UNIVERSITA' DEGLI STUDI DI VERONA

DIPARTIMENTO DI
BIOTECNOLOGIE

DOTTORATO DI RICERCA IN
BIOTECNOLOGIE MOLECOLARI, INDUSTRIALI ED AMBIENTALI

CICLO XXIII

STRUCTURAL STUDIES OF HUMAN PROTEINS OF MEDICAL INTEREST

S.S.D. BIO/11

Coordinatore: Prof. Roberto Bassi
Firma _____

Tutor: Dott. Massimiliano Perduca
Firma _____

Dottorando: Manuel Furlani
Firma _____

Index

Riassunto

L'obiettivo di questo lavoro di tesi era determinare la struttura tridimensionale di tre proteine umane, Heat shock protein 60 (HSP60), Activation-induced cytidine deaminase (AICDA) e Cholesterol 7-alpha-monooxygenase (CYP7A1), attraverso la tecnica di diffrazione di raggi X.

La chaperonina umana HSP60 è una proteina mitocondriale, espressa in maniera costitutiva. La proteina è stata espressa in *E. coli* e purificata tramite cromatografia di affinità immobilizzata dello ione metallo, utilizzando un tag di sei istidine inserito all'estremità N-terminale della proteina, e tramite cromatografia a esclusione molecolare. Le prove di cristallizzazione non danno risultati positivi, probabilmente per problemi di eterogeneità della proteina.

La AICDA umana è una citidina deaminasi, la quale è selettivamente espressa nei linfociti B e svolge un ruolo cruciale nell'ipermutazione somatica e nella ricombinazione per scambio di classe degli anticorpi. L'espressione della proteina è stata tentata in *E. coli* con differenti vettori ed è stata ottenuta con successo con il plasmide pGEX-4T-1, che permette di esprimere la GST all'estremità N terminale della proteina oggetto di studio; sfortunatamente, il protocollo di purificazione non è stato efficace.

La CYP7A1 è un enzima microsomale che catalizza la conversione del colesterolo a 7α-idrossicolesterolo, la prima reazione della sintesi degli acidi biliari, e questa è la reazione limitante la velocità della via metabolica; la proteina è espressa solamente nel fegato. L'espressione della proteina umana CYP7A1 in *E. coli* è stata difficoltosa e, sebbene siano stati testati differenti vettori, non è stato ottenuto un buon livello di espressione della proteina. Per superare questi problemi, la CYP7A1 di pesce zebra (*Danio rerio*) è stata espressa in *E. coli*, ma senza apprezzabili miglioramenti.

Un'altra parte della mia tesi ha riguardato la determinazione della struttura tridimensionale della Methionine aminopeptidase 1 umana con due differenti inibitori. La proteina catalizza l'eliminazione della metionina N-terminale dalle proteine nascenti. Inibitori contro questa proteina sono di grande interesse medico grazie al loro potenziale impiego come farmaci antitumorali.

Questa parte di tesi è stata svolta presso il laboratorio del Dr. L. Mario Amzel sotto la

supervisione della Dr. Sandra B. Gabelli - Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, School of Medicine (Baltimore, USA).

Abstract

The aim of this thesis work was to determine the three-dimensional structure of three human proteins, Heat shock protein 60 (HSP60), Activation-induced cytidine deaminase (AICDA) and Cholesterol 7-alpha-monooxygenase (CYP7A1), by X-ray diffraction of single crystals.

The human chaperonine HSP60 is a mitochondrial protein, expressed in a constitutive manner. The protein was expressed in *E. coli* and purified by immobilized metal ion affinity chromatography, using a histidine tag fused to the N-terminus of the protein, and by size exclusion chromatography. The crystallization trials do not give positive results, probably due to protein heterogeneity problems.

Human AICDA is a cytidine deaminase, which is selectively expressed in B lymphocytes and plays a crucial role in antibody somatic hypermutation and class switch recombination. Protein expression was attempted in *E. coli* with different vectors and was successfully achieved with the pGEX-4T-1 plasmid, that allows to express GST at the N-terminus of the target protein; unfortunately, the purification protocol was not effective.

CYP7A1 is a microsomal enzyme that catalyzes the conversion of cholesterol to 7α-hydroxycholesterol, the first reaction of bile acid synthesis and the rate-limiting step of the metabolic pathway; the protein is expressed only in the liver. The expression of human CYP7A1 in *E. coli* was troublesome and, although different vectors were tested, a good level of protein expression was not obtained. In order to overcome these problems, zebrafish (*Danio rerio*) CYP7A1 was expressed in *E. coli*, but without any substantial improvement.

Another part of this thesis work concerned the determination of the three-dimensional structure of the human Methionine aminopeptidase 1 with two different inhibitors. The protein catalyzes the removal of the N-terminal methionine from nascent proteins. Inhibitors against this protein are of great medical interest because of their potential employment as anticancer drugs.

This part of the thesis was performed at Dr. L. Mario Amzel's laboratory under the direct supervision of Dr. Sandra B. Gabelli - Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, School of Medicine (Baltimore, USA).

Abbreviations

HSP: Heat shock protein
AICDA: Activation-induced cytidine deaminase
CYP7A1: Cholesterol 7-alpha-monooxygenase
MetAP: Methionine aminopeptidase
LB: Luria-Bertani
IPTG: isopropyl β-D-1-thiogalactopyranoside
IMAC: Immobilized Metal Ion Affinity Chromatography
PMSF: phenylmethanesulfonyl fluoride
PEG: polyethylene glycol
EDTA: ethylenediaminetetraacetic acid
Ig: immunoglobulin
SHM: somatic hypermutation
CSR: class switch recombination
dsDNA: double strand DNA
ssDNA: single strand DNA
GST: Glutathione-S-transferase
TB: Terrific Broth
DTT: dithiothreitol
UDCA: ursodeoxycholic acid
CA: cholic acid
CDCA: chenodeoxycholic acid
CYP8B1: Sterol 12α-hydroxylase
CYP27A1: Sterol 27-hydroxylase
FXR: Farnesoid nuclear receptor
RXR: Retinoid X receptor
LXR: Liver X receptor

1. Heat shock protein 60

1.1 Introduction

1.1.1 Heat Shock Proteins

Many proteins can fold spontaneously in vitro to their native state under appropriate conditions. Nevertheless, inside a cell, during the protein folding process, the highly crowded macromolecular environment favours protein misfolding and aggregation; aggregation is a disordered association of polypeptide chains devoid of any biological activity.
A group of proteins, known as molecular chaperones, counteracts in vivo the tendency of non-native proteins to misfold or aggregate. These proteins are essential for protein folding to occur with high efficiency.

Heat Shock Proteins (HSP) act as molecular chaperones during both de novo protein synthesis and under conditions of cellular stress that may lead to unfolding and aggregation of preexisting proteins. Under physiological conditions, HSP are involved in mediating protein folding, protein assembly into oligomeric structures, protein transport, refolding of misfolded proteins and proteosome degradation. In addition, HSP are implicated in gene expression regulation, DNA replication, signal transduction, cell differentiation, apoptosis, cellular senescence and carcinogenesis [1].
Moreover, HSP can be induced in response to cellular stress such as changes in temperature, the presence of free oxygen radicals, heavy metals, ethanol, amino acid analogs, viral and bacterial infections, and ischaemia; this process is referred to as heat shock response [2]. While expression of some HSP is strictly inducible by stress, other HSP proteins are constitutively expressed at normal temperature and are only slightly induced by heat shock.
HSP are present in all cellular compartments (cytosol, endoplasmic reticulum, mitochondria and nucleus) and they are generally considered as long half-life proteins.

According to their molecular weight measured in kDa, HSP can be divided into different families: small HSP, HSP60, HSP70 and HSP90 [3]. HSP from the same family often show a significant amount of sequence homology and are structurally and

functionally correlated, while no such homology has been found between chaperones from different families.

The small heat shock proteins (sHSP) have a molecular weight ranging from 25 to 40 kDa; they are ATP-dependent and are thought to form oligomers through their C-terminal core domain. One sHSP complex can bind several substrate proteins; the subsequent refolding is performed by a different chaperone [3].

The members of the HSP60 family, also called chaperonins, are highly expressed in a constitutive manner and are only moderately induced by stress.
They act by forming a double ring complex with a central cavity enclosed in each ring.
Chaperonines can be divided into two different groups, which share a similar architecture but are distantly related in primary sequence. Group I has been found in bacteria, mitochondria and chloroplasts, while group II is present in Archea and in the cytosol of eukaryotes. Members of group I require HSP10 as a cofactor, which was suggested to serve as cavity lid, instead Group II chaperonins do not require HSP10.
The best characterized group I chaperonine is GroEL from *Escherichia coli*; its cofactor is GroES. The GroEL complex is constituted by two rings, each of them formed by seven GroEL molecules.

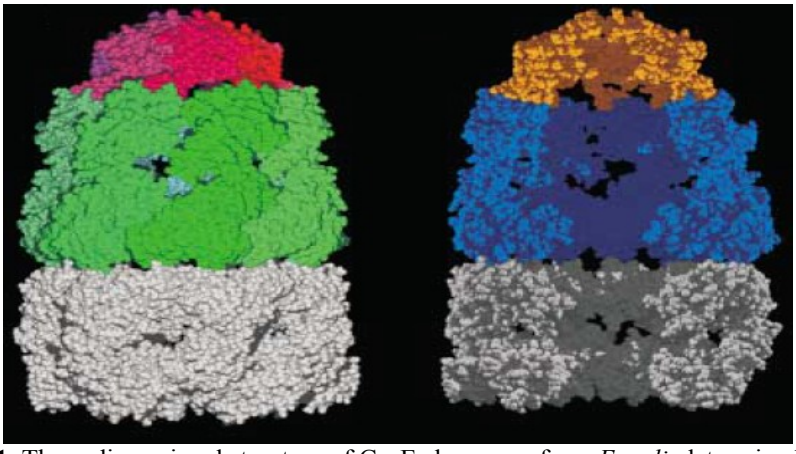


Figure 1.1: Three-dimensional structure of GroE chaperone from *E. coli*, determined by X-ray diffraction. Side view of GroE complex (left); distal GroEL ring is shown in grey, the seven subunits of the proximal GroEL are shown in shades of green, GroES is bound to the proximal ring and is shown in red. Cross-section of the GroE complex (right) [4].

Each subunit presents an equatorial ATP-binding domain, a hinge domain and an apical

substrate-binding domain. After binding of the substrate to one of the two rings, seven ATP molecules and GroES bind to that ring, thus determining the encapsulation of the substrate. As a result of GroES binding, GroEL undergoes a conformational change, with an enlargement of the cavity, providing an adequate environment for the substrate protein to fold without aggregating. ATP hydrolysis and ATP binding to the other ring determine GroES and protein substrate release. Many folding cycles are thought to be required for the substrate correct folding.

The HSP70 family represents the most highly conserved molecular chaperones and its members have been found in bacteria, Archea, eukaryotic cytosol, eukaryotic nucleus, mitochondria and chloroplasts.
The expression of some members of this family is strongly induced under stress conditions.
HSP70 are involved in both newly synthesized protein folding and in prevention of protein aggregation in several post-translational processes, such as protein degradation, protein targeting and membrane translocation. HSP70 chaperones act through an ATP-regulated cycle of substrate binding and release. The members of this family require cooperation with different cofactors to modulate their ATPase activity and for their correct targeting to specific cellular structures [3].

The HSP90 family plays an important role in conformational protein regulation and cell signaling. These chaperones are thought to form homodimers.
Substrate loading was suggested to be mediated by the protein p60/Hop, which interacts with both HSP70 and HSP90; after loading, HSP90 undergoes regulated cycles of ATP binding and hydrolysis and eventually releases the native protein.
Under stress conditions, this chaperone acts as a general storage compartment for unfolded proteins, which can subsequently be refolded by a different chaperone.
HSP90 is associated with a multi-chaperone machinery, which, depending on the substrate protein bound, includes several cofactors; some of these cofactors present enzymatic activity, such as peptidylprolyl isomerases and serine/threonine phosphatase 5. Other cofactors are located at specific cellular structures, where they act as docking partners for the chaperone; for example, one of these cofactors, Tom70, is located in the

outer mitochondrial membrane [3].

1.1.2 Human HSP60

The human chaperonine HSP60 is a mitochondrial protein; the HSP60 precursor has an N-terminal mitochondrial targeting sequence, which is cleaved during its import into the organelle [5]. But this protein has also been detected in other loci, such as cytosol, cell surface, extracellular space and circulation. HSP60 was detected on the surface of normal and tumor cells. An increased level of the protein on the cell surface is thought to be a danger signal for the immune system cells, able to induce activation and maturation of dendritic cells and generation of an antitumor T-cell response. HSP60 located in extracellular space is thought to be able to interact with several cell surface receptors, such as CD14, CD40 and Toll-like receptors; moreover, HSP60 can cause pro- and anti-inflammatory effects and can bind peptides, for example tumor derived peptide antigens, and present them to immune system cells. Additionally, HSP60 is present in the blood stream and can be detected in plasma of healthy subjects [6].

Innate immune system stimulation by HSP60

Besides being involved in protein folding, HSP60 exhibits immunoregulatory properties, i. e. it is thought to be a potent regulator of innate immune cell activity able to induce pro-inflammatory processes.

HSP60 and also HSP70 are present in detectable amounts in blood of healthy individuals; moreover, HSP serum levels were observed to be highly increased under several pathological conditions, such as hypertension, atherosclerosis and renal vascular diseases.

HSP60 is a potent activator of cells of the innate immune system, such as macrophages, dendritic cells and endothelial cells. Dendritic cells incubated with HSP60 show a significant raise in release of tumor necrosis factor α, interleukin-6 and chemokines RANTES and MIP-2. In innate immune cells, HSP60 stimulates the formation and the release of the pro-inflammatory mediators IL-1β, MCP-1, MIP-1α and the short-lived radical nitric oxide. In addition, production of the T helper 1 cell promoting cytokines IL-12 and IL-15 is increased [7].

Different HSP60 epitopes are responsible for binding and activation of innate immune cells. Using the macrophage line J774a.1, an epitope in the C-terminal region (481-500 aa) was identified [8]. With a different macrophage cell line, three completely different epitopes of HSP60 (241-260 aa, 391-410 aa and 461-480 aa) were found [9]. Therefore it was suggested that different regions of HSP60 are responsible for the interaction with different innate immune cell types.

Besides, a completely different epitope (354-365 aa) was proved to be responsible for the binding of HSP60 to LPS (lipopolysaccharide), a major component of the outer membrane of Gram-negative bacteria, which can strongly elicit immune response; the central motif of this epitope (LKGK) is thought to be crucial for the binding, as it is present also in factor C from *Limulus polyphemus*, another protein able to bind efficiently LPS. Interestingly, this motif is conserved in all mammalian HSP60, but not in prokaryotic ones [7].

Deficiency of HSP60

Since HSP60 is essential for cell survival, there are only few instances of diseases caused by lack of this protein.

Hereditary spastic paraplegia SPG13 is a genetic disorder caused by the missense mutation V72I; the mutated residue is a conserved valine located in the equatorial domain of HSP60.

This pathology is characterized by progressive weakness and spasticity of the lower limbs, caused by degeneration of the distal ends of long axons in the spinal cord. It is not clear why the cells affected by HSP60 dysfunction are the motoneurons of the lower limbs; one possible explanation is that the mutation affects the correct folding of a limited set of protein substrates which play an important role in this cellular type.

Since this is an autosomal dominant pathology, it was hypothesized that, when mutated HSP60 subunits are incorporated into the two-ring complex, this complex is not fully active [3, 10].

A patient was found affected by a systemic mitochondrial encephalomyopathy, which was caused by multiple deficiency in mitochondrial enzymes; the lower level of HSP60 detected in the patient was hypothesized to be the cause of the disease [11].

In a patient with congenital lactic acidaemia, a substantially decreased amount of HSP60 in cultured fibroblasts was reported; the disease was caused by insufficiency of the mitochondrial enzymes involved in cellular respiration, probably due to the low level of HSP60 [12].

Involvement of HSP60 in autoimmune diseases

The amino acid sequence of the HSP60 family is highly conserved throughout evolution. As a consequence, human and bacterial proteins share a sequence homology of about 50%; for example, human and *E. coli* HSP60 proteins share 52% sequence identity and 67% sequence similarity. It was suggested that immune response elicited in the human body against bacterial HSP60 may result in crossreactivity to the human HSP60, thus determining an autoimmune reaction [13].

For example, autoimmunity to HSP60 may be involved in some cases of infertility and embryo loss, caused by damage or occlusion of the fallopian tubes, resulting after chlamydial infections. Human cells chronically infected by *Chlamydia trachomatis*, an obligate intracellular parasite, synthesize high levels of chlamydial HSP60; indeed, during an infection, this pathogen greatly enhances HSP synthesis as a protective response against stressful environmental conditions, such as nutrient deprivation, changes in pH and exposure to oxygen radicals.

Human HSP60 is highly expressed during the pre-implantation and implantation stages of pregnancy by the maternal decidua and the embryo. Lymphocytes previously sensitized to chlamidyal HSP60 could be reactivated and release inflammatory and cytotoxyc mediators. As a result, immune regulatory mechanisms required for implantation and maintenance of the embryo may be perturbed, with deleterious consequences to embryo survival [2, 13].

Moreover, cross-reactions of antibodies and T cells against HSP60 of microbes and human beings may play a role in the development of atheroscerosis.

This pathology is a multifactorial disease, characterized by accumulation of lipids in the blood vessel wall, mononuclear cell infiltration and smooth muscle cell proliferation.

Inflammatory and immune mechanisms were suggested to be one of the first processes in the development of atherosclerosis, because the early stages of the pathology are characterized by a relative paucity of lipids and an abundance of inflammatory cells, such as activated T lymphocytes, mast cells and macrophages, and immunoglobulins.

Most of known risk factors for this disease, such as oxidized low-density lipoprotein, hypertension, infections and oxidative stress, determine an increase in HSP60 expression levels in endothelial cells, smooth muscle cells and macrophages, which are the main cellular constituents of atherosclerotic plaques.

Within atherosclerotic plaques, the presence of T lymphocytes responding to human HSP60 has been reported; moreover, the level of autoantibodies directed against HSP60 is increased in patients with atherosclerosis.

It was hypothesized that the immune reaction directed against microbial HSP60, owing to the high sequence identity of this protein across species, may be misdirected towards human HSP60 which is expressed in the stressed cells of the blood vessels, thus contributing to initiation and propagation of atherosclerosis [14].

Prion diseases and HSP60

Certain neurodegenerative diseases, including bovine spongiform encephalopathy in cattle and Creutzfeldt-Jakob disease in humans, are mediated by prions. The prion particle consists of PrP^{Sc}, a posttranslationally modified isoform of the host – encoded prion protein (PrP).

The interaction between human HSP60 and PrP was reported in a yeast two hybrid screen [15]. Additionally, it was shown that *E. coli* HSP60 homologous can convert PrP to the pathogenic isoform PrP^{Sc} in vitro [16]. Nevertheless, all these results were obtained only in vitro and HSP60 and PrP are not known to occur in the same cellular compartment.

Involvement of HSP60 in apoptosis and cancer

HSP60 was suggested to be involved in processes of apoptosis and tumorigenesis.

The pro-caspase-3, whose activation is regarded as a central event in the execution phase of apoptosis, was found to be present in the mitochondrial fraction of Jurkat T cells in a complex with HSP60 and HSP10. Induction of apoptosis with staurosporine

caused the activation of the pro-caspase-3 and its dissociation from the HSP60-HSP10 complex; moreover, the simultaneous release of HSP from mitochondria was observed, along with the release of other mitochondrial transmembrane space proteins, such as cytochrome c. Furthermore, HSP60 and HSP10 were found to associate with pro-caspase-3 and to accelerate its activation in vitro. Accordingly, it was suggested that the release of HSP60 from mitochondria may accelerate caspase activation in the cytoplasm of intact cells and thus HSP60 may play a pro-apoptotic role [17].

On the other hand, an anti-apoptotic role of cytosolic HSP60 was proposed in cardiac myocytes, a cell type that has abundant mitochondria; HSP60 was shown to interact with the pro-apoptotic Bax and Bak proteins, preventing the apoptotic process. A reduction in HSP60 level, obtained by the antisense technique, caused apoptosis and was associated with an increase in Bax protein and a decrease in BCL-2 protein. Additionally, it was demonstrated that cytosolic HSP60 forms a macromolecular complex with Bax and Bak in vitro; it was suggested that the complex formation with HSP60 may block the ability of Bax and Bak to effect apoptosis in vivo. Finally, a reduction in cytosolic HSP60 was associated with an increase in the number of Bax protein present in the mitochondrial fraction. Since Bax alone is sufficient to induce cytochrome c release and consequently apoptosis, the interaction between HSP60 and Bax was suggested to play a critical role in preventing apoptosis [1, 18].

The presence and levels of HSP60 have been examined in a variety of tumors from different tissues; the results obtained indicate that detection and quantification of HSP60 may be useful for histopathological diagnosis and for prognosis [6].

For example, in an immunohistochemical study of bladder carcinoma, HSP60 was found at low levels or absent in tumor specimens, while the protein was expressed in normal urothelium. The authors suggested that HSP60 may be a very useful marker for patients with superficial bladder carcinoma; in particular, low HSP60 levels are correlated with unfavourable prognosis after local treatment of superficial bladder carcinoma [19].

While HSP60 was scarce or absent in normal prostatic tissue, increased levels of the protein were reported in prostatic carcinogenesis; accordingly, HSP60 was suggested to be a marker for prostate malignancy [20].

1.2 Materials and Methods

1.2.1 The human HSP60 clone

The plasmid containing the full-length cDNA clone (IMAGE ID 2821157) of human HSP60 was purchased from RZPD (Deutsches Ressourcenzentrum fur Genomforschung). *E. coli* cells harbouring the plasmid vector pOTB7 were grown in 3 ml of Luria-Bertani (LB) medium containing 50 µg/ml chloramphenicol overnight at 37ºC.

Cells were harvested by centrifugation and the plasmid vector was purified using the GenElute™ Plasmid Miniprep kit (Sigma-Aldrich®).

1.2.2 Subcloning of HSP60 gene into pET15b plasmid

The coding sequence of the protein without the N-terminal 27 amino acids, corresponding to the mitochondrial targeting signal, was amplified by a PCR reaction using the following primers and parameters.

Primer forward:

5'-GGTGGTCATATGGCCAAAGATGTAAAATTTGGT-3'

Primer reverse:

5'-GGTGGTGGATCCTTAGAACATGCCACCTCCCAT-3'

CATATG - *NdeI* restriction site

GGATCC- *BamHI* restriction site

TTA - stop codon

Figure 1.2: The primers used to amplify hHSP60; the coding sequence is underlined.

STEP	TEMPERATURE	TIME	
Initial Denaturation	95ºC	10 min	
Denaturation	95ºC	1 min	35 cycles
Annealing	55ºC	1 min	
Extension	72ºC	1 min 40 sec	
Final Extension	72ºC	10 min	

Figure 1.3: The PCR parameters.

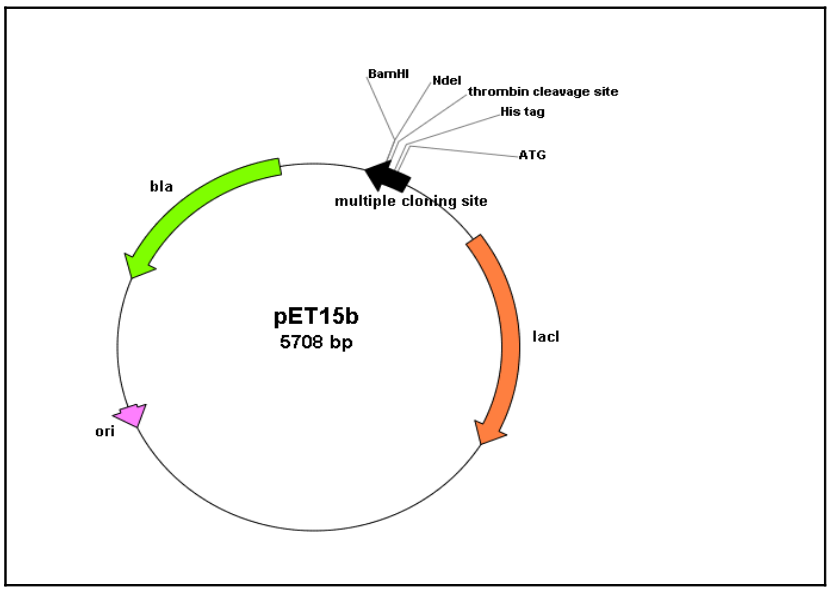


Figure 1.4: Schematic representation of pET15b.

The amplified coding sequence of the gene was quantified on a 0.8% agarose TBE gel and digested with *NdeI* and *BamHI* restriction enzymes; afterwards, the fragment was ligated into the *NdeI*/*BamHI* site of the pET15b plasmid (Novagen®), thus yielding HSP60-pET15b. The chosen expression vector allows the N-terminal portion of the protein to be fused to a 6-His tag, which is useful for protein purification and detection by Western blot; moreover, the plasmid contains a thrombin cleavage site to facilitate removal of the tag after purification.

The resulting recombinant plasmid was used to transform *E. coli* XL1 Blue competent cells by heat shock. Cells were plated on a LB-agar plate containing 100 µg/ml ampicillin and grown overnight at 37ºC. Colonies carrying the HPS60-pET15b plasmid were selected by colony PCR and subsequent digestion of the purified plasmid; finally, the vector was checked by automated DNA sequencing.

1.2.3 Expression screening of the recombinant protein

The recombinant plasmid was used to transform the *E. coli* BL21 (DE3) strain by heat shock; the resulting colonies were tested for protein expression. To test for protein expression, a single colony was inoculated in 3 ml of LB medium, grown at 37ºC adding 100 µg/ml ampicillin and then incubated at 37ºC for 4 hours and at 20ºC and 28ºC overnight after induction with IPTG (isopropyl β-D-1-thiogalactopyranoside) to a final concentration of 0.25 mM.

Cells were harvested by centrifugation at 9,000 rpm for 10 minutes at 4ºC and the pellet was resuspended in 2 ml of buffer A (20 mM TRIS-HCl pH 7.5, 0.5 M NaCl and 0.02% NaN_3). Cells were disrupted by sonication and then centrifuged at 9,000 rpm for 10 minutes at 4ºC. Samples from the total lysate and from the soluble fraction were loaded onto a SDS-PAGE gel (12% polyacrylamide resolving gel and 4% polyacrylamide stacking gel). Protein expression was evaluated by Western blot analysis using an anti-histidine tag monoclonal antibody (Sigma-Aldrich®).

The protein was present in all the samples, in both the total fraction and the soluble fraction; the best soluble protein yield was obtained with an induction at 37ºC for 4 hours.

1.2.4 Purification and crystallization trials

Immobilized Metal Ion Affinity Chromatography (IMAC)

E. coli BL21 (DE3) cells containing the plasmid HSP60-pET15b were inoculated in 10 ml of LB medium containing 100 µg/ml ampicillin and shaken overnight at 37ºC. The 10 ml overnight culture was transferred into 1 liter of LB medium; the culture was grown at 37ºC under shaking until an OD_{600} of 0.8-1.0 was reached. Protein expression was induced by adding IPTG to a final concentration of 0.25 mM and the culture was further shaken at 37ºC for 4 hours. Cells were recovered by centrifugation at 8,000 rpm for 10 minutes at 4ºC and the pellet was resuspended in 50 ml of buffer A plus 10 mM imidazole. The sample was stored at -20ºC.

