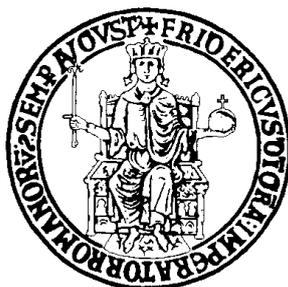


UNIVERSITY OF NAPLES “FEDERICO II”



DOCTORATE SCHOOL IN BIOLOGY

Cycle XXX

In vitro and *in vivo* studies of a DNA repair protein upon DNA alkylation damage.

Coordinator

Prof. Salvatore Cozzolino

Ph.D. student

Mario Serpe

Tutor

Prof. Ezio Ricca

Co-Tutor

Dr. Giuseppe Perugino

Academic Year 2016 - 2017

Riassunto Tesi

I danni da alchilazione a carico del DNA sono potenzialmente pericolosi per l'integrità e la stabilità del genoma. In particolare, l'alchilazione a carico dell'ossigeno in posizione 6 della guanina, se non riparata, porta alla transizione nel DNA dalla coppia G:C alla coppia A:T durante la fase di replicazione [Pegg *et al.*, 1991]. A tale scopo, l'evoluzione ha selezionato un enzima capace di riparare direttamente questo tipo di danno: l'*O*⁶-alchilguanina-DNA-alchiltransferasi (AGT, OGT, MGMT). Questa proteina ubiquitaria trasferisce irreversibilmente il gruppo alchilico dal DNA a una cisteina catalitica: la proteina alchilata è quindi riconosciuta non più funzionante dai sistemi di degradazione cellulari [Pegg, 2011].

Sebbene di ridotte dimensioni, le AGT sono costituite da due domini collegati da una regione non strutturata (*connecting loop*): analisi di struttura primaria e terziaria delle AGT da vari organismi mostrano una elevata identità del dominio C-terminale, coinvolto nel legame al DNA e nella reazione di catalisi, laddove il dominio N-terminale, meno conservato, sembrerebbe coinvolto nella regolazione, nella cooperatività di legame al DNA e nella generale stabilità della proteina. La più studiata in letteratura è quella umana (hAGT), perché la sua inaspettata over-espressione in alcuni tipi di neoplasie porta alla resistenza al trattamento chemioterapico a base di agenti alchilanti. Dopo numerosi studi è stato identificato uno degli inibitori più potenti: l'*O*⁶-benzil-guanina (*O*⁶BG) e suoi derivati, che, usato in combinazione con i trattamenti chemioterapici, ne ripristinano l'efficacia. Nonostante l'enorme caratterizzazione biochimica e strutturale su questa classe di proteine, alcuni dettagli dell'attività delle AGT non sono ancora chiari. In particolare, si suppone che l'alchilazione causi dei cambiamenti conformazionali che rendono le AGT instabili e identificabili dai sistemi di degradazione [Kanugula *et al.*, 1998]. Tuttavia, tali cambiamenti non sono stati così evidenti da essere identificati nelle strutture 3D della hAGT nella forma alchilata [Daniels *et al.*, 2000; Daniels *et al.*, 2004].

In questo studio, abbiamo rivolto la nostra attenzione all'OGT dell'arceobatterio ipertermoacidofilo *S. solfataricus* (*SsOGT*). Quest'organismo, vivendo a una temperatura intorno ai 75-80 °C e a estremi di pH pari a 2-4, è un utile modello per lo studio dei sistemi di riparo del DNA. Come atteso, la *SsOGT* si è mostrata molto stabile in svariate condizioni di reazione, rendendo possibile un'ampia caratterizzazione biochimica, attraverso un nuovo e semplice saggio di attività che utilizza derivati fluorescenti dell'inibitore *O*⁶BG [Perugino *et al.*, 2012]. La proteina, essendo irreversibilmente modificata in seguito alla reazione 1:1 col substrato, viene marcata con un gruppo fluorescente, e l'intensità può essere misurata con i metodi convenzionali di *gel imaging*. Questo nuovo metodo ha permesso di caratterizzare *in toto* questa OGT, evidenziandone la stabilità in svariate condizioni chimico-fisiche non permissive per la maggior parte delle proteine mesofile.

Attraverso esperimenti di termostabilità sulle proteine purificate, abbiamo potuto costatare e confermare una maggiore instabilità della forma metilata rispetto a quella libera alle temperature fisiologiche dell'organismo d'origine (> 70 °C), esattamente come avviene per tutte le AGT. Inoltre, mutanti della cisteina catalitica che "simulano" trasferimenti di gruppi alchilici ingombranti al sito attivo (C119L e C119F) sono risultanti ancor più instabili, suggerendo che l'instabilità generale della proteina è direttamente proporzionale all'ingombro sterico dei gruppi alchilici nel sito attivo. Tuttavia, la *SsOGT* è più stabile e versatile dell'omologo mesofilo umano alle temperature di quest'ultimo, rendendo possibile cristallizzarla nelle varie forme (*SsOGT* libera, in complesso con il DNA, e alchilata) e quindi poter analizzare gli eventuali cambiamenti conformazionali ipotizzati.

Dalla sovrapposizione della struttura della *SsOGT* in forma libera con quella metilata sono stati evidenziati due cambiamenti conformazionali importanti che coinvolgono soprattutto il C-terminale. Il primo coinvolge un'interazione tra un'arginina in posizione 133 (R133) sul dominio C-

terminale e un residuo di aspartato in posizione 27 (D27) sul dominio N-terminale: nella forma metilata la R133 si allontana sensibilmente dal D27. Siccome la distanza nella forma libera tra questi due residui è compatibile con un'interazione ionica, abbiamo ipotizzando che la perdita di questa interazione possa compromettere la stabilità generale della proteina. Attraverso la mutagenesi sito-diretta, abbiamo sostituito il residuo D27 con un'alanina o una lisina, in modo da eliminare la carica negativa o introdurre una positiva, stabilendo un'interazione di tipo repulsivo. Questi mutanti si sono rivelati estremamente instabili, confermando il ruolo di questa interazione nella stabilità generale della proteina. Inoltre, i saggi d'attività condotti su questi mutanti hanno chiarito il ruolo dell'interazione D27-R133 anche nella corretta comunicazione tra i domini durante il processo di riparazione [Perugino *et al.*, 2015]. Il secondo cambiamento conformazionale identificato è a carico del cosiddetto "K48 network", che coinvolge una rete di 5 residui amminoacidici (E44, K48, N59, R61 e E62) con al centro la lisina in posizione 48: dal confronto delle strutture tridimensionali delle tre forme della proteina (libera, legata al DNA e metilata) è stato osservato che questo *network* perde la propria interazione nelle forme legata al DNA e nella forma metilata. L'analisi biochimica ha dimostrato che la perturbazione del *K48 network*, tramite mutagenesi sito diretta, riguarda principalmente la stabilità della SsOGT, e non la sua attività di riparazione del DNA [Morrone *et al.*, 2017]. Questi risultati permettono di proporre un modello che mette in relazione i cambiamenti conformazionali e l'attività di riparazione della SsOGT sul DNA alchilato. L'attività e la stabilità ottimali della SsOGT richiedono un preciso coordinamento tra i due domini; infatti, numerose interazioni intramolecolari si trovano proprio all'interfaccia tra questi. Le due interazioni scoperte giocano un ruolo importante nella stabilità, e nell'attività catalitica. Queste interazioni agiscono letteralmente come "serrature" presenti in uno stato chiuso nella forma libera: se entrambe le serrature si trovano nello stato "aperto" come nel caso della forma alchilata, si ha la perdita totale d'interazione tra i domini N e C-terminale, innescando la destabilizzazione e la degradazione della SsOGT [Morrone *et al.*, 2016].

La reazione irreversibile dell'AGT con l'inibitore O^6 BG porta al legame covalente del gruppo benzilico al sito attivo della proteina, liberando la guanina. Sfruttando questa peculiare reazione, nel 2003 il gruppo del Prof. Johnsson ha sviluppato una versione dell'hAGT (definita *SNAP-tag*) incapace di legare il DNA, per utilizzarla come *protein-tag* nelle proteine di fusione per la marcatura covalente e specifica di proteine [Juillerat *et al.*, 2003]. Infatti, coniugando gruppi chimici d'interesse alla porzione benzilica dell' O^6 BG, è possibile marcare indirettamente una proteina d'interesse con quel gruppo chimico, sfruttando l'attività della *SNAP-tag*. Ad esempio, uno di questi derivati è la *SNAP-Vista Green*TM, in cui al gruppo benzilico dell' O^6 BG è stata coniugata la fluoresceina. Tuttavia, la *SNAP-tag technology* presenta una limitazione: essendo derivata dall'hAGT mesofila, molte condizioni estreme di reazione sono poco applicabili. In questo contesto, si può proporre un mutante della SsOGT (SsOGT-H⁵) che, sebbene simile alla *SNAP-tag* in termini d'attività e incapace di legare il DNA, tuttavia risulta simile alla relativa proteina WT, riguardo la stabilità agli agenti denaturanti chimici e fisici. Per valutare l'espressione eterologa del gene H⁵ in condizioni di alta temperatura, è stato scelto l'organismo modello termofilo *Thermus thermophilus* HB27, privo di alcuna attività AGT endogena: dopo SDS-PAGE degli estratti cellulari e il saggio d'attività, la banda fluorescente corrispondente all'altezza attesa era presente solo nelle cellule trasformate con il plasmide recante il gene di H⁵, dimostrando che la proteina è espressa eterologamente e funzionalmente attiva ad alte temperature [Vettone *et al.*, 2016]. Il mutante H⁵ è stato valutato anche come *protein-tag*, fondendo il gene di H⁵ a quello della termostabile beta-glicosidasi di *S. solfataricus* (Ss-βGly). Il saggio fluorescente per H⁵ e il calcolo delle costanti cinetiche dell'attività enzimatica della Ss-βGly chimerica comparabili a quelli della sola Ss-βGly, dimostrano che ciascuna proteina della fusione è attiva e non interferisce in alcun modo con l'attività dell'altra. Ma il principale vantaggio di questo *tag* termostabile consiste nella purificazione attraverso un trattamento al calore dell'estratto, che permette di eliminare così buona parte delle proteine contaminanti di *E. coli*, procedura non possibile con le *protein-tag* convenzionali mesofile.

Tutto ciò conferma che H⁵ è un ottimo candidato per ampliare la *SNAP-tag technology* anche in condizioni di reazione estreme [Vettone *et al.*, 2016].

Un successivo capitolo di questa tesi riguarda lo studio *in vivo* delle AGT del nematode ermafrodito *Caenorhabditis elegans*. Questo sistema modello presenta due geni, codificanti la *CeAGT-1* e la *CeAGT-2*: la prima sembra essere simile alle classiche AGT a livello di struttura primaria; la seconda invece è un'AGT in cui è presente il solo dominio C-terminale catalitico, fuso a un dominio *histone-like*. Data la peculiarità strutturale della *CeAGT-2*, ci siamo chiesti quale fosse il suo ruolo *in vivo*. In particolare, abbiamo caratterizzato i fenotipi del solo ceppo mutante disponibile della *CeAgt-2*. La maggiore sensibilità del ceppo *CeAGT-2* all'agente alchilante MMS (metil-metan-sulfonato) rispetto al WT, conferma che la *CeAGT-2* svolge un ruolo chiave nella risposta ai danni agli agenti alchilanti. Tuttavia, abbiamo sorprendentemente osservato in questo ceppo delle alterazioni rispetto al WT in condizioni fisiologiche, che hanno riguardato principalmente: *i*) una minore deposizione (circa il 20%), *ii*) una maggiore letalità embrionale e *iii*) una significativa presenza di fenotipi aberranti. Essendo questi fenotipi associati a mutazioni in geni coinvolti nella meiosi e nella formazione e/o produzione dei gameti, abbiamo quindi analizzato il ruolo della *CeAGT-2* durante la meiosi. In particolare, abbiamo focalizzato l'attenzione sulla presenza della proteina Rad-51 sui tagli del DNA durante la ricombinazione meiotica, eseguendo un immunolocalizzazione sui nuclei della gonade del nematode. Nel ceppo *CeAGT-2* si riscontra una maggiore persistenza e un maggior numero di *foci* per nucleo durante la fase di pachitene, suggerendo quindi un aumento fisiologico nel ceppo mutante di errori meiotici durante la formazione dei gameti femminili. Inoltre, l'analisi del numero di corpi apoptotici nel ceppo mutante ha rilevato anche un numero maggiore di questi rispetto al WT [Gartner *et al.*, 2008], indicando ancora un difetto nella selezione delle cellule destinate all'apoptosi. I risultati ottenuti nel complesso attribuiscono un ruolo chiave della *CeAGT-2* nelle vie metaboliche legate alla riparazione del DNA e allo sviluppo meiotico. Sono in corso vari esperimenti d'incroci genetici tra il ceppo *CeAGT-2* e geni coinvolti nello sviluppo, allo scopo di chiarire più approfonditamente il ruolo di questo gene *in vivo*.

Summary Thesis

The alkylation damage on DNA is potentially dangerous to the integrity and stability of the genome. In particular, the alkylation on the oxygen at position 6 of guanine, if not repaired, leads to transition from the G: C pair to the A: T pair during the replication phase [Pegg *et al.*, 1991]. For this purpose, evolution has selected an enzyme capable of directly repairing this type of damage: the *O*⁶-alkylguanine-DNA-alkyltransferase (AGT, OGT, MGMT). This ubiquitous protein irreversibly transfers the alkyl group from DNA to a catalytic cysteine and the alkylated protein, not active longer, is recognized by cellular degradation systems [Pegg, 2011].

Although small in size, AGTs consist of two domains linked by an unstructured region (*connecting loop*): primary and tertiary structure analysis of AGTs from various organisms show a high identity of C-terminal domain involved in DNA binding and in the catalysis reaction, whereas the less conserved N-terminal domain appears to be involved in regulation, binding cooperativeness, and overall protein stability. The most studied in the literature is human AGT (hAGT), because it unexpectedly involved in the resistance of some types of cancers to alkylating agents-based chemotherapy. After strong efforts, one of the most potent inhibitor was identified: the *O*⁶-benzyl guanine (*O*⁶BG) and relative derivatives, whose activity toward the protein restored the efficacy of the chemotherapy treatments. Despite the enormous biochemical and structural characterization of this protein class, some details of the activity of AGT are still unclear. In particular, degradation upon alkylation seemed to lead to conformational changes that make protein unstable and identifiable by degradation systems [Kanugula *et al.*, 1998]. However, the 3D structures of hAGT in the alkyl form did not show evident conformational changes [Daniels *et al.*, 2000; Daniels *et al.*, 2004].

In this study, we focused our attention on the OGT of the hyper-thermoacidophilus archaea *Sulfolobus solfataricus* (*Ss*OGT). This organism, living at the temperature around 75-80 °C and an extreme pH of 2-4, is a useful model to study DNA repair systems. Because of its nature, *Ss*OGT is very stable in various reaction conditions, making it possible a wide biochemical characterization [Perugino *et al.*, 2012]. We developed a new simple activity assay using fluorescence derivatives of the *O*⁶BG inhibitor: the protein is irreversibly modified by the 1:1 reaction with the substrate, labelling with a fluorescent probe, and the intensity can be measured by conventional *gel imaging* methods. This new method, evolved and optimised in this thesis, further allowed the fully characterization of *Ss*OGT, highlighting stability in various chemical-physical conditions that are not permissible for most mesophilic proteins.

Through thermostability experiments on purified proteins, we could observe and confirm a greater instability of the methylated form than the free form at their physiological temperatures (> 70 °C), as for other AGTs. In addition, catalytic cysteine mutants that "simulate" transfers of bulky alkyl groups to the active site (C119L and C119F) are even unstable, suggesting that the overall instability of the protein is directly proportional to the size of the groups which occupy the active site. However, *Ss*OGT is more stable and versatile than the hAGT at lower temperatures, allowing the crystallization in various forms (free, complex with DNA and alkylated form) and thus analysing all the eventual conformational changes.

From the superimposition of the structure of the *Ss*OGT in free form to the methylated form, we identified two significant conformational changes: the first involves an arginine residue in position 133 (R133) on the C-terminal domain and an aspartate residue in position 27 (D27) on the N-terminal domain. In the methylated form R133 moves away from D27: since the distance in the free form between these two residues could be compatible with an ionic interaction, we assumed that the loss of this interaction could compromise the overall protein stability. Through site-directed

mutagenesis, we replaced the D27 with an alanine or a lysine, in order to eliminate the negative charge or to introduce a positive one, respectively. The great instability and low activity of D27A and D27K mutants confirmed the role of D27-R133 interaction not only in the protein stability of SsOGT, but also in proper communication between domains during the DNA repair process [Perugino *et al.*, 2015]. The second conformational change identified has been termed *K48 network*, which involves five amino acid residues (E44, K48, N59, R61 and E62), in a network in which the K48 could have the main role. Comparing the three-dimensional structures of the three protein forms (free, bound to DNA and methylated form), it has been observed that this network loses its interaction during the binding to DNA and in the methylated form. Biochemical analysis of K48 mutants has shown that the perturbation of the *K48 network* mainly affects the stability of SsOGT, but not the DNA repair activity [Morrone *et al.*, 2017]. These results delineated a proposed model that correlates conformational changes and repair activity of SsOGT on alkylated DNA. The optimum activity of SsOGT and its stability require precise coordination between the two domains. The two discovered interactions play an important role in stability and catalytic activity, acting as closed "locks" in the free form. If the locks are in their "open" state upon alkylation, there is a total loss of interactions between the N-terminal and C-terminal domains, triggering the destabilization and degradation of the SsOGT [Morrone *et al.*, 2016].

The irreversible reaction of AGT with the O^6 BG inhibitor leads to the covalent bond of the benzyl group to the active site of the protein. In 2003, Prof. Johnsson's group developed an engineered version of the hAGT (defined *SNAP-tag*) unable to bind DNA. *SNAP-tag* can be used in fusion proteins for covalent and specific protein labelling [Juillerat *et al.*, 2003]. By derivatization of chemical groups of interest with the benzyl group of the O^6 BG, it is possible to label indirectly a protein of interest with a desired chemical group. For instance, one of these derivatives is the *SNAP-Vista Green*TM, in which a fluorescein moiety is conjugated to the benzyl group of O^6 BG. However, because of the mesophilic nature of the relative hAGT, the *SNAP-tag technology* is limited at mesophilic conditions, making extreme reaction conditions not applicable. In this context, we proposed a mutant of the SsOGT (SsOGT-H⁵). Similar to *SNAP-tag*, this mutant is unable to bind the DNA and possesses the same catalytic activity, but preserves the stability to the chemical and physical denaturing agents, as the relative WT protein. In order to test the heterologous expression of the H⁵ gene in the thermophilic model, we used *Thermus thermophilus* HB27 strain, which lacks of any AGT activity. By using the fluorescent substrate *SNAP-Vista Green*TM, we identified a fluorescent band corresponding to the expected molecular weight. This signal, being present only in cells transformed with the plasmid harbouring the H⁵ gene, confirmed that the protein is expressed and correctly folded at high temperatures [Vettone *et al.*, 2016]. The H⁵ mutant was also evaluated as a *protein-tag*, combining the H⁵ gene with the thermostable beta-glycosidase from *S. solfataricus* (SsβGly). The fluorescent assay and the calculation of the kinetic constants of the enzymatic activity of H⁵-SsβGly fusion, comparable to that of the SsβGly alone, demonstrated that the each moiety of the fusion protein does not interfere with the activity of the other one. It is very important to underline that the whole fusion can be treated at high temperature, in order to eliminate the most of *E. coli* proteins, during the purification phases. This is not applicable by using conventional mesophilic *protein-tag*. All this results led us to propose H⁵ as an excellent alternative for extending *SNAP-tag technology* even under extreme reaction conditions [Vettone *et al.*, 2016].

A further chapter of this thesis concerns the *in vivo* study of the AGT of the hermaphrodite nematode *Caenorhabditis elegans*. It coding two genes for CeAGT-1 and CeAGT-2: the first seems to be similar to the classical AGTs at the primary structure level; in the second only the catalytic C-terminal domain is present, fused to a histone-like domain. Due to the peculiar structure of CeAGT-2, we wonder its role *in vivo*. In particular, we have characterized the phenotypes of the only mutant strain of CeAGT-2 available. The hypersensitivity of CeAGT-2 strain to the MMS confirmed, as in other organisms, that this protein plays a key role in the response to damage to alkylating agents. Surprisingly, by following the phenotypic analysis of the CeAGT-2 strain, we observed alterations

with respect to the WT strain at physiological conditions: *i*) a lower deposition (about 20%); *ii*) major embryonic lethality and *iii*) significant presence of aberrant phenotypes. Because these phenotypes are often associated with mutations in genes involved in meiosis and in the formation and/or production of gametes, we analysed the role of *CeAGT-2* during meiosis. In particular, we focused attention on the presence of Rad-51 protein on DNA cuttings during meiotic recombination, by immunolocalization on nematode gonad nuclei. The high endurance and major number of *foci* observed during the *CeAGT-2* strain in the pachytene phase indicated meiotic defects during the formation of female gametes. Finally, we measured a higher number of germ cells undergoing to apoptosis in the mutant, respect to the physiological number in the WT [Gartner *et al.*, 2008], suggesting again a defect in the selection of cells destined for apoptosis. Overall results showed a key role of *CeAGT-2* in metabolic pathways related to DNA repair and meiotic development. Several genetic-crossing experiments between the *CeAGT-2* strain and the genes involved in worm development are in progress, in order to further clarify the role of this gene *in vivo*.

Index

1. Introduction	p. 8
1.1 Genomic integrity and its repair	p. 8
1.2 Direct DNA repair by alkylation damages	p. 11
1.3 The Alkylguanine-DNA-alkyltransferases	p. 12
1.3.1 The AGT 3D structure	p. 13
1.3.2 The activity of AGTs	p. 16
1.3.2.1 Recognition of lesion and DNA interaction	p. 16
1.3.2.2 The peculiar “suicide” reaction mechanism of AGTs	p. 17
1.3.2.3 The fate of hAGT after alkylation	p. 18
1.4 The role of hAGT in chemotherapy	p. 18
1.5 Methods to measure the activity of AGTs	p. 19
1.6 Biotechnological uses of AGTs	p. 21
1.7 The <i>Sulfolobus solfataricus</i> OGT (SsOGT)	p. 23
2. Purpose of the thesis	p. 25
3. Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein.	p. 26
4. Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase.	p. 43
5. Crystal structure of <i>Mycobacterium tuberculosis</i> O⁶-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA.	p. 55
6. A novel thermostable protein-tag: optimization of the <i>Sulfolobus solfataricus</i> DNA-alkyl-transferase by protein engineering.	p. 67
7. The role of O⁶-alkylguanine-DNA-alkyltransferase 2 in DNA repair and in the meiotic development of <i>Caenorhabditis elegans</i>.	p. 81
8. Final Conclusion	p. 100
9. Bibliography	p. 103

1. Introduction

1.1 Genomic integrity and its repair

The maintenance of genome integrity is essential for organism survival and for the heritage of genes set to offspring. The dynamic state of DNA and its relative transactions, like replication (copying of DNA prior to each cell division), recombination (exchanges between different DNA molecules in a cell) and so on, expose it continually to a myriad of types of damage, forcing the cells to set ingenious mechanisms for tolerating and repairing these damages. Genomic instability by DNA damage, aberrant DNA replication or uncoordinated cell division, as well as hydrolytic loss of DNA bases, base oxidations, non-enzymatic methylations and other chemical alterations, as well as because of multiple reactions with exogenous (environmental) and endogenous (intracellular) reactive species (see figure 1), if left unrepaired, may interfere with DNA metabolism, resulting in the accumulation of mutations and disturbance in cellular processes [Lindahl, 1993; Friedberg, 2003].

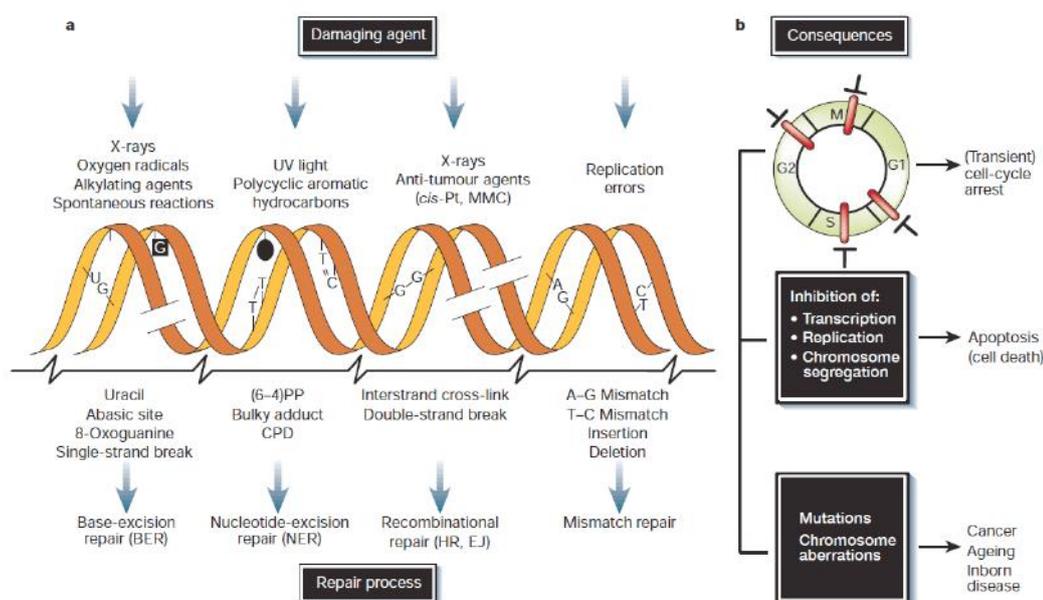


Figure 1: Schematic representation of the main types of damage, repair mechanisms and consequences of cellular response (from [Hoeijmakers, 2001]).

There are several strategies that allow the cell to survive the various types of damage. First of all, the genome has a mutation tolerance due to gene redundancy, semi-conservative replication and translesion DNA synthesis, that involves the replication machinery bypassing sites of damage, allowing normal DNA replication and gene expression to proceed downstream of the unrepaired damage [Goodman, 2002]. Biological responsiveness to genetic insult embraces more than tolerance of DNA damage. The cells have evolved complex signalling pathways to arrest the progression of the cell cycle in the presence of DNA damage, thereby increasing operate time for repair and tolerance mechanisms [Zhou *et al.*, 2000]. When the gravity of genomic damage is too large to be effectively met by the various

responses, cells are able to initiate programmed cell death (apoptosis), eliminating itself from the population that otherwise could suffer serious pathological consequences (Figure 1) [Newsheen *et al.*, 2012].

The exposure of the cell to DNA damaging agents results in the transcriptional upregulation of a large number of genes involved in different processes. The pathways of the DNA repair embraces, not only the direct repair of some types of damage (about we will discuss later in more detail), but also multiple distinct mechanisms for excising damaged bases that are termed nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR) (figure 2). The principle of all three mechanisms of repair involves splicing out the damaged region and inserting new bases to fill the gap, followed by ligation of the backbone [Fagbemia *et al.*, 2011]. The process of NER is biochemically complicated, involves as many as 30 distinct proteins, in human cells, that function as a large complex called the nucleotide excision repairosome. This 'repair machine' facilitates the excision of damaged nucleotides by generating incisions in the flanking regions and removing a fragment about 30 nucleotides in length. Damaged bases that are not recognized by the NER machines are corrected by BER, whereby the bases are excised from the genome as free bases by a different set of repair enzymes. In MMR, incorrect bases incorporated as a result of errors during DNA replication are excised as single nucleotides by yet a third group of repair proteins (see figure 2). Both NER and BER transpire by somewhat different mechanisms depending on whether the DNA damage is located in regions of the genome that are active gene expression (transcription-coupled repair) or are transcriptionally silent (global genome repair) [Yoshimoto *et al.*, 2012].

In addition to the various modes of excision repair that evolved to cope with damaged bases or mistakes during replication, the cell frequently can encounter breakage of one or both chains of the DNA duplex. Reactive oxygen molecules, cross-linking agent and ionizing radiation are prevalent sources of such damage. Double strand breaks (DSB) must be repaired to maintain genomic integrity because sever the chromosomes, interrupting genetic information, and are lethal unless repaired. Several mechanisms for the repair of DSB have been elucidated. One of these involves swapping equivalent regions of DNA between homologous chromosomes, a process called recombination. This type of exchange occurs naturally during meiosis, the special type of cell division that generates the germ cells. It can also be used to repair a damaged site on a DNA strand by using information located on the undamaged homologous chromosome. This process requires an extensive region of sequence homology between the damaged and template strands. Multiple proteins are required for DSB repair by recombination and deficiencies in this repair mechanism can cause cancer. An alternative mechanism for the repair of DSBs, called non-homologous end joining, also requires a multi-protein complex, and essentially joins broken chromosome ends in a manner that does not depend on sequence homology and may not be error free (see figure 1) [Kelley *et al.*, 2014]. A summary and schematic representation of the arguments above discussed until now is shown in Figure 1.

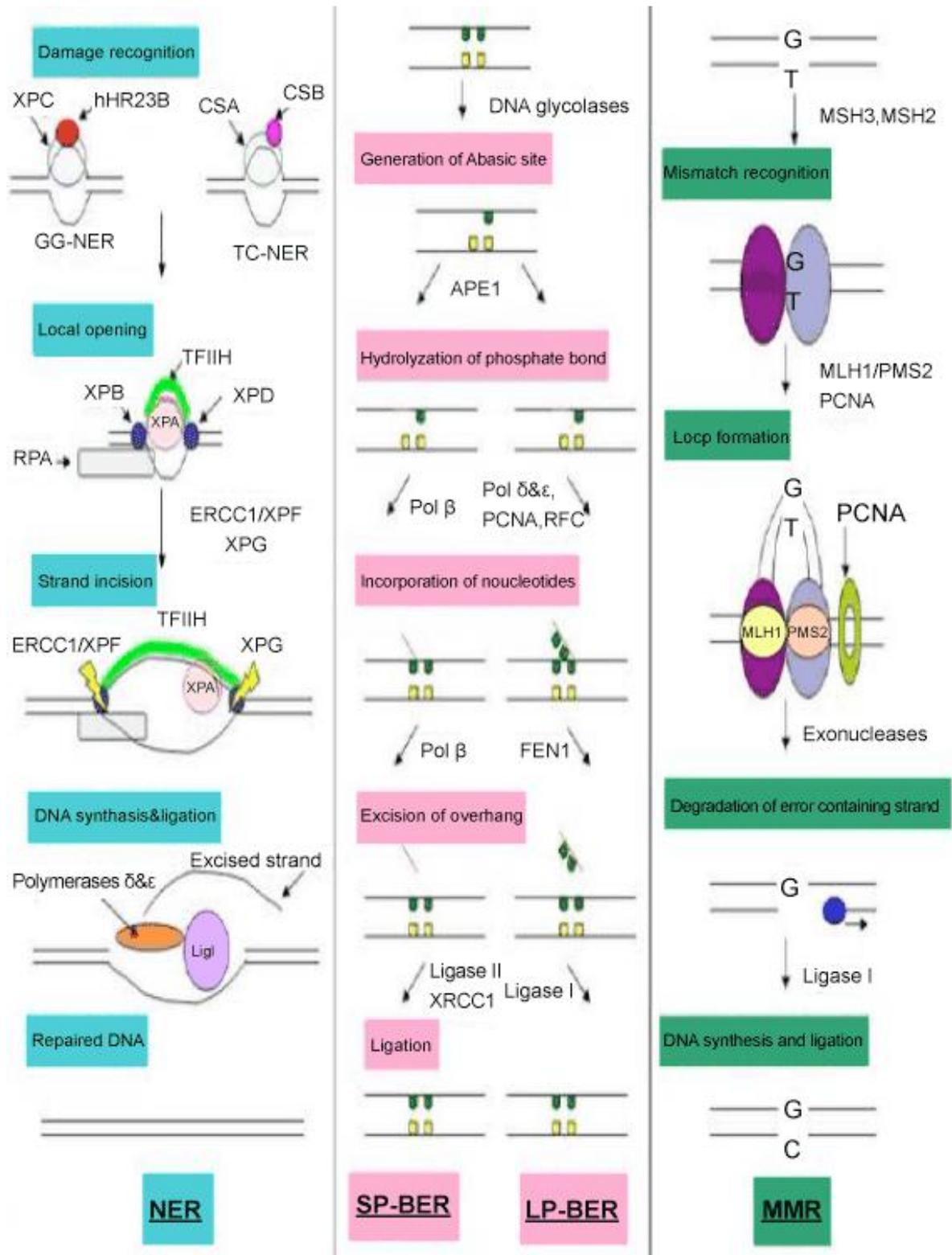


Figure 2: Schematic representation of the reaction mechanism of NER, BER and MMR.

1.2 Direct DNA repair by alkylation damages

Direct DNA repair mainly repair changes made by intracellular and extracellular chemical compounds, which can lead to the formation of covalent bonds with the bases or the backbone of the DNA. The alkylating agents are one group of chemicals compounds that can lead to DNA damage. They comprise a group of mutagens and carcinogens that can modify DNA by alkylation, some are not only widespread in the environment but are also produced intracellularly as subproducts of normal metabolism. Alkylation of bases can arrest replication, interrupt transcription, or blocking the activation of cell-cycle checkpoints or apoptosis. In mammals they could be involved in carcinogenesis, neurodegenerative disease and aging [Singer *et al.*, 1976].

Alkylating agents can introduce methyl or ethyl groups at all of the available nitrogen and oxygen atoms in DNA bases, producing a number of lesions (figure 3). The alkylation susceptibility of each site on the bases or backbone varies depending of the compound. The majority of evidence indicates that among the 11 identified base modifications two, N^3 -methyladenin (N^3 mA) and O^6 -methylguanine (O^6 mG), are mainly responsible for the biological effects of alkylation agents [Singer *et al.*, 1976]. Contribution of various lesions depends on the type of agent, its reaction mechanism and the secondary structure of the DNA target. Based on the reaction mechanism used, alkylating agents can be divided into two subgroups. The SN^1 reagents, such as N-methyl-N-nitrosourea (MNU) and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) use a monomolecular mechanism, while the SN^2 ones, that include methyl methanesulfonate (MMS), act by a bimolecular mechanism. The SN^1 type agents introduce alkyl adducts both at N and O atoms; in the case of the SN^2 subgroup, N-alkylation prevails, a summary of the possible alkylation site is shown in Figure 3. The major base modifications introduced in doubled stranded DNA by methylating agents are: N^7 -methylguanin (N^7 mG), N^3 mA, O^6 mG, while N^1 -methyladenin (N^1 mA), N^3 -methylcytosine (N^3 mC), N^7 -methyladenin (N^7 mA) and O^4 -methylthymine (O^4 mT) represent minor modifications. In single stranded DNA, N^1 mA and N^3 mC are more frequent than in double strand DNA. The most mutagenic adduct introduced into DNA by methylating agents is O^6 mG. This modification mispairs during DNA replication with thymine leading to GC→AT transition, which could be fixed in the genome [Falnes *et al.*, 2003]. The repair of DNA alkylation damage can be accomplished by some repair systems as BER, NER and MMR, but also directly by methyl transferase or oxidation demethylases. What mechanisms are most involved in the repair after exposure to alkylating agents is not yet clear: in principle it is assumed that the alkyltransferases and the MMR system are involved in *O*-alkylation repair, while BER and oxidative demethylases repair the *N*-alkylation [Chaney *et al.*, 1996].

The most common strategy used by nature to directly repair alkylation damage is to irreversibly transfer the alkyl groups to nucleophilic Cys residues in repair proteins. O^6 -Alkylguanine and O^4 -alkylthymine are repaired in such a manner by the sacrificial protein O^6 -alkylguanine-DNA alkyltransferase (AGT or MGMT). The AGT are evolutionary conserved DNA repair proteins that represent the main cellular mechanism responsible for alkylguanine repair [Pegg, 2011].

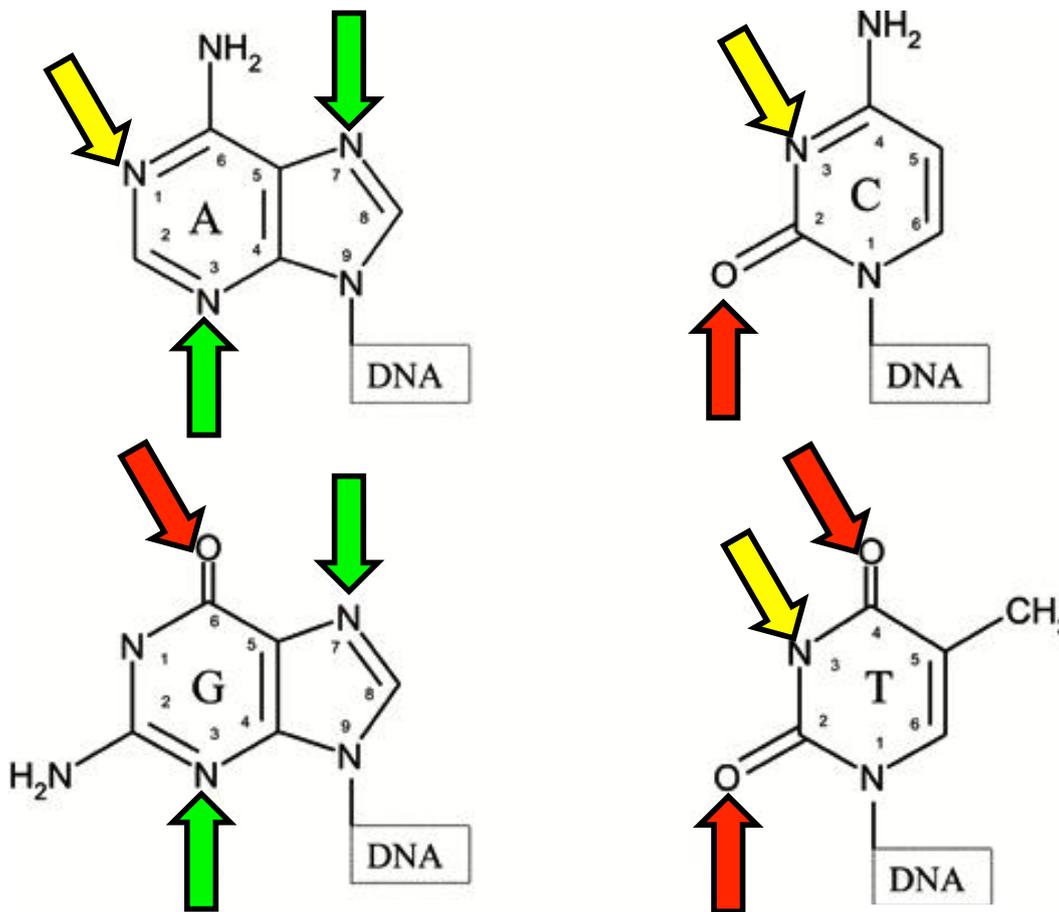


Figure 3: Possible Sites Base Alkylation. Green arrows indicate most frequent alkylation sites; red arrows indicate sites that are most frequently alkylated by SN^1 type mechanisms; yellow arrows indicate alkylated sites by SN^2 type mechanisms.

1.3 The Alkylguanine-DNA-alkyltransferases

The first AGT to be identified and characterized is the *E. coli* Ada protein [Olsson *et al.*, 1980], followed by the human AGT (hAGT) described by the Pegg's Group in 1991 [Pegg *et al.*, 1991]. Then, AGT from two important model organisms as *Drosophila melanogaster* [Kooistra *et al.*, 1999] and *Caenorhabditis elegans*, were characterised [Kanugula *et al.*, 2001]. It is important to note that it has been found both in thermophilic bacteria and hyperthermophilic Archaea, such as *Aquifex aeolicus* and *Archaeoglobus fulgidus*, respectively: however, they have been poor characterized because of their low solubility when expressed heterologously in *E. coli* [Kanugula *et al.*, 2003]. Although AGTs are widespread proteins, however, no AGT gene was found in plants, the *Schizosaccharomyces pombe* yeast, *Deinococcus radiodurans* and *Thermus thermophilus*.

From the analysis of the primary sequence, a very low homology of the N-terminal domain is evident, which also adds to functional diversification, whereas the high conservation of the C-terminal domain is expected for the presence of the main motifs, including the consensus sequence active site PCHRV, important for performing its proper

function. In the figure 4, it is highlighted the high homology of sequence that characterizes C-terminal domain and how this has been preserved during the course of evolution.

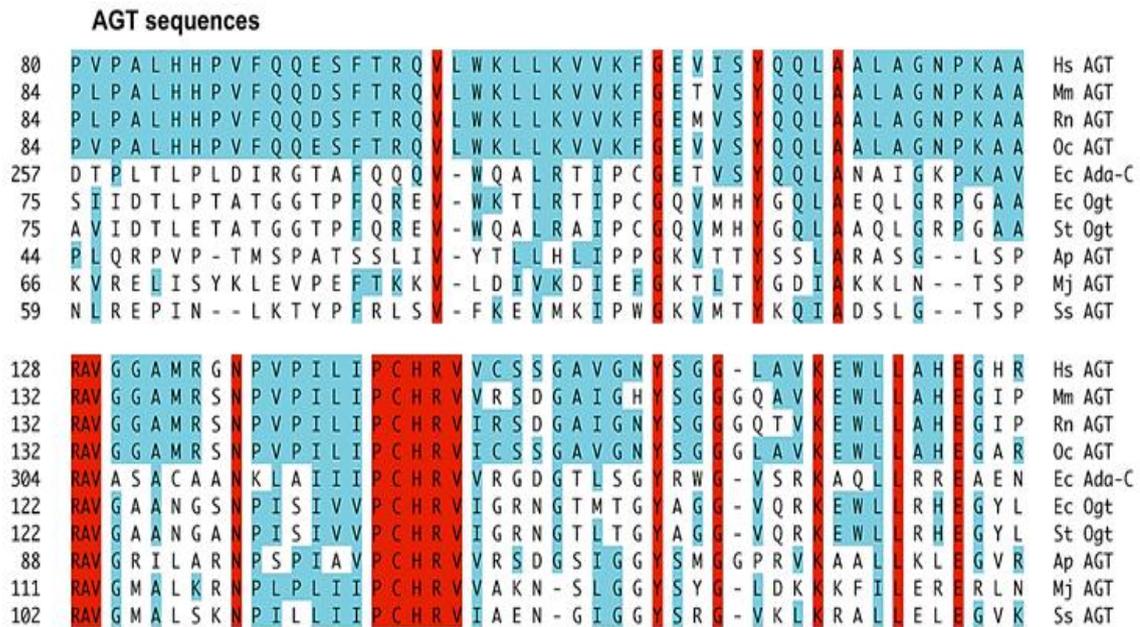


Figure 4: Alignment of amino acid sequences of C-terminal domains from AGT of different species. The AGT sequences are from *Homo sapiens* (Hs), *Mus musculus* (Mm), *Rattus norvegicus* (Rn), *Oryctolagus cuniculus* (Oc), *Escherichia coli* (Ec), *Salmonella typhimurium* (St), *Aeropyrum Pernix* (Ap), *Methanococcus jannaschii* (Mj), and *Sulfolobus solfataricus* (Ss). Highly conserved amino acids are shown in red. Similar amino acids or identical to human protein are indicated in blue (from [Pegg, 2011]).

1.3.1 The AGT 3D structure

The three-dimensional structure of these small proteins consists of two domains connected by a loop, as reported for the hAGT (figure 5) [Daniels *et al.*, 2000a]. The highly structurally conserved C-terminal domain contains the active site pocket that localizes the catalytic cysteine within the PCHR_V consensus sequence, adjacent to the DNA-binding Helix-Turn-Helix (HTH) motif. The poorly functionally characterized N-terminal domain displays a low degree of primary sequence and structural conservation between different species. The hAGT contains 207 amino acids (mw = ca. 22.0 kDa) with Cys145 as the reactive cysteine. Several X-ray crystallographic structures of unreacted AGT from human have revealed two-domain α/β fold.

The hAGT N-terminal domain (approximately residues 1-85) does not show significant primary structural homology between other AGTs [Daniels *et al.*, 2000a]: this domain contains a Zn²⁺ ion, which is not present in any of the bacterial structures, but is completely conserved among the known mammalian sequences. This suggests that the zinc site may be a structural role in higher eukaryotes [Daniels *et al.*, 2000b], binding the Cys5, Cys24, His29, and His85 residues with tetrahedral coordination of the N-terminal domain, and stabilizing the interface between the N- and C-terminal domains [Rasimas *et al.*, 2003a].

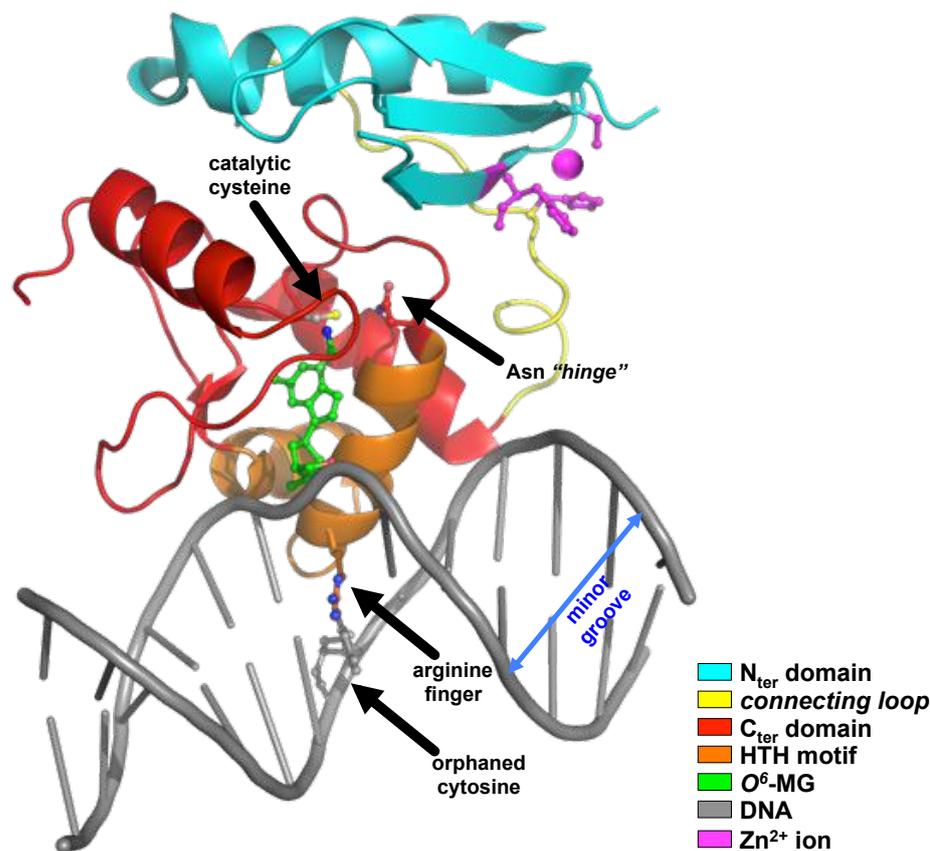


Figure 5: Three-dimensional structure of hAGT::DNA complex (PDB ID: 1T38) (adapted from [Daniels *et al.*, 2000]).

The topology of the C-terminal domain is absolutely conserved in all known AGT structures. This domain (approximately residues 86-207 in hAGT) contains the conserved active-site motif (PCHRV), the *O*⁶-alkylguanine binding channel, and the HTH motif. The latter element is very common in DNA-binding protein families, including DNA repair proteins [Daniels *et al.*, 2004]. Prior to the hAGT–DNA complex structures, HTH motifs of DNA repair proteins were thought to interact with the DNA major groove for sequence-specific recognition. Surprisingly, structural data revealed that the HTH motif of human AGT promotes non-classical minor groove DNA binding. The small and hydrophobic nature of the recognition helix residues allows this helix to pack closely with the DNA minor groove, minimizes sequence-specific interactions, and may also be advantageous for DNA repair and nucleotide flipping [Wibley *et al.*, 2000]. The hAGT–DNA complex structures, shown in figure 5, confirmed that to gain access to alkylated guanines, hAGT rotates the target base out from the DNA base stack into its active site pocket [Daniels *et al.*, 2004, Duguid *et al.*, 2005]. DNA binding does not cause evident structural change to the protein itself. However, it does result in significant structural changes to the DNA, causing bend roughly 15° away from the protein [Daniels *et al.*, 2004, Duguid *et al.*, 2005]. These changes may help to flip-out the damaged base from the DNA helix and to promote hAGT cooperative binding to facilitate damage recognition. Crystal structures of hAGT in complex with DNA suggest a mechanism for 3' phosphate rotation that induces nucleotide flipping, involving conserved residues

Tyr114 and Arg128 [Daniels *et al.*, 2004]. This residue, located at the beginning of the recognition helix of the HTH motif (figure 5), is called “arginine finger”, because it intercalates via the minor groove between the bases on either side of the substrate guanine [Daniels *et al.*, 2004], and interacts *via* a charged hydrogen bond with the orphaned cytosine (previously base paired to the flipped-out guanine) to stabilize the extrahelical DNA conformation (figure 5). It may also be involved in scanning for damaged bases by playing an active role in pushing out bases. Mutations at Arg128 revealed that the length of this side chain is directly proportional to the alkyl transfer rate for methylated DNA, but not for binding to the latter [Daniels *et al.*, 2000b].

hAGT binds to single- and double-stranded DNA in a highly cooperative manner [Rasimas *et al.*, 2003b]. It has been proposed that hAGT detects DNA lesions by searching for weakened and/or distorted base pairs rather than for the presence of adduct [Duguid *et al.*, 2005]. This promotes directional bias in repair and helps the search for damaged bases through localization of multiple hAGT molecules. In fact the protein repairs O^6 mG lesions at the 5' end of DNA faster than at the 3' end in single stranded oligonucleotides with lesions near both ends. The 3'-to-5' kinetic scanning bias, along with the cooperative DNA binding, suggest efficient hAGT binding at the 5' side of another DNA-binding hAGT molecule [Tessmer *et al.*, 2012]. These complexes contain overlapping protein molecules where there is little contact between the *n* protein and proteins *n*+1 and *n*+2, but the N-terminal surface of the *n* protein is positioned to contact the C-terminal surface of protein *n*+3. Such binding could facilitate rapid directional scanning and the efficient repair of lesions [Adams *et al.*, 2009].

