Crystal structure of *Mycobacterium tuberculosis* O^6-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 O^6-methylguanine-DNA methyltransferase (*Mt*OGT) contributes to protect the bacterial GC-rich genome against the pro-mutagenic potential of O^6-methylated guanine in DNA. Several strains of *M. tuberculosis* found worldwide encode a point-mutated O^6-methylguanine-DNA methyltransferase (OGT) variant (*Mt*OGT-R37L), which displays an arginine-to-leucine substitution at position 37 of the poorly functionally characterized N-terminal domain of the protein. Although the impact of this mutation on the *Mt*OGT activity has not yet been proved *in vivo*, we previously demonstrated that a recombinant *Mt*OGT-R37L variant performs a suboptimal alkylated-DNA repair *in vitro*, suggesting a direct role for the Arg^{37}-bearing region in catalysis. The 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 Arg^{37}-bearing random coil. This peculiar structural plasticity of *Mt*OGT could be instrumental to proper protein clustering at damaged DNA sites, as well as to protein–DNA complexes disassembling on repair.

Key words: co-operativity, crystal structure, DNA-binding protein, DNA repair, *Mycobacterium tuberculosis*, O^6-methylguanine-DNA methyltransferase.

INTRODUCTION

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

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

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

Parallel X-ray crystallography studies of the ligand-free form of *Mt*OGT showed that Arg^{37} belongs to a mainly random coiled region (residues 28–47) of the N-terminal domain, the sequence and overall structure of which significantly vary among OGTs from different species [21]. Moreover, Arg^{37} maps away from the protein active site and the DNA-binding motifs so far identified,

Abbreviations: AGT, O^6-alkylguanine-DNA alkyltransferase; E1X-dsDNA, N^1-O^6-ethano-2′-deoxyxanthosine-containing dsDNA; hAGT, human AGT; HTH, helix–turn–helix; *Mt*OGT, *Mycobacterium tuberculosis* OGT; OGT, O^6-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 Arg^{37}-to-Leu substitution produces a negligible impact on the protein conformation in the absence of ligands [21], underlining the need to obtain the structure of *Mt*OGT in alternative substrate-bound states in order to elucidate the molecular determinants of the observed suboptimal catalysis performed by the *Mt*OGT-R37L variant.

In the present study we describe the crystal structure of wild-type *Mt*OGT complexed with a modified dsDNA molecule, N^1-O^6-ethano-2′-deoxyxanthosine-containing dsDNA (*Mt*OGT::E1X-dsDNA), which reveals similar as well as peculiar traits when compared with the equivalent structure of human AGT [24]. Indeed, in the *Mt*OGT::E1X-dsDNA structure, we directly observed, for the first time, a possible mode of assembling three adjacent protein chains on to the same damaged DNA duplex. This allowed us to gain insight into the architecture of protein–DNA complexes that could explain the 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.

Crystallograpic studies

Crystallization

Wild-type *Mt*OGT was purified as previously described [21], with buffer exchanged against 20 mM Tris/HCl, pH 7.5, and 25 mM NaCl (PD10 column, GE Healthcare), mixed with E1X-dsDNA in equimolar ratio, and incubated for 18 h at 4 °C. The reaction mixture was concentrated (10-kDa 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 spare-matrix strategy using kits from Hampton Research and Qiagen. The initially obtained needle clusters were used as micro-seeds to inoculate 1 μl 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 μl 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 μl 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 μl 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 $P2_12_12_1$, with cell

Table 1 Data collection, phasing and refinement statistics

Values in parentheses refer to the highest resolution shell.

	*Mt*OGT::E1X-dsDNA	*Mt*OGT-R37K	*Mt*OGT-Y139F
Data collection			
Space group	$P2_12_12_1$	$P2_12_12$	$P2_12_12$
Wavelength (Å)	0.972	0.979	0.99
Resolution (Å)	3.0	2.3	2.6
Total reflections	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_{factor}/R_{free} (%)	19.4/26.5	18.4/22.5	20.4/27.6
Protein/DNA atoms	4257	1257	1258
Ligand atoms	6	12	4
Water molecules	8	61	9
RMSD bonds (Å)	0.011	0.009	0.014
RMSD angles (°)	1.42	1.05	1.75
Average B (Å^2)			
Protein	62.0	41.7	55.0
Solvent	27.2	40.9	50.8

