UNIVERSITA` DEGLI STUDI DI NAPOLI

FEDERICO II

Ph.D. School in Chemical sciences

XXVII course

Mycobacterium tuberculosis adaptive response to DNA methylation stress

Francesca De Maria

Tutor:

Prof. Angela Duilio

Supervisor:

Prof. Filomena Sica

Coordinator:

Prof. Luigi Paduano

A mia madre

Table of Contents

Summary

Summary

Tuberculosis (TB) is an infectious disease caused by the micro-organism *Mycobacterium tuberculosis* (MTB) that currently represents a global health emergency with around 9 million people affected and nearly 2 million deaths each year. The present therapeutic treatment for TB consists in the use of multiple anti-mycobacterial drugs such as rifampicin and isoniazid. This treatment is carried out for a long time leading to the onset of drug- and multidrug-resistant strains of MTB. Therefore new therapeutic models are needed to overcome this problem. During MTB infection, the host antimicrobial response generates several metabolically activated DNA alkylating agents leading to severe DNA-damaging injuries on MTB cells. Protection of the DNA molecule from chemical damages then strictly depends on the MTB repair mechanisms. Recent studies identified an adaptive response mechanism in MTB homologous to that already described in *E.coli*. Being the adaptive response a fundamental biological process for MTB viability whereas it is missing in human, this process may represent a putative therapeutic target to explore in search for new TB treatments. Previous studies in MTB identified four genes which encode the proteins constituent of the adaptive response mechanism homologous to the same process already defined in *E. coli*. Exposure of *E. coli* to sublethal concentrations of alkylating agents induces the expression of four genes (*ada*, *alkA*, *alkB* and *aidB*). The activation of these genes increases the resistance to the cytotoxic and mutagenic effects of alkylating molecules. Unlike *E. coli*, the DNA repair/protection systems in MBT have not been investigated and it is still largely unknown.

This thesis project focuses on the investigation of the adaptive response to DNA methylation stress in MTB, in order to identify possible innovative therapeutic targets against tubercular infection.

A non-pathogenic, reference strain of Mycobacteria, *M. smegmatis*, was used to perform all the experiments. The effect of methylating molecules on *M. smegmatis* was evaluated using MMS, a common laboratory methylating molecule.

First the effect of alkylating agents was investigated on *M. smegmatis* cells by evaluating the effect of MMS on cell survival, DNA alkylation and on the global proteomic response.

Second, structural and functional characterization of the recombinant Ada-Alka and Ogt proteins was carried out because of their involvement in the cellular response to alkylating agents.

Finally, functional proteomic experiments on *M.smegmatis* Ogt protein was performed in order to shine a light on its biological function through the identification of its protein partners *in vivo*.

The effect of methylating agents on cell survival was evaluated by monitoring the growth profile of bacterial cultures of *Mycobacterium smegmatis c*ells, grown in the presence and in the absence of different concentrations of MMS.

The effect of DNA methylation was quantitatively monitored by Multiple Reaction Monitoring tandem MS. N-7-methylguanine (7MeG), O6-methylguanine (7MeG) and N-3-methyladenine (3MeA) are the most abundant DNA base modifications by methyl methane sulfonate (MMS).

Further experiments were then designed in order to investigate the global proteomic response of *M.smegmatis* to methylating stress using bi-dimensional electrophoresis analysis.

Then, proteomic response to methylation stress was deeply investigated by performing quantitative proteomic analyses using the DIGE (Differential In Gel Electrophoresis) technology.

The identified proteins were functionally classified according to their known biological functions.

Results obtained in the presence of MMS showed the increasing of proteins involved in biofilm formation, explored by performing specific assays in the presence and in the absence of MMS to evaluate the amount of biofilm formed by *M.smegmatis*.

In the second part of this thesis project Ada-AlkA and Ogt proteins were produced in *E. coli* and characterised.

The genes were cloned in pGEX4T1 plasmid vector, producing proteins fused to Glutathione S-transferase (GST). The recombinant proteins were purified by affinity chromatography on immobilized glutathione beads and hydrolyzed in column to release GST. The primary structures of recombinant Ada-alkA and Ogt were validated by mass spectrometry methodologies. Circular dichroism (CD) analyses highlighted correctly folded products. The ability of the Ada-alkA and Ogt to bind DNA was probed by Elettrophoretic Mobility Shift Asing ssay (EMSA) experiments.

In the last part, *in* vivo isolation of Ogt-containing protein complexes was performed using functional proteomics approaches in order to shine a light on Ogt biological rol

Chapter 1:

Introduction

1. Tuberculosis

1.1 TB epidemiology

Tuberculosis (TB) is an acute infectious disease that is among the most ancient infection in the world. Signs of disease attributable to tuberculosis were identified in bones of North American bovids Pleistocene, demonstrating the presence of the disease in earlier times. This is then confirmed by recent studies of DNA sequencing for a bone sample from a bison dated to 17,000 years before the present .

Also, the presence of similar pathology in fossil bighorn sheep and musk ox suggests that bovids were the likely vehicle for spreading of what became known as the white plague. [1]

Despite such evidence, the etiology of this disease and the causative agent of TB remained unknown until 1882, when the German doctor Robert Koch isolated the bacterial pathogen *Mycobacterium tuberculosis* (MTB), also named Kock bacillus, by a new staining procedure that allowed him to observe the new organism in tuberculous lesions [2].

In the seventeenth century, in Europe, TB became the leading cause of death, due to the high population density, poor sanitation and migration.

Robert Kock has pioneered the diagnosis and therapy of TB, but although great progress made, this disease still represents a global health emergency.

In 1994-1995, in order to improve TB control programs, the DOTS (Directly Observed Treatment, Short-Course) strategy was launched. This strategy was based on five key essential components:

1) Political commitment with increased finance

2) Case detection among people presenting with symptoms in clinics through quality-assured bacteriology;

3) Standardized and supervised treatment along with patient support;

4) Effective drug supply and management system;

5) A standard monitoring and evaluation system

By 2004, more than 20 million patients had been treated in DOTS programmes worldwide and more than 16 million of them had been cured. Mortality due to TB has been declining and incidence diminishing or stabilizing in all world regions except sub-Saharan Africa and, to some extent, eastern Europe [3].

According to the World Health Organization (WHO) , in 2012 8.6 million people developed the disease, and 1.3 million people died. Respect to 1990, a 50% reduction in mortality has been achieved, but TB incidence is declining too slowly [4-5].

1.2 Mycobacterium tuberculosis and stage infection

The TB causative agent is *Mycobacterium tuberculosis*, a non sporogenous, aerobic bacterium, belonging to the family of *Mycobacteria* .

Mycobacteria are intracellular pathogens, characterized by high percentage of guanine and cytosine in their genome.

The main feature of all these micro-organisms is the presence of a complex wall much thicker than that in other bacteria, characterized by a high concentration of mycolic acids and fatty acids which confers enormous resistance to environmental factors and antibiotic substances (Fig 1).

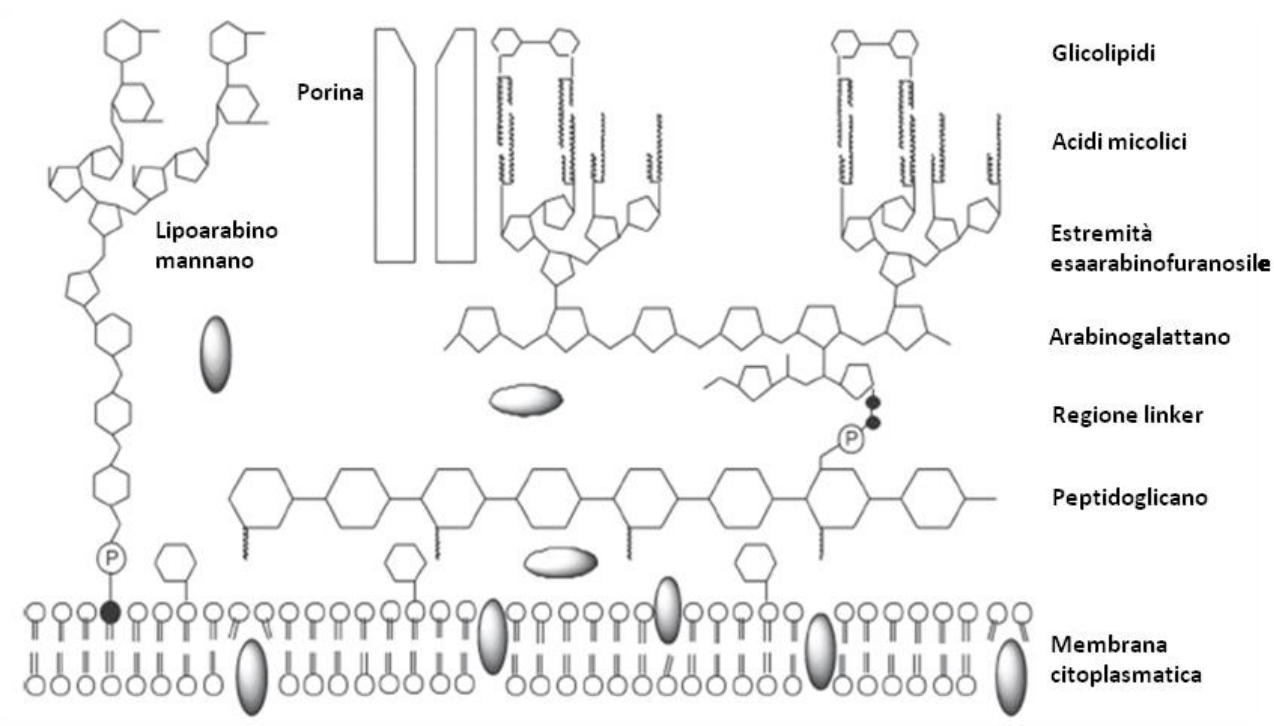

Figure 1: *Mycobacterium tuberculosis* cell wall

Depending on their pathogenicity, these species are divided into *Mycobacterium Other Than Tuberculosis* (MOTT) and *Mycobacterium Tuberculosis Complex* (MTC), such as *Mycobacterium africanum* (MTB like, isolated in countries of central Africa), *Mycobacterium bovis* (causative agent of bovine TB), *Mycobacterium leprae* (the causative agent of leprosy) and *Mycobacterium tuberculosis (*MTB), the causative agent of human TB (fig. 2).

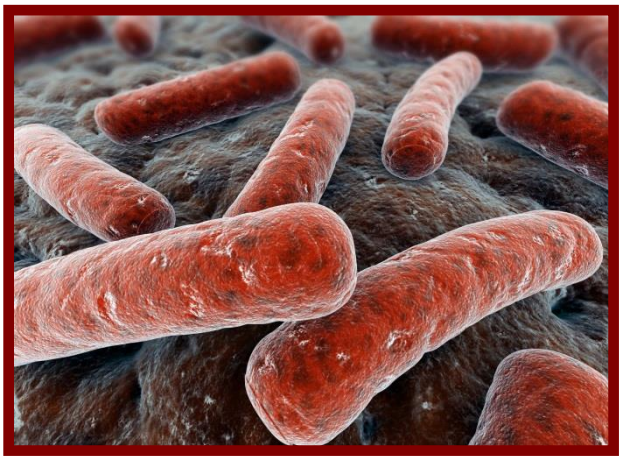

Figure 2: *Mycobacterium tuberculosis*

MTB has a slow generation time (12-18 hours), due to the complexity of the its cell wall. Its cell wall is composed of three covalently linked macromolecules: peptidoglycan, arabinogalctan and the hallmark of mycobacteria, myolic acids. Also it possess an important molecule, LAM (lipoarabinomannan), representing the virulence factor [6].

In particular, cell wall MTB exterior is primary characterised by the presence of a glycolipid molecule, trehalose dimycolate, also known as cord factor. These molecules bring to the arrangement of M. tuberculosis cells into long and thin structures named cord factors like the glycolipid component. This cellular formation is a key step in pathogenicity of tuberculosis infection because can affect immune response and granuloma production [7,8].

MTB is transmitted in aerosol droplets coughed out by infected individuals. The infection establishes itself in the lungs, where the macrophages start phagocytosis of the bacterium generating the primary infection. When the host's response is efficient, MTB can easily be removed, but if the host response is unsatisfactory, the bacterium may survive within

macrophages (fig. 3). In this complex enviroment, the mycobacteria are continuously exposed to both oxidative and nitrosative stress generated by the activated macrophages that they inhabit. When the active phase of infection is over, mycobacteria persist in the infected lung and inhibit the fusion of phagosomes and lysosomes causing a chronic infection. Constant antigenic stimulation and accumulation of T-cells leads to the formation of a granuloma (fig. 4), which can maintain viable bacilli for for a long time. So the infection may enter a stage of latency.

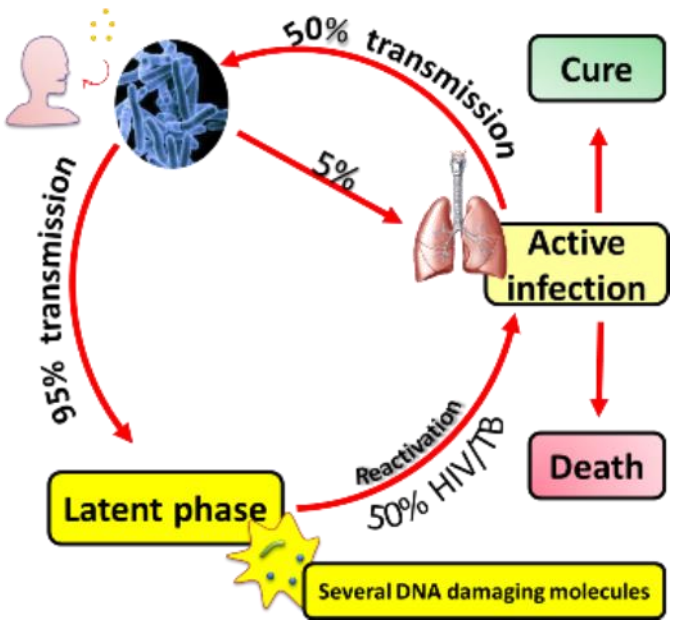

Figure 3: Stages of MTB infection

The structure of a human granuloma is characterized by a central region of large CD68+ epitheloid cells, surrounded by macrophages, T-lymphocytes (predominantly CD4+) and multi-nucleate giant Langerhans cells. [9]

The central region of the granuloma does not contain blood vessels and becomes hypoxic [10]. When the granuloma becomes necrotic, it may either be resolved by fibrosis and calcification or become liquefied, therefore leading to dissemination of any pathogens contained within it.

The state of latency can be very long until the breakage of the granuloma generates the reactivation of the disease. This phenomenon occurs for two main reasons: one of the causes is the essential

formation failure and the co-location signals maintenance from macrophages and T-cells, even if there are other events that suggest that T-cells remain effective. The other one is related to a decline in the host's immunity, which may derive from the host genetic susceptibility or from environmental reasons, such as co-infection with HIV.

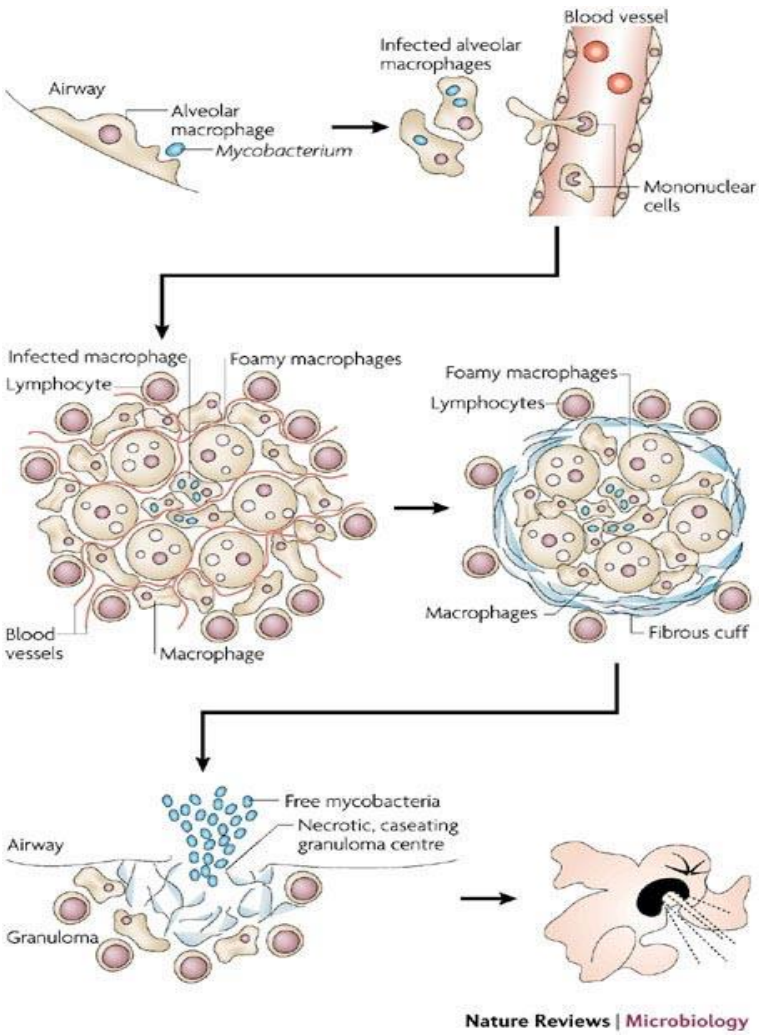

Figure 4: Granuloma formation

Current therapeutic treatments of TB disease consist in the use of multiple anti-mycobacterial drugs such as Rifampicine and Isoniazid. Both latent TB infection and TB disease can be treated with just one drug or more in combination [11]. Rifampicine seems to have a discreet success on both metabolic conditions. Isoniazid and Streptomycin are effective on metabolic active bacteria and Ethambutol, an Isoniazid derivative, has a bacteriostatic action. The results of such therapies are not very satisfactory: the treatment requires many months of administration and the

combination of multiple drugs can create a certain amount of side effects [12]. Particularly, the latent phase of TB is characterized by the absence of clinical signs and individuals with such latent TB are non-infectious [13]. In this phase there is a reduced response to antibiotic therapy and moreover it was found that HIV co-infection lead to a greater than 100-fold risk of developing active tuberculosis [14].

The term dormancy is used to refer to the state in which bacteria are found during clinical latency: they appear to remain viable, do not divide, but they are surely able to reactivate and resume cell division [15].

2. DNA damage

Because of its permanence in a hypoxic environment and exposure to selective antibiotics, during latency the bacterium undergoes DNA damage that result in the occurrence of drug-resistant, multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains.

An interesting approach to understanding the emergence of drug-resistant strains is the study of mutation rates of the bacterium.

It has been proved that the mutation rate of animal extracted Mycobacteria with active, latent and reactivated infection is quite similar in the three phases of the disease. MTB keeps acquiring mutations during latency [16]. The identified polymorphisms pattern suggest that the amount *in vivo* mutations derives from the DNA oxidative damage.

Thanks to the whole bacterium genome sequencing the mutation rate has been quantified and 14 single nucleotide polymorphisms (SNPs) have been identified. The mutation rate (μ) of a bacterium, *in vivo,* can be derived from the number of mutations (m) in a genome of known size (N) in time (T).

	Mutation rate μ
Latent infection	$2{,}71 \times 10^{-10}$

Reactivated infection	$3{,}03 \times 10^{-10}$
Active infection	$2{,}01 \times 10^{-10}$

Bacteria populations have the same number of mutations in time, independently of the number of bacterial replications. So, the polymorphisms identified are probably caused by an oxidative damage [17].

Therefore new therapeutic models are needed, to reduce treatment duration, decrease dosing frequency, improve the treatment of MDR and XDR, in order to overcome this problem [18]. Reactive species such as free radicals, one-electron oxidants, different chemicals, etc., can react with different components of DNA to produce a plethora of DNA lesions. During the infection, MTB is exposed to a multitude of hostile conditions, due to host defense systems and antibiotic treatments. Particularly MTB undergoes a variety of potentially DNA-damaging assaults, primarily from host-generated antimicrobial reactive oxygen intermediates (ROI) and reactive nitrogen-intermediates (RNI) [19] .

Microorganisms evolved in a world lacking of oxygen and rich with reduced iron. The subsequent oxygenation of the atmosphere by photosynthetic organisms created a crisis: oxygen is a reactive chemical, and organisms had to devise strategies to defend themselves. Molecular oxygen is small and non-polar, and it quickly diffuses across biological membranes and water. Consequently, cells cannot respire quickly enough to lower the intracellular oxygen concentration substantially below the concentration immediately outside the cell. Some microorganisms escape oxidative stress by residing in anaerobic micro habitats; all others must contend with intracellular molecular oxygen.

However, virtually all of these microorganisms suffer poor growth, elevated mutagenesis or even death when they are exposed to oxygen levels that exceed those of their native habitats.

The oxygen toxicity derives mostly from the formation of partially reduced oxygen species (ROS)[20]. The hydroxyl radical is a highly reactive oxygen species (ROS) that efficiently reacts with nearby biomolecules at diffusion-controlled rates of reaction. The reaction volume of radicals is less than 2 nm in cells and tissues; thus, it reacts essentially at the site of generation.

The most likely source of OH in cells is the Fenton reaction, which involves the reaction of reduced redox active metal ions, such as ferrous and cuprous ions, with metabolically produced H2O2 [21]. The oxidative damage greatly contributes to spontaneous mutation in bacterial cells [22].

Nitric oxide is a key participant in many physiological pathways; however, its reactivity gives it the potential to cause considerable damage to cells and tissues in its vicinity.

Formation of nitric oxide ($NO^{\cdot}$) catalysed by nitric oxide synthase can produce ONOO− due to its reaction with $O2^{\cdot}$-, which is very reactive. More interestingly, ONOO− itself is capable of generating other reactive species that are very reactive. For example, the conjugate acid of ONOO−, i.e. ONOOH on homolytic dissociation, can generate reactive O2 and OH radicals.

Furthermore, in the presence of carbon dioxide (CO2), ONOO− can generate nitrosoperoxycarbonate anion (ONOOCO2−), which on homolytic dissociation may yield $CO3^{\cdot}$− and $NO2^{\cdot}$ free radicals [23].

Nitric oxide can react with DNA via multiple pathways. Once produced, subsequent conversion of nitric oxide to nitrous anhydride and/or peroxynitrite can lead to the nitrosative deamination of DNA bases such as guanine and cytosine. Complex oxidation chemistry can also occur causing oxidative modifications on DNA bases and sugars [24].

The cytotoxic and mutagenic effect of ROIs and RNIs is based on a number of chemical reactions that modify DNA, including several metabolically activated DNA alkylating agents [23,25]. Specifically, alkylating agents are molecules having one or more alkyl groups able to react with nucleophilic sites on DNA bases causing covalent modifications known as adducts [26]. The most reactive sites to be attacked by alkylating molecules are the ring nitrogens (N) and the extracyclic oxygen (O) atoms of DNA bases.

The chemical lesions pattern depends on the kind of alkyl agents and on their features, such as the number of reactive sites, the type of group added (methyl, chloroethyl, etc.) and its reactivity (type of nucleophilic substitution) [27].

If alkylating agents contain one active chemical moiety, it can modify DNA in a single site while bifunctional species possessing two reactive groups can bind simultaneously two sites generating intra- or inter-strand crosslinks.

DNA alkylation on reactive groups can occur by both S_N2 mechanism by targeting ring nitrogen atoms and S_N1 mechanism modifying nitrogen and extra cyclic oxygen groups (fig 5).

Amongst alkylating species, monofunctional methylating agents induce preferentially the formation of N-7-methylguanine (7MeG) at a rate of 60-80% of the total alkylation lesions because of the high nucleophilic reactivity of the N7 position of guanine. Mono-methylating agents can also produce N-3-methyladenine (3MeA) with an amount of 10-20% of total methyl adducts. In addition oxygen at position 6 of guanine is the most reactive oxygen in DNA and it is modified by S_N1 alkylating agent to form O6-methylguanine (O6MeG).

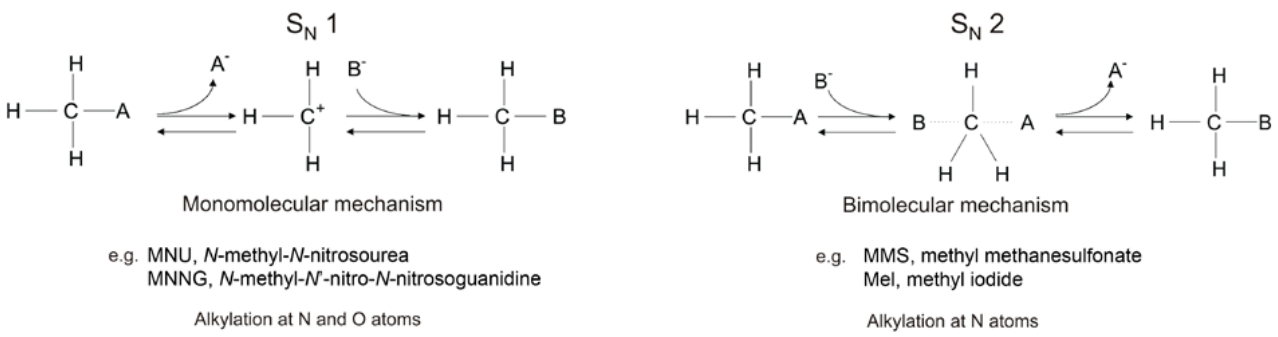

Figure 5: Schematic mechanism of DNA alkylation through nucleophilic substitutions: B is the nucleophilic base while the leaving group is donated by the alkylating agent.