The sample was thawed and 0.1 mM PMSF (phenylmethanesulfonyl fluoride) was added as protease inhibitor; then cell disruption was carried out by sonication and the

sample was centrifuged at 9,000 rpm for 10 minutes at 4ºC to remove cell debris.
The soluble fraction was directly loaded onto a Nickel Sepharose column (5 ml prepacked column, HisTrap™ FF, Amersham Biosciences) preequilibrated with buffer A plus 10 mM imidazole. After loading, the column was further washed with buffer A and 10 mM imidazole until the U.V. absorption at 280 nm returned to the baseline.
A gradient was applied to the column ranging from 10 mM to 500 mM imidazole. The fractions containing the target protein were collected and loaded onto a SDS-PAGE gel.

Size exclusion chromatography (SEC)
The protein was dialyzed against buffer B (20 mM TRIS-HCl pH 7.5, 0.15 M NaCl and 0.02% NaN_3) and 2.5 mM $CaCl_2$.
Subsequently, the protein was concentrated to 1.0 mg/ml in an Amicon ultrafiltration cell (Millipore™) and 15 units of thrombin were added; digestion was performed overnight at 20ºC.
The protein was concentrated to 2 ml in an Amicon cell and loaded onto a Superdex G200 gel filtration column, preequilibrated in buffer B. Gel filtration chromatography was performed to further purify the target protein and to remove thrombin and the 6-His tag. The effectiveness of thrombin cleavage was evaluated by Western blot.
The fractions containing the target protein were collected and loaded onto a SDS-PAGE gel.

Ion Exchange Chromatography (IEC)
A further purification step was performed with an anion exchange chromatography.
After Nickel Sepharose affinity chromatography, the protein was dialyzed against buffer C (20 mM TRIS-HCl pH 8.0 and 0.02% NaN_3). The sample was loaded onto an anion exchange column (Hi Trap™ DEAE FF, Amersham Biosciences). After loading, the column was further washed with buffer C until the U.V. absorption at 280 nm returned to the baseline. Elution was achieved by a step gradient increasing the salt concentration, from 0 to 0.5 M NaCl. The flow-through and the elution fractions were analyzed by SDS-PAGE.

Crystallization trials

The protein sample was dialyzed against buffer D (20 mM TRIS-HCl pH 7.5 and 0.02% NaN_3). The protein was concentrated to 15-20 mg/ml by ultrafiltration (Amicon and Centricon centrifugal filter devices, Millipore™) and crystallization trials were set up using the hanging drop vapor diffusion technique using 24-well plates (MULTIWELL, Falcon™). Drops were formed by mixing 1 µl of protein solution with 1 µl of precipitant solution. To screen 96 different crystallization solutions, a commercial kit was used (Structure Screens 1 and 2, Molecular Dimensions Limited).

1.3 Crystallization

1.3.1 Crystallization theory

Obtaining a crystallin form of the desired protein represents a crucial and often the most difficult step in order to determine the three-dimensional structure of a protein by X-ray diffraction.

Macromolecules are very complex systems, characterized by several physico-chemical parameters, that can change as a function of temperature, pH, ionic strength, solvent composition and other environmental variables; for this reason, the crystallization process is necessarily empirical.

As a general rule, the reduction of the solubility of a protein solution results in the formation of an amorphous precipitate; however, when appropriate conditions are selected, protein crystallization can occur.

The phase diagram represents the solubility of a protein as a function of the precipitant concentration.

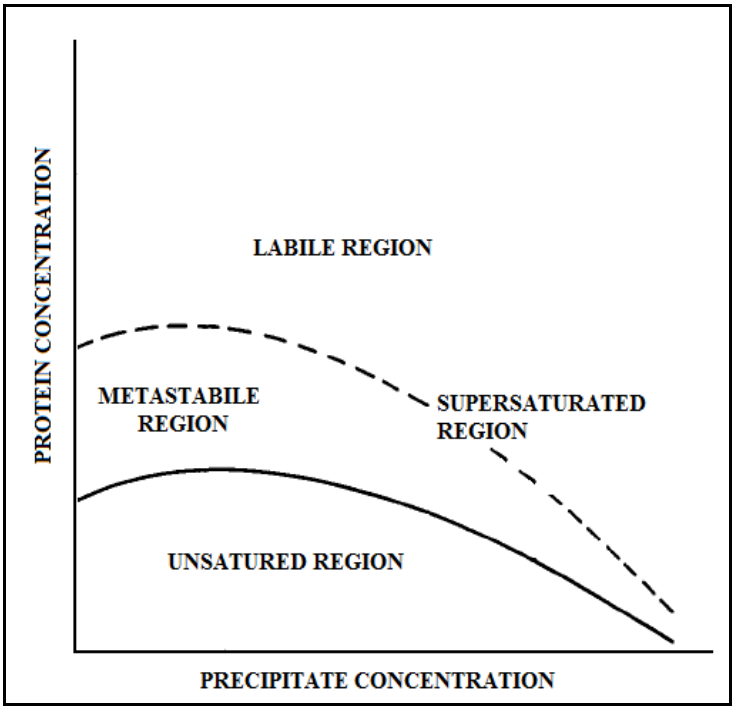


Figure 1.5: A phase diagram for a hypothetical protein.

In a saturated solution a solid phase exists in equilibrium with molecules in solution, i. e. crystals do not grow, since any increase in the proportion of solid phase is counterbalanced by an equivalent dissolution. Instead, crystal growth may occur in a supersaturated solution, in which a solid and a liquid phase are not at equilibrium and in which the solute moves from the liquid to the solid phase.
The supersaturated region can be divided into a labile region, where crystals can form and grow, and a metastable region, where crystals can only grow. If the system is pushed deeply into the labile region, the crystals obtained will be numerous, but small and afflicted with growth defects.
In order to obtain a few and big crystals, crystal growth should begin in the labile region just beyond the metastable region; in this way only a few crystals will form in the labile region and then, by depletion of the protein in solution, the system will return in the metastable region, where the few crystals previously formed will continue to growth.
Proteins in a crystal are completely solvated; indeed, protein crystals present a high water content, ranging from 30% to 80%.

Commonly used precipitants can be divided into: inorganic salts, organic solvents and long-chain polymers.
Inorganic salts are charged and consequently they can decrease the solubility of proteins by competition for water molecules. They also change the ionic strength of the solution. The ionic strength (μ) depends both on the concentration of the salt (c) and on its charge (z), according to the following relation:

$$\mu = 1 / 2 \Sigma c_i * z_i^2$$

At low salt concentration, protein solubility increases ("salting in" effect) as salt concentration increases, because interactions between proteins and water molecules become more favourable. Instead at high salt concentration, protein solubility decreases ("salting out" effect) as salt concentration increases, because protein and salt have to compete to interact with water molecules; the "salting out" effect can be exploited to create conditions suitable for protein crystallization.
Since ionic strength depends on the square of the charge, bivalent and trivalent ions are

preferable; another aspect to be taken into account is the solubility of the salt in water, which should be high enough to avoid salt precipitation.
Commonly used inorganic salts are sulfates, phosphates and citrates.

Organic solvents reduce protein solubility by lowering the dielectric constant of the solvent. Commonly used organic solvents are: ethanol, isopropanol, dioxane, DMSO (dimethyl sulfoxide) and MPD (2-methyl-2,4-pentanediol).

Polyethylene glycols (PEGs) are linear polymers produced in various lengths, containing from several to many hundred monomers. Like organic solvents, they lower the dielectric constant of the solvent; moreover, they show volume exclusion properties, forming a reticulum with water molecules, from which proteins are excluded.
All PEG sizes from M_r 400 to 20,000 have provided protein crystals, but the most commonly used are those in the range from 2,000 to 6,000; commonly PEGs are used in concentrations from 5% to 30%.

Other factors that could affect protein crystal growth are:

- homogeneity and purity of the protein solution. The presence of contaminants or even different isoforms of the same protein can hinder formation and reproducibility of protein crystallization; these contaminations may be revealed by SDS-PAGE or IEF (isoelectrofocusing) analysis;
- protein concentration can vary from 5 mg/ml to 50 mg/ml;
- pH; it is a very important parameter in protein crystal formation. The optimum pH value is searched taking into account the protein stability at various pHs and that protein solubility reaches a minimum at the protein isoelectric point; also the buffer chosen to maintain the pH constant can affect the crystallization process;
- temperature; crystallization trials are kept at constant temperature, usually 4°C or 20°C. Crystallization occurs slowlier at colder than at warmer temperatures;
- time; it is quite variable and can range from a few hours to several months;
- presence of substrates, inhibitors, coenzymes and ligands. Sometimes the addition of these compounds can favour crystallization; their effect is due to

fixing an enzyme or a protein in a more compact and stable form;

- metal ions.

In addition, microheterogeneity of the protein solution may prevent crystal formation; common sources of microheterogeneity are the following:

- presence, absence or variation in a bound prosthetic group, substrate, coenzyme, inhibitor or metal ion;
- variation in length or composition of the carbohydrate moiety of a glycoprotein;
- proteolytic modification of the protein during the course of isolation or crystallization;
- oxidation of sulfhydryl groups during isolation;
- reaction with heavy metal ions during isolation or storage;
- presence, absence or variation in post-translational side chain modifications such as methylation, amidation and phosphorylation;
- microheterogeneity in the amino or carboxy terminus or modification of termini;
- variation in the aggregation or oligomeric state of the protein due to association/dissociation;
- conformational instability due to the dynamic nature of the molecule;
- microheterogeneity due to the contribution of multiple but nonidentical genes to the coding of the protein, isozymes;
- partial denaturation of the sample;
- genetically different animals, plants or microorganisms that make up the source of protein preparations;
- bound lipids, nucleic acids or carbohydrates, or substances such as detergents used in the isolation [21].

1.3.2 Crystallization techniques

The most commonly used technique to reach the supersaturation state is the vapor diffusion method. This method is based on the transport of either water or some volatile agents between a microdrop of protein solution and a much larger precipitant solution (reservoir solution); the microdrop and the reservoir solution come to equilibrium through the vapor phase and the final equilibration conditions are those of the reservoir solution, due to its much larger volume.
The protein and the precipitant solution are usually mixed in a 1:1 ratio; reagents in the reservoir solution are more concentrated than in the drop and consequently water gradually evaporates from the microdrop. When a supersaturated state is reached in the drop, protein crystals may form.

The vapor diffusion method can be divided into the hanging drop and the sitting drop techniques.
In the first method a drop containing a mixture of protein and precipitant solution of a size ranging from 100 nl to 5 µl is dispensed on a microscope cover glass; the drop is then sealed with grease or mineral oil into a well containing the reservoir solution (usually 300 µl).

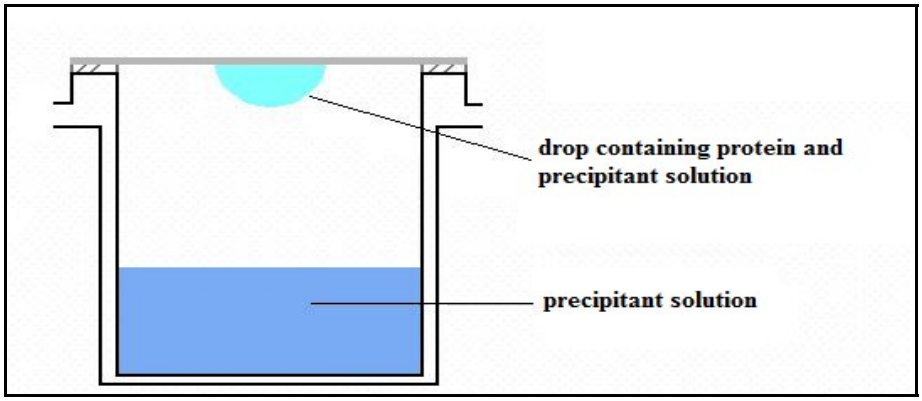


Figure 1.6: The hanging drop technique.

This method requires only small amounts of material and, as a consequence, it is ideal for screening a large number of conditions.

In the sitting drop method, the drop containing the protein and the precipitant solution mixed is dispensed on a microbridge, which is placed inside a well containing the reservoir solution; the well is then closed with a microscope cover glass and sealed by grease or mineral oil.

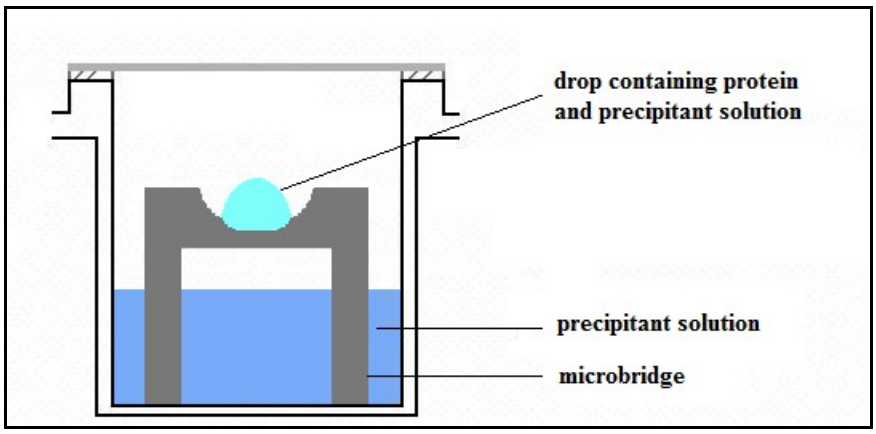


Figure 1.7: The sitting drop technique.

The sitting drop technique requires larger volumes of protein solution than the hanging drop technique, ranging from 5 to 40 µl, and it is usually used to improve crystallization conditions previously identified by the hanging drop technique.

Another crystallization method is the dialysis technique; it is based on diffusion and equilibration of small precipitant molecules through a semipermeable membrane as a means of slowly approaching the supersaturated state.

The protein solution is contained inside a dialysis cell with a semi-permeable membrane, that allows the passage of small molecules and ions and not of the protein molecules through its pores. The dialysis cell containing the protein solution is submerged into a much larger volume of precipitant solution. Following equilibration against the precipitant solution, the protein solution within the dialysis cell can reach supersaturation and thus eventually protein crystals may appear.

The main disadvantage of this technique is the relatively large amount of protein solution required (10-100 µl in each trial). An important advantage of the dialysis method is that the precipitant solution can be easily varied, simply by moving the dialysis tube from one solution to a different one.

1.3.3 Solution of the phase problem

X-ray diffraction is used to determine the 3D structure of a target protein. The electron density distribution within the unit cell can be described as the inverse 3D Fourier transform of the structure factors; structure factors are defined as the ratio of the radiation scattered by any real sample to that scattered by a single electron placed at the origin.

The structure factor, F(h,k,l), is a complex number describing amplitude ($|F|$) and phases ($e^{i\varphi}$) of a wave diffracted from crystal lattice planes characterised by Miller indices h, k, l. Both amplitude and phases must be known in order to calculate the electron density map. In X-ray data the intensity of each reflection is proportional to the square of the structure factor amplitude, but all phase information is systematically lost.
Commonly used ways to recover the phases are Molecular Replacement, Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction.

Molecular Replacement (MR)

The unknown phases of the target protein are obtained using phases from a known model. It is accepted that molecular replacement will be straightforward if the protein sequence identity of the two proteins is greater than 40%; it becomes progressively more difficult as the model shares less sequence identity.

To place the model structure in the correct orientation and position in a different unit cell, three rotational angles and three translational parameters are required; therefore, the molecular replacement is a 6-dimensional problem. However it is possible to separate this into two 3-dimensional problems. First, a rotation function is computed to find the three rotational angles and thus to orient the known structure within the unit cell; after that, a translation function is used to place the oriented model in the cell. The phases obtained from molecular replacement can be used together with the experimental intensities to calculate the electron density map.

Multiple Isomorphous Replacement (MIR)

MIR is based on the introduction of a small number of heavy atoms (atoms containing many electrons) into the protein crystal. The native and the heavy atom derivative

crystal should be isomorphus, that is crystal form and unit cell dimensions should be unchanged. Derivatives are usually prepared by soaking the crystal in a solution containing a heavy atom. The data from a native crystal and a derivate are compared and thus the differences in scattered intensities are measured; by using this information, a Patterson difference map can be computed and heavy atom coordinates in the unit cell can be calculated. Two derivate crystals are sufficient to determine the phases in theory; however, due to errors in the experimental results and imperfect isomorphism, more than two derivates are often required.

Multiwavelength Anamalous Diffraction (MAD)

The MAD technique exploits the ability of the inner electrons of atoms to absorb X-rays at particular wavelengths. Due to the anomalous dispersion effect, the X-rays reemitted from these scatterers have a small shift in both amplitude and phases. According to Friedel's law, structure factors related by opposite Miller indices, called Friedel mates, show equal reflection intensities:

$$\left| F(h,k,l) \right| = \left| F(-h,-k,-l) \right|$$

Indeed the resonant frequencies of most elements in biological molecules are far below the energies used for diffraction experiments. However, elements of atomic number from 24 to 92 have resonance frequencies consistent with X-ray diffraction experiments. If the frequencies of the X-rays fall near an absorption frequency of an atom, Friedel's law does not hold and the differences between Friedel's mates in scattered intensities can be used to calculate the position of the anomalous scatterer in the unit cell by Patterson methods.

MAD data collection is performed at two or more different wavelengths near the absorption edge of the scatterer. A commonly used atom for MAD technique is selenium, by replacing the natural sulphur containing amino acid methionine with selenomethionine.

1.4 Results and discussion

The human HSP60 gene was inserted into the pET15b plasmid and used to transform *E. coli* BL21 (DE3) strain cells.

According to the Western blot analysis, the protein was expressed in soluble form in appreciable quantity in each of the three induction conditions tested (37ºC for 4 hours, 20ºC overnight and 28ºC overnight).

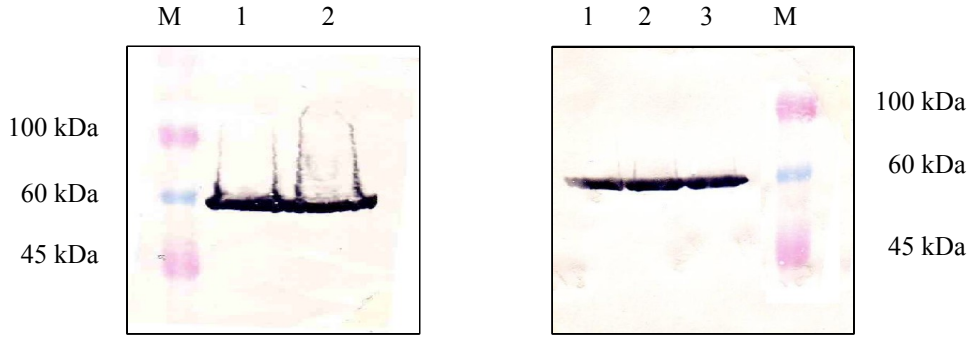


Figure 1.8: Western blot analysis of HSP60-pET15 *E. coli* BL21. The total fractions (left) are referred to induction at 20ºC overnight (lane 1) and 28ºC overnight (lane 2), while the soluble fractions (right) to induction at 37ºC for 4 hours (lane 1), 28ºC overnight (lane 2) and 20ºC overnight (lane 3).

When 1 liter of culture was grown, the induction was best carried out at 37ºC for 4 hours, with a yield of 15-20 mg of protein after IMAC chromatography. The elution pattern from the IMAC column showed two peaks; the protein was present in the second, corresponding to an imidazole concentration of about 200 mM. The protein showed a good purity grade, but further purification steps were still required.

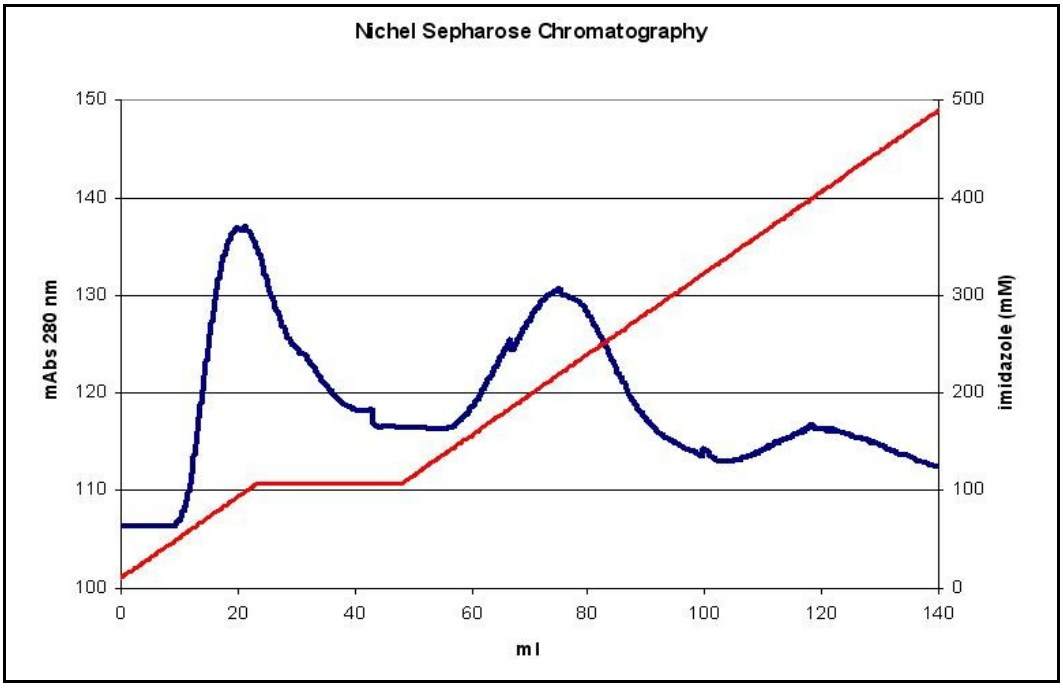


Figure 1.9: Elution profile of HSP60 from Nickel Sepharose affinity column.
The blue line corresponds to the UV absorbance and the red line to the imidazole gradient.

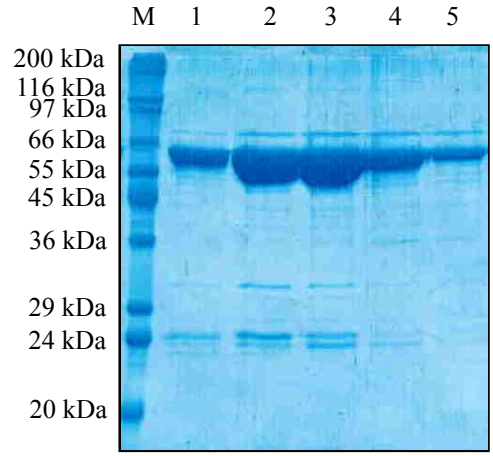


Figure 1.10: SDS-PAGE analysis of HSP60 obtained after Nickel Sepharose affinity chromatography.
The target protein is present in the second peak with a good purity grade.
Lanes 1 to 5: fractions corresponding to the second peak.

Proteolitic removal of the 6-His tag was performed by dialysis of the protein against a suitable buffer and the subsequent addition of thrombin; complete tag removal was confirmed by the disappearance of the band from a Western blot with an anti-histidine tag monoclonal antibody.

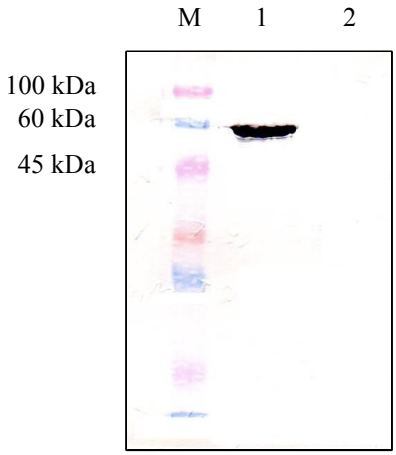


Figure 1.11: HSP60 before and after His-tag proteolytic removal (lane 1 and lane 2, respectively).

Then the protein was further purified by size exclusion chromatography. The protein was present in the main peak of the chromatogram.

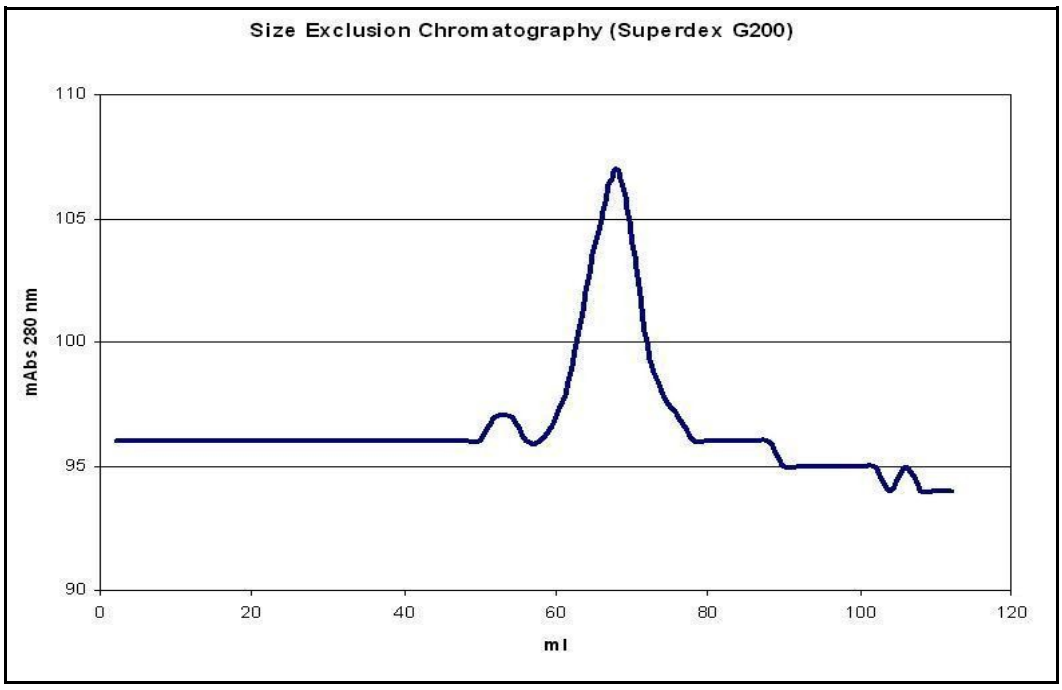


Figure 1.12: Elution profile of HSP60 from Superdex G200 size exclusion column. The blue line corresponds to the UV absorbance.

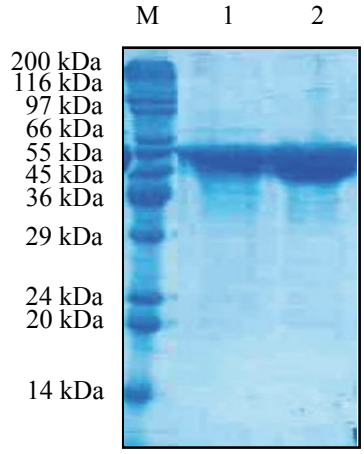


Figure 1.13: SDS-PAGE analysis of HSP60 obtained after size exclusion chromatography. Lane 1 and 2: fractions corresponding to the main peak.

The protein was concentrated to 20 mg/ml and crystallization trials were set up using the hanging drop vapor diffusion technique. None of the cystallization trials gave crystals suitable for X-ray data collection. This is probably due to the heterogeneity of the purified protein.

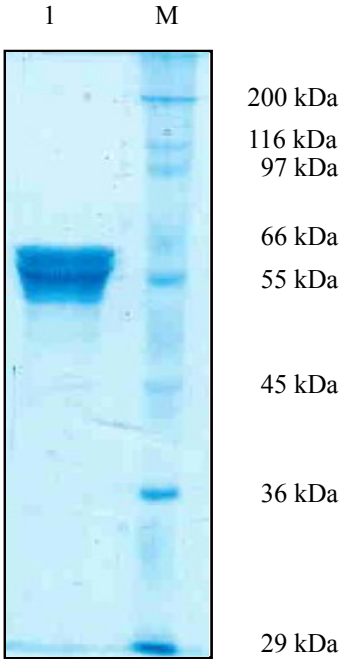


Figure 1.14: SDS-PAGE (10% resolving gel) analysis of HSP60 used in crystallization trials. Lane 1: 1 µl of HSP60 after concentration was loaded.