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 Tyr^{139} 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 Arg^{37} or Tyr^{139} 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 Arg^{37} in *Mt*OGT-mediated catalysis, we co-crystallized the wild-type protein in the presence of the 13-bp-long E1X-dsDNA, thus choosing the same experimental strategy first adopted by Daniels et al. [24] to solve the structure of wild-type hAGT covalently bound to a modified dsDNA (PDB code 1T39). Different from the equivalent structure of the human enzyme, in the *Mt*OGT::E1X-dsDNA crystal structure, three protein chains (A, B and C) and one E1X-dsDNA molecule are present in the asymmetrical unit, with chain A binding the E1X base at position 7 of the modified strand ($E1X_7$) (Figure 1a). It is of interest that, by applying crystallographic symmetry operators, a peculiar supramolecular assembly can be observed in the *Mt*OGT::E1X-dsDNA crystal lattice (Figure 1b). By focusing on a unit consisting of chain A bound to the $E1X_7$ base, and counting nitrogenous bases starting from the 5′-end of each strand, a symmetry equivalent of chain C (‘C sym. mate’) binds the deoxyadenosine residue at position 4 of the modified strand (dA_4), and a symmetry equivalent of chain B (‘B sym. mate’) does the same with the deoxyadenosine residue at position 5 of the complementary strand (dA_{18}). In all cases, the bound base adopts a fully extrahelical conformation, and is deeply inserted into the protein active site. Overall, the *Mt*OGT::E1X-dsDNA complex can be described as consisting of two co-oriented *Mt*OGT monomers sharing 1 bp of their 4-bp-long DNA-anchoring site on the damaged strand and displaying a reciprocal ‘N-to-C’ domain arrangement (‘chain A’ and ‘C sym. mate’), whereas the third chain (‘B sym. mate’), which binds the intact strand, shows a ‘C-to-C’ domain arrangement with respect to chain A (Figure 1c).

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

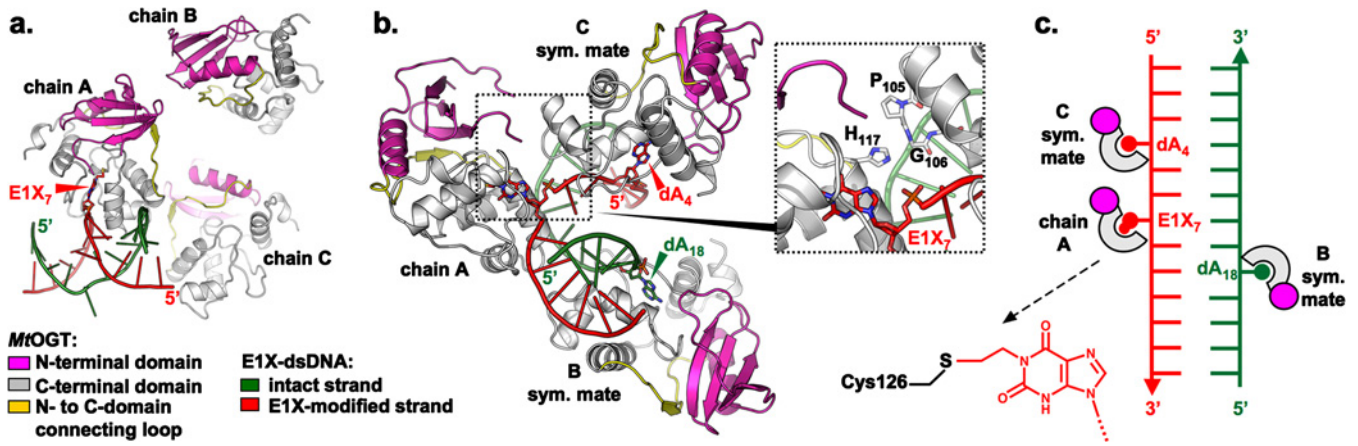

Figure 1 The overall structure of *Mt*OGT complexed with modified DNA

(**a**) Cartoon representation of the asymmetrical unit content of the *Mt*OGT::E1X-dsDNA crystal; the E1X-containing dsDNA is observed complexed with chain A, the C-terminal domain of which hosts the modified base (E1X, red arrowhead). (**b**) Cartoon representation of three *Mt*OGT chains assembled on to the same E1X-dsDNA molecule, resulting from applying crystal symmetry operators; extrahelical bases are rendered as sticks and 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 (Cys^{126}) 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 *Mt*OGT active site complexed with E1X-dsDNA

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

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

With the obvious exception of contacts involving $E1X_7$-specific positions, an almost identical bonding scheme is observable in the active site of chains B and C – which host the symmetry equivalent of the dA_{18} and dA_4 unmodified bases, respectively (Figure 2d), thus indicating that *Mt*OGT can efficiently bind nitrogenous bases independently of the presence of the alkyl adduct.

In principle, the insertion of an undamaged adenine residue into the *Mt*OGT ligand-binding pocket would not expose the DNA substrate to an increased risk of chemical modifications. In fact, the reactivity of the purine ring N^1 and N^6 positions, as they were observed in the active site of the B and C chains of the *Mt*OGT::$E1X_7$-dsDNA structure, does not appear significantly enhanced by the nearby catalytic cysteine, nor by the presence of the other residues 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 *Mt*OGT::E1X-dsDNA structure, the thymine base appears to be only partially inserted into the hAGT-B ligand-binding pocket (see Supplementary Figure S3b). We therefore speculated that *Mt*OGT could perform lesion searching through a non-selective base-flipping mechanism, with the flipped-out base fully inserted into the active site. If this assumption is correct, *Mt*OGT could not adopt a gate-keeping mechanism in discriminating between normal and damaged bases *in vivo*, different from what was hypothesized for the human counterpart [25,33].