The active site and HTH motif are stabilized by an extremely conserved “Asn-hinge” (Asn137 in hAGT) [Daniels *et al.*, 2000a] (figure 5). While the C-terminal domain of AGT contains the known necessary residues for DNA binding and alkyl transfer, the function of the N-terminal domain is still not clear. In an attempt to better understand the role of the N-terminal domain, the two domains of hAGT have been separately expressed and purified: interestingly, the C-terminal domain alone is not active in the absence of the N-terminal [Fang *et al.*, 2005]. These data suggest an important structural role for the N-terminal domain, in orienting the C-terminal domain for proper catalysis. Notably, two active AGT homologs, CeAGT-2 from *Caenorhabditis elegans* [Kanugula *et al.*, 2001] and AGTendoV from *Ferroplasma acidarmanus* [Kanugula *et al.*, 2005], lack the N-terminal domain. Indeed, CeAGT-2 and AGTendoV each possess a long C-terminal extension protruding from the AGT domain that is similar to histone 1C and endonuclease V, respectively. These C-terminal extensions may be important for proper catalytic orientation of the active sites in these homologs.

1.3.2 The activity of AGTs

1.3.2.1 Recognition of lesion and DNA interaction

One of the still obscure but interesting point is how AGTs recognize the lesion of a single DNA base between thousands and thousands of nucleotides. Three mechanisms have been proposed for this purpose: *i*) AGT migrates along the DNA and detects lesions by actively extruding each base on its active site; this scenario would be very expensive in terms of energy and reaction rates (Figure 6a); *ii*) AGT selectively detects intra-helical lesions resulting from non-canonical Watson-Crick couplings, "sensing" the resulting distortion and thus extruding only the damaged base (Figure 6b); *iii*) AGT captures a spontaneously extra-helical lesion due to the inability of the damaged base to form a stable joint with the base of the complementary filament, thus limiting the capture of the base damaged by this passive extrusion (Figure 6c).

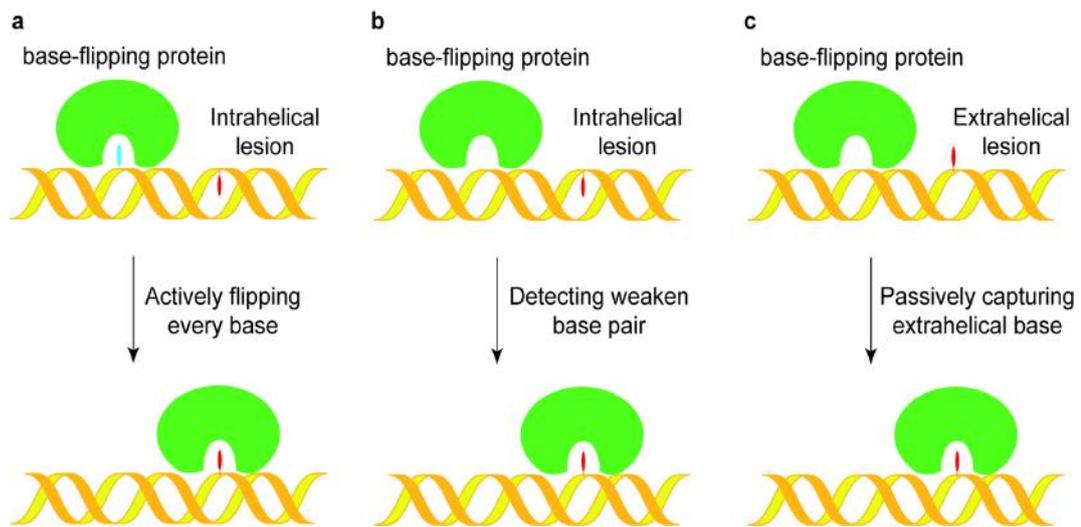


Figure 6: Three possible models to explain the AGT base recognition mechanism. A) In this model, the protein actively *flip-out* each base in the task of the active site, until it finds the lesion. B) Model proposed for damage detection by sensing weak sites or helix distortions due to non-canonical pairs. C) The DNA repair protein passively acquires a transiently extra-helical lesion (from [Yang *et al.*, 2009]).

Distinguishing between three scenarios described is not an easy work, because the direct methods of testing AGT do not good to discriminate between the various hypotheses. One of the few reported attempts aimed to solve the problem has been to use an EMSA assay combined with cross-linking substrates, consisting in appropriate oligonucleotides, in which there is a modified cytosine that after reaction covalently binding the protein to the substrate. To differentiate between different scenarios, the modified cytosine is coupled with the different oligonucleotides with different bases, so that different effects on the double filament can be analyzed. The results of this elaborated experiment suggest that while *E. coli* AGT (C-Ada) recognizes and captures extra-helical lesions, hAGT actively activates the flipping-out of the injured base, being able to sense the distortion caused by incorrect coupling [Duguid *et al.*, 2003]. However, kinetic and fluorescence experiments demonstrate that the latter

selectively pivoted the pair with unstable clinging [Zang *et al.*, 2005] and it was recently proposed that hAGT detects DNA lesions for weakening or distortion in the double helix, rather than for the actual presence of the intake [Duguid *et al.*, 2005].

1.3.2.2 The peculiar “suicide” reaction mechanism of AGTs

After nucleotide disposal in the active site, the transfer of the alkyl group to the catalytic cysteine occurs (figure 7). This residue remains *irreversibly alkylated* and the enzyme, upon completion of its task, leaving the double filament and releasing the repaired guanine, which falls into its usual intra-helical configuration. The selectivity for the nucleotide is due to hydrogen bonds and geometric constraints in the active site. The extra-helical base is inserted into a hydrophobic tasc, which is defined by the side chain of Met134 and a loop present on the active site [Vora *et al.*, 1998]. The conformation of the active site places the Cys145 thiolate almost in front of the alkyl group in O^6 . A Glu-His-water-Cys hydrogen bond network may increase reactivity of Cys145 in AGT, this hydrogen bond network is conserved in AGTs, the consensus sequence PCHRV [Daniels *et al.*, 2000a; 2000b].

The AGT hydrogen bond network also places its His146 side chain in a hydrophobic environment. It is thought that high reactivity of Cys145 is due to generation of a thiolate anion via proton transfer through the network. According to this mechanism for the dealkylation reaction, His146 acts as a water-mediated base, while Cys145 acts as the nucleophile. These interactions create a suitable environment for nucleophilic attack by the anionic thiolate on the carbon of the methyl group (figure 7) [Daniels *et al.*, 2000a; 2000b].

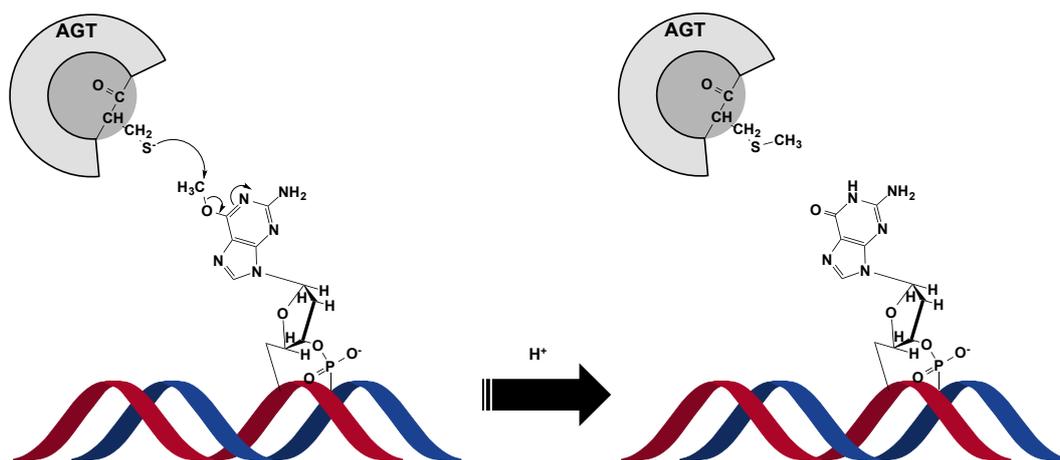


Figure 7: Reaction mechanism of AGTs. The alkyl group at the O^6 -position of guanine is removed in a one step reaction, by the irreversible transfer of the alkylic group onto a catalytic internal cystein (adapted from [Mishina, 2006], [Pegg, 2011] and [Perugino *et al.*, 2016]).

1.3.2.3 The fate of hAGT after alkylation

All AGTs are degraded upon alkylation. The molecular mechanisms to the base the AGT unfolding and degradation, which follow the alkylation of protein, are unclear. The conformational changes and modification of intramolecular interactions are likely to be associated to each step of the reaction cycle. Data on structure and properties of modified hAGT are limited; this is due to the fact that the intrinsic instability of alkylated AGT is an important limitation to structural and biochemical studies [Brunk *et al.*, 2013]. Once alkylated, the hAGT is unstable and is rapidly degraded via the ubiquitin/proteasomal system [Srivenuopal *et al.*, 1996, Xu-Welliver *et al.*, 2002]. The alkylation induce conformation modifications that cause distortion of the DNA binding surface, facilitating the dissociation from DNA and the degradation is need to continue repair since the alkylated form interfere with repair by active hAGT molecules [Edara *et al.*, 1999] The cause of ubiquitination is unknown, it has been postulated that the signal for this degradation is the disruption of a salt bridge between Asn137 and the carbonyl oxygen of Met134 due to steric clash of the new S-alkylcysteine [Daniels *et al.*, 2000b]. Crystal structures of the alkylated form of hAGT show a shift in the position of the guanine-binding loop 0.5-1.5 Å away from the protein core when compared to the native structure. Alkylation of hAGT at Cys145 leads to a sterically driven movement of a helical region that contains Met134 and Arg128 and an opening of the Asn137 hinge. These changes presumably reveal the sites for ubiquitination [Daniels *et al.*, 2000b]. Recently, two glycine residues (G131 and G132) were proposed to be implicated in the balance between stability and instability of hAGT, stabilizing the protein in the native form and triggering its destabilization upon alkylation, through a still unclear mechanism [Brunk *et al.*, 2013]. It is important to emphasize that the 3D structures of modified hAGT were obtained from native hAGT crystals flash-frozen upon soaking in solutions containing the alkyl adduct and not directly from the modified forms, due that of their strong instability upon alkylation [Daniels *et al.*, 2000b]. However, the modified 3D structures showed that C145 alkylation induces subtle conformational changes at the active site, but these 3D structures might not reflect the physiological conformation of the alkylated hAGT, adopted in solution, and it is likely that protein movements are restricted in crystals [Daniels *et al.*, 2000b].

1.4 The role of hAGT in chemotherapy

The importance of human AGT rises because in some types of tumor high levels of protein expression were highlighted, often associated with resistance to alkylating agents. For this reason, the human AGT has received attention as a potential target for the development of treatments to be integrated into current chemotherapy protocols based on such drugs [Sabharwal *et al.*, 2006]. As previously mentioned, many chemotherapeutic drugs used for their cytotoxic effects are based on alkylating agents [Chaney *et al.*, 1996]. For example, the active intermediate of temozolomide (TMZ) [Tentori *et al.*, 2009] interacts with DNA and generates a broad spectrum of methylated adducts, including N^7mG , N^3mA and O^6mG . The O^6mG lesion, although produced in low percentage (~ 8%), is generally considered to be the most toxic and mutagenic among those produced by methylating agents. The cytotoxicity is highly influenced, in human cells, by the enzymatic levels of hAGT. The number of O^6 -methyl-adducts repaired depends on the number of hAGT molecules present in the cell and from the *ex-novo* enzyme synthesis rate. When the cell expresses low concentrations of hAGT, the O^6mG adducts are not repaired and the cell starts the cell death or apoptosis

program [Tentori *et al.*, 1997]. Indeed, if the methyl adducts in the O^6 position of guanine is not removed by hAGT, O^6 mG forms altered interactions with cytosine or thymine, this resulting in activation of the MMR. It, while acknowledging this mismatch, is unable to complete the repair because it can remove the cytosine or thymine, but not methylated guanine. The DNA polymerase can't find the complementary base at O^6 mG, causing repeated activation of MMR. This futile and cyclic activation of MMR causes single and double filament breaks, stopping growth with consequent induction of the apoptotic process, this cycle is interrupted if high levels of hAGT are present in the tumor cell, reducing the efficacy of alkylating drugs [Horton *et al.*, 2009]. It has been reported that high levels of hAGT are present in neoplastic cells that make tumors less sensitive to methylating agents.

The antagonistic action between the effect of alkylating chemotherapies and the protection of hAGT is one of the main reasons because cancer cells have resistance to these types of drugs. For this reason, some studies suggest that the resistance to chemotherapy may be reduced by the inhibition of this class of methyltransferases. For this purpose, several approaches have been proposed, among which the most effective has been the combined use of alkylating chemotherapeutic agents and inhibitors pseudo-substrates of hAGT [McElhinney *et al.*, 2003]. Instead, it has been found that the protein is strongly inhibited by O^6 -benzyl-guanine (O^6 BG), because the benzyl group is more stabilized on the active site [Dolan *et al.*, 1990]. This compound mimics the damaged guanine on the DNA and reacts with the protein by covalent transfer of the benzyl group to the active site-cysteine, causing an irreversible inactivation of the enzyme. At therapeutic levels, O^6 BG is not toxic alone, but efficiently renders the tumor cells 2- to 14-fold more sensitive to alkylating agents. This establishes the potential therapeutic effect of O^6 BG as an enhancer of these drugs. It was proposed that the strongly inhibitory action of this compound could be greater if presented to the enzyme in the form of small oligonucleotides, using the ability of the protein to bind to nucleic acids. Oligonucleotides that containing more O^6 BG are potent inhibitors and they are a valid alternative to the use of free modified guanines to improve the activity of alkylating chemotherapeutic drug in the treatment of some classes of tumors [Luu *et al.*, 2002].

1.5 Methods to measure the activity of AGTs

High interest in AGTs is not only in the deepening of critical biological processes, such as DNA repair, but also for the development of new and simple, intuitive and economical assay to optimize inhibitory molecules of therapeutic interest in the treatment of tumors. Various assays for measuring AGT activity are reported in the literature. The first methods developed were based on the use of oligonucleotides in which the O^6 -alkylguanine group was labeled with radioactive isotopes (3 H or 14 C). Later with proteinase K it was done a proteolytic digestion of proteins and after an automatic amino acid analyzer were measured S-methylcysteine levels in lysate [Olsson *et al.*, 1980]. A very similar radioactive assay, but simpler and quicker, was just put on. In this procedure the terminal oligonucleotide was labeled with the radioactive phosphorus isotope and a modified guanine was inserted in a DNA sequence for a methylation-sensitive restriction enzyme. In this way, it was possible to investigate in which tumor and/or physiological tissue was removed the methyl and so to identify where the activity was present [Wu *et al.*, 1987]. A similar, but improved procedure, was used by the Ciaramella's group to identify for the first time the activity of the OGT of *Sulfolobus solfataricus* (*Ss*OGT): this assay has the advantage of analyzing digested fragment directly by electrophoresis on polyacrylamide gels [Perugino *et al.*, 2012]. In figure 8 is shown the experiment that demonstrated the alkyltransferase activity of *Ss*OGT. This method

was then made more precise, but less simple to do, by subsequent separation of digested oligonucleotides by HPLC. The separation allowed knowing the concentration of active AGT after measurement of the radioactivity of the peak corresponding to the digested fragment [Klein *et al.*, 1992]. Similarly, the Luu's group developed analysis of hAGT reaction products in 2002, based on HPLC separation. This assay investigated the degree of inhibition of oligonucleotides having O^6 mG or O^6 BG in different positions which ranged from end 3' to end 5' and whether they could be used as chemotherapeutic agents. IC_{50} values were obtained by quantifying the active protein remaining after radioactive DNA reaction. [Luu *et al.*, 2002]. All of these assays allowed for direct measurements of protein activity, however, the use of radioactive materials and chromatographic separations, although reliable and accurate, are long, laborious and unsafe methods.

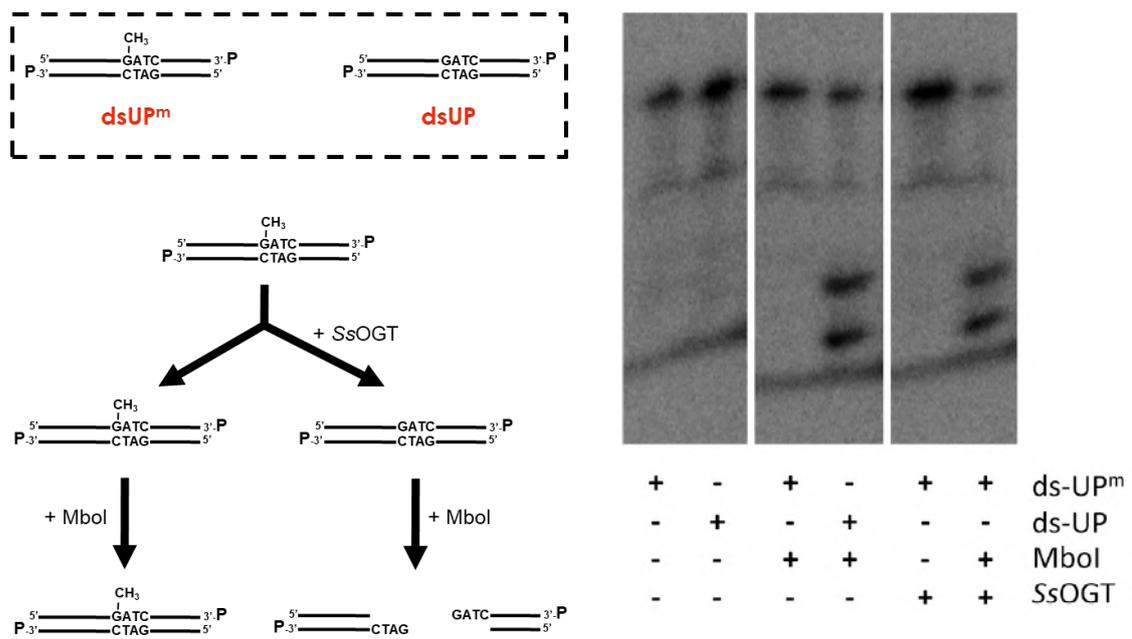


Figure 8: Radioactive assay for SsOGT. In this experiment, set up by the Ciaramella's group, the SsOGT alchyltransferase activity was addressed. Left panel, the schematic representation of experiment; on the right panel the electrophoresis on polyacrylamide gels is reported (from [Perugino *et al.*, 2012]).

1.6 Biotechnological uses of AGTs

The specific labeling of proteins in living cells with synthetic probes is an important approach to study protein function. One way to achieve such a labeling is through expression of the protein of interest as a fusion protein with an additional polypeptide, called *tag*, which mediates the labeling. For example, these *tags* allow to simplify and optimize purification methods (*affinity tags*), simplify and optimize detection procedures (by using specific antibodies) and allow following the proteins *in vivo* [Johnsson *et al.*, 2003]. Factors that affect the feasibility and attractiveness of such an approach are: *i*) the size of the *tag*; *ii*) the specificity and speed of the labelling; *iii*) the availability of a broad range of different probes and *iv*) properties of the *tag* that affect the function of the fusion protein, such as *tag*-dependent localization or stability. However, most of them do not satisfy all features listed first, for example the *Aequora victoria* green fluorescent protein (GFP) [Tsien, 1998] have some disadvantages like: relative big dimensions, insensibility to all possible changes of the cellular environment in terms of pH, hydrophobicity and ionic concentrations, incompatibility in all applications concerning anaerobic conditions and, although few examples of thermostable GFP variants were applied in thermophilic microorganisms [Cava *et al.*, 2008], the general use of most protein-tags is restricted only in mesophiles and in mild reaction conditions.

The unusual covalent linkage established between the AGT protein and its substrates has been exploited for biological applications, ranging from the specific quantification of hAGT to using in fusion proteins, for *in vivo* fluorescent labeling. All of the uses are made possible due to the high tolerance of hAGT toward transfer of various large alkyl-groups on the O^6 -position of guanine, without neglecting the advantage of their small protein size. The hAGT has been exploited for other applications completely removed from its DNA repairing function. In 2003, Kai Johnsson's group has pioneered the use of hAGT fusion proteins as *in vitro* and *in vivo* biotechnology tools, which led to the commercialization of these hAGT variants, namely *SNAP-tag*TM and *CLIP-tag*TM (New England Biolabs) [Juillerat *et al.*, 2003]. Evolved hAGT variants possess high reactivity for transferring large alkyl groups on the O^6 -position of guanine free base to the reactive Cys residue of the protein [Juillerat *et al.*, 2003]. Thus, bifunctional small molecules which contain O^6 BG conjugated through the benzyl group to a probe molecule can be prepared as good substrates for the hAGT mutants [Damoiseaux *et al.*, 2001]. The *SNAP-tag*TM can be fused with other proteins of interest: expressing the fusion protein inside cells followed by incubation with the bifunctional molecule leads to *in vivo* labeling of the fusion protein with the probe (figure 9) [Keppler *et al.*, 2003]. Fluorescent probes have been used to determine the location of the target protein [Keppler *et al.*, 2004]. The same concept has been used for *in vitro* immobilization of fusion proteins as well. In this case the alkyl group of O^6 -alkylguanine is a tether attached to a surface. Transfer of the alkyl group to hAGT anchors the fusion proteins to that surface [Kindermann *et al.*, 2003]. This offers a mild condition for fixing a wide range of proteins on a surface for evaluating their functions. However, being originated by the hAGT, a serious limitation of using these mesophilic *tags* is, the application to extremophilic organisms and/or to harsh reaction conditions.

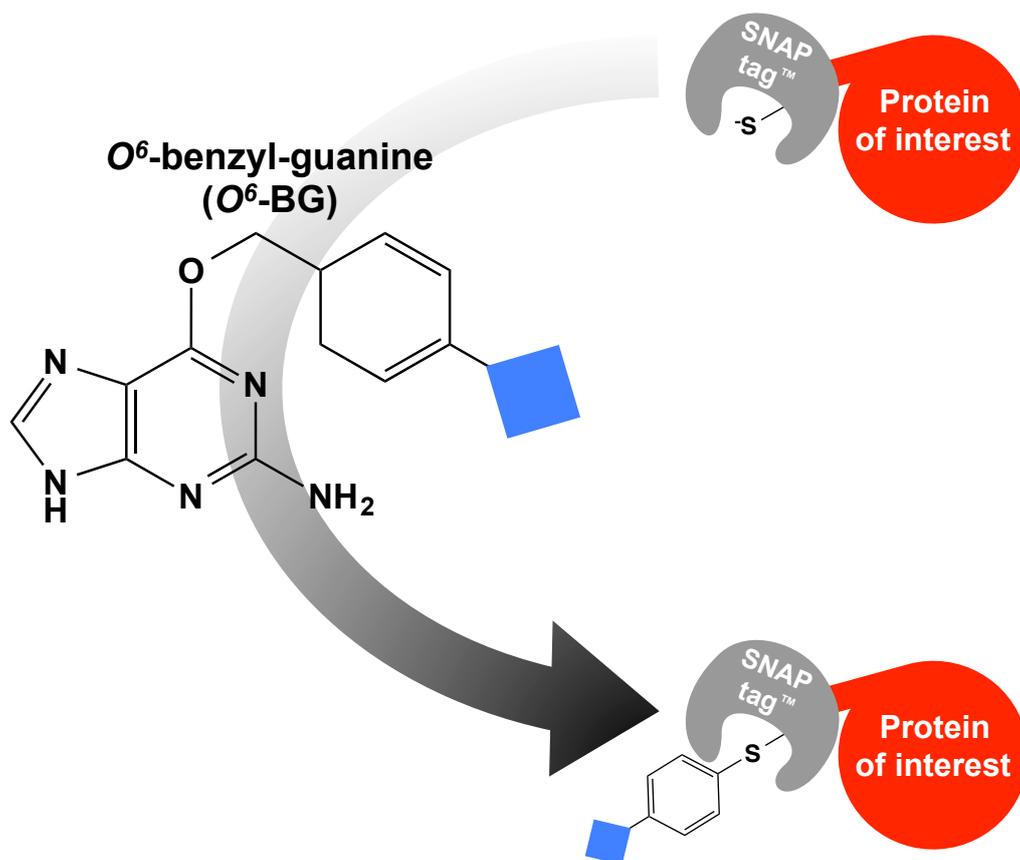


Figure 9: The *SNAP-tag*TM technology. *SNAP-tag*TM is an engineered version of hAGT, for the specific *in vivo* and/or *in vitro* labelling of fusion proteins with small molecules of interest (indicated with a blue square) conjugated with the O^6 BG moiety (adapted from [Keppler *et al.*, 2003]).

1.7 The *Sulfolobus solfataricus* OGT (*SsOGT*)

Among thermophilic archaea, the *Sulfolobus solfataricus* OGT (*SsOGT*) has been extensively characterized by the fact that, contrary to previous homologues, the heterologously expressed *E. coli* protein is soluble, very stable and easy to purify [Perugino *et al.*, 2012].

In *Sulfolobus solfataricus* the open reading frame (ORF) n. SSO2487 encodes for a protein of 151 amino acids, of a total molecular mass of 17 kDa expressed heterologously in *E. coli*, widely characterized and classified as *O*⁶-alkylguanine-DNA-alkyltransferase [Perugino *et al.*, 2012]. As expected, alignments of sequences of *SsOGT* with hAGT and homologous to other thermophilic organisms show a high sequence identity mainly in the C-terminal portion involved in catalytic activity and DNA binding.

SsOGT has a high thermal stability at 80 °C and the enzyme has a maximum activity at 80 °C, but it is important to note that it is active even at lower temperatures, as low as 25 °C. It is able to work under various reaction conditions such as pH extremes (5.0 - 8.0), high ionic strength (up to 2M NaCl) and in the presence of detergents such as Tween 20 (0.5%), Triton X-100 (1.0%) and Sarcosyl (0.5%) [Perugino *et al.*, 2012]. In addition, it can also function in the presence of EDTA, confirming the lack of any structural zinc ion [Perugino *et al.*, 2012].

The biochemical characterization of *SsOGT* was carried out through the development of an innovative assay, based on the use of a commercial fluorescein derivative of the *O*⁶BG inhibitor (*SNAP-Vista Green*TM) (figure 10a). Given the irreversible reaction of the protein, the stoichiometric ratio between protein and inhibitor is 1:1: so the amount of covalently bound probe (measured by fluorimetric techniques) is a direct measure of protein activity, and the complex can be analyzed under denaturing conditions. This made possible to calculate the kinetic constants for this complex (figure 10b) [Perugino *et al.*, 2012]. By following this approach, it was also possible measure the activity of *SsOGT* on its natural substrate (methylated DNA), using the latter as non-fluorescent competitor of the *SNAP-Vista Green*TM: the measured fluorescence reduction was a function of concentration of the competitor (figure 10c). The affinity for methylated DNA was determined through the rules of the classical competitive inhibition [Perugino *et al.*, 2012]. This new assay has also been used for the biochemical characterization of *Mycobacterium tuberculosis* (*MtOGT*) and relative mutants [Miggiano *et al.*, 2013]. A schematic representation of this set of experiment on *SsOGT* is reported in figure 10, describing a general method to measure the activity of all *O*⁶BG-sensitive AGTs.

In *SsOGT* highly conserved residues are involved in the repair of damaged DNA, such as residues forming the helix recognition and the role of Arg102 and Tyr90 in extrusion of the damaged base. The role of HTH in binding to DNA and the role of Arg102 in flipping-out have been confirmed by EMSA assays with *SsOGT* mutants. Indeed, H⁵ mutant, containing five mutations in HTH (S100A, R102A, G105K, M106T, K110E), shows a complete loss of binding to DNA, but maintaining its activity on free synthetic substrates [Perugino *et al.*, 2012]. The mutant R102A, lacking the "arginine finger", shows a reduced binding activity at the double helix, confirming its involvement in the flipping-out of the damaged base. These experiments also showed a cooperative manner in protein binding DNA [Perugino *et al.*, 2012]. Interesting results were obtained from the *in vivo* study of *SsOGT*. Although in *Sulfolobus* cells the levels of its mRNA increase following treatment with the alkylating agent, the Western Blot data using an α -*SsOGT* antibody show a decrease in intracellular protein, proportionally to time and at doses of treatment with the alkylating agent [Perugino *et al.*, 2012]. This experimental evidence show that *SsOGT* after alkylation, such as hAGT and

other examples of eukaryotic organisms, is irreversibly inactivated and then subjected to degradation.

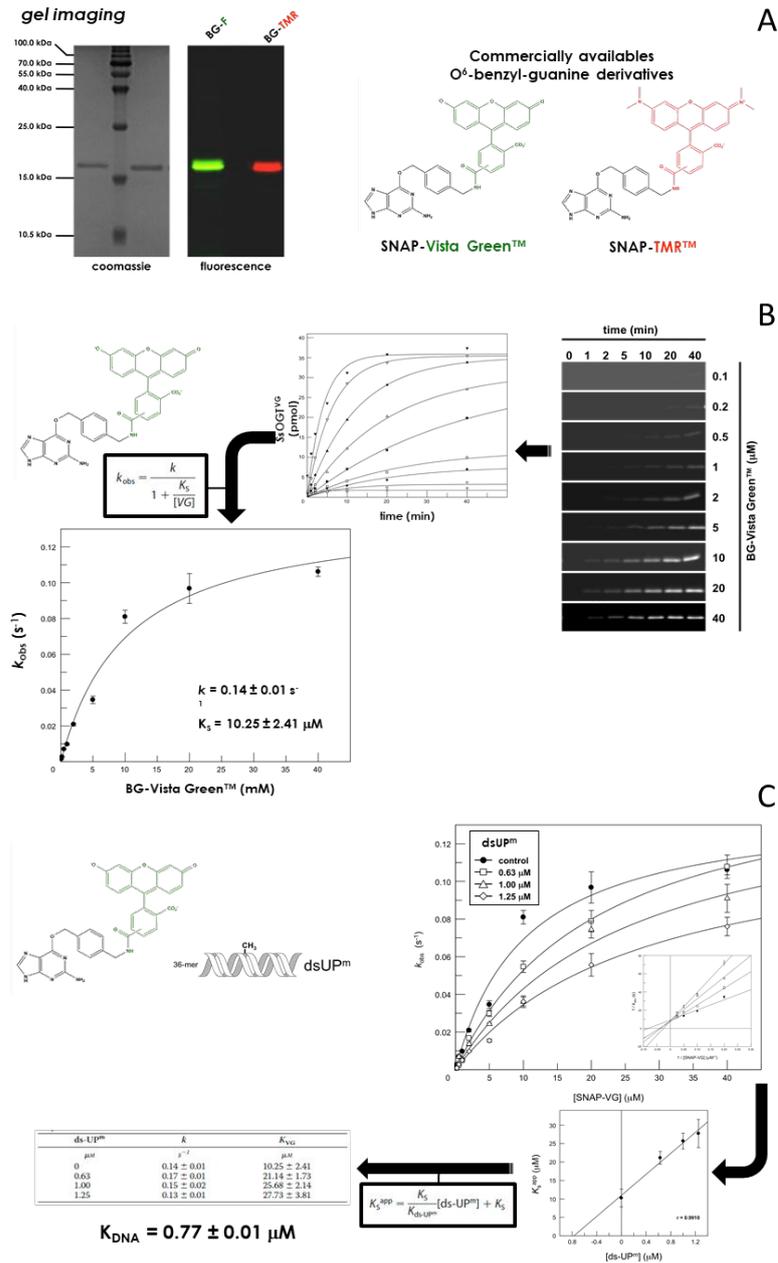


Figure 10: The new fluorimetric assay for AGTs. A) *In vitro* labelling of *Ss*OGT with commercially *SNAP*[™] substrates, which are *O*⁶BG-derivatives (in black) with different types of fluorophores. After the reaction with *Ss*OGT, it is possible to follow the labeling for gel imaging; B) The direct activity assay: fluorescence intensity data give a direct value of the amount of covalently modified protein (in pmol) and allow to calculate kinetic constants for the *Ss*OGT reaction. k_1 = rate of covalent linkage. K_S = dissociation constant for the free enzyme and free substrate during the first collision step, before covalent modification. C) The competition assay: using the natural and non-fluorescent substrate (methylated DNA) together with *SNAP-Vista Green*[™], it is possible to calculate the K_{DNA} inhibitory constant: all the linearized plots reached a similar intercept, but showing a rising slope with increasing concentration of ds-UP^m, confirming that both substates compete for the same active site. (adapted from [Perugino *et al.*, 2012]).

2. Purpose of the thesis

DNA alkyltransferases are conserved proteins that catalyze repair of alkylation damage by a mechanism involving the transfer of the alkyl group from O^6 mG to a cysteine residue in their own active site. AGTs are called *suicide proteins* because the alkylated form of the protein is irreversibly inactivated. In human cells, alkylated AGT is rapidly directed to the degradation ubiquitin pathway due to alkylation-induced conformational change that turns the protein into a proteolysis-sensitive form. Although AGT protects from mutagenic and carcinogenic effects of alkylating agents, in humans, it also plays a role in the onset of resistance to alkylating agents used in the treatment of several cancer types. Inhibition of human AGT is a potential strategy to improve the efficacy of anticancer alkylation chemotherapies but for this purpose is necessary know as much as possible on its mechanism of action. For this reason, biochemical and structural investigations on AGT and relative mutants, mimicking changes occurring *in vivo*, could provide new insights on the mechanisms of protein binding to DNA and its stability after alkylation.

In my present work, I show biochemical, structural and mutational analysis of two AGTs, from *Sulfolobus solfataricus* (SsOGT) and *Mycobacterium tuberculosis* (MtOGT).

SsOGT is intrinsically stable at high temperatures and to common denaturing agents [Perugino *et al.*, 2012]: for these reasons, it resulted a convenient model to unravel structure modifications after upon alkylation. Like the modified hAGT (obtained only by soaking protein crystals in substrate solutions [Daniels *et al.*, 2000b]) the methylated form of SsOGT was unstable at its physiological temperatures (> 70 °C) and precipitates in solution; however, the *Sulfolobus* AGT resulted soluble and stable at mesophilic temperatures (as high as 40 °C). This great stability allows deep analysis of the protein in its post-reaction form. In this way, the structure could maintain the physiological conformations upon protein alkylation, and tries to unravel those conformational changes, which were not possible to observe in hAGT since the movements are forced by the crystalline grind.

Furthermore, the identification and the characterization of this thermophilic AGT [Perugino *et al.*, 2012] could be a candidate to extend the *SNAP-tag*TM technology to extreme environments. To this aim, we consider the SsOGT-H⁵ mutant, because it lacks of any DNA binding activity and a marked catalytic efficiency with the fluorescent BG-derivatives. Furthermore, likely its relative wt enzyme, it displays a marked stability to wide range of temperatures, pH, ionic strength and to common denaturing agents [Perugino *et al.*, 2012].

The MtOGT encodes for a sequence (adaB; Rv1316c) part of an adaptive response operon and associated with the virulence of *Mycobacterium tuberculosis* strains. The amino acid sequence of this protein is very similar to hAGT and shares with this latter its mesophilic characteristics. In addition, certain AGT variants are associated with the virulence of *Mycobacterium tuberculosis* species and it is interesting that a number of multidrug-resistant strains are characterized by possess mutated OGT that carrying an amino acid substitution of N- terminal domain. It has been suggested that a defective alkylated-DNA repair could have played a role in the balance between genomic stability preservation and adaptability to host during the evolutionary history of the pathogen [Miggiano *et al.*, 2013].

From resolution of the 3D structure of MtOGT complexed with a modified DNA molecule, we can reveal similar or peculiar traits when compared with the equivalent structure of hAGT. The crystal structure of the complex could show features of the protein-protein and protein-DNA interactions occurring during alkylated DNA binding. Finally, biochemical and structural characterization of protein and relative variants at the N- and the C-terminal domains could highlight how this protein contributes to protect the bacterial GC-rich genome against the mutagenic lesion of O^6 mG in DNA.

Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein.

Giuseppe Perugino*¹, Riccardo Miggiano*², Mario Serpe¹, Antonella Vettone¹, Anna Valenti¹, Samarpita Lahiri², Franca Rossi², Mosé Rossi¹, Menico Rizzi², Maria Ciaramella¹.

¹Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80125 Naples, Italy.

²DiSF-Dipartimento di Scienze del Farmaco, University of Piemonte Orientale “A. Avogadro”, Via Bovio 6, 28100 Novara, Italy.

*The authors equally contributed to the present work.

In this paper, the study on the SsOGT is reported, taking advantage from its exceptionally high stability for performing structural and biochemical analysis in different stages of the enzyme activity: the free form, the DNA-bound complex and the post-alkylation form. This analysis provides insights in the structure-function relationships between the two domains of the protein during its activity. The molecular basis for DNA recognition, catalytic activity and fate upon alkylation was highlighted. This information by using a thermostable AGT gives hints on the general mechanism of alkylation-induced inactivation of this class of proteins.

My contribution was to express in E.coli, purify and characterize biochemically all the above-mentioned forms of the SsOGT and relative single mutants. The group of Prof. Menico Rizzi of the University of Novara solved all 3D structures.

Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein

Giuseppe Perugino^{1,†}, Riccardo Miggiano^{2,†}, Mario Serpe¹, Antonella Vettone¹, Anna Valenti¹, Samarпита Lahiri², Franca Rossi², Mosè Rossi¹, Menico Rizzi² and Maria Ciaramella^{1,*}

¹Institute of Biosciences and Bioresources, National Research Council of Italy, Via P. Castellino 111, 80125 Naples, Italy and ²DiSF-Dipartimento di Scienze del Farmaco, University of Piemonte Orientale 'A. Avogadro', Via Bovio 6, 28100 Novara, Italy

Received May 28, 2015; Revised July 09, 2015; Accepted July 18, 2015

ABSTRACT

Alkylated DNA-protein alkyltransferases repair alkylated DNA bases, which are among the most common DNA lesions, and are evolutionary conserved, from prokaryotes to higher eukaryotes. The human ortholog, hAGT, is involved in resistance to alkylating chemotherapy drugs. We report here on the alkylated DNA-protein alkyltransferase, SsOGT, from an archaeal species living at high temperature, a condition that enhances the harmful effect of DNA alkylation. The exceptionally high stability of SsOGT gave us the unique opportunity to perform structural and biochemical analysis of a protein of this class in its post-reaction form. This analysis, along with those performed on SsOGT in its ligand-free and DNA-bound forms, provides insights in the structure-function relationships of the protein before, during and after DNA repair, suggesting a molecular basis for DNA recognition, catalytic activity and protein post-reaction fate, and giving hints on the mechanism of alkylation-induced inactivation of this class of proteins.

INTRODUCTION

Alkylated DNA-protein alkyltransferases (AGTs, MGMTs or OGTs, EC 2.1.1.63) are conserved proteins that repair alkylation damage in DNA, mainly at position *O*⁶ of guanines. They use a peculiar single-step mechanism in which the direct repair of the alkylated base is coupled with irreversible alkylation of the catalytic cysteine in the protein active site. The trans-alkylated protein is permanently inactivated and prone to degradation both *in vivo* and *in vitro* (1–

4). Most knowledge these proteins comes from classic studies on Ada-C and OGT of *Escherichia coli*, as well as the human hAGT (1–4). In light of the observation that hAGT over-expression in tumor cells is frequently associated with resistance to alkylating agents, hAGT has received attention as a potential target for the development of treatments to be integrated into current chemotherapy protocols based on such drugs (5).

The current understanding of the AGTs molecular mechanism responsible for alkylated DNA recognition and repair is mainly based on the 3D structures of hAGT and its complex with double-stranded (ds) DNA molecules (6–9). The protein contacts the DNA minor groove through the helix-turn-helix (HTH) motif of its C-terminal domain. By adopting an extra-helical conformation stabilized by an 'arginine finger' (R128) of the HTH, the damaged base is deeply inserted in the protein active site, and the alkyl moiety is then transferred to the catalytic cysteine (C145). Data on structure and properties of alkylated AGTs are limited, because alkylation greatly destabilizes their folding; as for 3D structures, the only available are those of the methylated (hAGT^m) and benzylated (hAGT^b), which were obtained from native hAGT crystals flash frozen upon soaking in solutions containing *O*⁶-methyl- (*O*⁶-MG) and *O*⁶-benzyl-guanine (*O*⁶-BG), respectively (7). The hAGT^m and hAGT^b structures showed that C145 alkylation induces subtle conformational changes at the active site; however, these structures might not reflect the physiological conformation of the alkylated hAGT, since in the crystalline state the protein could not accurately display the conformation adopted in solution. Indeed, it is likely that protein movements are restricted in crystals, and larger rearrangements may lead to crystal decay (7).

Alkylation-induced instability of AGTs is interesting from both a biological and mechanistic point of view. AGTs

*To whom correspondence should be addressed. Tel: +390816132247; Fax: +39081632646; Email: maria.ciaramella@ibbr.cnr.it

†These authors contributed equally to the paper as first authors.

degradation is an important process in the organisms' response to alkylation damage; in human cells, hAGT physically associates and undergoes repair-mediated degradation with the BRCA2/FancD1 protein (10). Germline mutations of Brca2 are associated with cancer prone syndromes and Fanconi anemia, and absence of a functional BRCA2 protein induces increased cell sensitivity to DNA crosslinking agents (10); thus, hAGT activity and degradation might also affect other DNA repair pathways. The relation between active site modification and protein unfolding/degradation has been difficult to study due to the instability of alkylated AGT forms. *In vivo*, conformational changes might expose residues, which are target for ubiquitination, thus triggering the protein degradation; on the other hand, alkylated hAGT is also intrinsically unstable *in vitro* (7,11). It has been suggested that alkylation-induced conformation modifications induce distortion of the DNA binding surface, facilitating the dissociation of alkylated hAGT from DNA, while destabilizing the protein native fold (7,12,13). Recently, two glycine residues (G131 and G132) were proposed to be implicated in the balance between stability and instability of hAGT, stabilizing the protein in the native form and triggering its destabilization upon alkylation, through a still unclear mechanism (14).

AGTs are present in organisms from the three living domains (Eucarya, Bacteria, Archaea). In thermophilic bacteria and archaea, living at >80°C, alkylation damage is a serious harm since alkylated bases are unstable at high temperature and induce DNA ruptures (15,16). We have previously reported on OGT from the archaeon *Sulfolobus solfataricus* (*Ss*OGT), a protein with outstanding stability at high temperature, which, upon alkylation, becomes unstable and undergoes degradation *in vivo*, suggesting that it follows the same fate as hAGT (17,18).

We present here a biochemical, structural and mutational analysis of *Ss*OGT. The crystal structure of the ligand-free protein and its complex with an *O*⁶-MG containing dsDNA revealed overall similarities with the corresponding structures of hAGT, but also peculiarities, which were found to have functional significance. Moreover, in contrast to the corresponding hAGT^m, the methylated form of *Ss*OGT (*Ss*OGT^m) was soluble and relatively stable, thus allowing in-deep analysis of the protein in its post-reaction form. Structural and biochemical analysis of *Ss*OGT^m, as well as of a mutant mimicking the presence of a bulkier adduct in the active site (C119L), suggested a possible mechanism of alkylation-induced *Ss*OGT unfolding and degradation. Based on our data, we suggest a general model for the mechanism of post-reaction AGTs destabilization.

MATERIALS AND METHODS

DNA mutagenesis and protein purification

Site-directed mutants were obtained by using the oligonucleotide 'mut' and 'rev' pairs (Supplementary Table S1) and the GeneTailorTM Site-Directed Mutagenesis System (Invitrogen); the template was the *S. solfataricus* ogt gene cloned in the pQE31TM vector (17). N-terminally His-tagged proteins were expressed in the *E. coli* ABLE-C strain and purified as described (17).

Fluorescent assays for *O*⁶-alkyl-guanine alkyltransferase activity

*Ss*OGT activity was determined by a fluorescent assay, by using the SNAP-Vista GreenTM (BG-VG) substrate under the standard reaction conditions as previously reported (17,19). After SDS-PAGE, fluorescent protein bands were visualized by gel imaging using the VersaDoc 4000TM system (Bio-Rad). The fluorescence intensity of each band was corrected for the amount of protein loaded by Coomassie Blue staining. Determination of pseudo-first-order reaction rate values was carried out under standard conditions, taking protein aliquots at different times. Second-order rate constants were then obtained by dividing values by the substrate concentration (17,20).

Direct DNA repair activity

The efficiency of repair of ds oligonucleotides containing a single *O*⁶-MG was determined by incubation of fixed amounts of protein (5 μM) for different timespans at 25°C in the presence of increasing concentrations of the ds-Fwd^{m26} oligonucleotide. Then, 10 μM of BG-VG was added, incubation was prolonged for 2 h and samples analyzed by SDS-PAGE. Gel imaging and Coomassie correction was performed as described above.

Competition assay with non-fluorescent *O*⁶-methyl-guanine containing DNA

The efficiency of repair of ds oligonucleotides containing a single *O*⁶-MG was determined by using the half maximal inhibitory concentration (IC₅₀) method, where IC₅₀ is the concentration of methylated DNA needed to reduce by 50% the fluorescence intensity of the *Ss*OGT band in reactions containing fixed concentrations of the fluorescent BG-VG molecule (Supplementary Figure S1A). Reactions were incubated at fixed temperatures with increasing concentrations (0–10 μM) of the appropriate methylated oligonucleotide. Gel imaging and Coomassie correction were performed as described above; corrected data of fluorescence intensity were fitted with the IC₅₀ equation. For competitive inhibition, the affinity of the inhibitor (K_{DNA}) is related to the IC₅₀ value by the adapted Cheng-Prusoff Equation (1) (21):

$$K_{DNA} = IC_{50} / [1 + ([VG]/K_{VG})] \quad (1)$$

where K_{DNA} is the binding affinity of the inhibitor, IC₅₀ is the functional strength of the inhibitor, [VG] and K_{VG} are the VG concentration and the concentration of the BG-VG substrate at which enzyme activity is half maximal, respectively (17).

Electrophoretic mobility shift assay

A tetramethylrhodamine labeled dsDNA probe was prepared by annealing the oligonucleotides A⁺ and D⁻, as reported (Supplementary Table S1; 17). Different amounts of protein (0.0–25.0 μM) were incubated at 37°C for 10 min with the probe (0.2 μM), in a total volume of 10.0 μl, as described (17). After loading the sample on an 8% polyacrylamide native gel run in 1X TBE (90.0 mM Tris-HCl,

90.0 mM boric acid, 2.0 mM EDTA, pH 8.3), signals were measured by gel imaging, using a green LED/605 band-pass filter as excitation/emission parameters, respectively. For DNA binding assay with unlabeled dsDNA oligonucleotides containing *O*⁶-MG, samples were prepared as described above, except that, after electrophoresis, gels were stained with ethidium bromide (10 mg/ml) for 15 min at RT.

Protein stability assays

Thermal stability was determined by two different methods. (i) temperature-induced aggregation: 30 μ l aliquots of 5.0 μ M (0.1 mg/ml) of each protein were incubated for 20 min at different temperatures in PBS 1X buffer, centrifuged for 20 min at 16 000 \times g, and 20 μ l of supernatant were immediately loaded on 15% SDS-PAGE. The relative intensity of Coomassie-stained SsOGT bands were plotted as a function of temperature, considering as 100% the intensity of each protein band incubated at 25°C. Reported data are the mean \pm SD of three independent experiments; (ii) Differential Scan Fluorimetry (adapting the protocol described in 22): each protein (25 μ M; 0.5 mg/ml) was incubated in PBS 1X buffer and SYPRO Orange dye 1X, in a total volume of 30 μ l. Samples were heated from 20 to 95°C in a Real-Time Light Cycler (Bio-Rad, Milan, Italy). Thermal stability scans were performed at 0.2°C/min (5 min/cycle with an increase of 1°C/cycle). Data were normalized to the maximum fluorescence value within each scan. Relative fluorescence intensities were plotted as a function of temperature; the obtained sigmoidal curve describes a two-state transition, where the T_m value represents the inflection point of the transition curve, as described by the Boltzmann Equation (2),

$$y = LL + \frac{(UL - LL)}{1 + \exp\left(\frac{T_m - x}{a}\right)} \quad (2)$$

where the values of minimum and maximum fluorescent intensities are LL and UL, respectively, and a represents the slope of the curve within T_m (22). Data are the mean \pm SD of three independent experiments.

Data analysis

Corrected data were fitted to appropriate equations by using GraFit 5.0 Data Analysis Software (Erithacus Software) or Prism Software Package (GraphPad Software) (23).

Preparation of SsOGT^m

20 ml of 10 μ M of the SsOGT wt protein were incubated at 25°C for 24 h in the presence of *O*⁶-MG (10 mM, protein:inhibitor ratio 1:1000). To test the efficiency of methylation, an aliquot containing 50 pmols (1.0 μ g) of SsOGT from the methylation reaction was incubated with an equimolar amount of BG-VG at 70°C for 30 min and subsequently subjected to SDS-PAGE and fluorescence imaging: no fluorescent signal was detected, confirming the complete methylation of the protein. Unreacted inhibitor was removed using a HiTrap DesaltingTM (GE, Healthcare) column pre-equilibrated in PBS 1X.

Preparation of SsOGT-C119A::dsDNA^m complex

To obtain the modified dsDNA molecule used in co-crystallization experiment (dsDNA^m), the *O*⁶-methylguanine-containing oligonucleotide (5'-GCCATG[*O*⁶-MG]CTAGTA-3', Primm, Milan, Italy) was annealed to the complementary oligonucleotide 5'-TACTAGCCATGGC-3' (Eurofins MWG Operon). The resulting sample was mixed with the C119A protein solution (7 mg/ml in 20 mM phosphate buffer, pH 7.3 and 150 mM NaCl) at a protein:DNA molar ratio of 1:1.2, and incubated 1 h at room temperature before crystallization trials.