The heterogeneity of the sample may be explained with protein degradation due to the

presence of contaminant proteases; accordingly, the entire purification protocol was repeated adding a protease inhibitor cocktail (Complete, EDTA-free, Roche) to the cells just before sonication and adding EDTA (ethylenediaminetetraacetic acid) 5 mM in all the buffers used after Nickel Sepharose elution. Nevertheless, no appreciable improvement was achieved in the final purified protein; however, new crystallization trials were set up.

As a further purification step, an anion exchange chromatography was performed in order to obtain a more homogeneous sample; the protein obtained after the IMAC chromatography was previously dialyzed against a buffer with a slightly higher pH value and without NaCl and then loaded onto a DEAE column.

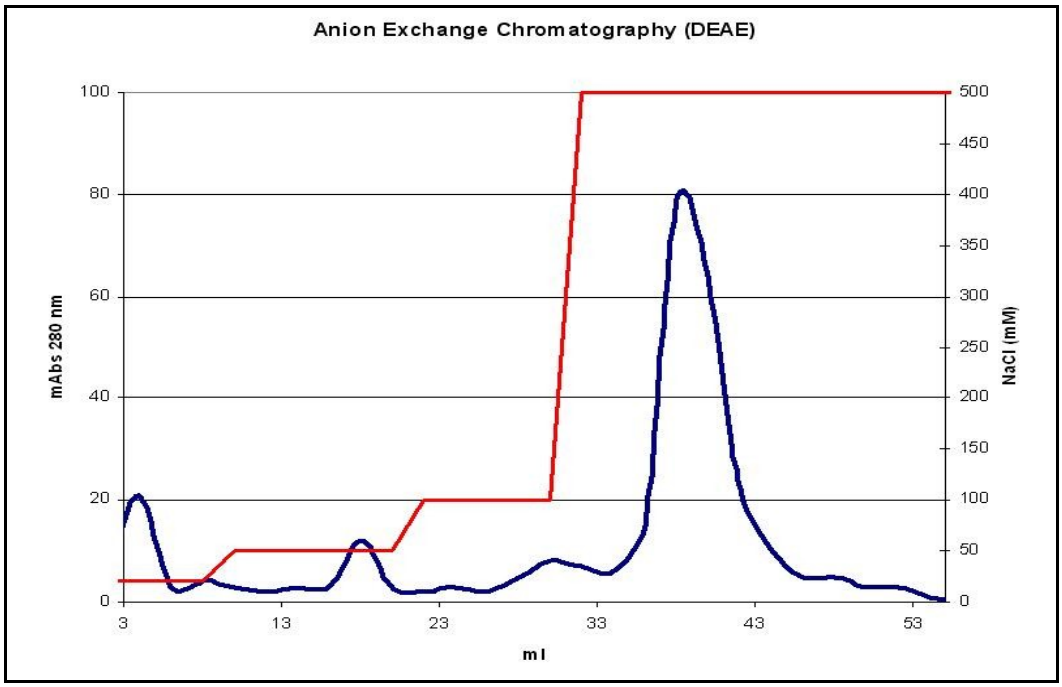


Figure 1.15: Elution profile of HSP60 from DEAE column.
The blue line corresponds to the UV absorbance and the red line to the NaCl gradient.

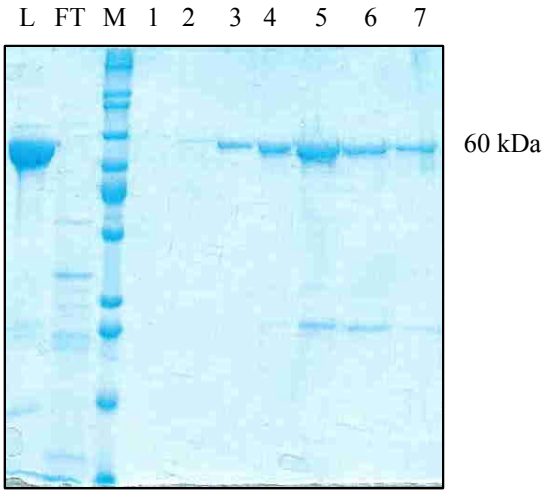


Figure 1.16: SDS-PAGE analysis of HSP60 obtained after anion exchange chromatography. L: sample loaded. FT: flow-through. Lane 1: elution with 20 mM NaCl. Lane 2: elution with 50 mM NaCl. Lane 3: elution with 100 mM NaCl. Lane 4 to 7: elution with 500 mM NaCl.

The protein was completely bound to the column and it was eluted at high values of salt concentration (100 and 500 mM NaCl). The fractions containing the pure protein were collected and crystallization trials were set up.

1.4 Conclusions and prospectives

Human HSP60 was expressed in the prokaryotic system *E. coli* and a purification protocol was set up. The protein was purified with an IMAC chromatography, followed by removal of the His-tag and a size exclusion chromatography. The protein was dialyzed in a suitable buffer, concentrated and crystallization trials were set up.

Nevertheless, no crystal suitable for X-ray experiments has yet been obtained. This is probably due to protein heterogeneity and degradation; the use of an inhibitor cocktail and EDTA has not avoided protein degradation.

The undergoing experiments are focused on improving the purification protocol; the addition of an anion exchange chromatography step seems encouraging to reach this goal.

1.5 References

[1] Czarnecka A.M., Campanella C., Zummo G., Cappello F. (2006), Mitochondrial chaperones in cancer: from molecular biology to clinical diagnostics. *Cancer Biol. Ther.* **5**: 714-720.

[2] Neuer A., Spandorfer S.D., Giraldo P., Dieterle S., Rosenwaks Z., Witkin S.S. (2000), The role of heat shock proteins in reproduction. *Hum. Reprod. Update* **6**: 149-159.

[3] Barral J.M., Broadley S.A., Schaffar G., Hartl F.U. (2004), Roles of molecular chaperones in protein misfolding diseases. *Semin. Cell. Dev. Biol.* **15**: 17-29.

[4] Walter S., Buchner J. (2002), Molecular chaperones--cellular machines for protein folding. *Angew. Chem. Int. Ed.* **41**: 1098-1113.

[5] Singh B., Patel H.V., Ridley R.G., Freeman K.B., Gupta R.S. (1990), Mitochondrial import of the human chaperonin (HSP60) protein. *Biochem. Biophys. Res. Commun.* **169**: 391-396.

[6] Cappello F., Conway de Macario E., Marasà L., Zummo G., Macario A.J. (2008), Hsp60 expression, new locations, functions and perspectives for cancer diagnosis and therapy. *Cancer Biol. Ther.* **7**: 801-809.

[7] Habich C., Burkart V. (2007), Heat shock protein 60: regulatory role on innate immune cells. *Cell. Mol. Life Sci.* **64**: 742-751.

[8] Habich C., Kempe K., Burkart V., Van Der Zee R., Lillicrap M., Gaston H., Kolb H. (2004), Identification of the heat shock protein 60 epitope involved in receptor binding on macrophages. *FEBS Lett.* **568**: 65-69.

[9] Habich C., Kempe K., Gomez F.J., Lillicrap M., Gaston H., van der Zee R., Kolb H.,

Burkart V. (2006), Heat shock protein 60: identification of specific epitopes for binding to primary macrophages. *FEBS Lett.* **580**: 115-120.

[10] Hansen J.J., Dürr A., Cournu-Rebeix I., Georgopoulos C., Ang D., Nielsen M.N., Davoine C.S., Brice A., Fontaine B., Gregersen N., Bross P. (2002), Hereditary spastic paraplegia SPG13 is associated with a mutation in the gene encoding the mitochondrial chaperonin Hsp60. *Am. J. Hum. Genet.* **70**: 1328-1332.

[11] Huckriede A., Agsteribbe E. (1994), Decreased synthesis and inefficient mitochondrial import of hsp60 in a patient with a mitochondrial encephalomyopathy. *Biochim. Biophys. Acta* **1227**: 200-206.

[12] Briones P., Vilaseca M.A., Ribes A., Vernet A., Lluch M., Cusi V., Huckriede A., Agsteribbe E. (1997), A new case of multiple mitochondrial enzyme deficiencies with decreased amount of heat shock protein 60. *J. Inherit. Metab. Dis.* **20**: 569-577.

[13] Ranford J.C., Henderson B. (2002), Chaperonins in disease: mechanisms, models, and treatments. *Mol. Pathol.* **55**: 209-213.

[14] Mandal K., Jahangiri M., Xu Q. (2004), Autoimmunity to heat shock proteins in atherosclerosis. *Autoimmun. Rev.* **3**: 31-37.

[15] Edenhofer F., Rieger R., Famulok M., Wendler W., Weiss S., Winnacker E.L. (1996), Prion protein PrPc interacts with molecular chaperones of the Hsp60 family. *J. Virol.* **70**: 4724-4728.

[16] DebBurman S.K., Raymond G.J., Caughey B., Lindquist S. (1997), Chaperone-supervised conversion of prion protein to its protease-resistant form. *Proc. Natl. Acad. Sci. USA* **94**: 13938-13943.

[17] Samali A., Cai J., Zhivotovsky B., Jones D.P., Orrenius S. (1999), Presence of a pre-apoptotic complex of pro-caspase-3, Hsp60 and Hsp10 in the mitochondrial fraction

of jurkat cells. *EMBO J.* **18**: 2040-2048.

[18] Kirchhoff S.R., Gupta S., Knowlton A.A. (2002), Cytosolic heat shock protein 60, apoptosis, and myocardial injury. *Circulation* **105**: 2899-2904.

[19] Lebret T., Watson R.W., Molinié V., O'Neill A., Gabriel C., Fitzpatrick J.M., Botto H. (2003), Heat shock proteins HSP27, HSP60, HSP70, and HSP90: expression in bladder carcinoma. *Cancer* **98**: 970-977.

[20] Johansson B., Pourian M.R., Chuan Y.C., Byman I., Bergh A., Pang S.T., Norstedt G., Bergman T., Pousette A. (2006), Proteomic comparison of prostate cancer cell lines LNCaP-FGC and LNCaP-r reveals heatshock protein 60 as a marker for prostate malignancy. *Prostate* **66**: 1235-1244.

[21] McPherson A. (1990), Current approaches to macromolecular crystallization. *Eur. J. Biochem.* **189**: 1-23.

2. Activation-induced cytidine deaminase

2.1 Introduction

2.1.1 Somatic hypermutation and class switch recombination

Vertebrates have evolved the immune system as a defence barrier against pathogens; it can be divided into innate and adaptative immune system, which is composed of B and T lymphocytes and is characterized by extreme specificity for foreign antigens.

The molecules produced by lymphocytes able to recognize pathogenic antigens in a very specific manner are antibodies and TCR (T cell receptors), synthesized by B cells and T cells, respectively. Antibodies are produced in a membrane-bound form, which is involved in B cell activation, and in a soluble form, which is responsible for antigen clearance.

The antibody, also called immunoglobulin (Ig), comprises two identical pairs of heavy chains (IgH) and of light chains (IgL) [1].

Each of the two light chains is covalently bound to one of the two heavy chains through a disulfide bond; the two heavy chains are also bound to each other by disulfide bonds. Both the heavy chains and the light chains contain immunoglobulin domains (Ig domains), composed of two sandwiched β-pleated sheets pinned together by a disulfide bridge. Both the heavy chains and the light chains are formed by an N-terminal variable (V) region and a C-terminal constant (C) region; in the light chains, both the V region and the C region comprise only one Ig domain, whereas in the heavy chains the C region has 3 or 4 Ig domains and the V region contains only one Ig domain [1].

The V regions of IgH and IgL contain the antigen-binding site; since the antibody consists of two IgL and two IgH, each antibody has two antigen binding sites. On the other hand, the C region of IgH determines the biological properties of the antibody, since it is involved in interactions with molecules and cells of the immune system.

There are two types of IgL, called k o λ; as a consequence, each antibody comprises two identical IgL k or two identical IgL λ [1].

The antibodies expressed in soluble form and in membrane-bound form differ in their amino acid sequence in the last Ig domain of IgH; the former, which are secreted into the blood and other extracellular fluids, contain charged and hydrophilic residues, while the latter, present in the membrane of B cells, have a certain number of hydrophobic

residues, followed by basic residues.

The extreme specificity for foreign antigens shown by the adaptative immune system is based on the generation of a myriad of receptors for antigens of virtually unlimited diversity; this is due to the somatic recombination of immunoglobulin genes of B lymphocytes. Germ line antigen receptor genes include an array of variable (V), diverse (D), joining (J) and constant (C) segments; D segments are present only in IgH locus. V, D and J segments encode for the variable region, while C segments encode for the constant region; there are three separate loci encoding for IgH, IgL k and IgL λ, located on chromosome 14, 2 and 22, respectively [1].

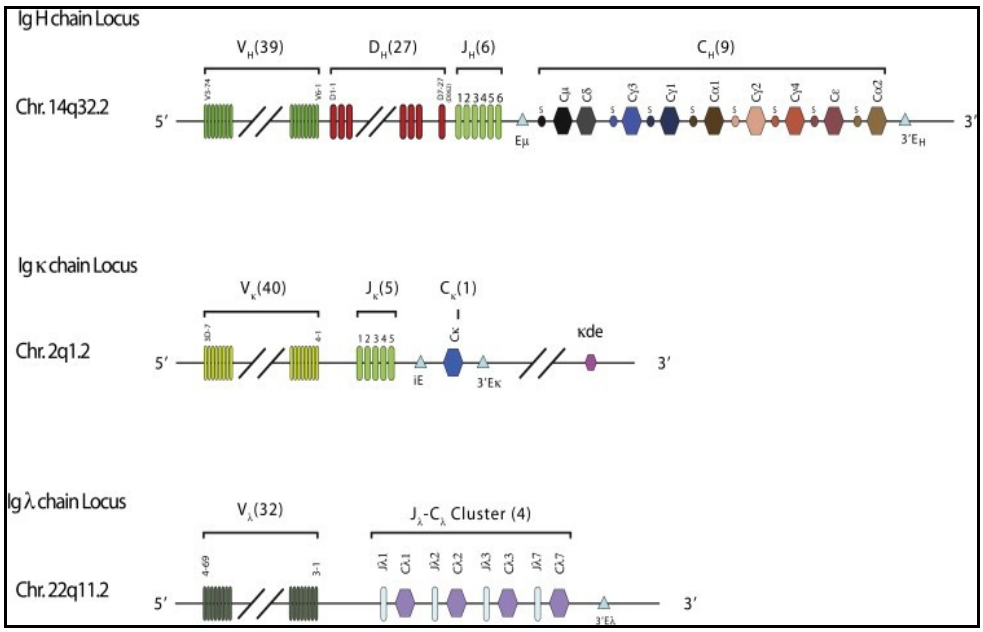


Figure 2.1: Chromosomal organization of the immunoglobulin H, k and λ loci [1].

During lymphocyte differentiation, a somatic recombination of a V, a D and a J segment generates the antigen recognition site of the receptor in each lymphocyte in a completely random process. First, a single D segment is joined to a single J segment and then this partially rearrenged D-J gene is joined to a single V segment. Moreover, junctional diversity is present in the junctions between V and D and between D and J in IgH and between V and J in IgL; this is due to the random insertion or deletion of some nucleotides in these points. The combinatorial diversity given by V, D and J segments and the junctional diversity allow the generation of a pletora of different antigen receptors, virtually able to recognize every possible antigen; the primary repertoire of

immunoglobulins is thus generated in an antigen-indipendent manner.
The IgH locus has a number of constant regions, each preceded by a high repetitive and non-coding sequence called switch region. Before encountering the antigen, the immunoglobulins produced by B cells are IgM or IgD. In particular, after leaving the bone marrow to colonize secondary limphoid organs, B cells express surface IgD of the same specificity as surface IgM; this is due to alternative splicing of a pre-messanger RNA comprising both heavy chain Cμ and Cδ regions.

After encountering its cognate antigen, B cells have another chance to further diversify their receptors; in germinal centers immunoglobulin genes undergo two indipendent mutational events, called somatic hypermutation (SHM) and class switch recombination (CSR). As a consequence, higher affinity immunoglobulins with different isotypes are generated [2, 3].

SHM consists in the introduction of nucleotide changes, usually point mutations, in the V(D)J segments of the variable region of the immunoglobulin gene. If the acquired mutations increase the receptor affinity for the antigen, the B cells will proliferate and differentiate into either plasma cells, that produce antibodies, or memory B cells. On the other hand, if the acquired mutations during SHM diminish the affinity or produce deleterious effects for the final protein, the B cells will be rendered anergic [2, 3].
Mutations are produced at a rate of 10^{-3}-10^{-4} per cell division, that is one million times greater than typical somatic cell mutation frequency. The mutations begin at about 200 bp downstream from the Ig gene transcription start site, reach a peak over the variable region, taper off at 2 kb downstream from the promoter and end before reaching the constant region [2].
Mutations are about equally divided between transcribed and non-transcribed strands. Of the 50% of mutations at G/C sites, 37% are within the WRC hot spot motif, where W is A or T and R is A or G; of the A/T mutations, 34% occur at A sites and 16% at T sites on the non-transcribed strand [2].

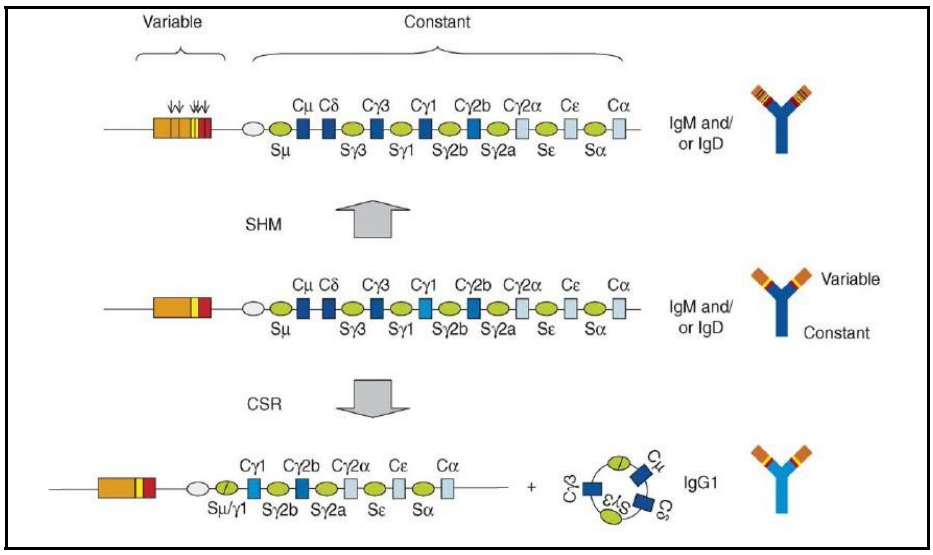


Figure 2.2: Generation of the secondary antibody repertoire in IgH locus. SHM implies the introduction of nucleotide substitutions (depicted as arrows) in the rearranged V(D)J region (orange, yellow and red, respectively). CSR consists in the exchange of the IgH Cμ segment for one of the downstream C segments (blue boxes), in the figure Cγ1 giving rise to IgG1; the recombination takes place between two switch regions (green ovals) [3].

CSR is a region-specific recombination reaction between two switch regions of the IgH locus. This reaction implies the replacing of the Cμ constant region by a downstream constant region; as a consequence, the Ig will acquire different effector functions. Double strand breaks are generated within the donor switch region (Cμ) and a recipient switch region of a downstream C region; the C regions between the two break sites are deleted as a circular DNA molecule and the VDJ region and the C region are joined together by non-homologous end joining [2, 3].

According to the constant region of the IgH, antibodies can be divided into five isotypes: IgA, IgD, IgE, IgG and IgM; additionally, IgA can be further divided in IgA1 and IgA2 and IgG in IgG1, IgG2, IgG3 and IgG4 [1].

IgMs are present in the membrane of B cells before antigen encounter; in the secretory form, they are pentameric and play an important role in the activation of the complement system [4].

IgDs are present as membrane-bound proteins in B cells before antigen encounter. Additionally, they can be found in the blood at low concentration; they are also present in nasal, lacrimal, salivary and bronchial fluids [5].

IgAs are secreted as monomers, dimers and trimers. IgA1s are produced by B cells in

bone marrow and they can be found in the serum, while IgA2s are produced by B cells in mucosae and are secreted into maternal milk, saliva and tears. In particular, IgAs represent the main element of the humoral immune response which provides protection against pathogens at mucosal surfaces [6].

IgEs are secreted as monomers and can be found in the blood at a very low concentration. They are involved in the immune response against parasitic worms and play a very important role in allergy [7].

IgGs are the most abundant antibodies in the body; they are secreted as monomers. They have the longest serum half-life of all immunoglobulin types. They can bind many kinds of pathogens, such as viruses, bacteria and fungi. The IgG subclasses exhibit different functional activity; IgG1s and IgG2s are the most abundant IgGs in the blood. IgG1s, IgG2s and IgG3s are involved in the activation of the complement cascade. IgG is the only isotype of immunoglobulin able to undergo transplacental transport, thereby providing protection to the fetus [1].

2.1.2 Human activation-induced cytidine deaminase

AICDA was first identified in 1999 as a gene specifically expressed in B cells after CSR activation [8]. Subsequently, in an experiment performed with AICDA-deficient mice, AICDA was proved to be essential for both CSR and SHM [9]. Moreover, mutations in the human AICDA gene were found to be associated with an immune deficiency called Hyper-IgM syndrome type 2; patients with the autosomal recessive form of the pathology showed three major abnormalities: the lack of immunoglobulin class switch recombination, the absence of immunoglobulin somatic hypermutation and lymph node hyperplasia due to the presence of giant germinal centers [10].

Interestingly, when AICDA was heterologously expressed in fibroblasts, it was sufficient to promote both CSR and SHM on transcribed substrates, thus proving that AICDA is the only B-cell specific factor required to initiate these reactions [11, 12].

AICDA is homologous to APOBEC1, a protein involved in the editing of C to U nucleotide in apolipoprotein B and neurofibromatosis-1 mRNA; AICDA was thus

suggested to be a mRNA deaminase involved in editing of a mRNA encoding for a protein, possibly an endonuclease, responsible for CSR and SHM. Some alternative hypotheses were proposed according to which AICDA would deaminate cytidines on double strand DNA (dsDNA), single strand DNA (ssDNA) or DNA/RNA hybrids.
At present the ssDNA deamination model is the most widely accepted. As a matter of fact, AICDA can deaminate ssDNA in vitro, with the same sequence preference observed in vivo in B cells, while dsDNA is refractory to the enzyme activity [13]. Additionally, AICDA was reported to interact with the switch regions of DNA in B cells undergoing CSR [14]. Before the identification of AICDA, SHM and CSR were known to require transcriptional activity of variable regions and switch regions, respectively; since transcription can, at least transiently, expose DNA in a single strand configuration, at present it is hypothesized that transcription of the immunoglobulin gene during germinal center reactions can allow AICDA to access its substrate ssDNA [2, 3].

The U:G mismatch generated by the AICDA catalyzed reaction can undergo different processes that lead to SHM:

- the mismatch is replicated over, thus producing a transition mutation at the C:G pair (phase 1A);
- uracil is removed by uracil-N-glycosylase (UNG), thus resulting in an abasic site; upon replication, both transitions and transversions at C:G pairs can occur (phase 1B);
- the mismatch is recognized by mismatch repair (MMR) factors, thus generating mutations at A:T pairs by error-prone patch repair (phase 2).

Alternatively, the abasic site generated by UNG can be substrate for an apyrimidinic endonuclease (AP), producing a nick in the DNA strand containing the abasic site; this nick can be repaired by the base-excision repair pathway, eliminating the DNA lesion caused by AICDA, or can be processed by non-homologous end joining (NHEJ), leading to CSR.

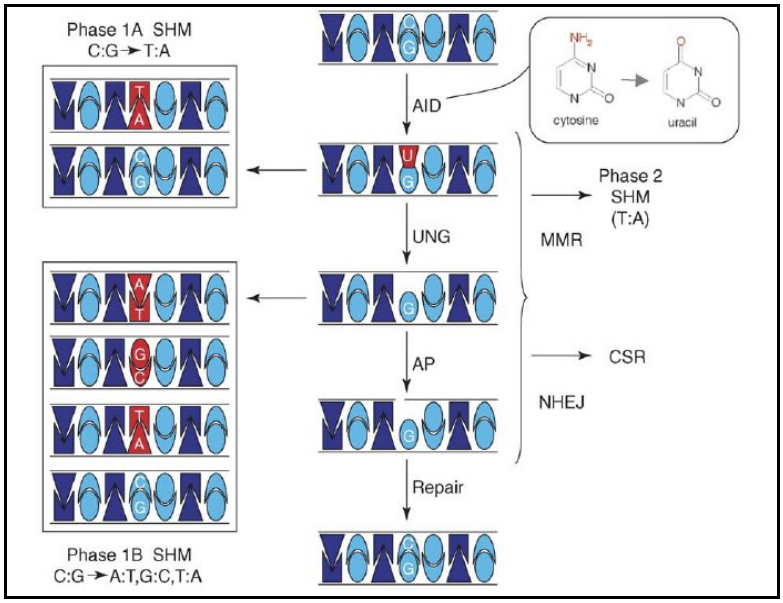


Figure 2.3: DNA deamination by AICDA in SHM and CSR. The transition is the exchange of a purine for a purine (A↔G) or a pyrimidine for a pyrimidine (C↔T), while the transversion is the exchange of a purine for a pyrimidine or a pyrimidine for a purine (A/G↔C/T) [3].

Transcriptional regulation

AICDA expression is restricted to activated B cells. The specificity and the levels of AICDA transcription are regulated by a complex combination of transcriptional activators and repressors.

Gene expression is dependent on signaling pathways that mediate the activation of B cells in germinal centers, such as those mimicked by lipopolysaccharide, IL4, transforming growth factor β and CD40 ligation; accordingly, janus kinase-signal transducer and activator of transcription and nuclear factor-kappa B (NF-kB) pathways seem to be involved [15].

An intronic E-box element present in the AICDA gene was reported to regulate positively AICDA transcription through the binding of the helix-loop-helix transcription factor E47; additionally, this intronic element downregulates AICDA transcription through the binding of the inhibitor of DNA binding helix-loop-helix protein ID3 [16].

AICDA transcription is also regulated by binding of paired box gene 5 (PAX5) and the inhibitor of DNA binding 2 (ID2) factors to the promoter region [17].

Additionally, binding sites for ubiquitous Sp transcription factors were identified in the AICDA promoter; Sp are a family of transcription factors involved in the regulation of

many different promoters through binding to GC box or GT motifs [18].

Post-transcriptional regulation

A microRNA previously known to play a crucial role in activation of B cells, miR-155, was found to be upregulated in murine B cells undergoing CSR; this microRNA targets a conserved site in the 3'-untranslated region of the mRNA encoding AICDA. Disruption of this site in vivo was observed to cause a deregulation of AICDA expression, increasing the level of both CSR and SHM [19]. Additionally, mutations in miR-155 binding site in the 3'-untranslated region of AICDA mRNA determined an increase in AICDA mRNA half-life and in the protein level [20]. Deregulation of AICDA activity resulted in mutations on non-target genes, such as the antiapoptotic gene *Bcl6*, and in an increase in *Myc-Igh* translocations, suggesting that miR-155 can act as a tumor suppressor by reducing potentially oncogenic translocations produced by AICDA [19, 20].

