UNIVERSITÀ DEL PIEMONTE ORIENTALE

Università degli Studi del Piemonte Orientale
"Amedeo Avogadro"

Dipartimento di Scienze del Farmaco

Dottorato di Ricerca in "Scienza delle Sostanze Bioattive"
XXVIII ciclo a.a. 2012-2015

BIOCHEMICAL, STRUCTURAL AND INTERACTION STUDIES OF *Mycobacterium tuberculosis* ALKYLATED-DNA REPAIR PROTEINS

Samarpita Lahiri

Università degli Studi del Piemonte Orientale
"Amedeo Avogadro"

Dipartimento di Scienze del Farmaco

Dottorato di Ricerca in "Scienza delle Sostanze Bioattive"
XXVIII ciclo a.a. 2012-2015

BIOCHEMICAL, STRUCTURAL AND INTERACTION STUDIES OF *Mycobacterium tuberculosis* ALKYLATED-DNA REPAIR PROTEINS

SAMARPITA LAHIRI

Supervised by
Dott.ssa Franca Rossi

PhD program co-ordinator Prof. Luigi Panza

....Bodhisattwa Gangopadhyay

Contents

Chapter 1

Introduction

Chapter 2

Outline of the Thesis

Chapter 3

“Crystal structure of *Mycobacterium tuberculosis* O6-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA”

Chapter 4

“Analyses of macromolecular interactions sustaining NER- and OGT-mediated alkylated-DNA repair in *Mycobacterium tuberculosis*”

Chapter 5

Concluding Remarks

List of publications

Acknowledgments

Introduction

1. Tuberculosis as a global disease

Tuberculosis (TB), one of the oldest recorded scourges of mankind, is still one of the biggest killers among the communicable diseases from a single infectious agent. In facts, despite the availability of effective short course chemotherapies (Directly Observed Treatment Short course-DOTS), the *Bacille Calmettee-Guerin* (BCG) vaccine and the use of modern antibiotics, TB ranks alongside the human immunodeficiency virus (HIV) as a leading cause of mortality worldwide. TB is a major global health problem that causes ill-health among millions of people each year, with a devastating social and financial impact.

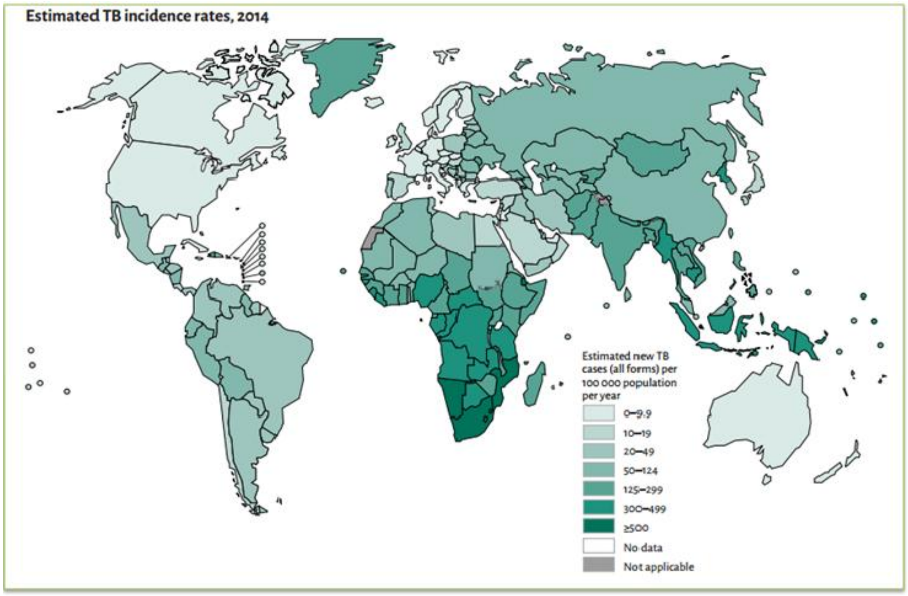

Estimated TB incidence rates, 2014
Estimated new TB cases (all forms) per 100 000 population per year
0–9.9
10–19
20–49
50–124
125–299
300–499
≥500
No data
Not applicable

Figure 1: Estimated new TB incident rates (TB cases per year per 100 000 population per year) in 2014 [1]

According to the latest Global Tuberculosis Report from World Health Organization (WHO) [1], in 2014 there were an estimated 9.6 million new TB cases, among which 5.4 million were men, 3.2 million were women and 1.0 million were children (Figure 1). Amidst of these new TB cases, 58% were found in South-East Asia and Western Pacific regions, and 20% in Africa. Notably countries like India, China, and Indonesia had the highest number of emerging TB cases that are 23%, 10% and 10% of the global total, respectively. There were also 1.5 million TB deaths which includes 1.1 million HIV-negative people and 0.4 million HIV-positive people**.**

The co-epidemics of TB and HIV is a TB issue that needs to be addressed, at the light of the observation that an estimated 12% of the 9.6 million of new TB cases were found HIV-positive [1].

TB is widely recognized in low-income countries, showing the greatest incidence rates in the poorer section of the community, where risk factors such as the lack of basic health services, poor nutrition, alcoholism and inadequate living conditions contribute to the spreading of the disease. Although TB strongly correlates with poverty, a main TB concern is the long-term multi-drug treatment regimen, leading to anticipated non-compliance of the patient and/or resulting in increased drug toxicity, with escalating mortality rates due to detrimental side-effects. This situation is further aggravated by the observation that, often as a consequence of premature termination of drug therapy, a steadily increasing number of drug resistant TB cases is reported, whose treatment is more challenging, thus allowing TB to reach pandemic proportions [2].

2. *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (MTB), the highly predominant causative agent of TB in humans, is a member of the so called "*Mycobacterium tuberculosis* complex" (MTBC), a group of genetically related mycobacterial species. MTBC also includes *M. bovis* (the etiological agent of TB in cattle), and its live-attenuated BCG strain, which is at present the only approved vaccine used in human TB prophylaxis [3].

Similar to the vast majority of mycobacteria, MTB is a non-motile, non-sporulating, acid-fast Gram-positive bacillus that appears microscopically as a straight or slightly curved rod, 1 to 4 µm in length and 0.3 to 0.6 µm wide. Other worth mentioning characteristics of MTB are its slow generation time (approx. 24 hours in synthetic medium or in animal models), which is mainly due to the presence of a tough, complex cell wall; the capability to enter in a dormant state that can last for long periods; the peculiar strategies adopted for intracellular survival and reactivation; the high GC-content of its genomic DNA and an overall high level of genetic stability [4].

In order to isolate and cultivate MTB *in vitro*, Middlebrook´s and Lowenstein-Jensen (L-J) media are most commonly used, at a growth optimum temperature of 37°C; in the case of growth on solid media, 4-5 weeks are needed for small and slightly buff colored MTB colonies becoming visible to the naked eye (Figure 2A). As other acid-fast bacteria, MTB could be easily visualized by Ziehl-Neelsen (or acid-fast) staining, whose pattern is characterized by a poor absorption quality and a high stain retention capacity, making the MTB bacilli bright red and emerging out clearly against the backdrop (Figure 2B) [5].

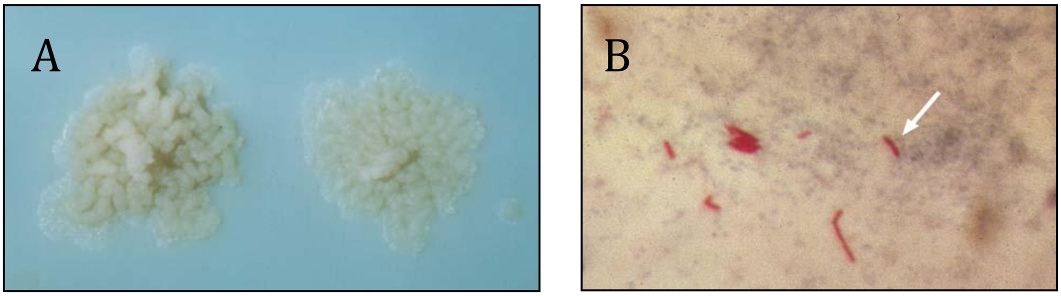


Figure 2: (A) Colonies of MTB on L-J medium against the blue backdrop. (B) Acid-fast MTB emerging out as bright red cells [5]

Many of the peculiar MTB features depend on its cell wall, which is different from both Gram-positive and Gram-negative bacteria. As shown in the Figure 3 (panel B), and moving from the inner plasma membrane towards the surface of the bacillus, three covalently linked layers can be identified in the MTB cell wall skeleton, built up by peptidoglycan, arabinogalactan and mycolic acids. In particular, the outer part of the mycolic acid layer (also called mycomembrane) contains various free lipids -like lipoarabinomannan (LAM), lipomannan, phenolic gycolipids, phthiocerol dimycocerosates, threalose dimycolate (cord factor), sulpholipids and phosphatidylinositol mannosides that are interposed with mycolic acids. The outermost layer of the MTB cell wall is represented by the capsule, which is mainly composed of polysaccharides [6].

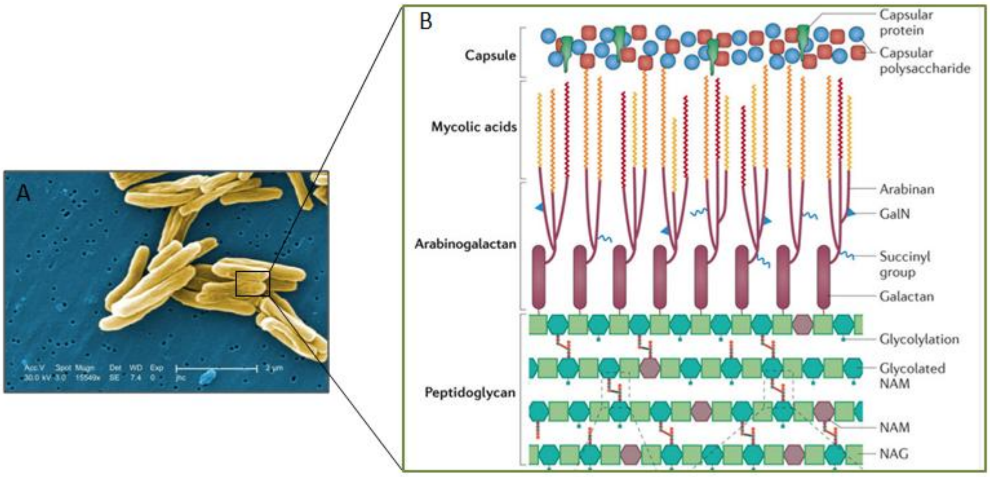

A
Acc.V Spot Magn Det WD Exp
30.0 kV 3.0 15549x SE 7.4 0
2 μm
B
Capsule
Mycolic acids
Arabinogalactan
Peptidoglycan
Capsular protein
Capsular polysaccharide
Arabinan
GalN
Succinyl group
Galactan
Glycolylation
Glycolated NAM
NAM
NAG

Figure 3: (A) Scanning electron microscope image of *M. tuberculosis;* (B) Schematic representation of mycobacterial cell wall [adapted from 5 and 8]

Among all the components of the mycobacterial waxy cell wall, three major fractions that deserves attention are the mycolic acids, cord factor (threalose-6,6-dimycolate) (Figure 4), and the wax-D, which together confer many of the unique characteristics of this *genus* such as acid-fastness, hindrance to crystal violet (differently from genuine Gram-positive bacteria), extreme hydrophobicity, resistance to drying, acidity/alkalinity, and many antibiotics, as well as distinctive immuno-stimulatory properties [5], therefore playing a crucial role in TB pathogenesis.

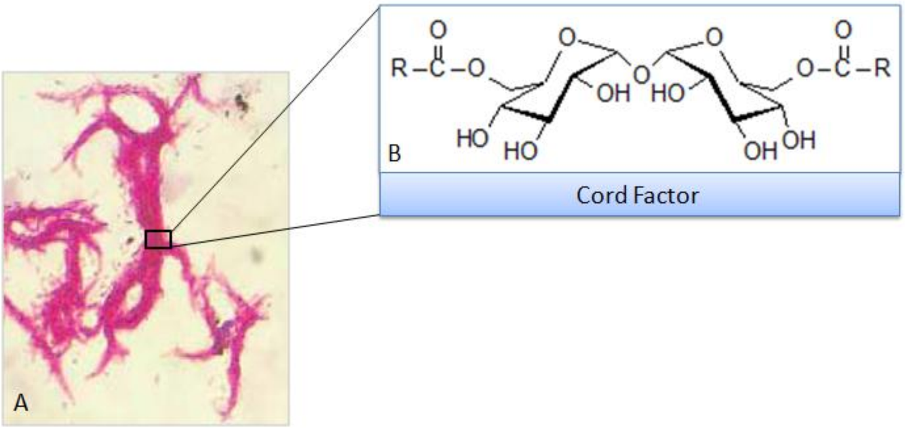


Figure 4: The cord factor is responsible for the unique feature of MTB known as serpentine cording, resulting in the formation of rope-like tangles (A) phenotype typically associated with virulent strain [9] (A) Acid-fast stained MTB cells showing serpentine cording (B) Chemical structure of cord factor (Tetrahalose 6,6´-dimycolate)[adapted from10,11].

3. Pathogenesis of TB in human

The pathogenesis process of TB in humans can be divided in several stages that are depicted below:

Transmission through aerosol droplets

MTB bacilli are carried in air-borne particles termed droplet nuclei of 1-5 µm in diameter and that are exhaled, in the form of aerosol, by individuals who have laryngeal or pulmonary active TB disease, by coughing, sneezing or talking (Figure 5A). Depending on the environmental conditions, these infected aerosol droplets can remain suspended in the air for several hours before inhalation by a healthy subject. Two important factors that affect the rate of survival of mycobacteria at this stage are desiccation and UV irradiation [12,13]. The success of the infection of exposed individuals also depends upon the frequency and the duration of the exposure, and the concentration and the size of the droplet nuclei. Indeed, once inhaled, the MTB-infected particles make their way through the upper respiratory tract (*i.e.* throat or nasal passages, where the infection is unlikely to develop); the smaller nuclei then traverse towards the bronchi to reach the alveoli of the lungs where the primary infection can begin [14].

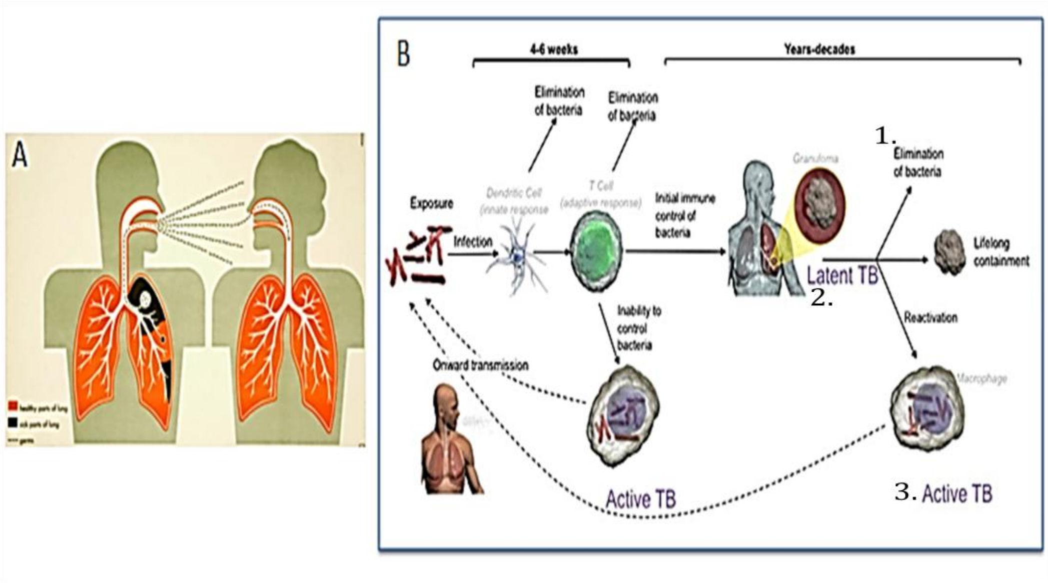

A
healthy parts of lung
sick parts of lung
germs
B
4-6 weeks
Years-decades
Elimination of bacteria
Elimination of bacteria
Exposure
Infection
Dendritic Cell (innate response
T Cell (adaptive response)
Initial immune control of bacteria
Granuloma
1.
Elimination of bacteria
Lifelong containment
Latent TB
2.
Reactivation
Inability to control bacteria
Onward transmission
Macrophage
Active TB
3. Active TB

Figure 5: (A) Transmission through aerosol droplet [15] (B) Three alternative fates of MTB following infection: 1) elimination by host immune system, 2) containment (as latent TB infection), 3) progression towards full-blown active disease [modified from 16].

Alveolar Macrophage infection

Inside the lung alveoli MTB bacilli are ingested by resident phagocytes (*i.e.*, the alveolar macrophages and dendritic cells), or, alternatively, by monocytes recruited from peripheral blood. MTB appears to gain entry into the macrophages *via* cell-surface molecules like complement receptor, mannose receptor, Fc and scavenger receptors [17]. Mainly due to the immune system status of the host, the initial infection can be either wiped out (resulting in an abortive infection), or contained as a latent TB (in the vast majority of the cases), or it can progress towards the full blown active disease, including the uncontrolled MTB bacilli proliferation and spread to other organs (Figure 5B) [18].

Macrophages are the primary defense against microbial invaders, due to their efficiency in internalizing microbes, engulfing them into phagosomes, and promoting the phagosome-lysosome fusion, that finally leads to the destruction of the cargo (Figure 6). Phagosome-confined pathogens experience stressful conditions like i) an increased acidification, ii) the exposure to highly reactive oxygen species (ROS) and reactive nitrogen intermediates (RNI) -generated by the host cell enzymes NADPH phagocyte oxidase and inducible nitric oxide synthase (iNOS)-, that can damage the captured microbes by modifying their DNA, lipids, proteins and active centers of metal-dependent proteins (like iron-containing siderophore), and iii) the release of hydrolytic enzymes and cationic antimicrobial peptides [19,20].

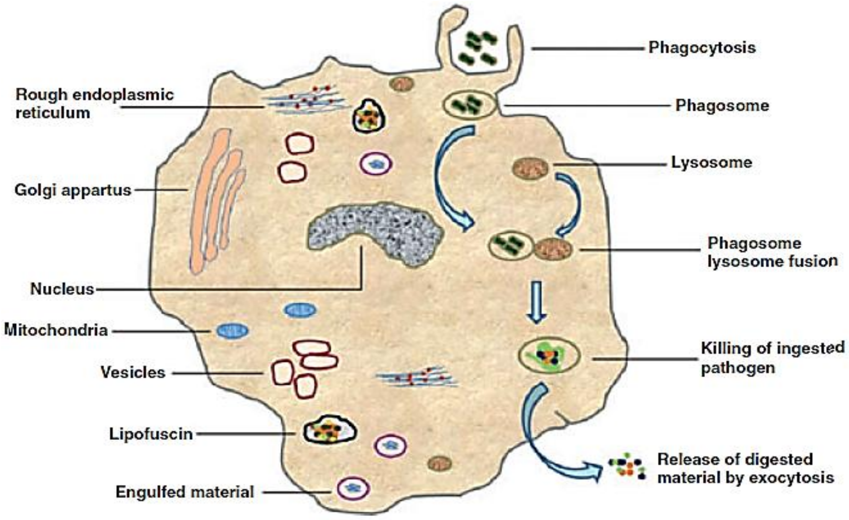
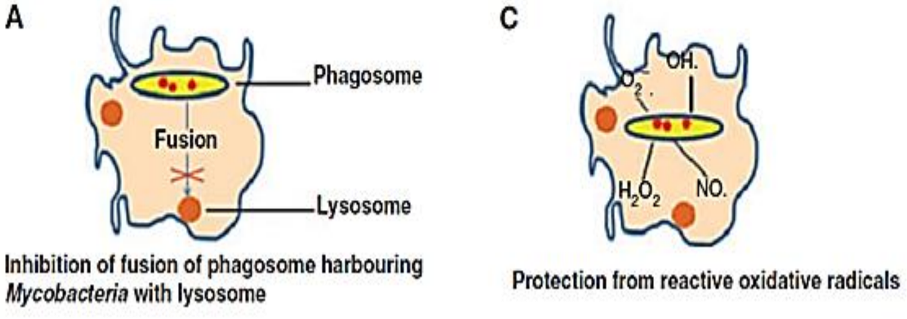

Phagocytosis
Rough endoplasmic reticulum
Phagosome
Lysosome
Golgi appartus
Phagosome lysosome fusion
Nucleus
Mitochondria
Killing of ingested pathogen
Vesicles
Lipofuscin
Release of digested material by exocytosis
Engulfed material

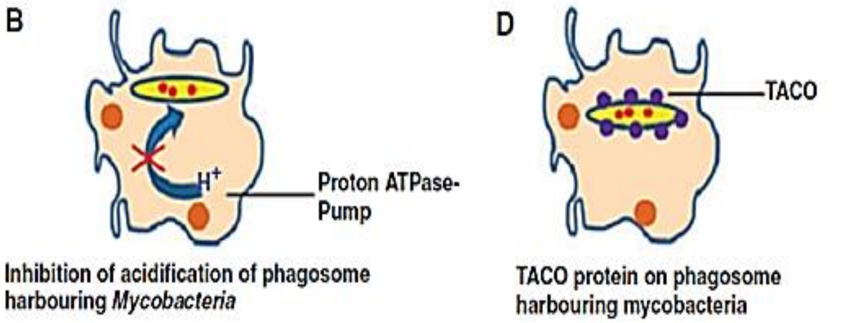
Figure 6: Schematic representation of macrophage showing the main steps of the process leading to microbial invaders elimination [21]

As a well-adapted intracellular pathogen, MTB evolved unique survival strategies to persist successfully within macrophages. In particular, inside the phagosome, MTB can fight the aforesaid hazardous acts at multiple levels. First, the bacilli interfere with the host intracellular trafficking pathways by modulating events in endosomal/phagosomal maturation pathway, and hampering the phagosome-lysosome fusion; the delivery to lysosomes is also prevented by the retention of the host TACO protein (tryptophan aspartate-containing coat protein) on the cytoplasmic surface of the MTB-harboring phagosome membrane. Moreover, MTB limits an excessive acidification by blocking the proton-ATPase pumps (normally reducing neutral pH to acidic pH), with the concomitant secretion of mycobacterial urease, an enzyme producing neutralizing ammonia from urea; finally MTB produces enzymes (like catalase-peroxidase and superoxide dismutase) for neutralizing the effects of oxidative radicals (Figure 7) [21,22]. Overall, these phenomena allow MTB to stably settle within the macrophages, which are among the most inhospitable cells of the body.



Inhibition of fusion of phagosome harbouring *Mycobacteria* with lysosome

Protection from reactive oxidative radicals



Inhibition of acidification of phagosome harbouring *Mycobacteria*

TACO protein on phagosome harbouring mycobacteria

Figure 7: Some of the survival strategies used by MTB inside activated macrophages to evade host cell defense [adapted from 21]

Formation of Granuloma

To contain the dissemination of the infection, infected macrophages, laden with replicating MTB, induce a pro-inflammatory response that recruits other cells, in a well-defined order; this process leads to the formation of the granuloma (or tubercle), the hallmark of TB, that provides a niche where the MTB bacilli could reside for long periods of time, evading the host immune response, and establishing the asymptomatic, non-contagious, latent phase of the disease (Latent Tuberculosis Infection, LTBI) [23]. This structure is maintained and stabilized by events mediated by a delicate cross-talk between the host and the pathogen. Moreover, the unfavorable conditions inside the granuloma, such as nutrient limitation and low oxygen tension, restricts the metabolic activities of subpopulations of MTB bacilli leading to their dormancy [18].

In humans, the TB granulomas show a high histological/morphological plasticity and three major types can be distinguished, which are not independent entities but forms a continuum. These formations are i) the solid granuloma, typifying the containment phase of the infection; ii) the necrotic granuloma, characterizing the early stages of active TB; and iii) the caseous granuloma, typically observed at the end-stage of the disease or in severe TB (Figure 8)**.**

The solid granulomas that prevail during LTBI, is composed of mononuclear phagacytes, dendritic cells along with T and B lymphocytes. The central part is predominated by mononuclear phagocytes, fibroblasts and dendritic cells while the lymphocytes form the outer ring. Solid granuloma is typically surrounded by a fibrotic wall, and has a low burden of MTB with low metabolic activity (dormant bacilli). The necrotic granuloma remains well-structured but forms necrotic center consisting of solid cell debris. This architecture has lower access to blood vessel inducing hypoxia and starvation conditions. In the caseous granuloma, the center liquefies, leading to cavity formation followed by reestablishing normal oxygen tension. Moreover, caseous material represent an abundant source of nutrients enhancing the resuscitation and the multiplication of the pathogen. This stage finally allows MTB bacilli to gain access to the alveolar space paving their way to dissemination [18,24].

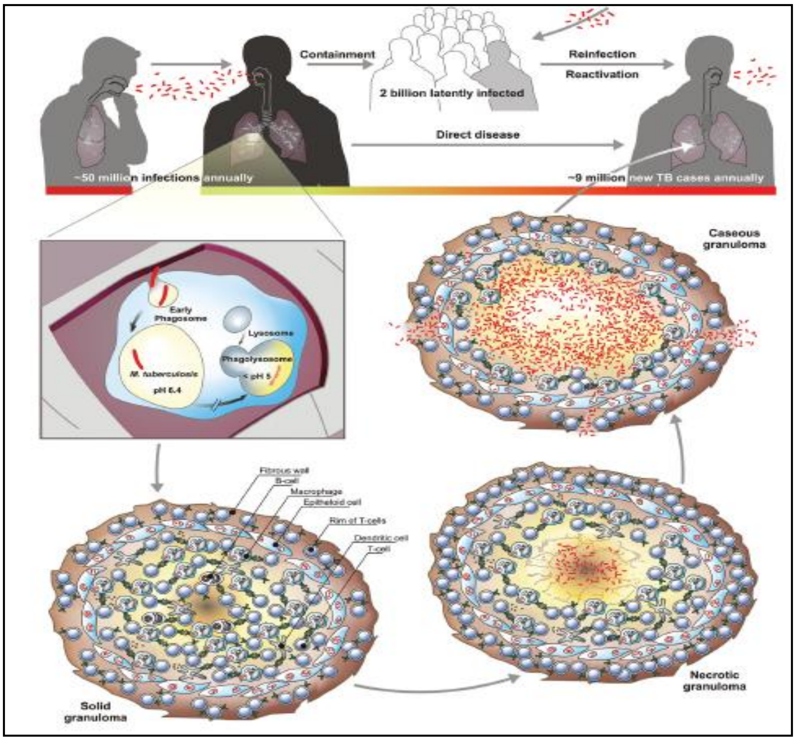

Containment
Reinfection
Reactivation
2 billion latently infected
Direct disease
~50 million infections annually
~9 million new TB cases annually
Caseous granuloma
Early Phagosome
Lysosome
Phagolysosome
M. tuberculosis
pH 6.4
< pH 5
Fibrous wall
B-cell
Macrophage
Epithelioid cell
Rim of T-cells
Dendritic cell
T-cell
Solid granuloma
Necrotic granuloma

Figure 8: The complete TB cycle, mainly focused on granuloma maturation stages [18]

Reactivation of infection

The containment of the infection within granuloma declines when the immune status of the host changes, as a consequence of old age, malnutrition, drug-induced immune-suppression, or during co-infection with HIV. This results in progression from LTBI to active TB. At this stage, caseating granuloma degenerate, followed by the rupture and spilling of thousands of viable infectious bacilli into the lung airways. The development of productive coughing promotes the release of MTB-infected aerosols into the environment and the potential spread of infectious bacilli to other people (Figure 8) [23]. The progression from LTBI to TB disease may occur at any time, from soon to many years later after infection.

Taken as a whole, the complexity of the MTB life cycle, and the multifarious ways the bacillus evolved to escape the host's immune defenses, explain the MTB considerable success as a human pathogen, while posing a great challenge to TB eradication.

4. TB diagnosis and therapy

The containment of TB spread is based on improved diagnostic methods and early antibiotic therapy, as the potency of BCG-vaccination against TB caters limited protection during childhood and also significantly varies among population **[25** and references therein**]**.

TB diagnosis

Last decade has seen significant advances in the methodology of TB diagnosis since the acid fast bacilli smear microscopy was insufficient to deal with the dual issue of drug-resistance TB and HIV-associated TB. There are several TB detection methods, like phenotypic methods (*e.g.* thin-layer agar, colour test, the microscopic observation drug susceptibility assay); genotypic methods (*e.g.* Real-time PCR, and sequencing); immunological methods (detection of antibodies/antigen); and also typical techniques like Interferon-Gamma Release Assays. Particular methodologies including automated real-time polymerase chain reaction (PCR)-based test truly revolutionizes the diagnosis of drug-resistant TB. However, little progress has been made in search of a true point-of-care test for TB. Thus, an impetus is needed for the development of rapid as well as accurate TB diagnostics for the lowest level of the health system, where the tuberculin test, radiographic analysis (chest X-ray) and sputum culture still represent the only means of diagnosis that are often not accurate [26,27].

Anti-TB drugs and treatment Regimen

As reported by WHO, multi-drug therapy is comprised of an initial intensive phase during which 3-4 first-line drugs i.e. rifampicin (RIF), isoniazid (INH), pyrazinamide (PYZ), ethambutol (ETB) and streptomycin- are administered daily for 2 months, followed by the continuation phase, consisting in the administration of RIF and INH for next 4 months, either daily or 3 times per week [28].

