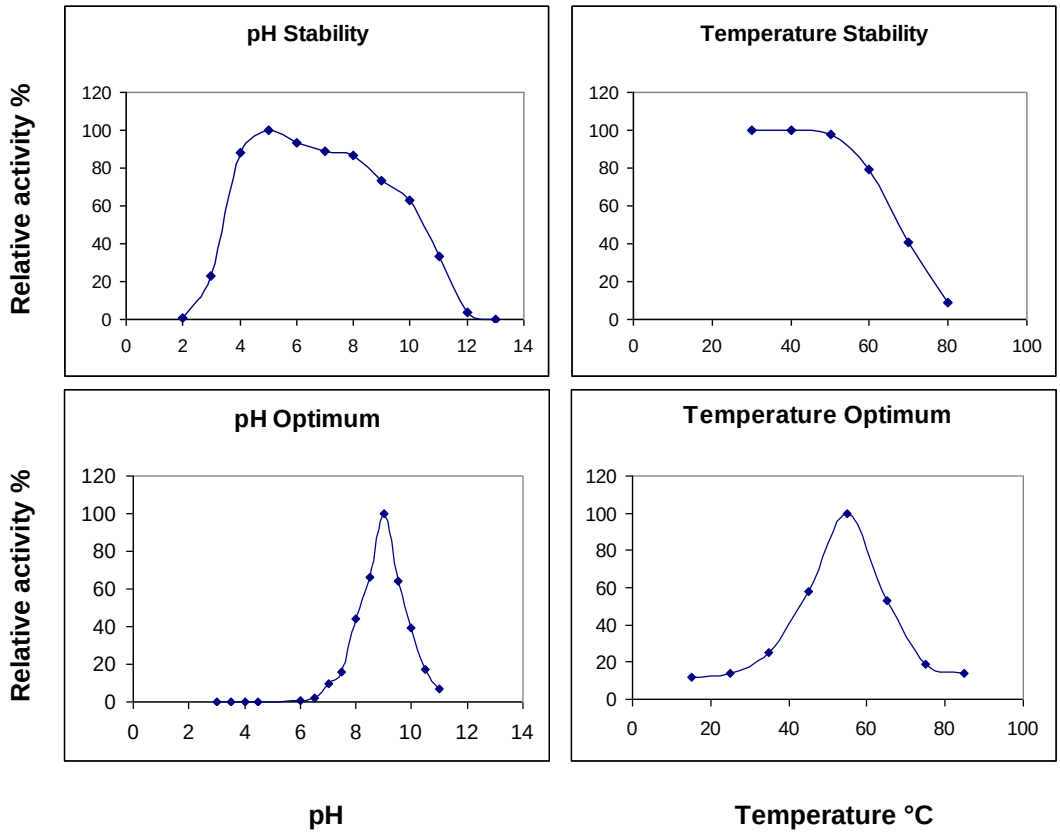


Figure. 2.2. Temperature stability, pH stability, optimal temperature, and optimal pH of the purified carboxylesterase.

The activity of PAO1 carboxylesterase was determined by measuring the hydrolysis of *p*-nitrophenyl acetate. The Lineweaver–Burk plot showed a linear response over the concentration range from 0.01 to 5 mM at pH 9.0 and at 25 ºC. The Michaelis constant K_m for *p*-nitrophenyl acetate was 0.43 ± 0.02 mM and the maximal velocity V_{max} was 12,500 ± 500 U mg).

The effect of some esterase inhibitors was also detected and is shown in Table 2.2. The enzyme was completely inhibited by phenylmethylsulfonyl fluoride (PMSF) at a concentration above 1 mM, thus suggesting the enzyme to be a serine-dependent acylhydrolase.

Dithiothreitol gave a lower inhibition at 1 mM, while EDTA did not have any effect even at a higher concentration (10 mM).

Enzyme specificity was tested at pH 9.0 and 25 ºC on α-naphthyl esters with different chain lengths at 5 mM concentration. Results reported in Table 2.3 showed that the activity rapidly decreased with the chain length of the acyl group.

The kinetic parameter k_{cat} was determined using some common acetylated substrates and the values are listed in Table 2.4.

The activity of the purified enzyme was also tested in the presence of various cations and the results are shown in Table 2.5. The carboxylesterase was sensitive to Co^{2+}, Cu^{2+}, Fe^{2+}, Fe^{3+}, Ni^{2+}, and Zn^{2+} metal ions.

Purification step[a]	Total protein (mg)	Total activity (U)b	Specific activity (U/mg)	Purification factor	Yield (%)
Cell-free crude extract	777.6	18,900	24	1	100
(NH4)2SO4 fractionation	516.6	17,200	33	1.37	91
Q Sepharose	45	14,700	320	13.3	77
Phenyl Sepharose	6	10,100	1,680	70	53
Sephacryl HR200	0.46	3,360	7,310	304	18

Table 2.1. Summary of the purification of carboxylesterase from 1.6 1 of *Pseudomonas aeruginasa* PAO1.
[a]See Materials and Methods for details.
[b]Measured with p-nitrophenyl acetate as the substrate.

Inhibitor	Concentration (mM)	Relative activity (%)
PMSF	0.01	94
	0.1	48
	1	1.5
DTT	0.1	84
	1	58
EDTA	1	100
	10	100

Table 2.2. Effects of inhibitors on carboxylesterase

α-Naphtyl ester	Relative activity (%)
α-Naphtyl acetate	100
α-Naphtyl butyrate	54
α-Naphtyl caproylate	14
α-Naphtyl laurate	9
α-Naphtyl oleate	2

Table 2.3. Influence of acyl chain length on carboxylesterase activity.

Substrate	K_{cat} (s^{-1})
α-Naphtyl acetate	3.0×10^3
α-Naphtyl butyrate	1.6×10^3
p-Nitrophenyl acetate	6.4×10^3
p-Nitrophenyl butyrate	1.7×10^3
4-Methylumbelliferyl acetate	8.5×10^3
Fluorescein diacetate	1.7×10^3
Carboxyfluorescein diacetate	2.8×10^3

Table 2.4. Kinetic parameters of purified carboxylesterase.

Cation	Relative activity (%)
$CaCl_2$	105
$CoCl_2$	34
$CuCl_2$	59
$FeCl_2$	40
$FeCl_3$	38
$MgCl_2$	112
$NiCl_2$	35
$ZnCl_2$	36

Table 2.5. Cation effects on carboxylesterase activity.

2.3.2 Identification of the PA3859 gene

The N-terminal amino acid analysis of the purified protein identified 11 amino acids: SEPLILDAPNA. The only ORF in the genome of *P. aeruginosa* PAO1 with this N-terminus sequence is PA3859, which was annotated as a probable carboxylesterase due to its similarity to carboxylesterase I (76%) and carboxylesterase II (75%) of *P. fluorescens*.

2.3.3 Heterologous protein purification

The recombinant enzyme was eluted from a HiTrap Chelating affinity column following a gradient, at a concentration between 220 and 250 mM imidazole. After the HiTrap Chelating column, the heterologous 6xHis carboxylesterase showed specific activity lower than the carboxylesterase purified from *P. aeruginosa* PAO1 (Tables 2.1 and 2.6). Due to the presence of some contaminant proteins and to the need of highly purified protein for structural and functional studies, the enzyme was further purified by ion-exchange Q Sepharose chromatography. Three peaks were identified and further

analyzed by EIS-MS: the first corresponded to a protein molecular mass of 24,691 Da and showed a specific activity in agreement with that of enzyme from *P. aeruginosa*; the second peak, corresponding to a protein molecular mass of 23,983 Da, showed only 10% of the expected specific activity; and the third peak corresponded to a protein molecular mass of 25,239 Da and did not show esterase activity.
Protein with a molecular mass of 24,691 Da was recovered, with an overall yield of 6.2 mg L^{-1} of cell culture (Table 6).

Purification step	Total protein (mg)	Total activity (U)	Specific activity (U/mg)	Purification factor	Yeld (%)
Cell-free crude extract	962	96130	100	1	100
Affinity chromatography	16.6	83730	5044	50	87
Q Sepharose	6.2	81840	13200	132	85

Table 2.6. Summary of purification of recombinant carboxylesterase from 1 L of *E. coli* pQE31 culture.

2.4 Discussion

As part of an ongoing research project aimed at the structural and functional characterization of biotechnological and biomedical relevant microbial enzymes, we have recently reported the identification of genes encoding for esterase activity. Namely, in the yeast *Saccharomyces cerevisiae* [4] the function of an esterase has been identified on the basis of its high sequence identity (41%) to S-formylglutathione hydrolase protein of *Paracoccus denitrificans*, known to be associated with the detoxification of formaldehyde [8].
Hemicellulose substrates (corncob powder or oat spelt xylan) supplied as the only carbon source in the growth medium induced an esterase activity in *B. pumilus*, which has been identified to be an acetylxylan esterase [5].
We now report on the isolation, purification, and identification of an enzyme produced by the *P. aeruginosa* PAO1 opportunistic pathogen showing esterase activity. Its N-terminal sequence revealed that this enzyme corresponds to the gene product of the open reading frame PA3859, the only one among the 15 possible acylhydrolases present in the PAO1 genome that has been annotated as a probable carboxylesterase (http://www.pseudomonas.com/index.html).
During the purification procedure one major peak of activity towards *p*-nitrophenyl acetate was identified which accounted for more than 70% of the total esterase activity of the crude extract. Part of the residual activity was found in another peak of the ion

exchange chromatography step, suggesting the presence of other esterases in *P. aeruginosa*, in accordance with the possible esterases reported in the PAO1 genome and with the four esterase activities reported in *P. fluorescens* [9].

The biochemical characterization of this enzyme showed that it hydrolyzes short-chain fatty acid esters with broad substrate specificity and is not inhibited by EDTA. Its hydrolytic activity is inhibited by PMSF and reduced by transition metal ions, suggesting that the catalytic histidine may be involved in their coordination shell.

PA3859 shows 20% identity with respect to a recently characterized carboxylesterase from *P. citronellolis* [3], which has a higher molecular mass of 41 kDa, a lower optimal temperature of 37–40 ºC, and a kcat for *p*-nitrophenyl acetate and *p*-nitrophenyl butyrate 80- and 40-fold lower, respectively.

A multiple sequence alignment, following a BLAST search based on the PA3859 sequence, is shown in Fig. 2.3 and includes proteins representative of diverse bacterial and eukaryotic species with sequence identity ranging between 25 and 65%. Sequence analysis revealed the presence of the conserved Ser His Asp catalytic triad and the serine-dependent acylhydrolase signature motif GXSXG [2].

The *in vivo* function has only been identified for the human acyl-protein thioesterase I (sp|O75608|) and involves the S-palmitoylation of target cysteine residues [7].

Three-dimensional structural information is presently available only for the human acyl-protein thioesterase (sp|O75608|). (PDB ID 1FJ2) [6] and for *P. fluorescens* carboxylesterase II (sp|Q53547|) (PDB ID 1AUO) [13]. They show 33% sequence identity and the atomic model reveals profound similarities. Both enzymes exist in solution and in solid state as homodimers.

In their respective crystal structures the active sites are unexpectedly occluded from the bulk solvent. Dissociation of the homodimer, probably induced by substrate binding, has been proposed as the mechanism for enzymatic activity regulation. It is note worthy that *P. aeruginosa* carboxylesterase exists in solution as an active monomer. Further structural and functional studies are required to elucidate the biological significance of this finding. The high sequence homology, 76%, found between the carboxylesterases from *P. fluorescens* and *P. aeruginosa* suggests that a high structural similarity should also exist between the human acyl-protein thioesterase and PA3859, and therefore a similar *in vivo* function is expected. The reported heterologous expression system has proven to be useful for the production of multimilligrams of recombinant *P. aeruginosa* carboxylesterase suitable for both structural and functional studies aimed at addressing some of the questions raised.

```
PA3859                  MSEPLILD.A PNADACIIWL HGLGADRTDF KPVAEALQMV LPSTRFILPQ
sp|Q53547|              MTEPLILQPA KPADACVIWL HGLGADRYDF MPVAEALQES LLTTRFVLPQ
sp|O75608|              TPLPAIVPAA RKATAAVIFL HGLGDTGHGW AEA.FAG.IR SSHIKYICPH
LmjF24.1840             RYGTLLEIGN KNPTGVVTLV HGLGDSAYGW ESVGHELLRR LPHLLFLLPT
G_YK2298                NGNPSIVSPR GEHKGTLIFL HGLGDQGHGW ADA.FKTEAK HDNIKFICPH
sp|P76561|              HDHFVVQSPD KPAQQLLLLF HGVGDNPVAM GEIGSWFAPL FP.DALVVSV
ttk003001929            .MSFIHRFEP GTSDTTLLLL HGTGGNEHDL IGLARDLSP. ..TAKLLSPR

PA3859                  APSQAVTVNG GWVMPSWYDI LAFSPAR... .AIDEDQLNA .SADQVIALI
sp|Q53547|              APTRPVTING GYEMPSWYDI KAMSPAR... .SISLEELEV .SAKMVTDLI
sp|O75608|              APVRPVTLNM NVAMPSWFDI IGLSPDS... .QEDESGIKQ .AAENIKALI
LmjF24.1840             APSRSVTING GMPMPAWYDI MDMCNSGLLR GRQDAASVRQ .SCDYVRSIA
G_YK2298                SSERPVTLNM GMRMPAWFDL FGLDPNA... .QEDEQGINR .ATQYVHQLI
sp|P76561|              GGAEPSGNPA GRQ...WFSV QGITEDN... .RQARVDAIM .PT.FIETVR
ttk003001929            G...KVLENG A...PRFFRR LAMGVF.... ...DEADLKA QAADLARFVQ

PA3859                  D.EQRAKGIA AERIILAGFS QGGAVVLHTA FRRYAQPLGG VLALSTYAPT
sp|Q53547|              E.AQKRTGID ASRIFLAGFS QGGAVVFHTA FINWQGPLGG VIALSTYAPT
sp|O75608|              D.QEVKNGIP SNRIILGGFS QGGALSLYTA LTTQQ.KLAG VTALSCWLPL
LmjF24.1840             HVATKKYGIP PQRVVYSGFS QGAAISLCTG LTAHIAP.AG IACMSGYLAA
G_YK2298                D.AEVAAGIP ASRIAVGGFS MGGALAIYAG LTYPQ.KLGG IVGLSSFFLQ
sp|P76561|              Y.WQKQSGVG ANATALIGFS QGAIMALESI KAEPGL.ASR VIAFNGRYAS
ttk003001929            .EAAERYSFD AGRVYALGYS NGANIAAALL LLHPEVL.AG GVLLRPVLPL
                                           *

PA3859                  FD.DL..ALD ERHKRIPVLH LHGSQDDVVD PALGRAAHDA LQAQG..VEV
sp|Q53547|              FGDEL..ELS ASQQRIPALC LHGQYDDVVQ NAMGRSAFEH LKSRG..VTV
sp|O75608|              RASFPQGPIG GANRDISILQ CHGDCDPLVP LMFGSLTVEK LKTLVNPANV
LmjF24.1840             CTDVL..P.R IVQKAVPITM FHGRQDPVVP ISAAKETKEI LEKDGGVAPI
G_YK2298                RTKFPGS..F TANNATPIFL GHGTDDFLVP LQFGQMSEQY IKKF..NPKV
sp|P76561|              LPE.......  TASTATTIHL IHGGEDPVID LAHAVAAQEA LISAG..GDV
ttk003001929            EPPTLP...D LS..GKAVFL AAGRRDPWAP SERVEALARW LEQAG..ANV
                                                    *

PA3859                  GWHDYP.MGH EVSLEEIHDI GAWLRKRL.
sp|Q53547|              TWQEYP.MGH EVLPQEIHDI GAWLAARLG
sp|O75608|              TFKTYEGMMH SSCQQEMMDV KQFIDKLL.
LmjF24.1840             SFLEYD.MDH STLPQEIDDI TSFLSRVL.
G_YK2298                ELHTYRGMQH SSCGEEMRDV KTFLSAHI.
sp|P76561|              TLDIVEDLGH AIDNRSMQFA LDHLRYTI.
ttk003001929            ELRWQ.DAGH ELHPEELEAA RAWLARWV.
                                *
```

Fig. 2.3. Sequence alignment of PA3859 and six selected proteins: sp|Q535487|, carboxylesterase 2 (esterase II family) from *Pseudomonas fluorescens*; sp|O75608|, acyl-protein thioesterase 1 (lysophospholipase I family) from *Homo sapiens*; LmjF24.1840, hypothetical protein from *Leishmania major*; G_YK2298, from *Caenorhabditis elegans*; sp|P76561|, hypothetical hydrolase from *Escherichia coli*; Ttk003001929, from *Thermus thermophilus*, obtained with T-Coffee [18]. The residues of the conserved catalytic traid are marked with an asterisk and the residues of the signature motif are boxed.

2.5 References

[1] Brady L, Brzozowski AM, Derewenda ZS, et al. A serineprotease triad forms the catalytic centre of a triacylglycerol lipase. *Nature* (1990) 343:767–770.

[2] Brenner S. The molecular evolution of genes and proteins: a tale of two serines. *Nature* (1988) 334:528–530.

[3] Chao YP, Fu H, Wang YL, Huang WB, Wang JY. Molecular cloning of the carboxylesterase gene and biochemical characterization of the encoded protein from *Pseudomonas citronellolis* ATCC 13674. *Res Microbiol*. (1988) 154:521–526.

[4] Degrassi G, Uotila L, Klima R, Venturi V. Purification and properties of an esterase from the yeast *Saccharomyces cerevisiae* and identification of the encoding gene. *Appl Environ Microbiol*. (1999) 65:3470–3472

[5] Degrassi G, Kojic M, Ljubijankic G, Venturi V. The acetyl xylan esterase of *Bacillus pumilus* belongs to a family of esterases with broad substrate specificity. *Microbiology*. (2000) 146:1585–1591.

[6] Devedjiev Y, Dauter Z, Kuznetsov SR, Jones TL, Derewenda ZS. Crystal structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 A. *Struct Fold Des* (2000) 8:1137–1146.

[7] Duncan JA, Gilman AG. A cytoplasmatic acyl-protein thioesterase that removes palmitate from G protein alpha subunits and p21 (RAS). *J BioI Chern* (1998) 273:15830–15837.

[8] Harms N, Ras J, Reijnders WNM, van Spanning RJM, Stouthhamer AH S-Formylglutathione hydrolase of Paracoccus denitrificans is homologous to human esterase D: A universal pathway for formaldehyde detoxification. *J Bacteriol* (1996) 178:6296–6299.

[9] Hong KH, Jang WH, Choi KD, Yoo OJ Characterization of Pseudomonas fluorescens carboxylesterase: Cloning and expression of the esterase gene in Escherichia coli. *Agr Biol Chem* (1991) 55:2839–2845.

[10] Jakoby WE, Ziegler DM. The enzymes of detoxification. *J Biol Chem* (1990) 265:20715–20718.

[11] Johnson LM, Talbot HW (1983) Detoxification of pesticides by microbial enzymes. Experientia 39:1236–1246.

[12] Kim YS, Lee HE, Choi KD, Park S, Yoo OJ. Cloning of *Pseudomonas fluorescens* carboxylesterase gene and characterization of its product expressed in *Escherichia coli*. *Biosci Biotech Biochem* (1994) 58:111–116.

[13] Kim KK, Song HK, Shin DH, Hwang KY, Choe S, Yoo OJ Sub SW. Crystal structure of carboxylesterase from Pseudomonas fluorescens, an alpha/beta hydrolase with broad substrate specificity. *Structure* (1997) 5:1571–1584.

[14] Kim HE, Lee IS, Kim JH, Hahn KW, Park VJ, Han HS, Park KR. Gene cloning, sequencing, and expression of an esterase from *Acinetobacter lwoffii* I6C-l. *Curr Microbiol* (2003) 46:291–295.

[15] McRee D. Protecting against cocaine, heroin, and sarin gas. *Chem Biol* (2003) 10:295–297.

[16] Laemmli VK. Cleavage of structural protein during the assembly of the head of bacteriophage T4. *Nature* (1970) 227:680–685.

[17] Morana A, Di Prizito N, Aurilia V, Rossi M, Cannio R. Acarboxylesterase from the hyperthermophilic archeon *Sulfolobus solfataricus*: Cloning of the gene, characterization of the protein. *Gene* (2002) 283:107–115.

[18] Notredame C, Higgins D, Heringa J. T-Coffee: A novel method for multiple sequence alignments. *J Mol Biol* (2000) 302:205–217.

[19] Petersen EI, Valinger G, Solkner B, Stubenrauch G, Schwab H. A novel esterase from Burkholderia gladioli which shows high deacetylation activity on cephalosporins is related to betalactamases and DD-peptidases. *J Biotechnol* (2001) 89:11–25.

[20] Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Press (2001).

[21] Sanishvili R, Yakunin AF, Laskowski RA, Skarina T, Evdokimova E, Doherty-Kirby A, Lajoie GA, Thornton JM, Arrowsmith CH, Savchenko A, Joachimiak A, Edwards AM. Integrating structure, bioinformatics, and enzymology to discover function: BioH, a new carboxylesterase from *Escherichia coli*. *J Biol Chem* (2003) 278:26032–26045.

[22] Satoh T, Hosokawa M. The mammalian carboxylesterases: From molecules to functions. *Annu RevPharmacol Toxicol* (1998) 38:257–288.

[23] Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-L Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Urn R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* (2000) 406:959–964.

[24] Sugihara A, Shimada Y, Nagao T, Iizumi T, Nakamura K, Fukase T, Tominaga Y. Purification and characterization of a carboxylesterase from *Pseudomonas* sp. KWI-56. *Biosci Biotechnol Biochem* (1994) 58:752–755.

Chapter 3

Crystallization, X-ray diffraction analysis and phasing of carboxylesterase PA3859 from *Pseudomonas aeruginosa* PAO1

Abstract. We have recently purified an intracellular carboxylesterase encoded by the open reading frame PA3859 of *Pseudomonas aeruginosa*. Among proteins showing a significant sequence homology with PA3859, the *in vivo* function is only known for the human acyl-protein thioesterase I that is involved in the deacylation of Gα proteins. The crystal structure determination of *P. aeruginosa* carboxylesterase is expected to provide insights into its physiological role. Therefore, the PA3859 gene was cloned and heterologously expressed in *Escherichia coli* as N-terminally 6xHis tagged recombinant protein. Here, we present the crystallization, X-ray diffraction analysis and phasing of this enzyme. Two crystal forms were obtained by the hanging drop vapor diffusion method. Crystals of form I belong to the space group $P2_1$ with cell dimensions of a =65.65, b =50.55, c =142.55 Å, b=92.9° and diffracted, upon flash annealing, up to a resolution of 2.9 Å.
Two dimers are present in the asymmetric unit. Crystals of form II belong to space group $P2_12_12$, with unit cell dimensions of a =96.42, b =96.36, c =68.04 Å and diffracted up to 2.1 Å resolution. One dimer is present in the asymmetric unit.

This chapter was adapted from:
Pesaresi A, Lamba D.
Crystallization, X-ray diffraction analysis and phasing of carboxylesterase PA3859 from Pseudomonas aeruginosa.
Biochim Biophys Acta. 2005 1752(2):197-201.

3.1 Crystallization, X-ray diffraction and phasing

Pseudomonas aeruginosa is an ubiquitous gram negative bacterium and is one of the major causes of opportunistic human infections [1]. The ability to degrade complex mixtures of organic compounds could provide this bacterium, characterized by the largest bacterial genome sequenced so far [2], with adaptive advantages to hostile environments.
This allows the *P. aeruginosa* to make use of a wide variety of organic compounds as carbon source, thus exploiting its exceptional ability to colonize ecological niches where nutrients are limited.
Among the enzymatic activities required to maintain such a complex catabolism, carboxylesterases are thought to play an important role. Indeed, the wide substrate tolerance shown by this class of enzymes has led to postulate that they have evolved to enable access to carbon sources [3].
Furthermore, carboxylesterases are known to be involved in the hydrolytic biotransformation of drugs [4] and in the detoxification of xenobiotics [5].
Carboxylesterases (E.C.3.1.1.1) are members of the serine esterase super-family [6] that hydrolyze carboxylic ester bonds with a relatively broad substrate specificity. They show a high degree of sequence similarity and share both conserved sequence motifs and secondary structure features with other members of the α/β hydrolase fold super-family [7].
We have recently reported the purification, the heterologous expression and the biochemical characterization of a novel 24 kDa esterase identified in *P. aeruginosa* PAO1 (see Chapter 2) [8]. The N-terminal sequence of the purified protein corresponded to the product of the open reading frame PA3859 that was already annotated as a putative carboxylesterase [http://www.pseudomonas.com].
In a genome-wide cloning exploration strategy and characterization of microbial esterases, PA3859 was among thirty open reading frames predicted to encode esterases and randomly selected from 13 sequenced bacterial chromosomes [9].
This enzyme, besides methyl-umbelliferyl acetate and fluorescein and carboxyfluorescein diacetate exhibited an ester-hydrolyzing activity also towards *p*-nitrophenyl and α-naphtyl esters of short fatty acid chains [8]. Overall, the biochemical characterization did not yet provide insights into its *in vivo* role. Furthermore, to our knowledge, the physiological substrates for most of the PA3859 homologues are also unknown. Among the proteins showing a significant sequence homology with PA3859, the *in vivo* function has been identified only for the human acyl-protein thioesterase I [10]. This enzyme shows a 35% sequence identity with *P. aeruginosa* carboxylesterase and was found to be involved in the deacylation of Gα proteins and other thioacyl protein substrates [11].

Structural information is available for human acylprotein thioesterase I (PDB ID 1FJ2) [10, 12] and for carboxylesterase II from *P. fluorescens* (PDB ID 1AUO)
[13, 12]. Besides the 33% identity in their primary sequence, they show important 3D similarities.
Both the human acyl-protein thioesterase I and the *P. fluorescens* carboxylesterase have been found to be present in solution as well in the solid state as active homodimers [10, 13]. Instead, PA3859 has been found to be present in solution as a monomer [8].
The high sequence homology of 76% between the *P. fluorescens* and *P. aeruginosa* carboxylesterases would suggest that close structural and functional similarities should also exist between PA3859 and the human acylprotein thioesterase I.
Despite their high sequence homology, PA3859 and *P. fluorescens* carboxylesterase II display to some extent a different substrate specificity. In fact, the former hydrolyzes methyl-umbelliferyl acetate but does not hydrolyze triacylglycerides, whereas the latter shows a reverse preference, indicating that the diversity in substrate specificity may in part reflect differences and/or flexibilities in the residues that encompass the catalytic sites and/or the neighbouring residues.
Furthermore, in the crystal structure of the *P. fluorescens* carboxylesterase II, a hydrogen bonding occurs at the dimer interface involving the main chain nitrogen of the catalytic histidine of one monomer and the carboxylate oxygen atoms of an aspartic residue which is adjacent to the catalytic aspartate in the facing subunit. Then, it is likely that dimerization may play an important role in determining the substrate specificity.
The aim of the present study on *P. aeruginosa* carboxylesterase is to provide the structural determinants that are expected to be essential in the elucidation of its substrate specificity as well as to shed light on its *in vivo* functional role.
As the first step, we describe here its crystallization, X-ray diffraction analysis and phasing.
PA3859 gene was cloned and the 6x-His tagged recombinant protein was expressed in *Escherichia coli* and purified as described in Chapter 2 [8].
Dynamic Light Scattering (DLS) has been proven to be a useful tool for detecting the aggregation state of proteins in solution prior to crystallization trials. DLS measurements were performed with a 2.2-mg ml^{-1} PA3859 solution in 50 mM NaCl, 200 mM ammonium sulphate, 100 mM MES pH 6.5, at 25 °C on a DynaPro MS/X instrument (Protein Solutions, USA). Samples were filtered through a 0.02-µm Anodisc membrane using the NanoFilter Kit (Wyatt Technology, USA) to remove large aggregates or dust particles and then injected in a 12-µl quartz MicroCUVETTE (Wyatt Technology, USA). Data collection and deconvolution were performed using the DYMAMICS 6.2.05 software (Protein Solutions, USA). DLS measurements performed on PA3859 indicated an hydrodynamic radius of 2.5 nm and a deduced molecular weight of 29.3 kDa, with a polydispersity of 25%. This result well agrees with the protein molecular weight obtained by electron ion spray-mass spectrometry of 24691

Da and estimated by reducing SDS-12% PAGE of 28 kDa, respectively [8]. Overall, DLS results indicate the existence in solution of monomer species.
Prior to crystallization screening, the purified protein was dialyzed against 10 mM sodium phosphate pH 7.5 and concentrated to 10 mg ml^{-1} using Centricon 3K concentrators (Millipore, USA).
Crystallization was performed by the hanging drop vapor diffusion method using the 24-well plates Crystallization Tool (Nextal, Canada). Initial crystallization conditions were established at 293 K using crystal screening kits Crystal Screen I and II (Hampton Research, USA) [14]. Each drop was prepared by mixing equal volumes (1 µl) of the protein and the precipitant buffer, and each hanging drop was equilibrated against 700 µl precipitant buffer.
Two different crystal forms were obtained from 200 mM $(NH_4)_2SO_4$, 100 mM MES pH 6.5, 30% w/v $PEG_{MME}5000$ (Fig. 3.1**A**, **B**). Long needles (form I) grew within 3 days. Squared bipyramidal crystals (form II) were obtained by adding to the same initial drop, 1 µl of milli-Q water. In this condition, only crystals of form II appeared within 7 to 10 days.

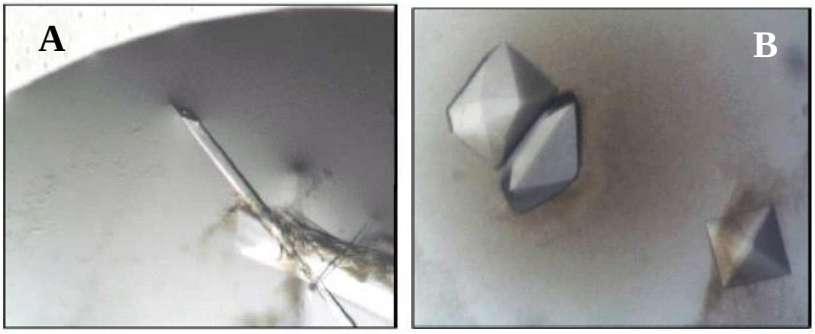


Figure 3.1. Crystals of PA3859 carboxylesterase. (**A**) Needles like: crystal form I; (**B**) Squared bipyramids: crystal form II.

For X-ray data collection, crystals were mounted on nylon loops and were directly flash cooled in a liquid nitrogen stream at 100 K, using an Oxford Cryosystems cooling device (Oxford Cryosystems Ltd., UK). A prior transferring to a cryo-protecting solution was not required for either crystal forms I and II (Fig. 3.2).
X-ray diffraction data were collected using a MAR CCD detector (MAR USA Inc., USA) at the XRD1 beam-line of the ELETTRA Synchrotron Light Laboratory (Trieste, Italy).
Crystal form I (0.2x0.2x0.8 mm) diffracted at first up to 5.0 Å resolution. It is worthy of note that a flash annealing step (crystal was warmed for a few second by blocking the

cold stream) produced a significant improvement in crystal diffraction up to 3.0 Å resolution [15,16]. Crystal form II (0.15x0.15x0.20 mm) diffracted to 2.1 Å resolution.
Data were indexed, integrated and scaled with DENZO and SCALEPACK [17] and further processed by the CCP4 package [18]. Data collection statistics are shown in Table 3.1. Assuming a molecular weight of 25 kDa, the unit cell volumes of the two PA3859 crystals forms are consistent with the occurrence of four (form I) and two molecules (form II) in their respective crystals asymmetric units. The Matthews coefficients are 2.4 $Å^3$ Da^{-1} (crystal form I) and 3.2 $Å^3$ Da^{-1} (crystal form II) corresponding to a solvent content of 48% and 62%, respectively [20].
The structures of either PA3859 crystal form I and II were determined by Patterson search techniques (data up to 3.5 Å resolution) using the monomer of the *P. fluorescens* carboxylesterase II (PDB ID 1AUO) [13,12] as a search model (76% amino acid sequence homology) after removal of all solvent molecules. Four (crystal form I) and two (crystal form II) rotation and translation function solutions were identified using the AMoRe suite [21] with a cumulative correlation coefficient and R factor of 0.52, 0.39 (crystal form I) and 0.49, 0.42 (crystal form II), respectively. The first noise peak had a cumulative correlation coefficient and R factor of 0.42, 0.43 (crystal form I) and of 0.24, 0.50 (crystal form II), respectively.
Rigid body refinement was subsequently performed using the CNS suite [22]. The R factor including reflections up to 2.98 Å (crystal form I) and 2.09 Å (crystal form II) was 0.33 (crystal form I) and 0.49 (crystal form II), respectively. In crystal form I, two non crystallographic dimers have been found in the asymmetric unit, while in crystal form II, one dimer is present.
This result is in contrast with the occurrence of a monomeric species in solution as prompted by the DLS analysis. Interestingly, dimers have been found in both solution and the crystal structures of human acyl-protein thioesterase I [11] and *P. fluorescens* carboxylesterase II [13].
A preliminary analysis of the crystal packing revealed that an almost identical dimer arrangement is observed in either crystal forms I and II that are characterized by the existence of two and four independent molecules in the respective crystals asymmetric units.
Instead, the observed dimer interface significantly differs from those being found in the crystal structures of the human acyl-protein thioesterase I [11] and *P. fluorescens* carboxylesterase II [13], respectively. The physiological implications of these findings are not clear.
Current efforts are focused on model building and structure refinement. The crystallization and the structure determination of PA3859 in complex with a variety of substrate analogs will be also attempted, leading to a structural-based elucidation of the enzyme mechanism of action and of its substrate binding specificities.