2.1 DNA repair

Living cells are evolved with intelligent enzymes that protect DNA from erroneous and hazardous effects by executing about 10^{16}–10^{18} repair events per cell per day [28].

Because of the technical difficulties in working with a slow-growing pathogen such as *M. tuberculosis*, the study of DNA repair systems in this organism has advanced more slowly than for other bacteria.

Base excision repair (BER), Nucleotide excision repair (NER), recombination and SOS repair system genes were identified in the MTB genome [29].

BER (base excision repair) is one of the most important DNA repair systems in nature. It can protects the bacterial genome from damage and harmful mutations, therefore it is essential for

the bacterium survival (fig. 6) [19]. The process to DNA repair starts off with the activation of DNA glycosylases that produces in DNA an AP-site (a location in DNA that has neither a purine nor a pyrimidine base). These sites occur as intermediates in base excision repair, but they can be also caused by spontaneous depurination . In this process, the DNA glycosylase recognizes a damaged base and cleaves the N-glycosidic bond to release the base. This event is followed by strand cleavage of the sugar-phosphate backbone by an AP-endonuclease, leaving 3' hydroxyl and 5' deoxyribosephosphate termini.

The process is completed by the actions of phosphonucleotide kinase or a 3'- or 5'deoxyribosephosphodiesterase. Later a DNA polymerase fill in a new base and a DNA ligase seals the gap. *M. tuberculosis* possesses two redundant DNA polymerases involved in replication: DnaE1 and DnaE2. DNA polymerase DnaE1 is a replicative polymerase whereas DNA polymerase DNAE2 is upregulated by UV-induced DNA damage. In Mycobacterium tuberculosis this pathway is induced when the bacterium is exposed to oxidative stress such as hydrogen peroxide (H202), superoxide and hydroxyl radicals from exogenous and endogenous sources.

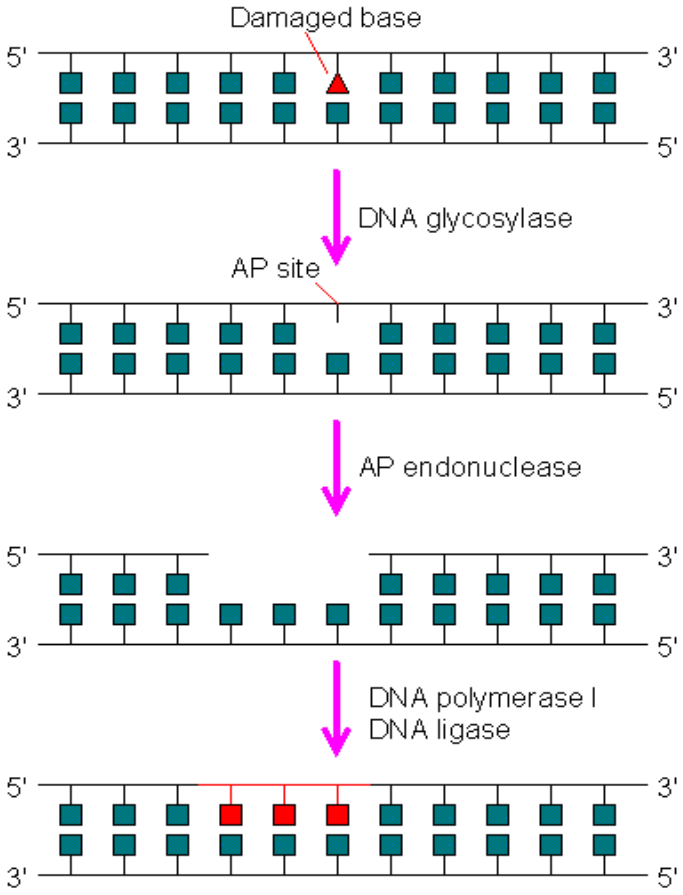

Figure 6: BER system

NER (nucleotide excision repair) is an alternative pathway to BER that acts on oligonucleotide excision (fig. 7). This system uses an endonucleases with low specificity therefore it is able to recognize a more broad range of damaged bases than the BER system. The NER pathway is composed by three components of the excinuclease ABC (UvrA,UvrB and UvrC), the DNA helicase II UvrD and the transcription repair-coupling factor (TRCF). According to studies, uvr gene expression is increased in *M. tuberculosis* isolated from human macrophages, demonstrating the importance of the uvr system for the bacterial survival upon infection [30].

In addition recent studies have demonstrated that uvr genes are involved in the response to H202 [31] and they are upregulated upon exposure to UV irradiation [32]. Furthermore, while most bacteria possess just one gene encoding DNA helicase II, M. tuberculosis has two putative gene encoding this protein: uvrD1 and uvrD2 [33]. The protein UvrD1 seems to be important to rapair DNA by UV damage and is upregulated by H202-induced stress whereas the protein

UvrD2 is essential for the survival of mycobacteria [34] NER is a complex process in which basically the following steps can be distinguished:

- ✓ recognition of a DNA lesion;
- ✓ separation of the double helix at the DNA lesion site;
- ✓ single strand incision at both sides of the lesion;
- ✓ excision of the lesion-containing single stranded DNA fragment;
- ✓ DNA repair synthesis to replace the gap
- ✓ ligation of the remaining single stranded nick.

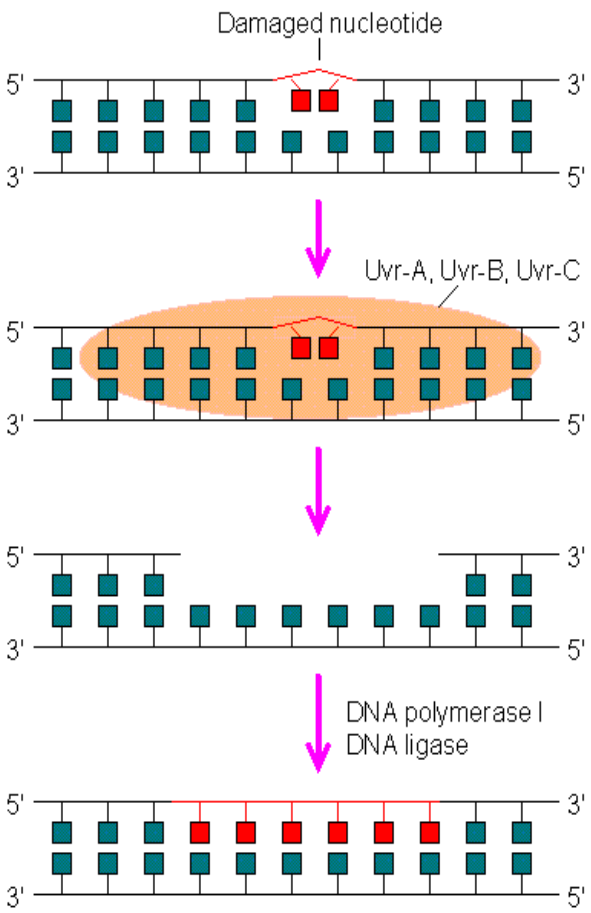

Figure 7: NER system

Breakage of the both strands of double helix (DSB) is another type of lethal damage to DNA forthe cells during replication or following exposure to DNA-damaging agents. There are two major DSB repair pathways: homologous recombination (HR) and nonhomologous end-joining (NHEJ) (fig. 8). HR is an homologous recombination where the second intact copy of the broken

chromosomal segment functions as a template for DNA synthesis across the break [35]. In this pathway, an 5'-exonuclease hydrolyzes the DNA after the break, generating sticky ends. Later the RecA protein promotes strand exchange and the resulting joint molecules can be processed by the RuvABC complex or RecG helicase. There are two major RecA-dependent pathway in initiating HR: the RecBCD and RecFOR pathway. RecBCD pathway is a bipolar helicase that separate the duplex into its component strands and digest them until it reaches a recombinational hotspot [36] whereas RecFOR pathway is important in postreplication daughter-strand gap repair [37]. According to studies, RuvA and RuvC, belonging to RuvABC complex, are upregulated upon UV damage in M. tuberculosis [32].

NHRJ joins together two ends of DNA after the break in the absence of a sequence or a chromosomal copy that can serve as a mold. The process starts off with the action of DNA-end-binding protein Ku, followed by the sealing of the broken strands by a specialized DNA ligase LigD [38]. LigD is a multifunctional enzyme formed by an ATP-dependent ligase (LIG) domain, a polymerase (POL) domain and a phosphoesterase (PE) domain. According to studies, NHEJ system is not conserved in all bacterial species but in mycobacterium it may be important during the latent phase of human tuberculosis. In fact this mechanism is resulted necessary for the mycobacterial survival during latent phase or reactivation from latency because when mycobacteria stay within microphages, they are continually exposed to the genotoxic defense mechanism of the host cell as the oxidative stress. Also this systems promotes mutagenesis so it results to be important for this mycobacterium to accumulate potential selective advantage under certain condition such as the acquisition of antibiotic resistance [35].

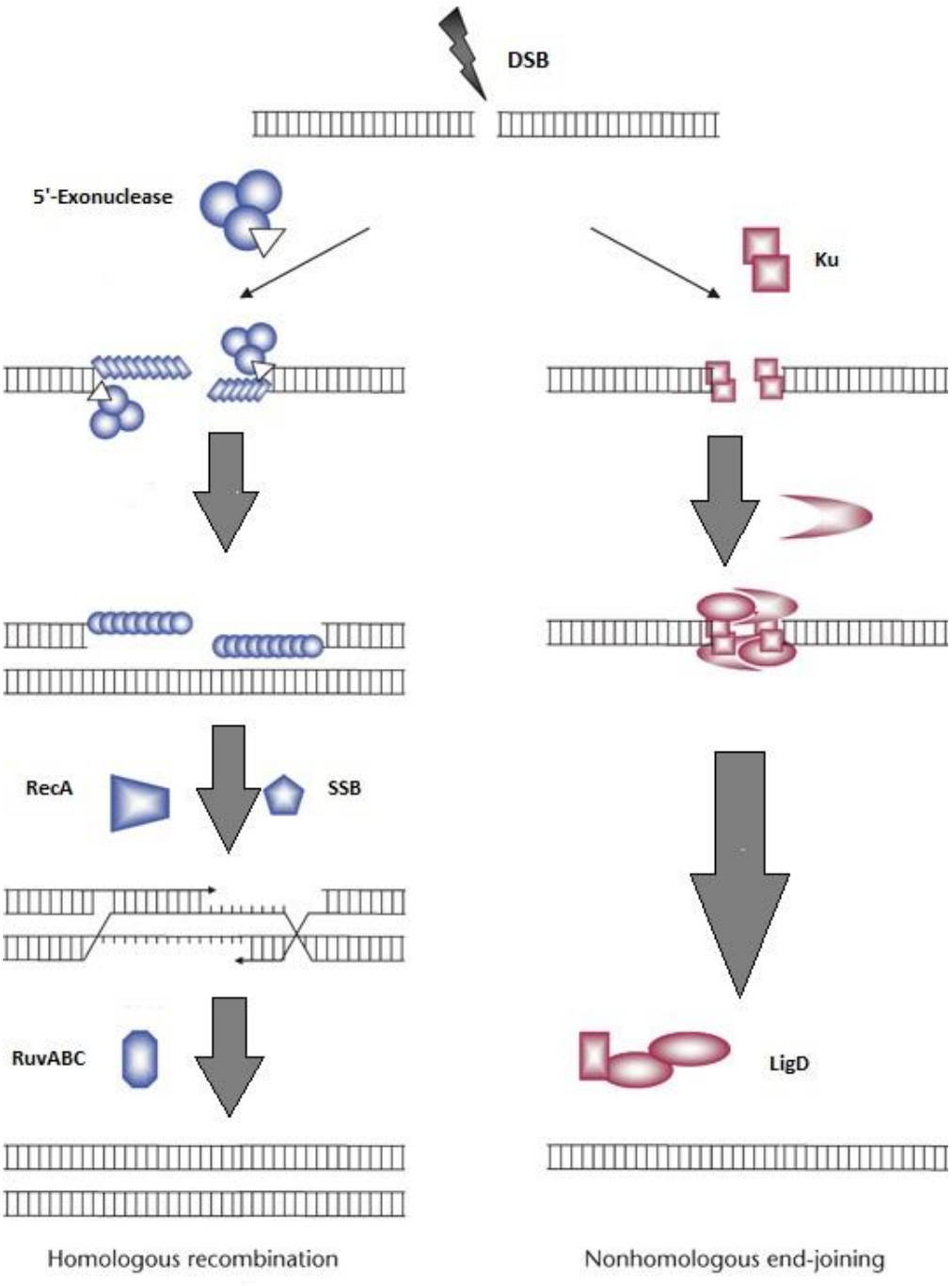

Fig.8 : Homologous recombination (a) and nonhomologous end-joining (b

2.2 Adaptive response

Among the repair mechanisms of DNA alkylation damage, some bacterial species have developed an adaptive response to alkylation stress [39] that is widely described in *E.coli*.

When *E. coli* is exposed to sublethal doses of alkylating agents it increases the cellular resistance to the mutagenic and cytotoxic effects of alkylation with increasing doses of damaging molecules.

The adaptive response set of genes is comprised of the *ada, alkA, alkB,* and *aidB* genes. Expression of these genes is regulated by Ada, (fig 9) that is a methyltransferase that also acts as positive regulator of the operon.

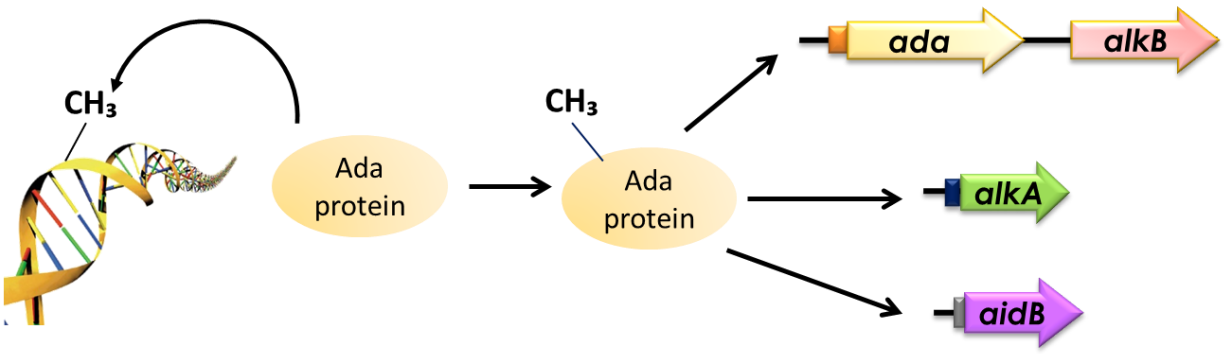

Figure 9 : Ada dependent mechanism

It has two cysteine residues, Cys-69 and Cys-321 that act as active methyl acceptor [40]: the first one is required to demethylate phosphomethyltriesters in the sugar phosphate backbone and the second one to remove methyl groups from either O6-methylguanine or O4-methylthymine. Both sites can be methylated from the specific substrates but only methylation by phosphates causes Ada conformational changes that induce the transcriptional regulator activity of the protein. Even though methylated phosphates are less dangerous than other lesions, these sites are the most quickly modified by methylating agents and thereby represent a sensitive regulatory signal leading to Ada operon induction. [41].

AlkA protein is a glycosylase that repair several damages such as N7-methylguanine and N3-methyl purines and O2-methyl pyrimidines. The mechanism of action comprises the removal of a damaged base from the sugar phosphate backbone through cleavage of the glycosylic bond leading to an abasic site.

AlkB protein is an oxidative demethylase capable of removing 1-methyladenine and 3-methylcytosine from DNA using an α ketoglutarate/Fe(II)-dependent mechanism coupled with the release of CO_2, succinate and formaldehyde [42].

Aidb is an Acyl-CoA dehydrogenase prevents DNA damage by alkylating agents and counteracts the block to transcription resulting from exposure to alkylating agents, preferentially acting on genes that are transcribed from promoters containing upstream (UP) elements [43].

Recent studies identified an adaptive response mechanism in MTB homologous to that already described in *E.coli*.

Due to bacterial evolution the genomic organization of the adaptive response is differently composed in E. coli and MTB (Fig 10).

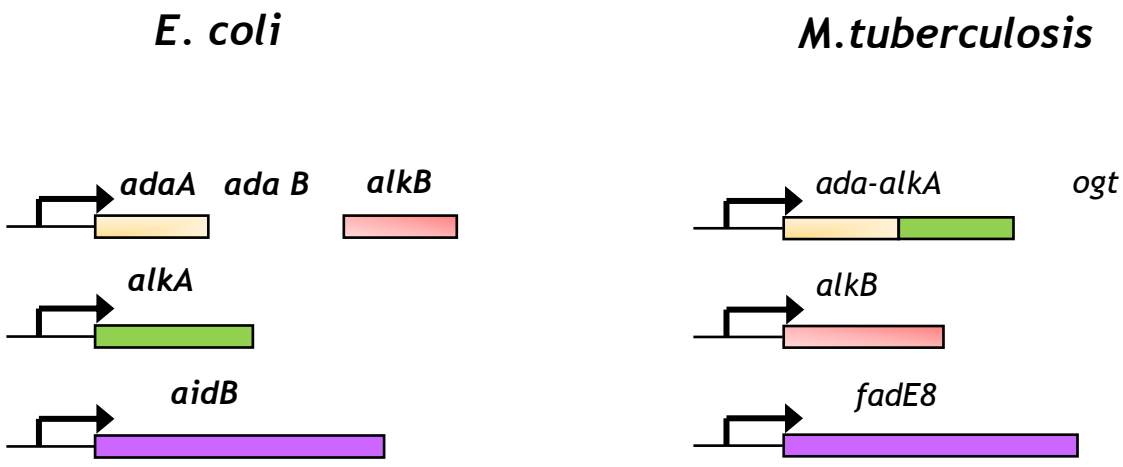

Figure 10: Schematic organization of chromosomal organization of the adaptive response genes *in E. coli* and MTB.

In MTB N-terminal domain of Ada (Ada A) is fused with AlkA and C-terminal domain (AdaB) constitutes a separate protein annotated as ogt. In both AdaA-AlkA and Ogt there are critics residues conserved.

The most interesting features includes the zinc finger of AdaA, the active site thiol of AdaA (Cys38 in *E. coli*, Cys34 in *M. tuberculosis*) and AdaB (Cys321 in *E. coli*, Cys126 in *M. tuberculosis*), the helix-hairpin-helix motifs (HhH) of AdaB and AlkA, the arginine finger of AdaB as well as the catalytic aspartic acid (Asp238 in *E. coli*, Asp441 in *M. tuberculosis*) of AlkA.

Exposure of MTB to a methylating molecule strongly increases transcription of adaA-alkA and Ogt, demonstrating an inducible response to methylating agents.

Based on homology the AdaA-AlkA protein is predicted to transfer methyl groups from methylphospotriesters of the DNA backbone to a conserved cysteine residue (Cys34) of the AdaA domain, suggesting that AdaA-AlkA is converted to a positive transcriptional regulator of the adaptive response to DNA alkylation damage in *M. tuberculosis* [44] while recent studies demonstrated that Ogt acts through a suicidal mechanism, by performing the stoichiometric transfer of the O6-alkyl group from the modified guanine to a strictly conserved cysteine residue (Cys 126) in the protein active site, which is hosted in the C-terminal domain of the protein (fig. 11). This covalent modification leaves OGT permanently inactivated and possibly more prone to degradation [45].

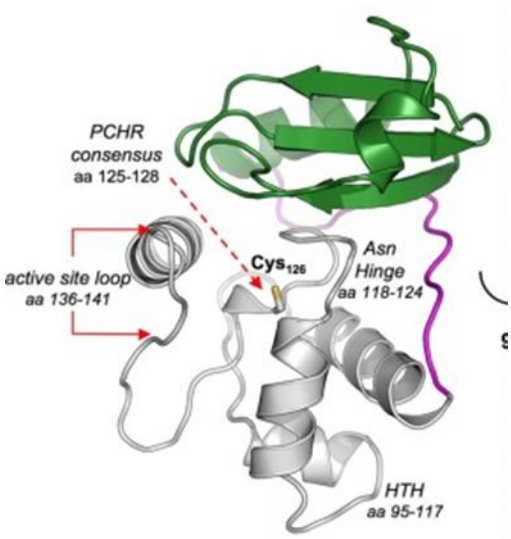

Figure 11: Ogt structure

The investigation of the DNA repair and protection mechanisms is really interesting in the tuberculosis context because of their involvement in the drug resistant development. Moreover, proteins involved in these mechanisms represent excellent therapeutic targets because of the fundamental role they play in bacterial survival and due to the absence of homologues in humans [46].

3. Aims

Tuberculosis (TB) is an infectious disease caused by the micro-organism *Mycobacterium tuberculosis* (MTB) that currently represents a global health emergency with around 9 million people affected and nearly 2 million deaths each year. Current therapeutic treatment for TB consists in the use of multiple anti-mycobacterial drugs such as rifampicin and isoniazid. This treatment is carried out for a long time leading to the onset of drug- and multidrug-resistant strains of MTB. Therefore new therapeutic models are needed to overcome this problem. During MTB infection, the host antimicrobial response generates several metabolically activated DNA alkylating agents leading to severe DNA-damaging injuries on MTB cells. Protection of the DNA molecule from chemical damages then strictly depends on the MTB repair mechanisms. Recent studies identified an adaptive response mechanism in MTB homologous to that already described in *E.coli*. Being the adaptive response a fundamental biological process for MTB viability whereas it is missing in human, this process may represent a putative therapeutic target to explore in search for new TB treatments. Previous studies in MTB identified four genes which encode the proteins constituent of the adaptive response mechanism homologous to the same process already defined in *E. coli*. Exposure of *E. coli* to sublethal concentrations of alkylating agents induces the expression of four genes (*ada*, *alkA*, *alkB* and *aidB*). The activation of these genes increases the resistance to the cytotoxic and mutagenic effects of alkylating molecules. Unlike *E. coli*, the DNA repair/protection systems in MBT have not been investigated and it is still largely unknown.

This thesis project focuses on the investigation of the adaptive response to DNA methylation stress in MTB.

In the first part of this project the effect of methylating agents on Mycobacterium smegmatis cells was investigated. It was evaluated on cell survival by monitoring the growth profile of bacterial cultures grown in the presence and in the absence of different concentrations of methylmethansulfonate (MMS),a common laboratory methylating molecule. The effect of DNA methylation was also quantitatively monitored by Multiple Reaction Monitoring tandem MS.

Then, the global proteomic response of *M.smegmatis* to methylation stress was investigated by performing a quantitative proteomic analyses in order to identify and functional analyze the differentially expressed proteins in the presence and in the absence of the methylating molecule.

The second purpose of this thesis is to characterize the proteins of the adaptive response in MTB.

Ada-AlkA and Ogt proteins were structurally and functionally characterized because of their key role in the cellular response to alkylating agents.

Lastly, functional proteomic experiments on *M.smegmatis* Ogt protein was performed in orderto shine a light on its biological function through the identification of its protein partners *in vivo.*

Chapter 2:

Experimental section

1. M. smegmatis growths

Mycobacterium smegmatis is an acid-fast bacterial species in the phylum Actinobacteria and the genus *Mycobacterium*. It is useful for the research analysis of other *Mycobacteria* species in laboratory experiments because of its being a "fast grower" and non-pathogenic. *M. smegmatis* is a simple model that requires only a biosafety level 1 laboratory. For these reasons, *M. smegmatis* is usually used as a model for mycobacterial species.

Wild type cells were grown 3 days in LB medium at 37 °C and they were diluted 1:100 in fresh medium containing 100µg/mL ampicillin, 0.05% tween 80. At an A600 nm of 0.4, the cultures were divided in aliquots and one of these was manteined as the untreated control while the others were supplemented with MMS in a 0.01-0.1% w/v range. Growing profiles were obtained by monitoring cells for 35 hours.

2. Genomic DNA extraction

Cellular pellets were treated as described by Telenti and co-workers to extract genomic DNA (Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. A Telenti, F Marchesi, M Balz, F Bally, E C Böttger, and T Bodmer).

3. LC-MS/MS analyses of methylated bases

The DNA samples were subjected to acidic hydrolysis in 0.1 M HCl at 80 °C for 30 min to obtain free bases. The samples were dried under vacuum and solubilized in 50 µL of 3% (v/v) methanol (MeOH)/0.1% (v/v) trifluoroacetic acid for methylated purine analysis. Standard methylated bases (Sigma-Aldrich) were solubilized in 3% (v/v) methanol (MeOH)/0.1% (v/v) trifluoroacetic acid [39]. Samples were then analyzed by reverse-phase HPLC associated with an Agilent Triple quadrupole 6420 used in the multiple reaction monitoring mode. The C18 column was eluted with a flow rate of 0.2ml/min starting with 3% (v/v) MeOH/0.1% (v/v) formic acid. The column was delivered at 3% MeOH for 2 minutes. The percentage of MeOH reached 75% in 5 min and rapidly back to 3% MeOH.

The amount of methylated bases was obtained by external calibration in the range of 10-80 pg.