Subcellular localization

AICDA activity is also regulated by controlling the protein subcellular localization; most AICDA protein is located in the cytoplasm, since AICDA was observed in the nucleus in only 3-11% of activated B cells [2]. Indeed, the protein possesses a nuclear export signal in the last C-terminal 10 amino acids; deletion of this export signal caused nuclear accumulation of the enzyme, but did not increase the number of mutations in immunoglobulin genes [2].

Post-translational modifications

AICDA activity is enhanced upon phosphorylation at Ser38 by protein kinase A (PKA). Mice expressing only S38A AICDA have 70% less CSR and SHM, thus suggesting that this phosphorylation is required for efficient levels of both SHM and CSR [21]. Ser-38 phosphorilation is required for AICDA to interact with replication protein A (RPA) [2]. Other three phosphorylation sites were identified (Thr27, Thr140 and Tyr184), but so far their physiological role remains unclear [21].

AICDA targeting to Ig locus

A not yet clarified point is the way AICDA specifically targets Ig genes and not other genes; moreover, the enzyme activity is restricted only in specific regions within Ig genes, that are the variable regions in SHM and the switch regions inside the constant regions in CSR.

Ig locus transcription is strictly required for AICDA activity and removal of Ig promoters dramatically decreases SHM; additionally, the rate of Ig transcription correlates with the rate of SHM and CSR [3].

It was hypothesized that this enzyme may be loaded on the transcription initiation complex, to follow RNA polymerase II during elongation and that it stochastically dissociates, creating the decaying pattern of mutations observed in SHM [21]. Nevertheless, transcription alone is not sufficient for targeting AICDA, since other genes transcribed at high levels in B cells are not mutated.

Ig cis-acting elements, such as the intronic enhancer (iE) and the matrix attachment regions (MAR) have been suggested to play a role in AICDA targeting to Ig locus [2].

Additionally, some epigenetic modifications may be involved. Hyperacetylation of histone 3 in switch regions correlates with induction of CSR and Ig transcription, suggesting that this modification may mark IgH locus for AICDA to initiate the CSR reaction. Histone 4 hyperacetylation of switch regions correlates with CSR in an AICDA-dependent manner, suggesting that this modification may be downstream of AICDA activity and play a role in recruiting double strand break repair factors [3].

AICDA in B-cell lymphomas

About 95% of diagnosed lymphomas are of B-cell origin and the majority of them are thought to originate from germinal-center or post germinal-center B cells. A common hallmark of B-cell lymphomas is the presence of chromosome translocation causing the juxtaposition of the Ig locus and a proto-oncogene. Some of these translocations are related to a specific type of B-cell lymphomas, such as BCL-1-Ig translocation in mantle-zone lymphoma, BCL-2-Ig translocation in follicular lymphoma and c-MYC-Ig translocation in Burkitt lymphoma; this chromosome fusion causes the oncogene to come under the control of Ig cis sequences and its deregulated and constitutive expression. Other common and characteristic features of particular B-cell malignancies

are mutations in tumor suppressor genes, such as *p53*, retinoblastoma (*RB*), *p19* and ataxia telangiectasia mutated (*ATM*) [3].

At physiological AICDA expression levels, c-MYC-IgH translocations are extremely rare; this suggests that in physiological situations cellular surveillance mechanisms prevent the generation of chromosome translocations [3]. Mice lacking DNA-damage response proteins, such as ATM, histone 2A family member X (H2AX) and p53 binding protein 1 (53BP1), are prone to accumulate chromosomal breaks and aberrations associated with the IgH locus; in the case of H2AX and 53BP1, these mutations were reported to depend on the presence of AICDA. Moreover, B cells lacking p53 show a higher AICDA-dependent c-MYC-IgH translocation frequency [3, 22, 23].

2.1.3 APOBEC family

AICDA belongs to the APOBEC family, which comprises enzymes able to insert mutations in RNA and single strand DNA through deamination of cytidine to uridine [24 - 26].

APOBEC1 (apoB editing catalytic subunit 1, Apo1) is expressed in the small intestine, where it is responsible for editing of *ApoB* pre-mRNA; the deamination of cytidine 6666 converts a glutamine codon (CAA) to a stop codon (UAA), thus generating a shorter form of the protein. The protein is the main component of the hydrophilic shell of the chylomicrons, which are the very low-density lipoproteins involved in trasport of triglycerides from the intestine to the tissues.

APOBEC2 (Apo2) is expressed in cardiac and skeletal muscle; the enzyme shows deaminase activity on cytidine and deoxycytidine in vitro, but its physiological role remains unknown.

The term APOBEC3 (Apo3) is referred to proteins (APOBEC-3A, -3B, -3C, -3DE, -3F, -3G, -3H) involved in the restriction of retrovirus and retrotransposons propagation. Apo3G was identified as a factor involved in HIV restriction. The protein is thought to be packed into the HIV virion and to exert its enzymatic activity on the nascent first DNA strand generated by reverse transcription in the target cell; as a consequence, Apo3G prevents the viral genome from integration into the cell's genome.

The APOBEC enzymes share a catalytic domain defined by a conserved cytidine deamination motif (H-X-E-X_{23-28}-P-C-X-X-C); the residues of this motif coordinate a zinc atom, which is required for the nucleophilic attack in the deamination reaction. Apo3B, Apo3DE, Apo3F and Apo3G have two catalytic domains containing the amino acids required for catalysis. Both domains of Apo3B are enzymatically active, while only the C-terminal domains of Apo3F and Apo3G are enzymatically active.

```
AICDA          ----------------------------------------MDSLLMNRRKFLYQFKNVR 19
APOBEC2        MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVE 60
APOBEC1        ---------------------------------MTSEKGPSTGDPTLRRRIEPWEFDVFY 27
                                                       .  :  .   ::*  .

AICDA          WAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNG-CHVELLFLR-YISDWDLDPGRCYR 77
APOBEC2        YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFN-TILP-AFDPALRYN 118
APOBEC1        DPRELRKEACLLYEIKWGMSRKIWRSSG----KNTTNHVEVNFIKKFTSERDFHPSMSCS 83
                .   .:          . . ..  . *    ::   *.*  *:.       :.*.

AICDA          VTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIA 137
APOBEC2        VTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQ-AALKKLKEAGCKLR 177
APOBEC1        ITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNR-QGLRDLVNSGVTIQ 142
               :**: * ***  *:  : . *    .: * * ..**:   : : .  .*: * .:*  :

AICDA          IMTFKDYFYCWNTFVEN---HERTFKAWEGLHENSVRLSRQLRRILLP------------ 182
APOBEC2        IMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK------------- 224
APOBEC1        IMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNH 202
               **  .:: : *..**:    ..  :  :  :       . :   :

AICDA          --LYEVDDLRDAFRTLGL----------------- 198
APOBEC2        ----------------------------------
APOBEC1        LTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR 236
```

Figure 2.4: Comparison of human AICDA, Apo2 and Apo1. The amino acids involved in coordination to the zinc atom and the conserved glutamate residues are shown in red and in yellow, respectively. Sequence alignment was performed using the program ClustalW 2.0 [27].

At present there are two crystal structures of APOBEC family members, Apo2 and the C-terminal catalytic domain of Apo3G (Apo3G-CD2) [28, 29].

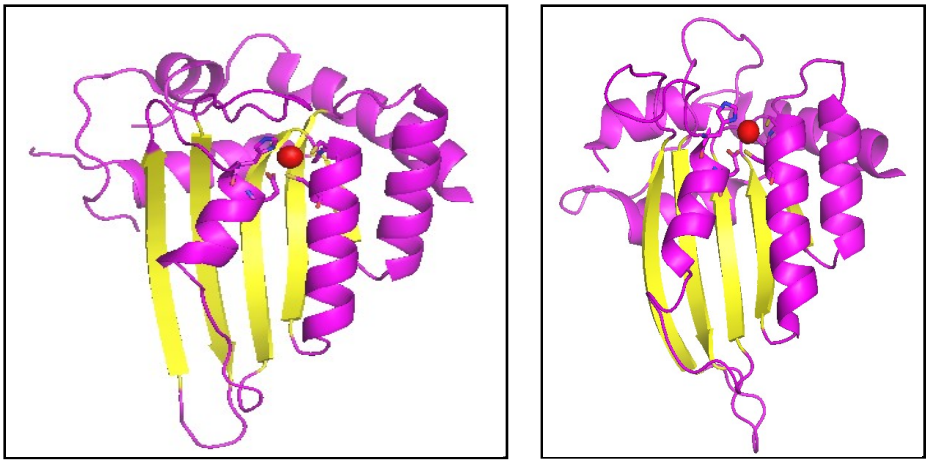


Figure 2.5: On the left an Apo2 (PDB ID 2NYT) monomer and on the right Apo3G-CD2 (PDB ID 3E1U) showing the catalytic residues and the zinc atom (red sphere) in the active site. The figures were generated using the program PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

The structures of the Apo2 monomer (residues 41-224) and of Apo3G-CD2 (residues 197-380) show a common β-sheet core composed of 5 β-strands surrounded by 6 α-helices.

In the active site of Apo2 and Apo3G-CD2, the active center zinc atom is coordinated by three conserved residues (H98, C128 and C131 in Apo2, H257, C288 and C291 in Apo3G-CD2); during the enzymatic reaction, a closely located water molecule acts as a hydrogen donor, while the conserved glutamate residue (E100 in Apo2 and E259 in Apo3G-CD2) serves as a proton shuffler.

2.2 Materials and Methods

2.2.1 The human AICDA clone

The plasmid with the full-lenght cDNA clone (IMAGE ID 4054915) of human AICDA was purchased from RZPD. *E. coli* cells harbouring the plasmid vector pOTB7 were grown in 3 ml of LB medium containing 50 µg/ml chloramphenicol overnight at 37ºC. After harvesting the cells by centrifugation, the GenElute™ Plasmid Miniprep kit was used to purify the plasmid.

2.2.2 Subcloning of AICDA gene

The protein coding sequence was subcloned into plasmids suitable for expression in *E. coli*.

First, a PCR reaction was used to amplify the coding sequence of the protein; the primers, the parameters of PCR and the plasmid maps are reported as follows.

pET15b:

Primer forward:

5'-GGTGGTCATATGGACAGCCTCTTGATGAAC-3'

Primer reverse:

5'-GGTGGATCCTTAAAGTCCCAAAGTACGAAA-3'

CATATG - *NdeI* restriction site

GGATCC- *BamHI* restriction site

TTA - stop codon

pQE50, pET22b, pET28:

Primer forward:

5'-GGTGGTGGATCCATGGACAGCCTCTTGATGAAC-3'

Primer reverse:

5'-GGTGGTAAGCTTGGAACCACGCGGAACCAGAAGTCCCAAAGTA

CGAAA-3'

GGATCC- *BamHI* restriction site

AAGCTT - *HindIII* restriction site

GGAACCACGCGGAACCAG – thrombin cleavage site

pGEX-4T-1:

Primer forward:

5'-GGTGGATCCATGGACAGCCTCTTGATGAACCGG-3'

Primer reverse:

5'-GGTGTCGACTCAAAGTCCCAAAGTACGAAATGC-3'

Primer reverse-His-tag:

5'-GGTGTCGACCTAATGATGATGATGATGATGAAGTCCCAAAGTACG-3'

GGATCC- *BamHI* restriction site

GTCGAC – *SalI* restriction site

TCA - stop codon

CTA - stop codon

ATGATGATGATGATGATG – His-tag

Figure 2.6: The primers used to amplify hAICDA; the coding sequence is underlined.

STEP	TEMPERATURE	TIME	
Initial Denaturation	95ºC	10 min	
Denaturation	95ºC	1 min	35 cycles
Annealing	58ºC	1 min	
Extension	72ºC	1 min	
Final Extension	72ºC	10 min	

Figure 2.7: The PCR parameters.

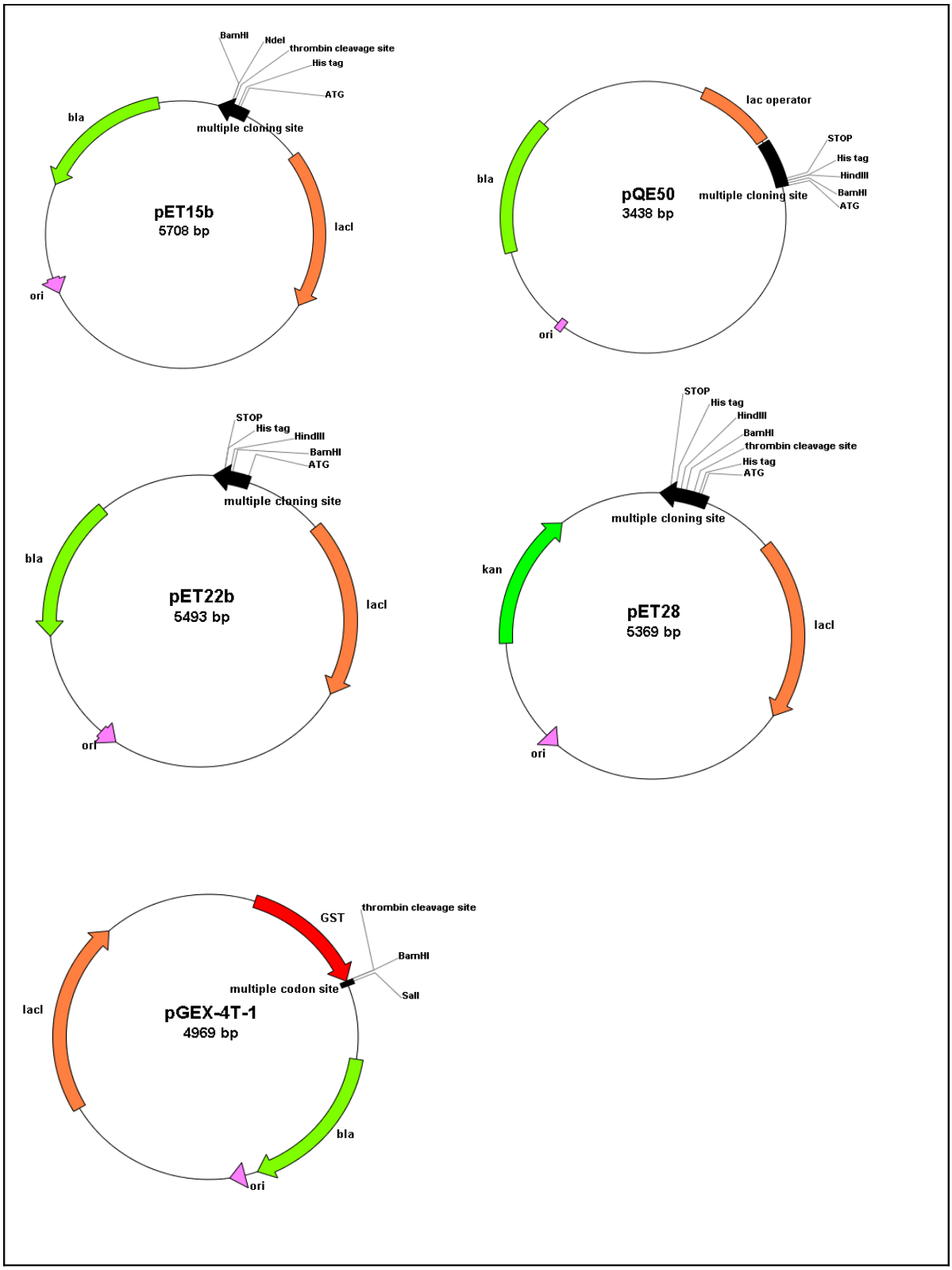


Figure 2.8: Schematic representation of the employed plasmids.

The plasmid pQE50 (QIAGEN) allows the insertion of a 6 histidine tag at the C-terminal portion of the protein, in order to facilitate target protein purification and detection by Western blot analysis. On the other hand, using pET15 and pET22 (Novagen®) it is possible to fuse the His-tag at the N- and C-terminal portion of the target protein, respectively, while in pET28 (Novagen®) the His-tag is inserted both at the N- and the C-terminus of the protein. In all the plasmids a thrombin cleavage site allows tag removal after protein purification.
The pGEX-4T-1 vector (Amersham Biosciences) permits to fuse the N-terminal portion of the target protein to the glutathione-S-transferase (GST), in order to help protein purification; a thrombin cleavage site is also placed to remove GST after protein purification. A second vector with GST was created, with the additional insertion of a His-tag at the C-terminal region of the target protein.

The amplified sequence was quantified on a 0.8% agarose TBE gel and digested with restriction enzymes. The fragment was then ligated into the *NdeI*/*BamHI* site of the pET15b plasmid, into the *BamHI*/*HindIII* site of the pQE50, pET22 and pET28 plasmids and into the *BamHI*/*SalI* site of pGEX-4T-1 plasmid.
The resulting recombinant plasmid was used to transform *E. coli* XL1 Blue competent cells by heat shock and the cells were plated on a LB-agar plate containing the appropriate antibiotic (100 μg/ml ampicillin or 50 μg/ml kanamycin); the cells were then grown overnight at 37ºC. The colonies were selected by colony PCR and subsequent digestion of the purified plasmids; the vectors were finally checked by automated DNA sequencing.

2.2.3 Expression screening of the recombinant protein

The pET15, pET22, pET28 and pGEX-4T-1 recombinant plasmids were used to transform *E. coli* BL21 (DE3) strain cells by heat shock, while pQE50 was inserted in *E. coli* SG13009. To test protein expression, a single colony was inoculated in 3 ml of LB medium, grown at 37ºC adding 100 μg/ml ampicillin or 50 μg/ml kanamycin and then incubated at 37ºC for 4 hours and at 20ºC and 28ºC overnight after induction with

IPTG to a final concentration of 0.25 mM.

Centrifugation at 9,000 rpm for 10 minutes at 4ºC was performed to harvest the cells, that were resuspended in 2 ml of buffer A (20 mM TRIS-HCl pH 7.5, 0.5 M NaCl and 0.02% NaN_3) and subsequently disrupted by sonication. Samples were collected from the total lysate and, after centrifugation at 9,000 rpm for 10 minutes at 4ºC, from the soluble fraction and loaded onto a SDS-PAGE gel (15% polyacrylamide resolving gel and 4% polyacrylamide stacking gel for pQE50, pET15, pET22 and pET28, while a 12% polyacrylamide resolving gel was used with pGEX-4T-1).

Protein expression was evaluated by Western blot analysis using an anti-His-tag monoclonal antibody for the recombinant vectors with His-tag, while for pGEX-4T-1 only the SDS-PAGE analysis was performed.

The expression screening was repeated using another culture medium (Terrific Broth, TB).

Additional expression strains, such as C41(DE3), Origami, ER2566, BL21 CodonPlus and Rosetta (DE3), were used to test protein expression with pET15, pET22, pET28 and pGEX-4T-1.

2.2.4 Protein purification

Glutathione affinity chromatography

E. coli BL21 (DE3) cells containing the plasmid AICDA-pGEX-4T-1 were grown overnight at 37ºC in 10 ml of LB medium containing 100 µg/ml ampicillin, the overnight culture was then transferred into 1 liter of LB medium and grown at 37ºC under shaking until an OD_{600} value of 0.8 was reached; after addition of IPTG to a final concentration of 0.25 mM to induce protein expression, the culture was shaken at 20ºC overnight. Cells were harvested by centrifugation at 8,000 rpm for 10 minutes at 4ºC and the pellet was then resuspended in 50 ml of buffer B (20 mM TRIS-HCl pH 7.5, 0.15 M NaCl and 0.02% NaN_3) and 1 mM DTT (dithiothreitol).

After addition of 0.1 mM PMSF as protease inhibitor, the cells were disrupted by sonication; subsequently, the sample was cetrifuged at 9,000 rpm for 10 minutes at 4ºC to remove cell debris.

The soluble fraction was directly loaded onto an immobilized glutathione resin (Pharmacia), previously preequilibrated in buffer B and 1 mM DTT. After loading, the column was further washed with buffer B and 1 mM DTT until the U.V. absorption at 280 nm returned to the baseline.
At the end, elution was achieved with buffer E (50 mM TRIS pH 8.0, 0.15 M NaCl, 10 mM glutathione and 0.02% NaN_3).

Immobilized Metal Ion Affinity Chromatography

The plasmid AICDA-pGEX-4T-1 with an additional insertion of a His-tag at the C-terminal region of the target protein was transformed into *E. coli* BL21 (DE3) cells. One liter LB culture was grown and induced according to the same procedure reported in the last paragraph.
The cells were recovered by centrifugation and resuspended in buffer A, 10 mM imidazole and 5 mM β mercaptoethanol, lysed by sonication and centrifuged.
Protein purification was performed using a Nickel Sepharose column (5 ml prepacked column, HisTrap™ FF); after loading the soluble fraction, the column was further washed until the U.V. absorption at 280 nm returned to the baseline.
Elution was achieved by applying a linear gradient ranging from 10 mM to 500 mM imidazole to the column.

2.3 Results and discussion

The human AICDA gene was subcloned into pQE50, pET15, pET22 and pET28 plasmids; the AICDA-pQE50 plasmid was used to transform *E. coli* SG13009 cells, while the other recombinant plasmids were used to transform *E. coli* BL21 (DE3) cells.
Human AICDA is a 198 amino acid chain, corresponding to a molecular weight of 24 kDa. Three induction conditions (37ºC for 4 hours, 20ºC and 28ºC overnight) and two different culture media (LB and TB) were tested.
According to the subsequent Western blot analysis, the protein was not detected in the total and in the soluble fractions in any of the conditions tested. Additional *E. coli* strains were tested, but no significant improvement in protein expression was observed; a different culture medium (TB) was also tested.

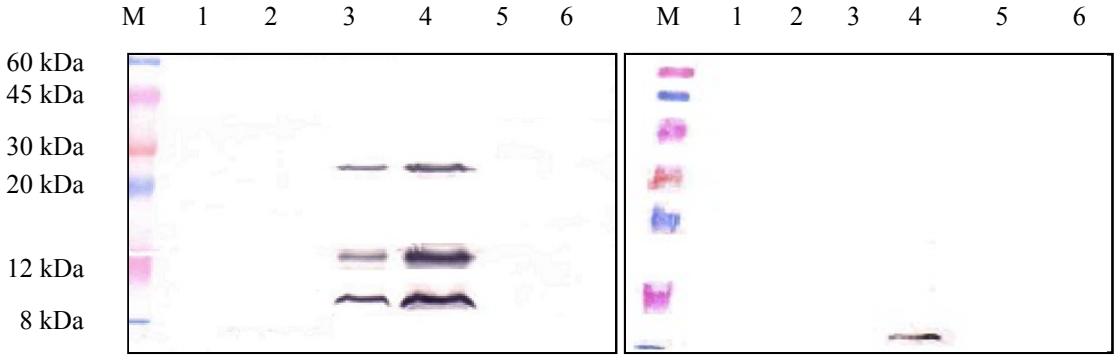


Figure 2.9: Western blot analysis of human AICDA in *E. coli* BL21 (DE3) with TB as a culture medium. On the left the total fractions of pET22, pET28 and pQE50 are shown in lane 1, 3 and 5 (induction at 37ºC for 4 hours) and 2, 4 and 6 (induction at 20ºC overnight); the corresponding soluble fractions are shown on the right.

The human AICDA coding sequence was also inserted into pGEX-4T-1, a construct which allows to fuse the N-terminal target protein to GST protein; a second GST-fusion construct was prepared, in which a His-tag was inserted at the C-terminus of the target protein.
One liter LB culture of *E. coli* BL21 (DE3) AICDA-pGEX-4T-1 was grown and induction was carried out at 20ºC overnight; the cells were lysed and the soluble fraction was loaded onto an immobilized glutathione resin.
In the SDS-PAGE gel, a band with the expected molecular weight for the fusion construct (51 kDa) is not clearly visible in the fraction eluted with glutathione.

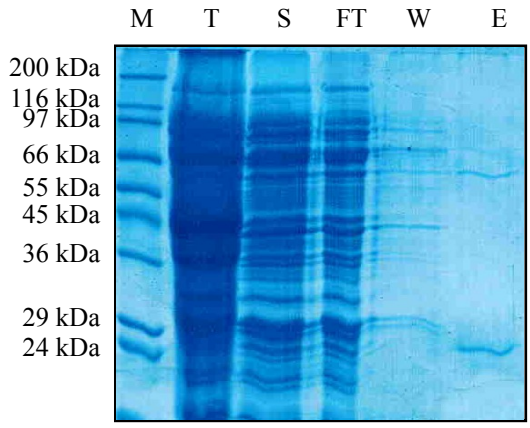


Figure 2.10: SDS-PAGE analysis of AICDA after glutathione resin purification.
T: total fraction. S: soluble fraction. FT: flow-through. W: wash. E: elution.

To check whether the protein was not expressed, not able to bind the glutathione resin or present in low amount in the elution, another bacterial culture was grown with the GST fusion construct containing an additional His-tag at the C-terminus and a Western blot analysis was performed.

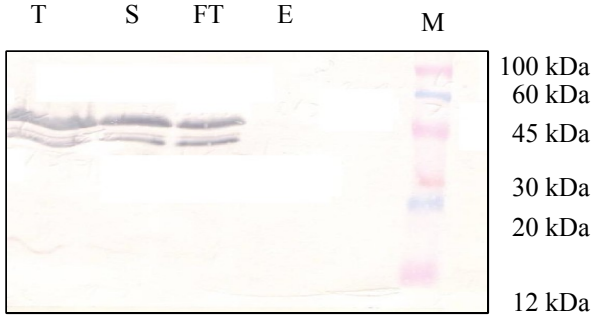


Figure 2.11: Western blot analysis of AICDA after glutathione resin purification.
T: total fraction. S: soluble fraction. FT: flow-through. E: elution.

The GST-AICDA-His-tag fusion protein is present in both the total and the soluble fraction, but not in the elution. This is due to lack of binding of the protein to the column; indeed the protein is entirely present in the flow-through.

The lack of binding of the fusion protein may be due to an incorrect folding of AICDA; according to this hypothesis, it is possible that the addition of zinc, which is known to interact with the active site residues of APOBEC family members, may positively affect

AICDA proper folding.

A new culture of *E. coli* BL21 (DE3) AICDA-pGEX-4T-1 was grown and, just before induction with ITPG, sterile $ZnCl_2$ 0.4 mM (final) was added; all the buffers used in the subsequent purification steps contained $ZnCl_2$ at the same concentration.

Nevertheless, again the fusion protein is entirely present in the flow-through and no protein was detected in the elution fraction.

After cell disruption, a DNA binding protein may bind to DNA in a non-specific manner; in particular, AICDA presents a high value of isoelectric point (9.50) and its substrate is thought to be a single strand DNA. As a consequence, the aspecifically bound nucleic acid contaminants may interfere with GST binding to the glutathione resin.

In order to destroy DNA and RNA, 5 µg/ml DNase I, 5 µg/ml RNase A and 5 mM (final) $MgCl_2$ were added to cells before sonication; after lysis, the cells were incubated for 30 minutes at 4ºC.

Additionally, to dissociate the target protein from nucleic acids, before the centrifugation to remove cell debris, 5.0 NaCl was added to the sample increasing the salt concentration to 1.0 M NaCl, before loading the sample onto the column.

In spite of this, the target protein did not bind to the affinity column.

Another one liter culture of *E. coli* BL21 (DE3) containing the plasmid AICDA-pGEX-4T-1 with a C-terminal His-tag was grown. The fusion protein does not bind to the Nickel Sepharose column and it is present completely in the flow-through.