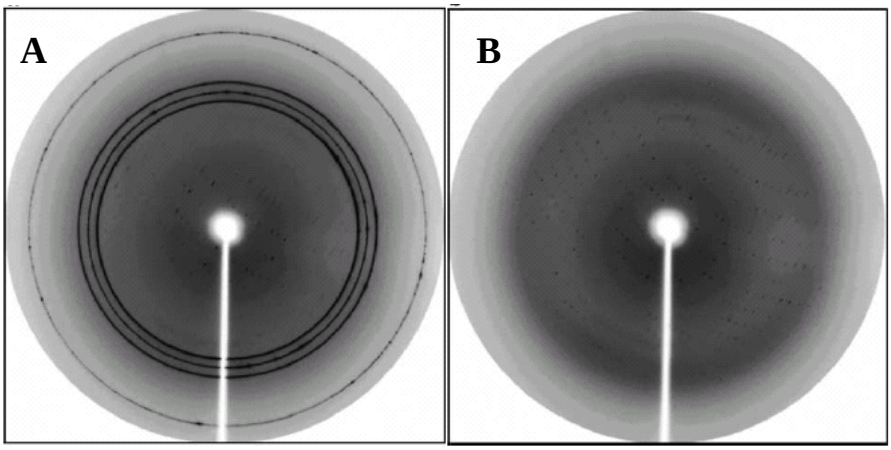


Figure 3.2. X-ray diffraction patterns of crystal form I before (**A**) and after (**B**) the flash annealing step.

Data set	Form I	Form II
X-ray source	ELETTRA-XRD1	ELETTRA-XRD1
Wavelength (Å)	1.200	1.278
Detector	marCCD	marCCD
Space group	$P2_1$	$P2_12_12$
Unit-cell parameters		
a (Å)	65.65	96.42
b (Å)	50.55	96.36
c (Å)	142.55	68.04
β (°)	92.94	-
Z	8	8
Total φ(°), oscillation range (°)	180, 1.0	240, 0.75
Mosaicity (°)	1.12	1.07
Resolution range (Å)	24.30-2.98 (3.00 – 2.98)	29.05 – 2.09 (2.10 – 2.09)
No. of measurements	204,325	1,219,326
No. of observed reflections $I \geq 0$	53,890 (1196)	197,535 (1085)
No. of unique reflections $I \geq 0$	18,812 (423)	36,512 (213)
Completeness (%)	98.3 (98.3)	96.7 (90.5)
Redundancy	2.9 (2.8)	5.4 (5.1)
$\langle I/\sigma(I) \rangle$ of measured data	3.9 (1.7)	6.4 (2.0)
[a] R_{sym}	0.13 (0.36)	0.06 (0.39)
[a] R_{meas}	0.16 (0.44)	0.07 (0.43)

Table 3.1. Crystal parameters, data collection and processing statistics. Values in parenthesis are for the highest resolution shell.
[a] $R_{sym}(I)=\Sigma_{hkl}\Sigma I\ |I_{hkl,i}- \langle I_{hkl}\rangle|\Sigma_{hkl}\Sigma I|I_{hkl,i}|$ with $\langle I_{hkl}\rangle$ mean intensity of the multiple $I_{hkl,i}$ observations from symmetry-related reflections.
[b] R_{meas} is the multiplicity weighted R_{sym}.

3.2 References

[1] V. Govan, V. Deretic, Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol. Rev.* 60 (1996) 539–574.

[2] K.C. Stover, X.Q. Pham, A.L. Erwin, S.D. Mizoguchi, P. Warrener, M.J. Hickey, F.S.L. Brinkman, O.W. Hufnagle, D.J. Kowalik, M. Lagrou, R.L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L.L. Brody, S.N. Coulter, K.R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G.K.-S. Wong, Z. Wu, I. Paulsen, J. Reizer, M.H. Saier, R.E.W. Hancock, S. Lory, M.V. Olson, Complete genome sequence of *Pseudomonas aeruginosa* PAO1: an opportunistic pathogen, *Nature* 406 (2000) 959–964.

[3] U.T. Bornscheuer, Microbial carboxyl esterases: classification, properties and application in biocatalysis, *FEMS Microbiol. Rev.* 26 (2002) 73–81.

[4] T. Satoh, M. Hosokawa, The mammalian carboxylesterases: from molecules to functions, *Annu. Rev. Pharmacol. Toxicol.* 38 (1998) 257–288.

[5] W.B. Jacoby, D.M. Zeigler, Detoxification of pesticides by microbial enzymes, *Experientia* 39 (1990) 1236–1246.

[6] T. Hotelier, L. Renault, X. Cousin, V. Negre, P. Marchot, A. Chatonnet, ESTHER, the database of the α/β hydrolase fold superfamily of proteins, *Nucleic Acids Res*. 32 (2004) D145– D147.

[7] D.L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S.M. Franken, M. Harel, S.J. Remington, I. Silman, J. Schrag, A. Goldman, The α/β hydrolase fold, *Protein Eng.* 5 (1992) 197– 211.

[8] A. Pesaresi, G. Devescovi, D. Lamba, V. Venturi, G. Degrassi, Isolation, characterization, and heterologous expression of a carboxylesterase of *Pseudomonas aeruginosa* PAO1, *Curr. Microbiol*. 50 (2005) 1 – 7.

[9] H.-S. Ro, H.P. Hong, B.H. Kho, S. Kim, B.H. Chung, Genome-wide cloning and characterization of microbial esterases, *FEMS Microbiol.Lett.* 233 (2004) 97–105.

[10] Y. Devedjiev, Z. Dauter, S.R. Kuznetsov, T.L. Jones, Z.S. Derewenda, Crystal Structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 Å, *Struct. Fold. Des.* 8 (2000) 1137– 1146.

[11] J.A. Duncan, A.G. Gilman, A cytoplasmatic acylprotein thioestease that removes palmitate from G protein alpha subunit and p21 (RAS), *J. Biol. Chem*. 173 (1998) 15830– 15837.

[12] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res*. 28 (2000) 235–242.

[13] K.K. Kim, H.K. Song, D.H. Shin, K.Y. Hwang, S. Choe, O.J. Yoo, S.W. Sush, Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an alpha/beta hydrolase with broad substrate specificity, *Structure* 5 (1997) 1571– 1584.

[14] J. Jancarik, S.H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, *J. Appl. Cryst*. 24 (1991) 409– 411.

[15] J.L. Yeh, W.G.J. Hol, A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase, *Acta Crystallogr., D Biol. Crystallogr*. 54 (1998) 479– 480.

[16] S. Kriminski, C.L. Caylor, M.C. Nonato, K.D. Finkelstein, R.E. Thorne, Flash-cooling and annealing of protein crystals, *Acta Crystallogr., D Biol. Crystallogr*. 58 (2002) 459– 471.

[17] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, in: C.W. Carter Jr., M. Sweet (Eds.), *Methods Enzymol., Macromolecular Crystallography, Part A,* vol.276, Academic Press, New York, 1997, pp. 307–326.

[18] Collaborative Computational Project, Number 4, The CCP4 Suite: Programs for Protein Crystallography, *Acta Crystallogr., D Biol. Crystallogr*. 50 (1994) 760–763.

[19] K. Diederichs, P.A. Karplus, Improved R-factors for diffraction data analysis in macromolecular crystallography, *Nat. Struct. Biol*. 4 (1997) 269–273.

[20] B.W. Matthews, Solvent content of protein crystals, *J. Mol. Biol*. 33 (1968) 491–497.

[21] J. Navaza, Implementation of molecular replacement in AMoRe, *Acta Crystallogr., D Biol. Crystallogr*. 57 (2001) 1367–1372.

[22] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography and NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., D Biol. Crystallogr*. 54 (1998) 905–921.

Chapter 4

PA3859 Carboxylesterase Structure

4.1 Structure refinement

The structures of the PA3859 carboxylesterase in crystal forms I and II were refined using the CNS suite [1]. 10% of the reflections were selected for the cross validation procedure. Simulated annealing was performed (slow cooling protocol), followed by positional and B-factor refinement and manual inspection of the resulting 2Fc-Fo and Fc-Fo maps. For the electron density map visualization the graphical program O was used [2]. Several cycles of positional and isotropic B-factor refinement and manual model building were performed. A search for water molecules was carried out as well. The water molecules were picked in the Fo-Fc electron density map according to the CNS protocol. Ligands such as MES (crystal form I) and PEG (crystal form II) were identified. Ligand coordinates and stereochemistry library files for CNS were taken from the HIC-UP server [3].
R and R_{free} values were computed after every step of the refinement procedure. The geometries at the present status were checked [4]. The refinement statistics for PA3859 carboxylesterase crystal forms I and II are given in the Table 4.1. According to the Ramachandran plot (Fig. 4.1) the serine at position 113 (the catalytic serine) had backbone angles outside the favored regions of the plot: this is a common feature of carboxylesterases. The sharp strand-turn-helix conformation of the catalytic elbow forces the serine to adopt unfavorable main chain torsion angles.

	Crystal form I	Crystal form II
Resolution range (Å)	24.30-2.98	29.05-2.09
Final R factor (%)	23.6	19.6
Final R free factor (%)	28.4	24.1
Rms deviation		
Bond length (Å)	0.007	0.009
Bond angles (°)	1.3	1.4
Average isotropic B-factors (Å^2)		
All protein	32.78	32.78
Water molecules	39.31	39.31
MES	52.89	-
PEG-dimer	-	60.53
PEG-nonamer	-	51.33
% of main-chain dihedral angles in Ramachandran plot found to be in		
fully allowed regions	84.6	88.8
additional allowed regions	14.6	10.7
generously allowed regions	0.3	0.5
disallowed regions	0.5	0

Table 4.1. Refinement statistics of PA3859 crystal form I and II.

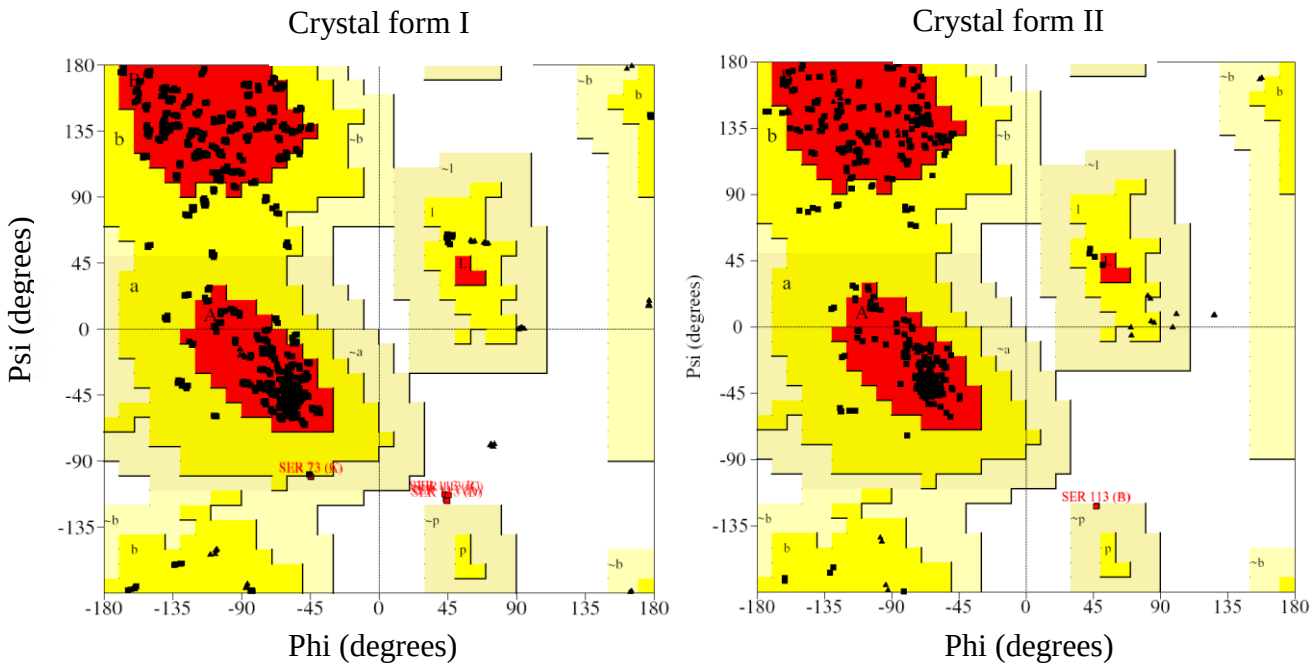


Figure 4.1. Ramachandran plot of PA3859 of crystal form I and II.

4.2 PA3859 structure description

4.2.1 PA3859 has a α/β hydrolase fold

PA3859 shows a typical α/β hydrolase fold with 7 β-strands (β_1, β_2, β_3, β_6, β_7, β_8, β_9) surrounded by 5 α-helices (α_1, α_3, α_4, α_5, α_6) (Fig. 4.2). With respect to the Prototypic Hydrolase Fold (see Chapter 1) the first N-terminal strand is missing and the helix between strand β_7 and β_8 is replaced by two short helices, η_3 and η_4, that do not interact with the β-sheet. Moreover an additional domain is present that encompasses the β-α-β motif made by strands β_4, β_5 and helix α_2, and the long loop which connects it to helix α_3. The catalytic triad is composed of Ser 113, located at the tip of the loop between β_6 and α_3, Asp 166, between β_8 and α_5 and His 197, in the loop that precedes α_6.
No significant differences are present between the structures of the PA3859 monomers in crystal forms I and II, the only significant change being a switch in the orientation of the Arg 212 and Trp 61 side chains (Fig. 4.3). In both crystal forms these residues are involved in the dimer-dimer interaction and the different orientations of their side chain can be considered as a consequence of the different crystal packing.

The enzyme's active site is located in a long cleft that is present on the protein surface (Fig.4.4). The left portion of the cleft encompasses helices α_1 and α_6 and is bodered by the hydrophobic residues Phe 29, Val 32, Leu 36, Val 39, Phe 112, Val 199 and Ile 204. The right part of the cleft is a rather polar cavity and is delimited between the α/β core (Tyr 140 and Val 168), and the β-α-β-loop extra domain (residues Ile 69 to Ile 78).

In crystal form I one molecule *per* dimer is inhibited by a MES molecule originated from the precipitant solution (fig. 4.5). The MES sulfur atom is covalently bound to the catalytic serine Oγ and the MES morpholino moiety is positioned in the polar cavity. One of the two sulfate group oxygen atoms points toward the oxyanion hole composed of Leu 22 and Gln 114 main chain nitrogen atoms.
No clear differences are present in the monomer structures or in the crystal packing that could explain why only one of the two molecule was inhibited by MES in each dimer.

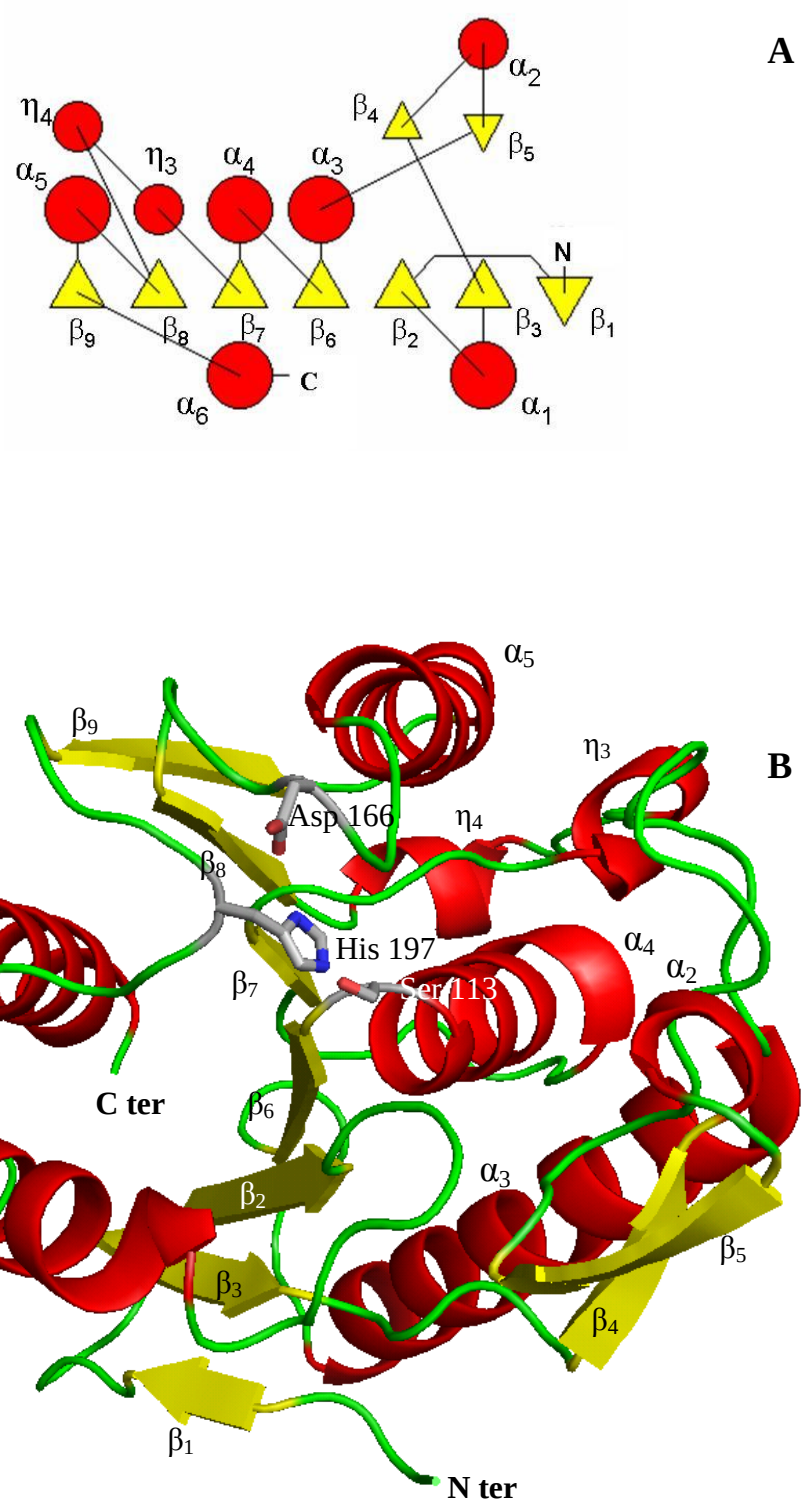


Figure 4.2. A, PA3859 topology. Yellow triangles: β-strand. Red circles, α-helix. **B**, PA3859 α/β fold. One extra β-α-β motive is present between strand 3 and helix 3. The catalytic triad residues are shown in stick format. This picture, as well as all the other ones present in this chapter, was made using the program PyMol (http://www.pymol.org).

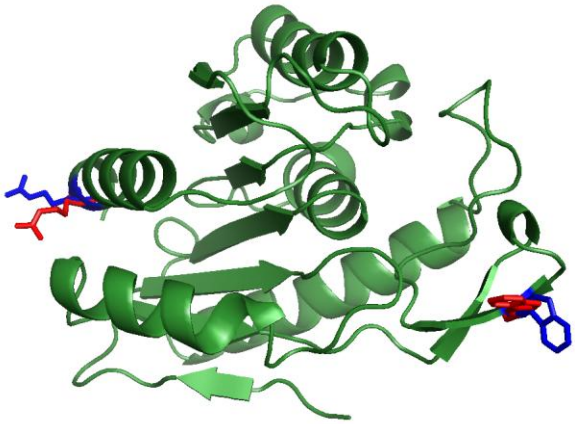


Figure 4.3. Different Trp 61 and Arg 212 side chains orientation. Red, crystal form I. Blue, crystal form II.

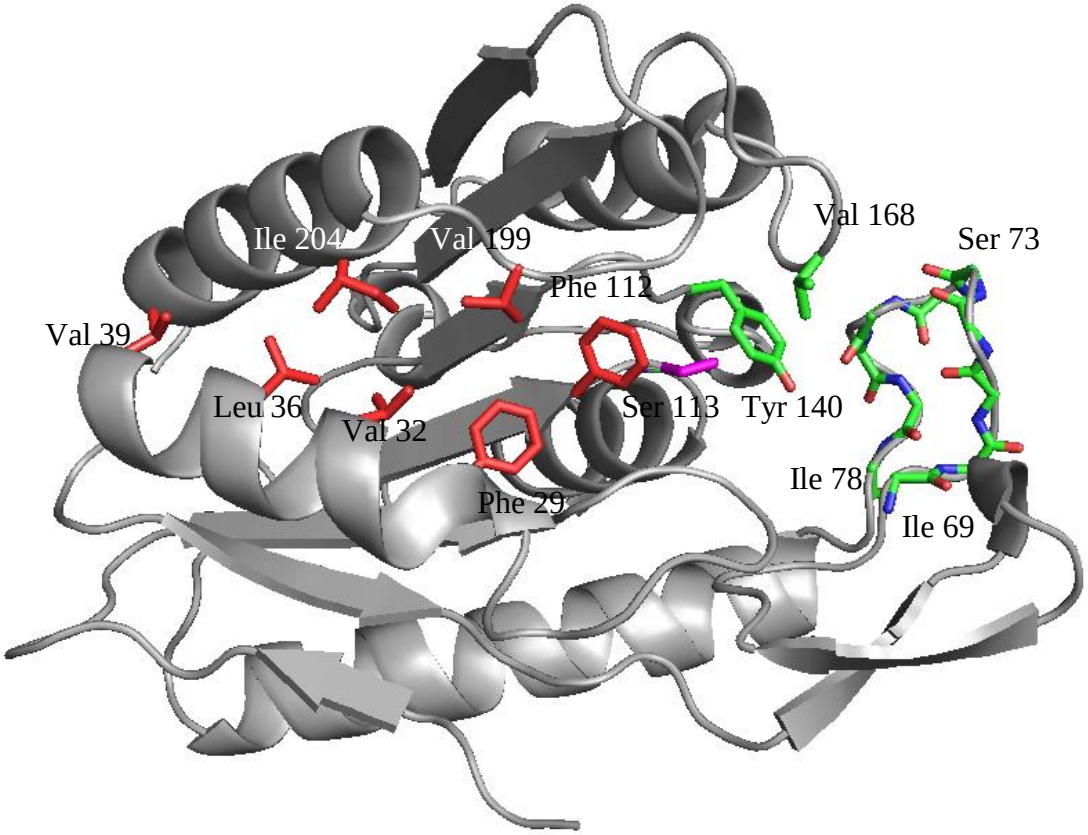


Figure 4.4. The PA3859 active site, here indicated by the position of the catalytic Ser 113 (magenta) is located roughly in the middle of a groove. The left part of the groove is composed of hydrophobic residues (red) and the right part is made up of the side chains of residues Tyr 140 and Val 168 of the α/β protein core, and the main chain of residues Ile 69 to Ile 78 located in the extra domain (green).

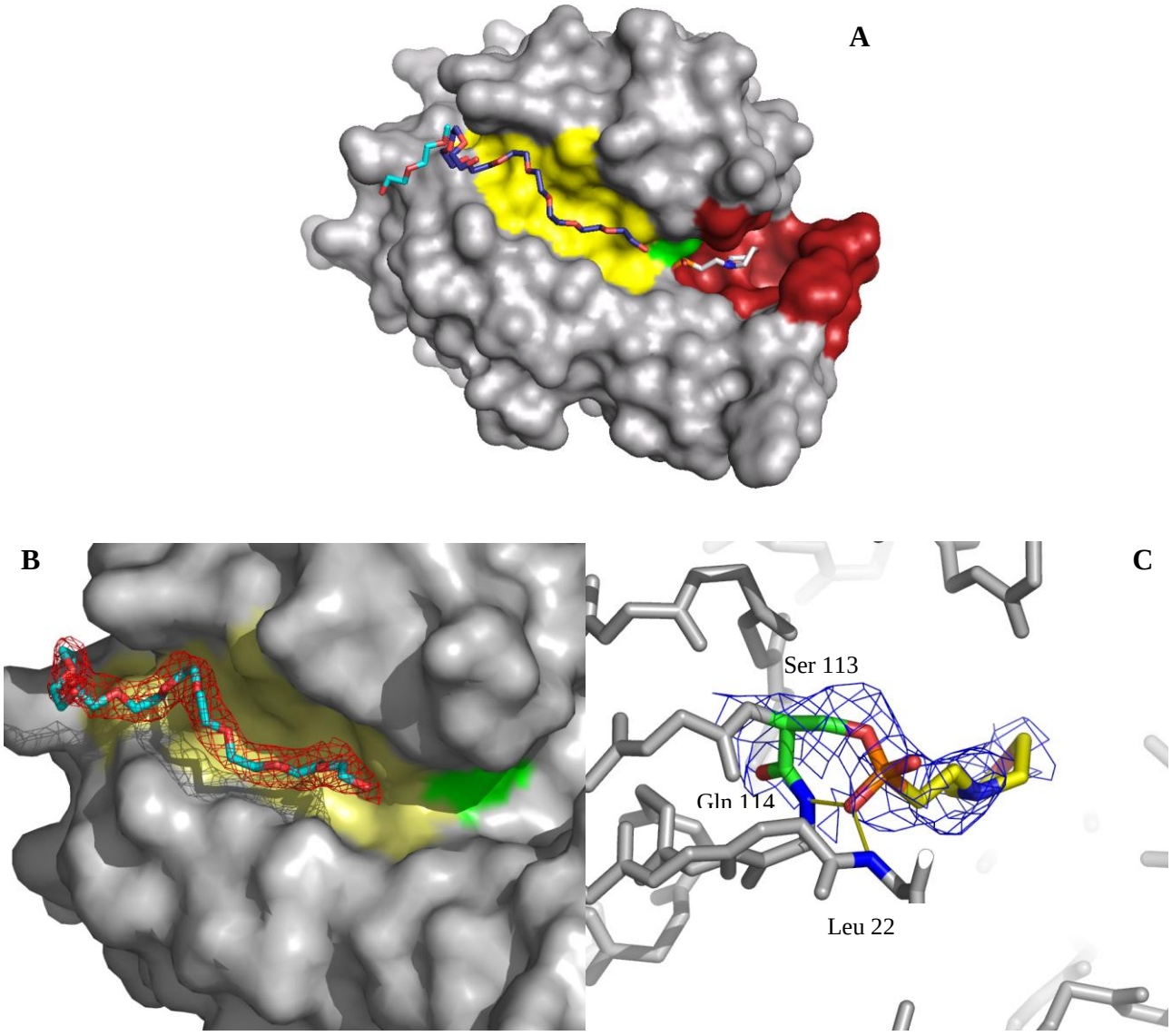


Figure 4.5. (**A**) PA3859 surface. Ser 113 (green), hydrophobic cleft (yellow), polar cavity (red). PEG nonamer (blue), PEG dimer (cyan), MES (white). 1σ level of the 2Fo-Fc electron density map is shown for the PEG nonamer in the hydrophobic cleft (**B**), and the MES molecule in the active site (**C**).

In crystal form II two PEG strands of different length are visible. PEG is a very flexible molecule and thus often is completely disordered within the crystal. Here, due to the strong interaction with the protein surface, the PEG strands are induced to assume an ordered structure that gives rise to clear electron densities. The PEG present in the precipitant solution had a mean molecular weight of 5 kDa that corresponds to a polymerization degree of 110. In the crystal structure a nonameric PEG strand is lining the hydrophobic portion of the cleft and a shorter strand, roughly a dimer, is located nearby the end of the cleft (fig. 4.5).

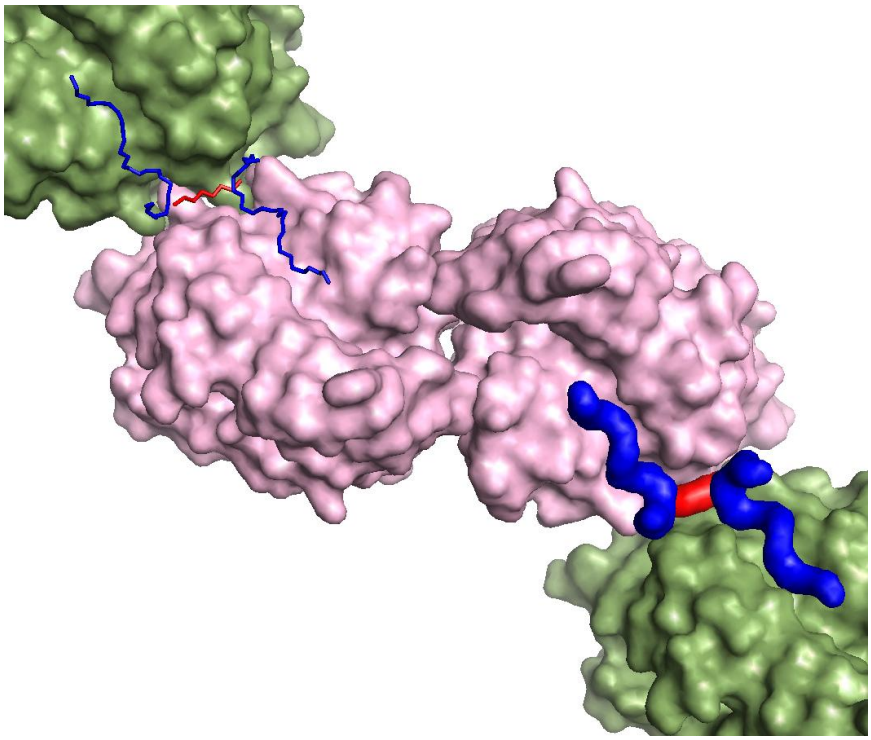


Figure 4.6. The Dimer-dimer interaction mediated by PEG strands. PEG nonamer (yellow), PEG dimer (red).

Interestingly interaction between dimers in adjacent unit cells is mediated by these PEG strands (Fig. 4.6): the nonamers protrude toward the dimer-dimer interface and the short strand makes a bridge between the long ones, as if PEG would have driven the protein crystallization.

4.2.2 Structures comparison

The Cα trace superimposition (Fig. 4.7) reveals that an almost identical fold is shared between PA3859 and both *P. fluorescens* carboxylesterase (76% sequence homology, RMSD = 0.96 Å) [5] and human thioesterase (33% sequence identity, RMSD = 1.47 Å) [6].
Among all of the three enzymes, the hydrophobic cleft is a very well conserved feature, while the extra domain, and thus the polar cavity, differ to a greater extent. In terms of sequence β_2, which forms the bottom of the cleft, is the overall most conserved region next to the active site loop, while α_1 and α_6 are much less conserved (Fig. 4.8). High sequence identity is also found around the β_4-α_2-β_5 motif that is present in the extra domain, while the final part of the long loop that connects it to α_3 displays a lower degree of homology:

the shape of the polar cavity displays the most structural divergence among the three enzymes and the most remarkable differences, corresponding to PA3859 residues 70 to 78, are found in this variable loop (Fig. 4.7 and Fig. 4.9). This finding probably accounts for the different substrate specificities displayed by these enzymes.

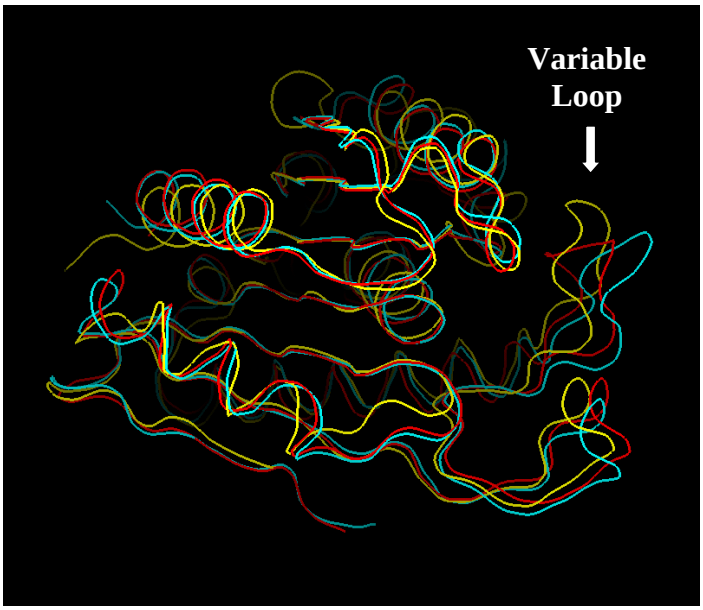


Figure 4.7. Cα trace superimposition of PA3859 (red), *P. fluorescens* carboxylesterase (cyan), and Human acyl-protein thioesterase (yellow).