*Mt*OGT undergoes structural rearrangements on DNA binding

The structure of *Mt*OGT complexed with E1X-dsDNA discloses a further unique feature of the mycobacterial protein, i.e. its conformational plasticity. In fact, the structural analyses of hAGT [22–25] and *Sulfolobus solfataricus* OGT [36], at different stages

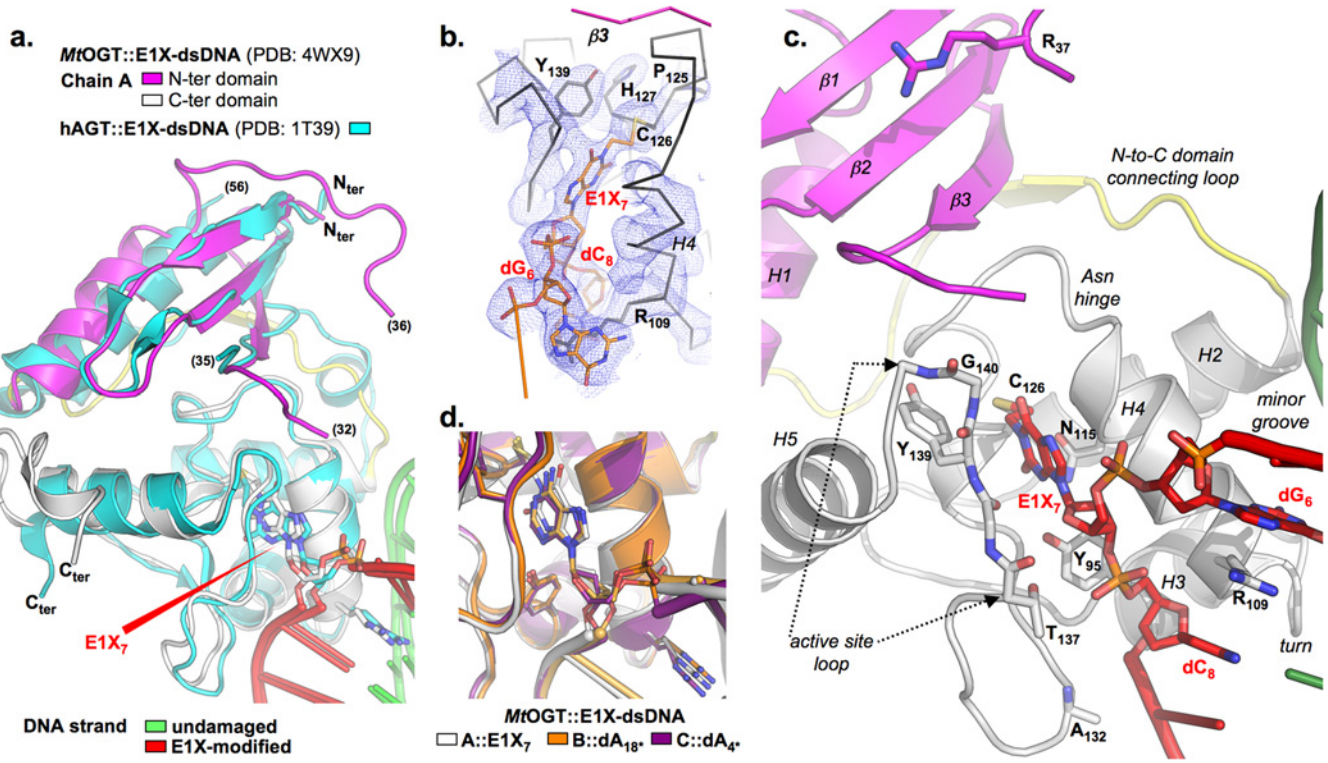

Figure 2 Structural analysis of the *Mt*OGT protein complexed with E1X-dsDNA

(**a**) Cartoon representation of the optimally superimposed structures of *Mt*OGT (chain A) and hAGT (PDB code 1T39), each in complex with the E1X-dsDNA substrate; the E1X base is rendered as sticks and coloured, applying the same colour codes used for the corresponding protein chain (shown on the top of the panel). (**b**) Close-up view of the active site of the *Mt*OGT::E1X-dsDNA chain A housing the modified base ($E1X_7$), with σ_A-weighted $2F_o$–F_c electron density contoured at 1.0 σ; the Cys^{126} thiol group is observed at a distance of 2.1 Å from $E1X_7$, the C11 atom; the protein backbone appears as a ribbon. (**c**) Close-up view of the *Mt*OGT chain A active site complexed with E1X-dsDNA; secondary structural elements and functional motifs are indicated in italic (the colour codes for protein domain/DNA strand identification appear in **a**). (**d**) Cartoon representation of the active site of optimally superimposed *Mt*OGT::E1X-dsDNA chains A, B and C (average RMSDs are 0.671 and 0.456 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 *Mt*OGT with the E1X-dsDNA substrate induces the repositioning of three solvent-exposed protein regions: a random coiled segment (residues 29–39) of the N-terminal domain, part of the active-site loop (residues 135–142) and the C-terminal tail (residues 156–165) (Figure 3a). As a consequence, each protein monomer in the *Mt*OGT::E1X-dsDNA complex appears more compact than the ligand-free protein (Figure 3b). These conformational changes are accompanied by the side-chain repositioning of a number of residues of both protein domains (Figure 3c). It is interesting that, in the *Mt*OGT::E1X-dsDNA structure, the segment encompassing residues 29–35 moves away from the three-stranded β-sheet that builds up the core of the N-terminal domain, and gets closer to the DNA-binding surface of the C-terminal domain, behaving as a flap that sees Arg^{37} as its pivotal point.