Crystallization and data collection

Initial crystallization conditions for wild type, C119L and SsOGT^m proteins and for the C119A::dsDNA^m complex were identified by means of a robot-assisted (Oryx4; Douglas Instruments), sitting-drop-based sparse-matrix strategy using screen kits from Hampton Research and Qiagen. Wild-type SsOGT crystals grew at 4°C in 4 μ l drops obtained by mixing equal volumes of 9 mg/ml purified protein solution and precipitant 0.35 M potassium nitrate and 1.6 M ammonium sulfate in a final droplet volume of 1 μ l. A single crystal was cryo-protected in precipitant solution containing 15% glycerol, mounted in a cryo-loop, and flash-frozen in liquid nitrogen at 100 K for further X-ray diffraction analysis. Single crystal diffracted at 1.85 Å of resolution at the ID23 synchrotron radiation ($\lambda = 0.87$ Å) (European Synchrotron Radiation Facility [ESRF], Grenoble, France). The diffraction data were indexed with XDS program (24), whose indexing score assigned crystal to the trigonal space-group R3 with the cell dimension $a = 94.72$ Å $b = 94.72$ Å $c = 76.70$ Å. Crystals of C119L mutant (8 mg/ml) were identified in the initial crystallization trials in the condition containing 4 M sodium formate as reservoir solution in a protein to reservoir ratio of 1:1, in a final droplet volume of 1 μ l. Single crystal suitable for X-ray diffraction was manipulated as previously describe for wild-type protein and it diffracted at 2.6 Å of resolution at 100K at the ID23 beam line ($\lambda = 1.89$ Å) (ESRF, Grenoble, France). Indexing process with XDS program assigned the crystal to the cubic I432 space-group with the dimension $a = b = c = 140.56$ Å. SsOGT^m, obtained as described above, was concentrated to 7 mg/ml. Well-shaped single crystals grew in a 1 μ l droplet containing equal volumes of protein and reservoir solution (0.1 M Bis-Tris pH 5.5, 0.1 M ammonium acetate, 17% w/v PEG 17000). Data collection was performed at BM30 synchrotron radiation ($\lambda = 0.979$ Å) (ESRF, Grenoble, France), under cryogenic condition using 15% glycerol as cryo protectant in reservoir solution. The best crystal diffracted at 2.8 Å of resolution and it was assigned to the orthorhombic space group P212121 with the cell dimensions $a = 48.49$ Å $b = 50.22$ Å $c = 142.02$ Å, containing 2 molecule for asymmetric unit, with a Matthews parameter and a solvent fraction of 2.28 Å³ Da⁻¹ and 45.98%, respectively. C119A::*O*⁶-MG-DNA single crystals grew in approximately 1 week in a 1 μ l droplet of equal volume of protein-DNA complex solution and reservoir solution composed by 0.2 M Na/K tartrate 20% (w/v) PEG 3350. Diffraction experiments were conducted at 100 K using synchrotron radiation ($\lambda = 0.972$

Å) at the ID29 (ESRF, Grenoble, France). Data collection was performed up to 2.7 Å of resolution. Indexing score with XDS assigned crystal to the orthorhombic space group P2₁2₁2₁ with the cell dimensions a = 41.76 Å b = 65.88 Å c = 97.52 Å. For all data sets described above, further data manipulations were carried out using COMBAT and SCALA from the CCP4 program suite (25). The data statistics of the solved structures are summarized in Table 1.

Structure determination, model building and refinement

The initial phases for wild-type SsOGT, C119L mutant and SsOGT^m structures were generated by molecular replacement with the program PHASER (26) of the PHENIX software suite (27) using *Sulfolobus tokodaii* OGT structure (PDB ID code:1WRJ) as the search model. The starting search model for C119A::O⁶-MeG-DNA complex structure consisted of *Sulfolobus tokodaii* OGT structure and the double stranded, methylated DNA molecule as crystallized in complex with hAGT (PDB ID code: 1T38) for the protein and DNA component, respectively. Initial model building was performed using AUTOBUILD of the PHENIX suite (28) followed by manual model building with the program COOT (29). Solvent molecules were added by ARP/wARP SOLVENT program from CCP4 program suite followed by structure refinement that was done with PHENIX (27). In the refined SsOGT^m structure we noticed a large difference between *R*_{work} and *R*_{free} values. We deeply investigated all possible space groups suggested by the XDS analysis, and acceptable statistics could only be obtained by processing data in the space group P2₁2₁2₁. Similarly an unambiguous solution of molecular replacement could only be found in P2₁2₁2₁ space group. All figures illustrating structural analyses were generated with PyMol (<http://www.pymol.org/>) (30).

Protein structure accession numbers

The atomic coordinates and structure factors of the SsOGT wild type, C119L mutant, SsOGT^m and C119A::dsDNA^m have been deposited in the Protein Data Bank (<http://www.rcsb.org>) under the accession codes 4ZYE, 4ZYH, 4ZYG, and 4ZYD, respectively

RESULTS

Structure-function analysis of free SsOGT

The crystal structure of SsOGT was solved at 1.8 Å resolution (Table 1). As observed for all AGTs structures present in the Protein Data Base (6–9,19,31–34), SsOGT folds in two domains joined by a long connecting loop (a.a. 54–69) (Figure 1A). The N-terminal domain (a.a. 1–53) consists of an anti-parallel β-sheet, connected to a conserved α-helix (H1) by a random-coiled region, which is stabilized at its N-side by a disulphide bridge established between the C29 and C31. The C-terminal domain (a.a 70–151) houses the functional elements required for DNA binding and repair: (i) the catalytic C119 residue within the conserved PCHR signature; (ii) the helix-turn-helix motif (HTH), which binds the DNA minor groove and holds the arginine finger (R102) that participates to the modified base flipping out from the

DNA base stacking; (iii) the ‘asparagine hinge’ that, together with the helix H4 of the HTH, defines one wall of the ligand-binding pocket; and (iv) the active site loop on the H4-facing side of the active site (Figure 1A). Previously, we demonstrated that mutation of the R102 residue reduces DNA binding efficiency, whereas mutation of five residues in the HTH motif abolishes the protein ability to form stable complexes with DNA, although both mutants are normally proficient in the trans-alkylation reaction (17).

An interesting feature, not previously reported for other AGTs, is the C29-C31 S-S bridge of the N-terminal domain. To test its role in the SsOGT activity and stability, we produced the C29A mutant, which resulted as active as the wild-type protein in repair of a O⁶-MG-containing ds oligonucleotide at 50°C (Supplementary Figure S1B). In contrast, the C29A mutant was significantly less thermostable than the wild-type (Figure 1B): incubation at increasing temperatures showed that, whereas the wild-type protein remained 100% soluble after 20 min at 70°C, the C29A protein aggregated above 60°C. Quantitative analysis by Differential Scanning Fluorimetry (DSF) allowed calculation of a *T*_m of 80°C for SsOGT and 60°C for C29A (Figure 1C). Thus, the S-S bond is not involved in DNA repair activity, but is an important structural element contributing to the impressive thermal stability of SsOGT.

Alkyltransferase-like proteins (ATLs) share structural similarity with AGTs, but are catalytically inactive and are believed to act as DNA damage sensors (35). Although the general fold is shared by hAGT and *Schizosaccharomyces pombe* ATL1, important differences were found in the catalytic loop and Asn hinge, resulting in larger size of the lesion-binding pocket in ATL1, which might account for its broad lesion recognition range (36). Superimposition of SsOGT, ATL1 and hAGT structures showed the same differences, confirming the higher similarity of SsOGT with hAGT as compared with ATL1 (Supplementary Figure S2).

Structure-function analysis of dsDNA-bound SsOGT

Different strategies have been adopted to trap the AGT-alkylated dsDNA complex and solve its structure: the wild-type hAGT has been crystallized crosslinked to oligonucleotides containing alkylated guanine analogues (8,9), and the hAGT C145S inactive mutant was co-crystallized with a more physiologic, O⁶-MG-containing, 13 base pair long oligonucleotide (dsDNA^m) (8). These DNA-bound hAGT structures were found essentially superposable to each other and to ligand-free hAGT, suggesting that binding to DNA does not affect the protein overall structure (8,9).

To solve the SsOGT DNA-bound crystal structure, we obtained the C119A mutant, carrying a substitution of the catalytic C119 to prevent protein alkylation and subsequent dissociation from the repaired DNA substrate (data not shown). The crystal structure of the C119A::dsDNA^m (Figure 2A) was solved at 2.7 Å resolution (Table 1), revealing that a single SsOGT monomer is able to occlude 4 base-pairs (bp) on dsDNA substrate, paralleling what observed for DNA-bound hAGT (37–39). The crystal structure of the ligand-free protein and of the C119A::dsDNA^m complex can be superposed with a 0.3 Å average root-mean-square deviation (r.m.s.d.) of Cα positions. Assuming that

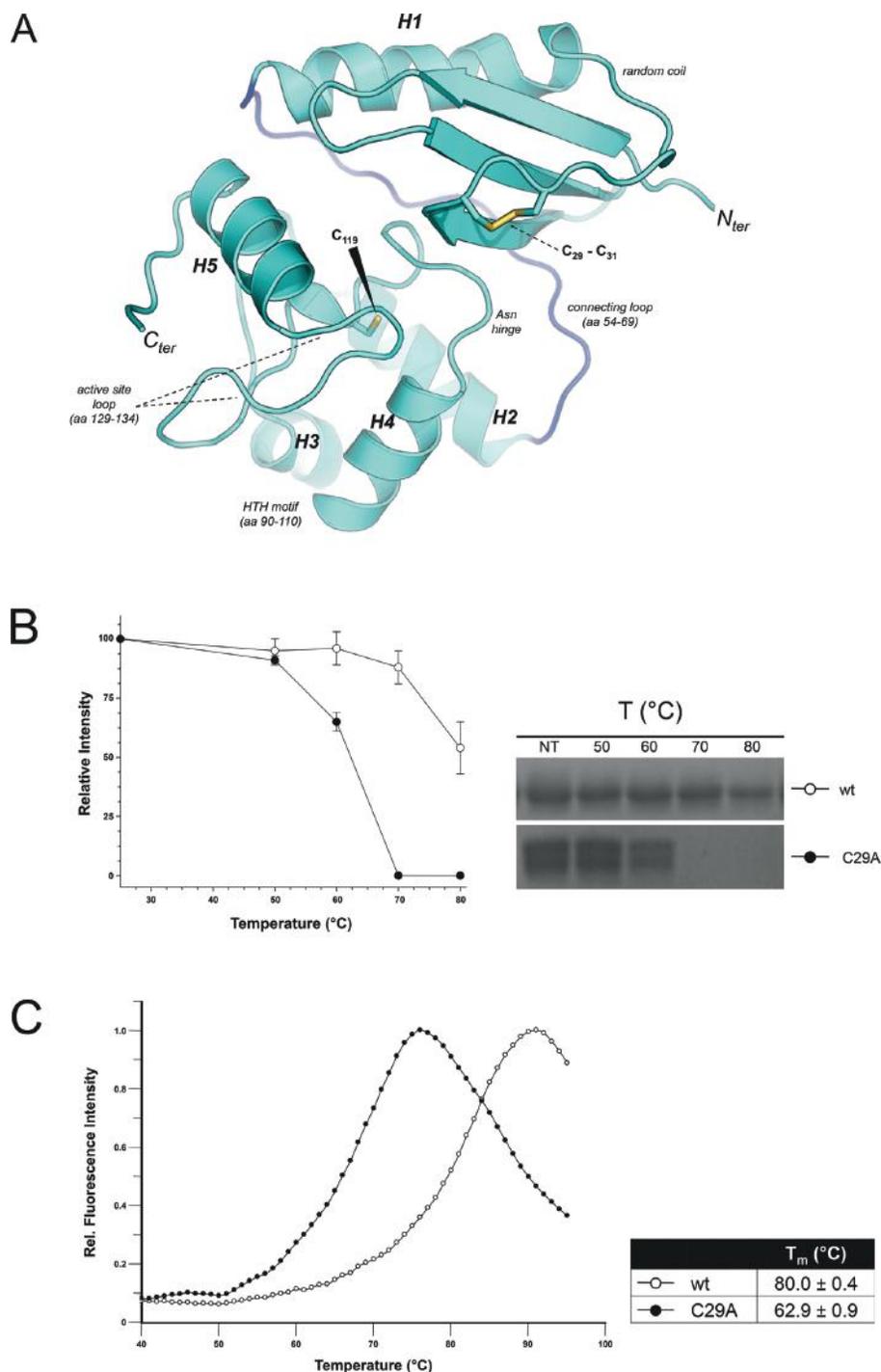


Figure 1. Structure–function of *SsOGT*. (A) Cartoon representations of the crystal structure of wild-type ligand-free *SsOGT*. Secondary structure elements and functional domains/motifs are indicated; residues commented through the text appears as sticks. (B) Thermal stability of the C29A mutant. The indicated proteins (0.1 mg/ml, 5 μ M) were incubated for 20 min at the indicated temperatures; after incubation, samples were centrifuged for 20 min at 16 000 \times *g* and supernatants were analyzed by SDS-PAGE (left); the relative intensity of the Coomassie-stained bands were plotted as a function of temperature (right), considering as 100% the reference value at 25°C for each protein; shown is the mean \pm SD from three independent experiments. (C) Differential Scanning Fluorimetry (DSF). The relative fluorescence intensity values of each protein as a function of temperature were measured and used to obtain the T_m values. Data were obtained from three independent experiments.

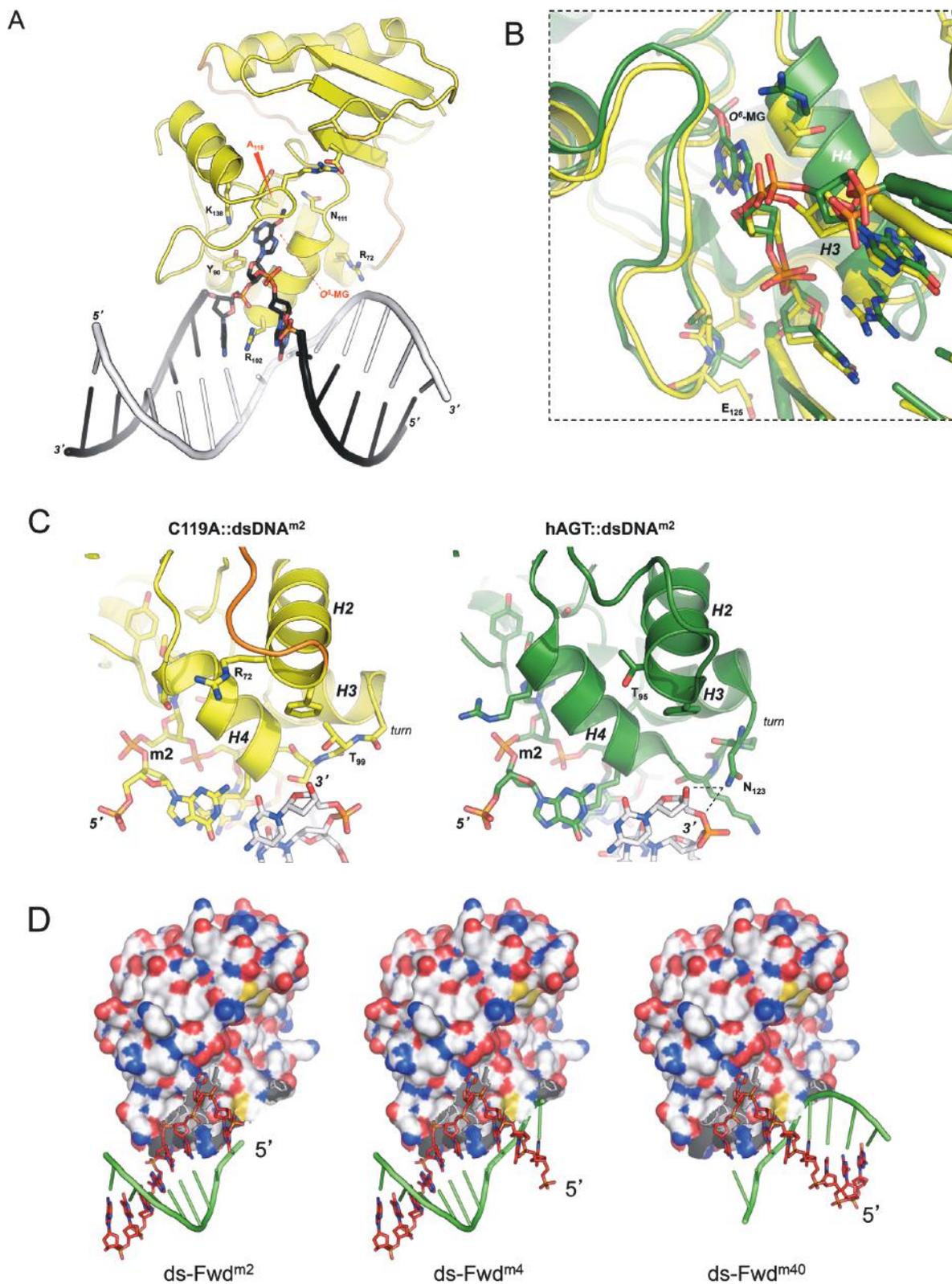


Figure 2. Structure-function of *SsOGT*-DNA complex. (A) Cartoon representations of the crystal structure of the C119A mutant in complex with dsDNA^m. (B) Detail of the interaction of superposed C119A (yellow) and hAGT (green) (PDB ID: 1T38) with the *O*⁶-MG ds DNA at the 3' side of the lesion; (C) left, detail of the C119A::ds DNA^{m2} interactions at the 5' side of the lesion; right, detail of the hAGT::ds DNA^{m2} interactions at the 5' side of the lesion; in both images the damaged DNA strand is colored as the corresponding protein chain. (D) Structure-based models of C119A::ds oligonucleotides carrying the lesion in different positions (see Supplementary Table S1).

Table 1. Data collection, phasing, and refinement statistics

	SsOGT	SsOGT ^m	SsOGT-C119L	C119A::dsDNA ^m
Data collection				
Space Group	R3	P212121	I432	P212121
Unit cell (Å)	a = b = 94.72	a = 48.49 b = 50.22 c = 142.02	a = b = c = 140.56	a = 41.76 b = 65.88 c = 97.52
	c = 76.70 α = β = γ = 90°	α = β = γ = 90°	α = β = γ = 90°	α = β = γ = 90°
Wavelength (Å)	0.873	0.979	1.89	0.972
Resolution (Å)	1.8	2.8	2.6	2.7
Total reflections	83580	35418	78410	41645
Unique reflections	21897	8764	7600	7960
Mean(I)/sd(I)	11.4 (2.43) ^a	10.1 (3.1) ^a	25.6 (5.3) ^a	10.62 (1.7) ^a
Completeness (%)	99.9 (99.8) ^a	97.5 (98.7) ^a	99.9 (100) ^a	98.5 (90.9) ^a
Multiplicity	3.8 (3.7) ^a	4.0 (4.1) ^a	10.3 (10.3) ^a	5.2 (5.2) ^a
R _{merge} (%)	6.7	12.9	7.3	8.1
R _{meas} (%)	7.8	14.8	7.7	9.0
Refinement				
R _{factor} /R _{free} (%)	16.0/18.2	19.0/29.0	17.5/24.2	21.3/27.1
Protein/DNA Atoms	1204	2380	1175	1707
Ligand atoms	24	-	-	-
Water molecules	211	12	45	6
R.m.s.d. bonds (Å)	0.007	0.010	0.009	0.012
R.m.s.d. angles (°)	0.98	1.27	1.31	1.29
Average B (Å ²)				
Protein	25.3	17.9	35.3	63.8
Solvent	38.9	11.0	39.4	49.6

^aValues in parentheses refer to the highest resolution shell.

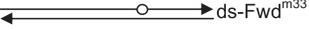
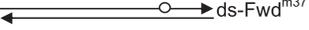
the structure of the C119A::dsDNA^m complex is similar to that of the physiologic complex formed by the wild-type protein, our analysis indicates that DNA binding does not substantially alter the protein architecture, as also shown for hAGT (8).

The C119A::dsDNA^m crystal structure was largely, but not completely, similar to that described for hAGT in complex with the same dsDNA^m substrate. The interaction network established by both proteins with the region of the modified strand at the 3' side of the methylated base appears conserved. Indeed, in the C119A::dsDNA^m complex the phosphodiester bond between bases +1 and +2 downstream of O⁶-MG is clamped between the positive dipole momentum at the N-side of the H3 helix of the HTH domain and the backbone nitrogen of E125, at a mean 3.0 Å distance (Figure 2B). From this analysis we predict that at least 2 bases downstream to the O⁶-MG are needed to establish correct interactions, as shown for hAGT (40). In contrast, the protein-dsDNA interactions at the 5' end of the O⁶-MG base appear less tight in the C119A::dsDNA^m structure compared to the human counterpart. Indeed, we found important differences in structure-based models of SsOGT and hAGT in complex with a dsDNA^{m2}, an oligonucleotide carrying the O⁶-MG lesion located one base from the 5' end of the damaged strand (Supplementary Table S1). In the C119A::dsDNA^{m2} complex, the phosphodiester bond joining the last two bases at the 3' end of the complementary intact strand is not engaged in contacts with the protein (Figure 2C, left panel), while it appears kept in place by the side chain of N123 in an equivalent model of the hAGT::dsDNA^{m2} complex (Figure 2C, right panel).

hAGT-catalyzed repair of dsDNA molecules containing a single O⁶-alkylguanine is directionally biased, as alkylated

bases at positions near the 3' end of the modified strand are repaired less efficiently than those located in the middle or at the 5' end (40). We reasoned that, if this behavior reflects the protein-DNA complex architecture, SsOGT might show a different bias toward the position of the lesion. To test this hypothesis we modified our previously developed competitive kinetic assay for O⁶-BG sensitive AGTs activity (17,19). Briefly, AGT proteins become covalently labeled when incubated with the fluorescent competitive inhibitor SNAP-Vista GreenTM (hereafter BG-VG); the fluorescence intensity of the protein bands is a direct measure of the protein trans-alkylation activity. Moreover, the fluorescence intensity obtained in competition assays with BG-VG and unlabeled alkylated DNA is an indirect measure of the efficiency of DNA repair (17). This method was successfully applied to determination of kinetic constants for DNA trans-alkylation reaction and DNA repair activity by SsOGT and *Mycobacterium tuberculosis* OGT (17,19). We have modified this assay to allow determination of DNA repair activity by SsOGT by measuring the protein fluorescence intensity in competition assays performed in the presence of fixed amounts of BG-VG and increasing amounts of ds oligonucleotides containing a single O⁶-MG. This method allows rapid determination of an IC₅₀, which can be converted to K_{DNA}, giving an indirect measure of the efficiency of O⁶-MG repair by SsOGT (Supplementary Figure S1A). The method was validated by determining the K_{DNA} of SsOGT for the ds-UP^m oligonucleotide (Supplementary Table S1), which was comparable to the value previously reported (17) (Supplementary Figure S1A). We thus applied this method to determine the efficiency of repair by SsOGT of 41 bp dsDNA substrates, each containing a single O⁶-MG at different posi-

Table 2. Polarity of lesion recognition by *Ss*OGT

oligonucleotide	IC ₅₀ (μ M)	K _{DNA} (μ M)
 ds-Fwd ^{m2}	6.38 ± 1.20	4.29 ± 0.39
 ds-Fwd ^{m4}	1.01 ± 0.08	0.68 ± 0.03
 ds-Fwd ^{m26}	1.23 ± 0.07	0.83 ± 0.02
 ds-Fwd ^{m33}	0.73 ± 0.11	0.49 ± 0.04
 ds-Fwd ^{m37}	0.89 ± 0.11	0.59 ± 0.04
 ds-Fwd ^{m39}	1.10 ± 0.12	0.74 ± 0.04
 ds-Fwd ^{m40}	84.3 ± 73.2	56.6 ± 23.8

The ds oligonucleotides are named according to the position of the O⁶-MG with respect to the 5' end of the upper strand. Arrows indicate the 3' end of each strand. Each IC₅₀ value was obtained from three independent experiments.

tions along the molecule (Supplementary Table S1; see also Figure 2D). The efficiency of repair was comparable when the methylated base was located in the middle or within 3 bases from either the 5' or the 3' end. In contrast, the efficiency of repair was significantly reduced (about 6-fold) when the damaged base was located one base from the 5' end, and dropped dramatically (more than 50-fold) when the lesion was located one base from the 3' end (Table 2). Solution binding and AFM studies clearly indicated that hAGT binds both single- and double-stranded substrates in a cooperative fashion (38,39,41–44); the wild-type *Ss*OGT also showed cooperative binding when analyzed in EMSA assays with short oligonucleotides (17). The C119A protein showed similar behavior and bound all methylated oligonucleotides with comparable efficiency (Supplementary Figure S3), thus suggesting that the observed differences reflect the efficiency of O⁶-MG recognition and/or removal, rather than of unspecific DNA binding.

These results show that *Ss*OGT, like hAGT, is highly inefficient in repairing lesions near the 3' end of the molecule, in line with the conserved network of interactions formed with the DNA strand on the 3' side of the lesion by both proteins (Figure 2B). In contrast, whereas two bases from the 5' end of the molecule are sufficient for efficient repair by hAGT, *Ss*OGT requires at least 4 bases between O⁶-MG and the 5' end for optimal activity (a situation illustrated by the central miniature in Figure 2D); in this latter case, positively charged residues of both helix H2 (e.g. R72 in Figure 2A) and helix H4 could efficiently contact the sugar-phosphate backbone of the complementary strand 3' end, fully restoring the protein-DNA association potential.

Effect of alkylation on *Ss*OGT stability

Methylated and benzylated forms of hAGT are highly unstable (37), and soaking of hAGT crystals in O⁶-alkyl-guanine substrates leads to crystals destabilization (7). Considering that *Ss*OGT is extremely stable at high temperature, but is also active at room temperature (17), we wondered whether it might be alkylated and remain relatively stable if kept at temperatures below its optimum. Consis-

tently with this assumption, we obtained a homogeneously methylated protein form (*Ss*OGT^m) by incubating purified *Ss*OGT with O⁶-MG at room temperature. Since the benzylated *Ss*OGT form (*Ss*OGT^b) was not soluble, to test the effect of the presence of larger adducts in the protein active site, we constructed two site-directed mutants carrying substitutions of the catalytic C119 with an F or an L residue, mimicking a benzylated or isopropylated protein, respectively. Whereas the corresponding C145F and C145L mutants of hAGT were extremely unstable when expressed in *E. coli* cells (7), both *Ss*OGT mutants resulted stable in the same expression system and could be purified to homogeneity (data not shown).

When tested in EMSA analysis, *Ss*OGT^m and mutants were able to bind dsDNA, although with slightly reduced affinity as compared with the wild-type protein (Figure 3A), in line with results reported for hAGT, whose dsDNA binding affinity decreases to little extent when the protein is alkylated *in vitro* (37). We then analyzed the stability of the mutants and *Ss*OGT^m at different temperatures. Whereas, as shown above, the wild-type protein was 100% soluble after 20 min incubation at 70°C, complete aggregation was observed for *Ss*OGT^m at 70°C and for C119F at 50°C (Figure 3B). DSF analysis demonstrated the instability of these proteins, showing a *T*_m value of 60°C for *Ss*OGT^m and ca. 45°C for both the C119L and C119F mutants, compared with the wild-type protein (*T*_m = 80°C) (Figure 3C). Thus, the presence of alkylated groups bound to the catalytic C119 leads to *T*_m destabilization, whose extent is dependent on the size of the adduct in the active site.

Effect of alkylation on *Ss*OGT structure

In order to elucidate the structural basis of the dramatic effect of alkylation on protein stability, we solved the crystal structure of the *Ss*OGT^m and *Ss*OGT-C119L proteins at 2.8 and 2.6 Å resolution respectively, whereas we failed to obtain suitable crystals of the C119F mutant, likely as a consequence of its intrinsic instability.

The analysis of the optimally superposed crystal structures of the ligand-free and *Ss*OGT^m proteins revealed conformational changes that occur in discrete regions of the molecule, upon the C119 methylation (Figure 4A). The only comparable structures available so far are those obtained by alkylating hAGT after crystallization (7); similarly to what observed in the human counterpart, in the *Ss*OGT^m structure we observed a 1.0 Å movement of the C-side of the recognition helix H4 and following Asn-hinge, a conformational change responsible for a modest increase of the ligand binding site size (Figure 4A). In addition, in the structure of *Ss*OGT^m the distance between the C-side of the conserved H2 helix and the H4 recognition helix is increased of approximately 3 Å with respect to the observed distance between the same structural elements in the ligand-free wild-type protein. Similarly, the distance between the active site loop and the H4 is increased of 1.9 Å in the *Ss*OGT^m structure. Alkylation of hAGT crystals induce shift of 0.5–1.5 Å C α of the recognition helix away from the N-terminal domain (7).

An unbiased difference distance matrix plot of *Ss*OGT^m versus *Ss*OGT revealed that many interactions between

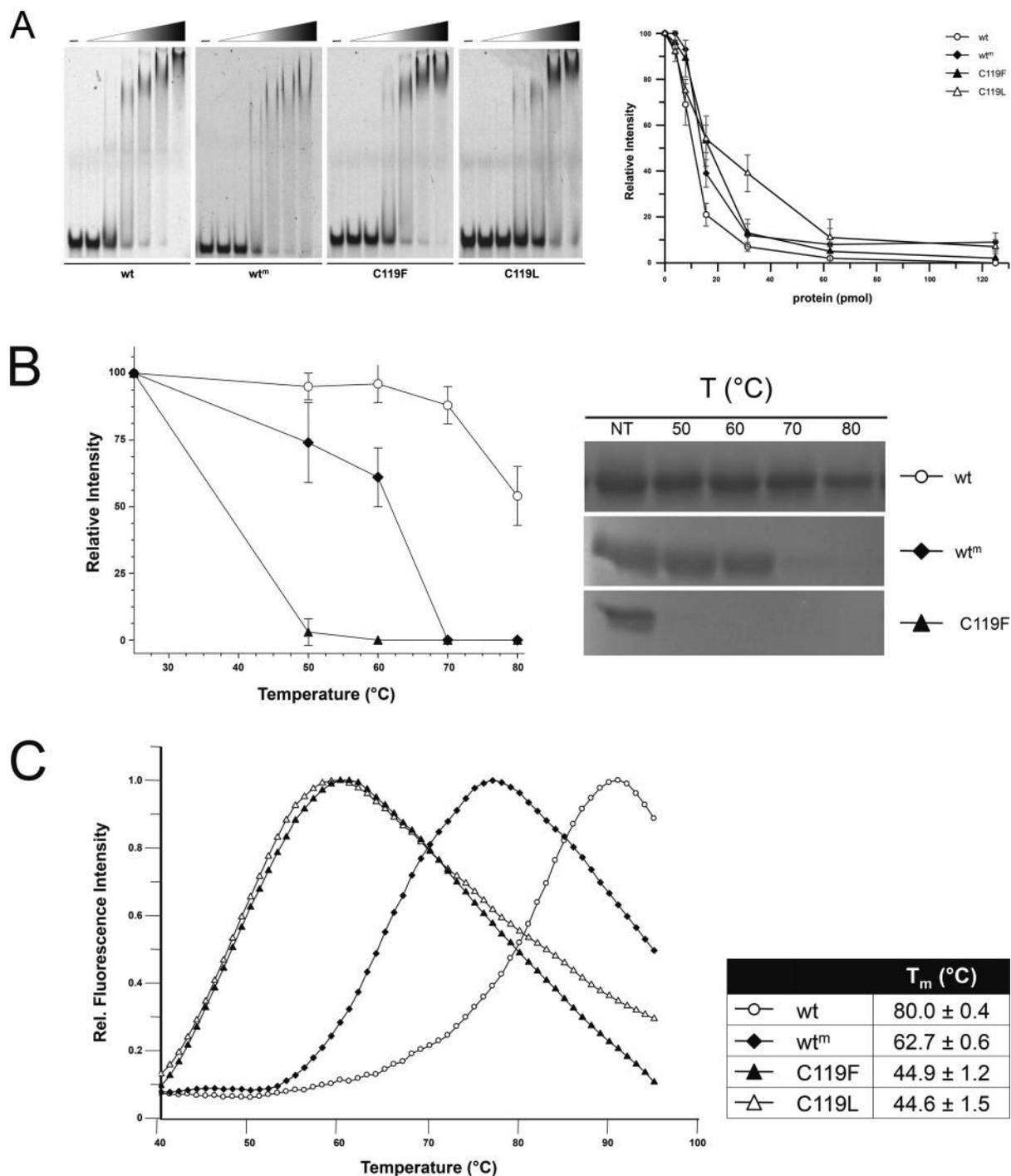


Figure 3. Effect of active site alkylation on *Ss*OGT activity and stability. **(A)** EMSA. Reactions (10.0 μ l) containing increasing amounts of each indicated protein (0–25.0 μ M) were incubated with the TAMRA-labeled ds A^+/A^- oligonucleotide (0.2 μ M; Supplementary Table S1) for 10 min at 37°C. Native polyacrylamide gels were analyzed by gel fluorescence imaging. The first lane of each gel is the no protein control. For quantification (right), the relative intensity of each band was plotted as a function of the protein concentration (pmol); data are from three independent experiments. **(B)** Thermal stability. Reactions were set and analyzed and quantified as described in the legend to Figure 1B. **(C)** DSF. Data were obtained from three independent experiments as described in the legend to Figure 1C.

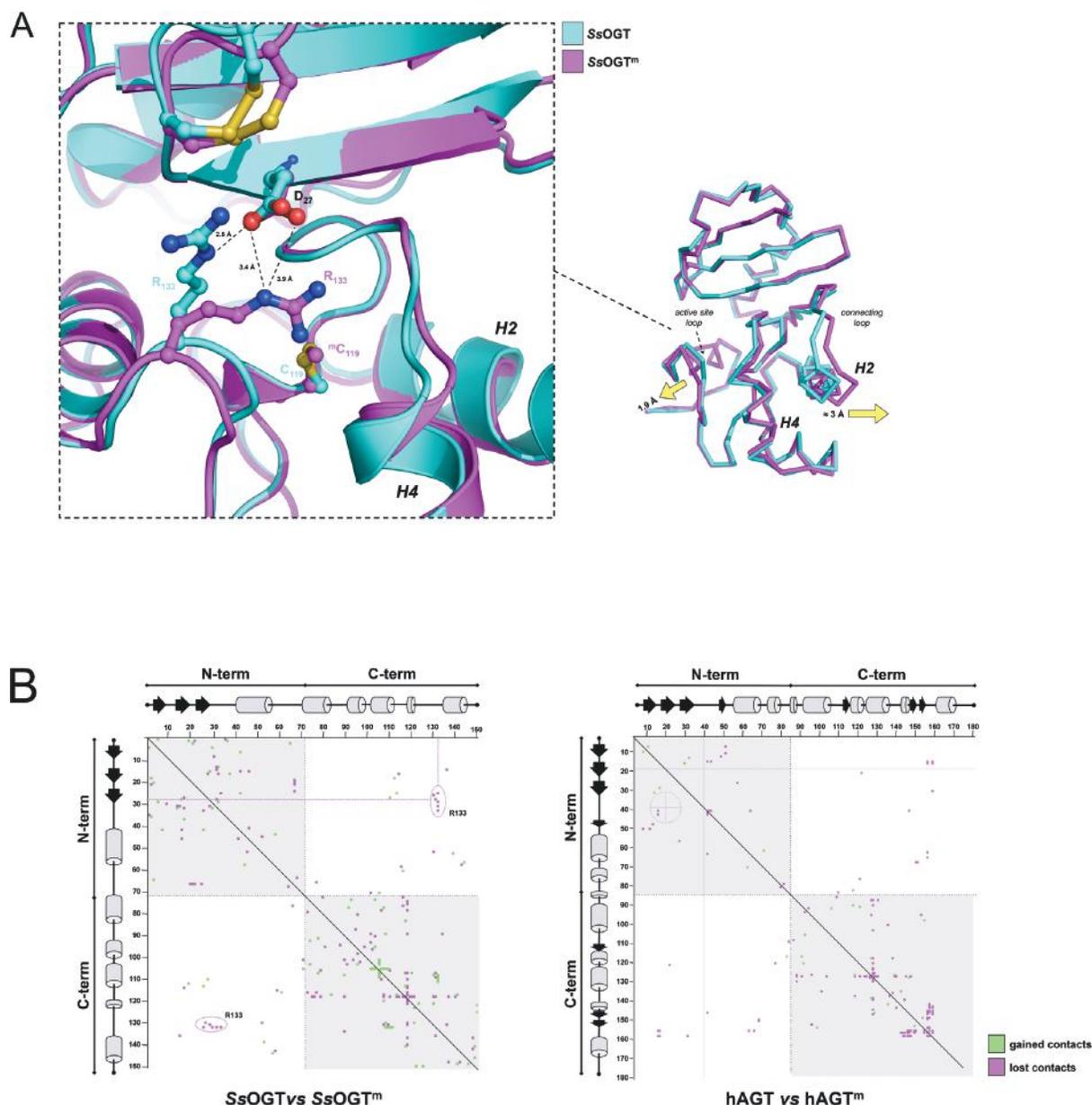


Figure 4. Comparative analysis of structures of free and methylated *SsOGT*. (A) Ribbon representation of the crystal structures of the two proteins upon optimal overlaying. Each protein chain is uniformly colored following the legend on the right of the picture. The arrows indicate the movements described in the text. An enlarged view of the D27 and R133 residues (drawn as sticks) appears in the inset on the left. (B) Distance difference matrix plot, obtained by the computation of Needleman-Wunsch sequence alignment by using the CMView v1.1.1 freeware, choosing amino acid side chain contact type and 8 Å cutoff. (left panel) methylated versus native *SsOGT*; (right panel) methylated (PDB ID: 1EH7) versus native hAGT (PDB ID: 1EH6).

aminoacid residues were lost or gained, suggesting overall reshaping of the molecule upon methylation (Figure 4B, left panel). Compared with analogous analysis performed on alkylated versus free hAGT (Figure 4B, right panel), many more interactions were affected upon *SsOGT* methylation, thus suggesting that crystallization of the protein in its reacted form allow us to observe the repositioning of higher number of residues than that observed in the hAGT methylated in crystallized form. In both proteins most rearrangements occurred within the N- and C-terminal domain, respectively (Figure 4B, gray quarters) rather than between domains (white quarters); however, in *SsOGT*^m a

group of inter-domain interactions were lost in correspondence of the R133 residue (Figure 4B). Interestingly, in the *SsOGT* structure, the R133 residue at the C-side of the active site loop of the C-terminal domain and the D27 residue at the C-side of the third β -strand of the N-terminal domain form an interaction at 2.6 Å distance (Figure 4A, inset). In the *SsOGT*^m structure the distance between the carboxylic group of D27 and the guanidinium group of R133 is increased by 0.6 Å and the R133 side chain is rotated by ca 60° with respect to its position in the unmodified protein on the same plane, overall resulting in a 1.9 Å movement of the active site loop toward the solvent (Figure 4A). Distance

increase from the D27 residue and rotation of the R133 residue was also observed in the superposition of the C119L structure with that of the wild type, although at lower extent (Supplementary Figure S4, inset), suggesting again that the modification of the D27–R133 interaction is a direct consequence of the active site modification. This observation prompted us to hypothesize that the C119 alkylation could have a role in the conformational change of the active site loop, which, in turn, might perturb the D27/R133 interaction, or alter the interaction geometry, ultimately triggering protein destabilization.

The active site expansion is not observed in the crystal structure of the C119L mutant, which rather shows a 1 Å movement of the H2 helix toward the Asn hinge (Supplementary Figure S4); therefore the C119L mutant is not as informative as the *Ss*OGT^m protein with respect to the description of structural rearrangements taking place upon trans-alkylation reaction.

Alkylation-induced structural changes affecting the connection between the two domains impair *Ss*OGT stability and activity

In order to directly test whether the D27/R133 interaction plays a role in *Ss*OGT stability, we prepared two mutants carrying substitutions of the D27 residue, namely D27A and D27K, whose positive charge should eventually enhance the mutation destabilizing effect. In thermal stability assays, the D27A mutant was indeed destabilized with respect to the wild type, showing complete aggregation at 70°C; moreover, the D27K protein was dramatically less stable and aggregated completely above 50°C (Figure 5A); the T_m determined by DSF was of 72°C for D27A and 44.7°C for D27K, respectively (Figure 5B). Thus, these data confirmed our prediction that the interaction between D27 and R133 plays a crucial role in *Ss*OGT stability.

We then characterized the D27A and D27K activity. Both mutants were effective in dsDNA binding, although the D27K mutant showed slightly reduced binding efficiency as compared with the wild-type and the D27A proteins (Figure 5C). In addition, mutated proteins were tested in their DNA repair activity by using the fluorescent assay with *O*⁶-MG containing ds oligonucleotides described in Supplementary Figure S1. While the D27A mutant showed slight reduction of repair efficiency, as compared with the wild-type (IC_{50} of 2.26 ± 0.13 versus 1.23 ± 0.07 μM at 50°C), we could not calculate an IC_{50} value for the D27K protein at either 50 or 25°C, thus suggesting that the D27K protein is unable to repair DNA, or this activity is greatly impaired. A slow DNA repair activity by the D27K protein could be detected in prolonged (6 h) incubation at 25°C with methylated oligonucleotides (Figure 5D; note that under the same conditions the wild-type *Ss*OGT completes the reaction within few minutes; data not shown); thus, the D27K protein is not completely inactive, yet its DNA repair activity is strongly compromised. This result suggests that the disruption of the D27–R133 interaction not only affects the protein stability, but also its DNA repair activity even at low temperature, thus not as a consequence of thermal induced denaturation. To understand the reason of the D27 mutants DNA repair defect, we then determined

the efficiency of the alkyl-transfer reaction independently of DNA binding and lesion recognition, using our previously developed assay based on the use of the synthetic substrate BG-VG (17). Interestingly, only marginal reduction of the catalytic efficiency of the covalent modification reaction at 25°C was observed for the D27K mutant, and no significant changes for the D27A, as compared with the wild-type protein (Figure 5E). Thus, the D27/R133 interaction does not play an important role in the alkyl-transfer reaction from free alkylated bases to the catalytic cysteine, while it greatly affects the efficiency of de-methylation of *O*⁶-MG bases in the context of dsDNA.

DISCUSSION

*Ss*OGT proved to be a convenient model to unravel the structure-function relations of AGTs. Indeed, the overall architecture of the free and DNA-bound protein (this work), as well as details of its reaction mechanisms (17) are conserved between *Ss*OGT and mesophilic counterparts. Most important, the peculiar thermal stability of *Ss*OGT allowed us to obtain the protein in a post reaction form suitable to both biochemical analysis and crystallization, overcoming the restrictions to experimental manipulation imposed by the extreme instability of alkylated hAGT forms.

One peculiar feature of *Ss*OGT is the disulphide bond in its N-terminal domain; this structural element is not found in the corresponding domain of hAGT, which instead contains a zinc atom coordinated by two cysteine and two histidine residues (7). Previously, it was reported that nitric oxide synthase undergoes a conformational change with release of a Zn ion coordinated by four cysteines, coupled to formation of a disulphide bond involving two of such cysteines (45), thus suggesting that an exchange between Zn ion coordination and S-S bond is in principle possible. Although in the nitric oxide synthase this change appears to have regulatory function, it might occur in *Ss*OGT as an artifact due to Zn depletion during purification of the His-tagged protein through metal-chelating columns, inducing formation of the S-S bond. However, this is unlikely to be the case for two reasons: (i) the S-S bond is also present in the OGT protein from the strictly related *Sulfolobus tokodai* species, showing 68% amino acid sequence identity with *Ss*OGT; although no functional data are available for this protein, its crystal structure (PDB entry 1WRJ) was obtained from the protein without any tag (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=1wrj&template=header_records.html&r=getheader); (ii) the *Ss*OGT structure does not reveal any possible canonical tetrad for Zn coordination around the C29–C31 residues. Thus, the S-S bond appears to be a structural peculiarity of the thermophilic protein, and we showed that its disruption significantly destabilizes *Ss*OGT; it is possible that the N-terminal domain is involved in protein activity, as previously suggested (19,46), as well as in the overall stability of AGTs, although by means of different structural elements in different proteins.

The efficiency of DNA repair by *Ss*OGT is highly dependent on the *O*⁶-MG position, and its optimal activity requires at least three bases at both sides of the lesion, consis-

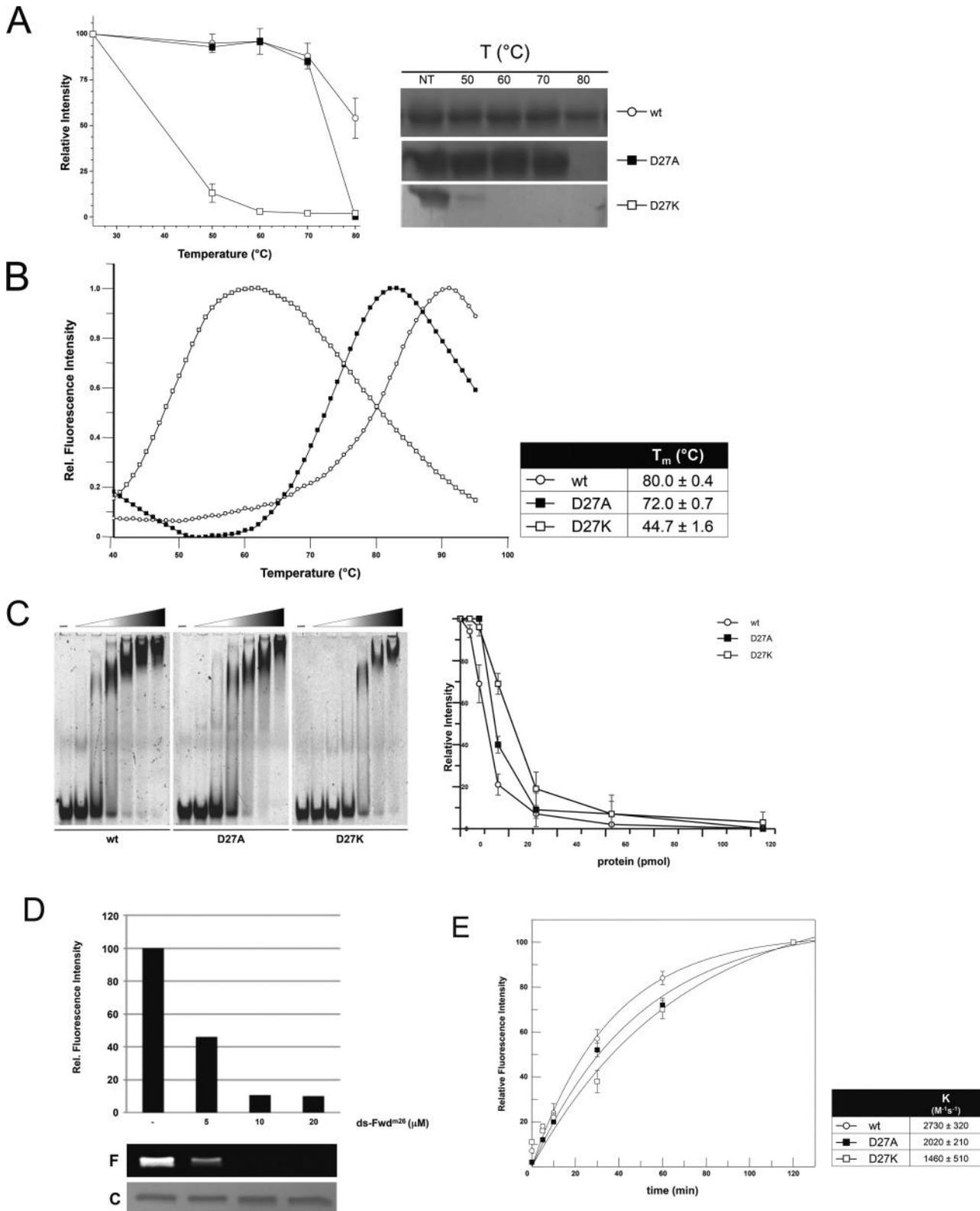


Figure 5. Effect of D27 residue mutations on *SsOGT* stability and activity. **(A)** Thermal stability (see legend to Figure 1B; mean \pm SD from three independent experiments). **(B)** DSF. (see legend to Figure 1C; data from three independent experiments). **(C)** EMSAs were performed and quantified as in the legend to Figure 3A. The first lane of each gel is the no protein control. **(D)** Direct DNA repair activity of *SsOGT* D27K mutant. The protein ($5 \mu M$) was incubated for 6 h at $25^\circ C$ in the presence of increasing amounts of the ds-Fwd^{m26} oligonucleotide, as indicated. Then, $10 \mu M$ of BG-VG was added and the incubation was prolonged for 2 h. The histogram reports the reduction of the fluorescence intensity of the protein band (shown in the gel labeled F) depending on the methylated oligonucleotide concentration in pre-incubation before BG-VG labeling. Equal amounts of protein were loaded in each lane, as shown by the Coomassie staining of the gel (C). **(E)** Trans-alkylation activity of D27 mutants. Second-order rate constants were obtained incubating the proteins at $25^\circ C$ with BG-VG for increasing time spans; data were fitted as previously reported (17).

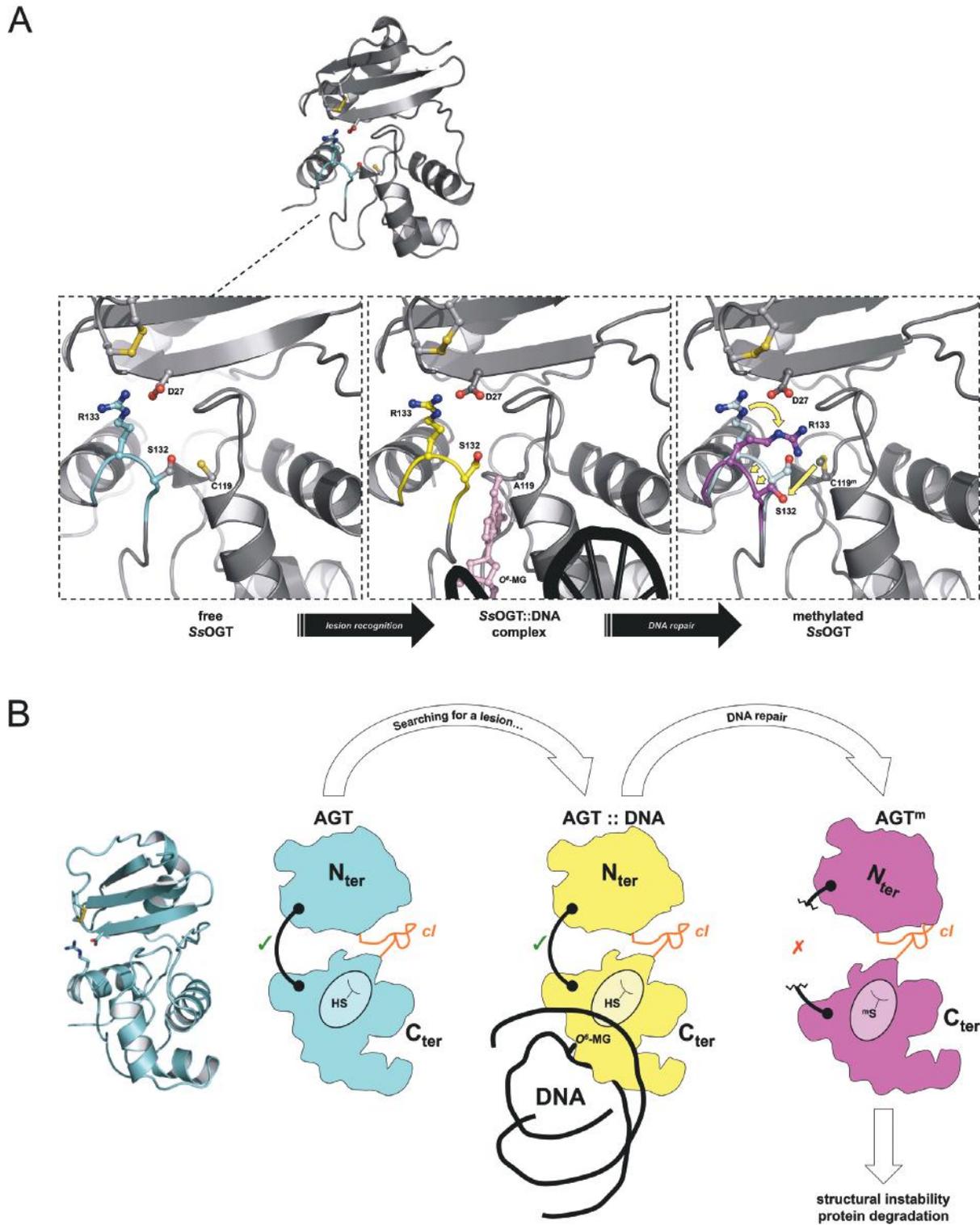


Figure 6. (A) Details of the D27–R133 interaction during the reaction cycle and its methylation-induced modification. (B) Model of the conformation modification induced by alkylation-induced perturbation of inter-domain interactions.

tent with the contacts between the protein and both DNA strands in crystal structure. Interestingly, whereas contacts formed with the DNA strand at the 3' side of the lesion are important for both hAGT and SsOGT, those formed with bases at the 5' side seem more important for the latter. One possible explanation for this difference is that SsOGT needs to form stronger contacts at both sides of the DNA substrate to assist stabilization of the DNA-protein complex at higher temperatures, and/or hold the bound DNA in a correct ds conformation, counteracting thermal denaturation and facilitating lesion removal.

