

---

**DEVELOPMENT OF A NEW ENZYME  
BASED BIOSENSOR FOR DETECTION OF  
DAMAGED DNA, BY USING THE  
*THERMO*SNAP-DISPLAY TECHNOLOGY**

---

**Rosanna Mattosovich**

Dottorato in Biotecnologie XXXIII ciclo

Università di Napoli Federico II







---

**DEVELOPMENT OF A NEW ENZYME  
BASED BIOSENSOR FOR DETECTION OF  
DAMAGED DNA, BY USING THE  
*THERMO*SNAP-DISPLAY TECHNOLOGY**

---

**Rosanna Mattossovich**

Dottorando: Rosanna Mattossovich  
Relatore: Prof. Marco Moracci  
Co-Relatore Dott. Giuseppe Perugino  
Coordinatore: Prof. Marco Moracci



*A Lulù*

*Se fossi ancora qui con*

*Ti farei vedere io che la lezione d'as*

*che mi hai inseg*

*Io l'ho imparata b*

*Sempre sarai nella tasca in a*

*Per sempre sarai in un sorriso inaspe*

# Index

---

<b>Summary</b>		<b>2</b>
<b>Abstract</b>		<b>8</b>
<b>Abrrreviations</b>		<b>9</b>
<b>Outiline of thesis</b>		<b>10</b>
<b>Chapter 1</b>	<b>General Introduction</b>	<b>11</b>
	Pollution and environmental safety and health	12
	<i>Genomic stability and DNA damages</i>	12
	<i>Mechanism of DNA repair</i>	17
	<i>Direct repair and O<sup>6</sup>alkyl-guanine</i>	19
	<i>DNA-alkyltransferase</i>	
<b>Chapter 2</b>	<b>Purposes of the thesis</b>	<b>32</b>
	<i>Analytical methods based on thermostable chimeras</i>	33
	<i>Production of a new AGT substrate based on a DNA triplex</i>	33
<b>Chapter 3</b>	<b>Development of a thermostable Chimera for the rapid determination of O<sup>6</sup>AGs in High-throughput</b>	<b>35</b>
	<i>Fast thermostable OGT</i>	39
	<i>Enhancing the thermostability of Halotag</i>	45
	<i>Construction and production of two different chimeras</i>	50
<b>Chapter 4</b>	<b>Production of a new AGTs substrate based on a DNA Triplex</b>	<b>63</b>
<b>References</b>		<b>81</b>
<b>Appendix I</b>	<b>Scientific Pubblications</b>	
<b>Appendix II</b>	<b>Communications</b>	
<b>Appendix III</b>	<b>Experiences in foreign laboratory</b>	
<b>Appendix IV</b>	<b>Papers</b>	

## Summary

Il DNA, l'unità di base dell'ereditarietà, è costantemente soggetto a modificazioni covalenti da parte di composti chimici che inducono cambiamenti molecolari come la formazione di legami covalenti tra i due filamenti del DNA, la loro rottura, la perdita o la modifica delle basi. Tra le sostanze chimiche più deleterie per il DNA ci sono gli agenti alchilanti, molecole molto reattive che legano gruppi chimici negli acidi nucleici, provocando alterazioni nelle loro funzioni. Gli agenti alchilanti esogeni sono prodotti principalmente da componenti alimentari, tabacco, fumo di sigaretta, combustione di biomasse, processi industriali e per molti anni, sono stati impiegati anche come agenti chemioterapici. La loro azione consiste nell'introdurre, generalmente, gruppi metilici o etilici su tutti gli atomi di ossigeno e azoto disponibili sulle basi del DNA, producendo diversi tipi di lesioni. La maggior parte dei dati a disposizione indica che, tra tutte le modifiche possibili, soltanto le N<sup>3</sup>-alchil-adenine e le O<sup>6</sup>-alchil-guanine hanno conseguenze sull'integrità dell'informazione genetica. In particolare, l'O<sup>6</sup>-metil-guanina si appaia erroneamente con la timina, e il suo mancato riparo porta alla mutazione per la transizione GC→AT del DNA durante il processo replicativo (figura 2). A tale scopo, l'evoluzione ha selezionato un enzima capace di riparare direttamente questo tipo di danno: l'O<sup>6</sup>-alchilguanina-DNA-alchiltransferasi (AGT, OGT, MGMT). Questa proteina catalizza la riparazione del DNA dal danno da alchilazione con un meccanismo *one-step*, che coinvolge il trasferimento irreversibile del gruppo alchilico dall'O<sup>6</sup>-alchil-guanina a un residuo conservato di cisteina nel sito attivo. La proteina alchilata è riconosciuta non più funzionante dai sistemi di degradazione cellulari e per tale motivo, le AGT sono definite proteine suicide o kamikaze (figura 3). L'AGT più studiata in letteratura è quella umana (MGMT), perché la sua inaspettata over-espressione in alcuni tipi di neoplasie, porta alla resistenza al trattamento chemioterapico a base di agenti alchilanti. Poiché l'attività e la variabilità dell'espressione delle AGT contrasta l'azione degli agenti alchilanti, questo progetto si propone di mettere a punto delle metodologie innovative, specifiche per l'analisi qualitativa e quantitativa del DNA danneggiato e quindi identificare il contenuto di basi azotate O<sup>6</sup>-alchil-guanina (O6-AG) in campioni di DNA provenienti da diverse fonti. In questo progetto

per raggiungere tale scopo, sono stati utilizzati due diversi approcci: il primo basato sulla produzione di chimere termostabili e il secondo sull'uso di oligonucleotidi opportunamente modificati (figura 9).

La chimera è composta da due frazioni: la prima è costituita dall'AGT che in presenza di DNA alchilato, rimuove la modifica alchilandosi irreversibilmente e di conseguenza non potrà più reagire con un substrato fluorescente (BG-FL) presente nella miscela di saggio. Ciò comporterà la riduzione del segnale di fluorescenza in modo direttamente proporzionale alla quantità di <sup>m</sup>DNA nel campione e con una linea di calibrazione appropriata (costruita con quantità note di <sup>m</sup>DNA), è possibile determinare il grado di alchilazione del DNA. L'altra parte della chimera è l'Halotag<sup>®</sup>, un aloalcano-deidrogenasi commerciale, modificata per legarsi covalentemente a un cloro-alcano che, legato a un supporto, permette l'immobilizzazione diretta e specifica della chimera. Va sottolineato che il collo di bottiglia di questo saggio è la necessità di eseguire la reazione a temperature moderatamente elevate (circa 60 ° C), in modo da eliminare le attività delle AGT endogene ancora presenti nei campioni di DNA. Per questi motivi, le chimere sono state opportunamente ingegnerizzate per essere tanto attive quanto stabili a queste temperature (Figura 10).

Per le motivazioni sopra descritte, mi sono occupata della ricerca e caratterizzazione di AGT termostabili alternative alla OGT proveniente dal *Saccharolobus solfataricus*, concentrando la mia attenzione sul OGT di *Pyrococcus furiosus* e del Batterio Gram-negativo *Thermotoga neapolitana*. Dalla caratterizzazione dell'AGT di *Thermotoga neapolitana* ho dimostrato che è completamente attiva sul substrato commerciale della benzil-guanina fluoresceina (SNAP-Vista Green<sup>®</sup>; BG-VG), mostrando un forte segnale fluorescente; ha un IC50 simile a quella ottenuta con l'enzima *S. solfataricus*, confermando così un'attività riparatrice. ; [Perugino et al., 2012; Vettone et al., 2016]) (figura 12), è molto attiva a basse temperature ostacolando la determinazione delle costanti del secondo ordine a temperature superiori a 50 ° C, poiché la sua velocità di reazione va oltre i limiti tecnici del saggio (tabella 1 e 2). L'unico importante svantaggio della proteina TnOGT è la degradazione durante la conservazione a -20°C e di conseguenza non è possibile utilizzarla per gli step successivi del progetto

(figura 13). Tuttavia, l'analisi eseguita sull'OGT proveniente da *P. furiosus* ha mostrato che, come previsto, questo enzima possiede una stabilità eccezionale ma una termofilia altrettanto eccezionale, essendo meno attivo di SsOGT sopra i 60 ° C. Per questi motivi anche questa proteina non può essere utilizzata per la costruzione della chimera. Di conseguenza si è scelto di utilizzare per la costruzione delle chimere la SsOGT.

Per permettere l'immobilizzazione diretta e specifica della chimera ad un supporto, ed eseguire il nuovo saggio è necessario ottenere una variante termostabile dell'Halotag. Questo aspetto del progetto è stato svolto presso il Dipartimento di Biologia Molecolare dell'Università Autonoma di Madrid (Spagna) in collaborazione con il gruppo del Dr. Aurelio Hidalgo. Durante la mia attività di ricerca in Spagna, ho clonato e ottimizzato l'espressione della proteina Halotag utilizzando il metodo definito come *folding interference principle*, che prevede l'uso di una variante termostabile della kanamicina nucleotidil transferasi (Kat) [Matsumura et al., 1985]. In questo metodo, Kat è fuso al Cterm della proteina di interesse e la fusione risultante è espressa nel batterio termofilo *Thermus thermophilus* HB27, uno dei microrganismi più importanti utilizzati per applicazioni biotecnologiche, perché ha la capacità di crescere in condizioni di laboratorio ed in termini di trasformazione è estremamente naturalmente competente (figura 14). Ho ottenuto mutanti più resistenti alla denaturazione termica della porzione di halotag costruendo librerie di mutanti halotag mediante mutagenesi casuale e di conseguenza ho selezionato alcune varianti termostabili dell'halotag (figura 15).

Poiché l'obiettivo finale nello sviluppo della chimera è che l'attività e la stabilità di ogni enzima (OGT e Halotag) non siano influenzate dalla presenza dell'altro, ho prodotto due diverse versioni: OGT-Halo (chimera 1) e Halo-OGT (chimera 2).

Purtroppo per quanto riguarda la chimera 1 non ho ottenuto una quantità sufficiente di proteina per consentire una completa caratterizzazione biochimica e sono in corso una serie di tentativi per potenziare e ottimizzare l'espressione di questa chimera. Invece la chimera Halo-OGT (circa 52 kDa) reagisce con i rispettivi substrati (Halotag<sup>TMR</sup> e SNAP-Vista green) e la sua efficienza, è paragonabile alle forme libere (circa 20 e 33 kDa rispettivamente). Questi risultati suggeriscono che SsOGT, anche

quando posizionato all'estremità C-term di una proteina, può essere etichettato in modo specifico ed efficiente con il suo substrato (BG-VG). Ho valutato anche l'attività catalitica della Halo-OGT su entrambi i substrati. E anche in questo caso la chimera risulta essere attiva mostrando una notevole capacità di eseguire la reazione a temperature molto basse come la forma libera dell'halotag (figura 16). L'attività OGT della chimera è stata misurata a 25 ° C e 60 ° C, risultando molto vicina alla SsOGT [Perugino et al., 2012,2015; Mattosovich et al 2020]. Di conseguenza la chimera può essere utilizzata a temperature di 50-60 ° C, come prevede lo scopo del progetto.

Inoltre la porzione OGT della chimera in presenza mDNA ha un'attività di riparo come enzima SsOGT [Perugino et al., 2015], confermando così un ruolo nella riparazione del DNA di questa proteina chimerica. L'ultima fase del progetto prevedeva *high-throughput* basato sulla chimera che non è stato possibile mettere a punto a causa della Pandemia.

La metil-transferasi umana (MGMT) protegge le cellule tumorali dall'effetto citotossico della chemioterapia a base di agenti alchilanti (figura 6). Per questi motivi, questo enzima è di fondamentale importanza clinica e lo sviluppo di inattivatori / inibitori di hMGMT da utilizzare in combinazione con questo tipo di chemioterapia è in continua evoluzione. Dato la forte rilevanza clinica, parte di questo progetto è stato investito per ottenere un test in grado di misurare l'attività della metil-transferasi in modo affidabile e rapido. Per la valutazione dell'attività metil-transferasica, in collaborazione con il Prof. Ricci dell'Università di Roma Tor Vergata, è stato progettato un DNA Triplex a singolo filamento, in grado di formare una struttura triplex intramolecolare tramite legami idrogeno (Interazioni Hoogsteen) tra un dominio a forcina (duplex) e un altro singolo filamento. Sono stati sintetizzati tre triplex che differiscono per la presenza o meno dei gruppi metili in posizione O6 nel dominio duplex a forcina (0, 1 e 2). I nanoswitch T, T1M, T2M sono anche marcati con una coppia FRET (Cy3 e Cy5). La presenza di anche un solo gruppo metile mina fortemente la formazione del triplex e di conseguenza, si osserva un segnale FRET che è coerente con una struttura triplex che non presenta gruppi metilici (controllo). Questi primi risultati hanno dimostrato la potenzialità di utilizzare questi nanoswitch triplex

come substrato adatto per il monitoraggio dell'attività metiltransferasica degli AGT (figura 19). Infatti, la rimozione enzimatica del gruppo metilico di O6-MG nei nanoswitch ristabilirebbe le condizioni ottimali per le interazioni di Hoogsteen e la formazione del triplex da parte del nanoswitch. Ciò è stato ottenuto con una serie di esperimenti. Utilizzando un derivato della fluoresceina di O6-benzil-guanina (BG-VG) ho confermato che T1M e T2M sono substrati naturali delle AGT. Per prima cosa, ho dimostrato che non si verifica alcuna interazione aspecifica tra il nanoswitch e il BG-VG in assenza della proteina. In seguito diverse AGT sono state pre-incubate con T, T1M e T2M e in seguito è stato aggiunto una quantità equimolare di BG-VG, che avrebbe dovuto reagire con gli enzimi liberi nella soluzione. Infatti nel caso di MGMT, è stata osservata una banda proteica fluorescente in presenza di T, indicando che l'enzima ha reagito con il BG-VG dimostrando che non si è verificata reazione di riparazione dato che questo oligonucleotide di DNA non presenta alcun metile. Al contrario, la reazione con il T1M e il T2M ha mostrato una completa assenza della banda fluorescente e sta a indicare che l'MGMT ha rimosso il gruppo metilico mediante la sua reazione irreversibile e di conseguenza per sua natura non è più disponibile per il BG-VG fluorescente, dimostrando così che i nanoswitch metilati sono substrati efficaci per MGMT (figura 20b). Gli stessi risultati sono stati ulteriormente confermati utilizzando l'enzima SsOGT e, principalmente, l'omologo di E. coli (Ada-C). In quest'ultimo caso, è degno di nota indicare che Ada-C è completamente insensibile ai derivati O6-BG, ostacolando qualsiasi analisi indiretta fluorescente con questo tipo di substrati. Infatti, l'assenza di bande fluorescenti, nonostante il corretto caricamento confermato dalla colorazione con coomassie, ha dimostrato la sua totale insensibilità a BG-VG. Quindi questi innovativi triplex di DNA potrebbero essere proposti come substrati di AGT universali.

I nanoswitch triplex del DNA metilato sono anche altamente specifici in quanto non hanno mostrato attività enzimatica quando incubati con H5, che in precedenza era stato segnalato come cataliticamente attivo sui derivati O6-BG, ma incapaci di legarsi e reagire con il DNA a doppio filamento (figura 20b). È stato quindi testato se l'attività di riparazione della AGT sui nostri nanoswitch potesse comportare uno switch conformazionale e un conseguente cambiamento misurabile nel segnale FRET (figura 21). L'MGMT

rimuove enzimaticamente il gruppo metile nella posizione O6 negli nanoswitch triplex e ripristina la loro capacità di formare una struttura triplex chiusa. Per dimostrare la possibilità di utilizzare questa piattaforma in terreni più complessi, abbiamo anche eseguito la misurazione dell'attività enzimatica di hMGMT nel 10% di siero e abbiamo osservato cambiamenti del segnale FRET ben distinguibili dall'esperimento di controllo in assenza di enzima. Questo approccio innovativo, quindi potrebbe essere utile per la determinazione del contenuto di <sup>m</sup>DNA, nonché per lo studio e la caratterizzazione di nuovi inibitori di MGMT come possibili farmaci candidati. In questo caso, questi nanoswitch sono stati testati per misurare l'attività dell'enzima umano in presenza di inibitori, inattivatori o pseudo-substrati (figura 22) . La forza di questo innovativo substrato è che fornisce un dosaggio diretto e in tempo reale; veloce, che consente di misurare l'attività AGT in pochi minuti.

Questi risultati potrebbero essere un punto di partenza per progettare razionalmente altri nanoswitch di DNA e essere utilizzati come strumenti di applicazione nella misurazione e determinazione di un'ampia gamma di enzimi coinvolti in altre attività di riparazione del DNA.

## Abstract

Cellular DNA is constantly subjected to covalent modifications by intracellular chemical compounds and coming from the external environment. The most dangerous types of molecular changes include the formation of covalent bonds between the strands of the DNA, or their breakage, and the loss or the bases' modification. Among these, alkylating agents are very reactive molecules that bind chemical groups in nucleic acids, causing alterations in their functions. The presence of alkylating agents in industrial waste, as well as in combustion and food products, is one of the main problems for the environment. The determination and the measure of damaged DNA is really important for: the identification and/or the optimization of DNA protective molecules against the action of alkylating agents, in cosmetic and pharmaceutical productions; the assessment of environmental pollution; the analysis of the quality of production, packaging and storage of food. Among the DNA modifications by alkylating agents, the most mutagenic adduct introduced into DNA by methylating agents is the  $O^6$ -MG. This modification alters during DNA replication with thymine leading to the GC  $\rightarrow$  AT transition, which could be fixed in the genome. The strategy used by nature to repair alkylation damage is the direct transfer of these alkyl groups to the active site of the  $O^6$ -alkylguanine-DNA alkyl-transferase (AGTs). AGTs are highly interesting not only for investigating critical biological processes, such as DNA repair, but also for the development of new and simple, intuitive and inexpensive assays to optimize inhibitory molecules of therapeutic and environmental interest. The aim of this project is the development of innovative analytical methods for the determination of the  $O^6$ -alkyl-guanine nucleobases ( $O^6$ -AGs) content in DNA samples from different sources, locations and cell treatments (mDNA). In general, this project proposes an innovative, rapid and reproducible method to identify particular DNA damages. The goal was pursued through two different approaches, the first based on the employing of particular enzymes that, unlike the complex repair mechanisms, are able to repair DNA with a single irreversible; and the second on the use of suitably modified oligonucleotides. The bottleneck of this assay is the need to perform the reaction at moderately high temperatures (ca. 60 °C), which avoids any endogenous AGTs' activities still present in DNA samples.

## Abbreviations

<b>6-4 PPs</b>	pyrimidone dimers of (6-4) pyrimidines
<b>AAF</b>	acetylated derivative N-acetyl 2 aminofluorene
<b>AF</b>	2 amino-fluorene
<b>AGT</b>	O <sup>6</sup> -alkylguanine-DNA-alkyl-transferases
<b>BER</b>	base excision repair
<b>BG-F</b>	fluorescent substrate
<b>c-PAHs</b>	polycyclic aromatic hydrocarbons
<b>C-Term</b>	C terminal domain
<b>cPDs</b>	cyclo-butane pyrimidines
<b>ds<sup>m</sup>DNA</b>	double stranded alkylated DNA oligonucleotide
<b>EMS</b>	ethyl methane sulfonate
<b>epPCR</b>	error prone PCR
<b>ETC</b>	electron transport chain
<b>FRET</b>	Forster Resonance Energy transfer
<b>GFP</b>	green fluorescent protein
<b>HTH</b>	helix -turn-DNA binding motif
<b>IR</b>	ionizing radiation
<b><sup>m</sup>DNA</b>	methylated DNA
<b>MGMT</b>	human AGT
<b>MMR</b>	misalignment repair
<b>MMS</b>	methyl methane sulfonate
<b>MNNG</b>	N-methyl-N'-nitro-N-nitrosoguanidine
<b>MNU</b>	methyl-nitroso-urea
<b>N-Term</b>	N terminal domain
<b>NER</b>	nucleotide excision repair
<b>O<sup>6</sup>-BG</b>	O <sup>6</sup> -benzyl-guanine
<b>O<sup>6</sup>-BGT</b>	O <sup>6</sup> -[4-bromothienyl]-guanine
<b>O<sup>6</sup>-MG</b>	O <sup>6</sup> -methyl-guanine
<b>O<sup>6</sup>AG</b>	O <sup>6</sup> -alkyl-guanine
<b>OH</b>	reactive hydroxyl radical
<b>PAH</b>	polycyclic aromatic hydrocarbon
<b>ROS</b>	reactive oxygen species
<b>scFV-SNAP</b>	anti-body fragments
<b>SPR</b>	surface plasmon resonance
<b>ssDNA</b>	single strand DNA
<b>T</b>	triplex
<b>T1M</b>	triplex with one O <sup>6</sup> MG
<b>T2M</b>	triplex with two O <sup>6</sup> MG
<b>TBA</b>	thrombin DNA aptamer
<b>TMZ</b>	Temozolomide

# Outline of the thesis

The overall aim of the present Thesis project is the development of innovative analytical methods for the determination of the O6-alkyl-guanine nucleobases (O6-AGs) content in DNA samples from different sources (water, environmental), locations (cosmetic industry) and cell treatments (chemotherapy).

Following these **Chapter 1** and **2** comprising a general introduction to the experimental part and the general purposes of thesis, **Chapters 3 - 4** will report a detailed experimental characterization of analytical methods based on thermostable chimeras and a production of a new AGT substrate based on a particular DNA structures (DNA Triplex).

A summary of the main chapters is reported below:

- **Chapter 3** describes the Construction, development and characterization of a thermostable chimera for the rapid determination of O6-AGs in a high-throughput assay.

- **Chapter 4** describes the design and characterization of a single-stranded DNA oligonucleotide-based DNA Triplex for the evaluation of methyltransferase activity (AGT).

# **CHAPTER 1**

## **General Introduction**

# General Introduction

## *Pollution and environmental safety and health*

Epidemiological studies conducted in metropolitan areas have consistently shown that pollution is one of the most serious global problems for public health and environmental safety. Respirable environmental particulate matter of aerodynamic diameter  $<10\ \mu\text{m}$  (PM<sub>10</sub>) comprises a complex mixture consisting in a large number of chemicals, many of which are toxic, including carcinogenic polycyclic aromatic hydrocarbons (c-PAHs) and inhalable particles, such as nitrogen oxides, hydrocarbons and carbon monoxide from vehicle exhausts. PM<sub>10</sub> can be higher along roads and can cause DNA damage to exposed organisms [Pope *et al.*, 1995; Risom *et al.*, 2005]. Many diseases caused by DNA mutations have recently been recognized as being related to toxic substances in the air where air pollution is severe. DNA damage research focuses primarily on the influence of specific pollutants on certain species or the effect of environmental pollution on humans.

Water pollution has also a strong impact on environmental safety and health, leading to the lack of safe drinking water for millions of people, the transmission of various diseases and thousands of deaths every day around the world [McDonald *et al.*, 2011]. In addition to natural phenomena, anthropogenic contaminants are the main sources of water pollution [Leeuwen *et al.*, 2000]. Every year, millions of tons of industrial and residential hazardous waste, including heavy metal ions, alkylating agents, anions, pesticides, and so on, are discharged into the environment, resulting in a wide range of contaminated water areas as a result of bioaccumulation in organisms of living species through the food chain. Therefore, continuous monitoring and management of water and environmental resources are necessary to maintain human health and environmental safety [Leeuwen *et al.*, 2000].

## *Genomic stability and DNA damages*

Genomic stability is one of fundamental points for survival organism and for the continuity and transmission of genetic heritage. If mutagenesis contributes to cancer, some human diseases and aging, on the other side it plays an extremely important role in the evolution and diversity of the genomic

sequence [Lindahl, 1993; Friedberg, 2003]. DNA, the basic unit of heredity, is known to be an intrinsically reactive molecule and is highly vulnerable to chemical modifications by endogenous and exogenous agents (Figure 2). Most endogenous DNA mutations result from errors of the replicative machinery, whereas damages are from spontaneous base deamination, a-basic sites, and DNA methylation. Hydrolytic and oxidative reactions with water and reactive oxygen species (ROS), typical by-products of the electron transport chain (ETC) during cellular respiration in aerobic organisms, are also derived from catabolic oxidases, anabolic processes and peroxisomal metabolism. The latter feed the development of hereditary diseases and sporadic tumours [Visconti and Grieco, 2009; Reuter *et al.*, 2010; Perrone *et al.*, 2016].

However, most DNA damage exogenous results from:

- Ionizing radiation (IR) (alpha, beta, gamma, neutrons and X-rays), is abundant in our environment, being developed from different sources ranging from rocks, soil, and radon, to cosmic radiation and medical devices. Ionizing radiation can damage the DNA either directly, or by indirect means, such as by radiolysis of the surrounding water to generate a cluster of highly reactive hydroxyl radicals ( $\bullet\text{OH}$ ) [Friedberg *et al.*, 2005; Omar *et al.*, 2015]. The presence of oxygen and other reactive species increases the formation of other DNA-reactive free radicals by IR [Wardman, 2009] and indirect DNA damage from ( $\bullet\text{OH}$ ) radicals accounts to about 65 % of the radiation-induced DNA damage [Vignard *et al.*, 2013]. Apart from causing base lesions, ionizing radiation also causes double strand break, formed from multiple damaged sites closely positioned on both DNA strands [Hutchinson, 1985; Iliakis, 1991] and single strand breaks with a unique signature, where the DNA breaks have 3' phosphate or 3'-phosphoglycolate ends rather than 3'-OH ends. [Henner *et al.*, 1982; 1983; Obe *et al.*, 1992].
- UV radiation is the principal cause of skin cancer in humans [Davies, 1995; Kiefer, 2007]. It is classified into three classes based on the wavelength range: UV-C (190–290 nm), UV-B (290–320 nm) and UV-A (320–400 nm). DNA absorbs the maximum UV radiation at 260 nm, beyond which photo absorption decreases dramatically. The effects of UV rays on matter extend in two ways: if the UV is absorbable, the molecules present in the matter are excited leading to their

photochemical alteration; if not, the transfer of energy from nearby molecules, called photosensitizers, indirectly affects matter. UV rays damage DNA both ways. Several studies have shown that UV-C damages DNA primarily by causing covalent bonds between two adjacent pyrimidines, leading to photoproducts, such as cyclo-butane pyrimidines (CPDs) and pyrimidone dimers of (6-4)pyrimidines ((6-4) PPs). These bulky dimers distort the helix, requiring TLS polymerase for replication beyond them, thus contributing to mutagenicity. If these lesions are not repaired or bypassed, they cause cytotoxicity. Other the exogenous chemical agents cause dangerous damage to DNA, as aromatic amines, Polycyclic aromatic hydrocarbon (PAH), environmental stresses and alkylating agents, on which we will focus our attention.

### *Aromatic amines*

Aromatic amines are primarily developed from cigarette smoke, fuel, coal, industrial dyes, pesticides and everyday high temperature cooking [Sugimura, 1986; Skipper *et al.*, 2010]. These compounds are activated by the P<sub>450</sub> monooxygenase system and converted into the carcinogenic (ester and sulfate) alkylating agents, that, in turn, attack the C<sup>8</sup> position of guanine [Hammons *et al.*, 1997; Naegeli, 1997]. Examples of aromatic amines are 2-amino-fluorene (AF) and its acetylated derivative N-acetyl-2-amino-fluorene (AAF). C<sup>8</sup>-guanine lesions formed from amino-fluorenes are known to form persistent lesions that ultimately give rise to base substitutions and frameshift mutations [Mah *et al.*, 1989; Heflich and Neft, 1994; Shibutani *et al.*, 2001]. The C<sup>8</sup>-guanine lesion can adopt two conformations on the DNA [Eckel and Krugh, 1994 a,b]: in the external conformation, it disturbs to Watson-Crick base pairing; in internal, the C<sup>8</sup>-guanine lesion and its partner cytosine are displaced into the minor groove, completely altering the geometry and acting as a very mutagenic substrate on the DNA [Kriek, 1992; Eckel and Krugh, 1994a; Eckel and Krugh, 1994b].

### *Enviromental stresses*

DNA damage in human cells can also be caused by environmental sources such as extreme heat or cold, hypoxia and oxidative stress [Gregory and Milner, 1994; Gafter-Gvili *et al.*, 2013; Luoto *et al.*, 2013; Neutelings *et al.*, 2013; Kantidze *et al.*,

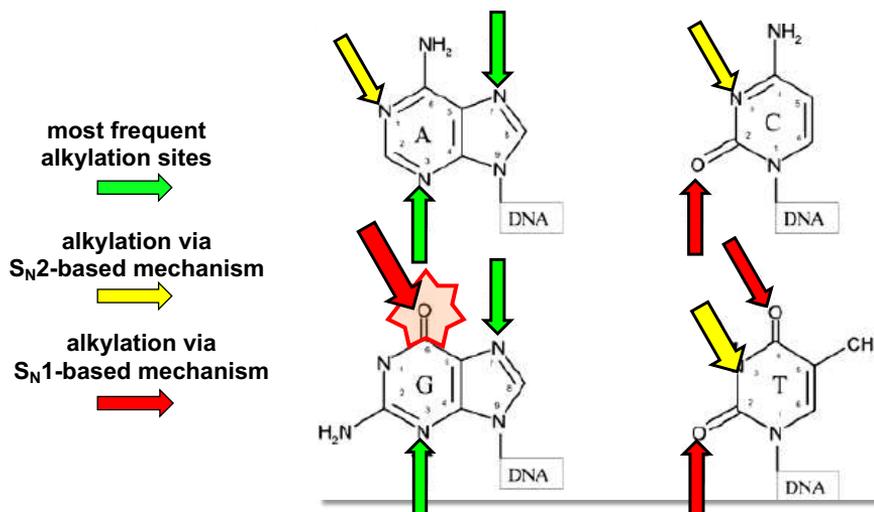
2016]. These stresses cause mutagenesis of trinucleotide repeats, which are implicated in the development of neurodegenerative disorders [Chatterjee *et al.*, 2015; 2016].

Everyday, most biological products have been increasingly related to DNA damage. For example, food preservatives, as sodium benzoate, potassium benzoate and potassium sorbate; and food additives, as citric acid, phosphoric acid, brilliant blue and sunset yellow are all known to cause DNA damage [Mamur *et al.*, 2010; Zengin *et al.*, 2011; Yilmaz *et al.*, 2014; Pandir, 2016]. Bisphenol A, found in cosmetics, pharmaceuticals, food and beverage processing, are linked to DNA damage in spermatozoa [Oishi, 2002; Meeker *et al.*, 2010a; Meeker *et al.*, 2010b; Meeker *et al.*, 2011] and have also been associated with DNA damage [Kasiotis *et al.*, 2012]. These products, and the related their consequences, underline the importance of global regulatory requirements on the use of chemicals that place human, animals and environment health in danger, as there may still be unknown chemicals that pose health risks.

### *Alkylating agents*

Exogenous alkylating agents are primarily produced from dietary components, tobacco and cigarette smoke, biomass burning, industrial processing, and for many years employed as chemotherapeutic agents [Lawley, 1966; Crutzen and Andreae, 1990]. The electrophilic alkylating agents react with intensified affinity to the highly nucleophilic base ring nitrogens, principally the  $N^7$  of guanine and  $N^3$  of adenine, and moderately with the oxygens. Examples of adducted DNA bases include modified adenine (at  $N^1$ ,  $N^3$ ,  $N^6$  and  $N^7$ ), guanine (at  $N^1$ ,  $N^2$ ,  $N^3$ ,  $N^7$  and  $O^6$ ), cytosine ( $N^3$ ,  $N^4$  and  $O^2$ ), thymine ( $N^3$ ,  $O^2$  and  $O^4$ ), and alkyl phosphates in the DNA backbone [Singer and Kusmieriek, 1982; Singer, 1986; Friedberg, 2005]. Alkylating agents can add the alkyl group by either two reactions: 1) an  $S_N1$  substitution reaction that progresses *via* the first order kinetics and involves a carbonium ion intermediate; or, 2) an  $S_N2$  substitution reaction that follows the second order kinetics, and in general produces adducts that are less mutagenic and carcinogenic than those of the  $S_N1$  pathway [Naegeli, 1997], although evidence has been presented that some  $S_N1$  alkylating agents may not proceed *via* the carbonium ion intermediate [Loechler, 1994] (Figure 1). Most common alkylating agents that are regularly used in labs, including methyl methane-sulfonate

(MMS), ethyl methane-sulfonate (EMS), N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) and methyl-nitroso-urea (MNU), react with DNA to generate mutagenic and carcinogenic lesions. In particular, MMS mainly produces the mutagenic  $N^7$ -methyl-guanine and  $N^3$ -methyl-adenine, both of which are susceptible to cleavage of the N-glycosidic bond, thereby generating AP sites, while MNNG and MNU produce  $O^6$ -methyl-guanine ( $O^6$ -MG), which pairs with T and induces G:C→A:T transversions [Loechler *et al.*, 1984; Beranek, 1990; Wyatt and Pittman, 2006].



**Figure 1:** Possible sites of base alkylation. Green arrows indicate most frequent alkylation sites; red arrows indicate sites that are most frequently alkylated by  $S_N1$  type mechanisms; yellow arrows indicate alkylated sites by  $S_N2$  type mechanisms.

Other classical examples of alkylating agents are sulfur and the nitrogen mustards, first used in World War I, and in many other conflicts since including the present-day in Syria. These agents follow  $S_N1$  reactions, and are bifunctional, in fact they carry two reactive groups, instead of one as in monofunctional alkylating agents, and consequently have the potential to react with two different sites on the DNA. Such bifunctional reactions result in DNA intra- and inter-strand crosslinks, and DNA-protein crosslinks, which block DNA metabolic activity [Lawley, 1966;]. The properties of alkylating agents have been also exploited in chemotherapy [DeVita and Chu, 2008], as the cyclophosphamide and the Temozolomide (TMZ), used in the treatment of lymphomas, leukaemia and solid tumours [Emadi *et al.*, 2009]. Another class of

chemotherapy crosslinking agents includes cisplatin, the first FDA approved platinum compound that is used to treat a wide variety of cancers [Kelland, 2007; Dasari and Tchounwou, 2014].

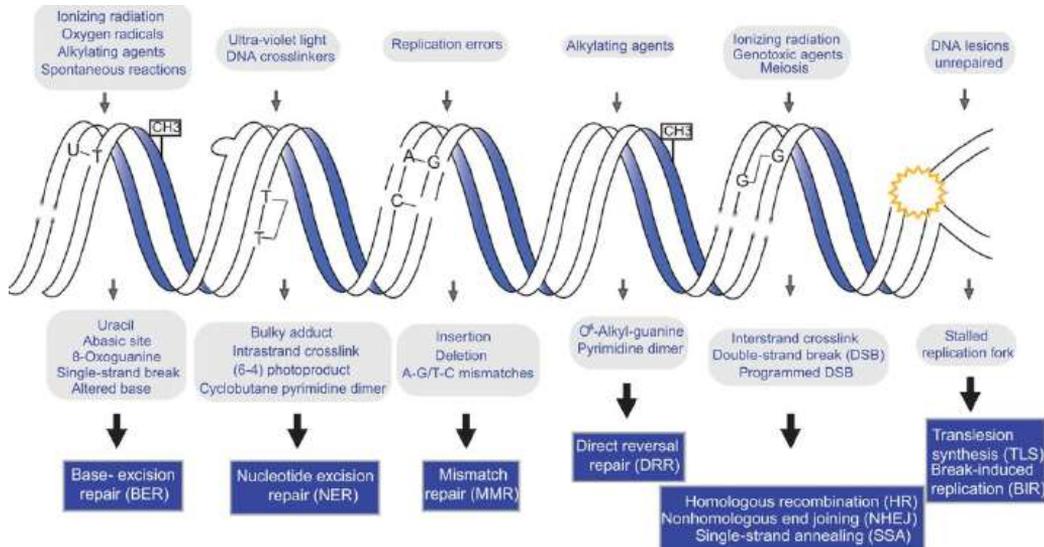
### *Mechanisms of DNA repair*

There are several approaches that cells evolutionary use to survive to various types of lesions. Indeed, they have developed complex signalling pathways to halt cell cycle progression in the presence of DNA damage, thus increasing the intervention time for repair and tolerance mechanisms, although during DNA replication and repair the introduction of potentially disadvantageous mutations into cells could occur [Zhou *et al.*, 2005]. Furthermore, in the presence of a massive genomic damage, cells start to a programmed cell death (*apoptosis*). In higher eukaryotes, there are five main translesion synthesis polymerases (TLS) - REV1, POL  $\zeta$ , POL  $\eta$ , POL  $\kappa$  and POL  $\iota$  - that bypass damages, by allowing replication and at same time integrating incorrect bases, that can be fixed as mutations in the subsequent replication cycle [Chatterjee, 2017]. DNA repair pathways include as a direct repair of certain types of lesions (discusses later in more detail), but also multi-enzymatic distinct mechanisms for the excision of damaged bases, which are called nucleotide excision repair (NER), base excision repair (BER) and misalignment repair (MMR) (Figure 2).

- NER is the pathway of choice to remove massive lesions such as PP (pyrimidine pyrimidone photoproducts) from UV radiation or damage from chemotherapeutic agents. The NER deficiency results in a number of different human syndromes: Xeroderma Pigmentosum, which is associated with a predisposition to skin cancers; Cockayne Syndrome; rare UV-Sensitive Syndrome; and Cerebro-Oculo-Facio-Skeletal syndrome [Friedberg, 2005; Vermeulen and Foustari, 2013]. NER contributes to the instability mechanisms in triplet repeat disorders [Lin *et al.*, 2006; Hubert *et al.*, 2011; Dion, 2014]. The process of NER is biochemically complicated and involves up to 30 distinct proteins in human cells that function as a large complex called *reparosome*, by excision of nucleotides. This “repair machine” facilitates the excision of damaged nucleotides by making incisions in the flanking regions and removing a fragment approximately 30 nucleotides in length.

- Damaged bases that are not recognized by NER machines are corrected by BER, whereby the bases are removed from the genome as free bases by a different set of repair enzymes. BER corrects those forms of oxidative, deamination, alkylation and a basic single base damage that are not perceived as significant distortions to the DNA helix. For BER transactions, chromatin remodelling at the DNA damage site is followed by lesion recognition by a DNA glycosylase [Odell *et al.*, 2013]. At least 11 different DNA glycosylases can recognize and excise a damaged base from undistorted helices, as well as ones flipped out from the major groove [Huffman *et al.*, 2005; Krokan and Bjoras, 2013]. In conclusion, both NER and BER use 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].
- MMR is an evolutionarily conserved repair pathway that activates post replication and contributes to replication fidelity by at least 100 times [Kunkel, 2009; Arana and Kunkel, 2010]. Typical substrates for the MMR pathway are base mismatches that occurred during replication and insertion deletion rings (IDLs) within repetitive DNA sequences that resulted from strand slip events [Friedberg, 2005; Jiricny, 2006]. This pathway is also implicated in other cellular processes including microsatellite stability, meiotic and mitotic recombination, DNA damage signalling, apoptosis, class change recombination, somatic hypermutation and triplet expansion [Jiricny, 2006; 2013; Chatterjee *et al.*, 2017]. Recent studies have shown that chromatin modifications pave the way for MMR proteins to access DNA lesion and initiate repair [Chatterjee *et al.*, 2017]. After the lesion recognition phase, the heterodimer complex that actively participates in the process moves along the DNA in an ATP-dependent manner to make way for downstream MMR components [Jiricny, 2013]. In addition to mismatch repair and other cellular functions, mismatch repair genes have recently been shown to be repressed in response to environmental stresses, such as hypoxia, benzo [a] pyrene, inflammation and even tumour microenvironment [Mihaylova *et al.*, 2003; Bindra and Glazer, 2007; Nakamura *et al.*, 2008;

Edwards *et al.*, 2009; Chen *et al.*, 2013]. It remains to be seen whether other exogenous stresses can also suppress the expression of MMR genes.



**Figure 2:** Schematic representation of the main types of damage and relative repair mechanisms.

## Direct repair and *O*<sup>6</sup>-alkyl-guanine-DNA-alkyl-transferase

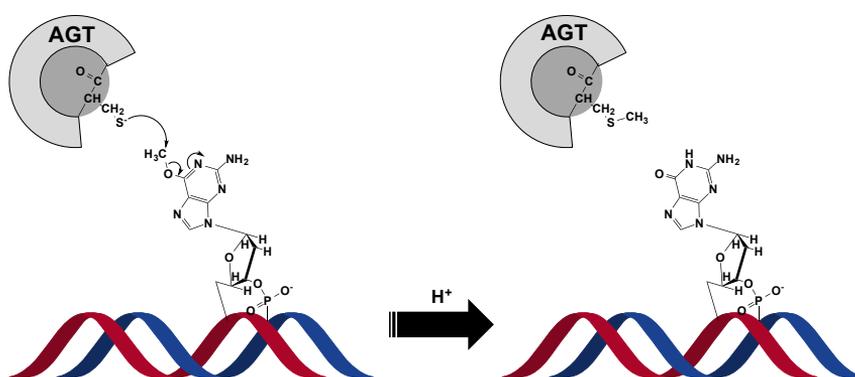
### *The irreversible reaction of AGTs*

Direct DNA repair primarily removes lesions made by intracellular and extracellular chemical compounds, which can lead to the formation of covalent bonds with the DNA bases or backbone [Singer *et al.*, 1976]. Alkylating agents for example, as mentioned above, are a group of chemical compounds that can cause these types of DNA damage. In fact, they include a group of mutagens and carcinogens that can modify DNA by alkylation.

The most mutagenic adduct introduced into DNA by methylating agents is the *O*<sup>6</sup>-MG. This modification alters during DNA replication with thymine leading to the GC → AT transition, which could be fixed in the genome [Falnes *et al.*, 2003]. The most common strategy used by nature to repair alkylation damage is the direct transfer of these alkyl groups to the active site of the *O*<sup>6</sup>-alkylguanine-DNA alkyl-transferases (AGT, OGT or MGMT; E.C. 2.1.1.63).