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

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

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

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

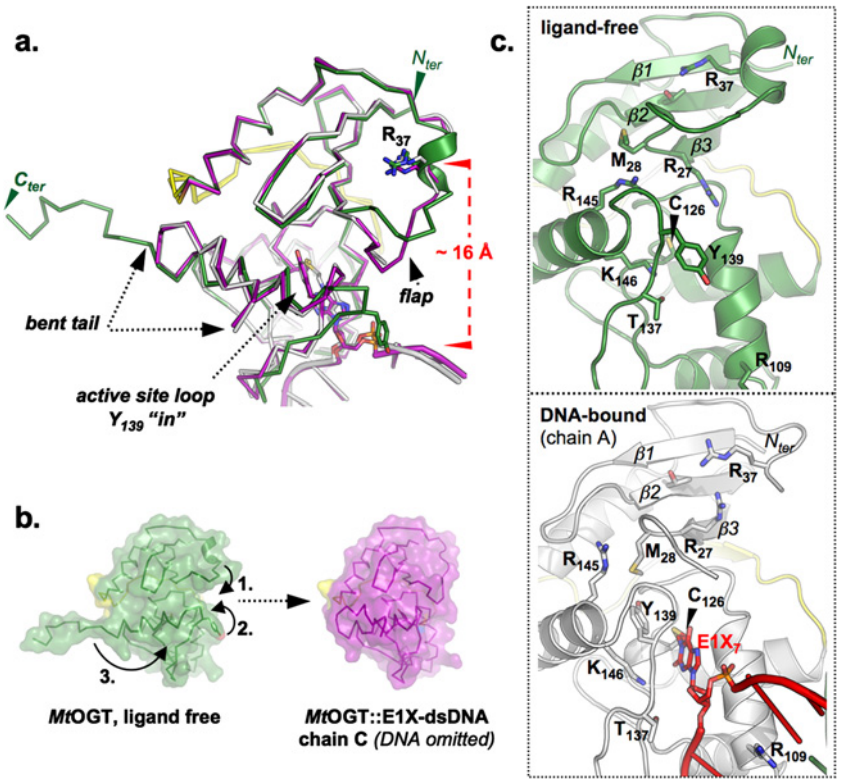

Figure 3 The conformation adopted by discrete protein regions differs in the ligand-free and DNA-bound *Mt*OGT structures

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

same VG-based assay [32] previously used to characterize the wild-type protein and the *Mt*OGT-R37L variant [21]. Our data (see Supplementary Table S2) show that the *Mt*OGT-R37E mutant exhibits a 5-fold lower affinity for the methylated duplex ($K_{DNA^{met}R37E} = 1.14 \pm 0.15\ \mu M$) with respect to *Mt*OGT ($K_{DNA^{met}wt} = 0.24 \pm 0.11\ \mu M$ [21]), whereas the more conservative Arg^{37}-to-Lys substitution produces a more limited effect on the $dsDNA^{met}$-binding constant ($K_{DNA^{met}R37K} = 0.38 \pm 0.2\ \mu M$).

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 *Mt*OGT-R37E protein reaches a plateau in band-shift activity at a DNA/protein molar ratio of 1:600 ($K_{DNA\text{-}R37E} = 41.4 \pm 1.1\ \mu M$), whereas both the wild-type *Mt*OGT and the *Mt*OGT-R37K variant induce a complete shift at a DNA/protein molar ratio in the range of 1:150 ($K_{DNA\text{-}wt} = 7.2 \pm 0.2\ \mu M$ [21], $K_{DNA\text{-}R37K} = 13.2 \pm 0.7\ \mu M$). 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 *Mt*OGT-R37L variant displays a 10-fold lower affinity towards the $dsDNA^{met}$ substrate compared with the wild-type *Mt*OGT, although the co-operativity of DNA binding is maintained [21]. Taken together, our previous study [21] and the present study confirm that the Arg^{37} residue, although not being directly involved in substrate binding, plays an active role during catalysis, a role that can be performed almost equally well by the positively charged lysine residue. By contrast, the presence of a hydrophobic or negatively charged side chain at position 37 of the *Mt*OGT protein, which characterizes the *Mt*OGT-R37L and *Mt*OGT-R37E proteins respectively, translates into less efficient DNA binding and repair.