Alkylation-induced instability of hAGTs is a well known process whose molecular mechanism is, however, poorly understood, mainly due to the instability of alkylated hAGT and its C145F and C145L mutants (7,37). Thanks to the relative stability of SsOGT^m, and of C119L and C119F mutants at mild temperatures, we could obtain direct and quantitative data on the protein stability in correlation with the active site status, as well as insights in the structural modifications occurring upon methylation in solution, thus in conditions more physiological than those in which the structures of alkylated hATG could be obtained (7). Indeed, although structural rearrangements were observed upon alkylation of hAGT crystals, Daniels and coworkers predicted that such rearrangements would be even larger in solution (7). Consistently, we found extensive remodeling of interactions between aminoacid residues upon methylation and larger movements in the backbone structure of SsOGT^m, as compared with alkylated hAGT. These data support the correlation between active site alkylation, conformational changes and protein unfolding (7). Our data are also in line with the observation that alkylated SsOGT undergoes degradation after treatment of *S. solfataricus* cells with alkylating agents (17): at the physiological growth temperature (75–80°C) alkylated SsOGT is destabilized, which might target the protein to degradation pathways, either directly or after some still unidentified post-translational modification.

Our structural and biochemical data show that the D27 residue of the N-terminal domain plays an important role in both SsOGT activity and stability, through the formation of an interaction with the R133 residue of the catalytic C-terminal domain (Figure 6A). Intriguingly, the remarkable extent of D27K destabilization is strikingly similar to that of the C119F and C119L mutants, showing that the same effect on protein stability is obtained by acting on two completely different residues. Moreover, the D27/R133 interaction is important not only to maintain SsOGT folding at high temperature, but also to allow its activity even at low temperature. Exploiting our different assays, which allow dissection of the SsOGT reaction, we showed that the D27–R133 interaction is not involved in the trans-alkylation reaction *per se*, whereas it is required for the protein to repair efficiently the alkylated base in the DNA context. An attractive hypothesis is that C119 alkylation-induced perturbation of the D27–R133 interaction weakens the contacts between the two domains, impairing optimal co-ordination between the N- and C-terminal protein domains, which in turn might be crucial both for DNA repair and maintenance of the correct protein folding.

We could also obtain the first crystal structure of an AGT mutant carrying a substitution of the catalytic cysteine with a leucine, mimicking an isopropyl group; the hAGT corresponding mutant was extremely unstable (7). The crystal structure of the C119L protein showed that the conformational modifications occurring in this mutant at the level of the D27–R133 interaction recall those occurring in the methylated protein. However, it is also important to note that not all the structural modifications observed in SsOGT^m are also present in the C119L protein, thus suggesting that the bulky adduct in the active site inserted by mutation may not completely reproduce the effect of the conformational changes triggered by alkylation. Indeed, alkylation takes place in the correctly folded protein, whereas the mutant accommodates the large adduct during its folding, which might affect the protein conformation in a different manner. These results suggest that structural and functional data obtained with substitution mutants should be carefully considered.

The inter-domain D27–R133 interaction is not conserved in the structures of other AGTs, as a consequence of poor sequence conservation at level of the N-terminal domain of AGTs from different species. However, in the case of *Thermococcus kodakaraensis* Tk-MGMT the two domains are connected by an ion-pair network formed by the R50–E93–R132 residues. Although no data are available on the effect of alkylation on this interaction, mutation to alanine of the E93 residue, located at the center of the network, destabilized the protein (32,47). In hAGT, alkylation causes disruption of the active site hydrogen-bond network and perturbs the hydrophobic packing between the N-hinge and the N-terminal domain, which gives an important contribution to the interface between the two protein domains (7). Upon alkylation, movements of the HTH helix H6 and collision of the N137 residue with the alkyl adduct disrupt three H-bonds formed by the N137 residue. Most important, mutation of N137 to alanine resulted in dramatic hAGT destabilization (48). Thus, although in three AGTs the connection between the two domains are provided by different structural elements, these observation suggest a common theme for alkylation-induced destabilization through perturbation of the connection between the two domains, triggering protein destabilization (Figure 6B).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

FIRB-Futuro in Ricerca RBFR12001G_002-Nematic and Merit RBNE08YFN3—Molecular Oncology. Funding for open access charge: FIRB-Futuro in Ricerca RBFR12001G_002-Nematic and Merit RBNE08YFN3—Molecular Oncology.

Conflict of interest statement. None declared.

REFERENCES

1. Pegg, A.E. (2000) Repair of O(6)-alkylguanine by alkyltransferases. *Mutat. Res.*, **462**, 83–100.

2. Daniels, D.S. and Tainer, J.A. (2000) Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O(6)-alkylguanine-DNA alkyltransferase. *Mutat. Res.*, **460**, 151–163.
3. 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*, **6**, 1100–1115.
4. 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.
5. Sabharwal, A. and Middleton, M.R. (2006) Exploiting the role of O6-methylguanine-DNA-methyltransferase (MGMT) in cancer therapy. *Curr. Opin. Pharmacol.*, **6**, 355–363.
6. 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.
7. 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.
8. Daniels, D.S., Mol, C.D., Arvai, A.S., Kanugula, S., 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.
9. 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.
10. Philip, S., Swaminathan, S., Kuznetsov, S.G., Kanugula, S., Biswas, K., Chang, S., Loktionova, N.A., Haines, D.C., Kaldis, P., Pegg, A.E. *et al.* (2008) Degradation of BRCA2 in alkyltransferase-mediated DNA repair and its clinical implications. *Cancer Res.*, **68**, 9973–9981.
11. 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.
12. Kanugula, S., Goodtzova, K. and Pegg, A.E. (1998) Probing of conformational changes in human O6-alkylguanine-DNA alkyltransferase protein in its alkylated and DNA-bound states by limited proteolysis. *Biochem. J.*, **329**, 545–550.
13. Rasimas, J.J., Dalessio, P.A., Ropson, I.J., Pegg, A.E. and Fried, M.G. (2004) Active-site alkylation destabilizes human O6-alkylguanine DNA alkyltransferase. *Protein Sci.*, **13**, 301–305.
14. Brunk, E., Mollwitz, B. and Rothlisberger, U. (2013) Mechanism to trigger unfolding in O(6)-alkylguanine-DNA alkyltransferase. *Chembiochem.*, **14**, 703–710.
15. Valenti, A., Napoli, A., Ferrara, M.C., Nadal, M., Rossi, M. and Ciaramella, M. (2006) Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.*, **34**, 2098–2108.
16. Valenti, A., Perugino, G., Nohmi, T., Rossi, M. and Ciaramella, M. (2009) Inhibition of translesion DNA polymerase by archaeal reverse gyrase. *Nucleic Acids Res.*, **37**, 4287–4295.
17. Perugino, G., Vettone, A., Illiano, G., Valenti, A., Ferrara, M.C., Rossi, M. and Ciaramella, M. (2012) Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J. Biol. Chem.*, **287**, 4222–4231.
18. Vettone, A., Perugino, G., Rossi, M., Valenti, A. and Ciaramella, M. (2014) Genome stability: recent insights in the topoisomerase reverse gyrase and thermophilic DNA alkyltransferase. *Extremophiles*, **18**, 895–904.
19. 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.
20. Gautier, A., Juillerat, A., Heinis, C., Corrèa, I.R. Jr, Kindermann, M., Beauvais, F. and Johnsson, K. (2008) An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.*, **15**, 128–136.
21. Cheng, Y. and Prusoff, W.H. (1973) Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I₅₀) of an enzymatic reaction. *Biochem. Pharmacol.*, **22**, 3099–3108.
22. Niesen, F.H., Berglund, H. and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.*, **2**, 2212–2221.
23. Leatherbarrow, R.J. (2004) *GraFit, version 5.0*, Erithacus Software Ltd., Staines.
24. Kabsch, W. (2010) XDS. *Acta Crystallogr. D*, **66**, 125–132.
25. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D*, **67**, 235–242.
26. 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.
27. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D*, **66**, 213–221.
28. Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J. and Adams, P.D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D*, **64**, 61–69.
29. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D*, **66**, 486–501.
30. DeLano, W.L. (2002) *The PyMOL molecular graphics system*. DeLano Scientific, Palo Alto, CA.
31. Hashimoto, H., Nishioka, M., Inoue, T., Fujiwara, S., Takagi, M., Imanaka, T. and Kai, Y. (1998) Crystallization and preliminary X-ray crystallographic analysis of archaeal O6-methylguanine-DNA methyltransferase. *Acta Crystallogr. D Biol.*, **54**, 1395–1396.
32. 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.
33. Roberts, A., Pelton, J.G. and Wemmer, D.E. (2006) Structural studies of MJ1529, an O6-methylguanine-DNA methyltransferase. *Magn. Reson. Chem.*, **44**, 71–82.
34. 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.
35. Tubbs, J.L. and Tainer, J.A. (2010). Alkyltransferase-like proteins: Molecular switches between DNA repair pathways. *Cell. Mol. Life Sci.*, **67**, 3749–3762.
36. Tubbs, J.L., Latypov, V., Kanugula, S., Butt, A., Melikishvili, M., Kraehenbuehl, R., Fleck, O., Marriott, A., Watson, A.J., Verbeek, B. *et al.* (2009) Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature*, **459**, 808–813.
37. 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.
38. 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.
39. 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.
40. Luu, K.X., Kanugula, S., Pegg, A.E., Pauly, G.T. and Moschel, R.C. (2002) Repair of oligodeoxyribonucleotides by O(6)-alkylguanine-DNA alkyltransferase. *Biochemistry*, **41**, 8689–8697.
41. Adams, C.A. and Fried, M.G. (2011) Mutations that probe the cooperative assembly of O⁶-alkylguanine-DNA alkyltransferase complexes. *Biochemistry*, **50**, 1590–1598.
42. 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.
43. Tessmer, I. and Fried, M.G. (2014) Insight into the cooperative DNA binding of the O⁶-alkylguanine DNA alkyltransferase. *DNA Repair*, **20**, 14–22.

44. Melikishvili, M. and Fried, M.G. (2012) Lesion-specific DNA-binding and repair activities of human O⁶-alkylguanine DNA alkyltransferase. *Nucleic Acids Res.*, **40**, 9060–9072.
45. Crane, B.R., Rosenfeld, R.J., Arvai, A.S., Ghosh, D.K., Ghosh, S., Tainer, J.A., Stuehr, D.J. and Getzoff, E.D. (1999) N-terminal domain swapping and metal ion binding in nitric oxide synthase dimerization. *EMBO J.*, **18**, 6271–6281.
46. Fang, Q., Kanugula, S. and Pegg, A.E. (2005) Function of domains of human O⁶-alkylguanine-DNA alkyltransferase. *Biochemistry*, **44**, 15396–15405.
47. Nishikori, S., Shiraki, K., Yokota, K., Izumikawa, N., Fujiwara, S., Hashimoto, H., Imanaka, T. and Takagi, M. (2004) Mutational effects on O(6)-methylguanine-DNA methyltransferase from hyperthermophile: contribution of ion-pair network to protein thermostability. *J. Biochem.*, **135**, 525–532.
48. 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.

Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase.

Castrese Morrone*¹, Riccardo Miggiano*², Mario Serpe¹, Alberto Massarotti², Anna Valenti¹, Giovanni del Monaco¹, Mosè Rossi¹, Franca Rossi², Menico Rizzi², Giuseppe Perugino^{¶1} and Maria Ciaramella^{¶1}.

¹Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80125 Naples, Italy.

²DiSF-Dipartimento di Scienze del Farmaco, University of Piemonte Orientale “A. Avogadro”, Via Bovio 6, 28100 Novara, Italy.

*The authors equally contributed to the present work.

¶Corresponding authors.

In this paper, a further step forward of the previous work is reported. We analysed the conformational changes and intramolecular interactions that are perturbed during binding of SsOGT to damaged DNA, before the alkylation reaction is completed. To aim this, we performed biochemical, structural, molecular dynamics and in silico analysis of ligand-free, DNA bound and mutated versions of the protein. We highlight a conformational changes occurring during lesion recognition and after the reaction, allowed us to identify novel interactions contributing to SsOGT stability. Taken as a whole, these results suggest a revision of model for conformational changes of the alkylated form, which leads to protein unfolding and degradation. This further confirming that coordination between the N- and C-terminal domains of SsOGT is important for protein activity and stability.

My contribution was to express in E.coli, purify and characterize biochemically all the above-mentioned forms of the SsOGT and relative single mutants. The group of Prof. Menico Rizzi of the University of Novara solved all 3D structures.



Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase



Castrese Morrone^{a,1}, Riccardo Miggiano^{b,1}, Mario Serpe^a, Alberto Massarotti^b, Anna Valenti^a, Giovanni del Monaco^a, Mosè Rossi^a, Franca Rossi^b, Menico Rizzi^b, Giuseppe Perugino^{a,*}, Maria Ciaramella^{a,*}

^a Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80125 Naples, Italy

^b DiSF-Dipartimento di Scienze del Farmaco, University of Piemonte Orientale 'A. Avogadro', Via Bovio 6, 28100 Novara, Italy

ARTICLE INFO

Article history:

Received 1 August 2016

Received in revised form 7 October 2016

Accepted 20 October 2016

Available online 22 October 2016

Keywords:

Alkylation damage

DNA repair

Fluorescent-based assay

Crystal structure

Conformational changes

ABSTRACT

Background: Alkylated DNA-protein alkyltransferases (AGTs) are conserved proteins that repair alkylation damage in DNA by using a single-step mechanism leading to irreversible alkylation of the catalytic cysteine in the active site. Trans-alkylation induces inactivation and destabilization of the protein, both *in vitro* and *in vivo*, likely triggering conformational changes. A complete picture of structural rearrangements occurring during the reaction cycle is missing, despite considerable interest raised by the peculiarity of AGT reaction, and the contribution of a functional AGT in limiting the efficacy of chemotherapy with alkylating drugs.

Methods: As a model for AGTs we have used a thermostable ortholog from the archaeon *Sulfolobus solfataricus* (SsOGT), performing biochemical, structural, molecular dynamics and *in silico* analysis of ligand-free, DNA-bound and mutated versions of the protein.

Results: Conformational changes occurring during lesion recognition and after the reaction, allowed us to identify a novel interaction network contributing to SsOGT stability, which is perturbed when a bulky adduct between the catalytic cysteine and the alkyl group is formed, a mandatory step toward the permanent protein alkylation.

Conclusions: Our data highlighted conformational changes and perturbation of intramolecular interaction occurring during lesion recognition and catalysis, confirming our previous hypothesis that coordination between the N- and C-terminal domains of SsOGT is important for protein activity and stability.

General significance: A general model of structural rearrangements occurring during the reaction cycle of AGTs is proposed. If confirmed, this model might be a starting point to design strategies to modulate AGT activity in therapeutic settings.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Alkylated-DNA protein alkyltransferases (called AGT, OGT, or MGMT, EC: 2.1.1.63) are specialized proteins that perform the direct repair of alkylation damage in DNA, mainly acting at position *O*⁶ of guanines. They use a peculiar one-step mechanism, in which a single trans-alkylation reaction catalyses the transfer of the alkyl group from DNA to a cysteine residue in the protein active site, restoring the correct DNA structure with no need for other factors or energy source [1,2].

Abbreviations: BG-VG, SNAP-Vista Green™ substrate; C_{ter}, C-terminal domain; DDMP, Difference Distance Matrix Plot; dsDNA^m, *O*⁶-methyl-guanine-containing double-strand DNA; DSF, Differential Scanning Fluorimetry; hAGT, human AGT; MD, molecular dynamics; N_{ter}, N-terminal domain; *O*⁶-BG, *O*⁶-benzyl-guanine; *O*⁶-MG, *O*⁶-methyl-guanine; RMSD, root-mean-square deviation; SsOGT, *S. solfataricus* *O*⁶-alkyl-guanine-DNA-alkyltransferase; SsOGT^m, methylated SsOGT.

* Corresponding authors.

E-mail addresses: giuseppe.perugino@ibbr.cnr.it (G. Perugino), maria.ciaramella@ibbr.cnr.it (M. Ciaramella).

¹ These authors contributed equally to the paper as First Authors.

AGTs are evolutionary highly conserved, and orthologs from humans, bacteria and archaea have been extensively studied [2,3]. In particular, the human protein (hAGT) has raised considerable interest because its activity antagonizes the effect of alkylating drugs widely used in chemotherapy, and its overexpression is frequently associated with resistance to such drugs [2]. Moreover, certain AGT variants have been associated with the virulence of *Mycobacterium tuberculosis* species [4,5].

Previous structural and biochemical studies elucidated details of DNA binding, lesion recognition and repair by AGTs [4–13]. Invariably, the protein architecture consists of a highly conserved C-terminal (C_{ter}) domain, containing both the DNA binding helix–turn–helix motif and the catalytic site; and a N-terminal (N_{ter}) domain, whose function is less understood and has been suggested to assist the correct folding/stability of the C_{ter} domain [14] and contribute to optimal catalysis [4,5]. Upon lesion recognition, the alkylated base is flipped out from the regular base stacking of the double helix, and inserted into the protein active site, allowing the transfer of the alkyl moiety to the catalytic cysteine. The drawback of this elegant reaction is that once alkylated, the protein is irreversibly inactivated and destabilized, both *in vitro*

and *in vivo* [15,16]. In human and yeast cells, the degradation of hAGT occurs through the ubiquitin-proteasome pathway [16] and might also affect other DNA repair pathways [17].

The molecular mechanisms underlying the AGT unfolding and degradation, which follow the alkylation of the active site cysteine, are only partially understood: conformational changes and modification of intramolecular interactions are likely to be associated to each step of the reaction cycle. Evidence of some changes occurring after hAGT methylation has been provided [18–20]; however, the intrinsic instability of alkylated AGTs has been an important limitation to structural and biochemical studies [19].

These limitations have been overcome by the availability of a thermostable DNA alkyl-transferase from the archaeon *Sulfolobus solfataricus* (SsOGT) [11], which proved to be a convenient model for AGTs. In this hyperthermophilic microorganism, alkylating agents action is potentiated by the high temperature of growth, resulting in an apoptotic-like effect, characterized by DNA fragmentation and degradation of key proteins involved in genome stability [3,21–24]. SsOGT is also degraded after treatment with lethal doses of alkylating agents, thus recalling its fate in human cells [11]. Although alkylation reduces the exceptional stability of SsOGT *in vitro*, the alkylated protein is stable enough to allow structural and biochemical analysis at operational temperatures [10,11]. We exploited this unique property to obtain direct structural and biochemical information on SsOGT in its free, DNA-bound and methylated post-reaction form [10]. Importantly, in contrast to the 3D structure of the methylated hAGT (obtained by soaking protein crystals in substrate solutions; [19]), the 3D structure of SsOGT was obtained by crystallizing the protein previously methylated in solution and purified, which might better reflect the physiological conformation of the alkylated protein, allowing free movements that could otherwise be restricted in crystals.

The available SsOGT structures are useful tools as they open the possibility of understanding how alkylation triggers the protein unfolding and describing the conformational changes associated to each step of the reaction. So far, this analysis showed significant changes occurring upon SsOGT active site cysteine (C119) methylation in discrete regions of the protein and allowed the identification of one interaction whose methylation-induced perturbation leads to protein destabilization [10] (see also below).

In this paper we report a further step forward, as we analyse the conformational changes and intramolecular interactions, which are perturbed during binding of SsOGT to methylated DNA, before the transmethylation reaction is completed. Moreover, we solved the crystal structure of the SsOGT-C119F mutant, gaining an indirect insight into the possible conformation adopted by the protein upon the removal of an aromatic and bulky adduct from the DNA. Taken as a whole, the structural, biochemical and molecular dynamics studies of ligand-free, methylated and mutated forms of SsOGT suggested a model for the alkylation-induced unfolding and degradation of this protein, which is of possible relevance also for other AGTs.

2. Materials and methods

2.1. Reagents

All chemicals were from Sigma, and synthetic oligonucleotides were from EuroFins (Milan, Italy). *E. coli* ABLE C and JM109 strains, as well as *Pfu* DNA polymerase were purchased from Stratagene (La Jolla, CA). The SNAP-Vista Green™ substrate (BG-VG) was from New England Biolabs (Ipswich, MA), and SYPRO Orange™ was from Thermo Fisher Scientific (NYSE, TMO).

2.2. DNA site-directed mutagenesis and protein production

Starting from the *S. solfataricus ogt* gene cloned in the pQE31™ vector [11] used as template, mutants were obtained by applying the

GeneTailor™ Site-Directed Mutagenesis System (Invitrogen) method, by using oligonucleotide pairs listed in the Supplementary Table S1. All proteins were heterologously expressed in the *E. coli* ABLE-C or JM109 strain and purified by affinity chromatography and desalting steps, as described [11].

2.3. Determination of the catalytic activity by fluorescent assay

The catalytic activity was measured by using a fluorescent assay we previously set up and validated, based on the use of derivatives of *O*⁶-benzyl-guanine (*O*⁶-BG), which is known to inhibit AGTs by covalent transfer of the benzylic group to the active site cysteine. Our assay employs a fluorescein-conjugated derivative of *O*⁶-BG (SNAP-Vista Green™, hereafter called BG-VG), which labels the protein covalently and stoichiometrically [4,5,10,11,25]. For the catalytic constants determination, fixed SsOGT and BG-VG concentrations were incubated for different time spans; the fluorescence intensity of the labelled protein band in SDS-PAGE was quantified and used to calculate second-order rate constants [11,26].

2.4. Determination of the DNA repair activity by the IC₅₀ method

The efficiency of repair of AGTs depends on their ability to bind DNA, recognize the lesion and repair it: all these functions have been analysed by a unique competition assay according to the IC₅₀ method, where a dsDNA oligonucleotide containing a single *O*⁶-MG was used as a competitor in the reaction of SsOGT with BG-VG. Finally, K_{DNA} constants were calculated from the obtained IC₅₀ values [10,27].

2.5. Differential Scanning Fluorimetry (DSF) assay for the protein stability determination

The stability against thermal denaturation was followed spectrofluorimetrically as described [10], adapting the method proposed by the group of Vedadi [28]. Because of the intrinsically high stability of SsOGT, thermal stability scans from 25.0 to 95.0 °C were performed at a scan rate of 0.2 °C min⁻¹ (5 min per cycle with an increase of 1.0 °C per cycle). Normalized data of the fluorescence intensities vs temperature from three independent experiments were fitted with the Boltzmann equation [10,28] allowing the determination of the T_m values.

2.6. Crystallization and data collection

The purified SsOGT-C119F mutated variant was subjected to robot-assisted crystallization trials (Oryx4; Douglas Instruments) using commercial screens from Hampton Research (Crystal Screen, Crystal Screen II, Peglon and Peglon II) and Qiagen (Classics Suite and Classics Suite II), applying the vapor diffusion method in sitting drop. Protein crystals, in the form of thin needles, grew to their maximum dimensions in three weeks at 4 °C in drops obtained by mixing 0.5 μL of a 5.3 mg mL⁻¹ protein solution and an equal volume of a precipitant solution containing 0.2 M ammonium phosphate monobasic and 20.0% W/V PEG 3350. The single crystals were cryo-protected in precipitant solution containing 15.0% glycerol, mounted in cryo-loop, and flash-frozen in liquid nitrogen at 100 K for subsequent X-ray diffraction analysis. Diffraction datasets were collected at the ID29 beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) equipped with a Pilatus 6M-F, at a wavelength of 1.008 Å. A data collection was performed up to 2.7 Å of resolution and the complete dataset was indexed with IMOSFLM program [29], allowing to assign the crystal to the orthorhombic space group *P*212121, with cell dimensions $a = 52.86$ Å $b = 85.85$ Å $c = 98.46$ Å, containing two molecules per asymmetric unit, with a Matthews coefficient and a solvent content of 2.94 Å³ Da⁻¹ and 58.19%, respectively. Further data analyses were carried out using the CCP4 suite programs COMBAT and SCALA [30].

2.7. Structure determination, model building and refinement

The initial phases for solving the structure of the SsOGT-C119F variant were generated by molecular replacement with the program PHASER [31] of the PHENIX software suite [32,33], using the structure of *Sulfolobus solfataricus* wild type OGT (PDB ID: 4ZYE) as the search model. In this structure, the catalytic cysteine at position 119, the water and the ligand molecules have been omitted. Model building was performed with the program COOT [34] and crystallographic refinement was carried out with PHENIX [32]. Data converged to the R_{factor} and R_{free} values of 0.172 and 0.237, respectively, with good geometry.

2.8. Protein structure accession numbers

The atomic coordinates and structure factors of the SsOGT-C119F mutated variant have been deposited in the Protein Data Bank (<http://www.rcsb.org>) under the accession code ID: 5LLQ.

2.9. Molecular dynamics (MD)

Crystallographic structure of ligand-free SsOGT (PDB ID: 4ZYE) and its mutants C119A (PDB ID: 4ZYD), C119L (PDB ID: 4ZYH) and C119F (PDB ID: 5LLQ) were used for the initial coordinates of the MD simulations. MD simulations were performed using GROMACS 5.1 software [10.1016/j.softx.2015.06.001] with the GROMOS 54a7 force field [10.1007/s00249-011-0700-9] implemented in Intel Xeon Octa Core processor with Linux environment. In our models, basic residues are protonated and acidic residues are unprotonated. Systems were neutralized and solvated in a periodic octahedric box containing a simple point charge (SPC) water model [31]. Two series of simulations were carried out: at a constant temperature of 353 K since SsOGT is a hyper-thermophile enzyme, and at a constant temperature of 500 K to analyse the protein unfolding. Before every MD simulation, the internal constraints were relaxed by energy minimization, followed by equilibration (100 ps at constant temperature and 100 ps at constant pressure) under position restraints of the carbon backbone atoms. During the MD runs, covalent bonds in the protein were constrained using the LINCS algorithm [10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H]. The SETTLE algorithm was used to constrain the geometry of water molecules [10.1002/jcc.540130805]. Berendsen's coupling algorithm was used to maintain the simulation under constant pressure and temperature [10.1063/1.448118]. Van der Waals forces were treated using a cutoff of 1.0 nm. Longrange electrostatic forces (r.1.0 nm) were treated using particle mesh implemented in the Verlet method [10.1103/PhysRev.159.98]. Through the production runs, the trajectory data were saved every 1 ps, and the total duration of the simulations was 200 ns.

Root-mean-square deviations (RMSD) were calculated taking the energy-minimized structure as a reference (Supplementary Fig. S1). Principal components analysis (PCA) of the protein motion was determined from the diagonalization of the covariance matrix of the interatomic fluctuation [10.1002/prot.340170408]. Average conformations were calculated from the variance-covariance matrix of all protein atoms during the equilibrium time of the run. Tools from the GROMACS package were used for the analysis of the data.

2.10. Data analysis

Prism Software Package (GraphPad Software) and GraFit 5.0 Data Analysis Software (Erithacus Software) were used for corrected data fitting using appropriate equations.

Difference Distance Matrix Plots (DDMPs) were obtained by using the CMView v1.1.1 freeware. All the structures used in this analysis (free SsOGT; SsOGT-C119A mutant bound to DNA; methylated SsOGT) were aligned by the computation of Needleman-Wunsch sequence

method, and choosing all types of amino acid contacts with a cut-off of 8.0 Å.

All the figures illustrating structural analysis were generated by Pymol freeware version 1.7 (Schrodinger, LLC., <http://www.pymol.org>) using standard code for amino acid atoms colouring.

3. Results and discussion

3.1. SsOGT conformational modifications associated with O^6 -MG recognition and binding

In order to describe the complete picture of conformational changes occurring throughout the DNA repair reaction, we performed unbiased DDMP analysis, by comparing the network of intramolecular interactions observed in the available 3D structures of SsOGT. A preliminary analysis previously performed on ligand-free vs SsOGT^m structures gave important hints on changes occurring upon protein methylation ([10]; see also below). We have now extended and further refined this analysis by comparing the free, DNA-bound and methylated SsOGT (SsOGT^m) structures, taking into account all intramolecular interactions, including those formed by the atoms of the protein backbone (Fig. 1).

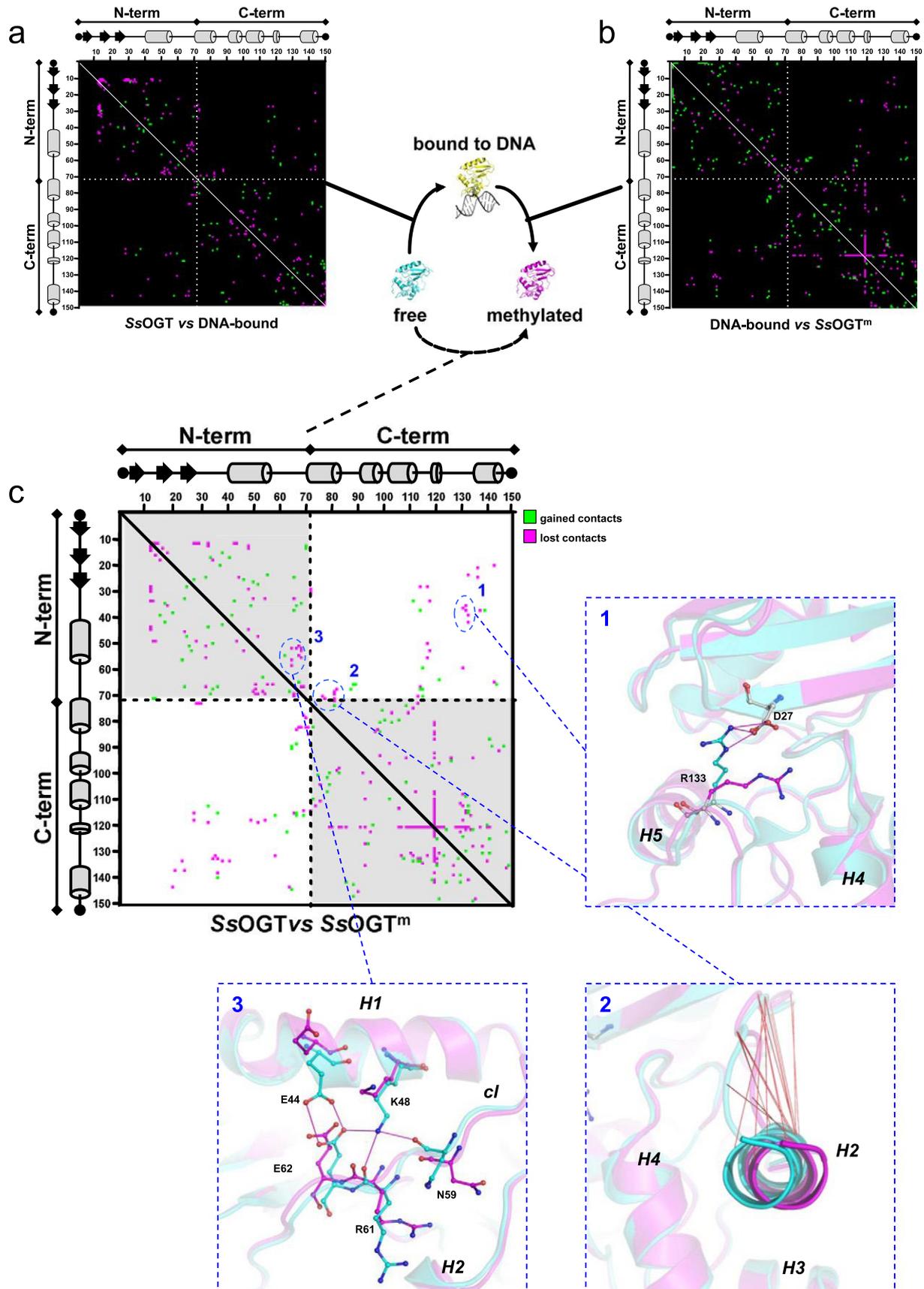
The crystal structure of the SsOGT inactive mutant, C119A, in complex with a methylated dsDNA^m oligonucleotide showed that substrate binding does not affect significantly the protein overall structure [10]; however, structural changes in specific protein regions are expected to occur, in order to accommodate the double helix as well as the methylated base, which adopts an extra-helical configuration in the complex. Indeed, the DDMP analysis comparing the free and DNA-bound protein forms showed an extensive reshaping of the molecule, with movements involving a number of residues and leading to a lot of lost and gained interactions when the protein contacts DNA and flips out the O^6 -MG into its active site (Fig. 1a). The DDMP analysis comparing the DNA-bound and SsOGT^m structures, which corresponds to the DNA-free protein in its post-reaction form, showed an almost symmetrical pattern, with most changes found in the DNA-bound structure no longer present (compare Fig. 1b and a); those changes, which correspond mainly to intra-domain interactions, are thus a consequence of the binding of the natural substrate, as expected from the induced fit model for a genuine enzyme.

However, like all AGTs, SsOGT is not a true enzyme, and irreversible methylation has dramatic consequences on its structure [10]; consistently, DDMP analysis comparing the SsOGT^m with the free and DNA-bound structures showed multiple modified contacts (Fig. 1c). We argued that interdomain interactions should have a major role in securing the overall protein stability, and could be reasonable candidates to trigger the protein destabilization; thus, we focused our attention on three clusters of interactions occurring at the interface between the N_{ter} and C_{ter} domains. Two such clusters were already identified in a previous analysis [10]: in particular, cluster 1 corresponds to perturbation of the interaction between the D27 residue in the N_{ter} domain and the R133 residue in the C_{ter} domain, due to a *ca.* 60° clockwise rotation of the latter (inset 1 in Fig. 1c; d); this modification leads to strong destabilization of the overall structure. The second cluster corresponds to a loss of interactions caused by the *ca.* 3.0 Å movement of the C-side of the conserved H2 helix away from the overall structure upon alkylation (Fig. 1c, inset 2; Fig. 1d). Both modifications were found in the methylated protein, but not in the DNA-bound protein, thus suggesting that they are a direct consequence of the active site alkylation (compare DDMPs in Fig. 1b and c).

In addition, our current analysis highlighted another region where conformational changes occur, in particular at the interface between the N_{ter} domain and the connecting loop (indicated in inset 3 of Fig. 1c; d) which, in contrast to clusters 1 and 2, is common to both DNA-bound and SsOGT^m. This cluster is characterized by a number of lost interactions at the level of a complex network involving the E44,

K48, N59, R61 and E62 residues (hereafter called *K48-network*). In the free SsOGT, these residues are close to each other (2.9 Å distance on average), consistent with an ionic/hydrogen bond network (Fig. 2 and

Table 1). In both the DNA-bound and methylated forms of the protein, the E44 and K48 appear to flip out, moving away from the body of the protein (Fig. 2). The former loses its hydrogen-bond with the E62,



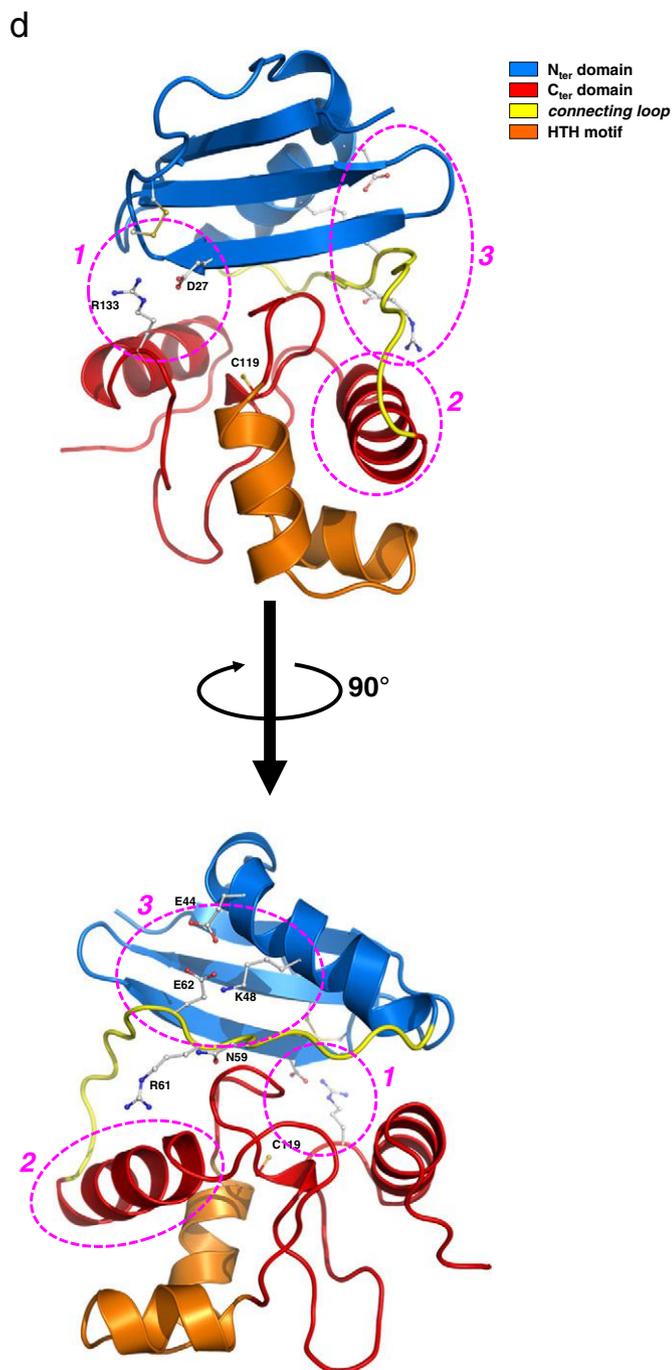


Fig. 1. The stages of the SsOGT activity analysed by DDMP. Available structures were compared in pairs, as indicated, leading to three distinct DDM plots (a, b and c). Clusters of lost (in magenta) and gained (in green) interactions were numbered and described in the text. Insets represent the position of the identified interactions of each corresponding cluster in the free protein 3D structure. SsOGT, the DNA-bound and SsOGT^m were represented in cyan, yellow and magenta colour, respectively. (d) Overall structure of the free SsOGT, where identified clusters are numbered as in (c) and indicated in magenta. Amino acid atoms are coloured according the CPK convention (carbon, in white or in the corresponding colour of each 3D structure; oxygen, in red; nitrogen, in blue, sulphur, in yellow).

whereas the conformational change of the K48 residue has a stronger impact, perturbing its interactions with the N59, R61 and E62 (Fig. 2 and Table 1).

Taken together, these observations suggest that lesion recognition and alkylation trigger distinct modifications of intramolecular interactions in SsOGT: whereas the changes at the level of cluster 1 and 2 occur only upon irreversible trans-alkylation of the catalytic cysteine,

those observed in the K48-network might be a consequence of lesion recognition/steric hindrance of the active site, since they are already found in protein in complex with the methylated DNA and are retained in the post-repair protein structure devoid of DNA. Importantly, all three clusters occur at the interface between the two domains, suggesting that the interdomain communication plays important role in multiple steps of the reaction.

3.2. Structural analysis of the SsOGT-C119F mutated variant

To confirm the hypothesis that the observed rearrangements of the K48-network are linked to the presence of chemical groups in the active site, we analysed the configuration of this network in the C119L mutant (PDB ID: 4ZYH), in which a leucine residue replaces the catalytic cysteine [10]. This mutant showed biochemical and structural features consistent with the hypothesis that the L119 residue mimics an “isopropylation” of the active site [10]. Similarly to what has been observed in the crystal structure of the DNA-bound and SsOGT^m, in the C119L mutant structure a displacement of the E44 residue was observed (Fig. 2); however, most distances among residues in the K48-network were not significantly affected (Fig. 2 and Table 1). The reasons for the difference with respect to the SsOGT^m structure are currently not clear. We previously observed that the C119L protein 3D structure does not completely overlap the SsOGT^m structure [10], suggesting that the C119L mutant might not be an optimal model of SsOGT alkylation. Therefore, we sought to obtain independent experimental support to our hypothesis.

We previously obtained the C119F mutant that, by virtue of the presence of a phenylalanine residue substituting the catalytic cysteine in the active site, mimics a “benzylated” form of the protein. Consistently, biochemical analysis showed that, similarly to the C119L mutant, this protein is also greatly destabilized, and preliminary crystallization trials failed [10]. Similar results were obtained with the C145F mutant of hAGT, which was extremely unstable when heterologously expressed in *E. coli* [19]. Nevertheless, new crystallization trials were successful and allowed the resolution of the SsOGT-C119F structure at 2.7 Å (Fig. 3 and Table 2). Interestingly, the structural analysis of SsOGT-C119F revealed elements of novelty compared to the available AGT structures that represent the protein at the different stages of the alkyltransferase reaction [4,7–10,13,19,36–38]. It has been suggested that alkylation of the active site could trigger movements of the conserved recognition helix belonging to the HTH motif, explaining the *in vivo* destabilization of the reacted form of the protein. However, a comparative structural analysis performed on the ligand-free and alkylated forms of the human enzyme did not reveal significant shifts of the correspondent recognition helix away from the N_{ter} domain of the protein [18,19]. By analysing the SsOGT-C119F structure, we directly observed for the first time a repositioning of the recognition helix H4, probably induced by the need to make room for the alkyl adduct in the active site pocket (i.e. the aromatic moiety of F119 side chain). Concomitantly, we observed a repositioning of the N111 of the highly conserved *Asn-hinge* of the protein (Fig. 3, inset), a residue that is crucial for protein stability; in fact, the mutation to alanine of the equivalent asparagine residue (N137) in the human ortholog resulted in a dramatic protein destabilization [35].

In this new conformation, N111 moves approximately 2.0 Å toward the active site loop that constitutes part of the opposite wall of the catalytic cavity, and establishes a hydrogen bond with the hydroxyl moiety of the S132 residue (Fig. 3). A partial movement of S132 was previously observed in the crystal structure of SsOGT^m and was associated with the loss of interaction between R133 and D27, triggering protein unfolding [10]. Moreover, the structural analysis of SsOGT-C119F confirms our hypothesis that alkylation of the active site induces further amino acid repositioning that alters the protein stability. Indeed, the peculiar arrangement of ionic bonds observed inside the K48-network of the ligand-free protein in the SsOGT-C119F structure is destroyed at the

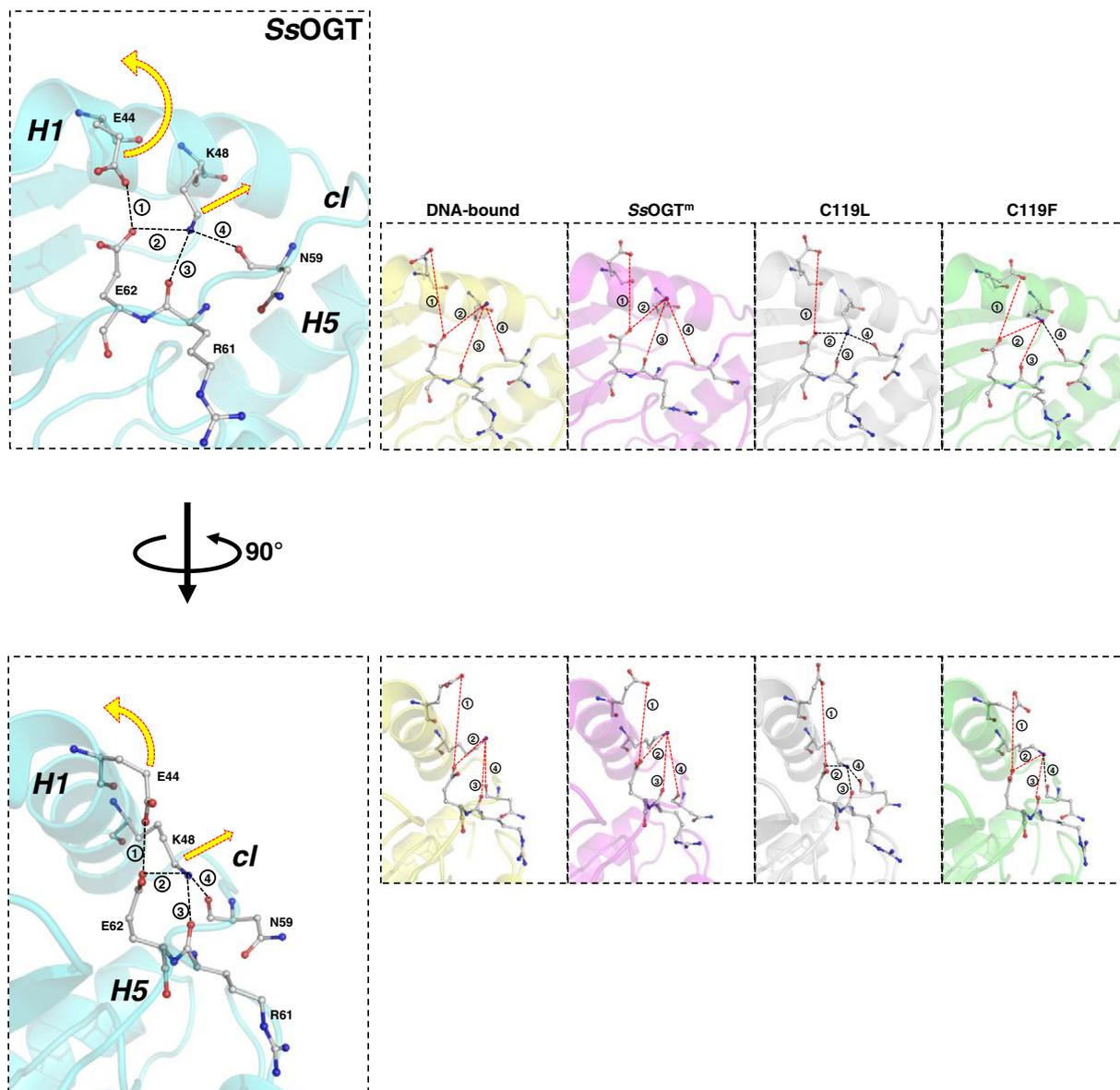


Fig. 2. The *K48*-network in SsOGT. 3D structures are in transparent *cartoon* representation, in a colour code as in Fig. 1. The *K48*-network residues are indicated and displayed in *ball and stick* format. Yellow arrows illustrate the movement of side chains on the structure, during the stages of the SsOGT activity and in the cysteine 119 mutants (C119L in grey, C119F in green). H1, H5 and cl stand for Helix 1, Helix 5 and the connecting loop, respectively. Amino acid atoms are coloured as in Fig. 1.

same extent as in the methylated form of the protein (Fig. 2 and Table 1). The extent of the movements detected in the *K48*-network as a consequence of the active site modification is summarized in Table 1.

As previously demonstrated, transalkylated AGTs are permanently inactivated and prone to degradation both *in vivo* and *in vitro* [1,2,10–12,19]. The opening of the tertiary structure that we observed in the SsOGT-C119F mutant may represent the mechanism for protein

destabilization and its consequent degradation *in vivo*. Overall, the analysis of the SsOGT-C119F structure allowed us to better describe the conformational modifications that characterize the SsOGT protein, following the entire alkyltransferase reaction.

The differences between the C119F and C119L structures at the level of the *K48*-network are currently difficult to interpret: for instance, although the interactions formed by the K48 residue were similar in the C119F and in the SsOGT^m, but not in the C119L, the reverse was true for the position of the E44 residue (Table 1; Figs. 2 and 3), showing that neither mutant could faithfully recapitulate all conformational changes triggered by physiological alkylation. These results should suggest caution, in general, in the interpretation of structure–function data obtained with synthetic mutant proteins. Indeed, post-translational modifications of a native protein might affect the protein conformation in a different manner as compared with amino acid substitutions, which are incorporated in the protein structure co-translationally. In our case, the availability of structural and biochemical data obtained from two independent mutants, along with the physiologically methylated protein, helped to partially overcome these limitations.

Table 1

Comparison of the distances among the *K48*-network residues in all available SsOGT 3D structures.

Entry ^a	Interaction	Distance (Å)				
		SsOGT	DNA-bound	SsOGT ^m	C119L	C119F
①	E44 (Oε1) ⇌ E62 (Oε2)	2.63	8.74	7.50	8.11	6.57
②	K48 (N) ⇌ E62 (Oε2)	3.02	5.33	4.82	3.12	4.60
③	K48 (N) ⇌ R61 (Oα)	3.05	6.67	6.39	3.21	5.01
④	K48 (N) ⇌ N59 (Oα)	2.87	5.08	6.73	3.06	3.80

^a The encircled numbers refer to the distances graphically represented in Fig. 2.