4. Bi-dimensional analyses of *M.smegmatis* cells

2D electrophoresis IEF (first dimension) was carried out on nonlinear wide-range immobilized pH gradients (pH 3-10; 18 cm long IPG strips; GE Healthcare, Uppsala, Sweden) and achieved using the Ettan IPGphor system (GE Healthcare, Uppsala, Sweden). Analytical-run IPGstrips were rehydrated with 60µg of total proteins in 350µl of lysis buffer and 0,2% (v/v) carrier ampholyte for 1h at 0 V and for 8h at 30 V, at 16°C. The strips were then focused according to the following electrical conditions at 16°C: 200 V for 1h, from 300 V to 3500 V in 30 min, 3500 V for 3h, from 3500 V to 8000 V in 30 min, 8000 V until a total of 80000 Vh was reached. For preparative gels 400µg of total proteins were used. After focusing, analytical and preparative IPG strips were equilibrated for 12 min in 6 M urea, 30% (V/V) glycerol, 2% (w/V) SDS, 0.05 M Tris-HCl, pH 6.8, 2% (w/V) DTE, and subsequently for 5 min in the same urea/SDS/Tris buffer solution but substituting the 2% (w/V) DTE with 2.5% (w/V) iodoacetamide. The second dimension was carried out on 10% polyacrylamide linear gradient gels at 40 mA/gel constant current until the dye front reached the bottom of the gel, according to Laemmli et al and Hochstrasser et al. Gels were stained with colloidal coomassie.

4.1 In situ digestion

Protein bands in 1D gels were reduced by incubation with 50 µL of 10 mM DTT in 0.1 M ammonium bicarbonate buffer (pH 7.5) and alkylated with 50 µL of 55 mM iodoacetamide in the same buffer. Enzymatic digestion was carried out with trypsin (12.5ng/µl) in 50mM ammonium bicarbonate pH 8.5 at 4°C for 4 hours. The buffer solution was then removed and a new aliquot of the enzyme/buffer solution was added for 18 hours at 37°C. A minimum reaction volume, enough for the complete rehydratation of the gel was used. Peptides were then extracted washing the gel particles with 20mM ammonium bicarbonate and 0.1% TFA in 50% ACN at room temperature and then lyophilized.

Spots of interest in 2D gels were excised, washed first with ACN, then with 0.1M ammonium bicarbonate and enzymatic hydrolysed by trypsin as just described.

4.2 LC-MS/MS Analyses

Tryptic peptide mixtures obtained from in situ digestions were analysed by LC/MS/MS using an HPLC-Chip/Q-TOF 6520 (Agilent Technologies). The peptide mixtures were injected by auto sampler. They were sent to the enrichment column of the chip at flow rate of 4 µL/min, in 98% water, 2% acetonitrile and 0.1% formic acid. Subsequently the peptides were eluted directly into the capillary column (C18 reversed phase), at a flow rate of 0.4 µL/min. The chromatographic separation was carried out with a linear gradient in 95% acetonitrile, 5% water and 0.1% formic acid. The eluate was then introduced in the ESI source for the tandem analysis. In this way each mass spectrum (range 300-2,400 m/z) was followed by one or more tandem mass spectra (range 100-2,000 m/z), obtained by fragmenting the most intense ions in each fraction eluted chromatographic. The acquired MS/MS spectra were transformed in Mascot generic file format and used for peptides identification with a licensed version of MASCOT (modular approach to software construction, operation and test, matrix science, USA), in a local database.

5. DIGE tecniques

DIGE experiments were performed on four biological replicates of MMS treated and four biological replicates of MMS untreated *M.smegmatis* as previously described [12]. Cells were homogenized in 0.5 ml of lysis buffer (7M urea, 2M thiourea, 4% chaps, 30mM Tris-HCl pH 7.5) using a Dounce homogenizer. The protein extract concentrations were determined and equal amounts of the protein lysates were then labeled in vitro using two different fluorescent cyanine minimal dyes (Cy3 and Cy5, respectively) differing in their excitation and emission wavelengths. A third cyanine dye (Cy2) was used to label a mixture of all samples as internal standard. The three differently labeled protein mixtures were pooled and subjected to isoelectric focusing through a pH range of 3-10 over a length of 24 cm. The reducing and alkylating steps were performed between the first and the second electrophoretic step.

Acrylamide strips were then transferred to the top of a classical SDS PAGE gel for a second orthogonal electrophoresis analysis. The Cy2, Cy3 and Cy5 images were obtained by scanning

each of the four DIGE gels at excitation/emission wavelength of 480/ 530 nm for Cy2, 520/590 nm for Cy3 and 620/680 nm for Cy5 using a Typhoon 9410 TM scanner (GE Healthcare). The semi-preparative gel, prepared in an identical fashion, was scanned with 480/633 nm wavelengths. After consecutive excitation at both wavelengths, the images from the preparative gel were overlaid and subtracted (normalized) from the samples, whereby only differences (up or down regulated proteins) between the two samples were visualized. By performing a high resolution image analysis on the six biological replicates, it was possible to visualize significant differences between numerous protein spots present on the gels. Differentially regulated spots were defined as having a variation higher than 1.2 ($p<0.05$) per previously established methods [13].The gels showed a high degree of similarity, with more than 80% of all spots superimposable. The remaining 20% showed variation and were further studied.

5.1 Proteomic analysis

The spots of interest were excised, hydrolyzed and the peptide mixtures analyzed by MALDI-MS and LC-MSMS mass spectrometry, on a 4800 Plus MALDI TOF/TOF™ (Applied Biosystems, Framingham, MA, USA) and a LC/MSD Trap XCT Ultra (Agilent Technologies, Palo Alto, CA) equipped with a 1100 HPLC system and a chip cube (Agilent Technologies) respectively. MALDI spectra were acquired in the positive ion reflector mode using delayed extraction in the 800 - 4000 Da mass range. LC-MS/MS analysis was performed using data-dependent acquisition of one MS scan followed by MS/MS scans of the three most abundant ions in each MS scan. Raw data analyses were converted into a Mascot format text to identify proteins using Matrix Science software. The protein search considered the following parameters: non-redundant protein sequence database (NCBInr), specificity of the proteolytic enzyme used for the hydrolysis (trypsin), taxonomic category of the sample, no protein molecular weight was considered, up to one missed cleavage, cysteines as S-carbamidomethylcysteines, unmodified N- and C-terminal ends, methionines both unmodified and oxidized, putative pyro-Glu formation by Gln, precursor peptide maximum mass tolerance of 200 ppm, and a maximum fragment mass tolerance of 200 ppm.

5.2 Bioinformatic analysis of identified proteins

Identified proteins were used to determine predicted interactions with other proteins. This functional protein association network for the each entry was obtained by searching the STRING online database (http://string-db.org). Proteins were analyzed with the STRING software.

6. Static biofilm assay

The wells of a sterile 96-well flat-bottomed polystyrene plate (Falcon) were filled with 90 mL of the appropriate medium containing or not containing the inhibitors. 10 mL of overnight bacterial cultures grown in LB was added into each well. The plates were incubated aerobically with or without the enzyme for 24 h at 37 °C in the presence of either 0.03% MMS. Growth was monitored by measuring the OD600, and after 24 h incubation the ability of the *M.smegmatis* strain to adhere to the polystyrene plates was tested. The content of the plates was then poured off and the wells washed with sterile distilled water. The plates were then stained with crystal violet for 5 min. Excess stain was rinsed off by placing the plate under running tap water. After the plates were air dried, the dye bound to the adherent cells was solubilized with 20% (v/v) glacial acetic acid and 80% (v/v) ethanol per well. The OD of each well was measured at 590 nm.

7. Construction of expression vectors

The *ogt* and *ada-alkA* genes of *M. tuberculosis* were amplified from genomic DNA by PCR. *ogt* amplification product was directly ligated with a pGEM Teasy vector. Then to obtain protein tagged with GST protein, the pGEM Teasy-ogt construct was digested with SmaI and BamHI and cloned into the pGEX4T1 vector. *ada-alkA* amplification product was directly hydrolysed with BamHI and XhoI and cloned into the pGEX4T1 vector. Plasmids construction was verified by automated DNA sequencing.

7.1 Plasmidic vectors

The pGEM®-T Easy Vectors is a linearized vector with a single 3´-terminal thymidine at both ends. The T-overhangs at the insertion site greatly improve the efficiency of ligation of PCR

products by preventing recircularization of the vector and providing a compatible overhang for PCR products generated by certain thermostable polymerases. The pGEM®-T Easy Vector is a high-copy-number vectors containing T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. Insertional inactivation of the α-peptide allows identification of recombinants by blue/white screening on indicator plates.

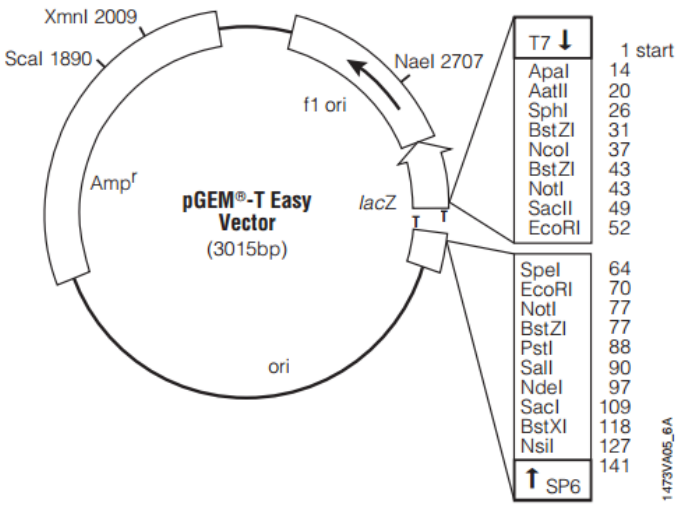

Figure A : pGEM®-T Easy Vector Map

The pGEX-4T-1 vector contains a tac promoter for chemically inducible, high-level expression of Glutathione S Transferase tagged recombinant protein. Moreover this vector permits to obtain Thrombin recognition site for cleaving the desired protein from the fusion product.

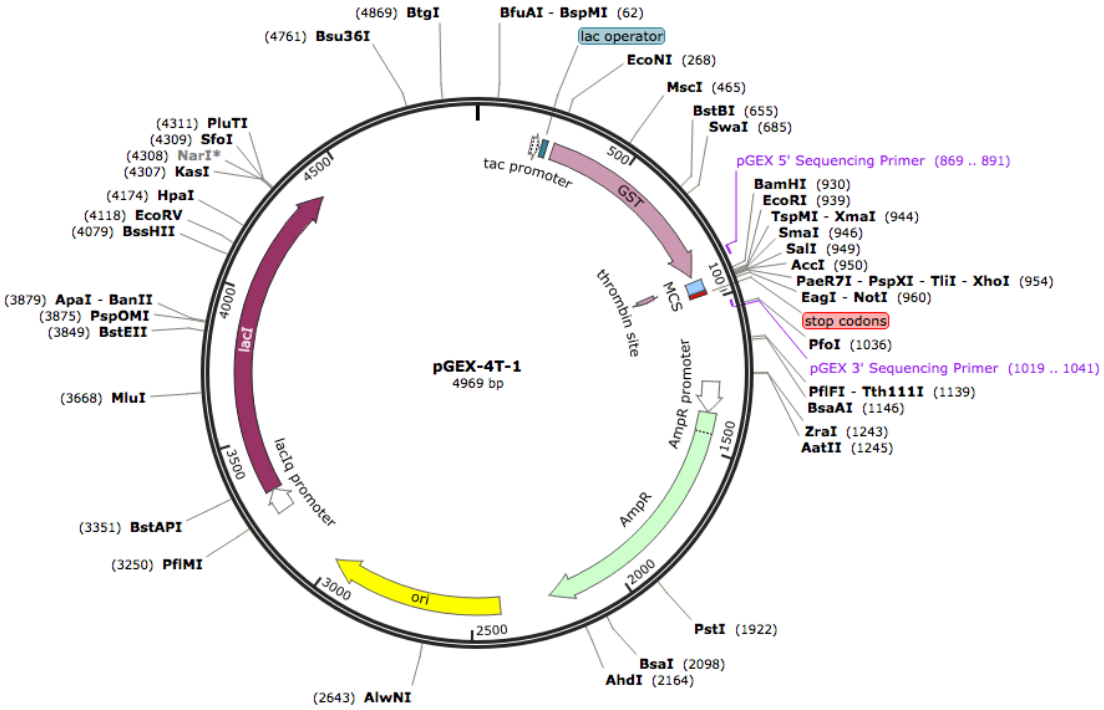

Figure B: pGEX-4T-1 vector map

8. Production and Purification of Recombinant Proteins

BL21 and C41 *E.coli* recombinant cells were grown at 37 °C to an OD (optical density) at 600 nm of about 0.5, at which time 0.1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) was added in order to express *ogt* and *ada-alkA* genes. Selective antibiotic was used at concentration of 100 µg/mL ampicillin. Cells were grown at 20°C for 16 hours to obtain recombinant expression of *ogt.* After incubation, cells were harvested by centrifugation at centrifugal force of 5,000 x g for 15 min at 4 °C, resuspended in Na_2PO_4 50mM, NaCl 150 mM, 1 mM PMSF pH = 7.4 and disrupted by passage through a French press and centrifuged at centrifugal force of 14000 rpm for 15 min at 4 °C. Recombinant proteins were purified by affinity chromatography on Glutathione-Agarose beads (Sigma). After 2 hours of incubation at 4 °C, the matrix was collected by centrifugation at centrifugal force of 11000 rpm for 1 min and washed three times with Na_2PO_4 50mM, NaCl 150 mM and Triton 1%. The recombinant proteins were eluted with Tris –

HCl buffer pH = 8 containing reduced glutathione 1 mM. Protein concentration was estimated with Bradford reagent (Bio-Rad protein assay) and protein content was checked by SDS-PAGE.

9. Circular Dichroism analyses

Proteins were equilibrated in 20 mM sodium phosphate buffer pH 7.4 and analysed. Circular dichroism (CD) measurements were recorded at 25 °C, using a J-750 spectropolarimeter equipped with a Peltier type temperature control system (Model PTC-348WI). Cells of 0.1 cm path length were used in all measurements.

10. Electrophoretic mobility shift assay

DNA probe of 28 bp having random sequence was used. Sense and antisense oligonucleotides were annealed by incubation at 95 °C for 5 min and successive gradual cooling to room temperature. Purified recombinant Ogt or Ada-AlkA were incubated with the probes for 20 min at room temperature in 20 µl of buffer Z (25 mM HEPES pH 7.6, 50 mM KCl, 12.5 mM MgCl2, 1 mM DTT, 20% glycerol, 0.1% triton). Protein–DNA complexes were separated on 1.5% agarose gel, containing Gel Red Dye 1x, in 0.5 × TBE (45 mM Tris pH 8.0, 45 mM boric acid, 1 mM EDTA) at 80 V at room temperature. Visualization of DNA samples was carried out by using UV radiation.

11. Pull down experiment

E.coli cells overproducing GST-Ogt were grown at 37 °C to an OD at 600 nm of about 0.5, at which time 0.1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) was added in order to express the gene. Cells were grown at 20°C for 16 hours to obtain recombinant expression of *ogt.* After incubation, cells were harvested by centrifugation at centrifugal force of 5,000 x g for 15 min at 4 °C, resuspended in Na_2PO_4 50mM, NaCl 150 mM, 1 mM PMSF pH = 7.4 and disrupted by passage through a French press and centrifuged at centrifugal force of 14,000 x g for 15 min at 4 °C. Recombinant proteins were immobilized by affinity chromatography on Glutathione-Agarose beads as the bait of pull down experiment. Simultaneously *M.smegmatis* wild type cells were grown in LB medium containing 100µg/mL ampicillin, 0.05% tween 80. At an

A600 nm of 0.4, the cultures were divided in two aliquots and one of these was maintained as the untreated control while the other was supplemented with 0.03% MMS. After 3 hours cells were recovered by centrifugation at 4000 rpm at 4°C for 15 minutes. Cells were disrupted Na_2PO_4 50mM, NaCl 150 mM, 1 mM PMSF pH = 7.4 by passage through a French press and centrifuged at 14000 rpm for 30 min at 4 °C. Cellular extracts from both samples were subjected to a pre-cleaning step by incubation with Glutathione-Agarose beads in order to isolate non specific binding proteins. Then treated and untreated *M.smegmatis* extracts were incubated with Glutathione-Agarose immobilized GST-Ogt protein to allow protein complexes formation. After 16 hours of incubation both the pre-cleaning and affinity chromatography resins were recovered and washed with Na_2PO_4 50mM, NaCl 150 mM and Triton 1%.

Ogt protein complexes were eluted with Tris – HCl buffer pH = 8 containing reduced glutathione 1 mM. Protein concentration was estimated with Bradford reagent and protein content was checked by SDS-PAGE and western blot analyses.

12. Western blot analyses

Protein complexes were resolved by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions using standard procedures (Sambrook & Russell, 2001). The proteins were transferred on to a polyvinylidene difluoride (PVDF) membrane using an electroblotting transfer apparatus (Trans-Blot Semi-Dry Transfer Cell, Bio-Rad). GST tagged proteins were detected using the anti-GST mouse monoclonal antibody and peroxidase-conjugated anti-mouse secondary antisera. The membrane was developed using SuperSignal West Femto Maximum Sensitivity Substrate detection kit (Pierce) according to the manufacturer'sinstructions.

Chapter 3:

Effect of methylating agents on *Mycobacterium smegmatis* cells

1. Effect of methylating agent on *Mycobacterium smegmatis* growth

The effect of methylating molecules on *M. smegmatis* cells was investigated in order to evaluate whether limited doses of methylating agent might affect *M. smegmatis* cell growth. Methyl methansulfonate (MMS) was used as methylating molecule.

To evaluate the lowest amount of reagent clearly affecting cells growth, *M. smegmatis* cells were grown in the presence and in the absence of different concentrations of MMS in a 0.01-0.1 % w/v range and the viability of bacterial cultures was monitored for 24 hours. Figure 12 clearly shows a decrease in cell survival with increasing doses of MMS in comparison with untreated cells (CN).

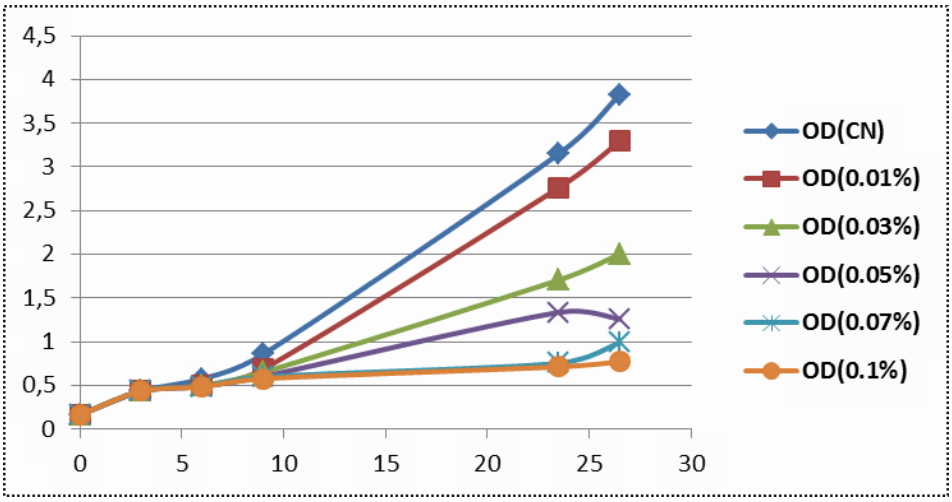

Figure 12: Growth profile of *M. smegmatis* cells in the presence and in the absence of several concentrations of MMS

These data demonstrated that 0.03% MMS concentration led to a decrease of about 50% in cell survival, indicating that in these conditions *M. smegmatis* cultures were affected by alkylation. This concentration was then selected for further experiments.

2. Quantitative analyses of DNA methylation in *M. smegmatis* cells

In order to examine the effect of methylation stress on DNA molecule, DNA methylation was quantitatively assessed by LC-MS/MS analysis in Multiple Reaction Monitoring mode, in the presence and in the absence of methylation stress [47].

N-7-methylguanine (7MeG), O6-methylguanine (6MeG) and N-3-methyladenine (3MeA) are the most abundant DNA base modifications by methyl methane sulfonate (MMS). This method is able to discriminate between the two isomeric structure O6-methylguanine and N7-methylguanine (7MeG) by the two different fragment ions at m/z 134 and 124 respectively, in order to identify and quantify the individual methylated DNA bases.

Individual calibration curves were constructed for each modified base using standard solutions.

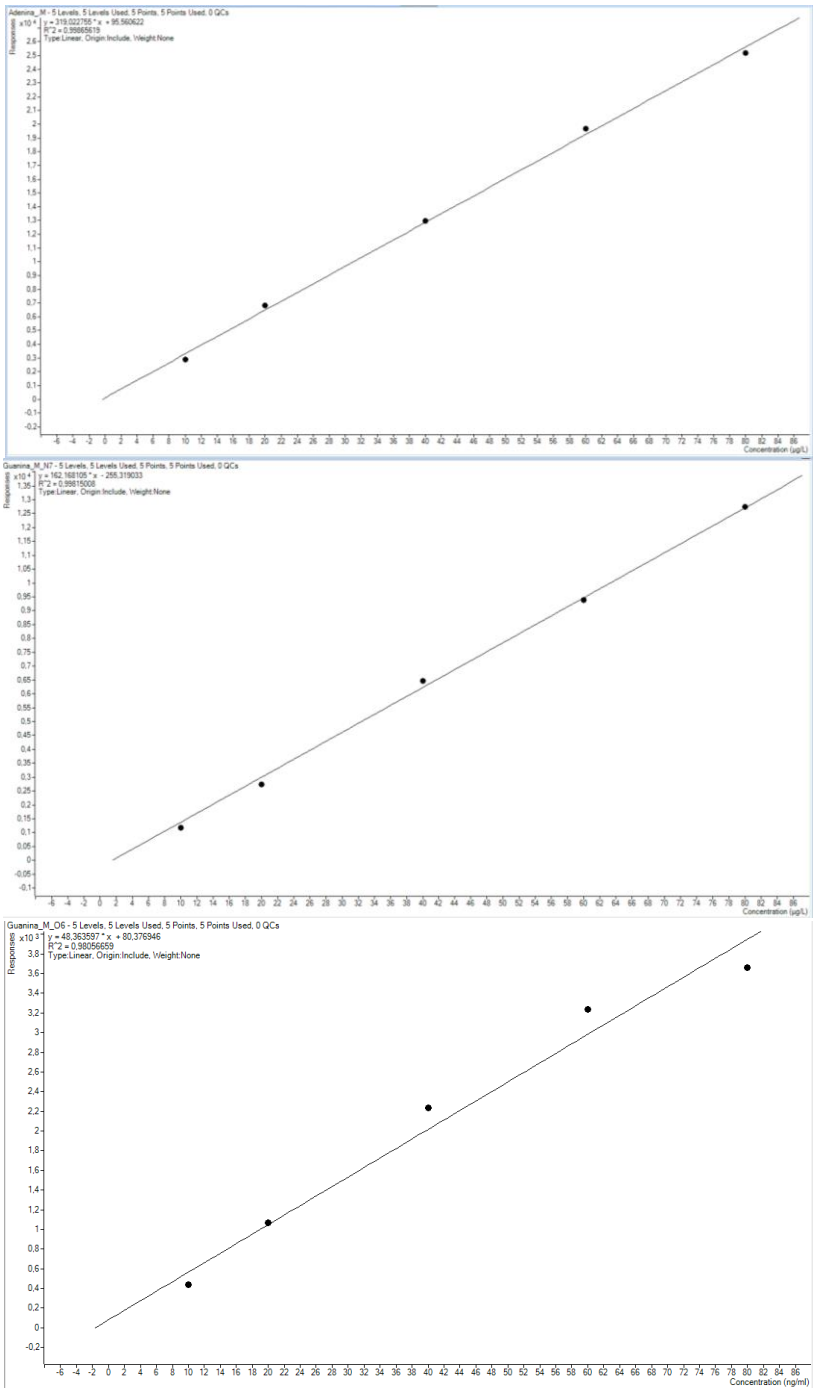

Figure 13. Calibration curves obtained by using 10, 20, 40, 60, 80 µg/L of methylated standard bases: 1) N-3-methyladenine, 2) N-7-methylguanine and 3) O6-methylguanine.

The MRM transitions used in the analytical method are reported in Table 1.

Multiple Reaction Monitoring (MRM)			
Methylated bases	**Precursor ion (m/z)**	**Fragment ions (m/z)**	**Loss**
O-6-methylguanine	166	134 139	CH3OH NH3
7-methylguanine	166	124 149	NH2CN NH3
3-methyladenine	150	108 123	CH3HCN HCN

Table 1: MRM transitions used for quantitative analyses

M.smegmatis cells were grown in the presence and in the absence of MMS. Cells were recovered at three different time (3, 6 and 9 hours) to extract genomic DNA samples that were acidic hydrolyzed and the released methylated bases submitted to LC-MS/MS analysis in MRM scan mode. Figure 14 reported DNA methylation data and cell survival at three different times upon exposure to MMS.