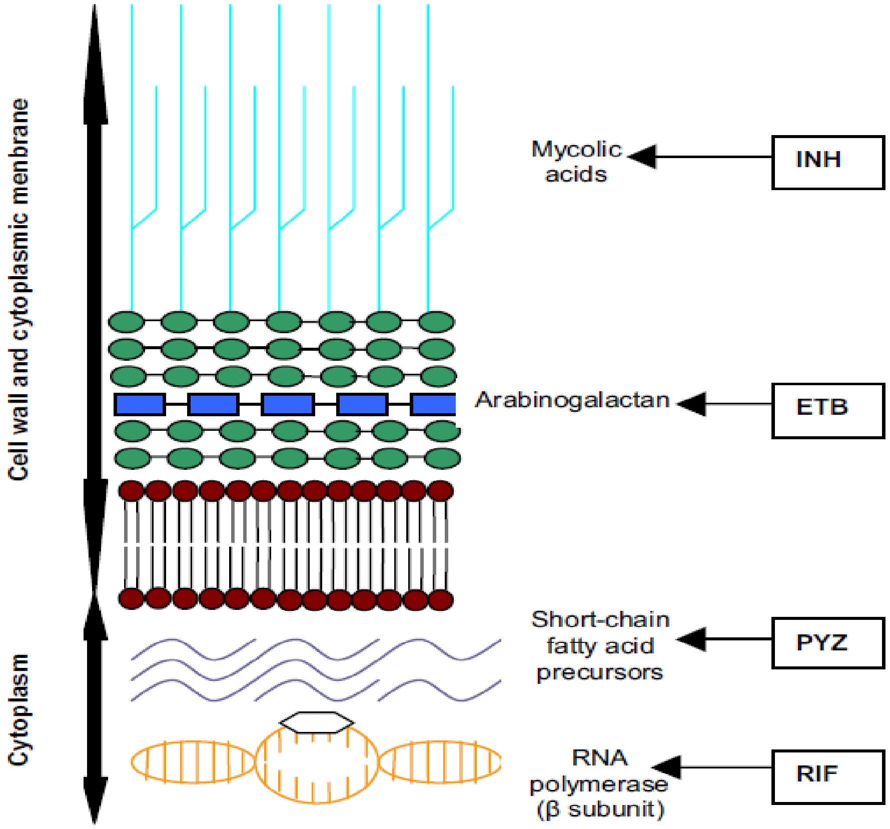

Cell wall and cytoplasmic menbrane
Cytoplasm
Mycolic acids
INH
Arabinogalactan
ETB
Short-chain fatty acid precursors
PYZ
RNA polymerase (β subunit)
RIF

Figure 9: Sites of action of principal anti-TB drugs [29]

INH together with streptomycin and ETB kills most of the rapidly replicating bacilli in the first 2 weeks of treatment. On the other hand, RIF eradicates slow or non-replicating organisms while the high sterilizing effect of PYZ serves to act on semi-dormant bacilli that remain unaffected by other anti-TB agents. The most potent anti -TB drug combination, INH and RIF, kill more than 99% of tubercular bacilli within 2 months of the initiation phase of the therapy. The sites of action of these principal anti-TB agents are shown in (Figure 9) [28,29].

The long-term multi-drug treatment regimen required for curing active TB, together with the absence of drugs targeting the latent disease, and therefore reducing the proportions of the MTB reservoir, contribute to make the global TB management and treatment far from being satisfactory, especially considering the warring phenomenon of MTB drug resistance.

5. MTB Drug-Resistance

Inadequate drugs with improper drug dosages, irregular treatment or even administration of a single drug to a failing regimen allows the selective growth of drug-resistant pathogens and/or acquisition of a drug resistant phenotype by a previously sensitive organism, thus imposing the drug resistance phenomenon as a “man-made” event.

Epidemiologically, drug resistance phenomenon in TB is classified into two types.

- *Primary drug resistance*: this situation is observed in untreated patients that are found to have drug-resistant bacilli, and is most frequently due to a new infection by a drug resistant strain.

- *Secondary drug resistance*: this situation develops during TB therapy because of inadequate, improper and irregular prescribed treatment regimen or incomplete capacitance to medication protocols; it could also be due to other conditions like drug mal-absorption or drug-drug interaction causing low serum levels [30].

As a consequence of improper administration of prescribed TB drug regimen, or due to abrupt augmentation of medicines by TB affected individuals, different drug-resistant strains of MTB develop leading to multi-drug resistant TB (MDR-TB), extensively-drug resistant TB (XDR-TB) and total drug resistant TB (TDR-TB).

- *MDR-TB*: is defined by the resistance to the two most commonly used drugs in first-line TB therapy, INH and RIF. The main causes of MDR-TB are poor patient management, non-adherence to the prescribed regimen, a poor national program or sometimes a combination of all these three [31]. According to the latest WHO report, in 2014 an estimated 190,000 people died of MDR-TB [1].

- *XDR-TB:* is defined by the resistance to any fluoroquinolone, and at least one of three injectable second-line drugs (capreomycin, kanamycin, and amikacin), in addition to INH and RIF, making XDR-TB medication extremely complicated, and representing an increasing health problem [32].

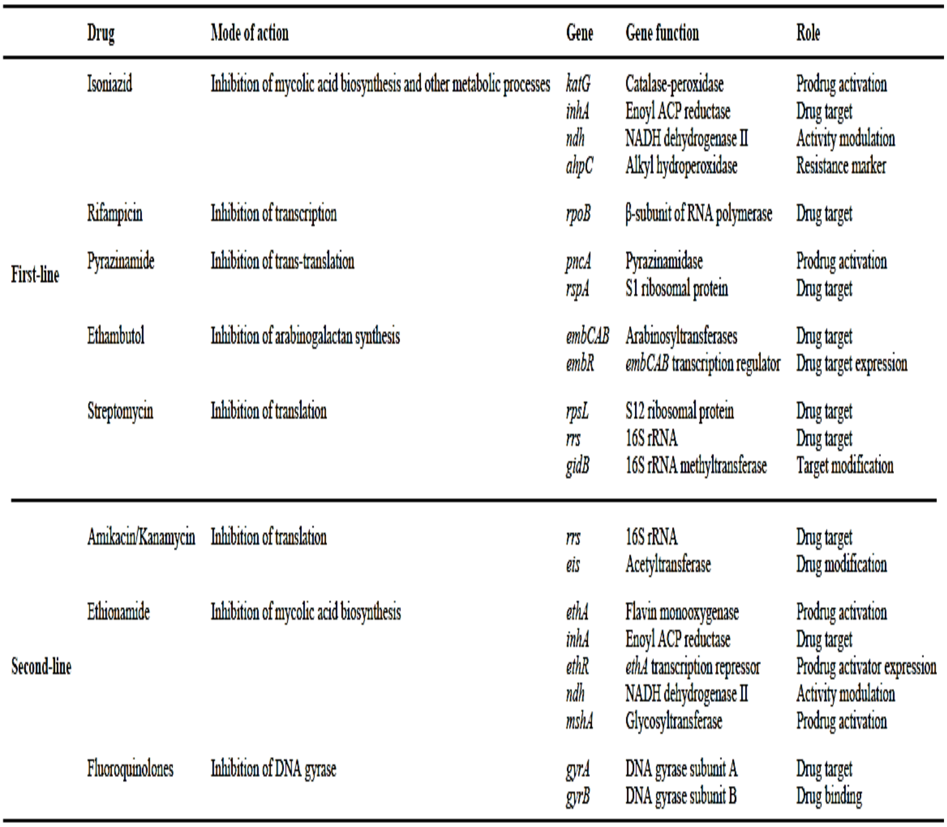
- *TDR-TB*: most recently MTB strains which are resistant to all kinds of first- and second-line drugs developed thus, is posing a major global health concern [33].

In many bacteria, the genetic determinants of the drug-resistance phenomenon are carried through mobile elements that, by promoting the horizontal transfer of genetic material, drive the evolution and adaptation of the microbe. On contrast, in MTB, horizontal transfer of resistance genes via plasmids or transposons has not been reported, and all currently well-studied acquired resistances are mediated through chromosomal mutations that arise spontaneously or under the selective pressure of antibiotics use. This kind of drug resistance is called “acquired resistance” and the MTB genes involved are shown in Table 1 [34,35].

Another type of drug resistance that is seen in MTB is the intrinsic antibiotic resistance. This kind of resistance can be offered either by the presence of the complex mycobacterial cell wall *per se* (*e.g.* the penetration of β-lactams through the MTB cell wall is hundreds fold slower than in *E. coli*), or by specialized mechanisms that allow the active neutralization/elusion of antibiotic actions [35].

Extensive MTB genomic analyses together with comparative studies reveal that the genes repertoire coding for proteins involved in DNA Replication, Recombination and Repair (the so called “3R-genes”), shows a level of sequence variability that is significantly higher than that characterizing the house-keeping genes, a feature that could promote the incidence and/or allow the fixation of spontaneous mutations elsewhere in the genome, finally contributing also to the emergence of drug resistance [36].

Table 1: Genes involved in acquired antibiotic resistance in MTB [35].



	Drug	Mode of action	Gene	Gene function	Role
First-line	Isoniazid	Inhibition of mycolic acid biosynthesis and other metabolic processes	*katG*	Catalase-peroxidase	Prodrug activation
			inhA	Enoyl ACP reductase	Drug target
			ndh	NADH dehydrogenase II	Activity modulation
			ahpC	Alkyl hydroperoxidase	Resistance marker
	Rifampicin	Inhibition of transcription	*rpoB*	β-subunit of RNA polymerase	Drug target
	Pyrazinamide	Inhibition of trans-translation	*pncA*	Pyrazinamidase	Prodrug activation
			rspA	S1 ribosomal protein	Drug target
	Ethambutol	Inhibition of arabinogalactan synthesis	*embCAB*	Arabinosyltransferases	Drug target
			embR	*embCAB* transcription regulator	Drug target expression
	Streptomycin	Inhibition of translation	*rpsL*	S12 ribosomal protein	Drug target
			rrs	16S rRNA	Drug target
			gidB	16S rRNA methyltransferase	Target modification
Second-line	Amikacin/Kanamycin	Inhibition of translation	*rrs*	16S rRNA	Drug target
			eis	Acetyltransferase	Drug modification
	Ethionamide	Inhibition of mycolic acid biosynthesis	*ethA*	Flavin monooxygenase	Prodrug activation
			inhA	Enoyl ACP reductase	Drug target
			ethR	*ethA* transcription repressor	Prodrug activator expression
			ndh	NADH dehydrogenase II	Activity modulation
			mshA	Glycosyltransferase	Prodrug activation
	Fluoroquinolones	Inhibition of DNA gyrase	*gyrA*	DNA gyrase subunit A	Drug target
			gyrB	DNA gyrase subunit B	Drug binding

6. MTB Genome stress and DNA Repair

At all stages of its life MTB is confronted by a variety of environmental and endogenous physical and chemical stresses that could produce genotoxic damage, posing a great threat to the integrity of the mycobacterial genome (Figure 10) [37].

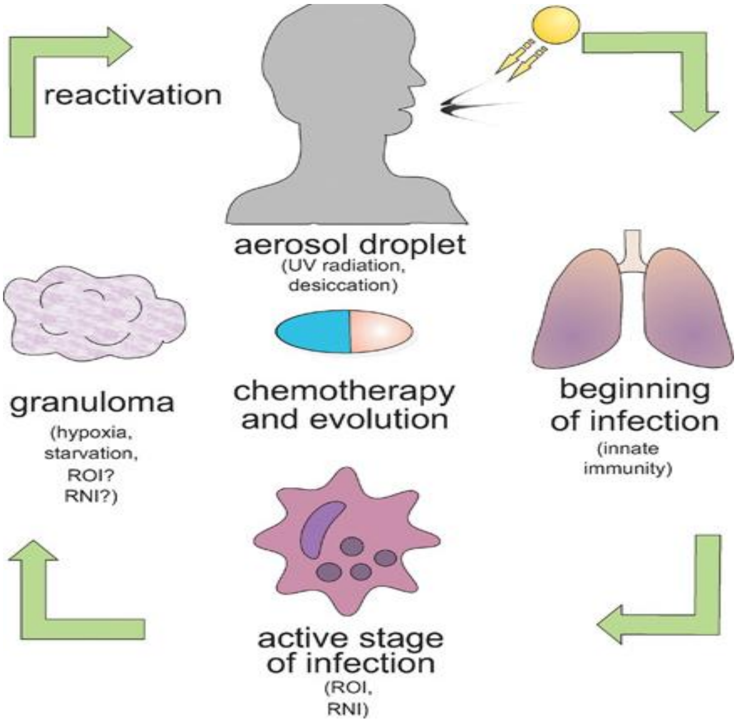

reactivation
aerosol droplet
(UV radiation, desiccation)
chemotherapy and evolution
granuloma
(hypoxia, starvation, ROI? RNI?)
beginning of infection
(innate immunity)
active stage of infection
(ROI, RNI)

Figure 10: MTB life cycle and associated stress [37]

On the opposite end, the occurrence of spontaneous mutations which could augment the fitness of the bacilli under specific selective pressures (*e.g.* mutations conferring drug resistance to anti-tubercular drug), is a prerequisite for adaptability to changing environment/conditions during the complex infection cycle. Therefore, while efficient DNA repair mechanisms must operate in MTB to avoid the deleterious and/or irreversible effects of DNA damage on chromosome stability, a certain degree of damage tolerance must be there to accommodate the evolutionary imperative to adaptation [38].

Extensive MTB genomic analyses, together with a more limited number of gene inactivation studies, have identified components of most of the DNA repair pathways which are active in other species, including multi-enzymatic systems like Nucleotide excision repair (NER), Base Excision Repair (BER) and recombination repair systems -with the notable exception of canonical Mismatch Repair (MMR) components (Figure 11), as well as proteins responsible for Direct Reversal of DNA-damage [37,39,40]. Together these systems allow the MTB bacilli to deal with the great majority of genotoxic and/or mutagenic insults they encounter during their life (Figure 10), that can produce DNA damage in the form of single nucleotide polymorphisms (SNPs, both synonymous and non-synonymous), frameshifts, insertion/deletion, oxidation, deamination, alkylation, and DNA strand breaks.

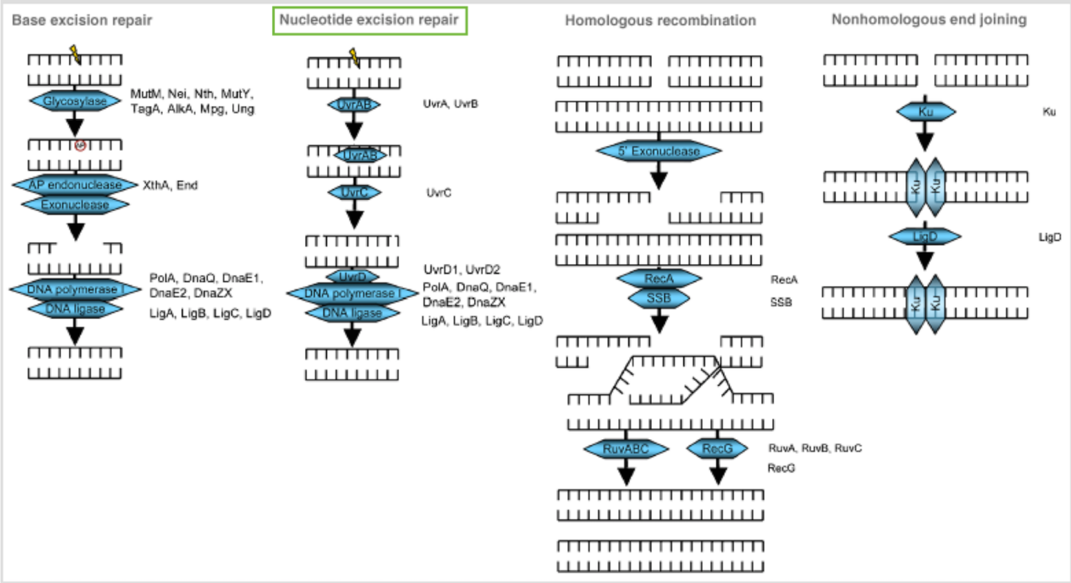

Base excision repair
Glycosylase
MutM, Nei, Nth, MutY, TagA, AlkA, Mpg, Ung
AP endonuclease
Exonuclease
XthA, End
DNA polymerase I
DNA ligase
PolA, DnaQ, DnaE1, DnaE2, DnaZX
LigA, LigB, LigC, LigD
Nucleotide excision repair
UvrAB
UvrA, UvrB
UvrAB
UvrC
UvrC
UvrD
DNA polymerase I
DNA ligase
UvrD1, UvrD2
PolA, DnaQ, DnaE1, DnaE2, DnaZX
LigA, LigB, LigC, LigD
Homologous recombination
5' Exonuclease
RecA
SSB
RecA
SSB
RuvABC
RecG
RuvA, RuvB, RuvC
RecG
Nonhomologous end joining
Ku
Ku
Ku
Ku
LigD
LigD
Ku
Ku

Figure 11: All multi-step repair pathway in MTB [adapted from 42]. The nucleotide Excision Repair pathway will be described in details in the following section.

In the form of contaminated aerosol droplets, at the first stage of infection, MTB is exposed to major genotoxic factors, namely UV irradiation and desiccation, affecting bacilli rate of survival. UV irradiation damages the DNA causing cyclobutane pyrimidine dimers which can be removed by NER; the desiccation of bacilli-infected droplets leads to double-stranded breaks (DSBs) that can be repaired mainly by homologous recombination (HR) and non-homologous end joining (NHEJ) [41]. During the following stage of TB pathogenesis, that is during the MTB persistence inside the infected macrophages inside the granuloma, the bacilli are mainly exposed to host-generated antimicrobial ROS and RNI, starvation and hypoxia, which induce DNA damages that can be counteracted by the concerted actions of all the DNA-repair systems (reviewed in [37,42] and reference therein).

However, our current knowledge of the reciprocal co-ordination of the mechanisms ensuring MTB genome integrity, and of the integration of the DNA repair toolkit encoded functions into an ensemble view of the DNA metabolism during the entire MTB life, is still somehow fragmentary.

7. The Nucleotide Excision Repair (NER) system

Much of the current comprehension of NER comes from the seminal work carried out in *Escherichia coli* [43], by Prof. Aziz Sancar who was honored by the Nobel Prize in Chemistry, 2015.

NER is a major DNA repair mechanism which is highly conserved among all biological systems, although performed by different proteins in prokaryotes and higher eukaryotes. NER importance is reflected by its broad substrate specificity, which includes UV-induced photoproducts, alkylated bases, DNA crosslinks, and bulky anti-cancer drug-DNA adducts.

In humans, three severe diseases are associated with defects in NER: xeroderma pigmentosum, Cockayne´s syndrome and trichothiodystrophy [44].

In Eubacteria and some Archaea, the first steps in NER (namely the damage tracking and recognition, followed by a dual incision at both side of the lesion), are carried out by the coordinated action of the UvrA, UvrB and UvrC proteins (often referred to as the “UvrABC” endonuclease). Based upon biochemical and structural data, the most recent models of the tracking complex from *Geobacillus stearothermophilus*, involve the assembling of an $UvrA_2$:$UvrB_2$ heterotetramer, which possesses the ability to locate sites of potential damage by virtue of the local DNA distortion, in an ATP-dependent UvrA-mediated process [45]. Once the lesion is recognized, the molecular architecture of the UvrA/UvrB/DNA complex undergoes a substantial rearrangement that frees the UvrA component from the ensemble, leaving UvrB stably associated to the modified DNA strand (pre-incision complex). The endonuclease UvrC is then recruited to the site of damage, where it performs two single-strand incisions, four and eight phosphodiester bonds away from the 3’- and 5'-side of the lesion, respectively. The dual incision creates a 12-residue long oligonucleotide containing the lesion. The DNA helicase II (UvrD) is then required to release UvrC and the excised oligonucleotide, while the DNA polymerase I fills the gap (using the complementary intact strand as the template), displacing the UvrB component during the synthesis. Finally the DNA ligase I, encoded by the *LigA* gene, seals the nick thus restoring the integrity of the double helix. This branch of NER is called Global Genome Repair (GGR) (Figure 12) [46].

Interestingly, the stalling of RNA polymerase (RNAP) at damaged DNA sites triggers an important branch of NER, the Transcription Coupled Repair (TCR) (Figure 12), whose activity is required to repair in real time actively transcribed genes. In most bacteria including MTB, TCR is mediated by the Mfd protein [47], a homolog of the eukaryotic transcription-repair coupling factor (TRCF). By binding the N-terminus of the RNA polymerase β-subunit [48], Mfd initiates a cascade of events which ensure the removal of the stalled RNA polymerase from the site of damage and the recruitment of the NER components to perform the necessary repair, thus allowing the effective resuming of transcription [38].

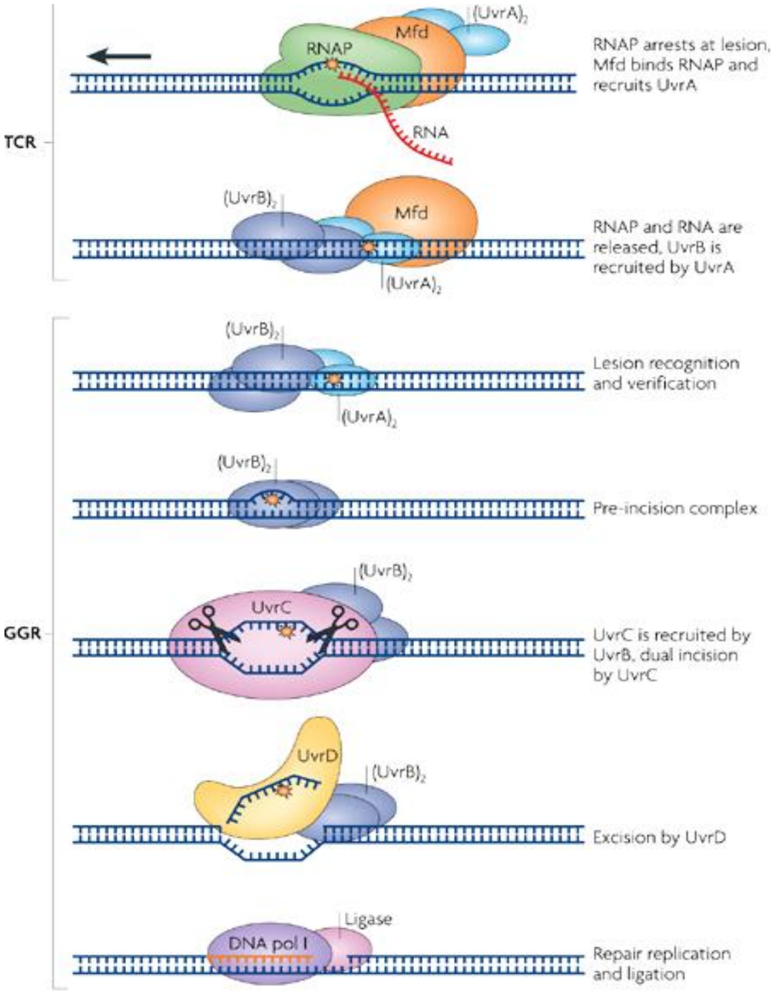

(UvrA)$_2$
Mfd
RNAP
RNA
RNAP arrests at lesion, Mfd binds RNAP and recruits UvrA
TCR
(UvrB)$_2$
Mfd
(UvrA)$_2$
RNAP and RNA are released, UvrB is recruited by UvrA
(UvrB)$_2$
(UvrA)$_2$
Lesion recognition and verification
(UvrB)$_2$
Pre-incision complex
(UvrB)$_2$
UvrC
GGR
UvrC is recruited by UvrB, dual incision by UvrC
UvrD
(UvrB)$_2$
Excision by UvrD
Ligase
DNA pol I
Repair replication and ligation

Figure 12: Steps of NER in eubacteria (TCR and GGR-NER) [49]

All NER components are conserved in MTB, highlighting the importance of this system in the maintenance of mycobacterial genome integrity. In particular concerning the UvrABC endonuclease, mycobacteria are found to up-regulate the expression level of the key players of NER like UvrA and UvrB, on exposure to hydrogen peroxide [49,51]. Moreover, a mutant MTB strain, in which the uvrA gene was inactivated, was found sensitive to various methylating agents. Finally a recombinant version of UvrA possesses the expected DNA-dependant ATPase activity and substrate specificities [52]. Similarly, UvrB was found to be important for bacterial survival using transposon mutagenesis screen. MTB deficient of UvrB also displayed sensitivity to acidified sodium nitrite, a major source of RNI. Further analysis disclosed that UvrB mutants displayed reduced survival rate within mouse model of infection [53]. UvrC gene expression level is enhanced within infected activated macrophages [54]. All the aforesaid observations, signal NER components as a useful drug target.

8. Direct Reversal of alkylated-DNA

Alkylating agents constitutes a wide class of highly reactive compounds of endogenous and exogenous origin that can damage the majority of cellular structures and macromolecules, including the DNA, and therefore being endowed with an extremely dangerous cytotoxic and mutagenic potential (Figure 13).

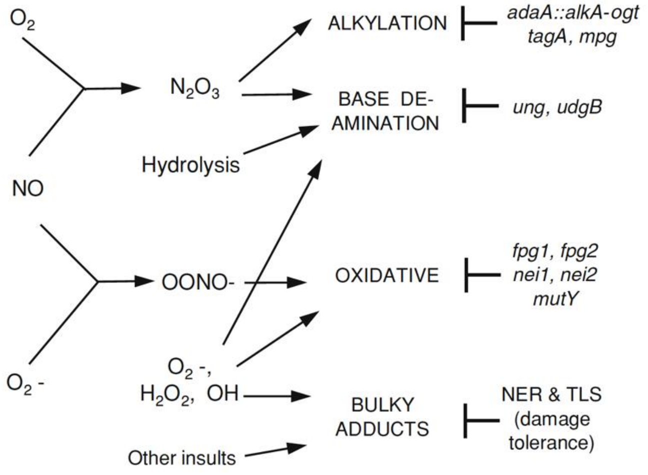

O_2
N_2O_3
ALKYLATION
adaA::alkA-ogt
tagA, mpg
BASE DE-AMINATION
ung, udgB
Hydrolysis
NO
OONO-
OXIDATIVE
fpg1, fpg2
nei1, nei2
mutY
O_2 -
O_2 -,
H_2O_2, OH
BULKY ADDUCTS
NER & TLS (damage tolerance)
Other insults

Figure 13: The type of DNA damage expected to be sustained in MTB and the genes implicated in damage repair and reversal [38]

In bacterial species, including MTB, DNA alkylation can be counteracted either by exploiting multi-enzymatic systems (*e.g.* NER and BER) or by mounting an adaptive response, the Ada response [55,40].

In *E. coli* the Ada operon consists of four genes: *ada* (depicted as *adaA*-*adaB* in Figure 14 A), *alkA*, *alkB* and *aidB*. The Ada protein is a methyltransferase that also acts as the positive regulator of the operon; the AlkA protein is a 3-methyl-adenine (3mA) DNA glycosylase; the AlkB protein is a Fe(II)-dependent dioxygenase that repairs lesions such as 1-methyladenine (1-mA) and 3methylcytosine (1-mC), by catalyzing oxidative demethylation; finally AidB is a DNA-binding protein predicted to catalyze direct repair of alkylated DNA[56].

More in details, *E. coli* Ada is composed by two domains, the N-terminal AdaA and the C-terminal AdaB domain. The 20 kDa AdaA domain transfers the methyl group of innocuous methylphosphotriesters in DNA onto the cysteine residue of its active site; this methylation converts the protein into a transcriptional activator with specific DNA binding affinity to genes having the ada operator sequence in their promoters, including the *ada* gene itself. The 19 kDa AdaB repairs the O^6-methylguanine (O^6-meG) and O^4-methylthymine (O^4-meT) bases by transferring the alkyl group to a strictly conserved cysteine of its active site, in an irreversible suicidal reaction [56,57].

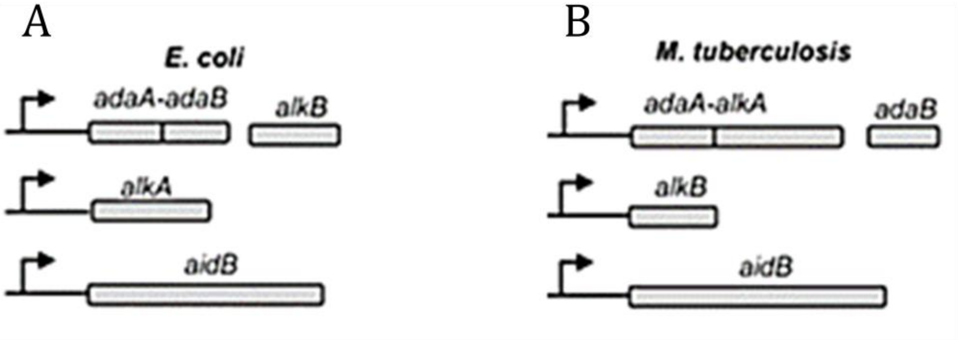


Figure 14: Comparison of Ada operon in *E.coli* and MTB [adapted from 40]

Although adaptive response is conserved among many bacterial species, the domains of Ada (namely AdaA and AdaB), AlkA and AlkB proteins exist in different combinations in different prokaryotes. In particular, MTB shows a gene fusion of *adaA* with *alkA* (*adaA-alkA*), and an independent *adaB* gene (Tuberculist code: Rv1316c; annotated in several databases as *ogt*, to put the accent onto the biochemical function of the corresponding protein, the O^6-methylguanine methyltransferase (OGT) (Figure 14 B).