By analysing the crystal structure of the *Mt*OGT-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 *Mt*OGT-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 Arg^{37} 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 *Mt*OGT::E1X-dsDNA structure.

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

As mentioned above, the active-site loop and the C-terminal tail of *Mt*OGT adopt different conformations, depending on the association of the protein with the DNA substrate (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

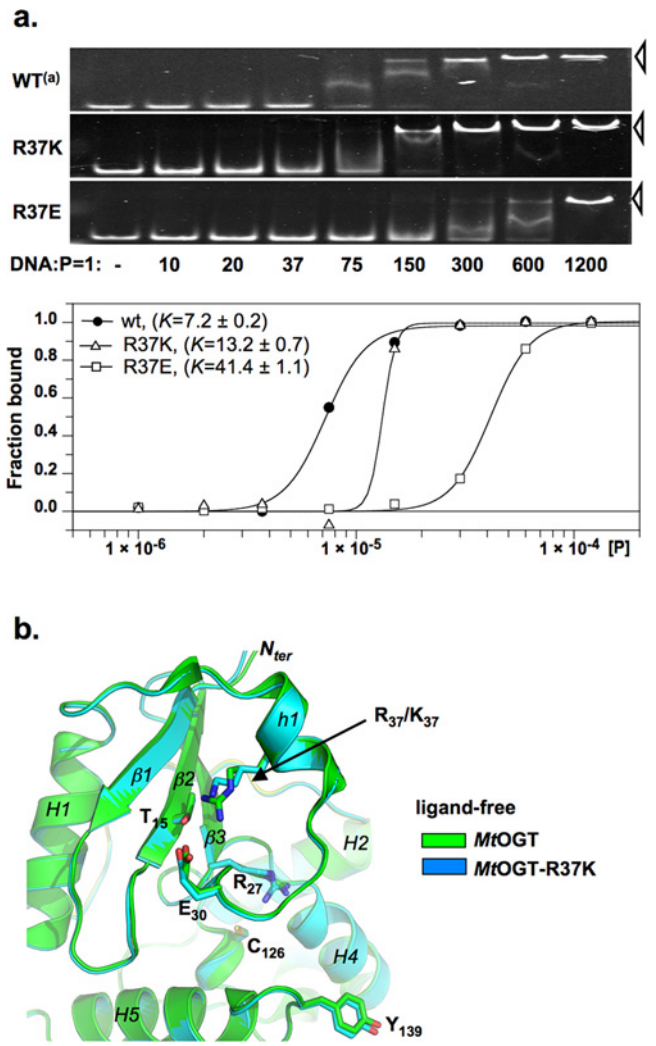

Figure 4 A positively charged residue at *Mt*OGT position 37 is required for optimal dsDNA binding