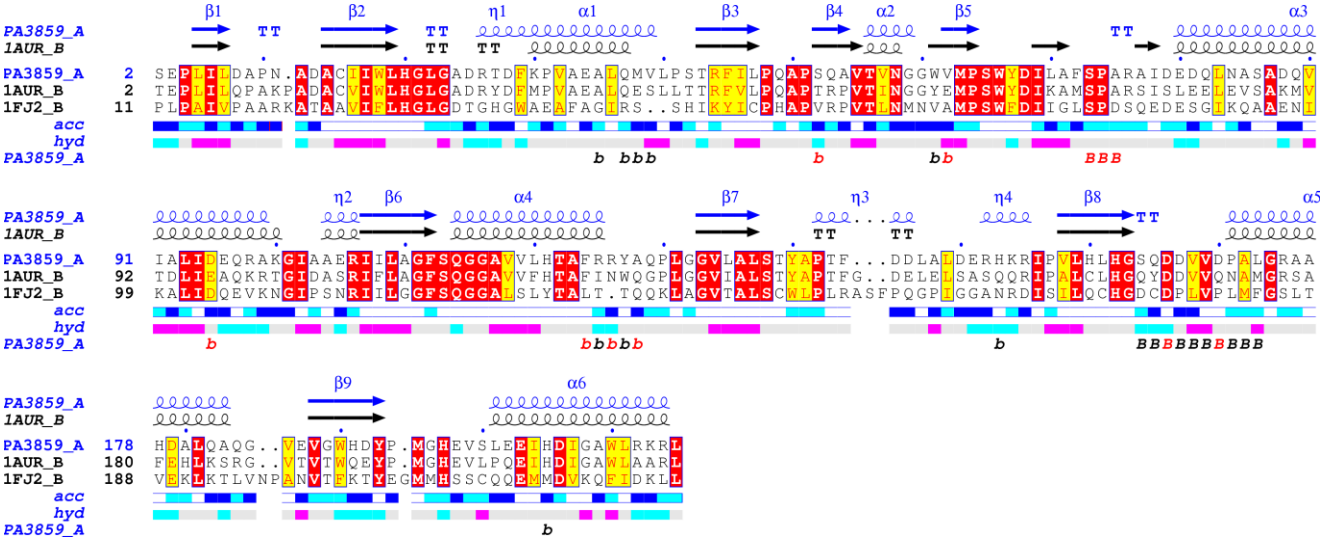


Figure 4.8. Multiple sequence alignmen produced by ENDscript (http://espript.ibcp.fr/ESPript/ENDscript/index.php) of PA3859, *P. fluorescens* carboxylesterase (1AUR) and Human acyl-protein thioesterase (1FJ2). Identity regions are highlighted in red, homology regions in yellow. The solvent accessibility (acc) and hydrophobicity (hyd) are also indicated.

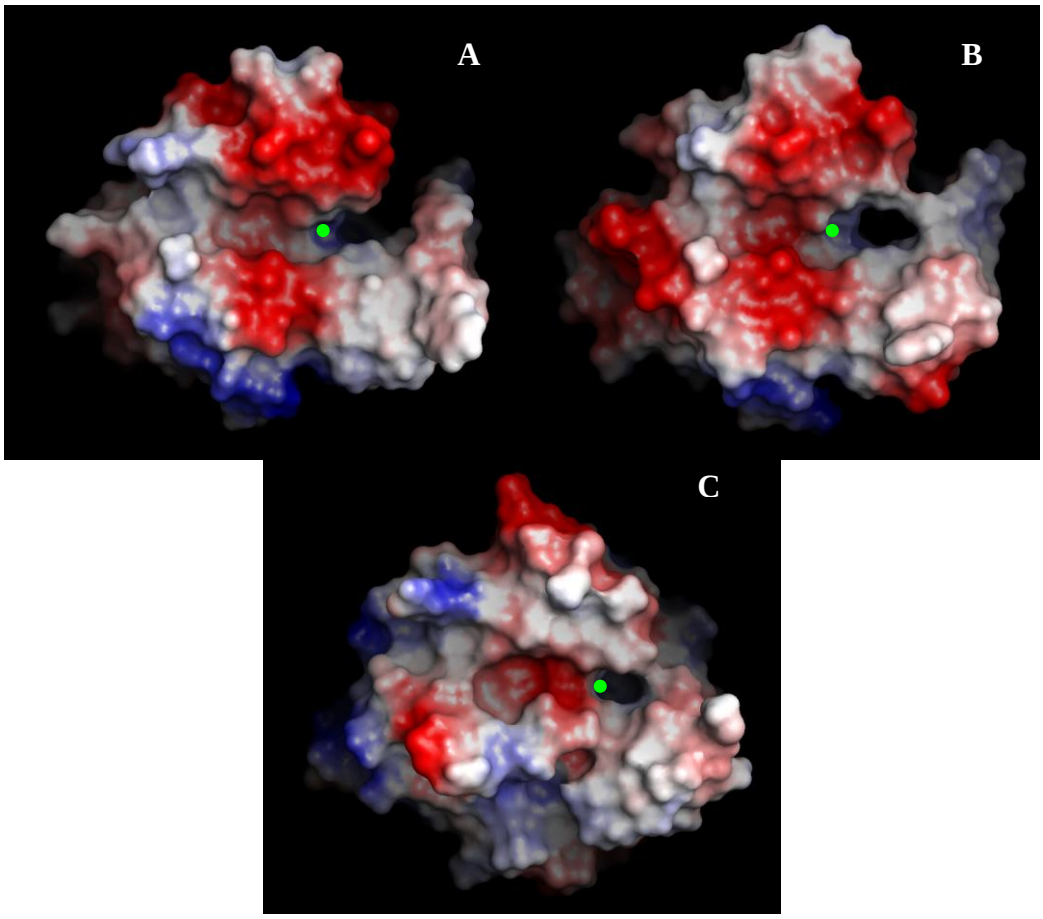


Figure 4.9. Protein surface electrostatic potential of PA3859 (**A**), *Pseudomonas fluorescens* carboxylesterase (**B**) and human thioesterase (**C**), computed by PyMol. Red, negatively charged residues. Blue, positively charged residues. The catalytic serines are highlighted by a green mark.

As is shown in Fig. 4.9, the protein surfaces of PA3859 and *P. fluoescens* CE display a very similar shape and electrostatic potential: in both molecules, two higly acidic regions are present above and below the hydrophobic cleft. In *P. fluorescens* CE an extra acidic region is found at the end of the hydrophobic cleft. Furthermore in this enzyme, the Met 73 that replaces the Phe 72 of PA3859, is in van der Waals' contact with the Val 170, thus making a bridge between the extra domain and the α/β core.
The protein surface of human thioesterase is significantly different from those of the bacterial enzymes: the hydrophobic cleft is much less outspoken and the polar cavity is reduced to a narrow hole with the catalytic serine positioned at the edge of this hole. Here the acidic regions that in the other two enzymes flank the hydrophobic cleft are missing.

4.2.3 Dimerization interface

The PA3859 dimerization interfaces, involving helix α_5 and the loop between β_5 and α_3, are very similar in the two crystal forms even though the Cα traces superimposition (Fig. 4.10) reveals some slight differences in the positioning of the molecules within the dimer. In crystal form II the α_5 helices interact more tightly than in crystal form I. This produces a rigid body-like rotation of one monomer with respect to the other and a more packed overall structure.

Figure 4.10. Dimerization interface. Crystal form II (Red), Crystal form I (White).

Despite their high sequence homology carboxylesterase from *P. aeruginosa* and *P. fluorescens* show a quite different dimeriaztion interface (Fig. 4.11 and Fig. 4.12). In the *P. fluorescens* dimer the catalytic sites are facing each other and, as a result, the dimerization interface is much larger than in PA3859. In both dimers however the substrate binding sites are exposed to the solvent.
The dimerization mode in the human acyl-thioesterase is still different: here both the catalytic serine and the hydrophobic cleft are buried. The interaction is much tighter than in the other two enzymes and dimer dissociation has been hypothesized to occur upon substrate binding [7].

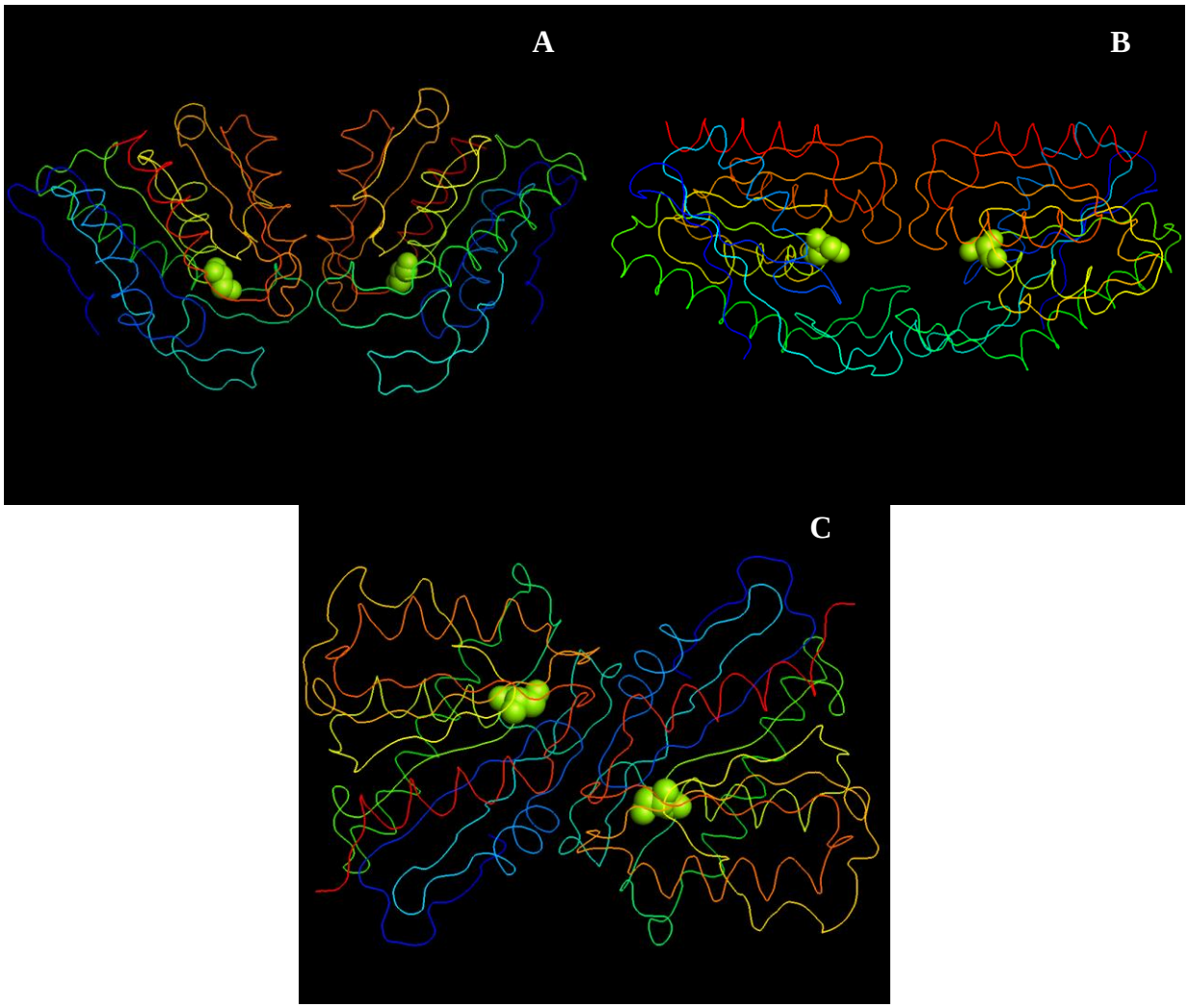


Figure 4.11. Dimerization modes: **A**, PA3859 dimer. **B**, *P. fluorescens* carboxylesterase dimer. **C**, PA3859 Human thioesterase dimer. The catalytic serines are depicted as spheres.

According to a structure analysis performed using PISA [8], in PA3859 only 3.8% of the protein surface is involved in the dimer formation (corresponding to 377 $Å^2$). In the carboxylesterase from *P. fluorescens* the portion of the protein surface involved in the dimerization is 6.9% (corresponding to 667 $Å^2$), and in the human thioesterase 11.9% (1225 $Å^2$) (Fig. 4.12).

It is worthy to remark however that while in carboxylesterase from *P. fluorescens* and in human acyl-protein thioesterase the dimerization mode could play a significant role in the accomplishment of the physiologic function, it should not have any functional relevance in the case of PA3859, since this protein was demonstrated to be present in solution as a monomer (see Chapter 2 and 3).

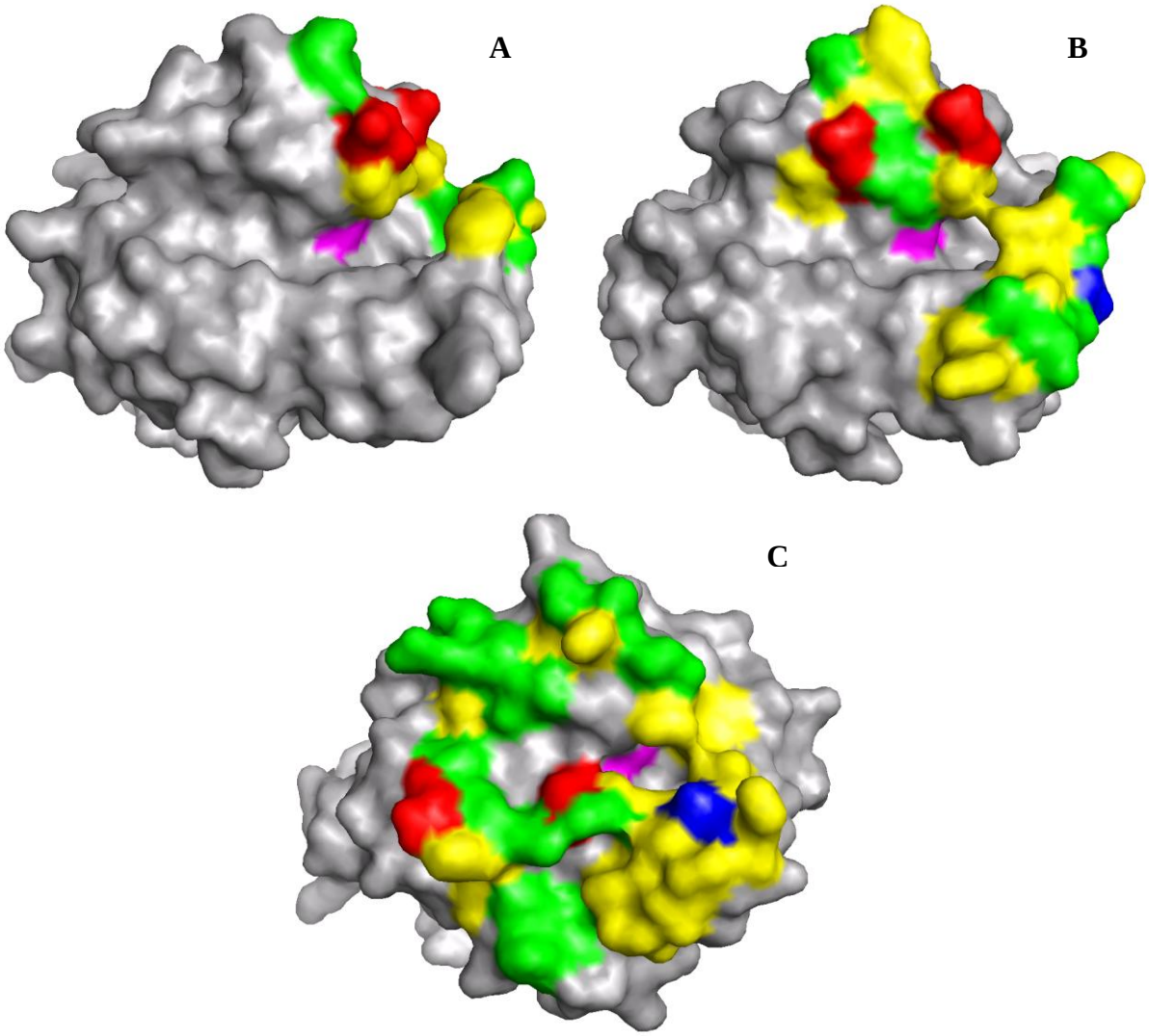


Figure 4.12. Protein surface involved in the dimerization determined by PISA: PA3859 (A), carboxylesterase from *P. fluorescens* (B), human thioesterase (C). Yellow, hydrophobic residues. Green, neutral residues. Red, negatively charged residues. Blue, positively charged residues. The catalytic serine is highlighted in magenta.

4.3 References

[1]] Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. (1998) *Acta Cryst*. **D54**, 905-921.

[2] Jones, T. A., Zou, J. Y., Cowan, S., Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. (1991). *Acta Cryst*. **A47**, 110-119.

[3] Kleywegt, G. J., Jones, T. A. Efficient rebuilding of protein structures. (1996) *Acta Crystallogr D Biol Crystallogr*. **52**, 829-32.

[4] Roman A. Laskowski, Malcolm W. MacArthur, David S. Moss and Janet M. Thornton (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291.

[5] Kim, K.K., Song, H.K., Shin, D.H., Hwang, K.Y., Choe, S., Yoo, O.J. Sush, S.W. Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an alpha/beta hydrolase with broad substrate specificity. (1997) *Structure* **5**, 1571– 1584.

[6] Devedjiev, Y., Dauter, Z., Kuznetsov, S.R., Jones, T.L., Derewenda, Z.S. Crystal Structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 Å, (2000) *Struct. Fold. Des*. **8**, 1137– 1146.

[7] Duncan, J.A., Gilman, A.G. A cytoplasmatic acyl-protein thioesterase that removes palmitate from G protein alpha subunits and p21 (RAS). (1998) *J. BioI. Chem.* **273**, 15830–15837.

[8] Krissinel, E. and Henrick K. Detection of Protein Assemblies in Crystals.. Berthold, M. R. *et.al.* (Eds.): CompLife 2005, LNBI 3695, pp. 163-174. Springer-Verlag, Berlin, Heidelberg.

Chapter 5

Docking simulations

5.1 Introduction

In PA3859 the location of Ser 113, the catalytic residue, in a cleft present on the protein surface prompted us to hypothesize that this cleft is the enzyme's substrate binding site. This hypothesis was also supported by the finding in crystal form I, of a MES molecule covalently bound to Ser 113 and positioned in the polar cavity, and in crystal form II of a nonameric PEG strand lining the hydrophobic cleft (Fig. 5.1).

The first step in the enzymatic catalysis is the recognition of the substrate and its binding to the enzyme. The highly specific substrate-enzyme interaction can only occur if a significant structural and chemical complementarity exists between the substrate and the enzyme active site. This implies that given the 3D structure of an enzyme whose function is unknown, following this "complementarity rule", it should be possible, at least in principle, to chase its physiologic substrate(s).
In order to obtain insights into the PA3859 *in vivo* substrates, and thus *in vivo* function, its active site was explored by means of automated docking simulations. Two different sets of simulations were performed to separately probe the binding properties of the hydrophobic cleft and polar cavity, using the automated docking AutoDock 3.05 program suite [1, 2], (see Appendix 2).

5.2 PA3859 Docking Simulation

The substrates PDBs were generated using the Dundee PRODRG2 Server (http://davapcl.bioch.dundee.ac.uk/programs/prodrg/). Atomic partial charges for both substrates (Gasteiger charges) [3] and protein (United Kollman charges) [4] were assigned using the Automated Docking Tool (ADT).
Protein atomic fragmental volume and atomic solvatation parameters were assigned with ADT. Substrate torsion values, defining bonds allowed to rotate, were assigned using the AutoTors feature in AutoDock.
Electrostatic and van der Waals' interaction energy grids were calculated by the AutoGrid module of Autodock. The grid was centered on the CD2 carbon atom of His 197, that is positioned roughly in the middle of the protein binding site. Grid dimensions were 65, 60 and 100 points for the X, Y and Z axes respectively (390,000 overall points). Grid point spacing was set to 0.375 Å, roughly a quarter of a carbon-carbon single bond length.

The searching space was explored using the Lamarckian Genetic Algorithm (LGA).
The docking of each substrate (docking job) is the result of 20 docking cycles. Each docking cycle was made up of 50 LGA runs with 2×10^6 energy evaluations *per* run. Thus overall 1000 LGA runs were conducted for each docking job (Fig. 5.2). The population size was set to 50 (default value). Crossover and mutation rate were set to 80 and 2 respectively (default values). The number of best individuals that automatically survive was set to 1.
The best docking solution produced in a cycle was used to seed the successive cycle: it means that the random populations produced at the beginning of the LGA runs were made up of 49 random solutions plus the best solution produced along with the precedent cycle.

The sulfur atom of the MES molecule found in crystal form I was used to validate the docking results: a substrate was considered to be docked if after repeated redockings its carbonylic carbon atom was within 1.5 Å from the MES sulfur atom covalently bound to the catalytic serine.
If after 10 LGA runs the obtained structure did not meet these criteria, the ligand was considered undockable.

Two sets of docking simulations were performed in order to separately assay the binding potentiality of the hydrophobic cleft and polar cavity (Fig 5.1). The same grid was used for both simulation sets.

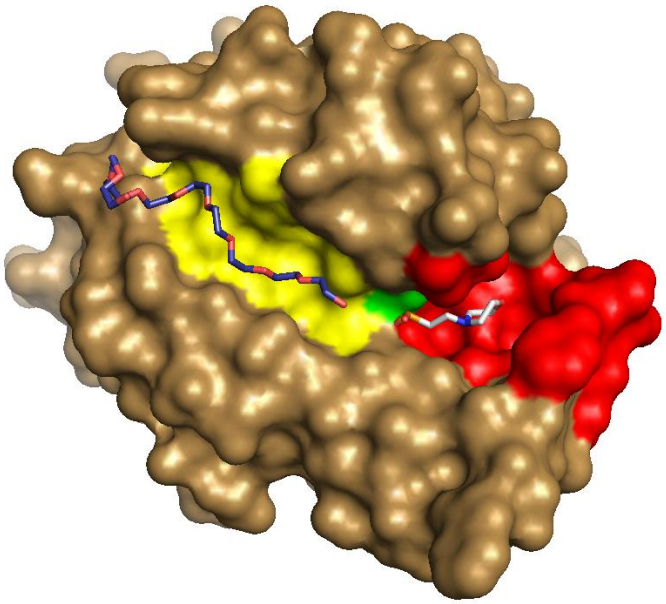


Figure 5.1. PA3859 surface. The hydrophobic cleft is highlightd in yellow and the polar cavity in red. The catalytic Ser 113 is highlighted in green. The sulfur atom of the MES molecule (white stick) was used for the validation of the docking results. The PEG nonamer found within the hydrophobic cleft is depicted as a blue stick.

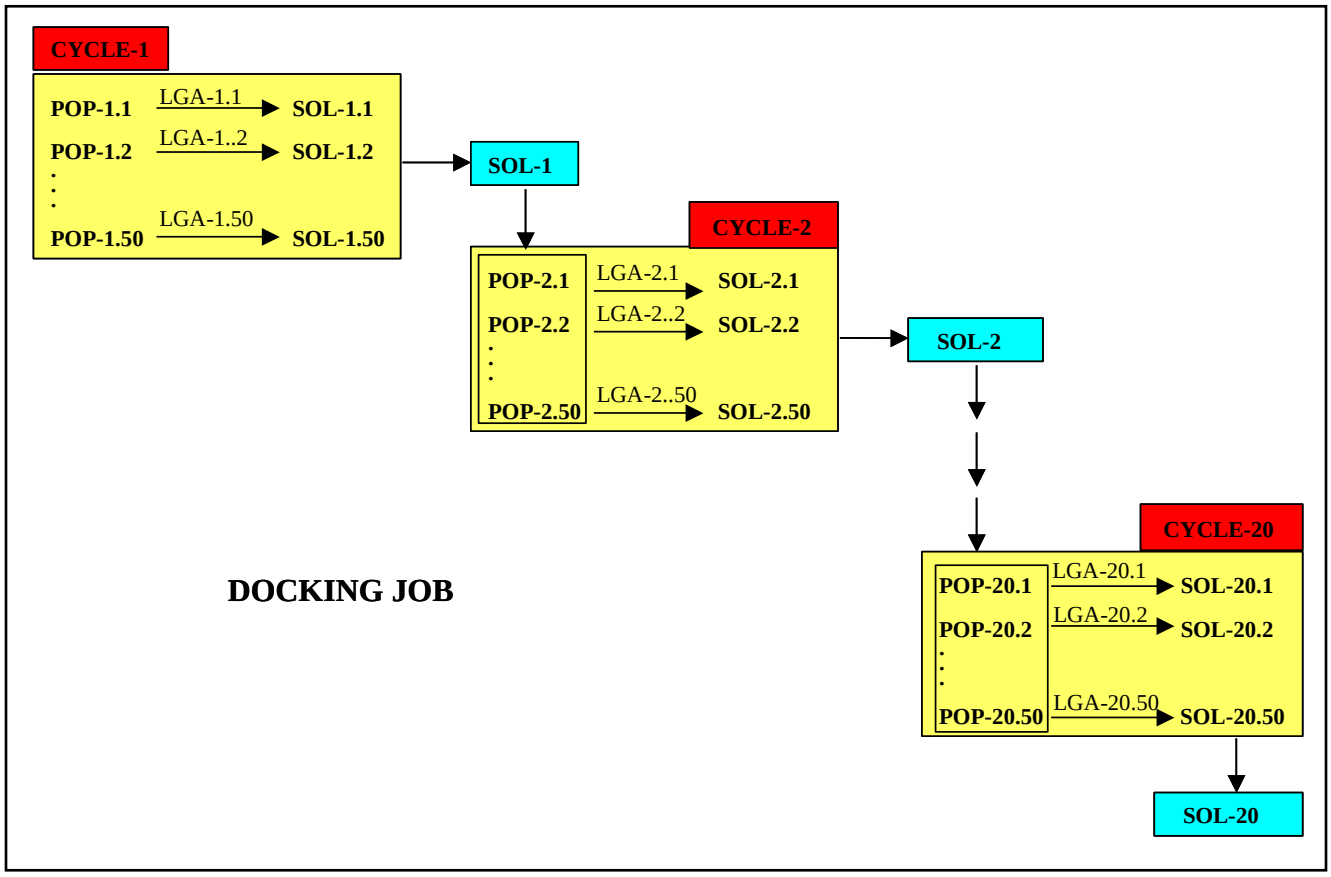
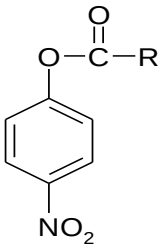


Figure 5.2. Schematic representation of the docking job. The docking of each substrate is the result of 20 cycles. In each cycle 50 LGA runs were performed. The LGA runs start with a population of 49 random solutions (50 for the first cycle) plus the best solution produced by the previous cycle. Thus the docking of each substrate is the result of 1000 LGA runs and overall $2x10^9$ energies evaluation.

5.2.1 Hydrophobic cleft

The geometry of the hydrophobic cleft and the finding in crystal form II of a PEG strand positioned into the cleft, prompted us to hypothesize that this cleft is the portion of the substrate binding site involved in the accommodation of the substrate alkyl chain.
To assay this possibility, the docking of *p*-nitrophenyl esters (*p*-NP) of several carboxylic acids with different chain length was attempted: *p*-NP derivatives are often used as non-physiologic substrates to determine the kinetic properties of hydrolases since the release of *p*-nitrophenol can easily be measured spectrophotometrically. Here these substrates were chosen in order to compare the automated docking results and the results of the kinetic study of PA3859 (see Chapter 6).