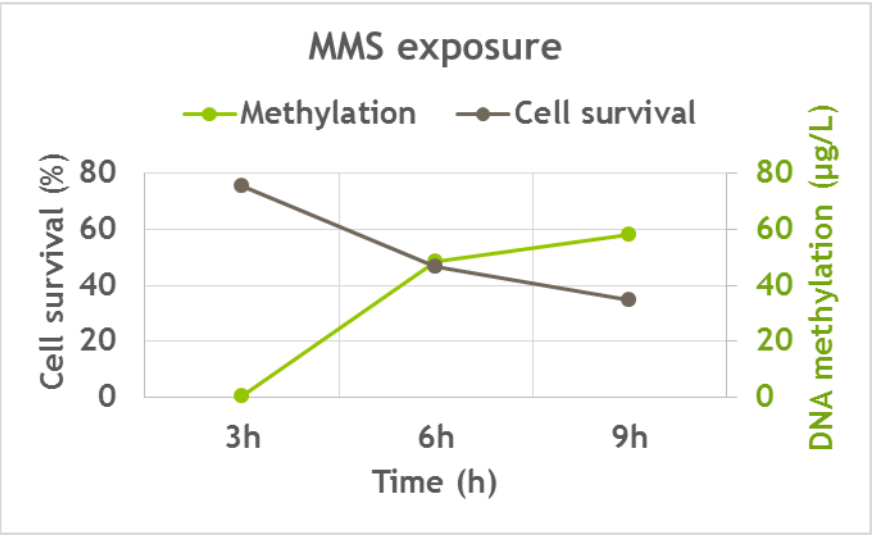

Figure 14: DNA methylation of M.smegmatis cells is reported in the presence of MMS in green while cellular survival is reported in grey over three different time: 3, 6 and 9 hours upon exposure to methylating agent.

Quantitative analyses of DNA methylation showed an increasing amount of base modification over time demonstrating that increasing DNA methylation induces a proportional widespread cellular death.

3. Bidimensional electrophoresis analyses

In order to investigate the global proteomic response to methylation stress, a basic two dimensional electrophoresis analysis was carried out.

M. smegmtis cells were grown in the absence and in the presence of 0.03% MMS and collected after 3 hours incubation. Cells were lysated and the protein extracts submitted to a bi-dimensional electrophoresis (2-D electrophoresis) analysis.

2-D electrophoresis is a method for the analysis of complex protein mixtures extracted from cells, tissues or other biological samples. Proteins are separated according to two independent and orthogonal fisico-chemical properties. The first-dimension is an isoelectric focusing (IEF) separation of proteins according to their isoelectric points (pI) whereas the second dimension consists in an SDS-polyacrylamide gel electrophoresis (SDS-PAGE) that separates proteins according to their molecular weights.

The two gels were stained with colloidal Coomassie and submitted to quantitative image analysis using a specific software to evaluate proteins up and down-regulated in the presence of MMS.

Protein spots corresponding to protein differently expressed in the two conditions were excised from the gel, *in situ* digested and the resulting peptide mixtures analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) approaches that generates data on both the accurate molecular mass and the sequence of individual peptides.

These data were used for database searches using a home version of the Mascot software leading to the identification of the proteins.

These informations were then used to search non-redundant protein databases leading to the identification of the protein components. Proteins identified in the proteomic experiments in

the absence of methylation stress are listed in table 2 where the protein name, the corresponding Swiss Prot code and the number of identified peptides are reported.

	Protein description	**Swiss prot**
1	Elongation factor Tu	A0QS98
2	Elongation factor Tu	A0QS98
3	60 kDa chaperonin 1	A0QQU5
4	ATP synthase subunit alpha	A0R202
5	Trigger factor	A0R199
6	Elongation factor Ts	A0QVB9
7	F420-dependent glucose-6-phosphate dehydrogenase	A0QQJ4
8	Ferrochelatase	A0QX29
9	ATP synthase subunit b	A0R204
10	NADP-dependent alcohol dehydrogenase C 2	P0CH37
11	DNA-directed RNA polymerase subunit alpha	A0QSL8
12	NADP-dependent alcohol dehydrogenase C 2	P0CH37
13	3-oxoacyl-[acyl-carrier-protein] reductase FabG	P71534
14	Acetylornithine aminotransferase	A0QYS9
15	3-oxoacyl-(Acyl-carrier-protein) synthase 1 KasA	A0R0B4
16	Chaperone protein DnaK	A0QQC8
17	Probable monoacyl phosphatidylinositol tetramannoside-binding protein LpqW	A0R2I8
18	Putative succinate-semialdehyde dehydrogenase [NADP(+)]	A0R4Q0
19	Aconitate hydratase	A0QX20

Table 2: Proteins identified in the bi-dimensional analysis.

About twenty differentially expressed proteins were identified in the two conditions. Most of the up-regulated proteins following MMS treatment were found to be involved in the cell wall components biosynthesis.

Particularly, were identified FabG and KasA proteins involved in the fatty acid biosynthesis and LpqW protein involved in the LAM synthesis, another cell wall components.
Also the protein KasA and another identified protein Ef- tu are related with biofilm formation.

4. DIGE (Differential In Gel Electrophoresis)

The previous interesting findings were deeply investigated by performing a quantitative proteomic analyses using the DIGE technology.
2D-DIGE is based on fluorescence pre-labeling of protein mixtures before 2D gel electrophoresis. Protein samples are each labeled with different dyes, cyanines, that have the same molecular weight, but different excitation/emission value. The labeled proteins are mixed, run together and separated simultaneously on the same 2D gel. The images are acquired at the respective excitation/emission values. Then the images are superimposed and the resulting map is analysed by specific software that carries out the detection and quantification of protein spots.
The ability to separate more than one sample on a single gel make 2D DIGE a highly reproducible technique. And then, the presence of an internal standard and the possibility to carry out an unique experiment for every biological replicate improve the accuracy and the sensitivity of this technique.
According to previous results, proteomic experiments were performed after 3 hours of MMS treatment. Each culture was divided into two aliquots: one of these was kept as control, and the other one was exposed to MMS 0.03%. Following 3 hours incubation, protein lysates from MMS treated and untreated cells were extracted and fractionated by bidimensional electrophoresis using a 3-10 pH gradient. The spots of interest were picked, *in situ* hydrolyzed with trypsin and the resulting peptide mixtures identified by LC-MS/MS analyses. Mass spectral data were used to search a non-redundant protein database by Mascot software.
A total of 71 differentially expressed proteins were identified in MMS treated *M.smegmatis* cells compared to controls, 28 down and 43 up regulated (table 3).

Fold Change	Protein	Gene	Swiss Prot

-1.14	DNA or RNA helicase of superfamily protein II	MSMEG_5706	A0R451
-1.25	1-deoxy-D-xylulose-5-phosphate synthase	dxs	A0QW19
-2,34	Alpha/beta hydrolase	MSMEG_0824	A0QQP0
-2,02	Phosphoenolpyruvate carboxylase	ppc	A0QWX4
-2,02	Protein translocase subunit SecA 2	secA2	A0QYG9
-1.90	Piperideine-6-carboxylic acid dehydrogenase pcd	MSMEG_1762	A0QT96
-1.79	Inosine-5'-monophosphate dehydrogenase	guaB	A0QSU3
-1,74	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
-1,61	Alpha-1,4-glucan:maltose-1-phosphate maltosyltransferase	glgE	Q9RP48
-1,61	Acyl-CoA synthase fad32	MSMEG_6393	A0R618
-1.60	Uroporphyrinogen decarboxylase	hemE	A0QW23
-1,6	Phospho-2-dehydro-3-deoxyheptonate aldolase	MSMEG_2799	A0QW41
-1,56	Sulfate adenylyltransferase subunit 2	cysD	A0R242
-1,56	6-phosphofructokinase PfkA	MSMEI_2306	I7G6C4
-1.55	Isocitrate dehydrogenase icd2	MSMEG_1654	A0QSZ3
-1,50	3-hydroxyacyl-CoA dehydrogenase fadB	MSMEG_5720	A0R465
-1,47	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
-1,39	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
-1.30	GTP-binding protein TypA	typA	A0R2J0
-1,28	Alpha/beta hydrolase	MSMEG_0824	A0QQP0
-1,25	Endopeptidase IV	sppA	A0QSH0
-1,25	Uroporphyrinogen-III synthase	hemD	0QR19
-1,17	F420-dependent oxidoreductase	MSMEG_5715	A0R461
-1,17	Adenosine deaminase	add	A0QT14
-1,14	1-pyrroline-5-carboxylate dehydrogenase	pruA	A0R2H8
-1,14	Fumarate reductase	sdhA	A0QT08
-1,13	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
-1,07	Acyl-CoA oxidase	MSMEG_4474	A0R0Q9
1,03	Adenosine deaminase	add	A0QT14
1,03	Putative phenylalanine aminotransferase	pat	A0R5X8
1,09	ABC-type sugar transport system ATPase component	MSMEG_6019	A0R505
1,15	Acyl-CoA synthase	MSMEG_4301	A0R090
1,15	Fumarate reductase	sdhA	A0QT08
1,16	Ribonuclease J rnj	MSMEG_2685	A0QVT2
1,28	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
1,29	Pyruvate dehydrogenase	MSMEG_3964	A0QZB5
1,29	Ribonuclease J rnj	MSMEG_2685	A0QVT2
1,29	60 kDa chaperonin 1	groL1	A0QQU5
1,29	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
1,30	3-ketoacyl-CoA thiolase fadA2	MSMEG_0373	A0QPE8
1,32	3-Hydroxyacyl-CoA dehydrogenase	MSMEG_5183	A0R2P1
1,35	D-alanyl-D-alanine dipeptidase vanX	MSMEG_5879	A0R4L7
1,49	Alanine dehydrogenase	ald	A0QVQ8

1,51	Inosine-5'-monophosphate dehydrogenase	guaB	A0QSU3
1,56	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
1,58	Inositol-3-phosphate synthase	ino1	A0R7G6
1,63	DNA gyrase subunit B	gyrB	A0QNE0
1,67	Divalent metal cation transporter MntH	MSMEG_5589	A0R3T6
1,68	IclR family transcriptional regulator	MSMEG_3335	A0QXK4
1,69	Bifunctional protein GlmU	glmU	A0R3C7
1,75	Alkanal monooxygenase alpha chain	MSMEG_5029	A0R293
1,75	4-hydroxy-2-oxovalerate aldolase 2 bbhI-2	MSMEG_5937	A0R4S6
1,78	Serine/threonine protein kinase	MSMEG_5513	A0R3L2
1,81	Fructose-1,6-bisphosphatase	glpX	A0R2U7
1,81	Acetyl-CoA acetyltransferase fadA3	MSMEG_5273	A0R2Y1
1,81	NADPH:quinone reductase	MSMEG_5164	A0R2M2
1,82	Ribonucleoside-diphosphate reductase subunit alpha 1 (nrdE1)	MSMEG_1019	P0CG99
2,17	Acyl-CoA ligase FadD31	MSMEI_4200	I7G4U9
2,18	Endopeptidase IV	sppA	A0QSH0
2,4	3-isopropylmalate dehydratase small subunit leuD	MSMEG_2388	A0QUZ0
2,4	Peptidyl-prolyl cis-trans isomerase ppiA	MSMEG_0024	A0QNF6
2,47	D-ribose-binding periplasmic protein	MSMEG_4172	A0QZW4
3,33	Glycine dehydrogenase (decarboxylating)	gcvP	A0QYF7
3,33	DNA polymerase III gamma/tau subunit dnaX	MSMEG_6285	A0R5R6
3,33	Alanine--tRNA ligase	alaS	A0QWQ4
4,61	Conserved hypothetical proline and alanine rich protein	MSMEG_0067	A0QNJ7
1,78	Ribonucleoside-diphosphate reductase subunit alpha 1 (nrdE1)	MSMEG_1019	P0CG99
1,78	LpqN	MSMEI_5361	I7G7W9
2,32	Chaperone protein DnaK	dnaK	A0QQC8
2,17	Acyl-CoA synthase	MSMEG_4301	A0R090
1,33	Carbohydrate kinase FGGY	MSMEI_1337	I7F889

Table 3: Differentially expressed proteins identified

The identified proteins were functionally classified according to their known biological functions in figure 15.

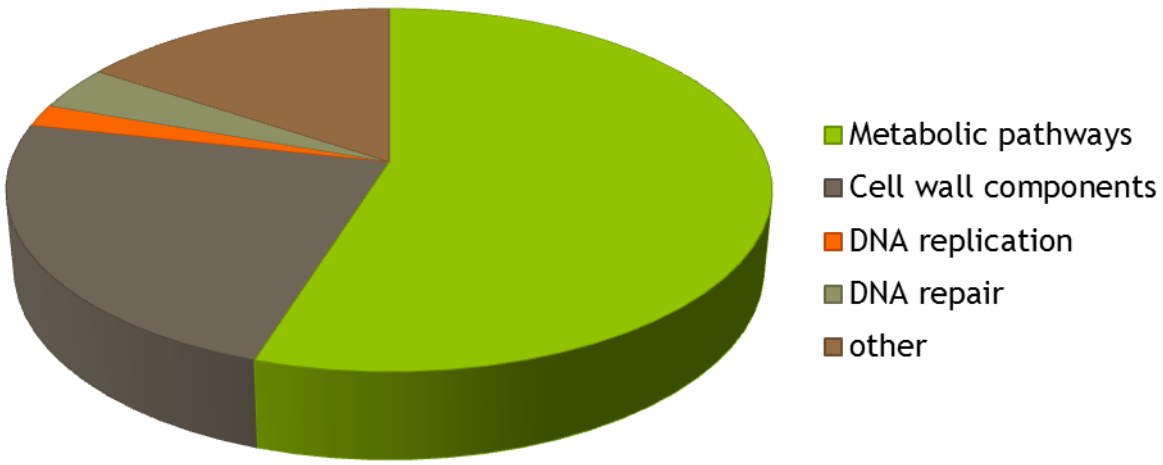

Figure 15: Clustered identified proteins are reported

Most of identified proteins gathered within cell wall processes and metabolic processes. The network distributions of 71 differentially expression proteins were explored using STRING software (fig 16).

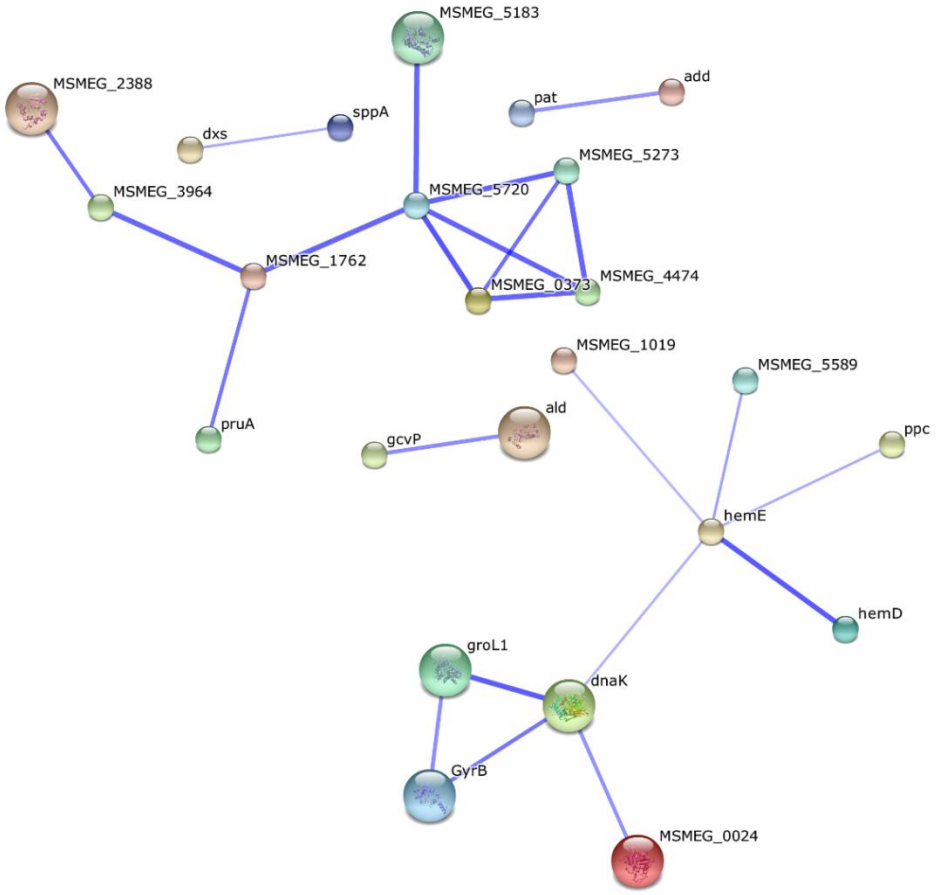

Figure 16: Network distribution of the 71 identified protein as obtained by STRING. The stronger associations are represented by thicker lines.

Functional investigation showed that some proteins are involved in the same pathways. Particularly in table 4 are listed the most interesting protein classes identified by literature studies.

Biofilm formation	**Fatty acid degradation**	**Cell wall biosynthesis**
dnaK	fadA2	ald
gyrB	fadA3	vanX
pknF	fadB2	alaS
ppiA	fadD31	glnU
groL1	fadD32	gcvP
leuD		

Table 4. Classification of protein identified in according to their biological functions.

Proteins involved in biofilm formation are all up regulated in the presence of MMS suggesting a physical response to prevent MMS access in the cell environment. The biofilm is an aggregate of cells adherent on a surface in a polymeric extracellular matrix. Biofilm formation requires two steps:

1. Adherence of bacterial cells to a surface
2. Accumulation of multilayered cell clusters

This bacterial attachment mechanism is required to adhere directly to host organism and it is involved in infectious processes for most of pathogen microorganisms.

5. Biofilm formation assay

The increasing of proteins involved in biofilm formation in the presence of methylation stress was explored by performing specific assay in the presence and in the absence of MMS to evaluate the amount of biofilm formed by *M.smegmatis*.

Static growth conditions were preliminary optimized in order to accurately evaluate biofilm formation. *M.smegmatis* strain was inoculated on wells in the presence and in the absence of 0.03% MMS. In vitro quantification of attached cells was determined after 24 hours of incubation at 37° C by crystal violet staining and subsequent measurement of the absorbance at 590 nm. The results are reported in figure 17.

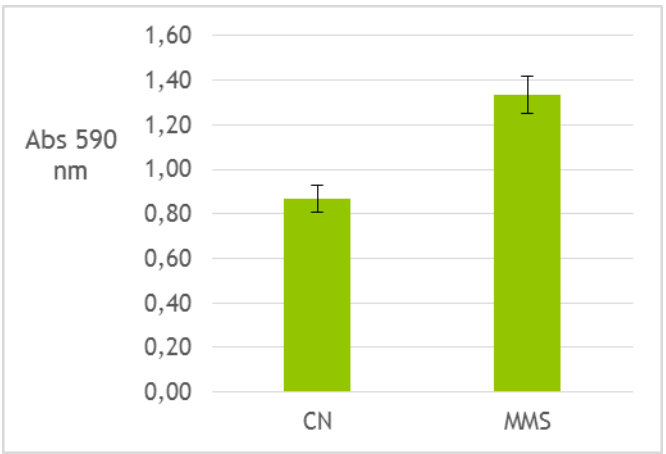

Figure 17. Biofilm formation assay in the presence and in the absence of MMS.

The presence of MMS increased biofilm production compared to untreated cells. This finding was in agreement with proteomic results that suggested under stress conditions *M. smegmatis* responses to damaging events increasing physical cellular defense mechanisms

Chapter 4:

Structural and functional analyses of proteins involved in the adaptive response

1. Homologous expression of Ada-Alka protein

Recombinant expression of the adaptive response protein was carried out homologously in *M.smegmatis.* Cloning strategy was performed in two steps, the first by ligating *ada-alkA* gene in an *E.coli* vector and the second by cloning into a *M.smegmatis* expression vector.

The selected system for cloning was pGEM®-T Easy Vector, a system used to clone directly PCR product. This vector is linearized with a single 3'-terminal thymidine at both 5' and 3' DNA double strand. The presence of two thymidine at the insertion site does not allow the circularization of the vector increasing the efficiency of ligation of PCR products. The *ada-alkA* gene was amplified by PCR from MTB genome in order to have a single 3'-terminal adenine at both 5' and 3' DNA double strand. The *ada-alkA* gene was hydrolyzed from the recombinant pGEM®-T-ada-alkA and cloned into a *M. smegmatis* vector pMV338.

M. smegmatis cells were finally transformed with the recombinant plasmid by electroporation and then grown to allow the production of recombinant protein obtained as Histidine tagged protein. The his-tagged Ada-AlkA was purified by Immobilized Metal Ion Affinity Chromatography, exploiting the interaction between the histidine tag on the recombinant protein and Ni2+ ions coupled to highly cross-linked agarose beads. SDS- PAGE analysis showed that the amount of soluble protein was too low to carry on with further experiments.

2. Heterologous expression of Ada-Alka and Ogt

Because the homologous expression yield was too low, Ada-AlkA was heterologously expressed in *E.coli*. The gene was cloned into different expression systems: pET22b, pET28a, pET16b and pGEX4T1. The highest amount of soluble protein was obtained in pGEX4T1 vector that produced Ada-AlkA fused to Glutathione S-transferase (GST). The ada-alkA gene was amplified from MTB genome through PCR (fig. 18) and cloned in pGEX4T1 vector containing sequence coding GST protein at the N-terminus of the protein.

Figure 18: PCR *ada-alkA* gene from MTB genome

Protein expression was carried out in BL21 *E.coli* cells in the presence of 0.1mM IPTG induction by growing bacterial cells at 20°C for 16 hours.

Cells were lysed and a specific protocol was used to separate insoluble fraction containing inclusion bodies from the soluble fraction. Production and cellular localization of recombinant Ada-AlkA were evaluated through SDS-PAGE analysis. Figure 19 shows the gel fractionation of total *E. coli* extract (lane 1), solubilized inclusion bodies (lane 2) and the soluble fraction (lane 3). Ada-AlkA was clearly produced in large amount but essentially confined in inclusion bodies.

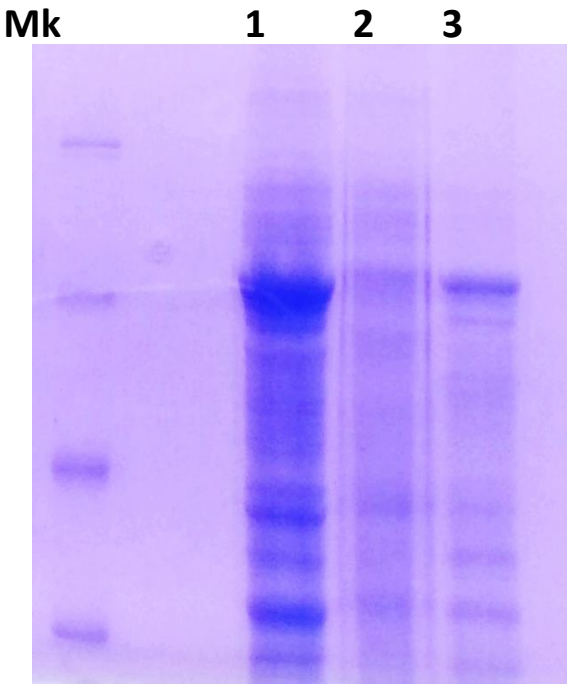

Figure 19 : SDS-PAGE of the Ada-AlkA production. Total cells extract is loaded in lane 1, inclusion bodies in lane 2 and soluble fraction in lane 3

The recombinant protein was purified by affinity chromatography on immobilized glutathione beads and hydrolyzed on column to release GST (fig. 20).

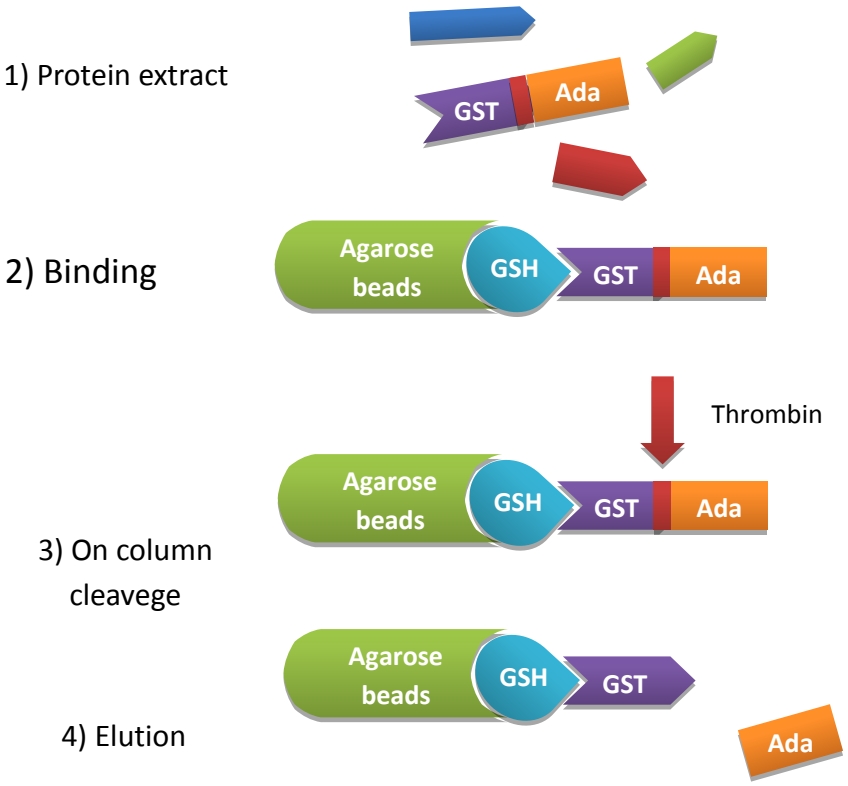

Figure 20: schematic representation of GST affinity chromatography

The eluted proteins were analysed by SDS-PAGE as reported in Figure 21.