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

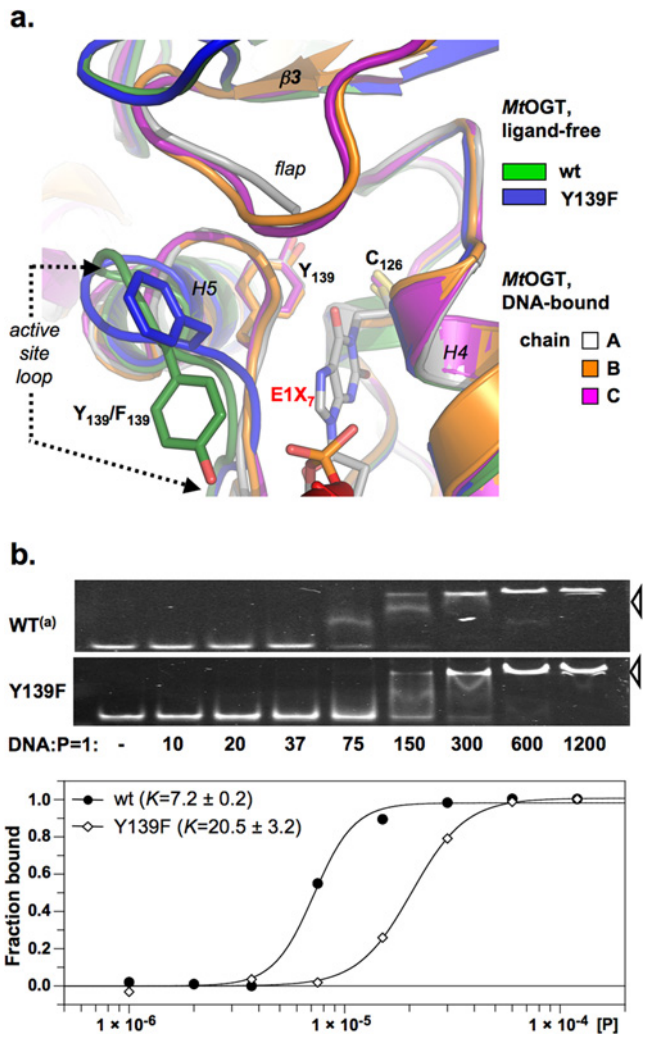

Figure 5 Tyr^{139} could help the active-site loop movements during DNA binding

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

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 *Mt*OGT and its mutated variants is invariably oriented towards the bulk solvent. This conformation is stabilized by contacts established between the conserved Tyr^{139} of the active-site loop and the stretched-out C-terminal tail of the closest symmetry mate within the crystal lattice [21] (and the present study). On the contrary, the C-side of the active-site loop of each protein chain that builds up the *Mt*OGT::E1X-dsDNA structure is bent inwards towards the catalytic pocket, where it participates in making the ligand-binding cavity fit the flipped-out base (Figure 5a). These observations raise the possibility that the active site of *Mt*OGT could exist in two alternative conformations ('ligand-free/active-site loop out' or 'DNA-bound/active-site loop in') also in a physiological context, displaying a degree of structural plasticity higher than that characterizing the equivalent region of hAGT.