Transcription of the *adaA-alkA* and *adaB* genes are induced in MTB when treated with methylating agent like N-methyl-N'-nitro-N-nitrosoguanidine, suggesting that MTB mounts an inducible response to methylating agents. Activity assays also disclose that both MTB AlkA and AdaB/OGT possess methyltransferase activity but not DNA glycosylase activity. Furthermore, the MTB AdaB/OGT expression in an *E. coli* Ada-OGT double-mutant, suppresses the hyper-mutator phenotype displayed upon alkylating agents treatment, while an analogous trans-complementation experiment using the MTB *adaA-alkA* gene does not produce the same effect. Collectively, these data indicate that MTB AdaB/OGT counteracts the mutagenic effect of O^6-mG while MTB AdaA-AlkA removes methyl groups from innocuous methylphosphotriesters [40]. As a matter of clarity through the following of the text the MTB AdaB/OGT protein will be indicated as *Mt*OGT.

Structural and biochemical studies of MTB O^6-methylguanine-DNA methyltransferase (MtOGT)

*Mt*OGT, invariably to other micro-organisms, acts through a suicidal mechanism by performing the irreversible transfer of O^6-alkyl group from the modified guanine to the strictly conserved cysteine residue within the protein active site (Cys^{126}) leading to its permanent inactivation (figure 15) [58].

Interestingly, a number of geographically widely distributed MTB strains (like MTB W-Beijing strains and multi-drug resistant isolates) have been identified which carry non-synonymous SNPs in their *ogt* genes [59,60]. These point-mutations result in amino acid substitution at position 15 (*Mt*OGT-T15S) or position 37 (*Mt*OGT-R37L) of *Mt*OGT, mapping at the poorly structurally conserved N-terminal domain of the protein.

Several researchers proposed that a defective OGT could lead to an increased mutation frequency, conferring to the corresponding strain a better adaptability to the host [54,60].

Recently, the high resolution crystal structure of *Mt*OGT and its mutated variant *Mt*OGT-R37L, together with the biochemical analysis of highly homogeneous recombinant versions of *Mt*OGT, *Mt*OGT-T15S and *Mt*OGT-R37L were reported by our laboratory [58].

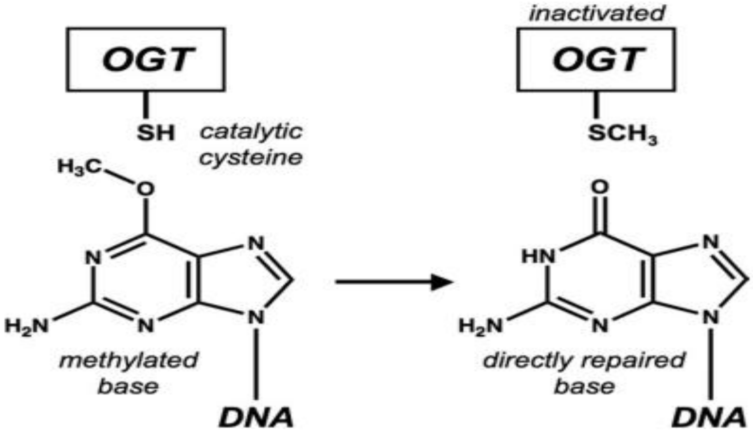


Figure 15: Alkyltransfer reaction from O^6-methylguanine to OGT active site cysteine residue [58]

Similar to structures of other prokaryotic [61-64] OGTs as well as of human AGT [65-68] in different liganded states, both *Mt*OGT and *Mt*OGT-R37L fold into a roughly globular molecular architecture built up by two domains connected by a long loop and ending in a 10-residue-long tail. The *Mt*OGT N-terminal domain consists of an anti-parallel three-stranded β-sheet and connecting loops sandwiched between a mainly randomly coiled region, containing a single helical turn at its middle, on one side and a structurally conserved α-helix on the opposite one.

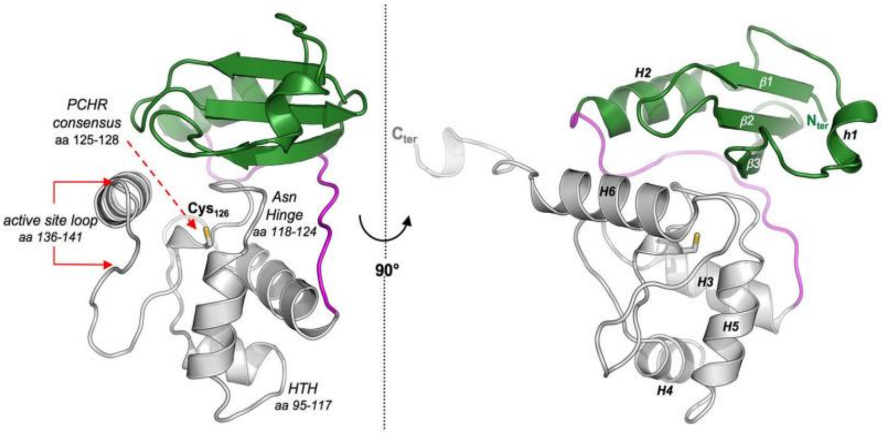


Figure 16: Crystal structure of *Mt*OGT. Shown are cartoon representations of the *Mt*OGT structure, as observed from two different points of view. In both images, the N-terminal domain, the C-terminal domain, and the connecting loop are colored in green, gray, and magenta, respectively. The catalytic cysteine residue (Cys^{126}) is invariably drawn as sticks. Functional element labeling and secondary structure element numbering appear in the image at the left and at the right, respectively. aa, amino acids [58].

The C-terminal domain adopts the typical all-α-fold and houses the highly conserved functional elements that previous biochemical and mutational studies demonstrated to be required to perform efficient catalysis: (i) the helix-turn-helix (HTH) motif, which is responsible for DNA binding at its minor groove and bears Arg^{109} acting as a temporary substitute for the modified base upon its flipping out from the regular base stacking; (ii) the “Asn hinge” building up one wall of the deep ligand binding cavity that accepts the modified base for repairing; (iii) the strictly conserved PCHR consensus motif surrounding the catalytic cysteine (Cys^{126}); (iv) the active-site loop that participates in the correct positioning of the alkylated base inside the ligand binding pocket; and (v) the structurally conserved H6 helix that, by shielding the ligand binding cavity on the opposite side of the Asn hinge, contributes essential residues for completing the modified-base bonding network (Lys^{146}) (Figure 16) [64-67,69 ,70].

Biochemical studies of wild-type *Mt*OGT and its *Mt*OGT-T15S and *Mt*OGT-R37L versions, revealed that *Mt*OGT-R37L displayed a ten-fold reduced affinity towards methylated double-stranded DNA (dsDNA) than the wild-type protein, although neither mutation affects the intrinsic alkyl-guanine-transferase reaction rate [58].

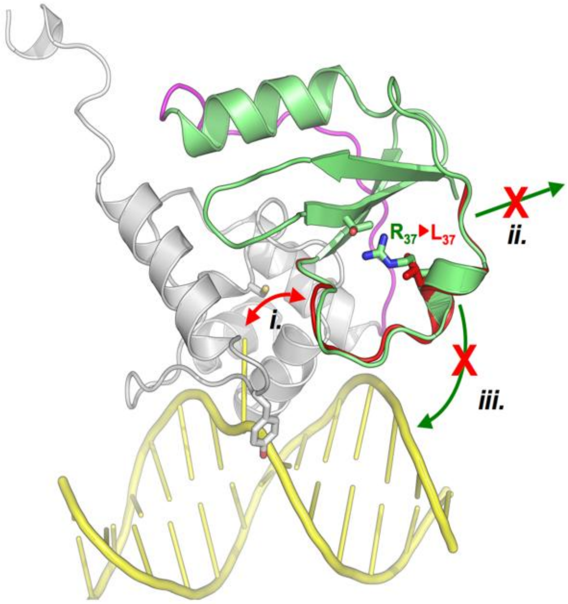


Figure 17: Schematic representation of the possible functional consequences of the replacement of Arg^{37} with Leu in *Mt*OGT. (i) Direct interference in the dealkylation reaction; (ii) reduced capability of establishing protein-protein interactions and tight monomer packing during cooperative DNA binding; (iii) direct participation of Arg^{37} in DNA binding. The N-terminal domain, the domain-connecting loop, and the C-terminal domain of the wild-type protein are colored green, magenta, and white, respectively. The *Mt*OGTR37L structure is uniformly rendered in red; the modeled dsDNA appears in yellow [58].

Interestingly, Arg^{37} seems to play a central role in coordinating a peculiar network of bonds established between the core β-sheets and the facing random coil of the N-terminal domain. This observation prompted us to propose three not mutually exclusive scenarios possibly explaining how the mutation of Arg^{37}, which maps far away from both the Cys^{126} and the currently identified DNA binding motifs could produce an impact on protein function. As shown in Figure 17, the R37L mutation could (i) induce conformational changes of the N-terminal random-coiled region that can affect catalysis, (ii) alter the assembly of protein complexes at the damaged site during cooperative DNA binding, and (iii) impair an unexpected participation of the Arg^{37}-bearing random coil to direct DNA binding [58].

Our most recent work [70] allowed us to discriminate between these hypotheses, since it led to the resolution of the crystal structure of *Mt*OGT in complex with DNA that, together with the characterization of further structure-based designed mutants of *Mt*OGT, adds novelty to the current description of this interesting class of proteins, as it will be described in the "Chapter 3" of my Thesis.

REFERENCES

1 Data extracted from WHO Global tuberculosis report 2015

2 Lienhardt, C., Vernon, A. and Raviglione, M.C., (2010). New drugs and new regimens for the treatment of tuberculosis: review of the drug development pipeline and implications for national programmes. *Current opinion in pulmonary medicine*, 16, 186-193.

3 Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G. and Yeboah-Manu, D., (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature genetics*, **45**, 1176-1182.

4 Sakamoto, K. (2012). The pathology of Mycobacterium tuberculosis infection. *Veterinary Pathology Online*,**49**, 423-439.

5 Data extracted from http://www.textbookofbacteriology.net/

6 Hett, E. C., & Rubin, E. J. (2008). Bacterial growth and cell division: a mycobacterial perspective. *Microbiology and Molecular Biology Reviews*, **72**, 126-156.

7 Abdallah, A.M., van Pittius, N.C.G., Champion, P.A.D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M., Appelmelk, B.J. and Bitter, W., 2007. Type VII secretion—mycobacteria show the way. *Nature reviews microbiology*, **5**, 883-891.

8 Kieser, K. J., & Rubin, E. J. (2014). How sisters grow apart: mycobacterial growth and division. *Nature Reviews Microbiology*, **12**, 550-562.

9 Glickman M.S., Cox J.S., Jacobs W.R. Jr. (2000). A novel mycolic acids cyclopropane synthetase is required for coding, persistence and virulence of Mycobacterium tuberculosis. *Molelular Cell*, **5**, 717-727.

10 Data extracted from www.studyblue.com

11 Data extracted from http://lipidlibrary.aocs.or

12 Peccia, J., & Hernandez, M. (2004). UV-induced inactivation rates for airborne Mycobacterium bovis BCG. *Journal of occupational and environmental hygiene*, **1**, 430-435.

13 Peccia, J. L., Werth, H. M., & Hernandez, M. T. (2000). Effects of relative humidity on the UV-induced inactivation of bacterial bioaerosols. *Journal of aerosol science*, **31**, 959-960.

14 Centre of Disease and Prevention - TB Pathogenesis (http://www.cdc.gov/tb/)

15 Data extracted from Dallas County Public Health- Epi Report-August 2015 (https://storify.com)

16 Data extracted from (http://www.oxfordimmunotec.com/)

17 Pieters, J. (2001). Entry and survival of pathogenic mycobacteria in macrophages. *Microbes and infection*, **3**, 249-255.

18 Gengenbacher, M., & Kaufmann, S. H. (2012). Mycobacterium tuberculosis: success through dormancy. *FEMS microbiology reviews*, **36**, 514-532.

19 Flannagan, R. S., Cosío, G., & Grinstein, S. (2009). Antimicrobial mechanisms of phagocytes and bacterial evasion strategies. *Nature Reviews Microbiology*, **7**, 355-366.

20 Pieters, J. (2001). Evasion of host cell defense mechanisms by pathogenic bacteria. *Current opinion in immunology*, **13**, 37-44.

21 Meena, L. S. (2010). Survival mechanisms of pathogenic Mycobacterium tuberculosis H37Rv. *FEBS Journal*, **277**, 2416-2427.

22 Adams, L. B., Dinauer, M. C., Morgenstern, D. E., & Krahenbuhl, J. L. (1997). Comparison of the roles of reactive oxygen and nitrogen intermediates in the host response to Mycobacterium tuberculosis using transgenic mice. *Tubercle and Lung Disease*, **78**, 237-246.

23 Russell, D. G. (2007). Who puts the tubercle in tuberculosis? *Nature Reviews Microbiology*, **5**, 39-47.

24 Saunders, B. M., & Britton, W. J. (2007). Life and death in the granuloma: immunopathology of tuberculosis. *Immunology and cell biology*, **85**, 103-111.

25 Flandrois, J. P., Lina, G., & Dumitrescu, O. (2014). MUBII-TB-DB: a database of mutations associated with antibiotic resistance in Mycobacterium tuberculosis. *BMC bioinformatics*, **15**, 107.

26 Boehme, C. C., Saacks, S., & O'Brien, R. J. (2013). The changing landscape of diagnostic services for tuberculosis. *Seminars in respiratory and critical care medicine*, **34**, 17-31.

27 Senol, G. (2013). Laboratory diagnosis of tuberculosis-latest diagnostic tools. *training*, **6**, 9-10.

28 Du Toit, L. C., Pillay, V., & Danckwerts, M. P. (2006). Tuberculosis chemotherapy: current drug delivery approaches. *Respiratory Research*, **7**, 118.

29 Joshi, J. M. (2011). Tuberculosis chemotherapy in the 21st century: Back to the basics. *Lung India: official organ of Indian Chest Society*, *28*, 193.

30 Long, R. (2000). Drug-resistant tuberculosis. *Canadian Medical Association Journal*, **163**, 425-428.

31 Ormerod, L. P. (2005). Multidrug-resistant tuberculosis (MDR-TB): epidemiology, prevention and treatment. *British medical bulletin*, **73**, 17-24.

32 Dorman, S. E., & Chaisson, R. E. (2007). From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nature medicine*, **13**, 295-298.

33 Udwadia, Z. F. (2012). MDR, XDR, TDR tuberculosis: ominous progression. *Thorax*, **67**, 286-288.

34 Koch, A., Mizrahi, V., & Warner, D. F. (2014). The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin?.*Emerging Microbes & Infections*, **3**, 17.

35 Smith, T., Wolff, K. A., & Nguyen, L. (2013). Molecular biology of drug resistance in Mycobacterium tuberculosis. In *Pathogenesis of Mycobacterium*

tuberculosis and its Interaction with the Host Organism (pp. 53-80). Springer Berlin Heidelberg.

36 Warner, D. F., & Mizrahi, V. (2013). Complex genetics of drug resistance in Mycobacterium tuberculosis. *Nature genetics*, **45**, 1107-1108.

37 Gorna, A., Bowater, R., & Dziadek, J. (2010). DNA repair systems and the pathogenesis of Mycobacterium tuberculosis: varying activities at different stages of infection. *Clinical Science*, **119**, 187-202.

38 Warner, D. F., Tønjum, T., & Mizrahi, V. (2013). DNA metabolism in mycobacterial pathogenesis. In *Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism* (pp. 27-51). Springer Berlin Heidelberg.

39 Mizrahi, V., & Andersen, S. J. (1998). DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Molecular microbiology*, **29**, 1331-1339.

40 Yang M., Aamodt R.M., 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.

41 Korycka-Machala, M., Brzostek, A., Rozalska, S., Rumijowska-Galewicz, A., Dziedzic, R., Bowater, R., & Dziadek, J. (2006). Distinct DNA repair pathways involving RecA and nonhomologous end joining in Mycobacterium smegmatis. *FEMS microbiology letters*, **258**, 83-91.

42 Dos Vultos, T., Mestre, O., Tonjum, T., & Gicquel, B. (2009). DNA repair in Mycobacterium tuberculosis revisited. *FEMS microbiology reviews*, **33**, 471-487.

43 Petit, C., & Sancar, A. (1999). Nucleotide excision repair: from E. coli to man. *Biochimie*, **81**, 15-25.

44 Rupp, W. D., Sancar, A., & Sancar, G. B. (1982). Properties and regulation of the UVRABC endonuclease. *Biochimie*, **64**, 595-598.

45 Pakotiprapha, D., Samuels, M., Shen, K., Hu, J. H., & Jeruzalmi, D. (2012). Structure and mechanism of the UvrA–UvrB DNA damage sensor. *Nature structural & molecular biology*, **19**, 291-298.

46 Truglio, J. J., Croteau, D. L., Van Houten, B., & Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical reviews*, **106**, 233-252.

47 Xu, S. Y., Prabha, S., Rao, D. N., & Nagaraja, V. (2011). Distinct properties of hexameric but functionally conserved Mycobacterium tuberculosis transcription-repair coupling factor. *PLoS One*, **6**, Article-ID.

48 Westblade, L. F., Campbell, E. A., Pukhrambam, C., Padovan, J. C., Nickels, B. E., Lamour, V., & Darst, S. A. (2010). Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic acids research*, gkq692.

49 Hanawalt, P. C., & Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nature reviews Molecular cell biology*,**9**, 958-970.

50 Graham, J. E., & Clark-Curtiss, J. E. (1999). Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proceedings of the National Academy of Sciences*, **96**, 11554-11559.

51 Rachman, H., Strong, M., Ulrichs, T., Grode, L., Schuchhardt, J., Mollenkopf, H., Kosmiadi, G.A., Eisenberg, D. and Kaufmann, S.H., (2006). Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis. *Infection and immunity*, **74**, 1233-1242.

52 Rossi, F., Khanduja, J. S., Bortoluzzi, A., Houghton, J., Sander, P., Güthlein, C. & Rizzi, M. (2011). The biological and structural characterization of

Mycobacterium tuberculosis UvrA provides novel insights into its mechanism of action. *Nucleic acids research*, **39**, 7316-7328.

53 Darwin, K. H., Ehrt, S., Gutierrez-Ramos, J. C., Weich, N., & Nathan, C. F. (2003). The proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide. *Science*, **302**, 1963-1966.

54 Kurthkoti, K., Kumar, P., Jain, R., & Varshney, U. (2008). Important role of the nucleotide excision repair pathway in Mycobacterium smegmatis in conferring protection against commonly encountered DNA-damaging agents. *Microbiology*, 154, 2776-2785.

55 Shrivastav N., Li D., Essigmann J.M. (2010). Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis*, **3**, 59 – 70.

56 Saget B.M., Walker G.C. (1994). The Ada protein acts as both a positive and a negative modulator of Escherichia coli's response to methylating agents. *Proceeding of national academy of Sciences,U.S.A.*,**91**, 9730–9734.

57 Sedgwick B. and Lindahl T. (2002). Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene*, 21,8886-94

58 Miggiano, R., 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. *Journal of bacteriology*, **195**, 2728-2736.

59 Olano J., Lopez B., Reyes A., Lemos MP., Correa N., Del Portillo P., Barrera L., Robledo J., Ritacco V., Zambrano MM. (2007). Mutations in DNA repair genes are associated with the Haarlem lineage of Mycobacterium tuberculosis independently of their antibiotic resistance. *Tuberculosis (Edinb.)*,**87**,502–508

60 Rad, M.E., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., Kreiswirth, B., Blazquez, J., Jouan, M., van Soolingen, D. and Gicquel, B.,

(2003). Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family. *Emerging infectious diseases*, **9**, 838.

61 Hashimoto H, Inoue T, Nishioka M, Fujiwara S, Takagi M, Imanaka T, Kai Y. (1999). Hyperthermostable protein structure maintained by intra and inter-helix ion-pairs in archaeal O6-methylguanine-DNA methyltransferase. *Journal of Molecular Biology* **292**,707–716. Roberts A, Pelton JG, Wemmer DE. 2006.

62 Structural studies of MJ1529, an O6-methylguanine-DNA methyltransferase. Magn. Reson. Chem. 44(Spec No):S71–S82. doi:10.1002/mrc.1823

63 Perugino, G., Miggiano, R., Serpe, M., Vettone, A., Valenti, A., Lahiri, S., Rossi, F., Rossi, M., Rizzi, M. and Ciaramella, M., (2015). Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic acids research*, p.gkv774.

64 Moore MH, Gulbis JM, Dodson EJ, Demple B, Moody PC. (1994). Crystal structure of a suicidal DNA repair protein: the Ada O6-methylguanineDNA methyltransferase from E. coli. *EMBO J.* **13**,1495–1501

65 Wibley JE, Pegg AE, Moody PC. (2000). Crystal structure of the human O(6)-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res.* **28**:393–401.

66 Daniels DS, Woo TT, Luu KX, Noll DM, Clarke ND, Pegg AE, Tainer JA. (2004). DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature Structural Molecular Biology*. **11**,714 –720.

67 Duguid EM, Rice PA, He C. (2005). The structure of the human AGT protein bound to DNA and its implications for damage detection. *Journal of Molecular Biology* 350:657–666.

68 Xu-Welliver M, Pegg AE. (2000). Point mutations at multiple sites including highly conserved amino acids maintain activity, but render O6- alkylguanine-DNA alkyltransferase insensitive to O6-benzylguanine. *Biochemical Journal* **347**,519 –526

69 Daniels DS, Mol CD, Arvai AS, Kanugula S, Pegg AE, Tainer JA. (2000). Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. EMBO J. **19**,1719 –1730.

70 Miggiano, R., Perugino, G., Ciaramella, M., Serpe, M., Rejman, D., Páv, O., Pohl, R., Garavaglia, S., Lahiri, S., Rizzi, M. and Rossi, F., (2016). Crystal structure of Mycobacterium tuberculosis O6-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA. *Biochemical Journal*, **473**, 123-133

Outline of the Thesis

A long history of co-evolution with the mankind makes MTB capable to deal with a plethora of physical and chemical stresses, and surviving and replicating within one of the most inhospitable cell of the body: the macrophage.

Despite this remarkable adaptation to the human host, during its entire life MTB is exposed to DNA-damaging assaults, whose consequences on genome stability could compromise the establishing of the primary infection, the persistence of the bacilli inside the macrophages - mainly in a dormant state that contributes to escape the host’s immune system surveillance-, as well as the correct execution of the reactivation program at resuscitation, that promotes the transition from the latent and asymptomatic phase to the active and contagious stage of the TB disease.

On the other end, the occurrence of new mutations inside the mycobacterial genomic DNA is considered the driving force to acquire new phenotypic traits, including specific drug resistance, for which horizontal gene transfer has rarely been documented. Therefore a sub-optimal DNA repair could result in the fixation of mutations that are compatible with the MTB life and that confer to the bacillus a superior adaptive capability under a positive selective pressure. Two main observations support this argument; first, inside the MTB genome, the level of sequence polymorphism of genes encoding DNA replication, recombination and repair functions is remarkably higher with respect to that of house-keeping genes. Secondly, a number of geographically widely spread MTB strains carry peculiar single-nucleotide polymorphisms in several DNA repair genes, leading to the hypothesis that the associated phenotypes, in principle more prone to cumulate gene mutations, could have represented a selective advantage to the corresponding strains during MTB evolution.

Although the key role of DNA repair in MTB biology is well established, the information about i) the actual contribution of each DNA repair pathway to the overall maintenance of the MTB genome integrity, ii) the cross-talk between the DNA repair systems and other machineries involved in MTB DNA metabolism, and iii) the reciprocal co-ordination of these activities during TB pathogenesis, is still somehow fragmentary.

Consequently, the study of DNA repair proteins in MTB is crucial not only to understand their functioning at a molecular level, but also to evaluate their impact upon the TB infection process.

This was the rationale of my PhD research activity, aimed at a better biochemical and structural description of proteins involved in alkylated-DNA repair in MTB, along with their interaction studies. The majority of living organisms deploy several strategies to defend their genome from the genotoxic and pro-mutagenic potential associated to DNA alkylation, including i) the direct reversal of the damage, through the one-step, irreversible and suicidal action of an alkylated DNA-protein alkyltransferase, such as OGT (also known as MGMT or AGT in other species including humans), and ii) the multistep excision of a short oligonucleotide containing the lesion, followed by DNA new synthesis (NER system).

My Thesis focuses on the *Mt*OGT protein and on its possible interaction with the *Mt*UvrA component of the lesion-sensing complex of the NER.

As detailed in Chapter 3, I participated to the determination of the crystal structure of *Mt*OGT in complex with a N^{1}-0^{6}-ethanoxanthosine (E1X)-modified double-stranded DNA molecule, which acts as a mechanistic inhibitor of the protein by irreversibly binding the C-terminal catalytic domain. The structural analysis of such a complex, together with the biochemical characterization of a restricted panel of *Mt*OGT mutated variants, allowed us to describe peculiar features of the mycobacterial protein compared to the pre-existing information on members this protein family. Indeed, in our structure, we directly observed for the first time the mode of assembling of three *Mt*OGT monomers to the same E1X-dsDNA molecule. This arrangement discloses details of the protein-protein and protein-DNA interactions sustaining the cooperative DNA-binding mechanism of *Mt*OGT. Another element of novelty is represented by the capability of the two protein monomers that are not engaged in binding the E1X base, to host an unmodified adenine in their active site, contributing further information to build a model of the alkylation damage detection process. Finally, we observed that discrete regions of both the N- and the C-terminal domains of *Mt*OGT display a high level of structural plasticity, a specific feature of the mycobacterial protein. Interestingly, several frequently isolated MTB strains contain a non-synonymous SNP in their OGT-encoding gene that leads to the substitution of the Arginine residue at position 37 (Arg^{37}) by a Leucine. Former studies at our research Unit revealed that a recombinant version of this mutated *Mt*OGT variant (*Mt*OGT-R37L) was impaired in catalysis, showing a 10-folds lower affinity for methylated DNA (see the *Introduction*). The results of the present work suggest that Arg^{37} could impose some restrictions to the structural flexibility of the N-terminal domain random coiled region, thus representing a plausible determinant of the optimal protein assembling at the alkylated site during DNA recognition and repair.

In the "*Unpublished Results*" section (Chapter 4) of my Thesis, I reported the still ongoing investigations aimed at characterizing a possible cross-talk between OGT-mediated alkylated-DNA repair and the NER system, pivoting on an unprecedented interaction between *Mt*OGT and *Mt*UvrA. The formation of *Mt*UvrA::*Mt*OGT complex was disclosed in the context of high throughput screenings of protein-protein interactions in mycobacteria (liquid chromatography-based affinity purification coupled with high resolution mass spectrometry), and found a preliminary confirmation in Surface Plasmon Resonance (SPR)-based interaction studies. To isolate and characterize the *Mt*UvrA::*Mt*OGT complex, I performed size-exclusion chromatography-based co-fractionation experiments, using the recombinant pure proteins, including the well-known protein partner of *Mt*UvrA, *i.e.* *Mt*UvrB. These analyses will aid us to better define the experimental setting to adopt in the future to further describe macromolecular interactions sustaining alkylated DNA repair in MTB.

Chapter 3

Crystal structure of *Mycobacterium tuberculosis* O6-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA

Riccardo Miggiano*, Giuseppe Perugino†, Maria Ciaramella†, Mario Serpe†, Dominik Rejman‡, Ondˇrej Pav´‡, Radek Pohl‡, Silvia Garavaglia*, Samarpita Lahiri*, Menico Rizzi*[1] and Franca Rossi*[1]

*DSF – Dipartimento di Scienze del Farmaco, University of Piemonte Orientale, 28100 Novara, Italy

†Institute of Biosciences and Bioresources, IBBR-CNR, 80125 Naples, Italy

‡IOCB-Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences v.v.i., 166 10 Prague 6, Czech Republic

Published in *Biochemical Journal*, **473**, 123-133 (2016)

Abstract

Mycobacterium tuberculosis O^6-methylguanine-DNA methyl-transferase (*Mt*OGT) contributes to protect the bacterial GC-rich genome against the pro-mutagenic potential of O^6-methylated guanine in DNA. Several strains of *M. tuberculosis* found worldwide encode a point-mutated O^6-methylguanine-DNA methyltransferase (OGT) variant (*Mt*OGT-R37L), which displays an arginine-to-leucine substitution at position 37 of the poorly functionally characterized N-terminal domain of the protein. Although the impact of this mutation on the *Mt*OGT activity has not yet been proved *in vivo*, we previously demonstrated that a recombinant *Mt*OGT-R37L variant performs a suboptimal alkylated-DNA repair *in vitro*, suggesting a direct role for the Arg^{37}-bearing region in catalysis. The herein reported crystal structure of *Mt*OGT complexed with modified DNA reveals details of the protein–protein and protein–DNA interactions occurring during alkylated-DNA binding, and the protein capability also to host unmodified bases inside the active site, in a fully extrahelical conformation. Our data provide the first experimental picture at the atomic level of a possible mode of assembling three adjacent *Mt*OGT monomers on the same monoalkylated dsDNA molecule, and disclose the conformational flexibility of discrete regions of *Mt*OGT, including the Arg^{37}-bearing random coil. This peculiar structural plasticity of *Mt*OGT could be instrumental to proper protein clustering at damaged DNA sites, as well as to protein–DNA complexes disassembling on repair.