AGTs are evolutionarily conserved, representing the main cellular mechanism responsible for the repair of DNA containing  $O^6$ -AGs. These enzymes show a peculiar reaction mechanism, in which the transfer of the alkyl group from the damaged guanine to the catalytic cysteine in the active site occurs in an *irreversible*  $S_N2$  single step mechanism. Upon alkylation, the repaired DNA is released, whereas the enzyme is no longer available for the subsequent transfer reaction [Pegg, 2011; Mattosovich *et al.*, 2020] (Figure 3 and Figure 6, *blue path*). These proteins are thus called “suicide proteins”, showing a 1:1 enzyme : substrate. Finally, their inactivation and the consequent destabilization trigger the recognition by cell systems and the degradation by the proteasome [Sri venugopal *et al.*, 1996, Xu-Welliver *et al.*, 2002]

AGTs are small enzymes (17-22 kDa) widely present in Bacteria, Archaea and Eukaryotes but apparently absent from plants, *Schizosaccharomyces pombe*, *Thermus thermophilus* and *Deinococcus radiodurans*. The first AGT to be discovered and characterized is the Ada protein of *E. coli* [Olsson *et al.*, 1980], followed by the human AGT (MGMT) described by the Pegg Group in 1991 [Pegg *et al.*, 1991]. Later, AGTs were characterized by two important model organisms such as *Drosophila melanogaster* [Kooistra *et al.*, 1999] and *Caenorhabditis elegans* [Kanugula *et al.*, 2001]. Importantly, it has been found in both thermophilic bacteria and hyperthermophilic Archaea, such as *Aquifex aeolicus* and *Archaeoglobus fulgidus* [Kanugula *et al.*, 2003].



**Figure 3:** the *irreversible* reaction of AGTs.

Despite the primary structures of AGTs analysed in Figure 4 are more different, all enzymes of this class show a typical

architecture, consisting of two domains separated by a *connecting loop* [Fang *et al.*, 2005]. The N-terminal domain (Nterm), which is very different among AGTs, has a role not well defined, likely involved in regulation, cooperative binding and stability [Daniel *et al.*,2000; Miggiano *et al.*,2013; Perugino *et al.*,2015]. On the contrary, the highly conserved C-terminal domain (Cterm; Figure 4b) contains all the elements responsible of the DNA repair activity: *i*) the helix-turn-helix DNA binding motif (HTH); *ii*) the *Asn hinge*, which precedes the extremely conserved amino acid sequence -V / IPCHRVV / I- containing the catalytic cysteine (except *Caenorhabditis elegans* AGT-2 which has the sequence -PCHP- [Kanugula *et al.*,2001;Serpe *et al.*, 2019]; *iii*) and the *active site cycle*, responsible for the specificity of the substrate.

### *Role of MGMT in chemotherapy*

Different living conditions take the action of DNA damage by alkylating agents and consequently the presence of AGT in most living organisms protects cells to killed by these agents. MGMT has a dual role: it protects healthy cells from the genotoxic effects of alkylating agents, but, at the same time, it protects cancer cells from lethal effect of the alkylating agents based chemotherapics.

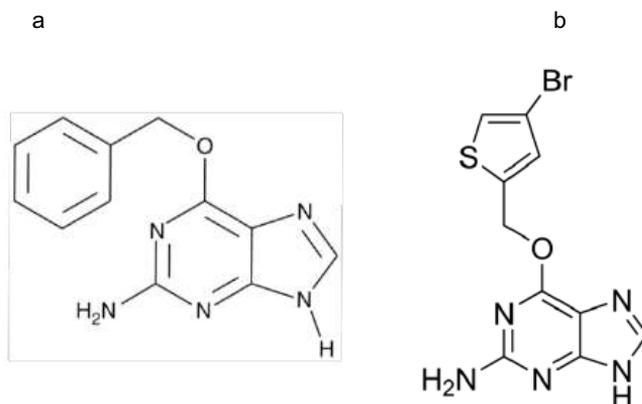
For these reason, MGMT is a target to be inactivated in chemotherapy protocols based on these agents [Gerson *et al.*,2004; Sabharwa *et al.*,2006; Zhong *et al.*,2010]. Among alkylating agents, the active intermediate of TMZ [Tintorè *et al.*, 2009] interacts with DNA and generates methylated adducts, including  $N^7$ -MG  $N^3$ -MA and  $O^6$ -MG. The latter lesion, although produced in a low percentage (~ 8%), is generally considered to be the most toxic and mutagenic of those produced by methylating agents.

The cytotoxicity is strongly influenced in human cells by the enzymatic levels of MGMT: the number of repaired  $O^6$ -methyl adducts depends on the number of MGMT molecules present in the cell and the speed of synthesis of the enzyme from scratch. When the cell expresses low concentrations of MGMT,  $O^6$ -MG adducts from the action of methylating agents are not repaired and the cell starts the *apoptosis*. Indeed, if methyl adducts in the  $O^6$  position of guanine are not removed from the MGMT, altered interactions with cytosine or thymine occur, resulting in activation of the MMR.



However, MMR is unable to complete the repair because it can remove cytosines or thymines, but not  $O^6$ -MGs. So, the DNA polymerase cannot find the complementary base to  $O^6$ -MG, and there causes a repeated activation of MMR, leading to single and double strand breaks, blocking growth with consequent induction of the apoptotic process. On the contrary, high levels of MGMT in the tumor cells interrupt this futile cycle, reducing the effectiveness of alkylating drugs [Mattosovich *et al.*2020]. The antagonistic action between the effect of alkylating chemotherapies and the cell protection by MGMT is one of the main reasons why some cancer cell lines have resistance to these types of drugs.

Resistance to chemotherapy can be decreased by the development of inhibitors/inactivators for these enzymes: as described above, after removing the lesion, the alkylated form of the protein is inactivated and enters the intracellular degradation pathways. In last decades, some pseudo-substrates of the MGMT were designed, namely,  $O^6$ -benzyl-guanine ( $O^6$ -BG) and the strong inactivator  $O^6$ -[4-bromothienyl]-guanine ( $O^6$ -BTG, Lomeguatrib) were used in combination with alkylating agents (Figure 5) [Khan *et al.*, 2007; Khaina *et al.*, 2019].



**Figure 5:** the chemical structure of  $O^6$ -BG and Lomeguatrib.

These elements mimic DNA-damaged guanine and react with the protein by covalent transfer of the alkyl adduct to the cysteine residue of the active site, thereby irreversibly inactivating the enzyme (Figure 6, *red path*). At therapeutic levels,  $O^6$ -BG is not

toxic alone, but efficiently provides tumor cells more sensitive to alkylating agents. This confirms the potential therapeutic effect of this inhibitor 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  $O^6$ -BGs 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 tumours [Luu et al., 2002; Mattossovich et al., 2020].

### *AGTs as biotechnological tool.*

In the study of protein function, specific labelling with synthetic probes has been a major advance in knowledge. One way to achieve such a labelling is through expression of the protein of interest as a fusion protein with an additional polypeptide, called *tag* [Keppler et al., 2003; Kindermann et al., 2003; Gronemeyer et al., 2006; Hinner et al., 2010]. For example, *tags* allow to simplify and optimize purification methods (*affinity tags*), detection procedures (by using specific antibodies) and allow following the proteins *in vivo* [Johnsson et al., 2003]. However, most factors 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.

The *Aequora victoria* green fluorescent protein (GFP), that it was the first development of a *tag*, allows *in vivo* localization of fusion proteins in cellular and molecular biology [Tsien, 1998]. However, it has some disadvantages, as the relative big dimensions, the insensibility to all possible changes of the cellular environment (in terms of pH, hydrophobicity and ionic concentrations), the incompatibility in all applications concerning anaerobic conditions, and the general use restricted only in mesophiles and in mild reaction conditions, although few examples of thermostable GFP variants were applied in thermophilic microorganisms [Cava et al., 2008].

The particular and unusual covalent linkage established between the AGT protein and its inhibitors/inactivators has been exploited

for biological applications, ranging from the specific quantification of MGMT to using in protein fusions, for *in vivo* fluorescent labelling. This is possible because the high tolerance of MGMT toward various large alkyl-groups on the O<sup>6</sup>-position of guanines, without neglecting the advantage of their small protein size. Then, MGMT has been exploited for other applications. In 2003, Kai Johnsson's group removed the DNA binding ability of MGMT, abolishing its DNA repair activity, pioneering MGMT as *in vitro* and *in vivo* biotechnology tool, which led to its commercialization, namely SNAP-tag<sup>®</sup> and CLIP-tag<sup>®</sup> (New England Biolabs) [Juillerat *et al.*, 2003;Mattossovich *et al.*,2020] (Figure 6, *green path*).

The expression of the fusion protein within the cells followed by incubation with suitable fluorescent derivatives leads to the *in vivo* labelling of the fusion proteins with the probe, which can be used for localization studies [Juillerat *et al.*, 2003]. The same principle has also been used for the *in vitro* immobilization of labelled fusion proteins [Hinner *et al.*, 2010]. This provides a gentle condition to better fix and target a wide range of proteins / enzymes on a surface. SNAP-tag<sup>®</sup> technology has been successfully applied to surface plasmon resonance (SPR) for the covalent immobilization of proteins of interest [Huber *et al.*, 2004].

Another interesting application of this *protein-tag* is the possibility of producing new antibody fragments (scFv-SNAP) to be used in the SPR analysis [Niesen *et al.*,2016].

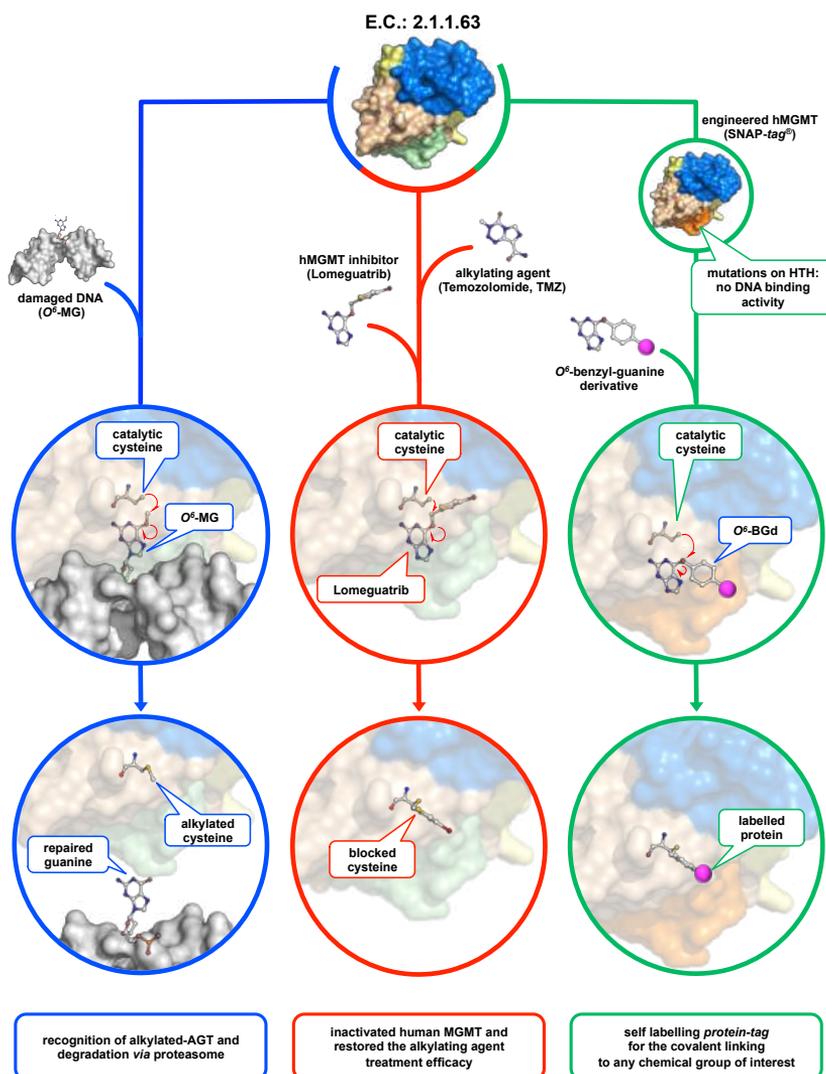
Despite the need to use a specific substrate, SNAP-tag has infinite applications: the possibility of covalently binding a desired chemical group (conjugated to O<sup>6</sup>-BG) to a protein of interest (genetically fused to it) makes it very advantageous, compared to traditional protein labels currently in use.

However, being originated by the MGMT, a limitation of using these mesophilic *tags* is, the application to extremophilic organisms and/or to harsh reaction conditions.

### *Thermostable AGTs*

As defined in the previous paragraphs, endogenous alkylating agents and environmental stresses such as high temperatures accelerate the alkylation process in the genome of thermophilic and hyperthermophilic organisms, leading to DNA breaks [Valenti *et al.*, 2006] and the formation of DNA alkylation products

[Kanugula *et al.*,2003]. Therefore, the presence of AGT and methylpurine glycosylase in hyperthermophilic organisms implies that they are naturally exposed to endogenous methylating agents [34], thus supporting the crucial role of AGTs [Leclere *et al.*,1998; Skorvaga *et al.*,1998].

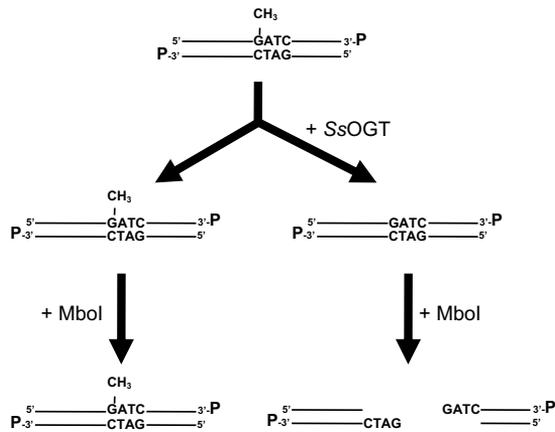


**Figure 6:** The AGT world. These enzymes are small enzymes composed of an Nterm domain (*in sky blue*), a Cterm domain (*in light brown*) connected by a loop (*in yellow*). In the Cterm, a helix-turn-helix motif (*in light green or orange* in the SNAP-tag<sup>®</sup>) is responsible for DNA binding activity. The peculiar irreversible reaction mechanism of these enzymes plays a fundamental role in the physiological repair of DNA (*blue path*), and has important repercussions in the treatment of cancer cells (*red path*) and in biotechnological applications (*green path*). Atoms are coloured by the CPK convention.

In addition to some Archaea studies using cell-free extracts, there are also some examples of biochemical studies of AGT from hyperthermophilic sources, such as enzymes from *Pyrococcus* sp. KOD1 [Leclere *et al.*, 1998; Mattossovich *et al.*, 2020] conducted by Imanaka and collaborators and enzymes of *Aquifex aeolicus* and *Archaeoglobus fulgidus* performed by Prof. Pegg's group in 2003. The latter, unfortunately, led to formation of inclusion bodies upon heterologous expression in *E. coli* [Kanugula *et al.*, 2003]. In the last decade, Perugino and co-workers introduced other thermostable OGTs from bacterial and archaea sources [Perugino *et al.*, 2012; Mattossovich *et al.*, 2020], taking advantage from their exceptional stability to shed light on the conformational changes upon the alkylation reaction [Perugino *et al.*, 2015; Morrone *et al.*, 2017], as well as to introduce innovative *protein-tags* to expand the SNAP-tag<sup>®</sup> technology [Vettone *et al.*, 2016; Visone *et al.*, 2017; Merlo *et al.*, 2019; Del Prete *et al.*, 2019; Merlo *et al.*, 2020].

### *Methods to measure the activity of AGTs*

AGTs are highly interesting not only for investigating critical biological processes, such as DNA repair, but also for the development of new and simple, intuitive and inexpensive assays to optimize inhibitory molecules of therapeutic interest and for identification and/or the optimization of DNA protective compounds against the action of alkylating agents (such as sulphur-containing antioxidants or sulfhydryl groups), in cosmetic and pharmaceutical productions [Trendowski *et al.*, 2015]. To date, AGT's activity can be measured with different methodologies. The first assay developed was based on the use of oligonucleotides possessing <sup>3</sup>H or <sup>14</sup>C-labelled O<sup>6</sup>-AGs groups. A proteolytic digestion of the proteins was then performed with proteinase K and, by an automatic amino acid analyser, the levels of radioactive S-methylcysteine in the lysate were measured [Olsson *et al.*, 1980]. In another system, a very similar but simpler and faster radioactive assay involved a terminal oligonucleotide flagged with <sup>32</sup>P isotope and an internal modified guanine positioned into a sequence for a methylation-sensitive restriction enzyme (such as *Mbo* I). In this way it was possible to investigate in which sample the methyl was removed and therefore to identify where the activity was present [Wu *et al.*, 1987]. A similar procedure was also used by Ciaramella's group to identify the activity of SsOGT for the first time [Perugino *et al.*, 2012] (Figure 7).



**Figure 7:** The determination of the activity of AGTs by a methylation-sensitive DNA digestion assay.

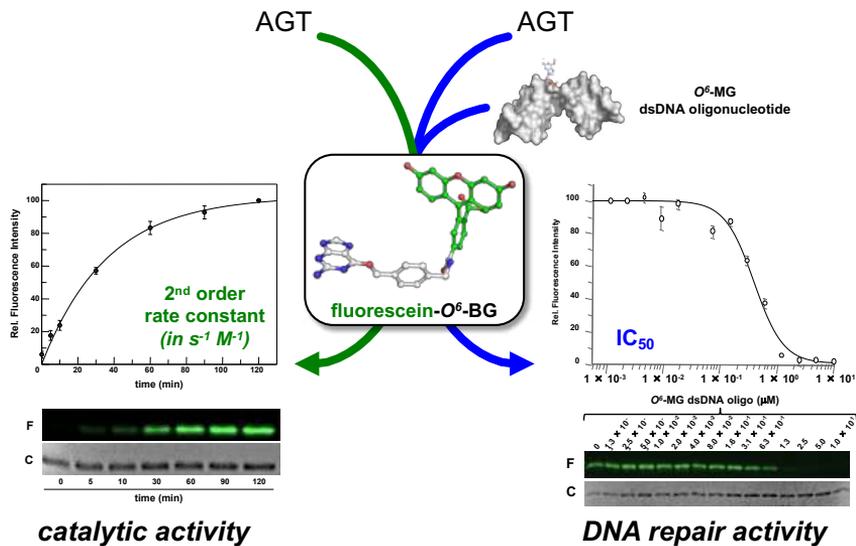
This test has the advantage of analysing the digested fragment directly by polyacrylamide gel electrophoresis [Klein et al.,1992]. Similarly, the Luu's group developed analysis of hMGMT reaction products in 2002, based on HPLC separation. This assay studied the degree of inhibition of oligonucleotides having  $O^6$ -MG or  $O^6$ -BG in different positions, which ranged from 3'- to 5'-end 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 tedious, laborious and unsafe methods.

An alternative approach was proposed in 2010 by Fàbrega's group, which developed an assay based on thrombin DNA aptamer (TBA), a single-stranded 15 mer DNA oligonucleotide, identified by Systematic Evolution of Ligands by Exponential Enrichment (SELEX), which in its G-quadruplex form binds to the thrombin protease with high specificity [Bock et al.,1992]. In this test, they put a fluorophore and a quencher at the TBA: the quadruplex structure of this oligonucleotide is compromised if one guanine of central four is an  $O^6$ -MG, preventing the formation of the G-quadruplex and, consequently, the two probes from staying closer together. The repair activity of an AGT on the oligonucleotide allows the folding of the quadruplex structure and the Förster

Resonance Energy Transfer (FRET) takes place, with a consequent decrease in the intensity of the fluorescence. This approach uses a single strand DNA: in this case, it is not the natural substrate of every AGTs (methylated DNA, <sup>m</sup>DNA).

The introduction of fluorescent derivatives of O<sup>6</sup>-BG (New England Biolabs) has made it possible to develop a new DNA alkyl transferase assay. Since AGTs covalently bind the benzyl-fluorophore portion of its substrate after the reaction, it is possible to immediately load the protein product onto an SDS-PAGE: fluorescence intensity from *gel-imaging* analysis provides a direct measure of activity of protein due to the 1:1 stoichiometry of protein : substrate (Figure 8). The signals of the fluorescent protein (after correction of the amount of loaded protein by Coomassie staining) obtained at different times are plotted by fitting a second order reaction rate [Miggiano *et al.*, 2013, 2016; Perugino *et al.*, 2012, 2015; Vettone *et al.*, 2016; Morrone *et al.*, 2017]. This fluorescent assay is very fast, cheap, reproducible, and allows a measure of the activity of AGTs on their natural substrate (<sup>m</sup>DNA): briefly, a double-stranded alkylated DNA oligonucleotide (ds<sup>m</sup>DNA) can be included in competition with the fluorescent substrate, leading to a decrease in fluorescence intensity in a manner directly proportional to the concentration of <sup>m</sup>DNA, allowing the determination of an IC<sub>50</sub> value (Figure 8) [Miggiano *et al.*, 2013, 2016; Perugino *et al.*, 2012, 2015; Vettone *et al.*, 2016; Morrone *et al.*, 2017]. However, this *end-point* method is indirect and can be applied to all AGTs sensitive to the O<sup>6</sup>-BG derivatives, with the exception, for example, of *E. coli* Ada-C [Elder *et al.*, 1994, Goodtzoaet *et al.*, 1997].



**Figure 8:** Innovative fluorescent AGT assay. The substrate could be used alone for the determination of the AGT catalytic activity, or in combination with a competitive non-fluorescent substrate (<sup>m</sup>DNA). 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 SsOGT reaction. In the latter case, an indirect measure of the DNA repair activity on natural substrates is determined. Competitive assay employing the fluorescent substrate in the presence of double strands (ds) oligonucleotides pairs (natural substrate), containing a single O<sup>6</sup>-methyl-guanine, was performed as described [Perugino *et al.* 2015] to determine the half maximal inhibitory concentration (IC<sub>50</sub>), that is the concentration of methylated DNA needed to reduce the fluorescence intensity of the OGT band by 50.0%

## **CHAPTER 2**

# **PURPOSES OF THE THESIS**



## *Analytical methods based on thermostable chimeras*

The starting point of my PhD project is the production of thermostable chimeras (Fig. 9a), composed by two moieties: the first is a particular enzyme (AGT, OGT, or MGMT) which recognizes and repairs O<sup>6</sup>-AG in the DNA sample. The peculiar irreversible reaction of this biocatalyst leads to an *alkylated enzyme*, which is no longer available to, again, irreversibly react with a fluorescent substrate (BG-F) also presents in the mixture. This will result in the reduction of the fluorescence signal in a directly proportional way to the amount of <sup>m</sup>DNA in the sample. With an appropriate calibration line (constructed by known amounts of <sup>m</sup>DNA), the degree of DNA alkylation can be determined. The other moiety of the chimera is the commercially available Halotag<sup>®</sup>, a modified halo-alkane dehydrogenase, which covalently binds to chloro-alkane substrate derivatives. Using a support opportunely derivatized with the latter, the chimera could be directly and specifically immobilized by the activity of this enzyme, and the AGT fluorescence can be easily separated from the context (crude cell extracts or other contaminants in the samples) and measured.

The bottleneck of this assay is the need to perform the reaction at moderately high temperatures (ca. 60 °C), which avoids any endogenous AGTs' activities still present in DNA samples. For these reasons, chimeras are opportunely engineered to be as active as stable at those temperatures.

## *Production of a new AGT substrate based on a DNA*

### *Triplex*

The second approach is focused on the characterization of an innovative *AGT assay, based on a DNA-Triplex*. Recently, single strand DNAs (ssDNAs) forming a Triplex structure have been further introduced into DNA nanotechnology as sequence-specific targeting agents or nanodevices [Amodio *et al.*, 2014]. A DNA Triplex consists of a normal DNA duplex *via* Watson–Crick base-pairing, and a third strand binding to the duplex in the major groove *via* Hoogsteen hydrogen bonds. While the T–A:T triad can readily form in aqueous solutions, the C–G:C<sup>+</sup> triad only when the cytosine nucleobase on the third strand is protonated at its N<sup>3</sup>

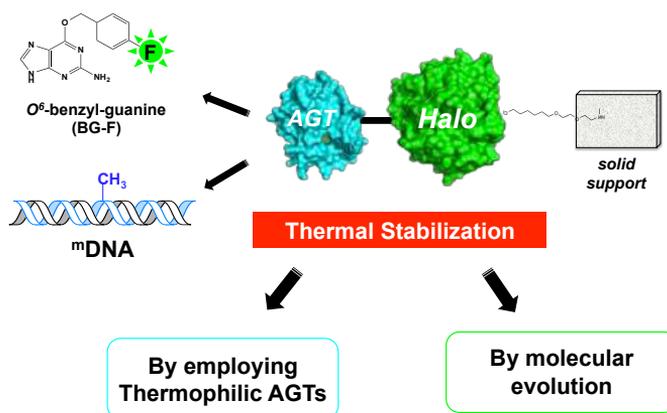
position occurs, allowing an Hoogsteen hydrogen-bond with a guanosine [Liu *et al.*, 2014]. For this reason, ssDNAs containing cytosines can only form a triplex structure at acidic pHs. The conformational switches between the duplex and the triplex form could be monitored through a FRET pair, e.g., a Cy3 and a Cy5, located in an internal position and at the 5' end respectively of the ssDNA [Amodio *et al.*, 2014] (Figure 9b). We noted that the presence of  $O^6$ -MGs in the oligonucleotide hampers the switching to the Triplex form. Thus, this methylated oligonucleotide should be the ideal fluorescent substrate for an innovative AGT's assay, and, consequently the starting point for the development of potential high-throughput platforms of innovative assays in the presence of <sup>m</sup>DNA.

## **CHAPTER 3**

# **DEVELOPMENT OF A THERMOSTABLE CHIMERA FOR THE RAPID DETERMINATION OF $O^6$ -AGS IN A HIGH-THROUGHPUT ASSAY**

# Development of a thermostable chimera for the rapid determination of $O^6$ -AGs in a high-throughput assay

Alkylation of guanine in the  $O^6$  position, as mentioned in chapter 1, is very harmful to the cell because  $O^6$ -alkyl-G incorrectly couples with thymine during replication, generating a mutation by transition from G:C to A:T [Daniels *et al.*, 2000; 2004 ]. To date, there are few specific and high-throughput methodologies for the qualitative and quantitative analysis of  $m$ DNA upon exposure to alkylating agents, present in environmental microorganisms and fresh foods (fruit, vegetables, meats, etc.), for the determination of the quality of the environment and food production. As already mentioned in the purposes of the thesis, the AGT moiety of the chimera recognizes and repairs the  $O^6$ -AGs in the DNA sample, while the *Halotag* part is used for direct and specific immobilization on the support, previously conjugated with its relative substrates (Figure 10).



**Figura 10:** Strategies for the thermal stabilization of the chimera AGT-Halo.

The need to perform the reaction at not mild temperatures to avoid any endogenous AGT activities still present in the DNA samples, led to the overstep the limitations of mesophilic AGTs for our high-throughput assay, focusing the attention to homologous activities from (hyper)thermophilic sources. In fact, thermophile and hyperthermophiles genome's is exposed to endogenous and exogenous agents and attacks. In particular, alkylation damage is

accelerated by high temperatures and leads to the formation of DNA breaks [Mattossovich *et al.*, 2020]. The presence of AGT homologous genes in thermophilic and hyper-thermophilic organisms has demonstrated the effective existence of methylating agents in extreme environments and the role essential for this enzyme in the survival of Archaea and Bacteria species [Chan *et al.*,1995; Koscielska *et al.*,2003; Miggiano *et al.*,2016]. Based on the 3D structures available in the Protein Data Bank, “thermoAGTs” show structural similarities to their mesophilic homologues, presenting two domains (N- and C-terminal) separated by a connecting loop (Figura 4). ThermoAGTs were first studied by Imanaka and collaborators on the archeon hyperthermophilic *Pyrococcus* sp. KOD145 [Leclere *et al.*,1998] and by Prof. Pegg on *Aquifex aeolicus* and *Archaeoglobus fulgidus* [Kanugula *et al.*,2003]

#### *The Saccharolobus solfataricus* OGT

In the last decade, a thermostable AGT has been introduced by the thermophilic archaeon *Saccharolobus solfataricus* (SsOGT). This enzyme, for its peculiarities, was used for the first time for SNAP-tag<sup>®</sup> technology [Juillerat *et al.*,2003;Keppler *et al.*,2003;Gronemeyer *et al.*,2006] at high temperatures, becoming an efficient thermostable *protein-tag* [Kindermann *et al.*,2003; Mattossovich *et al.*,2020]. *S. solfataricus* was discovered and isolated in volcanic hot springs in the Solfatara volcano (Pisciarelli-Naples, Italy) in 1980 [Ziling *et al.*,1980]. Living in harsh conditions, such as 80 °C and a pH 2.0–4.0 range, this microorganism has evolved efficient repair systems [Valenti *et al.*,2006]. The *S. solfataricus* ORF SSO2487 encodes a 151-aa polypeptide homologous to AGTs, which shows in primary structure alignments conserved motifs involved in the activity of the enzyme. The relative recombinant protein (SsOGT) is able to perform the repair of DNA from alkylation damages, mainly acting at position O<sup>6</sup> of guanines. This protein presented an excellent *in vitro* stability at high temperatures, showing a half-life of 165.1 ± 16.5 minutes and 18.7 ± 2.0 minutes at 70 °C and 80 °C respectively [Perugino *et al.*,2012] and proved optimal catalytic activity at 80 °C. As most thermophilic enzymes, SsOGT is competent to work in an ample range of reaction conditions, such as high ionic strength (up to 2 M NaCl), in the presence of detergents (0.5% Tween 20; 1.0% Triton X-100; 0.5% Sarcosyl), and common denaturing agents and

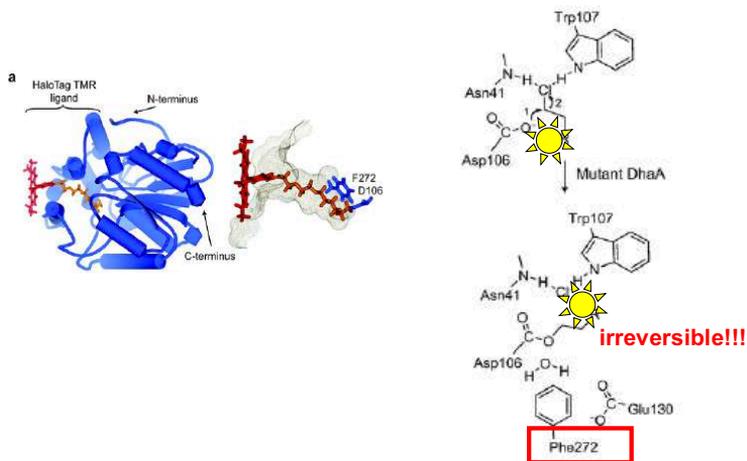
proteases [Perugino *et al.*,2012]. Interestingly, chelating agents do not influence the activity of this enzyme, whereas MGMT is strongly affected because of the presence of a structural Zinc ion in the N-terminal domain [Daniels *et al.*,2000] (Figure 4). Recently, the 3D SsOGT structure has been solved [Perugino *et al.*,2015], revealing that its Nterm domain (1-53 aa) lacks the structural zinc, so explaining the stability of the archaeal enzyme in the presence of chelating agents. On the other hand, the anti-parallel  $\beta$ -sheet in the Nterm domain is stabilized by the presence of a disulfide bond between two cysteine residues in positions 29 and 31 [Perugino *et al.*,2012].

However, SsOGT has an optimal activity at 80 °C, while at moderate high temperatures (50-60 °C; [Perugino *et al.*,2012]) its activity is much lower, limiting its utilization in our high-throughput assay, where also DNA thermal denaturation should be prevented. For this reason, part of the research was based to discover a new thermostable AGT with displays a higher activity at moderate high temperatures.

### *The commercial Halotag*

As the self-labelling SNAP-tag<sup>®</sup>, the Halotag<sup>®</sup> allows the indirect labelling of genetically fused proteins of interest with desired chemical groups. This enzyme is an engineered version of the bacterial haloalkane dehalogenase from *Rhodococcus rhodochrous* (EC 3.8.1.5), which is able to form a covalent bond with synthetic ligands (Figure 11). Ligands contain two components: 1) a common reactive linker that forms a covalent bond with the Halotag<sup>®</sup> protein, and 2) a functional reporter such as a fluorescent dye or an affinity handle such as biotin. Haloalkane dehalogenases remove halides from aliphatic hydrocarbons by means of a nucleophilic shift mechanism [Janssen *et al.*,2004]. A covalent ester bond is disposed during catalysis between an aspartate in the enzyme and the hydrocarbon substrate. In the wild type enzyme, the base-catalysed hydrolysis of this covalent intermediate, mediate by a conserved histidine in the active site, subsequently disengages the hydrocarbon as alcohol and regenerates the aspartate nucleophile for further catalysis cycles. In the Halotag<sup>®</sup>, a mutation of this histidine (His272Phe) led to the formation of a stable bond with 1,2-dibromoethane [Los et al. 2008] (Figure 11).

This is an approach so it allows you to easily create stable bonds with synthetic molecules added to a chloroalkane linker, thus allowing easy connection to a number of characteristics. The trapping of the covalent intermediate permits the irreversible attack of the chemical functionalities and the structure of the dehalogenase provides specificity and efficiency of the reaction.



**Figure 11:** the 3D model *Halotag*<sup>®</sup> protein and close-up of the ligand tunnel (outlined by a mesh Connolly surface) with covalently bound TMR ligand. The reaction mechanism of wild-type (WT) and mutant dehalogenase. Nucleophilic displacement of the terminal chloride with Asp106 leads to a covalent alkyl-enzyme intermediate. In the wild-type enzyme, His272 acts as a general base to catalyse hydrolysis of the intermediate, resulting in product release and regeneration of enzyme. In the mutant protein, the substituted Phe272 is ineffective as a base, thereby trapping the reaction intermediate as a stable covalent adduct.

*Halotag*<sup>®</sup> ligands have the same chloroalkane reactive linker, but the differences in functional reporter and the distance of the reporter from the linker create interchangeable labelling technology. For example, the *Halotag*-TMR<sup>®</sup> ligand is a cell permeable red emitting ligand, but unlike some red fluorescent proteins, it does not require tetramerization [Miller *et al.*, 2005; Los *et al.*, 2008;]. Green Emitting *Halotag*-Alexa Fluor<sup>®</sup> 488 and Biotin PEG are cell-impermeable ligands. As a result, the interchangeability of a wide range of ligands allows for a variety of functional studies of fusion proteins generated from a single genetic construct

Because of the mesophilic nature of the Halotag, the achievement of its thermostable and active version is mandatory. To aim this, I performed a molecular evolution by a selection of variants of this enzyme by expressed in the thermophilic bacterium *Thermus thermophilus*.

## **Fast thermostable OGTs**

### *Introduction*

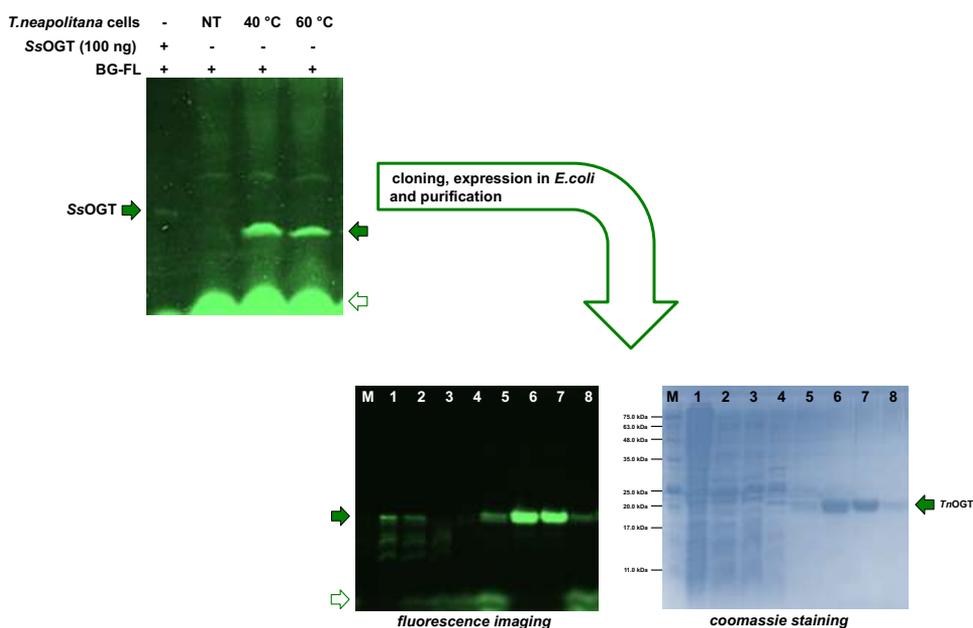
Following the Figure 4, I started my search of thermostable OGTs to be alternative to the archeal *Saccharolobus* counterpart, focusing my attention on the euryarchaeal *Pyrococcus furiosus* and on the Gram-negative bacterium *Thermotoga Neapolitana* [Terns *et al.*,2013;Kengen *et al.*, 2017;Mattossovich *et al.*, 2020;], an hyper-thermophilic microorganism, belonging to the *Thermotogales*, recently considered excellent models for biotechnological and genetic engineering applications [Jannasch *et al.*,1998;Connors *et al.*, 2006;Zhang *et al.*,2015;Fink *et al.*,2016]. It has an ORF (CTN1690) that shows a clear homology to O<sup>6</sup>-alkylguanine-DNA-alkyl-transferase enzyme class, encoding a polypeptide of 174 aa with a molecular weight of about 20 kDa.

### *Results and discussion*

To verify the above-mentioned hypothesis, lyophilized *T. neapolitana* cells were incubated at different temperatures with the fluorescein commercial benzyl-guanine derivative (SNAP-Vista Green®; BG-VG), showing a strong fluorescent signal with a molecular weight compatible to that of purified SsOGT by *gel-imaging* after SDS-PAGE (Figure 12).

The sensitivity to the benzyl guanine-derivatives was enough convincing to proceed with the cloning and the heterologous expression of this ORF in *E. coli* of the His-tagged version of this protein. After purification by affinity chromatography (Figure 12), the protein was confirmed to be fully active on BG-VG and subjected to biochemical characterization. The inhibition assay above described allowed the determination of an IC<sub>50</sub> value similar to that obtained with the *S. solfataricus* enzyme [Perugino *et al.* 2015], therefore attesting a role in DNA repair of this thermophilic protein, called hereinafter *TnOGT* (table 1).

Surprisingly, the enzyme from *T. neapolitana* displayed a very high activity at low temperatures (Table 1), in the same order of magnitude we achieved from the engineered SsOGT version (H<sup>5</sup>; [Perugino *et al.*, 2012; Vettone *et al.*, 2016]), hampering the determination of the second order constants at temperatures above 50 °C, since its reaction rate went beyond the technical limits of our assay. In H<sup>5</sup> mutant, a conserved serine residue in the active site loop (S132) was opportunely replaced with a glutamic residue, leading to a strong enhancement of the general activity on O<sup>6</sup>-BG derivatives, as observed in other examples (Juillerat *et al.*, 2003).



**Figure 12:** The OGT from *Thermotoga neapolitana*. Lyophilized *T. neapolitana* cells grown in the presence of CO<sub>2</sub>, were resuspended in PBS 1X buffer and BG-FL 5.0 μM and incubated 120 min at the indicated temperatures; NT, resuspended cells immediately loaded on SDS-PAGE. The *ogt* gene was expressed in *E. coli* and protein was purified by His-tag affinity chromatography, as described in the “Materials and methods” section. Lane M: molecular weight marker; lane 1: cell-free extract; lane 2: flow-through; lanes 3 and 4: column washing; lanes 5–8: eluted protein by imidazole gradient. Filled and empty green arrows indicate labelled proteins and the BG-VG substrate, respectively.

**Table 1:** DNA repair activity at 50 °C of OGTs by competitive inhibition studies in the presence of BG-FL (used as substrate) and a ds oligonucleotide containing an O<sup>6</sup>-methylated-guanine.

	IC <sub>50</sub> ( $\mu$ M)	notes
SsOGT	1.01 $\pm$ 0.08	from Perugino <i>et al.</i> 2015.
TnOGT	0.53 $\pm$ 0.13	this study; Mattossovich <i>et al.</i> 2020

Interestingly, an analysis of the superimposition between a TnOGT 3D model and the 3D structure of the free form of SsOGT (PDB ID: 4ZYE) showed the presence of this conserved serine in TnOGT (Figure 13). However, some residues involved in ionic interactions that play an important role in the stabilization of SsOGT at high temperatures [Mattossovich *et al.*, 2020] are missing in TnOGT. Likely, different stabilization mechanisms and alternative residues contribute to its exceptional catalytic activity at moderate temperatures and high thermal stability.

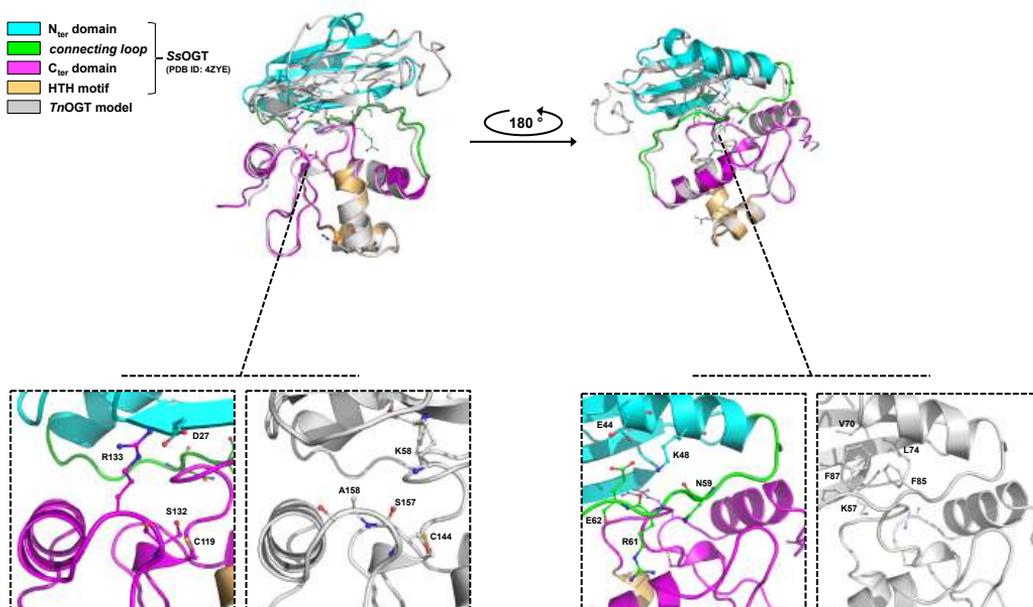
**Table 2:** Catalytic activities as a function of temperature of thermostable OGTs, expressed as second order rate constant values in the presence of the sole BG-VG substrate.

	T (°C)	k (s <sup>-1</sup> M <sup>-1</sup> )	notes
SsOGT	25	2.80 x 10 <sup>3</sup>	from Perugino <i>et al.</i> 2012.
	50	1.50 x 10 <sup>4</sup>	this study; Mattossovich <i>et al.</i> 2020
	70	5.33 x 10 <sup>4</sup>	from Perugino <i>et al.</i> 2012.
	80	ND <sup>a</sup>	this study; Mattossovich <i>et al.</i> 2020
TnOGT	25	4,65 x 10 <sup>4</sup>	this study; Mattossovich <i>et al.</i> 2020
	50	2.19 x 10 <sup>4</sup>	this study; Mattossovich <i>et al.</i> 2020
	70	ND	this study; Mattossovich <i>et al.</i> 2020
	80	ND	this study; Mattossovich <i>et al.</i> 2020

<sup>a</sup>ND, not determined.