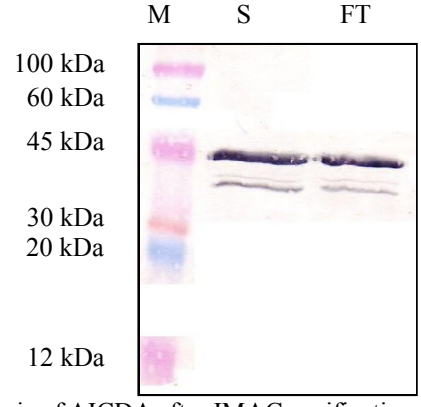


Figure 2.12: Western blot analysis of AICDA after IMAC purification. The soluble fraction (S) and the flow-through (F) are shown.

The lack of protein binding to the column may occur because the His-tag is not exposed to the solvent, but it is buried inside the protein.

2.4 Conclusions and prospectives

The expression of human AICDA was attempted in the prokaryotic system *E. coli* using different vectors.

The protein was not expressed using plasmids that allow to fuse a His-tag to the target protein; using the pGEX-4T-1 plasmid, which inserts GST at the N-terminus of the protein of interest, protein expression was obtained in the soluble fraction. Nevertheless, the purification protocol was not successful, due to lack of protein binding to the glutathione resin.

In the future, the protein coding sequence may be cloned into different vectors that allow to express AICDA together with a different fusion protein, preferentially at the N-terminus; in this way, it may be possible to obtain good expression levels of AICDA in the soluble fraction and then try new purification protocols.

2.5 References

[1] Schroeder H.W. Jr., Cavacini L. (2010), Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**: 41-52.

[2] Bransteitter R., Sneeden J.L., Allen S., Pham P., Goodman M.F. (2006), First AID (activation-induced cytidine deaminase) is needed to produce high affinity isotype-switched antibodies. *J. Biol. Chem.* **281**: 16833-16836.

[3] de Yébenes V.G., Ramiro A.R. (2006), Activation-induced deaminase: light and dark sides. *Trends Mol. Med.* **12**: 432-439.

[4] Kojouharova M., Reid K., Gadjeva M. (2010), New insights into the molecular mechanisms of classical complement activation. *Mol. Immunol.* **47**: 2154-2160.

[5] Chen K., Cerutti A. (2010), New insights into the enigma of immunoglobulin D. *Immunol. Rev.* **237**: 160-179.

[6] Snoeck V., Peters I.R., Cox E. (2006), The IgA system: a comparison of structure and function in different species. *Vet. Res.* **37**: 455-467.

[7] Erb K.J. (2007), Helminths, allergic disorders and IgE-mediated immune responses: where do we stand? *Eur. J. Immunol.* **37**: 1170-1173.

[8] Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K., Davidson N.O., Honjo T. (1999), Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**: 18470-18476.

[9] Muramatsu M., Kinoshita K., Fagarasan S., Yamada S., Shinkai Y., Honjo T. (2000), Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**: 553-563.

[10] Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O., Catalan N., Forveille M., Dufourcq-Labelouse R., Gennery A., Tezcan I., Ersoy F., Kayserili H., Ugazio A.G., Brousse N., Muramatsu M., Notarangelo L.D., Kinoshita K., Honjo T., Fischer A., Durandy A. (2000), Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**: 565-575.

[11] Okazaki I.M., Kinoshita K., Muramatsu M., Yoshikawa K., Honjo T. (2002), The AID enzyme induces class switch recombination in fibroblasts. *Nature* **416**: 340-345.

[12] Yoshikawa K., Okazaki I.M., Eto T., Kinoshita K., Muramatsu M., Nagaoka H., Honjo T. (2002), AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* **296**: 2033-2036.

[13] Pham P., Bransteitter R., Petruska J., Goodman M.F. (2003), Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* **424**: 103-107.

[14] Nambu Y., Sugai M., Gonda H., Lee C.G., Katakai T., Agata Y., Yokota Y., Shimizu A. (2003), Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* **302**: 2137-2140.

[15] Pérez-Durán P., de Yebenes V.G., Ramiro A.R. (2007), Oncogenic events triggered by AID, the adverse effect of antibody diversification. *Carcinogenesis* **28**: 2427-2433.

[16] Sayegh C.E., Quong M.W., Agata Y., Murre C. (2003), E-proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat. Immunol.* **4**: 586-593.

[17] Gonda H., Sugai M., Nambu Y., Katakai T., Agata Y., Mori K.J., Yokota Y., Shimizu A. (2003), The balance between Pax5 and Id2 activities is the key to AID gene expression. *J. Exp. Med.* **198**: 1427-1437.

[18] Yadav A., Olaru A., Saltis M., Setren A., Cerny J., Livák F. (2006), Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. *Mol. Immunol.* **43**: 529-541.

[19] Teng G., Hakimpour P., Landgraf P., Rice A., Tuschl T., Casellas R., Papavasiliou F.N. (2008), MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* **28**: 621-629.

[20] Dorsett Y., McBride K.M., Jankovic M., Gazumyan A., Thai T.H., Robbiani D.F., Di Virgilio M., Reina San-Martin B., Heidkamp G., Schwickert T.A., Eisenreich T., Rajewsky K., Nussenzweig M.C. (2008), MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* **28**: 630-638.

[21] Delker R.K., Fugmann S.D., Papavasiliou F.N. (2009), A coming-of-age story: activation-induced cytidine deaminase turns 10. *Nat. Immunol.* **10**: 1147-1153.

[22] Ramiro A.R., Jankovic M., Callen E., Difilippantonio S., Chen H.T., McBride K.M., Eisenreich T.R., Chen J., Dickins R.A., Lowe S.W., Nussenzweig A., Nussenzweig M.C. (2006), Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature* **440**: 105-109.

[23] Franco S., Gostissa M., Zha S., Lombard D.B., Murphy M.M., Zarrin A.A., Yan C., Tepsuporn S., Morales J.C., Adams M.M., Lou Z., Bassing C.H., Manis J.P., Chen J., Carpenter P.B., Alt F.W. (2006), H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell.* **21**: 201-214.

[24] Bransteitter R., Prochnow C., Chen X.S. (2009), The current structural and functional understanding of APOBEC deaminases. *Cell. Mol. Life Sci.* **66**: 3137-3147.

[25] Wedekind J.E., Dance G.S., Sowden M.P., Smith H.C. (2003), Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet.* **19**: 207-216.

[26] Conticello S.G. (2008), The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**: 229.1-229.10.

[27] Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. (2007), Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

[28] Prochnow C., Bransteitter R., Klein M.G., Goodman M.F., Chen X.S. (2007), The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **445**: 447-451.

[29] Holden L.G., Prochnow C., Chang Y.P., Bransteitter R., Chelico L., Sen U., Stevens R.C., Goodman M.F., Chen X.S. (2008), Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**: 121-124.

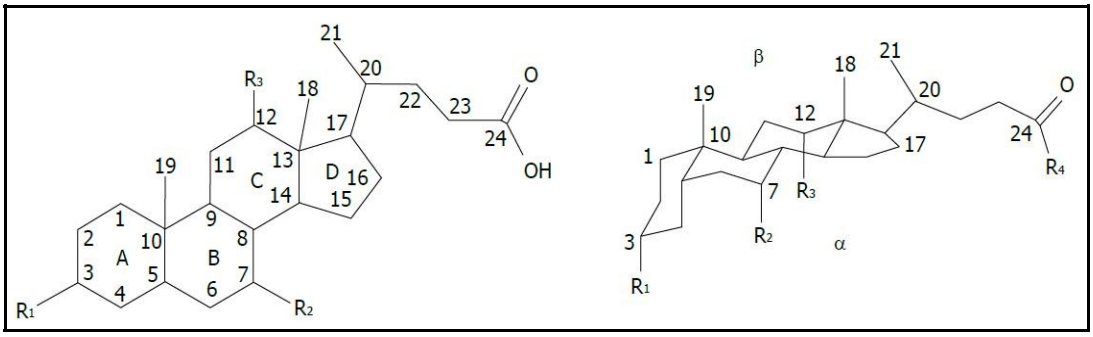
3. Cholesterol 7-alpha-monooxygenase

3.1 Introduction

3.1.1 Bile acids and their synthesis

Bile acids are a group of acidic steroids, synthesized in the liver and involved in emulsion and absorption of dietary fats and liposoluble vitamins [1].

They have 24 carbon atoms and their chemical structure consists of a saturated tetracyclic hydrocarbon perhydrocyclopentanophenanthrene system, called steroid nucleus.



Name	R_1	R_2	R_3	R_4
Cholanic acid	H	H	H	OH
Cholic acid	OH	OH	OH	OH
Chenodeoxycholic acid	OH (α)	OH (α)	H	OH
Deoxycholic acid	OH	H	OH	OH
Ursodeoxycholic acid	OH (α)	OH (β)	H	OH
Lithocholic acid	OH	H	H	OH
Glycocholate	OH	OH	OH	$NHCH_2COO^-$
Taurocholate	OH	OH	OH	$NHCH_2CH_2SO_3^-$

Figure 3.1: The most abundant bile acids in humans [1].

The steroid nucleus is constituted of three six-member rings (A, B and C) and a five-member ring (D); mammalian bile acids have a 5β-configuration, that is A/B ring junction is *cis* configuration, with hydroxyl groups at 3α, 7α and 12α. Bile acid

molecules are about 20 Å long, with an average radius of about 3.5 Å.

In human bile, these compounds are almost completely present in conjugated form with glycine (75%) or taurine (25%); conjugation increases water solubility of bile acids.

Bile acids are amphipathic, with a hydrophilic side (α-face) and a hydrophobic side (β-face). The carboxylic side chain and the hydroxyl groups, that are oriented towards the α-side, with the exception of C-7 in UDCA (ursodeoxycholic acid), confer hydrophilic character to the molecule; instead, the hydrophobic methyl groups at C-18 and C-19 are oriented towards the β-side.

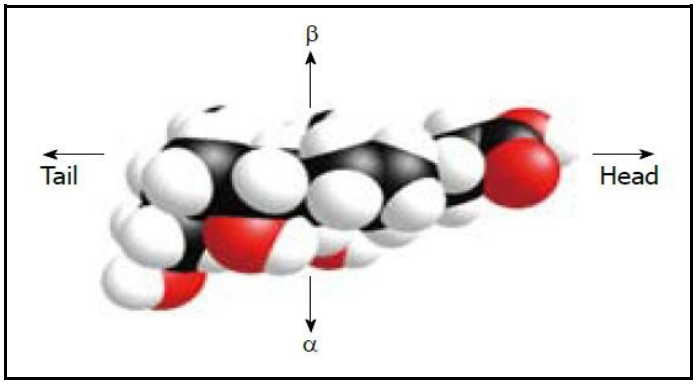


Figure 3.2: Structure of cholic acid [1].

Since they are amphipathic, bile acids show a great surface activity and in aqueous solution, when their concentration reaches a value above the critical micellar concentration, they form small micelles of usually less than 10 molecules.

Bile acid synthesis

Bile acid formation starts from cholesterol, through two biosynthetic pathways, the "classical" and the "alternative"; in humans, bile acid synthesis is the main catabolic pathway of cholesterol [1, 2].

The classical pathway occurs only in the liver and leads to the synthesis of CA (cholic acid) and CDCA (chenodeoxycholic acid), which are the two primary bile acids in humans; the enzymes involved in this pathway are located in different subcellular sites, that are microsomes, cytosol, mitochondria and peroxisomes. Since the intermediate metabolites are neutral sterols, this pathway is also called "neutral" pathway; in particular, the modification of the steroid nucleus occurs before the oxidative cleavage of its side chain.

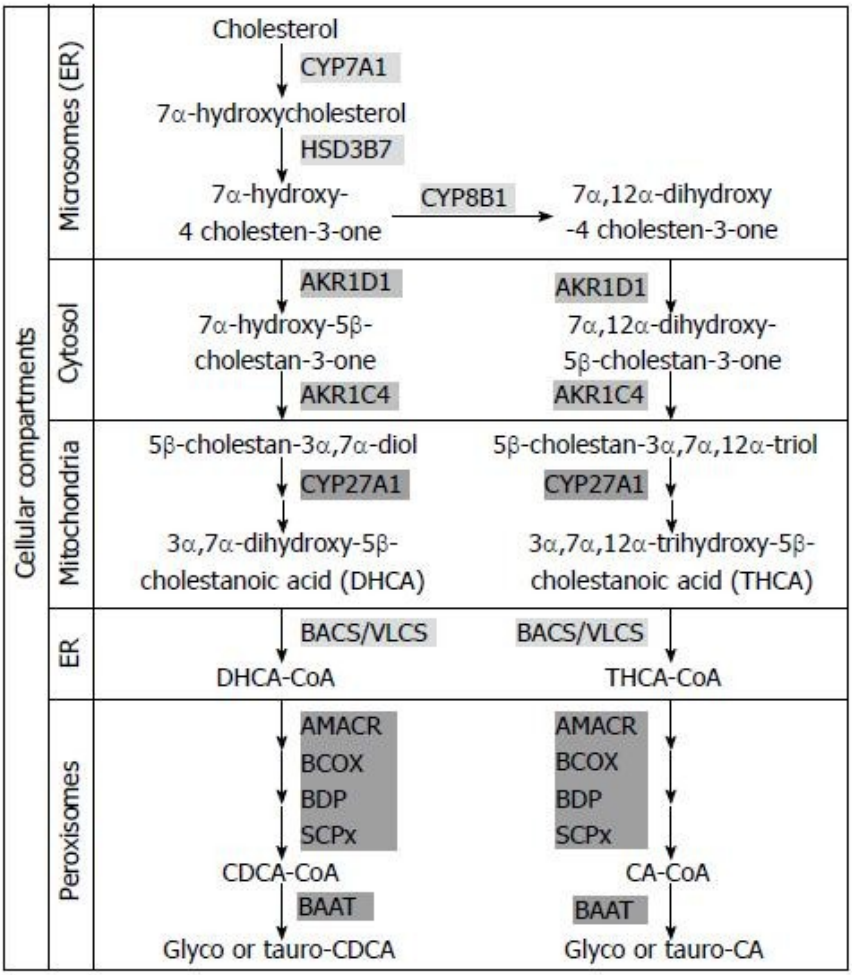


Figure 3.3: Bile acid biosynthesis by the classical neutral pathway [1].

The first step is catalyzed by the microsomal cholesterol 7α-hydroxylase (CP7A1) and consists in the hydroxylation of cholesterol at C-7; this enzyme is localized only in the liver and catalyzes the rate-limiting reaction of the whole pathway.

Another microsomal enzyme, 3β-hydroxy-Δ5-C27-steroid dehydrogenase/isomerase (HSD3B7), converts the 7α-hydroxycholesterol to 7α-hydroxy-4 cholesten-3-one.

The synthesis of CA, but not of CDCA, requires the further hydroxylation at C-12 position of 7α-hydroxy-4 cholesten-3-one, which is performed by another microsomal enzyme, sterol 12α-hydroxylase (CYP8B1).

The two following steps are carried out by cytosolic enzymes, Δ4-3-oxosteroid-5β-reductase (AKR1D1) and 3α-hydroxysteroid dehydrogenase (AKR1C4), which catalyze the reduction of the double bond, giving rise to 5β-cholestan-3α,7α-diol or 5β-

cholestan-3α,7α,12α-triol.
The side chains of these compounds are then oxidized by the mitochondrial sterol 27-hydroxylase (CYP27A1), thus introducing a hydroxyl group to the C-27 position; subsequently, the resulting precursors are oxidized to an aldehyde and then to a carboxylic acid.
The products, 3α,7α-dihydroxy-5β-cholestanoic acid and 3α,7α,12α-trihydroxy-5β-cholestanoic acid, are converted to coenzyme A-esters by either bile acid CoA synthetase (BACS) or very long chain acyl CoA synthetase (VLCS); both these enzymes are localized at the endoplasmic reticulum (ER).
The resulting precursors are then transported into peroxisomes, where their side chain is shortened through β-oxidation by four enzymes, alpha methylacyl-CoA racemase (AMACR), branched-chain acyl CoA oxidase (BCOX), D-bifunctional protein hydratase (BDP) and sterol carrier protein X (SCPx).
The final step is catalyzed by bile acid CoA: amino acid N-acyltransferase (BAAT) and consists in conjugation of the terminal side chain carboxylic acid with the amino acid glycine or taurine; this enzyme is located in peroxisomes, but a minor cytosolic fraction is also present.

The alternative biosynthetic pathway is also called “acidic” pathway, because acidic intermediates are generated; as a matter of fact, side chain oxidation of cholesterol occurs before steroid ring modification.
The first step consists in the oxidation of cholesterol to 27-hydroxycholesterol and it is catalyzed by CYP27A1; the following step leads to the formation of 7α,27-dihydroxycholesterol by oxysterol 7α-hydroxylase (CYP7B1), a microsomal enzyme present only in the acidic pathway.
Since CYP27A1 and CYP7B1 are expressed in various tissues, but only the liver has the required enzymes to complete bile acids biosynthesis, oxidized sterols are transported to the liver, where bile acids are formed. In the acidic pathway, CDCA is the main bile acid produced.
In humans, the relative contribution of the alternative pathway to overall bile acid synthesis is low (about 10%) under physiological conditions, but it may become significant in patients with liver diseases.

There are two more bile acid biosynthesis pathways, one initiated in the liver and the other in the brain, whose first reaction is 25-hydroxylase and 24-hydroxylase, respectively; however, the contribution of these two pathways to overall bile acid production seems minor.

Biotransformation by intestinal bacteria

During their intestinal transit, bile acids are modified by intestinal bacteria. In the small bowel these reactions involve mainly deconjugation and hydroxyl group oxidation; bile acids are very efficiently absorbed in the small bowel and only a small amount (<1 g/d) enters the large bowel.

In humans, the main reactions consist of 7α-dehydroxylation, deconjugation and oxidation and epimerization of hydroxyl groups at C-3, C-7 and C-12 [1]; while 7α-dehydroxylation is performed by a limited number of anaerobic bacteria, deconjugation and oxidation are carried out by a broad spectrum of anaerobes.

From CA and CDCA, dehydroxylation at C-7 position generates the secondary bile acids, deoxycholic acid (DCA) and lithocholic acid (LCA).

Deconjugation consists in enzymatic hydrolysis of the C-24 N-acyl amide bond, thus generating unconjugated bile acids and free glycine and taurine; some unconjugated bile acids are then absorbed by the intestine and return through the portal vein to the liver, where they are reconjugated.

Oxidation and epimerization of the 3-, 7- or 12-hydroxyl groups of bile acids are catalyzed by bacterial hydroxysteroid dehydrogenases; UDCA is generated from epimerization of CDCA.

Dehydroxylation at C-7 position is carried out only on deconjugated bile acids.

In the intestine CA, CDCA and DCA are reabsorbed and transported back to the liver, while most of the LCA is lost in feces; however, a small amount of LCA is circulated back to the liver, where it is sulfo-conjugated at the 3-hydroxy position by sulfotransferase (SULT2A1) and then secreted into bile. Sulfonation is a detoxification mechanism for hydrophobic bile acid, which are rather toxic [2].

Enterohepatic circulation

In humans, bile acids are synthesized in the liver, secreted into intestine during

digestion, reabsorbed and returned to the liver through the portal circulation. About 0.2-0.6 g of bile acids are produced daily by the human liver [1, 2].

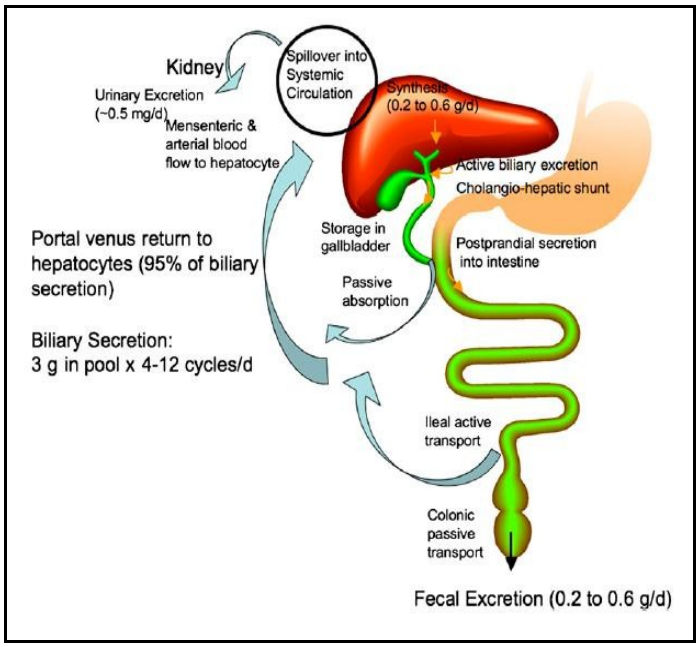


Figure 3.4: Enterohepatic circulation of bile acids [2].

The hepatocyte canalicular bile salt export pump (BSEP) secretes conjugated bile acids into the bile, which is stored in the gallbladder. When their concentration increases in the hepatocyte, bile acids are spilled over into sinusoid blood. Bile acids present in the blood are absorbed in the kidneys by the renal tubules and returned to the liver through the systemic circulation. In the cholangiohepatic shunt, some bile acids secreted in the bile duct are absorbed in the cholangiocytes and then sent back to the hepatocytes.

During digestion, bile acids are secreted into the intestinal tract upon gallbladder contraction; a small amount of bile acids is reabsorbed by passive diffusion in the upper intestine.

After execution of their digestive functions, most bile acids (95%) are captured by active transport in the intestine, especially in the ileum. At the apical membrane of the intestinal epithelial cells, bile acids are recovered in an active process by the apical sodium dependent bile acid transporter (ASBT), while the efflux from intestinal cells across the basal membrane is performed by the heterodimeric organic solute transporter alpha and beta (OSTα-OSTβ).

Bile acids bound to albumin reach the liver mainly through the portal blood, but also through the hepatic artery; subsequently, transport proteins placed at the sinusoidal membrane of hepatocytes recover bile acids. The uptake of conjugated bile acids is performed by the sodium taurocholate co-transport polypeptide (NTCP) in a sodium-dependent manner; sinusoidal bile acid uptake also occurs in a sodium-independent process, by members of the family of organic anion transporting polypeptides (OATP1B1 and OATP1B3).

The bile acid pool (about 3 g) is recycled 4-12 times a day; bile acids lost in feces (0.2-0.6 g) are replaced, to maintain a constant pool, by de novo synthesized bile acids in the liver.

Regulation of bile acids synthesis

Bile acid synthesis is mainly regulated through transcriptional modulation of CYP7A1, the first enzyme of the biosynthetic pathway [1 - 4].

Bile acids exert a negative feed back regulation on the cytochrome P450 enzymes CYP7A1, CYP8B1 and CYP27A1, involved in bile acid biosynthesis, through the farnesoid nuclear receptor (FXR). Hydrophobic bile acids, such as CDCA, can activate the FXR, which in turn induces the expression of the small heterodimer partner (SHP) transcriptional repressor. In particular, with a bile acid as a ligand, the FXR can dimerize with a nuclear hormone receptor, the retinoid X receptor (RXR); the complex FXR/RXR can induce the expression of SHP. SHP negatively regulates other transcription factors, the liver receptor homolog-1 (LRH-1) and the hepatocyte nuclear factor-4α (HNF-4α), which binds to the bile acid response elements (BARE) located in the promoter region of CYP7A1 and CYP8B1 and, consequently, in this way SHP can inhibit bile acid synthesis [5, 6].

Bile acids can inhibit CYP7A1 transcription by another FXR-dependent mechanism, that involves secreted fibroblast growth factor 19 (FGF-19) and its receptor (FGFR4) [7].

Cholesterol regulates its own catabolism to bile acids; in particular, oxysterols, oxidized metabolites of cholesterol, activate the liver X receptor (LXR), which, after dimerization with RXR, regulates positively the CYP7A1 transcription in rat hepatocytes. The LXR/RXR complex, after the binding of oxysterols to LXR, stimulates

the catabolic pathway of cholesterol to bile acids, thus lowering intracellular cholesterol levels and preventing excessive intracellular cholesterol accumulation. However, LXR has little or no effect on human CYP7A1, because the promoter of the human gene lacks an LXR response element [8].

Additionally, CYP7A1 shows diurnal variations, at both mRNA and protein levels. HNF-4α was proved to be required for the maintenance of CYP7A1 diurnal variation [9]. Also the circulating levels of FGF-19 undergo diurnal variation, in synchronicity with changes in CYP7A1 activity [10].

Some drugs can also influence bile acid synthesis; both phenobarbital and the antibiotic rifampicin repress CYP7A1 transcription through the nuclear receptor constitutive androstane receptor (CAR) and pregnane X receptor (PXR), respectively [11, 12].

Defects in cholesterol 7α-hydroxylase

Genetic defects in the CYP7A1 gene were reported to cause a decrease in bile acid production through the classical pathway; however, this decrease is compensated by the activation of the alternative pathway. Patients show an increase in hepatic cholesterol content and adult patients are usually affected by LDL hypercholesterolemia and cholesterol gallstones. However, patients do not usually show evidence of liver disease [13, 14].

3.1.2 Cytochrome P450 and cholesterol metabolism

In mammals, the most important pathways of cholesterol degradation are initiated by cytochrome P450 [15]; these enzymes bind a single heme group and their name derived from the characteristic absorption at 450 nm when reduced and in complex with CO [16, 17].

The cytochrome P450 family (CYP) has been found in all organisms, from bacteria to humans, and their members are involved in two main functional roles. First, they play a protective role against xenobiotics (compounds exogenous to the organism) being involved in their degradation or the attachment of polar handles for solubilization in preparation for excretion; secondly, they are involved in the synthesis of steroid

hormones, hydrophobic vitamin metabolism and the conversion of polyunsaturated fatty acids to biologically active molecules.

In general, P450 cytochrome substrates are hydrophobic or poorly water-soluble, but alcohols, phenols and other organic compounds have been reported to act as substrates.

The most important reaction catalyzed by P450 cytochrome is hydroxylation; these enzymes are called monooxygenases, because one atom of molecular oxygen is incorporated into the substrate and the other into a water molecule.

To perform the enzymatic reaction cytochrome P450 uses electrons donated from reduced nicotinamide adenine dinucleotide phosphate (NADPH). Type I cytochrome P450 is loosely associated to the inner mitochondrial membrane and receives electrons by an electron transport system composed of ferrodoxin (a 14 kDa iron-sulfur protein) and ferrodoxin reductase (a 54 kDa flavoprotein); type II cytochrome P450 is anchored to the inner membrane of the endoplasmic reticulum through an N-terminal hydrophobic segment of about 20 amino acids and receives electrons from cytochrome P450 oxidoreductase (an 82 kDa membrane-bound protein) [18, 19].

In humans, four major cytochrome P45Os use cholesterol as substrate to initiate different degradation pathways.

CYP7A1 is a microsomal enzyme that catalyzes the conversion of cholesterol to 7α-hydroxycholesterol, the first reaction of the classical pathway of bile acid synthesis; it is expressed only in the liver and it catalyzes the rate-limiting reaction of the overall pathway. Under physiological conditions, the human liver removes daily about 0.4-0.6 g of cholesterol.