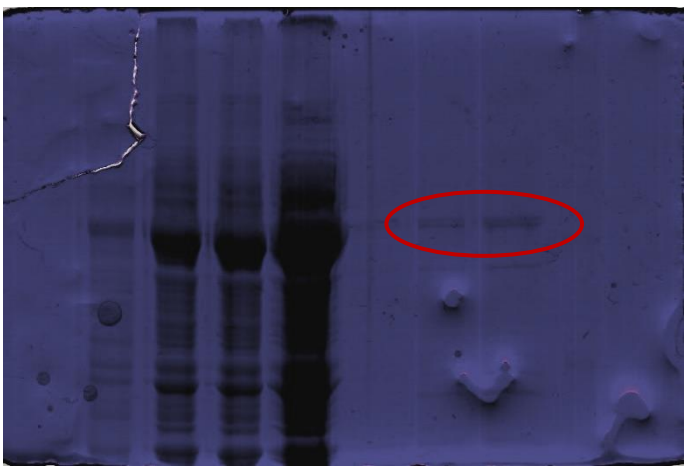

Figure 21: Protein purification verified by SDS-PAGE

Gel electrophoresis analysis showed that the protein was obtained in a good degree of purification, with purified protein concentration of 500 µg/ L of culture.

The *ogt* gene was amplified by PCR from MTB genome (fig. 22), cloned in pGEM-T easy system and then it was cloned in pGEX4T1 to obtain Ogt protein fused with Glutathione S-Transferase (GST).

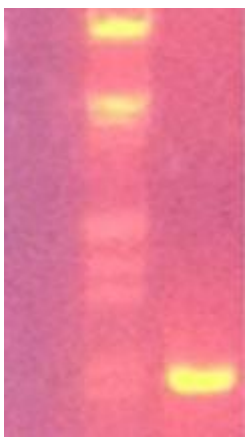

Figure 22 : PCR ogt gene from MTB genome

BL21 *E. coli* cells were transformed with recombinant plasmid in order to produce recombinant Ogt fused with GST. Bacterial cells were grown at 20°C for 16 hours after induction by 0.1mM IPTG. Cells were lysed and the insoluble fraction containing inclusion bodies was separated from the soluble fraction. Production and cellular localization of recombinant Ogt were evaluated through SDS-PAGE analysis (fig 23).

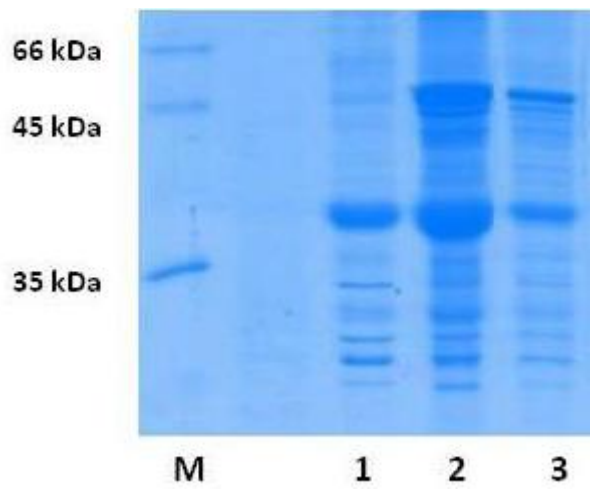

Figure 23 :Gel SDS PAGE. M = *marker*, 1 = protein extract, 2 = inclusion bodies 3 =soluble fraction

Gel image clearly showed that nevertheless the presence of GST, Ogt was essentially produced in inclusion body; however the amount of soluble fraction was enough to carry out further experiments.

Ogt was also purified by GST affinity chromatography and analysed by SDS-PAGE showed in figure 24.

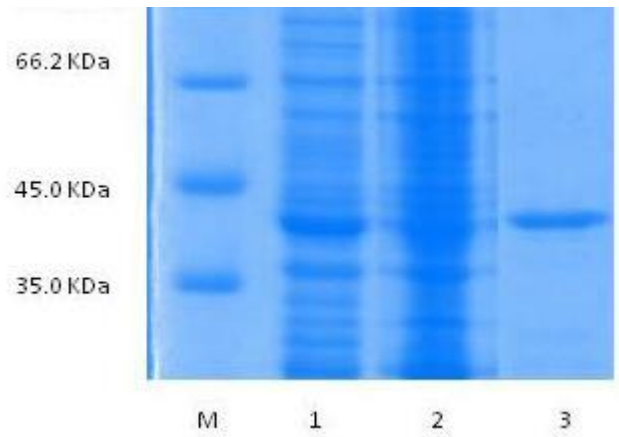

Figure 24 : Protein purification verified by SDS-PAGE: M = marker, 1 = cellular extract, 2 = unbound, 3 = 1 mM GSH elution

The lane 3 clearly shows a band at the expected electrophoretic mobility of about 40 KDa, corresponding to presence of Ogt fused with GST (18 KDa + 25 KDa). Bradford assay allowed to calculate purified protein concentration that was 3 mg/ L of culture.

3. Structural characterization of Ada-Alka protein

The aminoacidic sequence of recombinant Ada-Alka was validate using mass spectrometry methodologies by MALDI mass mapping.

Purified protein was analysed by SDS-PAGE and the corresponding protein band was excised from the gel, reduced with DTT and alkylated with iodoacetamide to irreversibly block the cysteine residues and *in situ* hydrolyzed with tripsin. The resulting peptide mixture was then directly analysed by MALDI-MS using a reflectron instrument. The obtained spectrum is reported in figure 25.

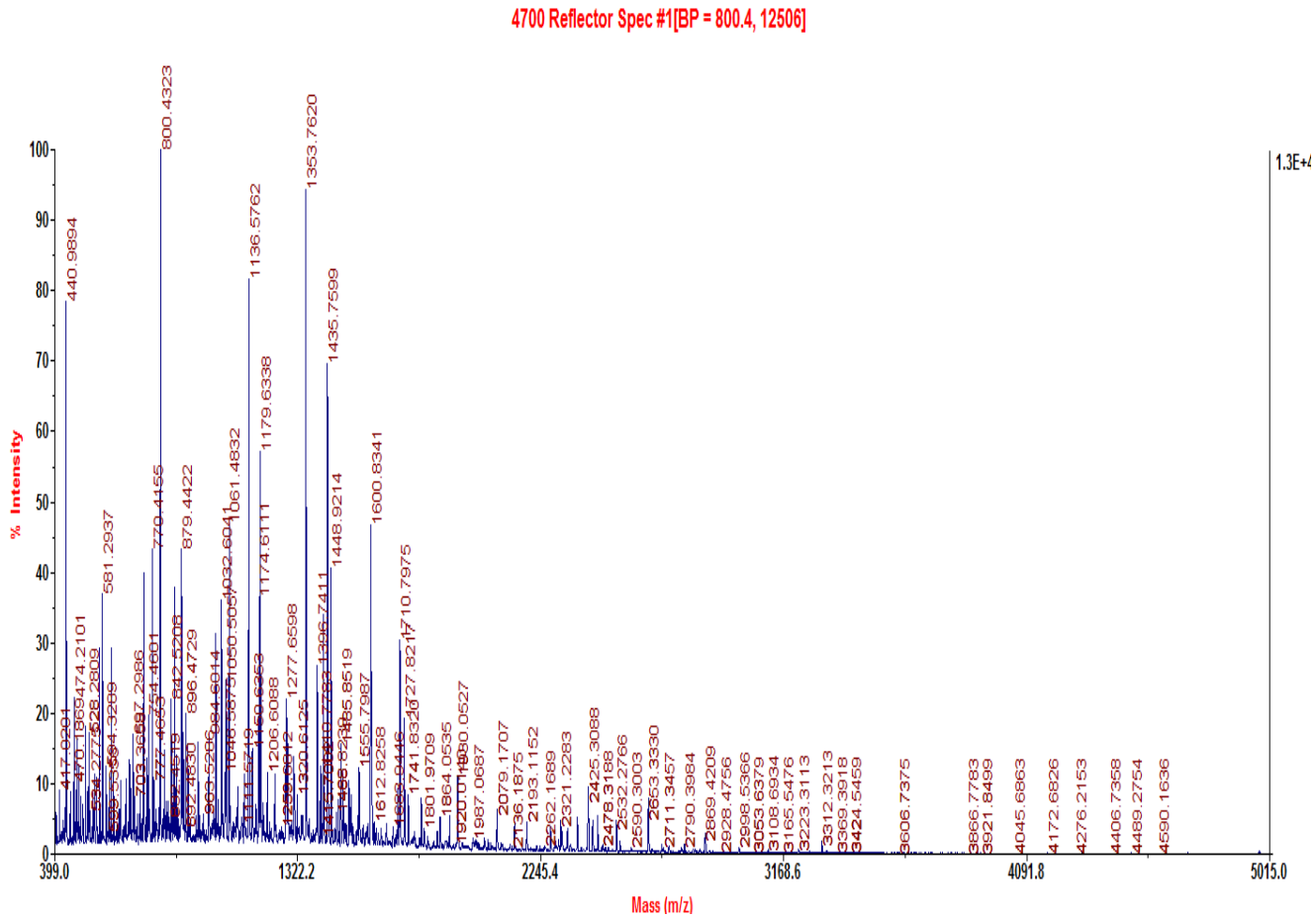

Figure 25. MALDI TOF spectrum of Ada-AlkA peptide mixture

The accurate mass values of the peptides in the mass spectrum was mapped onto the anticipated Ada-AlkA sequence leading to a coverage of the protein primary structure of 90%, as reported in figure 26.

Figure 26: Recombinant Gst-Ada-AlkA sequence is here reported. In yellow is indicated the validated primary structure.

In order to verify if the recombinant protein was correctly folded circular dichroism analyses was performed. Figure 27 shows spectrum obtained both Ada-Alka fused with GST and GST only.

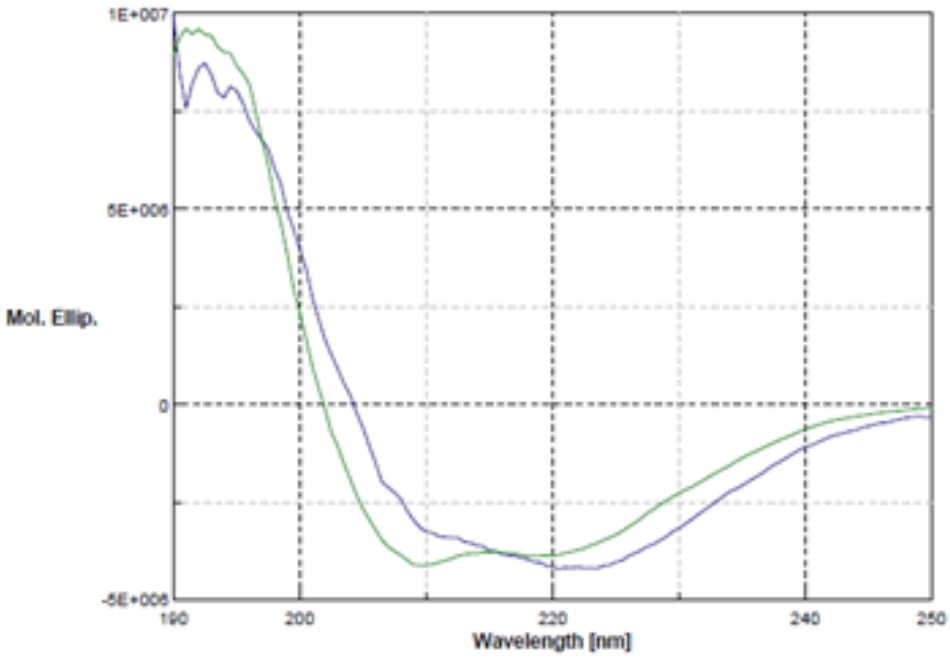

Figure 27. Circular dichroism analyses of Gst alone reported in green in comparison with GST-Ada-AlkA fusion protein in blu.

By comparing two spectra it was evident that the Ada-AlkA is properly structured despite the presence of GST, showing both α-helix to β-sheets secondary structure.

4. Structural characterization of Ogt protein

The correct primary structure of recombinant Ogt was verified by MALDI mass mapping. Purified Ogt was analysed by SDS-PAGE and the corresponding protein band excised from the gel, reduced with DTT and alkylated with iodoacetamide and *in situ* enzymatically digested with trypsin. The resulting peptide mixture was then directly analysed by MALDI-MS using a reflectron instrument.

The obtained spectrum is reported in figure 28.

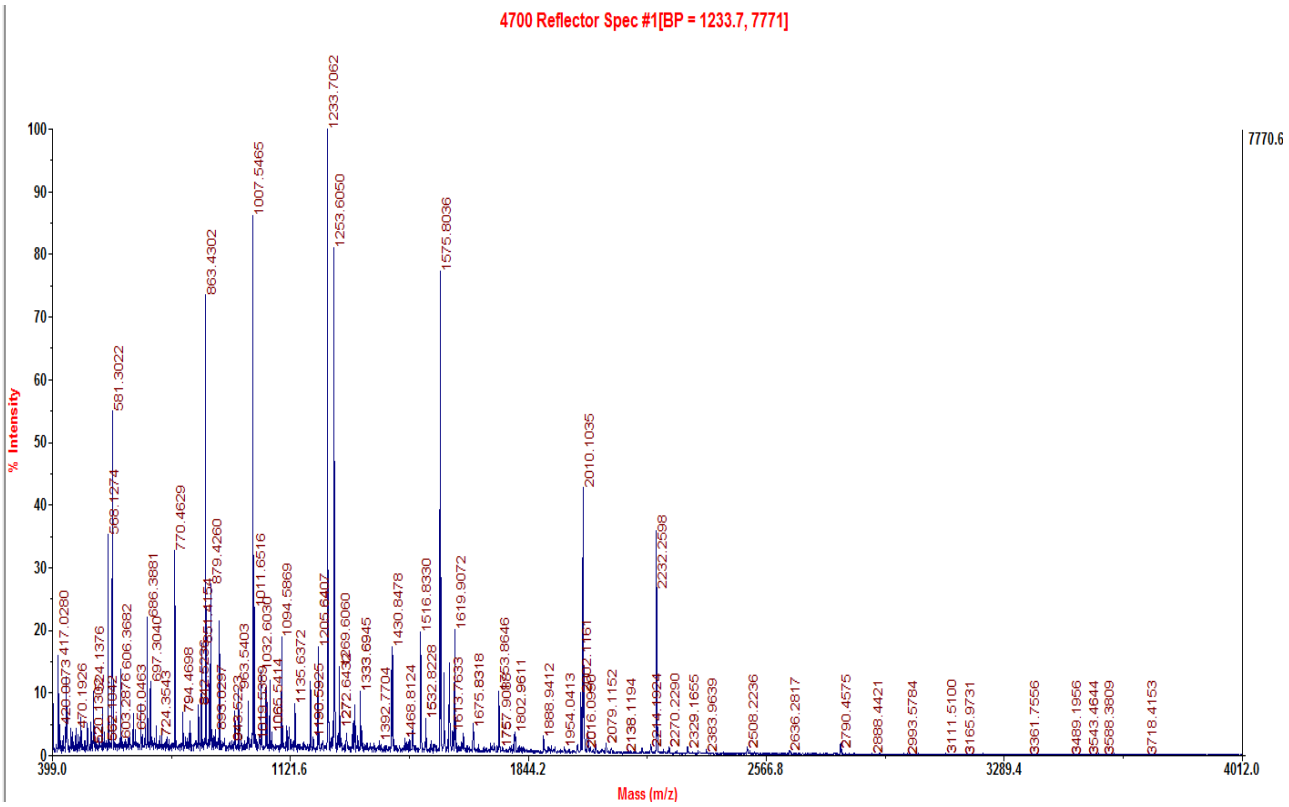

Figure 28. MALDI TOF spectrum of Ogt peptide mixture

The accurate mass values of the peptides in the mass spectrum was mapped onto the anticipated Ogt sequence leading to a coverage of the protein primary structure of 94%, as reported in figure 29.

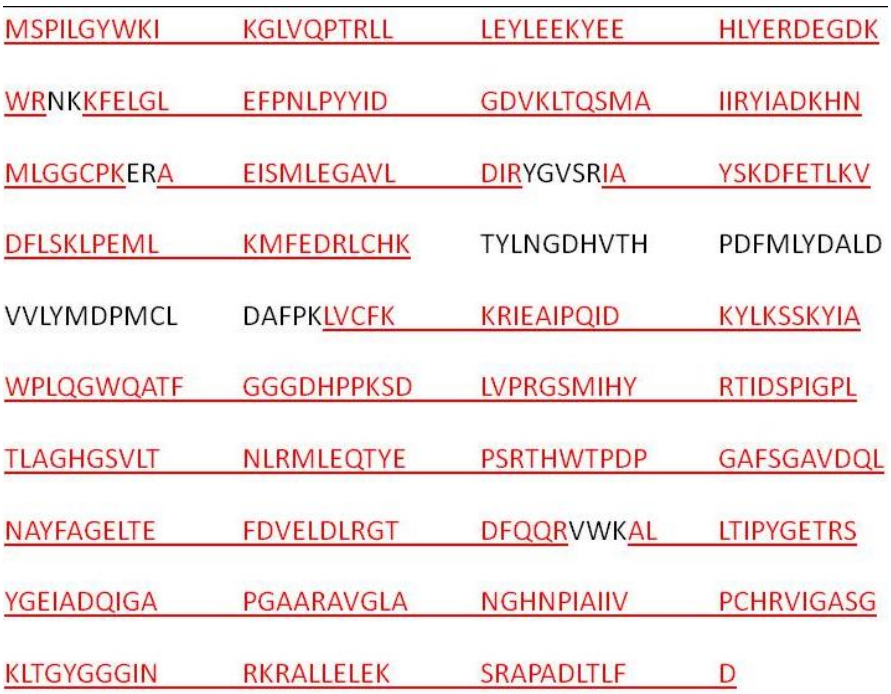

Figure 29. Recombinant Gst-Ogt sequence is here reported. In red is indicated the validated primary structure.

Circular dichroism analyses was performed for both Ogt fused with GST and GST only to verify the correct folding of recombinant protein.

Figure 30 showed that Ogt is correctly folded despite the presence of GST, showing both α-helix to β-sheets secondary structure.

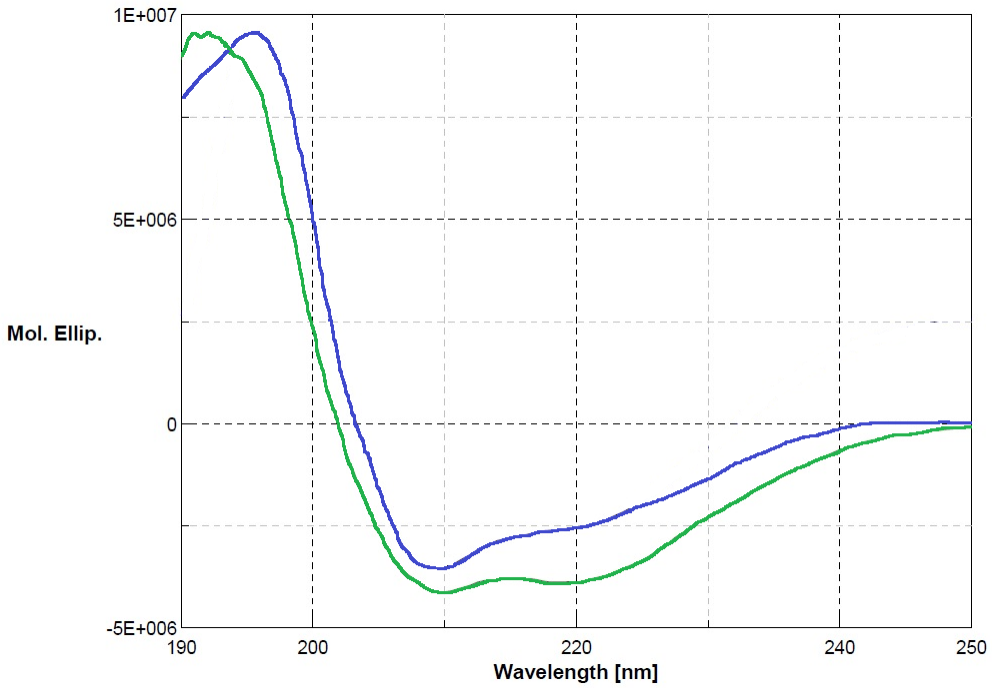

Figure 30: dichroism analyses of Gst alone reported in green in comparison with GST-Ogt fusion protein in blu.

The result obtained was in agreement with the literature data (reference).

5. Functional characterization of Ada-AlkA and Ogt

In order to investigate whether Ada-AlkA and Ogt were able to bind DNA, Elettrophoretic Mobility Shift Assay (EMSA) experiments were performed.

Different amounts of the recombinant proteins were incubated at room temperature with DNA double strand probe and protein complexes formation were analysed by agarose gel electrophoresis in comparison with the isolated probes.

The gel analysis clearly showed an Ada-AlkA-dependent shift in the electrophoresis mobility of the DNA-protein complex compared to the isolated probe (fig. 31). These data demonstrated that Ada-AlkA protein is capable to bind DNA with non sequence specificity.

Figure 31: EMSA assay performed on Ada-AlkA protein. DNA alone (0.2 µM) is loaded in the first line while DNA-protein complex is in the second one.

Ogt showed a similar behaviour to Ada-AlkA protein, demostrating to be also endowed with nonspecific DNA binding cability (fig. 32).

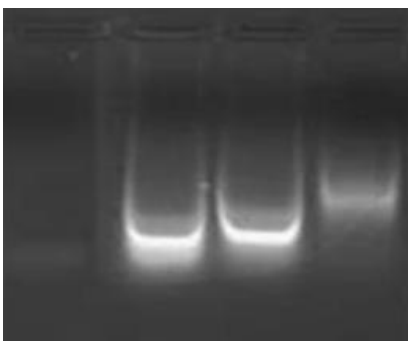

Figure 32: EMSA assay performed on Ogt protein. DNA alone (0.2 µM) is loaded in the first line while different ratio of protein (10, 50, 80 µM)-DNA complex are present respectively in lanes 2, 3 and 4.

It was clear that increasing the amount of Ogt increases the shift of the complex, compared to isolated probe. 80 mM protein concentration, corresponding to a ratio protein / probe equal to 400/1 was observed probe saturati

Chapter 5:

Functional proteomic analyses of the *M. smegmatis* Ogt protein

1. Molecular partners of the *Mycobacterium tuberculosis* Ogt protein

1.1 Isolation of Ogt complexes in *Mycobacterium smegmatis* upon exposure to MMS

The elucidation of Ogt biological role at the molecular level was investigated by functional proteomics approaches addressed to the identification of protein partners. In fact, multi-protein complexes are usually involved in relevant biological mechanisms, so the association of Ogt with partners belonging to a particular mechanism will be strongly suggestive of its biological function. Isolation of multi-protein complexes was performed using the GST-Ogt fusion protein as bait in the presence and in the absence of MMS as methylating agent.

In vivo isolation of Ogt containing protein complexes was performed in two steps. First, *M.smegmatis* cells were grown in the absence and in the presence of sub-inhibitory concentrations of the alkylating agent methyl methane sulfonate. Both cellular pellets were lysated in order to obtain Ogt interactors containing protein extracts. Then, *E.coli* cells producing the Ogt-Gst fusion protein were grown to obtain the tagged bait. Cells were lysated and the recombinant protein was immobilized on reduced glutathione beads as previously described. The total *M.smergmatis* extracts from the two samples were first submitted to a pre-cleaning step by incubation with reduced glutathione immobilized beads in order to remove non-specific binding proteins. Then, cellular extracts were incubated with the immobilized Ogt to allow the formation of functional protein complexes. After extensive washing the proteins specifically bound to Ogt bait were eluted with a buffer containing 1 mM reduced glutathione. Pre-cleaning samples were also eluted and used as control. All the samples were fractionated by SDS-PAGE and analysed by western blot. Specific incubation with anti Gst antibodies showed the presence of the Gst-Ogt bait both in the presence and in the absence of methylating molecules. The presence of the bait was not detected in the precleaning elutes representing an internal control for the correctness of the pull down experiment as shown in Figure 33.