On the other hand, *Tn*OGT was unusually prone to degradation during storage at -20 °C, showing a series of bands below the entire length, both in fluorescence imaging and in Coomassie staining. This liability hampered thermostability measurements from the DSF analysis. The reason for this degradation is unexpected: the recombinant enzyme is in its full-length form in *E. coli* cells, as well as after heat treatment of the cell-free extract, and finally after purification by affinity chromatography.

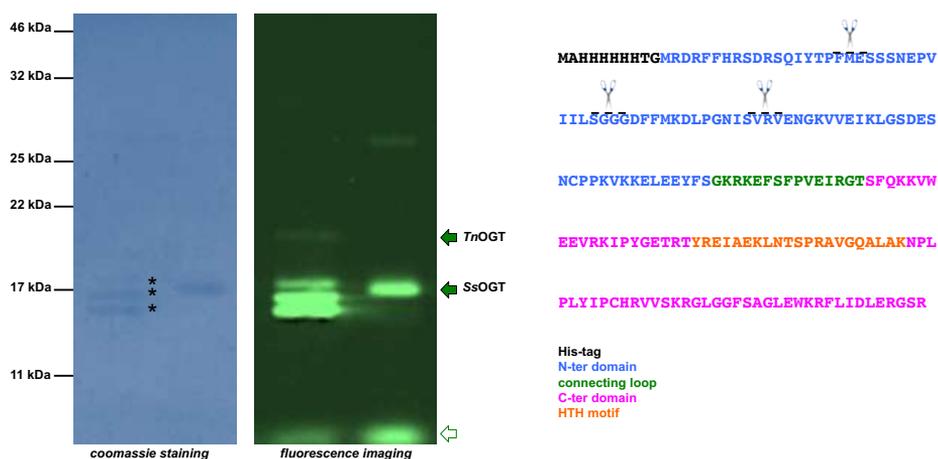
From Coomassie staining analysis, three main bands of approximately 17.5, 16.1 and 14.7 kDa were detected. Since they were also fluorescent (and, therefore, catalytically active), it is likely that these cuts are at the expense of the Nterm domain, still keeping the polypeptides together and making the enzyme catalytically active in solution. (Figure 13)



**Figure 12:** Superimposition between the SsOGT 3D structure (PDB ID: 4ZYE; coloured as described in the legend) and a model of the *Tn*OGT (*in gray*). Insets represent a zoom-view of local ionic interactions in SsOGT compared with the same positions in *Tn*OGT. Atoms are coloured according the CPK convention (carbon, in the corresponding colour of each 3D structure; nitrogen, *in blue*; oxygen, *in red*; sulphur, *in yellow*) From [Mattosovich et al., 2020].

Analysis performed on the OGT from *P. furiosus* OGT (*Pfu*OGT) showed that, as expected, this enzyme possesses an exceptional stability but an equally exceptional thermophilicity, being less active than SsOGT above 60 °C [Mattosovich *et al.*, 2020].

For the reasons described above, we decided to continue employing SsOGT as part of the chimera. We are currently studying our *Tn*OGT protein (stabilization solutes, mutagenesis, etc.), in order to decrease or abolish its degradation and make the enzyme suitable for our high-throughput assay.



**Figure 13:** Degradation of the *Tn*OGT. After reaction with the BG-VG substrate, the purified *Tn*OGT (lane 1) and SsOGT (lane 2) were loaded to the SDS-PAGE and subjected to the *fluorescence gel-imaging* and Coomassie staining, according to the protocol in Materials and Methods. Green filled arrows indicate the SsOGT and the full-length *Tn*OGT; asterisks indicate the fragments of the *Tn*OGT. Scissors on the *Tn*OGT primary structure indicate the probable sites of the cuts in the protein.

# ***Enhancing the thermostability of Halotag***

## *Introduction*

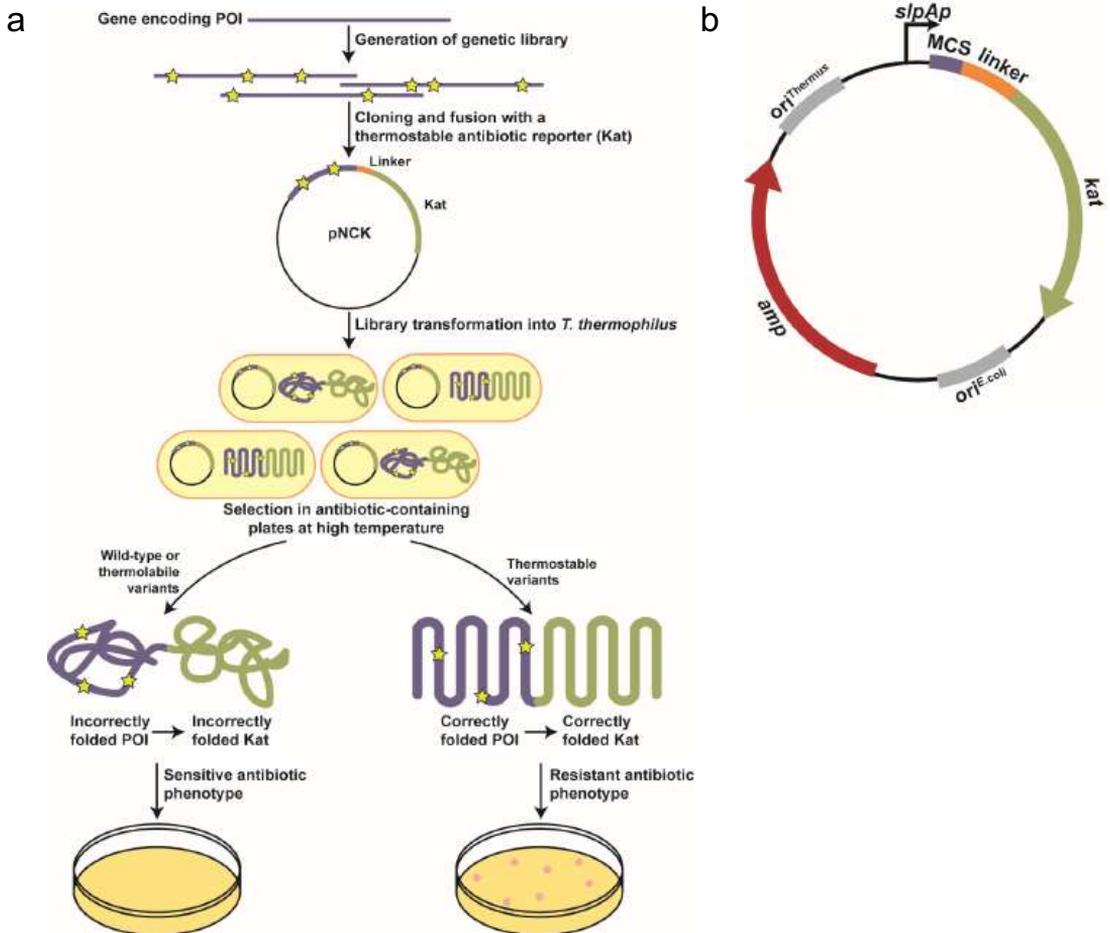
Direct evolution makes use of the principles of natural evolution, "remarking" the Darwinian principles of mutation and selection in the laboratory and reducing the time scale of evolution [Eigen et al.,1984; Mate et al.,2017]. The main object of any direct evolution experiment is to identify and restore variants from mutant libraries with the desired improvements in terms of fitness. In this project, a construction of a library of *Halotag* mutants by random mutagenesis and the next selection of its thermostable variant(s) will be performed.

This aspect of my PhD Project was carried out at the Department of Molecular Biology of the Autonomous University of Madrid (Spain) in collaboration with the group of Dr. Aurelio Hidalgo. For several years, in Hidalgo's laboratory, optimizations have been underway for the *ex vivo* expression of recombinant proteins through the use of high-temperature expression systems, purified from the bacterial strain *Thermus thermophilus* HB27 [Faraldo *et.al.*,1992; Schwarzenlander *et.al.*,2006;Cava *et al.*, 2009].

Among several strategies for analysing mutant libraries, particularly successful are those combining the improvement of the enzymatic stability and the survival of the host expression. Based on these premises, Chautard in 2007 developed an activity-independent selection system, called THR, a system for selecting thermostable variants decoupled from any biological function [Chautard *et al.*, 2007].

This *folding interference principle* method involves the use of a thermostable variant of the kanamycin nucleotidyl transferase (Kat) [Matsumura et al.,1985], although it could be replaced by any other thermostable antibiotic reporter. In this method, Kat is fused to the Cterm of the protein of interest and the resulting fusion is expressed in the thermophilic bacterium *Thermus thermophilus* HB27, one of the most important microorganisms used for biotechnological applications, because it has the ability to grow in laboratory conditions and it has an extremely efficient natural competence. Thus, cells expressing fusions with an unstable protein of interest show reduced resistance to kanamycin, due to a deleterious effect of the misfolded protein of interest on Kat folding. Therefore, the stability of the variants of the protein of interest is

linked to the correct folding of Kat and therefore to the growth of *T. thermophilus* on plates containing antibiotics at high temperatures (Figure 14a) [Chautard *et al.*, 2007].



**Figure 14:** a) Selection method based on the *folding interference principle*: the first step is the generation of a library of the protein of interest (POI) then cloned into a plasmid that allows fusion at the Cterm of an antibiotic resistance reporter (Kat) and transformed into *T. thermophilus*. At higher temperatures, and in the presence of an antibiotic, the not thermostable variants (or for unwanted non-stop codon incorporation) will not fold, leading to sensitive-antibiotic hosts (no colonies). On the contrary, the thermostable variants will fold correctly allowing, in turn, the correct folding of Kat and thus, the stable variants will be identified by the growth of the colonies. b) pNCK plasmid: promoter of the *slpA* gene (*slpAp*), multi-cloning site (MCS, containing the unique cloning sites *Nco* I, *Bam* HI and *Not* I), linker, thermostable kanamycin nucleotidyl transferase (*kat*), *E. coli* origin of replication (*ori<sup>E.coli</sup>*), ampicillin resistance gene (*amp*) and *Thermus* spp. origin of replication (*ori<sup>Thermus</sup>*).

To apply the THR method, an *E. coli*-*Thermus* spp. shuttle plasmid is employed for the construction of the library in *E. coli* and the expression of the fusion protein in *Thermus*. This plasmid, called pNCK (Figure 14b), is based on the pUC18 plasmid124 harbouring the *amp* gene, as well as both the origins of replication for *E. coli* and *T. thermophilus* [Lasa *et al.*,1992].The gene encoding Kat, the thermostable kanamycin nucleotidyltransferase from *Bacillus stearothermophilus* [Matsumura *et al.*,1985] were inserted under the control of the *slpA* strong promoter [Lasa *et al.*,1992], generating the *slpAp-kat* cassette.[Faraldo *et al.*,1992]This cassette was further modified by inserting the unique cloning sites *Nco* I, *Bam* HI and *Not* I and a linker (AAAGSSGSI) between the *slpA* promoter and the *kat* gene. The linker was designed to be hydrophobic, flexible and without any protease cleavage site [Chautard *et al.*, 2007].

Once the plasmid was constructed and the microorganism transformed, the folding interference hypothesis was validated to select thermostable variants. To this end, Chautard generated libraries for several proteins, notably human interferon alpha (IFN $\alpha$ ), beta (IFN $\beta$ ) and gamma (IFN $\gamma$ ), lipase A from *Bacillus subtilis* (LipA) and the formate dehydrogenase from *Pseudomonas* sp. This method proved effective for all the analysed proteins, obtaining thermostable variants in all cases.

### *Results and discussion*

By following the THR approach proposed by Chautard, where an improperly folded Nterm of a protein has a deleterious effect on the folding of its Cterm portion, I produced two constructs, one inserting the *halotag-kat* fused gene in pNCK, and the second *halotag-hyg* (the gene of the hygromycin from *Streptomyces hygroscopicus*) fused gene in pNCH.

After *T. thermophilus* transformation, I first verified that the expressed fused proteins allow the growth of the cells up to 60 °C. Furthermore, the transformed cells resulted active for the *Halotag*-TMR<sup>®</sup> substrate, indicating that the *Halotag*<sup>®</sup> is correctly folded and active and the cells are permeable with this kind of substrate. I also verified that the wt enzyme is compatible with the *Thermus* growth at 60 °C, whereas a higher temperature (ca. 63 °C) resulted in no growth. This was the starting point of the subsequent molecular evolution experiments.

The introduction of mutations into the *halotag* gene was carried out by error-prone PCR (epPCR), using different concentrations of  $Mn^{2+}$ . The amplified products were ligated into plasmids pNCH and pNCK,[Cava *et al.*,2007] followed by the *E. coli* transformation and plated on LB selective medium. The number of clones on the plated indicated a potential number of  $4.5 \times 10^4$  and  $2.3 \times 10^4$  clones for pNCK-*Halotag* and pNCH-*Halotag*, respectively. I picked up ca. 20 clones and I verified the mutation frequencies (technically 1-2 mutation per *halotag* gene), demonstrating that the optimal condition of the ePCR was 0.5 and 0.3 mM for the Kan and Hyg selection, respectively.

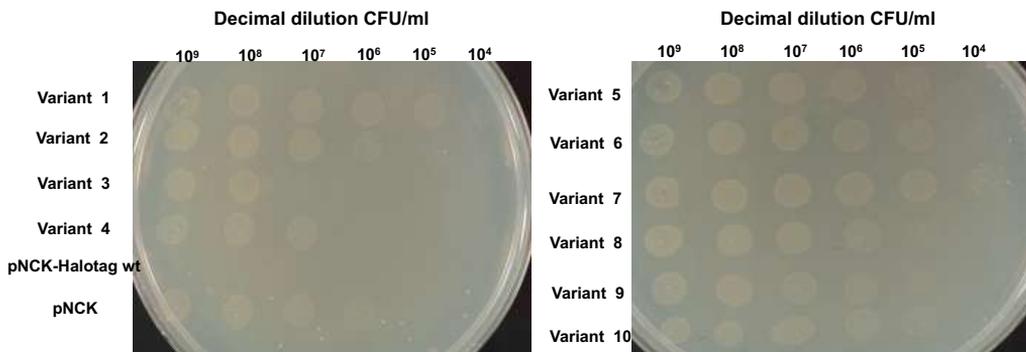
Successively, I transformed *T. thermophilus* with the obtained variant libraries, plated them at different concentrations of Kan/Hyg, and incubated at 60-63 °C. After 48 hours, I found colonies at 63 °C and 30 mg/mL of kanamycin for pNCK-*Halotag* and 100 mg/mL of Hygromycin for pNCH-*Halotag*. Then, I successfully verified the activity of Halotag variants from this first selection, by picking those colonies, growing them in liquid medium and incubating with the fluorogenic substrate, followed by *gel-imaging* analysis.

To verify the selection process, twenty randomly selected clones of these libraries were picked up and incubated at 65 °C for Kan-library and 67 °C for Hyg-library by the serial dilution assay. Some variants could grow under these conditions, whereas the *Halotag*<sup>®</sup> wt could not (Figure 15). Therefore, the next step should have to push on the temperature to 67 °C and 70 °C, by taking advantage of recombinant events by numerous techniques [Stemmer *et al.*,1994; Lehtonen *et al.*,2015; McCullum *et al.*,2010]. Unfortunately, all these aspects (including the DNA sequence of variants from the first selection) have not investigated, due to the pandemic situation and the return to Italy was mandatory.

## Conclusion

From this Spanish work experience I obtained a lot of information as:

- The acquisition of the manipulation of a new microorganism *T. thermophilus*;
- The generation of *Halotag* mutant libraries by ePCR and the subsequent selection in *T. thermophilus* by THR method. This allowed the recovering of some higher thermostable variants than the wt, also confirmed by the plate dilution assay; This was not possible due to the on-going pandemic. This activity will be part of my post-doc research.



**Figure 15:** Dilution assay on plate to compare thermostability of different mutants at 65 °C. *T. thermophilus* containing Halotag wt does not grow, while mutants showed different resistance degree.

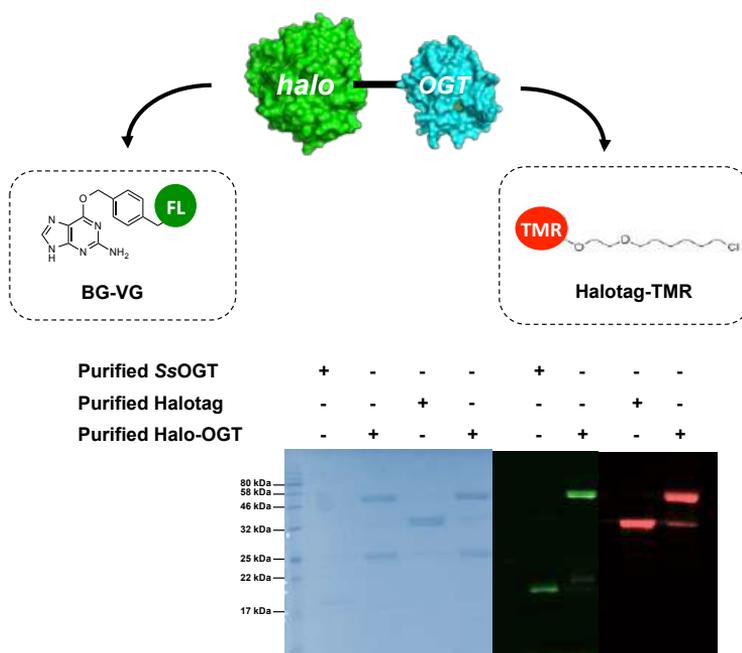
## ***Construction and production of two different chimeras: which is the best?***

Because the final scope in the development of the chimera is the achievement of a product in which the activity and the stability of each enzyme is not affected by the presence of the other one, I first produced two different version: OGT-Halo (chimera 1) and Halo-OGT (chimera 2). Furthermore, it is also important verify which of them is more suitable to the immobilization procedures on the support.

### ***Results and discussion***

In order to obtain OGT-Halo, I cloned into the pQE-*ogt* plasmid [Perugino et al., 2012] the *halotag* gene downstream and in-frame to the His-tagged *ogt* gene. Previous experiments producing the SsOGT wt at the N-terminal of fusion proteins containing different enzymes (as  $\beta$ -glycosidase, reverse gyrase and carbonic anhydrase) demonstrated that the total absence of any interference by SsOGT [Vettone *et.al.*,2016; Visone *et al.*, 2017; Merlo *et al.*, 2019]. I successfully expressed OGT-Halo in *E.coli* ABLE C strain and purified this protein by His-*tag* affinity chromatography, as well for the separated enzymes (see M&M). Unfortunately, I did not obtain a sufficient amount of protein to allow a complete biochemical characterization. A series of attempt are being performed to enhance and optimise the expression of this chimera, to finally verify its potential within this project.

To obtain the Halo-OGT version, we simply replaced the *Halotag* gene in place of the *lacS* gene in pQE-*ogt-lacS* [Vettone *et al.*, 2016], obtaining the pQE-*halo-ogt* plasmid. Again, for this second fusion protein, we used the same expression and purification procedures used for the chimera 1. As shown in Figure 16, both the enzymes in Halo-OGT (ca. 52 kDa) react with their respective substrates with efficiency comparable to their free counterparts (ca. 20 and 33 kDa, respectively). These results suggest that SsOGT also if positioned at the C-term of a *protein-tag*, it can be labelled specifically and efficiently with his substrate (BG-VG), as previously demonstrated for the ASL<sup>tag</sup> production [Merlo *et al.*, 2019; Del Prete *et al.*, 2019].



**Figure 16:** Orthogonal substrate specificity of the Halo-OGT chimera, compared with the purified single SsOGT and the *Halotag*<sup>®</sup>.

Initially, Halo-OGT were tested at a temperature of 60 °C for 10 minutes: under these conditions, it was not possible a quantitative determination of the fluorescent intensity, because the *Halotag* moiety completes its reaction after a few minutes, hindering the determination of the second order constants. For this reason, the assay was performed at 25 °C. As shown in the Table 3, Halo-OGT is as active as the *Halotag*<sup>®</sup>, showing a remarkable ability to perform the reaction at very low temperatures. The OGT activity of the chimera was measured at 25 °C and 60 °C, resulting very close to the SsOGT [Perugino *et al.*, 2012, 2015; Mattosovich *et al* 2020]. Thus encouraged us to use the chimera at temperatures of 50-60 °C, as the aim of the project provides.

The demonstration of the activity of this chimera is necessary, but not sufficient for the employing in the high-throughput assay: Halo-OGT activity on fluorescent substrate is strongly affected by the presence of <sup>m</sup>DNA (Table 4) as the SsOGT enzyme [Perugino *et al.*, 2015], thus confirming a role in DNA repair of this chimeric protein.

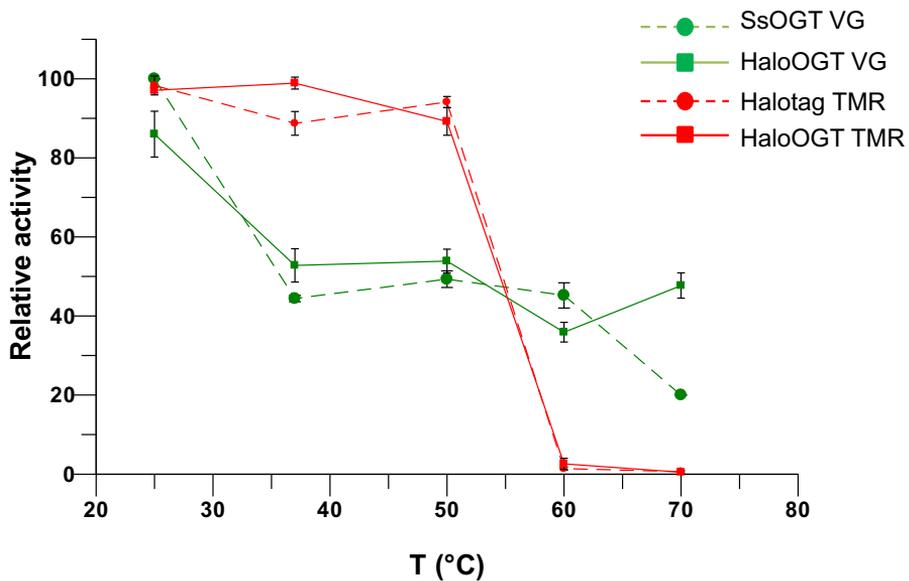
**Table 3:** Catalytic activities as a function of temperature of SsOGT, Halotag and HaloOGT, expressed as second order rate constant values in the presence of the BG-VG substrate and Halotag-TMR.

	T (°C)	BG-VG $k$ (s <sup>-1</sup> M <sup>-1</sup> )	HaloTag-TMR $k$ (s <sup>-1</sup> M <sup>-1</sup> )	notes
SsOGT	25	2.80 x 10 <sup>3</sup>	-	from Perugino et al. 2012
	60	1.02 x 10 <sup>4</sup>	-	this study
Halotag	25	-	5.7 x 10 <sup>5</sup>	this study.
	60	-	ND	this study
Halo-OGT	25	1.68 x 10 <sup>3</sup>	5.1 x 10 <sup>5</sup>	this study
	60	7.03 x 10 <sup>3</sup>	ND	this study

**Table 4** DNA repair activity of SsOGT and Halo-OGT at 50 °C by competitive inhibition studies in the presence of BG-VG (used as substrate) and a dsDNA oligonucleotide containing one O<sup>6</sup>-MG.

	IC <sub>50</sub> (μM)	notes
SsOGT	1.01 ± 0.08	from Perugino et al. 2015
Halo-OGT	0.95 ± 0.08	this study

As explained before, the chimera will be subjected to a reaction at > 60 °C to avoid any AGT endogenous activities. Preferentially, the final assay should be performed in ca. 10 min. To this aim, I verified the residual activity of the Halo-OGT chimera and its separated components, after a pre-treatment for 20 min at different temperatures, followed by an incubation with the relative fluorescent substrates for 120min at room temperature (Figure 17). Interestingly, each moiety shows as similar behavior compared with the separated counterparts, indicating that OGT from *S. solfataricus* retains half of its activity at > 60 °C, whereas the *Halotag* drops to the total inactivation at those temperatures. These results clearly demonstrated the need of a thermostable *Halotag* version.



**Figure 17:** Stability of SsOGT, Halotag and the Halo-OGT chimera, by measuring the residual activity after a pre-treatment at different temperatures, followed by the fluorescent activity assay at room temperature.

## Conclusion

In this first part of my PhD thesis, I have produced two chimera versions: both chimeras expressed in *E. coli* are as stable and active at 60 °C as for the need of our assay. We should optimize the expression and purification of OGT-Halo to be able to characterize it and choose which is actually the best fusion protein to use in our assay and high-throughput for the detection of alkylated DNA from samples such as blood, water and industrial food waste and contaminants.

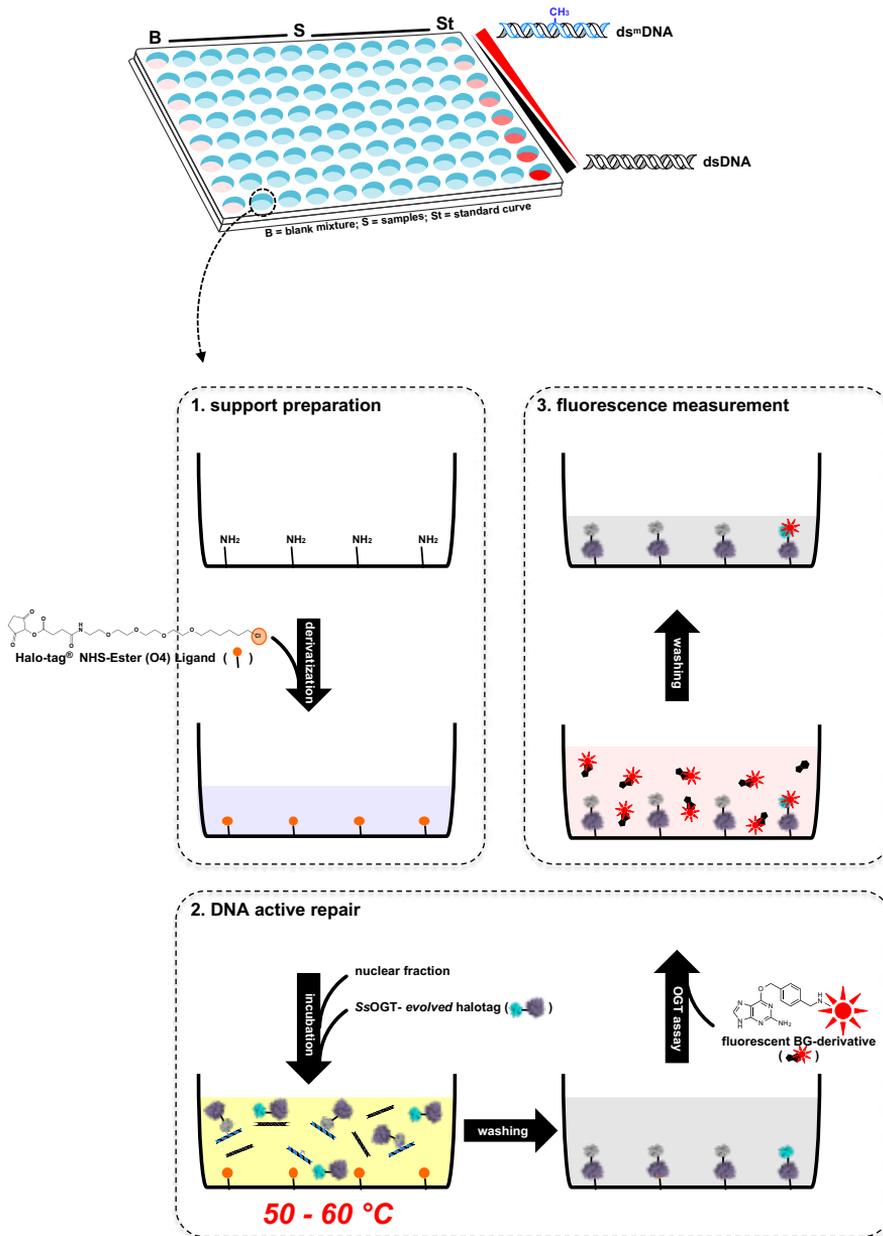
The last phase of the project, focused on the setting up of the high-throughput assay based on the chimeras (according to the Figure 18), had to be performed at the IXTAL srl (Novara, Italy), in the laboratory of Dr. Miggiano and Dr. Ferraris. This was not possible due to the on-going pandemic. This activity will be part of my post-doc research, when pandemic conditions will allow me to spend a period in IXTAL srl.

Briefly, samples containing <sup>m</sup>DNA are incubated at 60 °C in the presence of the chimera for an opportune time frame. Some part of the OGT in the chimera reacts with the <sup>m</sup>DNA (OGT *in gray* in

figure 18). At the same time, the activity of the *Halotag* moiety on the *Halotag* NHS-Ester (O4)<sup>®</sup> substrate makes possible the immobilisation of the chimera on the solid support, previously derivatised with NH<sub>2</sub> groups. This allows the following washing procedures, in order to eliminate all the extract from the samples.

Only the Halo-OGT molecules, which did not react with the <sup>m</sup>DNA (in blue in Figure 18), are available for the final reaction with a fluorescent BG-derivate substrate. After a washing, it is easy by fluorescent measurement to determine the amount of the labelled chimera, opportunely compared to a standard curve obtained by using dsDNA oligonucleotides containing specific amounts (in pmol) of O<sup>6</sup>-MG (Figure 18).

## The enzyme-based “mDNA detector”



**Figure 18:** the proposed high-throughput assay for the determination of mDNA, by the exploitation of thermostable chimeras harbouring AGT and Halotag activities.

## *Experimental procedures*

### Reagents

Fine chemicals were from Sigma-Aldrich, SNAP-Vista Green<sup>®</sup> fluorescent substrate (hereinafter BG-FL) was from New England Biolabs (Ipswich, MA). Synthetic oligonucleotides listed in Table 5 were from Eurofins (Milan, Italy); *Pfu* DNA polymerase was from NZYTech (Portugal). The Bio-Rad protein assay kit (Bio-Rad Pacific) was used for the determination of the protein concentration, using purified BSA as standard.

### DNA constructs.

The CTN1690 ORF, encoding a putative OGT, was amplified from the genomic DNA of *Thermotoga neapolitana*, using the Lig5-TnOGT and Lig3 TnOGT oligonucleotide pairs and cloned directly into the pHTP1 expression vector (NZYTech, Portugal), following the instructions described in the NZYEasy Cloning and Expression Kit Manual I (NZYTech, Portugal). The ligation mix was fully used to transform DH5 $\alpha$  cells from commercial *E. coli* (NZY5 $\alpha$  Competent Cells - NZYTech, Portugal) and the positive clones were confirmed by PCR. Subsequently, a DNA fragment from the resulting pHTP1-TnOGT was eliminated by digestion with Nco I restriction endonuclease and was replaced by a double-stranded oligonucleotide (NZY-His Fwd2 and NZY-His Rev2;) whose DNA sequence expresses a shorter His6- tag (MAHHHHHTG-), similar to that at the N terminus of the SsOGT protein (Perugino et al. 2012). To produce fusion proteins of genes of interest with SsOGT, the construction of a plasmid, the pQE-OgtMCS, was induced in which the amplified fragment of the *ogt* gene was cloned into the pQE31<sup>™</sup> vector and a small *Mul* Cloning Site was inserted (Perugino et al 2012). Using this plasmid as a recipient, the Halotag gene from *Rhodococcus rhodochrous* was cloned in the frame and upstream and downstream of the *ogt* gene. Resulting two plasmids pQEOGT-Halotag and pQEHaloOGT.

The halotag gene was also inserted in the plasmids of the pNCK and pNCH shuttles important for transformation into *Thermus thermophilus*. The oligonucleotides used for the DNA constructs are listed in Table 5. All the constructs described above were transformed into DH5 $\alpha$  cells and were confirmed by DNA sequence analysis.

name	sequence	note
Lig5-TnOGT	5'-tcagcaagggctcaggccatgqgagatcga-3'	Mattosovich <i>et al</i> 2020
Lig3-TnOGT	5'-cctcagcgggaagctgagggtatcgactacctgc-3'	Mattosovich <i>et al</i> 2020
NZY-His fwd2	5'-catggcaccatcaccatcaccatcggg-3'	Mattosovich <i>et al</i> 2020
NZY-His rv2	5'-catgcccgatggtgatggtgatggtgtgc-3'	Mattosovich <i>et al</i> 2020
Fwd <sup>m4</sup>	5'-ggcmgtaggcctagcatgacaatctgcattggtgatcacgg-3'	Perugino <i>et al.</i> 2015;
Rev4	5'-ccgtgatcaccatgagattgtcatgctaggcctaccgcc-3'	Perugino <i>et al.</i> 2015
Halo-Bam-fw	5'-ccatagcgaatccaatggcagaaatcggtactgg-3'	This study
Halo-Bam-rv	5'-agcattggatccccggaatctcgagcg-3'	This study
Halo-Pst-fw	5'-ccagaactgcagatggcagaaatcggtactgg-3'	This study
T7-rev	5'-gctagtattgctcagcg-3'	This study
Halo Hind fw	5'aaaaagccttgagggtgaggctaattggcagaaatcggtactg3'	This study
Halo Eco rv	5'aaagaattctaggaatctcgagcgtcgacag 3'	This study
Halo Nco fw	5'aaacatggcgcagaaatcggtactggc 3'	This study
Halo Not I rv	5'aaagcggccggaaatctcgagcgtcg 3'	This study

**Table 5** : list of oligonucleotides

## Protein purification

TnOGT ,OGT-Halotag and HaloOGT was expressed in the *E. coli* cells (BL21/ABLE C )grown overnight at 37 °C in Luria–Bertani (LB) selective medium supplemented with antibiotics (50 mg/L kanamycin and 30 mg/L chloramphenicol).and the protein expression was induced with 1.0 mM isopropyl-thio- $\beta$ -D-galactopyranoside (IPTG), when an absorbance value of 0.5–0.6 A600 nm was reached. Harvested cells were resuspended 1:3 (w/v) in Buffer A (50 mM phosphate, 300 mM NaCl; pH 8.0) supplemented with 1% Triton X-100 and stored overnight at – 20 °C. After this first step of lysis, the biomass was treated with lysozyme and DNase for 60 min in ice, followed by a sonication step. Finally, the lysate was centrifuged for 30 min at 60,000 $\times$ g in and the cell extract recovered. To remove *E. coli* contaminants, all cell extracts were incubated 20 20 min at 65 °C, followed by a centrifugation at 13,000 $\times$ g at 4.0 °C; the supernatant was diluted 1:2 (v/v) in purification Buffer A and applied to a Protino Ni–NTA

Column 1.0 mL (Macherey–Nagel) for His6-tag affinity chromatography. After two washing steps of 10 column volumes of Buffer A and 10 column volumes of Buffer A supplemented with 250 mM imidazole, the elution was performed in 20 column volumes of buffer A, by applying a linear gradient of 25.0–250.0 mM imidazole. The fractions containing the protein were collected and analysed by SDS-PAGE. The proteins were dialysed against PBS 1 × buffer (phosphate buffer 20 mM, NaCl 150 mM; pH 7.3), Finally, the proteins were concentrated and loaded on 15% SDS-PAGE gel to confirm its purity and stored at – 20.0 °C.

#### *In vitro alkyl-transferase and Halotag assays*

The fluorescent substrates BG-FL/HaloTagTMR were used for the determination of the catalytic activity, as previously described (Perugino et al. 2012, 2015; Miggiano et al. 2013; Vettone et al. 2016; Visone et al. 2017; Merlo et al. 2019; Del Prete et al. 2019). 5.0 µM of protein (ca. 0.1 mg/mL) was incubated with 10.0 µM of substrate in Fluo Buffer 1 × (50.0 mM phosphate, 100.0 mM NaCl, 1.0 mM DTT; pH 6.5) at different temperatures and times, as indicated; each reaction was stopped by adding a Laemmli buffer 1 × (formamide 95%; EDTA 20.0 mM; bromophenol 0.05%), followed by denaturation at 100.0 °C and the direct loading of the sample on SDS-PAGE. The gel was first analysed by fluorescence imaging on a VersaDoc 4000TM system (Bio-Rad) by applying as excitation/emission parameters a blue LED/530 bandpass filter, and then was stained by Coomassie. Assuming the irreversible mechanism with 1:1 BG-FL/OGT or HaloTagTMR/HaloTag ratio, fluorescence intensity data were corrected for the amount of loaded protein, and fitted by applying the second-order rate equation, to determine the relative amount of covalently modified protein in time- course experiments. (Gautier et al. 2008; Miggiano et al. 2013; Perugino et al. 2012, 2015)

### *Competitive assay and IC50 calculation*

Competitive assay employing the fluorescent substrate in the presence of double strands (ds) oligonucleotides pairs (Fwdm4: Rev4; Table 5 ), containing a single O6-methyl-guanine, was performed as described (Perugino et al. 2015) to determine the half maximal inhibitory concentration (IC50), that is the concentration of methylated DNA needed to reduce the fluorescence intensity of the OGT band by 50.0%. Reactions incubated at fixed temperatures with increasing concentrations (0.0–10.0  $\mu\text{M}$ ) of ds-Fwdm4 and keeping constant the BG-FL concentration (5.0  $\mu\text{M}$ ) were performed for 10 min at 50.0 °C. Corrected data of fluorescence intensity were fitted with the IC50 equation [Perugino et al. 2015; Morrone et al. 2017].

### *Microbiological methods : cultivation and transformation from *Thermus thermophilus**

*Thermus thermophilus* was grown in TB medium (Bacto-triptone 8g/L, yeast extract 4g/L, NaCl 3g/L, pH 7,5 in carbonate-rich mineral water) from 60 to 74 °C. Liquid cultures were shaken in an incubator at 180 rpm in flasks or tubes, without exceeding 1/5 of total volume. Incubation of *T. thermophilus* on plates was carried out in TB medium supplemented with 2% (w/v) of agar. To prevent desiccation, *T. thermophilus* plates were incubated for 48 h within wet chambers. To verify the temperature of the heater an iButton Thermo-S-kit-T thermochron Starter Kit 0-125°C (Measurement Systems LTD) was used. Selection of recombinant and transformed clones was carried out with antibiotics. For that purpose, media were supplemented with antibiotics to a final concentration 30  $\mu\text{g}/\text{mL}$  of kanamycin (Kan) or 100  $\mu\text{g}/\text{mL}$  of hygromycin B (HygB). Plates of *T. thermophilus* were kept at room temperature for short-term storage and pellets of cells in stationary phase were frozen at -20 °C for long-term storage. *T. thermophilus* was transformed by natural competence. An overnight culture was grown until it reached stationary phase. This pre-inoculum was used to inoculate a new culture with an OD600 of 0.05, which was grown to OD600 of 0.3-0.4. Then, 100 ng of DNA were added to 0.5 mL of this culture. After 4 h incubation at 60 °C, an appropriate dilution of cells were spread out on TB plates with the desired antibiotic and incubated for 48 h. Dilution assay on plate was

performed similarly; after transforming and incubating for 4 h, cells were serially diluted 10-fold and 10  $\mu$ L droplets were plated on selective TB plates.

### epPCR

The epPCR protocol modifies the conventional PCR conditions, e.g. by addition of MnCl<sub>2</sub>, higher concentration of MgCl<sub>2</sub> and unbalanced dNTP concentrations.<sup>177</sup> NZY<sup>Taq</sup> DNA polymerase (NZYTech) has been used, as Taq polymerase lacks a 3'→5' proofreading exonuclease activity, which is translated into a high intrinsic error rate (8x10<sup>-6</sup> mutations per bp per duplication).<sup>178</sup> Primers were designed to be around 20 bp upstream or downstream, respectively, of the gene to ensure an adequate subsequent DNA digestion of the PCR product. To carry out the epPCR on *halotag*, the oligonucleotides *ep\_pslpA* and *ep\_pNCH\_rv/ ep\_pNCK\_rv* were used.

Concentration of Mn<sup>2+</sup> in the reaction mix affects the mutation rate of the polymerase.<sup>102</sup> Therefore, for the gene *halotag*, three different Mn<sup>2+</sup> concentrations were checked, specifically, 0.2, 0.3 and 0.5 mM. The reaction mix and the thermocycler program for the epPCR are described in Table 6-7

Component	Final concentration
<b>Nuclease-free water</b>	Up to 100 $\mu$ l
<b>Mutagenesis buffer 10X</b>	1x
<b>dNTPs unbalanced</b>	0.35 mM A; 0.2mM G; 0.4 mM C; 1.35mM T
<b>MnCl<sub>2</sub></b>	0.2;/0.3/0.5
<b>Oligo fd</b>	0.5 $\mu$ M
<b>Oligo rv</b>	0.5 $\mu$ M
<b>Template DNA</b>	5 ng
<b>NZYTaq DNA polymerase</b>	5 U

Table 6 : Reaction mix for epPCR in *halotag* gene.

Cycle	Step	Temperature (°C)	Time (min)
1	Denaturation	95	3
2 to 31	Denaturation	95	1
	Annealing	59	1
	Elongation	72	1
32	Elongation	72	10

Table 7 Thermocycler program for epPCR

The halotag, which was encoded into the *E. coli*-*T. thermophilus* shuttle plasmid pNCK/H,179 was randomized by epPCR as mentioned above. The epPCR products were digested with Not I and Nco I, cloned in plasmid pNCK/h and transformed in *E. coli* XL10Gold. To obtain an overview of the mutational rate, 10 different colonies of each Mn<sup>2+</sup> concentration library were picked and sequenced with the primer pslpA\_fw.

### *Selection of improved variants*

Enzyme variants obtained by random mutagenesis were explored by a selection step at high temperature in *T. thermophilus* as a host. In this case, transformants expressing parental protein Halotag, could not grow. Thus, selection conditions, i.e. temperature and HygB/Kan concentration, were tested in order to identify thermostable clones. Once selection conditions were established, the resulting pNCK-epPCRhalotag or pNCH-epPCRhalotag library was transformed in *T. thermophilus* HB27EC.

To verify at which temperature and HygB/Kan concentration transformants expressing Halotag could not grow, three different temperatures (63, 65 and 67 °C) and four HygB/Kan concentrations (30, 45, 100 and 150 µg/mL) were evaluated.