Keywords: cooperativity, crystal structure, DNA-binding protein, DNA repair, *Mycobacterium tuberculosis*, O^6-methylguanine-DNA methyltransferase

INTRODUCTION

Mycobacterium tuberculosis displays a remarkable genetic stability despite the continuous exposure to potentially promutagenic and genotoxic stresses that could compromise the pathogen’s capability of establishing a latent infection in the human host and exiting from the dormant state at reactivation [1,2]. Generated by the *M. tuberculosis*-infected macrophages as part of the antimicrobial response, highly reactive oxygen and nitrogen intermediates can directly damage several mycobacterial targets, including DNA, and can trigger the endogenous synthesis of potent DNA-alkylating metabolites [3-5].

As observed in other organisms, *M. tuberculosis* repairs alkylated bases in DNA either by using multi-enzymatic systems or through the action of single proteins [6,7], such as the O^6-methylguanine-DNA methyltransferase (OGT, EC: 2.1.1.63). Genes encoding O^6-alkylguanine-DNA alkyltransferases (alternatively abbreviated as AGT or MGMT) have been identified in the genome of the most diverse organisms, and numerous studies aimed at the functional characterization of members of this protein family have been published (reviewed by Pegg [8,9]. These analyses reveal that alkyltransferases preferentially repair O^6-alkylated guanine in DNA, invariably performing the stoichiometric transfer of the alkyl group from the modified base to a conserved cysteine residue buried in their active site [10-12]. Much less is known about the cellular fate of the inactivated protein resulting from DNA repair, although it has been proposed that the irreversible alkylation of the catalytic cysteine could induce conformational changes, which might increase protein instability *in vitro* and its propensity to degradation *in vivo* [13,14].

The *M. tuberculosis* OGT (*Mt*OGT)-encoding gene is part of the mycobacterial adaptive response operon [15], and evidence was obtained pointing at *Mt*OGT as a main player in protecting the *M. tuberculosis* chromosome against the risk of G:C-to-A:T transition mutations associated with O^6-alkylated guanine in DNA [3,6,16,17]. It is interesting that a number of geographically widely distributed *M. tuberculosis* strains and multidrug-resistant isolates are characterized by point-mutated OGTs carrying an amino acid substitution at position 15 or 37 of the N-terminal domain (T15S and R37L), and it has been proposed that a defective alkylated-DNA repair could have played a role in tuning the balance between genome stability preservation and adaptability to the host during the evolutionary history of the pathogen [18-20]. Although the functional consequences of the presence of these *Mt*OGT variants on the biology of the corresponding strains have not yet been determined, we showed that a recombinant *Mt*OGT-R37L is significantly impaired in alkylated-DNA damage reversal *in vitro*, displaying a 10-fold lower affinity for methylated dsDNA ($dsDNA^{met}$) with respect to the wild-type protein [21].

Parallel X-ray crystallography studies of the ligand-free form of *Mt*OGT showed that Arg^{37} belongs to a mainly random coiled region (residues 28–47) of the N-terminal domain, the sequence and overall structure of which significantly vary among OGTs from different species [21]. Moreover, Arg^{37} maps away from the protein active site and the DNA-binding motifs so far identified, based on structural analyses of the human orthologue O^6-alkylguanine-DNA alkyltransferase (hAGT) [22-25]. Finally, the structural comparison of *Mt*OGT and *Mt*OGT-R37L showed that the Arg^{37}-to-Leu substitution produces a negligible impact on the protein conformation in the absence of ligands [21], underlining the need to obtain the structure of *Mt*OGT in alternative substrate-bound states in order to elucidate the molecular determinants of the observed suboptimal catalysis performed by the *Mt*OGT-R37L variant.

In the present study we describe the crystal structure of wild-type *Mt*OGT complexed with a modified dsDNA molecule, N^1-O^6-ethano-2′–deoxyxanthosine-containing dsDNA (*Mt*OGT::E1X-dsDNA), which reveals similar as well as peculiar traits when compared with the equivalent structure of human AGT [24]. Indeed, in the *Mt*OGT::E1X-dsDNA structure, we directly observed, for the first time, a possible mode of assembling three adjacent protein chains on to the same damaged DNA duplex. This allowed us to gain insight into the architecture of protein–DNA complexes that could explain the cooperative DNA-binding mechanism of *Mt*OGT, which was suggested by EMSA-based analyses [21] and the present study. It is interesting that, in the *Mt*OGT::E1X-dsDNA structure, the protein monomers that are not engaged in binding the modified base are equally observed to host an unmodified adenine in their active site, contributing further information to the vision of a mechanistic model of the alkylation damage detection process. Finally, discrete regions of both the N- and the C-terminal domains of *Mt*OGT display a high level of structural plasticity, a specific *Mt*OGT feature that could be required for proper protein assembly at the alkylated site during DNA repair, as also suggested by the biochemical and structural characterization of additional *Mt*OGT mutated variants.

EXPERIMENTAL

Chemicals

All reagents were obtained from Sigma-Aldrich unless otherwise specified.

Expression and purification of point-mutated *Mt*OGT variants

The pET-*Mt*OGT construct coding for the wild type MTB *O*^{6}-methylguanine methyltransferase (*orf*: Rv1316c) [21] was used as the DNA template in PCR-based site-directed mutagenesis experiments, using the QuikChange II site-directed mutagenesis kit reagents (Stratagene) and the primer pairs R37K5´/R37K3´, or R37E5´/R37E3´, or Y139F5´/Y139F3´ (Supplementary Table 1). The region encoding the corresponding point-mutated *Mt*OGT variant in each resulting expression construct (namely pET-*Mt*OGT-R37K, pET-*Mt*OGT-R37E and pET-*Mt*OGT-Y139F) was verified by sequencing (Eurofins MWG Operon). The expression and purification of the three new point-mutated versions of *Mt*OGT used in crystallization trials and activity assays were achieved by adopting the same procedure used for the wild-type protein [21]. All proteins are monomeric and display similar stability in solution (not shown).

Synthesis of the E1X-containing oligonucleotide

The E1X monomer [26] was prepared adopting the reaction scheme illustrated in Supplementary Figure S1. The full procedure used for the synthesis of the E1X-modified oligonucleotide (ON473 in Supplementary Table S1 and Supplementary Figure S2) appears in the Supplementary Methods section. The ON473 oligonucleotide was annealed to 1.2 molar equivalents of the complementary strand (anti-ON473 in Supplementary Table S1) in 20 mM Tris/HCl, pH 7.5, and 25 mM NaCl, resulting in the E1X-dsDNA used in crystallization trials.

Crystallograpic studies

Crystallization

Wild-type *Mt*OGT was purified as previously described [21], with buffer exchanged against 20 mM Tris/HCl, pH 7.5, and 25 mM NaCl (PD10 column, GE Healthcare), mixed with E1X-dsDNA in equimolar ratio, and incubated for 18 h at 4 °C. The reaction mixture was concentrated (10-kDa MWCO, Vivaspin, Vivascience, Fisher Scientific) and loaded on to a size exclusion chromatography column (Superdex 200 10/300, GE Healthcare). The *Mt*OGT::E1X-dsDNA complexes eluted in a broad peak corresponding to absorption maxima at wavelengths 280 nm and 260 nm; the corresponding fractions were pooled, and concentrated up to 5 mg/ml as described above. Crystallization conditions for the *Mt*OGT::E1X-dsDNA complex were identified by means of a robot-assisted (Oryx4, Douglas Instruments), sitting drop-based, spare-matrix strategy using kits from Hampton Research and Qiagen. The initially obtained needle clusters were used as micro-seeds to inoculate 1 *μ*l of freshly prepared *Mt*OGT::E1X-dsDNA complex mixed with an equal volume of reservoir solution (0.2 M ammonium acetate, 22 % PEG 3350 and 0.1 M Hepes, pH 7.5), and equilibrated in a hanging drop against 800 *μ*l of the reservoir solution at 4 °C. Single thin rod crystals grew up to their maximum dimensions of 0.05 mm in about 6 weeks. Crystals of the R37K or Y139F *Mt*OGT variants were grown using the hanging drop vapour diffusion method by mixing 2 *μ*l of the corresponding protein solution at 5 mg/ml with an equal volume of a reservoir solution containing 0.1 M Hepes, pH 7.5, 4 % PEG 8000, and either 4 % or 8 % ethylene glycol (for *Mt*OGT-R37K and *Mt*OGT-Y139F, respectively); the drops were equilibrated against 800 *μ*l of the corresponding reservoir solution at 4 °C until crystals reached their maximum dimensions of 0.2 mm in about 2 weeks.

Data collection

All crystals used in diffraction experiments were directly taken from the corresponding crystallization drop, rapidly equilibrated in the specific reservoir solution containing 15 % glycerol as cryoprotectant, and flash-frozen under liquid nitrogen. Diffraction experiments were conducted at 100K using synchrotron radiation at the ID-29 (*Mt*OGT::E1X-dsDNA complex) or ID14-EH4 (*Mt*OGT-R37K and *Mt*OGT-Y139F variants) beam lines (European Synchrotron Radiation Facility, Grenoble, France). Complete diffraction datasets were collected up to 3.0-, 2.3-and 2.6-Å resolution (1 Å = 0.1 nm) for crystals of the *Mt*OGT::E1X-dsDNA complex, and the *Mt*OGT-R37K and the *Mt*OGT-Y139F variants, respectively. For all data collections, diffraction intensities were integrated and scaled by using the CCP4 suite of programs [27].

Structure determination

Analysis of the *Mt*OGT::E1X-dsDNA diffraction dataset assigned the crystal to the orthorhombic space group $P2_12_12_1$, with cell dimensions a = 43.48 Å, b = 102.90 Å and c = 137.09 Å, containing three protein chains and one dsDNA molecule per asymmetrical unit, with a corresponding solvent content of 50 %. The structure of the *Mt*OGT::E1X-dsDNA complex was solved by molecular replacement using the program Phaser [28]. The starting search model for the protein component was the structure of *Mt*OGT (PDB accession code 4BHB) [21], edited to omit the Tyr^{139} residue of the active site loop and the C-terminal tail (residues 156–165); the starting search model for the DNA component was the E1X-dsDNA, as crystallized in complex with hAGT (PDB accession code 1T39) [24], omitting bases 12–13 and 14–15 of the duplex. The resulting electron density map was of good quality, allowing manual model rebuilding, using the program Coot [29]. The program PHENIX [30] and Refmac [27] were used for crystallographic refinement and to add water molecules. The structures of *Mt*OGT-R37K and *Mt*OGT-Y139F were solved by molecular replacement using the program Phaser [28] and the structure of wild-type *Mt*OGT as the search model (PDB accession code 4BHB) [21], omitting water/ligand molecules and either Arg^{37} or Tyr^{139} residues, respectively. In both cases the procedure yielded high-quality electron density maps. Manual model building, crystallographic refinement and solvent addition were performed as described above for the *Mt*OGT::E1X-dsDNA structure. The stereochemistry of the refined models has been assessed using the program PROCHECK [31]. Data collection and refinement statistics are summarized in Table 1. Structural superimpositions were performed with the Superpose program of the CCP4 suite [27]; figures were generated using PyMol [32].

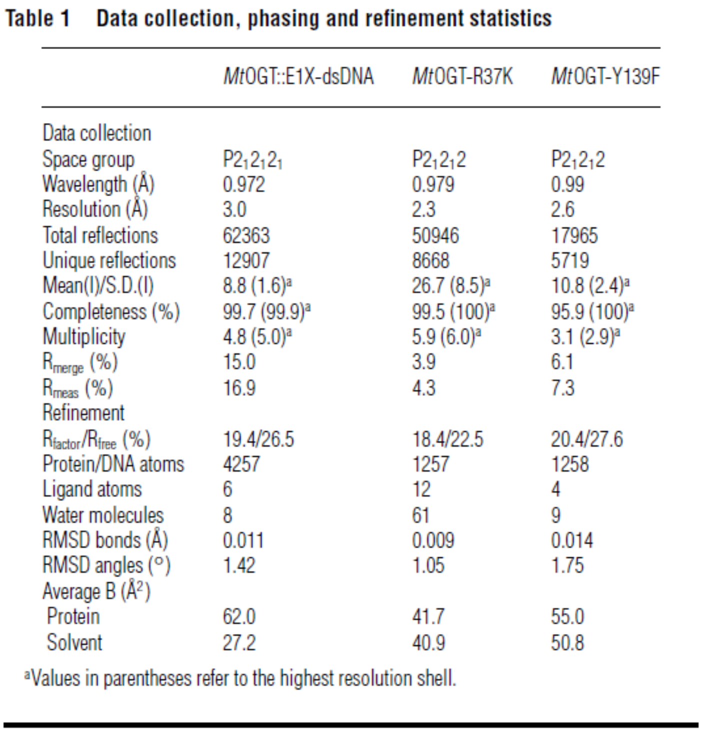


Table 1 Data collection, phasing and refinement statistics

	*Mt*OGT::E1X-dsDNA	*Mt*OGT-R37K	*Mt*OGT-Y139F
Data collection			
Space group	$P2_12_12_1$	$P2_12_12$	$P2_12_12$
Wavelength (Å)	0.972	0.979	0.99
Resolution (Å)	3.0	2.3	2.6
Total reflections	62363	50946	17965
Unique reflections	12907	8668	5719
Mean(I)/S.D.(I)	8.8 (1.6)[a]	26.7 (8.5)[a]	10.8 (2.4)[a]
Completeness (%)	99.7 (99.9)[a]	99.5 (100)[a]	95.9 (100)[a]
Multiplicity	4.8 (5.0)[a]	5.9 (6.0)[a]	3.1 (2.9)[a]
R_{merge} (%)	15.0	3.9	6.1
R_{meas} (%)	16.9	4.3	7.3
Refinement			
R_{factor}/R_{free} (%)	19.4/26.5	18.4/22.5	20.4/27.6
Protein/DNA atoms	4257	1257	1258
Ligand atoms	6	12	4
Water molecules	8	61	9
RMSD bonds (Å)	0.011	0.009	0.014
RMSD angles (°)	1.42	1.05	1.75
Average B (Å^2)			
Protein	62.0	41.7	55.0
Solvent	27.2	40.9	50.8

[a] Values in parentheses refer to the highest resolution shell.

Deposition

The atomic coordinates and structural factors of the *Mt*OGT::E1X-dsDNA complex, *Mt*OGT-R37K and *Mt*OGT-Y139F have been deposited in the Protein Data Bank(http://www.rcsb.org) under the PDB accession codes 4WX9, 4WXC and 4WXD, respectively.

Biochemical analyses

To measure the alkyltransferase activity of the new *Mt*OGT point-mutated variants.

*Mt*OGT-R37K, *Mt*OGT-R37E and *Mt*OGT-Y139F, competitive assays using the fluorescent SNAP-Vista Green reagent (VG; New England BioLabs) were performed as previously described [21,33]. Similarly, the EMSA-based analysis of the three mutated variants of *Mt*OGT was performed adopting the same protocol used to characterize the wild-type protein and its R37L and T15S mutated versions [21].

RESULTS

Overall structure of *Mt*OGT complexed with E1X-dsDNA

In order to clarify the functional role of the Arg^{37} in *Mt*OGT-mediated catalysis, we co-crystallized the wild-type protein in the presence of the 13-bp-long E1X-dsDNA, thus choosing the same experimental strategy first adopted by Daniels et al. [24] to solve the structure of wild-type hAGT covalently bound to a modified-dsDNA (PDB code 1T39). Different from the equivalent structure of the human enzyme, in the *Mt*OGT::E1X-dsDNA crystal structure, three protein chains (A, B and C) and one E1X-dsDNA molecule are present in the asymmetrical unit, with chain A binding the E1X base at position 7 of the modified strand ($E1X_7$) (Figure 1a). It is of interest that, by applying crystallographic symmetry operators, a peculiar supramolecular assembly can be observed in the *Mt*OGT::E1X-dsDNA crystal lattice (Figure 1b). By focusing on a unit consisting of chain A bound to the $E1X_7$ base, and counting nitrogenous bases starting from the 5´-end of each strand, a symmetry equivalent of chain C (‘C sym. mate’) binds the deoxyadenosine residue at position 4 of the modified strand (dA_4), and a symmetry equivalent of chain B (‘B sym. mate’) does the same with the deoxyadenosine residue at position 5 of the complementary strand (dA_{18}). In all cases, the bound base adopts a fully extrahelical conformation, and is deeply inserted into the protein active site. Overall, the *Mt*OGT::E1X-dsDNA complex can be described as consisting of two co-oriented *Mt*OGT monomers sharing 1 bp of their 4-bp-long DNA-anchoring site on the damaged strand and displaying a reciprocal ‘N-to-C’ domain arrangement (‘chain A’ and ‘C sym. mate’), whereas the third chain (‘B sym. mate’), which binds the intact strand, shows a ‘C-to-C’ domain arrangement with respect to chain A (Figure 1c).

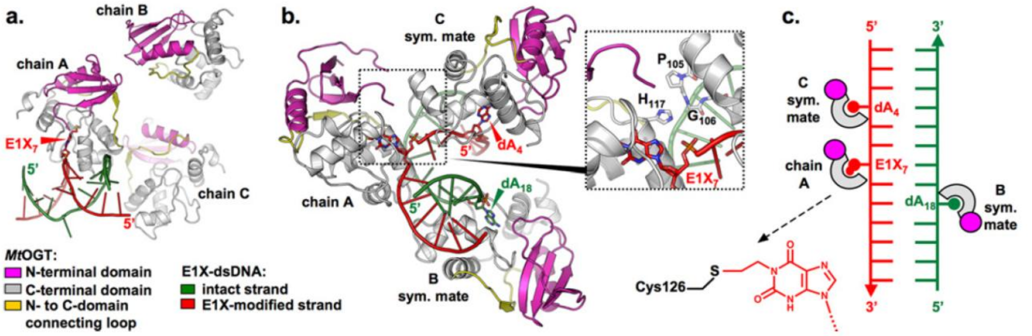

a.
chain B
chain A
$E1X_7$
5'
chain C
5'
*Mt*OGT:
N-terminal domain
C-terminal domain
N- to C-domain connecting loop
E1X-dsDNA:
intact strand
E1X-modified strand
b.
C sym. mate
P_{105}
H_{117}
G_{106}
5'
dA_4
$E1X_7$
dA_{18}
5'
chain A
B sym. mate
c.
5'
3'
C sym. mate
dA_4
chain A
$E1X_7$
dA_{18}
B sym. mate
Cys126
S
O
N
N
N
O
N
H
N
3'
5'

Figure 1: The overall structure of *Mt*OGT complexed with modified DNA

(**a**) Cartoon representation of the asymmetrical unit content of the *Mt*OGT::E1X-dsDNA crystal; the E1X-containing dsDNA is observed complexed with chain A, the C-terminal domain of which hosts the modified base (E1X, red arrowhead). (**b**) Cartoon representation of three *Mt*OGT chains assembled on to the same E1X-dsDNA molecule, resulting from applying crystal symmetry operators; extrahelical bases are rendered as sticks and signalled by an arrowhead (inset: zoomed view of the contact region between the co-oriented chains that bind bases of the modified DNA strand). (**c**) Representation of the reciprocal arrangement of the protein chains and DNA duplex illustrated in (b); the dashed arrow points to a scheme of the covalent adduct formed between the chain A catalytic cysteine (Cys^{126}) and the E1X base. The colour codes for protein domain and DNA strand identification appear at the bottom of the figure.

The association of each *Mt*OGT chain on to the E1X-dsDNA molecule is mainly stabilized by the strong protein–DNA interactions established by the helix-turn-helix (HTH) motif and a few conserved active site residues of each subunit, with the DNA minor groove and the flipped base, respectively (see below). In contrast, protein–protein interchain contacts are limited to a weak interaction engaging the co-oriented monomers ‘chain A’ and ‘C sym. mate’ (Figure 1b, inset). However, it must be noticed that, different from the crystal structures of wild-type *Mt*OGT and point-mutated variants of the protein in their ligand-free forms [21] (and the present study), in the *Mt*OGT::E1X-dsDNA structure no electron density was visible for chain A residues 33– 35 (omitted from the final model), and a poor electron density characterizes region 29–36 of the N-terminal domain random coil in each chain. For this reason, we cannot assume, under physiological conditions, that a higher number of contacts are established between the *Mt*OGT chain binding the alkylated base and the adjacent protein subunit locking the unmodified base at 4 bp upstream to the lesion ('chain A' and 'C sym. mate' in Figure 1).

The structure of the *Mt*OGT active site complexed with E1X-dsDNA

Ground-breaking X-ray crystallography-based studies on recombinant versions of hAGT complexed with modified dsDNA, containing either a physiologically relevant O^6-methylguanine residue [24] or base analogues carrying bulky substituting groups [24,25], disclosed the molecular details of the protein association with alkylated DNA. These results showed that hAGT invariably binds the dsDNA substrate at the level of its minor groove, by exploiting the conserved HTH motif of the protein C-terminal domain. In this peculiar mode of protein–dsDNA assembly, the modified nitrogenous base is flipped out from the regular base stacking and clamped into the enzyme active site pocket, thus resulting in proper placement of the reactive cysteine (Cys^{145} in hAGT) to catalyse the S_N2-like dealkylation reaction [24,25].

The architecture of the substrate-binding site of the three *Mt*OGT chains building up the *Mt*OGT::E1X-dsDNA crystal structure is quite similar to the one described for the human orthologue complexed with different dsDNA species (Figure 2a). Inspection of the active site of *Mt*OGT chain A reveals a continuous density signal contouring the catalytic Cys^{126} and the modified $E1X_7$ base (Figure 2b). Other close protein–DNA contacts involve: the strictly conserved 'arginine finger' (Arg^{109}) which, by invading the double helix from the minor groove side, and stacking between the planes of the dG_6 and dC_8 bases, structurally compensates for the flipped-out $E1X_7$ base; the carboxamide group of Asn^{115}, observed at a 2.9-Å mean distance from the $E1X_7$ O^2 position; the hydroxyl group of Tyr^{95}, standing at a 3.4-Å mean distance from both the N^3 atom and the deoxyribose moiety of the E1X$_7$ base; the active site loop residues Thr^{137} and Gly^{140}, the backbone oxygen and nitrogen atoms of which are observed at a distance of 2.8 and 2.7 Å from E1X$_7$ O^{4-} and O^6, respectively; and Tyr^{139} which contributes to narrowing of the active site and increasing the aromatic nature of the ligand-binding pocket. In addition, the positive charge at the N-side of helix H3 and the main-chain nitrogen atom of Ala^{132} appear to lock, from both sides, the sugar–phosphate backbone downstream of the lesion (Figure 2c).

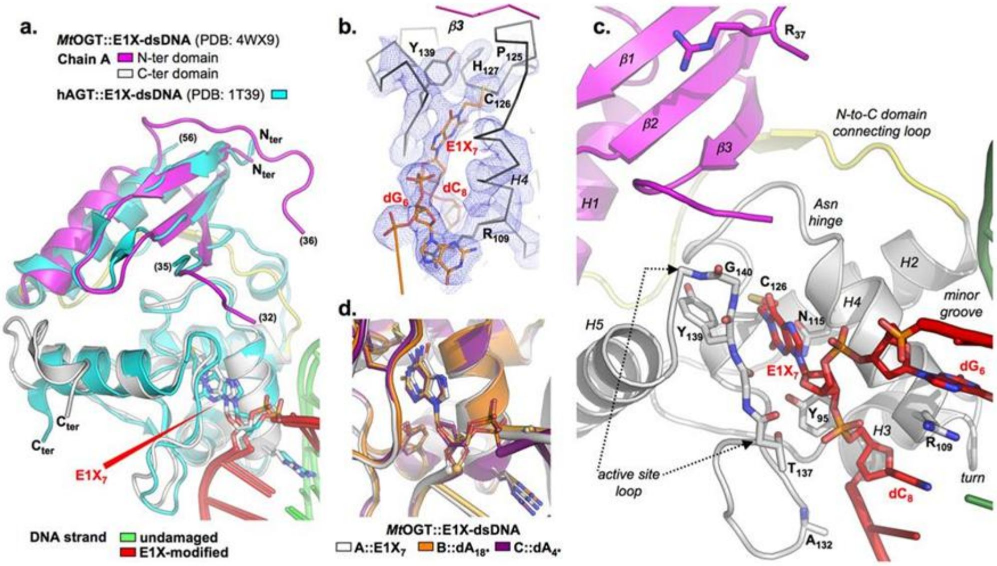

a.
*Mt*OGT::E1X-dsDNA (PDB: 4WX9)
Chain A 🟪 N-ter domain
☐ C-ter domain
hAGT::E1X-dsDNA (PDB: 1T39) 🟦
(56)
N$_{ter}$
N$_{ter}$
(36)
(35)
(32)
C$_{ter}$
C$_{ter}$
E1X$_7$
DNA strand 🟩 undamaged
🟥 E1X-modified
b.
β3
Y$_{139}$
P$_{125}$
H$_{127}$
C$_{126}$
E1X$_7$
dG$_6$
dC$_8$
H4
R$_{109}$
d.
*Mt*OGT::E1X-dsDNA
☐ A::E1X$_7$ 🟧 B::dA$_{18'}$ 🟪 C::dA$_{4'}$
c.
R$_{37}$
β1
β2
β3
N-to-C domain connecting loop
H1
Asn hinge
H2
G$_{140}$
C$_{126}$
H4
minor groove
H5
Y$_{139}$
N$_{115}$
E1X$_7$
dG$_6$
Y$_{95}$
H3
R$_{109}$
active site loop
T$_{137}$
turn
dC$_8$
A$_{132}$

Figure 2: Structural analysis of the *Mt*OGT protein complexed with E1X-dsDNA

(**a**) Cartoon representation of the optimally superimposed structures of *Mt*OGT (chain A) and hAGT (pdb code 1T39), each in complex with the E1X-dsDNA substrate; the E1X base is rendered as sticks and coloured, applying the same colour codes used for the corresponding protein chain (shown on the top of the panel). (**b**) Close-up view of the active site of the *Mt*OGT::E1X-dsDNA chain A housing the modified base (E1X7), with σA-weighted 2Fo − Fc electron density contoured at 1.0 σ; the Cys^{126} thiol group is observed at a distance of 2.1A° from E1X7, the C11 atom; the protein backbone appears as a ribbon. (**c**) Zoomed view of the *Mt*OGT chain A active site complexed with E1X-dsDNA; secondary structural elements and functional motifs are indicated in italic (the colour codes for protein domain/DNA strand identification appear in panel a). (**d**) Cartoon representation of the active site of optimally superimposed *Mt*OGT::E1X-dsDNA chains A, B and C (average RMSDs are 0.671 and 0.456A° for the couples B/A and C/A, respectively); DNA appears as a cartoon and coloured, applying the same colour codes used for the corresponding protein. Protein residues and DNA bases mentioned throughout the text are rendered as sticks.

With the obvious exception of contacts involving E1X$_7$-specific positions, an almost identical bonding scheme is observable in the active site of chains B and C – which host the symmetry equivalent of the dA$_{18}$ and dA$_4$ unmodified bases, respectively (Figure 2d) – thus indicating that *Mt*OGT can efficiently bind nitrogenous bases independently of the presence of the alkyl adduct.

In principle, the insertion of an undamaged adenine residue into the *Mt*OGT ligand-binding pocket would not expose the DNA substrate to an increased risk of chemical modifications. In fact, the reactivity of the purine ring N^1 and N^6 positions – as they were observed in the active site of the B and C chains of the *Mt*OGT::E1X$_7$-dsDNA structure – does not appear significantly enhanced by the nearby catalytic cysteine, nor by the presence of the other residues coordinating the base (see Supplementary Figure S3a).

Notably, one of the three protein chains (chain hAGT-B) building up the crystal structure of hAGT in complex with a dsDNA containing an *N*^{4}-alkylcytosine base (PDB code 1YFH) [25] binds the thymine base at the 3´-end of the modified strand. However, different from what we observed in the active site of the B and C monomers of the *Mt*OGT::E1X-dsDNA structure, the thymine base appears to be only partially inserted into the hAGT-B ligand-binding pocket (see Supplementary Figure S3b). We therefore speculated that *Mt*OGT could perform lesion searching through a non-selective base-flipping mechanism, with the flipped-out base fully inserted into the active site. If this assumption is correct, *Mt*OGT could not adopt a gate-keeping mechanism in discriminating between normal and damaged bases *in vivo*, different from what was hypothesized for the human counterpart [25,34].

*Mt*OGT undergoes structural rearrangements on DNA binding

The structure of *Mt*OGT complexed with E1X-dsDNA discloses a further unique feature of the mycobacterial protein, i.e. its conformational plasticity. In fact, the structural analyses of hAGT [22,23,24,25] and *Solfolobus sulfataricus* OGT [35], at different stages of the transalkylation reaction, suggest that the active site of the human and archaea proteins is largely pre-shaped to perform the catalysis, without requiring heavy structural rearrangements. On the contrary, the association of *Mt*OGT with the E1X-dsDNA substrate induces the repositioning of three solvent-exposed protein regions: a random coiled segment (residues 29–39) of the N-terminal domain, part of the active site loop (residues 135–142) and the C-terminal tail (residues 156–165) (Figure 3a). As a consequence, each protein monomer in the *Mt*OGT::E1X-dsDNA complex appears more compact than the ligand-free protein (Figure 3b). These conformational changes are accompanied by the side-chain repositioning of a number of residues of both protein domains (Figure 3c). It is interesting that, in the *Mt*OGT::E1X-dsDNA structure, the segment encompassing residues 29–35 moves away from the three-stranded β -sheet that builds up the core of the N-terminal domain, and gets closer to the DNA-binding surface of the C-terminal domain, behaving as a flap that sees Arg^{37} as its pivotal point.