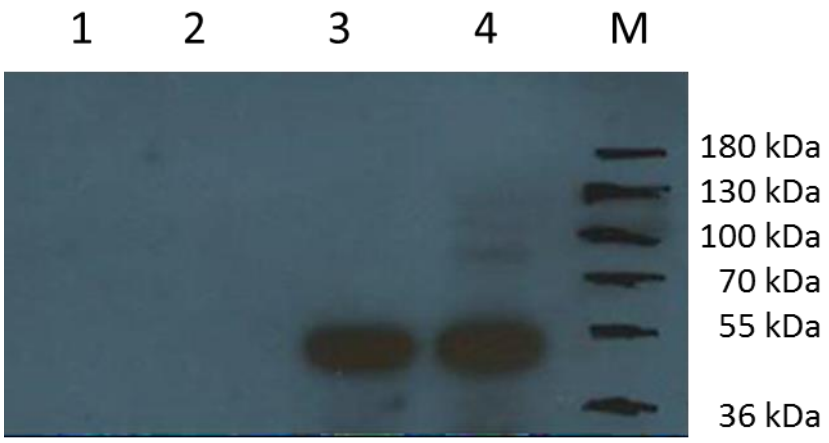

Figure 33. Western blot of Ogt pull down experiments. Lanes 1 and 2 precleaning eluates. Lanes 3 and 4 Ogt complexes in the absence and in the presence of MMS, respectively.

The pull down experiment resulting SDS-PAGE was stained with Coomassie Blue and is reported in Figure 34.

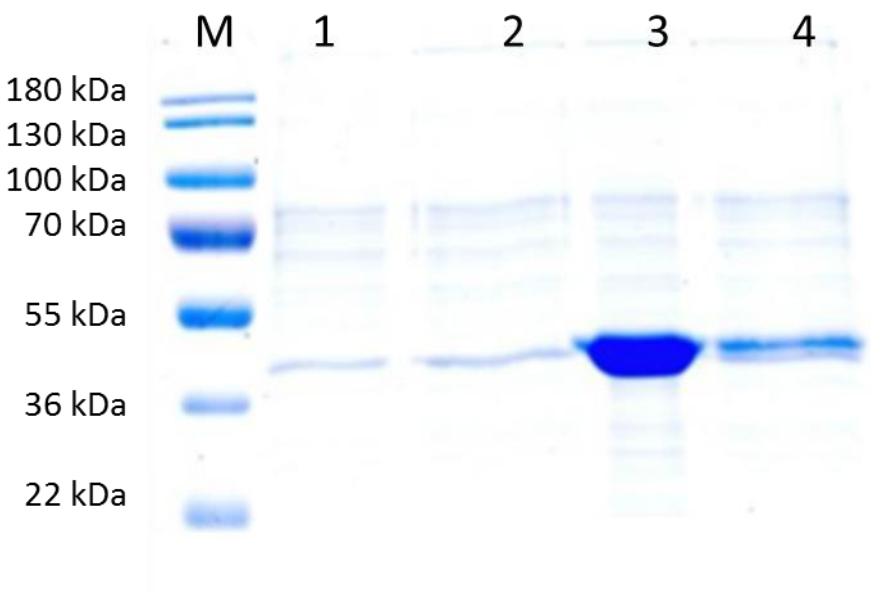

Figure 34. Fractionation of Ogt isolated complexes by SDS-PAGE. Lanes 1 and 2 precleaning eluates. Lanes 3 and 4 Ogt complexes in the absence and in the presence of MMS, respectively.

1.2 Identification of proteins specifically interacting with Ogt

In order to identify Ogt interactors, the entire lanes of the SDS-PAGE from both samples and controls were cut in 23 slices and each gel slice was in situ digested with trypsin and the corresponding peptide mixtures directly analysed by LC-MS/MS procedures. The analyses provided both accurate molecular weight and sequences of each peptide. These data were used for database search using a home version of the Mascot software leading to the identification of the proteins. Common proteins identified in both the sample and the control gel slices were ruled out and only those solely occurring in the samples were considered as putative Ogt interactors thus greatly decreasing the number of false positives.

Proteins identified in the proteomic experiments in the absence of methylation stress are listed in Tables 1 where the protein name, the corresponding Swiss Prot code and the number of identified peptides are reported.

Protein	**Gene**	**Swiss Prot**	**Peptides**
Transcriptional regulator, Ada family protein/DNA-3-methyladenine glycosylase II	MSMEG_4925	A0R1Z2	3(2)
Enoyl-CoA hydratase	MSMEG_4299	A0R088	2(1)
Transcriptional regulator WhiB	whiB1	A0QTP7	1(1)
Putative oxidoreductase	MSMEG_1968	A0QTU5	1(1)
60 kDa chaperonin 1	groL1	A0QQU5	5(2)
ATP-dependent zinc metalloprotease FtsH	ftsH	A0R588	2(1)
Pyruvate dehydrogenase E1 component	aceE	A0R0B0	3(1)
Uncharacterized protein	MSMEI_3028	I7G1A1	2(1)
Uncharacterized protein	MSMEG_1244	A0QRU7	3(1)
ABC transporter, permease/ATP-binding protein (exiT)	MSMEG_0018	A0QNF0	2(1)
Penicillin-binding protein 1A	ponA1	A0R7G2	1(1)
Bacterial regulatory protein, GntR family protein	MSMEG_0895	A0QQW0	4(2)
O-succinylhomoserine sulfhydrylase	metZ	A0QQI7	2(1)
Carboxyl transferase domain protein	MSMEG_2255	A0QUL9	1(1)
Cytochrome P450	cyp123	A0R5U2	2(2)
CheR methyltransferase, SAM binding domain protein	cheR	A0R3Q0	5(3)
Putative aminotransferase	MSMEG_6286	A0R5R7	3(1)

Mammalian cell entry protein	MSMEG_1147	A0QRK3	2(1)
Isocitrate lyase	aceA	A0QYL9	2(1)
Acyl-CoA oxidase	MSMEG_4474	A0R0Q9	6(3)
NADH-quinone oxidoreductase subunit B	nuoB	A0QU35	7(2)
Uncharacterized protein	MSMEG_5168	A0R2M6	7(1)
5-oxovalerate dehydrogenase	MSMEG_1158	A0QRL4	7(1)
50S ribosomal protein L7/L12	rplL	A0QS63	7(1)
30S ribosomal protein S4	rpsD	A0QSL7	6(1)
Uncharacterized protein	MSMEG_4413	A0R0K0	2(1)
50S ribosomal protein L18	rplR	A0QSG5	3(1)

Table 1. Proteins identified in the absence of MMS.

Proteins identified in the proteomic experiments in the presence of methylation stress are listed in Tables 2 where the protein name, the corresponding Swiss Prot code and the number of identified peptides are reported.

Protein	**Gene**	**Swiss Prot**	**Peptides**
Putative multifunctional enzyme siroheme synthase Cysg	cobA	A0QVL7	2(1)
Xylose isomerase	xylA	A0R507	1(1)
UPF0182 protein MSMEG_1959/MSMEI_1915	MSMEG_1959	A0QTT7	1(1)
Oxidoreductase, FAD-binding	MSMEG_0295	A0QP72	2(2)
Succinyl-CoA ligase [ADP-forming] subunit beta	sucC	A0R3M4	2(1)
Integral membrane protein	MSMEG_5066	A0R2C8	2(1)
Nitrilase	MSMEG_1120	A0QRH6	2(1)
Trans-aconitate 2-methyltransferase	tam	A0QQ52	4(1)
50S ribosomal protein L5	rplE	A0QSG1	2(1)
ABC transporter binding protein	MSMEG_4762	A0R1I1	1(1)
Nuclease	MSMEG_0370	A0QPE5	2(1)
Uncharacterized protein	MSMEI_3829	I7G3X2	2(1)
Enolase	eno	A0R3B8	4(1)
ATP-dependent Clp protease ATP-binding subunit ClpX	clpX	A0R196	1(1)
Endopeptidase IV	sppA	A0QSH0	2(2)
Error-prone DNA polymerase	dnaE2	A0QSX1	6(1)

Urease accessory protein UreG	ureG	A0QRF2	1(1)
Serine protease	MSMEG_6289	A0R5R9	1(1)
IS629 transposase orfB	MSMEG_1730	A0QT64	4(1)
Putative succinate-semialdehyde dehydrogenase [NADP(+)]	gabD2	A0R4Q0	2(1)
30S ribosomal protein S12	rpsL	A0QS96	1(1)
2Fe-2S iron-sulfur cluster binding domain protein	MSMEG_1885	A0QTL3	2(1)
Glucose-1-dehydrogenase	MSMEG_4419	A0R0K6	4(1)
Acyl-CoA dehydrogenase	MSMEG_4391	A0R0H7	2(1)
Linalool 8-monooxygenase	MSMEG_3524	A0QY38	1(1)
Uncharacterized protein	MSMEG_2266	A0QUM9	1(1)
Iron-dependent peroxidase	MSMEG_6567	A0R6J2	2(1)
LysR family regulatory protein	MSMEG_0708	A0QQC7	1(1)
50S ribosomal protein L1	rplA	A0QS46	8(4)
Uncharacterized protein	MSMEI_5400	I7GED7	7(5)
Transcriptional regulator, Ada family protein/DNA-3-methyladenine glycosylase II	MSMEG_1613	A0QSV3	3(2)
ABC polar amino acid transporter, inner membrane subunit	MSMEG_3665	A0QYI0	2(1)
Potassium transporter TrkA	MSMEG_2642	A0QVP1	3(1)
Uncharacterized protein	MSMEG_4701	A0R1C4	3(1)
Single-stranded DNA-binding protein	MSMEG_2107	A0QU76	1(1)
Uncharacterized protein	argS	Q9X5M0	3(1)
Arginine--tRNA ligase	MSMEG_6238	A0R5L9	1(1)
Putative two-component system sensor kinase	MSMEG_2487	A0QV88	1(1)
4-aminobutyrate aminotransferase	MSMEG_1613	A0QSV3	3(2)

Table 2. Proteins identified in the presence of MMS.

1.3 Classification of the identified proteins

A total of 67 putative partners were identified, 27 in the absence of MMS and 40 in the presence of methylating agent. Only the protein Ada-AlkA was identified in both condition while all the others were different. According to their reported biological activities, the putative interactors

were grouped into different functional categories belonging to metabolism, DNA binding and repair, ribosome and membrane components in the absence of MMS and they are listed in Table 3.

Metabolism	DNA binding and repair	Membrane components	Ribosome	Other
MSMEG_4299	MSMEG_4925	ftsH	rplL	groL1
aceE	whiB1	MSMEG_0018	rpsD	cheR
metZ	ftsH	ponA1	rplR	
cyp123	MSMEG_0018	nuoB		
aceA	MSMEG_0895	MSMEG_5168		
MSMEG_4474				
nuoB				
MSMEG_1158				

Table 3. Classification of protein partners identified in the absence of MMS in according to their biological functions

Protein partners classification is graphically reported in the pie chart in Figure 35.

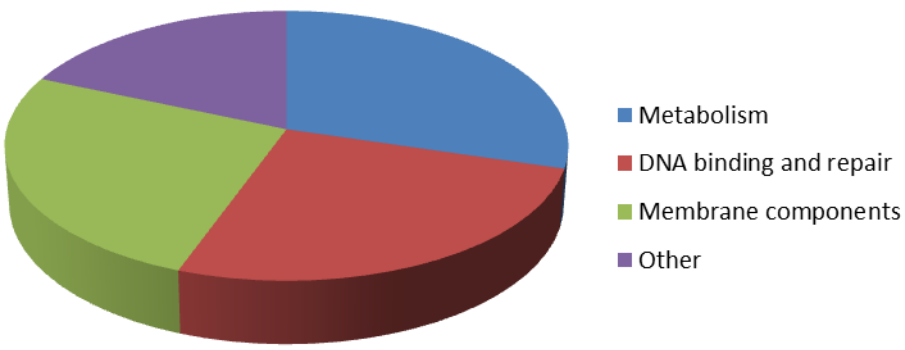

Figure 35. Pie chart showing classification of protein partners identified in the absence of MMS.

In the presence of methylation stress, functional classification of the identified partners showed a similar organization except ribosome to that obtained in the absence of MMS. Interactors were listed in table 4.

Metabolism	DNA binding and repair	Membrane components	Other
cobA	clpX	MSMEG_1959	rplE
xylA	dnaE2	MSMEG_5066	rpsL
sucC	ureG	MSMEG_4762	MSMEG_3524
MSMEG_1120	MSMEG_1730	sppA	rplA
eno	MSMEG_0708	MSMEG_1613	argS
gabD2	MSMEG_4925	MSMEG_3665	
otsB	MSMEG_4701	MSMEG_6238	
MSMEG_2487			

Table 4. Classification of protein partners identified in the presence of MMS in according to their biological functions.

Protein partners classification in the presence of MMS is graphically reported in the pie chart in Figure 36.

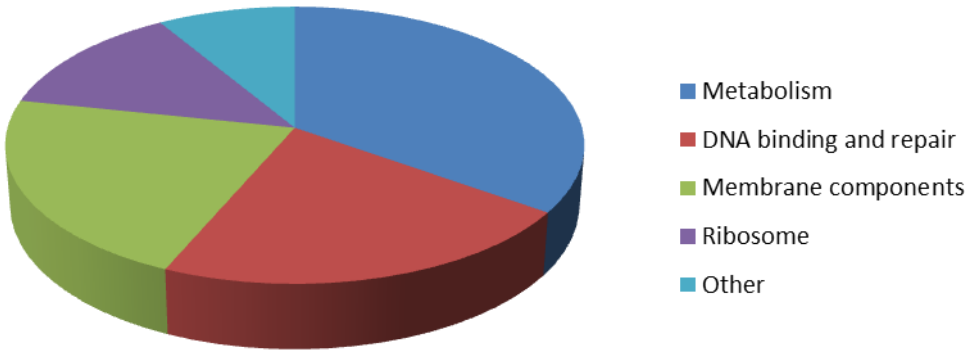

Figure 36. Pie chart showing classification of protein partners identified in the presence of MMS.

The network distributions of the 67 putative interactors were explored using STRING software. The top-ranked networks were in Ribosome (*p*-value = 6.86E-5), RNA degradation (*p*-value =9.98E-1) and Alanine, Aspartate and Glutamate metabolism (*p*-value = 1.26E-1). The obtained STRING network is reported in Figure 37.

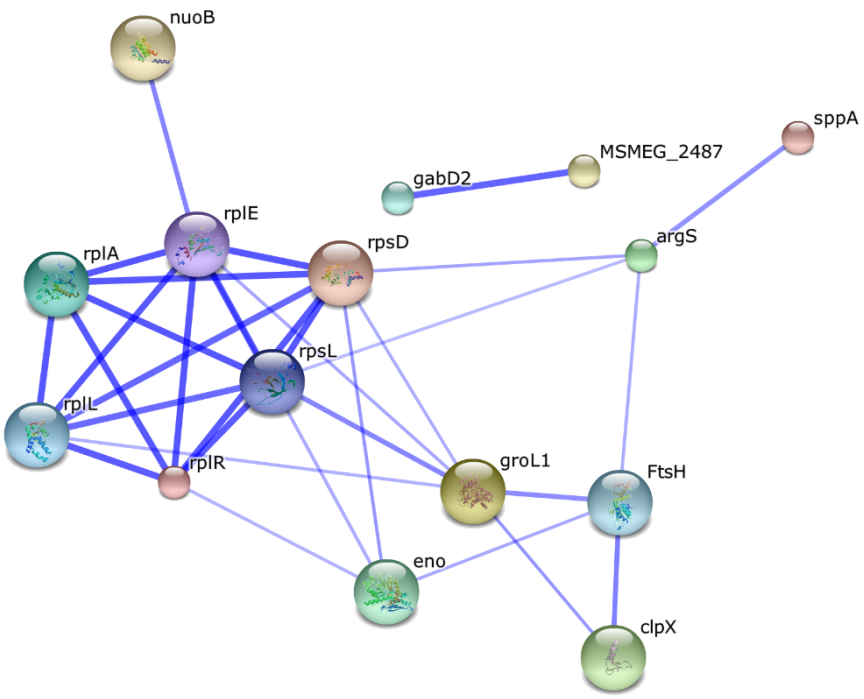

Figure 37. Network distribution of the 67 putative interactors identified as obtained by STRING. The stronger associations are represented by thicker lines

Chapter 6

Discussion

Discussion

Tuberculosis is an acute infection disease that over 130 years after discover of its causative agent MTB, still represents a global health emergency with around 9 million people affected and nearly 2 million deaths each year. Despite the progress made in the pathogenesis and therapy, TB is one of the leading causes of death worldwide from a single infectious agent [48].

Current therapeutic treatment for TB consists in the use of multiple anti-mycobacterial drugs such as rifampicin and isoniazid. This treatment is carried out for a long time (6-9 months) leading to the onset of drug- and multidrug-resistant strains of MTB, due to the occurrence of *in vivo* mutations derived from the DNA oxidative damage.

Therefore new therapeutic models are needed to overcome this problem.

During the infection, if the host immune response is not able to strongly counteract bacterial invasion, MTB cells are confined within macrophages and together with different cell types are enclosed in a granuloma, a latent infection status without evident clinical manifestations [49, 50]. During MTB infection, the host antimicrobial response generates several metabolically activated DNA alkylating agents leading to severe DNA-damaging injuries on MTB cells [51]. Protection of the DNA molecule from chemical damages then strictly depends on the MTB repair mechanisms. Recent studies identified an adaptive response mechanism in MTB homologous to the same process already described in *E.coli*. *In silico* analyses identified in MTB four genes which encode the proteins homologous to the components of adaptive response in *E. coli*. Exposure of *E. coli* to sublethal concentrations of alkylating agents induces the expression of four genes (*ada*, *alkA*, *alkB* and *aidB*) [52]. Expression of these genes is regulated by Ada, that is a methyltransferase that also acts as positive regulator of the operon.

The activation of the genes increases the resistance to the cytotoxic and mutagenic effects of alkylating molecules. Unlike *E. coli*, the DNA repair/protection systems in MBT have not been investigated and it is still largely unknown.

The investigation of the DNA repair and protection mechanisms is really interesting in the tuberculosis context because of their involvement in the drug resistant development. Moreover, proteins involved in these mechanisms represent excellent putative therapeutic targets to explore in search for new TB treatments because of the fundamental role they play for MTB viability and due to the absence of homologues in humans. The aim of this thesis project was the investigation of the adaptive response of *M.tuberculosis* to DNA methylation stress and it

was divided in three phases. In the first part, cellular response to DNA methylation damage was evaluated using *M.smegmatis* cells, a non pathogenic strain of *Mycobacteria*. This study was carried out by evaluating cellular survival, DNA methylation over time and the global proteomics changes in the presence of methylation stress. In the second part, the most interesting proteins of the adaptive response were cloned, heterologously expressed in *E. coli* and structurally and functionally characterized. Finally the last part was entirely focused on a functional proteomics study of Ogt protein.

M.smegmatis growths performed in the presence of increasing doses of methylating agent, methylmethane sulphonate, showed an extensive and proportional decrease in cell survival in comparison with untreated strain. Methylation stress affects cellular life quickly, just after six hours corresponding to two generation time. This finding probably is directly related to an extensive DNA methylation due to the presence of methylating molecule. Experimental measure of DNA bases methylation was performed in order to deeply analyze this point. Tandem mass spectrometry analyses were carried out in Multiple Reaction Monitoring mode to quantify methylated bases. Obtained data showed an increasing in DNA methylation over time upon exposure to methylating molecule with decreasing in *M.smegmatis* cell survival. This data suggest that the cellular responses to DNA alkylation induced by exogenous stress have a leading role in the survival and so in the pathogenicity of *Mycobacteria*. The effect of methylating stress on global proteome was investigated by quantitative differential approaches in order to investigate this interesting point. In fact over the past decade, proteomics has gained an instrumental role within biologic system studies, by enhancing our knowledge of the functions of biological networks through the generation of a large amount of information. The study of the proteome can be divided into profiling, functional and structural proteomics (53). The 2D DIGE (DIfferential In Gel Electrophoresis) technology belongs to the “second generation” proteomic techniques, that allow us to evaluate the relative abundance of protein species in two different conditions (54). 71 differentially proteins were identified between MMS treated and untreated cells. The most of them were up regulated in the presence of methylation stress. According to their annotated biological functions, all of them were grouped in some classes. Three categories were all up regulated in the presence of MMS gathered within biofilm formation, fatty acid degradation and cell wall biosynthesis. Fatty acid degradation involved fadA2, fadA3, fadB2, fadD31, fadD32 while cell wall biosynthesis involved ald, vanX, alaS, glnU,

gcvP. *M. tuberculosis* produces a large range of lipophilic molecules. These molecules comprise simple fatty acids such as palmitate and tuberculostearate and very-long-chain, highly complex molecules such as mycolic acids and the phenolphthiocerol alcohols. *Mycobacteria* contain examples of every known lipid biosynthetic system, including enzymes usually found in mammals and plants as well as the common bacterial systems. The biosynthetic capacity is less remarkable than degradative pathways and fatty acid oxidation systems. A total of about 250 distinct enzymes involved in fatty acid metabolism are present in *M. tuberculosis* compared with only 50 in *E. coli* (55). Interestingly, because of the wide variability and amount of lipids within mammalian cells and the tubercle, *Mycobacteria* are largely lipolytic rather than lipogenic (56). This thesis is also supported by the abundant presence of genes encoding components of fatty acid oxidation systems. Moreover the genome of *M. tuberculosis* contains several enzymes belonging to enoyl-CoA hydratase/isomerase superfamily, acyl-CoA synthases and enzymes that could catalyse the first step in fatty acid degradation, convert the 3-hydroxy fatty acid into a 3-keto fatty acid. In addition to this extensive set of dissociated degradative enzymes, the genome also encodes the canonical FadA/FadB β-oxidation complex.

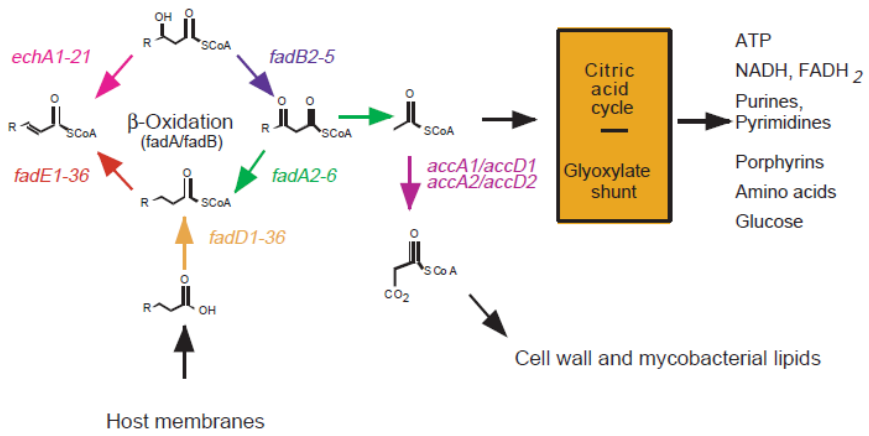

Figura 38. Schematic representation of metabolic processes involved in the degradation of fatty acids on the host cells together with cell wall biosynthesis

This information highlighted that degradation host-cell lipids is vital in the intracellular life of *M. tuberculosis.* In addition *M.tuberculosis* take precursors directly from the cell membranes of the host for many metabolic processes such as mycobacterial cell-wall constituents. This feature is

allowed by an enormous group of β-oxidative enzymes capable to produce acetyl CoA useful in several reaction of bacterial metabolism. According to above-mentioned literature data, the up regulation of both fatty acid degradation and cell wall biosynthesis might suggest that under methylation stress conditions *M.smegmatis* counteracts DNA damages providing fuel to increase cellular physical defense. This interesting feature is also support by the up regulation of proteins involved in biofilm formation. Bacteria generally exist in one of two types of population: planktonic, freely existing in bulk solution, and sessile, as a unit attached to a surface or within the confines of a biofilm (57). A biofilm consists of cells immobilised at a substratum and frequently embedded in an organic polymer matrix of microbial origin. A biofilm may be described as a microbially derived sessile community characterised by cells that attach to an interface, embedded in a matrix of exo-polysaccharide which demonstrates an altered phenotype. Biofilm growth is governed by a number of physical, chemical and biological processes. Fletcher [42] described the accumulation of microorganisms on a collecting surface as a process of three stages: (1) adsorption, or the accumulation of an organism on a collector surface i.e. substrate (deposition); (2) attachment, or the consolidation of the interface between an organism and a collector, often involving the formation of polymer bridges between the organism and collector; (3) colonisation, or growth and division of organisms on the collector's surface. For bacteria, the advantages of biofilm formation are numerous. These advantages include protection from antibiotics [58], disinfectants [59], and dynamic environments [60]. Intercellular communications within a biofilm rapidly stimulate the up and down regulation of gene expression enabling temporal adaptation such as phenotypic variation and the ability to survive in nutrient deficient conditions [61]. In GIGE analyses dnaK, gyrB, ppiA, groL1 and leuD were identified as involved in biofilm production.

The molecular chaperone DnaK is important in *E.coli* for biofilm formation and the chemical inhibition of its cellular function is capable to effectively prevent biofilm development. Genetic, microbial and microscopic analyses revealed that the lesion of the dnak gene markedly reduced the production of the extracellular functional amyloid curli, which contributes to the robustness of *E.coli* biofilm (62).

The molecular chaperone GroEL1 modulates synthesis of mycolates, long-chain fatty acid components of the mycobacterial cell wall, specifically during biofilm formation and physically associate with kasA, a key component of the type II fatty acid synthase involved in mycolic acid

synthesis. GroEL1 is specifically required for complete growth and biofilm maturation but not for initial growth at the liquid-air interface. The possibility of biofilm formation during *M.tuberculosis* infection of mammalian cells remains an unresolved question but could give rise to drug-tolerant organisms that persist for long durations of drug treatment (63).