Likewise, four temperatures (60, 63, 65 and 67 °C) and three HygB/Kan concentrations (30, 45, 100 and 150 µg/mL) were analysed to check that transformants expressing wild-type Halotag could not grow.

## **CHAPTER 4**

# **PRODUCTION OF A NEW AGTs' SUBSTRATE BASED ON A DNA TRIPLEX**

# PRODUCTION OF A NEW AGTs' SUBSTRATE BASED ON A DNA TRIPLEX

## *Introduction*

DNA damage caused by endogenous and exogenous mutagens, often associated to the insurgence of several diseases, is counteracted by cellular protection multi-enzymatic mechanisms, as the base excision repair (BER), the mismatch repair (MMR) and the nucleotide excision repair (NER) [Liu *et al.*, 2006], whereas a direct damage reversal activity of unique enzymes is performed by dioxygenases and AGTs [Zhong *et al.*, 2010; Pegg 2011; Mattosovich *et al.*, 2020]. The latter enzymes, as above described, indeed represent the major factor in contrasting the effects of alkylating agents that form such adducts [Pegg, 2011]. Recently, an unexpected role of AGTs in meiosis and early development under physiological conditions in *Caenorhabditis elegans* has reported [Serpe *et al.*, 2019]. Additionally, human methyltransferase (MGMT) protects cancer cells from the cytotoxic effect of alkylating agents-based chemotherapy [Kaina *et al.*, 2019]. For these reasons, this enzyme has crucial clinical importance, and the development of MGMT inactivators/inhibitors to be used in combination with this kind of chemotherapy is constantly evolving. Because of this clinical relevance, tests are needed that can measure methyltransferase activity reliably and rapidly. As explained in the Introduction Section, over the past forty years, assays based on the use of oligonucleotides carrying radioactive  $O^6$ -MG groups and chromatographic separations have been developed [Klein *et al.*, 1992; Luu *et al.*, 2002; Tintorè *et al.*, 2010]. These assays are very time-consuming and dangerous. Recently, however, alternative fluorescence-based tests have been developed that allow simple and rapid detection of MGMT activity. Among these, a chemo-sensor was designed that couples the fluorescence change to the bond-breaking phase that occurs during the repair activity. [Beharry *et al.*, 2016; Wilson *et al.*, 2018] Other approaches employing optically labelled DNA aptamers, [Tintorè *et al.*, 2010] DNA-based electrochemical sensors [Chen *et al.*, 2020] or double-stranded DNA oligonucleotides containing  $O^6$ -MG (dsDNA) competing with fluorescent substrates [Elder *et al.*, 1994; Goodtzova *et al.*, 1997; Vettone *et al.*, 2016; Morrone *et al.*, 2017; Mattosovich *et al.*, 2020] have also been proposed. The

latter tests have several advantages, including ease of use and high sensitivity, but are unable to test the activity of all AGTs and perform efficient monitoring based on the activity of DNA repair enzymes.

DNA Nanotechnology offers a promising approach to rationally design responsive DNA-based nanodevices and nanoswitches that can be applied for many different applications including sensing [Zhang *et al.*, 2018; Ranallo *et al.*, 2018]. Because of the high programmability and predictability of DNA/DNA interactions [Seeman *et al.*, 2018], the possibility to synthesize at low cost DNA strands modified with different signalling labels or recognition elements, it is, in fact, straightforward to design DNA sequences that can respond to chemical and biological stimuli and give a measurable optical [Krishnan *et al.*, 2011; Rossetti *et al.*, 2018;] or electrochemical signal [Kang *et al.*, 2017]. Many DNA-based devices have been developed in the recent years for the detection of complementary nucleic acid sequences [Chandrasekaran *et al.*, 2016; Das *et al.*, 2020], proteins [Farag *et al.*, 2021], antibodies [Ranallo *et al.*, 2015] and small molecules [Rossetti *et al.*, 2018]. Moreover, employing non-conventional DNA/DNA interactions, many groups have demonstrated the possibility to detect metal ions and pH variations using T-T mismatches [Porchetta *et al.*, 2013] and pH-responsive domains like cytosine-rich i-motifs [Xu *et al.*, 2017] and DNA triplexes [Hu *et al.*, 2017; Patino *et al.*, 2019; Chandrasekaran *et al.*, 2018]. Therefore the DNA nanotechnology offers the possibility to create DNA nanodevices to monitor DNA repair activity.

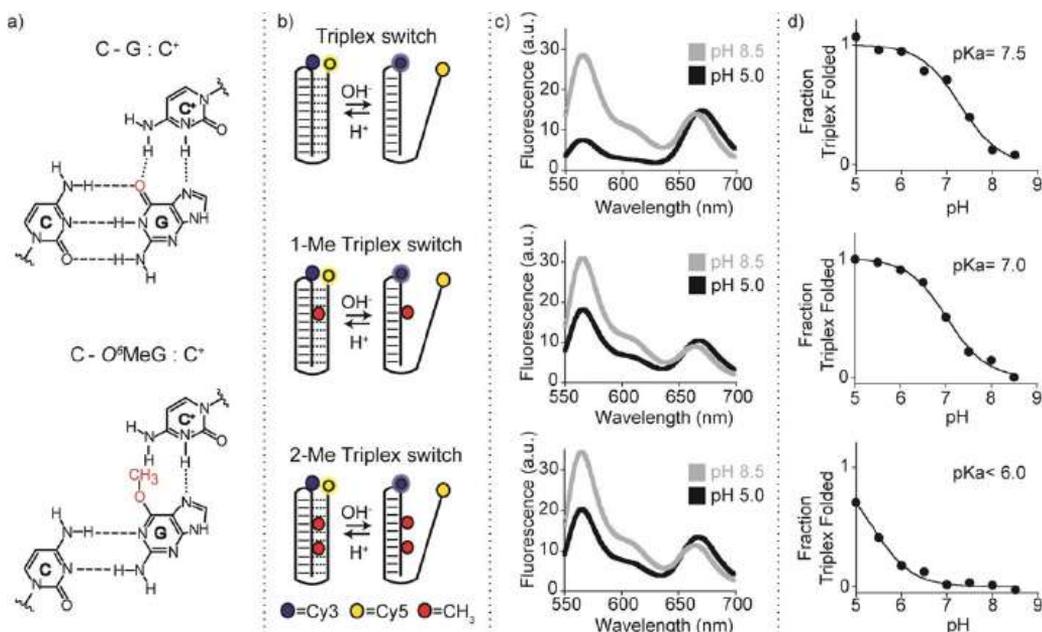
## *Results and discussion.*

For the evaluation of methyltransferase activity, in collaboration with Prof. Ricci of the University of Rome – Tor Vergata, a single-stranded DNA oligonucleotide-based DNA Triplex (defined as *nanoswitches*) was designed, able of forming an intramolecular triplex structure through hydrogen bonds (Hoogsteen interactions) between a hairpin duplex domain and a triplex-forming portion a single filament. [Idili *et al.*, 2014]

To this aim, three nanoswitches were synthesized showing the same triplex forming domains (i.e., 4 cytosines + 6 thymines) but different in O<sup>6</sup>-MG content in the hairpin duplex domain (0, 1 and

2). It was initially hypothesized that the presence of an  $O^6$ -MG on DNA nanoswitches (a natural substrate of AGT) could hinder the formation of the Hoogsteen hydrogen bond with cytosine on single-stranded DNA (Figure 19a), thus influencing the pH-dependent triplex folding/unfolding behaviour of the nanoswitch. The nanoswitches are also labelled with a pH-insensitive FRET pair to follow pH-dependent folding/unfolding (Figure 19b). Specifically, a cyanine-3 fluorophore (Cy3) is internally conjugated in the loop of the hairpin duplex DNA, and a cyanine-5 fluorophore (Cy5) is linked at the 3'- end of the triplex-forming DNA portion.

To prove the effect of  $O^6$ -MG on the triplex formation, we tested the pH-dependent folding/ unfolding behaviour of our three nanoswitches by measuring the fluorescence signal of our FRET pair at a fixed concentration of the nanoswitch and different pH values. Our control triplex nanoswitch (T, Figure 19, *top*), which lacks of any  $O^6$ -MG in its sequence, shows signals that are consistent with the formation of a folded triplex at acid pH (pH 5.0), suggesting the deployment of the triplex structure at basic pH values (pH 8.5) [Farang *et al.* 2021]. The pH of semi-protonation (here defined as pKa, the average pKa due to several interacting protonation sites) for this triplex nanoswitch is 7.5 (Figure 19 c,d *top*). The presence of one  $O^6$ -MG (T1M, Figure 19, *centre*) in the hairpin duplex (at position 7) strongly undermines triplex formation; as a result, we observe at pH 5.0 a FRET signal that is consistent with a partially unfolded triplex structure and we obtain a pKa of 7.0 (Figures 19 c,d *centre*) [Farang *et al.* 2021]. Finally, the nanoswitch containing two  $O^6$ -MG (T2M, at position 4 and 7; Figure 19, *bottom*) is even more destabilized and the pH titration curve with this switch does not reach a plateau at lower pH values, suggesting a pKa lower than 6.0 (Figures 19 c,d *bottom*) [Farang *et al.* 2021]. These results were confirmed by melting and urea denaturation experiments performed at different pHs e with CD spectra experiments [Farang *et al.* 2021].



**Figure 19:** Folding-upon-repair DNA nanoswitches. a) DNA parallel triplets formed between two cytosines and one guanine and involving Watson–Crick (three hydrogen bonds) and Hoogsteen interactions (two hydrogen bonds) require the protonation of the  $N^3$  of cytosine in the third strand (top, left). Methylation of the guanine in the  $O^6$ -position destabilizes the triplex conformation. b) Programmable DNA-based triplex nanoswitches designed to form an intramolecular triplex structure with 0 (T), 1 (T1M) or 2 (T2M)  $O^6$ -MG in the sequence. c) Fluorescence spectra were obtained for each nanoswitch at pH 5.0 and pH 8.5. d) pH-titration curves of the triplex nanoswitches. Triplex-to-duplex transition is monitored through a pH-insensitive FRET pair at the 3'-end (Cy5) and internally located (Cy3). The pH titration experiments were performed at 25 °C, [nanoswitch] = 50 nM by measuring the fluorescence signal at different pH values in PBS 1X buffer.

To better understand the effect that guanine methylation has on the stability of our triplex nanoswitches, an overall molecular dynamics (MD) simulations was performed. As expected, the control T is the most stable, followed by the T1M and then by the T2M. These results provide an atomic-level understanding of the triplex stability, indicating that the efficient Hoogsteen base pairing is critical C-G:C<sup>+</sup> triplet and shows that the  $O^6$ -methylation of guanine hampers the efficient interaction with the Hoogsteen pairing cytosine.

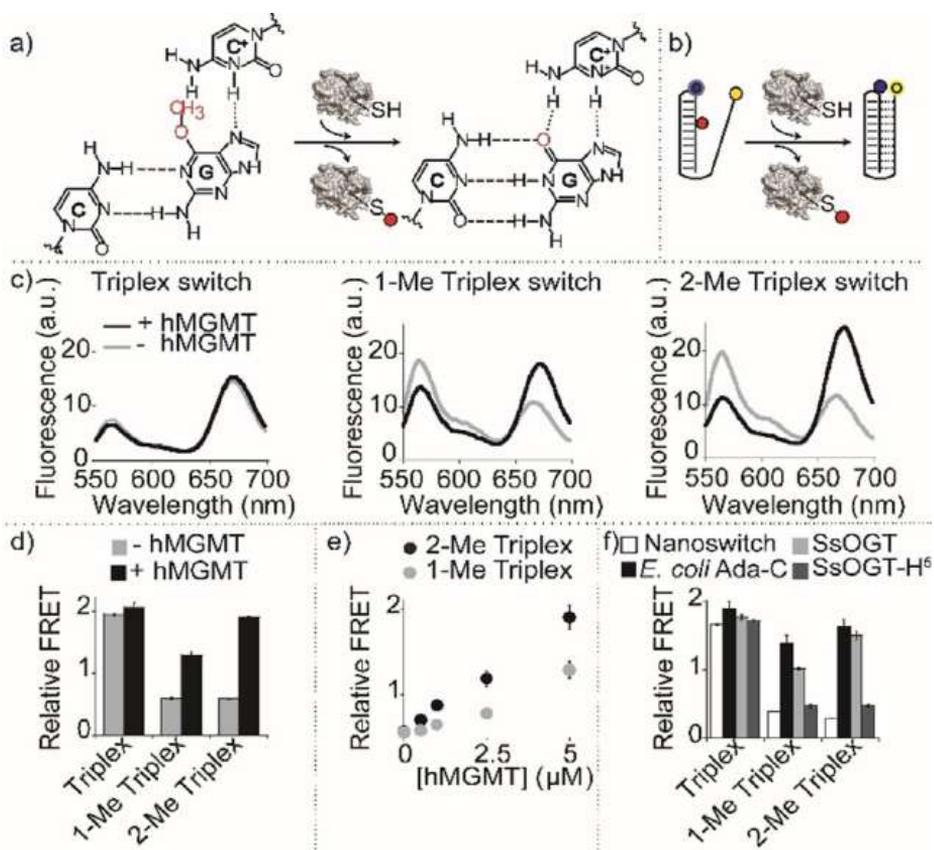
These results convinced us to utilise these triplex nanoswitches as suitable substrate for monitoring the methyltransferase activity of AGTs. Indeed, the enzymatic removal of the methyl group of  $O^6$ -

MG in the nanoswitches would re-establish the optimal conditions for the Hoogsteen interactions and the triplex formation by the nanoswitch. This was achieved with a series of experiments. By the above-mentioned indirect assay using a fluorescein derivative of O<sup>6</sup>-benzyl-guanine (BG-VG) [Perugino et al.,2012;Vettone et al.,2015; Mattossovich et al.,2020; 33], I confirmed that T1M and T2M are suitable substrates for AGTs (Figure 20a). I first demonstrated that no interaction occurs between the nanoswitch and the fluorescently labelled substrate in the absence of the protein. As expected, separate visible bands can be observed for both the BG-VG fluorescent derivative and the three DNA nanoswitches tested (Figure 20b). Several AGTs were pre-incubated with DNA nano-switches and then we added an equimolar amount of the BG-VG, which interacts with the free enzymes in the solution (Figure 20b). In the case of MGMT, a fluorescent protein band was observed in the presence of T, demonstrating that no repair reaction occurred on this DNA oligonucleotide (Figure 20b). Conversely, the reaction with the T1M and the T2M showed a complete absence of the fluorescent band (Figure 20b). This is due to the suicide nature of these class of enzymes, [Pegg 2011] since the methyl group is irreversibly transferred to the enzyme, which is no longer available for the fluorescent BG-VG, thus demonstrating that methylated nanoswitches are effective substrates for hMGMT. The same results were further confirmed using the SsOGT enzyme [Morrone et al.,2017; Mattossovich et al 2020; Perugino et al.,2012,2015] and, mainly, the homolog of *E. coli* (Ada-C). In the latter case, it is noteworthy indicate that *Ada-C is completely insensitive to O<sup>6</sup>-BG derivatives*, hampering any fluorescent indirect assay with these kind of substrates [Elder et al.,1994; Goodtzova et al.,1997]. The absence of any fluorescent bands, despite being correctly loaded confirmed by coomassie staining, demonstrated its total insensitivity to BG-VG (Figure 20b). However, because the migration of T2M was similar to T and T1M after the reaction with Ada-C, whereas in absence of any AGT the former has a migration different from the latter (Figure 20b), suggests that an enzymatic activity took place. So these innovative DNA Triplexes could be proposed as *universal AGTs' substrates*.

The methylated DNA triplex nanoswitches are also highly specific as they showed no enzymatic activity when incubated with H<sup>5</sup>, that was previously reported to be catalytically active on O<sup>6</sup>-BG



It was then tested whether methyltransferase repair activity on our nano-switches could result in a conformational switch and consequent measurable change in the FRET signal, recommending a means of direct detection of MGMT enzyme activity (Figure 21 b). As expected, no significant difference can be observed in the FRET signal of the control T, before ( $1.9 \pm 0.1$ ) and after ( $2.1 \pm 0.1$ ) the MGMT incubation (Figure 21c, *left*). Under the experimental conditions used, both fluorescence spectra suggest that the unmethylated triplex switch is fully bent (at pH 5.0). The same experiment conducted using T1M and T2M shows, however, a strong difference in FRET signals before and after MGMT incubation (Figure 21c, *centre, right*) confirming a partially explained configuration. The MGMT enzymatically removes the methyl group in the  $O^6$  position in the nano-triplex switches and restore their ability to form a triplex structure. It is particularly clear for the T2M: it shows a FRET signal ( $1.9 \pm 0.1$ ) after MGMT incubation i.e. within the error of the T ( $2.1 \pm 0.1$ ) (Figure 21 d). Both methylated triplex nanoswitches show a change in the relative FRET values that are linearly dependent on the concentration of MGMT in the range between 0.5 and 5 mM ( $R^2 = 0.99$ ,  $R^2 = 0.95$ ) (Figure 21 e). Similar experiments were also performed with other AGTs, (Figure 21 f). As expected,  $H^5$  produced no effect on the FRET signal of both nanoswitches (Figure 21 f), once again confirming the specificity of the folding-upon-repair mechanism. To demonstrate the possibility of using this platform in more complex media, we have also performed the measurement of the enzymatic activity of MGMT in 10% serum and observed FRET signal changes well distinguishable from the control experiment in absence of enzyme [Farag *et al.*, 2021]. Given that, this innovative approach could be useful for the determination of the <sup>m</sup>DNA content, as well as the study and characterization of novel MGMT inhibitors as possible drug candidates [Kaina *et al.*,2019]. In this case, these nanoswitches were tested to measure the activity of the human enzyme in the presence of inhibitors, inactivators or pseudo-substrates.



**Figure 21:** Detection of methyltransferase activity with triplex nano-switches. a) AGTs activity on T1M and T2M leads to folding-upon-repair of the triplex DNA structure. b) Enzymatic activity detection can be achieved by monitoring folding/unfolding of the Triplex nanoswitch by fluorescence FRET signalling. c) Spectra and d) relative FRET signals obtained with the Triplex nanoswitches before and after incubation with 5 mM (0.1 mg mL<sup>-1</sup>) MGMT. e) Relative FRET signals observed at different MGMT concentrations for T1M and T2M. f) Relative FRET signals with different AGTs. Spectra were obtained by excitation at 530 ± 5 nm and acquisition between 545 and 700 nm ± 10 nm. Relative FRET signals were obtained by first incubating the DNA nanoswitches (0.5 mM) in the absence or presence of AGTs (0.1 mg mL<sup>-1</sup>) in 10 mL solution of PBS 1X pH 7.5 at 30 °C for 60 minutes. The reaction mixtures were then diluted to 100 mL using PBS 1X at pH 5.0, and heat-inactivated for 2 minutes at 70 °C before performing the fluorescence spectra at 25 °C.

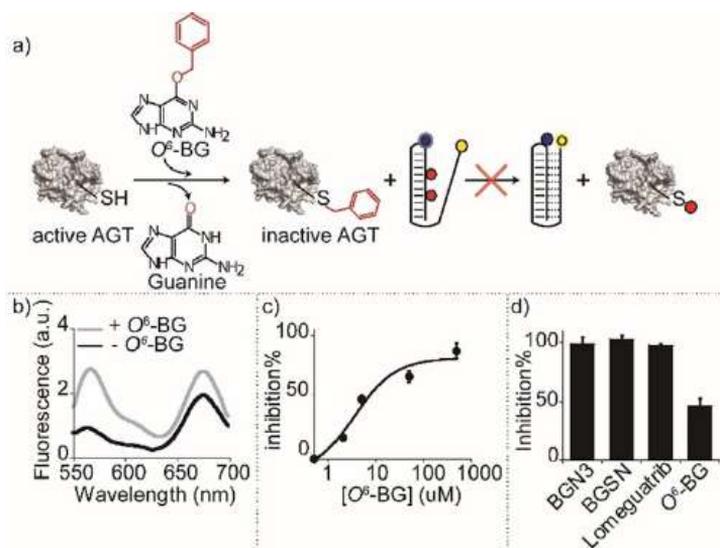
Adding an MGMT inhibitor to the reaction mixture should prevent the enzyme from repairing the methylated DNA and no significant change in the nanoswitch FRET signal should be observed (Figure 22 a). As a first evaluation, O<sup>6</sup>-BG, a widely characterized methyltransferase inactivator, was tested [Reinhard *et al.*,2001]. The relative FRET signal of the nanoswitch incubated with MGMT

and  $O^6$ -BG (Figure 22 b, *gray line*) is, at  $0.8 \pm 0.1$ , similar to the values observed in the absence of enzyme (Figure 21 b, *black line*) an effect consistent with the inactivation of MGMT activity by  $O^6$ -BG. The measured inactivation is concentration dependent, it increases with an increase in inhibitor concentration with a measured  $IC_{50} = 3.5 \pm 2$  mM (Figure 22 c). Furthermore, this platform is able to measure the inactivation efficiency of several molecules including 4-azido-N- (4- (hydroxymethyl) benzyl) butanamide (BGN3) (N- (4 - (((2-amino- 9H-purin -6-yl) oxy) -methyl) benzyl)-4-azidobutanamide (BGSN), and Lomeguatrib, *et al.*, 2000; Merlo *et al.*, 2019; Farag *et al.* 2021], demonstrating once again the versatility of this assay (Figure 22 d).

## Conclusions

Intramolecular DNA nanoswitches containing  $O^6$ -MG residues undergo the duplex-to-triplex conformational transition after enzymatic de-methylation. I demonstrate that the presence of a methyl group in the  $O^6$  position of a guanine in the duplex portion of the nanoswitch strongly affects the formation of the triplex. Molecular simulations were used to establish how methylation of guanine bases prevents the efficient formation of Hoogsteen interactions with cytosine in the triplex formation domain, resulting in a local unwinding of the triplex structure.

Triplex-based methylated nanoswitches are versatile tools for direct measurement of AGT activity. This innovative approach is applicable for all AGTs, as I successfully demonstrated by using evolutionarily distant AGTs from all the domains of life, which efficiently recognize these DNA oligonucleotides, leading to measurable changes in the FRET signal after the reaction. The fact that these enzymes show different reaction conditions and, mainly, different substrate specificities [Goodtzova, 1997; Pegg, 2011; Perugino, 2012], suggest that these DNA-based nanoswitches are *universal AGT substrates*. Finally, it was also measured the activity of MGMT in the presence of enzymatic inactivators, demonstrating the possibility of using our nanoswitches for the screening of new potential inactivators/pseudo-substrates/inhibitors of MGMT [Kaina, 2019].



**Figure 22:** Detection of AGTs inhibitors. a) Pre-incubation of the enzyme with an inhibitor (here O<sup>6</sup>-BG) leads to irreversible inhibition of the enzyme that is thus not able to demethylate the DNA nanoswitch. b) Spectra obtained with the T2M in the presence of MGMT (5 mM) before and after incubation with the inhibitor O<sup>6</sup>-BG (5 mM). c) % Inhibition plot obtained at different concentrations of O<sup>6</sup>-BG inhibitor. d) % Inhibition obtained with the T2M with an equimolar concentration of different enzymatic inhibitors and MGMT (5 mM). Spectra were obtained by excitation at  $530 \pm 5$  nm and acquisition between  $545$  and  $700$  nm  $\pm 10$  nm. Relative FRET signals were obtained by enzymatic incubation of 5 mM MGMT and inhibitor at 30 °C for 60 minutes in 10 mL PBS 1x at pH 7.5. T2M (0.5 mM) was then added to the reaction mixtures and incubated for another 60 min at 30 °C. The reaction mixtures were diluted to 100 mL using PBS 1X at pH 5.0, and heat-inactivated for 2 minutes at 70 °C before performing the fluorescence spectra at 25°C.

The triplex-based nanoswitch assay offers several advantages over other conventional methodologies currently used to test AGTs' activity, as largely explained in the Introduction Section. The strength of this innovative substrate is that it gives a *direct and in real-time assay*; quick, allowing AGT activity to be measured in minutes; safety; and it resemble to a natural AGT substrate (<sup>m</sup>DNA).

This results could be a *proof of concept*, suggesting to rationally design other fold-based DNA nanoswitches as application tools in the measurement and determination of a wide range of enzymes involved in other DNA repair activities. Since mutations on the nitrogenous bases could affect the DNA-DNA thermodynamic

interactions, it would, in fact, be simple to design programmable DNA nanoswitches that can undergo a conformational change in response to a specific DNA repair activity. As DNA repair activities are involved in many diseases, including cancer, these tools could be useful as high-throughput platforms for drug discovery.

## Experimental Procedures

### *Reagents and Materials*

All reagent-grade chemicals, including  $\text{Na}_2\text{HPO}_4$ ,  $\text{NaCl}$ , Phosphate buffered saline, Dimethyl sulfoxide (DMSO) (all from Sigma-Aldrich, Italy), were used as received. SNAP-Vista Green<sup>®</sup> fluorescent substrate was from New England Biolabs (Ipswich, MA). Synthetic oligonucleotides were from Eurofins (Milan, Italy); *Pfu* DNA polymerase was from NZYTech (Portugal). The Bio-Rad protein assay kit (Bio-Rad Pacific) was used for the determination of the protein concentration, using purified BSA as standard. The inhibitor *O*<sup>6</sup>-benzyl-guanine (*O*<sup>6</sup>-BG) was from Sigma-Aldrich, (Italy). The 4-azido-N-(4-(hydroxymethyl) benzyl) butanamide (BGN3) and the (N-(4-(((2-amino-9H-purin-6-yl)oxy)methyl)benzyl)-4-azidobutanamide (BGSN) were a gift of Prof. Alberto Minassi (University of Piemonte Orientale, Novara). Lomeguatrib (5 mg) was obtained as a dry white powder from Sigma-Aldrich, St. Louis, Missouri.

### *Oligonucleotides*

Oligonucleotides and DNA-based receptors employed in this work were synthesized, labelled, and purified (HPLC and reverse-phase) by IBA GmbH (Göttingen, Germany) and used without further purification. Unless otherwise stated the labeled oligonucleotides were dissolved in phosphate buffer at a concentration of 100  $\mu\text{M}$ . The final concentration of the oligonucleotides was confirmed using Tecan Infinite M200pro (Mannedorf, Switzerland) through the NanoQuant Plate. The sequences of the DNA constructs are reported in table.

### *DNA constructs, protein expression and purification*

The commercial MGMT cDNA (from a glycerol stock MGC Human MGMT Sequence-Verified Dharmacon) was cloned in the Quiagen pQE31TM (Hilden, Germany) as previously described for

SsOGT and its relative mutant H5,<sup>[1]</sup> by using specific synthesized primers (Eurofins):MGMT-fwd and MGMrv which possess the BamH I and Hind III (New England Biolabs, Ipswich, MA) cloning sites, respectively. The insertion of fragment (628 bp) in the pQE31 vector was in frame and downstream of a hexahistidine tag, for the subsequent purification procedures.

The cloning procedures for E. coli Ada-C gene followed a different strategy. The genomic DNA of E. coli strain DH5 $\alpha$  was used as the DNA template for the gene amplification by using primers Lig3/Lig5AdaC (table 8)

A DNA fragment of 580 bp was obtained and directly cloned into the expression vector pHTP1 (NZYTech, Portugal). The ligation mixture was entirely used to transform commercial E. coli DH5 $\alpha$  cells (NZY5 $\alpha$  Competent Cells- NZYTech, Portugal). For both the cloning procedures positive clones were confirmed by DNA sequencing. All AGT proteins were expressed and purified by following a previously described protocol.

name	sequence	note
<b>Triplex (T)</b>	5'-aaggaagaagttt(cy3)cttctcctctttgttcctcttc(cy-5)-3'	Farad et al 2021
<b>Triplex 1Me</b>	5'-aaggaa(O6-MG)aagttt(cy3)cttctcctctttgttcctcttc(cy5)-3'	Farad et al 2021
<b>Triplex 2Me</b>	5'-aag(O6-MG)aa(O6-MG)aagttt(cy3)cttctcctctttgttcctcttc(cy5)-3'	Farad et al 2021
<b>MGMT fwd</b>	5'-aatgatggatccaatggacaaggattgtg-3'	Farad et al 2021
<b>MGMT rv</b>	5' ttcgatcaagcttatcagtttcggccagcagg 3'	Farad et al 2021
<b>Lig5AdaC</b>	5'tcagcaagggctgaggccatggcggctaacaattcc-3'	Farad et al 2021
<b>Lig3AdaC</b>	5'-cctcagcgggaagctgaggttacctctcctcattttcagc-3'	Farad et al 2021

**Table 8** :list of oligonucleotides

### *SDS-PAGE gel-imaging AGT assay*

This method is used for the evaluation of AGTs activity with nanoswitches by using a fluorescein derivative of the O<sup>6</sup>-BG (SNAP-Vista Green<sup>®</sup>) as substrate. In each sample, 5.0  $\mu$ M of protein (0,1 mg/mL) was incubated at a specific temperature (30°C for hMGMT and Ada-C; 40°C for SsOGT and SsOGT-H<sup>5</sup>) with 10  $\mu$ M of each nanoswitch in a total volume of 10  $\mu$ L of Reaction Buffer (50 mM phosphate, 150 mM NaCl; pH 6.5) for 60 min. Then, SNAP-Vista Green<sup>®</sup> was added to the solution at a final concentration of 5  $\mu$ M and incubated again at the same temperature and time. Samples were boiled and directly loaded on 15% acrylamide SDS-PAGE. Bands were detected by direct gel-

*imaging* using the VersaDoc 4000™ system (Bio-Rad), performing a double acquisition by applying as excitation/emission parameters a BLUE LED/530 and GREEN LED/605 bandpass filters, for the determination of fluorescent-labeled proteins and Cy5-based triplexes, respectively. Then, gels underwent *Coomassie staining* for the determination and correction of the protein amount loaded.

### *Fluorescence measurements*

Fluorescence measurements were carried out on a Cary Eclipse Fluorimeter (Varian), setting excitation wavelength to  $\lambda_{\text{ex}} = 530$  nm ( $\text{slit}_{\text{ex}} = 5$  nm) and acquisition between 545 and 700 nm ( $\text{slit}_{\text{em}} = 10$  nm) using quartz cuvettes of microvolume (100  $\mu\text{L}$ ). All measurements were performed at  $T = 25$  °C in 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 250 mM NaCl at pH 5.0. For detection of methyltransferase activity, nanoswitches were first diluted in 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 150 mM NaCl at pH 7.5 to a concentration of 1  $\mu\text{M}$ . Then in a 10  $\mu\text{L}$  solution of 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 150 mM NaCl, we prepared an enzymatic reaction mixture of 0.5  $\mu\text{M}$  nanoswitches in the absence and presence of 5  $\mu\text{M}$  AGTs (0.1  $\mu\text{g}/\mu\text{l}$ ). The reaction mixtures were incubated at a certain temperature and time for each enzyme; (60 minutes at 30°C for MGMT and *E. coli* Ada-C) and (30 minutes at 70 °C for the thermostable SsOGT and SsOGT-H<sup>5</sup>). All the mixtures were heat-inactivated after the enzymatic reaction by incubating at 70 °C for 2 minutes (when using the thermostable SsOGT and SsOGT-H<sup>5</sup> heat inactivation was performed at 90 °C for 10 minutes). Fluorescence measurements were conducted at 25 °C by diluting the reaction mixtures to 100  $\mu\text{l}$  using 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 250 mM NaCl at pH 5.0. Test of inhibitor activity was performed by first dissolving all the inhibitors in DMSO to a concentration of 10 mM. Inhibition reactions were conducted in an equimolar concentration of MGMT (0.125  $\mu\text{g}/\mu\text{l}$ ) and different inhibitors (5  $\mu\text{M}$ ) at 30 °C for 60 minutes in a 10  $\mu\text{l}$  solution of 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 150 mM NaCl at pH 7.5. 0.5  $\mu\text{M}$  of 2-Me triplex nanoswitch was then added to the reaction mixtures to interact with the remaining active MGMT and incubated for another 60 minutes at 30 °C. The reaction mixture was heat-inactivated after the enzymatic reaction by incubating at 70 °C for 2 minutes. Fluorescence measurements were conducted at 25 °C by diluting the reaction mixtures to 100  $\mu\text{l}$  using 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 250 mM NaCl at pH 5.0.

### Fluorescence data analysis

The ratiometric FRET has been calculated as follows:

$$\text{Rat. FRET} = F_{\text{Cy5}} / F_{\text{Cy3}}$$

Where  $F_{\text{Cy5}}$  is the maximum fluorescence emission of Cy5 ( $\lambda_{\text{em}} = 670 \text{ nm}$ ) and  $F_{\text{Cy3}}$  is the maximum fluorescence emission of Cy3 ( $\lambda_{\text{em}} = 565 \text{ nm}$ ). The pH titration curves were obtained by plotting

$$\text{Triplex Fraction} = \text{Ratiometric FRET} + \left( \frac{[\text{H}^+] * (\text{Rat. FRET}_{\text{Triplex}} - \text{Rat. FRET}_{\text{Duplex}})}{[\text{H}^+] * K_D} \right)$$

Rat.FRET vs pH, and fitting the data with the following Langmuir-type equation:

Where  $\text{Rat.FRET}_{\text{Triplex}}$  and  $\text{Rat.FRET}_{\text{Duplex}}$  represent the FRET signal of the Triplex Switch in the triplex (closed, pH = 5.0) and duplex state (open, pH = 8.5), respectively, and where  $[\text{H}^+]$  represents the total concentration of hydrogen ions and  $K_D$  is the observed acid constant for the switch.

Inhibition percentage for different inhibitors was calculated as follows:

$$\% \text{inhibition} = \left[ 1 - \frac{(\text{Rat. FRET}_{\text{inh}} - \text{Rat. FRET}_{2\text{-Me Triplex}})}{(\text{Rat. FRET}_{2\text{-Me Triplex, unmethylated}} - \text{Rat. FRET}_{2\text{-Me Triplex}})} \right] \%$$

Where  $\text{Rat.FRET}_{\text{inh}}$  is the FRET signal of the 2MT nanoswitch measured after the pre-incubation of different inhibitors with (0.1  $\mu\text{g}/\mu\text{l}$ ) MGMT at pH 5.0,  $\text{Rat.FRET}_{2\text{-Me triplex, unmethylated}}$ , and  $\text{Rat.FRET}_{2\text{-Me Triplex}}$  represent the FRET signal of the 2MT switch at pH = 5.0) after the incubation with (0.1  $\mu\text{g}/\mu\text{l}$ ) MGMT and in the absence of MGMT at pH = 5.0, respectively.

## Molecular Dynamics (MD) simulations

### Simulation protocol

All simulated model systems were composed of a single DNA molecule sequence 5'-AAGGAAG [TTT] CTTCTTCCTT [CTTTG] TTCTTCTTC-3' placed in 8 nm x 8 nm x 13 nm rectangular box, solvated with ~27,250 water molecules and sodium ions to neutralize the system. Based on this sequence, four triplex molecules were considered, composed of (a) conventional DNA bases; (b) protonated cytosines (at positions shown in italics and underlined in the sequence above); (c) protonated cytosines and *O*<sup>6</sup>-methylated guanine at position 7; (d) protonated cytosines and *O*<sup>6</sup>-methylated guanines at position 4 and 7 (bold, underlined). The initial structure of single-stranded DNA was obtained using the x3DNA package.<sup>[3]</sup> The DNA has been represented with Amber ff12SB force field, which includes the ff99bsc0 corrections for DNA.<sup>[4,5]</sup> This is a well-established force field for nucleic acid, which we have successfully employed in several recent articles.<sup>[6,7]</sup> The TIP3P model was used to represent water.<sup>[8]</sup> The electrostatic potential around each of the modified nucleobases (protonated cytosine and *O*<sup>6</sup>-methylated guanine) were computed from the geometries optimized at HF/6-31G\* level of theory with Merz-Kollman ESP fitting with Gaussian 09.<sup>[9]</sup> Next, this potential was used to compute atomic partial charges of the bases with Amber 16 package.<sup>[10]</sup>

MD simulations were performed with Gromacs 5 package<sup>[11]</sup> with Plumed 2.2.3 plugin.<sup>[12]</sup> The simulations were performed in NPT ensemble with the temperature kept at 310 K with v-rescale thermostat<sup>[13]</sup> and pressure at 1 bar with Parrinello-Rahman algorithm.<sup>[14]</sup>

Periodic boundary conditions were applied and the particle mesh Ewald algorithm<sup>[15]</sup> was used to compute long-range electrostatic interactions with a real-space cut-off of 1 nm. All covalent bonds including hydrogen were restrained with the LINCS algorithm.<sup>[16]</sup> The leap-frog verlet algorithm was used to integrate equations of motion with a time step of 2 fs.

### *Single-stranded DNA folding to triplex conformation*

The structure of the DNA triplex was generated through *de-novo* design over ~60 ns of steered MD simulation of single-stranded DNA (ssDNA) with deprotonated cytosines. As a reference structure, the example of triplex topology from x3DNA was used. DNA folding was performed in four consecutive steps: initially the root mean squared deviation (RMSD) with respect to the duplex structure was used as a reaction coordinate for heavy atom positions of nucleobases 1 to 10 and 14 to 23. During the first ~20 ns of the simulation, the center of a one-sided harmonic potential was moved with a constant speed from the initial value of the reaction coordinate (2.2 nm) to 0 nm with a force constant gradually increased from 500 kJ nm<sup>-2</sup> to 7000 kJ nm<sup>-2</sup>. At the same time, the RMSD to the initial position of nucleobase heavy atoms of residues 14 to 38 was restrained using a harmonic biasing potential with a force constant of 7000 kJ nm<sup>-2</sup>. Next, the restraint for nucleobases 14 to 38 was removed and RMSD with respect to the nucleobase heavy atom position in the triplex structure was used for residues 1 to 10, 14 to 23 and 29 to 31 was used as a new reaction coordinate. In the next 8 ns of the simulation, the center of a one-sided harmonic potential was moved with a constant speed from the initial value of the reaction coordinate (2.3 nm) to 0 nm with a force constant gradually increased from 500 kJ nm<sup>-2</sup> to 7000 kJ nm<sup>-2</sup>. This step was further repeated for residues 1 to 10, 14 to 23, 29 to 34 and for residues for residues 1 to 10, 14 to 23, 29 to 38. Finally, additional 16 ns of simulation with restraint was produced in order to relax the position of the DNA backbone. The comparison of the final structure obtained with steered MD simulations to the reference triplex structure from x3dna resulted in RMSD for the nucleobase heavy atom position below 0.08 nm.

### *Conventional MD simulations and analysis*

The initial structures of the protonated and methylated triplex switches were obtained by substituting the modified bases in the final structure of deprotonated triplex and subsequent minimization and 1 ns long simulation, during which RMSD of the initial position of nucleobase heavy atoms for residues 1 to 10, 14 to 23 and 29 to 38 was restrained using a harmonic biasing potential with a force constant of 7000 kJ nm<sup>-2</sup>. For each of the considered systems, 3

replicas of 500 ns conventional MD simulation were performed. DNA structural parameters from the trajectories were computed using do\_x3dna plugin<sup>[17]</sup> to VMD. All molecular images were created using VMD.<sup>[18]</sup> The graphs were prepared using matplotlib library of python.<sup>[19]</sup>

### *Free energy simulations*

The free energy profiles for the triplex to duplex transition were computed in four considered systems (i.e., protonated unmethylated triplex; protonated 1-Me triplex; protonated 2-Me triplex; deprotonated unmethylated triplex) using replica-exchange umbrella sampling (REUS).<sup>[20]</sup> We used as a reaction coordinate (RC) the distance between two center of mass (COM): (i) the heavy atoms of 3'-terminal cytidine phosphate and (ii) the nucleobases that co-form the triplex plane with the terminal cytosine (panel A, Figure S13). The initial configurations for REUS simulations were obtained by performing ~600 ns-long steered-MD simulations, in which the distance between terminal cytidine and the nucleobases was increased from 1.15 nm to 8.35 nm using a one-sided harmonic potential with a force constant of 2500 kJ nm<sup>-2</sup>. The reaction coordinate was divided into 38 windows, with 0.1 nm spacing in the region of the RC from 1.15 to 1.35 nm, and with 0.2 nm spacing in the remaining region of the RC. A force constant of 1000 kJ nm<sup>-2</sup> was used to restrain the RC at the given distance. 500 ns long trajectories were obtained for each REUS window. For each window, the first ~100 ns were discarded from the analysis (being part of the system's equilibration) and the final free energy profiles were obtained considering the equilibrated runs using the weighted histogram analysis method (WHAM).<sup>[21]</sup> Monte Carlo bootstrap method was used to estimate the uncertainties of the free energy profiles taking into account time series correlations.

To characterize the impact of both protonation and methylation on the stability of the triplexes, the free energy profiles for the triplex to duplex transition have been computed for (i) protonated unmethylated triplex; (ii) protonated and 1-Me triplex; (iii) protonated and 2-Me triplex and (iv) deprotonated unmethylated triplex (Figure S13). For the protonated unmethylated triplex (black line), the unfolding process is strongly disfavored in the whole range of RC. In the 1-Me (green) and 2-Me protonated (blue) triplexes, we observe an initial increase of the free energy (i.e., at

RC  $\sim$ 1.2 nm), corresponding to the cost for the contact break between the 3' terminal cytosine and the corresponding guanine. Then, the free energy profile reaches a plateau up to RC  $\sim$ 2.7 nm, corresponding to the unfolding process. This suggests that the unfolding could proceed spontaneously in this region. Next, above RC  $\sim$ 4.3 nm, the free energy rapidly increases with the RC for the 1-Me triplex (green), showing that the process is strongly disfavored, similarly to the protonated triplex (black). On the other hand, in the case of the 2-Me system (blue), we observe a lower increase of the free energy with the RC, suggesting that further unfolding is easier up to RC  $\sim$ 6.15 nm. For the deprotonated unmethylated triplex (red), we initially observe a drop in free energy up to RC  $\sim$ 1.6 nm, which suggests that the unfolding is initially favored, as the deprotonated terminal cytosine does not form efficient Hoogsteen interactions with the corresponding guanine. The subsequent region of the free energy profile shows similarity to the 2-Me triplex (blue), yet the free energy cost is lower. These results indicate that the stability of the triplex systems is as it follows: the most stable is the protonated unmethylated triplex (black), followed by the protonated 1-Me (green), then by the protonated 2-Me triplex (blue), while the least stable is deprotonated unmethylated triplex (red).