General structure of a *p*-NP ester

The docking of twelve substrates with the carboxylic acid acyl chain "R" spanning from 1 carbon atom (*p*-NP acetate, C2) to 21 (*p*-NP behenoate, C22) was attempted.
Among the tested substrates only *p*-NP acetate could not be successfully docked. In table 5.1 are reported the interaction energies and the binding constants for each solution and in Figure 5.3 the docked substrates conformations are shown. In all of the found solutions the *p*-nitrophenyl moiety is positioned in the polar cavity and the acyl chain lays, completely or partially, within the hydrophobic cleft. Distances between the substrate carbonylic carbon atom ad the MES sulfur atom ranges between 1.33 and 1.48 Å and the carbonylic oxygen atoms always point toward the oxyanion hole (see Chapter 4 and Fig. 5.3.**G**).
Two different conformations are found for the *p*-NP moiety: in substrates C4, C6, C8, C12, C18, C20 and C22 hydrogen bonding occurs between the nitro group oxygen and the Tyr 140 hydroxyl, whose distances range from 2.93 to 3.49 Å. For substrate C10, C14 and C16 in the overall minimum energy solution the aromatic ring seems to have a less tight interaction with the polar cavity and the distance between the nitro group oxygen and Tyr 140 hydroxyl is roughly 6 Å. Solutions with the aromatic ring positioned deeper into the polar cavity were also found.
The most valuable information produced by this set of docking simulations however arise from the analysis of the interaction mode of the acyl chains with the hydrophobic cleft: for substrates C4, C6, C8 and C12 an almost identical conformation is observed, with the chains that assume an extended conformation perfectly fitted into the cleft (5.3.**A**,

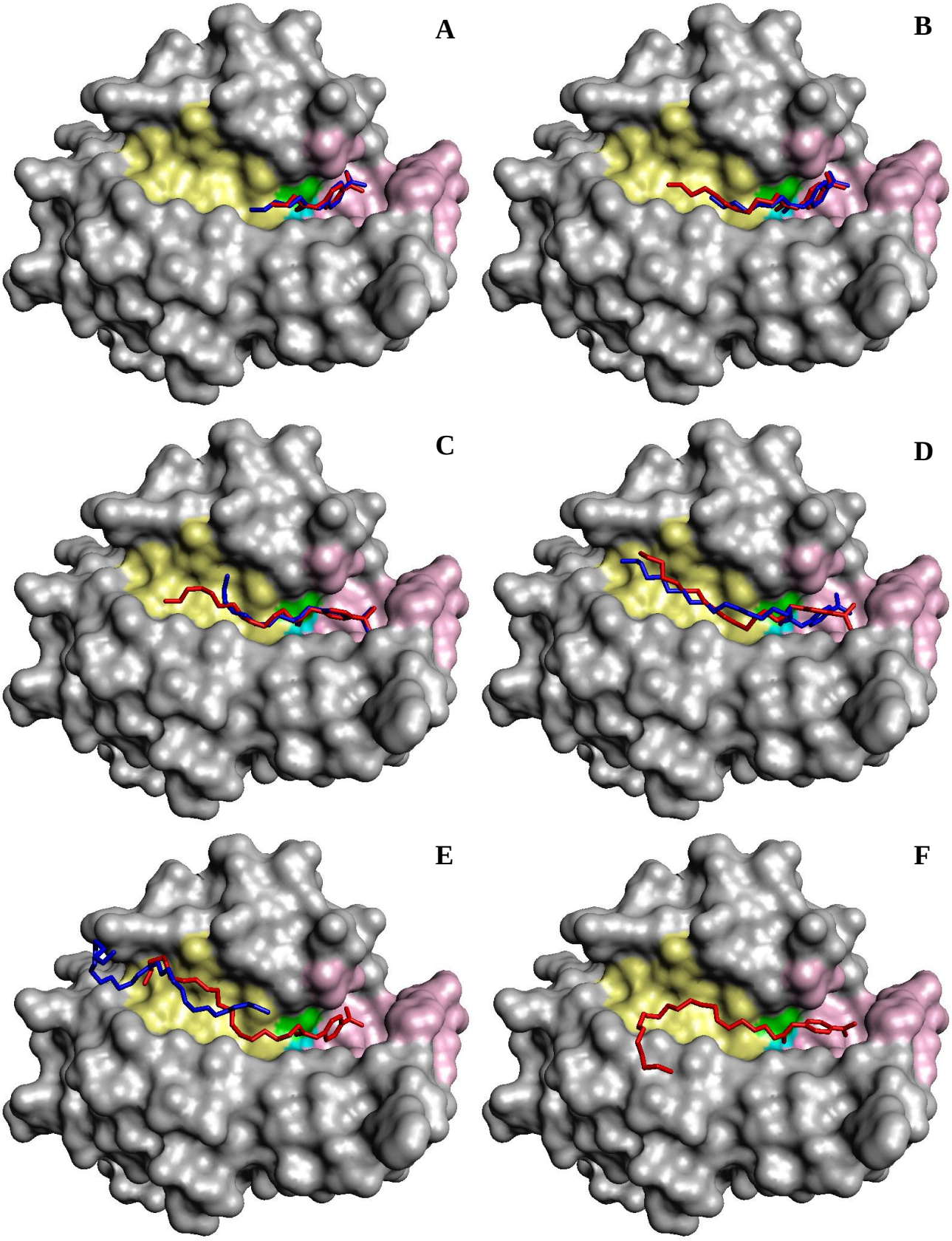
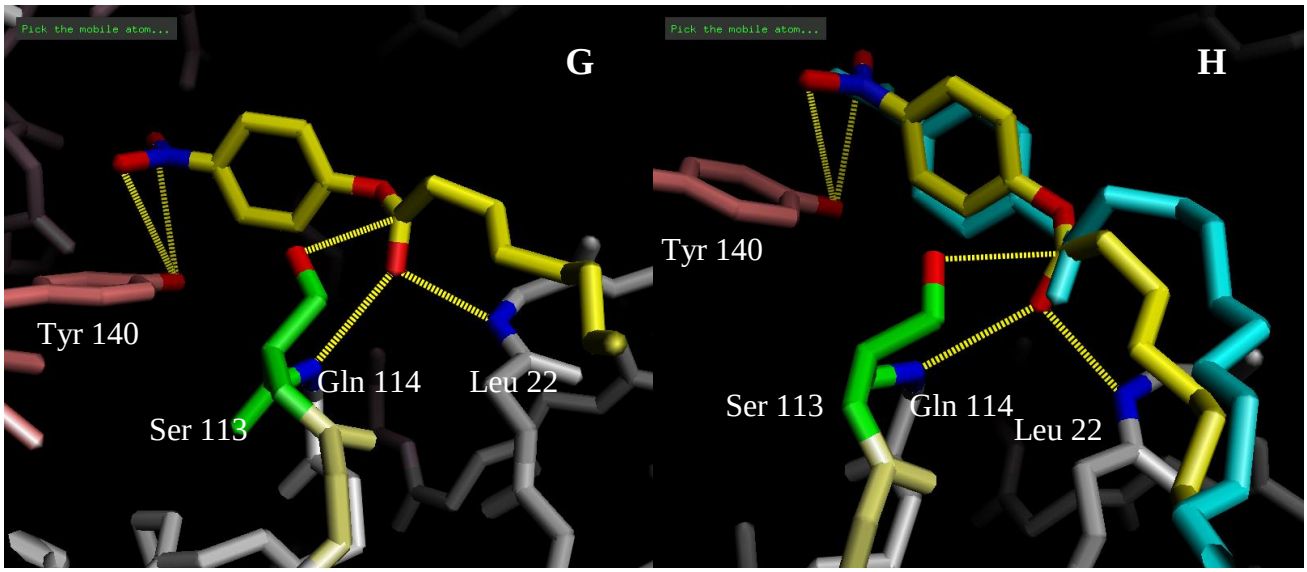


Figure 5.3. Docked *p*-nitrophenyl esters.The catalytic Ser 113 is highlighted in green, the oxyanion hole in cyan. Hydrophobic cleft is highlighted in yellow and the polar cavity in pink. **A,** *p*-NP C4 (red) and *p*-NP C6(blue). An almost identical conformation is found for these substrates, with the *p*-nitrophe



nitrophenyl moiety positioned into the polar cleft and the acyl chain into the cleft. **B,** *p*-NP C8 (red) and *p*-NP C12 (blue). The conformation is very similar to that found for the shorter substrates in **A.** The alkyl chains perfectly fit the hydrophobic cleft. **C**, *p*-NP C10 (red) and *p*-NP C14 (blue). The *p*-nitrophenyl moieties have a conformation that differs from those of shorter substrates and that gives rise to a weaker interaction with the polar cavity. In *p*-NP C10 the tail makes a turn to interact with Phe 112 aromatic ring. **D**, *p*-NP C16 (red), *p*-NP C18 (blue). The proximal parts of the acyl chains display a torsion that allows the entire molecules to fit within the cleft. **E**, *p*-NP C20 (red), PEG nonamer (blue). *p*-NP C20 chain undergoes a torsion in order to fit the cleft. Its final conformation is similar to that of the PEG nonamer found in crystal form II. **F**, *p*-NP C22 (red). The chain is too long to be entirely accommodated in the cleft. **G**, details of the interaction between *p*-NP C8 (yellow) and the PA3859 active site: the substrate carbonylic carbon atom accepts a hydrogen bond from the catalytic Ser 113, the carbonylic oxygen interacts with the main chain nitrogen atoms of the oxyanion hole residues Leu 22 nad Gln 114 and the nitro group oxygen atoms of the nitrophenyl moiety interact with the Tyr 140 hydroxyl. **H**, comparison of the conformations of the docked *p*-NP C8 (yellow) and *p*-NP C18 (cyan) While the *p*-nitrophenyl moiety and the carbonylic group of the two substrates assume an almost identical position within the PA3859 active site, the conformation of the acyl chains differ significantly: in comparison to *p*-NP C8 acyl chain, the chain of *p*-NP C18 assume a less tight interaction with the protein.

B). In C10 and C14 the different positioning of the *p*-NP moiety does not interfere with the chain conformation, because it is the proximal part, which perfectly superimposes to that of the shorter chains. However the C10 tail does not lay for all of its length at the bottom of the cleft but makes a turn that is stabilized by an interaction with the aromatic ring of Phe 112 (Fig. 5.3.**C**). For substrates C16 and C18 a weaker interaction between the proximal part of the chain and the cleft is found: in C16 the first three atoms of the chain are positioned furhter apart from the protein surface if compared to those of shorter substrates. In C18 this less tight interaction is extended up to the sixth chain carbon atom (Fig. 5.3, **H**): this bending allows the chain, that otherwise would be too long, to be positioned entirely within the cleft (Fig. 5.3.**D**). The conformational deviation from the short chain substrates conformation becomes larger for C20 and C22:

	Intermolecular Energy (Kcal mol^{-1})	Ligand Internal energy (Kcal mol^{-1})	Torsional Free energy (Kcal mol^{-1})	Binding Free Enrgy (Kcal mol^{-1})	Binding Constant (μM^{-1})
p-nitrophenyl C4	-6.41	-0.06	1.56	-4.85	276
p-nitrophenyl C6	-7.41	-0.16	2.18	-7.56	148
p-nitrophenyl C8	-8.32	-0.14	2.80	-5.52	89
p-nitrophenyl C10	-9.38	-0.31	3.42	-5.96	43
p-nitrophenyl C12	-10.13	-0.35	4.05	-6.08	35
p-nitrophenyl C14	-10.80	-0.22	4.67	-6.13	32
p-nitrophenyl C16	-11.25	0.13	5.29	-5.96	43
p-nitrophenyl C18	-11.79	-0.10	5.91	-5.88	49
p-nitrophenyl C20	-12.55	0.13	6.54	-6.01	39
p-nitrophenyl C22	-12.61	0.78	7.16	-5.45	101

Table 5.1. Interaction energies and binding constants for *p*-nitrophenyl esters of different chain length carboxylic acids. The estimated free energy of binding is the summation of the Intermolecular Energy and the Torsional Free Energy that accounts for the loss of ligand conformational degree of freedom occurring upon binding.

here the interaction of the proximal part of the chains with the protein surface is very similar to that found for substrates C4 - C10 but the tails assume a completely different conformation: in order to be fitted into the cleft the chain of C20 assumes a cork like conformation, similar to those of the PEG strand found in crystal form II, and the tail makes a sharp hook. The C22 chain is however too long to be accommodated into the cleft: the chain torsion that is required to let it fit within the cleft would probably produce a too high internal energy and in all of the obtained docking solutions its 8 or 9 tail carbon atoms are located outside of the groove (Fig. 5.3.**D**). This constrained conformation gave rise for both C20 and C22 substrates to a positive internal energy (Tab. 5.1).

In Fig. 5.4.**A** the binding constants estimated by Autodock are reported for each substrate as a function of the alkyl chain length (see also Tab. 5.1). For the substrates with chain length from C10 to C20 the binding constant values were found to be within the range 32 to 49 μM^{-1}. In this region of the plot the chain length does not affect significantly the strength of the binding. For substrates either shorter than C10 or longer than C20 a rapid increase in the binding constants is observed. It is worthy to note that this plot is in clear disagreement with that reported in Fig. 5.4.**B**: here the energetic contribution of the acyl chain to the binding, averaged over the number of carbon atom present in the chain (specific energy of interaction), are reported. These values can be considered as a goodness quantification of the acyl chain – protein interaction.

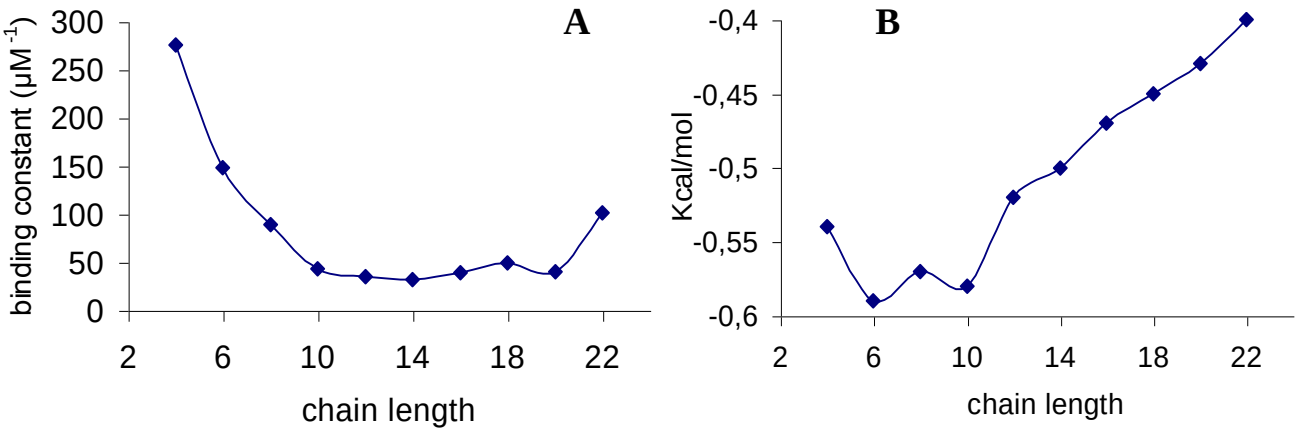


Figure 5.4. In **A** the estimated binding constants are plotted against the alkyl chain length: the lowest values are obtained for substrates between C10 and C20. For substrates with the alkyl chain either shorter than C10 or longer than C20, the binding free energy, and thus the binding constant, increases rapidly.
In **B** the mean energy contribution to the binding of the aliphatic carbon atoms is plotted as a function of the chain length: lowest values are found for substrates C6, C8 and C10.

For each substrate the mean energy of interaction *per* methylene group is plotted against the chain length. The overall minimum value, corresponding to –0.59 Kcal mol^{-1} *per* methylene group is obtained for substrate C6. Similar low energies, -0.57 and -0.58 Kcal mol^{-1}, were found for substrates C8 and C10 respectively. From substrate C12 to C22 the values increase linearly with the chain length. The maximum specific energy value is obtained for C22 and corresponds to –0.40 Kcal mol^{-1} *per* methylene group.
This result is in agreement with the acyl chain – protein interaction mode that was found to be tight for short substrates and to become weaker and weaker with the increasing chain length (Fig. 5.3). On the other hand it seems to be in contrast with the binding constant plot (**A**): for example C6, the substrate with the overall minimum energy value, has a binding constant that is significantly higher than substrates C10 – C20, and C20 that shows a small binding constant has a high mean energy *per* methylene group.
To explain this apparent paradox the torsional free energy loss that occurs up on ligand binding has to be taken into account (Tab. 5.1): upon binding the ligand is forced to assume one specific conformation that is chosen among all the possible ones. Thus a dramatic loss of the ligand torsional degree of freedom occurs. Since the torsional degree of freedom increases exponentially with the alkyl chain length, the extent of the binding ordering effect will also increase exponentially with the chain length. This entropic effect, according to the AutoDock free energy function, corresponds to +0.31 Kcal mol^{-1} *per* methylene group. Thus, even thought the interaction between the substrate acyl chain and protein becomes weaker as the chain length increases, as long

as the specific energy of interaction is larger than the entropic loss per methylene group, the increase in the substrate chain lengths will end up in stronger binding (Fig. 5.5).

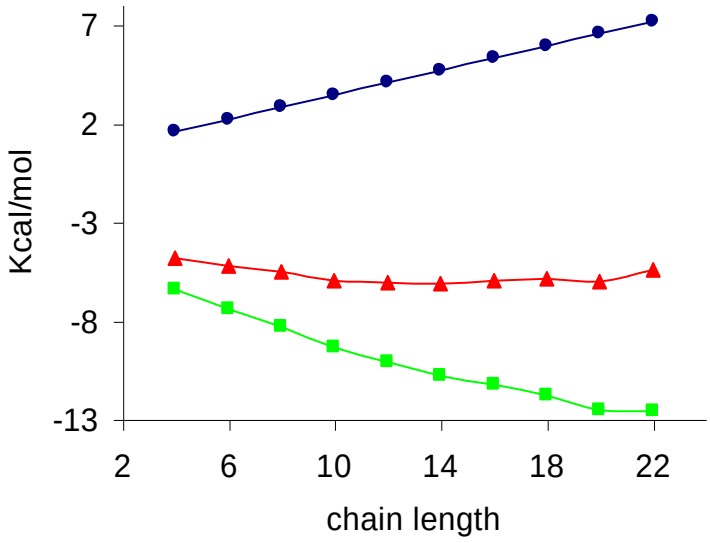
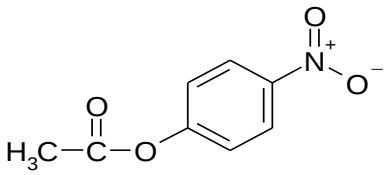


Figure 5.5. The estimated free energies are plotted against the acyl chain lengths: the free energy of binding (red triangles) is the combination of the internal energy (green squares) and the torsional free energy (blue circles). As long as the internal energy decreases faster than the increase of the torsional free energy, the increase of the chain length will end up in a stronger ligand – protein interaction.

Among the twelve substrates whose docking has been attempted, only those for the *p*-NP C2 failed. Actually a strong binding to PA3859 active site was also achieved for this substrate but it was not possible to dock it in the active conformation: in all of the found docking solutions, strong hydrogen bonding interaction occurs between the substrate's nitro group oxygen and the active site hydrogen bond donors, Ser 113 hydroxyl and main chain nitrogen atom of the oxyanion hole residues Leu 22 and Gln 114. This is not surprising since the negatively charged nitro group oxygen atoms are a much stronger hydrogen bond acceptors than a carbonyl oxygen atom.



Structure of *p*-NP acetate

Interestingly, even though substrate C4 and C6 could be docked in the active conformation, the overall lowest binding energies achieved were obtained for this non-active conformation, with the nitro group oxygen positioned in the catalytic site. The shorter substrate whose lowest binding energy was obtained for its active conformation is *p*-NP C8, suggesting that an acyl chain of at least 7 carbon atom is required to stabilize the productive positioning of *p*-NP substrates within the active site.

5.2.2 Polar cavity

While the function of the PA3859 hydrophobic cleft could be inferred from the significant structural similarity shared with the human acyl-protein thioesterase, no information were available that could suggest the binding specificity of the polar cavity.
The only indication came from the PAO1 genome sequence: the PA3859 locus was found to be located in the vicinity of two genes involved in the lipid/phospholipid metabolism: PA3857 has been demonstrated to be a phosphatidylcholine synthase [5] and PA3860 is hypothesized to be involved in lipid biosynthesis due to its 59% sequence homology to a long chain fatty acid – CoA ligase of *Archaeaglobus fulgidus* (http://www.pseudomonas.com).
Since genes sharing a similar function are often clustered in the same chromosomic region, this finding prompted us to hypothesize that PA3859 could also be involved in the metabolism of phospholipids. Furthermore, the presence in this protein of an hydrophobic binding site whose structural features are suited for the accommodation of only one acyl chain, suggested lysophospholipids to be the possible PA3859 physiologic substrates (see Chapter 1).
To assay this possibility the docking of the six polar head groups found in the most common phospholipids was attempted. Since the aim of this docking simulation was the evaluation of the binding properties of the enzyme polar cavity, the long lysophospholipid alkyl chain was replaced by a methyl group in order to speed up the searching space exploration.
The structures of the six 1-acetyl-*sn*-glycerophospholipids used in these docking simulations, namely 1-acetyl-*sn*-glycerophosphatidyl -serine (Ac-PS), -choline (Ac-PC), -ethanolammine (Ac-PE), -glycerol (Ac-PG), -inositol (Ac-PI) and 1-acetyl-*sn*-glycerophosphatidic acid (Ac-PA), are reported in Fig. 5.6.

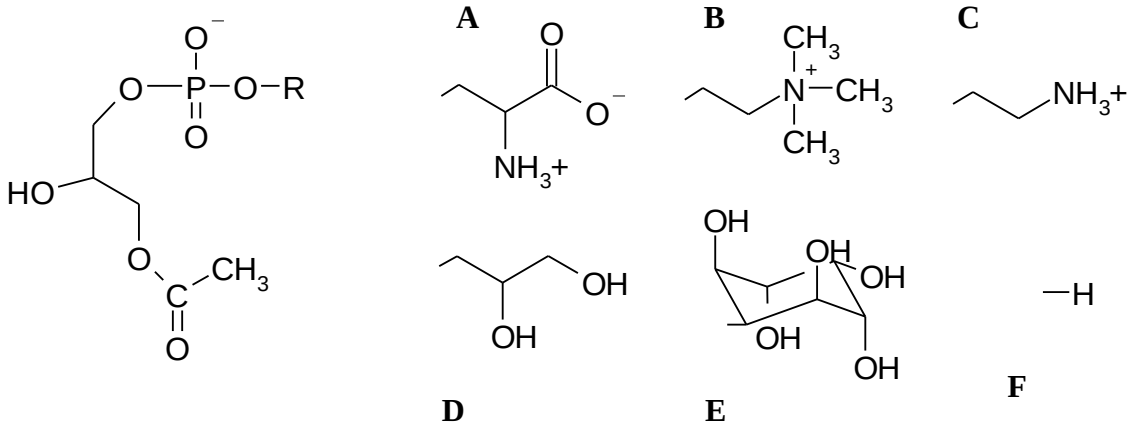


Figure 5.6. Structure of the 1-acetyl-*sn*-glycerolphosphatidates used to assay the polar cavity binding properties. R is replaced by the headgroup A in acetyl-phosphatidylserine (Ac-PS), B in acetyl-phosphatidylcholine (Ac-PC), C in acetyl-phosphatidylethanolammine (Ac-PE), D in acetyl phosphatidylglycerol (Ac-PG), E in acetyl-phosphatidylinositol (Ac-PI), F in acetyl-phosphatidic acid.

Acetyl-phosphatidyl -serine, -ethanolamine, -choline and -glycerol were successfully docked while it was not possible to dock phosphatidic acid and the inositol derivate.
The binding energies for the docked substrates are reported in Tab. 5.2 and in Fig. 5.7 their binding within the PA3859 polar cavity are shown.
The distance between the reference MES sulfur atom and the substrates carbonylic carbon range between 1.21 and 1.31 Å, and the carbonylic oxygen is always found to interact with the nitrogen main chain of the oxyanion hole residues Leu 22 and Gln 114.

In the polar cavity a deep gorge is present that is delimited by the Tyr 140 hydroxyl and main chain carbonyl groups of residues Ala 71, Pro 74, Ala 75 and Ala 77. As is shown in Fig. 5.7.**A**, the polar head of Ac-PE is positioned within this gorge, where strong hydrogen bonding occurs between its positively charged amino group and the main chain carbonyl oxygen atoms of Ala 71, Pro 74 and Ala 77. The phosphate group is located exactly in the middle of the cavity and is stabilized by the phosphorus - protein van der Waals' interaction. An hydrogen bond is also present between the catalytic Ser 113 Oγ and the free hydroxyl of the glycerol moiety: this interaction was indeed found in all of the four successfully docked substrates.

The overall structure of the docked Ac-PS is similar to that of Ac-PE (Fig. 5.7.**B**), but the positive Ac-PS amino group, being the polar cavity gorge occupied by the serine carboxylate, is positioned at the edge of the hole, where it is stabilized by hydrogen bonding with Ala 71 and Phe 72 main chain carbonyls. Within the hole hydrogen bonds are present between the substrate carboxylate oxygen atoms and Ala 77 main chain nitrogen and Tyr 140 hydroxyl. Due to the disadvantageous phosphate oxygen atoms – protein electrostatic interaction, the binding contribution of the phosphate moiety is estimated to be only of -0.14 Kcal mol^{-1}.

An almost identical structure with respect to that of Ac-PS is found for Ac-PG (Fig. 5.7.**C**), where the *sn*-2 and *sn*-3 hydroxyls of the glycerol headgroup replace the Ac-PS amino and carboxyl group respectively. Here the absence of a charged polar head capable to form strong hydrogen bonds with the polar cavity results in a weaker protein – substrate interaction (Tab. 5.2).

Being the positively charged choline headgroup too big to be accommodated within the polar cavity hole, a completely different conformation is adopted by Ac-PC (Fig. 5.7.**D**). Here the phosphate group is stabilized by the formation of a strong hydrogen bond with Tyr 140 hydroxyl, while the choline headgroup is positioned in the center of the polar cavity, mainly stabilized by a cation-π stacking with Tyr 140 aromatic ring and van der Waals' interaction with Val 168 and Phe 72 main chains.
The choline headgroup interaction in most of the phosphatidylcholine binding protein is driven by the cation-π stacking: typically, the quaternary ammonium is placed between two aromatic rings at a distance of 4.5 - 5 Å. [6].

In PA3859 this stabilization mode could arise from the rotation of Phe 72 side chain. In this modified orientation, the Phe 72 and Tyr 140 aromatic rings face one each other at a distance of 10 Å and the choline ammonium is located in the middle, roughly at the same distance between each of them.
In order to evaluate the effect of this protein structure modification on the binding of Ac-PC, a modified protein PDB file was generated using the program O, where the Phe 72 aromatic ring was rotated as shown in Fig. 5.7.**D**. A further docking job was then carried out using this modified PA3859 structure. The final solution displayed a very small conformational deviation from the previously obtained conformation, corresponding to a RMSD of 0.78 Å, with the quaternary ammonium displaced of 0.17 Å but the binding energy dropped from -4.61 Kcal mol^{-1} to -5.34 Kcal mol^{-1}, with a gain in term of binding constant from 416 μM^{-1} to 122 μM^{-1} (Tab. 5.2).

	Intermolecular Energy (Kcal mol^{-1})	Ligand Internal energy (Kcal mol^{-1})	Torsional Free energy (Kcal mol^{-1})	Binding Free Enrgy (Kcal mol^{-1})	Binding Constant (μM^{-1})
Ac-PS	-9.01	-1.39	3.11	-5.90	48
Ac-PE	-8.73	-0.69	2.80	-5.93	45
Ac-PG	-8.23	-0.23	3.11	-5.11	179
Ac-PC	-7.72	-0.45	3.11	-4.61	416
Ac-PC*	-8.78	-0.33	3.11	-5.34	122

Table 5.2. Binding energies and constants for acetyl-phosphatidates. The strongest binders resulted to be Ac-PE and Ac-PS. Phe 72 side chain re-orientation allowed a significant drop in the Ac-PC binding free energy (Ac-PC*).

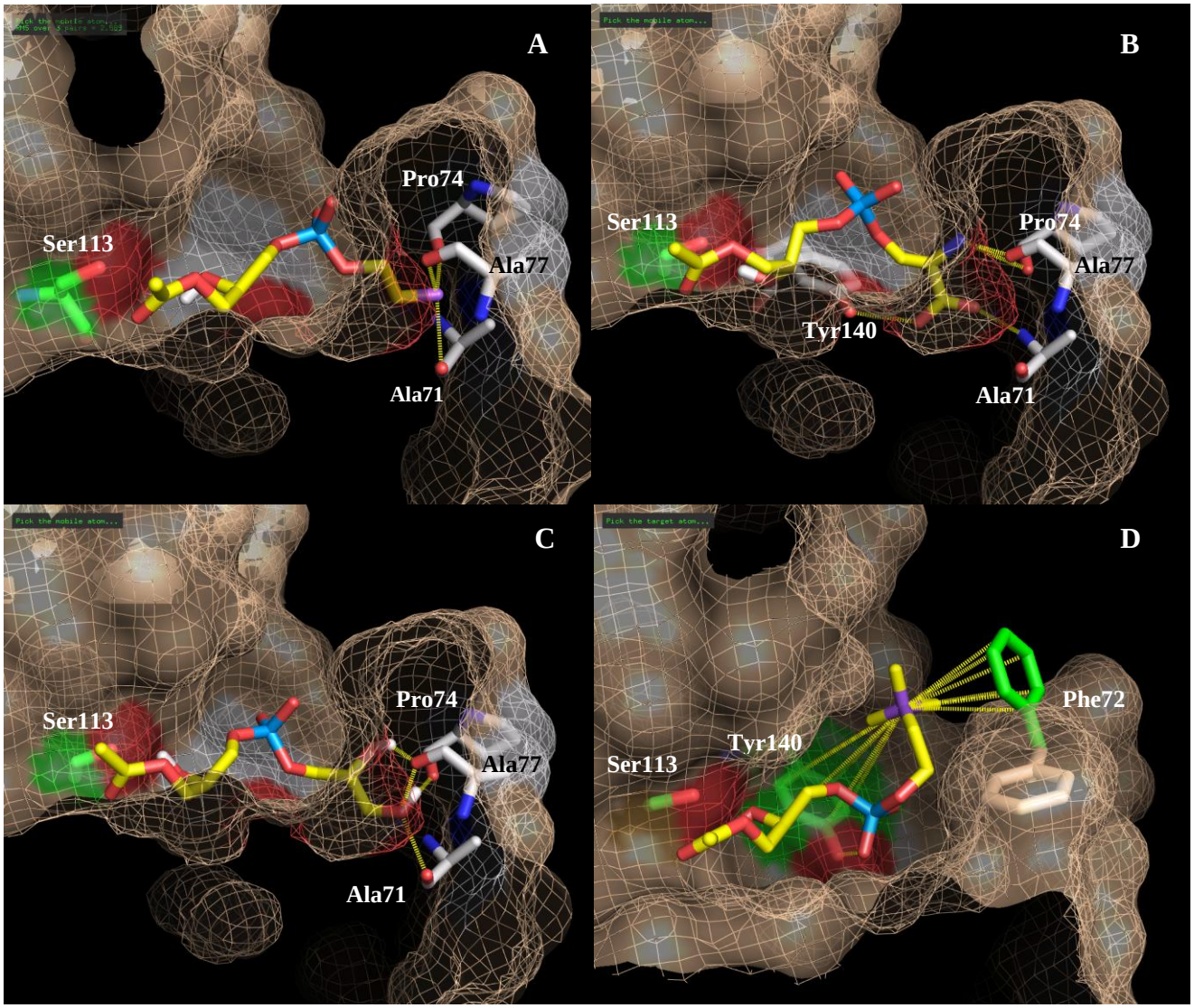


Figure 5.7. Docking of the acetyl phosphatidates. For all of the docked substrates (yellow) the carbonylic carbon is placed in the catalytic site (Ser 113 is shown as green stick) with the carbonyl oxygen oriented toward the oxyanion hole. **A**: the positively charged amino headgroup of Ac-PE is located into a deep gorge present in the bottom of the polar cavity and is stabilized by the formation of hydrogen bonds (highlighted in yellow) with the main chain carbonyl oxygen atom of Ala 71, Phe 72, Pro 74 and Ala 77. An overall similar interaction mode is also achieved for both Ac-PS and Ac-PG. In **B** the Ac-PS binding is shown: the headgroup carboxyl is positioned withino the hole, where it accepts hydrogen bonds from Ala 77 main chain nitrogen and Tyr 140 hydroxyl. The serine moiety amino group is instead located at the border of the hole and is stabilized by the hydrogen bond formation with Ala 71 and Phe 72 main chain carbonyl oxygen atoms. In **C** the binding mode of Ac-PG is reported: the conformation is almost identical to those achieved for Ac-PS: the *sn*-2 hydroxyl of the glycerol headgroup moiety is positioned at the border of the hole and forms hydrogen bond with Phe 72 and Pro 74, while the *sn*-3 hydroxyl is located into the hole where forms hydrogen bond with Ala 71 and Pro 74.
The completely different binding mode obtained for Ac-PC* is shown in **D**: the positively charged quaternary ammonium is positioned in the middle of the polar cavity, roughly 5.5 Å apart from the Tyr 140 aromatic ring. The docking run carried out using a modified structure of PA3859, in which the Phe 72 was rotated (rotated Phe is highlighted in green) produced a very similar solution (RMSD 0.78 Å) but with a significantly lower binding energy. The choline headgroup is mainly stabilized by

cation-π stacking with the aromatic ring of Tyr 140 and Phe 72 that in this modified structure are roughly 10 Å far one from the other. The distance between ammonium and Phe 72 aromatic ring is roughly 4.5 Å.
Furthermore an hydrogen bond is formed between Tyr 140 hydroxyl and the phosphate group oxygen.

5.3 Conclusions

According to the results produced by AutoDock, acyl chain as long as 20 carbon atoms can be bound by PA3859 hydrophobic cleft, even thought the cleft seems to be suited to accommodate somewhat shorter chains.
Moreover these docking simulations have provided insights into the binding mechanism of the polar cavity: the lipid polar head are mainly stabilized by hydrogen bonding between their carboxyl and/or amino group and the protein main chain carbonyl oxygen atoms and nitrogen atoms. In the case of choline headgroup the binding is likely to be achieved by the classical cation-π staking between quaternary ammonium and two aromatic residues side chains. The lack of a strong interaction between binding site and the substrate phosphate moiety is probably the reason of the unsuccessful docking of Ac-phosphatidyc acid, while phosphatydilinositol polar head is likely to be simply too big to fit within the protein polar cavity.

Overall, the docking simulations so far presented suggested that PA3859 protein binding site indeed possesses the structure feature required to likely be a lysophospholipase.

5.4 References

[1] Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Comp. Chem.* (1998) **19**:1639-1662.

[2] Goodsell, D. S., Morris, G. M. and Olson, A. J. Docking of Flexible Ligands: Applications of AutoDock. *J. Mol. Recognition*. (1996) **9**: 1-5.

[3] Gasteiger, J., Marsili, M. *Tetrahedron*. (1980), ***36***: 3219-3228.

[4] Kollman, R. *JACS* (1976) **98**: 3335-7811.

[5] Wilderman, P. J., Vasil, A. I., Martin, W. E., Murphy, R. C., Vasil, M. L. *Pseudomonas aeruginosa* synthesizes phosphatidylcholine by use of the phosphatidylcholine synthase pathway. *J. Bacteriol*. (2002) **184**: 4792-4799.

[6] Anbazhagan, V., Swamy, M. J. Thermodynamics of phosphorylcholine and lysophosphatidylcholine binding to the major protein of bovine seminal plasma, PDC-109. *FEBS Letters*. (2005) **579**: 2933-2938.

Chapter 6

Enzymatic characterization

6.1 Introduction

A wealth of information on PA3859 based on the *P. aeuginosa* genome sequence, its 3D X-ray crystal structure and the results of automated docking simulations with possible substrates, led to hypothesize PA3859 to be a lysophospholipase rather than a carboxylesterase, with a preference for ethanolamine and serine over choline and glycerol polar headgruop and an optimal substrate alkyl chain between 16 and 20 carbon atom.
In order to confirm the real function and to shed light on this enzyme specificity and mechanism of action, enzymatic assays were performed.

As a first step better targeted studies compared to those reported in Chapter 2 were carried out using *p*-nitrophenyl esters in order to empirically evaluate the enzyme specificity with respect to the substrate acyl chain length.

The ability of PA3859 to release free fatty acid from lysophospholipids was assayed by TLC analysis and the kinetic parameters for the hydrolysis of lysophopshatidyl choline were then determined.

6.2 Hydrolysis of *p*-nitrophenyl esters

In the preliminary PA3859 biochemical characterization reported in Chapter 2, the influence of the substrate acyl-chain length was evaluated simply by incubating the enzyme with the α-naphtyl esters of five carboxylic acids and measuring their overall velocity of hydrolysis (Tab. 2.3). The information provided by these experiments however, even though useful to make comparison with other carboxylesterases, do not allow a reliable evaluation of the enzyme substrate specificity.