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

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

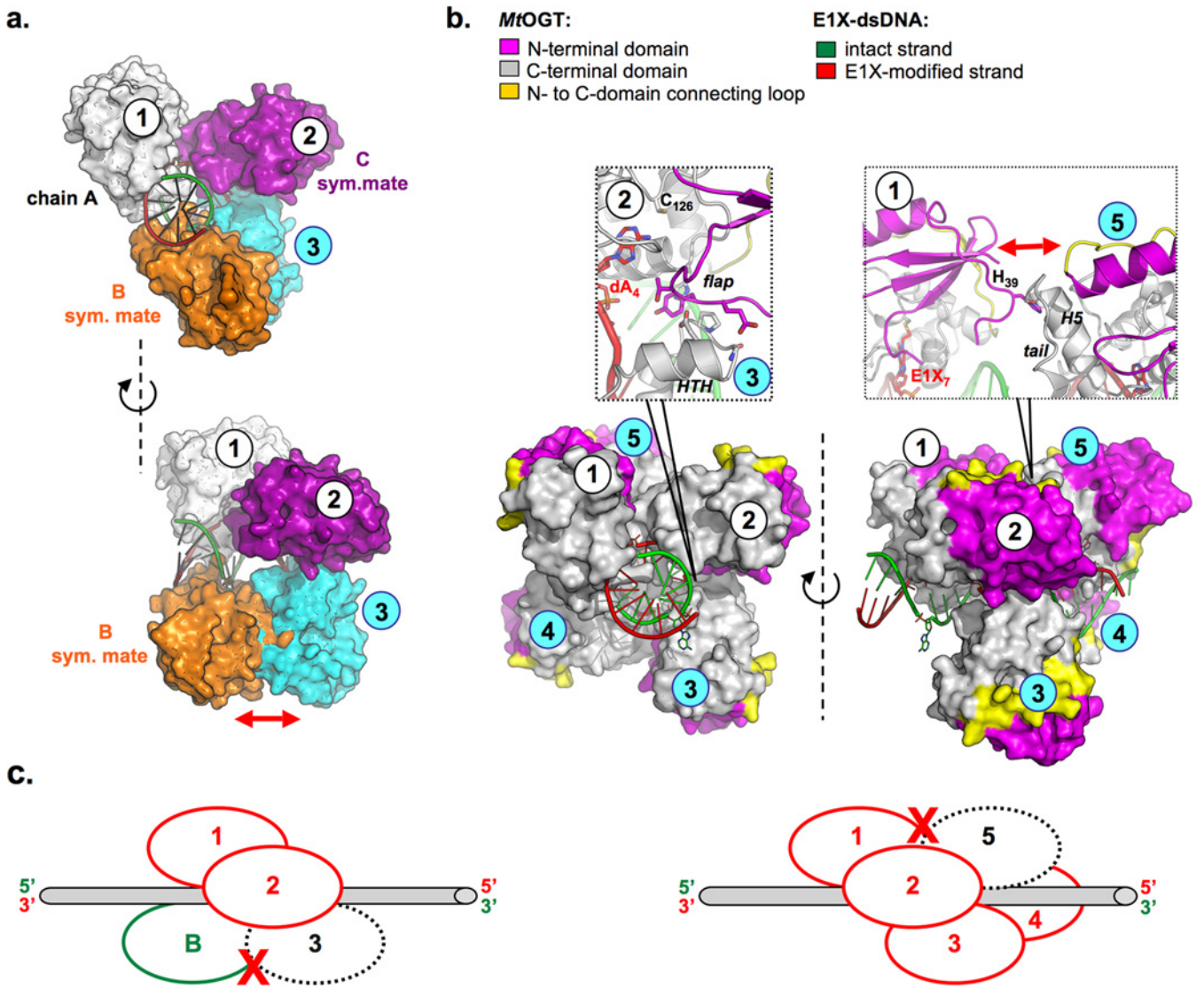

Figure 6 Structure-based model of *Mt*OGT clustering on to a DNA duplex

(**a**) The possible negative effect of the presence of a *Mt*OGT monomer ('B sym. mate'), which is bound to the complementary strand opposing the lesion, on the growth of the protein cluster towards the 5′-end of the damaged strand. (**b**) Structure-based model of five co-oriented *Mt*OGT monomers assembled on to the same dsDNA molecule; the 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 *Mt*OGT monomers that bind bases of the damaged or intact strand are depicted as red or green ovals, respectively; the dashed ovals represent the *Mt*OGT subunit that would come into collision with pre-assembled monomers.

($K_{DNA}{}^{met}{}_{Y139F} = 2.19 \pm 0.5\ \mu M$; $K_{DNA}{}^{met}{}_{wt} = 0.24 \pm 0.11\ \mu M$ [21]), confirming the *Mt*OGT requirement of a tyrosine residue at position 139 for optimal repair of an O^6-methylated guanine in dsDNA. Instead, the ability of *Mt*OGT to bind unmodified dsDNA appears less affected by the Tyr^{139}-to-Phe substitution; in fact, when analysed in EMSA, both proteins reach a plateau in band-shift activity at a DNA/protein molar ratio of approximately 1:150, displaying an affinity towards the unmodified probe that differs 3-fold ($K_{DNA\text{-}wt} = 7.2 \pm 0.2\ \mu M$ [21], $K_{DNA\text{-}Y139F} = 20.5 \pm 3.2\ \mu M$; Figure 5b). Therefore, we speculate that Tyr^{139} could play a role not only in properly fixing the base inside the protein active site on DNA binding, as proposed for hAGT Tyr^{158} [24,25,38–40], but also in making *Mt*OGT able to discriminate between intact and alkylated dsDNA molecules, albeit through a molecular mechanism that will need further study for elucidation.

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

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

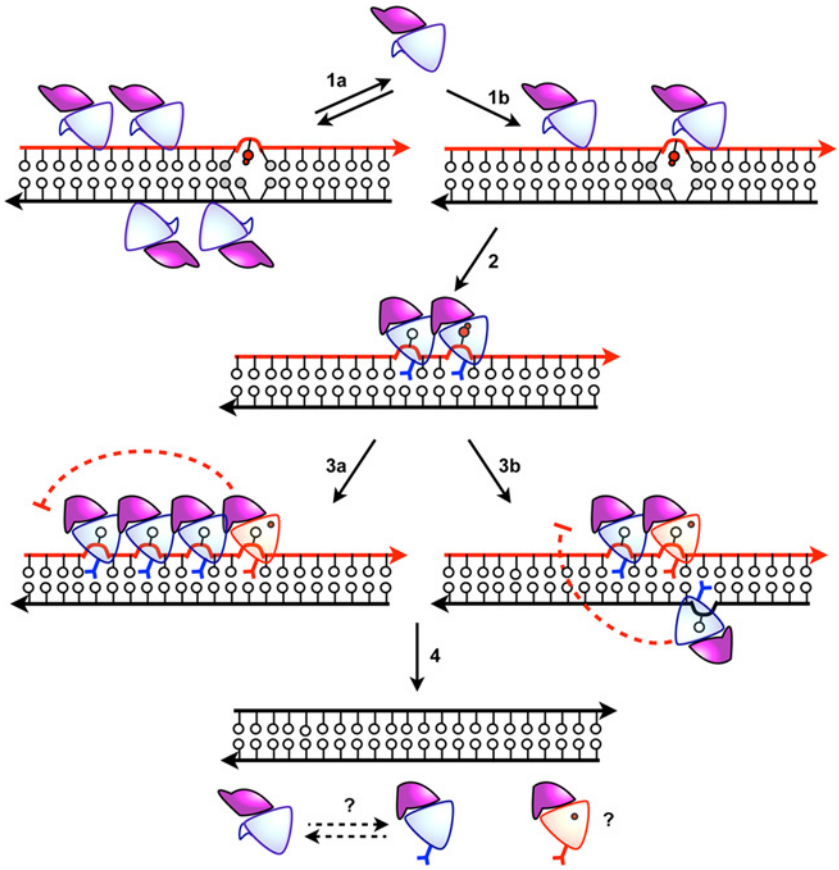

Figure 7 Preliminary model of *Mt*OGT-mediated direct alkylated DNA repair

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 *Mt*OGT protein clusters larger in size than the one characterizing the *Mt*OGT::E1X-dsDNA crystal structure. To verify this hypothesis, we tried to model further DNA-bound *Mt*OGT monomers towards the 5′-end of the modified strand, starting from the experimental 'C sym. mate' (monomer '2' in Figure 6), and using the 'chain A/C sym. mate' dimer as the moving unit (monomers '1' and '2' in Figure 6).