It is important to note that the computed values of the free energy profiles cannot be considered as absolute, due to the limitations of the force fields in describing single stranded DNA and the inability of the simulations to capture the protonation-deprotonation equilibria for cytosine. Nevertheless, the relative energetics depict a reasonable scenario in consistency with the experiments reported here. Accordingly, while the unfolding of the protonated triplex is highly disfavored, the cost for the unfolding of 1-Me triplex is moderately low in the region of RC  $\sim$ 1 to  $\sim$ 3.5 nm and the further unfolding is strongly disfavored. On the other hand, the free energy profiles for the 2-Me triplex and the deprotonated unmethylated triplex are comparable with each other, showing the lowest stability in the whole region of RC.

## REFERENCES

1. Amodio A, Zhao B, Porchetta A, Idili A, Castronovo M, Fan C, Ricci F. Rational design of pH controlled DNA strand displacement. *J. Am. Chem. Soc.* **2014**
2. Arana ME, Kunkel TA. Mutator phenotypes due to DNA replication infidelity. *Semin Cancer Biol.* **2010**
3. Beharry AA., Nagel ZD., Samson LD., Kool ET. Fluorogenic Real-Time Reporters of DNA Repair by MGMT, a Clinical Predictor of Antitumor Drug Response. *PLoS One.***2016**
4. Beranek DT. Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat Res.* **1990**
5. Bindra RS., Glazer PM. Co-repression of mismatch repair gene expression by hypoxia in cancer cells: role of the Myc/Max network. *Cancer Lett.* **2007**
6. Bock L.C.; Grin L.C.; Latham J.A.; Vermaas E.H.; Toole J.J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **1992**.
7. Cava F., de Pedro MA., Blas-Galindo E., Waldo GS., Westblade LF., Berenguer J. Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environ Microbiol* **2008**
8. Cava, F.; Hidalgo, A.; Berenguer, J. *Thermus thermophilus* as Biological Model. *Extremophiles* **2009**.
9. Chan M.K.; Mukund S.; Kletzin A.; Adams M.W.W.; Rees D.C. Structure of a hyperthermo-philic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **1995**.
10. Chandrasekaran AR., Wady H., Subramanian HK. Nucleic Acid Nanostructures for Chemical and Biological Sensing. *Small.* **2016**
11. Chatterjee N., Lin Y., Santillan BA., Yotnda P., Wilson JH. Environmental stress induces trinucleotide repeat mutagenesis in human cells. *Proc Natl Acad Sci U S A.* **2015**.

12. Chatterjee N., Lin Y., Wilson JH. Mismatch repair enhances convergent transcription-induced cell death at trinucleotide repeats by activating ATR. *DNA Repair*. **2016a**
13. Chatterjee N., Lin Y., Yotnda P., Wilson JH. Environmental Stress Induces Trinucleotide Repeat Mutagenesis in Human Cells by Alt-Nonhomologous End Joining Repair. *J Mol Biol*. **2016b**
14. Chatterjee N., Walker G.C. Mechanisms of DNA damage, repair and mutagenesis. *Environ Mol Mutagen*. **2017**.
15. Chautard H.; Blas-Galindo E.; Menguy T.; Grand'Moursel L.; Cava F.; Berenguer J.; Delcourt M. An Activity-Independent Selection System of Thermostable Protein Variants. *Nat. Methods* **2007**.
16. Chen Y., Huang C., Bai C., Gao H., Ma R., Liu X., Dong Q. Benzo[alpha]pyrene repressed DNA mismatch repair in human breast cancer cells. *Toxicology*. **2013**.
17. Connors S.B., Mongodin E.F., Johnson M.R., Montero C.I., Nelson K.E., Kelly R.M. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS Microb. Rev*. **2006**
18. Crutzen P.J., Andreae MO. Biomass burning in the tropics: impact on atmospheric chemistry and biogeochemical cycles. *Science*. **1990**.
1. Daniels D.S.; Mol C.D.; Arvai A.S.; Kanugula S.; Pegg A.E.; Tainer J.A. Active and alkylated human AGT structures: A novel zinc site, inhibitor and extrahelical base binding. *EMBO J*. **2000**
2. Daniels D.S.; Woo T.T.; Luu K.X.; Noll D.M.; Clarke N.D.; Pegg A.E.; Tainer J.A. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol*. **2004**
3. Das J, Kelley SO. High-Performance Nucleic Acid Sensors for Liquid Biopsy Applications. *Angew Chem Int Ed Engl*. **2020**
4. Dasari S., Tchounwou PB. Cisplatin in cancer therapy: molecular mechanisms of action. *Eur J Pharmacol*. **2014**
5. Davies RJ. Royal Irish Academy Medal Lecture. Ultraviolet radiation damage in DNA. *Biochem Soc Trans*. **1995**.

6. Del Prete, S.; Merlo, R.; Valenti, A.; Mattosovich, R.; Rossi, M.; Carginale, M.; Supuran, C.T.; Perugino, G.; Capasso, C. Thermostability enhancement of the carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* by using the anchoring-and-selflabelling-protein-tag system (ASL<sup>tag</sup>). *J. Enz. Inhib. Med. Chem.* **2019**.
7. DeVita VT Jr, Chu E. A history of cancer chemotherapy. *Cancer Res.* **2008**
8. Eckel LM., Krugh TR. 2-Aminofluorene modified DNA duplex exists in two interchangeable conformations. *Nat Struct Biol.* **1994a**.
9. Eckel LM., Krugh TR. Structural characterization of two interchangeable conformations of a 2- aminofluorene-modified DNA oligomer by NMR and energy minimization. *Biochemistry.* **1994b**
10. Edwards RA., Witherspoon M., Wang K., Afrasiabi K., Pham T., Birnbaumer L., Lipkin SM. Epigenetic repression of DNA mismatch repair by inflammation and hypoxia in inflammatory bowel disease-associated colorectal cancer. *Cancer Res.* **2009**
11. Eigen, M.; Gardiner, W. Evolutionary Molecular Engineering Based on RNA Replication. *Pure Appl. Chem.* **1984**
12. Elder, R.H.; Margison, G.P.; Rafferty, J.A. Differential inactivation of mammalian and *Escherichia coli* O<sup>6</sup>-alkylguanine-DNA alkyltransferases by O<sup>6</sup>-benzylguanine. *Biochem. J.* **1994**
13. Emadi A., Jones RJ., Brodsky RA. Cyclophosphamide and cancer: golden anniversary. *Nat Rev Clin Oncol.* **2009**
14. Falnes PO., Klungland A., Alseth I. Repair of methyl lesions in DNA and RNA by oxidative demethylation. *Neuroscience.* **2007**
15. Fang Q. Kanugula, S.; Pegg, A.E. Function of domains of human O<sup>6</sup>-alkyl-guanine-DNA alkyltransferase. *Biochemistry* **2005**
16. Farag N, Mattosovich R, Merlo R, Nierzwicki Ł, Palermo G, Porchetta A, Perugino G, Ricci F. Folding-upon-Repair DNA Nanoswitches for Monitoring the Activity of DNA Repair Enzymes. *Angew. Chem. Int. Ed.* **2021**

17. Faraldo, M. M.; De Pedro, M. A.; Berenguer, J. Sequence of the S-Layer Gene of *Thermus thermophilus* HB8 and Functionality of Its Promoter in *Escherichia coli*. *J. Bacteriol.* **1992**
18. Fink M., Trunk S., Hall M., Schwab H., Steiner K. Engineering of TM1459 from *Thermotoga maritima* for increased oxidative alkene cleavage activity. *Front. Microb.* **2016**
19. Friedberg E.C DNA damage and repair. *Nature.* **2003**
20. Friedberg E.C., GCW S., Wolfram W., Richard D., Schultz R., Ellenberger T. DNA Repair and Mutagenesis. 2. *ASM Press.* **2005**.
21. Gafter-Gvili A., Zingerman B., Rozen-Zvi B., Ori Y., Green H., Lubin I., Malachi T., Gafter U., Herman- Edelstein M. Oxidative stress-induced DNA damage and repair in human peripheral blood mononuclear cells: protective role of hemoglobin. *PLoS One.* **2013**
22. Gerson S.L. MGMT: Its role in cancer aetiology and cancer therapeutics. *Nat. Rev. Cancer.* **2004**
23. Goodtzova K.; Kanugula S.; Edara S.; Pauly G.T.; Moschel R.C.; Pegg A.E. Repair of O<sup>6</sup>-benzylguanine by the *Escherichia coli* Ada and Ogt and the human O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *J. Biol. Chem.* **1997**
24. Gregory CD., Milner AE. Regulation of cell survival in Burkitt lymphoma: implications from studies of apoptosis following cold-shock treatment. *Int J Cancer.* **1994**
25. Gronemeyer, T.; Chidley, C.; Juillerat, A.; Heinis, C.; Johnsson, K. Directed evolution of O<sup>6</sup>-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng. Des. Sel.* **2006**
26. Hammons GJ., Milton D., Stepps K., Guengerich FP., Tukey RH., Kadlubar FF. Metabolism of carcinogenic heterocyclic and aromatic amines by recombinant human cytochrome P450 enzymes. *Carcinogenesis.* **1997**
27. Heflich RH., Neft RE.. Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutat Res.* **1994**

28. Henner WD., Grunberg SM., Haseltine WA. Sites and structure of gamma radiation-induced DNA strand breaks. *J Biol Chem.* **1982**
29. Hinner M.J., Johnsson K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **2010**
30. Huber W.; Perspicace S.; Kohler J.; Müller F.; Schlatter D. SPR-based interaction studies with small molecular weight ligands using hAGT fusion proteins. *Anal. Biochem.* **2004**
31. Huffman JL., Sundheim O., Tainer JA. DNA base damage recognition and removal: new twists and grooves. *Mutat Res.* **2005**
32. Hutchinson F. Chemical changes induced in DNA by ionizing radiation. *Prog Nucleic Acid Res Mol Biol.* **1985**
33. Idili A., Vallée-Bélisle A., Ricci F., Programmable pH-Triggered DNA Nanoswitches *J. Am. Chem. Soc.* **2014**
34. Iliakis G. The role of DNA double strand breaks in ionizing radiation-induced killing of eukaryotic cells. *Bioessays.* **1991**
35. Jannasch H.W., Huber R., Belkin S., Stetter K.O. *Thermotoga neapolitana* sp. nov. of the extremely thermophilic, eubacterial genus *Thermotoga*. *Arch. Microbiol.* **1988**.
36. Janssen D. B. Evolving haloalkane dehalogenase, *Curr. Opin. Chem. Biol.* **2004**
37. Jiricny J. Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol.* **2013**
38. Jiricny J. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol.* **2006**
39. Johnsson N., Johnsson K. A fusion of disciplines: chemical approaches to exploit fusion proteins for functional genomics. *ChemBiochem.* **2003**
40. Juillerat A.; Gronemeyer T.; Keppler A.; Gendreizig S.; Pick H.; Vogel H.; Johnsson K. Directed evolution of O<sup>6</sup>-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem. Biol.* **2003**

41. Kaina B., Christmann M. DNA repair in personalized brain cancer therapy with temozolomide and nitrosoureas. *DNA Repair* .**2019**
42. Kang D, Sun S, Kurnik M, Morales D, Dahlquist FW, Plaxco KW. New Architecture for Reagentless, Protein-Based Electrochemical Biosensors. *J Am Chem Soc.* **2017**
43. Kantidze OL, Velichko AK, Luzhin AV, Razin SV. Heat Stress-Induced DNA Damage. *Acta Naturae* **2016**
44. Kanugula S.; Pegg, A.E. Alkylation damage repair protein O<sup>6</sup>-alkyl-guanine-DNA alkyltransferase from the hyperthermophiles *Aquifex aeolicus* and *Archaeoglobus fulgidus*. *Biochem. J.* **2003**
45. Kanugula S.; Pegg, A.E. Novel DNA Repair Alkyltransferase from *Caenorhabditis elegans*. *Env. Mol. Mut.* **2001**
46. Kasiotis KM., Kyriakopoulou K., Emmanouil C., Tsantila N., Liesivuori J., Souki H., Manakis S., Machera K. Monitoring of systemic exposure to plant protection products and DNA damage in orchard workers. *Toxicol Lett.* **2012**
47. Kelland L. The resurgence of platinum-based cancer chemotherapy. *Nat Rev Cancer.* **2007**
48. Kengen SWM. *Pyrococcus furiosus*, 30 years on. *Microb Bio- technol.* **2017**
49. Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **2003**
50. Khan O.; Middleton M.R. The therapeutic potential of O<sup>6</sup>-alkylguanine DNA alkyltransferase inhibitors. *Expert Opin. Investig. Drugs* **2007**.
51. Kiefer JR. Effects of Ultraviolet Radiation on DNA. Günter Obe, V., editor. Berlin; New York: Springer; **2007**.
52. Kindermann M.; George N.; Johnsson N.; Johnsson K. Covalent and selective immobilization of fusion proteins. *J. Am. Chem. Soc.* **2003**

53. Klein S.; Oesch F. Assay for O<sup>6</sup>-alkylguanine-DNA-alkyltransferase using oligonucleotides containing O<sup>6</sup>-methylguanine in a BamHI recognition site as substrate. *Anal. Biochem.* **1992**
54. Kooistra R., Zonneveld JB., Watson AJ., Margison GP., Lohman PH., Pastink A. Identification and characterisation of the *Drosophila melanogaster* O<sup>6</sup>-alkylguanine-DNA alkyltransferase cDNA. *Nucleic. Acid .Res.***1999**
55. Koscielska-Kasprzak, K.; Cierpicki, T.; Otlewski, J. Importance of alpha-helix N-capping motif in stabilization of betabetaalpha fold. *Protein Sci.* **2003**
56. Kriek E. Fifty years of research on N-acetyl-2-aminofluorene, one of the most versatile compounds in experimental cancer research. *J Cancer Res Clin Oncol.* **1992**.
57. Krishnan Y, Simmel FC. Nucleic acid based molecular devices. *Angew Chem Int Ed Engl.* **2011**
58. Krokan HE., Bjoras M. Base excision repair. *Cold Spring Harb Perspect Biol.* **2013**
59. Kunkel TA. Evolving views of DNA replication (in)fidelity. *Cold Spring Harb Symp Quant Biol.* **2009**
60. Lasa I.; Castón J. R.; Fernández-Herrero L. A.; de Pedro M. A.; Berenguer J. Insertional Mutagenesis in the Extreme Thermophilic Eubacteria *Thermus thermophilus* HB8. *Mol. Microbiol.* **1992**
61. Lawley PD. Effects of some chemical mutagens and carcinogens on nucleic acids. *Prog Nucleic Acid Res Mol Biol.* **1966**
62. Lawley PD. Effects of some chemical mutagens and carcinogens on nucleic acids. *Prog Nucleic Acid Res Mol Biol.* **1966**
63. Leeuwen F.X. Safe drinking water: the toxicologist's. *Food Chem. Toxicol.* **2000**
64. Lindahl T. Instability and decay of the primary structure of DNA. *Nature.* **1993**

65. Liu Z Chengde Mao. Reporting transient molecular events by DNA strand displacement. *Chem Comm.***2014**
66. Loechler EL. A violation of the Swain-Scott principle, and not SN1 versus SN2 reaction mechanisms, explains why carcinogenic alkylating agents can form different proportions of adducts at oxygen versus nitrogen in DNA. *Chem Res Toxicol.* **1994**
67. Loechler EL., Green CL., Essigmann JM. In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. *Proc Natl Acad Sci U S A.* **1984**
68. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Bio.***2008**
69. Luoto KR., Kumareswaran R., Bristow RG. Tumor hypoxia as a driving force in genetic instability. *Genome Integr.* **2013**
70. Luu K.X., Kanugula S., Pegg A.E., Pauly G.T., Moschel R.C. Repair of oligodeoxyribonucleotides by O(6)-alkylguanine-DNA alkyltransferase. *Biochemistry* **2002**
71. Mah MC., Maher VM., Thomas H., Reid TM., King CM., McCormick JJ. Mutations induced by aminofluorene-DNA adducts during replication in human cells. *Carcinogenesis.* **1989**
72. Mamur S., Yuzbasioglu D., Unal F., Yilmaz S. Does potassium sorbate induce genotoxic or mutagenic effects in lymphocytes? *Toxicol In Vitro.* **2010**
73. Mate, D. M.; Gonzalez-Perez, D.; Mateljak, I.; Gomez de Santos, P.; Vicente, A. I.; Alcalde, M. The Pocket Manual of Directed Evolution: Tips and Tricks. In *Biotechnology of Microbial Enzymes: Production, Biocatalysis and Industrial Applications*; **2017**
74. Matsumura, M.; Aiba, S. Screening for Thermostable Mutant of Kanamycin Nucleotidyltransferase by the Use of a Transformation System for a Thermophile, *Bacillus stearothermophilus*. *J. Biol. Chem.* **1985**

75. Mattosovich R., Merlo R., Fontana A., d'Ippolito G., Terns M.P., Watts, E.A., Valenti A., Perugino G. A journey down to hell: New thermostable protein-tags for biotechnology at high temperatures. *Extremophiles* **2020**.
76. Mattosovich R., Merlo R., Miggiano R., Valenti A., Perugino G. O6-alkylguanine-DNA Alkyltransferases in Microbes Living on the Edge: From Stability to Applicability. *Int. J. Mol. Sci.* **2020**
77. McDonald R.; Douglas I.; Revenga C.; Hale R.; Grimm N.; Grönwall J.; Fekete B. Global urban growth and the geography of water availability, quality, and delivery. *Ambio*. **2011**
78. Meeker JD., Calafat AM., Hauser R. Urinary bisphenol A concentrations in relation to serum thyroid and reproductive hormone levels in men from an infertility clinic. *Environ Sci Technol.* **2010a**
79. Meeker JD., Ehrlich S., Toth TL., Wright DL., Calafat AM., Trisini AT., Ye X., Hauser R. Semen quality and sperm DNA damage in relation to urinary bisphenol A among men from an infertility clinic. *Reprod Toxicol.* **2010b**
80. Meeker JD., Yang T., Ye X., Calafat AM., Hauser R. Urinary concentrations of parabens and serum hormone levels, semen quality parameters, and sperm DNA damage. *Environ Health Perspect.* **2011**
81. Merlo R, Caprioglio D, Cillo M, Valenti A, Mattosovich R, Morrone C, Massarotti A, Rossi F, Miggiano R, Leonardi A, Minassi A, Perugino G. The SNAP- tag technology revised: an effective chemo-enzymatic approach by using a universal azide-based substrate. *J Enzyme Inhib Med Chem.* **2021**
82. Merlo, R.; Del Prete, S.; Valenti, A.; Mattosovich, R.; Carginale, V.; Supuran, C.T.; Capasso, C.; Perugino, P. An AGT-based protein-tag system for the labelling and surface immobilization of enzymes on *E. coli* outer membrane. *J. Enz. Inhib. Med. Chem.* **2019**.
83. Middleton M. R., Kelly J., Thatcher N., Donnelly D. J., McElhinney R. S., McMurry T. B. H., McCormick J. E., Margison G. P. O6-(4-bromothienyl)guanine improves the therapeutic index of temozolomide against A375M melanoma xenografts. *Int. J. Cancer* **2000**
84. Miggiano R.; Casazza V.; Garavaglia S.; Ciaramella M.; Perugino G.; Rizzi M.; Rossi F. Biochemical and structural studies of the

*Mycobacterium tuberculosis* O<sup>6</sup>-methylguanine methyltransferase and mutated variants. *J. Bacteriol.* **2013**

85. Miggiano R.; Perugino G.; Ciaramella M.; Serpe M.; Rejman D.; Páv O.; Pohl R.; Garavaglia S.; Lahiri S.; Rizzi M. Crystal structure of *Mycobacterium tuberculosis* O<sup>6</sup>-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA. *Biochem. J.* **2016**
86. Mihaylova VT., Bindra RS., Yuan J., Campisi D., Narayanan L., Jensen R., Giordano F., Johnson RS., Rockwell S., Glazer PM. Decreased expression of the DNA mismatch repair gene Mlh1 under hypoxic stress in mammalian cells. *Mol Cell Biol.* **2003**
87. Miller L.W., Cai Y., Sheetz M.P., Cornish V.W. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag, *Nat. Methods.* **2005**
88. Morrone C.; Miggiano R.; Serpe M.; Massarotl A.; Valenti A.; Del Monaco G.; Rossi M.; Rossi F.; Rizzi M.; Perugino G. Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase. *Biochim. Biophys. Acta* **2017**
89. Naegeli H. DNA structure: Inherent instability and genomic reactions. *Springer* **1997**.
90. Nakamura H., Tanimoto K., Hiyama K., Yunokawa M., Kawamoto T., Kato Y., Yoshiga K., Poellinger L., Hiyama E., Nishiyama M. Human mismatch repair gene, MLH1, is transcriptionally repressed by the hypoxia-inducible transcription factors, DEC1 and DEC2. *Oncogene.* **2008**
91. Neutelings T., Lambert CA., Nusgens BV., Colige AC. Effects of mild cold shock (25 degrees C) followed by warming up at 37 degrees C on the cellular stress response. *PLoS One* **2013**
92. Niesen, J.; Sack, M.; Seidel, M.; Fendel, R.; Barth, S.; Fischer, R.; Stein, C. SNAP-tag technology: A useful tool to determine affinity constants and other functional parameters of novel anti-body fragments. *Bioconjug. Chem.* **2016**
93. Obe G., Johannes C., Schulte-Frohlinde D. DNA double-strand breaks induced by sparsely ionizing radiation and endonucleases as critical

lesions for cell death, chromosomal aberrations, mutations and oncogenic transformation. *Mutagenesis*. **1992**.

94. Oishi S. Effects of butyl paraben on the male reproductive system in mice. *Arch Toxicol*. **2002**
95. Olsson, M.; Lindahl, T. Repair of alkylated DNA in *Escherichia coli*. Methyl group transfer from O<sup>6</sup>-methylguanine to a protein cysteine residue. *J. Biol. Chem*. **1980**
96. Omar Desoukya ND.; Guangming Z. Targeted and non-targeted effects of ionizing radiation. *Journal of Radiation Research and Applied Sciences*. **2015**.
97. Pandir D. DNA damage in human germ cell exposed to the some food additives in vitro. *Cytotechnology*. **2016**
98. Pegg AE. Factors influencing carcinogenesis by alkylating agents. *Results Cancer Res*. **1991**
99. Pegg, A.E. Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. *Chem. Res. Toxicol*. **2011**
100. Perrone S.; Lotti F.; Geronzi U.; Guidoni E.; Longini M.; Buonocore G. Oxidative Stress in Cancer-Prone Genetic Diseases in Pediatric Age: The Role of Mitochondrial Dysfunction. *Oxid Med Cell Longev*. **2016**
101. Perugino G.; Miggiano R.; Serpe M.; Vettone A.; Valenti A.; Lahiri S.; Rossi F.; Rossi M.; Rizzi M.; Ciaramella M. Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic Acids Res*. **2015**
102. Perugino G.; Vettone A.; Illiano G.; Valenti A.; Ferrara M.C.; Rossi M.; Ciaramella M. Activity and regulation of archaeal DNA alkyltransferase: Conserved protein involved in repair of DNA alkylation damage. *J. Biol. Chem*. **2012**.
103. Poltronieri P, Mezzolla V, Primiceri E, Maruccio G. Biosensors for the Detection of Food Pathogens. *Foods*. **2014**

104. Pope C.A.; Thun M.J.; Namboodiri M.M.; Dockery D.W.; Evans J.S.; Speizer F.E.; Heath C.W.; Particulate air pollution as a predictor of mortality in a prospective study of US adults. *Am. J. Resp. Critic. Care Med.* **1995**
105. Porchetta A., Vallée-Bélisle A., Plaxco K. W., Ricci F. Allosterically Tunable, DNA-Based Switches Triggered by Heavy Metals. *J. Am Chem. Soc.* **2013**
106. Ranallo S., Rossetti M., Plaxco K. W., Vallée-Bélisle A., Ricci F, A Modular, DNA-Based Beacon for Single-Step Fluorescence Detection of Antibodies and Other Proteins *Angew. Chem. Int. Ed.* **2015**
107. Reinhard J, Eichhorn U., Wiessler M., Kaina B. Inactivation of O<sup>6</sup>-methylguanine-DNA methyltransferase by glucose-conjugated inhibitors. *Int. J. Cancer.***2001**
108. Reuter S.; Gupta SC.; Chaturvedi MM.; Aggarwal BB. Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic Biol Med.* 49 **2010**.
109. Risom L.; Moller P.; LofT S. Oxidative stress-induced DNA damage by particulate air pollution. *Mutat. Res.* 592 **2005**
110. Rossetti M., Ippodrino R., Marini B., Palleschi G., Porchetta A., Antibody-Mediated Small Molecule Detection Using Programmable DNA-Switches. *Anal. Chem.* **2018**
111. Sabharwal A.; Middleton M.R. Exploiting the role of O<sup>6</sup>-methylguanine-DNA-methyltransferase (MGMT) in cancer therapy. *Curr. Opin. Pharmacol.* **2006**
112. Schwarzenlander C.; Averhoff B. Characterization of DNA Transport in the Thermophilic Bacterium *Thermus thermophilus* HB27. *FEBS J.* **2006**
113. Serpe M., Forenza C., Adamo A., Russo N., Perugino G., Ciaramella M., Valenti A. The DNA Alkylguanine DNA Alkyltransferase-2 (AGT-2) Of *Caenorhabditis Elegans* Is Involved In Meiosis And Early Development Under Physiological Conditions. *Sci.Rep.* **2019**
114. Serpe M.; Forenza C.; Adamo A.; Russo N.; Perugino G.; Ciaramella M.; Valenti A. The DNA alkylguanine DNA alkyltransferase-2

- (AGT-2) of *Caenorhabditis elegans* is involved in meiosis and early development under physiological conditions. *Sci. Rep.* **2019**.
115. Shibutani S., Suzuki N., Tan X., Johnson F., Grollman AP. Influence of flanking sequence context on the mutagenicity of acetylaminofluorene-derived DNA adducts in mammalian cells. *Biochemistry.* **2001**
  116. Singer B. O-alkyl pyrimidines in mutagenesis and carcinogenesis: occurrence and significance. *Cancer Res.* **1986**
  117. Singer B., Kusmierek JT. Chemical mutagenesis. *Annu Rev Biochem.* **1982**
  118. Skipper PL., Kim MY., Sun HL., Wogan GN., Tannenbaum SR. Monocyclic aromatic amines as potential human carcinogens: old is new again. *Carcinogenesis.* **2010**.
  119. Srivenugopal KS., Yuan XH., Friedman HS., Ali-Osman F. Ubiquitination-dependent proteolysis of O<sup>6</sup>-methylguanine-DNA methyltransferase in human and murine tumor cells following inactivation with O<sup>6</sup>-benzylguanine or 1,3-bis(2-chloroethyl)-1-nitrosourea. *Biochemistry.* **1996**
  120. Sugimura T. Past, present, and future of mutagens in cooked foods. *Environ Health Perspect.* **1986**.
  121. Terns R.M., Terns M.P. The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochem. Soc. Trans.* **2013**
  122. Tintoré M.; Aviñó A.; Ruiz F.M.; Eritja R.; Fábrega C. Development of a novel fluorescence assay based on the use of the thrombin binding aptamer for the detection of O<sup>6</sup>-alkyl-guanine-DNA alkyltransferase activity. *J. Nuc. Ac.* **2010**.
  123. Trendowski M, Christen TD, Zoino JN, Fondy TP. Effects of alkylation and immunopotentialization against Ehrlich ascites murine carcinoma in vivo using novel tetra-O-acetate haloacetamido carbohydrate analogs. *Eur. J. Med. Chem.* **2015**.
  124. Tsien R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **1998** .

125. Valenti A.; Napoli A.; Ferrara M.C.; Nadal M.; Rossi M.; Ciaramella M. Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* **2006**
126. Vermeulen W, Fousteri M. Mammalian transcription-coupled excision repair. *Cold Spring Harb Perspect Biol.* **2013**
127. Vettone A.; Serpe M.; Hidalgo A.; Berenguer J.; del Monaco G.; Valenti A.; Ciaramella M.; Perugino G. A novel thermostable protein-tag: Optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering. *Extremophiles* **2016**
128. Vignard J., Mirey G., Salles B. Ionizing-radiation induced DNA double-strand breaks: a direct and indirect lighting up. *Radiother Oncol.* **2013.**
129. Visconti R.; Grieco D. New insights on oxidative stress in cancer. *Curr Opin Drug Discov Devel.* **2009.**
130. Visone V.; Han W.; Perugino G.; Del Monaco G.; She Q.; Rossi M.; Valenti A.; Ciaramella M. In vivo and in vitro protein imaging in thermophilic archaea by exploiting a novel protein tag. *PLoS ONE* **2017**
131. Wardman P. The importance of radiation chemistry to radiation and free radical biology (The 2008 Silvanus Thompson Memorial Lecture). *Br J Radiol.* **2009**
132. Wilson D. L., Beharry A. A, Srivastava A., O'Connor T. R., Kool E.T. Fluorescence Probes for ALKBH2 Allow the Measurement of DNA Alkylation Repair and Drug Resistance Responses *Angew. Chem. Int. Ed.* **2018**
133. Wu, R.S.; Hurst-Calderone, S.; Kohn, K.W. Measurement of O6-alkylguanine-DNA-alkyltransferase activity in human cells and tumor tissues by restriction endonuclease inhibition. *Cancer Res.* **1987**
134. Wyatt MD., Pittman DL. Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem Res Toxicol.* **2006**
135. Xu X, Saw PE, Tao W, Li Y, Ji X, Yu M, Mahmoudi M, Rasmussen J, Ayyash D, Zhou Y, Farokhzad OC, Shi J. Tumor

Microenvironment-Responsive Multistaged Nanoplatform for Systemic RNAi and Cancer Therapy. *Nano Lett.* **2017**

136. Xu-Welliver M., Pegg A.E. Degradation of the alkylated form of the DNA repair protein,  $O^6$ -alkylguanine-DNA alkyltransferase. *Carcinogenesis* **2002**
137. Yilmaz S., Unal F., Yuzbasioglu D., Celik M. DNA damage in human lymphocytes exposed to four food additives in vitro. *Toxicol Ind Health.* **2014**
138. Zengin N., Yuzbasioglu D., Unal F., Yilmaz S., Aksoy H. The evaluation of the genotoxicity of two food preservatives: sodium benzoate and potassium benzoate. *Food Chem Toxicol.* **2011**
139. Zhang J., Shi H., Xu L., Zhu X., Li X. Site-directed mutagenesis of a hyperthermophilic endoglucanase Cel12B from *Thermotoga maritima* based on rational design. *PLoS ONE* **2015**.
140. Zhong Y.; Huang Y.; Zhang T.; Ma C.; Zhang S.; Fan W.; Chen H.; Qian J.; Lu D. Effects of  $O^6$ -methylguanine-DNA methyltransferase (MGMT) polymorphisms on cancer: A meta-analysis. *Mutagenesis* **2010**
141. Zhou T, Lee JW, Tatavarthi H, Lupski JR, Valerie K, Povirk LF. Deficiency in 3'-phosphoglycolate processing in human cells with a hereditary mutation in tyrosyl-DNA phosphodiesterase (TDP1). *Nucleic Acids Res.* **2005**
142. Zillig W., Stetter K.O., Wunderl S., Schulz W., Priess H., Scholz I. The Sulfolobus-“Caldariella” group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.* **1980**

## APPENDIX I

### SCIENTIFIC PUBLICATIONS in PhD.

1. Lo Gullo G, **Mattossovich R**, Perugino G, La Teana A, Londei P, Benelli D. *Optimization of an In Vitro Transcription/Translation System Based on Sulfolobus solfataricus Cell Lysate*. *Archea* (2019) 10.1155/2019/9848253.
2. Del Prete S, Merlo R, Valenti A, **Mattossovich R**, Rossi M, Carginale V, Supuran CT, Perugino G, Capasso C. *Thermostability enhancement of the  $\alpha$ -carbonic anhydrase from Sulfurihydrogenibium yellowstonense by using the anchoring-and-self-labelling-protein-tag system (ASL tag)*. *J Enzyme Inhib Med Chem.*( 2019).10.1080/14756366.2019.1605991
3. Merlo R, Del Prete S, Valenti A, **Mattossovich R**, Carginale V, Supuran CT, Capasso C, Perugino G. *An AGT-based protein-tag system for the labelling and surface immobilization of enzymes on E. coli outer membrane*. *J Enzyme Inhib Med Chem.*( 2019) 10.1080/14756366.2018.1559161
4. **Mattossovich R**<sup>#</sup>, Merlo R<sup>#</sup>, Fontana A, d'Ippolito G, Terns MP, Watts EA, Valenti A, Perugino G. *A journey down to hell: new thermostable protein-tags for biotechnology at high temperatures*. *Extremophiles* (2020). 10.1007/s00792-019-01134-3
5. **Mattossovich R**<sup>#</sup>, Merlo R<sup>#</sup>, Miggiano R, Valenti A, Perugino G. *O<sup>6</sup>-alkylguanine-DNA Alkyltransferases in Microbes Living on the Edge: From Stability to Applicability*. *Review Int J Mol Sci.*2020. 10.3390/ijms21082878.
6. Merlo R, Caprioglio D, Cillo M, Valenti A, **Mattossovich R**, Morrone C, Massarotti A, Rossi F, Miggiano R, Leonardi A, Minassi A, Perugino G. *The SNAP-tag technology revised: an effective chemo-enzymatic approach by using a*

*universal azide-based substrate*. J Enzyme Inhib Med Chem (2021). 10.1080/14756366.2020.1841182.

7. Farag N<sup>#</sup>, **Mattosovich R**,<sup>#</sup> Merlo R, Nierzwicki Ł, Palermo G, Porchetta A, Perugino G, Ricci F. *Folding-upon-Repair DNA Nanoswitches for Monitoring the Activity of DNA Repair Enzymes*. Angew.Chem.Int.Ed.Eng (2021) 10.1002/anie.202016223

. #These authors contribute equally to this work

## **OTHER SCIENTIFIC PUBLICATION**

**Mattosovich R**, Iacono R, Cangiano G, Cobucci-Ponzano B, Istatico R, Moracci M, Ricca E. *Conversion of xylan by recyclable spores of Bacillus subtilis displaying thermophilic enzymes*. Microb Cell Fact. (2017). 10.1186/s12934-017-0833-3

## **APPENDIX II COMMUNICATIONS**

**Mattosovich R**, Merlo R, Del Prete R, Vincenzo Carginale V, Capasso C, Valenti A, Andreas Jaekel, Barbara Saccà D'Ippolito G, Fontana A, Ciaramella M, Perugino G. *AGTs from hot sources: from stability to applicability*. 12th International Congress of Extremophiles .September 2018.Ischia.

**Mattosovich R**, Merlo R, Valenti A, Porchetta A, Ricci F ,Perugino G. *Development of a innovative bio-sensors chimera for the analytical determination of environmental o<sup>6</sup>-alkyl-guanine containing and dna production of a new agt substrate based on a DNA triplex*. BIO-10 . Maggio 2019-Napoli.

**Mattossovich R**, Merlo R, Fontana A, d'Ippolito G, Ricci F, Porchetta A, Forenza C, Valenti A, Perugino G. *DNA repair enzymes: from basic research to bio- and nano-technological applications*. SIB2019 September Lecce.

## **Books**

**Mattossovich R<sup>#</sup>** and Merlo R<sup>#</sup>. *New Biotech tool from Hot Sources: thermostable self-labelling protein-tags near to the boiling water*. Elsevier. Accepted.

Merlo R<sup>#</sup>, **Mattossovich R<sup>#</sup>**, Valenti A, Perugino G. *Recent Advances in the Use of the SNAP-tag® in the Modern Biotechnology*. 2020/BP/8226D

## Other

**Mattossovich R**. *part of the Organization of the II Industrial Biotechnology Congress*. October 2019. University of Naples Federico II. Naples.

**Mattossovich R**, Iacono R, Cobucci-Ponzano B, Isticato R, Cangiano G, Moracci M, Ricca E. *Single-step conversion of xylan into xylose by thermoacidophilic enzymes adsorbed on Bacillus subtilis spores*. XXXII SIMGBM Congress • University of Palermo, September 17-20, 2017.

# APPENDIX III



## Departamento de Biología Molecular

Madrid, July 10th, 2019

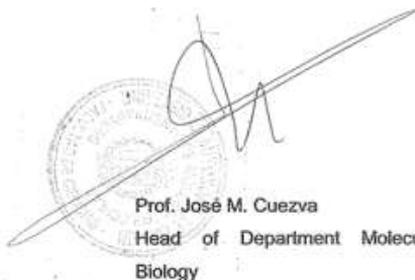
To whom it may concern:

Dr. Aurelio Hidalgo Huertas, Assistant Professor in the Department of Molecular Biology of the Autonomous University of Madrid,

Accepts the incorporation of Ms. Rosanna Mattosovich, of Italian Citizenship and ID number AX5674969 to his research group between October 1<sup>st</sup> 2019 and March 31<sup>st</sup> 2020 to carry out research tasks in the frame of her Ph. D. work.



Dr. Aurelio Hidalgo  
PI



Prof. José M. Cuezva  
Head of Department Molecular  
Biology

# Folding-upon-Repair DNA Nanoswitches for Monitoring the Activity of DNA Repair Enzymes

Nada Farag<sup>+</sup>, Rosanna Mattossovich<sup>+</sup>, Rosa Merlo, Łukasz Nierzwicki, Giulia Palermo, Alessandro Porchetta,\* Giuseppe Perugino,\* and Francesco Ricci\*

**Abstract:** We present a new class of DNA-based nanoswitches that, upon enzymatic repair, could undergo a conformational change mechanism leading to a change in fluorescent signal. Such folding-upon-repair DNA nanoswitches are synthetic DNA sequences containing *O*<sup>6</sup>-methyl-guanine (*O*<sup>6</sup>-MeG) nucleobases and labelled with a fluorophore/quencher optical pair. The nanoswitches are rationally designed so that only upon enzymatic demethylation of the *O*<sup>6</sup>-MeG nucleobases they can form stable intramolecular Hoogsteen interactions and fold into an optically active triplex DNA structure. We have first characterized the folding mechanism induced by the enzymatic repair activity through fluorescent experiments and Molecular Dynamics simulations. We then demonstrated that the folding-upon-repair DNA nanoswitches are suitable and specific substrates for different methyltransferase enzymes including the human homologue (hMGMT) and they allow the screening of novel potential methyltransferase inhibitors.

## Introduction

The genetic information inside cells is protected against DNA damage by multi-enzymatic DNA repair mechanisms, such as base excision, nucleotide excision repair,<sup>[1]</sup> or a direct damage reversal by *O*<sup>6</sup>-methylguanine-DNA-methyltransferases (abbreviated here as AGTs).<sup>[2–4]</sup> The latter is a class of evolutionarily conserved biocatalysts, able to directly and irreversibly remove alkyl groups at the *O*<sup>6</sup>-position of guanines on DNA in a single S<sub>N</sub>2-like reaction mechanism.

These enzymes, indeed, represent the major factor in contrasting the effects of alkylating agents that form such adducts.<sup>[2]</sup> On the other hand, human methyltransferase (hMGMT) activity impacts the alkylating agent-based chemotherapy in cancer cells.<sup>[2,5]</sup> For this reason, different hMGMT inactivators/inhibitors are usually employed in combination with this kind of chemotherapy.<sup>[2,6]</sup>

Because of the clinical relevance of AGTs, assays able to measure their activity in a reliable and rapid way are needed.<sup>[7]</sup> Most of the assays developed so far are based on the use of oligonucleotides carrying radioactive *O*<sup>6</sup>-methyl-guanine (*O*<sup>6</sup>-MeG) groups and chromatographic separations.<sup>[8–11]</sup> These are, however, time-consuming, tedious, and unsafe. Alternative fluorescence-based assays that allow simple and fast detection of hMGMT activity have been recently developed. For example, Kool and co-workers designed a chemosensor that couples fluorescence change to the bond-breaking step occurring during repair activity.<sup>[12,13]</sup> Other approaches employing optically-labelled DNA aptamers,<sup>[11]</sup> DNA-based electrochemical sensors<sup>[14]</sup> or *O*<sup>6</sup>-MeG-containing double strands DNA (dsDNA) oligonucleotides in competition with fluorescent substrates<sup>[4,15–18]</sup> were also proposed. While these latter systems provide several advantages including ease of use and high sensitivity, there is still an urgent need for finding new strategies to achieve efficient activity-based monitoring of DNA repair enzymes that can be versatile enough to be suitable for a wide range of repair activities.

Recently, DNA nanoswitches have emerged as a new class of programmable probes that allow the sensitive and rapid detection of a wide range of molecular targets.<sup>[19]</sup> DNA nanoswitches usually undergo a binding-induced conformational change that can provide a measurable output in the presence of a specific input. A number of strategies employing DNA nanoswitches have been developed for the detection of different targets including pH,<sup>[20,21]</sup> metal ions,<sup>[22]</sup> small molecules,<sup>[23,24]</sup> proteins,<sup>[25,26]</sup> or specific antibodies.<sup>[27,28]</sup> Critically, because their signaling is linked to a change in the physics of the DNA probe induced by the recognition with a target, such conformational switching sensors are generally highly specific and selective.<sup>[29]</sup>

Motivated by the above considerations, we demonstrate here the rational design of a new class of DNA nanoswitches, here named folding-upon-repair DNA nanoswitches, that can be conveniently applied for monitoring DNA repair activity. As an applicative example, we initially focused on methyltransferase enzymes and designed a fluorescent-labelled DNA nanoswitch containing *O*<sup>6</sup>-MeG that, upon enzymatic

[\*] N. Farag,<sup>[1]</sup> Dr. A. Porchetta, Prof. F. Ricci  
Department of Chemistry, University of Rome Tor Vergata  
Via della Ricerca Scientifica, 00133 Rome (Italy)  
E-mail: alessandro.porchetta@uniroma2.it  
francesco.ricci@uniroma2.it

R. Mattossovich,<sup>[1]</sup> R. Merlo, Dr. G. Perugino  
Institute of Biosciences and BioResources  
National Research Council of Italy  
Via Pietro Castellino 111, 80131 Naples (Italy)  
E-mail: giuseppe.perugino@ibbr.cnr.it  
Dr. Ł. Nierzwicki, Prof. G. Palermo  
Department of Bioengineering, University of California Riverside  
900 University Avenue, Riverside, CA 52512 (USA)

Prof. G. Palermo  
Department of Chemistry, University of California Riverside  
900 University Avenue, Riverside, CA 52512 (USA)

[†] These authors contributed equally to this work.