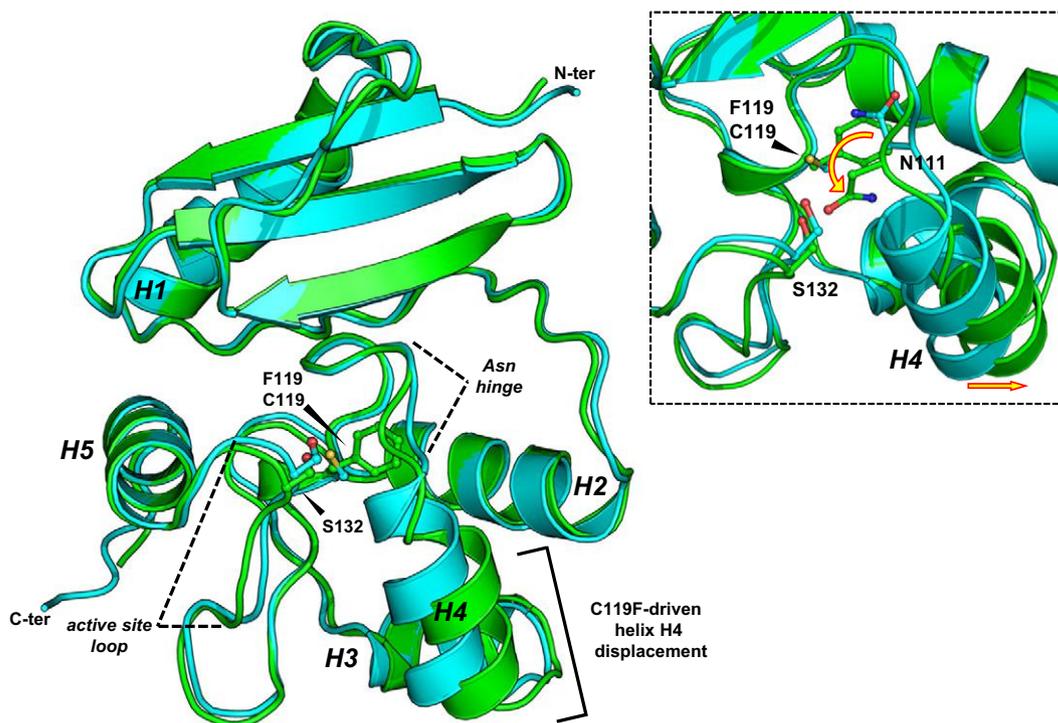


Fig. 3. Structural analysis of SsOGT-C119F. Cartoon representation of the crystal structure of the ligand-free form of wild-type SsOGT (pdb: 4ZYE; in cyan) and of the C119F mutated variant (pdb: 5LLQ; in green), upon optimal superposition. A zoomed view of the conformational changes observed in the C119F variant active site appears in the inset on the right. The yellow arrows indicate the relevant movements of the N111 residue and of the recognition helix H4, as described in the main text. For the colour code of amino acid atoms, refer to the description in Fig. 1.

3.3. Molecular dynamics (MD) simulations

In order to obtain independent information on the role of SsOGT residues in the protein stability, we followed the mobility of SsOGT residues by MD simulations. SsOGT wild type, C119A, C119L, and C119F mutants were simulated for 200 ns at 353 K (80.0 °C). To assess the simulation stability, some geometrical properties such as the average values of the radius of gyration (R_G), the number of intra-protein hydrogen bonds (HB_{intra}) and the solvent-accessible surface area (SASA), both apolar and polar, were monitored as a function of time; all these

properties indicated that the structures were stable during the entire simulation (Table 3). Information about protein equilibration is provided by convergence of RMSD of the atomic position, with respect to the initial structure, calculated as function of time. The comparison of these values (Table 3) and the RMSD plots (Fig. S1b) indicated that free SsOGT requires less time to equilibrate and its RMSD is the lowest. The resulting average conformation for each simulation was used to perform a second MD simulation for 200 ns at 500 K (227.0 °C, Supplementary Table 2). Thus, the MD simulation at 353 K was used to analyse the protein structure at physiological temperature, while the one at 500 K was used to interpret the effect of mutations on structural changes.

In order to obtain further insights into the dynamics of SsOGT structure, the principal component analysis (PCA) was used to dissect out cooperative inner motions. In free SsOGT (Fig. 4), at 353 K we observed movements of a few residues; at 500 K the internal molecular motion was obviously increased, especially in the C_{ter} domain, with significant movements of residues in helices H3, H4 and H5. By contrast, the N_{ter} domain showed higher stability, which can be explained in terms of rigid body-like behavior, since all its secondary structures moved together in the same direction, and the whole domain behaved like a single object. In the SsOGT mutants (Fig. S1a), the motions in the MD simulations performed at 353 K showed that overall molecular internal motion increases: in particular, the *K48-network* region was more

Table 2
Data collection, phasing, and refinement statistics of the SsOGT-C119F mutant.

<i>Data collection</i>	
Space group	P212121
Wavelength (Å)	1.008
Resolution (Å)	2.7
Total reflections	51,418
Unique reflections	12,411
Mean (I)/sd (I)	10.26 (3.35) ^a
Completeness (%)	96.5 (97.7) ^a
Multiplicity	4.1 (4.3) ^a
R_{merge} (%)	10.2
R_{meas} (%)	11.6
<i>Refinement</i>	
R_{factor}/R_{free} (%)	17.2/23.7
Protein atoms	2380
Ligand atoms	12
Water molecules	30
R.M.S.D. bonds (Å)	0.016
R.M.S.D. angles (°)	1.38
Average B (Å ²)	
Protein	32.40
Solvent	30.20
Ligands	47.40

Table 3
Selected structural properties of SsOGT structures at 353 K.

System	R_G (nm)	HB_{intra}	Apolar SASA (nm ²)	Polar SASA (nm ²)	C α RMSD (nm)	RMSD (nm)
SsOGT	1.53 ± 0.02	100 ± 6	42 ± 2	50 ± 2	0.13 ± 0.03	0.23 ± 0.04
C119A	1.56 ± 0.01	101 ± 6	37 ± 1	54 ± 1	0.16 ± 0.04	0.52 ± 0.17
C119L	1.54 ± 0.02	96 ± 6	37 ± 2	53 ± 2	0.14 ± 0.03	0.39 ± 0.19
C119F	1.56 ± 0.02	97 ± 5	38 ± 2	54 ± 2	0.15 ± 0.03	0.39 ± 0.16

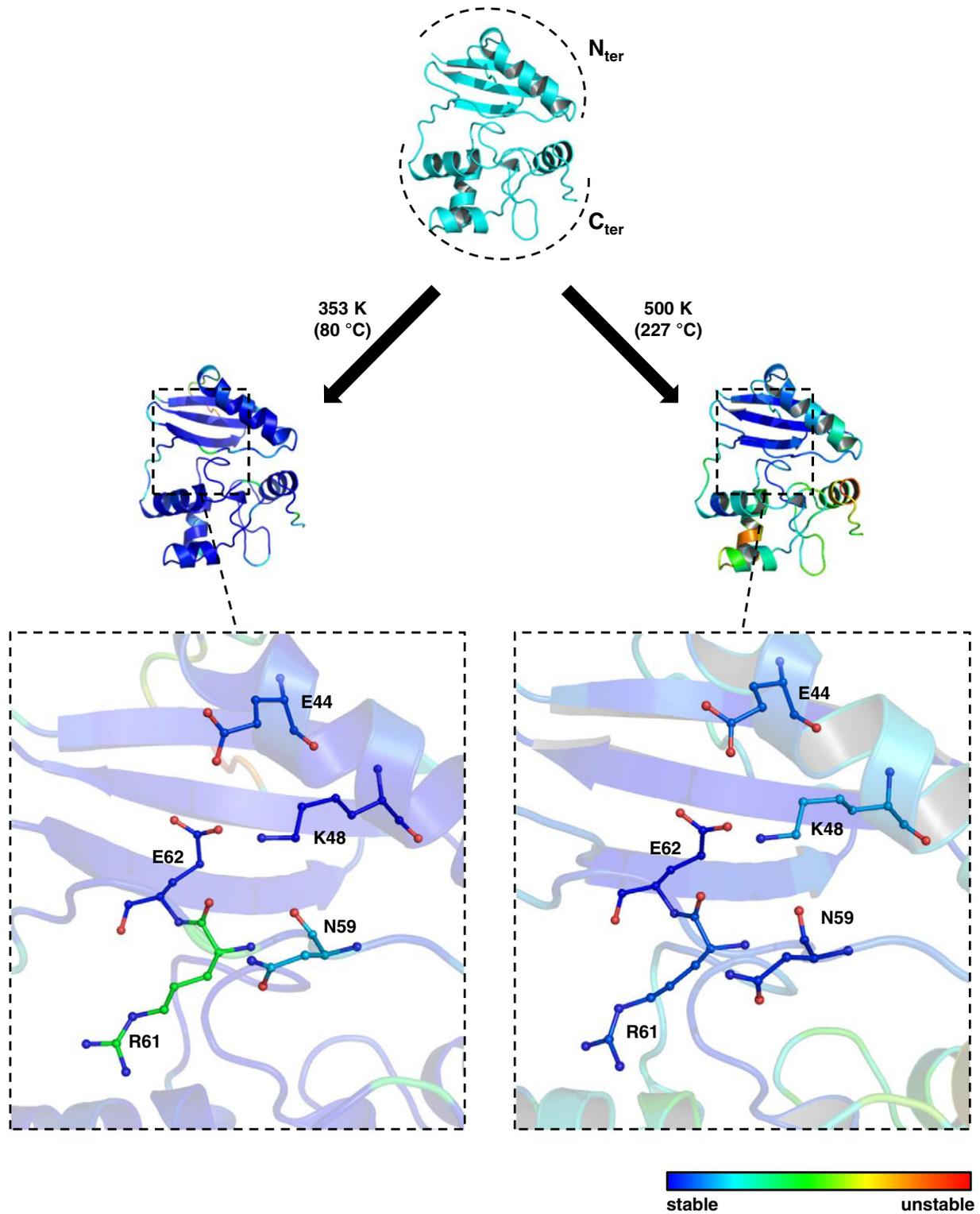


Fig. 4. PCA analysis of wild type SsOGT MD simulations. PCA analysis was performed to analyse the dynamics of residues during the MD simulations performed at 353 and 500 K. A colour scale is used to indicate the stable (blue) and unstable (red) portions of the protein structures. Detail of *K48-network* region of both MD simulation results is depicted. Amino acid atoms are coloured as described in Fig. 1.

unstable in all mutant structures, as compared with the wild-type. The dramatic effect of mutations on protein stability was even more evident in MD simulations at 500 K (Supplementary Table 2). The mutant structures showed overall larger movements than wild type SsOGT: these movements were mainly in the C_{ter} domain of the protein, consistently with our previous analysis which identified in this domain nuclei of destabilization that could provide explanation of protein degradation

in vivo upon alkyltransferase reaction, *i.e.* destroyed interactions and the displacement of conserved α -helices (H2 and H4) [10]. In all mutants, strong movements were also observed in the N_{ter} domain, which cannot be considered as rigid-body in these structures.

The above analysis confirmed that SsOGT stability depends on chemical modification of the catalytic cysteine, and all the mutations that mimic different adducts at the C119 residue have dramatic effects,

Table 4
Biochemical characterization of the activities and the stability of the *K48-network* mutants, in comparison with the wild type SsOGT.

	Catalytic activity BG-VG (M ⁻¹ s ⁻¹) ^a	DNA repair activity K _{DNA} (μM) ^b	Thermal stability T _m (°C) ^c	Note
SsOGT	2730 ± 320	0.83 ± 0.02	80.0 ± 0.4	From ref. [10]
E44L	1960 ± 410	0.66 ± 0.13	80.7 ± 0.1	This study
K48A	3090 ± 240	1.61 ± 0.30	72.0 ± 0.3	This study
K48L	2880 ± 330	1.59 ± 0.23	70.4 ± 0.3	This study

^a Second-order rate constants of the trans-alkylation activity at 25.0 °C on the fluorescent substrate from three independent experiments, accordingly to the conditions previously reported [10] (see **Materials and methods**).

^b Competition assay with BG-VG and a methylated-dsDNA oligonucleotide for 10 min at 50.0 °C. IC₅₀ values were from three independent experiments; K_{DNA} values were obtained as shown in **Materials and methods**.

^c Data obtained from three independent DSF experiments, as described in **Materials and methods**.

decreasing the overall protein stability and enhancing residue movements at high temperature. Increased movements were observed in the two protein domains, as well as at the level of the interdomain interface, where the D27-R133 and *K48-network* interactions are perturbed.

3.4. Analysis of the functional role of the *K48-network*

In order to establish the role of the *K48-network* in SsOGT stability and activity, we prepared three site-directed mutants carrying substitutions of crucial residues in the network, namely E44L, K48A and K48L.

We analysed the catalytic and DNA repair activity of these mutants by using the fluorescence based assays previously developed [10,11]. The former was determined by measuring the efficiency of the alkyl-transfer reaction with the synthetic substrate BG-VG; at 25.0 °C, the catalytic activity of mutants was not significantly different from that of the wild-type protein (Table 4). DNA repair activity assays were performed by using a dsDNA^m as a competitor of the fluorescent substrate, thus

giving a measure of the overall protein activity [10,11]. As shown in Table 4, whereas the efficiency of lesion repair by the E44L mutant was similar to that of the wild-type SsOGT, both K48 mutants showed a slight reduction of the repair efficiency (K_{DNA} of 1.6 vs 0.8 μM at 50.0 °C).

In thermal stability assays, the E44L mutant showed the same resistance to thermal denaturation as the wild type; instead, the K48A/L mutants were significantly destabilized, and aggregated above 65.0 °C (data not shown). As expected, T_m values determined by DSF showed a reduction of ca. 8.0 and 10.0 °C for K48A and K48L, respectively (Table 4). Thus, these data, consistently with the structural observations, demonstrated that the integrity of the *K48-network* contributes significantly to the protein stability; moreover, substitution of the E44 residue has minor impact on protein stability with respect to the K48 residue, in agreement with the central position and multiple intramolecular interactions established by this latter residue within the network in the free protein.

In conclusion, our biochemical analysis demonstrated that perturbation of the *K48-network* affects mainly SsOGT stability, and at lesser extent its DNA repair activity, consistent with the notion that AGT alkylation has a minor effect on the DNA binding activity [10,39]. Taken together, these results support the hypothesis that the main role of the *K48-network* is contributing to stabilizing connection between the N_{ter} domain and the connecting loop.

4. Conclusions

Our combined structural, biochemical, mutational and molecular dynamics analysis demonstrate that the *K48-network*, along with the previously identified D27-R133 interaction, contributes to maintain the correct folding of SsOGT and is perturbed after lesion recognition.

Based on our previous and present data, we propose a model of the conformational changes and fate of the SsOGT upon the repair of alkylated DNA (Fig. 5). The optimal activity and stability of SsOGT require coordination between the N_{ter} and C_{ter} domains of the protein; in fact, several intramolecular interactions are found at the interface

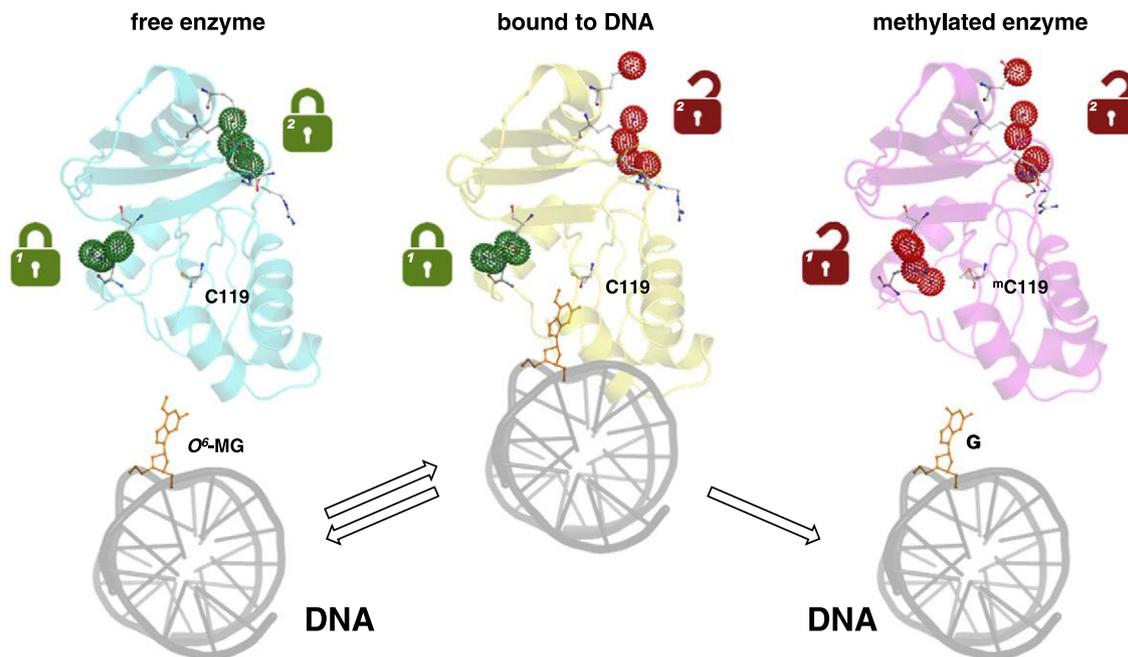


Fig. 5. Model of the conformational changes and fate of the SsOGT upon repair of alkylated DNA. The N_{ter} and C_{ter} domains of the free enzyme (in cyan; PDB ID: 4ZYE) are stabilized and coordinated by the presence of two locks (coloured dots), namely the D27-R133 interaction [10] and the *K48-network*, respectively. SsOGT binds the DNA (in yellow, and in gray, respectively; PDB ID: 4ZYD) and recognizes the damaged guanine (O⁶-MG, in orange ball and stick format), led to the opening only of the lock 2: this state could still allow the enzyme to dissociate the DNA and restore the *K48-network*. Upon the alkylation of the catalytic cysteine (^mC119), both the locks change to the open state, causing the irreversible conformational changes in charge of the alkylated-enzyme (in magenta; PDB ID: 4ZYG), and subsequently its destabilization. All the important residues for the activity of SsOGT are drawn as ball and stick, and coloured as described in Fig. 1.

between the two domains, which are likely to contribute to interdomain communication. Our previous [10] and present work identified two groups of interactions, namely the D27-R133 pair and the *K48-network*, respectively, playing important role in protein stability, as well as in communicating the state of active site. We suggest that these interactions act as “locks”: in the ligand-free protein, both *locks* are in their “closed state”, thus ensuring the correct folding of the protein for its optimal stability and ready to perform the reaction. Binding to DNA and recognition of the damaged guanine lead to opening of *lock 2* only (through perturbation of the *K48-network*), destabilizing the link between the N_{ter} domain and the connecting loop. We hypothesize that this modification is reversible as long as the active site remains unmodified: if the protein dissociates from DNA without performing the repair reaction, the integrity of the *K48-network* is restored and the structural stability of the protein is preserved (Fig. 5). However, once the DNA repair reaction is completed, alkylation of the catalytic cysteine induces irreversible conformational changes, fixing the *lock 1* (formed by the D27-K133 ion pair) in its “open state” (Fig. 5). The loss of coordination between the N_{ter} and C_{ter} domains triggers the SsOGT destabilization and, subsequently, its degradation.

To the best of our knowledge our studies provided the most detailed description of alkylation-induced conformational changes occurring in a protein of this class. The unavailability of a similarly detailed analysis for the human as well as other AGTs impairs extensive comparison of these results. By analysing the available structures of human AGT [7,13,19] we notice that the region corresponding to that containing the *K48-network* in SsOGT, is characterized by the presence of hydrophobic residues that stabilize contacts between the α -helix belonging to the N-terminal domain and the random coiled loop connecting the two protein domains. Although this hypothesis could not be tested directly, due to the instability of alkylated hAGT structures, these hydrophobic interactions might play a role similar to the SsOGT *K48-network*. Thus, our model might be extended to other AGTs, assuming that, whatever their nature, interdomain communication and coordination play a key role in maintaining proper folding and respond to alkylation triggering destabilization and degradation.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbagen.2016.10.020>.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgments

We gratefully thank Dr. Andrea Strazzulli for assistance in some experiments. This work was supported by: FIRB-Futuro in RicercaRBF12001G_002-Nematic.

References

- [1] A.E. Pegg, Repair of O(6)-alkylguanine by alkyltransferases, *Mutat. Res.* 462 (2000) 83–100.
- [2] A.E. Pegg, Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools, *Chem. Res. Toxicol.* 24 (2011) 618–639.
- [3] A. Vettone, G. Perugini, M. Rossi, A. Valenti, M. Ciaramella, Genome stability: recent insights in the topoisomerase reverse gyrase and thermophilic DNA alkyltransferase, *Extremophiles: life under extreme conditions* 18 (2014) 895–904.
- [4] R. Miggiano, V. Casazza, S. Garavaglia, M. Ciaramella, G. Perugini, M. Rizzi, F. Rossi, Biochemical and structural studies of the *Mycobacterium tuberculosis* O⁶-methylguanine methyltransferase and mutated variants, *J. Bacteriol.* 195 (2013) 2728–2736.
- [5] R. Miggiano, G. Perugini, M. Ciaramella, M. Serpe, D. Rejman, O. Pav, R. Pohl, S. Garavaglia, S. Lahiri, M. Rizzi, F. Rossi, Crystal structure of *Mycobacterium tuberculosis* O⁶-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA, *Biochem. J.* 473 (2016) 123–133.
- [6] D.S. Daniels, J.A. Tainer, Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O(6)-alkylguanine-DNA alkyltransferase, *Mutat. Res.* 460 (2000) 151–163.
- [7] D.S. Daniels, T.T. Woo, K.X. Luu, D.M. Noll, N.D. Clarke, A.E. Pegg, J.A. Tainer, DNA binding and nucleotide flipping by the human DNA repair protein AGT, *Nat. Struct. Mol. Biol.* 11 (2004) 714–720.
- [8] E.M. Duguid, P.A. Rice, C. He, The structure of the human AGT protein bound to DNA and its implications for damage detection, *J. Mol. Biol.* 350 (2005) 657–666.
- [9] M.H. Moore, J.M. Gulbis, E.J. Dodson, B. Dimple, P.C. Moody, Crystal structure of a suicidal DNA repair protein: the Ada O⁶-methylguanine-DNA methyltransferase from *E. coli*, *EMBO J.* 13 (1994) 1495–1501.
- [10] G. Perugini, R. Miggiano, M. Serpe, A. Vettone, A. Valenti, S. Lahiri, F. Rossi, M. Rossi, M. Rizzi, M. Ciaramella, Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein, *Nucleic Acids Res.* 43 (2015) 8801–8816.
- [11] G. Perugini, A. Vettone, G. Illiano, A. Valenti, M.C. Ferrara, M. Rossi, M. Ciaramella, Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage, *J. Biol. Chem.* 287 (2012) 4222–4231.
- [12] J.L. Tubbs, A.E. Pegg, J.A. Tainer, 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* 6 (2007) 1100–1115.
- [13] J.E. Wibley, A.E. Pegg, P.C. Moody, Crystal structure of the human O⁶-alkylguanine-DNA alkyltransferase, *Nucleic Acids Res.* 28 (2000) 393–401.
- [14] Q. Fang, S. Kanugula, A.E. Pegg, Function of domains of human O⁶-alkylguanine-DNA alkyltransferase, *Biochemistry* 44 (2005) 15396–15405.
- [15] J.J. Rasimas, P.A. Dalessio, I.J. Ropson, A.E. Pegg, M.G. Fried, Active-site alkylation destabilizes human O⁶-alkylguanine DNA alkyltransferase, *Protein science: a publication of the Protein Society* 13 (2004) 301–305.
- [16] M. Xu-Welliver, A.E. Pegg, Degradation of the alkylated form of the DNA repair protein, O⁶-alkylguanine-DNA alkyltransferase, *Carcinogenesis* 23 (2002) 823–830.
- [17] S. Phillip, S. Swaminathan, S.G. Kuznetsov, S. Kanugula, K. Biswas, S. Chang, N.A. Loktionova, D.C. Haines, P. Kaldis, A.E. Pegg, S.K. Sharan, Degradation of BRCA2 in alkyltransferase-mediated DNA repair and its clinical implications, *Cancer Res.* 68 (2008) 9973–9981.
- [18] E. Brunk, B. Mollwitz, U. Rothlisberger, Mechanism to trigger unfolding in O⁶-alkylguanine-DNA alkyltransferase, *ChemBiochem. Eur. J. Chem. Biol.* 14 (2013) 703–710.
- [19] D.S. Daniels, C.D. Mol, A.S. Arvai, S. Kanugula, A.E. Pegg, J.A. Tainer, Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding, *EMBO J.* 19 (2000) 1719–1730.
- [20] S. Kanugula, K. Goodtzova, A.E. Pegg, 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) (1998) 545–550.
- [21] G. Perugini, A. Valenti, A. D’Amaro, M. Rossi, M. Ciaramella, Reverse gyrase and genome stability in hyperthermophilic organisms, *Biochem. Soc. Trans.* 37 (2009) 69–73.
- [22] A. Valenti, A. Napoli, M.C. Ferrara, M. Nadal, M. Rossi, M. Ciaramella, Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*, *Nucleic Acids Res.* 34 (2006) 2098–2108.
- [23] A. Valenti, G. Perugini, T. Nohmi, M. Rossi, M. Ciaramella, Inhibition of translesion DNA polymerase by archaeal reverse gyrase, *Nucleic Acids Res.* 37 (2009) 4287–4295.
- [24] V. Visone, A. Vettone, M. Serpe, A. Valenti, G. Perugini, M. Rossi, M. Ciaramella, Chromatin structure and dynamics in hot environments: architectural proteins and DNA topoisomerases of thermophilic archaea, *Int. J. Mol. Sci.* 15 (2014) 17162–17187.
- [25] A. Vettone, M. Serpe, A. Hidalgo, J. Berenguer, G. del Monaco, A. Valenti, M. Rossi, M. Ciaramella, G. Perugini, A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering, *Extremophiles: life under extreme conditions* 20 (2016) 1–13.
- [26] A. Gautier, A. Juillerat, C. Heinis, I.R. Correa Jr., M. Kindermann, F. Beauflis, K. Johnsson, An engineered protein tag for multiprotein labeling in living cells, *Chem. Biol.* 15 (2008) 128–136.
- [27] Y. Cheng, W.H. Prusoff, Relationship between the inhibition constant (K₁) and the concentration of inhibitor which causes 50 per cent inhibition (I₅₀) of an enzymatic reaction, *Biochem. Pharmacol.* 22 (1973) 3099–3108.
- [28] F.H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nat. Protoc.* 2 (2007) 2212–2221.
- [29] T.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G. Leslie, iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 67 (2011) 271–281.
- [30] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 67 (2011) 235–242.
- [31] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, *J. Appl. Crystallogr.* 40 (2007) 658–674.
- [32] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 66 (2010) 213–221.
- [33] T.C. Terwilliger, R.W. Grosse-Kunstleve, P.V. Afonine, N.W. Moriarty, P.H. Zwart, L.W. Hung, R.J. Read, P.D. Adams, Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 64 (2008) 61–69.

- [34] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 66 (2010) 486–501.
- [35] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.), *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, Springer Netherlands, Dordrecht 1981, pp. 331–342.
- [36] H. Hashimoto, T. Inoue, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, Y. Kai, Hyperthermostable protein structure maintained by intra and inter-helix ion-pairs in archaeal O^6 -methylguanine-DNA methyltransferase, *J. Mol. Biol.* 292 (1999) 707–716.
- [37] H. Hashimoto, M. Nishioka, T. Inoue, S. Fujiwara, M. Takagi, T. Imanaka, Y. Kai, Crystallization and preliminary X-ray crystallographic analysis of archaeal O^6 -methylguanine-DNA methyltransferase, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 54 (1998) 1395–1396.
- [38] A. Roberts, J.G. Pelton, D.E. Wemmer, Structural studies of MJ1529, an O^6 -methylguanine-DNA methyltransferase, *Magnetic Resonance in Chemistry: MRC*, 44 Spec No 2006, pp. S71–S82.
- [39] J.J. Rasimas, A.E. Pegg, M.G. Fried, DNA-binding mechanism of O^6 -alkylguanine DNA alkyltransferase. Effects of protein and DNA alkylation on complex stability, *J. Biol. Chem.* 278 (2003) 7973–7980.

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

Riccardo Miggiano², Giuseppe Perugino¹, Maria Ciaramella¹, Mario Serpe¹, Dominik Rejman³, Ondrej Pav³, Radek Pohl³, Silvia Garavaglia², Samarpita Lahiri², Menico Rizzi² and Franca Rossi².

¹Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80125 Naples, Italy.

²DiSF-Dipartimento di Scienze del Farmaco, University of Piemonte Orientale 'A. Avogadro', Via Bovio 6, 28100 Novara, Italy.

³IOCB-Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences v.v.i., 166 10 Prague 6, Czech Republic.

In the present study we described the crystal structure of wild-type MtOGT in complex with a modified DNA. This protein contributes to protect the bacterial GC-rich genome against the mutagenic lesion of O⁶mG in DNA. The crystal structure of the complex revealed features of the protein-protein and protein-DNA interactions occurring during alkylated DNA binding, and the protein capability to host unmodified bases inside the active site, in a extra-helical conformation. Finally, biochemical and structural characterization of protein mutated variants the N- and the C-terminal domains of MtOGT displayed a structural plasticity that could be required for proper protein assembly at the alkylated site during DNA repair.

My contribution to the collaboration was performing EMSA analyses of the MtOGT and related mutants.

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‡, Ondřej Páv‡, 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

Mycobacterium tuberculosis O⁶-methylguanine-DNA methyltransferase (*Mt*OGT) contributes to protect the bacterial GC-rich genome against the pro-mutagenic potential of O⁶-methylated guanine in DNA. Several strains of *M. tuberculosis* found worldwide encode a point-mutated O⁶-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 Arg³⁷-bearing region in catalysis. The crystal structure of *Mt*OGT complexed with modified DNA solved in the present study 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 Arg³⁷-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.

Key words: co-operativity, crystal structure, DNA-binding protein, DNA repair, *Mycobacterium tuberculosis*, O⁶-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 O⁶-methylguanine-DNA methyltransferase (OGT, EC 2.1.1.63). Genes encoding O⁶-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 O⁶-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 O⁶-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 Arg³⁷ 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, Arg³⁷ maps away from the protein active site and the DNA-binding motifs so far identified,

Abbreviations: AGT, O⁶-alkylguanine-DNA alkyltransferase; E1X-dsDNA, N¹-O⁶-ethano-2'-deoxyxanthosine-containing dsDNA; hAGT, human AGT; HTH, helix–turn–helix; *Mt*OGT, *Mycobacterium tuberculosis* OGT; OGT, O⁶-methylguanine-DNA methyltransferase; VG, SNAP-VISTA Green reagent.

¹ Correspondence may be addressed to either of these authors (email menico.rizzi@uniupo.it or franca.rossi@uniupo.it).

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 Arg³⁷-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 co-operative 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 *M. tuberculosis* *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 R37K fwd/R37K rev, R37E fwd/R37E rev, and Y139F fwd/Y139F rev (see Supplementary Table S1). 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 (results 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.

Crystallographic 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 molecular mass cut-off, 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 sparse-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 100 K using synchrotron radiation at the ID-29 (*Mt*OGT::E1X-dsDNA complex) or ID14-EH4 (*Mt*OGT-R37K and *Mt*OGT-Y139F variants) beamlines (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 *P*2₁2₁, with cell

Table 1 Data collection, phasing and refinement statistics

Values in parentheses refer to the highest resolution shell.

	<i>Mt</i> OGT::E1X-dsDNA	<i>Mt</i> OGT-R37K	<i>Mt</i> 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	62 363	50 946	17 965
Unique reflections	12907	8668	5719
Mean(I)/S.D.(I)	8.8 (1.6)	26.7 (8.5)	10.8 (2.4)
Completeness (%)	99.7 (99.9)	99.5 (100)	95.9 (100)
Multiplicity	4.8 (5.0)	5.9 (6.0)	3.1 (2.9)
R_{merge} (%)	15.0	3.9	6.1
R_{meas} (%)	16.9	4.3	7.3
Refinement			
$R_{\text{factor}}/R_{\text{free}}$ (%)	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

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 Tyr¹³⁹ 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 programs 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 Arg³⁷ or Tyr¹³⁹ 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 (<http://www.pymol.org>).

Deposition

The atomic co-ordinates 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 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 [21,32]. 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 Arg³⁷ 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 (E1X₇) (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 E1X₇ 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 (dA₄), and a symmetry equivalent of chain B ('B sym. mate') does the same with the deoxyadenosine residue at position 5 of the complementary strand (dA₁₈). 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).

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

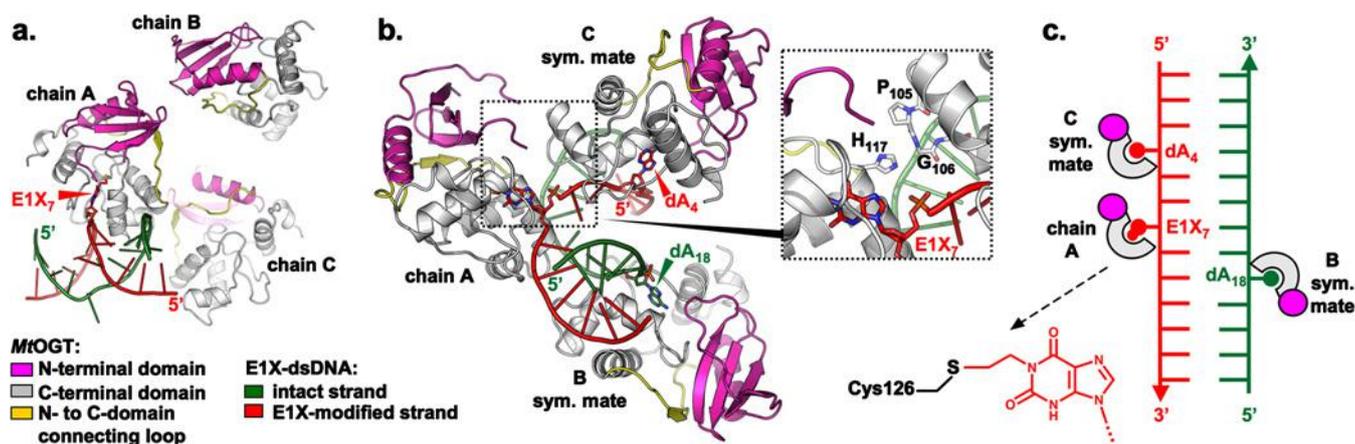


Figure 1 The overall structure of *MtOGT* complexed with modified DNA

(a) Cartoon representation of the asymmetrical unit content of the *MtOGT*::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 *MtOGT* chains assembled on to the same E1X-dsDNA molecule, resulting from applying crystal symmetry operators; extrahelical bases are rendered as sticks and indicated by an arrowhead (inset: close-up 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 broken arrow points to a scheme of the covalent adduct formed between the chain A catalytic cysteine (Cys¹²⁶) and the E1X base. The colour codes for protein domain and DNA strand identification appear at the bottom of the figure.

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 *MtOGT* 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 (Cys¹⁴⁵ in hAGT) to catalyse the S_N2-like dealkylation reaction [24,25].

The architecture of the substrate-binding site of the three *MtOGT* chains building up the *MtOGT*::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 *MtOGT* chain A reveals a continuous density signal contouring the catalytic Cys¹²⁶ and the modified E1X₇ base (Figure 2b). Other close protein–DNA contacts involve: the strictly conserved 'arginine finger' (Arg¹⁰⁹) which, by invading the double helix from the minor groove side, and stacking between the planes of the dG₆ and dC₈ bases, structurally compensates for the flipped-out E1X₇ base; the carboxamide group of Asn¹¹⁵, observed at a 2.9-Å mean distance from the E1X₇ O² position; the hydroxy group of Tyr⁹⁵, standing at a 3.4-Å mean distance from both the N³ atom and the deoxyribose moiety of the E1X₇ base; the active-site loop residues Thr¹³⁷ and Gly¹⁴⁰, the backbone oxygen and nitrogen atoms of which are observed at a distance of 2.8 and 2.7 Å from E1X₇ O⁴ and O⁶, respectively; and Tyr¹³⁹ 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 Ala¹³² appear to

lock, from both sides, the sugar–phosphate backbone downstream of the lesion (Figure 2c).

With the obvious exception of contacts involving E1X₇-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 dA₁₈ and dA₄ unmodified bases, respectively (Figure 2d), thus indicating that *MtOGT* can efficiently bind nitrogenous bases independently of the presence of the alkyl adduct.

In principle, the insertion of an undamaged adenine residue into the *MtOGT* ligand-binding pocket would not expose the DNA substrate to an increased risk of chemical modifications. In fact, the reactivity of the purine ring N¹ and N⁶ positions, as they were observed in the active site of the B and C chains of the *MtOGT*::E1X₇-dsDNA structure, does not appear significantly enhanced by the nearby catalytic cysteine, nor by the presence of the other residues co-ordinating 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 *MtOGT*::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 *MtOGT* 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, *MtOGT* 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 [25,33].

MtOGT undergoes structural rearrangements on DNA binding

The structure of *MtOGT* 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 [22–25] and *Sulfolobus solfataricus* OGT [36], at different stages

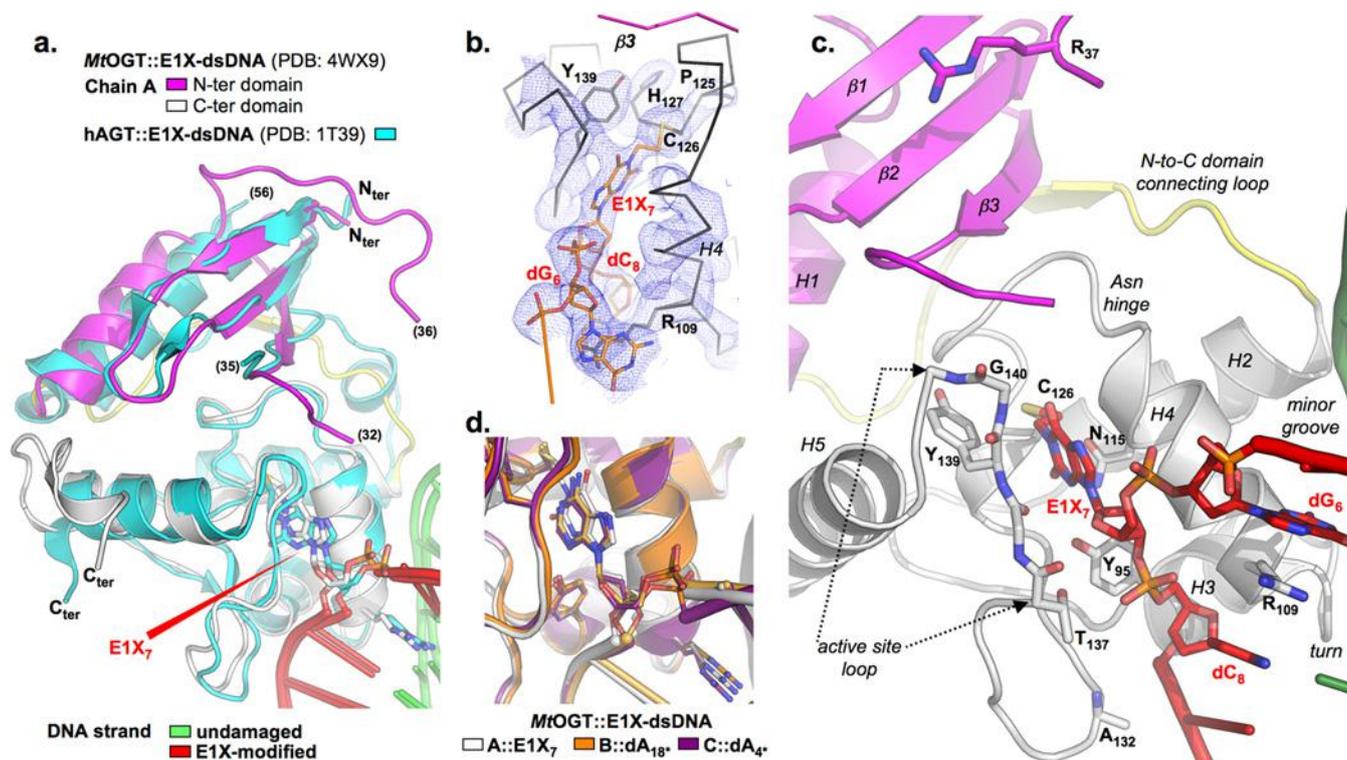


Figure 2 Structural analysis of the *MtOGT* protein complexed with E1X-dsDNA

(a) Cartoon representation of the optimally superimposed structures of *MtOGT* (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 *MtOGT*::E1X-dsDNA chain A housing the modified base (E1X₇), with σ_A -weighted $2F_o - F_c$ electron density contoured at 1.0 σ ; the Cys¹²⁶ thiol group is observed at a distance of 2.1 Å from E1X₇, the C11 atom; the protein backbone appears as a ribbon. (c) Close-up view of the *MtOGT* 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 a). (d) Cartoon representation of the active site of optimally superimposed *MtOGT*::E1X-dsDNA chains A, B and C (average RMSDs are 0.671 and 0.456 Å 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.

of the transalkylation reaction, suggest that the active site of the human and archaeal proteins is largely pre-shaped to perform the catalysis, without requiring heavy structural rearrangements. On the contrary, the association of *MtOGT* 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 *MtOGT*::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 *MtOGT*::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 Arg³⁷ 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 *MtOGT*::E1X-dsDNA crystal structure, and taking into account that the B and C monomers host a nitrogenous base in their active site (Figure 2d), we propose that an equivalent structural

repositioning of the flap might also occur in the *MtOGT* subunit binding the modified base.

The Arg³⁷-containing random coil could participate in the co-operative assembly of protein clusters on to the dsDNA substrate

The analysis of the *MtOGT*::E1X-dsDNA crystal structure seems to exclude direct participation of Arg³⁷ in DNA binding, because the protein residue and the sugar-phosphate backbone of the E1X-dsDNA substrate are observed at a distance of >16 Å (Figure 3a). Instead, we propose that Arg³⁷ 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 *MtOGT*-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 *MtOGT* clusters at the damaged DNA sites.

To test this hypothesis, we expressed and purified two new mutated versions of *MtOGT* (*MtOGT*-R37K and *MtOGT*-R37E), and analysed their dsDNA^{met}-repairing activity, by adopting the

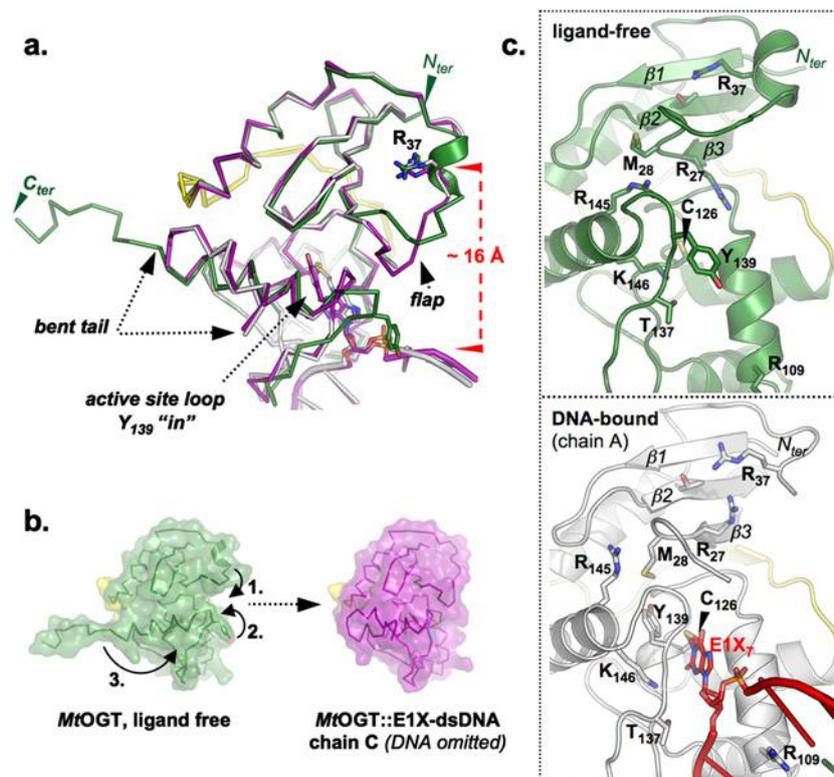


Figure 3 The conformation adopted by discrete protein regions differs in the ligand-free and DNA-bound *MtOGT* structures

(a) Structural superposition of *MtOGT* 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 *MtOGT* 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) close-up 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) *MtOGT*.

same VG-based assay [32] previously used to characterize the wild-type protein and the *MtOGT*-R37L variant [21]. Our data (see Supplementary Table S2) show that the *MtOGT*-R37E mutant exhibits a 5-fold lower affinity for the methylated duplex ($K_{\text{DNA}}^{\text{met-R37E}} = 1.14 \pm 0.15 \mu\text{M}$) with respect to *MtOGT* ($K_{\text{DNA}}^{\text{met-wt}} = 0.24 \pm 0.11 \mu\text{M}$ [21]), whereas the more conservative Arg³⁷-to-Lys substitution produces a more limited effect on the dsDNA^{met}-binding constant ($K_{\text{DNA}}^{\text{met-R37K}} = 0.38 \pm 0.2 \mu\text{M}$).

In parallel, we performed direct EMSA-based experiments (Figure 4a), using a (carboxy)tetramethylrhodamine (TAMRA)-labelled, non-alkylated dsDNA probe (see Supplementary Table S1). It is interesting that the *MtOGT*-R37E protein reaches a plateau in band-shift activity at a DNA/protein molar ratio of 1:600 ($K_{\text{DNA-R37E}} = 41.4 \pm 1.1 \mu\text{M}$), whereas both the wild-type *MtOGT* and the *MtOGT*-R37K variant induce a complete shift at a DNA/protein molar ratio in the range of 1:150 ($K_{\text{DNA-wt}} = 7.2 \pm 0.2 \mu\text{M}$ [21], $K_{\text{DNA-R37K}} = 13.2 \pm 0.7 \mu\text{M}$). The results of the EMSA-based analysis match well with those obtained from the VG-based competitive assays, and are consistent with previously published data showing that the recombinant *MtOGT*-R37L variant displays a 10-fold lower affinity towards the dsDNA^{met} substrate compared with the wild-type *MtOGT*, although the cooperativity of DNA binding is maintained [21]. Taken together, our previous study [21] and the present study confirm that the Arg³⁷ 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 *MtOGT* protein, which characterizes the *MtOGT*-R37L and *MtOGT*-R37E proteins respectively, translates into less efficient DNA binding and repair.

By analysing the crystal structure of the *MtOGT*-R37K variant (Figure 4b and 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 *MtOGT*-R37L protein, the presence of leucine at position 37 destroys this bonding scheme [21]. However, in both of 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 Arg³⁷ could play a role in co-ordinating 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 *MtOGT*::E1X-dsDNA structure.

The intrinsic flexibility of the *MtOGT* active site loop

As mentioned above, the active-site loop and the C-terminal tail of *MtOGT* adopt different conformations, depending on the association of the protein with the DNA substrate (Figure 3). Different from what has been reported for all OGTs for which the crystal structure has so far been solved [22–25,34–36], but

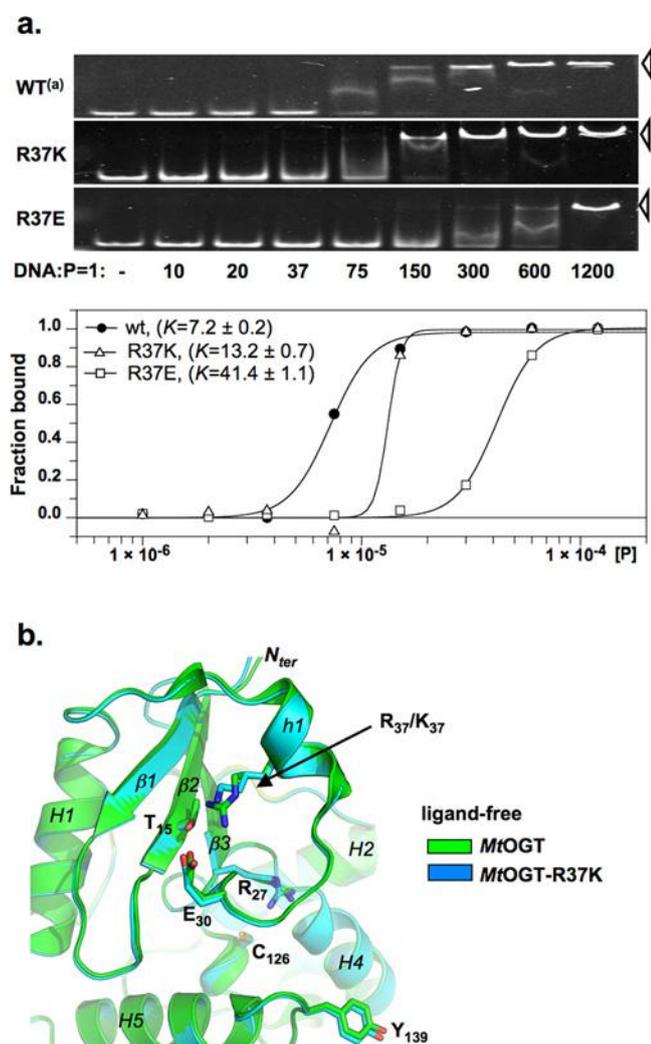


Figure 4 A positively charged residue at *MtOGT* position 37 is required for optimal dsDNA binding

(a) Upper image: EMSA-based analysis of wild-type *MtOGT* (WT) [21], and of the indicated point-mutated proteins, performed by using 1 pmol of TAMRA-labelled dsDNA (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 by EMSA (upper image); [P], protein concentration (M); *K*, dissociation constant (μ M). (b) Close-up of the N-terminal domain and part of the of the active site of the ligand-free *MtOGT* (PDB code 4BHB) [21] and *MtOGT*-R37K (PDB code 4WXD), on optimal structural superimposition; residues mentioned in the text appear as sticks; secondary structure elements are labelled in italic.

reminiscent of what was observed in the OGT structure of *Methanococcus jannaschii* in solution [37], the C-side region of the active-site loop (residues 136–141) of the ligand-free structures of *MtOGT* and its mutated variants is invariably oriented towards the bulk solvent. This conformation is stabilized by contacts established between the conserved Tyr¹³⁹ 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 *MtOGT*::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

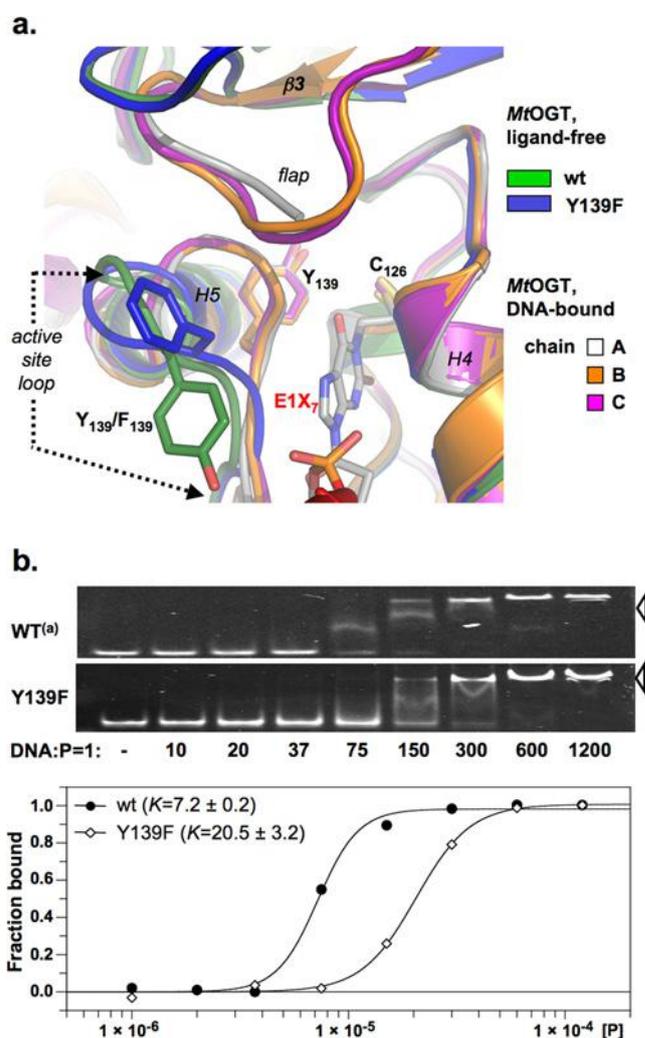


Figure 5 Tyr¹³⁹ could help the active-site loop movements during DNA binding

(a) Close-up view of the active site of *MtOGT* (PDB code 4BHB) [21] and *MtOGT*-Y139F (PDB code 4WXD), both crystallized in ligand-free form, and of the A, B and C chains building up the *MtOGT*::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 labelled in italic. (b) Upper image: EMSA-based analysis of *MtOGT* (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 by EMSA (upper image); [P], protein concentration (M); *K*, dissociation constant (μ M).

active site of *MtOGT* 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 Tyr¹³⁹ of the active-site loop of *MtOGT* performed exactly the same molecular tasks highlighted for the equivalent residue of the human protein (Tyr¹⁵⁸), namely narrowing of the ligand-binding pocket and providing an aromatic environment for the alkyl adduct [24,25], then the replacement of Tyr¹³⁹ by a phenylalanine should have little effect on catalysis.