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

On the other hand, the *Mt*OGT::E1X-dsDNA crystal structure itself does not provide any indication regarding the dynamics of the alkylated-DNA damage reversal performed by *Mt*OGT. Therefore, we can also assume that the binding of co-oriented *Mt*OGT monomers on to the modified strand could be favoured on assembly of protein clusters with components that display unbiased binding to both strands. To visualize the former situation, we reiterate the superimposition procedure mentioned above, by omitting the chain B symmetry mate. We obtained a model of *Mt*OGT protein clusters (Figure 6b) that proves to be more compact compared with the one proposed for hAGT [45,46], possibly due to the structural plasticity of *Mt*OGT, which could allow more crowded protein assembling on to DNA. It is interesting that the DNA binding-associated repositioning of the flap (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 *Mt*OGT–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 *Mt*OGT::E1X-dsDNA crystal structure could provide a potential solution to an inconsistency present in the literature concerning the DNA-binding mechanism of O^6-alkylguanine-DNA alkyltransferases. Indeed, in spite of the fact that previous X-ray crystallography-based studies revealed a 1:1 protein/modified DNA stoichiometry [24,25,34]: (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. Samarpita Lahiri performed robot-assisted crystallization trials and optimized crystal-growth conditions. All authors approved the final version of the manuscript.

ACKNOWLEDGEMENTS

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

FUNDING

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

REFERENCES

1 Mizrahi, V. and Andersen, S.J. (1998) DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? Mol. Microbiol. **29**, 1331–1339 CrossRef PubMed
2 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
3 Durbach, S.I., Springer, B., Machowski, E.E., North, R.J., Papavinasasundaram, K.G., Colston, M.J., Bottger, E.C. and Mizrahi, V. (2003) DNA alkylation damage as a sensor of nitrosative stress in *Mycobacterium tuberculosis*. Infect. Immun. **71**, 997–1000 CrossRef PubMed
4 Warner, D.F. and Mizrahi, V. (2006) Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. Clin. Microbiol. Rev. **19**, 558–570 CrossRef PubMed
5 Drablos, F., Feyzi, E., Aas, P.A., Vaagbo, C.B., Kavli, B., Bratlie, M.S., Pena-Diaz, J., Otterlei, M., Slupphaug, G. and Krokan, H.E. (2004) Alkylation damage in DNA and RNA – repair mechanisms and medical significance. DNA Repair (Amst.) **3**, 1389–1407 CrossRef PubMed
6 Gorna, A.E., Bowater, R.P. and Dziadek, J. (2010) DNA repair systems and the pathogenesis of *Mycobacterium tuberculosis*: varying activities at different stages of infection. Clin. Sci. (Lond.) **119**, 187–202 CrossRef PubMed
7 Kurthkoti, K. and Varshney, U. (2012) Distinct mechanisms of DNA repair in mycobacteria and their implications in attenuation of the pathogen growth. Mech. Ageing Dev. **133**, 138–146 CrossRef PubMed
8 Pegg, A.E. (2011) Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. Chem. Res. Toxicol. **24**, 618–639 CrossRef PubMed
9 Pegg, A.E. (2000) Repair of O^6-alkylguanine by alkyltransferases. Mutat. Res. **462**, 83–100 CrossRef PubMed
10 Fang, Q., Kanugula, S., Tubbs, J.L., Tainer, J.A. and Pegg, A.E. (2010) Repair of O4-alkylthymine by O6-alkylguanine-DNA alkyltransferases. J. Biol. Chem. **285**, 8185–8195 CrossRef PubMed
11 Tubbs, J.L., Pegg, A.E. and Tainer, J.A. (2007) DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. DNA Repair (Amst.) **6**, 1100–1115 CrossRef PubMed
12 Tessmer, I., Melikishvili, M. and Fried, M.G. (2012) Cooperative cluster formation, DNA bending and base-flipping by O6-alkylguanine-DNA alkyltransferase. Nucleic Acids Res. **40**, 8296–8308 CrossRef PubMed
13 Kanugula, S., Goodtzova, K. and Pegg, A.E. (1998) Probing of conformational changes in human O6-alkylguanine-DNA alkyl transferase protein in its alkylated and DNA-bound states by limited proteolysis. Biochem. J. **329**, 545–550 CrossRef PubMed
14 Xu-Welliver, M. and Pegg, A.E. (2002) Degradation of the alkylated form of the DNA repair protein, O^6-alkylguanine-DNA alkyltransferase. Carcinogenesis **23**, 823–830 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., Rauzier, 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., Rauzier, 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., 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 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 O^6-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 O^6-alkylguanine alkyltransferase-DNA complex using N^1,O^6-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 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 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 Ciaramella, 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 O^6-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 O^6-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 O^6-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

Received 24 July 2015/19 October 2015; accepted 28 October 2015
Accepted Manuscript online 28 October 2015, doi:10.1042/BJ20150833