We underline that this analysis was mainly conducted by inspecting the conformation adopted by chains B and C, because, different from chain A, their α-carbon backbone at the level of the flap is fully defined. However, given the minimal average RMSD resulting from superimposing the three protein chains building up the *Mt*OGT::E1X-dsDNA crystal structure, and taking into account that the B and C monomers host a nitrogenous base in their active site (see Figure 2d), we propose that an equivalent structural repositioning of the flap might also occur in the *Mt*OGT subunit binding the modified base.

The Arg^{37}-containing random coil could participate in the cooperative assembly of protein clusters on to the dsDNA substrate

The analysis of the *Mt*OGT::E1X-dsDNA crystal structure seems to exclude direct participation of Arg^{37} in DNA binding, because the protein residue and the sugar–

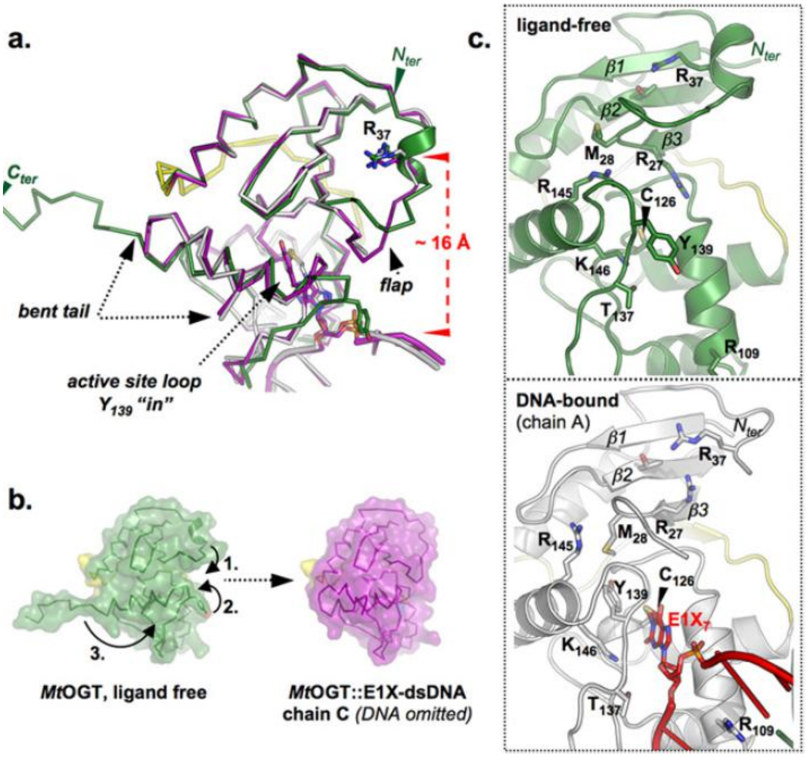

a.
N_{ter}
C_{ter}
R_{37}
~ 16 Å
flap
bent tail
active site loop
Y_{139} "in"
b.
1.
2.
3.
*Mt*OGT, ligand free
*Mt*OGT::E1X-dsDNA
chain C *(DNA omitted)*
c.
ligand-free
$\beta 1$
$\beta 2$
$\beta 3$
N_{ter}
R_{37}
M_{28}
R_{27}
R_{145}
C_{126}
Y_{139}
K_{146}
T_{137}
R_{109}
DNA-bound
(chain A)
$\beta 1$
$\beta 2$
$\beta 3$
N_{ter}
R_{37}
R_{145}
M_{28}
R_{27}
C_{126}
Y_{139}
$E1X_7$
K_{146}
T_{137}
R_{109}

Figure 3: The conformation adopted by discrete protein regions differs in the ligand-free and DNA-bound *Mt*OGT structures

(**a**) Structural superposition of *Mt*OGT in its apo form (PDB code 4BHB, green coloured) and complexed with E1X-dsDNA (PDB code 4WX9), highlighting the main structural rearrangements characterizing the DNA-bound chain A (in white) and chain C (in violet). (**b**) Surface representation of a *Mt*OGT monomer in ligand-free and DNA-bound states; the arrows indicate the direction of the movements of the flap (1), the active site loop (2) and the tail (3) of the protein on DNA binding. (**c**) Zoomed views of selected residues, with a side-chain conformation that differs between the superimposed structures of ligand-free (upper panel) and E1X-dsDNA-bound (lower panel) *Mt*OGT.

phosphate backbone of the E1X-dsDNA substrate are observed at a distance of >16 Å (Figure 3a). Instead, we propose that Arg^{37} could function as a hinge limiting the conformational plasticity at the C-side of the flap, by participating to keep it in contact with the bulk core of the N-terminal domain, and also on the formation of the protein–DNA complex (Figure 3c). In principle, the absence of such an anchoring site – as exemplified by the *Mt*OGT-R37L variant characterizing a number of frequently isolated *M. tuberculosis* strains – could affect the capability of the flap to undergo discrete movements. In turn, the resulting unrestrained flexibility of the N-terminal domain random coil could hamper the correct assembly of *Mt*OGT clusters at the damaged DNA sites.

To test this hypothesis, we expressed and purified two new mutated versions of *Mt*OGT (*Mt*OGT-R37K and *Mt*OGT-R37E), and analysed their dsDNA^{met}-repairing activity, by adopting the same VG-based assay [33] previously used to characterize the wild-type protein and the *Mt*OGT-R37L variant [21]. Our data (see Supplementary Table S2) show that the *Mt*OGT-R37E mutant exhibits a 5-fold lower affinity for the methylated duplex ($K_{DNA}{}^{met}{}_{R37E}$ = 1.14 ± 0.15μM) with respect to *Mt*OGT ($K_{DNA}{}^{met}{}_{wt}$ = 0.24 ± 0.11μM [21]), whereas the more conservative Arg^{37}-to-Lys substitution produces a more limited effect on the dsDNA^{met}-binding constant($K_{DNA}{}^{met}{}_{R37K}$ = 0.38 ± 0.2 μM).

In parallel we perform EMSA-based experiments (Figure 4a) using TAMRA (tetramethylrhodamine)-labelled non-alkylated dsDNA probe (see Supplementary Table S1). It is interesting that the *Mt*OGT-R37E protein reaches a plateau in band-shift activity at a DNA:protein molar ratio of 1:600 ($K_{DNA}{}^{met}{}_{R37E}$ = 41.4 ± 1.1 μM), whereas both the wild-type *Mt*OGT and the *Mt*OGT-R37K induce a complete shift at DNA:protein molar ratio of 1:150 ($K_{DNA\text{-}wt}$ = 7.2 ± 0.2μM [21], $K_{DNA\text{-}R37K}$ = 13.2 ± 0.7 μM). The results of the EMSA-based analysis well with those obtained from the VG-based competitive assays, and are consistent with previously published data showing that the recombinant *Mt*OGT-R37L variant displays a 10-fold lower affinity towards the dsDNA^{met} substrate compared with the wild-type *Mt*OGT, although the cooperativity of DNA binding is maintained [21]. Taken together, our past studies [21] and the present study confirm that the Arg^{37} residue, although not being directly involved in substrate binding, plays an active role during catalysis, a role that can be performed almost equally well by the positively charged lysine residue. By contrast, the presence of a hydrophobic or negatively charged side chain at position 37 of the *Mt*OGT protein – which characterizes the *Mt*OGT-R37L and *Mt*OGT-R37E proteins, respectively – translates into less efficient DNA binding and repair.

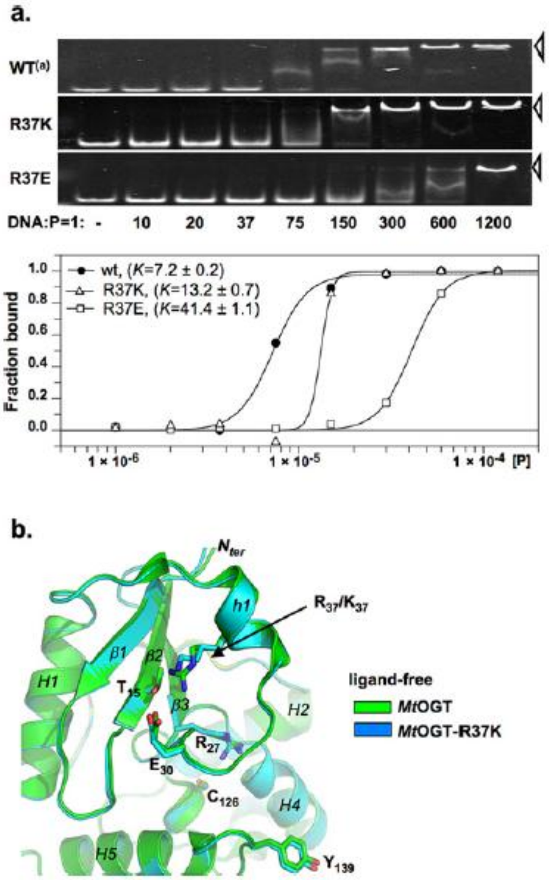


Figure 4: A positively charged residue at *Mt*OGT position 37 is required for optimal dsDNA binding

(**a**) Upper image: EMSA-based analysis of wild-type *Mt*OGT (WT) [21], and of the indicated point-mutated proteins, performed by using 1 pmol of TAMRA-labelled dsDNA (see Table 1) as the probe (DNA); lanes 2–9: increasing amounts of protein (P) incubated in the presence of the probe at the indicated DNA:protein molar ratio; in each panel the open arrowheads point to the shifted DNA probe. Lower image: plot of the DNA-bound protein fraction at each DNA:protein molar ratio tested in EMSA (upper image); [P], protein concentration (M); K , dissociation constant (μM). **(b**) Close-up of the N-terminal domain and part of the of the active site of the ligand-free *Mt*OGT (PDB code 4BHB) [21] and *Mt*OGT-R37K (PDB code 4WXD), on optimal structural superimposition; residues mentioned in the text appear as sticks; secondary structure elements are labelled in italic.

By analyzing the crystal structure of the *Mt*OGT-R37K variant (Figure 4b and see Table 1), we noticed that the lysine residue could partially substitute for arginine, in terms of charge and size, inside the peculiar network of contacts established between the second β -strand and the facing random coiled region of the N-terminal domain. In the structure of the loss-of-function *Mt*OGT-R37L protein, the presence of a leucine at position 37 destroys this bonding scheme [21]. However, in both these ligand-free structures, no relevant changes of the local fold are observable. Given the requirement of a positively charged group at position 37 of the protein for a fully efficient catalysis, we speculate that Arg^{37} could play a role in coordinating the repositioning of the flap during proper DNA recognition and binding, thus optimizing molecular contacts between adjacent monomers assembled on to the damaged DNA, as observed in the *Mt*OGT::E1X-dsDNA structure.

The intrinsic flexibility of the *Mt*OGT active site loop

As mentioned above, the active site loop and the C-terminal tail of *Mt*OGT adopt different conformations, depending on the association of the protein with the DNA substrate (see Figure 3). Different from what has been reported for all OGTs for which the crystal structure has so far been solved [22-25,35-37], but reminiscent of what was observed in the OGT structure of *Methanococcus jannaschii* in solution [38], the C-side region of the active site loop (residues 136–141) of the ligand-free structures of *Mt*OGT and its mutated variants is invariably oriented towards the bulk solvent. This conformation is stabilized by contacts established between the Tyr^{139}-conserved residue of the active site loop and the stretched-out C-terminal tail of the closest symmetry mate within the crystal lattice [21] (and the present study). On the contrary, the C-side of the active site loop of each protein chain that builds up the *Mt*OGT::E1X-dsDNA structure is bent inwards towards the catalytic pocket, where it participates in making the ligand-binding cavity fit the flipped-out base (Figure 5a). These observations raise the possibility that the active site of *Mt*OGT could exist in two alternative conformations (‘ligand-free/active site loop out’ or ‘DNA-bound/active site loop in’) also in a physiological context, displaying a degree of structural plasticity higher than that characterizing the equivalent region of hAGT.

However, if the Tyr^{139} residue of the active site loop of *Mt*OGT performed exactly the same molecular tasks highlighted for the equivalent residue of the human protein (Tyr^{158}), namely narrowing of the ligand-binding pocket and providing an aromatic environment for the alkyl adduct [24,25], then the substitution of Tyr^{139} by a phenylalanine should have little effect on catalysis.

Data from VG-based assays (see Supplementary Table S2) reveal that a *Mt*OGT-Y139F variant displays a 10-fold lower affinity for dsDNA^{met} compared with wild-type *Mt*OGT ($K_{DNA}{}^{met}{}_{Y139F}$ = 2.19±0.5 μM; $K_{DNA}{}^{met}{}_{wt}$ = 0.24±0.11 μM [21]), confirming the *Mt*OGT requirement of a tyrosine residue at position 139 for optimal repair of an O^6-methylated guanine in dsDNA. Instead, the *Mt*OGT ability to bind unmodified dsDNA appears less affected by the Tyr^{139}-to-Phe substitution; in fact, when analysed in EMSA, both proteins reach a plateau in band-shift activity at a DNA:protein molar ratio of approximately 1:150, displaying an affinity towards the unmodified probe that differs 3-fold ($K_{DNA\text{-}wt}$ = 7.2±0.2 μM [21], $K_{DNA\text{-}Y139F}$ = 20.5±3.2 μM; Figure 5b). Therefore, we speculate that Tyr^{139} could play a role not only in properly fixing the base inside the protein active site on DNA binding, as proposed for hAGT Tyr^{158} [24,25,39-41], but also in making *Mt*OGT able to discriminate between intact and alkylated dsDNA molecules, albeit through a molecular mechanism that will need further study for elucidation.

The *Mt*OGT::E1X-dsDNA crystal structure provides insights into cooperative DNA binding

The architecture of *Mt*OGT in a stable complex with the E1X-dsDNA substrate could be regarded as a snapshot of a potential reaction step at which the modified base has already been recognized and bound by one monomer (chain A), whereas two other subunits (the chain B and C symmetry mates) occlude available binding

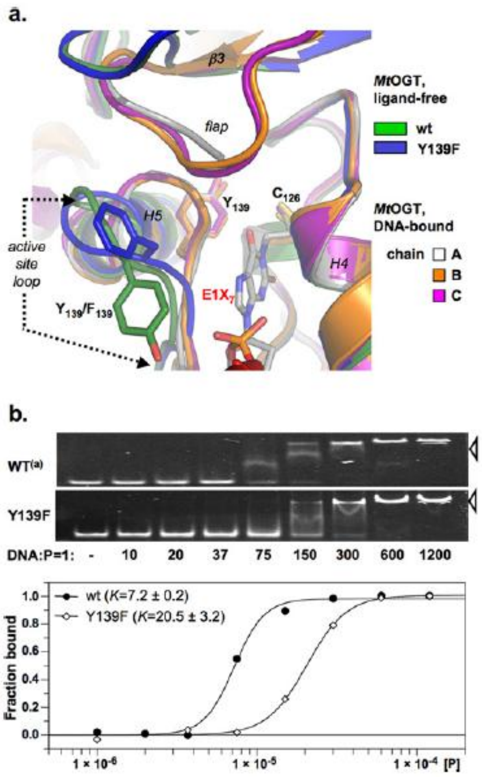


Figure 5: The Tyr^{139} residue could help the active site loop movements during DNA binding

(**a**) Zoomed view of the active site of *Mt*OGT (PDB code 4BHB) [21] and *Mt*OGT-Y139F (PDB code 4WXC) – both crystallized in ligand-free form – and of the A, B and C chains building up the *Mt*OGT::E1X-dsDNA complex (PDB code 4WX9), resulting from optimal superimposition of the corresponding structures (colour codes for protein/chain identification appear on the right); residues mentioned in the text appear as sticks; secondary structure elements are labeled in italic. (**b**) Upper image: EMSA-based analysis of *Mt*OGT (WT) [21] and its Y139F mutated variant, performed as detailed in Figure 4. Lower image: plot of the DNA-bound protein fraction at each DNA:protein molar ratio tested in EMSA (upper image); [P], protein concentration (M); K , dissociation constant (μM).

sites on both strands of the dsDNA substrate, at the highest possible density allowed in the close proximity of the lesion, by housing unmodified nucleobases in their active site (see Figure 1). From this standpoint, the supramolecular assembly revealed by our structure could represent a model of the *Mt*OGT clustering on a monoalkylated dsDNA molecule.

However, we cannot rule out the possibility that longer, more physiological DNA substrates might sustain the nucleation of *Mt*OGT protein clusters larger in size than the one characterizing the *Mt*OGT::E1X-dsDNA crystal structure. To verify this hypothesis, we tried to model further DNA-bound *Mt*OGT monomers towards the 5′-end of the modified strand, starting from the experimental ‘C sym. mate’ (monomer ‘2’ in Figure 6), and using the ‘chain A/C sym. mate’ dimer as the moving unit (monomers ‘1’ and ‘2’ in Figure 6).

It is of interest that the unprecedented association of a *Mt*OGT monomer with the region of the intact DNA strand facing the alkylated base (i.e. the ‘B sym. mate’-binding dA_{18}) would hamper the recruitment of additional protein subunits at the 5′ -side of the damaged base (Figure 6a).

On the other hand, the *Mt*OGT::E1X-dsDNA crystal structure itself does not provide any indication of the dynamics of the alkylated-DNA damage reversal performed by *Mt*OGT. Therefore, we can also assume that the binding of co-oriented *Mt*OGT monomers on to the modified strand could be favoured on assembly of protein clusters with components that display unbiased binding to both strands. To visualize the former situation, we reiterate the superimposition procedure mentioned above, by omitting the chain B symmetry mate. We obtained a model of *Mt*OGT protein clusters (Figure 6b) that proves to be more compact compared with the one proposed for hAGT [46,47], possibly due to the structural plasticity of *Mt*OGT, which could allow more crowded protein assembling on to DNA. It is interesting that the DNA binding-associated repositioning of the flap (see Figure 3) enables additional contacts between adjacent subunits. In particular, residues 32–34 of chain 2 are clamped between the active site entrance of the same chain, and the turn element of the HTH motif of adjacent chain 3 (residues 100–105) (Figure 6b, inset on the left), thus strengthening our hypothesis of a possible direct contribution of the flap to cooperative DNA binding. Moreover, by considering chains 1 and 5 in our model (Figure 6b, inset on the right) we noticed that the N-terminal domain, random-coiled region of chain 1 collides with the tail

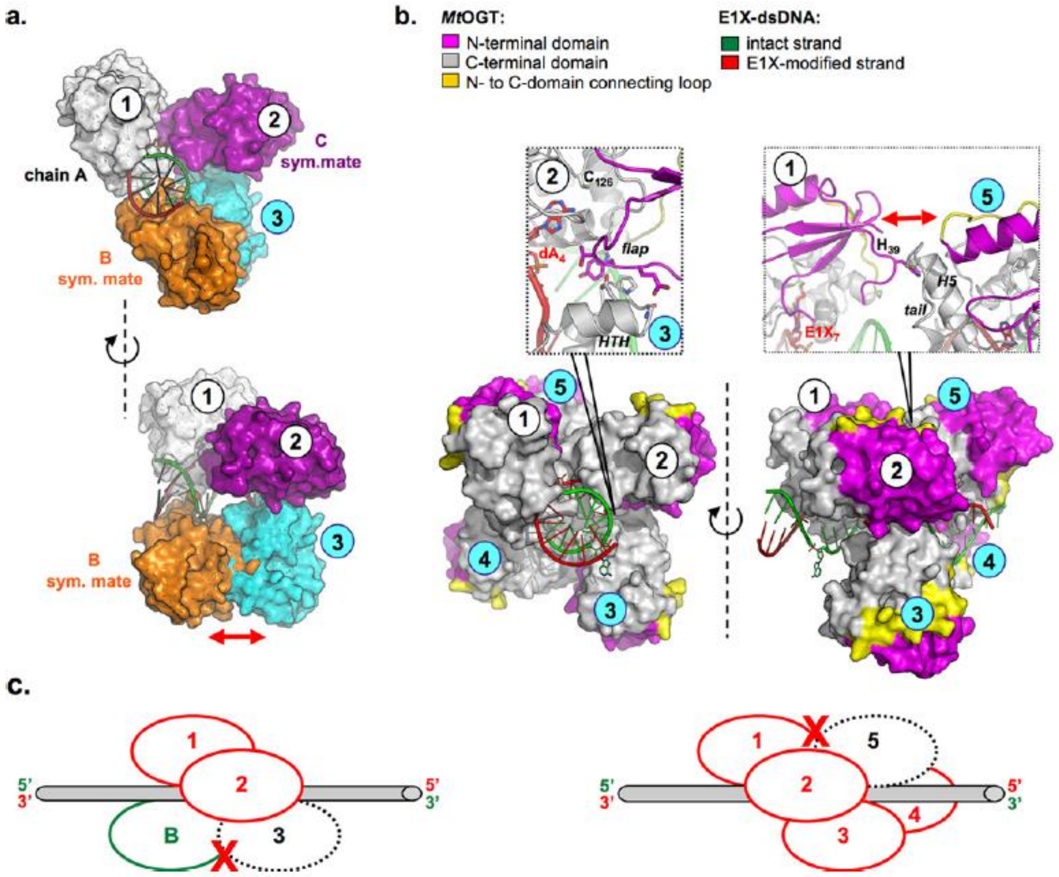

a.
1
2
C
sym.mate
chain A
3
B
sym. mate
1
2
3
B
sym. mate
b.
MtOGT:
N-terminal domain
C-terminal domain
N- to C-domain connecting loop
E1X-dsDNA:
intact strand
E1X-modified strand
2
C126
flap
dA4
HTH
3
1
5
H39
H5
tail
E1X7
5
1
2
4
3
1
5
2
4
3
c.
1
2
5'
3'
5'
3'
B
X
3
1
X
5
2
5'
3'
5'
3'
4
3

Figure 6: Structure-based model of *Mt*OGT clustering on to a DNA duplex

(**a**) The possible negative effect of the presence of a *Mt*OGT monomer (‘B sym. mate’), which is bound to the complementary strand opposing the lesion, on the growth of the protein cluster towards the 5_-end of the damaged strand. (**b**) Structure-based model of five co-oriented *Mt*OGT monomers assembled on to the same dsDNA molecule; the zoomed images show the interchain contacts observed at the chain 2/chain 3 (left) and chain 1/chain 5 (right) interfaces; the colour codes for protein domain and DNA strand identification appear at the top of the panel. In (a) and (b), the protein chains are rendered as on the surface and the DNA appears as a cartoon. (**c**) Schematic representations of the protein–DNA assemblies depicted in (a) and (b), viewed perpendicular to the dsDNA axis; the *Mt*OGT monomers that bind bases of the damaged or intact strand are depicted as red or green ovals, respectively; the dashed ovals represent the *Mt*OGT subunit that would come into collision with preassembled monomers.

of chain 5. This analysis suggests that both short- and long-range steric hindrance phenomena could play a role in regulating *Mt*OGT–DNA association and dissociation, resulting in protein clusters that are capable of self-limiting their own size, similar to what has been experimentally determined by direct atomic force microscopy studies of hAGT [48].

DISCUSSION

The architecture of the protein–DNA complexes revealed by the *Mt*OGT::E1X-dsDNA crystal structure could provide a potential solution to an inconsistency present in the literature concerning the DNA-binding mechanism of O^6-alkylguanine-DNA alkyltransferases.

Indeed, in spite of the fact that previous X-ray crystallography-based studies revealed a 1:1 protein:modified DNA stoichiometry [24,25,35]: (i) the cooperative binding of DNA has been demonstrated as a bona fide activity of hAGT [42-47], (ii) structure-based models of hAGT/dsDNA supramolecular complexes have been built and experimentally tested in solution [46] and, more importantly, (iii) the protein assembling into discrete clusters on physiological DNA substrates was directly observed by AFM-based experiments [48].

Several authors have analysed the possible functional benefits of performing alkylated-DNA direct repair in a cooperative manner. It has been pointed out that cooperative assembly of protein–DNA complexes might contribute to the efficiency of lesion search and removal, by concentrating the repair activity on the DNA substrate at a higher density than that expected in a non-cooperative DNA-binding model [45]. Furthermore, a small protein cluster size could allow tracking of a lesion wherever short regions of free DNA were made available, i.e. during DNA replication and transcription, on chromatin remodelling [49]. Moreover, an inherent capability of the protein of limiting its own distribution on DNA could influence the rates of binding to and dissociation from the target, and hence the kinetics of the lesion search; in fact, the repositioning of a subunit placed in the middle of a single long protein cluster should probably be slower than the repositioning of subunits mapping at the ends of many short clusters [45,48].

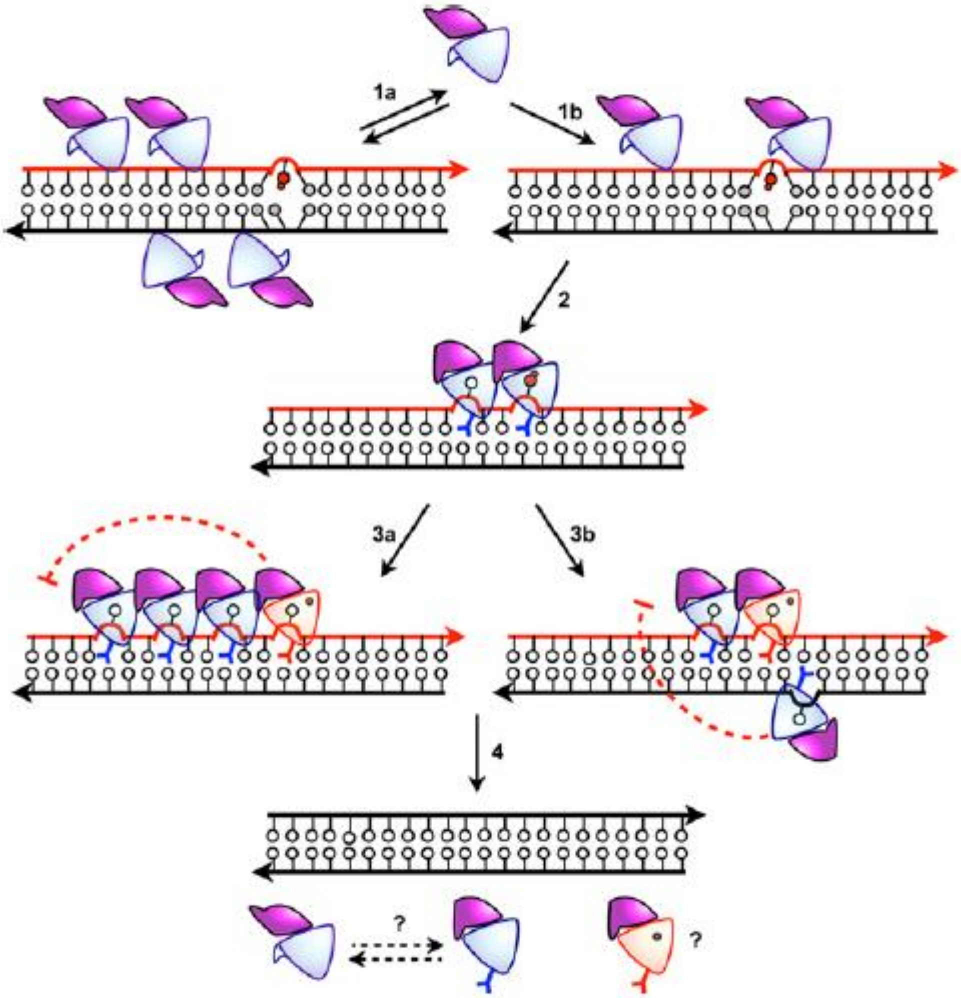


Figure 7: Preliminary model of *Mt*OGT-mediated direct alkylated DNA repair

Schematics of possible modes of DNA binding, protein cluster assembly and protein–DNA complex dismantling in *M. tuberculosis* emerging from crystallographic studies. The single steps are described under 'Discussion'.

These arguments could work well with our results to outline a preliminary model of alkylated-DNA recognition and repair performed by *Mt*OGT (Figure 7). In principle, to guarantee efficient scouting of alkylated bases inside the genome, both intact and damaged DNA molecules should represent a ligand for *Mt*OGT. However, the binding of the protein to an adduct-free DNA region could be less efficient compared with the binding to an alkylated substrate, or the protein could form suboptimal protein clusters (Figure 7, step 1a), leading to a weaker assembly (or an easier dissociation) of the protein–DNA complex. Our structural studies show that the insertion of a modified base inside the *Mt*OGT active site triggers conformational modifications of solvent-exposed regions of the protein which could act as a signal that a lesion has been encountered (Figure 7, step 1b). Additional protein subunits could now be tightly packed, also by virtue of their peculiar structural flexibility, at the 5´-side of the lesion, where they undergo the same structural rearrangements to host extrahelical nucleobases in their active site. In this way, the GC-rich mycobacterial DNA could be scanned, at a fixed space interval, for the presence of other alkylated sites (Figure 7, step 2).

Our present results suggest that, although an *O*-alkylated guanine is a potential substrate for the catalysis (leading to permanent protein inactivation), both purine bases could behave as reversible protein ligands; as a consequence, an unmodified guanine temporarily occupying the *Mt*OGT active site could be safely checked for chemical modifications. This corroborates the concept that identifies, in the alkylated-DNA, direct damage reversal performed by OGTs, a smart system to oversee genome quality [49]. Finally, the conformational changes induced in *Mt*OGT by its association with DNA appear to directionally bias protein–protein interaction towards the 5´-end of the modified strand; however, the harmful sequestration of *Mt*OGT in the form of a continuous coat on the DNA substrate could be limited either by steric hindrance phenomena involving co-oriented monomers (Figure 7, step 3a) or by the association of an *Mt*OGT subunit with the intact strand region opposing the alkylated base (Figure 7, step 3b), leading to the release of the unreacted monomers into the free protein pool (Figure 7, step 4), ready to initiate a new cycle.