As reviewed in the milestone paper by Berg *et al* "Interfacial Enzymology: the Secreted Phospholipase A_2-Paradigm" [1], the accurate and detailed studies of a lipolytitc enzyme kinetic are far more complex than those of the classical homogeneous-phase catalysis in which water-soluble substrate are implicated: difficulties arise from the presence of a two phase system, with reactions occurring at the water-lipid interface. The resolution of all the kinetic details required to fully describe the enzyme – substrate interaction in such system are thus out of the scope of the present work.
However, in order to gain deeper inside into the PA3859 substrate specificity, and thus into its physiological function, a further kinetic study was attempted using *p*-nitrophenyl esters as substrates.

The PA3859 ability of hydrolyzing *p*-nitrophenyl ester of acetic (C2) to stearic acid (C18) was assayed. While acetic, propionic and butyric acid display a water-solubility large enough to allow the K_{cat} and K_m determination in a homogeneous phase system, the low solubility of substrate carrying a longer acyl chain required the presence of a detergent in the reaction mixture. Thus, the hydrolysis of *p*-nitrophenylesters of carboxylic acids longer than C4 occurred mainly at the water-micelle interface. This implies a completely different mechanism of action, making problematic a meaningful comparison between the kinetic parameter found for the two classes of substrates. Furthermore, as will be discuss later on, the presence of a detergent in the reaction mixture where soluble substrates were assayed, produced a non-competitive-like inhibition of PA3859.

For the soluble substrates *p*-nitrophenyl acetate, propionate and butyrate, the obtained data fit the Michaelis-Menten equation. The Lineweaver-Burk plot was found to be linear in the range between 0.01 to 5 mM for all three substrates.
For the water-insoluble substrates (C5 to C18) the classical Michaelis-Menten model is not adequate to describe the reaction mechanism. Since the enzyme-substrate

interaction occurs at the water-micelle interface, the reaction rate can not be considered to be simply dependent on the substrate concentration in the bulk solution. According to the micellar surface dilution model [2, 3], to describe this micellar catalysis, one additional step has to be included in the classical Michaelis-Menten model:

$$E + M \underset{K_{-1}}{\overset{K_1}{\rightleftharpoons}} EM$$

$$EM + S \underset{K_{-2}}{\overset{K_2}{\rightleftharpoons}} EMS \xrightarrow{K_3} P$$

$$V = \frac{V_{max}[M]\chi_s}{Km\,Ks + Km[M] + [M]\chi_s} \qquad (1)$$

The first step consists in the interaction of the enzyme (E) with the micelle surface (M) and the formation of the EM complex. Once the EM complex, also called the penetration complex, has formed, the enzyme-substrate interaction takes place, with the formation of the active complex EMS that will end up in the hydrolysis of the substrate and the release of the product P. Thus, the Michaelis-Menten equation results modified as reported in equation 1, where [M] represents the total molar concentration of the mixed micelle (detergent + substrate) in the bulk solution and χ_s is the molar fraction of the substrate on the micelle surface. Calculation of the amount of the detergent in the micelle has to take into account the amount of detergent present in solution (for Triton X-100 the critical micelle concentration is assumed to be 240 μM). The fitting parameter K_s represents the dissociation constant (K_{-1}/K_1) for the micelle-enzyme complex, K_m is the Michaelis constant ((K_3+K_{-2})/K_2) on the micellar surface and V_{max} is the catalytic turnover rate (K_3).

It is useful to consider the expression for V as a function of [M] when χ_s is held constant (2) and its reciprocal form (3):

$$V = \frac{[M]\dfrac{V_{max}\chi_s}{Km + \chi_s}}{[M] + \dfrac{Ks\,Km}{Km + \chi_s}} \qquad (2)$$

$$\frac{1}{V} = \frac{Ks\,Km}{V_{max}\,\chi_s}\frac{1}{[\mathrm{M}]} + \frac{1}{V_{max}}\left(1 + \frac{Km}{\chi_s}\right) \quad (3)$$

Equation 3 shows that if the initial rate studies are performed under condition in which the substrate molar fraction within the mixed micelle χs is held constant while the total micellar concentration [M] is varied, the plot of 1/*V versus* 1/[M] gives linear curves (Fig. 6.1.**A**). Replot of the these slopes and intercepts as a function of χs allows the determination of V_{max}, K_s and K_m. Specifically, a plot of the 1/*V* intercepts obtained in Fig. 6.1**A** for each value of χ_s plotted as a function of $1/\chi_s$, is linear and the intercept on the 1/*V* axis is $1/V_{max}$ and the intercept on the $1/\chi_s$ axis corresponds to $-1/K_m$ (Fig. 6.1.**B)**. The plot of the slope obtained for each χs value in Fig. 6.1.**A** as a function of 1/[M] is also linear and the slope is K_sK_m/V_{max} (Fig. 6.1.**C**). Thus the value of K_s can be determined from this curve slope and the values of K_m and V_{max} can be derived from the plot in Fig. 6.1.**B**.

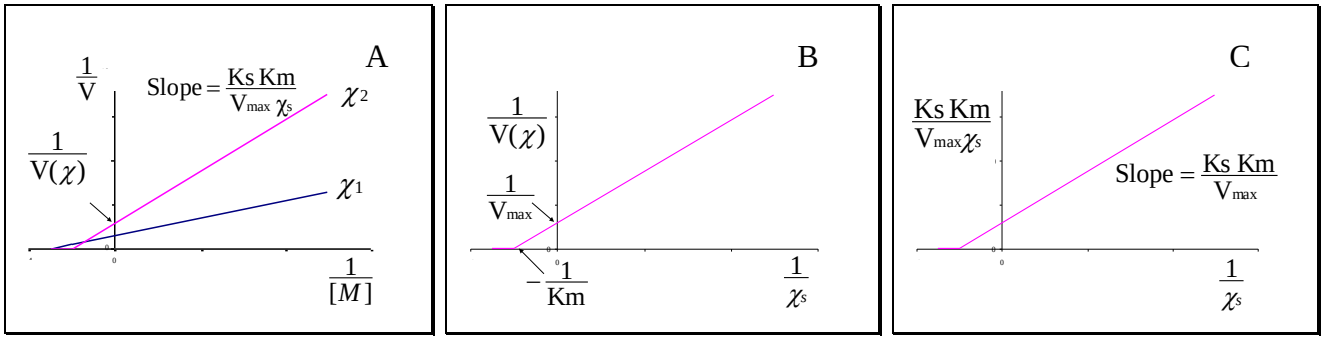


Figure 6.1. **A**, plot of 1/*V* as a function of 1/[M] measured for constant substrates molar fraction (χs). **B**, Intercepts on the 1/*V* axis obtained in A for each χ_s value are plotted as a function of $1/\chi_s$. **C**, The slopes of the curves obtained in A for each χ_s value are plotted as a function of $1/\chi_s$.

For the water-soluble substrates *p*-NP C2, C3 and C4 the increase of the acyl chain length gives rise to a rapid decrease of the K_m value (Tab. 6.1, Fig. 6.2**A**). The presence in the reaction mixture of a 0.2% Triton X-100 does not affect significantly the enzyme affinity for these substrates but produce a reduction of the reaction rate of roughly 70%, mimicking a non competitive-like inhibition (Fig. 6.2.**B**). Being these short chain substrates not expected to interact with the hydrophobic interior of the micelles, the reduction of the observed K_{cat} suggests that a strong interaction between PA3859 and the micelle surface occurs, thus reducing the real concentration of the available enzyme active sites.

Substrate	*Km* (μM)	*Ks* (μM)	*Kcat* (S^{-1})	*Kcat*/*Km* ($M^{-1}S^{-1} \times 10^6$)
p-NP C2	486 ± 116	-	6400 ± 880	13.3 ± 2.1
p-NP C2 +Tx-100	475 ± 154	-	2130 ± 430	4.4 ± 0.9
p-NP C3	315 ± 93	-	5700 ± 920	18.1 ± 2.8
p-NP C3 +Tx-100	332 ± 82	-	1900 ± 310	6.0 ± 1.2
p-NP C4	233 ± 74	-	1700 ± 350	7.3 ± 0.8
p-NP C4 +Tx-100	246 ± 61	-	700 ±120	3.0 ± 0.7
p-NP C5	179 ± 42	84 ± 13	1900 ± 190	10.8 ± 1.6
p-NP C6	113 ± 55	98 ± 17	4500 ± 320	39.4 ± 3.1
p-NP C8	75 ± 26	83 ± 12	4600 ± 280	61.3 ± 3.8
p-NP C10	80 ± 32	75 ± 13	1800 ± 100	22.8 ± 3.4
p-NP C12	85 ± 35	96 ± 19	180 ± 40	1.5 ± 0.2
p-NP C14	91 ± 34	106 ± 32	120 ± 20	1.2 ± 0.3
p-NP C16	96 ± 21	91 ± 30	50 ± 20	0.5 ± 0.1
p-NP C18	98 ± 32	76 ± 11	50 ± 20	0.5 ± 0.2

Table 6.1. PA3859 kinetic parameters for the hydrolysis of *p*-nitrophenyl esters. For soluble substates *p*-NP C2, C3 and C4, the data obtained in the presence of 0.2% Triton X-100 are also reported. K_s values corresponds to the micellar concentration exceeding the Triton X-110 CMC (240 μM): a *Ks* of roughly 90 thus corresponds to a micellar concentration of 330 μM.

For substrates longer than C4, due to their low water-solubility, the assays were conducted in the presence of a detergent according to the surface dilution model described above. Triton X-100 was used to allow the substrate solubilization. The highest K_{cat}/K_m ratio value is achieved by *p*-NP C8 that displays the lower K_m and the highest K_{cat}. K_m values do not differ significantly for substrates with chain length ranging between C8 to C18, while a dramatic reduction in the K_{cat} values are found for substrates longer than C10 (Fig. 6.2). The affinity for the micelle, as estimated by the K_s parameter, is also independent from the substrate chain length and roughly corresponds to a overall micellar concentration of 340 μM. However, being the critical micelle concentration (CMC) for Triton X-100 in the assay condition equal to 240 μM, the relevant value for K_s is roughly 100 μM (Tab. 6.1).

These data are consistent with those found for many other lipolytic enzymes [4, 5, 6, 7], where the optimal substrate chain length was found to be 8 carbon atom. Interestingly, to the best of ours knowledge, when the kinetic studies of lipases or phospholipases were carried out using *p*-nitrophenyl esters in a detergent mixed micelle system, the optimal substrate chain length has never been found to be longer than C8, suggesting

that some kind of artifact is likely to occur. This hypothesis is also supported by another consideration: the fatty acids found in the membrane are long at least 16 carbon atoms, thus lipids and phospholipids carrying 8 carbon atom alkyl chains are not expected to have any physiological relevance,. Furthermore, according to the docking simulation presented in Chapter 5, the PA3859 hydrophobic cleft seems to be suited to interact with substrates much longer than C8. All this suggest that the K_{cat}/K_m ratio determined for the mixed micelle system is not likely to provide a reliable evaluation of the *in vivo* enzymes substrate specificity.

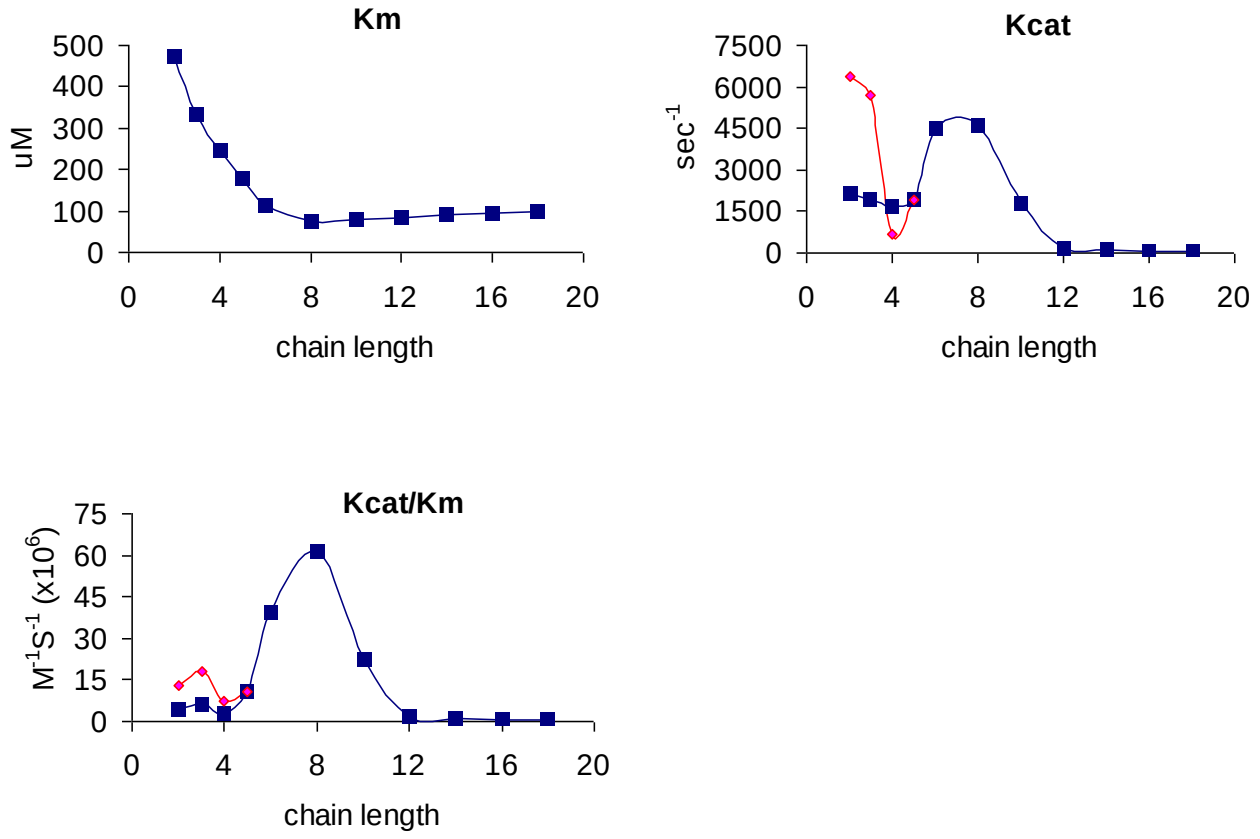


Figure 6.2. PA3859 kinetic parameters for the hydrolysis of *p*-nitrophenyl ester plotted as a function of the substrate chain length. The parameters concerning the hydrolysis of soluble substrates determined in the absence of Triton X-100 are shown in red.

6.3 PA3859 hydrolysis of lysophospholipids

The kinetic study presented above, even though is likely to be affected by an artifact that does not allow to clearly asses the protein preference for the substrates chain length, showed that PA3859 is able to hydrolyze esters of fatty acids as long as 18

carbon atoms. Furthermore, the binding site study carried out by means of automated docking simulations suggested PA3859 to be a lysophospholipase. In order to assay its ability to hydrolyze lysophospholipids the enzyme was extensively incubated with the following lysophospholipids: 1-palmitoyl-*sn*-phosphatidyl -choline, -serine, -ethanolamine, -glycerol, -inositol and 1-palmitoyl-*sn*-phosphatidic acid. To evaluate the enzyme activity against phospholipids the hydrolysis of 1,2-dipalmitoyl phosphatidylcholine was also assayed. The release of palmitic acid was evaluated by TLC analysis. As it is shown in Fig. 6.3, a significant release of palmitic acid was detected when the enzyme was incubated with lysophosphatidyl ethanolammine, serine, choline and glycerol while no activity was found with lysophosphatidic acid, lysophosphatidylinositol, and phosphatidylcholine as substrates.

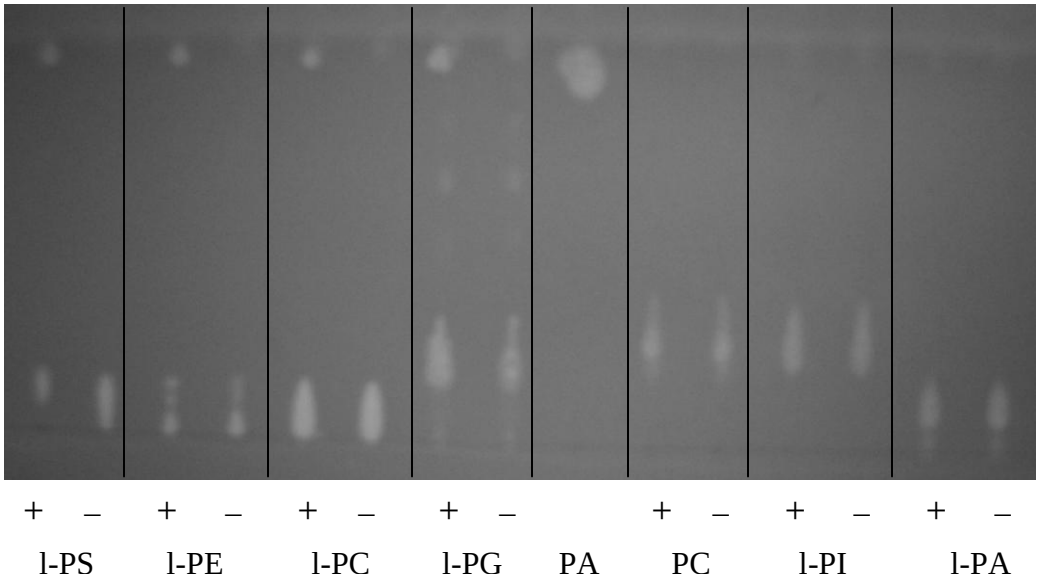


Figure 6.3. TLC analysis of the PA3859 lysophospholipolytic activity: l-PS, lysophosphatidylserine. l-PE, lysophosphatidylethanolamine. l-PC, lysophosphatidylcholine. l-PG, lysophosphatidylglycerol. PA, palmitic acid. PC, phosphatidylcholine. PI, lysophosphatidylinositol. l-PA, lysophosphatidic acid. Each substrate was incubated for 5 hours at 37 °C with 1 μg of PA3859 (lanes +) and without the enzyme (lanes –).

This result shows that PA3859 is indeed able to hydrolyze lysophospholipids and suggest a specificity for the polar headgroups that well agrees with the docking simulation results. This experiment however, being merely qualitative, does not provide any information on the kinetic. Thus it is not possible to evaluate the enzyme specificity for the different phospholipid polar headgroups nor the reaction kinetic parameters.

To further investigate the enzyme hydrolysis of lysophospholipids a preliminary kinetic study was perfomed using radiolabeled 1-palmitoyl-*sn*-phosphatidylcholine as substrate

(see paragraph 6.4.2). With respect to long chain *p*-nitrophenyl esters, lysophospholipids displays a higher water-solubility and, more important, can form pure micelle with no need for the addition of a detergent. This allowed the determination of the kinetic parameter according to the classical Michaelis-Menten model. As shown in Fig. 6.4, the PA3859 catalytic activity against lysophosphatidylcholine exhibits saturation kinetics. However, the data did not fit well the Michaelis-Menten equation (red curve). There seems to be some cooperativity effect, since the data fit to the Hill equation with n=1.58 ± 0.12, V_{max} 0.76 ± 0.05 μmol $min^{-1}mg^{-1}$ protein and substrate concentration at half-maximal saturation is 24.0 ± 2.3 μM. Furthermore, there is no dramatic increase in the enzyme activity near the CMC of the substrate (7 μM), indicating that the enzyme hydrolyzes the substrate both in monomeric and micellar form.

To determine the dependence of the enzyme activity on the substrate surface concentration, measurements with Triton X-100 - lysophosphatidyl choline mixed micelle were also performed. When the bulk l-PC concentration is held constant and the detergent concentration is varied, the enzyme activity increases as the molar fraction of the substrate within the micelle increases and the data fit the Michaelis-Menten equation with a V_{max} of 0.92 ± 0.11 μmol $min^{-1}mg^{-1}$.

Furthermore, the surface dilution kinetic model was also tested. The parameter values obtained are: V_{max} = 0.83 ± 0.06 μmol/min/mg, K_m = 21.1 ± 3.7 μM and K_s = 85.4 ± 4.4 μM.

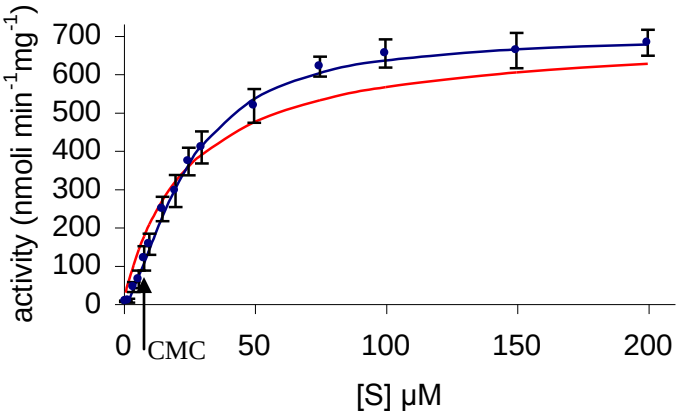


Figure 6.4. PA3859 activity as a function of l-PC concentration. Data were fitted to Michaelis-Menten equation (red curve) and to Hill model (blue curve). Each point represents the average of five independent experiments, each in triplicate. The l-PC critical micelle concentration (CMC) is indicated by the arrow.

Overall the kinetic data so far reported are consistent with the following three conclusions: first, even though the V_{max} is achieved for a substrate concentration above the CMC, no interfacial activation seems to occur. This finding is in agreement with the lack in PA3859 structure of a lid domain as those typically found in the real lipases (see

Chapter 1) that are though to be responsible for their interfacial activation mechanism. Second, the fact that a similar value of K_s fits all the experimental data suggests that PA3859 binds to the mixed-micelles with the same affinity regardless of the micellar composition. Third, the dependence of the apparent V_{max} on l-PC molar fraction within the mixed micelle is consistent with a single binding site for the substrate. This 1:1 stoichiometry of substrate-enzyme complex suggests that the apparent cooperativity observed in the kinetics of PA3859 on pure l-PC micelles results from the change in the substrate presentation (monomers *vs* premicellar aggregates *vs* micelles) rather than a cooperative binding of two or more substrate molecules to catalytic or allosteric sites [8].

6.4 Materials and methods

6.4.1 Hydrolysis of *p*-nitrophenylesters

Reaction were carried out at room temperature in Tris-HCl 50 mM, pH 9.0 in the presence of 2 µg ml^{-1} of PA3859. Formation of the reaction product *p*-nitrophenol was followed measuring the increase in the absorbance at 400 nm using a ULTRASPEC 3000 spectrophotometer (Pharmacia Biotech).
To measure the activity as a function of the overall micelle concentration [M] and at a constant molar fraction (χ_s), substrates were first dissolved in chloroform and than aqueous solutions in Tris-HCl 50 mM, pH 9.0 and Triton X-100 5 mM were prepared. For each substrate 10 different solutions were prepared with a substrate molar fraction χ_s spanning the range 0.2% to 10% (corresponding to a substate concentration between 10 µM and 500 µM). Reactions were started adding to the reaction mixture 10 µl of a 0.2 µg μl^{-1} PA3859 solution.
Substrates were purchased from Sigma Chemicals, Triton-X100 from Fluka.

6.4.2 Hydrolysis of lysophospholipids

To assay the PA3859 ability to hydrolyze lysophospholipds, 200 µg of substrates were incubated with 1 µg of enzyme in 80 µl of Tris-HCl 50 mM, pH 9.0, at 37 °C for 5 hours. Lipids were than extracted in an organic solvent according to the Bligh protocol in order to remove buffer and protein [9]. The chloroform phase was subsequently spotted on the TLC plate (Silica Gel Plates, Merck). Before use, the plate was wetted with a 2.3% (w/v) boric acid solution and then dried for 15 min. at 100 °C in an oven. The TLC plate was developed in a tank containing a solvent mixture of chloroform-

methanol-water in ratio 65:25:4 (vol/vol/vol). For lipid visualization the plate was sprayed with a primuline solution (5 mg/100ml in acetone-water 80:20 (v/v)) and lipids were detected under UV light as fluorescence quenched spot.
Lysophosphatidyl -serine, -ethanolamine, -glycerol and -inositol were purchased from Avanti Polar Lipids. Lysophosphatidylcholine, phosphatidyl choline, phosphatidic acid, palmitic acid and primuline were from Sigma Chemicals.

To determine the PA3859 lysophospholipase activity the radiolabeled substrate 1-[^{14}C]palmitoyl-*sn*-phosphatidyl choline was used (Amersham Biosciences, 112 µCi mg^{-1}). Reactions were carried out at room temperature in 50 µl Tris-HCl 50 mM, pH 9.0. The assay was started by adding 50 ng of enzyme to the reaction mixture. The released fatty acid was extracted by the Dole method [10]: the reaction was stopped by adding to the reaction mixture 250 µl of a isopropanol:heptane:sulfuric acid (25:85:1 v/v). The solution was heated at 60 °C for 3 min and then chilled on ice for 2 min. Subsequently 150 µl of heptane and 125 µl of water were added. The mixture was then vortexed, centrifuged for 1 min at 13000 rpm in a bench centrifuge and 350 µl of the upper organic phase containing the released radiolabeled palmitic acid was finally recovered. To quantify the content of palmitic acid, the recovered organic phase was mixed with 5 ml of scintillation liquid Ready Safe (Beckman) and the β emission measured using a TRI-CARB TR β-counter (Packard).

6.5 References

[1] Berg, O., G., Gelb, M., H., Tsai, M., Jain, M., K. Interfacial Enzymology: the Secreted Phospholipase A_2-Paradigm, *Chem. Rev.* (2001) **101**: 2613-2653.

[2] Deems, R., A., Eaton, B., R., Dennis, E., A. Kinetic Analysis of Phospholipase A_2 Activity toward Mixed Micelles and Its Implication for the Study of Lipolytic Enzymes. *J. Biol. Chem.* (1975) **250**: 9013-9020.

[3] Burdette, R., A., Quinn, D., M. Interfacial Reaction Dynamics and Acyl-enzyme Mechanism for Lipoprotein Lipase-catalyzed Hydrolysis of Lipid *p*-nitrophenyl Esters. *J. Biol. Chem.* (1986) **261**: 12016-12021.

[4] Burdette, R., A., Quinn, D., M. Interfacial Reaction Dynamics and Acyl-enzyme Mechanism for Lipoprotein Lipase-catalyzed Hydrolysis of Lipid *p*-Nitrophenyl Esters. *J. Biol. Chem.* (1986) **261**:12016-12021.

[5] Alquati, C., De Gioia, L., Santarossa, G., Alberghina, L., Fantucci, P., Lotti, M. The cold-active lipase of *Pseudomonas fragi*. *Eur. J. Biochem.* (2002) **269**: 3321-3328.

[6] Santarossa, G., Gatti Lafranconi, P., Alquati, C., De Gioia, L., Alberghina, L., Fantucci, P., Lotti, M. Mutations in the "lid" region affect chain length specificity and thermostability of a *Pseudomonas fragi* lipase. *FEBS*. (2005) **579**: 2383-2386.

[7] Redondo, O., Herrero, A., Bello, J., F., Roig, M., G., Calvo, M., V., Plou, F., J. Burguillo, F., J. Comparative kinetic study of lipase A and B from *Candida rugosa* in the hydrolysis of lipid *p*-nitrophenyl esters in mixed micelles with triton X-100. *Biochim. Biophys. Acta*. (1995) **1243**: 15-24.

[8] Wang, A., Yang, H., Friedman, P., Johnson, C., A., Dennis, E., A. A specific human lysophospholipase: cDNA cloning, tissue distribution and kinetic characterization. *Biochim. Biophys. Acta.* (1999) **1437**: 157-169.

[9] Bligh, E., G., Dyer, W., J. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology.* (1959) **37:** 911-917.

[10] Dole, V., P. J. A relation between non-esterified fatty acids in plasma and the metabolism of glucose. *J. Clin. Invest*. (1956) **270**:150-154.

Chapter 7

Study on the PA3859 *in vivo* function

7.1 Introduction

The enzymatic assays described in Chapter 6 demonstrated that PA3859 is indeed capable to hydrolyze lysophospholipids *in vitro*. In order to assess the role played *in vivo* by PA3859, further experiments were performed using a *Pseudomonas aeruginosa* mutant in which the expression of the PA3859 gene was suppressed (knock out mutant). Moreover, being lysophospholipids and lysophospholipases known to be involved in the bacterial infective process (see Chapter 1), a collaboration was initiated with an external laboratory in order to evaluate the possible involvement of PA3859 as a virulence factor. Here the preliminary results are reported.

The Pseudomonas aeruginosa PA3859 knock out (K.O.) mutant was purchased from the University of Washington Genome Center: within their "Pseudomonas Aeruginosa Project" a near-saturated library of mutant PAO1 *strains was generated, each with a unique gene suppression, by means of a high-throughput methods for producing sequence-defined transposon insertion mutation (http://www.genome.washington.edu/UWGC/pseudomonas/index.cfm). All the produced mutants are freely available.*

7.2 PA3859 involvement in the phospholipids metabolism

The fact that the PA3859 knock out is viable lead to the first conclusion that this protein does not accomplish any essential function or that in *P. aeruginosa* other enzymes with similar activity exist.
Anyway, if as expected PA3859 acts *in vivo* as a lysophospholipase, the knock out mutant should display some deficiency in the phospholipids metabolism. This possibility was assayed by growing both the PAO1 wild type and the K.O. mutant in a minimal medium containing phospholipids as the unique carbon source. Bacteria were first grown in LB at 37 °C to an $O.D._{600}$ of 0.4 and diluited 1:10 in a minimal medium (M9) containing the required inorganic nutrients and 0.2% Asolectin (Sigma Chemicals) as carbon source, and incubated for 32 hours at 37 °C. Asolectin is a mixture of phospholipids extracted from soybean containing roughly an equal proportion of phosphatidyl -choline, -serine and -inositol. As shown in Fig. 7.1, almost no difference was found in the growth rate on LB (**A**). On the contrary when bacteria were fed with phospholipids as the sole carbon source, a clear reduction in the growth rate of the PA3859 K.O. mutant was found with respect to the PAO1 wild type (**B**), suggesting a role of PA3859 in the phospholipids metabolism.

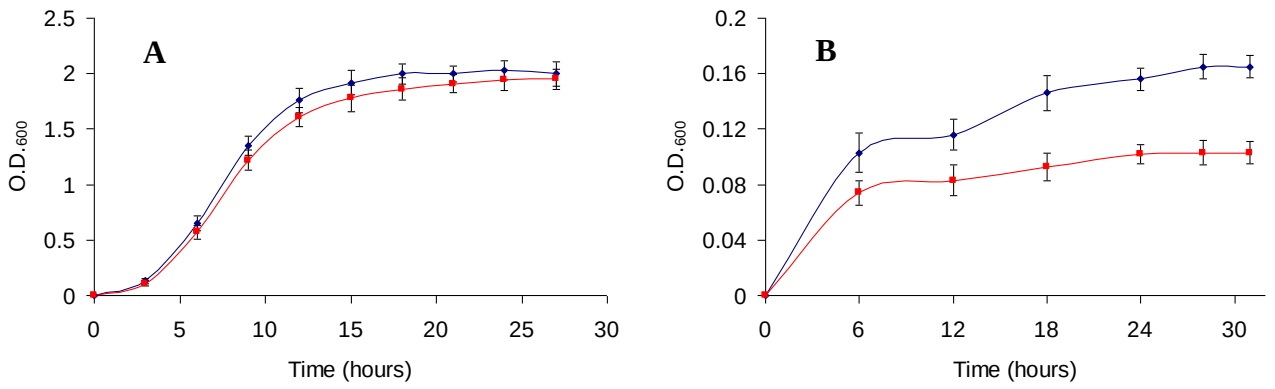


Figure 7.1. Growth curve of wild type PAO1 (blue curve) and PA3859 knock out mutant (red curve) on LB (**A)** and on a minimal medium containing phospholipids as the sole carbon source (**B**). The density of the bacterial cultures was quantified measuring the light scattering at 600 nm. The plots are the mean of three independent experiments.

Furthermore the final bacterial density reached by the wild type culture was roughly 1.5 fold higher than the K.O. mutant. A possible explanation for this finding could reside in the cytolytic effect of lysophospholipid: lysophospholipids are known to be toxic due to their membrane-perturbing properties. The incorporation of these molecules into a phospholipidic bilayer provokes a membrane destabilization that lead to free ion permeation and subsequent osmotic rupture of the cell [1]. Indeed, it has been suggested that one of the roles of lysophospholipases, besides to the implication in the

phospholipids catabolism, is the maintenance of the lysophospholipids at safe levels [2, 3]. Thus the lower density reached by the PA3859 K.O. mutant culture could be due to its impossibility to scavenge the lysophospholipids produced along with the phospholipid catabolism.