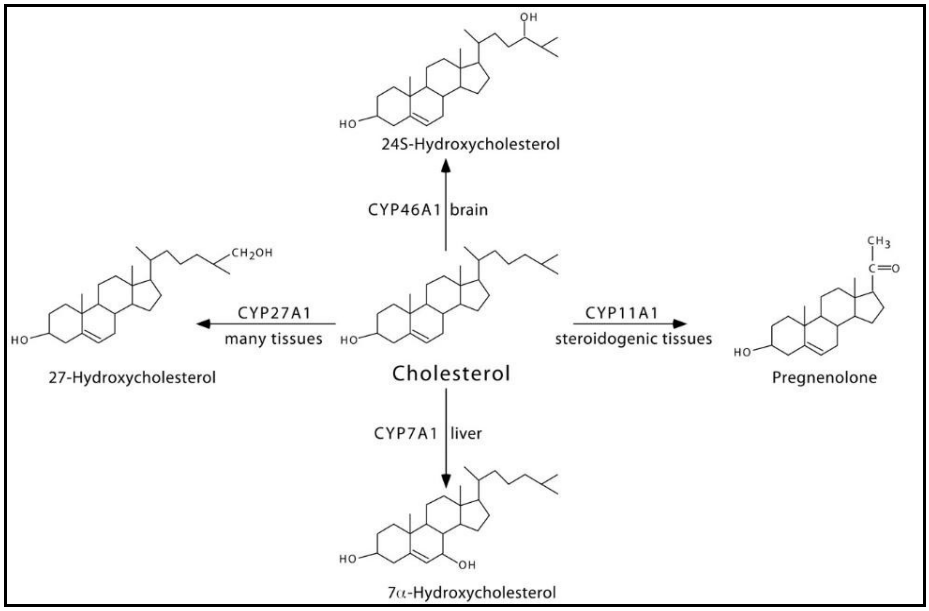


Figure 3.5: Cytochrome P450 involved in cholesterol catabolism [15].

CYP27A1 is a mitochondrial enzyme, expressed in many tissues and it catalyzes the conversion of cholesterol to 27-hydroxycholesterol; this is the first reaction of the alternative pathway of bile acid synthesis, which is initiated in extrahepatic tissues and eliminates daily about 20 mg of cholesterol.

In the liver this enzyme is also required in the classical pathway of bile acid synthesis and in the kidney it can hydroxylate vitamin D_3 [20].

In humans, genetic mutations of the CYP27A1 gene cause a disease called cerebrotendinous xanthomatosis (CTX), characterized by abnormal deposition of cholesterol and cholestan (5α-saturated analog of cholesterol) in the brain and tendons. The disease leads to a reduced level of bile acid synthesis, with a compensatory increase in CYP7A1 activity and a resulting accumulation of 7α-hydroxylated bile acid precursors. Oral supplementation with chenodeoxycholic acid halts CTX progression, by normalizing the activity of CYP7A1 [21].

CYP11A1 is a mitochondrial enzyme that metabolizes cholesterol to pregnenolone; this is the first reaction in steroid hormone biosynthesis and occurs in steroidogenic tissues, such as adrenal glands, ovaries, testis, placenta and brain [18].

About 40-50 mg of cholesterol are daily utilized to produce steroid hormones.

CYP46A1 is a microsomal enzyme expressed in neurons and it is responsible for

cholesterol turnover in the vertebrate central nervous system. The enzyme converts cholesterol to 24-hydroxycholesterol; in contrast to cholesterol, the product of this enzymatic reaction can cross the blood-brain barrier, enter the circulation and be delivered to the liver, where it undergoes further degradation [22].

This mechanism eliminates 6-7 mg of cholesterol daily.

3.2 Materials and Methods

3.2.1 The CYP7A1 clones

The human full-lenght cDNA clone was kindly provided by Dr. D. Russell (University of Texas Southwestern Medical Center).

The plasmid containing the full-length cDNA clone (IMAGE ID 6790722) of zebrafish CYP7A1 was purchased from RZPD. *E. coli* cells harbouring the plasmid vector pCMV-SPORT6.1 were grown in 3 ml of LB medium with 100 µg/ml ampicillin overnight at 37ºC.
Cells were harvested by centrifugation and the plasmid vector was purified using the GenElute™ Plasmid Miniprep kit.

3.2.2 Subcloning of CP7A1 gene into expression vectors

The coding sequence of the human and zebrafish proteins both lacking the N-terminal 25 amino acids were amplified by a PCR reaction using the following primers and parameters.

pQE50, pET28 (human):
Primer forward:
5'-GGTGGTGGATCCAGAAGGCAAACGGGTGAA-3'
Primer reverse:
5'-AATAAGCTTGGAACCACGCGGAACCAGCAAATGCTTGAATTTATA-3'
GGATCC- *BamHI* restriction site
AAGCTT - *HindIII* restriction site
GGAACCACGCGGAACCAG – thrombin cleavage site

pET15(human):

Primer forward:

5'-GGTGGTCATATGAGAAGGCAAACGGGTGAA-3'

Primer reverse:

5'-GGTGGTGGATCCTCACAAATGCTTGAATTT-3'

CATATG - *NdeI* restriction site

GGATCC- *BamHI* restriction site

TCA - stop codon

pET22 (zebrafish):

Primer forward:

5'-GGTGGTCATATGCGCAGAAGACATCCTGCA

Primer reverse:

5'-GGTGGTCTCGAGGAGAGATTTGAGTCTGTA

CATATG - *NdeI* restriction site

CTCGAG - *XhoI* restriction site

Figure 3.6: The primers used to amplify human and zebrafish CYP7A1; the coding sequence is underlined.

STEP	TEMPERATURE	TIME	
Initial Denaturation	95ºC	10 min	
Denaturation	95ºC	1 min	35 cycles
Annealing	56ºC	1 min	
Extension	72ºC	1 min 40 sec	
Final Extension	72ºC	10 min	

Figure 3.7: The PCR parameters.

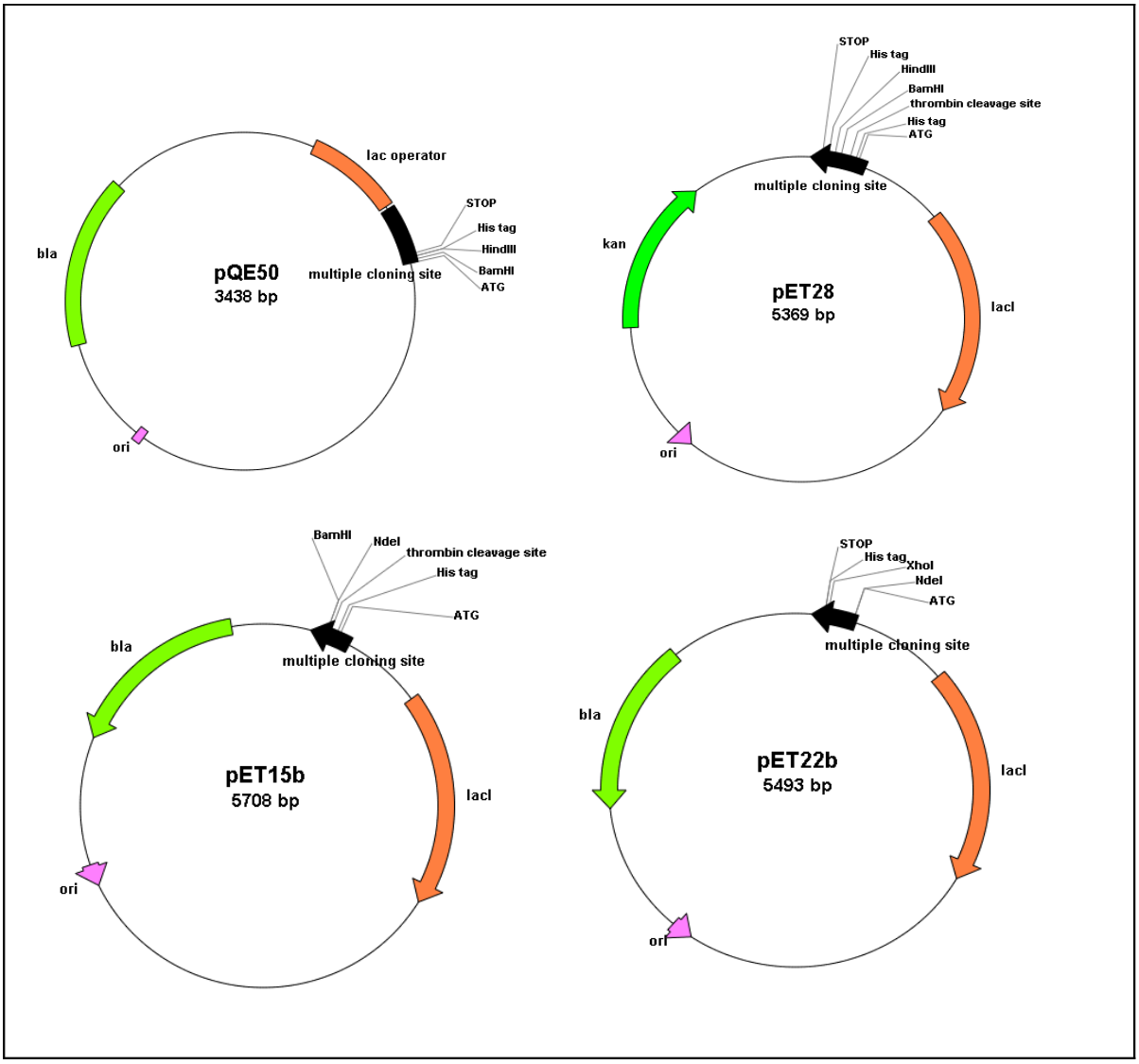


Figure 3.8: Schematic representation of the employed plasmids.

To perform protein detection and purification, the human CYP7A1 coding sequence was cloned into vectors that allow to insert a 6 histidine tag at the N-terminus (pET15), at the C-terminus (pQE50), and at both N- and C-termini of the protein (pET28). A thrombin cleavage site was also inserted to remove the tag after protein purification.
The coding sequence of the zebrafish CYP7A1 was inserted into pET22, that fuses the 6 histidine tag at the C-terminus of the target protein.

The amplified sequences were quantified on a 0.8% agarose TBE gel and digested with restriction enzymes; subsequently, the fragments were ligated into the *BamHI*/*HindIII*

site of the pQE50 and of pET28 vectors, into the *NdeI*/*BamHI* site of the pET15b plasmid and into *NdeI/XhoI* site of pET22.

E. coli XL1 Blue competent cells were then transformed with the ligated plasmids by heat shock and the cells were plated on a LB-agar plate containing 100 μg/ml ampicillin or 50 μg/ml kanamycin; the cells were then grown overnight at 37ºC. The colonies were selected by colony PCR and then by digestion of the purified plasmids; at the end, the vectors were checked by automated DNA sequencing.

3.2.3 Expression screening of the recombinant protein

The recombinant plasmids were used to transform *E. coli* expression strains by heat shock and the resulting colonies were tested for protein expression; hAICDA-pQE50 was inserted in *E. coli* SG13009, BL21 (DE3) and C41 (DE3), while hAICDA-pET15, hAICDA-pET28 and zAICDA-pET22 were used to transform *E. coli* BL21 (DE3), C41(DE3), Origami, ER2566, BL21 CodonPlus and Rosetta (DE3) strains and protein expression was tested by Western blot analysis.

A single colony was inoculated in 3 ml of LB medium, grown at 37ºC adding 100 μg/ml ampicillin or 50 μg/ml kanamycin and then incubated at 37ºC for 4 hours and at 20ºC and 28ºC overnight after induction with 0.25 mM IPTG.

After centrifugation, the pellet was resuspended in 2 ml of buffer A (20 mM TRIS-HCl pH 7.5, 0.5 M NaCl and 0.02% NaN_3), cells were disrupted by sonication and centrifuged at 9,000 rpm for 10 minutes at 4ºC. Samples from the total lysate and from the soluble fraction were loaded onto an SDS-PAGE gel (12% polyacrylamide resolving gel and 4% polyacrylamide stacking gel). Protein expression was evaluated by Western blot analysis using an anti-histidine tag monoclonal antibody.

3.3 Results and discussion

The human CYP7A1 gene was subcloned into pQE50, pET15 and pET28 plasmids, while the zebrafish CYP7A1 gene was inserted into pET22 vector; protein expression was tested by Western blot analysis. The human CYP7A1 gene encodes a 504 amino acid protein with a molecular weight of 57.6 kDa, while the zebrafish CYP7A1 protein comprises 512 residues with a molecular weight of 58.7 kDa; the sequence identity between the two proteins is 64%.

Using the pQE50 plasmid, protein expression was observed with *E. coli* SG13009 only in the total fraction, while *E. coli* BL21 (DE3) and C41 (DE3) did not show any bands, even in the total fraction.

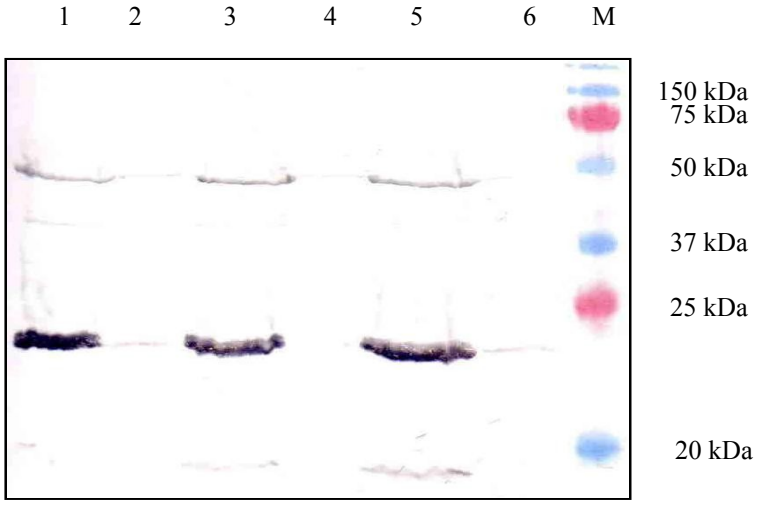


Figure 3.9: Western blot analysis of human CYP7A1-pEQ50 *E. coli* SG13009. The total and the soluble fractions are shown for different induction conditions: 37ºC for 4 hours (lane 1 and 2), 28ºC overnight (lane 3 and 4) and 20ºC overnight (lane 5 and 6).

Two bands are clearly visible in the total fractions of all the tested induction conditions; the band at high molecular weight shows the expected molecular weight, but has a much lower intensity than the other band at low molecular weight, which is likely due to protein degradation.

With pET28, several *E. coli* expression strains were tested: BL21 (DE3), C41 (DE3), BL21 CodonPlus, Rosetta (DE3) and ER2566.

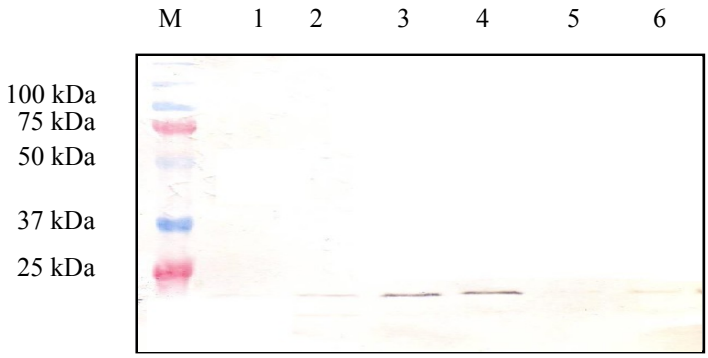


Figure 3.10: Western blot analysis of human CYP7A1-pET28. The soluble fraction of *E. coli* BL21 (DE3) are shown in lane 1 (induction at 37ºC for 4 hours), lane 2 (induction at 28ºC overnight) and lane 3 (induction at 20ºC overnight), while the soluble fractions of *E. coli* C41 (DE3) are shown in lane 4 (induction at 20ºC overnight), lane 5 (induction at 28ºC overnight) and lane 6 (induction at 37ºC for 4 hours).

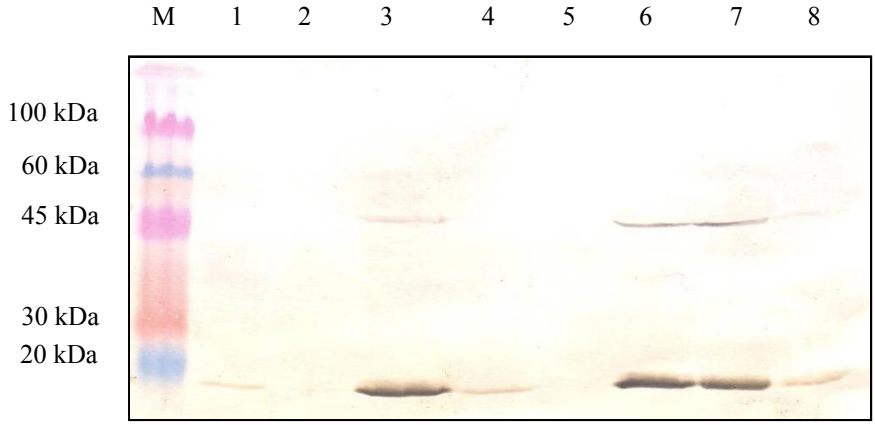


Figure 3.11: Western blot analysis of human CYP7A1-pET28.
The soluble fractions of *E. coli* BL21 CodonPlus are shown in lane 1 (induction at 20ºC overnight) and lane 2 (induction at 37ºC for 4 hours). The fractions of *E. coli* Rosetta (DE3) are shown in lane 3, 4 (total and soluble fraction with 20ºC overnight induction) and 5 (soluble fraction with induction at 37ºC for 4 hours). *E. coli* ER2566 samples are shown in lane 6, 7 (total and soluble fraction with 20ºC overnight induction) and 8 (soluble fraction with induction at 37ºC for 4 hours).

The best protein expression was observed with *E. coli* ER2566 with induction at 20ºC overnight; in particular, while a major band is visible at a lower molecular weight, a faint band at the expected molecular weight can be observed in both the total and the soluble fractions.

Target protein expression was also observed with pET15 in *E. coli* BL21 (DE3) and C41 (DE3).

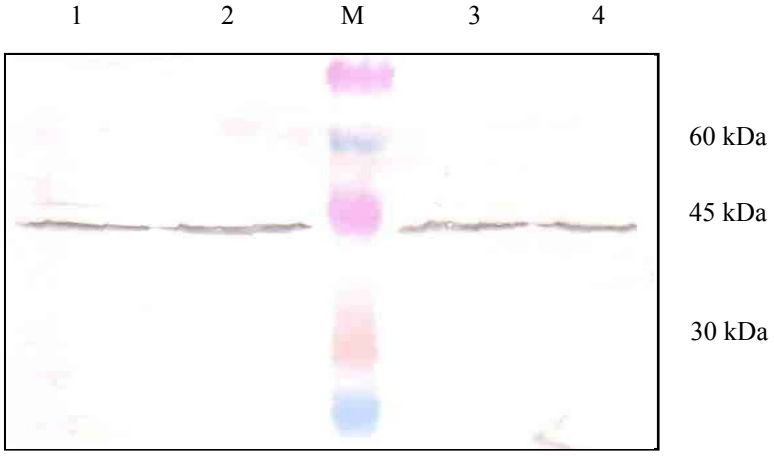


Figure 3.12: Western blot analysis of human CYP7A1-pET15. The *E. coli* BL21 (DE3) soluble fractions are shown for different induction conditions: 37ºC for 4 hours (lane 1) and 20ºC overnight (lane 2); the corresponding samples with *E. coli* C41 (DE3) strain are shown in lane 3 and 4.

A band, which corresponds to a lower molecular weight than expected, is visible in the soluble fractions of the tested conditions.

Finally, the zebrafish CYP7A1 gene was subcloned into the pET22 vector and protein expression was tested in different *E. coli* expression strains and induction conditions, but target protein expression was not observed in any tested samples.

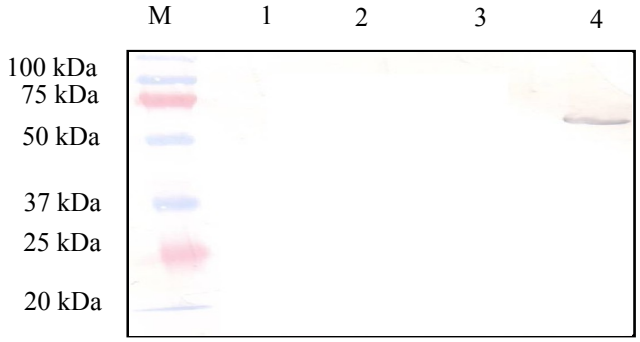


Figure 3.13: Western blot analysis of zebrafish CYP7A1-pET22 in *E. coli* BL21 (DE3). The soluble fractions are shown for different induction conditions: lane 1 (20ºC overnight), lane 2 (28ºC overnight) and lane 3 (37ºC for 4 hours);
a total fraction is shown in lane 4, referred to induction at 37ºC for 4 hours.

3.4 Conclusions and prospectives

Human CYP7A1 was expressed in the prokaryotic system *E. coli* using different vectors and expression conditions. Unfortunately, the protein was expressed only in the total form, for example in pQE50, or showed a lower molecular weight than expected, probably due to degradation.
The best expression conditions were obtained with *E. coli* ER2566 when induction was performed at 20ºC overnight; though most protein was degraded, a small band was still visible in the soluble fraction. Consequently, 1 liter LB culture was grown and induced and purification attempts were carried out by IMAC chromatography; the eluted fractions were cecked by SDS-PAGE gel and Western blot analysis, but no protein was detected. This is probably due to the very low level of expression of the target protein.

To overcome the problems encountered with protein expression in *E. coli*, different expression systems should be tested, such as the eukaryotic systems *Pichia pastoris*, *Saccharomyces cerevisiae* and insect cells transfected with recombinant baculovirus.

A different possible strategy consists in the study of the homologous protein from a different organism; the zebrafish CYP7A1 was chosen and its expression tested in *E. coli*, but no significant improvement was obtained.

To increase protein expression, it may be useful to add δ-aminolevulinic acid (δALA), a heme precursor, to the bacterial culture just before IPTG induction; this compound has been used with success in cytochrome P450 expression in *E. coli* [23]. Presumably enhancing heme biosynthesis may increase protein expression and stability.

3.5 References

[1] Monte M.J., Marin J.J., Antelo A., Vazquez-Tato J. (2009), Bile acids: chemistry, physiology, and pathophysiology. *World J. Gastroenterol.* **15**: 804-816.

[2] Chiang J.Y. (2009), Bile acids: regulation of synthesis. *J. Lipid Res.* **50**: 1955-1966.

[3] Redinger R.N. (2003), The coming of age of our understanding of the enterohepatic circulation of bile salts. *Am. J. Surg.* **185**: 168-172.

[4] Gilardi F., Mitro N., Godio C., Scotti E., Caruso D., Crestani M., De Fabiani E. (2007), The pharmacological exploitation of cholesterol 7alpha-hydroxylase, the key enzyme in bile acid synthesis: from binding resins to chromatin remodelling to reduce plasma cholesterol. *Pharmacol. Ther.* **116**: 449-472.

[5] Pandak W.M., Vlahcevic Z.R., Heuman D.M., Redford K.S., Chiang J.Y., Hylemon P.B. (1994), Effects of different bile salts on steady-state mRNA levels and transcriptional activity of cholesterol 7 alpha-hydroxylase. *Hepatology* **19**: 941-947.

[6] Lew J.L., Zhao A., Yu J., Huang L., De Pedro N., Peláez F., Wright S.D., Cui J. (2004), The farnesoid X receptor controls gene expression in a ligand- and promoter-selective fashion. *J. Biol. Chem.* **279**: 8856-8861.

[7] Holt J.A., Luo G., Billin A.N., Bisi J., McNeill Y.Y., Kozarsky K.F., Donahee M., Wang D.Y., Mansfield T.A., Kliewer S.A., Goodwin B., Jones S.A. (2003), Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis. *Genes Dev.* **17**: 1581-1591.

[8] Goodwin B., Watson M.A., Kim H., Miao J., Kemper J.K., Kliewer S.A. (2003), Differential regulation of rat and human CYP7A1 by the nuclear oxysterol receptor liver X receptor-alpha. *Mol. Endocrinol.* **17**: 386-394.

[9] Inoue Y., Yu A.M., Yim S.H., Ma X., Krausz K.W., Inoue J., Xiang C.C., Brownstein M.J., Eggertsen G., Björkhem I., Gonzalez F.J. (2006), Regulation of bile acid biosynthesis by hepatocyte nuclear factor 4alpha. *J. Lipid Res.* **47**: 215-227.

[10] Lundåsen T., Gälman C., Angelin B., Rudling M. (2006), Circulating intestinal fibroblast growth factor 19 has a pronounced diurnal variation and modulates hepatic bile acid synthesis in man. *J. Intern. Med.* **260**: 530-536.

[11] Miao J., Fang S., Bae Y., Kemper J.K. (2006), Functional inhibitory cross-talk between constitutive androstane receptor and hepatic nuclear factor-4 in hepatic lipid/glucose metabolism is mediated by competition for binding to the DR1 motif and to the common coactivators, GRIP-1 and PGC-1alpha. *J. Biol. Chem.* **281**: 14537-14546.

[12] Li T., Chiang J.Y. (2005), Mechanism of rifampicin and pregnane X receptor inhibition of human cholesterol 7 alpha-hydroxylase gene transcription. *Am. J. Physiol. Gastrointest. Liver Physiol.* **288**: G74-84.

[13] Pullinger C.R., Eng C., Salen G., Shefer S., Batta A.K., Erickson S.K., Verhagen A., Rivera C.R., Mulvihill S.J., Malloy M.J., Kane J.P. (2002), Human cholesterol 7alpha-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. *J. Clin. Invest.* **110**: 109-117.

[14] Beigneux A., Hofmann A.F., Young S.G. (2002), Human CYP7A1 deficiency: progress and enigmas. *J. Clin. Invest.* **110**: 29-31.

[15] Pikuleva I.A. (2006), Cholesterol-metabolizing cytochromes P450. *Drug Metab. Dispos.* **34**: 513-520.

[16] Denisov I.G., Makris T.M., Sligar S.G., Schlichting I. (2005), Structure and chemistry of cytochrome P450. *Chem. Rev.* **105**: 2253-2277.

[17] Johnson E.F., Stout C.D. (2005), Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.* **338**: 331-336.

[18] Guo I.C., Shih M.C., Lan H.C., Hsu N.C., Hu M.C., Chung B.C. (2007), Transcriptional regulation of human CYP11A1 in gonads and adrenals. *J. Biomed. Sci.* **14**: 509-515.

[19] Nakayama K., Puchkaev A., Pikuleva I.A. (2001), Membrane binding and substrate access merge in cytochrome P450 7A1, a key enzyme in degradation of cholesterol. *J. Biol. Chem.* **276**: 31459-31465.

[20] Araya Z., Hosseinpour F., Bodin K., Wikvall K. (2003), Metabolism of 25-hydroxyvitamin D3 by microsomal and mitochondrial vitamin D3 25-hydroxylases (CYP2D25 and CYP27A1): a novel reaction by CYP27A1. *Biochim. Biophys. Acta* **1632**: 40-47.

[21] Björkhem I., Hansson M. (2010), Cerebrotendinous xanthomatosis: an inborn error in bile acid synthesis with defined mutations but still a challenge. *Biochem. Biophys. Res. Commun.* **396**: 46-49.

[22] Russell D.W., Halford R.W., Ramirez D.M., Shah R., Kotti T. (2009), Cholesterol 24-hydroxylase: an enzyme of cholesterol turnover in the brain. *Annu. Rev. Biochem.* **78**: 1017-1040.

[23] White M.A., Mast N., Bjorkhem I., Johnson E.F., Stout C.D., Pikuleva I.A. (2008), Use of complementary cation and anion heavy-atom salt derivatives to solve the structure of cytochrome P450 46A1. *Acta Crystallogr D Biol. Crystallogr.* **64**: 487-495.

4. Inhibitors of Methionine aminopeptidase 1

4.1 Introduction

4.1.1 Proteolytic removal of N-terminal methionine

Protein synthesis starts at an AUG codon, which is translated as the amino acid methionine. This initiator methionine is cotranslationally removed from nascent proteins by methionine amino peptidases (MetAPs). Between 55 and 70 % of proteins are subject to the proteolytic removal of the N-terminal methionine. This process is essential for the maturation of many proteins. The physical size of the second amino acid determines the substrate specificity of MetAPs. The N-terminal methionine will only be cleaved if the second residue is uncharged and small, leaving a no bulky N-terminal residue such as Ala, Cys, Gly, Pro, Ser, Thr or Val, while the initiator methionine will be retained with remaining residues with larger side chains [1].