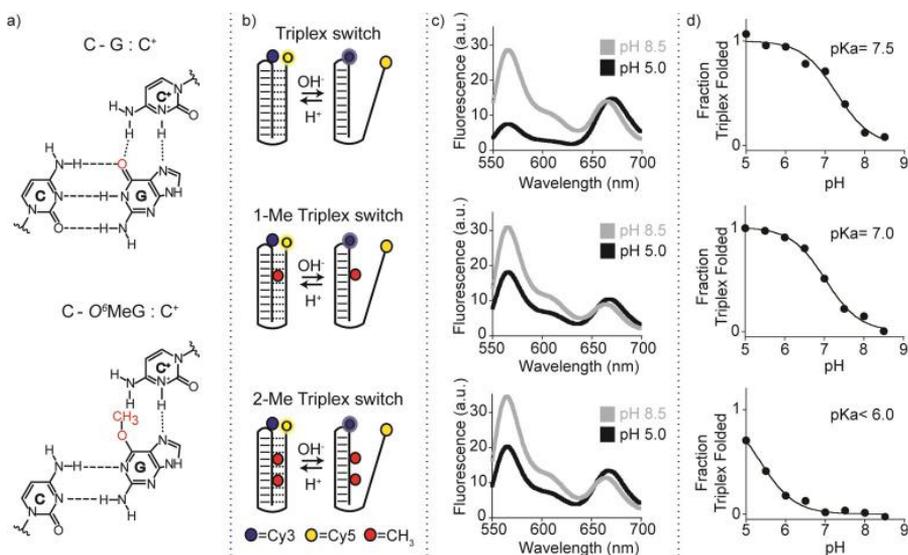
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:  
<https://doi.org/10.1002/anie.202016223>.

repair, could undergo a conformational change associated with a change in the output signal. This approach allows the direct measurement of the activity of, in principle, any methyltransferase enzyme thus opening the possibility to develop an easy methodology for the high-throughput determination of methyltransferase enzyme inhibitors.

## Results and Discussion

Our strategy to design programmable nucleic acid nanoswitches for the detection of methyltransferase activity is based on the use of a single-stranded DNA capable of forming an intramolecular triplex structure through hydrogen bonds (Hoogsteen interactions) between a hairpin duplex domain and a single-strand triplex-forming portion.<sup>[30]</sup> More specifically, we have designed three nanoswitches displaying the same triplex-forming domains (i.e., 4 cytosines + 6 thymines) but differing in the content of *O*<sup>6</sup>-MeGs in the hairpin duplex domain (0, 1, and 2). The idea underlying the molecular design of the triplex nanoswitches is that the methyl group at the guanine *O*<sup>6</sup> position should affect the formation of the Hoogsteen hydrogen bond with the cytosine on the single-strand DNA (Figure 1 a), thus influencing the pH-dependent triplex folding/unfolding behaviour of the nanoswitch. The nanoswitches are also labelled with a pH-insensitive FRET pair

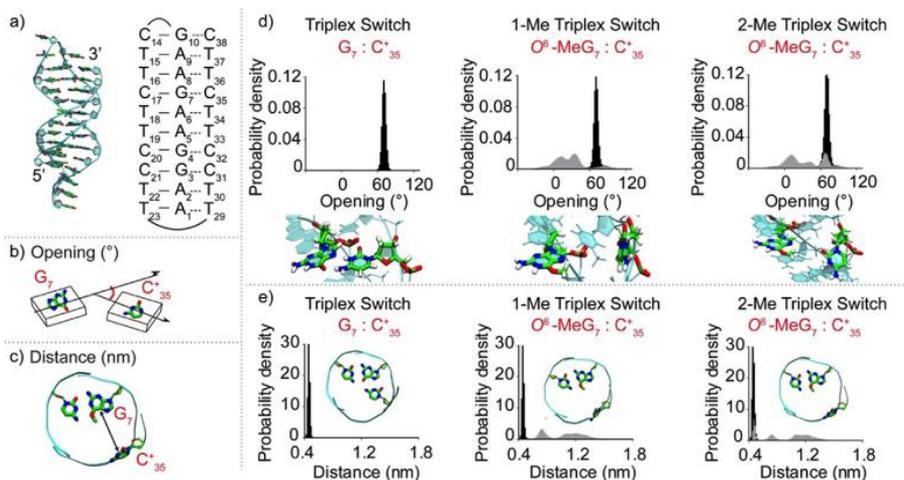
pair to follow pH-dependent folding/unfolding (Figure 1 b). To demonstrate the effect of *O*<sup>6</sup>-MeG on the triplex formation, we have first tested the pH-dependent folding/unfolding behaviour of our three nanoswitches by measuring the fluorescence signal of our FRET pair at a fixed concentration of the nanoswitch and different pH values. As expected, our control triplex nanoswitch (Triplex switch, Figure 1 b, top) that lacks any *O*<sup>6</sup>-MeG in its sequence, shows signals that are consistent with the formation of a folded triplex at acidic pH (pH 5.0) and suggest the unfolding of the triplex structure at basic pH values (pH 8.5). The pH of semi-protonation (here defined as *pK*<sub>a</sub>, the average *pK*<sub>a</sub> due to several interacting protonation sites) for this triplex nanoswitch is 7.5 (Figure 1 c,d top, and S1). The presence of one *O*<sup>6</sup>-MeG (1-Me Triplex switch, Figure 1 b, center) in the hairpin duplex (at position 7) strongly destabilizes triplex formation; as a result, we observe at pH 5.0 a FRET signal that is consistent with a partially unfolded triplex structure and we obtain a *pK*<sub>a</sub> of 7.0 (Figures 1 c,d center, and S1). Finally, the nanoswitch containing two *O*<sup>6</sup>-MeG (2-Me Triplex switch, at position 4 and 7, Figure 1 b, bottom) is even more destabilized and the pH titration curve with this switch does not reach a plateau at lower pH values suggesting a *pK*<sub>a</sub> lower than 6.0 (Figures 1 c,d bottom, and S1). The formation of a folded triplex structure of the Triplex switch was confirmed by melting and urea denaturation experiments performed at different pHs (Figures S2, S3). Also, CD spectra of the three nanoswitches further support the formation of pH-dependent Hoogsteen interactions (Figure S4).



**Figure 1.** Folding-upon-repair DNA nanoswitches. a) DNA parallel triplets formed between two cytosines and one guanine and involving Watson-Crick (three hydrogen bonds) and Hoogsteen interactions (two hydrogen bonds) require the protonation of the N<sup>3</sup> of cytosine in the third strand (top, left) and thus are only stable at acidic pH values (average *pK*<sub>a</sub> of protonated cytosines in C-G:C triplet is ≈ 6.5). Because it affects Hoogsteen base pairing efficiency, methylation of the guanine in the *O*<sup>6</sup>-position destabilizes the triplex conformation. b) Programmable DNA-based triplex nanoswitches designed to form an intramolecular triplex structure with 0 (Triplex), 1 (1-Me Triplex) or 2 (2-Me Triplex) *O*<sup>6</sup>-MeG in the sequence. c) Fluorescence spectra were obtained for each nanoswitch at pH 5.0 and pH 8.5. d) pH-titration curves of the triplex nanoswitches. Triplex-to-duplex transition is monitored through a pH-insensitive FRET pair at the 3'-end (Cy5) and internally located (Cy3). The pH titration experiments were performed at 25 °C, [nanoswitch] = 50 nM by measuring the fluorescence signal at different pH values in 50 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaCl buffer. Spectra were obtained by excitation at 530 (± 5) nm and acquisition between 545 and 700 nm (± 10) nm.

at different pHs (Figures S2, S3). Also, CD spectra of the three nanoswitches further support the formation of pH-dependent Hoogsteen interactions (Figure S4).

To better understand the effect that guanine methylation has on the stability of our triplex nanoswitches we performed overall molecular dynamics (MD) simulations. We first considered the non-methylated nanoswitch (Triplex switch) where all cytosines involved in Hoogsteen base pairing are protonated (Figure 2a). Then, we introduced the *O*<sup>6</sup>-MeG at position 7 (creating the 1-Me Triplex switch), and at both positions 4 and 7 (2-Me Triplex switch). The obtained trajectories encompassed > 4.5 μs of sampling, including multiple replicas and providing solid statistics for the analysis of the DNA triplex structural changes. We focused on the



**Figure 2.** Molecular dynamics simulations of triplex nanoswitches. a) Structure (left) and schematic representation (right) of the Triplex switch with numeration of DNA nucleotides. Graphical representation of b) opening angle and c) distance parameters for the  $G_7:C_{35}^+$  pair. d) Opening angle distribution for the Triplex switch (left), 1-Me Triplex switch (center), and 2-Me Triplex switch (right). Below each graph, side-view snapshots of the analysed base pairs are reported. The opening distribution for 1Me- and 2Me- Triplex switches are shown in grey. The distribution of the non-methylated Triplex switch (black) is also reported in each graph as a comparison. e) Distance parameter for the  $G_7:C_{35}^+$  Hoogsteen interaction for the Triplex switch, 1-Me Triplex switch, and 2-Me Triplex switch. In each graph, a top-view snapshot of the triplet most probable configuration is shown. The distance distribution for 1Me- and 2Me- triplex switches is shown in grey. The distribution of the non-methylated Triplex switch (black) is also shown in each graph as a comparison.

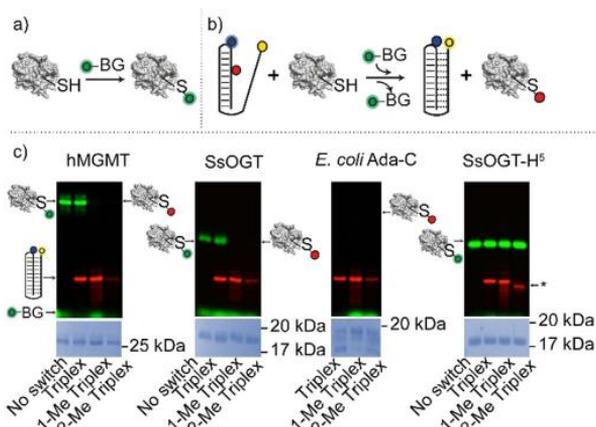
$C_{17}-G_7:C_{35}^+$  triplet, as this triplet share the methylated guanine in both the 1-Me and 2-Me Triplex switches. Our analysis took into consideration two parameters: 1) the base-pair opening angle, which describes the in-plane opening between  $G_7:C_{35}^+$  paired Hoogsteen-forming bases, as defined by Lu and Olson<sup>[31]</sup> (Figure 2b; see Tables S1–S4 for all DNA base-pair parameters); 2) the distance between the centers of mass of the atoms of the  $G_7:C_{35}^+$  Hoogsteen-forming bases at the interface with each other (Figure 2c).

The narrow distributions of both the opening and distance parameters for the  $C_{17}-G_7:C_{35}^+$  triplet in the Triplex switch (with an average value of  $73^\circ$  and 0.45 nm, respectively) correspond to a well-defined triplex structure (Figure 2d,e, left and Table S1). Conversely, in the case of the 1-Me and 2-Me Triplex switches, distributions of both parameters are broader and have multiple maxima, showing that the same triplet adopts multiple conformational states (Figure 2d,e, center, right). Remarkably, in case of 1-Me and 2-Me Triplex switches the distance distribution shifts towards higher values. This clearly shows that the  $O^6$ -methylation of guanine hampers the efficient interaction with the Hoogsteen pairing cytosine, leading to a local unwinding of the triplex (see representative structures, Figure 2d, and highlights on Figure 2e, center, right and Figures S5–8). Similar shifts and broadening of opening and distance distributions were also observed in the case of  $C_{20}-G_4:C_{32}^+$  triplet in 2Me-Triplex switch, whereas in the case of 1Me-Triplex switch, these distributions were virtually the same as in case of non-methylated Triplex switch (Figures S9–11). Although at a lower extent, we also observed the broadening of the opening and distance distributions for the  $C_{17}-G_7:C_{35}^+$  triplet during addi-

tional  $\approx 1.5 \mu\text{s}$  long simulations of the Triplex switch in the deprotonated form (Figure S12), further supporting the notion that methylation strongly destabilizes the formation of the triplet. Finally, to achieve more meaningful structural information on the triplex destabilization induced by the methylation, we obtained free energy profiles that can be associated with the increase of the distance between the two fluorophores on each switch. As expected, under protonated conditions the control Triplex is the most stable, followed by the 1-Me Triplex and then by the 2-Me Triplex (Figure S13). These results provide an atomic-level understanding of the triplex stability, indicating that the efficient Hoogsteen base pairing is critical

for the triplex formation and both cytosine deprotonation at high pH and guanine methylation should reduce the propensity of DNA hairpin to adopt triplex conformation.

Because the methyl group at the guanine  $O^6$  position prevents the efficient formation of Hoogsteen interactions with the respective cytosine in the triplex-forming domain, the  $O^6$ -MeG-containing triplex nanoswitches characterized above could be used as suitable probes for the monitoring of methyltransferase activity. The enzymatic removal of the methyl group of  $O^6$ -MeG in the nanoswitches would restore the optimal conditions for Hoogsteen interactions and triplex formation by the nanoswitch. To achieve this goal, we initially performed an indirect assay using a fluorescein derivative of  $O^6$ -benzyl-guanine ( $O^6$ -BG)<sup>[4,15,32,33]</sup> (Figure 3a) to demonstrate that  $O^6$ -MeG-containing triplex nanoswitches are suitable substrates for methyltransferase enzymes (Figure 3b). We first demonstrated that no interaction occurs between the nanoswitch and the fluorescent-labelled substrate in the absence of the protein. As expected, visible separated bands for both the  $O^6$ -BG fluorescent derivative and the three DNA nanoswitches tested can be observed (Figure S14). It is worth noting that the nanoswitch containing two  $O^6$ -MeGs (2-Me Triplex) shows a slightly, but significant, higher mobility (marked by an asterisk in Figure 3c) compared to the single-methylated triplex (1-Me Triplex) and the non-methylated triplex (Triplex) (Figure S14). Although speculative, a possible explanation of the different migration of the 2-Me Triplex could be that this oligonucleotide adopts a partial folding in denaturing conditions. We have pre-incubated the methyltransferase enzymes with our nanoswitches and then added an equimolar amount of the  $O^6$ -BG



**Figure 3.** Characterization of triplex nanoswitches as suitable substrates of methyltransferase enzymes. a) Suicide reaction of an AGT enzyme with a fluorescein derivative of  $O^6$ -BG. b) The alkyl transfer suicide reaction between the enzyme and a DNA methylated triplex hampers the covalent bond between the enzyme and the fluorescein group. c) SDS-PAGE images of reactions in which different enzymes (i.e. hMGMT, SsOGT, *E. coli* Ada-C, SsOGT-H<sup>5</sup>) after pre-incubation with the DNA nanoswitches and then reacted with the  $O^6$ -BG fluorescent derivative. The enzymes that have reacted with the fluorescein derivative appear as green bands on the gel. The absence of any fluorescent bands indicates that the enzyme has not reacted with the fluorescent substrate. Fluorescence-labelled triplex nanoswitches appear as red bands. The asterisk shows the higher mobility of the 2-Me Triplex nanoswitch compared to the other Triplex switches (see also Figure S14). The experiments were performed by pre-incubating for 60 min the different AGT enzymes with the relevant triplex nanoswitches and then adding the  $O^6$ -BG fluorescent derivative. Samples were then loaded on 15% acrylamide SDS-PAGE. Gels underwent Coomassie staining for the determination and correction of the protein amount loaded.

fluorescent derivative that interacts with the free enzymes in the solution (Figure 3 b). In the case of hMGMT, a fluorescent protein band was observed in the presence of the non-methylated Triplex, suggesting that no repair reaction on this DNA occurred (Figure 3c). On the contrary, 1-Me Triplex nanoswitch and 2-Me Triplex nanoswitch, showed a complete absence of the fluorescent band (Figure 3c). This is likely due to the suicidal nature of methyltransferase enzymes,<sup>[3]</sup> as the methyl group is irreversibly transferred to the enzyme, which is no longer available for the  $O^6$ -BG fluorescent derivative, thus demonstrating that methylated nanoswitches are effective substrates for hMGMT. The same results were further confirmed using the thermostable *Saccharolobus solfataricus* homologue (SsOGT), an enzyme responsible for the direct repair of  $O^6$ -alkylguanine in double-stranded DNA at high temperatures<sup>[4,16,32,33]</sup> and the *Escherichia coli* homologue (Ada-C), which is reported to be insensitive to  $O^6$ -BG derivatives.<sup>[17,18]</sup> In the latter case, the absence of the enzyme fluorescent bands despite being correctly loaded as confirmed by the Coomassie staining analysis (Figure 3c, *E. coli* Ada-c), and the similar migration of all the triplexes, confirm that the DNA-repair occurred. The methylated DNA triplex nanoswitches are also highly specific as they showed no enzymatic activity when incubated with a mutant of SsOGT (SsOGT-H<sup>5</sup>)

that was previously reported to be catalytically active on  $O^6$ -BG derivatives, but unable to bind and react with ds-DNA.<sup>[4,32–35]</sup>

Prompted by the results described above, we tested whether the methyltransferase repair activity on our nanoswitches could result in a conformational switch, and a consequent measurable FRET signal change, providing a means of direct detection of enzymatic activity (Figure 4a,b). Initially, we focused on the detection of the activity of hMGMT; as expected, no significant difference in the FRET signal of the control nanoswitch (non-methylated Triplex) before ( $1.9 \pm 0.1$ ) and after ( $2.1 \pm 0.1$ ) hMGMT incubation can be observed (Figure 4c, left). Under the experimental conditions used, both the fluorescence spectra suggest that the non-methylated triplex switch is completely folded (at pH 5.0). The same experiment carried out using the methylated triplex nanoswitches shows, instead, a strong difference in FRET signals before and after hMGMT incubation (Figure 4c, center, right). For both nanoswitches, the FRET signals observed before hMGMT incubation confirm a partially unfolded configuration (1-Me Triplex,  $0.6 \pm 0.1$ , 2-Me



**Figure 4.** Detection of methyltransferase activity with triplex nanoswitches. a) Methyltransferase enzymatic activity on DNA methylated nanoswitches leads to folding-upon-repair of the triplex DNA structure. b) Enzymatic activity detection can be achieved by monitoring folding/unfolding of the Triplex nanoswitch by fluorescence FRET signaling. c) Spectra and d) relative FRET signals obtained with the Triplex nanoswitches before and after incubation with 5  $\mu$ M ( $0.1 \mu\text{g} \mu\text{L}^{-1}$ ) hMGMT. e) Relative FRET signals observed at different hMGMT concentrations for 1-Me and 2-Me Triplex nanoswitches. f) Relative FRET signals with different methyltransferase enzymes. Spectra were obtained by excitation at  $530 (\pm 5)$  nm and acquisition between 545 and 700 nm ( $\pm 10$ ) nm. Relative FRET signals were obtained by first incubating the DNA nanoswitches ( $0.5 \mu\text{M}$ ) in the absence or presence of AGTs ( $0.1 \mu\text{g} \mu\text{L}^{-1}$ ) in 10  $\mu\text{L}$  solution of 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 150 mM NaCl, pH 7.5 at 30°C for 60 minutes. The reaction mixtures were then diluted to 100  $\mu\text{L}$  using 50 mM  $\text{Na}_2\text{HPO}_4$  buffer, 250 mM NaCl at pH 5.0, and heat-inactivated for 2 minutes at 70°C before performing the fluorescence spectra at 25°C.

Triplex,  $0.6 \pm 0.1$ , Figure 4d). Upon hMGMT incubation, the enzymatic demethylation of  $O^6$ -MeG in the triplex nanoswitches restores their ability to form a triplex structure. This is particularly clear for 2-Me triplex nanoswitch that shows a FRET signal ( $1.9 \pm 0.1$ ) after hMGMT incubation that is within the error of the control non-methylated nanoswitch ( $2.1 \pm 0.1$ ) (Figure 4d). Of note, the folding dynamics of these switches are extremely rapid ( $K_{\text{folding}}$  and  $K_{\text{unfolding}} = 10 \text{ s}^{-1}$  and  $2 \text{ s}^{-1}$ , respectively)<sup>[36]</sup> and thus the rate-determining step in these measurements is given by the enzyme-catalyzed reaction. We found out that with hMGMT a saturation of signal is observed after 15 minutes of enzymatic reaction (Figure S15). Both methylated triplex nanoswitches show a change in the relative FRET values that are linearly dependent on the concentration of hMGMT in the range between 0.5 and  $5 \mu\text{M}$  ( $R^2_{2\text{-Me Triplex}} = 0.99$ ,  $R^2_{1\text{-Me Triplex}} = 0.95$ ) (Figure 4e). Similar experiments were also performed with other methyltransferase enzymes, both methylated triplex nanoswitches upon incubation with SsOGT and *E. coli* Ada-C enzymes gave FRET signal changes consistent with triplex folding suggesting efficient enzymatic activity (Figure 4f). The same experiment performed with SsOGT-H<sup>5</sup> (a DNA repair defective mutant, but catalytically active) produced no effect on the FRET signal of both nanoswitches (Figure 4f) once again confirming the specificity of the folding-upon-repair mechanism. To demonstrate the possibility of using this platform in

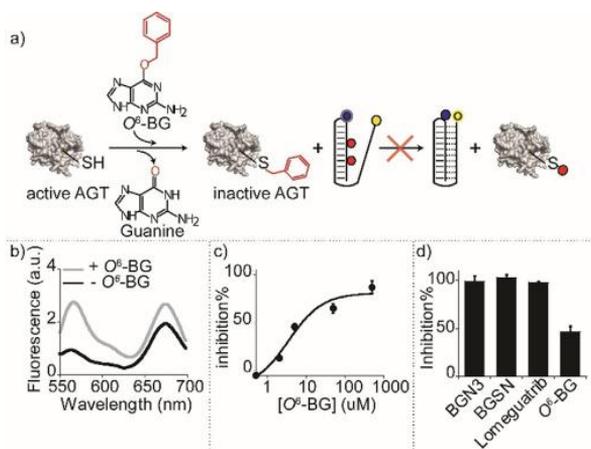
more complex media, we have also performed the measurement of the enzymatic activity of hMGMT in 10% serum and observed FRET signal changes well distinguishable from the control experiment in absence of enzyme (Figure S16).

To demonstrate the utility of our platform for the study and characterization of new hMGMT inhibitors as possible drug candidates,<sup>[37]</sup> we used our nanoswitches to measure the activity of the human enzyme in the presence of inhibitors, inactivators, or pseudo-substrates. The addition of an hMGMT inhibitor in the reaction mixture should prevent the enzyme from repairing the methylated DNA, and no significant change of the nanoswitch FRET signal should be observed (Figure 5a). We initially tested  $O^6$ -BG, a widely characterized methyltransferase inactivator.<sup>[38,39]</sup> The relative FRET signal of the nanoswitch incubated with hMGMT and  $O^6$ -BG (Figure 5b, grey line) is, at  $0.8 \pm 0.1$ , similar to the values observed in the absence of enzyme (Figure 5b, black line) an effect consistent with the inactivation of hMGMT activity by  $O^6$ -BG. The measured inactivation, as expected, follows a concentration-dependent behaviour with a measured  $\text{IC}_{50} = 3.5 \pm 2 \mu\text{M}$  (Figure 5c). Moreover, our platform is able to measure the inactivation efficiency of different molecules including 4-azido-*N*-(4-(hydroxymethyl) benzyl) butanamide (BGN3), *N*-(4-(((2-amino-9H-purin-6-yl)oxy)methyl)benzyl)-4-azidobutanamide (BGSN), Lomeguatrib,<sup>[40,41]</sup>  $O^6$ -Benzylguanine ( $O^6$ -BG),<sup>[42]</sup> demonstrating once again its versatility (Figure 5d).

## Conclusion

Here, we have designed folding-upon-repair DNA nanoswitches containing  $O^6$ -MeG nucleobases that undergo a conformational switch from a duplex to a triplex conformation upon enzymatic demethylation. We have first demonstrated that the presence of a methyl group at the guanine  $O^6$  position in the duplex portion of the nanoswitch strongly affects the triplex formation. Molecular simulations were used to determine how the  $O^6$ -methylation of guanine bases prevents the efficient formation of Hoogsteen interactions with the cytosine in the triplex-forming domain, resulting in a local unwinding of the triplex structure. Such methylated triplex-based nanoswitches are versatile tools for the direct measurement of methyltransferase activity as they form the natural enzymatic substrate duplex conformation when methylated and, upon enzymatic demethylation, they can fold into a measurable triplex structure. We showed that these methylated nanoswitches were efficiently recognized by these enzymes leading to measurable FRET signal changes upon repair activity. Finally, we have also measured hMGMT activity in the presence of several enzyme inactivators, demonstrating the possibility of using in a high-throughput system our nanoswitches for the screening of novel potential inactivators/pseudo-substrates/inhibitors of hMGMT.<sup>[37]</sup>

Our folding-upon-repair DNA nanoswitches are reagentless, highly specific, and versatile. These features make our proposed approach convenient and easy to perform. Although the current experimental protocol requires a change in the pH of the solution after the enzymatic reaction making



**Figure 5.** Detection of methyltransferase inhibitors. a) Pre-incubation of the enzyme with an inhibitor (here  $O^6$ -BG) leads to irreversible inhibition of the enzyme that is thus not able to demethylate the DNA nanoswitch. b) Spectra obtained with the 2-Me Triplex nanoswitch in the presence of hMGMT ( $5 \mu\text{M}$ ) before and after incubation with the inhibitor  $O^6$ -BG ( $5 \mu\text{M}$ ). c) % Inhibition plot obtained at different concentrations of  $O^6$ -BG inhibitor. d) % Inhibition obtained with the 2-Me Triplex nanoswitch with an equimolar concentration of different enzymatic inhibitors and hMGMT ( $5 \mu\text{M}$ ). Spectra were obtained by excitation at  $530 (\pm 5) \text{ nm}$  and acquisition between 545 and  $700 \text{ nm} (\pm 10) \text{ nm}$ . Relative FRET signals were obtained by enzymatic incubation of  $5 \mu\text{M}$  hMGMT and inhibitor at  $30^\circ\text{C}$  for 60 minutes in  $10 \mu\text{L}$  solution of  $50 \text{ mM Na}_2\text{HPO}_4$  buffer,  $150 \text{ mM NaCl}$  at pH 7.5. 2-Me Triplex nanoswitch ( $0.5 \mu\text{M}$ ) was then added to the reaction mixtures and incubated for another 60 min at  $30^\circ\text{C}$ . The reaction mixtures were diluted to  $100 \mu\text{L}$  using  $50 \text{ mM Na}_2\text{HPO}_4$  buffer,  $250 \text{ mM NaCl}$  at pH 5.0, and heat-inactivated for 2 minutes at  $70^\circ\text{C}$  before performing the fluorescence spectra at  $25^\circ\text{C}$ .

it less amenable to automatization, we believe that the principle we have described together with the ease with which these probes are designed make the approach suitable to develop a suite of activity-based DNA nanoswitches for other DNA repair enzymes.

## Acknowledgements

This work was supported by the European Research Council, ERC (project n.336493) (FR), by Associazione Italiana per la Ricerca sul Cancro, AIRC (project n. 14420) (FR), by the Italian Ministry of Health (project n. GR-2010-2317212) and by the Italian Ministry of University and Research (Project of National Interest, PRIN, 2017YER72K). Computational work has been supported by the National Science Foundation under Grant No. CHE-1905374 and by the National Institute of Health under Grant No. R01-EY027440 (GP). Computer time for MD simulations has been awarded by XSEDE via the Grant No. TG-MCB160059 (GP). We would like to thank Mateusz Kogut for valuable suggestions regarding DNA triplex folding simulations. F.R., G.P., and A.P. would like to thank all the authors for their efforts in writing this work, technical assistance, but mainly for human support during the difficult and delicate period of stay-at-home following the COVID-19 outbreak.

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** conformational change mechanism · DNA nanoswitches · DNA nanotechnology · DNA repair enzymes · triplex DNA

- [1] E. C. E. T. Madders, J. L. Parsons in *Advances in Experimental Medicine and Biology*, Vol. 1241 (Eds.: W. E. Crusio, J. D. Lambri, H. H. Radeke, N. Rezaei), Springer, Cham, **2020**, pp. 59–75.
- [2] S. L. Gerson, *Nat. Rev. Cancer* **2004**, *4*, 296–307.
- [3] A. E. Pegg, *Chem. Res. Toxicol.* **2011**, *24*, 618–639.
- [4] R. Mattosovich, R. Merlo, R. Miggiano, A. Valenti, G. Perugino, *Int. J. Mol. Sci.* **2020**, *21*, 2878.
- [5] D. Fu, J. A. Calvo, L. D. Samson, *Nat. Rev. Cancer* **2012**, *12*, 104–120.
- [6] S. L. Gerson, *J. Clin. Oncol.* **2002**, *20*, 2388–2399.
- [7] Z. D. Nagel, A. A. Beharry, P. Mazzucato, G. J. Kitange, J. N. Sarkaria, E. T. Kool, L. D. Samson, *PLoS One* **2019**, *14*, e0208341.
- [8] S. Hurst-Calderone, K. W. Kohn, *Cancer Res.* **1987**, *47*, 6229–6235.
- [9] S. Klein, F. Oesch, *Anal. Biochem.* **1992**, *205*, 294–299.
- [10] K. X. Luu, S. Kanugula, A. E. Pegg, G. T. Pauly, R. C. Moschel, *Biochemistry* **2002**, *41*, 8689–8697.
- [11] M. Tintoré, A. Aviñó, F. M. Ruiz, R. Eritja, C. Fàbrega, *J. Nucleic Acids* **2010**, *2010*, 632041.
- [12] A. A. Beharry, Z. D. Nagel, L. D. Samson, E. T. Kool, *PLoS One* **2016**, *11*, e0152684.
- [13] D. L. Wilson, A. A. Beharry, A. Srivastava, T. R. O'Connor, E. T. Kool, *Angew. Chem. Int. Ed.* **2018**, *57*, 12896–12900; *Angew. Chem.* **2018**, *130*, 13078–13082.
- [14] S. Chen, J. Su, Z. Zhao, Y. Shao, Y. Dou, F. Li, W. Deng, J. Shi, Q. Li, X. Zuo, S. Song, C. Fan, *Nano Lett.* **2020**, *20*, 7028–7035.
- [15] A. Vettone, M. Serpe, A. Hidalgo, J. Berenguer, G. del Monaco, A. Valenti, M. Rossi, M. Ciaramella, G. Perugino, *Extremophiles* **2016**, *20*, 1–13.
- [16] C. Morrone, R. Miggiano, M. Serpe, A. Massarotti, A. Valenti, G. del Monaco, M. Rossi, F. Rossi, M. Rizzi, G. Perugino, M. Ciaramella, *Biochim. Biophys. Acta* **2017**, *1861*, 86–96.
- [17] R. H. Elder, G. P. Margison, J. A. Rafferty, *Biochem. J.* **1994**, *298*, 231–235.
- [18] K. Goodtzova, S. Kanugula, S. Edara, G. T. Pauly, R. C. Moschel, A. E. Pegg, *J. Biol. Chem.* **1997**, *272*, 8332–8339.
- [19] S. G. Harroun, C. Prévost-Tremblay, D. Lauzon, A. Desrosiers, X. Wang, L. Pedro, A. Vallée-Bélisle, *Nanoscale* **2018**, *10*, 4607–4641.
- [20] S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor, Y. Krishnan, *Nat. Nanotechnol.* **2009**, *4*, 325–330.
- [21] T. Patino, A. Porchetta, A. Jannasch, A. Lladó, T. Stumpp, E. Schäffer, F. Ricci, S. Sánchez, *Nano Lett.* **2019**, *19*, 3440–3447.
- [22] A. Porchetta, A. Vallée-Bélisle, K. W. Plaxco, F. Ricci, *J. Am. Chem. Soc.* **2013**, *135*, 13238–13241.
- [23] M. Hansen-Bruhn, L. D. F. Nielsen, K. V. Gothelf, *ACS Sens.* **2018**, *3*, 1706–1711.
- [24] M. Rossetti, R. Ippodromo, B. Marini, G. Palleschi, A. Porchetta, *Anal. Chem.* **2018**, *90*, 8196–8201.
- [25] L. Liu, H.-C. Wu, *Angew. Chem. Int. Ed.* **2016**, *55*, 15216–15222; *Angew. Chem.* **2016**, *128*, 15440–15446.
- [26] N. A. W. Bell, U. F. Keyser, *J. Am. Chem. Soc.* **2015**, *137*, 2035–2041.
- [27] S. Ranallo, M. Rossetti, K. W. Plaxco, A. Vallée-Bélisle, F. Ricci, *Angew. Chem. Int. Ed.* **2015**, *54*, 13214–13218; *Angew. Chem.* **2015**, *127*, 13412–13416.
- [28] R. Arts, S. K. J. Ludwig, B. C. B. Van Gerven, E. M. Estirado, L. G. Milroy, M. Merckx, *ACS Sens.* **2017**, *2*, 1730–1736.
- [29] A. Vallée-Bélisle, F. Ricci, K. W. Plaxco, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13802–13807.
- [30] A. Idili, A. Vallée-Bélisle, F. Ricci, *J. Am. Chem. Soc.* **2014**, *136*, 5836–5839.
- [31] X. J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- [32] G. Perugino, A. Vettone, G. Illiano, A. Valenti, M. C. Ferrara, M. Rossi, M. Ciaramella, *J. Biol. Chem.* **2012**, *287*, 4222–4231.
- [33] G. Perugino, R. Miggiano, M. Serpe, A. Vettone, A. Valenti, S. Lahiri, F. Rossi, M. Rossi, M. Rizzi, M. Ciaramella, *Nucleic Acids Res.* **2015**, *43*, 8801–8816.
- [34] F. Rossi, C. Morrone, A. Massarotti, D. M. Ferraris, A. Valenti, G. Perugino, R. Miggiano, *Biochem. Biophys. Res. Commun.* **2018**, *500*, 698–703.
- [35] R. Merlo, S. Del Prete, A. Valenti, R. Mattosovich, V. Carginale, C. T. Supuran, C. Capasso, G. Perugino, *J. Enzyme Inhib. Med. Chem.* **2019**, *34*, 490–499.
- [36] D. Mariottini, A. Idili, M. A. D. Nijenhuis, G. Ercolani, F. Ricci, *J. Am. Chem. Soc.* **2019**, *141*, 11367–11371.
- [37] B. Kaina, M. Christmann, *DNA Repair* **2019**, *78*, 128–141.
- [38] J. A. Quinn, S. X. Jiang, D. A. Reardon, A. Desjardins, J. J. Vredenburg, J. N. Rich, S. Gururangan, A. H. Friedman, D. D. Signer, J. H. Sampson, R. E. McLendon, J. E. Herndon, A. Walker, H. S. Friedman, *J. Clin. Oncol.* **2009**, *27*, 1262–1267.
- [39] J. Reinhard, U. Eichhorn, M. Wiessler, B. Kaina, *Int. J. Cancer* **2001**, *93*, 373–379.
- [40] M. R. Middleton, J. Kelly, N. Thatcher, D. J. Donnelly, R. S. McElhinney, T. B. H. McMurry, J. E. McCormick, G. P. Margison, *Int. J. Cancer* **2000**, *85*, 248–252.
- [41] M. Ranson, M. R. Middleton, J. Bridgewater, S. M. Lee, M. Dawson, D. Jowle, G. Halbert, S. Waller, H. McGrath, L.

Gumbrell, R. S. McElhinney, D. Donnelly, T. B. H. McMurry,  
G. P. Margison, *Clin. Cancer Res.* **2006**, *12*, 1577–1584.  
[42] R. Merlo, D. Caprioglio, M. Cillo, A. Valenti, R. Mattosovich,  
C. Morrone, A. Massarotti, F. Rossi, R. Miggiano, A. Leonardi,

A. Minassi, G. Perugino, *J. Enzyme Inhib. Med. Chem.* **2021**, *36*,  
85–97.

Manuscript received: December 7, 2020

Accepted manuscript online: January 8, 2020

Version of record online: February 17, 2021

---



# A journey down to hell: new thermostable protein-tags for biotechnology at high temperatures

Rosanna Mattosovich<sup>1</sup> · Rosa Merlo<sup>1</sup> · Angelo Fontana<sup>2</sup> · Giuliana d'Ippolito<sup>2</sup> · Michael P. Terns<sup>3</sup> · Elizabeth A. Watts<sup>3</sup> · Anna Valenti<sup>1</sup> · Giuseppe Perugino<sup>1</sup>

Received: 20 June 2019 / Accepted: 13 September 2019  
© Springer Japan KK, part of Springer Nature 2019

## Abstract

The specific labelling of proteins in recent years has made use of self-labelling proteins, such as the SNAP-tag<sup>®</sup> and the Halotag<sup>®</sup>. These enzymes, by their nature or suitably engineered, have the ability to specifically react with their respective substrates, but covalently retaining a part of them in the catalytic site upon reaction. This led to the synthesis of substrates conjugated with, e.g., fluorophores (proposing them as alternatives to fluorescent proteins), but also with others chemical groups, for numerous biotechnological applications. Recently, a mutant of the OGT from *Saccharolobus solfataricus* (H<sup>5</sup>) very stable to high temperatures and in the presence of physical and chemical denaturing agents has been proposed as a thermostable SNAP-tag<sup>®</sup> for in vivo and in vitro harsh reaction conditions. Here, we show two new thermostable OGTs from *Thermotoga neapolitana* and *Pyrococcus furiosus*, which, respectively, display a higher catalytic activity and thermostability respect to H<sup>5</sup>, proposing them as alternatives for in vivo studies in these extreme model organisms.

**Keywords** (Hyper)thermophiles · Thermostable proteins · Protein-tag · Biotechnology

---

Communicated by M. Moracci.

---

This manuscript is part of a special issue of Extremophiles journal for the 12th International Congress of Extremophiles (Extremophiles 2018) that was held on 16–20 September 2018 in Ischia, Naples, Italy.

---

Rosanna Mattosovich and Rosa Merlo equally contributed to the present work.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00792-019-01134-3>) contains supplementary material, which is available to authorized users.

---

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

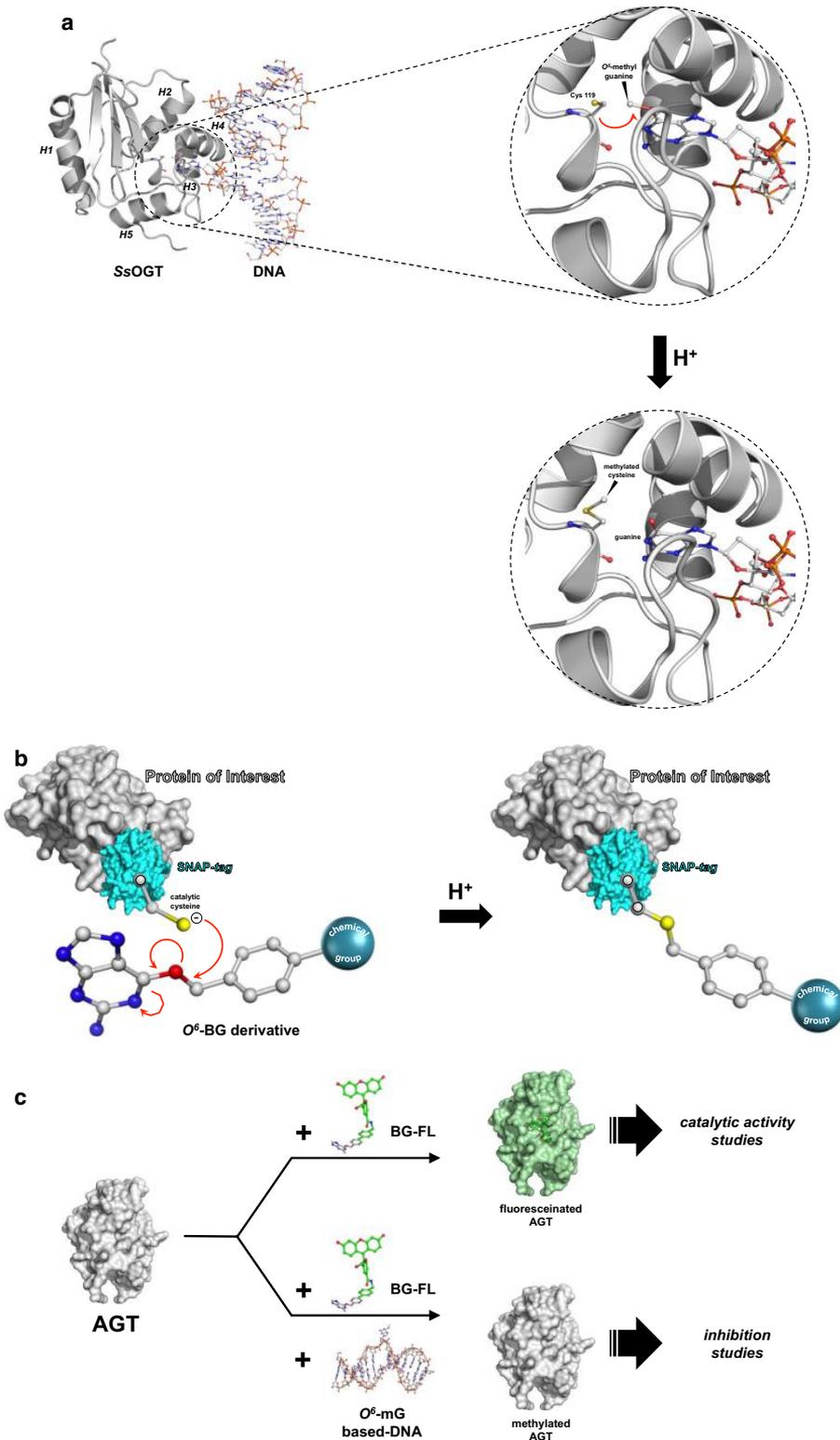
<sup>1</sup> Institute of Biosciences and BioResources, National Council of Research of Italy, Via P. Castellino 111, 80131 Naples, Italy

<sup>2</sup> Institute of Biomolecular Chemistry, National Council of Research of Italy, Via Campi Flegrei, 34, 80078 Pozzuoli, NA, Italy

<sup>3</sup> Departments of Biochemistry and Molecular Biology, Genetics, and Microbiology, University of Georgia, Athens, GA, USA

## Introduction

Protein-tags are short or long peptide sequences genetically fused to recombinant proteins for various purposes, as affinity purification, protein localization and general labelling procedures. In some cases, the presence of these tags enhances the solubilization of proteins and enzymes expressed in chaperone-deficient species such as *Escherichia coli*, to assist in the proper folding of proteins and avoid precipitation. These include commercially available thioredoxin (TRX), poly(NANP), maltose-binding protein (MBP), and glutathione S-transferase (GST). The discovery of Fluorescent Proteins (FPs) has revolutionized the world of cell and molecular biology, allowing several applications as reporter gene in fluorescence microscopy (Chalfie et al. 1994; Tsien 1998; Aliye et al. 2015). Although FPs are intrinsically fluorescent without the addition of any external substrate, they have some disadvantages. Indeed, their relatively large dimensions and the sensitivity to environmental changes (pH or the absence of O<sub>2</sub>) affect the formation of the internal natural fluorophore (Ashby et al. 2004; Campbell and Choy 2000).