To assay the possible involvement of PA3859 in the lysophospholipid detoxification, both PAO1 wild type and the PA3859 K.O. mutant were plated on LB-agar containing up to 500 µM phosphatidylcholine and grown overnight at 37 °C.
If PA3859 function concerned the lysophospholipid detoxification a much slower growth, or even the complete impossibility to grow, would have been expected. No differences were observed in the growth of wild type and the K.O. mutant. This result however does not mean that PA3859 is not involved in the detoxification of lysophospholipids: in the *P. aeruginosa* genome in fact, the ORF PA2856 (*http://www.pseudomonas.com*) codifies for a protein that has been supposed to be a lysophospholipase due to its 65% sequence identity with the acetyl-CoA thioesterase I/lysophopsholipase of *E. coli* [4]. This protein is secreted in the periplasm and could be responsible for the detoxification of the exogenous lysophospholipids. Thus, besides its involvement in the phospholipids catabolism, a role of PA3859 in the detoxification of endogenous lysophospholipid is still possible.

7.3 PA3859 as a possible *P. aeruginosa* virulence factor

Pathogen microorganisms are known to secrete lytic enzymes. Among these, phospholipase play a predominant role, being involved in the partial digestion of the host cell plasma membrane [5,6]. Specifically, in *Pseudomonas aeruginosa*, one of the most important secreted virulence factor, ExoU, has a phospholipase/lysophospholipase activity [7, 8, 9]. Furthermore, among the enzymatic activity required for the *P. aeruginosa* persistence in the lungs of the infected host, an essential role is also played by phospholipases of type C and D, suggesting an involvement of the phospholipids metabolism in the bacterial infectivity [10, 11], (see Chapter 1).
To evaluate the possible contribute of PA3859 to the *P. aeruginosa* pathogenicity, we started a collaboration with a laboratory involved in the *in vivo* study of the *P. aeruginosa* - host interaction (Dr. Roger C. Levesque group at Facultè de Mèdecine, Universitè Laval, Stefoy, Quebèc, Canada). The ability of the PA3859 knock out mutant to compete with the wild type PAO1 was examined in a model of infection. The rat lung model of chronic infection was chosen for these studies [12]: young adult Sprague-Dawley rats were used. The animals were anaesthetized using isofluorane and inoculated into the left lobe of the lungs with 100 µl of a suspension of agar beads containing 10^5 - 10^6 bacterial cells. A healthy rat typically clears the *P. aeruginosa* infection within two weeks. After 7 days lungs were removed from sacrificed rats and homogenized tissues were plated on agar with and without chloramphenicol: the

transposon used for PA3859 K.O. mutantation carries the resistance for chloramphenicol. Thus the number of POA1 wild type and K.O. mutant could be evaluated by plate counts on media with and without the antibiotic.
Nine rats were infected for this experiment and, on average, a five fold higher amount of wild type bacteria were found in the rat lungs with respect to the PA3859 K.O. mutant. This result, even thought still preliminary, demonstrates that the mutant is reduced in its capacity to persist in the rat in competition with the PAO1 parental strain, suggesting that PA3859 is indeed necessary for *Pseudomonas aeruginosa* to thrive in the host lungs and thus for its pathogenicity to be exploited.

7.4 References

[1] Weltzien, H., U. Cytolytic and Membrane-Perturbing properties of Lysophosphatidylcholine. *Biochim. Biophys. Acta* (1979) **559**: 259-287.

[2] Garsetti, D., Holtsberg, F., Steiner, M., R., Egan, R., W., Clark, M., A. Butyric acid-induced differentiation of HL-60 cells increases the expression of a single lysophospholipase. *Biochem. J.* (1992) **288**: 831-837.

[3] Flieger,A., Neumeister, B., Cianciotto, N., P. Characterizaion of the Gene Encoding the Major Secreted Lysophospholipase A of *Legionella pneumophila* and Its Role in Detoxification of Lysophosphatidylcholine. *Infection and Immunity*. (2002) 6094-6106.

[4] Karasawa, K., Kudo, I., Kobayashi, T., Homma, H., Chiba, N., Mizushima, H. Lysophospholipase L_1 from Escherichia coli k-12 overproducer. *J. Biochem*. (1991) **109**: 288-293.

[5] Scmiel, D., H., Miller, V., L. Bacterial phospholipase and pathogenesis. *Microbes and Infections*. (1999) **1**: 1103-1112.

[6] Gannoum, M., A. Potential Role of Phospholipase in Virulence and Fungal Pathogenesis. *Clin.l Microbiol.Rev.*. (2000) 122-143.

[7] Tamura, M., Ajayi, T., Allmond, L., R., Moriyama, K., Wiener-Kronish, J., P., Sawa, T. Lysophospholipase A activity of *Pseudomonas aeruginosa* type III secretory toxin ExoU. *Biochem. Biophys. Res. Commun*. (2004) **316**: 323-331.

[8] Pankhaniya, R., R., Tamura, M., Allmond., L., R., Moriyama, K., Ajayi, T., Wiener-Kronish, J., P., Sawa, T. *Pseudomonas aeruginosa* causes acute lung injury via the catalytic activity of the patatin-like phospholipase domain ExoU. *Crit. Care Med*. (2004) **32**: 2293-2299.

[9] Sato, H., Frank, D., W. ExoU is a potent intracellular phospholipase. *Mol. Microbiol.* (2004) **53**: 1279-1290.

[10] Barker, A., P., Vasil, A., I., Flloux, A., Ball, G., Wilderman, P., J., Vasil, M., L. A novel extracellular phospholipase C of *Pseudomonas aeruginosa* is required for phospholipids chemotaxis. *Mol. Microbiol*. (2004) **53**: 1089-98.

[11] Wilderman, P., J., Vasil, A., I., Johnson, Z., Vasil, M., L. *Mol. Microbiol*. (2001) **39**: 291-303.

[12] Cash, H., A., Woods, D., E., McCullough, B., Johanson, W., G., Bass, J., A. A rat model of chronic respiratory infection with *Pseudomonas aeruginosa*. *Am. Rev. Respir. Dis.* (1979) **119**: 453-459.

Chapter 8

Discussion and final remarks

8.1 Introduction

The broad substrate specificity typically shown by microbial carboxylesterases prompted to hypothesize that these enzymes have evolved in order to allow the utilization of a wider variety of organic compounds as carbon source [1]. However, it does not necessarily implicate that the physiological function of the carboxylesterases is limited to this trophyc role.
The hydrolysis of an ester is a relatively simple reaction. For instance, esters undergo a spontaneous hydrolysis under basic condition due to the nucleophilic attach by hydroxyls to the carbonylic carbon. Furthermore, a nucleophile as hydroxylamine is able to increase the carboxylesterase-catalyzed reaction rate by providing an alternative route for the substrate hydrolysis, further confirming the relatively intrinsic instability of carboxyl esters [2]. The relatively broad substrate specificity often displayed by carboxylesterases could thus be considered as a somewhat unavoidable consequence of the substrates intrinsic instability rather than an enzyme feature evolved to accomplish a specific function.

Among the PA3859 homologous sequences many others carboxylesterases have been found. It does not come as a surprise of course, but it is worthy of note that for none of them the physiological function has been unambiguously assessed. All of these enzymes are simply considered to be “general carboxylesterases”. This is also true for the carboxylesterase II of *Pseudomonas fluorescens*, that was extensively studied and whose 3D structure was determined [3, 4, 5]. Besides a sequence homology of 76%, this protein shares an extremely similar structure with PA3859 (see Chapter 4), suggesting that the same *in vivo* function is also shared.

Since experimental data are missing, so far the above mentioned hypothesis about carboxylesterases function has a merely speculative meaning. However it is not unlikely that the broad substrate specificity often assigned to carboxylesterases rather than defining an enzyme feature represents a statement of the lack of information about their physiological substrate(s).

The work reported in this thesis represents an example of how the 3D structure determination of an enzyme can help to unveil its function: the presence on the PA3859 surface of a groove that encompasses the catalytic site was in fact supposed to be the possible enzyme binding site. Its binding properties were evaluated by means of automated docking simulations and the obtained results were subsequently validated by enzymatic assays. Furthermore, experiments carried out with a *Pseudomonas aeruginosa* mutant in which the expression of PA3859 was suppressed, allowed the determination of the enzyme role *in vivo*.
Thus, starting from the enzyme 3D structure determination, an approach involving *in silico*, *in vitro* and *in vivo* assays lead to the reliable determination of the PA3859 physiological function.

8.2 PA3859 *in silico* and *in vitro* study

The PA3859 binding site can be considered as build up by two distinct part, a long groove mainly hydrophobic, here called the "hydrophobic cleft" and a polar part, named the "polar cavity", with the catalytic Ser 113 positioned exactly in between these two elements (Fig. 8.1). The finding in PA3859 crystal form I of a MES molecule covalently bound to Ser 113 and positioned within the polar cavity and in crystal form II of a PEG strand lying within the hydrophobic cleft, further suggested this grove to be the enzyme binding site (See Chapter 4).

According to the docking simulations, the hydrophobic cleft should be involved in the interaction with the substrates acyl chain. The cleft seems to be suited for the binding of alkyl chain as long as 18 - 20 carbon atoms, chains longer than 20 carbon atoms in fact could not be docked completely into the cleft. The docking simulation results are in a good agreement with the enzymatic assays. As shown in Fig. 8.2, the K_m values for the hydrolysis of *p*-nitrophenyl esters experimentally determined and the theoretical binding constants, produced by the docking simulation, display a very similar dependence on the substrate chain length.

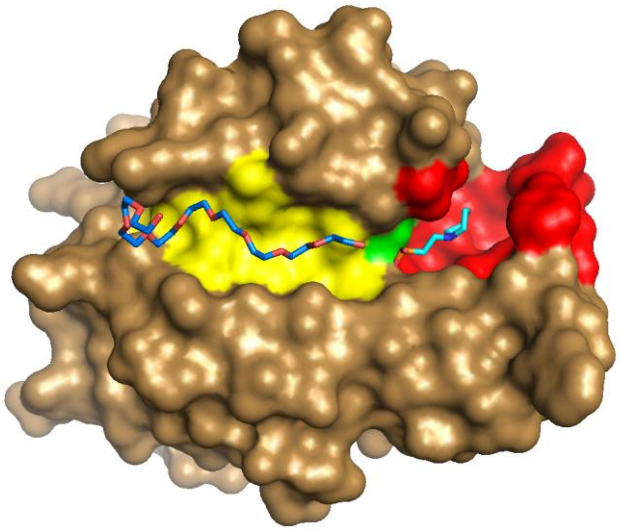


Figure 8.1. PA3859 surface. Hydrophobic cleft, yellow. Catalytic Ser 113, green. Polar cavity, red. The PEG nonamer positioned into the polar cavity is shown in blue stick format and the MES molecule bound to Ser 113 is depicted as cyan stick.

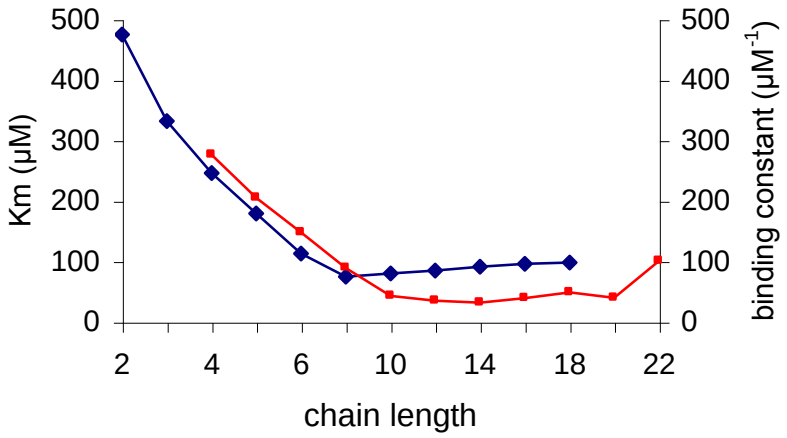


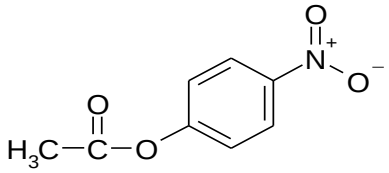
Figure 8.2. The K_m values for the hydrolysis of *p*-nitrophenyl esters (blue curve) and the theoretical binding constants produced by the docking simulation (red curve) are plotted as a function of the substrate alkyl chain length.

The almost perfect numerical overlap between these values however has to be considered as occasional due to the different conceptual meaning of binding constant and K_m: binding constant is a thermodynamic parameter, while K_m, even though representative of the enzyme affinity for the substrate, results from a combination of

kinetic constants. The former represents the ratio between the concentration of products and reagents once that the thermodynamic equilibrium is achieved, the latter corresponds to the substrate concentration that give rise to a reaction rate equal to 50% of the maximal velocity.
The K_m minimum value is obtained for *p*-NP C8 (75 μM). For longer substrates K_m values slightly increase up to 98 μM obtained for *p*-NP C18 (see Chapter 5). Theoretical binding constant values of roughly 40 μM^{-1} are instead obtained for substrate *p*-NP C10 to C20, with a minimum of 32 μM^{-1} achieved for *p*-NP C14 (see Chapter 6). Thus in both curves a plateau is observed, where the measured (or computed) values dependence from the substrates chain length is reduced.

The enzymatic hydrolysis of water-insoluble substrate embedded in a micelle is a quite complex process. For example, it could be argued that if the enzyme-substrate interaction requires the substrate to be extracted from the micelle into the bulk solvent, then the measured K_m values could be affected by this "extraction process" and thus would be no longer representative of the real enzyme affinity for the substrate. The extraction of a *p*-nitrophenyl ester from a micelle to the water solvent is an unfavourable process whose positive ΔG becomes larger and larger as the substrate chain length increases [6]. Thus in such case, an increase in the apparent K_m values would be expected for the longer chain substrates that is regardless of the real enzyme-substrate affinity. Moreover, a completely different dependence on the substrate chain length would be expected for the measured K_m and the theoretical binding constant values. In the AutoDock Free Energy Function indeed, the binding energies are evaluated considering a process in which the substrate, monomolecularly dispersed in the bulk solution, binds to the enzyme (see Appendix 2). Thus, being the complications that arise from the presence of the substrate in the micelle completely ignored by the docking simulations, the final result is indicative of the real enzyme-substrate affinity. In conclusion, the almost perfect agreement between K_m and the theoretical binding constants, suggests that the measured K_m are really indicative of the PA3859 affinity for the *p*-nitrophenyl esters. Furthermore, it suggests that a tight interaction occurs between enzyme and micelle and allows the direct transfer of the substrate from the micelle to the enzyme binding site without the formation of a substrate solvated intermediate.

For substrates *p*-nitrophenyl C2 to C6, K_m values much higher than those of longer chain substrates were found (Fig. 8.2). Also in this case the automated docking provides a possible explanation. The PA3859 oxyanion hole, whose function consists in the stabilization of the oxyanion intermediate formed along with the catalytic process (see Chapter 1) obviously displays a high affinity for negatively charged oxygen atoms. Thus, when the docking of short chain *p*-nitrophenyl esters was attempted, in the lowest



Structure of *p*-NP acetate

energy solutions, the substrate nitro group oxygen atom, rather than the carbonylic carbon, was positioned in the active site (see Chapter 5 and Fig. 8.3). Indeed, for this reason *p*-NP C2 could not be docked in the productive conformation. The shorter substrate whose lowest binding energy was obtained for its active conformation is *p*-NP C8, suggesting that an alkyl chain of at least 7 carbon atom is required to stabilize the productive positioning of *p*-NP substrates within the active site. Thus, the low K_m values for substrate C2 to C6 could be the consequence of their preferential non-productive interaction with the PA3859 active site that would mimic a reduced substrate concentration.

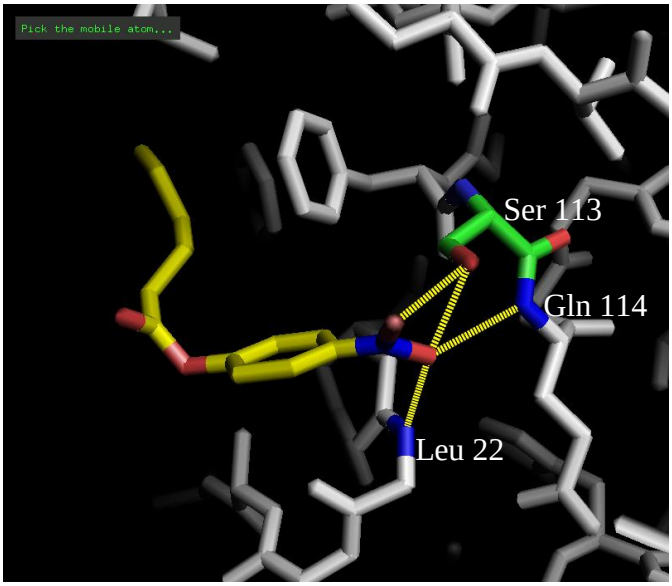


Figure 8.3. The overall minimal energy docking solution for the *p*-NP C6: the substrate is positioned with its nitro group into the active site, where hydrogen bonds occur between the nitro group oxygen atoms and the Ser 113 (green) and the main chain nitrogen atom of the oxyanion hole residues Leu 22 and Gln 114 (blue).

Another interesting question, that arose by the study of PA3859 hydrolysis of *p*-nitrophenyl esters, has been the enzyme specificity as a function of the substrate chain length. According to the experiments reported in Chapter 6, the higher PA3859 catalytic efficiency, quantified by the K_{cat}/K_m ratio, is achieved for esters of octanoic acid and decreases sharply for substrates with alkyl chain either shorter or longer than 8 carbon atoms (Fig. 8.4).

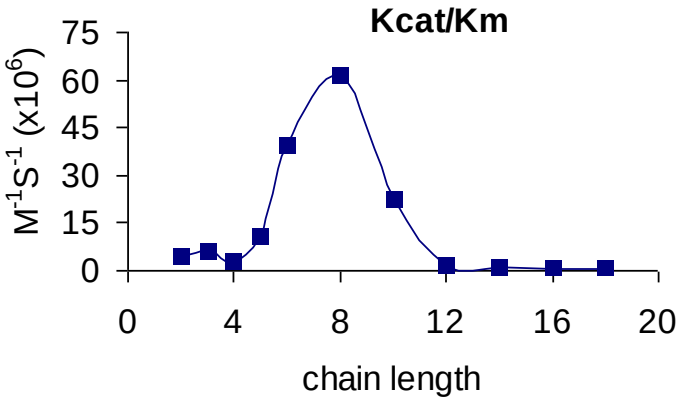


Figure 8.4. Plot of the PA3859 K_{cat}/K_m ratio for the hydrolysis of *p*-nitrophenyl ester as a function of the substates alkyl chain length.

This result is in a clear disagreement with the PA3859 structure feature and with the results of the docking simulation: the hydrophobic cleft present on the enzyme surface, that is believed to be involved in the accommodation of the substrate alkyl chain, is in fact long enough to accommodate chains as long as 18 - 20 carbon atoms, suggesting its physiological substrates alkyl chain to be much longer than octanoic acid. Furthermore, naturally occurring lipids and phospholipids typically derive from the esterification of fatty acid at least 16 carbon atom long (palmitic acid). Thus a hypothetical phospholipid derived from octanoic acid is not likely to have any physiological relevance.

When the specificity of lipolytic enzymes has been assayed using mixed micelles containing *p*-nitrophenyl esters as substrate, an optimal alkyl chain length of 8 carbon atoms is often found [7, 8, 9, 10] Actually, to the best of our knowledge, these experiments have never lead to the finding of an optimal substrate chain length longer than 8 carbon atom, suggesting that some kind of artifacts probably occurs. A possible explanation is the following: the enzymatic hydrolysis of a *p*-nitrophenyl ester is a two step process (see Chapter 1). The first step leads to the release of the *p*-nitrophenol and to the formation of the acyl-enzyme. Subsequently, the acyl-enzyme is hydrolyzed and the free fatty acid is released. This second step has been demonstrated to be the rate limiting step. The release of the fatty acid from the enzyme hydrophobic cleft into the bulk solvent is an energetically disadvantageous process, whose positive ΔG is expected to increase as the fatty acid chain length increases. As a consequence, the release of long chain fatty acids is likely to be slower than the release of shorter ones, resulting in a considerable reduction of the K_{cat} values for the longer chain substrates (Fig. 8.4). This hypothesis however is only valid if the leaving fatty acid would be released into the aqueous phase rather than directly into the micelle.

A further possible explanation has also been suggested. For many interfacial enzymes a so called “scouting” mode of action has been proposed (see Chapter 1) [11]. According to this mechanism the enzymes associate to the micelle interface and then processively carry out the catalytic turnover without detaching from the surface. In this case the free fatty acid reaction product is directly released into the micelle.
A limitation of the detergent-substrate mixed-micelle method for the study of such lipolytic enzymes is that if the substrate replenishment rate, that is the rate for the substrate to exchange between micelles, is not much faster than the turnover rate, the reaction rate will be limited by the replenishment rate rather than by the catalytic efficiency of the enzyme. The replenishment rate for a long chain *p*-nitrophenyl ester is exceedingly slow via the exchange mechanism where the substrate desorbs into the aqueous phase because the aqueous concentration is in the subnanomolar range [12, 13]. Thus the observed K_{cat} rather than a measure of the enzymatic process is a measure of the micelle component exchange.

When the hydrolysis of the soluble *p*-nitrophenyl ester was carried out in the presence of a detergent, a non-competitive like inhibition was clearly observed (see Chapter 6). This finding can be interpreted as a consequence of a tight interaction between PA3859 and the micelle. The trapping of the enzyme on the micelle surface can be supposed to reduce the concentration of the available active sites and thus the overall reaction rate. On the other hand, as will be discussed later, PA3859 is not an interfacial activated enzyme and is able to hydrolyze its substrates in both the aggregated and the monodispersed form. Furthermore, while both lipase and phospholipase, when loaded on a gel filtration chromatography column together with a detergent, tipically coelute with the micelles [2], the PA3859 elution was not found to be affected by the presence of Triton X-100 or phospholipids. These results suggest that PA3859 interaction with the micelle surface is much weaker than those of real interfacial enzymes.

Thus it is possible that both the above suggested mechanisms could contribute to the observed PA3859 substrate specificity. The disadvantageous release of the free fatty acid in the aqueous phase could affect the turnover rate of the enzyme molecules that are not bound to the micelle surface, and the turnover of the enzyme bound to the micelle could be limited by the substrate replenishment rate.
In any case, the apparent optimal substrate chain length of 8 carbon atom found for PA3859, rather than being representative of a real substrate specificity is likely to be merely the consequence of the substrates intrinsic properties.
These considerations also suggest that for lipolytic enzymes, unless a very careful and detailed kinetic study is performed, the substrate alkyl chain specificity can be inferred with more reliability from the protein binding site structure than from a preliminary enzymatic assay.

Studies carried out on the 14-kDa family of secreted phospholipase A_2, have lead to the identification of the residues involved in the protein-micelle surface interaction. The portion of the enzyme surface that interacts with the micelle, called i-face, is mainly made by hydrophobic or positively charged residues [14, 15]. The i-face cationic residues Lys and Arg have been demonstrated to be directly involved in the protein micelle interaction [16, 17, 18, 19]. Specifically, the positively charged side chains of these residues are thought to be involved in the interaction with the negatively charged phosphate moiety of the phospholipids.

In PA3859, the portion of the protein surface around the binding site is mainly made by hydrophobic residues, however two negatively charged patches positioned next to the hydrophobic cleft are also present (fig. 8.5). Study to explicitly evaluate the interaction of PA3859 with the micelles surface have not been performed, however kinetic data indicated that this enzyme interacts with the micelle and that this interaction is weaker than the one displayed by the prototypic phospholipases A_2 [15, 18]. This finding allow to hypothesize that the interaction between PA3859 and the micelle surface could be mediated by the negatively charged residues, thus involving the cationic phospholipids headgrup exposed on the micelle surface rather than the anionic phosphate moiety positioned deeper inside. This different interaction mode would cause PA3859 to "float" on the micelle surface, thus explaining the observed weaker protein-micelle interaction.

In a model for membrane binding of human pancreatic phospholipase A_2, a Trp residue present on the enzyme i-face is suggested to stabilize the membrane anchoring of the protein by both non-ionic interaction with the lipid alkyl chains and by hydrogen bonding between the imino group of the Trp indole ring and carbonyl oxygen atoms of membrane lipids [16]. Interestingly, in PA3859 the Trp 61 is positioned next to the active site and its side chain protrudes from the protein surface, suggesting that similarly to what has been found for the human pancreatic phospholipase A_2, the protein anchoring to the micelle could also be mediated by this Trp residue (Fig. 8.5).

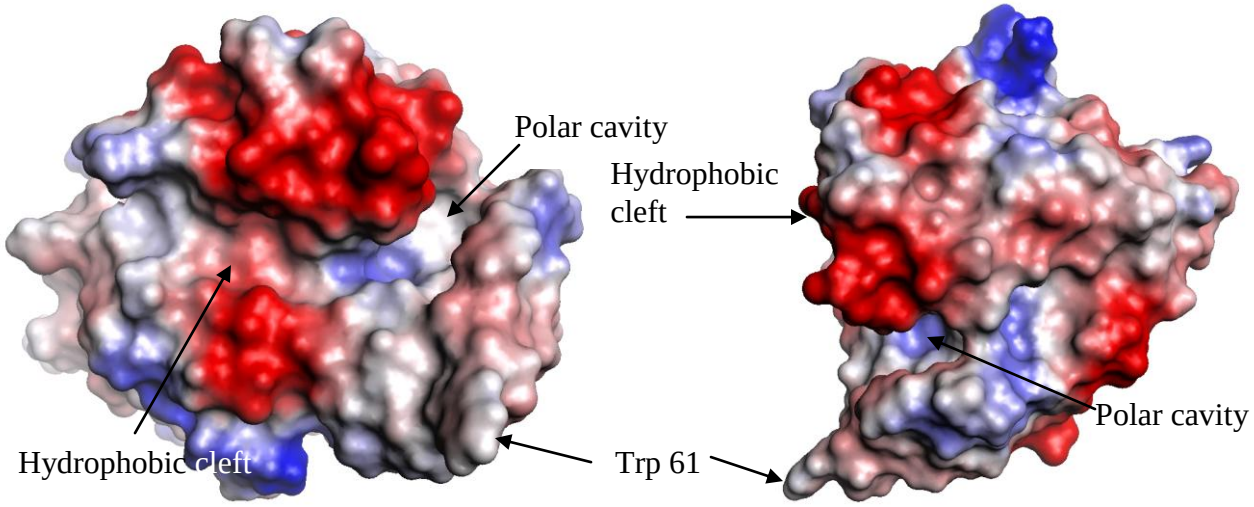


Figure 8.5. PA3859 electrostatic potential surface: red, negatively charges, blue, positively charges. The hydrophobic cleft is flanked by two acidic regions, composed by Asp 167, Asp 170 and Glu 198 (the bigger one) and Asp 25 and Asp 289 (the smaller one). The portion of the protein surface around the hydrophobic cleft is rather flat, and residue Trp 61 sticks out sharply into the solvent. Figure produced with PyMol (http://www.pymol.org).

In *Pseudomonas aeruginosa* PAO1 genome the PA3859 locus is positioned in the vicinity of two genes involved in the phospholipids metabolism: PA3857 was demonstrated to be a phosphatidylcholine synthase [20] and PA3860 is hypothesize to be involved in the lipid biosynthesis, due to its 59% sequence homology with a long chain fatty acid – CoA ligase of *Archaeaglobus fulgidus* (see Chapter 5) (http://www.pseudomonas.com).This finding prompted to hypothesize PA3859 to be implicated in the phospholipids metabolism. The ability of the polar part of the PA3859 binding site, named the "polar cavity", to bind to the phospholipids polar headgroup was checked by automated docking. These simulations suggested the enzyme to be specific for phosphoryl -serine, -ethanolamine, -choline and -glycerol headgroups. Subsequently, the PA3859 activity against 1-palmitoyil-*sn*-phosphatidyl -serine, ethanolamine, -glycerol and choline was confirmed by an enzymatic assay and TLC analysis (see Chapter 6).

Of a certain interest is the finding that PA3859 is able to hydrolyze choline based lysophospholipid: phosphatidylcholine in fact, up to few years ago, was thought to be present only in the membrane of eukaryotic cells. Now it is known that several highly specialized groups of bacteria, mainly photosynthetic bacteria containing extensive internal membrane structures or those living in association with eukaryotes, such as pathogen or symbiotic bacteria, contain phosphatidylcholine as a membrane lipid [21]. Even though the requirement of phosphatidylcholine in these bacteria has not been

clearly determined, it is believed to play an important role in the bacterium-eukaryotic cell interaction [22].
In most of the phosphatidylcholine binding protein, the choline headgroup stabilization is achieved by cation-π stacking: typically the quaternary ammonium is placed between two aromatic rings at a distance of 4.5 - 5 Å, [23]. As suggested by automated docking, in PA3859 this interaction mode could arise from the rotation of Phe 72 side chain (see Chapter 5). As will be discussed later on, the binding site of carboxylesterase II from *Pseudomonas fluorescens* is very similar the one of PA3859, suggesting that these enzymes are probably involved in the same physiologic function. However, in the polar cavity of carboxylesterase II from *P. fluorescens*, Phe 72 is replaced by a methionine, (Met 73). Thus the classical choline stabilization can not be exploited, suggesting that this enzyme is probably unable to efficiently hydrolyze phosphatidylcholine. This hypothesis is in agreement with the different physiology of these bacteria: despite their phylogenetic closeness, *P. fluorescens* is not a pathogen and, as far as it is known, its plasmatic membrane does not contain phosphatidylcholine.

The PA3859 hydrolysis of lysophospholipid was further characterized by using radiolabeled lysophosphatydilcholine as substrate. The result of this study suggests that PA3859 is indeed a lysophospholipase, being its kinetic parameters in agreement with those found for other lysophospholipases (see Chapter 6) [24, 25, 26].
Furthermore, this study has shown that the enzyme is not interfacially activated, that is able to hydrolyze its substrates in both monodisperse and micellar form and that its affinity for the micelle surface does not depend significantly from the micelle composition, suggesting that the protein-micelle interaction is mainly due to aspecific hydrophobic interaction.

The availability of a *P. aeruginosa* mutant in which the expression of PA3859 was suppressed allowed to set up some preliminary experiment to assay the role of this enzyme *in vivo*.

As reported in Chapter 2, PA3859 was firstly purified from *P. aeruginosa* PAO1 grown on LB, with no need for any special induction. Thus the enzyme is constitutively expressed, suggesting that it is probably involved in a so called “house-keeping” process. When the wild type and mutant bacteria growth rates on LB were compared, no significant difference was found (see Chapter 7), indicating that PA3859 is not an essential enzyme, or possibly that in *P. aeruginosa* other enzymes with similar function are present. However, when bacteria were grown on a minimal medium containing a mixture of phospholipids as the unique carbon source, the PA3859 KO mutant was

found to grow much slower than the wild type, suggesting an involvement of PA3859 in the phospholipids metabolism (see Chapter 7).

Lysophospholipids are known to be toxic due to the destabilization effect that they exert on plasma membranes [27].
The observation that the final density reached by the PA3859 KO mutant culture is roughly equal to 60% of the cell density reached by the wild type culture is probably the consequence of the mutant incapacity to scavenge the toxic lysophospholipids that are produced along with the bacterial growth on the phospholipids containing medium. Thus the lower culture density reached by the PA3859 knock out mutant could be due to an increased cytolysis rate that would be produced by the accumulation of intracellular lysophospholipids (see Chapter 7, Fig. 7.1)

Given these interesting results, the possible contribution of PA3859 in the *Pseudomonas aeruginosa* lung infection was also assayed.
Infection of mammalian airways by *Pseudomonas aeruginosa* leads to varied pathological responses and lung injury. At one extreme there is the acute nosocomial necrotizing pneumonia that is associated with a high incidence of mortality despite antibiotic therapy, due in part to epithelial destruction and bacterial invasion of the pulmonary vasculature with resultant bacteremia and its sequellae. At the opposite extreme there is the persistent airway infection in individuals with cystic fibrosis, where microorganisms remain confined to the airway as biofilms, triggering a chronic neutrophilic inflammatory response (see Chapter 1).
To assay the possible role played by PA3859 along with the *Pseudomonas aeruginosa* lung infection, the rat chronic lung infection model was used: a healthy rat typically is able to clear a *P. aeruginosa* infection within two weeks. In order to study the pathogen-host interaction a model was developed that allows the lung infection to continue for longer time [28]. Bacteria, embedded in agar beads, are directly inoculated in the rat lungs. The beads will then provide a source for “fresh” bacteria that is able to let the infection persist for longer than one month.
The experiment, carried out in the laboratory of Dr. Roger C. Levesque at Facultè de Mèdecine, Universitè Laval, Stefoy, Quebèc, Canada, consisted in the inoculation in the rat lungs of en equal amount of wild type and PA3859 knock out mutant. After 7 days from inoculation, the lungs were recovered and the amount of both wild type and mutant bacteria was quantified. On average, the amount of the recovered mutant bacteria resulted to be 5 fold smaller than the wild type, indicating that PA3859 contributes to the ability of *P. aeruginosa* PAO1 to persist in a chronic pulmonary infection in rats model.