Further experiments, using different DNA substrates and crystallization-independent techniques, have been undertaken to verify this model.

AUTHOR CONTRIBUTIONS

R. Miggiano, M. Rizzi and F. Rossi designed experiments, analysed structural and biochemical data, and wrote the manuscript. R. Miggiano performed structural and biochemical characterization experiments. G. Perugino and M. Ciaramella designed experiments for biochemical characterization and analysed biochemical data. M. Serpe performed EMSA analysis. D. Rejman designed experiments for E1X-dsDNA synthesis and analysed mass and NMR spectra. D. Rejman, O. Pav´ and R. Pohl developed and performed a new synthetic procedure for E1X-dsDNA preparation. S. Garavaglia performed crystal data collections analysis. S. Lahiri performed robot-assisted crystallization trials and optimized crystal-growth conditions. All authors approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities and assistance in using the ID-29 and ID14-EH4 beamlines.

FUNDING

This work was supported by the European Community (Project ‘SysteMTb’ [HEALTH-F4-2010-241587).

REFERENCES

1 Mizrahi, V. and Andersen, S. J. (1998) DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Mol Microbiol*. **29**, 1331-1339

2 Dos Vultos, T., Mestre, O., Tonjum, T. and Gicquel, B. (2009) DNA repair in Mycobacterium tuberculosis revisited. *FEMS Microbiol Rev*. **33**, 471-487

3 Durbach, S. I., Springer, B., Machowski, E. E., North, R. J., Papavinasasundaram, K. G., Colston, M. J., Bottger, E. C. and Mizrahi, V. (2003) DNA alkylation damage as a sensor of nitrosative stress in Mycobacterium tuberculosis. *Infect Immun*. **71**, 997-1000

4 Warner, D. F. and Mizrahi, V. (2006) Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. *Clin Microbiol Rev*. **19**, 558-570

5 Drablos, F., Feyzi, E., Aas, P. A., Vaagbo, C. B., Kavli, B., Bratlie, M. S., Pena-Diaz, J., Otterlei, M., Slupphaug, G. and Krokan, H. E. (2004) Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst)*. **3**, 1389-1407

6 Gorna, A. E., Bowater, R. P. and Dziadek, J. (2010) DNA repair systems and the pathogenesis of Mycobacterium tuberculosis: varying activities at different stages of infection. *Clin Sci (Lond)*. **119**, 187-202

7 Kurthkoti, K. and Varshney, U. (2012) Distinct mechanisms of DNA repair in mycobacteria and their implications in attenuation of the pathogen growth. *Mech Ageing Dev*. **133**, 138-146

8 Pegg, A. E. (2011) Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. *Chem Res Toxicol*. **24**, 618-639

9 Pegg, A. E. (2000) Repair of O(6)-alkylguanine by alkyltransferases. *Mutat*

Res. **462**, 83-100

10 Fang, Q., Kanugula, S., Tubbs, J. L., Tainer, J. A., and Pegg, A. E. (2010) Repair of O4-alkylthymine by O6-alkylguanine-DNA alkyltransferases. *J Biol Chem.* **285**, 8185-8195

11 Tubbs, J. L., Pegg, A. E., and Tainer, J. A. (2007) DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair (Amst).* **6**, 1100-1115

12 Tessmer, I., Melikishvili, M., and Fried, M. G. (2012) Cooperative cluster formation, DNA bending and base-flipping by O6-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res.* **40**, 8296-8308

13 Kanugula, S., Goodtzova, K., and Pegg, A. E. (1998) Probing of conformational changes in human O6-alkylguanine-DNA alkyl transferase protein in its alkylated and DNA-bound states by limited proteolysis. *Biochem J.* **329** (Pt 3), 545-550

14 Xu-Welliver, M., and Pegg, A. E. (2002) Degradation of the alkylated form of the DNA repair protein, O(6)-alkylguanine-DNA alkyltransferase. *Carcinogenesis.* **23**, 823-830

15 Yang, M., Aamodt, R. M., Dalhus, B., Balasingham, S., Helle, I., Andersen, P., Tonjum, T., Alseth, I., Rognes, T., and Bjoras, 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

16 Boshoff, H. I., Myers, T. G., Copp, B. R., McNeil, M. R., Wilson, M. A., and Barry, C. E., 3rd. (2004) The transcriptional responses of Mycobacterium tuberculosis to inhibitors of metabolism: novel insights into drug mechanisms of action. *J Biol Chem.* **279**, 40174-40184

17 Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., Dolganov, G., Efron, B., Butcher, P. D., Nathan, C., and Schoolnik, G. K. (2003) Transcriptional Adaptation of Mycobacterium tuberculosis within

Macrophages: Insights into the Phagosomal Environment. *J Exp Med.* **198**, 693-704

18 Ebrahimi-Rad, M., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., Kreiswirth, B., Blazquez, J., Jouan, M., van Soolingen, D., and Gicquel, B. (2003) Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family. *Emerg Infect Dis.* **9**, 838-845

19 Olano, J., Lopez, B., Reyes, A., Lemos, M. P., Correa, N., Del Portillo, P., Barrera, L., Robledo, J., Ritacco, V., and Zambrano, M. M. (2007) Mutations in DNA repair genes are associated with the Haarlem lineage of Mycobacterium tuberculosis independently of their antibiotic resistance. *Tuberculosis (Edinb).* **87**, 502-508

20 Mestre, O., Luo, T., Dos Vultos, T., Kremer, K., Murray, A., Namouchi, A., Jackson, C., Rauzier, J., Bifani, P., Warren, R., Rasolofo, V., Mei, J., Gao, Q., and Gicquel, B. (2011) Phylogeny of Mycobacterium tuberculosis Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS One.* **6**, e16020

21 Miggiano, R., Casazza, V., Garavaglia, S., Ciaramella, M., Perugino, G., Rizzi, M., and Rossi, F. (2013) Biochemical and structural studies of the Mycobacterium tuberculosis O6-methylguanine methyltransferase and mutated variants. *J Bacteriol.* **195**, 2728-2736

22 Daniels, D. S., Mol, C. D., Arvai, A. S., Kanugula, S., Pegg, A. E., and Tainer, J. A. (2000) Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *EMBO J.* **19**, 1719-1730

23 Wibley, J. E., Pegg, A. E., and Moody, P. C. (2000) Crystal structure of the human O(6)-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res.* **28**, 393-401

24 Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., and Tainer, J. A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat Struct Mol Biol.* **11**, 714-720

25 Duguid, E. M., Rice, P. A., and He, C. (2005) The structure of the human AGT protein bound to DNA and its implications for damage detection. *J Mol Biol.* **350**, 657-666

26 Noll, D. M., and Clarke, N. D. (2001) Covalent capture of a human O(6)-alkylguanine alkyltransferase-DNA complex using N(1),O(6)-ethanoxanthosine, a mechanism-based crosslinker. *Nucleic Acids Res.* **29**, 4025-4034

27 Collaborative Computational Project, Number 4 (CCP4) (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* **50**, 760-763

28 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658-674

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

30 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* **66**, 213-221

31 Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* **26**, 283-291

32 DeLano, W. L. (2002) The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA.

33 Perugino, G., Vettone, A., Illiano, G., Valenti, A., Ferrara, M. C., Rossi, M., and Ciaramella, M. (2013) Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation

damage. *J Biol Chem.* **287**, 4222-4231

34 Hu, J., Ma, A., and Dinner, A. R. (2008) A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc Natl Acad Sci U S A.* **105**, 4615-4620

35 Perugino, G., Miggiano, R., Serpe, M., Vettone, A., Valenti, A., Lahiri, S., Rossi, F., Rossi, M., Rizzi, M., and Ciaramella, M. (2015) Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic Acids Res.* doi: 10.1093/nar/gkv774

36 Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B., and Moody, P. C. (1994) Crystal structure of a suicidal DNA repair protein: the Ada O6-methylguanine-DNA methyltransferase from E. coli. *EMBO J.* **13**, 1495-1501

37 Hashimoto, H., Inoue, T., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., and Kai, Y. (1999) Hyperthermostable protein structure maintained by intra and inter-helix ion-pairs in archaeal O6-methylguanine-DNA methyltransferase. *J Mol Biol.* **292**, 707-716

38 Roberts, A., Pelton, J. G., and Wemmer, D. E. (2006) Structural studies of MJ1529, an O6-methylguanine-DNA methyltransferase. *Magn Reson Chem.* **44** Spec No, S71-82

39 Crone, T. M., Goodtzova, K., and Pegg, A. E. (1996) Amino acid residues affecting the activity and stability of human O6-alkylguanine-DNA alkyltransferase. *Mutat Res.* **363**, 15-25

40 Edara, S., Goodtzova, K., and Pegg, A. E. (1995) The role of tyrosine-158 in O6-alkylguanine-DNA alkyltransferase activity. *Carcinogenesis.* **16**, 1637-1642

41 Xu-Welliver, M., Leitao, J., Kanugula, S., and Pegg, A. E. (1999) Alteration of the conserved residue tyrosine-158 to histidine renders human O6-alkylguanine-DNA alkyltransferase insensitive to the inhibitor O6-benzylguanine. *Cancer Res.* **59**, 1514-1519

42 Fried, M. G., Kanugula, S., Bromberg, J. L., and Pegg, A. E. (1996) DNA

binding mechanism of O6-alkylguanine-DNA alkyltransferase: stoichiometry and effects of DNA base composition and secondary structure on complex stability. *Biochemistry*. **35**, 15295-15301

43 Rasimas, J. J., Pegg, A. E., and Fried, M. G. (2003) DNA-binding mechanism of O6-alkylguanine-DNA alkyltransferase. Effects of protein and DNA alkylation on complex stability. *J Biol Chem.* **278**, 7973-7980

44 Rasimas, J. J., Kar, S. R., Pegg, A. E., and Fried, M. G. (2007) Interactions of human O6-alkylguanine-DNA alkyltransferase (AGT) with short single-stranded DNAs. *J Biol Chem.* **282**, 3357-3366

45 Melikishvili, M., Rasimas, J. J., Pegg, A. E., and Fried, M. G. (2008) Interactions of human O(6)-alkylguanine-DNA alkyltransferase (AGT) with short double-stranded DNAs. *Biochemistry.* **47**, 13754-13763

46 Adams, C. A., Melikishvili, M., Rodgers, D. W., Rasimas, J. J., Pegg, A. E., and Fried, M. G. (2009) Topologies of complexes containing O6-alkylguanine-DNA alkyltransferase and DNA. *J Mol Biol.* **389**, 248-263

47 Adams, C. A., and Fried, M. G. (2011) Mutations that probe the cooperative assembly of O(6)-alkylguanine-DNA alkyltransferase complexes. *Biochemistry*. **50**, 1590-1598

48 Tessmer, I., and Fried, M. G. (2014) Insight into the cooperative DNA binding of the O(6)-alkylguanine DNA alkyltransferase. *DNA Repair (Amst).* **20**, 14-22

49 Begley, T. J and Samson, L. D. (2004) Reversing DNA damage with a directional bias. *Nat Struct Mol Biol.* **11**, 688-690

SUPPLEMENTARY MATERIAL

Crystal structure of Mycobacterium tuberculosis O6-methylguanine-DNA methyltransferase protein clusters assembled onto damaged DNA

Riccardo Miggiano, Giuseppe Perugino, Maria Ciaramella, Mario Serpe, Dominik Rejman, Ondřej Páv, Radek Pohl, Silvia Garavaglia, Samarpita Lahiri, Menico Rizzi and Franca Rossi

TABLE of CONTENTS:

Supplementary Methods 76-80
Supplementary Table 1 81
Supplementary Table 2 82
Supplementary Figure 1 83
Supplementary Figure 2 84
Supplementary Figure 3 85

Supplementary Methods

Synthesis of eX and of the oligonucleotide d(GCC ATG eX CTA GTA) (ON473)

Unless otherwise stated, all solvents used were anhydrous. TLC was performed on TLC plates precoated with silica gel (silica gel/TLC-cards, UV 254, Merck). Compounds were detected using UV light (254 nm), heating (for the detection of dimethoxytrityl group; orange colour). The purity of the prepared compounds was determined by LC-MS using a Waters AutoPurification System with 2545 Quarternary Gradient Module and 3100 Single Quadrupole Mass Detector using Luna C18 column (100 x 4.6 mm, 3 µm, Phenomenex) at a flow rate of 1 ml/min. Mass spectra were collected on an LTQ Orbitrap XL (Thermo Fisher Scientific) using ESI ionisation. NMR spectra were collected in DMSO-*d*6 or CDCl3 on a Bruker AVANCE 400 (1H at400.0 MHz, 13C at 100.6 MHz, and 31P at 162.0 MHz) and/or Bruker AVANCE 500 (1H at 500.0 MHz, 13C at125.7 MHz, and 31P at 202.3 MHz) NMR spectrometers. Chemical shifts (in ppm, δ scale) were referenced to the residual DMSO-*d*6 signal (2.5 ppm for 1H and 39.7 ppm for 13C). 31P NMR spectra were referenced to H3PO4 (0 ppm) as an external standard. Coupling constants (*J*) are given in Hz.

3’,5’-Bis-O-(tert-butyldimethylsilyl)-O6-[(2,4,6-trimethylphenyl)sulfonyl]-2‘-deoxyguanosine 2

The title compound was synthesized according to published procedure [Hayakawa Y, Hirose M, & Noyori R(1993) O-Allyl protection of guanine and thymine residues in oligodeoxyribonucleotides. *J Org Chem*, *58*(20):5551-5555] in 82% yield.

3’,5’-Bis-*O*-(*tert*-butyldimethylsilyl)-6-*O*-hydroxyethyl-2’-deoxyguanosine 3

DABCO (9.5 g, 84.7 mmol) was added to the stirred solution of 2 (14.36 g, 21.18

mmol) in THF (200 ml). The reaction mixture was stirred at rt for 3 h. Ethylenglycol (12 ml, 215.17 mmol) and DBU (9.5 ml, 63.59 mmol) were added. The mixture was stirred at rt 1 h and evaporated. The final product was obtained by chromatography on silica gel (elution with a gradient of ethyl acetate in toluene) in 50% yield (5.77 g, 10.69 mmol) in the form of yellowish oil.

1H NMR (499.8 MHz, DMSO-*d*6): 0.03, 0.04, 0.105, 0.107 (4 × s, 4 × 3H, CH3Si); 0.86, 0.89 (2 × s, 2 ×9H, (CH3)3CSi); 2.55 (ddd, 1H, *J*gem = 13.2, *J*2'b,1' = 6.0, *J*2'b,3' = 3.3, H-2'b); 2.73 (ddd, 1H, *J*gem = 13.2, *J*2'a,1' =7.6, *J*2'a,3' = 5.7, H-2'a); 3.64 (dd, 1H, *J*gem = 11.0, *J*5'b,4' = 4.4, H-5'b); 3.72 (dd, 1H, *J*gem = 11.0, *J*5'a,4' = 6.0, H-5'a); 3.75 (bq, 2H, *J* = 5.2, OCH2C**H2**OH); 3.82 (ddd, 1H, *J*4',5' = 6.0, 4.4, *J*4',3' = 2.7, H-4'); 4.41 (m, 2H, OC**H2**CH2OH); 4.51 (ddd, 1H, *J*3',2' = 5.7, 3.3, *J*3',4' = 2.7, H-3'); 4.89 (bt, 1H, *J* = 5.2, OH); 6.20 (dd, 1H, *J*1',2'= 7.6, 6.0, H-1'); 6.42 (bs, 2H, NH2); 8.05 (s, 1H, H-8).

^{13}C NMR (125.7 MHz, DMSO-*d*$_6$): -5.33, -5.29, -4.77, -4.59 (CH_3Si); 17.90, 18.15 (**C**(CH_3)$_3$); 25.87, 25.96 ((**C**H_3)$_3$C); 38.88 (CH_2-2'); 59.44 (OCH_2**C**H_2OH); 62.92 (CH_2-5'); 67.64 (O**C**H_2CH_2OH); 72.34 (CH-3'); 82.49 (CH-1'); 87.14 (CH-4'); 114.06 (C-5); 137.54 (CH-8); 154.15 (C-4); 159.92 (C-2); 160.64 (C-6).

HR-ESI $C_{24}H_{46}O_5N_5Si_2$ $[M+H]^+$ calcd 540.30320, found 540.30323

1-*N*-6-*O*-Ethano-2’-deoxyxanthosine 6

1M TBAF in THF (3 ml) was added to the solution of **5** (1.12 g, 2.14 mmol) in THF (8 ml). The reactionmixture was stirred at rt for 45 min. The reaction mixture was diluted with ethanol (20 ml) and Dowex 50 in TEA form (5 g) was added. The suspension was stirred 5 min and filtrated. The filtrate was evaporated and the final product was obtained by chromatography on silica gel (elution with a gradient of 0–100% ethanol/acetone/water/ethyl acetate 1:1:1:4 in ethyl acetate) in 87% yield (0.55 g, 1.87 mmol) in the form of white amorfous solid.

1H NMR (600.1 MHz, DMSO-*d*6): 2.24 (ddd, 1H, *J*gem = 13.2, *J*2'b,1' = 6.2, *J*2'b,3' = 3.2, H-2'b); 2.55 (ddd, 1H, *J*gem = 13.2, *J*2'a,1' = 7.6, *J*2'a,3' = 5.9, H-2'a); 3.50 (ddd, 1H, *J*gem = 11.8, *J*5'b,OH = 5.5, *J*5'b,4' = 4.9, H-5'b); 3.57(dt, 1H, *J*gem = 11.8, *J*5'a,4' = *J*5'a,OH = 4.9, H-5'a); 3.97 (td, 1H, *J*4',5' = 4.9, *J*4',3' = 2.9, H-4'); 4.21 (m, 2H, NC**H2**CH2O); 4.35 (m, 1H, *J*3',2' = 5.9, 3.2, *J*3',OH = 4.1, *J*3',4' = 2.9, H-3'); 4.77 (m, 2H, OC**H2**CH2N); 4.94(bdd, 1H, *J*OH,5' = 5.5, 4.9, OH-5'); 5.30 (d, 1H, *J*OH,3' = 4.1, OH-3'); 6.16 (dd, 1H, *J*1',2' = 7.6, 6.2, H-1'); 8.12(s, 1H, H-8).

13C NMR (150.9 MHz, DMSO-*d*6): 39.42 (CH2-2'); 42.64 (N**C**H2CH2O); 61.81 (CH2-5'); 68.06(O**C**H2CH2N); 70.86 (CH-3'); 83.55 (CH-1'); 88.03 (CH-4'); 119.45 (C-5); 137.56 (CH-8); 148.95 (C-4); 154.82 (C-2); 157.92 (C-6).

HR-ESI C12H14O5N4Na [M+Na]+ calcd 317.08564, found 317.08561

5’-*O*-Dimethoxytrityl-1-*N*-6-*O*-ethano-2’-deoxyxanthosine 7

DMTrCl (0.76 g, 2.24 mmol) was added in two portions to the solution of **6** (0.55 g,1.87 mmol) in pyridine(40 ml). The reaction mixture was stirred at rt overnight. The solvent was removed in vacuo and the product was obtained by purification by chromatography on silica gel (elution with a gradient of 0–10% ethanol in chloroform) in 45% yield (0.5 g, 0.84 mmol) in the form of white foam.

1H NMR (499.8 MHz, DMSO-*d*6): 2.57 (ddd, 1H, *J*gem = 13.5, *J*2'b,1' = 6.8, *J*2'b,3' = 4.4, H-2'b); 2.65 (dt, 1H, *J*gem = 13.5, *J*2'a,1' = *J*2'a,3' = 6.4, H-2'a); 3.09 (dd, 1H, *J*gem = 10.2, *J*5'b,4' = 3.5, H-5'b); 3.26 (dd, 1H, *J*gem = 10.2, *J*5'a,4' = 6.7, H-5'a); 3.719, 3.725 (2 × s, 2 × 3H, CH3O-DMTr); 3.97 (dt, 1H, *J*4',5' = 6.7, 3.5, *J*4',3' = 3.5, H-4'); 4.19 (m, 2H, NC**H2**CH2O); 4.33 (m, 1H, *J*3',2' = 6.4, 4.4, *J*3',OH = 4.5, *J*3',4' = 3.5, H-3'); 4.74 (m, 2H, OC**H2**CH2N); 5.35 (d, 1H, *J*OH,3' = 4.5, OH-3'); 6.21 (dd, 1H, *J*1',2' = 6.8, 6.4, H-1'); 6.80, 6.83 (2 × m, 2 × 2H, H-*m*-C6H4-DMTr); 7.17-7.26 (m, 7H, H-*o*-C6H4-DMTr, H-*m,p*-C6H5-DMTr); 7.34 (m, 2H, H-*o*-C6H5- DMTr); 8.00 (s, 1H, H-8).

13C NMR (125.7 MHz, DMSO-*d*6): 38.91 (CH2-2'); 42.58 (N**C**H2CH2O); 55.16, 55.18 (CH3O-DMTr); 64.47 (CH2-5'); 68.00 (O**C**H2CH2N); 70.88 (CH-3'); 83.53 (CH-1'); 85.59 (C-DMTr); 86.23(CH-4'); 113.20, 113.24 (CH-*m*-C6H4-DMTr); 119.72 (C-5); 126.75 (CH-*p*-C6H5-DMTr); 127.86, 127.87 (CH-*o,m*-C6H5-DMTr); 129.86, 129.90 (CH-*o*-C6H4-DMTr); 135.71, 135.78 (C-*i*-C6H4-DMTr); 137.77 (CH-8); 145.10 (C-*i*- C6H5-DMTr); 148.98 (C-4); 154.81 (C-2); 157.70 (C-6); 158.16, 158.20 (C-*p*-C6H4-DMTr). C33H33O7N4 [M+H]+ calcd 597.2349, found 597.2353

3’-*O*-(2-Cyanoethyl-*N*,*N*-diisopropyl)phosphoramidite-5’-*O*-dimethoxytrityl-1-*N*-6-*O*-ethano-2’- deoxyxanthosine 8

2-Cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (0.33 ml; 1.5 mmol) was added to a solution of **7** (0.3 g; 0.5 mmol mmol) and DIPEA (0.52 ml; 3.0 mmol) in DCM (5 mL). The reaction mixture was stirred 2 h at rt After that, the solution was diluted with ethyl acetate (50 ml) and extracted with saturated solution of sodium hydrogencarbonate (3 x 20 ml). The organic layer was dried over anhydrous sodium sulfate and evaporated. The product was purified by chromatography on silica gel (elution with a gradient of 0–50% acetone in toluene; silica gel was buffered with TEA). Lastly, phosphoramidite was freeze-dried from benzene.

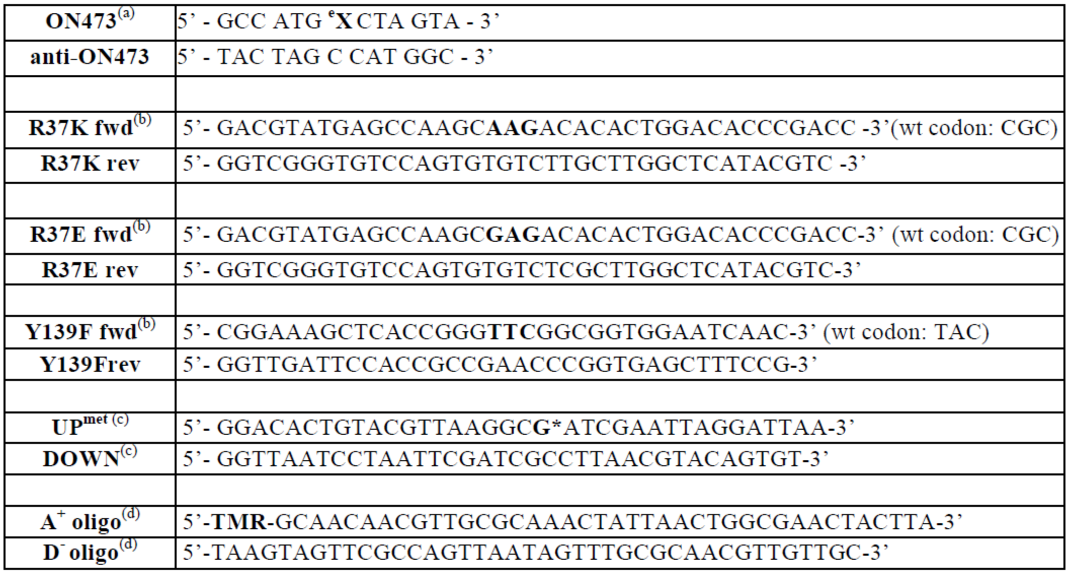
Yield 0.25 g (62%).

31P (162 MHz**;** d6-benzene) 148.92; 148.90.

Synthesis and purification of the ON473 oligonucleotide

Oligonucleotide ON473 (Supplementary Table 1 and Supplementary Figure 2**)** was synthesized as DMT OFF using monomers with base-labile nucleobase protecting groups (PAC) and standard phosphoramidite condensation protocol. Synthesis was performed on a 0.5 μmol scale on 5'-dimethoxytrityl-*N*-benzoyl-2'-deoxyadenosine-3'-succinoyl-LCAA CPG using GenSyn V02 DNA/RNA synthesizer. Modified units eX were also incorporated using standard phosphoramidite condensation protocol. After the solid-phase synthesis, oligonucleotide ON473 was deprotected using gaseous ammonia. The column was inserted into the pressure vessel and treated with gaseous ammonia (0.7 MPa) for 1.5 h to remove acyl protecting groups and to release the final product from the solid support. After that, the deprotected DNA was washed off of the column by a 0.1 M TEAA buffer and purified using ion exchange chromatography. Isolated yield 65 %.

Supplementary Table 1 - Oligonucleotides used in the present study



ON473(a)	5’ - GCC ATG ^{e}**X** CTA GTA - 3’
anti-ON473	5’ - TAC TAG C CAT GGC - 3’
R37K fwd(b)	5’- GACGTATGAGCCAAGC**AAG**ACACACTGGACACCCGACC -3’(wt codon: CGC)
R37K rev	5’- GGTCGGGTGTCCAGTGTGTCTTGCTTGGCTCATACGTC -3’
R37E fwd(b)	5’- GACGTATGAGCCAAGC**GAG**ACACACTGGACACCCGACC-3’ (wt codon: CGC)
R37E rev	5’- GGTCGGGTGTCCAGTGTGTCTCGCTTGGCTCATACGTC-3’
Y139F fwd(b)	5’- CGGAAAGCTCACCGGG**TTC**GGCGGTGGAATCAAC-3’ (wt codon: TAC)
Y139Frev	5’- GGTTGATTCCACCGCCGAACCCGGTGAGCTTTCCG-3’
UP^{met} (c)	5’- GGACACTGTACGTTAAGGC**G***ATCGAATTAGGATTAA-3’
DOWN(c)	5’- GGTTAATCCTAATTCGATCGCCTTAACGTACAGTGT-3’
A^{+} oligo(d)	5’-**TMR-**GCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTA-3’
D^{-} oligo(d)	5’-TAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGC-3’

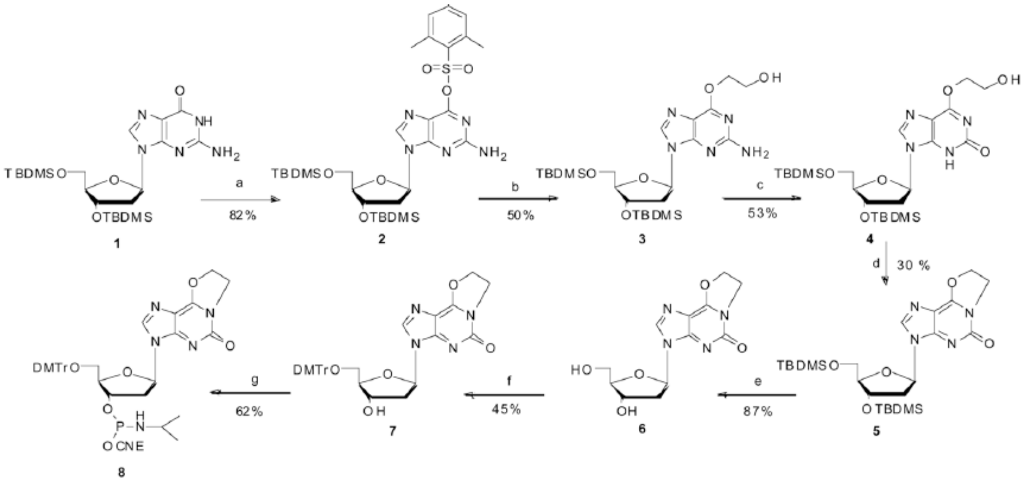
(a) The modified base appears in bold; E1X-dsDNA used in co-crystallization was obtained by annealing the ON473 and anti-ON473 oligonucleotides. (b) Triplets encoding the substituting aminoacid in the “sense” oligonucleotide appears in bold. (c) The ds-DNA^{met} substrate used in VG-based assays was obtained by annealing the UP^{met} (the methylguanine is labeled by an asterisk) and DOWN oligonucleotides. (d) The unmodified dsDNA probe used in EMSA analyses was obtained by annealing the A^{+} (“TMR” indicates the TAMRA^{TM} fluorescent group) and D^{-} oligonucleotides

Supplementary Table 2

	ds-DNA^{met} (μM)	**Mean *k*** (s^{-1}) ± SD	**Mean K_{VG}** (μM) ± SD	**Mean K_{DNA}^{met}** (μM) ± SD
***Mt*OGT-R37K**	0	0.07 ± 0.01	0.42 ± 0.1	**0.38 ± 0.2**
	0.63	0.07 ± 0.01	0.88 ± 0.2	
	1.00	0.07 ± 0.01	0.96 ± 0.5	
	1.25	0.06 ± 0.01	1.43 ± 0.5	
***Mt*OGT-R37E**	0	0.03 ± 0.01	1.72 ± 0.4	**1.14 ± 0.15**
	0.63	0.04 ± 0.01	2.81 ± 0.6	
	1.00	0.05 ± 0.01	3.47 ± 0.2	
	1.25	0.04 ± 0.02	3.61 ± 0.4	
***Mt*OGT-Y139F**	0	0.02 ± 0.02	0.69 ± 0.3	**2.19 ± 0.5**
	0.63	0.03 ± 0.02	0.73 ± 0.4	
	1.00	0.02 ± 0.01	0.93 ± 0.4	

Kinetic constants of the reaction catalysed by the indicated *Mt*OGT variants - The concentration value of dsDNA^{met} refers to the amount of mono-methylated double stranded DNA used in the SNAP Vista-green^{TM} reagent (VG)-based competitive assays. *k* and K_{VG} correspond to the rate of covalent linkage and the dissociation constant for the free enzyme and free VG reagent, respectively. K_{DNA}^{met} is the dissociation constant of the protein for the dsDNA^{met} substrate.