In prokaryotes, mitochondria and chloroplasts protein synthesis starts with a formylated methionine residue. A dedicated enzyme, $MetRNA^{fMet}$ transformylase, adds a one-carbon unit to the Met-tRNA; this reaction is thought to facilitate the recognition of translation factors, thus increasing the rate of protein synthesis. The formyl group is cleaved from most nascent proteins by peptide deformylase (PDF), thus facilitating the action of MetAP [2].

The enzymatic removal of N-terminal methionines occurs early in translation and is completed before the nascent protein is fully synthesized. Indeed, in mature proteins the N-terminal amino acid may be buried within the three-dimensional structure of the protein or not available for enzymatic cleavage due to protein folding or location. The proteolytic removal of the initiator methionine is a prerequisite for the control of protein half-life, protein sub cellular localization and posttranslational modifications.

4.1.2 Methionine amino peptidases

While at least two genes are known in eukaryotes, most, but not all, prokaryotes have a

single gene encoding methionine amino peptidase. However, prokaryotes with more than one methionine amino peptidase have been found; for example, the *Plasmodium falciparum* genome encodes four MetAPs [3]. This family of enzymes is evolutionally conserved and essential, as demonstrated by knock out of the genes in both bacteria and yeasts [4, 5]. For example, in the budding yeast *Saccharomyces cerevisiae* two genes are present in the genome and a single deletion results in a slow-growth phenotype compared with the wild-type strain, but the disruption of both genes is lethal [6].

There are two classes of MetAPs. The first class (MetAP1s) occurs in Eubacteria and in the cytoplasm and organelles of eukaryotes, whereas the second class (MetAP2s) has been found in Archea and in the cytoplasm of eukaryotes. MetAP1 is further divided into MetAP1a, MetAP1b and MetAP1c. MetAP1b is present in eukaryotes, while MetAP1a and MetAP1c are present in prokaryotes.

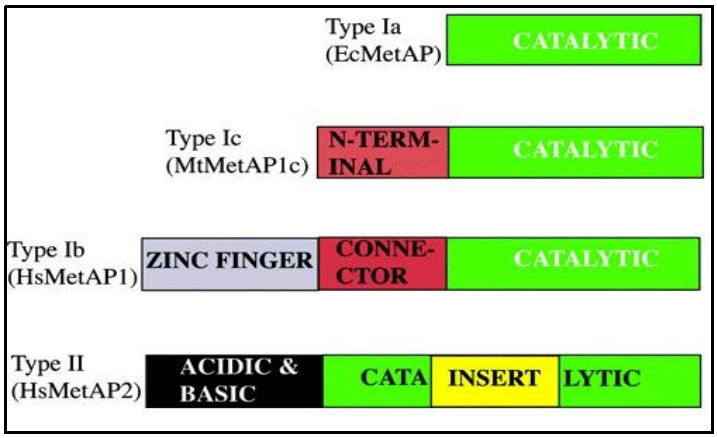


Figure 4.1: Domain organization of MetAPs. Enzymes named in parentheses are from *Escherichia coli, Mycobacterium tuberculosis* and *Homo sapiens* [7].

The main difference between MetAP1 and MetAP2 is a 60-amino acid insertion in the latter. MetAP1b and MetAP1c have an N-terminal extension, whereas no such extension is present in MetAP1a. In MetAP1b this extension contains two zinc finger motifs and a connector region. Deletion of the zinc finger domain of yeast MetAP1 has no effect on enzymatic activity in vitro, but determines a slow growth phenotype in vivo; consequently, the zinc finger domain is thought to be important for the function of the protein in vivo, being probably involved in the interaction with the ribosome [8]. Eukaryotic MetAP2s have an N-terminal extension not required for enzymatic activity

and found disorganized in the crystal structure of the human enzyme [9]. MetAP2s present in Archea contain only catalytic and insertion domains.

4.1.3 Human MetAP1

MetAPs are metalloproteases that require divalent cations; in vitro activity has been observed in the presence of several divalent metals such as Fe^{2+}, Co^{2+}, Mn^{2+} and Zn^{2+}. Generally prokaryotic MetAPs are thought to be iron enzymes. However, human MetAP2 seems to be a manganese enzyme, while the nature of the metal cation remains elusive for human MetAP1. MetAPs have five conserved amino acids involved in metal binding (1 His, 2 Asp and 2 Glu) and similar topology of the active site, including additional conserved histidine and glutamate residues.

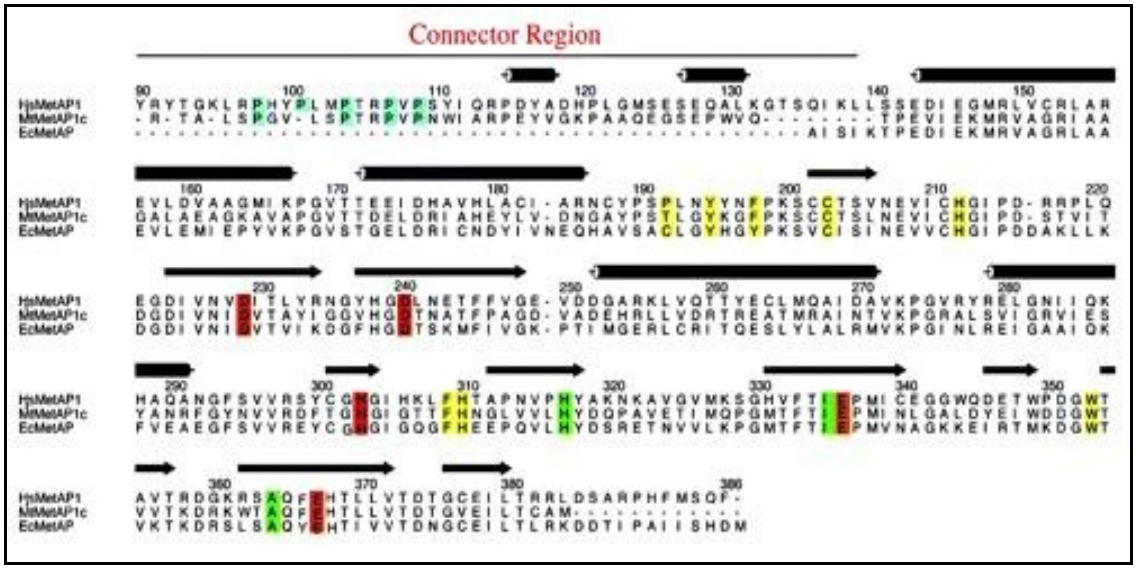


Figure 4.2: Sequence alignment of *Hs*MetAP1, *Mt*MetAP1c and *Ec*MetAP.
The prolines residues in the connector region are shown in blue,
amino acids that coordinate metals are shown in red,
residues present in S1 and S1' substrate -binding subsites are shown in yellow and green [7].

The three-dimensional structure of a truncated version (Δ 1 – 80), which lacks the zinc finger motifs, of human MetAP1 was determined by X – ray diffraction [7].

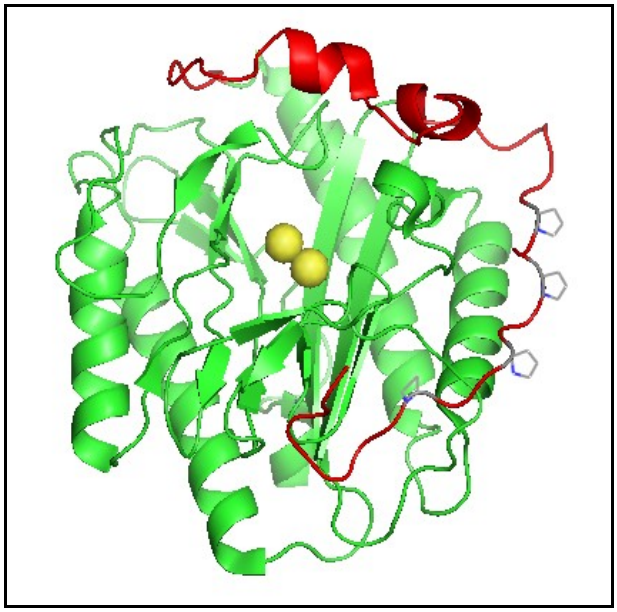


Figure 4.3: Crystal structure of human tMetAP1. In red N-terminal extension, in green catalytic domain and in yellow cobalt ions [7]. The figure was generated using the program PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). (PDB ID 2B3K)

The holoenzyme contains two cobalt atoms in the active site. Apart from small differences in the conformations of the amino acids involved in interactions with metal cations, the apo and the holo forms are essentially identical. In the presence of an excess of cobalt ions, a third cobalt atom is bound to the protein, in two alternative sites; the holo and the trimetallo structures are essentially identical.

Three distinct Pro - X - X - Pro motifs are present in the N-terminal region of human MetAP1 (residues 98, 101, 104 and 107); they are thought to participate in binding to the ribosome.

Both the full-length and the truncated protein can be activated by Co^{2+}, but neither protein is activated by Mn^{2+} or Ni^{2+}. Only the full-length protein can be activated by a high concentration of Zn^{2+}, while neither protein is activated at low concentrations of Zn^{2+}.

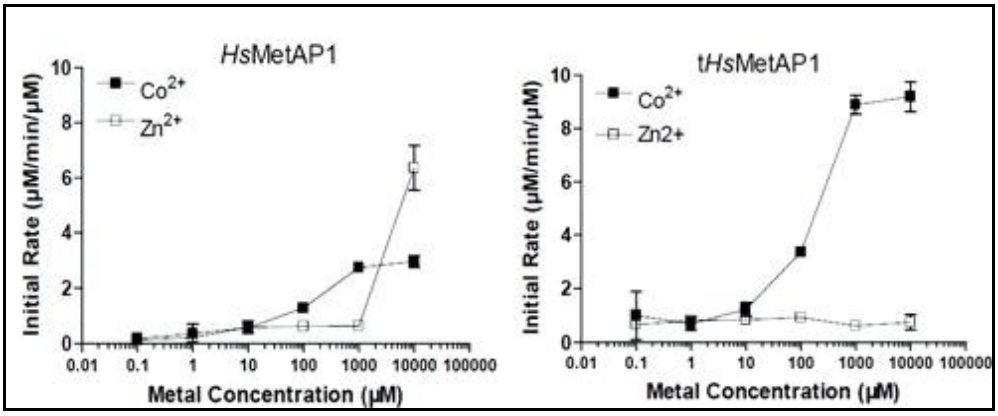


Figure 4.4: Activation of full-length MetAP1 and truncated MetAP1 by divalent cations Co^{2+} (filled squares) and Zn^{2+} open squares) [7].

The Michaelis constant (K_m) is lower for the wild type enzyme. Truncated MetAP1 shows a much higher value of turn over rate (K_{cat}). The rate constant for the enzymatic reaction (K_{cat}/K_m) is higher for the truncated protein, indicating that tMetAP1 can process the substrate more efficiently than the wild type protein.

	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m ($M^{-1}\,min^{-1}$)
t*Hs*MetAP1	1.8 ± 0.13	25.5 ± 0.9	14 167
*Hs*MetAP1	0.74 ± 0.08	6.7 ± 0.2	9114

Table 4.1: Kinetic constants of truncated and full-length human MetAP1 [7].

Since the truncated protein still shows enzymatic activity, the N-terminal region of the protein (1- 80 residues) seems not to be essential for catalytic activity.

4.1.4 MetAP inhibitors and their therapeutic use

Given that the N-terminal methionine excision is considered an essential mechanism conserved both in prokaryotes and in eukaryotes, MetAPs have been suggested to be good candidates for drug design; inhibitors of MetAPs are under development for a variety of therapies, including anti-cancer, anti-infective and anti-arthritis drugs.

Fumagillin and its anti-angiogenesis activity

Human MetAP2 has been identified as the primary target of fumagillin, a potent inhibitor of angiogenesis; this natural compound is a sesquiterpene diepoxide, and it is produced by the fungus *Aspergillus fumigatus*. Fumagillin and its derivatives, such as ovalicin and TNP-470, covalently bind and inactivate MetAP2. Human MetAP2 has been suggested to play an important role in endothelial cell proliferation and to mediate inhibition of endothelial cells by fumigillin and related analogs.

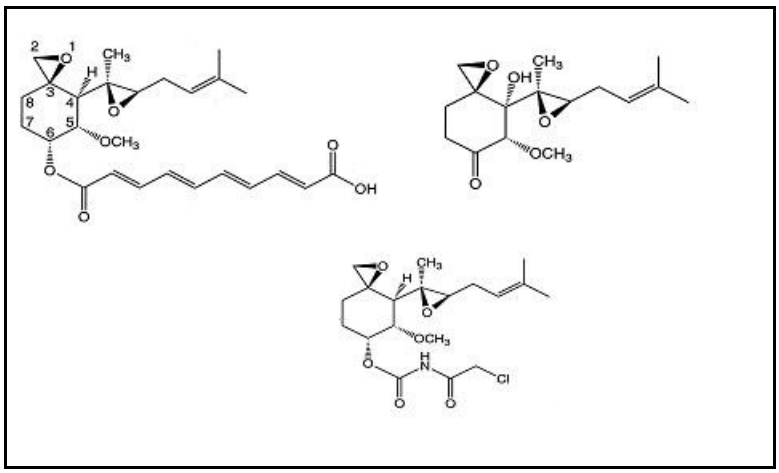


Figure 4.5: Chemical structures of fumagillin, ovalicin and TNP-470, respectively [10].

The common core structure of this class of inhibitors is a cyclohexane ring substituted with an epoxide at C2-C3, an epoxide-containing hydrophobic side chain at C4 and a methoxy group at C5. Variations are possible at C4 and C6 positions. Fumagillin contains a monoester of decatetraenedioic acid at C6, while ovalicin has a keto group at that position; on the other hand, TNP-470 contains a chloro-acetylcarbamoyl group at C6 position. The irreversible inhibition is due to the covalent bond formed by fumagillin with an active site histidine residue of MetAP2, which is involved in coordination of the

catalytic metal cation.

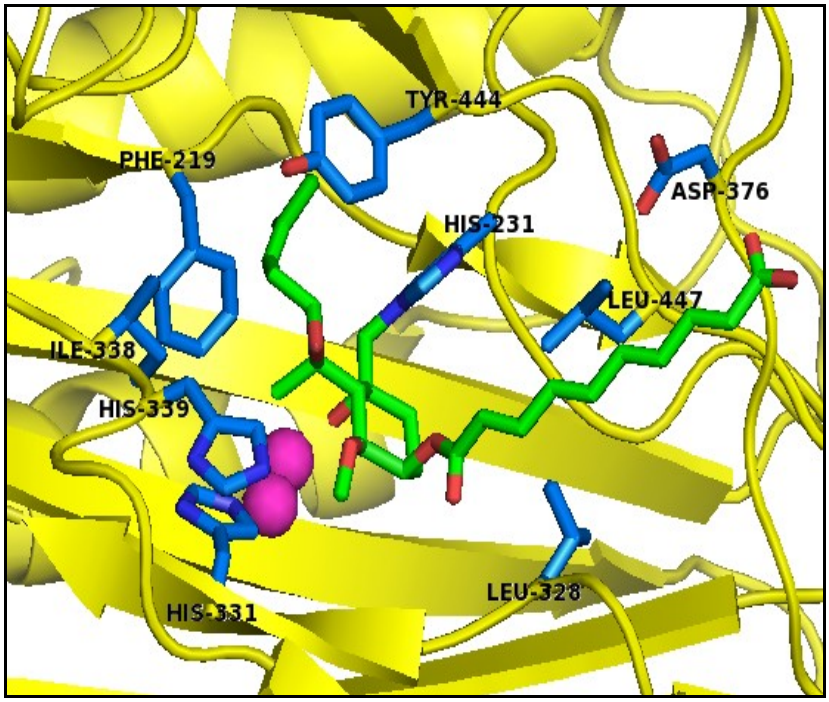


Figure 4.6: Interactions in the active site of human MetAP2 and fumagillin.
In yellow MetAP2, in blue inhibitor and in magenta cobalt ions [9]. (PDB ID 1BOA)

Fumagillin is covalently attached via C2 to His-231, a conserved residue in methionine aminopeptidase family. The compound completely occludes the active site of the enzyme, thus preventing the binding of the substrate. The C4 hydrophobic side chain interacts with several residues of the putative methionine binding pocket (Phe-219, His-331, Ile-338, His-339 and Tyr-444); the extended substituent at C6 position contacts Leu-328, Leu-447 and Asp-376.

Fumagillin and its analogs are potent anti-angiogenesis agents and they are able to prevent the vascularization and the metastasis of tumors. For example, TNP-470 was tested in clinical trials against several cancers, including Kaposi's sarcoma, retinoblastomas, hemangiomas, renal carcinoma, breast, brain and prostate cancers [11]. Fumagillin and its analogs can inhibit MetAP2 with more than a million-fold higher affinity than MetAP1. The structure of the complex between truncated MetAP1 (Δ 1-80) and ovalicin, an analog of fumagillin, was determined by X-ray diffraction [12].

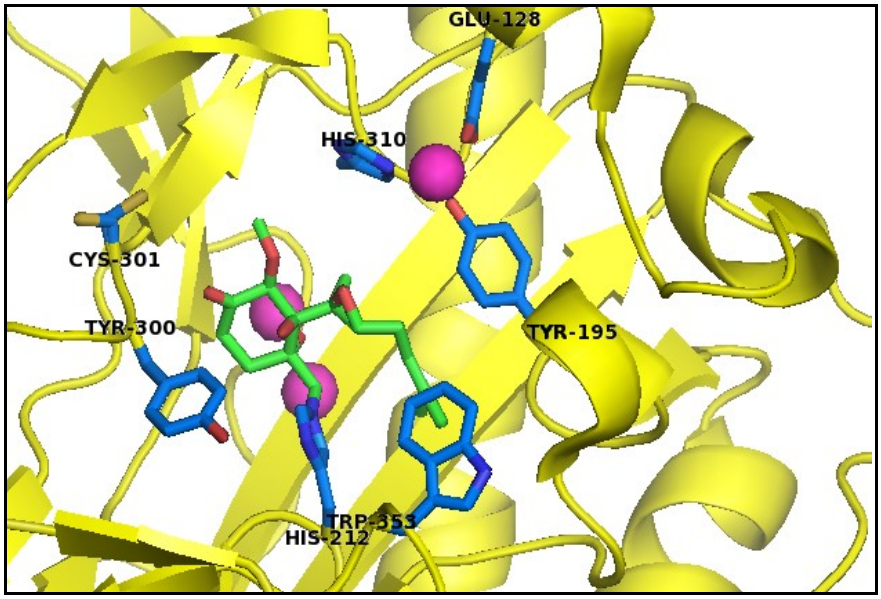


Figure 4.7: Crystal structure of tMetAP1 and ovalicin [12]. (PDB ID 2GZ5)

For both MetAP1 and MetAP2 the inhibitor makes a covalent adduct with the corresponding histidine residues. The lower affinity of the compound for MetAP1 is due to the small size of the active site of the enzyme; indeed, several active site residues move "outward" in order to make room for the inhibitor. Another contributing factor is a different orientation of the side chain of His-310 of MetAP1 compared with His-339 in MetAP2. In the new position His-310 coordinates a third cobalt ion that may contribute to the binding of ovalicin to MetAP1 and thus may influence the affinity of this class of inhibitors toward the different MetAPs.

Bengamides as inhibitors of both human MetAP1 and MetAP2

Both human MetAPs were identified as the binding targets of bengamides, a class of compounds isolated from marine sponges; they inhibit the growth of human tumors at concentrations in the low nanomolar range.

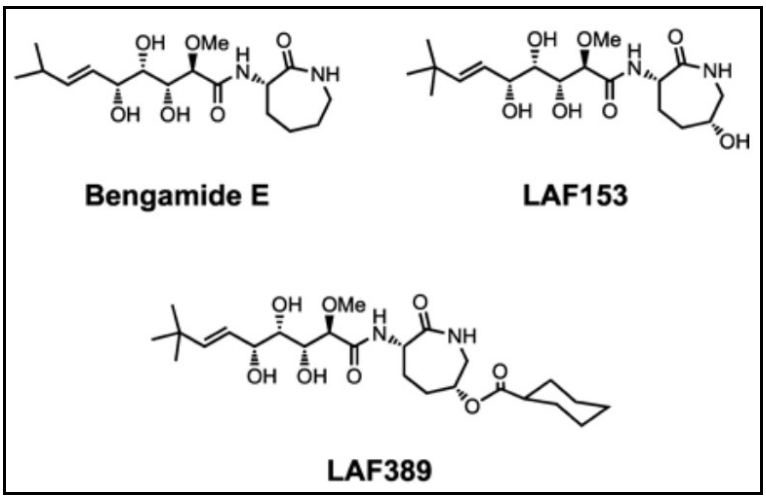


Figure 4.8: Bengamide E and two synthetic analogs (LAF153 and LAF389) [2].

Bengamides bind and inhibit both human MetAP1 and MetAP2 with similar K_d values for both enzymes. The structure of MetAp2 co-crystallized with LAF153 was determined [13]. Unlike fumagillin, which binds covalently a histidine of the active site, bengamides bind in a manner that mimics the expected polipeptide substrate. The hydrophobic alkenyl linkage of LAF153 with attached *tert*-butyl alcohol mimics the methionine side chain and occupies the deeply buried P1 pocket, which is thought to recognize the N-terminal methionine. The hydroxymethyl group of the inhibitor extends into the hydrophobic P1' pocket, which is thought to accommodate the penultimate residue. The caprolactam ring of the inhibitor is coordinated in the P2' region, a solvent-exposed surface.

The central hydroxyl groups of LAF153 coordinate the two cobalt ions present in the active site.

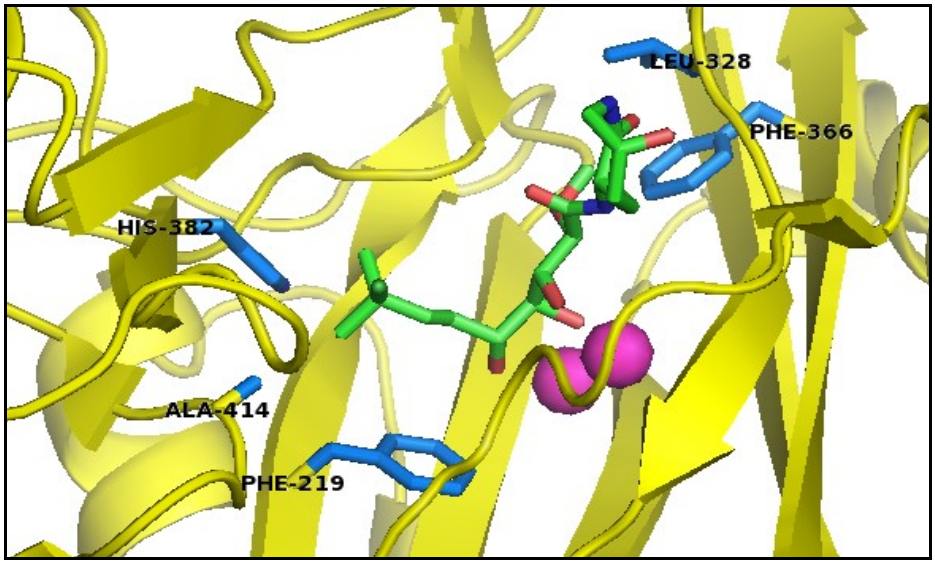
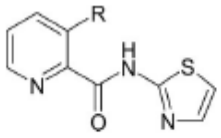


Figure 4.9: Structure of MetAP2 in complex with LAF153 [13]. (PDB ID 1QZY)

This mode of binding of bengamides is completely different from that of the previously known inhibitor fumagillin. Indeed, the latter forms a covalent binding with a histidine in the active site and makes only modest contacts with the dinuclear metal center.
Bengamides are not specific only for MetAP2, but they have similar effects as fumagallin in vivo, with a strong inhibitory activity in endothelial cells.

Pyridine-2-carboxylic acids as a new class of inhibitors specific for human MetAP1

The first class of inhibitors specific only for MetAP1 was that of pyridine-2-carboxylic acids [14].
Compounds of this class were tested in vitro against MetAP1 and MetAP2, in the presence of Co^{2+}.



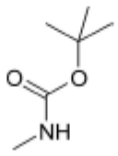
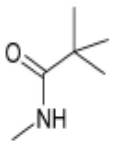
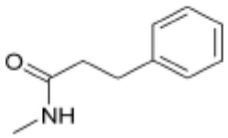
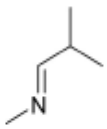
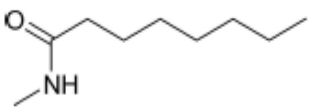
Compound	R	IC_{50}, μM hMetAP1	IC_{50}, μM hMetAP2	Ratio MetAP1/MetAP2	IC_{50}, μM HeLa	IC_{50}, μM HT1080
1		1.5 ± 0.2	>300	<0.005	0.6 ± 0.1	0.3 ± 0.1
2		2.9 ± 0.1	>500	<0.006	3.9 ± 0.2	2.5 ± 0.3
3		3.5 ± 0.2	>500	<0.007	5.1 ± 0.3	4.3 ± 0.1
4		6.4 ± 0.6	>1,000	<0.006	8.9 ± 0.5	9.5 ± 0.8
5		>500	>500	N.A.	>50	>50

Table 4.2: Structures of pyridine-2-carboxylic acids and their inhibition of MetAP1, MetAP2 and cell proliferation [14].

The compounds tested were able to potently inhibit MetAP1 activity with IC_{50} values in the low micromolar range, while they have no effects on MetAP2 activity. For MetAP2 the experiments were repeated in the presence of Mn^{2+}, which is the suggested physiological metal ion for this enzyme, and no inhibition was reported.

The inhibitory effect of pyridine-2-carboxylic acids was tested on cell proliferation using a [^{3}H] thymidine incorporation assay. The IC_{50} values obtained were in the low micromolar range, with both cells tested (Hela and HT-1080). The crystal structure of truncated MetAP1 co-crystallized with inhibitor 1 and 2 was determined [14].

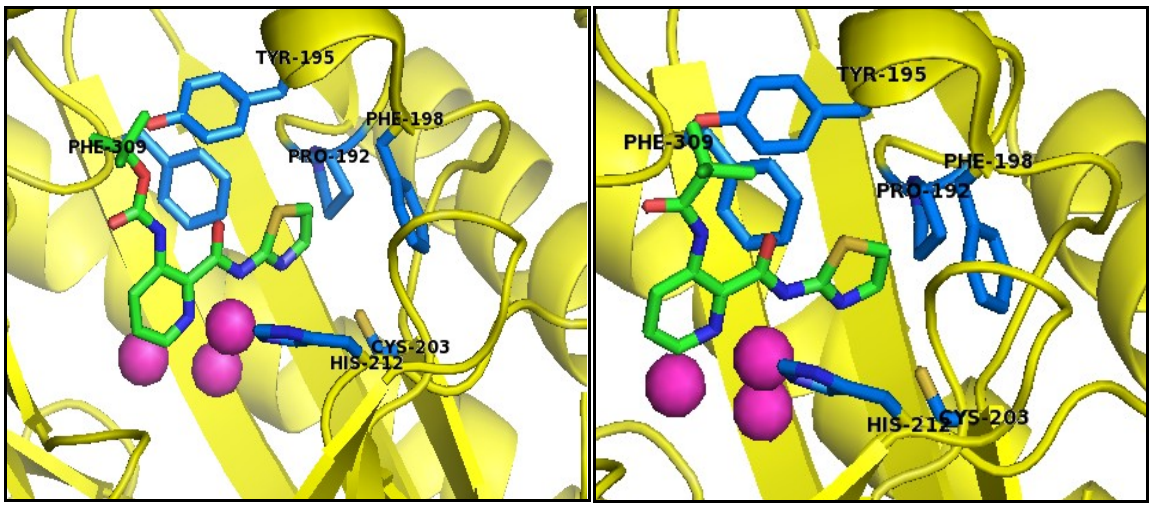


Figure 4.10: Structure of tMetAP1 in complex with inhibitory compounds 1 (left) and 2 (right) [14]. (PDB ID 2NQ6 and 2NQ7)

The structure confirms the presence of two Co^{2+} ions in the active site of the enzyme. Interestingly, a third Co^{2+} ion was observed in the active site; this additional metal ion mediates the interaction between His-212 and the inhibitor. The tiazole ring of the inhibitor is buried deeply in the active site pocket, and it is surrounded by Cys-203 and several hydrophobic residues (Pro-192, Tyr-195, Phe-198, Phe-309). Compounds 1 and 2 were able to cause a significant cell delay during G2/M phase, thus suggesting that MetAP1 may play an important rule in this phase transition; as a consequence, MetAP1 may be a useful target for the discovery and the development of new anticancer agents. On the other hand, fumagillin and its analogs, which inhibit MetAP2, are able to arrest the cell cycle at G1/S phase.