Along with the *Pseudomonas aeruginosa* lung infection, at least 5 different virulence determinants are secreted, that are known to possess a phospholipase and/or lysophospholipase activity [29]. ExoU is a phospholipase/lysophospholipase that is directly injected into the host cell cytoplasm via the type III secretion system [25, 30, 31]. It is involved in the inhibition of the eukaryotic cell survival pathways [32] and in the activation of the proapoptotic pathways [33].
Three phospholipase C (PlcH, PlcN and PlcB) secreted trough the inner membrane via the arginine translocase pathway [34], and a phospholipase D, named PldA [35], are also known to be involved in the P. *aeruginosa* lung infection. PlcB is required for the bacterium phospholipids chemotaxis, but details about the role of the PlcH, PlcN and PldA in the pulmonary infection have not been yet provided. A possibility is that these enzymes are involved in the host cell membrane degradation and possibly in the hydrolysis of the phosphatidylcholine that is present in the lungs as a surfactant.

Overall an involvement of phospholipases/lysophospholipases and phospholipids metabolism in the *P. aeruginosa* chronic lung infection has been clearly demonstrated. Even though so far it is not possible to unambiguously asses the role of PA3859 in the bacteria-host interaction, some hypothesis can be made. As suggested by the growth rate of PA3859 knock out mutant on the minimal medium containing phospholipids as the unique carbon source, a possibility is that PA3859 is required to scavenge the lysophospholipids produced by the partial digestion of the host cell membrane or to hydrolyse the phosphatidylcholine present in the lungs.
Moreover, being PA3859 involved in the phospholipids metabolism, in the knock out mutant the phosphatidylcholine metabolism is likely to be altered. Since phosphatidylcholine is required for an optimal bacterium-eukaryotic cell interaction [22] the reduced persistence of the knock out mutant with respect to the wild type PAO1 in the chronic lung infection model, could be the consequence of a modified lipid content of the bacterium cell membrane.

8.2 Conclusion and perspectives

Overall, the results reported in the present work, demonstrate that the putative carboxylesterase from PA3859 of *Pseudomonas aeruginosa* is indeed a lysophospholipase. The experiments carried out with the PA3859 knock out mutant indicate that this enzyme is involved in the phospholipids metabolism and perhaps in the detoxification of lysophospholipids. Furthermore, it has been demonstrated that PA3859 is necessary for the bacteria to thrive in the rat lungs and thus a role for this enzyme as a virulence factor can be hypothesized.

These results also suggest that PA3859 could be a possible therapeutical target protein for the development of new anti – *P. aeruginosa* antibiotics. Further investigation however would be required in order to determine the enzyme specificity with respect to the lysophospholipid polar headgroup and to provide details into its *in vivo* role and its implication in the *Pseudomonas aeruginosa* pathogenicity.

8.3 References

[1] Bornscheuer, U., Y. Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Micr. Rew.* (2002) **26**: 73-81.

[2] Burdette, R., A., Quinn, D., M. Interfacial Reaction Dynamics and Acyl-enzyme Mechanism for Lipoprotein Lipase-catalyzed Hydrolysis of Lipid *p*-Nitrophenyl Esters. *J. Biol. Chem.* (1986) **261**: 12016-12021.

[3] Hong, K., H., Jang, W., H., Choi, K., D., Yoo, O., J. Characterization of *Pseudomonas fluorescens* carboxylesterase: cloning and expression of the esterase gene in *Escherichia coli*. *Agric. Biol. Chem*. (1991) **55**: 2839-2845.

[4] Kim, K., K., Suh, S., W. Crystallization and preliminary X-ray crystallographic analysis of carboxylesterase from *Pseudomonas fluorescens*. *Arch. Biochem. Biophys*. (1993) **302**: 417-419

[5] Kim, K., K., Song, H., K., Shin, D., H., Hwang, K., Y., Choe, S., Yoo, O., J., Suh, S., W. Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an acyl-hydrolase with broad substrate specificity. *Structure* (1997) **5**: 1571-1584.

[6] Harberland, M., E., Reynolds, J., A. J. *Biol. Chem.* (1975) **250**: 6636-6639.

[7] Burdette, R., A., Quinn, D., M. Interfacial Reaction Dynamics and Acyl-enzyme Mechanism for Lipoprotein Lipase-catalyzed Hydrolysis of Lipid *p*-Nitrophenyl Esters. *J. Biol. Chem.* (1986) **261**:12016-12021.

[8] Alquati, C., De Gioia, L., Santarossa, G., Alberghina, L., Fantucci, P., Lotti, M. The cold-active lipase of *Pseudomonas fragi*. *Eur. J. Biochem*. (2002) **269**: 3321-3328.

[9] Santarossa, G., Gatti Lafranconi, P., Alquati, C., De Gioia, L., Alberghina, L., Fantucci, P., Lotti, M. Mutations in the "lid" region affect chain lenght specificity and thermostability of a *Pseudomonas fragi* lipase. *FEBS*. (2005) **579**: 2383-2386.

[10] Redondo, O., Herrero, A., Bello, J., F., Roig, M., G., Calvo, M., V., Plou, F., J. Burguillo, F., J. Comparative kinetic study of lipase A and B from *Candida rugosa* in the hydrolysis of lipid *p*-nitrophenyl esters in mixed micelles with triton X-100. *Biochim. Biophys. Acta*. (1995) **1243**: 15-24.

[11] Berg, O., G., Jain, M., K. Interfacial Enzyme Kinetics. Wyley, London. 2002.

[12] Nichols, J., W. *Biochemistry*. (1985) **24**: 6390.

[13] Nichols, J., W., Pagano, R., E. *Biochemistry* (1982) **21**:1720.

[14] Ramirez, F., and Jain, M. K. *Proteins: Struct. Funct. Genet.*(1991) ***9***: 229-39.

[15] Bahnson, B., J. Structure, function and interfacial allosterism in phospholipase A_2: insight from the anion-assisted dimer. Arch. Biochem. Biophys. (2005) **433**: 96-106.

[16] Yu, B. Z., Poi, M. J., Ramagopal, U. A., Jain, R., Ramakumar, S., Berg, O. G., Tsai, M. D., Sekar, K., and Jain, M. K. *Biochemistry.* (2000) ***39***: 12312-12323.

[17] Quin, S., Pande,A., H., Nemec, K., N., Tatulian, S., A. The N-terminal α-Helix of Pancreatic Phospholipase A_2 Determines Productive-mode Orientation of the Enzyme at the Membrane Surface. *J. Mol. Biol.* (2004) **344**: 71-89.

[18] Yu, B., Rogers, J., Tsai, M., Pidgeon, C., Jain, M., K. Contributions of Residues of Pancreatic Phospholipase A_2 to Interfacial Binding, Catalysis and Activation. *Biochemistry* (1999) **38**: 4875-4884.

[19] Yu, B., Poi, M., J., Ramagopal, U., A., Jain, R., Ramakumar, S., Berg, O., G., Tsai, M., Sekar, K., Jain, M., K. Structural Basis of the Anionic Interface Preference and k_{cat} Activation of Pancreatic Phospholipase A_2. *Biochemistry*. (2000). **39**:12312-12323.

[20] Wilderman PJ, Vasil AI, Martin WE, Murphy RC, Vasil M. L. *Pseudomonas aeruginosa* synthesizes phosphatidylcholine by use of the phosphatidylcholine synthase pathway. *J Bacteriol*. (2002) **184**: 4792-4799.

[21] Goldfine, H. Lipids of procaryotes: structure and distribution. *Curr. Top. Membr. Transp.* (1982) **17:** 1-43.

[22] de Rudder, K. E., I. M. Lopez-Lara, and O. Geiger. Inactivation of the gene for phospholipid *N*-methyltransferase in *Sinorhizobium meliloti*: phosphatidylcholine is required for normal growth. *Mol. Microbiol*. (2000) **37**: 763-772.

[23] Anbazhagan, Musti J. Swamy, Thermodynamics of phosphorylcholine and lysophosphatidylcholine binding to the major protein of bovine seminal plasma, PDC-109. *FEBS Letters* (2005) **579:** 2933-2938.

[24] Wang, A., Yang, H., Friedman, P., Johnson, C., A., Dennis, E., A. A specific human lysophospholipase: cDNA cloning, tissue distribution and kinetic characterization. *Biochim. Biophys. Acta.* (1999) **1437**: 157-169.

[25] Tamura, M., Ajayi, T., Allmond, L., R., Moriyama, K., Wiener-Kronish, J., Sawa,T. Lysophospholipase A activity of *Pseudomonas aeruginosa* type III secretory toxin ExoU. *Biochem. Biophys. Res. Comm*. (2004) **316**: 323-331.

[26] Phillips, R., M., Six, D., A., Dennis, E., A., Ghosh, P. In vivo Phospholipase Activity of the *Pseudomonas aeruginosa* Cytotoxyn ExoU and Protection of Mammalian Cells with Phospholipase A_2 Inhibitors. J. *Biol. Chem*. (2003). **278**: 4136-4132

[27] Weltzien, H., U. Cytolytic and Membrane-Perturbing properties of Lysophosphatidylcholine. *Biochim. Biophys. Acta* (1979) **559**: 259-287.

[28] Cash, H., A., Woods, D., E., McCullough, B., Johanson, W., G., Bass, J., A. A rat model of chronic respiratory infection with *Pseudomonas aeruginosa*. *Am. Rev. Respir. Dis.* (1979) **119**:453-459.

[29] Lau, G., W., Hassett, D., J., Britigan, B., E. Modulation of lung epithelial functions by Pseudomonas aeruginosa. *TRENDS in Microbiology*. (2005) **13**: 389-397.

[30] Sato, H., Frank, D., W. ExoU is a potent intracellular phospholipase. *Mol. Microbiol*. (2004) **53**:1279-1290.

[31] Pankhaniya, R., R., Tamura, M., Allmond., L., R., Moriyama, K., Ajayi, T., Wiener-Kronish, J., P., Sawa, T. *Pseudomonas aeruginosa* causes acute lung injury via the catalytic activity of the patatin-like phospholipase domain ExoU. *Crit. Care Med*. (2004) **32**: 2293-2299.

[32] Sato, H. The mechanism of action of the *Pseudomonas aeruginosa*-encoded type III cytotoxin ExoU. *EMBO J*.(2003) **22**: 2959-2969.

[33] Shaver, C., M., Hauser, A., R. Relative contributions of *Pseudomonas aeruginosa* ExoU, ExoS and ExoT to virulence in the lung. *Infect. Immun*. (2004) **72**: 6696-6977.

[34] Barker, A., P., Vasil, A., I., Flloux, A., Ball, G., Wilderman, P., J., Vasil, M., L. A novel extracellular phospholipase C of *Pseudomonas aeruginosa* is required for phospholipids chemotaxis. *Mol. Microbiol*. (2004) **53**: 1089-98.

[35] Wilderman, P., J., Vasil, A., I., Johnson, Z., Vasil, M., L. *Mol. Microbiol*. (2001) **39**: 291-303.

Appendix 1

Protein X-ray crystallography

The first published observation of the crystallization of protein was authored by Hünefeld in 1840. The protein was haemoglobin from the earthworm. Flat, plate-like crystals were obtained when the blood of an earthworm was pressed between two slides of glass and allowed to dry very slowly. Until the 1930s, and indeed for many years beyond that, the motivation for crystallising proteins, particularly enzymes, was to purify a specific macromolecule from a complex extract, or to demonstrate, in the classical chemist's sense, the homogeneity of a preparation. Throughout this period, crystallinity was associated with purity. In the late 1930s, however, X-ray diffractionists, such as Atsbury, Bernal, Crowfoot, Fankuchen, Kendrew and Perutz, turned their attention to protein crystals as a source of structural information concerning biological macromolecules [1].

A1.1 Crystal definition

Crystal is a solid having regularly repeating (periodic) internal arrangements of atoms, molecules or ions and is a result of their cooperative, self-promoted, three-dimensional ordering [2, 3].

A1.2 The uniqueness of macromolecular crystals

Protein crystals, as all macromolecular crystals, rarely exceed a millimeter on an edge and are generally much smaller.
They are extremely fragile and have very weak mechanical properties.

Most, but not all, macromolecular crystals demonstrate only weak birefringence. The extent of optical activity can serve as preliminary but not definite test, whether a crystal is of a macromolecule or of some other mother liquor component such as a salt or a buffer.
Most protein crystals exhibit rather poor diffraction properties in comparison with organic or inorganic small molecule crystals. The number and the relative strengths of interactions securing macromolecules in the crystal lattice are far fewer and significantly weaker in proportion to the molecular mass (hydrogen-bonds, van der Waals contacts, salt-bridges). For that reason the macromolecules in the lattice tend to be less well ordered. They do not generally produce Bragg reflections to resolution approaching the theoretical limit, but decline substantially as a function of the scattering angle. Macromolecular crystals, depending on the mass of protein, usually have cell dimensions in the range of thirty to several hundred angstroms. Because the number of unit cells in a crystal (the effective scattering units) is inversely proportional to the volume of the unit cell, a macromolecular crystal produces far less scattering intensity than conventional crystals of the same size. Moreover, proteins consist mainly of C, N, O, which are light elements with only a few electrons (6-8) per atom that scatter X-rays much more weakly than do heavier elements.
The symmetry groups available are limited, because macromolecules lack inversion symmetry (see paragraph 4).
The protein crystals have high solvent content, at minimum this is usually no less than 30% by volume, for most cases it is 40% to 60%, and in the extreme cases it can be as high as 90%. Thus, macromolecules in a crystal are like islands, isolated from one another in a sea of water, but all interconnected by a network of delicate contacts. Macromolecules in the crystal bind their ligands just as they would in solution, except in those rare cases where lattice interactions interfere. This is what makes macromolecular crystals such an outstanding system to perform biochemical experiments. It is this feature of protein crystals that has made them invaluable tools in enzymology and for structural based drug design [1].

A1.3 Protein crystallization

A1.3.1 The phase diagram

If a protein-solvent system is at a concentration of the solute below its solubility limit for a particular set of conditions, it is considered to be in the undersaturation state. If it is exactly at the limit, then the solution is called saturated. At this unique concentration

of the material in solution, the rates of loss and gain of both the solid and solution phases are equal, and the system is at the equilibrium. Crystals, therefore, cannot grow from a solution that is simply saturated. When there are more protein molecules in a solution than solubility limit would allow, the solution is called a supersaturated one (fig. A1.1). We would expect molecules in solution to rejoin the solid state due to a thermodynamic driving force that pushes the system back to the solubility limit. In this way the system will tend to reestablish the equilibrium. This is from non-equilibrium, saturated solutions that crystals are grown. If conditions are adjusted in a tranquil manner, in the absence of any preexisting solid state, and if no sudden impulses of energy are provided to the system that could raise anomalously its overall energy and generate critical nuclei, then supersaturation will result. Certain amount of activation energy is required to initiate the formation of the solid state. There are two approaches to helping the system to overcome the activation barrier: putting energy into the system by increasing the supersaturation of the solution, or effectively lowering the energy barrier by seeding the solution with preexisting crystals. If no adequate amount of energy becomes available, the system will remain in metastable, non-equilibrium state of supersaturation. No nuclei will form and no crystals will grow. In the metastable zone crystals may grow only if nuclei are present. The high probability of spontaneous critical nuclei or amorphous precipitate formation in solution is present in the extreme supersaturation zone called labile region. The latter corresponds to an increased energy state of the system (fig. A1.1).

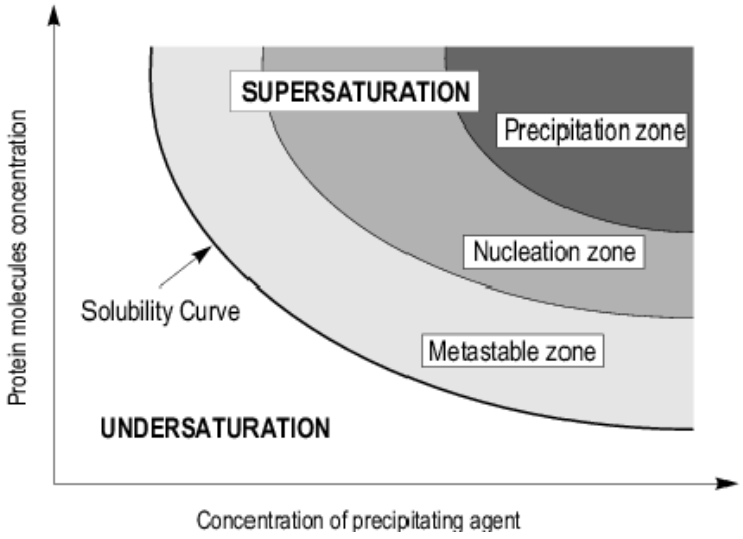


Figure A1.1 Phase diagram.

The ideal approach for protein crystallization would be to induce nuclei at the lowest level of labile region. As these few nuclei begin to grow, the supersaturation will fall and the solution will gradually enter the metastable region where the crystal growth will be slow and ordered. A few and larger crystals will be produced. No competition for new nuclei or precipitate will be observed [1].

A1.3.2. Main factors influencing supersaturation

Protein crystallization is mainly a trial-and-error procedure in which the protein is allowed slowly to precipitate from its solution as a result of the achievement of the supersaturation. Generally the protein supersaturation is obtained by a slow increase in the concentration of some precipitants, or by changes in pH and/or in temperature. The solubility effects of a precipitant and the protein behavior (electrostatic character, surface features, conformation) depend on the pH and on the temperature at which the experiment is performed. Proteins may be crystallized from a solution equally well, by increasing the precipitant concentration at constant pH and temperature (Ks fractionation), or at constant precipitant concentration by variation of pH and temperature (β fractionation). In practice, mutually unrelated premixed crystallization conditions are used (sparse matrix kits) for an initial screening of successful conditions to grow crystals. During the optimization, the precipitation point of a protein is determined at sequential precipitant concentrations with a given pH value. Different pH values are also checked keeping the precipitant concentration constant. The procedure is repeated at different temperatures (4 °C – 20 °C), and then the effect of different precipitating agent is examined. The precipitating (crystallization) agents could be divided in four main categories: salts (sulfates, phosphates), organic solvents (ethanol, acetone, propanol), long-chain polymers (polyethylene glycol 1000, 3350, 6000) and low-molecular-weight polyethylene glycols (PEG-400) and non-volatile organic alcohols (2-methyl-2,4-pentanediol, hexanediol). The role of macromolecular precipitants is to increase the effective protein concentration by dehydrating proteins through competition for water molecules (mainly salts, salting-out/in effects), by the reduction of the electrostatic screening effectiveness of the solvent (mainly alcohols), or by size-exclusion or crowded-out effect (PEG) [1].

A1.3.3 Methods for promoting supersaturation

Vapor diffusion

The method of vapor diffusion is the most widely used approach. It has produced more crystallized macromolecules than all other methods combined [7]. To initiate the trial, the sample drop is dispensed onto a siliconized surface, in this way it is spatially separated from a reservoir solution. When the drop is supported by some surface, the

method is called sitting drop (5 µl to 25 µl drop); when the drop is suspended from some surface- hanging drop (1 µl to 5 µl drop). The drop is prepares by mixing the same amount of protein and reservoir solution (fig. A1.2).

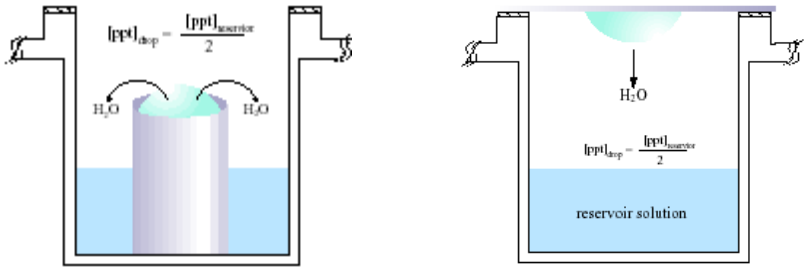


Figure A1.2 Sitting and hanging drop vapor diffusion method.

The reservoir solution is larger in volume (1 ml usually) and contains the precipitant and a buffer. The crystallization agent in the drop is in lower amount than that required for the formation of crystals. The entire system must be sealed out from air, so no solvent is lost to the environment and the only transport of water or any other volatile component is between droplet and reservoir solution. Over time, equilibration takes place through the vapor phase. It will be complete when the concentration of the precipitant in the droplet is equal to that in the reservoir. In the case of non-volatile precipitant, this occurs by the evaporation of water from the protein drop and its incorporation into the reservoir. When the precipitant is a volatile one, equilibrium takes place through the transport of solvent from the reservoir to the drop, due to its absence in the latter at the beginning of the experiment. Because the volume of the reservoir is much bigger than that of the droplet, the ultimate concentration of the precipitant in the drop of mother liquor will correspond to the initial, experimentally established conditions of the reservoir. When water is lost by the droplet to the reservoir, there is a volume decrease of the drop with an increase of protein and non-volatile components concentration (fig. A1.3). Vapor diffusion simultaneously provides at least two ways to drive the system toward supersaturation: increase in precipitant concentration and increase in protein concentration too.

In almost any vapor diffusion experiment, there will be some volume of air in the system between protein sample and reservoir. One of the first things to happen when the experiment is initiated is air enrichment in water. Thus, in the initial stages there will be a particularly pronounced loss of water and subsequent fast increase of supersaturation. Hanging-drop vapor diffusion is among the most popular methods now in use [1].

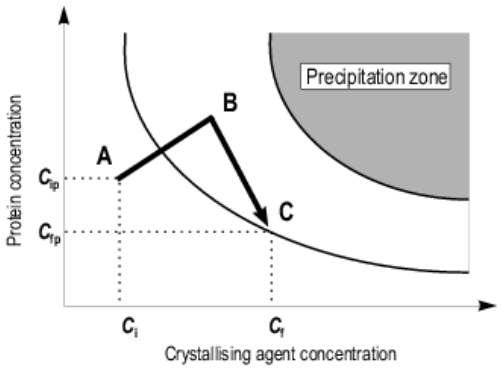


Figure A1.3. Phase diagram based on the vapor diffusion method.

Batch crystallization

This is one of the oldest techniques used for the crystallization of proteins [1]. It consists of direct mixing of an undersaturated protein solution with a precipitant. The latter alters the protein solubility or the electrolyte properties of the solution giving immediately saturated mother liquor. Thus, the protein and the precipitating agent are mixed at their final concentrations at the start of the experiment. There is less exploration of the phase diagram (fig. A1.4), and several microbatch trials may be required to replace a single vapor diffusion experiment. The introduction of the microbatch technique, where the crystallization samples are dispensed as small drops under oil, has overcome the requirement for large quantities of material (protein solution and precipitant).

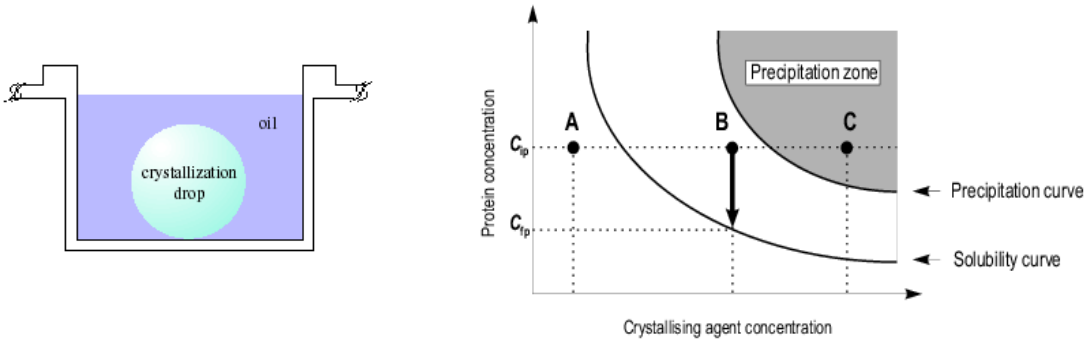


Figure A1.4. Microbatch method.

In fact, numerous crystallization conditions could be screened as 1-2 µl trials. Setting up microbatch trials is simpler and speedier than vapor diffusion trial even when performed manually because it is mechanically less complicated. The oil, also, protects the samples from evaporation (small-volume trials), airborne contamination and

physical shock. It was observed, as well, that the presence of oil might act as a stimulant for the crystallization of membrane proteins (chlorophyll binding protein 43 of the photosystem II membrane protein complex from spinach): the oil slowly absorbs the detergent from the aqueous drop, encouraging the protein to crystallize gradually [4]. The microbatch is not useful when volatile organic compounds are involved in the protein crystallization due to their interaction with the oil. The most frequently used oil is the paraffin one (a purified mixture of liquid saturated hydrocarbons obtained from petroleum). In a typical microbatch experiment, crystallization takes place within a couple of weeks or a month. During this time the drop is well sealed by the paraffin oil. Since oil and water are essentially immiscible, evaporation during this time is negligible. However, slow evaporation could occur (there is not absolute immiscibility), which can proceed until the drop dries out. Such slow evaporation cannot take place using a diffusion trial. It is, also, possible to control the rate of the vapor diffusion, even in the microbatch setup, introducing different types of oil (silicone, Al's oil). Silicone oil consist of polymers of repeating dimethylsiloxane units, that permit a freely diffusion of water. A mixture of paraffin and silicon oil (Al's oil: 50% paraffin and 50% silicon oil) allows partial diffusion. This is what is called modified microbatch. Combinations of the microbatch and diffusion methods are beginning to emerge that resulted very useful in the optimization of the crystallization conditions. Another example of possible combination is to place an oil barrier (mixture of paraffin and silicon oils in different ratio, 0.3 ml to 0.7 ml final volume) over a reservoir of a vapor diffusion trial in order to slow down the equilibration rate and thus to approach supersaturation more slowly [5, 1].

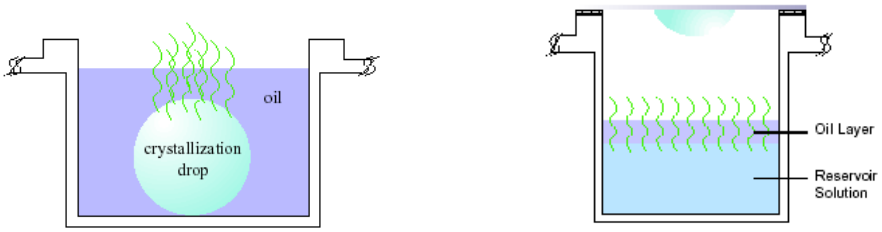


Figure A1.5. Modified microbatch and hanging drop methods.

A1.4 Protein crystal structure determination

The crystal structure analysis by X-ray or neutron diffraction is a method unique in providing an unambiguous and complete three-dimensional representation of the atoms in a crystal. In order to "see" the fine details of a molecular structure, it is necessary to use radiation of a wavelength comparable to, or smaller than, the distance of the atoms. That's why X-rays, an electromagnetic radiation with wavelengths of 10^{-7} m to 10^{-11} m (1000 Å to 0.1 Å) are used for the crystal structure determination. A succesfull high-resolution diffraction analysis requires that the sample is organized or prepared as an ordered array. It is because of this regular arrangement that the crystals diffract X-rays. The condition that determines the existence of the diffracted beam is provided by the Bragg's law:

$$\mathbf{n\lambda = 2d \sin\theta}$$

with **λ**, the wavelength of the radiation used, **n** an integer, **nλ** is the path difference between waves scattered from adjacent lattice planes with equivalent indices (hkl), **d**, perpendicular spacing between the lattice planes (hkl) in the crystal, and **θ**, angle of incidence of the X-ray beam (fig. A1.6). The diffraction maxima (hkl) occur when the path difference between waves scattered from adjacent parallel lattice planes are integral number of wavelength, or when the scattered waves are in phase. The angular distribution of the scattered radiation could be described as considering that diffracted beams behave like if they are reflected from planes passing through points of the crystal lattice.

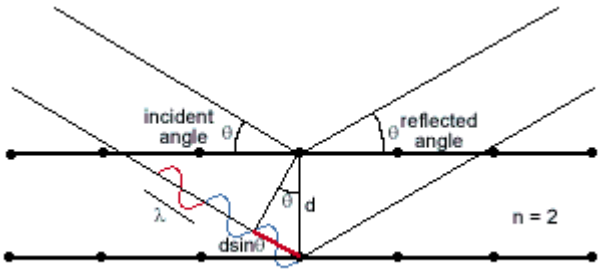


Figure A1.6. "Reflection" of X-rays by imaginary planes (hkl) through points in the crystal lattice. hkl are called Miller indices.

According to the Bragg's equation, if a monochromatic beam of X-rays hits a single crystal (particular sets of lattice spacing are present), diffraction maxima will occur only

for discrete values of the angle of incidence of the X-rays beam. If the crystal is rotated in the beam, additional diffracted beams will originate at given angles. In practice, more reflections could be observed due to the static disorder present in the real crystals (as if composed of minute blocks of unit cells, each one misaligned with respect to its neighbors), and due to the fact that the X-ray radiation is never truly monochromatic. Diffraction pattern is the output from the diffraction experiment and consists in a distribution of spots (diffraction maxima). The diffraction pattern of a crystal is arranged as a new lattice related to the lattice of the crystal (reciprocal lattice).
X-rays are scattered by the electrons in atoms and unfortunately can not be focused by presently known experimental technique to give an appropriately magnified image of the scattering matter. It is possible, however, to simulate the recombination of the scattered X-rays (transformation of the intermediate reciprocal space image to the real space image), just like the lens does, by a mathematical tool; the Fourier synthesis.

$$\rho(xyz) = 1/V_c x \Sigma\Sigma\Sigma_{all\ hkl}\ F(hkl)exp[-2\pi i(hx+ky+lz)]$$

$$F(hkl)=|F(hkl)|exp[i\alpha(hkl)]$$

$$\rho(xyz) = 1/V_c x \Sigma\Sigma\Sigma_{all\ hkl}|F(hkl)|exp[-2\pi i(hx+ky+lz)+ i\alpha(hkl)]$$

ρ(xyz) is the number of electrons per unit volume or electron density at any point x,y,z. **Vc** is the volume of the unit cell and **F(hkl)** is the structure factor for the particular set of indices **hkl**. The diffracted radiation has a particularly combination of amplitude (**|F(hkl)|**) and phase (**α(hkl)**) known as the structure factor **F(hkl)**. The diffracted (scattered) X-radiation is intercepted by a detecting system. In this way only the intensities (**I**) of the diffracted waves could be measured and the phases are lost. The amplitude of the diffracted beam is easily derived from the measured intensity.

$$I(hkl)= K|F(hkl)|^2 (Lp)(ABS)$$

For the intensity general expression, **Lp** is Lorentz-polarization factor, **ABS** is the absorption factor and **K** is a scale factor.
In order to compute the Fourier synthesis, the relative phase of the diffracted beams should be determined (solving the "phase problem"). The aim of an X-ray diffraction study is to obtain the electron density map of the structure, using the measured amplitudes for each of the diffracted waves and their respective phase angles derived by one of the available methods. Then, assuming atomic nuclei to be at the centers of peaks maxima, we would know the entire structure. Since a crystal is build up of an extremely large number of regularly stacked unit cells, each of which has identical content, the

problem of determining the structure of the crystal is reduced to that of determining the spatial arrangement of the atoms within a single unit cell, or within the asymmetric unit. Thus, the required calculations for the electron density map estimation will be greatly simplified. The calculation of the Matthews coefficient [6] allows estimating the number of protein molecules present in an asymmetric unit and the percentage of its solvent content. The asymmetric unit is the smallest part of the crystal structure from which the complete structure can be obtained by applying the space group symmetry operations. The 230 space groups represent the distinct ways of arranging identical objects on one of the 14 Bravais lattices (corresponding to 7 different crystal systems) by the use of symmetry operations. From the 230 possible space groups, only the 65 non-centrosymmetric space groups are possible for protein crystals, because of the chirality of the amino acids presented in proteins (L-enantiomers). Thus, the allowed symmetry operations are rotation, screw axis and translation, but not mirror plane and inversion center [3].
The basic steps towards the structural determination are:
1 Collection, processing and reduction of X-rays diffraction data;
2 Phase determination;
3 Interpretation of electron density maps (model building);
4 Refinement of the molecular model.