Supplementary Figure 1



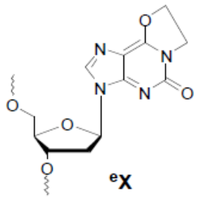
a) MSC, TEA, DMAP, DCM; b) 1. DABCO, THF 2. DBU, ethylenglycol, THF; c) $NaNO_2$, AcOH, H_2O, acetone; d) TsCl, pyridine; e) TBAF, THF; f) DMTrCl, pyridine; g) 2-cyanoethyl N,N-diisopropylchlorophosphoramidite, DIPEA, DCM.

Reaction scheme for the synthesis of *N*^{1}-*O*^{6}-ethano-2’-deoxyxanthosine phosphoramidite monomer (^{e}X)

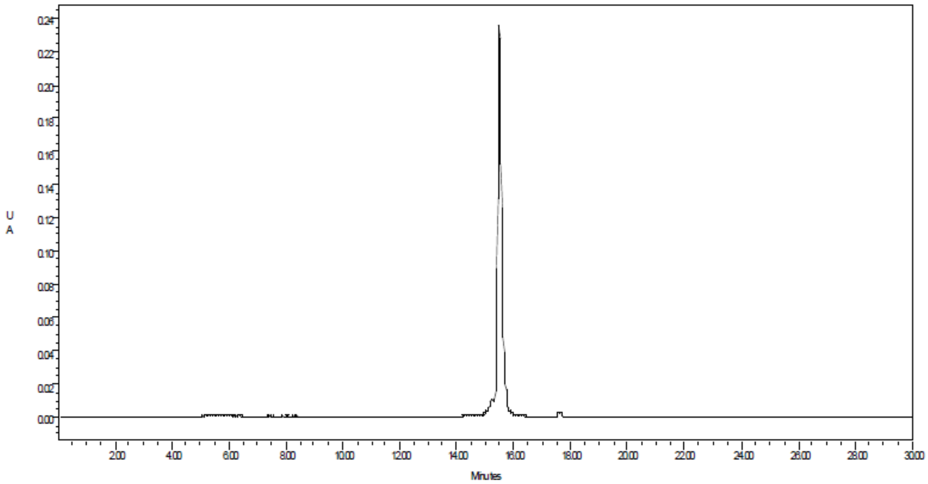
6-*O*-Mesitylenesulfonyl derivative **2** was prepared according to published procedure [Hayakawa, Y., Hirose, M., and Noyori, R. (1993) O-Allyl protection of guanine and thymine residues in oligodeoxyribonucleotides. *J Org Chem* **58**, 5551-5555] in 82% yield. 6-*O*-Ethyleneglycol derivative **3** was prepared via DABCO quarternary intermediate in 50% yield. 6-*O*-hydroxyethylxanthosine derivative **4** was obtained by $NaNO_2$ mediated deamination in 53% yield. Subsequent five-membered ring closure was accomplished by treatment with tosylchloride in pyridine affording derivative **5** in 30% yield. Desilylation (87%) followed by dimethoxytritylation (45%) afforded protected ^{e}X nucleoside **7**. Finally, nucleoside **7** was converted to ^{e}X phosphoramidite monomer employing standard procedure with 62% yield.

Supplementary Figure 2

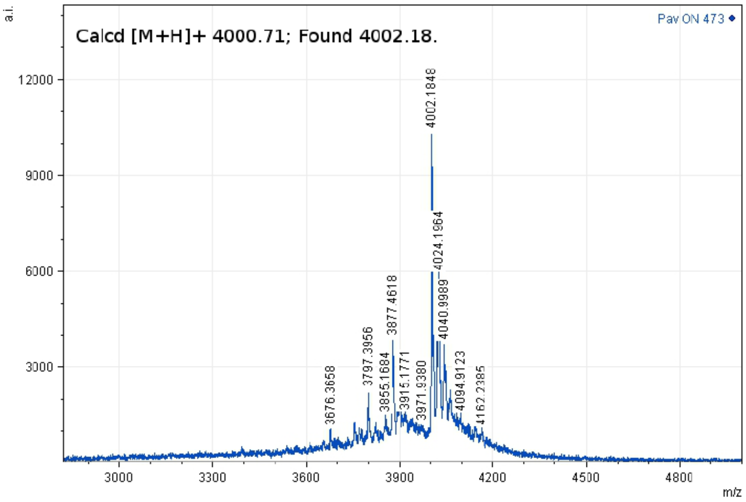
Analyses of the d(GCC ATG eX CTA GTA) oligonucleotide (ON473)



^{e}X

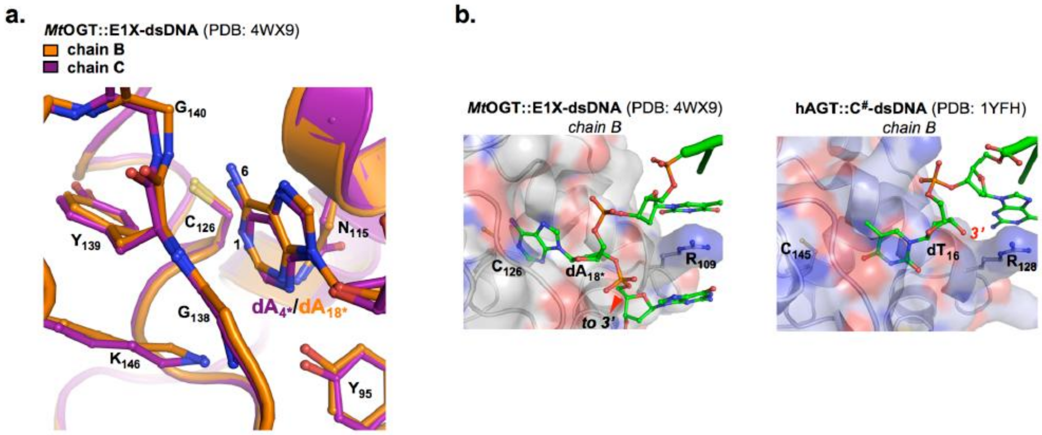


Chromatogram of ON473 after purification using ion exchange chromatography. Gradient 0-30%B30min 55 oC [A = 20mM AcONa, 20mM NaCl, 10% ACN; B = 20mM AcONa, 1.5M NaCl, 10% ACN]



MALDI MS of ON473

Supplementary Figure 3


a.
MtOGT::E1X-dsDNA (PDB: 4WX9)
chain B
chain C
G140
6
C126
Y139
N115
1
G138
dA4*/dA18*
K146
Y95
b.
MtOGT::E1X-dsDNA (PDB: 4WX9)
chain B
C126
dA18*
R109
to 3'
hAGT::C#-dsDNA (PDB: 1YFH)
chain B
C145
dT16
3'
R128

***Mt*OGT is able to host unmodified extra-helical DNA bases in its active site - a.** Close-up of the active site of superposed chains B and C of the *Mt*OGT::E1X-dsDNA structure; the adenine bases are colored applying the same color code used for the corresponding protein chain. **b.** Surface representation of the active site of optimally superimposed *Mt*OGT::E1X-dsDNA (monomer B) and hAGT::C$^{\#}$-dsDNA (monomer B), which are observed to bind an unmodified nitrogenous base in the corresponding crystal structure; as a matter of clarity, images are represented side by side.

Analyses of macromolecular interactions sustaining NER- and OGT-mediated alkylated-DNA repair in *Mycobacterium tuberculosis*

Samarpita Lahiri[a], Riccardo Miggiano[a], Menico Rizzi[a] and Franca Rossi[a].

[a]Dipartimento di Scienze del Farmaco, University of Piemonte Orientale “A. Avogadro”, Novara, Italy.

Unpublished results

BACKGROUND

In order to combat the potential genotoxic and mutagenic effects of DNA-alkylation by endogenous and exogenous compounds, all living species, including *Mycobacterium tuberculosis* (MTB) [1] deploy different strategies, depending upon the entity and the nature of the damage, and the physiologic state of the cell. These repairing mechanisms include i) the removal of a short, lesion-containing oligonucleotide followed by DNA new synthesis, through the action of the Nucleotide Excision Repair (NER) system [2]; and ii) the conversion of alkylated-nucleobases back to their unmodified original state, through a single step reaction catalyzed by an alkylated-DNA protein alkyltransferase protein (referred to as AGT, MGMT or OGT) [3,4].

Bacterial NER is capable to recognize and repair a broad range of damaged-DNA substrates by utilizing three core proteins (UvrA, UvrB and UvrC), which act co-operatively to form the damage detection/verification, pre-incision and incision complexes that are involved in first events of the multi-step repairing cascade [2]. UvrA, the first enzyme acting in bacterial NER, is an ABC ATPase responsible for the lesion sensing activity (which is triggered by DNA substrates characterized by a certain degree of geometry distortion), and the recruitment of UvrB at the damaged

DNA site. Moreover, UvrA is also a key player in the Transcription Coupled NER pathway, by virtue of its association to the Mfd protein, the bacterial functional equivalent of the Transcription Coupled Repair Factor (TCRF) [5].

Recent investigations carried out in our laboratory led to the biochemical characterization of MTB UvrA (*Mt*UvrA) and the X-ray crystallography-based structural description of the protein in its ligand-free state [6]. As observed in other structures of UvrAs reported so far (*e.g.* the ADP-bound UvrA [7] and the UvrA/UvrB complex [8] from *Bacillus stearothermophilus*, as well as the *Thermotoga maritima* ortholog in complex with a modified double-stranded DNA [9]), the protein architecture is organized around a globular catalytic core, hosting two nucleotide binding sites, and exposing the DNA-contacting Zinc-finger on its ventral side (Figure 1).

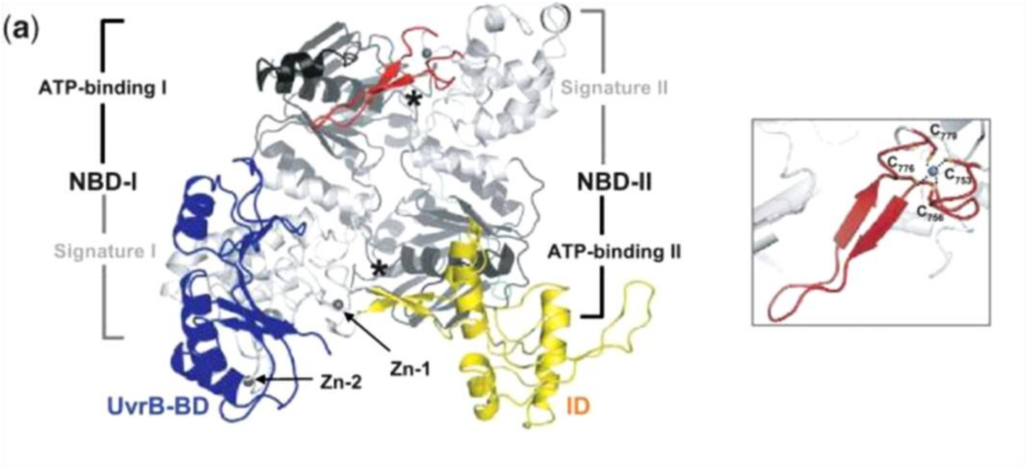


FIGURE 1: (**a**) Cartoon representation of a *Mt*UvrA subunit; the asterisks label the position of the nucleotide binding sites; a close-up view of the C-terminal Zn finger appears in the inset. The UvrB binding domain (UvrB-BD) and the insertion domain (ID) are colored blue and yellow, respectively and the two Zn modules are indicated [modified from 6].

In sharp contrast, the UvrB-binding domain (UvrB-BD) and the insertion domain (ID) (distinguishing UvrAs from the other ABC ATPases family members), are observed in a quite different conformation. In particular, both of the UvrB-BD of the functional *Mt*UvrA dimer appear 90° rotated towards the ventral side of the protein compared to the position occupied by the equivalent region in *B. stearothermophilus* and *T. maritima* UvrA structures (Figure 2).

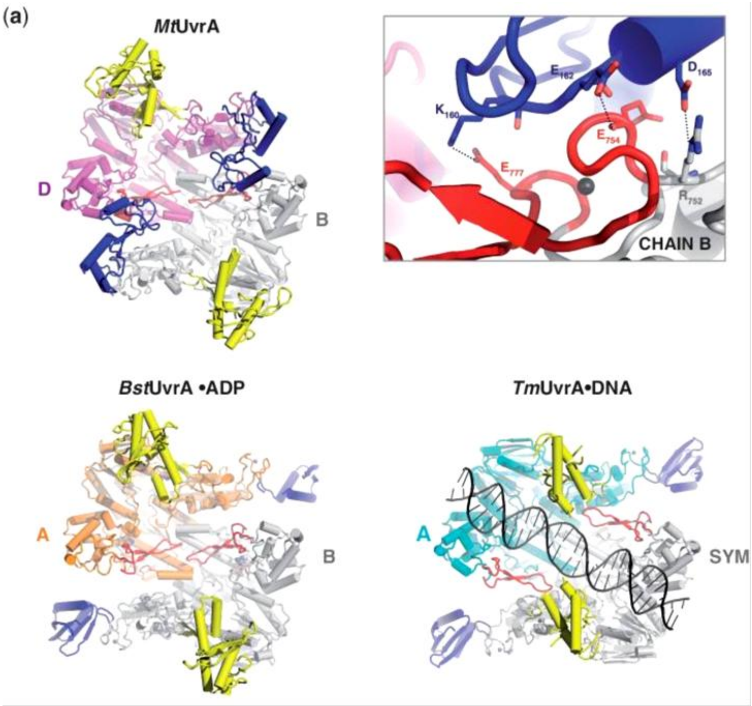


FIGURE 2: *Mt*UvrA dimer organization. (**a**) Comparison between the functional dimeric assemblies of *Mt*UvrA (upper side, left), *Bst*UvrA (lower side, left) and *Tm*UvrA (lower side, right) as viewed from their ventral surface, upon optimal superposition; the two catalytic cores in each dimer are coloured magenta and white in *Mt*UvrA, orange and white in *Bst*UvrA and cyan and white in *Tm*UvrA; the UvrB-BD, the ID and the C-terminal Zn finger are invariably coloured blue, yellow and red, respectively; DNA appears in black. The inset depicts a close-up view of the peculiar interactions observed in the *Mt*UvrA dimer on its ventral side; residues participating in main stabilizing contacts are shown as sticks [6].

In principle, the unprecedented conformation characterizing the MTB UvrA dimer could still allow the simultaneous binding of two UvrB monomers and DNA substrate, although we cannot exclude that these functional complexes could display a molecular architecture or a stoichiometry different from the one reported

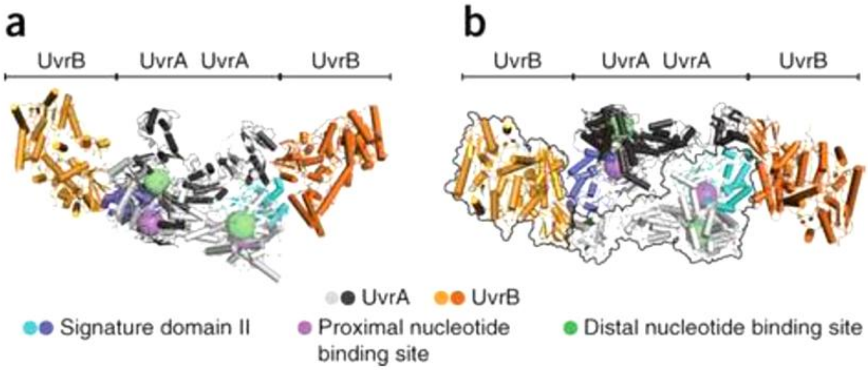

a
b
UvrB
UvrA UvrA
UvrB
UvrA
UvrB
Signature domain II
Proximal nucleotide binding site
Distal nucleotide binding site

FIGURE 3: The AB damage sensor is shown as a cartoon in two different views. The view in **a** is related to that in **b** by a 90° rotation about the horizontal *x* axis. The two protomers of UvrA are shown in different shades of gray, with the signature domain II in cyan and blue. UvrB molecules are shown in different shades of orange. The positions of proximal and distal nucleotide binding sites are denoted by semi-transparent magenta and green spheres, respectively. In panel **b**, the boundaries of UvrA and UvrB are outlined modified from[8].

for the *B. stearothermophilus* $UvrA_2$::$UvrB_2$ assembly (figure 3) [8].

Notably, studies in other bacterial and lower eukaryotic species revealed that alkyltransferase-like proteins (ATL) -which display structural similarity to genuine OGT, but lack the catalytic cysteine and the N-terminal domain-, strongly enhance the repair by NER of mutagenic adducts (including alkylated-guanine), which are otherwise poor substrates of this system [reviewed in 10,11], by directly interacting with UvrA (Figure 4).

Overall these data highlight the importance of analyzing not only individual DNA-repair pathways in a given species, but also their reciprocal coordination when approaching the study of such a complex biochemical problem in a systematic way. These considerations represent the *rationale* that prompted us to undertake studies aimed at characterizing and/or disclosing macromolecular cross-talks between the different machineries controlling DNA-repair in MTB.

As part of the EU 7FP collaborative project “Syste*MTb*” the study of the “interaction network” of selected MTB proteins was performed, including crucial

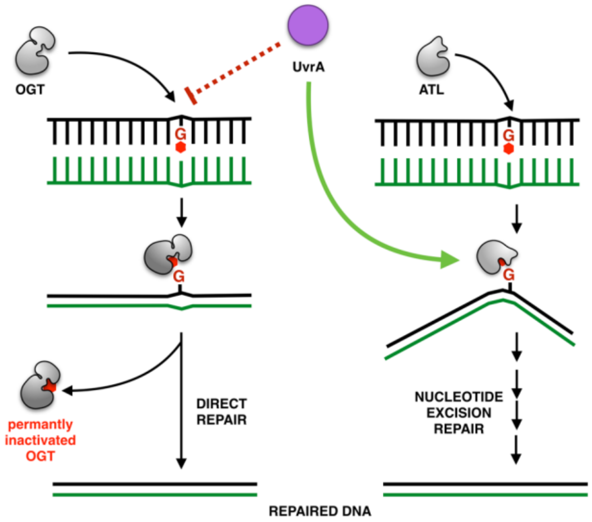


FIGURE 4: OGTs directly repair alkylated bases through a suicidal reaction; ATLs (which are absent in higher eukaryotes) behave as molecular adaptors for UvrA, to allow the NER-dependent repair of non-helix distorting damaged bases.

players in DNA-repair. Preliminary high throughput (HT) interaction analyses based on a liquid chromatography-based single-step affinity purification method followed by high-resolution mass spectrometry (LC-MS/MS), confirmed the capability of *Mt*UvrA to directly associate with UvrB [12], while no interaction was detected between *Mt*UvrA and ATL in the same experimental setting.

Unexpectedly, these analyses disclosed an unprecedented interaction between *Mt*OGT and UvrA (unpublished), signaling that, in MTB, a coordination between alkylated-nucleobase direct repair (by *Mt*OGT) and NER could exist.

Preliminary data obtained from Surface Plasmon Resonance (SPR)-based experiments, using the Biacore^{TM} platform, confirmed the direct association between pure preparations of recombinant *Mt*UvrA -bearing a poly-His tag at its N-terminus [6], and *Mt*OGT (tag-free) [13]. The kinetic analysis of the *Mt*UvrA::*Mt*OGT complex formation allowed the determination of the dissociation constant (KD) of the complex that ranks in the low micromolar range ($1.42*10^{-6}$

M), revealing the transient nature of this interaction. Notably, this value is comparable to that found for the well-described UvrA::UvrB complex from *B. stearothermophilus*.

I describe hereafter the unpublished experiments we performed 1) to start to investigate the biochemical properties and the structural features of the *Mt*UvrA::*Mt*UvrB complex, and 2) to provide further independent confirmation to the direct assembling of the newly identified *Mt*OGT::*Mt*UvrA complex.

MATERIALS AND METHODS

Chemicals

All reagents were obtained from Sigma-Aldrich unless otherwise specified.

Construction of a plasmid driving the over-expression of *Mt*UvrB in *E. coli* cells

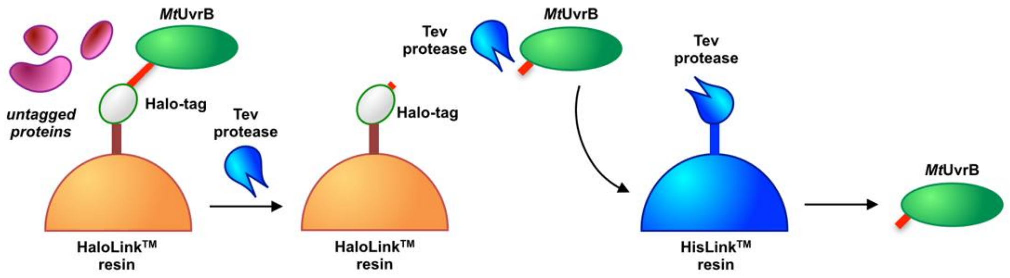
The complete open reading frame encoding UvrB of *Mycobacterium tuberculosis* H37Rv strain (Tuberculist entry: Rv1633) was amplified by PCR using 40 ng of MTCY06H11.01 bacmid DNA (Institut Pasteur, Paris, France), the Hotstar High fidelity polymerase (Qiagen) and the primers: UvrB_for (5'-ATTAGCGATCGCCGTGCGCGCCGGCGGTCAC-3'-, *Sgf* I site is underlined) and UvrB_rev (5'- AATTGTTTAAACTCACTTCAGGCCGGCCGCG- 3', *Pme* I site is underlined). The *Sgf* I/*Pme* I-double digested PCR product was inserted into the pFN18A Halotag T7Flexi vector (Promega), linearized by the same restriction enzymes, adopting standard procedures [14], resulting in the p*Mt*HaloUvrB expression plasmid, whose *Mt*UvrB-encoding region was verified by sequencing (Eurofins MWG Operon).

Expression and purification of *Mt*UvrB

Escherichia coli strain BL21(DE3) competent bacteria were transformed with p*Mt*HaloUvrB expression construct and spread onto 2XTY-agar medium added by ampicillin (50µg/L). Freshly transformed colonies were dispersed in 1 L of 2XTY-ampicillin medium in a 5 L flask, to reach a starting optical density at 600 nm (O.D.$_{600}$) of 0.1, and grown under vigorous shaking till an O.D.$_{600}$ of 0.8 at 37 °C. Upon a cold shock on an ice-water bath for 15 minutes, the bacterial culture was added of 0.1 mM Isopropyl β-D-1-thiogalactopyranoside, brought at 17 °C, and further grown overnight under vigorous shaking. The bacterial pellet obtained (8 g) was resuspended in 60 mL of lysis buffer (50 mM Hepes pH 7.5, 150 mM NaCl, 1 mM DTT) and the cells were disrupted using Ultrasonic processor (Vibra-Cell). The clear supernatant was obtained after a 30 min. centrifugation at 16,000 *g*. 20 mL of a 25% HaloLink^{TM} resin slurry (Promega) (Scheme 1) (corresponding to 5 ml of settled resin) were equilibrated with the same lysis buffer and then incubated with the clear supernatant, under slight rotation for 2 hours. Upon binding, the protein-laden resin was extensively washed in same lysis buffer, and incubated in the presence of the Tev protease (330 unit/mL resin bed volume) to trim the Halo-tag from the recombinant protein, under gentle rotation for 2 hours. Following this step, the resin was packed in a column and fractions containing tag-free *Mt*UvrB were recovered, pooled and incubated with HisLink ^{TM} resin (Promega) (Scheme 1) to remove the Tev protease, following manufacturer's instructions. *Mt*UvrB fractions from this last step were pooled and concentrated up to 12 mg/mL and stored in aliquots at -80 °C, if not immediately used. The expression/purification procedure yields 6 mg of pure tag-free *Mt*UvrB per liter of induced bacterial culture.

Optimization of *Mt*UvrA Expression and purification

The procedure to express and purify *Mt*UvrA was slightly modified from [6]. *E. coli* BL21(DE3) colonies harboring the plasmid p*Mt*HisUvrA were dispersed in 4 L



SCHEME 1: Procedure to obtain tag-free *Mt*UvrB from a clarified lysate of *E. coli* cells transformed with p*Mt*HaloUvrB

of 2XTY-ampicillin medium to obtain a starting O.D.$_{600}$ of 0.15. This culture was further grown at 37 °C under vigorous shaking until an O.D.$_{600}$ of 1.0 was reached. Upon a 15 minutes cold shock and the addition of 0.15 mM IPTG, the bacterial culture was cultivated at 17 °C overnight for the expression of the His-tagged *Mt*UvrA. Upon an immobilized metal affinity chromatography (IMAC) purification step using the Ni-NTA resin (Qiagen) -as described in [6]-, the pure protein sample was loaded onto a column pre-packed with heparin-functionalized resin (HiTrap Heparin HP, GE healthcare) and pre-equilibrated with buffer A (20 mM Tris-HCl pH 8.0). The bound protein was then eluted using linear 0-2 M NaCl gradient. The fractions corresponding to the eluted DNA-free protein, were pooled and concentrated by ultrafiltration (membrane NMWCO=100 kDa, Vivascience) up to 1 mg/mL. Preparative SEC was performed using a Sephacryl S-200 High Resolution 16/60 (GE healthcare), by fluxing buffer B (20 mM Tris-HCl, pH 8.0, 500 mM NaCl) as the mobile phase. The *Mt*UvrA containing fractions were pooled, concentrated up to 5 mg/mL and stored in small aliquots at -80 °C, if not immediately used. The optimized expression/purification procedure yields 0.6 mg of DNA-free His-tagged *Mt*UvrA protein per liter of induced bacterial culture.

Protein co-fractionation experiments using SEC

SEC-based protein co-fractionation experiments were performed using an Akta FPLC instrumentation and the Superdex 200 5/150 GL or Sephacryl S-200 High Resolution 16/60 columns (GE healthcare) -as specified in the figures-, pre-calibrated with standard proteins following manufacturer’s instructions, and using the buffer B (see above) as the mobile phase. Co-elution experiments of *Mt*UvrA and *Mt*OGT in the presence of single stranded DNA (ssDNA) were conducted using the oligonucleotide (5'-TTAAGTTTAAACGAACCCGACCTGCCGC ATCA-3').

RESULTS AND DISCUSSION

High yield expression and purification of a recombinant, tag-free version of *Mt*UvrB

The expression vector p*Mt*HaloUvrB was designed to drive in *E. coli* the over-expression of a full-length *Mt*UvrB recombinant protein, which bears, at its N-terminus, a Halo-tag moiety that could be removed by Tev-mediated proteolysis (Scheme 1).

Several trials, systematically exploring the temperature and time of bacterial growth and induction of the p*Mt*HaloUvrB-transformed *E. coli* BL21(DE3) strain, as well as the concentration of IPTG, were carried out in order to obtain the protein at high yields in the soluble fraction of the bacterial lysate. The best expression conditions in terms of protein yield and solubility correspond to an overnight induction of the protein at 17 °C by adding 0.1 mM IPTG (Figure 5). The purified *Mt*UvrB protein is highly homogeneous, and behaves as a probable dimer in solution (Figure 6), in agreement with published data studying the *E. coli* UvrB [15].

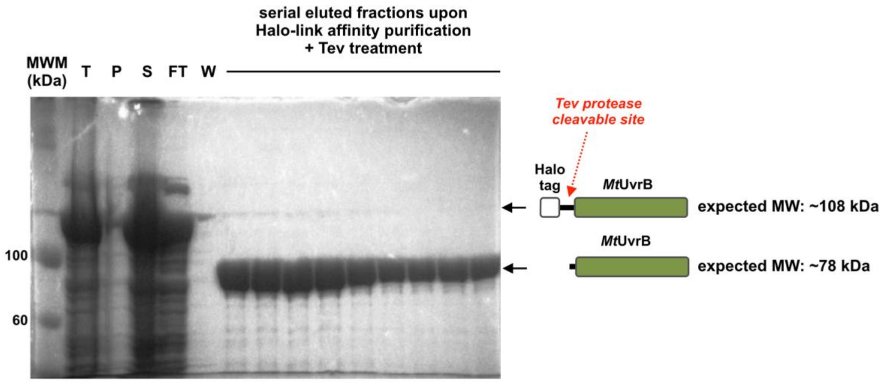

serial eluted fractions upon
Halo-link affinity purification
+ Tev treatment
MWM
(kDa)
T
P
S
FT
W
100
60
Tev protease
cleavable site
Halo
tag
*Mt*UvrB
expected MW: ~108 kDa
*Mt*UvrB
expected MW: ~78 kDa

FIGURE 5: SDS-PAGE analysis, using a 7.5 % acrylamide/bis-acrylamide gel, of the **T**: total lysate of the induced p*Mt*HaloUvrB-transformed bacteria, **P**: insoluble fraction of the lysate upon centrifugation, **S**: soluble fraction of the lysate upon centrifugation. **FT** (flow-through) and **W** (wash) refer to proteins in the **S** fraction that do not bind or do not specifically associate to the HaloLink resin, respectively. **MWM**: molecular weight standards. The gel was stained in Coomassie Brilliant Blue (CBB staining).