So biofilm formation could represent a cellular defense of *M.smegmatis* responding to methylation stress. For this reason, biofilm formation assays were performed in the presence and in the absence of MMS. Obtained results showed an increasing in biofilm formation upon exposure to methylation stress demonstrating a strong correlation between DNA methylation and biofilm production.

In the second part the corresponding gene of Ada-AlkA and Ogt were heterologously expressed in *E.coli*. They were structurally characterized demonstrating the correctness of recombinant products in order to use proteins in further functional investigations. In *E.coli*, Ada acts as positive regulator of the mechanism demethylating phosphomethyltriesters in the sugar phosphate backbone representing a sensitive regulatory signal leading to Ada operon induction. Considering that N-terminal domain of Ada (Ada A) is fused with AlkA and C-terminal domain (AdaB) constitutes a separate protein annotated as Ogt, it was interesting investigate whether proteins maintained a correlated function. For this reason the protein sensor must be able to bind DNA so recombinant products were tested to assess the DNA binding cability by EMSA experiments. Obtained results demonstrated that both Ada-AlkA and Ogt were endowed with non specific DNA binding capability suggesting that they could work together as the adaptive response sensor in the methylation stress.

In the last part, the biological role of Ogt was analysed by a functional proteomics approach. Complete description of the complex network of cellular mechanisms and use of the network to predict the full range of cellular behaviors are major goals of systems biology. A key role in contemporary biology can be played by functional proteomics, which focuses on the elucidation of protein functions and the definition of cellular mechanisms at the molecular level. The attainment of these targets is strictly dependent on the identification of individual proteins within functional complexes in vivo (64).

In vivo isolation of Ogt containing protein complexes was performed in order to explore its possible role of biological sensor of the mechanism. The experiments was carried out growing *M.smegmatis* cells in the presence and in the absence of sub-inhibitory concentration of MMS.

A total of 67 proteins was identified by the proteomic procedure, 40 of which were found under methylating stress conditions. The most evident result was the individuation of proteins completely different between treated and untreated conditions. The unique protein identified both in MMS treated and untreated was Ada-AlkA, part of the adaptive response. This result highlighted that Ogt needs to bind Ada-AlkA not only to counteract methylation damages but probably also to detect DNA methylation as the sensor of mechanism. Nevertheless bacteria evolution separated Ogt and Ada-AlkA genes these findings suggest that the proteins are complexed in the adaptive response to methylation stress.

According to their reported biological activities, identified interactors were grouped into different functional categories: Metabolism, DNA binding and repair, Membrane components and Ribosome.

In the presence of MMS, a number of putative interactors involved in pathways strictly connected with stress response mechanisms were identified. Particularly, the transcriptional regulator TamR, 2 ABC transporter binding protein and Error prone DNA polymerase dnaE were identified.

TamR (trans-aconitate methyltransferase regulator) is a transcriptional regulator that in Streptomyces functions in the oxidative stress response to regulate a key step in central metabolism (65). Citrate isomerization is an important reaction in both the citric acid and glyoxylate cycles, in which citrate is converted to isocitrate by aconitate hydratase (aconitase) via the intermediate cis-aconitate (fig. 39).

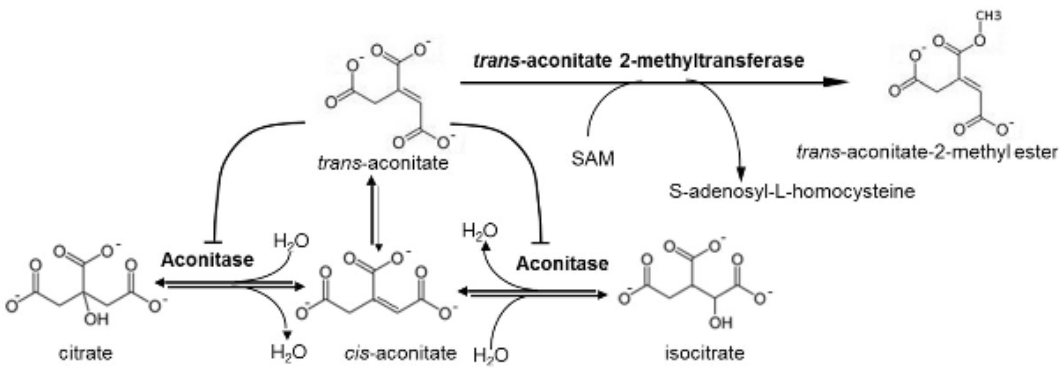

Figure 39: Citrate isomerization

When cis-aconitate is released from the enzyme-substrate complex, it can spontaneously be converted to the more stable isomer trans-aconitate. This product represents an inhibitor of aconitase, that is a toxic product and its accumulation leads to the inhibition of the citrate isomerization step. Trans-aconitate may accumulate during the oxidative stress because the aconitase contains a [4Fe-4S] cluster that is essential for catalytic activity, and it is disassembled under oxidative stress conditions rendering the enzyme nonfunctional. TamR acts to prevent accumulation of trans-aconitate by using it as a substrate to catalyse a methyl group transfer from S-adenosyl-methionine (SAM), resulting in formation of the trans-aconitate methylester. Esterification of trans-aconitate significantly attenuates its ability to inhibit aconitase and to interfere with key steps in central metabolism. Thus, TamR functions in oxidative stress responses to alleviate the consequences of aconitase inactivation, probably contributing to restore the catalytic function of aconitase by participating in regulating its gene activity and by preventing inhibition of functional enzyme by removing the inhibitor trans-aconitate [65].

The ABC transporter binding proteins are integral membrane proteins comprising two transmembrane domains and two cytoplasmic nucleotide-binding domains; they bind and hydrolyse ATP providing energy for uptake or export of substrates across cell membranes. Functions include the uptake of nutrients into the cells and the export of virulence factors and toxins [66]. The ABC transporter binding protein is able to transfer substances from the outside to the inside of the cell and vice versa. It has been reported that in *M. smegmatis* this protein is up-regulated in deficiency of oxygen [67]. Moreover ABC transporter is essential as other transporter systems also in the bacterial resistance. In fact, *M.tuberculosis* is naturally resistant to many antibiotics causing drug treatments very difficult. This resistance derived not only from the highly hydrophobics cell envelope that represent a permeability barrier but also from other structural determinants such as drug efflux system like ABC transporter.

The Error prone DNA polimerase dnaE is involved in DNA repair. In MTB, under stress conditions, DNAE2 is the primary enzyme responsible for adaptive mutagenesis, a mechanism that occurs under stress conditions increasing the frequency of appearance of heritable mutations can allow the organism to adapt to the environment and relieve selection pressure. It has been suggested that pathogenic bacterial strains utilize adaptive mutagenesis to develop resistance against therapeutic agents. In fact, it was reported that in MTB, according to this mechanism , DNAE2 increases the frequency of appearance of mutations allowing pathogenic bacteria to develop

drug resistance (68). DNAE2 was found up-regulated in the presence of DNA-damaging agents and the loss of this protein reduces survival of MTB after UV irradiation [69].

Functional proteomics results highlighted that in the presence of methylating agent Ogt interacts with several proteins involved in pathways responding to stress conditions, suggesting that the protein could play an essential role also in different cellular responses to methylation stress.

References

References

1. Bruce M. Rothschild, Larry D. Martin, Galit Lev, Helen Bercovier, Gila Kahila Bar-Gal, Charles Greenblatt, Helen Donoghue, Mark Spigelman, and David Brittain, 2001, Mycobacterium tuberculosis Complex DNA from an Extinct Bison Dated 17,000 Years before the Present. Clinical Infectious Diseases, 33:305–11.

2. E. Cambau and M. Drancourt, 2014, Steps towards the discovery of Mycobacterium tuberculosis by Robert Koch, 1882. Clinical Microbiology and Infection. 20 (3): 196–201.

3. World Health Organizzation (WHO), 2006, The Stop TB Strategy Building on and enhancing DOTS to meet the TB-related Millennium Development Goals

4. J. Chakaya and M. Raviglione, 2014, Quality Tuberculosis Care All Should Adopt the New International Standards for Tuberculosis Care. AnnalsATS 11 Number 3; 397-398

5. G. Sulis, A. Roggi, A. Mattelli, M. Raviglione, 2014, Tuberculosis: Epidemiology and Control. Mediterranean journal of hematology and infectious disease, 6(1)

6. Delphi chatterjee, 1997, The mycobacterial cell wall , byosinthesis and sites of drug action. Biopolymers 1: 579-588

7. McMullen AM, Hwang SA, O'Shea K, Aliru ML, Actor JK, 2013, Evidence for a unique species-specific hypersensitive epitope in Mycobacterium tuberculosis derived cord factor. Tuberculosis (Edinb). 2013 Dec;93 Suppl:S88-93.

8. Welsh KJ1, Hunter RL, Actor JK, 2013, Trehalose 6,6'-dimycolate--a coat to regulate tuberculosis immunopathogenesis. Tuberculosis (Edinb). 2013 Dec;93 Suppl:S3-9.

9. Adams LB,Dinauer, Alina E; Bowater, Richard P; Dziadek, Jaroslaw, 2010, DNA repair systems and the pathogenesis of Mycobacterium tuberculosis: varying activities at different stages of infection. Clinical science, 119(5):187-202.

10. Tsai, M. C., Chakravarty, S., Zhu, G., Xu, J., Tanaka, K., Koch, C., Tufariello, J., Flynn, J. and Chan J., 2006, Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. Cell. Microbiol. 8, 218–232.

11. Tubercolosis (TB) treatment, Center of Desease Control and Prevention , Atlanta, 24/7 saving life protectin people

12. Warner, D. F. and Mizrahi, V., 2006, Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. Clin. Microbiol. Rev

13. A. Cobat, M. Orlova, L.F. Barrera, E. Schurr, 2013, Host Genomics and Control of Tuberculosis Infection. Public Health Genomics, 16:44–4.

14. Saltini C., 2006, Chemotherapy and diagnosis of tuberculosis. Respiratory Medicine 100, 2085–2097.

15. Michael C. Chao, Eric J. Rubin, 2010, Letting Sleeping dogs Lie: Does Dormancy Play a Role in Tuberculosis?. Annual review of microbiology, 64:293-311.

16. Sherman D.R., Gagneux S., 2011, Estimating the mutation rate of Mycobacterium tuberculosis during infection. Nature

17. C. B. Ford, P. Ling Lin, M.R. Chase, R. R Shah, O. Iartchouk, J. Galagan, N. Mohaideen, T. R. Ioerger, J. C. Sacchettini, M. Lipsitch, J. L Flynn, S. M. Fortune, 2011, Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nature

18. Koul, Anil; Arnoult, Eric; Lounis, Nacer; Guillemont, Jerome; Andries, Koen, 2011, The challenge of new drug discovery for tuberculosis. Nature. 469(7331):483-90

19. Dos Vultos, Tiago; Mestre, Olga; Tonjum, Tone; Gicquel, Brigitte, 2009, DNA repair in Mycobacterium tuberculosis revisited. FEMS microbiology reviews. 33(3):471-87

20. James A. Imlay, 2013, The molecular mechansms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat Rev Microbiol. 11(7): 443–454

21. Jean Cadet1,2 and J. Richard Wagner, 2013, DNA base damage by reactive oxygen species, oxidizing agents and UV radiation. Cold Spring Harb Perspect Biol 5:a012559

22. Patricia L. Foster, 2007, Stress-Induced Mutagenesis in Bacteria. Crit Rev Biochem Mol Biol. 42(5): 373–397

23. NR JENA, 2012, DNA damage by reactive species: Mechanisms, mutation and repair. J. Biosci. 37(3), 503–517

24. Burney, Caulfield JL, Niles JC, Wishnok JS, Tannenbaum SR, 1999, The chemistry of DNA damage from nitric oxide and peroxynitrite. Mutat Res. 8;424(1-2):37-49

25. Adams LB,Dinauer MC, Morgenstern DE &Krahenbuhl JL (1997) Comparison of the roles of reactive oxygen and nitrogen intermediates In the host response to Mycobacterium tuberculosis using transgenic mice. Tubercle lung Dis 78:237-246

26. Hasan Koc, James A. Swenberg. Applications of mass spectrometry for quantitation of DNA adducts. J. Chromatogr. B 778, 323–343.

27. Dragony Fu, Jennifer A. Calvo & Leona D. Samson, 2002, Balancing repair and tolerance of DNA damage caused by alkylating agents. Nat Rev Cancer 12, 104-120

28. Schärer OD 2003, Chemistry and biology of DNA repair. Angew. Chem. Int. Ed. 42 2946–2974

29. Mizrahi V, Andersen SJ, 1998, DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? Mol Microbiol, 29(6):1331-9.

30. Graham JE & Clark-Curtis JE, 1999, Identification of Mycobacterium tubercolosis RNAs synthesized in response to phagocitosis by human macrophages by selective capture of transcribed sequenze (SCOTS). P Natl Acad Sci USA 96: 11554-11559

31. Cabusora L, Sutton E, Fulmer A & Forst CV, 2005, Differential network expression during drug and stress response. Bioinformatics 21: 2898-2905

32. Boshoff HI, Reed MB, Barry CE III & Mizrahi V, 2003, DnaE2 polymerase contributes to in vivo survival and the emerge of drug resistance in Mycobacterium tuberculosis. Cell 113:183-193

33. Cole ST, Brosch R, Parkhill J et al., 1998, Deciphering the biology of *Mycobacterium tubercolosis* from the complete genome sequence. Nature 393: 537-544

34. Sinha KM, Stephanou NC, Unciuleac MC, Glickman MS & Shuman, 2008, Domain requirements for DNA unwinding by mycobacterial UvrD2, an essential DNA helicase. Biochemistry 47:9355-9364

35. Shuman S & Glickman MS, 2007, Bacterial DNA repair by nonhomologous end joining. Nat Rev Microbiol 5: 852-861

36. Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC & Wigley DB, 2004, Cristal structure of RecBCD enzyme reveals a machine for processing DNA breaks. Nature 432: 187-193

37. Morimatsu K & Kowalczykowski SC, 2003, RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand Exchange:a universal step of recombinational repair. Mol Cell 11: 1337-1347

38. Della M, Palmbos PL, Tseng HM, Tonkin LM, Daley JM, Topper LM, Pitcher RS, Tomkinson AE,Wilson TE & Doherty AJ, 2004, Mycobacterial Ku and ligase proteins costitute a two-components NHEJ repair machine. Science 306:683-685

39. Volkert M. R. and Dinh C. Nguyen, 1984, Induction of specific Escherichia coli genes by sublethal treatments with alkylating agents. Proc. Natl. Acad. Sci. USA Vol. 81, pp. 4110-4114.

40. Lindahl, T., B. Demple, and P. Robins, 1982, Suicide inactivation of the *E. coli* O6-methylguanine-DNA methyltransferase. EMBO J. 1:1359–1363.

41. Landini P., VOLKERT M.R., 2000, Regulatory Responses of the Adaptive Response to Alkylation Damage: a Simple Regulon with Complex Regulatory Features. Journal of bacteriology p. 6543–6549.

42. Jadwiga Nieminuszczy and Elżbieta Grzesiuk, 2007, Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA. Acta biochimica Polonica, 54(3):459-68

43. Rippa V, Amoresano A., Esposito C., Landini P., Volkert M.R., DuilioA., 2010, Specific DNA Binding and Regulation of Its Own Expression by the AidB Protein in *Escherichia coli*. Journal of Bacteriology, p. 6136–6142.

44. Yang M, Aamodt RM, Dalhus B, Balasingham S, Helle I, Andersen P, Tønjum T, Alseth I, Rognes T, Bjørås M., 2011, The ada operon of *Mycobacterium tuberculosis* encodes two

DNA methyltransferases for inducible repair of DNA alkylation damage. DNA Repair (Amst). 10:595-602.

45. Miggiano R., Casazza V., Garavaglia S., Ciaramella M., Perugino G., Rizzi M., Rossia F., 2013, Biochemical and Structural Studies of the *Mycobacterium tuberculosis* O6-Methylguanine Methyltransferase and Mutated Variants. Journal of bacteriology 195(12):2728.

46. Kurthkoti, Krishna; Varshney, Umesh, 2012, Distinct mechanisms of DNA repair in mycobacteria and their implications in attenuation of the pathogen growth. Mechanisms of ageing and development. 133(4):138-46.

47. Hu, C.W., Chen, C.M., Ho, H.H., Chao, M.R., 2012, Simultaneous quantification of methylated purines in DNA by isotope dilution LC-MS/MS coupled with automated solid-phase extraction. Anal. Bioanal. Chem 402, 1199-208

48. Kochi A., 1994, Tuberculosis: distribution, risk factors, mortality. Immunobiology. Oct 91(4- 5): p.325-36.

49. Hamill MJ1, Jost M, Wong C, Bene NC, Drennan CL, Elliott SJ, 2012, Electrochemical Characterization of *Escherichia coli* Adaptive Response Protein AidB. Int J Mol Sci. 13(12):16899-915.

50. Shell SS1, Prestwich EG, Baek SH, Shah RR, Sassetti CM, Dedon PC, Fortune SM, 2013, DNA methylation impacts gene expression and ensures hypoxic survival of Mycobacterium tuberculosis. PLoS Pathog. 2013;9(7):e1003419

51. Miggiano R1, Casazza V, Garavaglia S, Ciaramella M, Perugino G, Rizzi M, Rossi F. 2013, Biochemical and structural studies of the Mycobacterium tuberculosis O6-methylguanine methyltransferase and mutated variants. J Bacteriol. 2013 Jun;195(12):2728-36.

52. Salmelin C, Vilpo J, 2003. Induction of SOS response, cellular efflux and oxidative stress response genes by chlorambucil in DNA repair-deficient *Escherichia coli* cells (ada, ogt and mutS). Mutat Res. 522(1-2):33-44

53. Figeys D., 2003, Proteomics in 2002: a year of technical development and wide-ranging applications. Anal. Chem. 75, 2981–2905.

54. Caterino M., Ruoppolo M., Fulcoli G., Huynth T., Orru S., Baldini A., Salvatore F., 2009, Transcription Factor TBX1 Overexpression Induces Downregulation of Proteins Involved in Retinoic Acid Metabolism: A Comparative Proteomic Analysis. Journal of proteome research. J. Proteome Res. 1515-26.

55. S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston1, K. Taylor, S. Whitehead & B. G. Barrell, 1998, Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393, 537-544

56. Wheeler, P. R. & Ratledge, C., 1994, Tuberculosis: Pathogenesis, Protection, and Control. Am. Soc. Microbiol. 353–385

57. Garrett T. R., Bhakoo M., Zhang Z., 2008, Bacterial adhesion and biofilms on surfaces, Progress in Natural Science 1049–1056

58. Goldberg J., 2002, Biofilms and antibiotic resistance: a genetic linkage. Trends Microbiol, 10:264.

59. Peng JS, Tsai WC, Chou CC., 2002 ,Inactivation and removal of Bacillus cereus by sanitizer and detergent. Int J Food Microbiol, 77:11–8.

60. Chen MJ, Zhang Z, Bott TR., 1998, Direct measurement of the adhesive strength of biofilms in pipes by micromanipulation. Biotechnol Tech, 12:875–80.

61. Fletcher M., 1980, Microbial adhesion to surfaces. Chichester Ellis Horwood

62. Morioka A., Yamanaka K., Mizunoe Y., Ogura T, Sugimoto S, 2015, Novel startegiy for biofilm inhibition by using small molecules targeting molecular chaperone DnaK. Antimicrob Agents Chemother. 59(1):633-41.

63. Ojha A., Anand M, Bhatt A, Kremer L, Jacobs WR Jr, Hatfull GF., 2005, GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. Cell. 123(5):861-73.

64. Monti M, Cozzolino M, Cozzolino F, Vitiello G, Tedesco R, Flagiello A, Pucci P., 2009, Puzzle of protein complexes in vivo: a present and future challenge for functional proteomics. Expert Rev Proteomics., 6(2):159-69.

65. Huang H., Grove A. 2013. The transcriptional regulator TamR from Streptomyces coelicolor controls a key step in central metabolism during oxidative stress. Molecular Microbiology 87(6):1151-66

66. Spivevy V.L. , Whalan R.H. 2013. An attenuated mutant of the Rv1747 ATP-binding cassette transporter of Mycobacterium tuberculosis and a mutant of its cognate kinase, PknF, show increased expression of the efflux pump-related iniBAC operon). FEMS Microbiol. Lett.

67. Murugasu-Oei B, Tay A, Dick T. 1999, Upregulation of stress response genes and ABC transporters in anaerobic stationary-phase Mycobacterium smegmatis. Mol Gen Genet pp 677-682.

68. Sharma A., Nair D.T. 2012. sDpo4-a DinB Homolog from Mycobacterium smegmatis-Is an Error-Prone DNA PolymeraseThat Can Promote G:T and T:G Mismatches. Journal of nucleic acids 012:285481.

69. Boshoff HI1, Reed MB, Barry CE 3rd, Mizrahi V., 2003, DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance inMycobacterium tuberculosis. Cell. 18;113(2):183-93.

J. Chem. Chem. Eng. 7 (2013) 876-884

Molecular Partners of *Escherichia coli* Transcriptional Modulator AidB

Pamela Di Pasquale, Angela Amoresano, Francesca De Maria and Angela Duilio*

Department of Chemical Sciences, University Federico II of Naples, Naples 80126, Italy

Received: May 29, 2013 / Accepted: June 17, 2013 / Published: September 25, 2013.

Abstract: The AidB protein is involved in the adaptive response to DNA alkylation damages in *Escherichia coli*. Functional proteomic experiments were designed to elucidate AidB biological functions in the presence and in the absence of methyl methanesulfonate as methylating agent. Several proteins were identified in both conditions and according to their reported biological activities, the inter-actors were grouped into three different functional categories: stress response, energetic metabolic pathways and nucleic acid metabolism. Particularly, the interaction between AidB and UvrA, a member of the UvrABCD nucleotide excision system, suggested a new interesting putative role for AidB.

Key words: AidB protein, adaptive response, UvrA protein, DNA alkylation damages, functional proteomics.

1. Introduction

DNA modifications by alkyl molecules can cause cytotoxic and mutagenic lesions in all living organisms. Several repair mechanisms able to remove alkyl groups and restore genetic information occur in microorganisms to counteract DNA chemical damages. Furthermore, bacteria exhibit the adaptive response by which cellular resistance enhances with increasing doses of methylating agents [1]. This response in *Escherichia coli* involves the presence of the acyl-CoA dehydrogenase AidB, among others. AidB consists of two domains, the N-terminal region responsible of the dehydrogenase activity, and the C-terminal domain exhibiting DNA binding capability [2]. Unlike all the other proteins of the adaptive response, AidB seems to act as a DNA protective protein and it is not endowed with DNA repair capabilities.

Moreover, in the presence of methylating agents, AidB allows efficient transcription from promoters containing an UP element, AT rich transcription enhancer sequences and protects downstream genes from chemical damages [3, 4]. Even though the mechanism of DNA protection exerted by AidB is still obscure, this protein might prevent alkyl damage either by binding and physically hiding the DNA molecule or by inactivating alkylating agents thus reducing their local concentration.

This paper focused on the elucidation of AidB biological role at the molecular level by functional proteomic approaches addressed to the identification of AidB protein partners. Since it is now clear that relevant biological mechanisms involve multi-protein complexes, the association of AidB with partners belonging to a particular mechanism will be strongly suggestive of its biological function. Isolation of AidB multi-protein complexes was performed *in vivo* by pull down strategies using a His-tagged form of the protein as bait in the presence and in the absence of methyl-methane sulfonate as methylating agent.

Authors identified several proteins involved in different biological mechanisms including various response complexes thus suggesting that AidB is endowed with different functions indicative of new cellular strategies to counteract alkylation stresses.

*Corresponding author: Angela Duilio, Associate Professor, research fields: microbiology, DNA repair mechanisms and M. tubercolosis. E-mail: anduilio@unina.it.

2. Experimental Sections

Proteomic grade trypsin, DTT (dithiothreitol), HEPES, KCl, $MgCl_2$, glycerol, ammonium bicarbonate and triton were purchased from Sigma. All used solvents were of the highest purity available from Romil.

2.1 Escherichia Coli Growths and Cell Extraction Preparation

E. coli cells growths were transformed with the construct pET22b-AidB. Bacterial culture was grown overnight in LB medium at 37 °C and it was diluted 1:100 in fresh medium containing ampicillin (100 μg/mL) and riboflavin (100 μM). At an $A_{600\ nm}$ of 0.4, the culture was divided in two aliquots and one of these was supplemented with 0.04% MMS (methyl methane sulfonate) that has been shown to induce the adaptive response. After one cell duplication cellular pellets were collected. The cells were resuspended in 20 mM Na_2HPO_4, 20 mM Imidazole, 500 mM NaCl, 1 mM PMSF (phenil methane sulphonyl fluoride) (pH = 7.4), disrupted by passage through a french press and centrifuged at centrifugal force of 14,000 x g for 15 min at 4 °C. The supernatant was collected and protein concentration was determined with the Bio-Rad protein assay, using bovine serum albumine as standard.