Data from VG-based assays (see Supplementary Table S2) reveal that a *MtOGT*-Y139F variant displays a 10-fold lower affinity for dsDNA^{met} compared with wild-type *MtOGT*

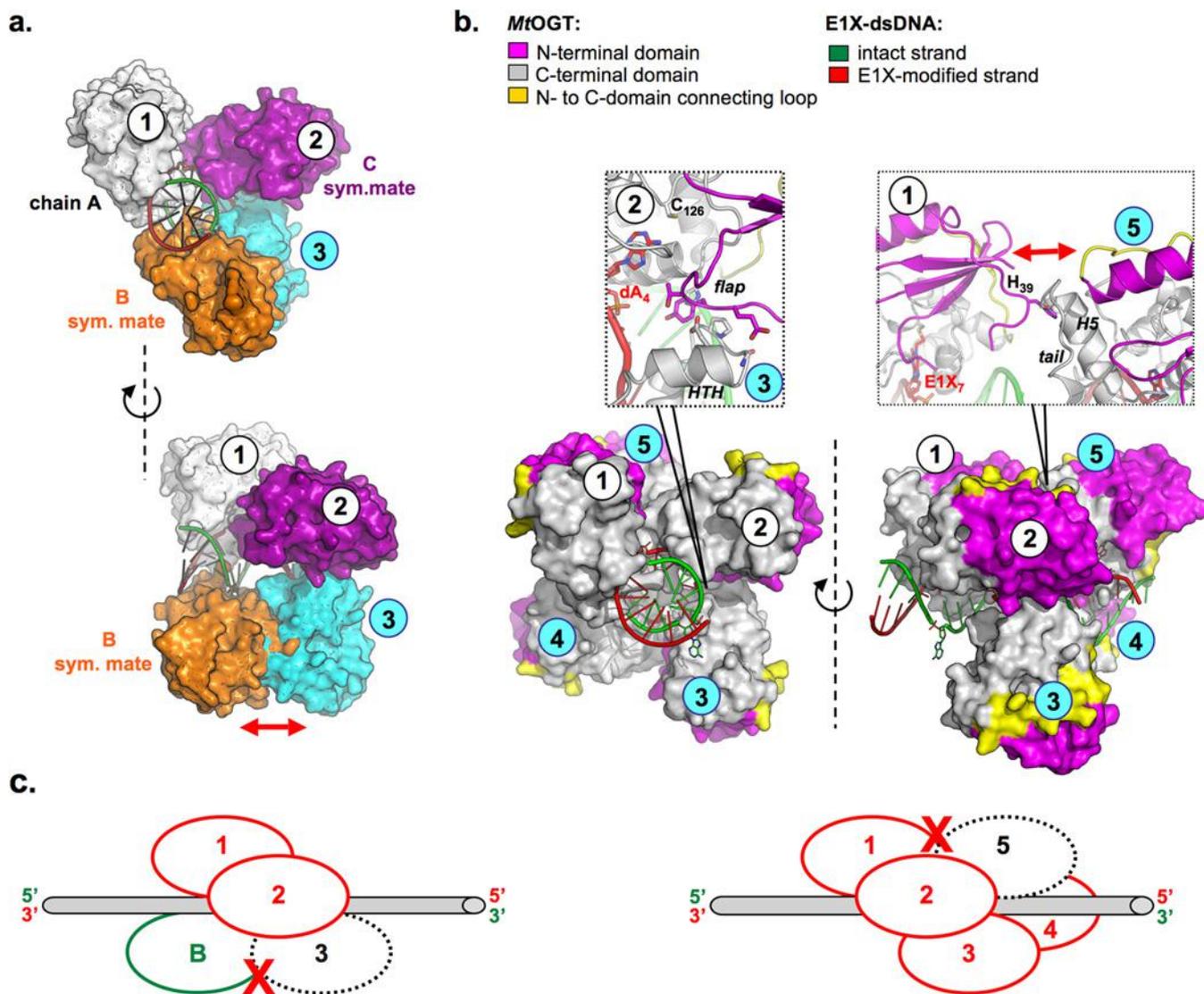


Figure 6 Structure-based model of *MtOGT* clustering on to a DNA duplex

(**a**) The possible negative effect of the presence of a *MtOGT* 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 *MtOGT* monomers assembled on to the same dsDNA molecule; the close-up 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 *MtOGT* monomers that bind bases of the damaged or intact strand are depicted as red or green ovals, respectively; the dashed ovals represent the *MtOGT* subunit that would come into collision with pre-assembled monomers.

($K_{\text{DNA}}^{\text{met}_{\text{Y139F}}} = 2.19 \pm 0.5 \mu\text{M}$; $K_{\text{DNA}}^{\text{met}_{\text{wt}}} = 0.24 \pm 0.11 \mu\text{M}$ [21]), confirming the *MtOGT* requirement of a tyrosine residue at position 139 for optimal repair of an O^6 -methylated guanine in dsDNA. Instead, the ability of *MtOGT* to bind unmodified dsDNA appears less affected by the Tyr¹³⁹-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 ($K_{\text{DNA-wt}} = 7.2 \pm 0.2 \mu\text{M}$ [21], $K_{\text{DNA-Y139F}} = 20.5 \pm 3.2 \mu\text{M}$; Figure 5b). Therefore, we speculate that Tyr¹³⁹ could play a role not only in properly fixing the base inside the protein active site on DNA binding, as proposed for hAGT Tyr¹⁵⁸ [24,25,38–40], but also in making *MtOGT* able to discriminate between intact and alkylated dsDNA molecules, albeit through a molecular mechanism that will need further study for elucidation.

The *MtOGT*::E1X-dsDNA crystal structure provides insights into co-operative DNA binding

The architecture of *MtOGT* 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 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 (Figure 1). From this standpoint, the supramolecular assembly revealed by our structure could represent a model of the *MtOGT* clustering on a monoalkylated dsDNA molecule.

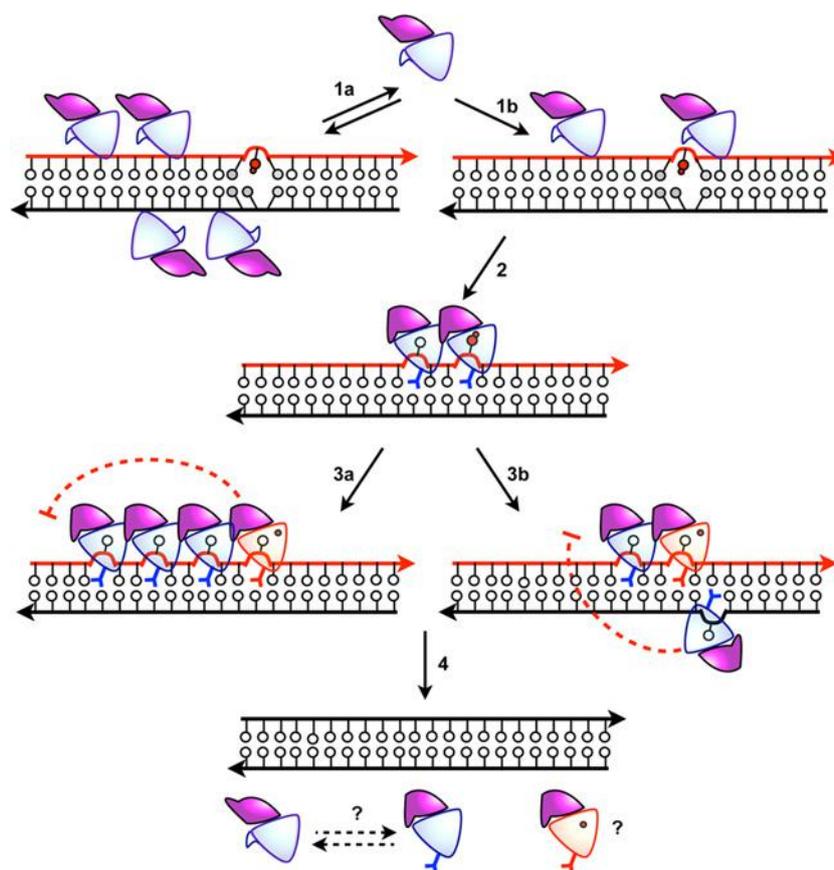


Figure 7 Preliminary model of *MtOGT*-mediated direct alkylated DNA repair

Schematic diagrams 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 in the Discussion.

However, we cannot rule out the possibility that longer more physiological DNA substrates might sustain the nucleation of *MtOGT* protein clusters larger in size than the one characterizing the *MtOGT*::E1X-dsDNA crystal structure. To verify this hypothesis, we tried to model further DNA-bound *MtOGT* 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 *MtOGT* monomer with the region of the intact DNA strand facing the alkylated base (i.e. the 'B sym. mate'-binding dA₁₈) would hamper the recruitment of additional protein subunits at the 5'-side of the damaged base (Figure 6a).

On the other hand, the *MtOGT*::E1X-dsDNA crystal structure itself does not provide any indication regarding the dynamics of the alkylated-DNA damage reversal performed by *MtOGT*. Therefore, we can also assume that the binding of co-oriented *MtOGT* 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 *MtOGT* protein clusters (Figure 6b) that proves to be more compact compared with the one proposed for hAGT [45,46], possibly due to the structural plasticity of *MtOGT*, which could allow more crowded protein assembling on to DNA. It is interesting that the DNA binding-associated repositioning of the

flap (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 co-operative 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 of chain 5.

This analysis suggests that both short- and long-range steric hindrance phenomena could play a role in regulating *MtOGT*–DNA association and dissociation, resulting in protein clusters that are capable of self-limiting their size, similar to what has been experimentally determined by direct atomic force microscopy (AFM)-based studies of hAGT [47].

DISCUSSION

The architecture of the protein–DNA complexes revealed by the *MtOGT*::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 [24,25,34]: (i) the co-operative binding of DNA has been demonstrated as a *bona fide* activity

of hAGT [41–46], (ii) structure-based models of hAGT/dsDNA supramolecular complexes have been built and experimentally tested in solution [45], and, more importantly, (iii) the protein assembling into discrete clusters on physiological DNA substrates was directly observed by AFM-based experiments [47].

Several authors have analysed the possible functional benefits of performing alkylated-DNA direct repair in a co-operative manner. It has been pointed out that co-operative 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-co-operative DNA-binding model [44]. 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 [48]. 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 [44,47].

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 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 [48]. 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 CONTRIBUTION

Riccardo Miggiano, Menico Rizzi and Franca Rossi designed experiments, analysed structural and biochemical data, and wrote the paper. Riccardo Miggiano performed structural and biochemical characterization experiments. Giuseppe Perugino and Maria Ciaramella designed experiments for biochemical characterization and analysed biochemical data. Mario Serpe performed EMSAs. Dominik Rejman designed experiments for E1X-dsDNA synthesis and analysed mass and NMR spectra. Dominik Rejman, Ondřej Páv and Radek Pohl developed and performed a new synthetic procedure for E1X-dsDNA preparation. Silvia Garavaglia performed crystal data collection analysis. Samarrita 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

- 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 [CrossRef PubMed](#)
- Dos Vultos, T., Mestre, O., Tonjum, T. and Gicquel, B. (2009) DNA repair in *Mycobacterium tuberculosis* revisited. *FEMS Microbiol. Rev.* **33**, 471–487 [CrossRef PubMed](#)
- Durbach, S.I., Springer, B., Machowski, E.E., North, R.J., Papavinasundaram, 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 [CrossRef PubMed](#)
- 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 [CrossRef PubMed](#)
- Drablos, F., Feyzi, E., Aas, P.A., Vaagbo, C.B., Kavli, B., Brattlie, 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 [CrossRef PubMed](#)
- 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 [CrossRef PubMed](#)
- 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 [CrossRef PubMed](#)
- 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 [CrossRef PubMed](#)
- Pegg, A.E. (2000) Repair of O⁶-alkylguanine by alkyltransferases. *Mutat. Res.* **462**, 83–100 [CrossRef PubMed](#)
- 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 [CrossRef PubMed](#)
- 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 [CrossRef PubMed](#)
- 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 [CrossRef PubMed](#)
- Kanugula, S., Goodtzova, K. and Pegg, A.E. (1998) Probing of conformational changes in human O6-alkylguanine-DNA alkyltransferase protein in its alkylated and DNA-bound states by limited proteolysis. *Biochem. J.* **329**, 545–550 [CrossRef PubMed](#)
- Xu-Welliver, M. and Pegg, A.E. (2002) Degradation of the alkylated form of the DNA repair protein, O⁶-alkylguanine-DNA alkyltransferase. *Carcinogenesis* **23**, 823–830 [CrossRef PubMed](#)

- 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 [CrossRef](#) [PubMed](#)
- 16 Boshoff, H.I., Myers, T.G., Copp, B.R., McNeil, M.R., Wilson, M.A. and Barry, 3rd, C.E. (2004) The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *J. Biol. Chem.* **279**, 40174–40184 [CrossRef](#) [PubMed](#)
- 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. et al. (2003) Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J. Exp. Med.* **198**, 693–704 [CrossRef](#) [PubMed](#)
- 18 Ebrahimi-Rad, M., Bifani, P., Martin, C., Kremer, K., Samper, S., Razuzier, J., Kreiswirth, B., Blazquez, J., Jouan, M., van Soolingen, D. et al. (2003) Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerg. Infect. Dis.* **9**, 838–845 [CrossRef](#) [PubMed](#)
- 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 [CrossRef](#) [PubMed](#)
- 20 Mestre, O., Luo, T., Dos Vultos, T., Kremer, K., Murray, A., Namouchi, A., Jackson, C., Razuzier, J., Bifani, P., Warren, R. et al. (2011) Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS One* **6**, e16020 [CrossRef](#) [PubMed](#)
- 21 Miggiano, R., Casazza, V., Garavaglia, S., Ciarabella, 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 [CrossRef](#) [PubMed](#)
- 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 [CrossRef](#) [PubMed](#)
- 23 Wibley, J.E., Pegg, A.E. and Moody, P.C. (2000) Crystal structure of the human O⁶-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res.* **28**, 393–401 [CrossRef](#) [PubMed](#)
- 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 [CrossRef](#) [PubMed](#)
- 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 [CrossRef](#) [PubMed](#)
- 26 Noll, D.M. and Clarke, N.D. (2001) Covalent capture of a human O⁶-alkylguanine alkyltransferase-DNA complex using N¹,O⁶-ethanoxanthosine, a mechanism-based crosslinker. *Nucleic Acids Res.* **29**, 4025–4034 [PubMed](#)
- 27 Collaborative Computational Project, Number 4 (CCP4) (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 [CrossRef](#) [PubMed](#)
- 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 [CrossRef](#) [PubMed](#)
- 29 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 [CrossRef](#) [PubMed](#)
- 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. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 [CrossRef](#) [PubMed](#)
- 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 [CrossRef](#)
- 32 Perugino, G., Vettone, A., Illiano, G., Valenti, A., Ferrara, M.C., Rossi, M. and Ciarabella, M. (2012) Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J. Biol. Chem.* **287**, 4222–4231 [CrossRef](#)
- 33 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 [CrossRef](#) [PubMed](#)
- 34 Perugino, G., Miggiano, R., Serpe, M., Vettone, A., Valenti, A., Lahiri, S., Rossi, F., Rossi, M., Rizzi, M. and Ciarabella, M. (2015) Structure–function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic Acids Res.* **43**, 8801–8816 [CrossRef](#) [PubMed](#)
- 35 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 [PubMed](#)
- 36 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 [CrossRef](#) [PubMed](#)
- 37 Roberts, A., Pelton, J.G. and Wemmer, D.E. (2006) Structural studies of MJ1529, an O6-methylguanine-DNA methyltransferase. *Magn. Reson. Chem.* **44**, S71–S82 [CrossRef](#) [PubMed](#)
- 38 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 [CrossRef](#) [PubMed](#)
- 39 Edara, S., Goodtzova, K. and Pegg, A.E. (1995) The role of tyrosine-158 in O6-alkylguanine-DNA alkyltransferase activity. *Carcinogenesis* **16**, 1637–1642 [CrossRef](#) [PubMed](#)
- 40 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 [PubMed](#)
- 41 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 [CrossRef](#) [PubMed](#)
- 42 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 [CrossRef](#) [PubMed](#)
- 43 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 [CrossRef](#) [PubMed](#)
- 44 Melikishvili, M., Rasimas, J.J., Pegg, A.E. and Fried, M.G. (2008) Interactions of human O⁶-alkylguanine-DNA alkyltransferase (AGT) with short double-stranded DNAs. *Biochemistry* **47**, 13754–13763 [CrossRef](#) [PubMed](#)
- 45 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 [CrossRef](#) [PubMed](#)
- 46 Adams, C.A. and Fried, M.G. (2011) Mutations that probe the cooperative assembly of O⁶-alkylguanine-DNA alkyltransferase complexes. *Biochemistry* **50**, 1590–1598 [CrossRef](#) [PubMed](#)
- 47 Tessmer, I. and Fried, M.G. (2014) Insight into the cooperative DNA binding of the O⁶-alkylguanine DNA alkyltransferase. *DNA Repair (Amst.)* **20**, 14–22 [CrossRef](#) [PubMed](#)
- 48 Begley, T.J. and Samson, L.D. (2004) Reversing DNA damage with a directional bias. *Nat. Struct. Mol. Biol.* **11**, 688–690 [CrossRef](#) [PubMed](#)

A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering.

Antonella Vettone*¹, Mario Serpe*¹, Aurelio Hidalgo², José Berenguer², Giovanni del Monaco¹, Anna Valenti¹, Mosé Rossi¹, Maria Ciaramella¹, Giuseppe Perugino¹.

¹Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80125 Naples, Italy.

²Centro De Biología Molecular Severo Ochoa, Nicolás Cabrera 1, Cantoblanco, 28049 Madrid, Spain.

*The author equally contributed to the present work.

*A powerful protein-tag for the in vivo and in vitro specific labelling of proteins was proposed as a valid alternative to the new SNAP-tag™ technology, which is based on the irreversible reaction of the human hAGT in the presence of benzyl-guanine derivatives. However, this technology could be applied only in mild reaction conditions, given the mesophilic nature of the hAGT. For these reasons, this new approach is not completely applicable to extremophilic organisms and/or in all the applications requiring harsh reaction conditions. Here, we introduce an engineered variant of SsOGT (namely H⁵), which displays a high catalytic activity comparable to the SNAP-tag™, but keeping the stability behaviours of the wt counterpart. The successful heterologous expression obtained in the thermophilic model organism *Thermus thermophilus* makes SsOGT-H⁵ a valid candidate as protein-tag for extreme environments.*

*My contribution was the cloning of the ogt and the ogt-H5 genes in a plasmid for the heterologous expression of the protein in *T.thermophilus*.*

A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering

Antonella Vettone¹ · Mario Serpe¹ · Aurelio Hidalgo² · José Berenguer² · Giovanni del Monaco¹ · Anna Valenti¹ · Mosé Rossi¹ · Maria Ciaramella¹ · Giuseppe Perugino¹

Received: 4 August 2015 / Accepted: 3 October 2015 / Published online: 23 October 2015
© Springer Japan 2015

Abstract In the last decade, a powerful biotechnological tool for the in vivo and in vitro specific labeling of proteins (SNAP-tagTM technology) was proposed as a valid alternative to classical protein-tags (green fluorescent proteins, GFPs). This was made possible by the discovery of the irreversible reaction of the human alkylguanine-DNA-alkyl-transferase (hAGT) in the presence of benzyl-guanine derivatives. However, the mild reaction conditions and the general instability of the mesophilic SNAP-tagTM make this new approach not fully applicable to (hyper-)thermophilic and, in general, extremophilic organisms. Here, we introduce an engineered variant of the thermostable alkylguanine-DNA-alkyl-transferase from the Archaea *Sulfolobus solfataricus* (SsOGT-H⁵), which displays a catalytic efficiency comparable to the SNAP-tagTM protein, but showing high intrinsic stability typical of proteins from this organism. The successful heterologous expression obtained in a thermophilic model organism makes SsOGT-H⁵ a valid candidate as protein-tag for organisms living in extreme environments.

Keywords Archaea · *Sulfolobus solfataricus* · Thermostable proteins · Protein-tag · Biotechnology

Communicated by S. Albers.

A. Vettone and M. Serpe equally contributed to the present work.

✉ Giuseppe Perugino
giuseppe.perugino@ibbr.cnr.it

¹ Institute of Biosciences and BioResources, National Council of Research of Italy, Via P. Castellino 111, 80131 Naples, Italy

² Centro De Biología Molecular Severo Ochoa, Nicolás Cabrera 1, Cantoblanco, 28049 Madrid, Spain

Introduction

Labeling of proteins with synthetic probes in living cells is an important approach to study protein function (Grone-meyer et al. 2006). However, radioactive and fluorescent probes need complex chemical modification and protein purification procedures, leading only in few cases to a specific labeling of proteins. The introduction of fusion polypeptides (tags), which allow to simplify and optimize purification methods (affinity tags) and/or detection procedures (by using specific antibodies) has been fundamental. Nevertheless, most tags do not satisfy the need of the in vivo specific detection of proteins of interest. The discovery of autofluorescent proteins as the *Aequorea victoria* green fluorescent protein (GFP) made possible the expansion of the studies on fusion proteins in several applicative fields of cellular and molecular biology (Chalfie et al. 1994; Tsien 1998). Despite their wide use in the last decades, GFP tag and its derivatives have some disadvantages and/or limitations: (1) their relative big dimensions (ca. 230 amino acids) in some cases might affect the target protein's function and localization; (2) the isolation of the natural fluorophore in the active site makes GFPs and variants partially sensitive to environmental changes, as pH (Ashby et al. 2004; Campbell and Choy 2000); (3) because the formation of the internal fluorophore is O₂-dependent, GFPs are not suitable tags in applications requiring anaerobic conditions (anaerobic organisms and tumor cells in particular stages); (4) the general use of most protein-tags is restricted to mesophiles and mild reaction conditions. Recently, thermostable GFPs variants have been developed as tags for thermophilic microorganisms (Aliye et al. 2015; Cava et al. 2008; Pédelacq et al. 2006); however, these tools, again, suffer from most of the same limitations listed above for their mesophilic counterparts.

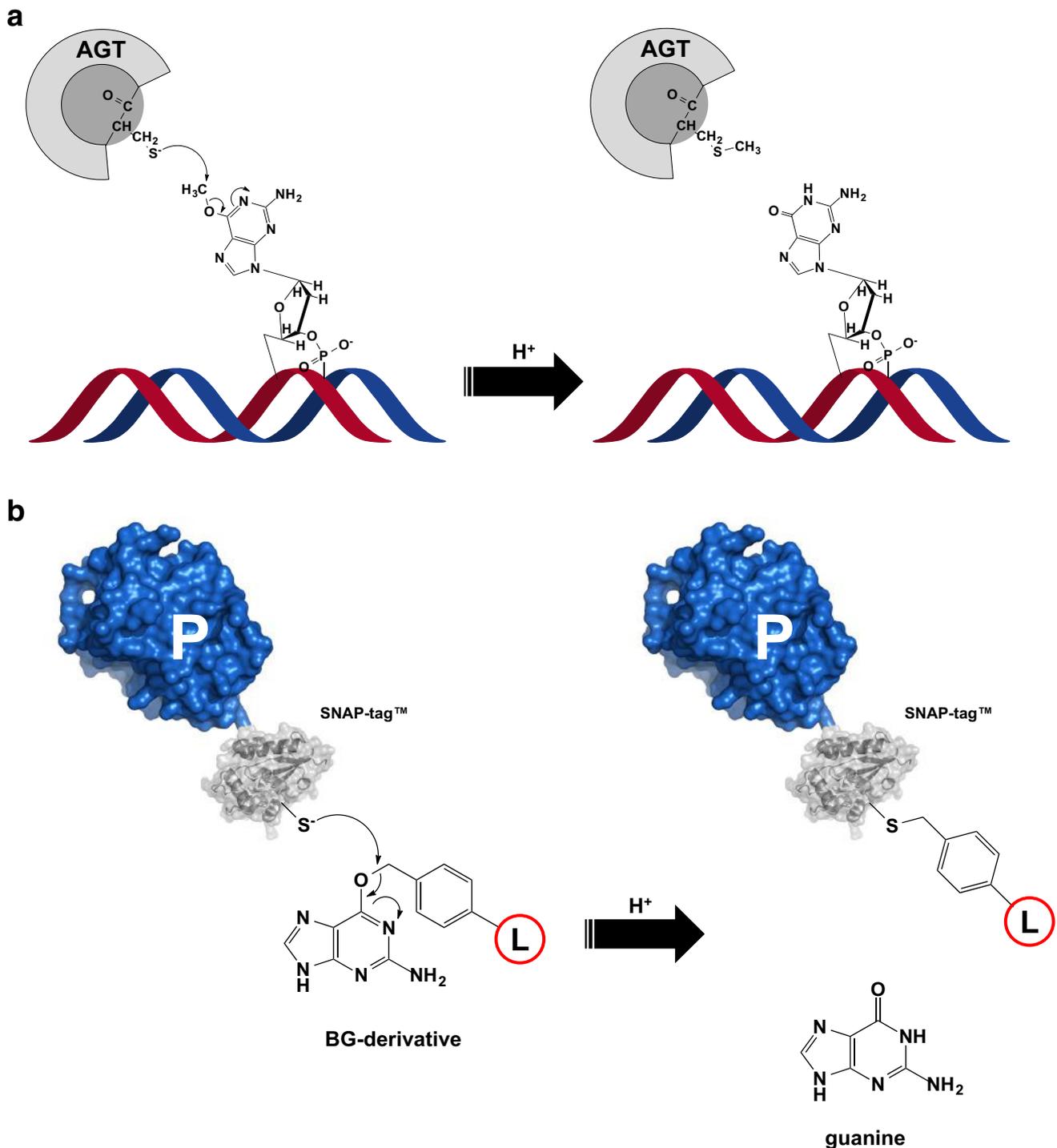


Fig. 1 **a** Direct repair of alkylation DNA damage by AGTs: reaction mechanism of the irreversible transfer of the alkyl group from the damaged guanine to a cysteine in the active site. **b** A variant of

*O*⁶-alkylguanine-DNA-alkyl-transferases (AGTs or MGMTs, EC: 2.1.1.63) are ubiquitous proteins involved in the direct repair of alkylation damage in DNA. By a one-step S_N2 reaction mechanism, they specifically transfer the alkyl group from the damaged base (*O*⁶-alkylguanine or

the human AGT (SNAP-tag™) can react in vitro and/or in vivo with BG-derivative substrates, leading to the labeling of a fused protein of interest (*P*) with a desired chemical group (*L*)

*O*⁴-alkylthymine) to a cysteine residue in their own active site (Daniels et al. 2000, 2004; Fang et al. 2005; Pegg 2011; Tubbs et al. 2007; Yang et al. 2009) (Fig. 1a). AGTs are called suicide or kamikaze proteins because of the 1:1 stoichiometry of their reaction, leading to an irreversible

inactivation and destabilization of the alkylated form of the protein (Fang et al. 2005; Yang et al. 2009). In 2003, the group of Kai Johnsson, studying the human AGT (hAGT), investigated the possibility to employ this protein as protein-tag, using benzyl-guanine-derivative substrates, also accepted by the active site of the protein (Keppler et al. 2003). The covalent and irreversible nature of the alkylated form of hAGT and its small size (ca. 20.0 kDa) allowed new biotechnological approaches, making it an interesting tool for in vivo and in vitro specific labeling of proteins, when fused to a protein of interest (Gautier et al. 2008; Keppler et al. 2003, 2004). A commercial variant of the hAGT (SNAP-tag™, New England Biolabs) is already available to be covalently labelled with a large number of chemical groups conjugated with O⁶-benzyl-guanine (BG-derivative) (Fig. 1b).

However, no thermophilic AGTs have been developed as biotechnological tools to date, making the SNAP-tag™ technology suitable only for mesophilic organisms and mild reaction conditions. Hence, a thermostable version of AGT could be useful for the development of protein-tags for organisms growing under harsh reaction conditions. We previously characterized an AGT from the thermoacidophilic archaeon *Sulfolobus solfataricus* (*SsOGT*), which has the same catalytic efficiency of hAGT with fluorescent BG-derivatives, but displays a marked stability over a wide range of temperature, pH, ionic strength and to common denaturing agents (Perugino et al. 2012). Here, we present the use of an engineered variant of the thermostable *SsOGT* protein, which lacks DNA binding activity (Perugino et al. 2012), as a protein-tag for extremophilic organisms.

Materials and methods

Reagents

All chemicals were purchased from Sigma-Aldrich, SNAP-Vista Green™ and CLIP-FI™ fluorescent substrates

(referred to BG-FL and BC-FL throughout, respectively) were from New England Biolabs (Ipswich, MA). Synthetic oligonucleotides were from Primm (Milan, Italy) and listed in Table 1, *Pfu* DNA polymerase were from Stratagene (La Jolla, CA), as well as *Escherichia coli* ABLE C strain. *Thermus thermophilus* HB27^{EC} strain is a derivative of DSM 7039 with enhanced natural competence because of a knockout mutation in the argonaute protein (Swarts et al. 2014). Protein concentration was determined by the BioRad protein assay kit (Bio-Rad Pacific), using BSA as standard.

DNA constructs

The DNA fragment of the beta-glycosidase from *S. solfataricus* gene (*lacS*) was PCR-amplified using the pGEX-K-Gly construct as template (Moracci et al. 1996), and by using the LacS-fwd and LacS-rev oligonucleotides (Table 1), which possess an internal *Pst* I site. This allowed the ligation of this gene downstream and in frame with the *ogtH⁵* gene in the pQE-*ogtH⁵* construct (Perugino et al. 2012), and leading to the pQE-*ogtH⁵-lacS* plasmid for the heterologous expression of the *SsOGT-H⁵-Ssβgl* fusion protein. pMK-*ogtH⁵* for the heterologous and constitutively expression in *T. thermophilus* of the *SsOGT-H⁵* mutant (referred to H⁵ throughout) was obtained by multiple rounds of PCR amplification: the PlspA-fwd/OP-rev and PO-fwd/Ogt3' oligonucleotides pairs were first used to amplify the PlspA promoter and the *ogtH⁵* gene DNA fragments, respectively. In a final round of PCR, the former two DNA fragments were fused to each other by the complementary nature of the PO-fwd and OP-rev oligonucleotides (Table 1), and amplified by using the external PlspA-fwd/Ogt3' oligonucleotides pair. Finally, the obtained DNA fragment was ligated in the multi-cloning site of the pMK184 (Cava et al. 2007) vector by using the *EcoR* I/*Hind* III sites. For all the obtained constructs, regions encoding the cloned genes were verified by DNA sequencing (Primm, Milan, Italy).

Table 1 List of oligonucleotides used in this study

Name	Sequence	Notes
LacS-fwd	5'-CCAGA <u>ACTGCAGATGTACTCATTCCAAATAGC</u> -3'	<i>Pst</i> I site underlined
LacS-rev	5'-CCAGA <u>ACTGCAGTTAGTGCCTTAATGGCTTTACTGG</u> -3'	<i>Pst</i> I site underlined
OP-rev	5'-GGTGATGGTGATGGTGAGATCCTCTC <u>ATatgcctcacacctccttaaggg</u> -3'	PlspA promoter sequence in lowercase; <i>ogt</i> gene sequence in uppercase; <i>Nde</i> I site underlined
PlspA-fwd	5'-aatggagaattcggggatcgtatccccggg-3'	<i>EcoR</i> I site underlined
PO-fwd	5'-cccttaaggaggtgtgaggcatATGAGAGGATCTCACCATCACCATCACC-3'	PlspA promoter sequence in lowercase; <i>ogt</i> gene sequence in uppercase; <i>Nde</i> I site underlined
Ogt-3'	5'-TATTAAGCTTTTATTCTGGTATTTGACTCCCTCTAGT-3'	Perugino et al. (2012); <i>Hind</i> III site underlined

Protein expression and purification

H⁵ and H⁵-Ssβgly were expressed in *E. coli* ABLE C and purified by His₆-tag affinity chromatography, as described (Perugino et al. 2012, 2015), with the exception of a further purification step for the fusion protein: briefly, to remove *E. coli* contaminants, the pool of the eluted fractions from the affinity chromatography was incubated for 20 min at 70 °C, and centrifuged for 30 min at 30,000g. The soluble fraction was dialysed against PBS 1× buffer (phosphate buffer 20 mM, NaCl 150 mM, pH 7.3) and aliquots stored at −20 °C. To assess the purity of the protein samples and determine their concentrations, SDS-PAGE and Bio-Rad protein assay were performed, respectively.

In vitro and in vivo alkyl-transferase assay

The activity of the purified proteins wild-type SsOGT and H⁵, alone or in fusion with Ssβgly, was measured by a fluorescent-based assay using BG-FL and BC-FL substrates in standard conditions, as previously described (Miggiano et al. 2013; Perugino et al. 2012, 2015). Briefly, in a total volume of 10 μL, 5.0 μM of protein (0.1 mg mL^{−1}) was incubated with 20 μM of the substrate in 1× Fluo Reaction Buffer (50 mM phosphate, 0.1 M NaCl, 1.0 mM DTT, pH 6.5) at different temperatures and times, as indicated. Reactions were stopped by denaturing and loading samples on SDS-PAGE, followed by fluorescence imaging analysis on a VersaDoc 4000™ system (Bio-Rad) by applying as excitation/emission parameters a blue LED/530 bandpass filter, respectively. For the determination of the second-order rate constants, assuming the irreversible mechanism with 1:1 substrate/enzyme binding, fluorescence intensity data were corrected for the amount of loaded protein (by coomassie staining analysis), in order to estimate the relative amount of covalently modified protein in time-course experiments. (Gautier et al. 2008; Miggiano et al. 2013; Perugino et al. 2012, 2015). For the in vivo assay, transformed cells of *T. thermophilus* HB27^{EC} strain with the pMK184 and pMK-*ogtH*⁵ plasmids were grown at 60 °C in TB selective medium (tryptone 8 g L^{−1}, yeast extract 4 g L^{−1}, NaCl 3 g L^{−1}, kanamycin 30 mg L^{−1}; in mineral water, pH 7.5) as late as stationary phase (O.D._{600 nm} >1.5) (Cava et al. 2009). Cell pellets from an appropriate volume (typically 1.0 mL) were resuspended in 0.1 mL of TB medium in the presence of 3.0 μM of BG-FL and incubated at several temperatures for different times, as indicated. After the reaction, cells were first washed twice with 1.0 mL of TB medium, then denatured for 10 min at 110 °C in O'Farrell 1× buffer supplemented

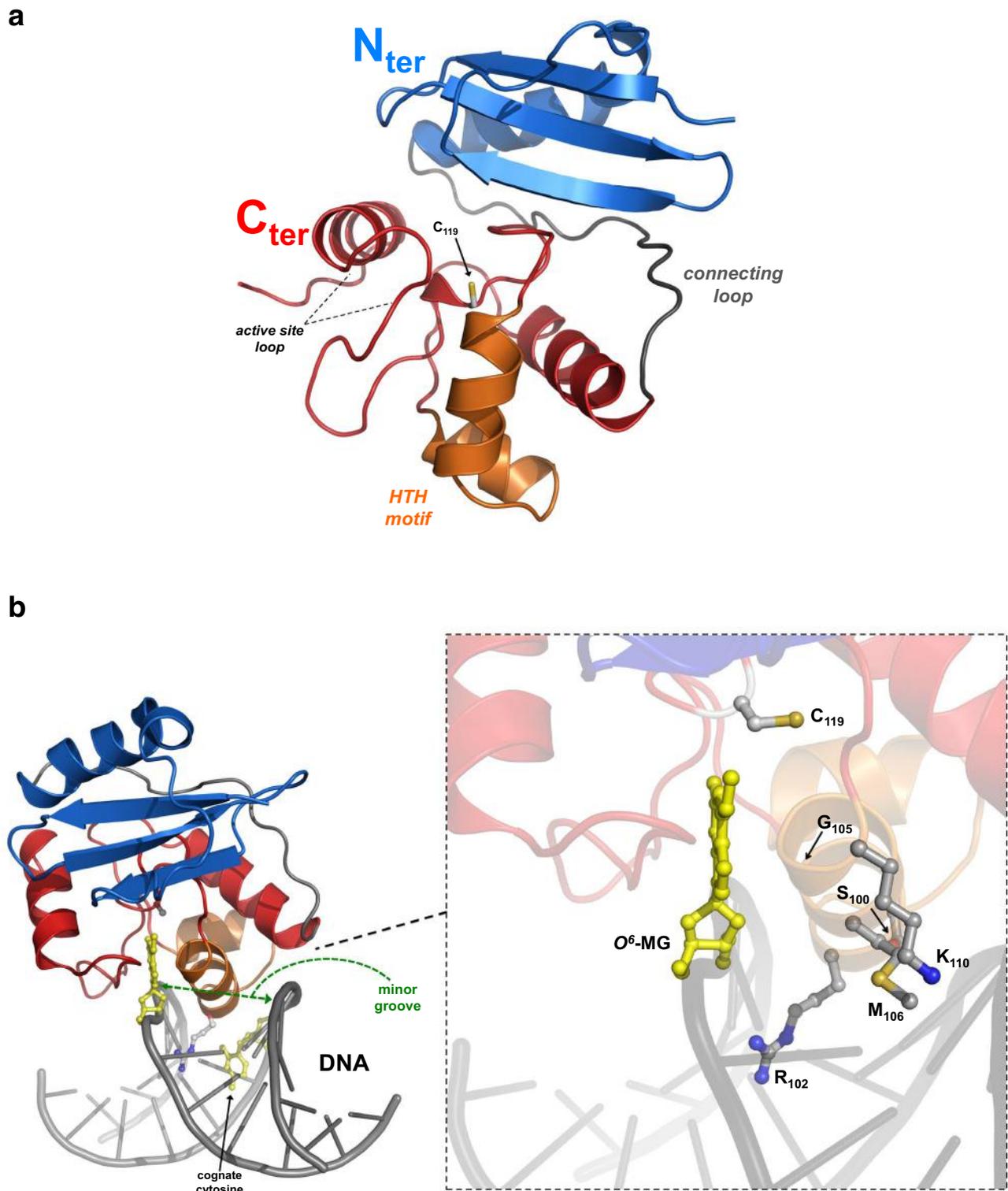
with EDTA 10 mM, and finally loaded on SDS-PAGE for the analysis as described above.

Protein stability analysis

The stability of the wild-type and H⁵ was determined by Differential Scan Fluorimetry (DSF), following an adaptation of a protocol previously described (Niesen et al. 2007; Perugino et al. 2015): samples containing 25 μM of each protein (0.5 mg mL^{−1}) in PBS 1× buffer and SYPRO Orange dye 1× were subjected to a temperature scanning from 20 to 95 °C at 0.2 °C min^{−1} (5 min per cycle with an increase of 1 °C per cycle) in a Real-Time Light Cycler™ (Bio-Rad). Relative fluorescence data were normalized to the maximum fluorescence value within each scan. Finally, plots of fluorescence intensity as a function of temperature showed increasing fluorescence sigmoidal curves (which described a two-state transition), leading to the determination of the inflection points, indicating T_m values, as described by the Boltzmann equation (Niesen et al. 2007). Stability was also tested by protease attack resistance and by in vivo protein degradation in *T. thermophilus* cells. In the first analysis, in a total volume of 10 μL, 1.0 μM of each protein (0.02 mg mL^{−1}) was incubated for 2 h at 25 °C in the presence of trypsin protease at different substrate/enzyme weight ratio, as indicated. Reactions were then denatured for 5 min at 100 °C in O'Farrell 1× buffer and immediately loaded on SDS-PAGE (0.2 μg per lane). Gels were subjected to Western Blotting analysis by using an anti-SsOGT antibody, as previously described (Perugino et al. 2012). For the preparation of the fluoresceinated form of the wild-type and H⁵, before the trypsin treatment, a reaction with BG-FL for 2.0 h at 25 °C was performed, followed by a gel fluorescence imaging to assess the complete labeling. To determine the stability of alkylated SsOGT in *T. thermophilus* HB27^{EC} strain, we followed the same procedure described above to assay the protein activity; after incubation with BG-FL and washing, cells were incubated at 70 °C for different times, followed by denaturation, SDS-PAGE and fluorescence imaging analysis.

Beta-glycosidase assay

Glycoside hydrolytic activity of H⁵-Ssβgly was detected in vivo on LB Agar selective medium in the presence of the chromogenic substrate X-gal (Sambrook et al. 1989), as well as by calculating the steady-state kinetic constants with 4-nitrophenyl-β-D-galactopyranoside (4Np-gal) and 4-nitrophenyl-β-D-glucopyranoside (4Np-glc), by using a Cary E1 spectrophotometer (Varian) and following the



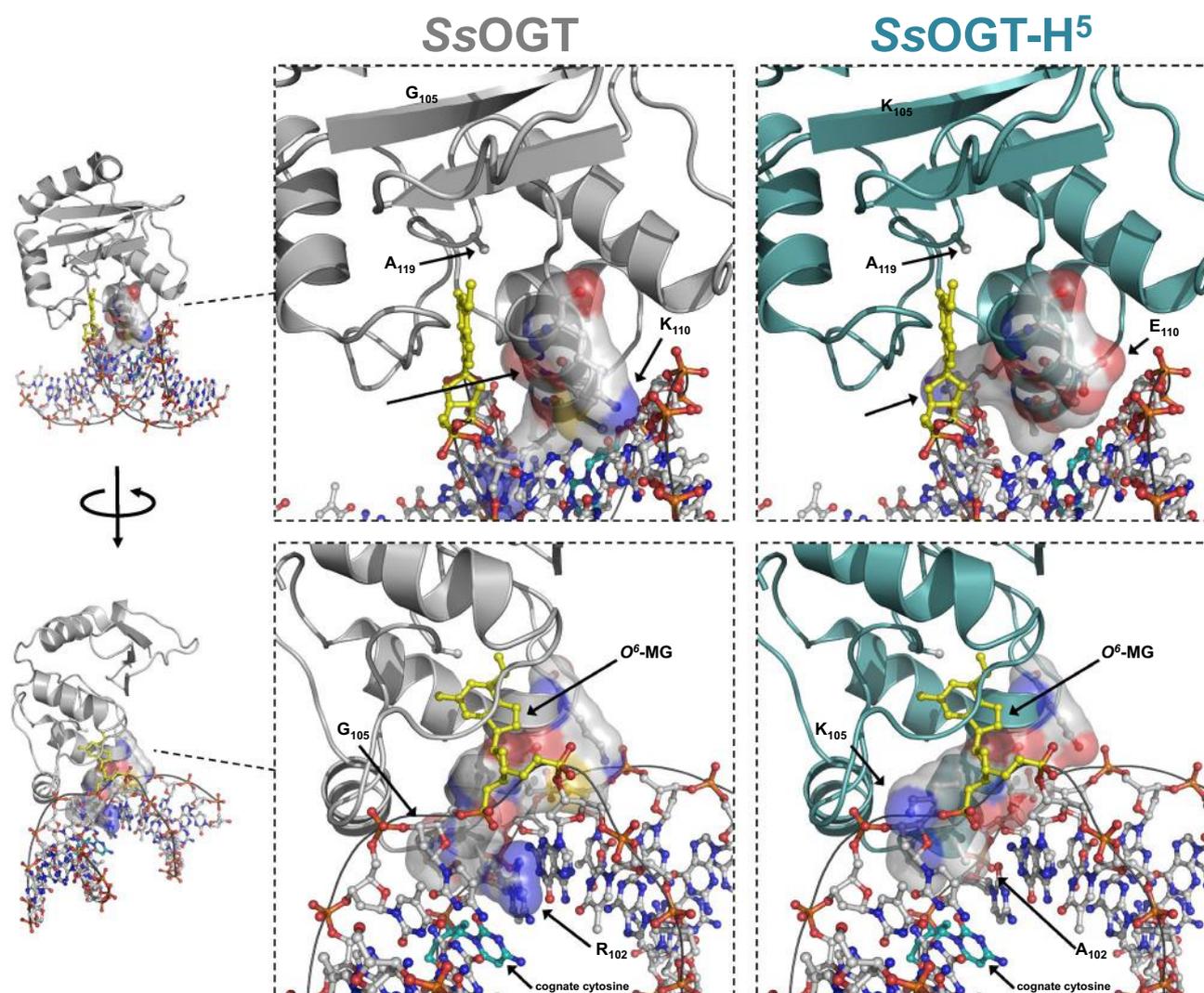


Fig. 3 Comparison of the crystal structure of methylated-DNA::SsOGT-C119A complex of the SsOGT-C119A (gray) with a 3D model of the same DNA molecule in complex with the H⁵ mutant (cyan): DNA, O⁶-MG and residues involved in the mutagenesis are represented as ball and sticks. A spheric representation in transpar-

ency in the SsOGT-C119A and H⁵ for the five residues involved in the mutagenesis is also shown. Atoms are colored by the CPK convention, whereas O⁶-MG and the unpaired cytosine are colored in yellow and cyan, respectively

same procedure used for the Ssβgly enzyme (Moracci et al. 1996, 1998).

Data analysis and softwares

Analysis of the solved 3D structures of the wild-type and the methylated DNA::SsOGT-C119A complex, as well as the modeling of the methylated DNA::H⁵ complex, were performed by using the MacPyMOL freeware (DeLano Scientific FCC). Data from activity and stability assays were fitted to appropriate equations by using the Prism Software Package (GraphPad Software) and GraFit 5.0 Data Analysis Software (Erithacus Software).