**Fig. 1** **a** A cartoon of the AGTs irreversible reaction mechanism is shown as example: the *Ss*OGT enzyme (PDB ID: 4ZYE) recognizes the methylated guanine on DNA (from PDB ID: 4ZYD) and proceeds to the irreversible transfer of the methyl group from the guanine to the catalytic cysteine, inactivating the protein (from PDB ID: 4ZYG) (Perugino et al. 2015). **b** The SNAP-tag<sup>®</sup> technology is based of an engineered variant of the hAGT, which is able to recognize BG-derivative substrates, and irreversibly transferring in its active site a desired chemical group conjugated to the benzyl moiety. Consequently, this allows the labelling of a protein of interest expressed as fusion protein to the SNAP-tag<sup>®</sup>. **c** The use of commercially available fluorescent AGT substrates led to the setting up of new safety AGT's assay, for the determination of the catalytic activities, or inhibition studies in the presence of natural substrates (e.g., alkylated DNA) or AGT's inhibitors

Recently, studies on *O*<sup>6</sup>-alkylguanine-DNA-alkyl-transferases (AGTs, OGTs or MGMTs; EC: 2.1.1.63) led to the proposal of alternative protein-tags with new behaviours, useful for several applications in many fields. In nature, these ubiquitous small proteins have a crucial role in the direct repair of DNA by alkylating agents (Mishina et al. 2006; Serpe et al. 2019) and the human homolog (hAGT) is target of chemotherapeutic drugs for its crucial role in many types of tumours (Sun et al. 2018; Aoki and Natsume 2019). Their catalytic activity is based on the recognition of the damage on DNA (an *O*<sup>6</sup>-alkylguanine or an *O*<sup>4</sup>-alkylthymine) and, by a *one-step* reaction mechanism (via SN<sub>2</sub> type), the alkyl group from the damaged base is irreversibly transferred to a cysteine residue in their own active site (Daniels et al. 2000, 2004; Fang et al. 2005; Tubbs et al. 2007; Pegg 2011; Yang et al. 2009) (Fig. 1a). For these reasons, they are called suicide or kamikaze proteins, with a 1:1 stoichiometry of their reaction with the natural substrate. However, some AGTs resulted very reactive with a strong inhibitor, the *O*<sup>6</sup>-benzyl-guanine (*O*<sup>6</sup>-BG), which was used in combination with chemotherapeutic drugs (Pegg et al. 2001; Luu et al. 2002; Coulter et al. 2007). This information led professor Kai Johnsson and his group to put their effort for the production of a new protein-tag from a variant of the hAGT, introducing the SNAP-tag<sup>®</sup> technology in the “biotech” scenario (Keppler et al. 2003, 2004; Gronemeyer et al. 2006; Gautier et al. 2008; Mollwitz et al. 2012): upon the irreversible reaction, hAGT keeps the benzyl moiety of the substrate covalently linked to its catalytic cysteine. This makes possible the labelling of this protein (and a relative protein of interest, if expressed as fusions with it) when any chemical group conjugated to the benzyl moiety of *O*<sup>6</sup>-BG (Keppler et al. 2003, 2004; Gautier et al. 2008) (Fig. 1b). Furthermore, the commercially available fluorescent *O*<sup>6</sup>-BG derivatives (as the SNAP-Vista Green<sup>®</sup> and the SNAP-Cell<sup>®</sup> TMR Star; New England Biolabs) allowed the development of new applications in the fluorescent microscopy field, and recently of a new fluorescent assay for all the BG-sensitive AGTs, overcoming the traditional, long and unsafe assays

for this class of proteins based on HPLC procedures and radioactive substrates (Hishiguro et al. 2008; Perugino et al. 2012). The presence of methylated-DNA in this assay, led to the determination of the activity of AGTs on their natural substrates, by applying the classical enzyme inhibition approaches (Fig. 1c) (Perugino et al. 2012, 2015; Miggiano et al. 2013, 2017; Vettone et al. 2016; Morrone et al. 2017).

All the above-mentioned protein-tags have the disadvantage to be employed in mild reaction conditions and in mesophilic organisms. This was successfully overcome by the introduction of a thermostable OGT from the hyperthermophilic archaea *Saccharolobus solfataricus* (formerly *Sulfolobus solfataricus*). Starting from the studies of the wild-type enzyme (*Ss*OGT), which not only displays the same behaviours of the hAGT on BG derivatives, but is also characterized by an exceptional stability at extremes of temperature, pH, ionic strength and the presence of denaturing agents (Perugino et al. 2012), we developed a DNA-bindingless variant (called H<sup>5</sup>), which resulted in a strong “thermostable SNAP-tag<sup>®</sup>”. H<sup>5</sup> was successfully employed in *in vivo* heterologous expression in thermophilic organisms as *Thermus thermophilus* HB27 and *Sulfolobus islandicus* E233S1 (Vettone et al. 2016; Visone et al. 2017), in fusion with a thermostable β-glycosidase and a *S. solfataricus* reverse gyrase (Vettone et al. 2016; Visone et al. 2017), as well as under extreme reaction conditions as gene reporter in *in vitro* transcription/translation systems based on *Sulfolobus* cell lysates (Lo Gullo et al. 2019). Furthermore, H<sup>5</sup> was successfully fused to the N-terminal domain of the ice nucleation protein (INPN) from the Gram-negative bacterium *Pseudomonas syringae*, a transmembrane protein useful for the one-step heterologous expression and the *in vivo* immobilization of proteins of interest on the outer membrane of *E. coli* (Samuelson et al. 2002). This led to the development of the new Anchoring-and-self-labelling-protein-tag (ASL<sup>tag</sup>) (Merlo et al. 2019), which simultaneously allows the immobilization (by INPN) and the quantitative determination of the yield of an immobilized protein (by the fluorescent assay using H<sup>5</sup>). Surprisingly, the presence of the thermostable H<sup>5</sup> between the INPN and a protein of interest resulted in an enhancement of the stability of the latter, as it was the case for the *Sulfurihydrogenibium yellowstonense* α-carbonic anhydrase (Del Prete et al. 2019).

The growing demand to use recent technologies at very extreme conditions in thermophilic bacteria and archaea, such as the *in vivo* CRISPR-Cas immune systems, leads us to search for new protein-tags with very high activity and thermostability. In this regard, the present work is focussed on the characterization of two new OGTs, from the hyperthermophilic organisms *Thermotoga neapolitana* and *Pyrococcus furiosus*. The purified proteins were compared to *Ss*OGT, showing respect to its exceptional characteristics in terms of enzymatic activity and thermostability.

These results open promising perspectives in the development of new protein-tags to employ in these extreme model organisms.

## Materials and methods

### Reagents

Fine chemicals were from Sigma-Aldrich, SNAP-Vista Green<sup>®</sup> fluorescent substrate (hereinafter BG-FL) was from New England Biolabs (Ipswich, MA). SYPRO Orange 5000× (Invitrogen). Synthetic oligonucleotides listed in Table 1 were from Eurofins (Milan, Italy); *Pfu* DNA polymerase was from NZYTech (Portugal). The Bio-Rad protein assay kit (Bio-Rad Pacific) was used for the determination of the protein concentration, using purified BSA as standard.

### DNA constructs

The cloning procedures for the construction of *E. coli* expression plasmids were the same for both the proteins: the ORF CTN1690/PF1878, encoding a putative OGT, was amplified from genomic DNA from *Thermotoga neapolitana* DSMZ 4359<sup>T</sup>/*Pyrococcus furiosus* JFW02 strain genomic DNA, using Lig5:Lig3 *Tn*OGT/*Pfu*OGT oligonucleotides pairs (listed in Table 1) and directly cloned into the expression vector pHTP1 (NZYTech, Portugal), following the instructions described in the NZYEasy Cloning and Expression kit I (NZYTech, Portugal) manual. The ligation mixture was entirely used to transform commercial *E. coli* DH5 $\alpha$  cells (NZY5 $\alpha$  Competent Cells-NZYTech, Portugal) and positive clones were confirmed by PCR. Subsequently, a DNA fragment from the resulted pHTP1-*Tn*OGT/pHTP1-*Pfu*OGT was removed by digestion with Nco I restriction endonuclease, and was replaced by a double-stranded oligonucleotides (NZY-His Fwd2 and NZY-His Rev2; Table 1), whose DNA sequence expresses a shorter His<sub>6</sub>-tag (MAHHHHHTG-), similar to that at the N-terminal of the *Ss*OGT protein (Perugino

et al. 2012). Positive clones after transformation of the ligation mixture in *E. coli* KRX competent cells were confirmed by DNA sequencing.

### Protein purification

*Tn*OGT and *Pfu*OGT were expressed in the *E. coli* BL21 (DE3) cells grown overnight at 37 °C in Luria–Bertani (LB) selective medium supplemented with 50 mg/L kanamycin and 30 mg/L chloramphenicol, and the protein expression was induced with 1.0 mM isopropyl-thio- $\beta$ -D-galactopyranoside (IPTG), when an absorbance value of 0.5–0.6 A<sub>600 nm</sub> was reached. Harvested cells were resuspended 1:3 (w/v) in Buffer A (50 mM phosphate, 300 mM NaCl; pH 8.0) supplemented with 1% Triton X-100 and stored overnight at –20 °C. After this first step of lysis, the biomass was treated with lysozyme and DNase for 60 min in ice, followed by a sonication step. Finally, the lysate was centrifuged for 30 min at 60,000 $\times$ g in and the cell extract recovered. To remove *E. coli* contaminants, all cell extracts were incubated 20 min at 70.0 °C and 20 min at 65 °C, respectively, followed by a centrifugation at 13,000 $\times$ g at 4.0 °C; the supernatant was diluted 1:2 (v/v) in purification Buffer A and applied to a Protino Ni–NTA Column 1.0 mL (Macherey–Nagel) for His<sub>6</sub>-tag affinity chromatography. After two washing steps of 10 column volumes of Buffer A and 10 column volumes of Buffer A supplemented with 25.0 mM imidazole, the elution was performed in 20 column volumes of buffer A, by applying a linear gradient of 25.0–250.0 mM imidazole. The fractions containing the protein were collected and analysed by SDS-PAGE. *Tn*OGT was dialysed against PBS 1 $\times$  buffer (phosphate buffer 20 mM, NaCl 150 mM; pH 7.3), whereas the fractions of the *Pfu*OGT protein were pooled, concentrated and subjected to a further gel-filtration chromatography, using a Superdex 75 10/300 GL column (GE Healthcare Life Sciences). Finally, both the proteins were concentrated and loaded on 15% SDS-PAGE gel to confirm its purity and stored at –20.0 °C.

**Table 1** List of oligonucleotides used in this study

Name	Sequence	Notes
Lig5- <i>Pfu</i> OGT	5'-TCAGCAAGGGCTGAGGCCATGGTATTGGAAGTTAGG-3'	Nco I site underlined
Lig3- <i>Pfu</i> OGT	5'-CCTCAGCGGAAGCTGAGGTTAGCTTGTCCATCCTTCC-3'	
Lig5- <i>Tn</i> OGT	5'-TCAGCAAGGGCTCAGGCCATGGGAGATCGA-3'	Nco I site underlined
Lig3- <i>Tn</i> OGT	5'-CCTCAGCGGAAGCTGAGGTTATCGACTACCTCGC-3'	
NZY-His fwd2	5'-catgGCACACCATCACCATCACCATACGGG-3'	Inserting an His <sub>6</sub> -tag (MAHHHHHTG-; underlined) upstream the <i>Pfu</i> OGT and <i>Tn</i> OGT sequence
NZY-His rev2	5'-catgCCCCTATGGTGATGGTGATGGTGTGC-3'	
Fwd <sup>m4</sup>	5'-ggcMgtaggcctagcatgacaatctgcattggtgatcacgg-3'	From Perugino et al. (2015); M = O <sup>6</sup> -methyl-guanine
Rev4	5'-ccgtgatcaccaatgcagattgtcatgctaggcctaccgcc-3'	From Perugino et al. (2015)

## In vitro alkyl-transferase assay

The fluorescent substrate BG-FL was used for the determination of the catalytic activity of all thermostable enzymes analysed, as previously described (Perugino et al. 2012, 2015; Miggiano et al. 2013; Vettone et al. 2016; Visone et al. 2017; Merlo et al. 2019; Del Prete et al. 2019). Briefly, 5.0  $\mu\text{M}$  of protein (ca. 0.1 mg/mL) was incubated with 10.0  $\mu\text{M}$  of BG-FL in Fluo Buffer 1  $\times$  (50.0 mM phosphate, 100.0 mM NaCl, 1.0 mM DTT; pH 6.5) at different temperatures and times, as indicated; each reaction was stopped by adding a Laemmli buffer 1  $\times$  (formamide 95%; EDTA 20.0 mM; bromophenol 0.05%), followed by denaturation at 100.0  $^{\circ}\text{C}$  and the direct loading of the sample on SDS-PAGE. The gel was first analysed by fluorescence imaging on a VersaDoc 4000<sup>TM</sup> system (Bio-Rad) by applying as excitation/emission parameters a blue LED/530 bandpass filter, and then was stained by Coomassie. Assuming the irreversible mechanism with 1:1 BG-FL/OGT ratio, fluorescence intensity data were corrected for the amount of loaded protein, and fitted by applying the second-order rate equation, to determine the relative amount of covalently modified protein in time-course experiments. (Gautier et al. 2008; Miggiano et al. 2013; Perugino et al. 2012, 2015).

## Competitive assay and IC<sub>50</sub> calculation

Competitive assay using the fluorescent substrate in the presence of double strands (ds) oligonucleotides pairs (Fwd<sup>m4</sup>: Rev4; Table 1), containing a single O<sup>6</sup>-methyl-guanine, was performed as described (Perugino et al. 2015) to determine the half maximal inhibitory concentration (IC<sub>50</sub>), that is the concentration of methylated DNA needed to reduce the fluorescence intensity of the OGT band by 50.0%. Reactions incubated at fixed temperatures with increasing concentrations (0.0–10.0  $\mu\text{M}$ ) of ds-Fwd<sup>m4</sup> and keeping constant the BG-FL concentration (5.0  $\mu\text{M}$ ) were performed for 10 min at 50.0  $^{\circ}\text{C}$ . Corrected data of fluorescence intensity were fitted with the IC<sub>50</sub> equation (Perugino et al. 2015; Morrone et al. 2017).

## Protein stability analysis

The stability at several conditions of the thermostable OGTs was analysed by the differential scan fluorimetry method (DSF), adapted by a protocol previously described for the SsOGT and relative mutants (Niesen et al. 2007; Perugino et al. 2015; Vettone et al. 2016; Morrone et al. 2017). Triplicates of each condition containing 25.0  $\mu\text{M}$  of enzyme (ca. 0.5 mg/mL) in PBS 1 $\times$  buffer and SYPRO Orange dye 1 $\times$  were subjected to a scan of 70 cycles at temperatures from 25.0 to 94.0  $^{\circ}\text{C}$  for 10 min/ $^{\circ}\text{C}$   $\times$  cycle, and analysed in a Real-Time Light Cycler<sup>TM</sup> (Bio-Rad). Relative fluorescence

data were then normalized to the maximum fluorescence value within each scan. Obtained plots of fluorescence intensity vs temperature displayed sigmoidal curves (typical of a two-state transition), which allowed the determination of the inflection points ( $T_m$  values) by fitting the Boltzmann equation (Niesen et al. 2007; Perugino et al. 2015; Vettone et al. 2016; Morrone et al. 2017).

## Data analysis and softwares

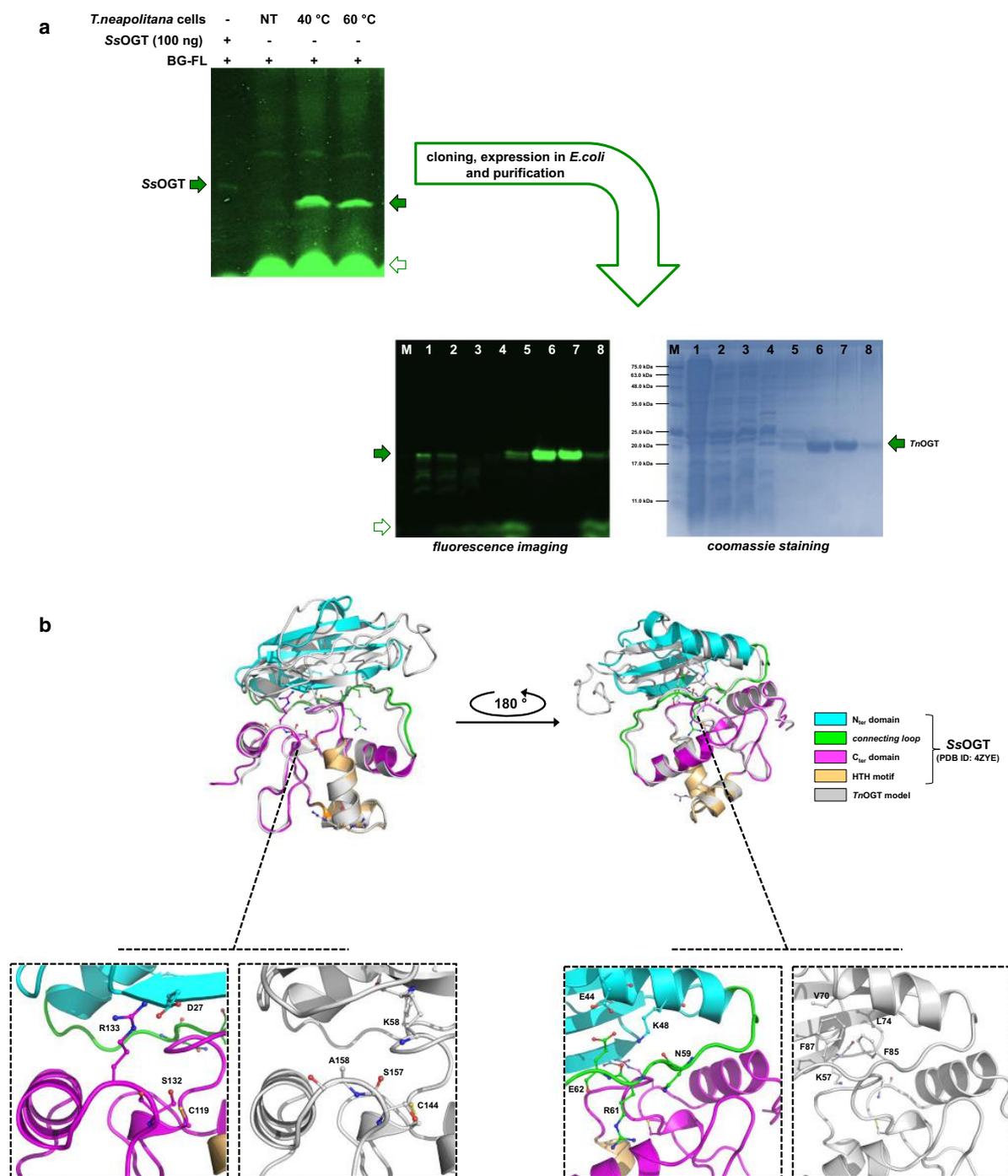
Prism Software Package (GraphPad Software) and GraFit 5.0 Data Analysis Software (Erithacus Software) were used for the data analysis from activity, competitive inhibition and stability assays.

## Results and discussion

### The OGT from *Thermotoga neapolitana*

*Thermotoga neapolitana* is a hyperthermophilic Gram-negative bacterium of the order *Thermotogales* (Belkin et al. 1986; Jannasch et al. 1988) that include good candidates for genetic engineering and biotechnological applications (Conners et al. 2006; Zhang et al. 2015; Fink et al. 2016; Donaldson et al. 2017; Han and Xu 2017). Several *Thermotogales* have been studied for the fermentative production of hydrogen (H<sub>2</sub>) by [FeFe] hydrogenases with yield close to the theoretical Thauer limit of four moles of H<sub>2</sub> per mole of consumed sugar (d'Ippolito et al. 2010; Pradhan et al. 2015, 2016). *T. neapolitana* also shows a novel, anaerobic process named capnophilic lactic fermentation (CLF) that leads to the synthesis of lactic acid (LA) without affecting H<sub>2</sub> production under CO<sub>2</sub> trigger (Di Pasquale et al. 2014; d'Ippolito et al. 2014; Pradhan et al. 2017).

The ORF CTN1690 of *T. neapolitana* encodes a putative 174-aa polypeptide, with a calculated molecular weight of 19.9 kDa, and a clear homology to the O<sup>6</sup>-alkylguanine-DNA-alkyl-transferases. Furthermore, lyophilised *T. neapolitana* cells incubated with the BG-FL substrate showed a strong fluorescent signal close to that of SsOGT by gel imaging after SDS-PAGE (Fig. 2a). The observed molecular weight and, mostly the sensitivity to the benzyl-guanine derivative BG-FL substrate, led to the cloning and the heterologous expression in *E. coli* of the His-tagged version of this protein. After purification by affinity chromatography, the protein was fully active on BG-FL (Fig. 2a) and was further subjected to a biochemical characterization. The inhibition assay in the presence of methylated-dsDNA (Table 2) and the fluorescent substrate BG-FL allowed the determination of an IC<sub>50</sub> value similar to that obtained with the *S. solfataricus* enzyme, thus confirming a role in DNA repair of this thermophilic



protein, hereinafter properly named *TnOGT* (Table 2; Perugino et al. 2015). Surprisingly, the enzyme from *T. neapolitana* displayed a very high activity at low temperatures (Table 3), similar to that shown by the SsOGT-H<sup>5</sup> mutant (Perugino et al. 2012; Vettone et al. 2016). This

feature hampered the determination of second order constants at temperatures above 50.0 °C, since its reaction rate went beyond the technical limits of our assay. Such a high activity at moderate temperatures was obtained with the H<sup>5</sup> mutant by replacing the conserved S132 with

**Fig. 2** The OGT from *Thermotoga neapolitana*. **a** Lyophilized *T. neapolitana* cells grown in the presence of CO<sub>2</sub>, were resuspended in PBS 1× buffer and BG-FL 5.0 μM and incubated 120 min at the indicated temperatures; NT, resuspended cells immediately loaded on SDS-PAGE. The *Tn*OGT gene was expressed in *E. coli* and protein was purified by His-tag affinity chromatography, as described in the “Materials and methods”. Lane M: molecular weight marker; lane 1: cell-free extract; lane 2: flowthrough; lanes 3 and 4: column washing; lanes 5–8: eluted protein by imidazole gradient. Filled and empty green arrows indicate labelled proteins and the BG-FL substrate, respectively. **b** Superimposition between the *Ss*OGT 3D structure (PDB ID: 4ZYE; coloured as described in the legend) and a model of the *Tn*OGT (in gray). Insets represent a zoom-view of local ionic interactions in *Ss*OGT compared with the same positions in *Tn*OGT. Atoms are coloured according the CPK convention (carbon, in the corresponding colour of each 3D structure; nitrogen, in blue; oxygen, in red; sulphur, in yellow)

a glutamic acid residue. This replacement was also performed on the SNAP-tag<sup>®</sup>: the substitution led to a strong enhancement of the activity of both these engineered OGTs towards the *O*<sup>6</sup>-BG derivative substrates (Juillerat et al. 2003; Perugino et al. 2012). However, the superimposition analysis between a *Tn*OGT model (constructed by the i-TASSER freeware; <https://zhanglab.ccmh.med.umich.edu/I-TASSER/>) and the 3D structure of the free form of *Ss*OGT (PDB ID: 4ZYE) revealed the presence of the conserved serine in *Tn*OGT (Fig. 2b): evidently, other residues in the active site contribute to the exceptional catalytic activity of this enzyme.

On the other hand, *Tn*OGT was unusually prone to degradation during the storage at –20.0 °C, showing a series of bands below the full length, both in fluorescence imaging and in coomassie staining (Fig. S1a). This lability hampered measurements of thermostability from the DSF analysis. The reason for this degradation is unexpected: the recombinant enzyme is in its full-length form in *E. coli* cells, as well as after heat treatment of the cell free extract, and finally after purification by affinity chromatography (see “Materials and methods”; Fig. 2a). From coomassie staining analysis, three main bands of approx. 17.5, 16.1 and 14.7 kDa were detected (Fig. S1). Since they were also fluorescent (and, therefore, catalytically active), it is probable that these cuts are at the expense of the N-terminal domain, still keeping the polypeptides joined together, and making the enzyme catalytically active in solution. Again, from the superimposition, in *Tn*OGT some residues that play an important role in the stabilization of *Ss*OGT are missing. In particular, the ionic interactions that have a crucial role in the stability of the *Saccharolobus* enzyme at high temperatures, as the R133-D27 pair (Perugino et al. 2015) and the K-48 network (Morrone et al. 2017), are mainly replaced by hydrophobic residues in the *Thermotoga* homolog.

## The *Pyrococcus furiosus* *O*<sup>6</sup>-alkylguanine-DNA-alkyl-transferase

*Pyrococcus furiosus* is one of the best-studied representatives among microorganisms able to thrive above the boiling point of water (Kengen 2017). It was originally isolated anaerobically from geothermally heated marine sediments of Porto Levante (Vulcano Island, Italy), and described in 1986 by Karl Stetter and Gerhard Fiala (Fiala and Stetter 1986). This microorganism in the last decades was a source of thermostable enzymes, with potential applications in various industrial processes: the most famous example is the DNA polymerase I described in 1991 (Lundberg et al. 1991), possessing an associated 3′–5′ exonuclease activity (Kengen 2017). Recently, the discovery of the CRISPR-Cas systems in *P. furiosus* provides fundamental knowledge for new biomedical and biotechnological applications (Hael et al. 2009; Terns and Terns 2013).

In 1998, Margison and co-workers demonstrated the presence of an OGT activity in *P. furiosus*, identifying a 22.0 kDa size band by an SDS-PAGE fluorography assay

**Table 2** DNA repair activity of OGTs by competitive inhibition studies in the presence of BG-FL (used as substrate) and a ds oligonucleotide containing an *O*<sup>6</sup>-methylated-guanine (as inhibitor; see Table 1)

	IC <sub>50</sub> (μM)	Notes
<i>Ss</i> OGT	1.01 ± 0.08	From Perugino et al. (2015)
<i>Tn</i> OGT	0.53 ± 0.13	This study
<i>Pfu</i> OGT	0.88 ± 0.10 <sup>a</sup>	This study

<sup>a</sup>Performed at 65 °C

**Table 3** Catalytic activities as a function of temperature of thermostable OGTs, expressed as second-order rate constant values in the presence of the sole BG-FL substrate

	T (°C)	K (s <sup>-1</sup> M <sup>-1</sup> )	Notes
<i>Ss</i> OGT	25	2.80 × 10 <sup>3</sup>	From Perugino et al. (2012)
	50	1.50 × 10 <sup>4</sup>	This study
	70	5.33 × 10 <sup>4</sup>	From Perugino et al. (2012)
	80	ND	This study
<i>Tn</i> OGT	25	4.65 × 10 <sup>4</sup>	This study
	50	2.19 × 10 <sup>4</sup>	This study
	70	ND	This study
	80	ND	This study
<i>Pfu</i> OGT	25	ND	This study
	50	1.80 × 10 <sup>3</sup>	This study
	70	2.30 × 10 <sup>3</sup>	This study
	80	1.20 × 10 <sup>4</sup>	This study
	90	1.50 × 10 <sup>5</sup>	This study

ND not determined

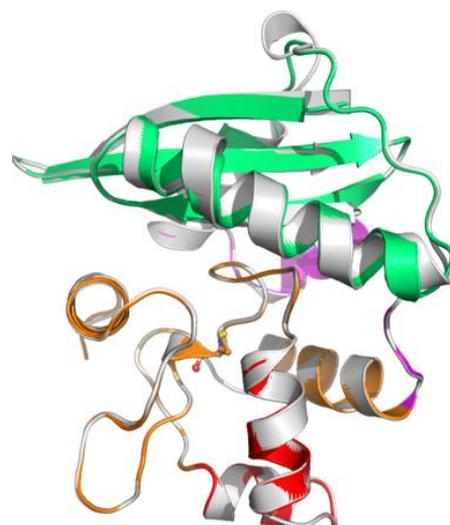
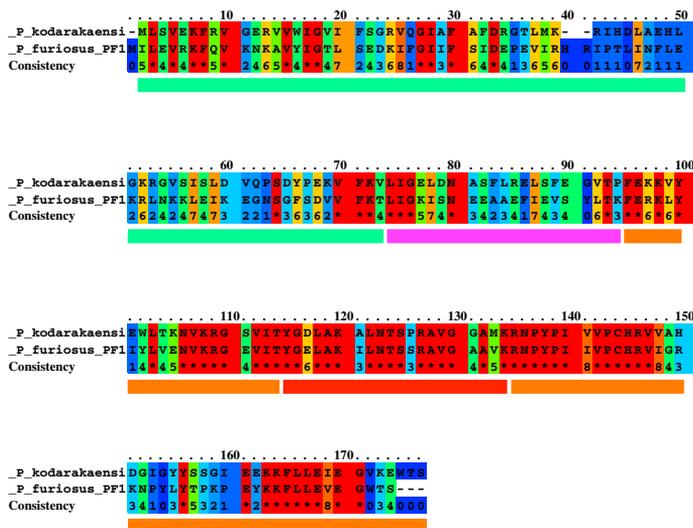
(Skorvaga et al. 1998). Furthermore, this activity was completely abolished by the treatment with the  $O^6$ -benzyl-guanine ( $O^6$ -BG) inhibitor. This information opened the possibility of employing a hyper-thermostable enzyme as SNAP-tag<sup>®</sup> in in vivo CRISPR-Cas system-based applications. The ORF PF1878 is relative to a 174-aa polypeptide, with an expected m.w. of 20.1 kDa: the primary structure is closed to that from the MGMT from *Thermococcus kodakarensis* (*Tk*-MGMT), a well-known enzyme, in terms of structure and thermal stability (Fig. 3) (Leclere et al. 1998; Hashimoto et al. 1999; Nishikori et al. 2005). *Tk*-MGMT is a very thermostable enzyme, and from its solved 3D structure (PDB ID: 1MGT) emerges that a lot of intra-helix ion-pairs contribute to reinforce stability of  $\alpha$ -helices, whereas the presence of inter-helix ion-pairs stabilize internal packing of tertiary structure (Hashimoto et al. 1999).

The cloning of the ORF PF1878 and the subsequent expression and purification of the relative protein allowed to a complete characterization of this new enzyme. Likewise *Tn*OGT, the enzyme from *P. furiosus* is fully active on BG-FL substrate and displayed a clear ability to repair methylated DNA, as shown in the IC<sub>50</sub> experiment listed in Table 2. On the other hand, *Pfu*OGT is a strong thermophilic enzyme, displaying a measurable catalytic activity only at very high temperatures (Table 4), whereas at moderate

temperatures the rate of the reaction is slow, making difficult to perform the fluorescent assay. For this reason, the competitive inhibition in the presence of methylated DNA was performed at 65.0 °C instead of the standard procedure at 50.0 °C (Perugino et al. 2015; Morrone et al. 2017); however, at this temperature, the activity of *Pfu*OGT on single- and/or double-methylated DNA cannot be excluded.

The stability of this enzyme was compared with that of *Ss*OGT by the Differential Scan Fluorimetry analysis: the latter was previously treated using a scan rate of 5 min/°C × cycle (Perugino et al. 2015; Vettone et al. 2016) instead of the classical 1 min/°C × cycle (Niesen et al. 2007). Due to its very high thermal stability, a further increase of the time (10 min/°C × cycle) was necessary for *Pfu*OGT, to obtain the sigmoidal curve to fit with the Boltzmann equation (Niesen et al. 2007). In these new conditions, the  $T_m$  value of *Ss*OGT drops by approx. 13.0 °C ( $67.9 \pm 1.1$ ; Table 4), whereas *Pfu*OGT displayed a  $T_m$  value over 80.0 °C. The stability of *Pfu*OGT was also tested in the presence of perturbants, as high ionic strength or detergents. In the first case, we tested the importance of ionic interactions involved in the maintenance of the structure: as expected, the presence of 4.0 M NaCl strongly affected the stability of *Pfu*OGT, whereas *Ss*OGT has even shown an increase of the  $T_m$  value. Nevertheless, *Ss*OGT is more sensitive to the SDS,

Unconserved 1 2 3 4 5 6 7 8 9 10 Conserved



█ N<sub>ter</sub> domain  
█ connecting loop  
█ C<sub>ter</sub> domain  
█ HTH motif  
█ PFOGT model

**TkMGMT**  
(PDB ID: 1MGT)

**Fig. 3** Primary structure alignment and superimposition between the *Pfu*OGT and the OGT from *T. kodakarensis* (*Tk*-MGMT; PDB ID: 1MGT). Conservation of residues and protein domains are coloured on the basis of the respective legends

**Table 4** Protein thermal stability by DSF method.

	Conditions	$T_m$ (°C)	Rate (min/°C×cycle)	Notes
SsOGT	PBS 1×	80.0±0.4	5	From Perugini et al. (2015)
	PBS 1×	67.9±1.1	10	This study
	PBS 1×; NaCl 1.0 M	79.6±0.3	10	This study
	PBS 1×; NaCl 4.0 M	82.9±0.4	10	This study
	PBS 1×; SDS 0.01%	46.5±1.7	10	This study
PfuOGT	PBS 1×	78.8±0.4	10	This study
	PBS 1×; NaCl 1.0 M	83.7±0.3	10	This study
	PBS 1×; NaCl 4.0 M	50.4±2.1	10	This study
	PBS 1×; SDS 0.01%	71.8±0.3	10	This study

$T_m$  values were obtained by plotting the relative fluorescence intensity as a function of temperature. Data were achieved from three independent experiments

showing a collapsed  $T_m$  value up to 22.0 °C, while PfuOGT drops by about 10.0 °C. Taken together, the data obtained clearly indicate a difference in the strategies of protein stabilization adopted by these two enzymes (Table 4).

## Conclusion and perspectives

The modification of AGTs to produce new protein-tags for use in the “SNAP-tag® technology” offers a lot of advantages for the specific labelling of a protein of interest with an innumerable number of chemical groups conjugated to a classical inhibitor of this class of enzymes, the  $O^6$ -BG (Fig. 1b) (Hinner and Johnsson 2010). A further step forward was to adapt to this new biotechnology to a thermostable OGT from a microorganism that lives at high temperatures. After the production of SsOGT-H<sup>5</sup> mutant, here we propose two new thermostable OGTs, which will be modified in the future to abolish their ability to bind DNA, without, however, decreasing their activity and stability to heat and to denaturing agents. The new protein-tags can be used in *T. neapolitana* and *P. furiosus*, to analyze in vivo the functions of proteins and enzymes of interest in these model systems.

**Acknowledgements** This work was supported by Fondazione CARIPLO (Ricerca biomedical condotta da giovani ricercatori, project 2016-0604).

## 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
- Aoki K, Natsume A (2019) Overview of DNA methylation in adult diffuse gliomas. *Brain Tumor Pathol* 36:84–91
- Ashby MC, Ibaraki K, Henley JM (2004) It's green outside: tracking cell surface proteins with pH-sensitive GFP. *Trends Neurosci* 27:257–261
- Belkin S, Wirsén CO, Jannasch HW (1986) A new sulfur-reducing, extremely thermophilic eubacterium from a submarine thermal vent. *Appl Environ Microbiol* 51:1180–1185
- Campbell TN, Choy FYM (2000) The effect of pH on green fluorescent protein: a brief review. *Mol Biol Today* 2:1–4
- Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. *Science* 5148:802–805
- Connors SB, Mongodin EF, Johnson MR, MonteroCI Nelson KE, Kelly RM (2006) Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS Microb Rev* 30:872–905
- Coulter R, Blandino M, Tomlinson JM, Pauly GT, Krajewska M, Moschel RC, Peterson LA, Pegg AE, Spratt TE (2007) Differences in the rate of repair of  $O^6$ -alkylguanines in different sequence contexts by  $O^6$ -alkylguanine-DNA alkyltransferase. *Chem Res Toxicol* 20:1966–1971
- d'Ippolito G, Dipasquale L, Vella FM, Romano I, Gambacorta A, Cutignano A, Fontana A (2010) Hydrogen metabolism in the extreme thermophile *Thermotoga neapolitana*. *Int J Hydrog En* 35:2290–2295
- d'Ippolito G, Dipasquale L, Fontana A (2014) Recycling of carbon dioxide and acetate as lactic acid by the hydrogen-producing bacterium *Thermotoga neapolitana*. *Chemsuschem* 7:2678–2683
- 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
- Del Prete S, Merlo R, Valenti A, Mattossovich R, Rossi M, Carginale M, Supuran CT, Perugini G, Capasso C (2019) Thermostability enhancement of the  $\alpha$ -carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* by using the anchoring-and-self-labelling-protein-tag system (ASL<sup>ts</sup>). *J Enz Inhib Med Chem* 34:946–954
- Dipasquale L, d'Ippolito G, Fontana A (2014) Capnophilic lactic fermentation and hydrogen synthesis by *Thermotoga neapolitana*: an unexpected deviation from the dark fermentation model. *Int J Hydrog En* 39:4857–4862

- Donaldson T, Iozzino L, Deacon LJ, Billones H, Ausili A, D'Auria S, Dattelbaum JD (2017) Engineering a switch-based biosensor for arginine using a *Thermotoga maritima* periplasmic binding protein. *An Biochem* 525:60–66
- Fang Q, Kanugula S, Pegg AE (2005) Function of domains of human  $O^6$ -alkyl-guanine-DNA alkyltransferase. *Biochemistry* 44:15396–15405
- Fiala G, Stetter KO (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100 °C. *Arch Microb* 145:56–61
- Fink M, Trunk S, Hall M, Schwab H, Steiner K (2016) Engineering of TM1459 from *Thermotoga maritima* for increased oxidative alkene cleavage activity. *Front Microb*. <https://doi.org/10.3389/fmicb.2016.01511>
- Gautier A, Juillerat A, Heinis C, Corrêa IR Jr, Kindermann M, Beauflis 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  $O^6$ -alkylguanine-DNA alkyltransferase for applications in protein labeling. *Prot Eng Des Sel* 19:309–316
- Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139:945–956
- Han D, Xu Z (2017) Development of a pyrE-based selective system for *Thermotoga* sp. strain RQ7. *Extremophiles* 21:297–306
- Hashimoto H, Inouel T, Nishioka M, Fujiwara S, Takagi M, Imanaka T, Kai Y (1999) Hyperthermostable protein structure maintained by intra and inter-helix ion-pairs in archaeal  $O^6$ -methylguanine-DNA methyltransferase. *J Mol Biol* 292:707–716
- Hinner MJ, Johnsson K (2010) How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol* 21:766–776
- Ishiguro K, Shyam K, Penketh PG, Sartorelli AC (2008) Development of an  $O^6$ -alkyl-guanine-DNA alkyltransferase assay based on covalent transfer of the benzyl moiety from [benzene-3H] $O^6$ -benzylguanine to the protein. *Anal Biochem* 383:44–51
- Jannasch HW, Huber R, Belkin S, Stetter KO (1988) *Thermotoga neapolitana* sp. nov. of the extremely thermophilic, eubacterial genus *Thermotoga*. *Arch Microbiol* 150:103–104
- 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(4):313–317
- Kengen SWM (2017) *Pyrococcus furiosus*, 30 years on. *Microb Biotechnol* 10:1441–1444
- Keppler 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
- 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* 10:9955–9959
- Leclere MM, Nishioka M, Yuasa T, Fujiwara S, Takagi M, Imanaka T (1998) The  $O^6$ -methylguanine-DNA methyltransferase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1: a thermostable repair enzyme. *Mol Gen Genet* 258:69–77
- Lo-Gullo G, Mattosovich R, Perugino G, La Teana A, Londei P, Benelli D (2019) Optimization of an in vitro transcription/translation system based on *Sulfolobus solfataricus* cell lysate. *Archaea* 1:1–10
- Lundberg KS, Shoemaker DD, Adams MWW, Short JM, Sorge JA, Mathur EJ (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* 108:1–6
- Luu KX, Kanugula S, Pegg AE, Pauly GT, Moschel RC (2002) Repair of oligodeoxyribonucleotides by  $O^6$ -alkylguanine-DNA alkyltransferase. *Biochemistry* 41:8689–8697
- Merlo R, Del Prete S, Valenti A, Mattosovich R, Carginale V, Supuran CT, Capasso C, Perugino G (2019) An AGT-based protein-tag system for the labelling and surface immobilization of enzymes on *E. coli* outer membrane. *J Enz Inhib Med Chem* 34:490–499
- 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
- Miggiano R, Valenti A, Rossi F, Rizzi M, Perugino G, Ciaramella M (2017) Every OGT is illuminated ... by fluorescent and synchrotron lights. *Int J Mol Sci* 18:2613–2631
- Mishina Y, Duguid EM, He C (2006) Direct reversal of DNA alkylation damage. *Chem Rev* 106:215–232
- 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
- Morrone C, Miggiano R, Serpe M, Massarotti A, Valenti A, del Monaco G, Rossi M, Rossi F, Rizzi M, Perugino G, Ciaramella M (2017) Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase. *Biochem Biophys Acta* 1861:86–96
- 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
- Nishikori S, Shiraki K, Fujiwara S, Imanaka T, Takagi M (2005) Unfolding mechanism of a hyperthermophilic protein  $O^6$ -methylguanine-DNA methyltransferase. *Biophys Chem* 116:97–104
- 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
- Pegg AE, Goodtzova K, Loktionova NA, Kanugula S, Pauly GT, Moschel RC (2001) Inactivation of human  $O^6$ -alkylguanine-DNA alkyltransferase by modified oligodeoxyribonucleotides containing  $O^6$ -benzylguanine. *J Pharmacol Exp Ther* 296:958–965
- 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. *Nucleic Acids Res* 43:8801–8816
- Pradhan N, Dipasquale L, d'Ippolito G, Panico A, Lens PNL, Esposito G, Fontana A (2015) Hydrogen production by the thermophilic bacterium *Thermotoga neapolitana*. *Int J Mol Sci* 16:12578–12600
- Pradhan N, Dipasquale L, d'Ippolito G, Fontana A, Panico A, Lens PNL, Pirozzi F, Esposito G (2016) Kinetic modeling of fermentative hydrogen production by *Thermotoga neapolitana*. *Int J Hydrog Energy* 41:4931–4940
- Pradhan N, Dipasquale L, d'Ippolito G, Panico A, Lens PNL, Esposito G, Fontana A (4359T) Hydrogen and lactic acid synthesis by the wild-type and a laboratory strain of the hyperthermophilic bacterium *Thermotoga neapolitana* DSMZ 4359T under capnophilic lactic fermentation conditions. *Int J Hydrog Energy* 42:16023–16030
- Samuelson P, Gunneriusson E, Nygren PA, Ståhl S (2002) Display of proteins on bacteria. *J Biotechnol* 96:129–154
- Serpe M, Forenza C, Adamo A, Russo N, Perugino G, Ciaramella M, Valenti A (2019) The DNA Alkylguanine DNA

- Alkyltransferase-2 (AGT-2) of *Caenorhabditis Elegans* is involved in meiosis and early development under physiological conditions. *Sci Rep* 9:6889
- Skorvaga M, Raven NDH, Margison GP (1998) Thermostable archaeal  $O^6$ -alkylguanine-DNA alkyltransferases. *Proc Natl Acad Sci USA* 95:6711–6715
- Sun G, Zhao L, Zhong R, Peng Y (2018) The specific role of  $O^6$ -methylguanine-DNA methyltransferase inhibitors in cancer chemotherapy. *Future Med Chem* 10:1971–2199
- Terns RM, Terns MP (2013) The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochem Soc Trans* 41:1416–1421
- 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
- Vettone A, Serpe M, Hidalgo A, Berenguer J, del Monaco G, Valenti A, Rossi M, Ciaramella M, Perugino G (2016) A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering. *Extremophiles* 20:1–13
- Visone V, Han W, Perugino G, del Monaco G, She Q, Rossi M, Valenti A, Ciaramella M (2017) In vivo and in vitro protein imaging in thermophilic archaea by exploiting a novel protein tag. *PLoS ONE* 12:e0185791
- Yang CG, Garcia K, He C (2009) Damage detection and base flipping in direct DNA alkylation repair. *ChemBioChem* 10:417–423
- Zhang J, Shi H, Xu L, Zhu X, Li X (2015) Site-directed mutagenesis of a hyperthermophilic endoglucanase Cel12B from *Thermotoga maritima* based on rational design. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0133824>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Research Article

# Optimization of an In Vitro Transcription/Translation System Based on *Sulfolobus solfataricus* Cell Lysate

Giada Lo Gullo,<sup>1</sup> Rosanna Mattosovich,<sup>2</sup> Giuseppe Perugino ,<sup>2</sup> Anna La Teana,<sup>3</sup> Paola Londei ,<sup>1</sup> and Dario Benelli <sup>1</sup>

<sup>1</sup>Department of Cellular Biotechnologies and Haematology, Sapienza University of Rome, Via Regina Elena 324, 00161 Rome, Italy

<sup>2</sup>Institute of Biosciences and BioResources, National Research Council of Italy, Via Pietro Castellino 111, 80131 Naples, Italy

<sup>3</sup>Department of Life and Environmental Science, Polytechnic University of Marche, Via Breccie Bianche, 60131 Ancona, Italy

Correspondence should be addressed to Dario Benelli; [benelli@bce.uniroma1.it](mailto:benelli@bce.uniroma1.it)

Received 27 July 2018; Accepted 5 November 2018; Published 11 February 2019

Academic Editor: Antonio Ventosa

Copyright © 2019 Giada Lo Gullo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A system is described which permits the efficient synthesis of proteins *in vitro* at high temperature. It is based on the use of an unfractionated cell lysate (S30) from *Sulfolobus solfataricus* previously well characterized in our laboratory for translation of pretranscribed mRNAs, and now adapted to perform coupled transcription and translation. The essential element in this expression system is a strong promoter derived from the *S. solfataricus* 16S/23S rRNA-encoding gene, from which specific mRNAs may be transcribed with high efficiency. The synthesis of two different proteins is reported, including the *S. solfataricus* DNA-alkylguanine-DNA-alkyl-transferase protein (SsOGT), which is shown to be successfully labeled with appropriate fluorescent substrates and visualized in cell extracts. The simplicity of the experimental procedure and specific activity of the proteins offer a number of possibilities for the study of structure-function relationships of proteins.