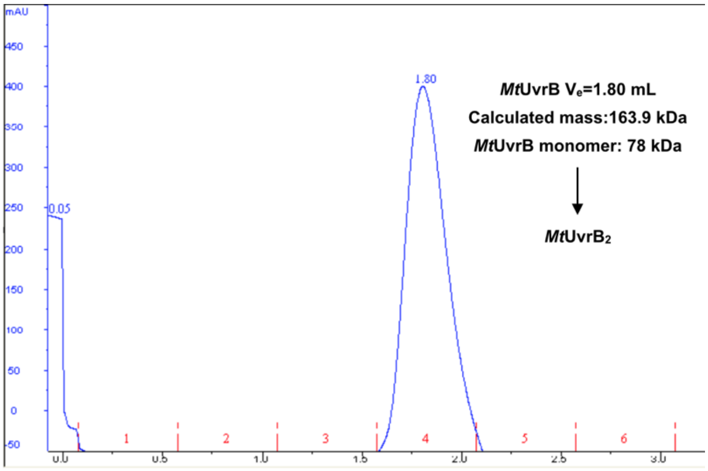


FIGURE 6: SEC analysis of *Mt*UvrB at the end of the purification procedure, using a pre-calibrated Superdex 200 5/150 GL column.

Modified procedure for expression and purification of *Mt*UvrA

The expression conditions to obtain the His-tagged recombinant variant of *Mt*UvrA were modified with respect to the published procedure [6] in order to increase the percentage of the recombinant protein in the soluble fraction of the bacterial lysate after cell disruption.

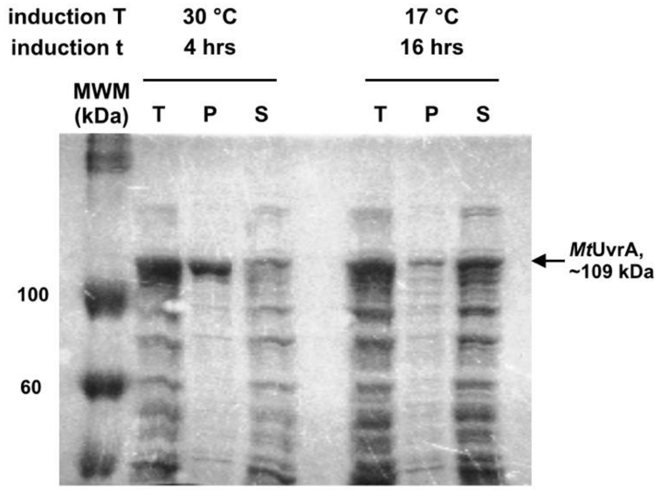

induction T
induction t
30 °C
4 hrs
17 °C
16 hrs
MWM
(kDa)
T
P
S
T
P
S
100
60
*Mt*UvrA,
~109 kDa

FIGURE 7: SDS-PAGE analysis (7.5 % acrylamide/bis-acrylamide gel, CBB staining), of the **T**: total lysate, **P**: insoluble fraction of the lysate upon centrifugation, and **S**: soluble fraction of the lysate upon centrifugation, of p*Mt*UvrA-transformed bacteria induced (by the addition of 0.15 mM IPTG) at the temperature and for the time indicated. **MWM**: molecular weight standards.

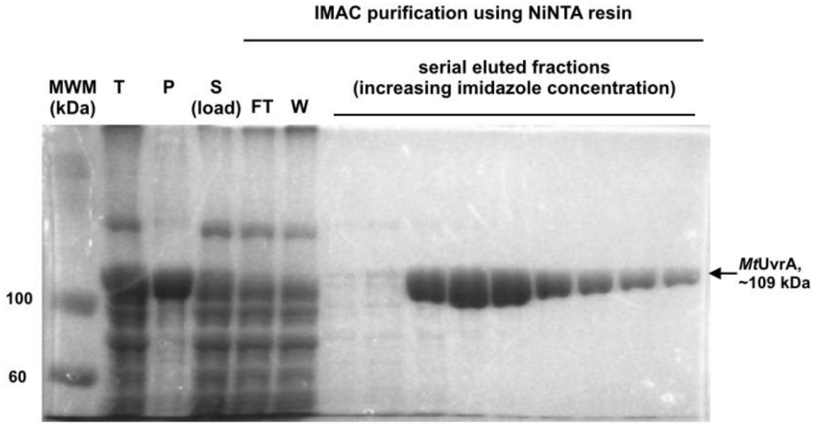

IMAC purification using NiNTA resin
serial eluted fractions
(increasing imidazole concentration)
MWM
(kDa)
T
P
S
(load)
FT
W
100
60
*Mt*UvrA,
~109 kDa

FIGURE 8: SDS-PAGE analysis (7.5 % acrylamide/bis-acrylamide gel, CBB staining), of the **T**: total lysate of the p*Mt*UvrA-transformed bacteria induced to express the *Mt*UvrA protein by following the modified procedure, **P**: insoluble fraction of the lysate upon centrifugation, **S**: soluble fraction of the lysate upon centrifugation. **FT** (flow-through) and **W** (wash) refer to proteins in the loaded sample that do not bind or do not specifically associate to the Ni-NTA resin, respectively. **MWM**: molecular weight standards.

Indeed, at each step of the purification protocol a large proportion of the protein is lost; therefore, by augmenting the amount of the starting soluble protein we were able to add two further FPLC-based purification steps, namely a pseudo-affinity chromatography using a heparin-functionalized resin, followed by SEC. According to the new protocol, *E. coli* BL21(DE3) cells harboring the plasmid p*Mt*HisUvrA were grown overnight at 17 °C after protein induction instead of 4 hours at 30 °C as previously described [6]. As shown in Figure 7, the percentage of the protein in the soluble fraction upon centrifugation of the disrupted bacteria is increased, without affecting the behavior of the protein during Ni-NTA affinity purification (Figure 8).

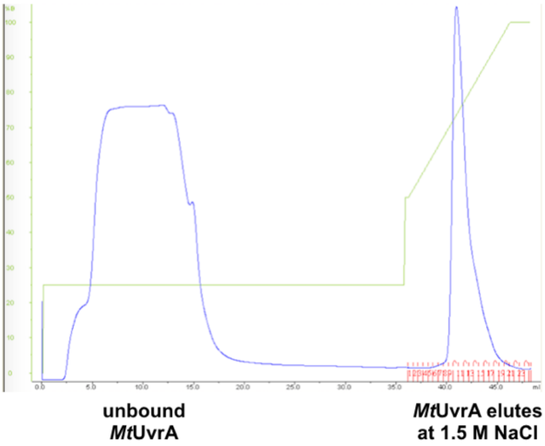


FIGURE 9: A. chromatogram showing the *Mt*UvrA elution in a sharp peak from the HiTrap Heparin HP column, at 2/3 of the linear salt gradient, corresponding to a 1.5 M NaCl molar concentration. The blue trace corresponds to the reading of the absorbance at 280 nm; the green graph follow the NaCl gradient.

We previously observed that *Mt*UvrA tends to be contaminated by small traces of damaged/fragmented genomic DNA produced during cell disruption; this phenomenon, while not affecting the “crystallizability” of the protein [6], could interfere with protein-protein and protein-DNA studies. The pure *Mt*UvrA protein obtained from the IMAC step (Figure 8), was therefore subjected to a further chromatographic purification using a heparin-functionalized resin. In this case the *Mt*UvrA protein elutes in a DNA-free form at a high salt concentration (1.5 M NaCl), where the high ionic strength of the elution buffer could exert a stabilizing effect upon the the ligand-free protein (Figure 9).

In order to reach a salt concentration in solution compatible with protein-protein interaction studies, *Mt*UvrA was subjected to SEC in the presence of fixed decreasing concentrations of NaCl, in order to reveal the minimum salt concentration required to maintain the solubility of the protein. These analyses revealed that below a 500 mM NaCl concentration in the SEC mobile phase *Mt*UvrA tends to aggregate and precipitate either during the injection or into the column (not shown). Therefore, a buffer composed of 20 mM Tris-HCl pH 8.0 and 500 mM NaCl was selected as the mobile phase during SEC-based interaction studies. The prevailing oligomeric state of the *Mt*UvrA protein, purified following the optimized protocol, has been confirmed to be a dimer (Figure 10) [6].

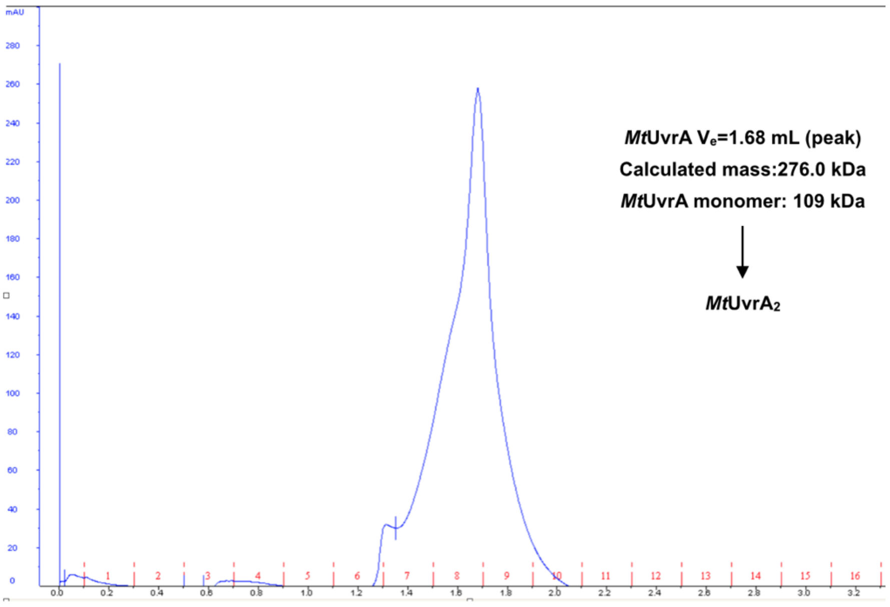


FIGURE 10: SEC analysis of *Mt*UvrA at the end of the optimized purification procedure, using a pre-calibrated Superdex 200 5/150 GL column.

***Mt*UvrA$_2$::*Mt*UvrB$_2$ complexes formation in solution in the absence of DNA substrate**

A huge number of most recent studies in other bacterial species, using a wide array of technologies ranging from X-ray crystallography to cutting-edge single molecule imaging techniques [16], shed light on the structural details and the dynamics of macromolecular interactions supporting the first UvrABC endonuclease-mediated events of the NER cascade.

In particular, the stoichiometry of the damage tracking assembly has recently been confirmed through the analysis of the crystal structure of the *B. stearothermophilus* UvrA$_2$::UvrB$_2$ complex [8].

In order to disclose the capability of *Mt*UvrA to *Mt*UvrB to associate in solution in the absence of any DNA substrate (as suggested by the the LC-MS/MS-based HT screening [12]) we used the homogenous preparations of the two proteins in SEC-based co-fractionation experiments on small-scale (Figure 11A).

The recombinant versions of *Mt*UvrA and *Mt*UvrB co-elute from the column in a single peak, showing a V_e = 1.53 mL. This value corresponds to a 527 kDa molecular mass of the eluted material, which could in principle be contributed by more than one dimer of each protein. We subsequently performed a SEC experiment exploiting a Sephacryl S-200 High Resolution 16/60 column (Figure 11B) In this case the calculated molecular mass of the eluted material suggests the formation of a *Mt*UvrA$_2$::*Mt*UvrB$_2$ complex in solution**.**

SEC-based analyses of *Mt*OGT/*Mt*UvrA association

To confirm the direct interaction between *Mt*OGT and *Mt*UvrA, which was disclosed by the HT LC-MS/MS experiments and confirmed by SPR-based assays (not shown), we performed the same set of SEC-based analyses, using a mixture of the two pure proteins at a 2:1 molar ratio. The resulting data (Figure 12A) cannot be unambiguously interpreted; in facts, although the chromatogram show two partially overlapping absorption peaks (peaks 1 and 2 in Figure 12A), the SDS-PAGE analysis of the eluted fractions revealed the presence of a significant amount of *Mt*UvrA in the second peak that corresponds to a V_e=2.47 mL, and to a calculated molecular mass of less than 10 kDa, even lower than the one expected for a single *Mt*OGT monomer.

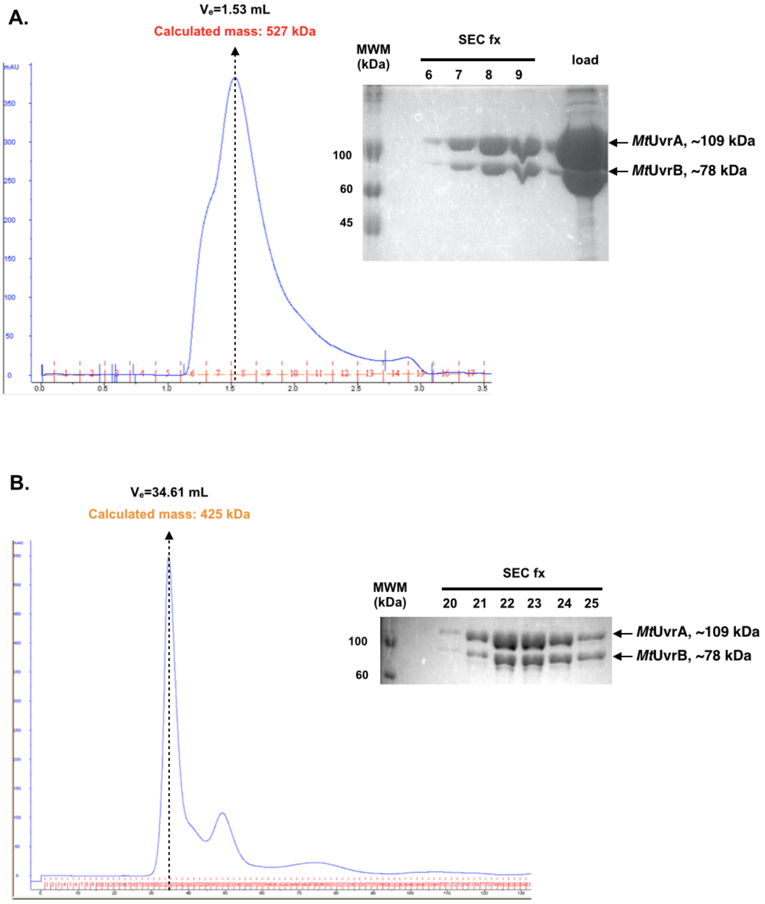


FIGURE 11: SEC analysis of the *Mt*UvrA/MtUvrB direct association in solution, using a pre-calibrated **A.** Superdex 200 5/150 GL column, and **B.** Sephacryl S-200 High Resolution 16/60 column. In both panels the inset shows the SDS-PAGE analysis of the main peak-corresponding eluted fractions (**fx #**) using 7.5% acrylamide/bisacrylamide gel stained by coomassie brilliant blue (CBB).**load:** *Mt*UvrA+*Mt*UvrB sample before SEC-fractionation; **MWM**: molecular weight standards.

.

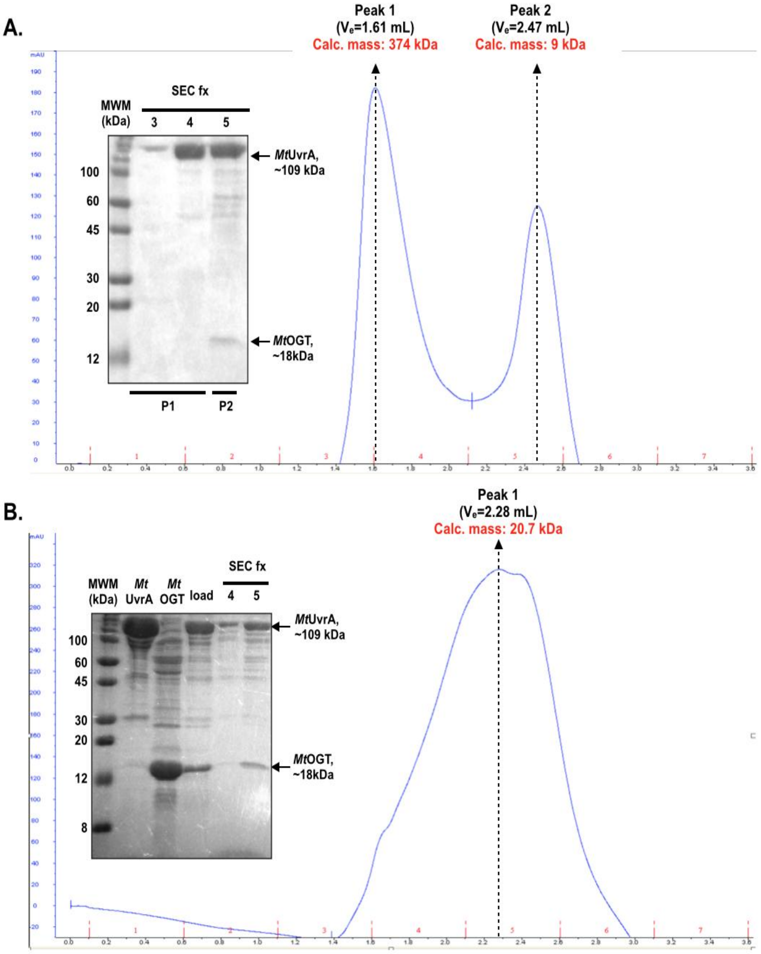


FIGURE 12: SEC analysis of **A.** a 2:1 *Mt*OGT:*Mt*UvrA mixture; **B.** a 1:1.4:2 *Mt*UvrA:ssDNA:*Mt*OGT mixture, using a using a pre-calibrated Superdex 200 5/150 GL column. In both panels the inset shows the SDS-PAGE analysis (15% acrylamide/bis-acrylamide gel, CBB staining) of the eluted fractions (**fx #**); **MWM**: molecular weight standards. In **B.** aliquots of the individual proteins (***Mt*UvrA, *Mt*OGT**) are also shown; **load:** *Mt*UvrA+*Mt*OGT+ssDNA sample before SEC-fractionation.

We also preincubate *Mt*UvrA in the presence of a potential ssDNA substrate, followed by the addition of *Mt*OGT, prior to performing the SEC analysis described above (Figure 12B). However, also in this case, the three components of the injected mixture (namely *Mt*UvrA, ssDNA and *Mt*OGT) co-elute in a broad peak, of very poor quality, and the analysis of the corresponding chromatogram did not allow us to make speculations on the oligomerization state of the complex. Similar contradictory data were obtained using the Sephacryl S-200 High Resolution 16/60 column (Figure 13).

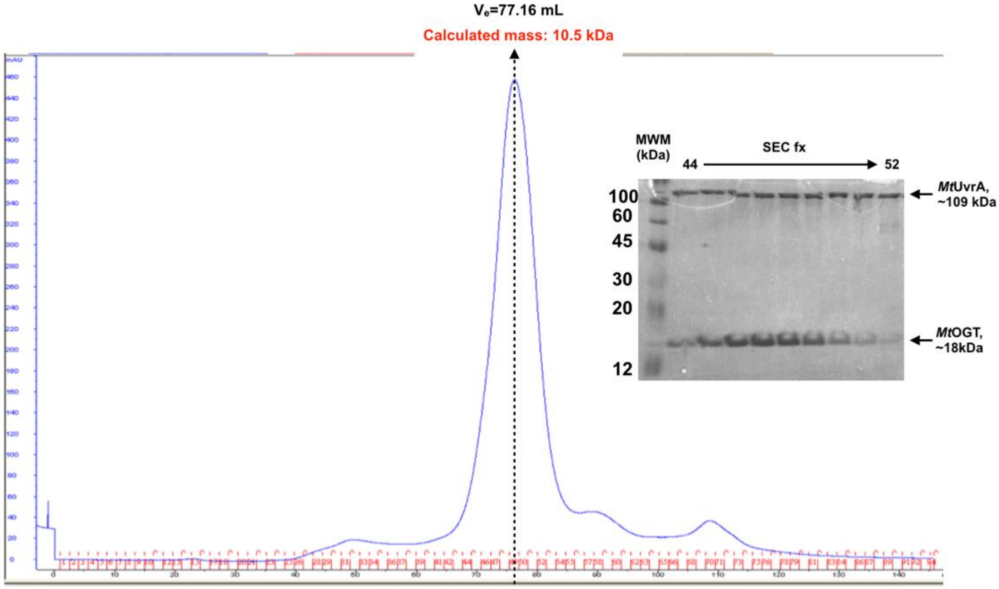


FIGURE 13: SEC analysis of a 1:1.4:2 *Mt*UvrA:ssDNA: *Mt*OGT mixture, using a using a pre-calibrated Sephacryl S-200 High Resolution 16/60 column. The inset shows the SDS-PAGE analysis (15% acrylamide/bis-acrylamide gel, CBB staining) of the eluted fractions corresponding to the main peak (**fx #**); **MWM**: molecular weight standards.

Further analyses, exploiting different techniques characterized by a higher resolution potential (*e.g.* analytical ultracentrifugation), will be performed to confirm the subunit composition of the *Mt*UvrA/*Mt*UvrB-containing assemblies (also in the presence of known ligands of both proteins, such as ATP/ADP or short DNA molecules), as well as to obtain additional information about the unprecedented *Mt*OGT/*Mt*UvrA association.

REFERENCES

1 Gorna, A., Bowater, R., & Dziadek, J. (2010). DNA repair systems and the pathogenesis of Mycobacterium tuberculosis: varying activities at different stages of infection. *Clinical Science*, **119**, 187-202.

2 Truglio, J. J., Croteau, D. L., Van Houten, B., & Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical reviews*, **106**, 233-252.

3 Mishina, Y., Duguid, E.M., & He, C. (2006). Direct Reversal of DNA Alkylation Damage. *Chemical reviews*, **106**, 215-232.

4 Pegg, A.E. (2011). Multifaceted Roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy and research tools. *Chemical research in toxicology*, **24**, 618-639.

5 Savery, N.J.(2007). The molecular mechanism of transcription-coupled DNA repair. *Trends in Microbiology*, **15**, 326-333

6 Rossi, F., Khanduja, J. S., Bortoluzzi, A., Houghton, J., Sander, P., Güthlein, C. & Rizzi, M. (2011). The biological and structural characterization of Mycobacterium tuberculosis UvrA provides novel insights into its mechanism of action. *Nucleic acids research*, **39**, 7316-7328.

7 Pakotiprapha, D., Inuzuka, Y., Bowman, B. R., Moolenaar, G. F., Goosen, N., Jeruzalmi, D., & Verdine, G. L. (2008). Crystal structure of Bacillus stearothermophilus UvrA provides insight into ATP-modulated dimerization, UvrB interaction, and DNA binding. *Molecular cell*, **29**, 122-133.

8 Pakotiprapha, D., Samuels, M., Shen, K., Hu, J. H., & Jeruzalmi, D. (2012). Structure and mechanism of the UvrA–UvrB DNA damage sensor. *Nature structural & molecular biology*, **19**, 291-298.

9 Jaciuk, M., Nowak, E., Skowronek, K., Tańska, A., & Nowotny, M. (2011). Structure of UvrA nucleotide excision repair protein in complex with modified DNA. *Nature structural & molecular biology*, **18**, 191-197.

10 Tubbs, J. L., & Tainer, J. A. (2010). Alkyltransferase-like proteins: molecular switches between DNA repair pathways. *Cellular and molecular life sciences*, **67**, 3749-3762.

11 Tubbs, J.L., Latypov, V., Kanugula, S., Butt, A., Melikishvili, M., Kraehenbuehl, R., Fleck, O., Marriott, A., Watson, A.J., Verbeek ,B., McGown, G., Thorncroft, M., Santibanez-Koref, M.F., Millington, C., Arvai, A.S., Kroeger, M.D., Peterson, L.A., Williams, D.M., Fried, M.G., Margison, G.P., Pegg, A.E. & Tainer, J.A. (2009). Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature*, **459**, 808-813

12 Płociński, P., Laubitz, D., Cysewski, D., Stoduś, K., Kowalska, K., & Dziembowski, A. (2014). Identification of protein partners in mycobacteria using a single-step affinity purification method. *PloS one*, 9 (3), e91380

13 Miggiano, R., 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. *Journal of bacteriology*, **195**, 2728-2736.

14 Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning: a laboratory manual.* (Vol. 2, pp. 14-9). New York: Cold Spring Harbor Laboratory press.

15 Hildebrand, E.L. & Grossman, L. (1999). Oligomerization of the UvrB nucleotide excision repair protein of *Escherichia coli*. *Journal of Biological Chemistry*, **274**, 27885–27890.

16 Mueller, M., Jenni, S. and Ban, N., (2007). Strategies for crystallization and structure determination of very large macromolecular assemblies. *Current opinion in structural biology*, **17**, 572-579.

Concluding Remarks

MTB lives and survives inside the human macrophages, conserving an overall high level of genetic stability despite the continuous exposure to host-generated DNA damaging assaults. This observation indirectly signals that efficient DNA-repairing functions are present in MTB. Most of these mechanisms require a direct and dynamic association between multiple proteins; therefore, studies aimed at characterizing these intra- and inter-pathway macromolecular interactions could represent the basis to the description of mycobacterial DNA repair as a whole, and in the broader context of MTB biology, also including potential cross-talks with the host cell in the course of TB pathogenesis.

The research presented in this Thesis already allowed us to gain new insights into biochemical and structural aspects of single proteins involved in MTB alkylated-DNA repair (the suicidal *Mt*OGT protein and the NER components *Mt*UvrA and *Mt*UvrB) and of their reciprocal association. In order to obtain a detailed description of the association dynamics and the oligomerization state of the macromolecular complexes revealed by these studies, also in the presence of different DNA substrates and/or specific small-molecule effectors, we planned to include in our future analyses experiments based on alternative bio-physical technique, such as bio-layer interferometry. Moreover, the design and/or the optimization of reproducible procedures to express in *E. coli* and purify to the homogeneity and at high yield key components of the MTB DNA-repair toolkit, as described in this Thesis, will speed up their further structural characterization. In this context, we are aware that obtaining crystals suitable for X-rays crystallography analysis of huge, multi-domain, flexible proteins and/or macromolecular complexes is challenging; however, different actions could be implemented to overcome this possible obstacle. First we could try crystallization of the proteins in the presence of different substrates and/or stabilizing chemical species, as they will be revealed by thermal shift analysis. Moreover, preliminary low resolution structural characterization of selected protein-protein and protein-DNA complexes could be carried out in solution by Small-Angle X-ray Scattering (SAXS) experiments using synchrotron radiation. Finally, due to the large size of the expected macromolecular complexes, alternative molecular imaging techniques (such as Atomic Force Microscopy or Single Molecule Electron Microscopy) will be considered.

List of publications

1. Miggiano, R., Perugino, G., Ciaramella, M., Serpe, M., Rejman, D., Páv, O., Pohl, R., Garavaglia, S., **Lahiri, S**., Rizzi, M. and Rossi, F., (2016). Crystal structure of *Mycobacterium tuberculosis* O^6-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA. *Biochemical Journal*, **473**, 123-133
2. Perugino, G., Miggiano, R., Serpe, M., Vettone, A., Valenti, A., **Lahiri, S**., Rossi, F., Rossi, M., Rizzi, M. and Ciaramella, M., (2015). Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic Acids Research*,**43**, 8801-8816.
3. McVey, C.E., Ferreira, M.J., Correia, B., **Lahiri, S**., de Sanctis, D., Carrondo, M.A., Lindley, P.F., de Sá Nogueira, I., Soares, C.M. and Bento, I., (2014). The importance of the Abn2 calcium cluster in the endo-1, 5-arabinanase activity from *Bacillus subtilis*. *JBIC Journal of Biological Inorganic Chemistry*, **19**, 505-513

Acknowledgments

I express my heartiest gratitude and indebtedness to Professor Menico Rizzi for giving me the opportunity to pursue my doctoral degree and to have the great experience through-out these years.

I am obliged to my PhD coordinator, Professor Luigi Panza, for motivating and guiding me all these years.

I would like to express my sincere thanks and gratitude to my PhD supervisor, Dr. Franca Rossi, for her continuous inspiration, precious guidance and helpful advice without which, the completion of this Thesis would not have been possible.

I would also like to thank Dr. Silvia Garavaglia for all her useful suggestions related to protein crystallographic programs.

A special thanks to my colleague, Riccardo for providing me scientific ideas as well as support through-out and huge thanks to all other lab-mates... Davide, Stefano, Andrea, Valentina and Serena. I am very grateful for all the help I have received over the years and I am glad that I got to know all of you.

Thanks to our collaborator (Maria Ciaramella’s Group at IBBR-CNR, Napoli, Italy).

I would like to take the opportunity to express my gratitude to Dr. Philip Jackson, Dr. Isabel Bento, Dr. Sanjit Dey and Prof. Samrat Mukherjee who did guide me in important phases of my student life paving the way towards this day.

Special thanks goes to my family for their all-time support and inspiration (my parents, my grandparents, parents-in-law, and all my cousins).

Thanks to all my friends for sharing with me moments of joy and support.