Results

Mutations at the HTH motif affect the DNA-binding by SsOGT

The 3D structure of SsOGT, recently solved by X-ray crystallography, is overall very similar to that of all other AGTs and consists of two domains connected by a loop (Perugino et al. 2015) (Fig. 2a). The role of the N_{ter} domain of these proteins is not completely understood and is likely involved in the protein overall stability (Perugino et al. 2015) as well as activity (Miggiano et al. 2013), whereas the highly conserved C_{ter} domain contains all the elements for the

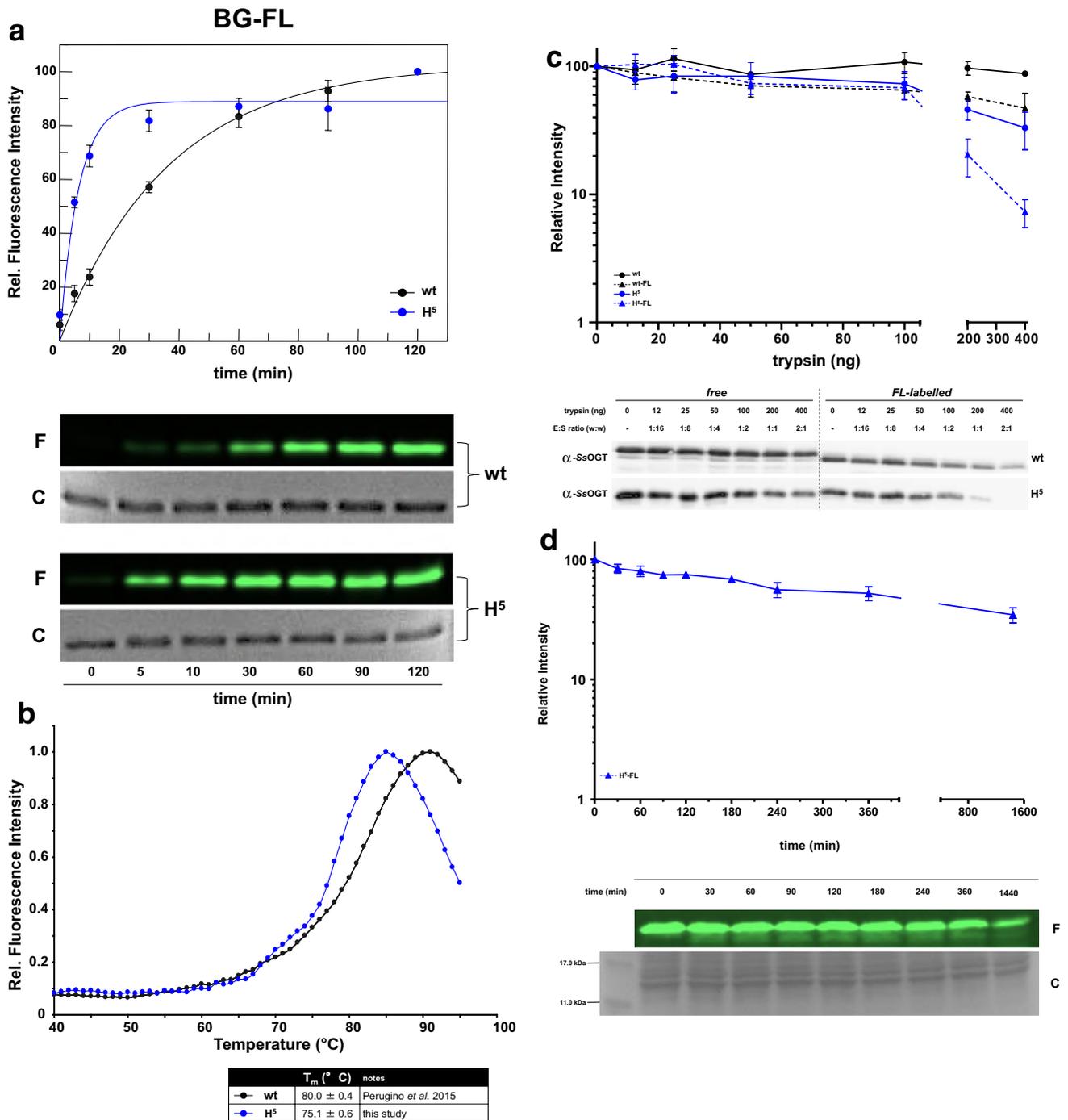


Fig. 4 **a** Time-course of wild-type (Perugino *et al.* 2012) and H⁵ under standard conditions with 20 μ M of SNAP-Vista Green™ (BG-FL) at 25 °C. Analysis and quantification of protein activity, as well as second-order rate constant values determination were performed as previously described (Perugino *et al.* 2012), and shown in Table 2. **b** Stability of H⁵ in comparison with the wild-type protein by DSF analysis, as described in “Materials and methods”. T_m values were obtained by measuring the relative fluorescence intensity of each protein from three independent experiments, as a function of temperature. **c** Resistance to protease attack of the wild-type SsOGT and H⁵: proteins (ca. 0.2 μ g per lane) were incubated for 2 h at 25 °C in the

presence of different amounts of trypsin, as indicated. After SDS-PAGE, gels were analyzed by fluorescence imaging and then subjected to western blot analysis using a polyclonal anti-SsOGT antibody (Perugino *et al.* 2012). Fluorescein-labelled forms of wild-type and H⁵ are indicated as wt-FL and H⁵-FL, respectively. **d** In vivo stability of fluoresceinated H⁵: after protein labeling procedure with BG-FL (see “Materials and methods”), intact *T. thermophilus* cells were incubated at 70 °C for different times as indicated. Acrylamide gel was analyzed by fluorescence imaging and coomassie staining (to correct fluorescence for the amount of loaded proteins)

Table 2 Second order rate constant values of wild-type and H⁵ under standard conditions on the BG-FL substrate

<i>T</i> (°C)	<i>k</i> (s ⁻¹ M ⁻¹)	References
Wild-type		
70	5.33 × 10 ⁴	Perugino et al. (2012)
25	2.80 × 10 ³	Perugino et al. (2012)
H ⁵		
70	3.78 × 10 ⁴	Perugino et al. (2012)
25	1.60 × 10 ⁴	This study

Incubation temperatures are indicated

recognition and the repair of the damaged guanine. In contrast to most DNA binding proteins, AGTs contact the DNA via the minor groove through their helix turn helix (HTH) motif (Daniels et al. 2004) (Fig. 2b).

We previously prepared a mutant version of SsOGT (H⁵), whose mutations in five conserved residues of the HTH motif (S100A, R102A, G105K, M106T, K110E; zoom in Fig. 2b) impair dramatically the DNA binding capacity of the protein (Perugino et al. 2012). The availability of the crystal structure of the methylated DNA bound SsOGT (Perugino et al. 2015) allowed us to obtain a reliable model of the H⁵-DNA complex 3D structure (Fig. 3), which showed that: i) the substitution of the positively charged K110 with a glutamic acid residue could result in repulsive effects toward DNA phosphate backbone; ii) the replacement of the arginine finger R102 by an alanine residue might hamper the stabilization of the transiently unpaired cytosine opposite the methylated G (Kanugula et al. 1995; Liu et al. 2007; Perugino et al. 2012; Zang et al. 2005); iii) the lysine in place of the G105 residue could produce a dramatic steric hindrance and collide with the phosphate backbone of the DNA strand close to the O⁶-MG. This analysis provided a structural explanation of the complete inability of H⁵ to bind DNA, which is an important pre-requisite for its application as a protein-tag.

Activity and stability of H⁵

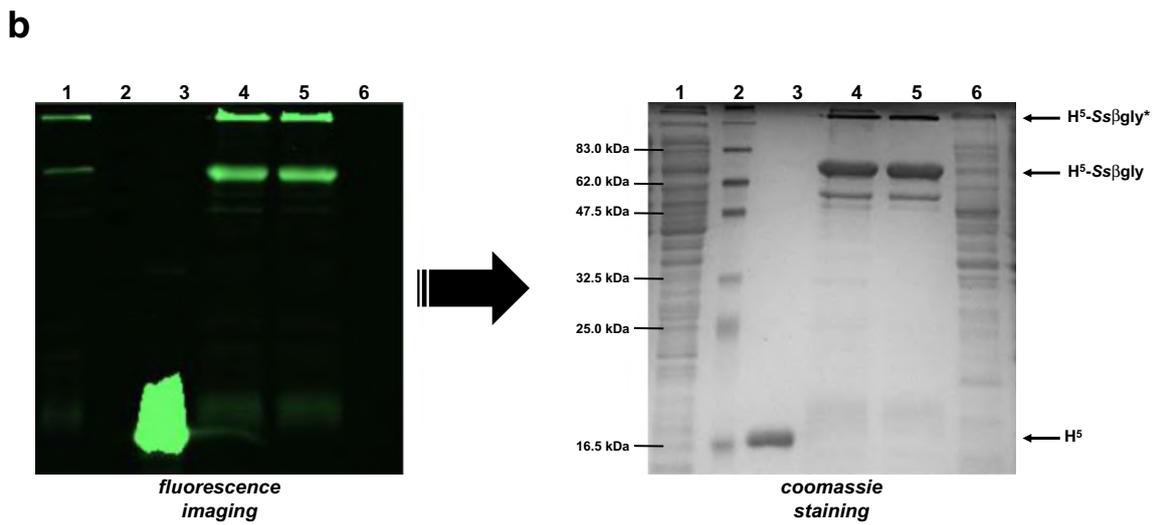
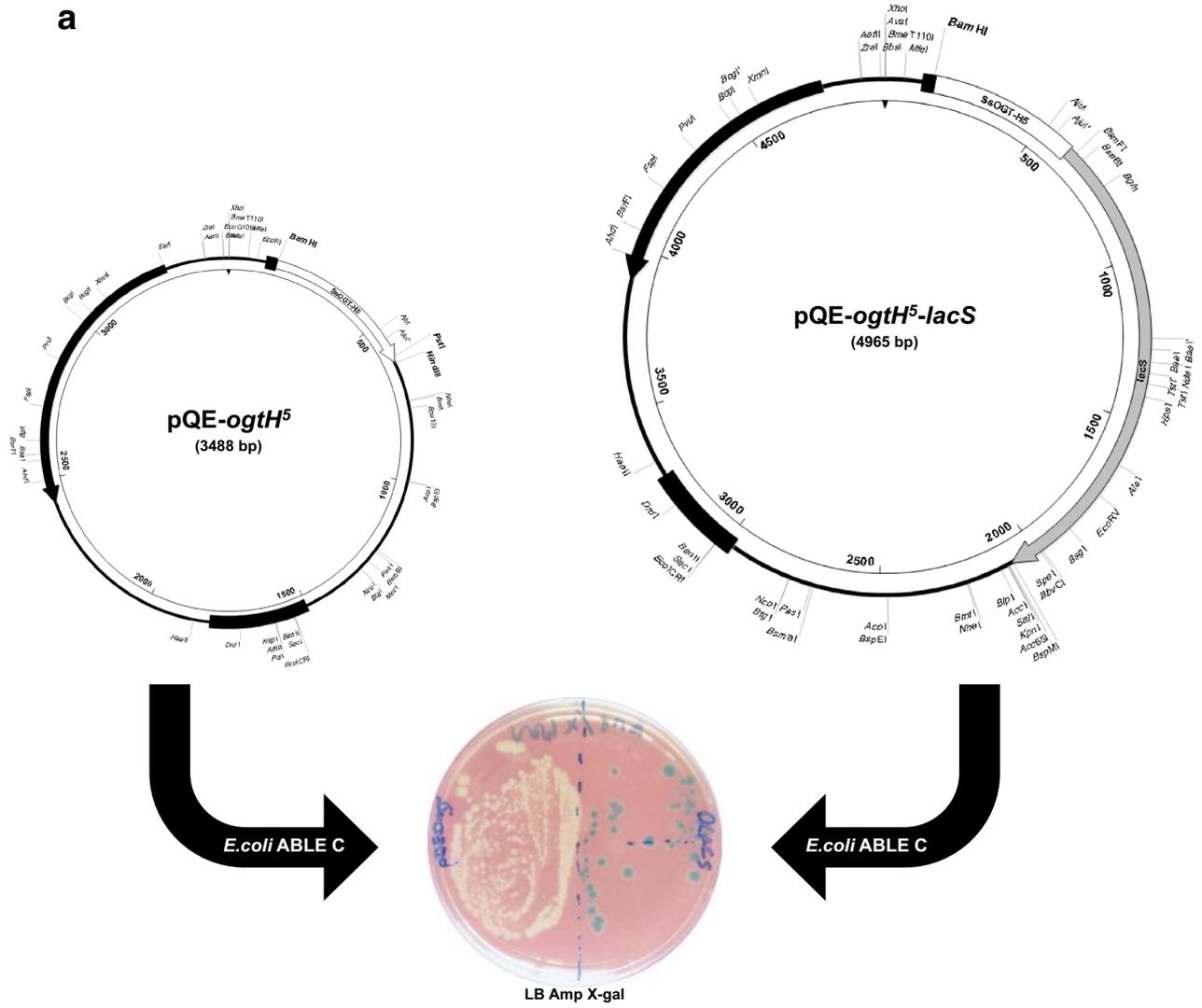
Previous biochemical analysis suggested that H⁵ might be a good candidate thermostable protein-tag because: i) it retained the catalytic activity on the fluorescent BG-FL substrate, and, like SsOGT, showed significant activity over a broad range of reaction conditions (Perugino et al. 2012); ii) the absence of any structural Zn²⁺ ion in the SsOGT structure (Perugino et al. 2015), makes it insensitive to the presence of chelating agents (EDTA up to 10 mM; Perugino et al. 2012). Furthermore, H⁵ displays a significant increase of the activity at any temperature tested, if compared to the wild-type SsOGT (Perugino et al. 2012): the second-order rate constant value at 25 °C on a

Fig. 5 **a** *E. coli* ABLE C strain was transformed by using pQE-*ogtH⁵* (Perugino et al. 2012) and the pQE-*ogtH⁵-lacS* plasmids. Transformed cells were plated on LB agar in the presence of ampicillin and the X-gal chromogenic substrate. **b** SDS-PAGE of the expression and purification by His₆-tag affinity chromatography of the H⁵-Ssβgly fusion protein: 1 cell free extract of *E. coli* ABLE C/pQE-*ogtH⁵-lacS*; 2 protein molecular weight marker; 3 2.0 μg of H⁵ protein; 4 pool of fractions eluted by imidazole; 5 soluble fraction of the sample of lane 4, after thermal treatment (see “Materials and methods”); 6 whole *E. coli* ABLE C cells transformed with the pQE31™ empty vector (Qiagen). All samples, except for the protein marker, were incubated with the BG-FL substrate under standard conditions, and loaded on SDS-PAGE. After, the gel was exposed for the fluorescence imaging analysis (left) and then stained by coomassie blue (right), as described in “Materials and Methods”. The high molecular weight band marked by an asterisk correspond to a partially denatured form of the fusion protein, which is particularly resistant to denaturation, as previously shown for the tetrameric Ssβgly (Moracci et al. 1996, 1998)

BG-derivative substrate was one order of magnitude higher than the wild-type (Fig. 4a; Table 2). This result suggests that the replaced residues in the HTH motif mutated in H⁵ play a role in the overall flexibility of the protein during the protein-substrate complex formation.

The substrate specificity of the wild-type and H⁵ proteins was evaluated by reactions in the presence of a benzyl-cytosine-derivative substrate (BC-FL): the failure of complete labeling hampered the determination of second-order rate constants (data not shown), confirming that alkylated cytosines are poor substrates for natural AGTs.

Previous experiments showed that H⁵ was only marginally less stable than the wild-type (Perugino et al. 2012). To compare the stability of the H⁵ protein with that of the SNAP-tag™, we performed quantitative analysis by DSF. We first followed the protocol used for SNAP-tag™ (ramping of 1 %: 1 min per cycle with an increase of 1 °C per cycle; Mollwitz et al. 2012), leading to a *T_m* value for the SNAP-tag™ protein of ca. 67 °C. However, these conditions were not sufficient to induce the H⁵ protein denaturation, since no fluorescence change was obtained up to 95 °C, as also observed for the wild-type protein (data not shown). Melting temperature values were achieved only increasing the incubation times by heating samples from 20 to 95 °C with a ramping of 0.2 % (5 min per cycle with an increase of 1 °C per cycle). Under these conditions, we calculate a *T_m* value of 75 °C (Fig. 4b; Perugino et al. 2015). Although this value is comparable to that obtained for SNAP-tag™ (67 °C; Mollwitz et al. 2012), the stronger conditions used in the DSF analysis for H⁵ clearly indicate that this protein is more stable than its mesophilic counterpart. Stability was also analyzed by the intrinsic resistance to protease attack of free and labelled form of both wild-type and H⁵. An incubation for 2 h at 25 °C in the presence of different amounts of trypsin protease was performed (Fig. 4c). Due to their intrinsic stability, both proteins were unaffected by the action of the trypsin in most of the conditions tested (from 1:16 to 1:2 *E:S* ratio), notably higher than



those used for hAGT (Kanugula et al. 1998; Mollwitz et al. 2012). To observe a reasonable reduction of the western blot signals, a very high amount of protease was needed (1:1 and 2:1 *E:S* ratio): as expected, alkylated (fluoresceinated) forms of both proteins were less stable than their unlabelled counterparts. Again, a slight difference between the wild-type and H⁵ proteins was found at high *E:S* ratios; overall data confirmed the general stability and resistance to protease attack of these thermostable AGTs.

Application of H⁵ as protein-tag

In order to test the possibility to use H⁵ as a protein-tag, we used as a model the *S. solfataricus lacS* gene coding for a thermophilic β -glycosidase (*Ss* β gly; Moracci et al. 1996); previous studies demonstrated that tags linked at the N-terminus of *Ss* β gly (such as a GST-tag) did not affect the overall folding of the latter (Moracci et al. 1996, 1998). The *lacS* gene was fused downstream to and in frame with the *ogtH⁵* gene, for the expression of the H⁵-*Ss* β gly fusion protein in the *E. coli* ABLE C *lac*⁽⁻⁾ strain. Blue colonies on X-gal LB agar solid medium appeared, indicating the presence of a hexogenous β -galactosidase activity, and confirming that *Ss* β gly is active also as fusion protein even at mesophilic temperatures (Fig. 5a). Purification of the fusion protein by affinity chromatography through a nickel column was performed, exploiting the presence of a His₆-tag at the N-terminus of the H⁵ moiety. Interestingly, the presence of *Ss* β gly did not interfere with the H⁵ activity, making it possible to follow all purification steps by SDS-PAGE and fluorescence imaging of the samples, previously incubated with the BG-FL substrate. Finally, we were able to calculate the kinetic constants of the hydrolytic activity of H⁵-*Ss* β gly at high temperatures, which were comparable to those previously described for the *Ss* β gly enzyme (Table 3) (Moracci et al. 1996). Importantly, the utilization of the thermostable protein-tag

enabled us to proceed with a thermoprecipitation treatment to further purify the pool of affinity chromatography eluted fractions (Fig. 5b, lane 5), and to perform kinetic analysis on a thermophilic enzyme at high temperatures without the need to remove the tag.

Heterologous expression of H⁵ in *T. thermophilus*

Thermus thermophilus HB27^{EC} was chosen as model thermophilic cell host to test our protein-tag because it is a naturally *agt*⁽⁻⁾ strain, thus lacking any endogenous DNA-alkyl-transferase activity. It encodes a single AGT-like protein (ATL), which is inactive in DNA repair, but involved in the recruiting of nucleotide excision repair (NER) proteins (Morita et al. 2008). The *ogtH⁵* gene was cloned in the pMK184 plasmid (Fig. 6a) (Cava et al. 2007), a shuttle vector containing: i) both *E. coli* and *T. thermophilus* HB27 replication origins; ii) the *kat* gene for growth in kanamycin selective medium; iii) a multi-cloning site (MCS) upstream of the α -*lacZ* gene, allowing *E. coli* white/blue colony screening on X-gal containing plates. As in *E. coli* cells (Perugino et al. 2012), it was possible to check the in vivo activity of H⁵, by directly incubating intact cells in the TB medium supplemented with the BG-FL substrate. To test the permeability of *T. thermophilus* (a bacterium with a complex cell envelope, including an outer membrane) to this substrate, we incubated at different temperatures and also in the presence of non-toxic organic solvents, such as DMSO, to improve cell permeabilization. Figure 6b shows fluorescent signals at the same molecular weight of the purified H⁵ protein in whole cell extracts from *Tth*HB27/pMK-*ogtH⁵* transformants, whereas no signals were seen in cells transformed with the empty plasmid (*Tth*HB27/pMK184), confirming the lack of any endogenous alkyl-transferase activity. The maximal activity was obtained at 60 and 70 °C, likely due to the higher permeability of the cells at their physiological temperatures concomitant with an increase of activity of the thermophilic H⁵ protein (Fig. 6, lanes 5, 6 and 9). The presence of organic solvents impaired the protein activity and/or the substrate permeability, as suggested by the absence of the unreacted substrate at the bottom of the gel at the highest concentration of DMSO (lane 8 in Fig. 6). As for most of model organisms used, BG-FL molecule resulted non-toxic for *T. thermophilus* HB27^{EC} cells, as indicated by a TB agar plate (“spot”) assay (data not shown) (Wilkinson et al. 2012).

Moreover, the fluoresceinated form of H⁵ was reasonably stable in vivo in *T. thermophilus* cells, showing a slow fluorescence signal decay: ca. 50 % of the protein was still detectable after 24 h at high temperatures (Fig. 4d), a time span compatible with most of the experiments performed by using the SNAP-tagTM technology in vivo.

Table 3 Steady state kinetic constants of *Ss* β gly and the H⁵-*Ss* β gly fusion protein for the hydrolysis of chromogenic substrates at 65 °C in 50 mM phosphate buffer pH 6.5

Substrate	K_M (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_M (s ⁻¹ mM ⁻¹)	References
<i>Ss</i> β gly				
4Np-gal	1.17 ± 0.06	275.0 ± 4.0	235.0	Moracci et al. (1996)
4Np-glc	0.30 ± 0.04	240.0 ± 7.0	800.0	Moracci et al. (1996)
H ⁵ - <i>Ss</i> β gly				
4Np-gal	2.13 ± 0.50	205.2 ± 14.4	96.5	This study
4Np-glc	0.04 ± 0.02	88.5 ± 7.1	2328.2	This study

if conjugated to a benzyl-guanine, a classical inhibitor of hAGT and a suitable substrate of its commercially available variant (Hinner and Johnsson 2010).

The discovery of a BG-sensitive thermostable OGT from the thermoacidophilic Archaea *S. solfataricus* opened the possibility to widen the SNAP-tag™ technology to organisms which thrive in extreme environments. Because of the intrinsic nature of proteins from hyperthermophiles, SsOGT displayed a strong stability under several harsh conditions, as extremes of pH, ionic strength, temperatures, presence of organic solvents (Perugino et al. 2012), and protease treatments. Mutagenesis at the expense of the HTH motif allowed the construction of a variant of this enzyme (H⁵), which retained most of the advantages of the native protein in terms of stability, and showed even enhanced activity at lower temperatures. We showed that H⁵ is correctly folded, expressed, functionally active and stable in both mesophilic *E. coli* and thermophilic *T. thermophilus* hosts. In order to obtain a smaller version of SsOGT, a preliminary attempt to reduce the wt protein by limited proteolysis was performed, obtaining a truncated polypeptide (ca. 14.0 kDa) which was still active in the presence of BG-FL substrate. Gel purification and N-terminal sequence analysis revealed the loss of the first 36 aminoacids. However, this “mini SsOGT” protein could not be obtained by heterologous expression in *E. coli* (data not shown), suggesting that the first 36 aminoacids are important for the correct folding of the protein during translation. Interestingly, inspection of the 3D structure revealed the presence of a disulfide bond at the N_{ter} (C²⁹-C³¹), which is important for the protein stability (Perugino et al. 2015).

The analysis of the H⁵-Ssbgly fusion protein demonstrated that it is possible to study both mesophilic and thermophilic proteins/enzymes fused to H⁵ under their own physiological conditions, without the need to remove the tag. These properties, together with the complete abolition of binding to DNA, make H⁵ a robust alternative as protein-tag for the application of the SNAP-tag™ technology for in vivo (protein–protein interaction, in situ localization, FRET experiment in combination with thermostable GFP variants, etc.) and in vitro studies (heterologous expression and purification of proteins of interest in extremophilic organisms). Finally, the high substrate specificity toward BG-derivatives makes H⁵ an interesting starting point to be modified by molecular evolution in order to obtain a thermostable variant active on BC-derivative substrates (like the commercial CLIP-tag™): this orthogonal substrate specificity will allow simultaneously and specifically the labeling of different molecular targets in living cells (Gautier et al. 2008).

Acknowledgments We gratefully thank Castrese Morrone for help in some experiments. This work was supported by: (1) Short

Term Mobility Program of the National Research Council of Italy; (2) FIRB-Futuro in Ricerca RBFR12001G_002-Nematic and Merit RBNE08YFN3-Molecular Oncology; (3) BIO2013-44953-R Project.

References

- Aliye N, Fabbretti A, Lupidi G, Tsekoa T, Spurio R (2015) Engineering color variants of green fluorescent protein (GFP) for thermostability, pH-sensitivity, and improved folding kinetics. *Appl Microbiol Biotechnol* 99:1205–1216
- Ashby MC, Ibaraki K, Henley JM (2004) It's green outside: tracking cell surface proteins with pH-sensitive GFP. *Trends Neurosci* 27:257–261
- Campbell TN, Choy FYM (2000) The effect of pH on green fluorescent protein: a brief review. *Mol Biol Today* 2:1–4
- Cava F, Laptenko O, Borukhov S, Chahlafl Z, Blas-Galindo E, Gómez-Puertas P, Berenguer J (2007) Control of the respiratory metabolism of *Thermus thermophilus* by the nitrate respiration conjugative element NCE. *Mol Microbiol* 64:630–646
- Cava F, de Pedro MA, Blas-Galindo E, Waldo GS, Westblade LF, Berenguer J (2008) Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environ Microbiol* 10:605–613
- Cava F, Hidalgo A, Berenguer J (2009) *Thermus thermophilus* as biological model. *Extremophiles* 13:213–231
- Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. *Science* 5148:802–805
- 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
- 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. *Nat Struct Mol Biol* 11:714–720
- Fang Q, Kanugula S, Pegg AE (2005) Function of domains of human O⁶-alkyl-guanine-DNA alkyltransferase. *Biochemistry* 44:15396–15405
- Gautier A, Juillerat A, Heinis C, Corrêa IR Jr, Kindermann M, Beaufile F, Johnsson K (2008) An engineered protein-tag for multi-protein labeling in living cells. *Chem Biol* 15:128–136
- Gronemeyer T, Chidley C, Juillerat A, Heinis C, Johnsson K (2006) Directed evolution of O⁶-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Prot Eng Des Sel* 19:309–316
- Hinner MJ, Johnsson K (2010) How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol* 21:766–776
- Kanugula S, Goodtzova K, Edara S, Pegg AE (1995) Alteration of arginine-128 to alanine abolishes the ability of human O⁶-alkyl-guanine-DNA alkyltransferase to repair methylated DNA but has no effect on its reaction with O⁶-benzylguanine. *Biochemistry* 34:7113–7119
- Kanugula S, Goodtzova K, Pegg AE (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* 329:545–550
- Keppeler A, Gendreizig S, Gronemeyer T, Pick H, Vogel Johnsson K (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* 21:86–89
- Keppeler A, Pick H, Arrivoli C, Vogel H, Johnsson K (2004) Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci USA* 10:9955–9959
- Liu L, Watanabe K, Fang Q, Williams KM, Guengerich FP, Pegg AE (2007) Effect of alterations of key active site residues in O⁶-alkylguanine-DNA alkyltransferase on its ability to modulate

- the genotoxicity of 1,2-dibromoethane. *Chem Res Toxicol* 20:155–163
- Miggiano R, Casazza V, Garavaglia S, Ciaramella M, Perugino G, Rizzi M, Rossi F (2013) Biochemical and structural studies of the *Mycobacterium tuberculosis* O^6 -methylguanine methyltransferase and mutated variants. *J Bacteriol* 195:2728–2736
- Mollwitz B, Brunk E, Schmitt S, Pojer F, Bannwarth M, Schiltz M, Rothlisberger U, Johnsson K (2012) Directed evolution of the suicide protein O^6 -alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* 51:986–994
- Moracci M, Capalbo L, Ciaramella M, Rossi M (1996) Identification of two glutamic acid residues essential for catalysis in the β -glycosidase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Protein Eng* 12:1191–1195
- Moracci M, Trincone A, Perugino G, Ciaramella M, Rossi M (1998) Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* 37:17262–17270
- Morita R, Nakagawa N, Kuramitsu S, Masui R (2008) An O^6 -methylguanine-DNA methyltransferase-like protein from *Thermus thermophilus* interacts with a nucleotide excision repair protein. *J Biochem* 144:267–277
- Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2:2212–2221
- Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotech* 24:79–88
- Pegg AE (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
- Perugino G, Vettone V, Illiano G, Valenti A, Ferrara MC, Rossi M, Ciaramella M (2012) Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J Biol Chem* 287:4222–4231
- Perugino G, Miggiano R, Serpe M, Vettone A, Valenti A, Lahiri S, Rossi F, Rossi M, Rizzi M, Ciaramella M (2015) Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucl Ac Res*. doi:10.1093/nar/gkv774
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York, pp 1.123–1.125
- Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, Wang Y, Patel DJ, Berenguer J, Brouns SJJ, van der Oost J (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507:258–261
- Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* 67:509–544
- Tubbs JL, Pegg AE, Tainer JA (2007) DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O^6 -alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair* 6:1100–1115
- Wilkinson OJ, Latypov V, Tubbs JL, Millington CL, Morita R, Blackburn H, Marriott A, McGown G, Thorncroft M, Watson AJ, Connolly BA, Grasby JA, Masui R, Hunter CA, Tainer JA, Margison GP, Williams DM (2012) Alkyltransferase-like protein (At1) distinguishes alkylated guanines for DNA repair using cation- π interactions. *Proc Natl Acad Sci* 109:18755–18760
- Yang CG, Garcia K, He C (2009) Damage detection and base flipping in direct DNA alkylation repair. *ChemBioChem* 10:417–423
- Zang H, Fang Q, Pegg AE, Guengerich FP (2005) Kinetic analysis of steps in the repair of damaged DNA by human O^6 -alkylguanine-DNA alkyltransferase. *J Biol Chem* 280:30873–30881

The role of O⁶-alkylguanine-DNA alkyltransferase 2 in DNA repair and in the meiotic development of Caenorhabditis elegans.

Serpe M.¹ et al.

¹Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80125 Naples, Italy.

In this study, the in vivo analysis of the AGT of the hermaphrodite nematode Caenorhabditis elegans is reported. This worm encodes two genes, CeAGT-1 and CeAGT-2: the first seems to be similar to the classical AGT at the primary structure level; the second is an AGT, harbouring only the catalytic C-terminal domain, fused to a histone-like domain. Due to the peculiar primary structure of CeAGT-2, I tried to analyse its role in vivo. C. elegans is a suitable model organism to investigate and dissect each component of the various DNA repair pathways in response to DNA damage, but also to study the role of DNA repair in different developmental stages. In particular, to understand the possible role of the protein, I have characterized the phenotypes of CeAGT-2 mutant strain by subjecting it to various experimental procedures e.g. testing the susceptibility of the CeAGT-2 strain to the alkylating agent treatment. Furthermore, C. elegans offers the possibility to observe multiple phenotypes related to the embryonic/larval development of the nematode, by analysing living features. In this latter case, the embryonic lethality and the aberrant phenotypes observed are associated with mutations in genes involved in meiosis and in the formation and/or production of gametes, so I analysed the role of CeAGT-2 during gametogenesis. Overall results show a key role of CeAGT-2 in metabolic pathways related to DNA repair and meiotic development.

Introduction

Caenorhabditis elegans as a model organism

Integrity of genetic information is essential to both individual health and reproductive capacity. The *Caenorhabditis elegans* model system provides an opportunity not only to investigate the components of the various repair pathways and the pathway interactions, but also to study the role of repair in different developmental stages. As an animal system provides a useful model to more fully understand the integration of the types of repair that occur in response to DNA damage. Sydney Brenner in 1963 introduces for the first time *Caenorhabditis elegans* as model system. It is a member of the Nematode phylum, belonging to the Rhabditidae family. It is a small animal, about 1 mm long when adult. Its normal habitat is in soil, surviving on bacteria and fungi. *C. elegans* exists either as hermaphrodite or as male, characterized by clear different structures when adult. The number of sexual chromosomes determines the worm's sex. Six chromosomes are present: five pairs of autosomes (chromosomes I, II, III, IV, and V) and one pair of sex chromosomes X. Hermaphrodites have two X chromosomes (designated XX), while males have one X chromosome (designated XO), the hemizygous state [Brenner, 1974]. Spontaneous males appear in the wild type population with a very low frequency (0.2%), due to the loss of one sexual chromosome for non-disjunction. Males can cross-fertilize hermaphrodites and give a mixed progeny of 50% males and 50% hermaphrodites. The phenotype in which the frequency of males is higher than the wild type is named Him phenotype (high incidence of males) indicating segregation defects of sexual chromosomes [Hodgkin *et al.*, 1979].

C. elegans genome has been the first multicellular eukaryote genome to be sequenced in 1998 (*The C. elegans Sequencing Consortium*). Its genome is 97 Mb and codify for more than 19.000 proteins, sharing an high degree of gene homology with other superior eukaryotes. About 60% of genes have an orthologue in mouse and human. Many physiological pathways and damage/stress response pathways are conserved [Lai *et al.*, 2000]. All this features makes *C. elegans* a competitive model system to study biological processes in a multicellular organism compared to longer cell cycle organisms with a lower progeny. In the next figure is reported the life cycle of *C. elegans*, in which embryonic and larval development is schematized.

Germline in C. elegans

Over the past twenty years, *C. elegans* has emerged as a major model organism for investigating meiotic mechanisms. Several features of *C. elegans* biology have contributed to facilitate this kind of studies. The worm germ line is especially amenable to high-resolution cytological analysis of chromosome and nuclear organization in the context of whole mount preparations that preserve 3D nuclear architecture. Importantly, each germ line contains a complete time course of meiosis, with nuclei organized in a temporal/spatial gradient corresponding to the stages of meiotic prophase [Crittenden *et al.*, 1994]. In the distal region, named mitotic zone, nuclei undergoes the mitotic S phase [Hirsh *et al.*, 1976]. From the mitotic zone to the proximal region it is possible to observe the different prophase stages I. In transition zone chromosomes acquires crescent-shaped form and

nuclei undergoing Leptotene and Zygotene. At this stage chromosomes pair and recombination starts with cuts due to SPO-11 topoisomerase [Dernburg, 1998]. The number of Spo11-induced DSBs in meiosis is regulated: there are twelve cuts per nucleus [Mets *et al.*, 2009]. The crescent-shaped form of the nuclei is needed to approach the chromosomes, in *C. elegans* only one chromosomal end bound the nuclear membrane [Goldstein, 1982]. The chromosomal pairing starts at the “Pairing Center” (PC), a region mapped at the chromosomal ends. The PC have many repeated sequences recognized from Zinc finger proteins, ZIM-1, ZIM-2, ZIM-3 e HIM-8 [Villeneuve, 1994]. The synapsis starts from the PC to the other chromosomal end [MacQueen *et al.*, 2005]. The synaptonemal complex (SC) formation is necessary for recombination using the homologous chromosome as a repair template [Colaiacovo *et al.*, 2003; MacQueen *et al.*, 2002].

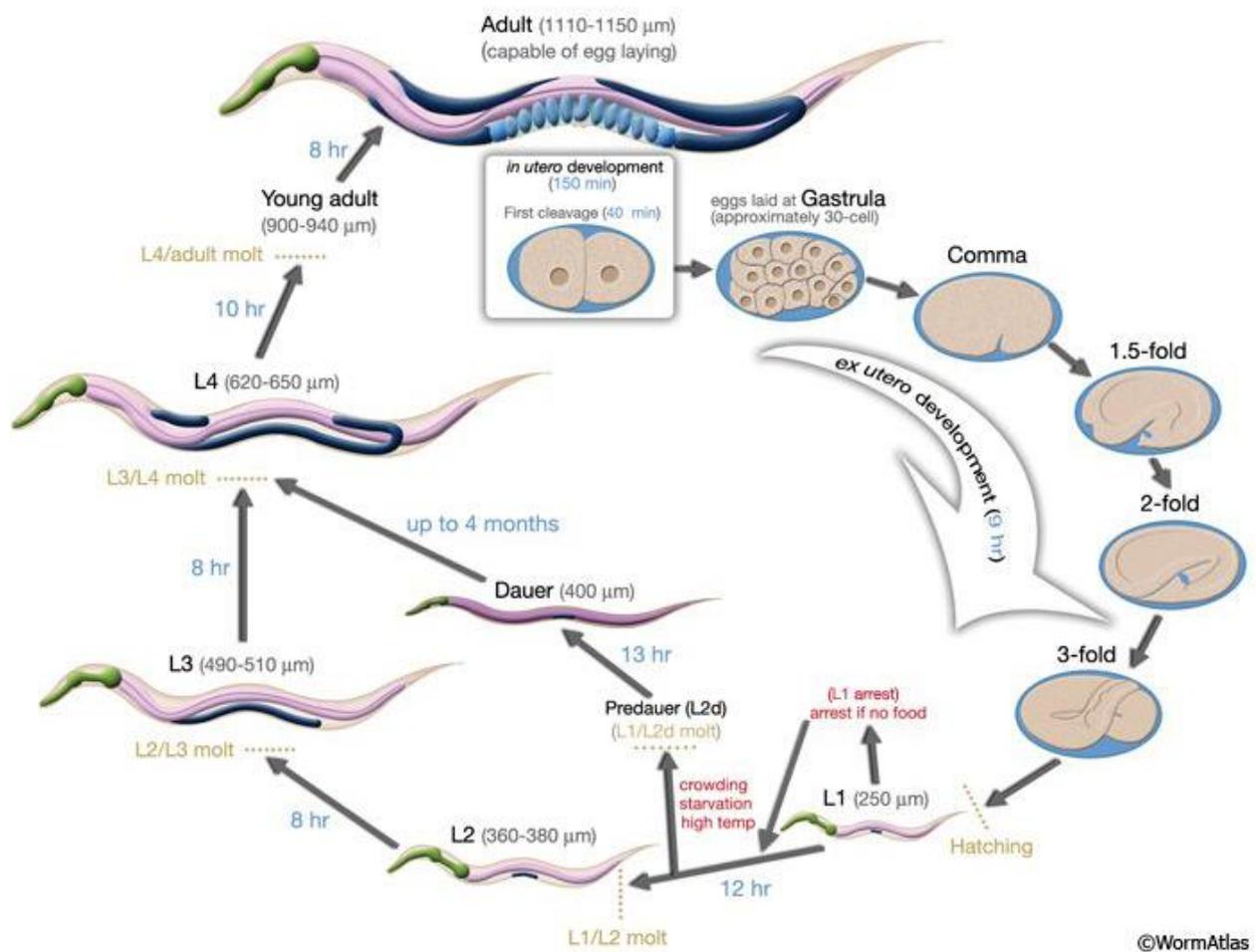


Figure 1: *C. elegans* life cycle at 20° C (from [Altun and Hall 2009; WormAtlas])

In *C. elegans*, unlike yeast, plants and mammals, the SC can be assembled without DSBs formation [Dernburg, 1998]. The key factor in the meiotic step of strand invasion, Dmc1, is missing in *C. elegans*, and RAD-51 is the only RecA-like enzyme present at this step. In WT *C. elegans*, RAD-51 foci representing nascent meiotic DSBs arise during zygotene and early/middle pachytene stages, and then decrease in number during the late pachytene stage as meiotic DSB repair progresses [Colaiacovo *et al.*, 2003]. Increase and/or persistence of such foci are indicative of DNA repair

defects. At early pachytene, after SC formation chromosomes are distributed at nuclear periphery. Most DSBs are repaired in early pachytene. The meiotic machinery ensures that each chromosome receives at least and just one crossing-over (CO). The sites of the maturing crossover become visible in diakinesis [Colaiacovo *et al.*, 2003]. By the end of pachytene, about half of the oocytes in the *C. elegans* germ line undergo physiological cell death [Gumienny *et al.*, 1999]. Among the end of pachytene and diplotene stages the SC disassemble and chromosomes assuming a cruciform structure [MacQueen *et al.*, 2002]. The SC disassembly in a non-symmetric way involving pro-crossover protein, ZHP-3 [Nabeshima *et al.*, 2005]. The chiasmata is the physical display of CO between homologues at the end of recombination and produces the right tension to allowed correct chromosomal segregation to the pole. During diakinesis stage, meiocytes leave the syncytium and cellularize. At diakinesis stage is possible to observe six pairs of bivalent held together by chiasmata [Villeneuve, 1994]. Defects in homologous pairing, SC assembly or CO formation leads to twelve separated univalents and an improper segregation in meiosis. In the proximal part of the gonad, the oocytes arrest until the end of their maturation, that is a fundamental step for their fertilization [Gumienny *et al.*, 1999]. In the next figure is reported a cartoon representation of organization of female germline in *C. elegans*.

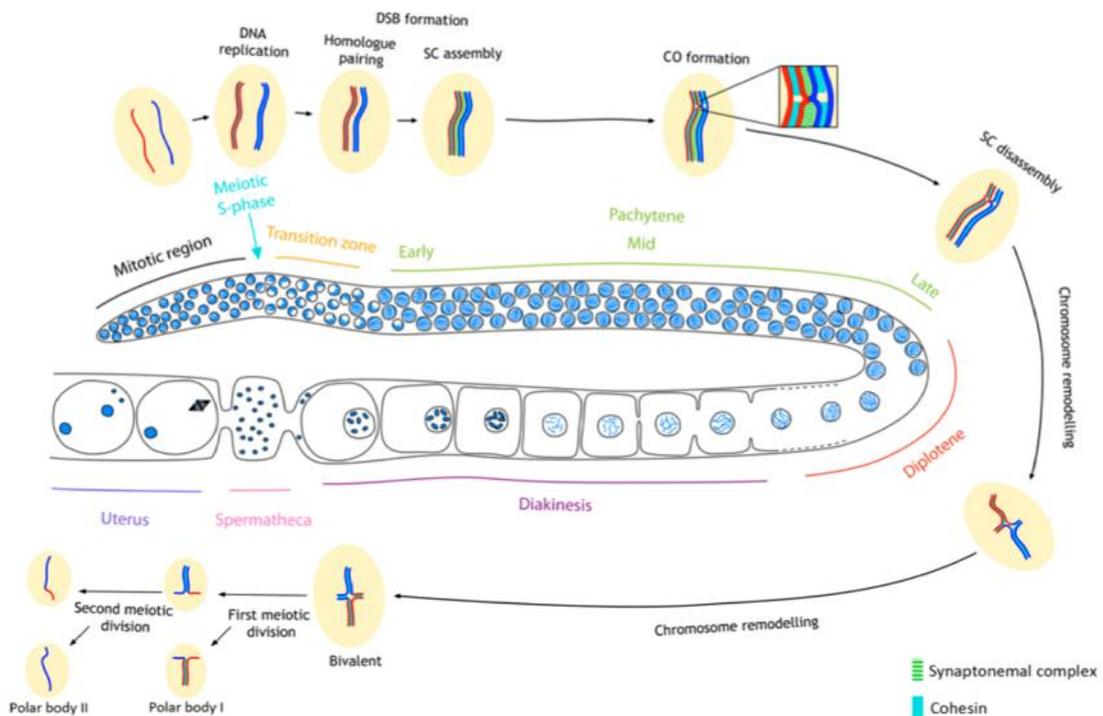


Figure 2: Organization of the female germline in *C. elegans*

Apoptotic Pathway in C. elegans

During the embryonic development, in the hermaphrodite worms it is possible to notice that always the same 131 cells, out of the 1090 generated, are doomed to die by apoptosis [Sulston *et al.*, 1983]. This process has been deeply characterized in *C. elegans*, leading to the identification of the main genes, which drive apoptosis pathway: *ced-3* (*C. elegans* abnormal death 3), *ced-4* (orthologous to mammalian Apaf1), *ced-9* (orthologous to mammalian Bcl2), *egl-1* (egg laying defective 1). The *ced-3*, *ced-4* and *egl-1* genes are pro-apoptotic factors, and their knockout leads to the survival of the 131 somatic cells that normally die [Ellis *et al.*, 1986]. The *ced-9* gene has instead an anti-apoptotic function, so a gain of function mutation causes a block of apoptosis, while a loss of function mutation determines a precocious death during early development, due to the mis-regulation of the apoptotic pathway [Hengartner *et al.*, 1992]. In order to elucidate the interaction existing among the “deathgenes”, epistasis studies have been performed, highlighting the mechanism at the base of the apoptotic mechanism execution. When death signals trigger the activation of the apoptotic program, the caspases belonging protein CED-3 is switched from an inactive state (procaspase) to an active state (caspase) by the CED-4 protein. The sequestration of CED-4 by CED-9 maintains the state of CED-3 inactive. Under normal conditions, when no death signals are present, CED-4 cannot activate CED-3, since it is associated in an inactive state with CED-9, but when apoptosis is triggered, another factor plays its role, EGL-1. EGL-1 binds CED-9, which this way is unable to block CED-4, and therefore CED-4 can associate and activate CED-3, achieving the apoptotic program. The identification of these key genes regulating cell death program in *C. elegans* that are conserved in humans has represented a significant advance in the knowledge about apoptosis. At critical times when cells progress from one stage to the next, the quality of the DNA is assessed by cell cycle checkpoints. If the chromosomes don't pass quality control, the checkpoint may generate a ‘Wait’ signal. In the germ-line, damage can lead to apoptosis (non programmed cell death), cell cycle arrest, or heritable mutation. The *C. elegans* ortholog of the transcription factor P53, CEP-1, which is a key regulator of the DNA damage checkpoint, is required for DNA damage-induced apoptosis in the *C. elegans* germ-line, but not for programmed cell death occurring during worm development nor physiological (radiation-independent) germ cell death [Gartner *et al.*, 2008].

The AGT protein in Caenorhabditis elegans

Alkylation at the O^6 -position of guanine and the O^4 -position of thymine in DNA makes these bases highly mutagenic, causing GC to AT and AT to GC transitions [Singer *et al.*, 1976]. Many organisms have developed an effective mechanism to repair these mutagenic lesions through the action of the ubiquitous AGTs. All these proteins contain a -PCHRV- consensus sequence in the active site and repair the alkylation damage to DNA by transferring the alkyl group from the O^6 -position of guanine to the active site cysteine residue in this sequence, thus restoring the DNA to its

unmodified form in a single step. The AGT contain a helix-turn-helix DNA binding domain and that the target substrate base is flipped out of the DNA helix into a binding pocket that contains the cysteine acceptor residue [Pegg *et al.*, 2011].

The process of identifying specific genes of a particular function in a species has changed drastically with rapidly emerging complete analysis of genomes from many different organisms. The presence of an AGT gene has been confirmed in almost all the organisms with the -PCHRV- motif providing a key recognition point. The search of the *C. elegans genome database* [*C. elegans Sequencing Consortium*, 1999] for any AGT protein identified a clone Y62E10A.5 containing the consensus sequence and many other features common to AGT, and is termed *CeAGT-1* [Kanugula *et al.*, 2001]. A detailed examination of the *C. elegans genome database* also revealed the presence of a hypothetical protein, clone F09E5.13, which has some similarities to AGT. This sequence codes for a 274 amino acid protein. In particular, it has some similarities to known AGT in the region from residues 62 to 96 (corresponding to residues 137-172 in hAGT), where a cysteine acceptor site within the -PCHRV- motif was identified: it is therefore termed *CeAGT-2* [Kanugula *et al.*, 2001]. Although in several bacteria, including *E. coli*, are known to express two AGT proteins, where one of them is highly inducible in response to exposure to alkylating agents [Samson, 1992; Sekiguchi *et al.*, 1996], this is the first time that the presence of two AGTs in a eukaryote were identified. The cDNA of *CeAGT-2* was present in a mixed-stage expression library and the mRNA corresponding to this gene product was detected at all stages, especially at the egg and adult stages of the organism in a comprehensive analysis of gene expression [Kanugula *et al.*, 2001].

The *CeAGT-2* shows a number of highly novel features: *i*) the primary sequence displayed that, among known AGTs, the presumed active site cysteine is located much closer to the amino-terminus of the protein. The available 3D structures of AGT show that the protein is made up of a two-domain structure, where all residues that have been identified as part of the active site and DNA binding region are in the C-terminal domain. The N-terminal domains have a roughly similar structure but virtually no amino acid homology between species [Daniels *et al.*, 2000]. This N-terminal domain is totally absent in *CeAGT-2*; *ii*) *CeAGT-2* has a much longer C-terminal sequence, no similar to any known AGT, rather this region shows about 25% sequence homology with human histone 1C. This region is very rich in proline and basic amino acids and there are several sequences with similarity to known nuclear location signals (figure 3). These may be needed to allow nuclear uptake of the 274 amino acid protein, which is much larger than other eukaryotic AGT [Kanugula *et al.*, 2001]; *iii*) although *CeAGT-2* has several of the highly conserved residues in the active-site (including Pro-144, Cys-145, His-146, Val-148, an aromatic residue at position 158, Leu-168, Leu-169, and Glu-172), several other conserved residues are not present: residues Asn-137 and Arg-147, which are totally conserved in all other known AGTs, are replaced by Gln and Pro, respectively; *iv*) the amino acid sequence of the presumed DNA binding domain of *CeAGT-2* is strikingly different from known AGTs. Few of the residues in the helix-turn-helix motif made up of residues 113 to 136 in hAGT that form this domain [Daniels *et al.*, 2000; Wibley *et al.*, 2000] are identical to those in *CeAGT-2*. Moreover, the fully conserved residues Tyr-114 and Arg-128, which are clearly involved in AGT reaction, are replaced by the conservative substitutions Phe and Lys, respectively [Kanugula *et al.*, 2001].

Previously, the *CeAGT-2* gene was cloned and expressed as truncated form lacking the Histone-like domain, termed *CeAGT-2tr*, and show that it does possess DNA repair AGT activity despite not canonical structural features [Kanugula *et al.*, 2001]. The *CeAGT-2tr* is able to remove methyl groups from both O^6 -methylguanine and O^4 -methylthymine. The good activity toward the later

substrate indicates that it is more nearly similar in its repair capabilities to the *E. coli* Ogt (which repairs O^4 -methylthymine efficiently) than to mammalian AGTs, which have a very poor ability to repair this minor methylation product [Samson *et al.*, 1997].

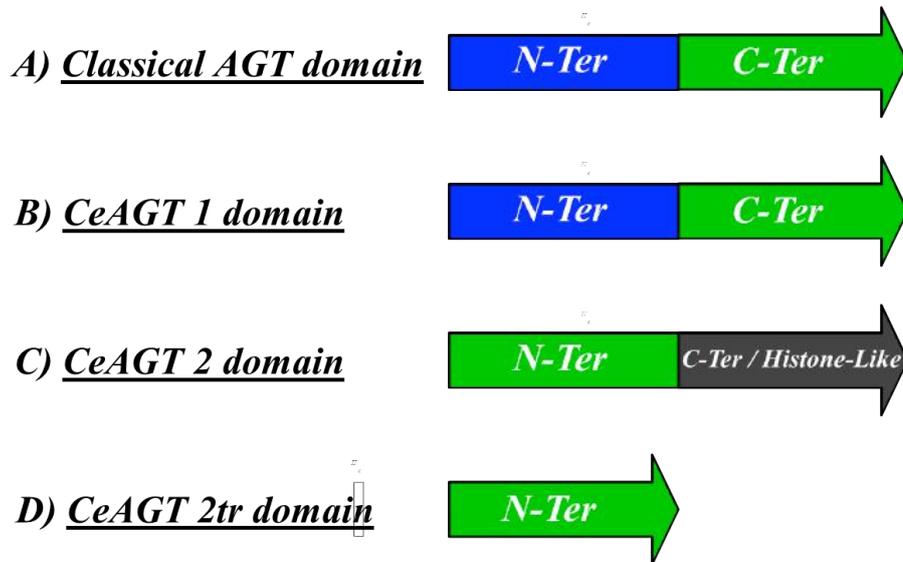


Figure 3: Schematic representation of the AGT domains: A) Scheme of the peculiar primary structure of the N and C-terminal domains of AGTs; B) Scheme of the primary structure of the N and C-terminal domains of *CeAGT-1*; C) Scheme of the primary structure of the N and C-terminal domains of *CeAGT-2*; D) Schematic portion of the N-terminal domain characterized [Kanugula *et al.*, 2001]. Typical N-terminal domain, C-terminal domain and histone-like domain are represented in blue, green and gray rectangle, respectively.

Purpose

C. elegans encodes two distinct orthologs of AGTs, known as *CeAGT-1* and *CeAGT-2*. Whereas no information is available about the former, the *CeAGT-2* protein was shown to have DNA alkyltransferase activity *in vitro* and in bacterial cells [Kanugula *et al.*, 2001]. Interestingly, *CeAGT-2* has a peculiar structure, if compared with other AGT: in this protein the domain containing the DNA binding region and active site is fused to a domain resembling histone 1C, the function of this domain has not been established (figure 3). For this peculiar primary structure, it has drawn our attention for its possible role *in vivo*.