A1.4.1 Collection processing and reduction of X-rays diffraction data

Once a diffraction pattern was visualized and a sufficient number of frames collected, data processing is carried out. It consists of: assigning an index (hkl) to each spot (indexing) and calculation of the cell dimensions with a preliminary identification of the crystal class and Bravais lattice. Intensity estimation for each diffraction maximum (integration). After the intensity adjustment on a common scale is derived, the symmetry equivalent reflections are merged, according to the crystal symmetry. The space group may be determined at this point, on the basis of merging statistics (Rsym) and the list of systematic absences. R_{sym} is an indicator of the intrinsic data quality. It is the reability factor of intensities.

$$\mathbf{R_{sym}(I)= \Sigma_{hkl}\Sigma_i| I_i(hkl)-Imean(hkl)|/\Sigma_{hkl}\Sigma_i|I_i(hkl)|}$$

The result of the X-ray diffraction data collection is a list of relative intensity, **I**, for each reflection with indices h, k, l and an error estimate. Only after the intensities have been put on absolute scale the amplitudes for all reflections measured can be computed as square root of I.
Intensities can be put on an absolute scale by using a scaling factor and the temperature factor **B** obtained from the Wilson plot. The temperature factor or Debye-Waller factor

is a consequence of the atomic motion in the crystal caused by temperature-dependent vibration of the atoms in the structure.

$$\mathbf{B=8\pi^2<u^2>}$$

The mean square displacement of the atomic vibration is indicated as $\mathbf{<u^2>}$. Taking into account the thermal vibrations, the atomic scattering factor and F(hkl) are expressed as follows:

$$\mathbf{f_j^2 = exp(-2B_j sin^2\theta/\lambda^2)(f_j^0)^2}$$

$$\mathbf{F(hkl) = \Sigma_i fi\ exp\ [2\pi_i(hx+ky+lz)]}$$

where **j** is a specific atom, and $\mathbf{f_j^0}$ is the scattering factor of the atom j at rest.

A1.4.2 Phase determination

To compute **ρ** for all values of x, y, z and to plot **ρ(xyz)** in order to obtain a three-dimensional electron density map, not only the **|F(hkl)|** values should be known, but the phase angle **α(hkl)** as well. The four existing techniques for solving the phase problem in protein X-rays crystallography are:

- Direct methods

Statistical and probabilistic techniques for deriving an approximate set of phases from which the electron density map can be calculated. The atomicity and the fact that the electron density in a real crystal can not be negative are used. Required characteristics are 1.1-1.2 Å data, 100% completeness in the last shell, the number of non-H atoms in the asymmetric unit should be less than 200. This is a method indicated especially for small molecule crystallography [2, 7].

- Isomorphous replacement

Requires covalent or non covalent binding of heavy atoms to the protein molecules (heavy atom derivatives). The crystal space group, cell dimensions and the arrangements of the protein atoms should be essentially identical (isomorphism). In practice, the cell dimensions of two isomorphic crystals should not differ more than 1.0 %. Differences in the diffraction intensity pattern are due to the presence of the heavy

atom. Once the position of the heavy atom is determined (by difference native-derivative Patterson map), the difference in the intensities for the two isomorphs can be used to refine the initial phase angle for each reflection. At least two pairs of isomorph compounds are required.

- Multiple wavelength anomalous dispersion

The effect of anomalous scattering (dispersion) is an intensity difference between reflection with indices hkl and those with indices -h-k-l. It is due to absorption of X-rays on an atom (anomalous scatter) in the structure and results in phase and intensity change for the X-rays scattered from that atom. This is the equivalent of advancing or delaying the radiation for a short time. The intensity difference due to the presence of anomalous scatters in the protein molecule could be used for the phase angle determination. Usually the diffraction pattern is measured at wavelength near to and far from the absorption edges of an anomalous scatter presented in the structure.

- Molecular replacement (MR)

Particular attention is dedicated to this method, because of its application in the *Pseudomonas aeruginosa* PA3859 structural determination. The method is based on the observation that homologue proteins show a similar folding. If the structure of a homologue protein is known, it could be used as a model for the refinement procedure. Thus, the estimates of the phases can be computed and electron density calculated. The success of MR depends largely on the similarity between the model and the protein of interest. The model should be complete and has to share at least 40 % sequence homology with the unknown structure. The molecular replacement is a six dimensional search, in which the model is placed (orientated and positioned), in the target unit cell. In practice, it involves two steps: rotation and translation. In the rotational step, the spatial orientation of known (search model) and unknown molecule with respect to each other is determined, while in the second step the translation that superimposes the already correctly oriented model to the molecule is calculated. These steps are carried out by Patterson search technique. The Patterson map is basically a Fourier map calculated with the square of the structure factor amplitudes and phases set to zero.

$$\mathbf{P(u,v,w) = 1/V_c x \Sigma\Sigma\Sigma_{all\ hkl} |F(hkl)|^2 cos[2\pi(hx+ky+lz)]}$$

The function is evaluated at each point u,v,w, of a three-dimensional grid that fills a space with the size and shape of the unit cell [3].
The Patterson map gives a vector representation of the atomic contents of the unit cell. It is a signature for a protein structure. The peaks in the map occur at points whose distances from the origin correspond in magnitude and direction with distance between atoms in the crystal. The vectors in the Patterson map can be divided in two categories:

intramolecular (fig. A1.7, from one atom in the molecule to another atom in the same molecule) and intermolecular (between atoms of symmetry related molecules). The intramolecular vectors depend only on the orientation of the molecule and are involved in the rotation function.

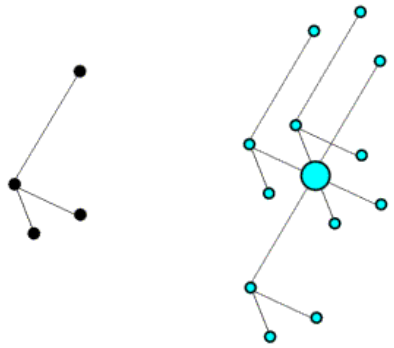


Figure A1.7. A molecule and its Patterson intramolecular vector representation.

The observed Patterson map can be computed from the experimental diffraction intensities. The model Patterson map, on the other hand, can be calculated from the search model structure factors, and compared with the observed Patterson by superposition. The model is placed in a P1 unit cell. In this way it is represented in the model Patterson map only by peaks arising from intramolecular vectors and from those intermolecular vectors that connect copies of the molecule generated from the lattice translations. Since the two vector types are spatially intermixed (fig. A1.8A, [8]), they could be separated placing the search model in the artificially enlarged unit cell (fig. A1.8B). The intramolecular vectors are thus concentrated around the origin of the Patterson map and the intermolecular vectors are concentrated around the other lattice points. It is possible to cut out the spherical region around the origin that contains the isolated intramolecular vectors and to use it as a partial Patterson map in the rotational search (fig. A1.8C). The radius of the spherical region is chosen to be similar to the largest intramolecular vector. To improve the signal-to-noise ratio, the region around the large Patterson origin peak is omitted.

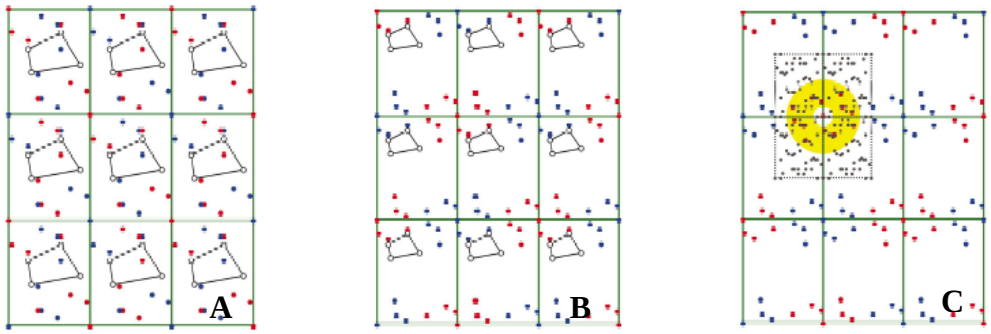


Figure A1.8. **A**. A molecule and the corresponding Patterson peaks translated to each lattice point; **B**. The molecule and the corresponding Patterson peaks in an artificially enlarged unit cell. The molecule is scaled down instead of enlarging the cell; **C**. Superposition of the model and calculated Patterson maps [8].

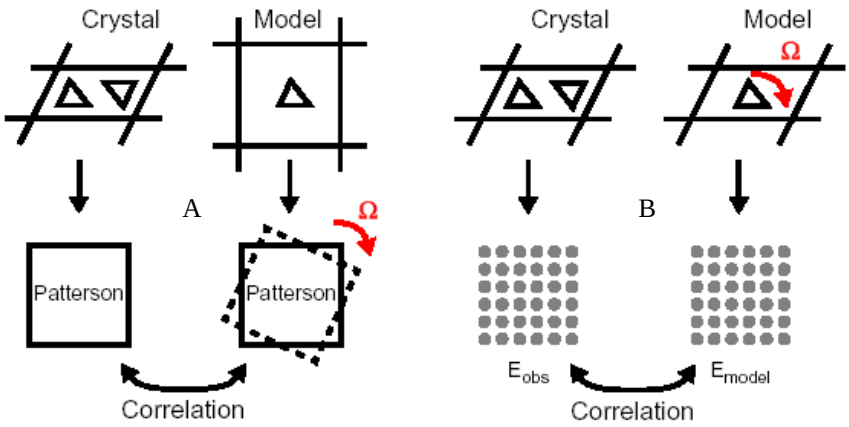


Figure A1.9. Traditional rotation search, A, and direct rotation search, B [8].

In the direct rotation search (CNS), on the other hand, the model is rotated directly and the structure factor calculations are carried out for each sampled angular orientation (fig. A1.9).

To describe an orientation, three parameters are required. In the kappa/phi/psi description, the rotation axis is defined in the x,y,z coordinate system by the angles psi (y) and phi (z). The angle kappa gives the rotation around the rotation axis. Another convention makes use of Eulerian angles. The coordinate system is rotated by an angle α around the original z axis, then by an angle β, around the new y axis and in the end by

an angle γ around the final z axis. In CNS package both a "traditional" and "direct" approach to the rotational search are possible. Traditional rotation search is carried out in the real space and the rotation **Rot(Ω)**, which has to be maximized, is a Patterson correlation integral. Given a specific angular orientation Ω, Rot(Ω) assumes the following form:

$$\mathbf{Rot(\Omega) = \int_U P_o\,(u) P_c(\Omega, u) du}$$

Where U is the Patterson space u is a location vector and P_o is the observed Patterson function and P_c is the search-model Patterson function.

In direct rotation search, which is carried out in reciprocal space, the rotation function is evaluated as the linear correlation coefficient of the observed and calculated normalized structure factors. For a given angular orientation Ω of the search model CC(Ω) assumes the following form:

$$\mathbf{CC(\Omega)=[\Sigma_{hkl}\,(X_{hkl,\,o} - mean\ X_o)(X_{hkl,\,\Omega} - mean\ X_\Omega)]/\{[\,\Sigma_{hkl}(X_{hkl,\,o} - mean\ X_o)^2]^{1/2}\,[\Sigma_{hkl}\,(X_{hkl,\,\Omega} - mean\ X_\Omega)^2]^{1/2}\}}$$

Where $X=|E|^2$.

Reciprocal and direct space approaches do not normally differ in results in absence of approximations, nonetheless the much more computationally intensive direct search gives, in general, better results, especially in difficult cases, because of the lack of the approximations which are instead normally used in traditional rotational search (limited numbers of Patterson peaks considered and grid points interpolation).

Usually a Patterson correlation (PC) refinement is carried out before performing the translation search. It could be useful in enhancing the discrimination between correct and incorrect rotation function points, improving the search model, and in enhancing the signal-to-noise ratio. A drastically reduction in the number of noise peaks in the translation function could be obtained [8].
Once the angular orientation of the search model is assumed to be known, the location of the oriented search model should be determined with respect to the symmetry elements. The unit cell is subdivided in a regular grid and the search model is moved to each grid point in turn. The translation search is evaluated by the standard linear correlation coefficient (modified for translation) of the structure factor amplitudes and by the crystallographic factor **R**.

$$R = 100x\ \Sigma_{hkl}\ |(|Fo| - k|Fc|)|/\ \Sigma\ (|Fo|)$$

In order to evaluate the **R** factors and correlation coefficients for the translation search, structure factors need to be calculated for the search model, with a given orientation, at different positions in the unit cells. In the case of the correct solution, the correlation coefficient should be as high as possible, and the R-factor as low as possible.
In the case of the traditional translation search, the correct position of a search model in the crystal unit cell should lead to the maximal overlap of both the intra- and intermolecular vectors (maximal overlap between the observed and the calculated Patterson maps) throughout the entire unit cell.

$$T = \int_{\Omega} P_o(u) P_c(u) du$$

The results of the calculations are on the arbitrary scale and this makes difficult to compare results from different calculations. The R factor or the correlation coefficient can be calculated for the top peaks to place the results on the absolute scale. The intramolecular vectors can be removed from the Patterson maps by subtracting, after appropriate scaling, the structure factors calculated from individual molecules.
When the orientation and the position of the search model within the unit cell is know, the phases and the electron density map could be calculated.

A1.4.3 Model building

This is an interpretation of the calculated electron density map, consisting in tracing protein backbone of the model through the electron density map and in insertion of the ordered water molecules as well.

A1.4.4 Refinement

By a three-dimensional Fourier synthesis of the observed structure factor amplitudes with a starting phases that is an approximation of the true electron density map can be calculated. Once the approximate positions and identities of the atoms in the asymmetric unit are known (model building), the amplitudes and the phases of the structure factors can be calculated. The calculated amplitudes |**Fc**| may be compared with the observed ones |**Fo**|. If the structure model is a correct one and the experimentally observed data are reasonably accurate, the agreement should be good. The refinement is a systematical variation of the positional parameters (x, y, z) and of the temperature factors (B) for all atoms in the structure (except hydrogen atoms) in order to obtain the best possible agreement of the observed structure factor amplitudes

with those calculated for the proposed structure. Some additional information are also incorporated in the refinement process like stereochemical data derived from high resolution small molecule structures, assuming that bond length and angles determined with high precision in amino acids and small peptides are also valid in proteins. The refinement is performed in successive cycles until a convergence is reached. This is a stage at which the shifts from cycle to cycle in the parameters being refined are negligible with respect to the energy gradient. If the starting model is not too different from the real structure the refinement converges easily to the correct solution. If the distance between the atoms in the model and in real structure is large, the refinement may be trapped in a local minimum instead of reaching the true, global energy minimum of the target:

$$\mathbf{E = E_{chem} + w_{x\text{-}ray} E_{x\text{-}ray}}$$

$\mathbf{E_{chem}}$ comprises empirical information about chemical interactions. It is a function of all atomic positions, describing covalent and not bonded interactions. The energy for strereochemical terms is evaluated by considering the difference between the actual value and the ideal value in a way that if the atom deviate more from the ideal geometry the energy increases. The crystallographyc term, $\mathbf{E_{x\text{-}ray}}$ comes from the difference between observed and calculated data; $\mathbf{w_{x\text{-}ray}}$ is a chosen weight.

To avoid a local (false) minimum, an optimization technique called simulated annealing, that uses molecular dynamics, is carried out. In molecular dynamics, the dynamic behavior of a system of particles is simulated. In the simulated annealing technique, the system temperature is raised sufficiently high for the atoms to overcome energy barriers and then they are cooled slowly to approach the energy minimum. Large errors in the starting model could be corrected and this will speed up the refinement procedure. Usually the simulated annealing protocol is performed at the beginning of the refinement. Model modifications that are needed to give better agreement with observation, can be deduced just by comparing the calculated maps with **|Fo-Fc|** and **|2Fo–Fc|** coefficients respectively. The maps are inspected visually and model corrections are done manually in agreement with the observation. A **Fo-Fc** difference Fourier map (difference between the observed and calculated structure features) will show positive electron density at atom sites not included in the model and negative density at atom sites present in the starting model, but missing in the real structure.

The water molecules are added automatically in the end of the refinement process by analyzing both the **2Fo-Fc** and **Fo-Fc** maps.

One measure of the correctness of the structure is the so-called discrepancy index, or crystallographic **R** value, defined as:

$$\mathbf{R = \Sigma_{hkl} |(|Fo| - k|Fc|)| / \Sigma (|Fo|) \qquad (hkl)}$$

It is a measure of how closely the observed structure factor amplitudes are matched by the values calculated for the model. A more objective version of this quality indicator is obtained if the calculation extends over a set of (5 - 10)% of the reflections only, chosen randomly and excluded from refinement (**R_{free}**). If the structure is improving during the refinement, the R and Rfree will decrease. For a structure determination with data resolution up to 2.0 Å, the final R factor should be close to 20% with Rfree a few percent higher. The calculated r.m.s.d. value is an estimation of the accuracy of the results. It is an assessment of the covalent geometry of the model by comparing the bond lengths and angles to a library of ideal values [2, 3, 9].

The purpose of refinement and rebuilding is to detect and fix the errors in order to obtain the best possible final model. Some of the model quality indicators like the crystallographic R factor, R_{free} and r.m.s.d for the covalent bond angles and lengths, have been already discussed. They are global indicators used to obtain an impression of the overall quality of the model and the overall fit of the model to the data. The Ramachandran plot, on the other hand, is one of the local model quality indicators. Ramachadran plot investigates the stereochemistry of the main chain folding and especially of the main-chain dihedral angles. The angles φ and ψ for each amino acid residue are plotted in 2D diagram. The Ramachadran plot classifies the residue in one of the three categories: "allowed", "partially allowed" and "disallowed" conformations. The "partially allowed" conformations give rise to modestly unfavorable repulsion between non bonded atoms and might be overcome by attractive effects such as hydrogen bonds. The "disallowed" interactions give highly unfavorable non-bonded interatomic distances. Since the Ramachadran plot can identify unacceptable clusters of φ-ψ angles, it is used to reveal possible errors made during model building and refinement. As opposed to covalent bond angles and bond lengths, the main-chain dihedral angles are usually weakly restrained during X-ray refinement and therefore can be used to validate the structure model independently. In practice, the Ramachandran plot is one of the simplest most sensitive tools for assessing the quality of a protein model [9].

A1.5 References

[1] McPherson, A. (1999). *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, New York.

[2] Drenth, J. (1999). *Principals of protein X-Ray Crystallography*. Springer-Verlag, New York.

[3] Glusker, J. P.,Trueblood, K. N. (1985). *Crystal Structure Analysis. A primer*. Second Edition. Oxford University Press, New York.

[4] Chayen, N.E. (1998). *Acta Cryst.* **D54**, 8-15.

[5] Chayen, N.E. (1999). *J. Cryst. Growth* **196,** 434-441.

[6] Matthews, B. W. (1968). *J. Mol. Biol*. **33**, 491-497.

[7]Gilmore, C.J. (2000). *Acta Cryst*. **D56,** 1205-1214.

[8] Grosse-Kunstleve, R.W., Adams, P.D. (2001). *Acta Cryst*. **D57**, 1390-1396.

[9] Rossman, M.G., Arnold, E. (2001). *International Tables for Crystallography* vol F. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Appendix 2

Automated docking simulations

In any protein-ligand docking scheme two conflicting requirements must be balanced: the desire for a robust and accurate procedure and the desire to keep the computational demands at a reasonable level. The ideal procedure should find the global minimum for the energy of interaction between the substrate and the target protein, by exploring all available degrees of freedom for the system. However, it must also run on a computer within an acceptable amount of time. In order to meet these demands a number of techniques simplify the docking procedure. The method developed for AutoDock 3.05 [1, 2] tackles the search space exploration and the energy evaluation separately. Configurational exploration is performed by a genetic algorithm (see paragraph A2.3) and energies are evaluated using grid-based molecular affinity potentials (see paragraph A2.1). It thus combines the advantages of exploring a large search space and a robust energy evaluation.

A2.1 Energy evaluation

Rapid energy evaluation is achieved by precalculating atomic affinity potentials for each atom type in the substrate. The protein is embedded in a three-dimensional grid and a probe atom is placed at each grid point (Fig. A2.1). The energy of interaction of

this single atom with the protein atoms is assigned to the grid point. An affinity grid is calculated for each type of atom in the substrate. The energetics of a particular substrate configuration is then found by interpolation of affinity values of the eight grid points surrounding each of the atoms in the substrate.
The electrostatic interaction is evaluated similarly, by interpolating the values of the electrostatic potential and multiplying it by the atom charge.
The time to perform an energy calculation using the grids is proportional only to the number of atoms in the substrate, and is independent of the number of atoms in the protein. Each point within the grid map stores the potential energy of a probe atom that is due to all the atoms in the macromolecule.

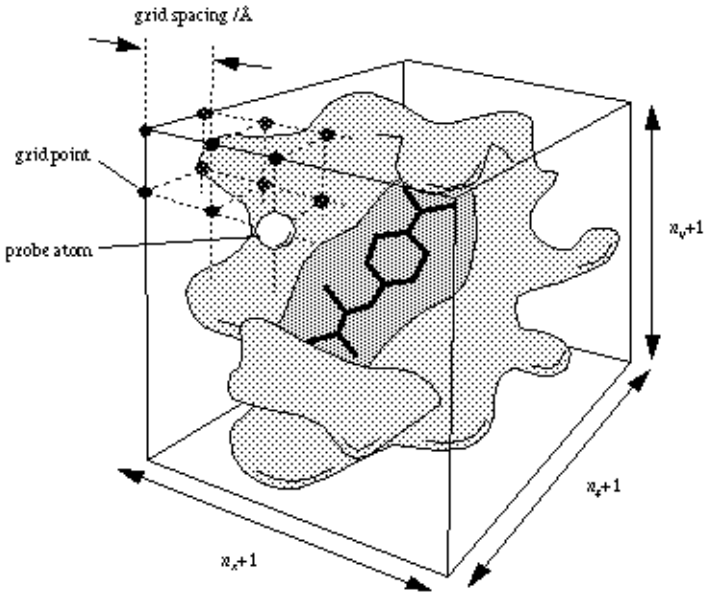
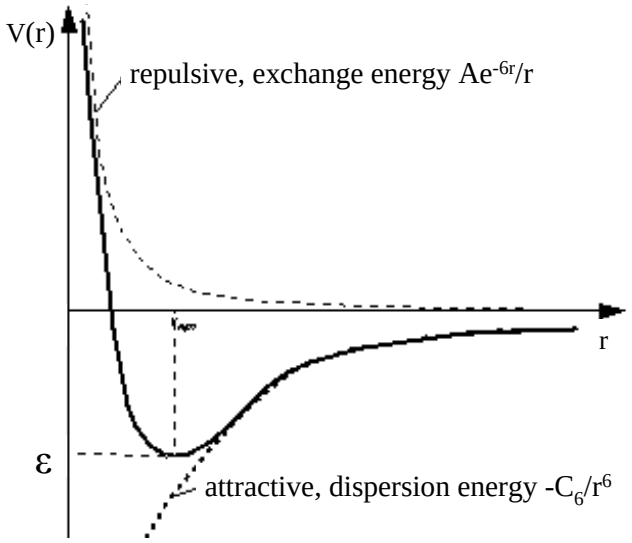


Figure 5.1. Grid map main features: the ligand is in the centre of the grid map, buried inside the active site of the protein. In this case, the grid map encompasses the whole protein.

The pairwise potential energy, V(r) between two non-bonded atoms can be expressed as a function of internuclear separation, r, as follows,

$$V(r) = \frac{Ae^{-6r}}{r} - \frac{C_6}{r^6}$$

Graphically, if r_{eqm} is the equilibrium internuclear separation and ε is the well depth at r_{eqm}, then:



The exponential, repulsive, exchange energy is often approximated to,

$$\frac{A}{r}e^{-6r} \cong \frac{C_{12}}{r^{12}}$$

hence pairwise-atomic interaction energies can be approximated using the following general equation,

$$V(r) \cong \frac{C_n}{r^n} - \frac{C_m}{r^m}$$

where m and n are integers, and C_n and C_m are constants whose values depend on the depth of the energy well and the equilibrium separation of the two atoms' nuclei. Typically the 12-6 Lennard-Jones parameters (n =12, m =6) are used to model the Van der Waals' forces experienced between two instantaneous dipoles.

In addition to the atomic affinity grid maps, AutoDock requires an electrostatic potential grid map. Partial atomic charges must be assigned to the macromolecule. The electrostatic grid can be generated by the program AutoGrid, which calculates Coulombic interactions between the macromolecule and a probe of charge e, $+1.60219\times10^{-19}$ C; there is no distance cutoff used for electrostatic interactions. A sigmoidal distance-dependent dielectric function is used to model solvent screening.

$$\varepsilon(r) = A + \frac{B}{1 + Ke^{-KBr}}$$

Where $B = \varepsilon_0 - A$; ε_0 is the dielectric constant of bulk water at 25 °C; A, l and K are constants.

The atomic partial charges are written in PDBQ format that is an augmented form of the standard PDB format, in which an extra column is used to store the partial atomic charges (hence the "Q" in "PDBQ"). In addition a new utility program is used to specify the atomic fragmental volume and atomic solvation parameters for each atom in the macromolecule. The macromolecule solvation parameters are stored in a PDBQS file. This resembles a PDBQ file, but in addition gives the solvation parameters for each atom. Default solvation parameter values are instead assigned to the ligand probe atoms.

A2.2 Overview of the Free Energy Function

A scoring function based on the principles of QSAR (quantitative structure-activity relationships) is applied in order to rank the docking solutions. It was parameterized using a large number of protein-inhibitor complexes for which both the structures and inhibition constants, or K_i, were known.

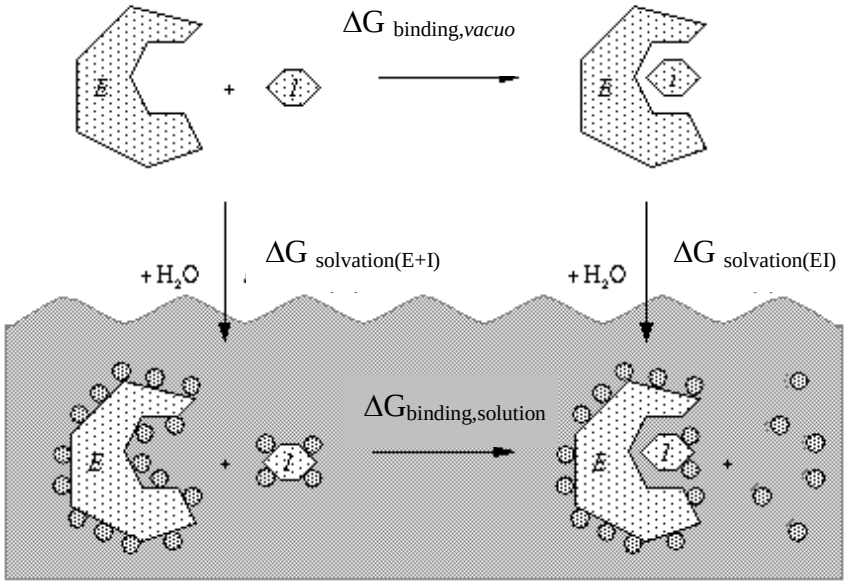


Figure A2.2. Cicle for enzyme substrate binding.

According to the Gibbs free energy equation, the change in the free energy for a given process is given by $\Delta G = \Delta H - T\Delta S$.
Figure A2.2 shows the thermodynamic cycle for the binding of an enzyme, E, and a ligand, I , in both the solvated phase and in *vacuo* . The solvent molecules are indicated

by filled circles: they tend to be ordered around the larger molecules, but when E and I bind, several solvent molecules are liberated and become disordered.
This entropic effect is the basis of the hydrophobic interaction. The solvent ordering around E and I in either the bound and unbound state is strongly influenced by hydrogen bonding: hydrogen bonds between solvent and E, and solvent and I, contribute an enthalpic stabilization that can be estimated by Autodock.
According to the Hess law of heat summation, the change in free energy between two states will be the same, no matter what the path. So the free energy of binding in solution we can be calculated by the following equation:

$$\Delta G_{binding,solution} = \Delta G_{binding,\textit{vacuo}} + \Delta G_{solvation(EI)} - \Delta G_{solvation(E+I)}$$

Since $\Delta G_{binding,\textit{vacuo}}$ can be calculated from the docking simulation, and is possible to estimate the free energy change upon solvation for the separate molecules E and I and for the complex, EI , $\Delta G_{solvation(EI)}$ and $\Delta G_{solvation(E+I)}$ respectively, then it is also possible to calculate the free energy change upon binding of the ligand to the enzyme in solution, $\Delta G_{binding,solvation}$. Thus, the binding constant, K_i, for the ligand, I can be estimated.
In this free energy scoring function various terms in the molecular mechanics energy function are scaled by coefficients derived using a linear regression model that most closely fit the experimentally observed inhibition constant data. Furthermore, a term for the evaluation of the ligand loss of conformational degrees of freedom up on binding is also applied.

A2.3 Search Space Exploration

The set of feasible docking solutions among which the desired solution resides is called search space. Each point in the search space is a possible protein-ligand interaction mode. Each solution is characterized by a defined free energy of binding that can be evaluated as described above. Thus the docking simulation goal is to find the overall minimum energy point within the search space.
In the case of a docking problem the search space is extremely wide and its exhaustive exploration is not feasible due to the huge amount of calculations, and thus time, that it would require. A number of different methods have been developed in order to speed up the search of the overall minimum. In Autodock a Genetic Algorithm (GA) is used.

Genetic algorithms are a part of evolutionary computing, which is a rapidly growing area of artificial intelligence: they are inspired by Darwin's theory of evolution.

Each docking solution is codified in a numeric string in which the ligand position, orientation and conformation are stored. In analogy with biologic systems, the string can be regarded as a chromosome and the single variables within the string as genes. The algorithm begins with a set of random solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population by means of crossover. This is motivated by a hope, that the new population will be better than the old one. The probability for a solution to transmit its genotype to the next generation is a function of its fitness, the higher is the fitness (low free energy of binding) the more chances it has to reproduce. Thus the offspring arise from a random combination of translation, rotation and dihedral values of the best individuals. Furthermore, the new solution can be subjected to a random mutation: this feature has been demonstrated to be important in order to avoid the system getting trapped in a local minimum.
The user has to specify 4 parameters:
crossover probability, that is how often crossover will be performed. If there is no crossover, offsprings are exact copies of the parents. If there is crossover, offsprings are made from parts of both parent's chromosomes. If crossover probability is 100%, then all offsprings are made by crossover. If it is 0%, whole new generation is made from exact copies of chromosomes from the old population. Crossover is made in hope that new chromosomes will contain good parts of old chromosomes and therefore the new chromosomes will be better. However, it has been shown that leaving some part of old population survive to next generation is necessary in order to guarantee a big enough genetic diversity and thus the possibility, for future generations, to produce better individuals.

When the crossover rate is high, the probability of loosing the best solution is also high. To avoid this a feature called eliticity is used, and the user has to specify how many of one generations' best individuals have to be directly copied in the next generation.

Mutation probability: how often parts of the chromosome will be mutated. If there is no mutation, offspring are generated immediately after crossover (or directly copied) without any change. If mutation is performed, one or more parts of a chromosome are changed. If mutation probability is 100%, whole chromosome is changed, if it is 0%, nothing is changed.

Mutation generally prevents the GA from falling into local minima. It should not occur very often, otherways the GA will turn in a random search.

Population size: big population size usually does not improve performance of GA. The best population size, that is the size that allows the fastest search, depends on the size of

the encoded string (chromosome). In AutoDock it has been evaluated to be between 50 and 100 individuals, depending on the ligand size and torsional degree of freedom.

In AutoDock version 3.05 a new hybrid search technique is employed that implements an adaptative global optimizer with a local search [3]. The global search method is a genetic algorithm with a 2-point cross over (Fig. 5.3) and random mutation. The local search method, that is made to improve the fitting of a single docking solution rather than the entire population, is based on the optimization algorithm of Solis and Wets [4]. The local searcher modifies the phenotype, which in turn updates the genotype: clearly this contravenes Mendelian genetics observed in nature, but it does improve the overall performance of the method. This hybrid genetic algorithm with phenotypic local search has been called Lamarckian Genetic Algorithm (LGA), since it utilizes the discredited Lamarckian notion that an adaptation of an individual to its environment can be inherited by its offspring.

Chromosome 1	ABCD | EFGH | ILMN
Chromosome 2	ABCD | EFGH | ILMN
Offspring 1	ABCD | EFGH | ILMN
Offspring 2	ABCD | EFGH | ILMN

Figure A2.3. Example of a 2 point crossover.

A2.4 References

[1] Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Comp. Chem.* (1998) **19**:1639-1662.

[2] (Goodsell, D. S., Morris, G. M. and Olson, A. J. Docking of Flexible Ligands: Applications of AutoDock. (1996), *J. Mol. Recognition*, **9**: 1-5.

[3] William Hart's doctoral thesis describes this hybrid global-local method, and can be found on http://www.cs.sandia.gov/~wehart/abstracts_html/thesis.html).

[4] Solis, F.J., Wets. R. J. Minimization by random search techniques. *Mathematical Operations Research*. (1981) **6**: 19-30.

Aknowledgments

I would like to thank:

Francois Sanschagrin and Roger Levesque from University of Laval, Canada, for the experiment with the *Pseudomonas aeruginosa* PA3859 knock out mutant in the chronic lung infection model in rats.

Giulia Devescovi, Guliano Degrassi and Vittorio Venturi from the bacteriology unit of ICGEB, Trieste, for their collaboration part of the work reported in Chapter 2.

Professor Bauke W. Dijkstra, for accepting the task of the external examiner.

Doriano Lamba, for having supervisioned me.