2.2 Pull-Down Experiments

Isolation of AidB partners complex was performed by using His-SelectTM Nickel (Sigma) containing Ni^{2+} ions immobilized to bind His-tagged AidB. A control was carried out in order to discriminate between proteins that interact specifically with the Ni^{2+} compared to those that bind in a nonspecific manner to the resin. For this reason, the stripping of the resin was executed by washing in 20 mM sodium phosphate, 0.5 M NaCl and 50 mM EDTA, for the purpose of removing Ni^{2+} ions. In this way, the resin lost the ability to interact specifically with the tag of histidines, but it was still able to establish nonspecific interactions. At this point, the resin was washed with 20 mM sodium phosphate, 0.5 M NaCl and 20 mM imidazole pH 7.4. The two protein extracts (2.5 mg) were incubated for 16 h at 4 °C with 100 μL of resin without nickel ions in the precleaning step. The extracts were then recovered and incubated with His-SelectTM Nickel resin for 16 h at 4 °C to bind AidB by tag of histidines and to isolate its complexes. Both the precleaning and affinity chromatography resins were recovered and washed with 20 mM sodium phosphate, 0.5 M NaCl and 20 mM imidazole pH 7.4. The elution was performed with sample buffer. The samples were then subjected to SDS-PAGE (SDS-polyacrilamide gel electrophoresis).

2.3 In Situ Digestion

Protein bands stained with Coomassie brilliant blue were excised from the gel and destained by repetitive washes with 0.1 M NH_4HCO_3 (pH 7.5) and acetonitrile. Samples were reduced by incubation with 50 μL of 10 mM DTT in 0.1 M NH_4HCO_3 buffer (pH 7.5) and alkylated with 50 μL of 55 mM iodoacetamide in the same buffer. Enzymatic digestion was carried out with trypsin (12.5 ng/μL) in 10 mM ammonium bicarbonate (pH 7.8). Gel pieces were incubated at 4 °C for 2 h. Trypsin solution was then removed and a new aliquot of the digestion solution was added; samples were incubated for 18 h at 37 °C. A minimum reaction volume was used as to obtain the complete rehydratation of the gel. Peptides were then extracted by washing the gel particles with 10 mM ammonium bicarbonate and 1% formic acid in 50% acetonitrile at room temperature.

2.4 LC/MS/MS (Liquid Chromatography Tandem Mass Spectrometry) Analyses

Tryptic peptide mixtures obtained from *in situ* digestions were analysed by LC/MS/MS using an HPLC-Chip/Q-TOF 6520 (Agilent Technologies).

The peptide mixtures were injected by auto sampler.

They were sent to the enrichment column of the chip at flow rate of 4 μL/min, in 98% water, 2% acetonitrile and 0.1% formic acid. Subsequently the peptides were eluted directly into the capillary column (C18 reversed phase), at a flow rate of 0.4 μL/min. The chromatographic separation was carried out with a linear gradient in 95% acetonitrile, 5% water and 0.1% formic acid. The eluate was then introduced in the ESI source for the tandem analysis. In this way each mass spectrum (range 300-2,400 m/z) was followed by one or more tandem mass spectra (range 100-2,000 m/z), obtained by fragmenting the most intense ions in each fraction eluted chromatographic. The acquired MS/MS spectra were transformed in Mascot generic file format and used for peptides identification with a licensed version of MASCOT (modular approach to software construction, operation and test, matix science, USA), in a local database.

2.5 Construction of Expression Vectors

The bacterial strains and plasmids used in this work are all reported in Table 1. The UvrA, DeaD, RecA, TnaA and Ada genes of *E. coli* K12 were amplified from genomic DNA by PCR (polymerase chain reaction). To obtain proteins tagged with c-myc epitope, the corresponding amplification products were digested with BamHI and XhoI and cloned into the pET22b-c-myc vector [4], respectively. All plasmids (Table 1) containing the coding sequence for the corresponding recombinant protein fused to a 6X histidine tag to facilitate protein purification by Ni^{2+} affinity chromatography. Plasmids construction was verified by automated DNA sequencing.

2.6 Production and Purification of Recombinant Proteins

Recombinant cells were grown at 37 °C to an OD (optical density) at 600 nm of about 0.5, at which time 0.05 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) was added in order to express UvrA, DeaD, RecA, TnaA and Ada genes. Selective antibiotic was used at concentration of 100 μg/mL ampicillin. After incubation, cells were harvested by centrifugation at centrifugal force of 5,000 x g for 15 min at 4 °C, resuspended in 50 mM Na_2HPO_4, 20 mM Imidazole, 500 mM NaCl, 1 mM PMSF (pH 7.4), disrupted by passage through a French press and centrifuged at centrifugal force of 14,000 x *g* for 30 min at 4 °C. Recombinant proteins were purified by affinity chromatography on His-Select Nickel Affinity Gel (Sigma). After 1 min of incubation at 4 °C, the matrix was collected by centrifugation at centrifugal force of 11,000 x g for 1 min and washed three times with same equilibration buffer. The recombinant proteins were eluted with buffer containing 500 mM imidazole in 20 mM Na_2HPO_4, pH 7.4, 0.5 M NaCl.

Protein concentration was estimated with Bradford reagent (Bio-Rad protein assay) and protein content

Table 1 Bacterial strains and plasmids used.

Strains/plasmids	Description	Reference or source
Strains		
C41 (DE3)	Strain that derives from BL21 [F⁻*ompT hsdSB* (r_B-m_B-) *gal dcm* (DE3)]. This strain has at least one uncharacterized mutation that prevents cell death associated with expression of many toxic recombinant proteins.	Ref. [3]
Plasmids		
pET22b(+)	Carries an N-terminal *pelB* signal sequence for potential periplasmic localization, plus an optional C-terminal His-tag sequence	Novagen
pET22b-Ada	pET22bΔ(BamHI-XhoI)Ω(Ada gene)	This work
pET22b-DeaD	pET22bΔ(BamHI-XhoI)Ω(DeaD gene)	This work
pET22b-TnaA	pET22bΔ(BamHI-XhoI)Ω(TnaA gene)	This work
pET22b-UvrA	pET22bΔ(BamHI-XhoI)Ω(UvrA gene)	This work
pET22b-RecA	pET22bΔ(BamHI-XhoI)Ω(RecA gene)	This work

was checked by SDS-PAGE.

2.7 Co-Immunoprecipitation and Western Blotting

For co-immunoprecipitations, *E. coli* strain C41 (DE3) was transformed with the following constructs: pET22b-c-myc-Ada, pET22b-c-myc-TnaA, pET22b-c-myc-DeaD, pET22b-c-myc-RecA and pET22b-c-myc-UvrA. After expression of the recombinant genes without induction, cells were harvested, suspended in 50 mM Na_2HPO_4 (pH 7.4), disrupted by passage through a French press and centrifuged at centrifugal force of 14,000 x g for 30 min at 4 °C. The supernatants were used for the co-immunoprecipitation experiments.

Cell lysates (0.5 mg) were incubated with agarose-linked c-myc antibody (Bethyl) and with agarose beads only (control of the experiment) at 4°C overnight. The beads were then collected by centrifugation. Precipitates were washed several times, the bound proteins were eluted with 1 × SDS-PAGE sample buffer and subjected to SDS-PAGE followed by Western Blot Analysis that was performed by using anti-AidB antibody (Primm, Milano Italy) and anti-c-myc mouse antibody (Calbiochem) as first antibodies and anti-mouse IgG conjugated to peroxidase as a secondary antibody (Calbiochem).

3. Results and Discussion

3.1 Isolation of AidB Complexes in E. coli Upon Exposure to MMS

In vivo isolation of AidB containing protein complexes were performed by transforming *E. coli* C41 strain with the pET22b-AidB construct carrying the AidB gene fused to a six histidine tags. The strain was grown in the absence and in the presence of sub-inhibitory concentrations of the alkylating agent MMS. Isolation of AidB complexes was performed by IMAC (ion metal affinity chromatography). The total protein extracts from the two samples were first submitted to a pre-cleaning step by incubation with His-Select beads lacking nickel ions in order to remove non-specific proteins. Eluates from the pre-cleaning were then recovered and AidB containing complexes were isolated by IMAC on His Select beads. After extensive washing, the proteins specifically bound to AidB bait were eluted with a strong ionic buffer containing 0.5 M imidazole and fractionated on SDS-PAGE stained with coomassie blue. Pre-cleaning samples were also eluted and used as control. Fig. 1 shows the obtained Coomassie black stained gel.

3.2 Identification of Proteins Specifically Interacting with AidB

The entire lanes from both samples (3, 4) and controls (1, 2) were cut in 24 slices and each gel slice was *in situ* digested with trypsin and the corresponding peptide mixtures directly analysed by LC/MS/MS procedures. Tandem mass spectral analyses provided both the accurate molecular mass and sequence information from the daughter ion spectra of each peptide. These data were used for database searches using a home version of the Mascot

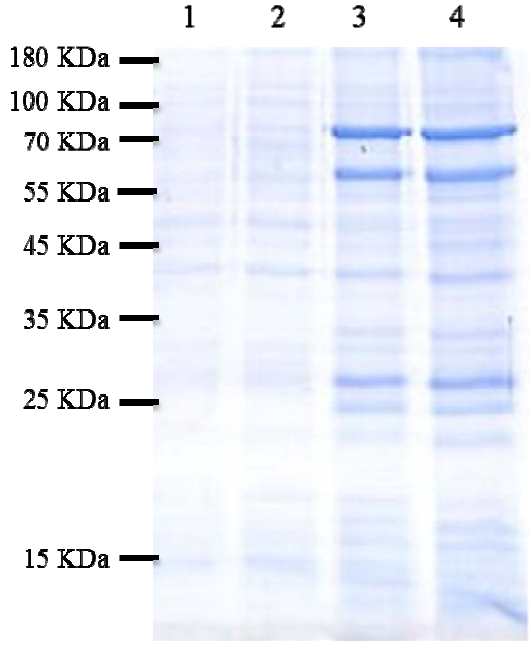

Fig. 1 SDS-PAGE fractionation of AidB complexes. Lanes 1 and 2 precleaning eluates. Lanes 3 and 4 AidB complexes in the absence and in the presence of MMS, respectively.

software leading to the identification of the proteins.

Common proteins identified in both the sample and the control gel slices were ruled out and only those solely occurring in the samples were considered as putative AidB inter-actors thus greatly decreasing the number of false positives. Proteins identified in the proteomic experiments are listed in Tables 2 and 3 where the protein name, the corresponding Swiss Prot code and the number of identified peptides are reported. The presence of the AidB bait in both lists constituted a sort of internal control indicating the correctness of the pull down experiment.

A total of 73 kinds of proteins were identified by the proteomic procedure, 17 of which were found both in the presence and in the absence of the methylating proteomic experiments are summarized in Tables 2 and 3. According to their reported biological activities, these identified inter-actors were grouped into different functional categories: metabolic pathways including several FAD and NAD^+ dependent dehydrogenases, stress response and transcription, translation and processing of DNA/RNA. Among others, we focused our attention on the stress response proteins for further investigations.

3.3 Validation of Protein-protein Interactions by Co-immunoprecipitation Experiments

Putative protein-protein interactions detected by the proteomic experiments were validated by co-immunoprecipitation experiments. Proteins involved in pathways strictly connected with DNA repair and protection mechanisms were firstly examined. Each putative protein partner was recombinantly expressed as c-myc-tagged protein in *E. coli* C41 cells and the cell extracts were immunoprecipitated with anti-c-myc-conjugated antibody. Immunoprecipitates were fractionated by SDS-PAGE and stained by Western blot analysis

Table 2 Proteins identified in the control sample.

In the absence of MMS	Swiss prot code	Peptides
2-oxoglutarate dehydrogenase E1 component (SucA)	P0AFG3	2
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex (AceF)	P06959	6
Phosphoenolpyruvate synthase (PpsA)	P23538	2
Bifunctional polymyxin resistance protein arnA (ArnA)	P77398	42
Glucosamine-fructose-6-phosphate aminotransferase [isomerizing] (GlmS)	P17169	10
Phosphoenolpyruvate-protein phosphotransferase (PtsI)	P08839	3
Protein AidB	P33224	20
Alkyl hydroperoxide reductase subunit F (AhpF)	P35340	7
Glycogen synthase (GlgA)	POA6U8	11
UDP-N-acetylmuramate: L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase (mpl)	P37773	2
NADP-specific glutamate dehydrogenase (gdhA)	P00370	12
3-oxoacyl-[acyl-carrier-protein] synthase 2 (fabF)	P0AAI5	3
Transcriptional activator protein (lysR)	P03030	9
Ribosomal small subunit pseudouridine synthase A (rsuA)	P0AA43	11
UPF0011 protein yraL (yhbJ)	P67087	6
Enoyl-[acyl-carrier-protein] reductase [NADH] (fabI)	P0AEK4	2
Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase (lpxA)	P0A722	2
FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase (fklB)	P0A9L3	6
Catabolite gene activator (crp)	P0ACJ8	22
UPF0011 protein yraL (yraL)	P67087	2
Uncharacterized protein yqjI (yqjI)	P64588	3
Uncharacterized protein ybgA (ybgA)	P24252	2
Ferric uptake regulation protein (fur)	P0A9A9	6
50S ribosomal protein L17 (rplQ)	P0AG44	3

Table 3 Proteins identified in the sample treated with 0.04% MMS.

In the presence of MMS	Swiss prot code	Peptides
UvrABC system protein A (UvrA)	P0A698	5
Aldehyde-alcohol dehydrogenase (AdhE)	P0A9Q7	2
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex (AceF)	P06959	5
Ribonucleoside-diphosphate reductase 1 subunit alpha (NrdA)	P00452	3
Maltodextrin phosphorylase (malP)	P00490	2
Bifunctional polymyxin resistance protein arnA (arnA)	P77398	37
Glucosamine-fructose-6-phosphate aminotransferase [isomerizing] (glmS)	P17169	20
Phosphoenolpyruvate-protein phosphotransferase (ptsI)	P08839	7
Cold-shock DEAD box protein A (DeaD)	P0A9P6	6
Succinate dehydrogenase flavoprotein subunit (sdhA)	P0AC41	3
GTP-binding protein typA/BipA (typA)	P32132	2
L-aspartate oxidase (NadB)	P10902	2
Chaperone protein hscA (hscA)	P0A6Z1	2
D-lactate dehydrogenase (dld)	P06149	2
Protein aidB (AidB)	P33224	19
Alkyl hydroperoxide reductase subunit F (ahpF)	P35340	9
Glucose-6-phosphate 1-dehydrogenase (zwf)	P0AC53	3
Glycogen synthase (glgA)	POA6U8	10
UDP-N-acetylmuramate: L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase (mpl)	P37773	3
Tryptophanase (TnaA)	P0A853	2
NADP-specific glutamate dehydrogenase (gdhA)	P00370	24
3-oxoacyl-[acyl-carrier-protein] synthase 2 (fabF)	P0AAI5	4
3-oxoacyl-[acyl-carrier-protein] synthase 1 (fabB)	P0A953	5
Succinylornithine transaminase (astC)	P77581	13
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (murA)	P0A749	11
Glutamate-1-semialdehyde 2,1-aminomutase (hemL)	P23893	4
S-adenosylmethionine synthetase (metK)	P0A817	5
Protein hflK (hflK)	P0ABC7	4
Isocitrate dehydrogenase [NADP] (icd)	P08200	2
Maltose/maltodextrin import ATP-binding protein malK (malK)	P68187	4
Transcription termination factor rho (rho)	P0AG30	3
ATP-dependent Clp protease ATP-binding subunit clpX (clpX)	P0A6H1	3
Cysteine desulfurase (iscS)	P0A6B8	3
Regulatory protein ada (ada)	P06134	9
Glycerol dehydrogenase (glda)	P0A9S5	3
Acetylornithine deacetylase (arge)	P23908	4
Lactose operon repressor (laci)	P03023	3
Glutamate-1-semialdehyde 2,1-aminomutase (gsa)	P23893	2
USG-1 protein (usg)	P08390	2
Riboflavin biosynthesis protein RibD (RibD)	P25539	4
P-protein (pheA)	P0A9J8	3
UDP-4-amino-4-deoxy-L-arabinose—oxoglutarate aminotransferase (ArnB)	P77690	3
Glyceraldehyde-3-phosphate dehydrogenase A (gabA)	P0A9B2	13
Elongation factor Ts (tsf)	P0A6P1	8
UPF0042 nucleotide-binding protein yhbJ (yhbJ)	P0A894	6
Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (accD)	P0A9Q5	2
Aspartate carbamoyltransferase catalytic chain (pyrB)	P0A786	2
Ribosomal small subunit pseudouridine synthase A (rsuA)	P0AA43	12

(Table 3 continued)

In the presence of MMS	Swiss Prot code	Peptides
Uncharacterized HTH-type transcriptional regulator yeiE (yeiE)	P0ACR4	10
UPF0042 nucleotide-binding protein yhbJ (yhbJ)	P0A894	11
Formyltetrahydrofolate deformylase (purU)	P0A440	10
UPF0011 protein yraL (yraL)	P67087	7
Enoyl-[acyl-carrier-protein] reductase [NADH] (fabI)	P0AEK4	2
Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase (lpxA)	P0A722	3
D-methionine-binding lipoprotein metQ (metQ)	P28635	2
FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase (fklB)	P0A9L3	6
GTP cyclohydrolase 1 (folE)	P0A6T5	2
Catabolite gene activator (crp)	P0ACJ8	21
Translation initiation factor IF-3 (infC)	P0A707	2
Uncharacterized protein yqjI (yqjI)	P64588	3
UPF0227 protein ycfP (ycfP)	P0A8E1	3
50S ribosomal protein L6 (rplF)	P0AG55	2
UPF0304 protein yfbU (yfbU)	P0A8W8	2
Ferric uptake regulation protein (fur)	P0A9A9	8
50S ribosomal protein L27 (rpmA)	P0A7L8	2
30S ribosomal protein S15 (rpsO)	P0ADZ4	3

using an anti-AidB antibody. Interaction of the individual partner with AidB was confirmed by the presence of a positive signal revealed by the western blot analysis.

As an example, Fig. 2 shows the Western Blot Analysis performed on the immunoprecipitate from *E. coli* cells expressing c-myc tagged UvrA. Fig. 2a shows the total cell extract (lane 1) and the corresponding immunoprecipitate (lane 3) immunorevealed by anti c-myc antibody demonstrating that UvrA was expressed by the recombinant cells and immunoprecipitated by the antibody. Fig. 2b shows the positive signal detected when the UvrA immunoprecipitate was incubated with the anti-AidB antibody, demonstrating the presence of AidB in the sample and confirming the interaction.

Positive interactions of AidB with UvrA, DeaD and TnaA were identified, whereas no positive bands in the western blot could be detected when Ada and RecA proteins were tested.

It should be underlined that both Cold-shock DEAD box protein A (DeaD) and Tryptophanase (TnaA) had already been identified in complex with AidB in the transcriptional machinery gathered at the *E. coli* arnB P1 promoter in the presence of MMS [4].

3.4 Discussion

The biological mechanism underlying the adaptive response to DNA alkylation damages in *E. coli* was extensively described. Sub inhibitory concentration of methylating agents induces the expression of four

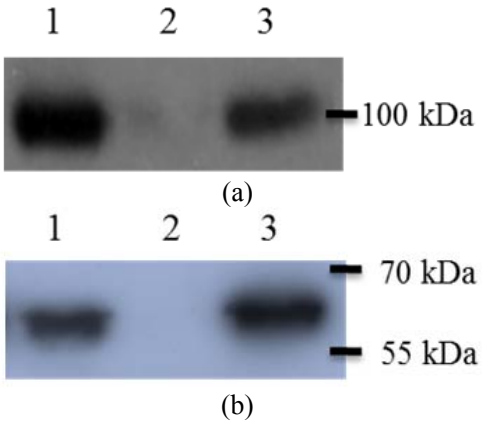

Fig. 2 (a) Western blot analysis of the total cell extract from *E. coli* C41 cells producing c-myc-UvrA (lane 1) and the UvrA containing immunoprecipitate (lane 3). Lane 2 contains the precleaning; (b) Western blot analysis of the total cell extract from *E. coli* C41 cells (lane 1) and the UvrA immunoprecipitate revealed by the anti-AidB antibody (lane 3). Lane 2, precleaning.

genes (Ada, AlkA, AlkB, AidB) involved in the direct repair of DNA alkylation damages. While the role of the first three proteins, Ada, AlkA and AlkB, was clearly defined [5], the involvement of AidB in this system has long been known but its specific role in the protection/repair of DNA is still obscure. Previous data demonstrate that AidB interacts with DNA very likely to protect the nucleic acid from alkylating molecules but it is not able to repair the DNA molecule following alkylation [4]. In addition, the protective action of AidB is preferentially expressed on DNA regions containing upstream elements. This observation led to the hypothesis that AidB might belong to a putative pathway of degradation of alkylating agents through its FAD dependent dehydrogenase activity. Alternatively, AidB might protect these DNA regions by physically interact with them thus impairing the dangerous action of alkylating agents.

Proteomic approaches were designed to shed some light on the mechanism of action of AidB through the identification of its protein partners *in vivo*. AidB partners were isolated by immunoprecipitation procedures both in the absence and in the presence of MMS as alkylating agent and the individual protein components identified by mass spectrometry. Several proteins were identified in both conditions, although the number of AidB molecular partners is considerably higher in the presence than in the absence of MMS.

Proteins identified under methylating stress conditions were grouped in three large categories according to their reported biological activities, stress response, energetic metabolic pathways, and nucleic acid metabolism (transcription, processing and translation). In particular, AidB was found to interact with UvrA whose expression is under the control of the SOS response system involved in DNA damages response [6]. This interaction was also validated by co-immunoprecipitation experiments confirming proteomic data. Interaction of these two proteins is very interesting since UvrA is part of the UvrABCD nucleotide excision system involved in removing modified nucleotides as a result of several different DNA modifications including formation of covalent bonds, local unfolding, abnormal folds and variations in charge distribution [7]. UvrA works in a multienzyme complex with the specific role of examining the DNA molecule in search for modifications in order to allow the other proteins of the complex to perform the excision of the damaged nucleotides [8]. This protein is generally present at very low concentrations within the cell but its expression strongly increases under stress conditions [9]. Interaction of AidB with UvrA might then indicate that AidB is involved in different response complexes other than the Ada-dependent adaptive mechanism, suggesting new cellular strategies to minimize DNA damages.

In the presence of MMS, several other proteins belonging to metabolic pathways were identified suggesting a general mechanism of energy production developed by *E. coli* to counteract stress conditions. Among these, many AidB inter-actors are involved in the biosynthesis of fatty acids, indicating the occurrence of a possible mechanism of cell wall repair. It is likely, in fact, that besides DNA modifications, alkylating agents might produce damages to cell wall components thus inducing a specific response from the microorganism that enhances the metabolic synthesis of fatty acids in an attempt to repair or replace damaged membrane components thus strengthening the physical defenses of the cell.

4. Conclusions

In conclusion, the data reported in this paper identified a novel interaction between AidB and UvrA which is a component of the UvrABCD system involved in DNA repair under several stress conditions delineating a new possible role for the AidB protein. Moreover, the identification of several proteins belonging to the fatty acid biosynthetic

pathway as AidB partners pointed out to a mechanism of cell wall consolidation as a new defense strategy of the cell against the poisonous effect of methylating agents.

Acknowledgments

This paper was in part supported by PON01_01802 and MIUR FIRB "Italian Human Proteome Net" RBRN07BMCT grants.

References

[1] Landini, P.; Hajec, L. I.; Volkert, M. R. Structure and Transcriptional Regulation of the *E. coli* Adaptive Response Gene AidB. *J. Bacteriol* **1994,** *176*, 6583-6589.

[2] Bowles, T.; Metz, A. H.; Quin, J. O.; Wawrzak, Z; Eichman, B. F. Structure and DNA Binding of Alkylation Response Protein AidB. *PNAS* **2008,** *105*, 15299-15304.

[3] Rippa, V.; Duilio, A.; Pasquale, P.; Amoresano, A.; Landini, A.; Volkert, M. R. Preferential DNA Damage Prevention by the *E. coli* AidB Gene: A New Mechanism for Protection of Specific Genes. *DNA Repair* **2011,** *10*, 934-941.

[4] Rippa, V.; Amoresano, A.; Esposito, C.; Landini, P.; Volkert, M.; Duilio, A. Specific DNA Binding and Regulation of Its Own Expression by the AidB Protein in *Escherichia coli*. *J. Bacteriol* **2010,** *192(23)*, 6136-6142.

[5] Miroux, B.; Walker, J. E. Over-Production of Proteins in *Escherichia coli*: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* **1996,** *260*, 289-298.

[6] Sedgwick, B.; Lindahl, T. Recent Progress on the Ada Response for Inducible Repair of DNA Alkylation Damage. *Oncogene* **2002,** *21*, 8886-8894.

[7] Croteau, D. L.; Della, V. M. J.; Perera, L.; Van, H. B. Cooperative Damage Recognition by UvrA and UvrB: Identification of UvrA Residues That Mediate DNA Binding. *DNA Repair* **2008,** *7(3)*, 392-404.

[8] Wagner, K.; Moolenaar, G. F.; Goosen, N. Role of the Insertion Domain and the Zinc-Finger Motif of *Escherichia coli* UvrA in Damage Recognition and ATP Hydrolysis. *DNA Repair* **2011,** *10(5)*, 483-496.

[9] Gu, C.; Zhang, Q.; Yang, Z.; Wang, Y.; Zou, Y.; Wang, Y. Recognition and Incision of Oxidative Intrastrand Cross-Link Lesions by UvrABC Nuclease. Biochemistry 2006, 45(35), 10739-10746.