4.2 Materials and methods

4.2.1 Cloning of MetAP1 gene

The gene encoding for human MetAP1 lacking the first 80 amino acids (ΔMetAP1, 81-386) was cloned into the plasmid pET28a(+) (Novagen®) with a 6-histidine tag at the N-terminus, as reported [7]. The plasmid, pET-28a+ΔHsMetAP1, was transformed into *E. coli* BL21 (DE3) cells. Single clones were saved as glycerol stocks for inoculation.

4.2.2 Expression and purification

A glycerol stock of ΔHsMetAP1-pET-28a+ was used to inoculate 50 ml of LB medium containing 30 μl/ml kanamycin and shaken at 37°C overnight. 5 ml of the overnight culture were transferred into one liter of LB medium with 30 μl/ml kanamycin and shaken at 37°C. When the $O.D._{600}$ reached 1.0, IPTG was added to a final concentration of 1.0 mM and induction was carried out for about 16 hours at 20°C. Cells were harvested by centrifugation for 15 minutes at 5,000 rpm at 4° C and the cell pellet was stored at -80°C.

The cell pellet was resuspended in buffer F (50 mM HEPES pH 8.0, 0.5 M KCl, 5 mM imidazole, 10% glycerol and 0.1% Triton X-100). The cells were lysed by microfluidization. The sample was centrifuged for 30 minutes at 12,000 rpm at 4°C to clear cell debris.

Ni^{2+} affinity beads (QIAGEN) were added to the soluble fraction, which was then shaken for 2 hours at 4°C. Beads were then collected and washed with buffer F until the U.V. absorption at 280 nm returned to the baseline. The beads were washed with further 10 ml of buffer G (50 mM HEPES pH 8.0, 0.5 M KCl and 5 mM imidazole).

The pure target protein was eluted with 10 ml of buffer H (50 mM HEPES pH 8.0, 0.5 M KCl and 100 mM imidazole) into a 50 ml tube containing 0.2 ml of 0.5 M EDTA pH 8.0.

The protein was dialyzed into freshly made buffer J (25 mM HEPES pH 8.0, 150 mM KCl and 5 mM methionine), concentrated to 3 mg/ml and stored at -80° C.

4.2.3 Crystallization

The protein was thawed and concentrated to 6 mg/ml. Crystals of ΔHsMetAP1 were grown by hanging drop vapor diffusion with 1 ml reservoirs consisting of 14 – 20 % PEG monomethyl ether 2000, 100 mM MES pH 5.4 – 6.2. 1 µl of protein solution was added to the same volume of the reservoir solution. The crystals grew at 20° C, and appeared after 3 days. Native crystals were soaked with 1 mM inhibitor 1 and 1 mM $CoCl_2$ for 1 hour and 30 minutes. Native crystals were soaked in 10 µl of reservoir solution for 2 days with 0.3 mM inhibitor 2 and 0.3 mM $CoCl_2$ added.

Data collection and data processing

Data from the ΔMetAP1 crystals were collected on a FRE+ generator (Rigaku) as the X-ray source with an R-AXIS IV image plate detector at the X-ray facility of the Department of Biophysics and Biophysical Chemistry of the Johns Hopkins University School of Medicine. Indexing and data reduction were carried out with HKL2000 [15]. The crystals of ΔMetAP1 are monoclinic, space group $P2_1$ and contain one molecule in the asymmetric unit. Data collection statistics are listed in Table 4.3.

Structure determination and refinement

The structure of the ΔMetAP1 was determined by direct refinement using the coordinates of the native enzyme (PDB ID 2B3L). The initial model was rebuilt using Coot [16] and refined using Refmac5 [17]. Refinement statistics for the data are summarized in Table 4.3.

4.3 Results and discussion

Human ΔMetAP1 was cloned into the pET-28a+ plasmid and transformed in *E. coli* BL21 (DE3) strain. The protein was purified in a single step with Ni^{2+} beads; the 6-histidine tag at the N-terminus of the protein was not cleaved.

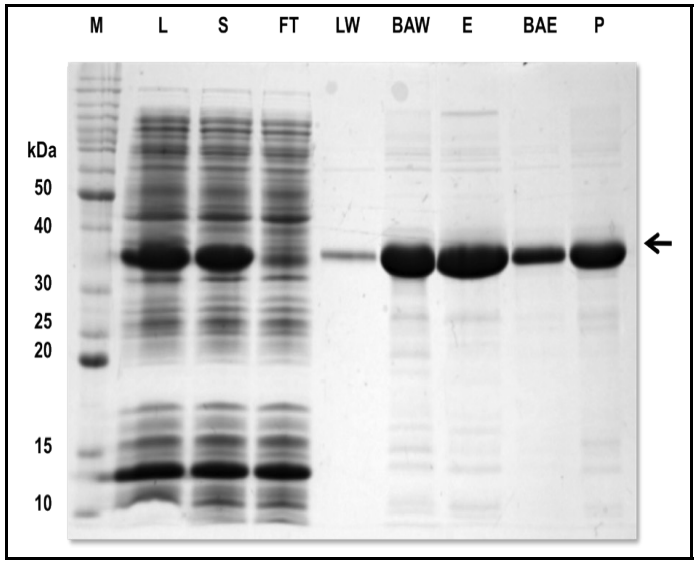


Figure 4.11: Coomassie blue staining of 15% SDS-PAGE: M, BenchMarker, 5 µl; L, whole cell lysate, 1/3600; S, supernatant after centrifugation, 1/3600; FT, flow-through from QIAGEN beads, 1/3600; LW, last wash, 17 µl; BAW, beads after wash, 1/900; E, elute, 1/1000; BAE, beads after elution; P, precipitation after overnight on ice and concentrating.

The protein was dialyzed in a suitable buffer, concentrated to 6 mg/ml and crystallization trials were set up.

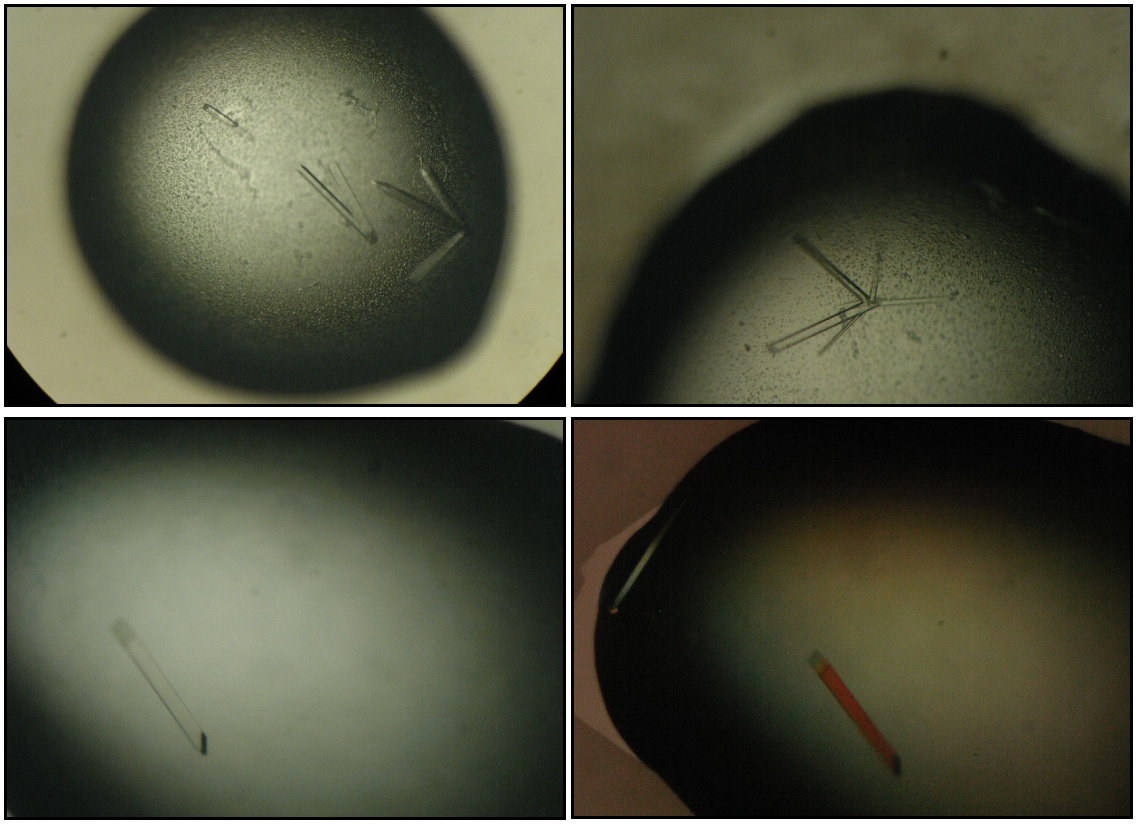


Figure 4.12: Microphotographs of ΔMetAP1 crystals.

Crystal	ΔMetAP1 native	ΔMetAP1 inhibitor 1	ΔMetAP1 inhibitor 2
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions	a = 47.3 Å b = 77.2 Å c = 47.9 Å β = 91.7°	a = 47.3 Å b = 77.3 Å c = 47.8 Å β= 91.5°	a = 47.5 Å b = 77.4 Å c = 48.0 Å β = 90.9°
Data Collection Statistics			
X-ray Source	FRE/SaturnCCD	FRE/RaxisIV	FRE/RaxisIV
Wavelength (Å)	1.54178 Å	1.54178 Å	1.54178 Å
Resolution (Å) (HighRes shell)	50-2.20 (2.28-2.20)	50-1.90 (1.97-1.90)	50-2.09 (2.16-2.09)
Measured Reflections	53,078	74,840	71,126
Unique Reflections	16,929	24,981	20,226
I/σ	16.3 (5.5)	28.8 (3.3)	29.6 (2.8)
Completeness (%)	96.1 (80.5)	91.6 (59.7)	97.9 (86.6)
R_{merge} (%)	12.2 (27.7)	5.5 (27.6)	11.7 (49.0)
Refinement			
R_{cryst} (%)		0.175 (0.24)	0.20 (0.27)
R_{free}		0.225 (0.31)	0.26 (0.32)
R.m.s deviations			
Bond length (Å)		0.012	0.010
Angle (°)		1.291	1.192
Monomer in ASU		1	1
Total Atoms		2,661	2,586
Protein atoms		2,400	2,400
Water molecules		227	152
Ligand		34	34
Bfactor (MAP) $Å^2$		29.6	41.8
Bfactor (ligand) $Å^2$		30.9	62.0
Bfactor (H_20) $Å^2$		42.9	55.0

Table 4.3: Data collection and refinement statistics of the ΔHsMetAP1 free enzyme, and complexes with inhibitor 1 and 2.

The apo form of the crystal was checked to assess the absence of any compound in the active site pocket and it was not completely refined.

Description of the ΔHsMetAP1 structure

The position of the two inhibitors inside the active site pocket is similar and is shown here.

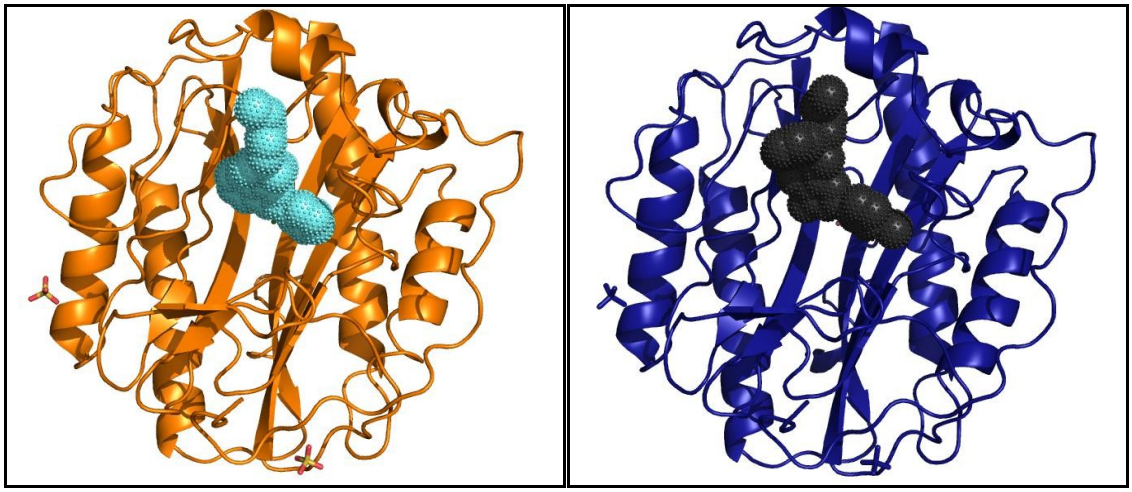


Figure 4.13: Inhibitor 1 (left) and inhibitor 2 complexes (right) with ΔHsMetAP1.

The two inhibitors compete for the same active site and show a similar behavior in the binding.

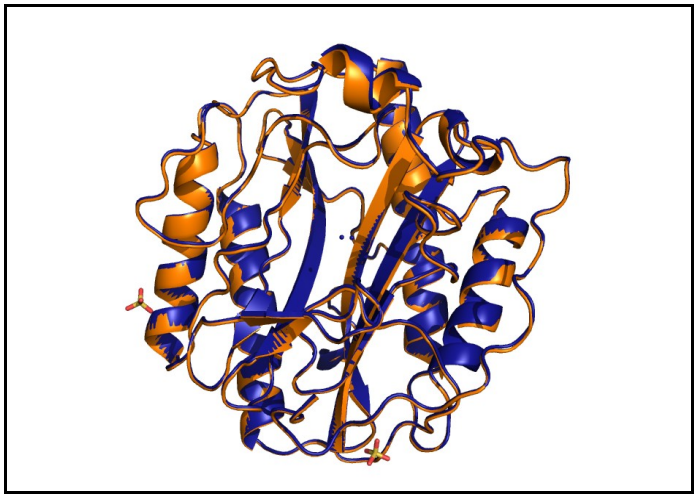


Figure 4.14: Superimposition of ΔHsMetAP1 in complex with inhibitor 1 and inhibitor 2; the protein is shown in orange and blue, respectively.

As the two figures show, there is no change in the Cα chain, thus indicating that the conformation of the protein binding the two inhibitors is very similar.

Since the natural substrate (methionine) is hydrophobic, the active site has several hydrophobic residues.

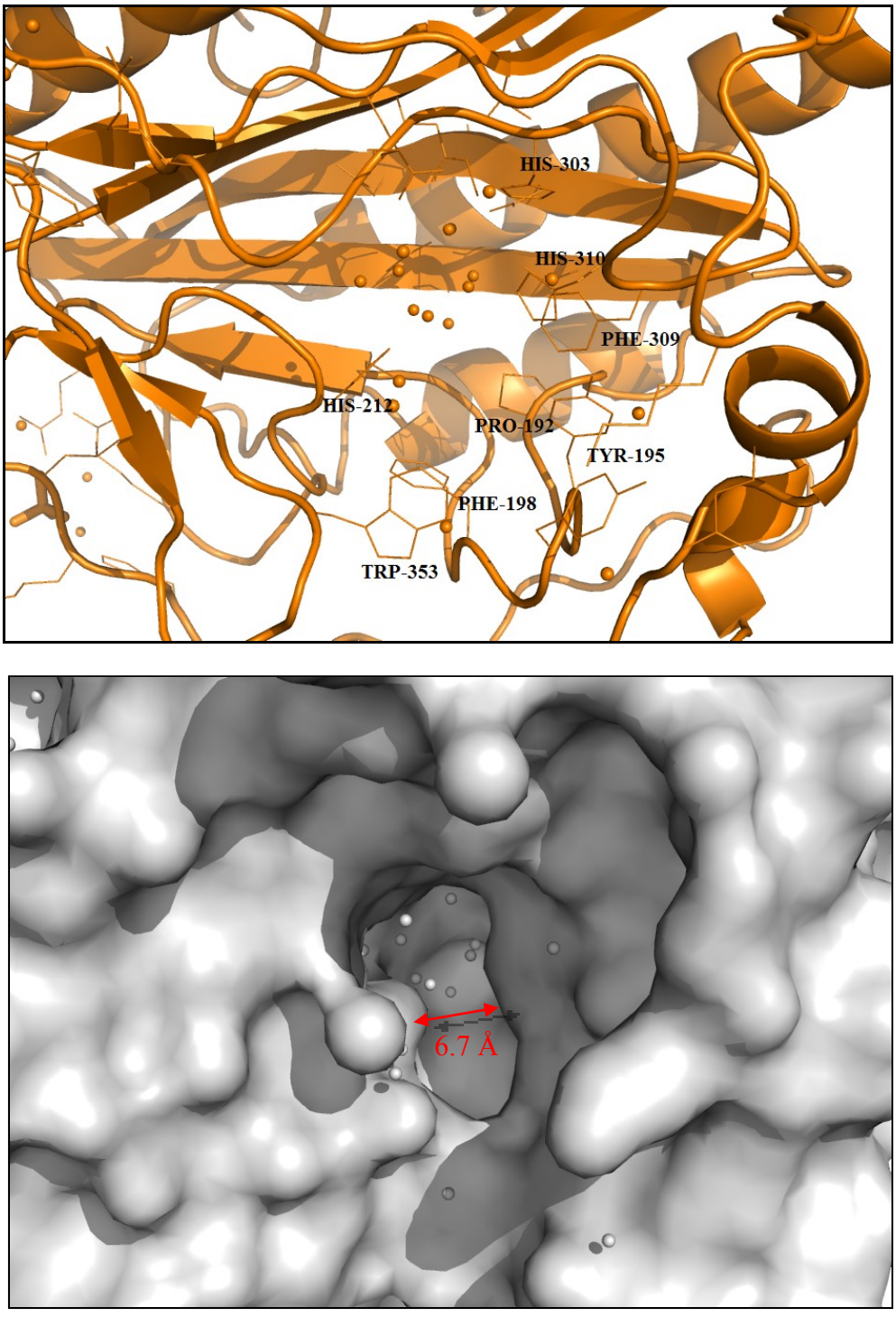


Figure 4.15: Active site (above) and active site surface (under) of ΔHsMetAP1 in complex with inhibitor 1 (not shown).

The inhibitors are located inside the active site pocket and completely block this cavity, thus avoiding substrate access.

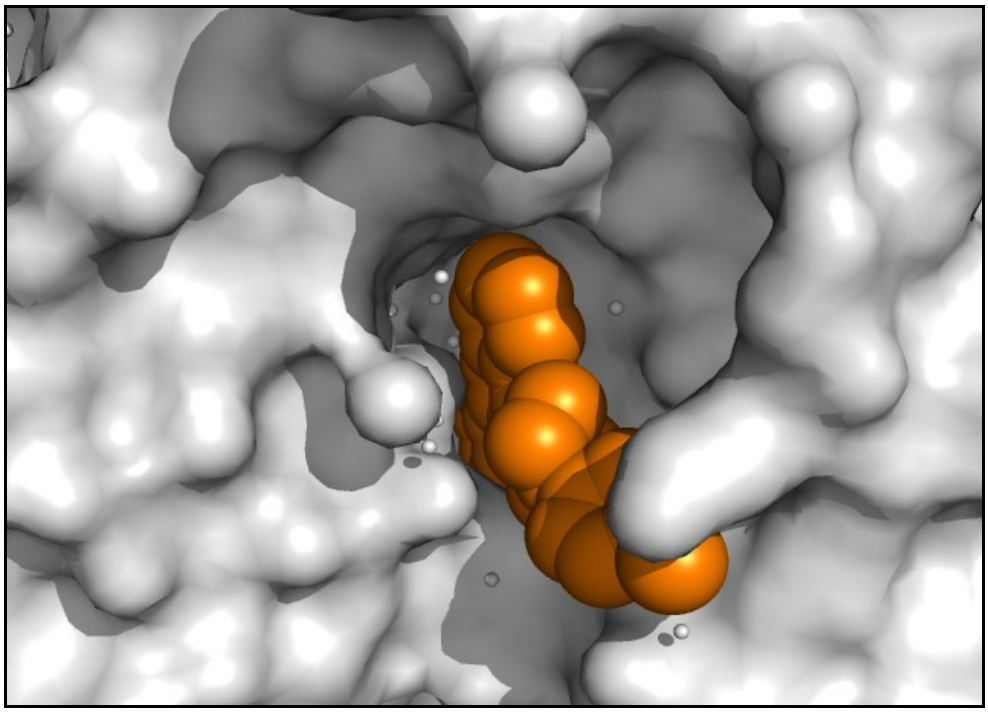


Figure 4.16: Active site surface of ΔHsMetAP1 in complex with inhibitor 1.

4.4 Conclusions

Human ΔHsMetAP1 protein has been expressed in the prokaryotic system *E. coli*, purified with a single affinity step and crystallized. The protein crystals were then soaked with two different inhibitors; the structure with inhibitor 1 and inhibitor 2 were determined to a resolution of 1.9 Å and 2.1 Å, respectively.
The inhibitors bind to the active site pocket with a similar mechanism.
The present study will be helpful in the design of new inhibitors of this protein.

5. References

[1] Kendall R.L., Bradshaw R.A. (1992), Isolation and characterization of the methionine aminopeptidase from porcine liver responsible for the co-translational processing of proteins. *J. Biol. Chem.* **267**: 20667-20673.

[2] Giglione C., Boularot A., Meinnel T. (2004), Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* **61**: 1455-1474.

[3] Chen X., Chong C.R., Shi L., Yoshimoto T., Sullivan D.J. Jr., Liu J.O. (2006), Inhibitors of *Plasmodium falciparum* methionine aminopeptidase 1b possess antimalarial activity. *Proc. Natl. Acad. Sci. U S A* **103**: 14548-14553.

[4] Chang S.Y., McGary E.C., Chang S. (1989), Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. *J. Bacteriol.* **171**: 4071-4072.

[5] Miller C.G., Kukral A.M., Miller J.L., Movva N.R. (1989), pepM is an essential gene in *Salmonella typhimurium*. *J. Bacteriol.* **171**: 5215-5217.

[6] Li X., Chang Y.H. (1995), Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl. Acad. Sci. U S A* **92**: 12357-12361.

[7] Addlagatta A., Hu X., Liu J.O., Matthews B.W. (2005), Structural basis for the functional differences between type I and type II human methionine aminopeptidases. *Biochemistry* **44**: 14741-14749.

[8] Zuo S., Guo Q., Ling C., Chang Y.H. (1995), Evidence that two zinc fingers in the methionine aminopeptidase from *Saccharomyces cerevisiae* are important for normal growth. *Mol. Gen. Genet.* **246**: 247-253.

[9] Liu S., Widom J., Kemp C.W., Crews C.M., Clardy J. (1998), Structure of human methionine aminopeptidase-2 complexed with fumagillin. *Science* **282**: 1324-1327.

[10] Lowther W.T., Matthews B.W. (2000), Structure and function of the methionine aminopeptidases. *Biochim. Biophys. Acta* **1477**: 157-167.

[11] Kruger E.A., Figg W.D. (2000), TNP-470: an angiogenesis inhibitor in clinical development for cancer. *Expert Opin. Investig. Drugs* **9**: 1383-1396.

[12] Addlagatta A., Matthews B.W. (2006), Structure of the angiogenesis inhibitor ovalicin bound to its noncognate target, human Type 1 methionine aminopeptidase. *Protein Sci.* **15**: 1842-1848.

[13] Towbin H., Bair K.W., DeCaprio J.A., Eck M.J., Kim S., Kinder F.R., Morollo A., Mueller D.R., Schindler P., Song H.K., van Oostrum J., Versace R.W., Voshol H., Wood J., Zabludoff S., Phillips P.E. (2003), Proteomics-based target identification: bengamides as a new class of methionine aminopeptidase inhibitors. *J. Biol. Chem.* **278**: 52964-52971.

[14] Hu X., Addlagatta A., Lu J., Matthews B.W., Liu J.O. (2006), Elucidation of the function of type 1 human methionine aminopeptidase during cell cycle progression. *Proc. Natl. Acad. Sci. U S A* **103**: 18148-18153.

[15] Z. Otwinowski and W. Minor, "Processing of X-ray Diffraction Data Collected in Oscillation Mode", *Methods in Enzymology* **276**: Macromolecular Crystallography, part A, p.307-326, 1997, C. W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York).

[16] Emsley P., Cowtan K. (2004), Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2126-2132.

[17] Vagin A.A., Steiner R.A., Lebedev A.A., Potterton L., McNicholas S., Long F., Murshudov G.N. (2004), REFMAC5 dictionary: organization of prior chemical

knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2184-2195.

Buffer compositions

buffer A: 20 mM TRIS-HCl pH 7.5, 0.5 M NaCl and 0.02% NaN_3

buffer B: 20 mM TRIS-HCl pH 7.5, 0.15 M NaCl and 0.02% NaN_3

buffer C: 20 mM TRIS-HCl pH 8.0 and 0.02% NaN_3

buffer D: 20 mM TRIS-HCl pH 7.5 and 0.02% NaN_3

buffer E: 50 mM TRIS pH 8.0, 0.15 M NaCl, 10 mM glutathione and 0.02% NaN_3

buffer F: 50 mM HEPES pH 8.0, 0.5 M KCl, 5 mM imidazole, 10% glycerol and 0.1% Triton X-100

buffer G: 50 mM HEPES pH 8.0, 0.5 M KCl and 5 mM imidazole

buffer H: 50 mM HEPES pH 8.0, 0.5 M KCl and 100 mM imidazole

buffer J: 25 mM HEPES pH 8.0, 150 mM KCl and 5 mM methionine

Medium compositions

Luria-Bertani (LB) medium:

- 1% tryptone
- 0.5% yeast extract
- 0.5% NaCl

Terrific Broth (TB) medium:

- 12 g tryptone
- 24 g yeast extract
- 4 ml glycerol
- 17 mM KH_2PO_4
- 72 mM K_2HPO_4

Acknowledgements

I would like to show my gratitude to Prof. Hugo Monaco and Dr. Massimiliano Perduca, for giving me the opportunity to work in a scientific environment and for their encouragement.

I would like to thank all the people of the lab, for the wonderful time spent together, for their jokes and their support: Sacco, Manu, Bovi, Sean Faggion, Laura Fin, Lucy, Divina, Giacy, Ilaria, Valentina, Fabio, Leonardo, Michela, Laura Civiero, Pedro, Fabio De Simone, Stefano, Genny, Pelle, Lissete, Elisa, Marghe.

I would like to thank Prof. L. Mario Amzel for the wonderful experience in his lab.
I am grateful to Dr. Sandra B. Gabelli for her support and her help.