The presence of two AGT may indicate a genomic duplication phenomenon resulting in functional divergence of the two genes. So the non-canonical structure of the second AGT could indicate the acquisition of additional functions. In *C. elegans*, high-resolution microscopy combined with genetic tools provides a unique opportunity for the detailed analysis of DNA repair pathways on an organismal level and in true tissues. In particular, the *C. elegans* gonad is an extremely powerful toolkit for the study of these processes, because the spatiotemporal organization of germ cells allows following their progress in specific stages of gametogenesis, from mitotic divisions up to the end of meiosis. In *C. elegans* no *ceagt-1* mutant and *ceagt-1/ceagt-2* double mutant be found in *C. elegans* banks, whereas a null *ceagt-2* mutant was available (tm 6462), although no information has been reported on it.

Results

The sensitivity of the CeAGT-2 mutant to alkylating agents

Because of the absence of one of the two AGTs in the *CeAGT-2* null mutant, this was first analysed in terms of drug sensitivity to alkylating agents: young adult worms were treated with the appropriate drug, in this case 300 μ M final concentration of methyl-methane-sulfonate (MMS) in for 16 h in liquid culture. Then the worms were isolated and the laid eggs were scored for 42h. For all multiple independent replicas and for data analysis and significance, the Student Test was applied. The progeny was also scored for embryonic lethality (unhatched eggs). In every experiment, untreated controls were screened at the same time points. Preliminary results indicated that the *CeAGT-2* mutant strain was more sensitive to alkylating agents than the WT strain, as was appreciable from the chart in figure 4. It suggested that *CeAGT-2* is involved in repairing alkylation damage, indicating a enancher interaction with *CeAGT-1*.

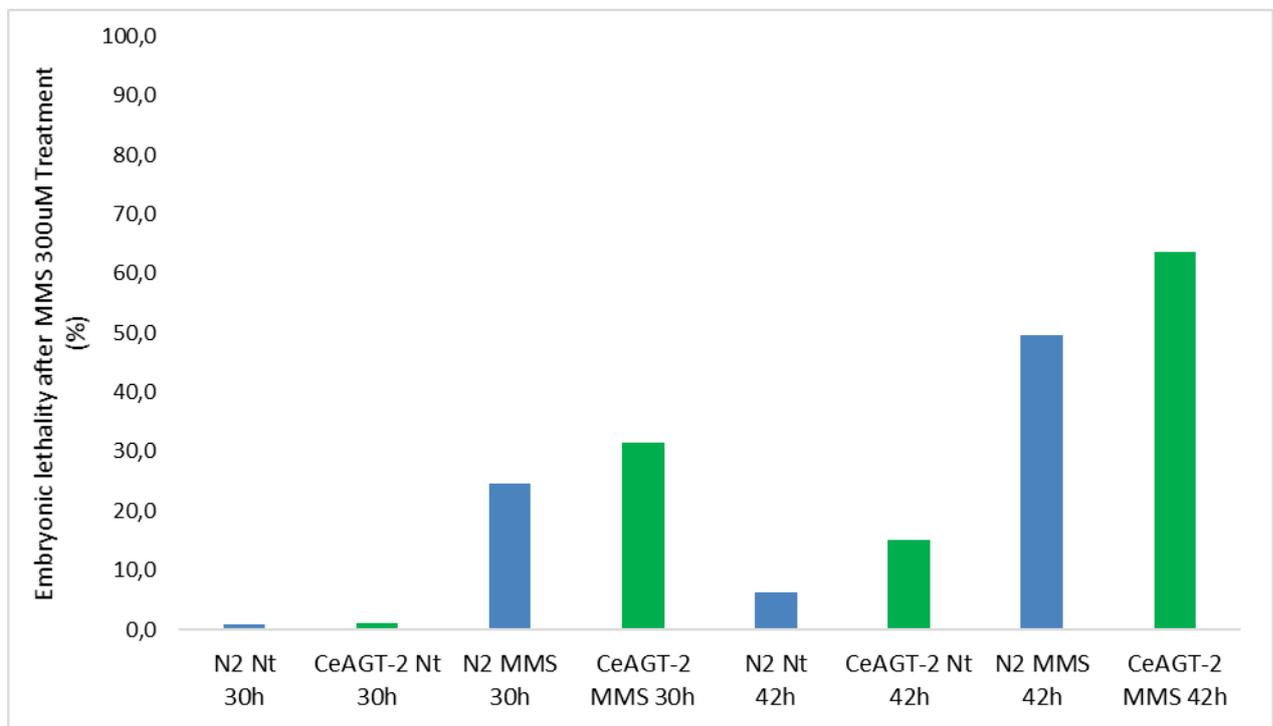


Figure 4: Embryonic survival at different time points after treatment with 300 μ M MMS (methyl methane sulfonate) for the indicated genotype.

Defect in CeAGT-2 lineage

The second step of this analysis was to investigate about the type of defects observed in the *CeAGT-2* homozygous mutant population (Figure 5). Phenotypic screenings consist of young adult worms individually cloned on freshly untreated seeded plates. Each worm was transferred to a fresh plate every 24 h, and the eggs laid after 48 and 72 h will be scored. The progeny was scored for embryonic lethality (unhatched eggs), percentage of males, developmental defects and larval arrests after 24, 48 and 72 h. We observed a significantly higher frequency of embryonic lethality, developmental defects and larval arrests in *CeAGT-2* mutant compared to wild type control. The developmental defects and larval arrest observed in mutant *CeAGT-2* especially fall in the alterations of cell lineage. These phenotypes are caused by alteration of post-embryonic lineages, suggesting that meiotic or somatic mutations in genes that regulate these lineages cause the formation of aberrant phenotypes.

	N2	CeAGT-2
Parental	20	28
Laid Eggs	5341	5769
Dead Embryos	27	70
Hatched Eggs	5314	5699
Males	3	2
Brood Size	267	206
% Dead Embryos	0,5%	1,2%
% Males	0,06%	0,04%
Larval Arrest	0	54
% Larval Arrest	0%	0,9%
Develop. Defects	2	48
% Develop. Defects	0,04%	0,8%

Figure 5: Main phenotypic characteristics of *CeAGT-2* strain compared with N2 strain

The embryonic lethality in the *CeAGT-2* strain may be caused by mutations in meiotic genes that regulate cell lineage or alterations in cell time division and fate during embryonic development. Both these results indicated that the lacking of this AGT protein hampers the correct germline

progression. This result is the starting point for a question: can *CeAGT-2* be involved in the physiological repair processes needed to properly perform meiosis and hence for a proper maturation of gametocytes? So, we had to focus our attention on the body responsible for these processes: the gonad of *C. elegans* contains germ cells, whose physical location corresponds to a particular stage in meiotic progression. This facilitates the study of the effect of gene depletions at various stages of meiotic prophase.

Immunolocalization of RAD-51 in the gonad

The *C. elegans* germline also offers an excellent system to study the kinetics of formation and repair of double strand breaks (DSB) that occur in all cells after certain types of DNA damage, but they are physiologically generated by the topoisomerase-like protein SPO-11 during meiosis and are essential for the correct execution of meiosis [Colaiacovo M. *et al.* 2003; MacQueen *et al.* 2002]. Meiotic DSB are preferentially repaired by RAD-51 recombinase, through homologous recombination, by using a parental homolog to form inter-homolog crossovers that are essential for accurate chromosome segregation at the first meiotic division [Colaiacovo *et al.*, 2003]. In WT *C. elegans*, *foci* formed at meiotic DSB arise during the early and middle pachytene stages, and then rapidly decrease in number during late pachytene stage, as meiotic DSB repair progresses [Colaiacovo *et al.*, 2003]. So, by following RAD-51 recombinase *foci* by immunolocalization, it is possible to monitor the proper continuation of meiotic processes in the *CeAGT-2* gonad, establishing if the protein is involved and maturation of gametocytes.

Rad-51 *foci* had a different localization pattern in *CeAGT-2*, when compared with the WT: in the latter, the Rad-51 *foci* number in the nuclei increase in early pachytene stage but decrease in middle pachytene; finally they disappear in late pachytene stage. In the *CeAGT-2* mutant we noted, however, not only an increase in the number of *foci* per nucleus, but also a major persistence. This is highlighted by a high number of *foci* still present in late pachytene stage: in the figure 6, photos of immunolocalizations performed on the WT and mutant and the relative graph obtained by quantification of the number of *foci* per nucleus are shown. Once again, we got indications of the possible involvement of *CeAGT-2* in the physiological repair processes needed to perform a correct meiosis.

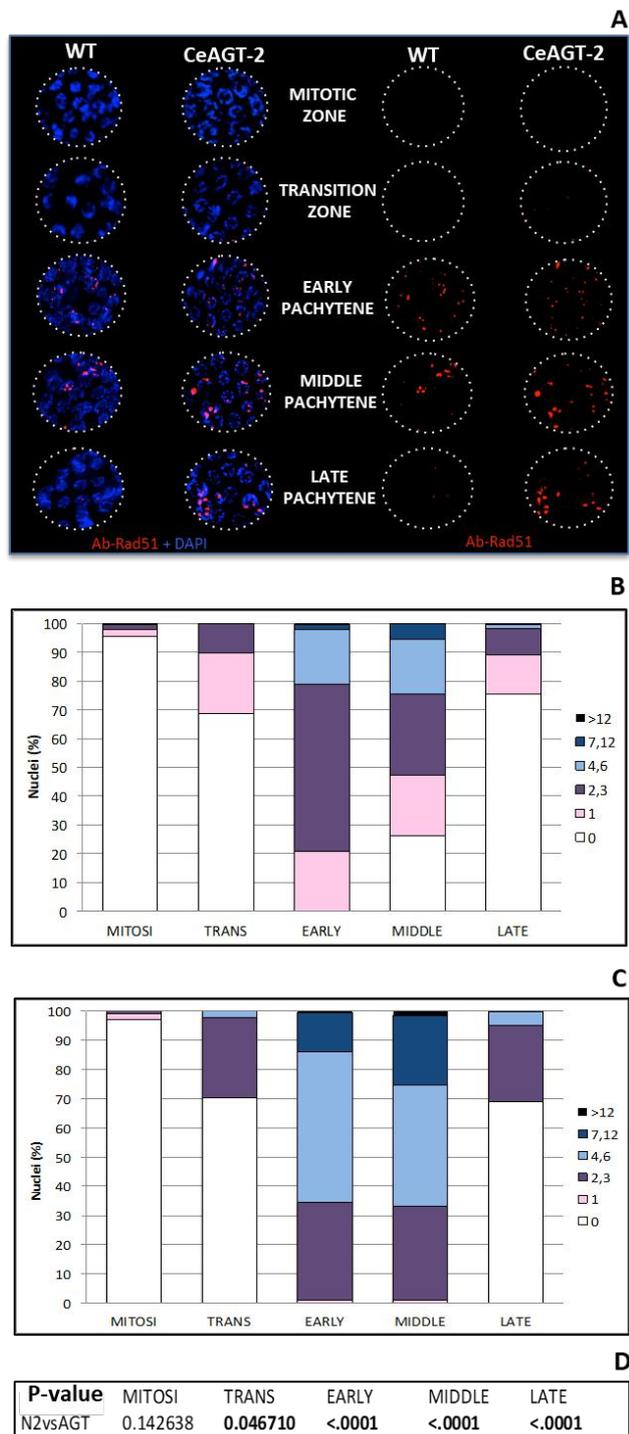


Figure 6: A) Representative images of gonade zones nuclei (indicated by circles) stained with anti-RAD-51 antibodies (red) and DAPI (blue). B) Quantification of RAD-51 *foci* in germlines of the WT. C) Quantification of RAD-51 *foci* in germlines of the *CeAGT-2*. D) Statistical analyses.

Number of apoptotic nuclei in the gonad

Finally, we evaluated the apoptosis levels in the *CeAGT-2* mutant strain. It has been widely demonstrated that in *C. elegans* under physiological conditions, a physiological number of germ cells undergo the apoptotic process, generally an average of three for gonad [Gartner *et al.*, 2008]. We measured an increase in the germline apoptotic level, probably caused by the persistence of DNA lesions in the *CeAGT-2* mutant germline, which triggers the DNA damage checkpoint and inducing apoptosis. As shown in the figure 7, the number of apoptotic bodies in the *CeAGT-2* strain was significantly more than the WT. So, *CeAGT-2* mutant worms show elevated basal levels of apoptosis of late pachytene nuclei, suggesting a possible defect in meiotic processes or defect in the selection of cells destined for apoptosis.

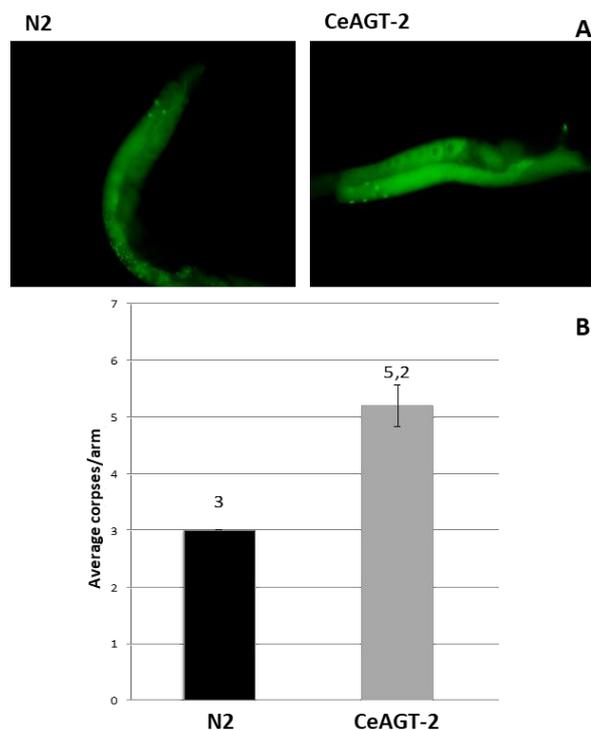


Figure 7: A) Representative image of a WT and *CeAGT-2* mutant apoptotic corps in the germline stained with viable dye Syto-12. B) Graph of the average of apoptotic corpses for gonad arm in the indicated strains.

Materials and methods

Solutions

Nematode Growth Medium (NGM) plates: NaCl 0.3%, Peptone 0.25%, and Agar 2%. After sterilization add: cholesterol (5mg/ml), CaCl₂ (1mM), MgSO₄ (1mM), and NaKPO₄ (25mM pH6).

Lysogeny Broth (LB) solution: NaCl 1%, Yeast extract 0.5%, Bacto tryptone 1%.

M9 1X solution: 0.3% H₂PO₄, 0.6% Na₂HPO₄, 0.5% NaCl and 1mM MgSO₄.

Bleaching solution: NaOH 1.6M, NaClO 0.6%

TAE solution: 40mM Tris HCl pH8, 20mM Acetic Acid and 1mM EDTA pH 8.

Lysis buffer: 20000U/ml proteinase K in 10 mM Tris pH 8.2, 50 mM KCl, 2.5 mM MgCl₂, 0.45% Tween 20.

Ab buffer: 1% BSA, 0.1% Tween20, 0.05% sodium azide in 1x PBS

Maintaining of the strains

Worms are grown at 20°C on NGM plates containing *Escherichia coli* (*E. coli*) OP50, a uracil-requirement mutant to prevent overgrowth of the bacterial layer. OP50 bacteria are grown in an LB solution at room temperature overnight. The maintaining of the worms was carried out as described by Sulston and Hodgkin in 1988.

Worms synchronization with Alkaline Hypochlorite solution (bleaching)

Worms are sensitive to bleach while the egg shell protects embryos from it. After treatment with alkaline hypochlorite solution, only embryos are able to hatch and they will grow synchronous.

1. Grow worms until adult stage.
2. Recover gravid adults in 15 ml tubes by washing plates with NaCl 0.1M.
3. Pellet worms by centrifuging for 2' at 1500 rpm at 4C° and discard supernatant.
4. Perform 3 washes with NaCl 0.1M until the buffer appears clear of bacteria.
5. Add some volume of bleaching solution and agitate for 10'.
6. Stop reaction by adding NaCl 0.1M to fill the tube.
7. Quickly centrifuge (since treatment may still be active) for 1' at 1500 rpm at 4C° and discard supernatant.

8. Wash pellet 3 times by filling the tube with NaCl 0.1M.
9. Place the eggs to seeded NGM plates with OP50.

Phenotypically Screening

Each strain worm are cloned during the L4 larval state on single Petri plates and keep at 20°C, producing and laying eggs for 4 days.

Transfer every 12 hours the laying worms onto fresh plates until the deposition of non-fertilized oocytes.

Monitor each plate for 24/72 hours to analyze three different parameters: embryonic lethality, the presence of males and aberrant phenotypes among the progeny.

The value of embryonic lethality is calculated as the ratio of unviable eggs to laid eggs. The percentage of males and of aberrant phenotypes is calculated as the ratio of males/aberrant phenotype on the hatched eggs.

Aberrant phenotypes are scored until the progeny reached the adulthood.

Aberrant phenotypes are classified as:

dpy: Dumpy, short and fat worms;

unc: Uncoordinated, worms that have problems in movement;

vul: Vulvaless sterile worms with no vulva;

no gonad: Worms with no gonadal arms;

bag: Bag of worms, worms defective in laying eggs and the embryos hatch inside the mother who soon dies;

sma: Small, these animals tend to be more proportional in shape than dpy animals;

rol: Roller, worms that tragically twist in place about their long axis;

long: Worms that have the body length half longer than WT;

lin: Cell lineage defects

no tail: Tailless worms;

larval arrest: Worms that do not reach adulthood.

Treatment with genotoxic agents (MMS)

Prepare NGM plates with MMS 300 µM.

Store at 4°C at dark for maximum 7 days.

Plates are seeded with fresh OP50 before use.

Treat young adult worms for 24 hours, adding fresh OP50 every 12 hours.

Transfer laying worms onto fresh MMS plates each 12 hours up to 72 hours.

Monitor the plates for 24/72 hours to analyze embryonic lethality.

The value of embryonic lethality is calculated as the ratio of unviable eggs to laid eggs at 24-36 and 36-72 hours after treatment.

Quantitative analysis of DAPI-stained bodies at diakinesis stage

Adult hermaphrodites were picked out for quantitative analysis of DAPI-staining bodies in diakinesis nuclei.

1. Transfer the worms in 15µl M9 1X solution on glass slides. Sacrifice 20 worms per strain to observed nuclei from both gonadal arms.

2. Permeabilize the samples and fix through 15µl of EtOH 100% two times.

3. Add 15µl of the 4', 6'-diamidino-2-phenylindole hydrochloride (DAPI) (2ng/µl) diluted in M9 1X to visualize the DNA in the fixed animals.

Perform quantitative analysis on z series of images acquired using a Leica DM6 fluorescence microscope, Leica DFC365 FX Digital camera under the control of Leica LAS X software. Optical sections were collected at 0.50µm increments.

Apoptosis assay

Staining is performed with the syto-12, a vital dye, that directly permits the recognizing of cells that undergo apoptosis [Gumienny et al., 1999].

1. Suspend adult worms in M9 1x solution.

2. Incubating with 33 µM of syto-12 for 1 hours and half at 20 °C in the dark.

3. Transfer the worms to seeded plates to allow stained bacteria to be purged from the gut for 30' in the dark.

4. Mount the worms on 2% agarose pads in levamisole 2mM.

The quantitative analysis was performed on z series of images acquired using a Leica DM6 fluorescence microscope, Hamamatsu camera under the control of Leica LAS AF 6000 software. The estimation of apoptotic levels was calculated as the average number of apoptotic nuclei per arm screened for each genotype.

Immuno-staining of meiotic nuclei

1. Dissect adult hermaphrodites in 15 μ l M9 1X solution on poly-lysine coated glass slides. Add silicized coverslips and put the slides on liquid nitrogen for 30". Lift the silicized coverslips from the glass slide.
2. Permeabilize the sample and fix through 3 steps at -20°C in methanol, methanol/acetone (1:1), and acetone respectively 5'.
3. Wash 3 times for 5' in 1x PBS.
4. Block fixation with 50 μ l of 3% BSA (BSA 10mg/ml) in PBS 1X under a parafilm coverslip for 30' at 37°C in a humid chamber.
5. Rinse off the coverslip in PBS 1x.
6. Dilute primary antibody in Ab buffer: abRAD-51 in rabbit 1:200.
7. Incubate slides with 60 μ l the primary Ab under a parafilm coverslip for 90' at room temperature in a humid chamber.
8. Rinse off coverslips in PBS 1x.
9. Wash 3 times for 5' in 1x PBS.
10. Dilute secondary antibody in Ab buffer.
11. Incubate with 60 μ l Ab Texas red for 60' at room temperature DARK
12. Wash 3 times for 5' in 1x PBST.
13. Mount with 6 μ l Prolong Gold Antifade reagent with DAPI (life technologies).

Quantitative analysis of RAD-51 foci was performed on z series of images acquired using a Leica DM6 fluorescence microscope, L camera under the Leica DFC365 FX Digital camera control of LASX software. Optical sections were collected at 0.18 μ m increments. The quantitative analyses of RAD-51 were performed by dividing the germ line into 7 zones (mitotic tip, mitotic zone, transition zone, early pachytene, middle pachytene, late pachytene and diplotene), in accordance with their cytological features.

Statistical tools

Statistical analyses of DAPI stained bodies in diakinesis nuclei, apoptosis levels and RAD-51 foci patterns were computed through t-Student test for independent samples using the VassarStats software (<http://faculty.vassar.edu/lowry/VassarStats.html>).

The level comparison of embryonic lethality and percentage of aberrant phenotypes of the different genotypes were computed through chi-square test using the VassarStats software (<http://faculty.vassar.edu/lowry/VassarStats.html>).

Discussion

In this chapter we focused on the *in vivo* study of one of two AGTs of the nematode *C. elegans*. To date, this is the first *in vivo* analysis of *CeAGT-2*, since only a little biochemical characterization is present in literature.

The *C. elegans* offers the opportunity not only to investigate the components of the various repair pathways and pathways interactions in response to DNA damage, but also to study the role of DNA repair at different stages of development. Its genome shares a high degree of homology with other complex eukaryotes: about 60% of the genes have an orthologous in the mouse and human. Many physiological pathways and damage/stress response pathways are conserved [Lai *et al.*, 2000]. In *C. elegans*, genetic screening together with high-resolution microscopy offers an opportunity to also observe numerous progeny and multiple phenotypes of the progeny. All of these features make it a good system to study biological pathway in a multicellular organism.

C. elegans has two genes coding for *CeAGT-1* and *CeAGT-2*, the first seems to be similar to the classical AGTs at the primary structure level. *CeAGT-2* shows three fundamental differences [Kanugula *et al.*, 2001]: *i*) the predicted active site is located much nearer to the N-terminus respect any other known AGTs. The canonical N-terminal domain is totally absent in *CeAGT-2*; *ii*) *CeAGT-2* has a much longer C-terminal sequence for which there is no equivalent in any known AGTs, and this region shows a primary sequence histone-like (figure 3). Previously, the *ceagt-2* gene was expressed as a truncated form of this gene without the histone-like moiety, termed *CeAGT-2tr*: the relative protein still possesses DNA repair AGT activity on both *O*⁶methylguanine and *O*⁴methylthymine [Kanugula *et al.*, 2001].

The peculiar primary structure of *CeAGT-2* capture our attention asking its *in vivo* role. In *C. elegans* the *ceagt-1* single mutant is not available, whereas only a null *ceagt-2* mutant is present in the worm banks, although no information has been reported on this strain. So, we characterized the *CeAGT-2* deleted strain, taking into account the presence of the *CeAGT-1*.

This mutant is more sensitive to the alkylating agent than the WT, showing a higher embryonic lethality correlated to a greater number of unhatched eggs. This confirmed that *CeAGT-2* could plays a key role in the response to DNA damage to alkylating agents as already noted for other different organisms. Surprisingly, exploiting the phenotypic analysis on the *CeAGT-2* strain in physiological condition, I observed other significantly alterations respect to the WT strain, that are often associated with genes involved in meiosis and in the formation and/or production of gametes. So, we analyzed the role of AGT during the maturation of gametocytes, performing the immunolocalization of RAD-51 *foci* in the gonad of *CeAGT-2* mutant, because the RAD-51 recombinase is a suitable marker for following the optimal continuation of meiotic processes. We found an increase of number and increase of persistence of RAD-51 *foci*, suggesting meiotic defects during the formation of germ line in *CeAGT-2*. Furthermore, we found also an increased number of apoptotic bodies in the *CeAGT-2* strain respect to the WT, probably because the persistence of DNA lesions in the *C. elegans* germline activates the DNA damage checkpoint and induces increase in the germline apoptotic level. These results show a possible *in vivo* role of *CeAGT-2* in pathways related to DNA repair and meiotic development. In order to assert with certainty the involvement of the protein in the above described pathways are needed additional experiments, mainly achieving

double and triple mutants of the *Ce*AGT-2 strain with genes involved in the development of the worm.

8. Final Conclusions

SsOGT resulted a convenient model to unravel structure-function relationships of AGTs. Taken as a whole, structural, biochemical and molecular dynamics studies of the enzyme and relative mutants at every stage of the DNA repair activity, suggested a model for the *in vivo* function and degradation of this class of proteins. Although the architecture of the free and DNA-bound AGT, as well as details of its reaction mechanisms, is conserved among the class, the peculiar stability of this archaeal AGT allowed us to obtain 3D structural information in the post reaction form, overcoming the severe limitations of extreme instability of the alkylated forms of the human homolog [Daniels *et al.*, 2000b].

The architecture of the DNA-protein complexes highlighted by the X-ray crystallography and biochemical approaches (new fluorescent assay), provided a potential solution to a few information present in the literature about the DNA-binding mechanism. In particular, we determined the O^6 -mG position dependent-efficiency of DNA repair by *SsOGT*, requiring for its optimal activity at least three bases at both sides of the lesion. These results are in coherence with the observed contacts between the protein and both DNA strands in 3D structure. Furthermore, contacts formed with the DNA strand at the both sides of the lesion are important for *SsOGT*, whereas those formed at the 3' side are more crucial for hAGT. The explanation for this difference could be that *SsOGT* needs to form more and stronger contacts on the DNA to assist stabilization of the DNA-protein complex at higher temperatures, preventing thermal denaturation and facilitating lesion removal [Perugino *et al.*, 2015].

Alkylation-induced instability of hAGTs is a well-known process whose molecular mechanism is, however, poorly understood, mainly due to the failure in the purification of alkylated hAGT and its C145F and C145L mutants [Daniels *et al.*, 2000b]. Again, the stability of *SsOGT* enabled us to purify it in the methylated form, as well as the C119L and C119F mutants at mild temperatures. We acquire direct and quantitative data on the protein stability in correlation with the active site occupancy and, for the first time, we could observe structural modifications occurring upon methylation in solution, whereas small structural rearrangements concerning hAGT upon alkylation were obtained in not optimal physiological conditions [Daniels *et al.*, 2000b]. Our data are also coherent with the observation that alkylated *SsOGT* undergoes degradation after treatment of *S. solfataricus* cells with alkylating agents [Perugino *et al.*, 2012], highlighting that at the physiological growth temperature (80 °C) alkylated *SsOGT* behaves as hAGT, being target the of degradation pathways *in vivo* [Perugino *et al.*, 2015].

By combined structural, biochemical, mutational and molecular dynamics analysis we demonstrated that the two main interactions contribute to maintain the correct folding of *SsOGT*: the D27-R133 pair and the *K48-network*. Before DNA-binding, then post-reaction alkylation induced a perturbation of these interactions: based on our data, we proposed a model for the conformational changes and fate of the *SsOGT* upon alkylation [Morrone *et al.*, 2017]. The optimal activity and stability of *SsOGT* require coordination between the N-terminal and C-terminal domains of the protein. The two clusters of interactions identified play an important role in protein stability and in communication the state of active site. We proposed that these cluster act as “locks”: in the ligand-free protein, both locks are in their “closed state”, ensuring the correct folding of the protein for its optimal stability and optimal activity. The protein binding to DNA and recognition of the damaged guanine leads to opening of the first lock only (the *K48-network*), destabilizing the

link between the N-terminal domain and the connecting loop. We presume that this modification is reversible until the active site remains unmodified: in this way, if the protein dissociates from DNA without any repair activity, the integrity of the *K48-network* is restored and the structural stability of the protein preserved. However, once the DNA repair reaction is completed, irreversible alkylation of the catalytic site induces irreversible conformational changes, switching the second lock (formed by the D27-R133 ion pair) in its “open state”. The loss of coordination between the N-terminal and C-terminal domains starts the *SsOGT* destabilization and its degradation [Perugino *et al.*, 2015].

By superimposition analysis of the available 3D structures, we noted that the region corresponding to the *K48-network* in *SsOGT*, is characterized by the presence of hydrophobic residues that stabilize contacts between the N-terminal and C-terminal domain in hAGT [Daniels *et al.*, 2000b]. Although this hypothesis could not be tested directly, we think that these hydrophobic interactions might play a similar role in hAGT. So, our model might be extended to other AGTs, assuming that interdomain communication and coordination plays a key role in maintaining folding and respond to alkylation triggering destabilization and degradation [Perugino *et al.*, 2015; Morrone *et al.*, 2017].

Regarding the mode of lesion recognition, several authors have analysed the possible functional advantages of performing the direct repair in a co-operative manner. It was underlined that co-operative assembly of protein complexes on the DNA might contribute to the efficiency of lesion search and removal [Melikishvili *et al.*, 2008]. The co-operative binding of hAGT on DNA was directly observed by AFM experiments and demonstrated the protein assembling into discrete clusters on DNA substrates [Tessmer *et al.*, 2014]. These arguments are not in contrast with our results, aimed to describe a model of lesion recognition and repair performed by *MtOGT*. In principle, to guarantee efficient recognition of alkylated bases, both intact and damaged DNA strand should represent a ligand for *MtOGT* [Miggiano *et al.*, 2016]. The insertion of a modified base inside the *MtOGT* active site allows conformational modifications of some regions of the protein, which could act as a signal, that a lesion has been encountered. Additional protein subunits could now be added, also thanks to of their 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 DNA could be scanned, at a fixed space interval, for the presence of other alkylated sites [Miggiano *et al.*, 2016].

The discovery that *SsOGT* is sensitive to BG-derivates, opened the possibility to expand the *SNAP-tag*TM technology, limited to mild reactions and environments. Because of the intrinsic nature of this thermostable protein, it displays a strong stability under several hard conditions, as extremes of pH, ionic strength, temperatures and presence of organic solvents [Perugino *et al.*, 2012]. Mutagenesis on the HTH motif abolished the bonding ability, allowed the construction of a variant of this enzyme (*SsOGT-H*⁵), which conserves most of the advantages of the native protein in terms of stability, but showing major activity at lower temperatures, as the commercial *SNAP-tag*TM. We successfully express and purify *SsOGT-H*⁵ in both mesophilic *E. coli* and thermophilic *T. thermophilus* hosts. We demonstrated that it is possible manipulate a thermostable protein/enzyme fused to the *SsOGT-H*⁵ in harsh conditions (as the *H*⁵-*Ssβgly* fusion protein; [Vettone *et al.*, 2016]) without the need to remove the tag. These properties make *SsOGT-H*⁵ a valid alternative as protein-tag for all the *in vivo* and *in vitro* applications of the *SNAP-tag*TM technology where extreme conditions are required.

Finally, we analyse the *CeAGT-2* in the nematode *C. elegans*. This protein shows a particular feature: the predicted active site is located much nearer to the N-terminus of the protein respect of the other AGTs. The canonical N-terminal domain is totally absent in *CeAGT-2*, whereas the C-terminal domain behaves to a histon-like domain [Kanugula *et al.*, 2001]. The peculiar primary structure of *CeAGT-2* attracts our attention, asking on the possible role of this AGT *in vivo*. To this aim, we characterized a mutant strain of *CeAGT-2* by genetics and high-resolution microscopy, providing a unique opportunity for the detailed analysis of DNA repair pathways at organismal level and in true tissues. As already noted for other organisms, we confirmed a major sensitivity to alkylating agent treatment of this mutant respect to the WT strain. However, by phenotypic analysis on the *CeAGT-2* strain, I observed significant alterations of other phenotypes (embryonic lethality and aberrant phenotypes), which are often associated with mutations in genes involved in meiosis and in the formation and/or production of gametes. So, we analysed the role of *CeAGT-2* during the maturation of gametocytes, taking advantage that nematode gonad is an extremely powerful toolkit for the study of these processes, because the spatio-temporal organization of germ cells allows following their progress in specific stages of gametogenesis, from mitotic divisions up to the end of meiosis. In particular, we performed a immunolocalization on nematode gonad nuclei following the RAD-51 recombinase *foci*, which localize on DNA cuttings during meiotic recombination, generated by the topoisomerase-like protein SPO-11 during meiosis, essential for the correct execution of crossing-over [Colaiacovo *et al.*, 2003]. In WT nematode, the RAD-51 *foci* formed at meiotic DSB: they appear during the early step of pachytene stage and then rapidly repaired during last step of pachytene stage [Colaiacovo *et al.*, 2003]. In the gonad of *CeAGT-2* mutant, we found not only an increase of number of RAD-51 *foci*/nucleus in first steps of pachytene stages, but also an increase of persistence in late pachytene stage, suggesting meiotic defects during the formation of germ line. This statement requires further experiments to understand if the *CeAGT-2* is really involved in maturation of gametocytes. Another aspect of this study was the evaluation of apoptosis levels in the *CeAGT-2* strain. It is important to say that in *C. elegans* WT strain, under physiological conditions, a certain number of germ cell suffer the apoptotic process, leading to a physiological number of apoptotic bodies (an average of 3/gonad arm) [Gartner *et al.*, 2008]. A defect in genes involved in apoptotic process or involved in repair pathway, which hampers a correct resolution of meiotic DSB, it is cause of a variation of the number of apoptotic bodies. In particular, the persistence of DNA lesions in the *C. elegans* germline triggers the DNA damage checkpoint and induces increase in the germline apoptotic level. This was the case for the *CeAGT-2* strain, in which the above-mentioned number was significantly higher than the WT (an average of 5/gonad arm). Although is premature to assert with certainty the involvement of the *CeAGT-2* in the pathways above described (additional experiments on several double mutants strains are needed), these result suggest a possible role of the protein in meiotic processes and in metabolic pathways related to DNA repair during the meiotic development.

9. Bibliography

- Adams C.A., Melikishvili M., Rodgers D.W., Rasimas J.J., Pegg A.E, Fried M.G. (2009).** Topologies of complexes containing O^6 -alkylguanine-DNA alkyltransferase and DNA. *J. Mol. Biol.* 389, 248-263.
- Brenner S. (1974).** The genetics of *Caenorhabditis elegans*. *Genetics.* 77, 71-94.
- Brunk E., Mollwitz B., Rothlisberger U. (2013).** Mechanism to trigger unfolding in O^6 -alkylguanine-DNA alkyltransferase, *Chembiochem. Eur. J. Chem. Biol.* 14, 703-710.
- Budirahardja Y., Gönczy P. (2008).** PLK-1 asymmetry contributes to asynchronous cell division of *C. elegans* embryos. *Development.* 135, 1303-13.
- Cava F., de Pedro M.A., Blas-Galindo E., Waldo G.S., Westblade L.F., Berenguer J. (2008).** Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environ Microbiol.* 10, 605-613.
- C. elegans Sequencing Consortium. (1999).*** Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 285, 1493.
- Chaney S.G., Sancar A. (1996).** DNA repair: enzymatic mechanisms and relevance to drug response. *J. Natl. Cancer Inst.* 88, 1346-60.
- Crittenden S.L., Troemel E.R., Evans T.C., Kimble J. (1994).** GLP-1 is localized to the mitotic region of the *C. elegans* germ line. *Development.* 120, 2901-2911.
- Colaiácovo M.P., MacQueen A.J., Martinez-Perez E., McDonald K., Adamo A., La Volpe A., Villeneuve A.M. (2003).** Synaptonemal complex assembly in *C. elegans* is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. *Dev Cell. Sep.* 5, 463-74.
- Damoiseaux R., Keppler A., Johnsson K. (2001).** Synthesis and applications of chemical probes for human O^6 -alkylguanine-DNA alkyltransferase. *ChemBioChem.* 2, 285-7.
- Daniels D.S., Tainer J.A., (2000a).** Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O^6 -alkylguanine-DNA alkyltransferase, *Mutat. Res.* 460, 151-163.
- Daniels D.S., Mol C.D., Arvai A.S., Kanugula S., Pegg A.E., Tainer J.A. (2000b).** Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *EMBO J.* 19, 1719-1730.
- Daniels D.S., Woo T.T., Luu K.X., Noll D.M., Clarke N.D., Pegg A.E., Tainer J.A. (2004).** DNA binding and nucleotide flipping by the human DNA repair protein AGT, *Nat. Struct. Mol. Biol.* 11, 714-720.
- Dernburg A.F. (1998).** Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Science.* 94, 387-398.
- Dolan M.E., Moschel R.C., And Pegg A.E. (1990).** Depletion of mammalian O^6 -alkylguanine-DNA alkyltransferase activity by O^6 -benzylguanine provides a means to evaluate the role of this protein in protection against carcinogenic and therapeutic alkylating agents. *Proc. Natl. Acad. Sci. U.S.A.* 87, 5368-5372.

- Duguid** E.M., Mishina Y., and He C.J. (2003). How do DNA repair proteins locate potential base lesions? A chemical crosslinking method to investigate *O*⁶-alkylguanine-DNA Alkyltransferases. *Chem. Biol.* 10, 827-35.
- Duguid** E. M., Rice P. A., and He C.J. (2005). The structure of the human AGT protein bound to DNA and its implications for damage detection. *Mol. Biol.* 350, 657-66.
- Edara** S., Kanugula S., and Pegg A.E. (1999). Expression of the inactive C145A mutant human *O*⁶-alkylguanine-DNA alkyltransferase in *E. coli* increases cell killing and mutations by N-methyl-N'-nitro-N-nitrosoguanidine. *Carcinogenesis*. 20, 103-108.
- Ellis** H.M. and Horvitz H.R. (1986). Genetic control of programmed cell death in the nematode *C. elegans*. *Cell*. 44, 817-29.
- Fagbemia** A.F., Orellia B., and Schärer O.D. (2011). Regulation of endonuclease activity in human nucleotide excision repair DNA Repair. *Amst.* 10, 722-729.
- Falnes** P.O., Rognes T. (2003). DNA repair by bacterial AlkB proteins. *Res Microbiol.* 154, 531-538.
- Fang** Q., Kanugula S., and Pegg A.E. (2005). Function of domains of human *O*⁶-alkylguanine-DNA-alkyltransferase. *Biochemistry*. 44, 15396-405.
- Friedberg** E.C. (2003). DNA damage and repair. *Nature*. 421, 436-440.
- Gartner** A., Boag P.R., Blackwell T. K. (2008). Germline survival and apoptosis. *WormBook*.
- Goldstein** P. (1982). The synaptonemal complexes of *Caenorhabditis elegans*: pachytene karyotype analysis of male and hermaphrodite WT and him mutants. *Chromosoma*. 86, 577-593.
- Goodman** M.F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* 71, 17-50.
- Gumienny** T.L., Lambie E., Hartweg E., Horvitz H.R., Hengartner M.O. (1999) Genetic control of programmed cell death in the *Caenorhabditis elegans* hermaphrodite germline. *Development*. 126, 1011-1022.
- Hengartner** M.O., Ellis R.E., Horvitz H.R. (1992). *Caenorhabditis elegans* gene *ced-9* protects cells from programmed cell death. *Nature*. 356, 494-9.
- Hirsh** D., Oppenheim D., Klass M. (1976). Development of the reproductive system of *Caenorhabditis elegans*. *Developmental biology*. 49, 200-219.
- Hodgkin** J., Horvitz H.R., and Brenner S. (1979). "Non-disjunction Mutants of the Nematode *Caenorhabditis elegans*". *Genetics*. 91, 67-94.
- Hoeijmakers** J.H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature*. 411, 366-74.
- Horton** T.M., Jenkins G., Pati D., Zhang L.M. (2009). Poly(ADP-ribose) polymerase inhibitor ABT-888 potentiates the cytotoxic activity of temozolomide in leukemia cells: influence of mismatch repair status and *O*⁶-methylguanine-DNA-methyltransferase activity. *Mol. Cancer Ther.* 8, 2232-42.

Johnsson N. and Johnsson K. (2003). A fusion of disciplines: chemical approaches to exploit fusion proteins for functional genomics. *Chembiochem.* 4, 803-810.

Juillerat A., Gronemeyer T., Keppler A., Gendreizig S., Pick H., Vogel H., Johnsson K. (2003). Directed evolution of *O*⁶-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem. Biol.* 10, 313-7.

Kanugula S., Goodtzova K. and Pegg A.E. (1998). Probing of conformational changes in human *O*⁶-alkylguanine-DNA alkyl transferase protein in its alkylated and DNA-bound states by limited proteolysis. *Biochem. J.* 329, 545-550.

Kanugula S. and Pegg A.E. (2001). Novel DNA repair alkyltransferase from *Caenorhabditis elegans*. *Environmental and Molecular Mutagenesis.* 38, 235-43.

Kanugula S. and Pegg A.E. (2003). Alkylation damage repair protein *O*⁶-alkylguanine-DNA-alkyltransferase from the hyperthermophiles *Aquifex aeolicus* and *Archaeoglobus fulgidus*. *Biochem. J.* 375, 449-55.

Kanugula S., Pauly G.T., Moschel R.C., Pegg A.E. (2005). A bifunctional DNA repair protein from *Ferroplasma acidarmanus* exhibits *O*⁶-alkylguanine-DNA alkyltransferase and endonuclease V activities, *Proc. Natl. Acad. Sci. U.S.A.* 102, 3617-3622.

Kelley M.R., Logsdon D., and Fishel M.L. (2014). Targeting DNA repair pathways for cancer treatment: what's new? *Future Oncol.* 10, 1215-1237

Keppler A., Gendreizig S., Gronemeyer T., Pick H., Vogel H., Johnsson K. (2003). A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* 21, 86-9.

Keppler A., Pick H., Arrivoli C., Vogel H., Johnsson K. (2004). Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9955-9.

Kindermann M., George N., Johnsson N., Johnsson K. (2003). Covalent and selective immobilization of fusion proteins. *J. Am. Chem. Soc.* 125, 7810-23.

Klein S., and Oesch F. (1992). Assay for *O*⁶-alkylguanine-DNA-alkyltransferase using oligonucleotides containing *O*⁶-methylguanine in a BamHI recognition site as substrate. *Anal. Biochem.* 205, 294-299.

Kooistra R., Zonneveld J.B.M., Watson A.J., Margison G.P., Lohman P.H.M. and Pastink A. (1991). Identification and characterisation of the *Drosophila melanogaster* *O*⁶-alkylguanine-DNA-alkyltransferase cDNA. *Nucleic Acids Research.* 27, 1795-1801.

Lai C.H., Chou C.Y., Ch'ang L.Y., Liu C.S. Lin W. (2000). Identification of novel human genes evolutionarily conserved in *Caenorhabditis elegans* by comparative proteomics. *Genome research* 10, 703-713.

Lindahl T. (1993). Instability and decay of the primary structure of DNA. *Nature.* 362, 709-715.

Luu K.X., Kanugula S., Pegg A.E., Pauly G.T., and Moschel R.C. (2002). Repair of oligodeoxyribonucleotides by *O*⁶-alkylguanine-DNA-alkyltransferase. *Biochemistry.* 41, 8689-97.

MacQueen A.J., Colaiacovo M.P., McDonald K., Villeneuve A.M. (2002). Synapsis-dependent and -independent mechanisms stabilize homolog pairing during meiotic prophase in *C. elegans*. *Genes & development*. 16, 2428-2442.

MacQueen A.J., Phillips C.M., Bhalla N., Weiser P., Villeneuve A.M., Dernburg A.F. (2005). Chromosome sites play dual roles to establish homologous synapsis during meiosis in *C. elegans*. *Cell*. 123, 1037-1050.

McElhinney R.S., McMurry T.B., Margison G.P. (2003). *O*⁶-alkylguanine-DNA-alkyltransferase inactivation in cancer chemotherapy. *Mini. Rev. Med. Chem.* 3, 471-85.

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.

Mets D.G., Meyer B.J. (2009). Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. *Cell*. 139, 73-86.

Miggiano R., Casazza V., Garavaglia S., Ciaramella M., Perugino G., Rizzi M. and Rossi F. (2013). Biochemical and structural studies of the *Mycobacterium tuberculosis* *O*⁶-methylguanine methyltransferase and mutated variants. *J. Bacteriol.* 195, 2728-2736.

Nabeshima K., Villeneuve A.M., Colaiacovo M.P. (2005). Crossing-over is coupled to late meiotic prophase bivalent differentiation through asymmetric disassembly of the SC. *J Cell. Biol.* 168, 683-689.

Newshean S. and Yang E.S. (2012). The intersection between DNA damage response and cell death pathways. *Exp Oncol.* 34, 243-254.

Olsson M. and Lindahl T. (1980). Repair of alkylated DNA in *Escherichia coli*. Methyl group transfer from *O*⁶-methylguanine to a protein cysteine residue. *The Journ. Of Biol. Chem.* 255, 10569-71.

Pegg A.E., Wiest L., Mummert C. and Dolan M.E. (1991). Production of antibodies to peptide sequences present in human *O*⁶-alkylguanine-DNA-alkyltransferase and their use to detect this protein in cell extracts. *Carcinogenesis*. 12, 1671-7.

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

Perugino G., Vettone A., Illiano G., Valenti A., Ferrara M. C., Rossi M., and Ciaramella M. (2012). Activity and regulation of archaeal DNA alkyltransferase, conserved protein involved in repair of DNA alkylation damage. *J.B.C.* 287, 4222-4231.

Rasimas J.J., Kanugula S., D'alesio P.M., Ropson I.J., Fried M.G., Pegg A.E. (2003a). Effects of zinc occupancy on human *O*⁶-alkylguanine-DNA alkyltransferase. *Biochemistry*. 42, 980-990.

Rasimas J.J., Pegg A.E., Fried M.G. (2003b). DNA-binding mechanism of *O*⁶-alkylguanine-DNA alkyltransferase. Effects of protein and DNA alkylation on complex stability. *J. Biol. Chem.* 278, 7973-7980.

Sabharwal A. and Middleton M.R. (2006). Exploiting the role of *O*⁶-methylguanine-DNA-methyltransferase (MGMT) in cancer therapy. *Curr. Opin. Pharmacol.* 6, 355-363.

- Samson L. (1992).** The suicidal DNA repair methyltransferases of microbes. *Mol Microbiol.* 6, 825- 831.
- Samson L, Han S, Marquis JC, Rasmussen LJ. (1997).** Mammalian DNA repair methyltransferases shield O^6 -MeT from nucleotide excision repair. *Carcinogenesis.* 18, 919 -924.
- Sekiguchi M, Nakabeppu Y, Sakumi K, Tuzuki T. (1996).** DNA-repair methyltransferase as a molecular device for preventing mutation and cancer. *J. Cancer Res. Clin. Oncol.* 122, 199 -206.
- Singer B. (1976).** All oxygens in nucleic acids react with carcinogenic ethylating agents. *Nature.* 264, 333-9.
- Srivenugopal K.S., Yuan X.H., Friedman H.S., Ali-Osman F., (1996).** Ubiquitination-dependent proteolysis of O^6 -methylguanine-DNA methyltransferase in human and murine tumor cells following inactivation with O^6 -benzylguanine or 1,3-bis-(2-chloroethyl)-1-nitrosourea. *Biochemistry.* 35, 1328-1334.
- Sulston J.E., Schierenberg E., White J.G., Thomson J.N. (1983).** The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol.* 100, 64-119.
- Tentori L., Orlando L., Lacal P.M., Benincasa E., Faraoni I., Bonmassar E., D’Atri S., Graziani G. (1997).** Inhibition of O^6 -allylguanine DNA-alkyltransferase or poly (ADP-ribose) polymerase increases susceptibility of leukemic cells to apoptosis induced by temozolomide. *Mol. Pharmacol.* 52, 249-58.
- Tentori L., Graziani G. (2009).** Recent approaches to improve the antitumour efficacy of temozolomide. *Curr. Med. Chem.* 16, 245-57.
- Tessmer I., Melikishvili M. and Fried M.G. (2012).** Cooperative cluster formation, DNA bending and base-flipping by O^6 -alkylguanine-DNA alkyltransferase. *Nucleic Acids Research.* 40, 8296-8308.
- 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.
- Tsien R.Y. (1998).** The Green Fluorescent Protein. *Annu. Rev. Biochem.,* 67, 509-44.
- Villeneuve A.M. (1994).** A cis-acting locus that promotes crossing over between X chromosomes in *Caenorhabditis elegans*. *Genetics.* 136, 887-902.
- Vora R.A., Pegg A.E. and Ealick S.E. (1998).** A new model for how O^6 -methylguanine-DNA-methyltransferase binds DNA. *Proteins.* 32, 3-6.
- Wibley J.E.A., Pegg A.E., Moody P.C.E. (2000).** Crystal structure of the human O^6 -alkylguanine-DNA alkyltransferase. *Nucleic Acids Res.* 28, 393-401.
- Wu R. S., Hurst-Calderone S. and Kohn K. W. (1987).** Measurement of O^6 -alkylguanine-DNA-alkyltransferase activity in human cells and tumor tissues by restriction endonuclease inhibition. *Cancer Res.* 47, 6229-35.
- Xu-Welliver M., Pegg A.E. (2002).** Degradation of the alkylated form of the DNA repair protein O^6 -alkylguanine-DNA alkyltransferase. *Carcinogenesis.* 23, 823-830.

Yang C.G., Garcia K., and He C. (2009). Damage detection and base flipping in direct DNA alkylation repair. *ChemBioChem*. 10, 417-23.

Yoshimoto K., Mizoguchi M., Hata N., Murata H., Hatae R., Amano T., Nakamizoand A. and Sasaki T. (2012). Complex DNA repair pathways as possible therapeutic targets to overcome temozolomide resistance in glioblastoma. *Frontiers in oncology*. 389, 248-263.

Zang H., Fang Q., Pegg A.E., Guengerich, F.P. (2005). Kinetic analysis of steps in the repair of damaged DNA by human *O*⁶-alkylguanine-DNA-alkyltransferase. *The J.B.C.* 280, 30873-81.

Zhou B.B.S. and Elledge S.J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature*. 408, 433-439.