## 1. Introduction

Cell-free protein synthesis (CFPS) systems have been used initially to investigate certain fundamental aspects of cell biology, such as deciphering the structure of the genetic code or elucidating the basic features of transcriptional and translational control [1–3]. Later, CFPS systems turned out to be also powerful tools to produce high amounts of proteins for a wide range of applications ranging from pharmaceutical use to protein structure analysis [4, 5].

The simplest forms of these systems consist of whole cell lysates (S30 extracts) containing all the necessary elements for transcription, translation, protein folding, and energy metabolism. Typically, CFPS systems are programmed for expression of proteins using two different substrates: RNA templates for translation only or DNA templates for coupled transcription/translation [6, 7].

The advantages of CFPS systems over *in vivo* methods are manifold. One can dispense with all the procedures required to support cell viability and growth; moreover, handling

cellular extracts instead of whole cells facilitates the active monitoring, rapid sampling, and direct manipulation of the protein synthesis process. Last but not least, the simplicity and low cost of preparing cellular extracts make the system a preferential choice among the available tools for the synthesis of proteins of interest.

The most commonly used cell-free translation systems consist of *Escherichia coli* (ECE) extracts, rabbit reticulocytes (RRL), wheat germ (WGE), and insect cells (ICE), each of them with peculiar characteristics [8–10]. *E. coli* CFPS is the most convenient economically, since extract preparation is simple and inexpensive and the required proteins can be produced in high yields. However, CFPS derived from extracts of eukaryotic cells may be the best choice when the scope is the production of some types of complex proteins or when eukaryotic posttranslational modifications are required.

In our laboratory, we have developed a CFPS from the thermophilic archaeon *S. solfataricus*, which we have successfully used to decipher a number of aspects of archaeal

translation and high-temperature translation [11, 12]. However, our standard system uses only pretranscribed RNA templates, while CFPs from hyperthermophiles allowing a coupled transcription/translation based exclusively on endogenous components of the adopted system have not so far, to the best of our knowledge, been described.

Yet, to develop such a system is highly desirable for a number of reasons. First of all, it represents a powerful tool to expand our understanding of the molecular mechanisms governing coupled transcription-translation in archaea. Moreover, the expression of recombinant proteins in thermophilic conditions similar to the native ones could facilitate the identification of associated factors. Furthermore, although mesophilic hosts such as *Escherichia coli* have been used to produce thermostable proteins for biochemical and crystallographic characterization [13], many hyperthermophilic proteins correctly fold only under physiological conditions of high temperature or in the presence of their native posttranslational modifications [14, 15].

We report here the development of a coupled *in vitro* transcription/translation system for cell-free protein synthesis from the thermophilic archaeon *S. solfataricus*. The system works with a plasmid vector obtained by cloning the strong promoter derived from the *S. solfataricus* 16S/23S rRNA-encoding gene upstream of a previously well-characterized *Sulfolobus* gene [16]. A preliminary assessment of the various parameters and components that affect the rate and yield of protein synthesis was performed. With this system, we obtained the *in vitro* expression of two different proteins, one of which was also shown to be enzymatically active at the temperature of 70°C.

## 2. Materials and Methods

**2.1. Preparation of Cell Extracts and Total tRNA.** Cell lysates competent for *in vitro* translation were prepared according to a method described previously with slight precautions [17]. Briefly, about 2 g of frozen cells were ground by hand with a double amount of alumina powder and adding gradually about less than 2 volumes (relative to the weight of the cell pellet) of lysis buffer (20 mM Tris-HCl (pH 7.4), 10 mM Mg(OAc)<sub>2</sub>, 40 mM NH<sub>4</sub>Cl, and 1 mM DTT). The procedure was performed by placing the mortar on ice and working in a cold room for no more than 15 min. Cell debris and alumina were removed by spinning the mix twice at 30,000 ×g for 30 min and taking care to withdraw only about two-thirds of the supernatant. Aliquots of the cell lysate (0.05 ml) were stored at -80°C, and total protein concentration, determined by Bradford assay, was in the range of about 20–25 mg/ml accordingly. Unfractionated tRNA from *S. solfataricus* was prepared by performing a phenol extraction of the crude S100 fraction and precipitating the aqueous phase with 2.5 volumes of 95% ethanol. The RNA pellet was resuspended in 10 mM glycine (pH 9.0), and the solution was incubated for 2 h at 37°C to achieve alkaline deacylation of the tRNA therein contained. Lastly, the RNA was again precipitated and the resulting pellet was dissolved in an adequate volume of 10 mM Tris-HCl (pH 7.5).

**2.2. Gene Constructs and In Vitro Transcription.** We used the plasmid pBluescript-SK(+) as a starting point for our subsequent constructs. Two synthetic DNA oligomers of 48 nucleotides were designed on the sequence of a 16S/23S rRNA operon promoter described elsewhere [18] whose sequence is identically conserved in all *S. solfataricus* species: promoter rRNA SSO forward 5'-CGAAGTTAGAT TTATATGGGATTTTCAGAACAATATGTATAATGGGT AC-3' and promoter rRNA SSO reverse 5'-CCATTATAC ATATTGTTCTGAAATCCCATATAAATCTAACTTCG GTAC-3'. Both primers contained at their 5' ends a sequence corresponding to the protruding cohesive 5' end of the *Kpn* I restriction site, and both were phosphorylated in separate 25 μl reaction mixtures containing 70 mM Tris-HCl (pH 7.6), 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, 1 mM ATP, 4 μM DNA, and 10 units of T4 polynucleotide kinase (New England BioLabs). After incubation at 37°C for 1 h, the reaction mixtures were combined and the kinase was heat-inactivated at 70°C for 10 min. Annealing of the two oligomers was obtained by heating this mixture at 100°C for 4 min and slowly cooling down to 37°C. The integrity of the double-stranded 16S/23S rRNA promoter fragment DNA was checked by agarose gel electrophoresis. One pmol of the purified double-stranded fragment was incubated with 0.25 pmol of *Kpn* I digested pBS-SK(+) plasmid in the presence of 10 units of T4 DNA ligase (New England BioLabs) in 25 μl of 50 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 1 mM ATP, and 25 μg/ml bovine serum albumin for 20 h at 16°C. One-tenth of this reaction mixture was then used directly for transformation of *E. coli* top 10 competent cells. Transformants harbouring plasmid DNA were screened for the presence of the insert using a *Kpn* I restriction analysis of purified plasmid DNA. The clone harbouring the construct with the insert in the correct orientation was selected after DNA sequencing and termed pBS-rRNA<sub>p</sub>. Successively, a fragment of 393 bp containing the gene termed ORF 104 with its Shine-Dalgarno (SD) motif was amplified from the construct pBS800 [12] by PCR using the following primers: Prom-104 *Xho* I 5'-TTTTTTTATCTCGAGCCGGAATAGTTGAATT AACAATGAAGC-3' (underlined sequence corresponds to the *Xho* I site) and Prom-104 *Pst* I 5'-CATGGTATGCTGCA GTCATTTGCTTCACCTCTTTAATAAACTCC-3' (underlined sequence corresponds to the *Pst* I site). The fragment was inserted into the *Xho* I-*Pst* I digested plasmid pBS-rRNA<sub>p</sub>, yielding the construct termed pBS-rRNA<sub>p</sub>-104. To generate the construct termed pBS-rRNA<sub>p</sub>-*ogt*, we excised the fragment *Xho* I-*Pst* I from the previous plasmid and inserted a DNA fragment of 533 bp amplified from the construct pQE-*ogt* by PCR with the following primers: forward rRNA/SsOGT *Xho* I 5'-TTTTTCTCGAGTGAGGTGAAAT GTAATGAGAGGATCTCACCATCACC-3' (underlined sequence corresponds to the *Xho* I site) and reverse rRNA/SsOGT *Pst* I 5'-TTTTTCTGCAGTCATTCTGG TATTTTGACTCCC-3' (underlined sequence corresponds to the *Pst* I site). Also in this case, the plasmid was designed to have the SD motif 7 nucleotides upstream of the *ogt* start codon.

TABLE 1: Experimental conditions adopted for reactions with S30 *S. solfataricus*.

	<i>In vitro</i> transcription adopted from [16]	<i>In vitro</i> transcription under our conditions	Coupled <i>in vitro</i> transcription and translation	<i>In vitro</i> translation
KCl (mM)	—	10	10	10
Tris-HCl (mM)	50 (pH 8.0)	20 (pH 6.8)	20	20
Mg(OAc) <sub>2</sub> (mM)	25	20	20	20
ATP (mM)	2	2	1.5	1.8
CTP (mM)	1	1	1.5	—
GTP (mM)	1	1	1.5	0.9
UTP (mM)	0.6	0.5	1.5	—
( $\alpha$ - <sup>32</sup> P) UTP ( $\mu$ M)	100	100	—	—
EDTA (mM)	1	—	—	—
DTT (mM)	1	—	—	—
Total tRNA ( $\mu$ g)	—	—	3,3	3,3
S30 ( $\mu$ g)	100–150	100–150	100–150	100–150
T (°C)	60	70	70	70

**2.3. Analysis of Transcriptional Activity of *Sulfolobus solfataricus* Lysate by *In Vitro* Labelling with <sup>32</sup>P-UTP.** The transcriptional activity of the *S. solfataricus* cell-free extract was tested by <sup>32</sup>P-UTP incorporation in two different reaction conditions using an aliquot of the lysate corresponding to 100  $\mu$ g of total proteins. The first reaction protocol was adopted from a previous study [16]: the cell-free extract was incubated in a reaction volume of 50  $\mu$ l, in the presence of 50 mM Tris-HCl (pH 8.0), 25 mM MgCl<sub>2</sub>, 1 mM EDTA, 1 mM dithiothreitol, 2 mM ATP, 1 mM GTP, 1 mM CTP, 0.6 mM UTP, and 100  $\mu$ M ( $\alpha$ -<sup>32</sup>P) UTP (4 Ci/mmol) in a reaction volume of 50  $\mu$ l. The reaction was carried out at 60°C for 30 min. The second protocol was based on the *in vitro* translation experiments carried out in our laboratory [12, 17]: *S. solfataricus* cell-free extract was incubated in a reaction volume of 50  $\mu$ l, in the presence of 10 mM KCl, 20 mM Tris-HCl (pH 6.8), 20 mM Mg(OAc)<sub>2</sub>, 2 mM ATP, 1 mM CTP, 1 mM GTP, 0.5 mM UTP, and 100  $\mu$ M ( $\alpha$ -<sup>32</sup>P) UTP (4 Ci/mmol). The reaction, in this case, was carried out at 70°C for 30 min. At the end of both reactions, 20 U of DNase I were added and incubation was extended for 30 min at 37°C. DNase I was added to remove any trace of plasmidic DNA that could alter the results of the next qRT-PCR analysis. The products of the reactions were extracted by phenol pH 4.7 and precipitated with 2.5 volumes of 95% ethanol. The pellets were resuspended in an adequate volume of DEPC-treated water and divided into two aliquots. RNase A (20  $\mu$ g) was added to one of them and both aliquots were incubated at 37°C for 30 min. The newly synthesized RNA was separated by 8.5% of nondenaturing polyacrylamide gels and detected using both an Instant Imager apparatus (Packard) and autoradiography film (Kodak XAR-5).

**2.4. *In Vitro* Translation and Coupled *In Vitro* Transcription-Translation.** The transcription-translation activity was measured in a final volume of 25  $\mu$ l and contained 10 mM KCl, 20 mM Tris-HCl (pH 6.8), 20 mM Mg(OAc)<sub>2</sub>,

1.5 mM ATP, 1.5 mM CTP, 1.5 mM GTP, 1.5 mM UTP, 3.3  $\mu$ g of bulk *S. solfataricus* tRNA, 5  $\mu$ l of 20–25 mg/ml *S. solfataricus* S30 extract (preincubated for 10 min at 70°C), and 0.5  $\mu$ l of L-(<sup>35</sup>S)-methionine (S.A. 1175 Ci mmol<sup>-1</sup> at 11 mCi ml<sup>-1</sup>, PerkinElmer). After mixing all components, 4  $\mu$ g of the desired mRNA or different amounts of the various plasmids were added and the mixtures were incubated for the indicated time at 70°C. Whole cell lysates were programmed for *in vitro* translation with transcripts of *S. solfataricus* genes ORF 104 and SsOGT cloned in the pBS-SK(+) plasmid downstream of the T7 RNA polymerase promoter under conditions described in Table 1. Before transcription, the plasmids were linearized with *Pst* I. The experimental conditions were the same as described above except for the absence of CTP and UTP and the presence of ATP and GTP to the final concentration of 1.8 and 0.9 mM, respectively. The analysis of the translation products was performed by loading 15  $\mu$ l of the incubation mixture in 16% polyacrylamide/SDS gels; after the run, the gels were dried and autoradiographed.

**2.5. qPCR and RT-PCR SsOGT Labelling.** At the end of *in vitro* transcription or coupled *in vitro* transcription-translation, total RNA was purified from the reactions by phenol extraction at pH 4.7 and precipitated by adding of 2.5 volumes of 95% ethanol. The pellets were resuspended in an adequate volume of DEPC-treated water and treated with 2 U of DNase I, RNase-free (Thermo Fisher Scientific) in an appropriate buffer at 37°C for 45 min. The residual products were reextracted by phenol pH 4.7 and precipitated with 2.5 volumes of 95% ethanol. 0.5  $\mu$ g of total RNA was retrotranscribed for relative qRT-PCR analysis (SensiFAST™ cDNA Synthesis Kit, Bioline). qPCR was performed with the Applied Biosystem StepOne Real-Time PCR System (Thermo Fisher Scientific) using 1/20 of cDNA and 10  $\mu$ l of GoTaq® qPCR Master Mix (Promega) in a final volume of 20  $\mu$ l. Cycling parameters were 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 3 sec, and

annealing/extension at 60°C for 30 sec. The relative amount of each mRNA was calculated by the  $2^{-\Delta\Delta Ct}$  method and normalized to endogenous aIF6 mRNA. Primer sequences used for qPCR were as follows: forward pBS 5'-TGGTAACAG GATTAGCAGAG-3' and reverse pBS 5'-ACCAAATACTG TCCTTCTAGTG-3'; aIF6 forward 5'-ATAAGCGGTAA CGATAACGG-3' and aIF6 reverse 5'-AATCCCTAGA TTCTCCTCAG-3'.

By performing RT-qPCR as described above, we measured the absolute amount of RNA transcribed from the plasmid pBS-rRNA<sub>p</sub>-104 following incubation in the *in vitro* transcription-translation system. Specifically, we compared the Ct values obtained from these samples with a standard curve plotted with Ct values obtained from serial dilutions of 1 µg of *in vitro* transcribed RNA (pBS-rRNA<sub>p</sub>-104). For semiquantitative RT-PCR, total RNA was extracted from the mix reaction as described above. 2 µg of total RNA was retrotranscribed in a final volume of 25 µl with 200 U M-MLV reverse transcriptase in 20 µl of mixture reaction for 1 h at 42°C according to the instructions of the supplier (Promega). The reaction contained 1 µM of the following reverse primer: 5'-GGTTTCCCGACTGGAAAGCGG GCAG-3'. At the end of the reaction, the final volume of the mixture reaction was adjusted to 50 µl and one-tenth of the RT reaction was PCR amplified with Taq DNA polymerase (Promega) for 30 sec at 95°C, 30 sec at 60°C, and 45 sec at 74°C (25 cycles) with a final extension step for 7 min at 74°C. Reverse primers for PCR amplification were the same used in the RT reaction coupled with the following forward primer: 5'-CGAATTCCTGCAGCCCCGGGGATCC-3'. The products of the reactions were separated by agarose gel electrophoresis and detected by ethidium bromide staining.

Controls correspond to reactions performed on RNA purified from samples in the absence of the plasmid and from RT minus cDNA reactions.

**2.6. SsOGT *In Vitro* Labeling.** The activity of *in vitro*-expressed SsOGT was analysed incubating 8 µg of pBS-rRNA<sub>p</sub>-ogt plasmid or 200 ng of recombinant SsOGT OGT with 200 µg of *S. solfataricus* whole cell extract under the experimental conditions described above for coupled *in vitro* transcription/translation and in the presence of BG-FL substrate (2.5 µM). The mix reaction was incubated at 70°C for 60 min. Reactions were stopped by denaturation, and samples were subjected to SDS-PAGE, followed by fluorescence imaging analysis using a VersaDoc 4000™ system (Bio-Rad Laboratories Inc.) by applying as excitation/emission parameters a blue LED bandpass filter. For western blot analysis, proteins were transferred onto PVDF filters (Bio-Rad Laboratories Inc.) using the Trans-Blot® Turbo™ Blotting System (Bio-Rad Laboratories Inc.). The presence of SsOGT protein was revealed using polyclonal antibodies raised in rabbit against *S. solfataricus* OGT as primary antibodies, the goat anti-rabbit IgG-HRP (Pierce) as secondary antibody, and the Amersham Biosciences ECL Plus kit. Filters were incubated, washed, and developed according to the manufacturer's instructions. Chemiluminescent bands were revealed using a VersaDoc apparatus (Bio-Rad Laboratories Inc.).

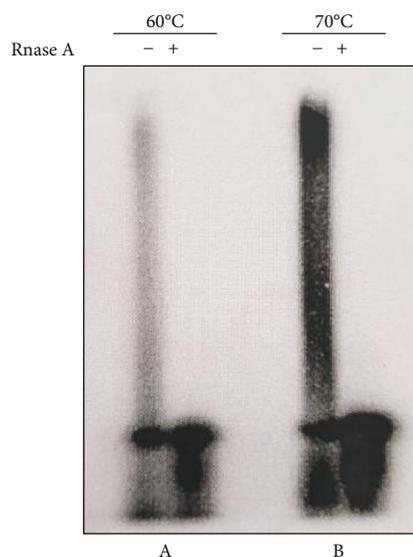


FIGURE 1: Transcriptional activity of *S. solfataricus* whole cell extracts. *In vitro* transcription reactions were performed using *S. solfataricus* S30 fractions with ( $\alpha$ -<sup>32</sup>P) UTP in different experimental conditions as described in Material and Methods and Table 1. Reaction A was incubated at 60°C while reaction B was incubated at 70°C. Total RNA was extracted from the reaction mixes, and an aliquot of the samples was treated with RNase A at 37°C for 30 min. The products of *in vitro* transcription were subjected to nondenaturing polyacrylamide gel electrophoresis, and those incorporating ( $\alpha$ -<sup>32</sup>P) UTP were visualized by autoradiography.

### 3. Results and Discussion

**3.1. Analysis of *In Vitro* Transcription in the S30 Fraction of *S. solfataricus*.** To prepare an S30 extract capable of efficient coupled transcription-translation, we performed preliminary experiments to verify whether the whole cell lysate of *S. solfataricus* prepared according to our described protocols [17] was competent for *in vitro* transcription. Specifically, we compared the transcriptional activity of our system with that of a previously described *Sulfolobus* *in vitro* transcription assay [16], testing the capacity of the S30 extract to incorporate  $\alpha$ -<sup>32</sup>P-UTP. Salt and temperature conditions of the reactions are summarized in Table 1 and described in detail in Materials and Methods. In both cases, we implemented the reactions with the nucleoside triphosphates at the final concentration of 1 mM each (except ATP to 2 mM) and the S30 fraction was prepared omitting DNase I treatment of lysate, unlike previously published protocols where DNase I was added to remove genomic DNA [17]. As shown in Figure 1, both S30 extracts showed the ability to recruit labeled uridine triphosphate supporting the idea that endogenous RNA polymerase was active. However, the extract prepared according to our protocol had a higher efficiency of uridine triphosphate incorporation.

Successively, based on a study characterizing the promoter for the single-copy 16S/23S rRNA gene cluster of the extremely thermophilic archaeobacterium *Sulfolobus* [18], we

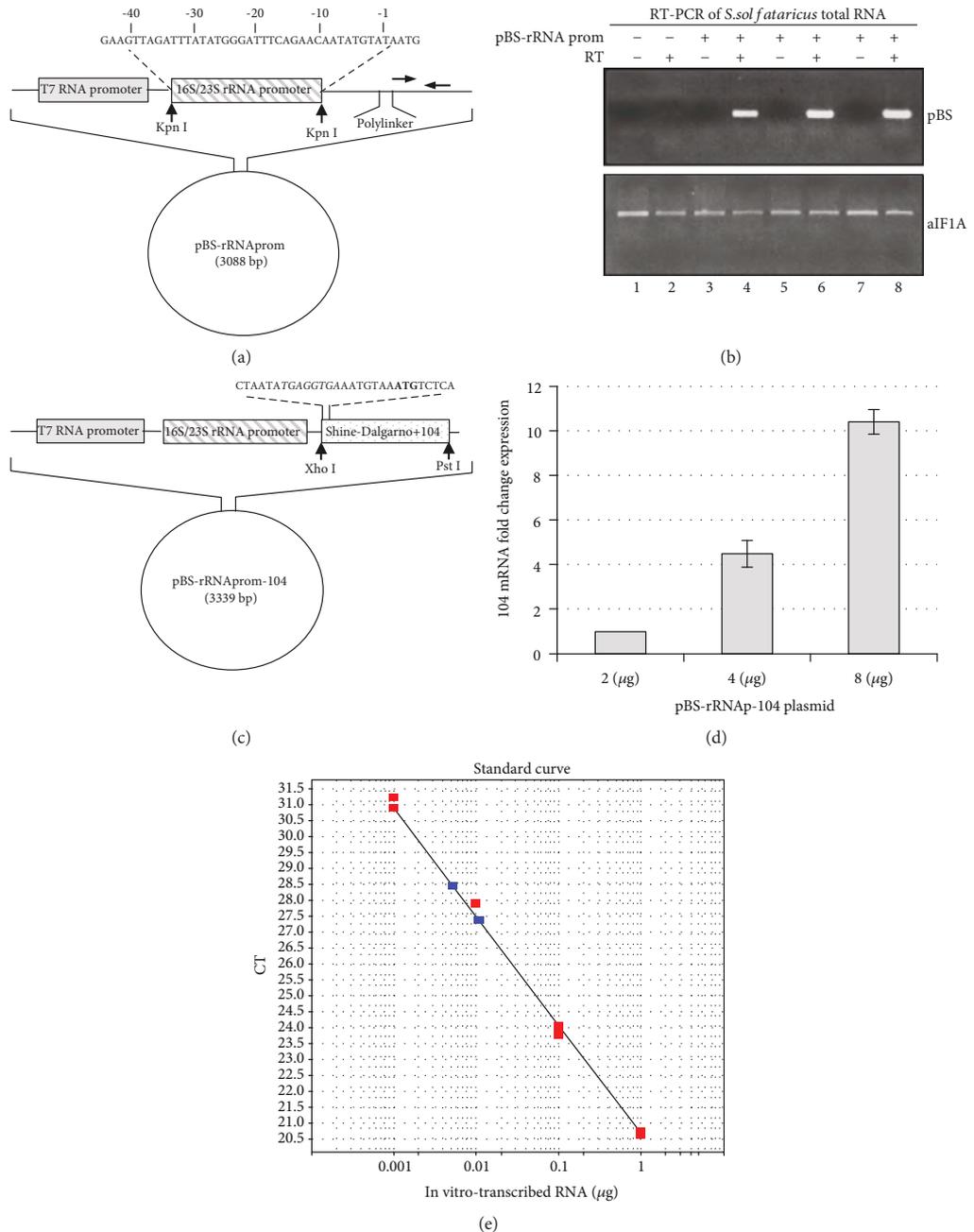


FIGURE 2: In vitro transcription of plasmids containing the 16S/23S rRNA promoter. (a) Schematic representation of the pBS-rRNA<sub>p</sub> construct. Horizontal arrows indicate the position of primers used for RT-PCR analysis. (b) RT-PCR on total RNA extracted from S30 of *S. solfataricus* previously incubated with 2, 4 and 8 µg of pBS-rRNA<sub>p</sub> plasmid and corresponding, respectively, to the lanes 4, 6 and 8 of the image. The product of the reaction is shown by the amplified fragment of 346 bp. Also shown is the RT-PCR of an mRNA encoding the translation factor aIF1A, used as an endogenous control to normalize the reactions. (c) Schematic representation of the pBS-rRNA<sub>p</sub>-104 plasmid. The SD motif is evidenced in italic, while the start codon is shown in bold. (d) Relative amount of RNA transcribed by the pBS-rRNA<sub>p</sub>-104 plasmid incubated into *S. solfataricus* S30 extract at 70°C for 1 h. (e) Absolute quantification of the pBS-rRNA<sub>p</sub>-104 transcript using the standard curve method. The absolute quantities of the standards were obtained measuring the concentration of T7 in vitro-transcribed pBS-rRNA<sub>p</sub>-104 RNA. Serial dilutions of the in vitro transcript were obtained and their Ct values (red dots) were compared to those unknown (blue dots) extrapolating the amount of copies expressed.

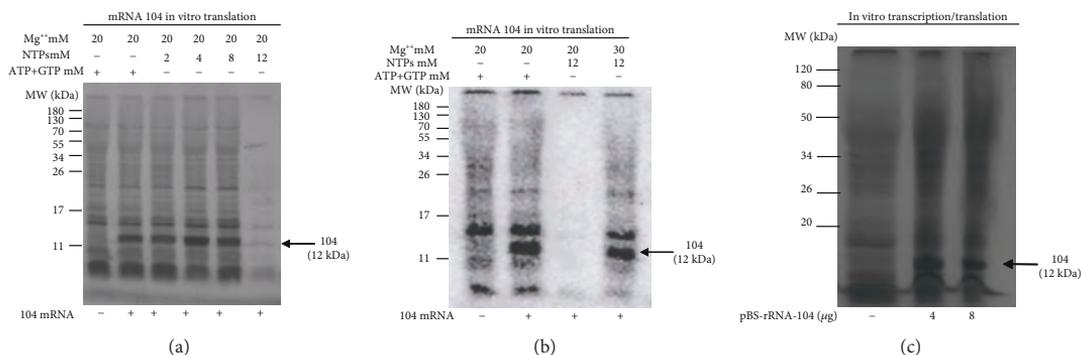


FIGURE 3: In vitro expression of ORF 104 under different experimental conditions. 4  $\mu\text{g}$  of in vitro-transcribed 104 mRNA were translated at different concentrations of NTPs (a) and  $\text{Mg}^{2+}$  (b) for 1 h in 25  $\mu\text{l}$  of reaction. (c) Different amounts of pBS-rRNA<sub>p</sub>-104 plasmid were incubated with *S. solfataricus* whole cell extract for 60 min at 70°C in a final volume of 25  $\mu\text{l}$ .

cloned this promoter into the pBS-SK(+) plasmid, as described in Materials and Methods. The construct contained the region of DNA upstream from the transcription start site of the 16S/23S rDNA gene spanning from -1 to -40 bp. The structure of the construct, termed pBS-rRNA<sub>p</sub>, is shown schematically in Figure 2(a). The plasmid was incubated with the S30 extract, and its transcription was analysed by RT-PCR using primers annealing to a specific region of the plasmid downstream of the cloned gene, thus excluding amplification of the endogenous target. The results showed an efficient transcription of the plasmid following incubation at 70°C (Figure 2(b)).

Starting from this construct, we cloned a previously well-characterized *Sulfolobus* gene encoding a putative ribosomal protein [12], under the transcriptional control of the 16S/23S rDNA promoter. The structure of this plasmid, termed pBS-rRNA<sub>p</sub>-104, is shown schematically in Figure 2(c); analysis by qPCR showed that it was also transcribed (Figure 2(d)). Finally, the pBS-rRNA<sub>p</sub>-104 construct was transcribed in vitro with T7 RNA polymerase, and known amounts of the corresponding purified RNA were used to draw a calibration curve, which was used to quantify the transcription reactions (Figure 2(e)). This analysis permitted us to assess the amount of *in vitro*-transcribed RNA to an order of magnitude corresponding to ng of RNA for  $\mu\text{g}$  of plasmid used, in 25  $\mu\text{l}$  of reaction.

#### 4. Optimization of *In Vitro* Translation Conditions with respect to NTPs and $\text{Mg}^{++}$ Ions

Next, we investigated whether the conditions adopted for *in vitro* transcription with the *S. solfataricus* S30 extract could affect its translational activity. Specifically, we sought to define an optimal concentration of NTPs since it is well known that free nucleotides chelate a proportional number of  $\text{Mg}^{++}$  ions, whose presence in a well-defined range of concentration is essential for translation [19]. For this purpose, we incubated the S30 extract with pretranscribed 104 mRNA in the absence or presence of different concentrations of

NTPs and determined its translational efficiency. Indeed, increased levels of NTP in the mix reactions were detrimental for *in vitro* translation (Figure 3(a)). However, this could be in part compensated by increasing the concentration of  $\text{Mg}^{++}$  ions as shown in Figure 3(b). On the other hand, the absence of NTPs in the mix reaction completely inhibited the activity of the system, since exogenous ATP and GTP are required as an energy source (data not shown). Overall, based on the results of Figures 3(a) and 3(b), we chose to strike a balance between NTP and  $\text{Mg}^{++}$ , setting them at the final concentration of 6 and 20 mM, respectively.

**4.1. Transcription and Translation-Coupled Protein Synthesis.** We then proceeded to verify whether the previously established experimental conditions allowed coupled transcription and translation. This question was addressed by incubating different amounts of the pBS-rRNA<sub>p</sub>-104 plasmid with the lysate at 70°C for 1 h under the conditions summarized in Table 1. As said before, the transcription of this construct from a strong rRNA promoter was expected to yield an mRNA encoding a ribosomal protein (ORF 104). The predicted mRNA was endowed with a 5'-UTR containing SD motif 7 nucleotides upstream from the AUG start codon of ORF 104. As shown in Figure 3(c), the reaction yielded a main protein band of about 12 kDa, corresponding to the expected size of the ORF 104.

To extend the above results to other *S. solfataricus* genes, we subcloned the O<sup>6</sup>-DNA-alkyl-guanine-DNA-alkyl-transferase gene (SsOGT) from the pQE-*ogt* construct, previously characterized by Perugino et al. [20]. The product of this gene is a ubiquitous protein of about 17 kDa, evolutionary involved in the direct repair of DNA lesions caused by the alkylating agents. SsOGT is a peculiar protein for its suicidal catalytic reaction: the protein irreversibly transfers the alkyl group from the DNA to a catalytic cysteine in its active site. The use of fluorescent derivatives of a strong inhibitor, the O<sup>6</sup>-benzyl-guanine (O<sup>6</sup>-BG), leads to an irreversible fluoresceinated form of this protein. This thermophilic variant of the so-called SNAP-tag<sup>TM</sup> [21] represents an alternative to the classical GFP-based systems and is eligible for our choice.

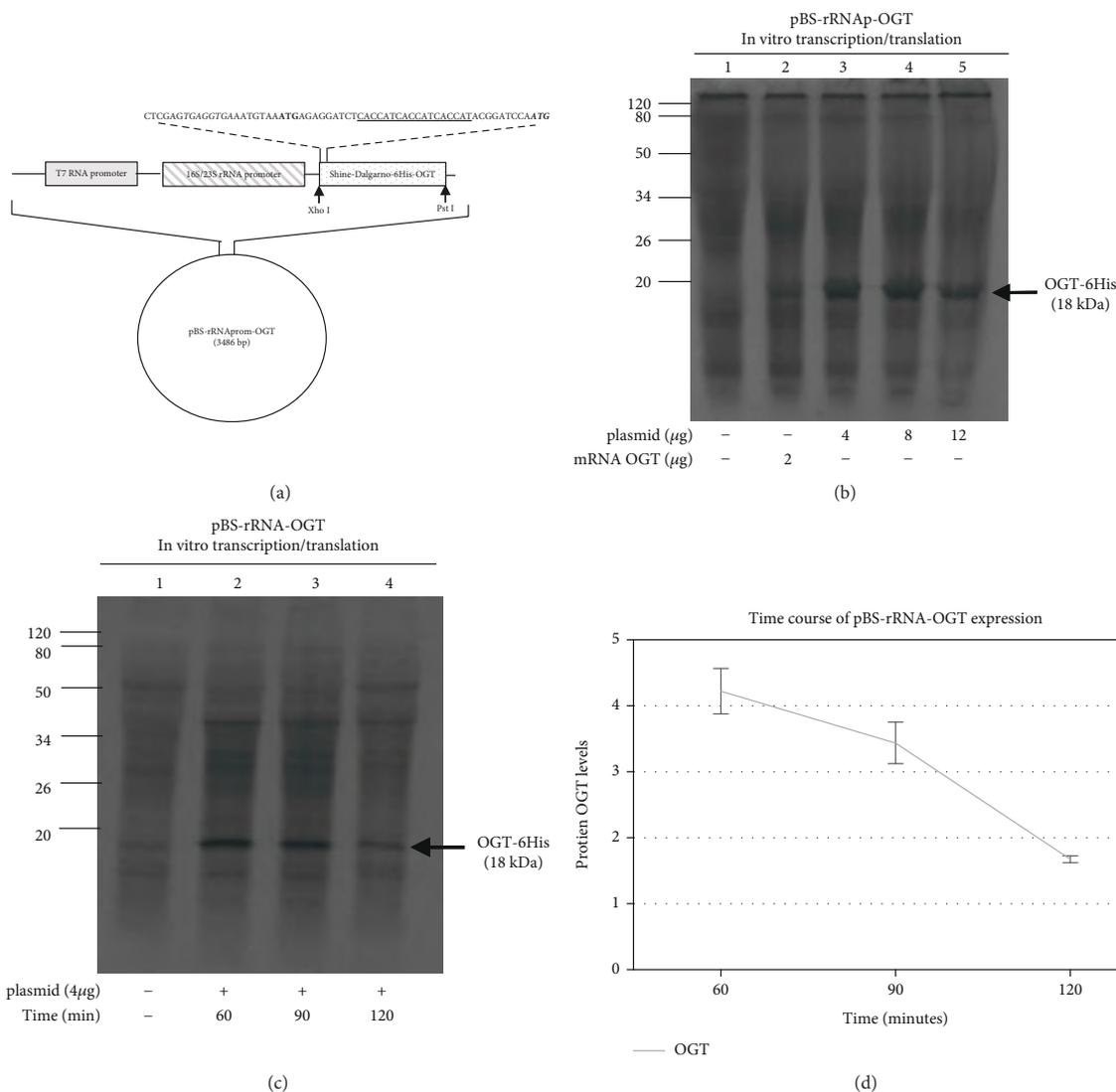


FIGURE 4: In vitro expression of OGT. (a) Schematic representation of the pBS-rRNA<sub>p</sub>-*ogt* plasmid. It was designed by introducing a DNA fragment of 522 bp containing the *ogt* gene into the *Xho* I-*Pst* I sites replacing ORF 104. The coding region starts with an AUG codon (bold letters) preceding a DNA region coding for six histidines (underlined letters) placed to the amino terminal region of the OGT protein (bold and italic letters). The DNA insert contains an SD motif (italic letters) retained from the ORF 104 and located 7 nucleotides upstream from the coding region. (b) Increased amounts of the pBS-rRNA<sub>p</sub>-*ogt* plasmid were incubated with *S. solfataricus* whole cell extract for 60 min at 70°C in a final volume of 25 μl, and the products of expression were resolved by 16% denaturing polyacrylamide gel electrophoresis. (c) Time course of OGT expression: 4 μg of pBS-rRNA<sub>p</sub>-*ogt* plasmid were incubated with *S. solfataricus* whole cell extract at 70°C and equal aliquots of the reaction were withdrawn from the mixture at the indicated times. (d) A graph is plotted with the values of the band intensity corresponding to the OGT protein shown in (c) and quantified using ImageJ software (NIH). The values represent the average of three independent experiments. All error bars indicate SD.

The construct was obtained by substituting the gene 104 from the construct pBS-rRNA-104 with the *ogt* gene, as described in Materials and Methods. The structure of the construct termed pBS-rRNA<sub>p</sub>-*ogt* is shown schematically in Figure 4(a).

Specifically, the strong SD motif 7 nucleotides upstream from the AUG start codon were retained, and 6 His-coding triplets were placed upstream of the *ogt* open reading frame.

As the results in Figure 4(b) show, the gene was expressed producing a main protein band of about 18 kDa, corresponding to the expected size of the ORF SsOGT-6His. As a positive control, we employed an *ogt* mRNA transcribed *in vitro* from the T7 promoter (lane 2), which, as expected, was translated less efficiently than the mRNA directly transcribed in the reaction mix. This is possibly due to the different 5'-UTR of the two mRNAs, but it is also conceivable that

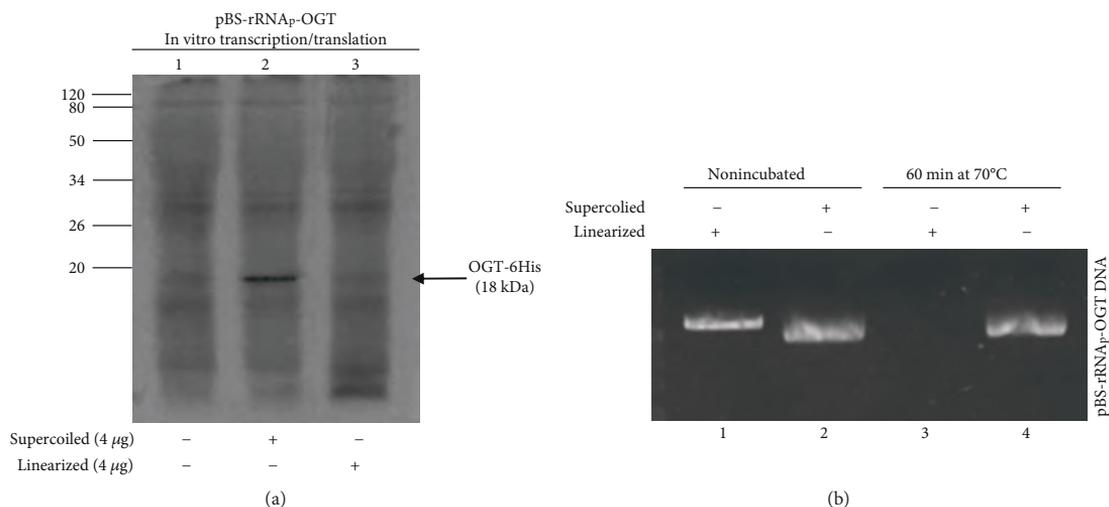


FIGURE 5: In vitro expression of *ogt* from linearized plasmid. (a) Supercoiled and linear pBS-rRNA<sub>p</sub>-*ogt* plasmids were incubated with *S. solfataricus* whole cell extract for 60 min at 70°C with <sup>35</sup>S-Met in a final volume of 25 μl, and the products of expression were resolved by 16% denaturing polyacrylamide gel electrophoresis. (b) Survival of supercoiled and linear pBS-rRNA<sub>p</sub>-*ogt* plasmid after incubation in the S30-coupled system. The constructs were incubated for 60 min at 70°C under standard conditions and then analysed on a 1% agarose gel. Lane 1, nonincubated linear pBS-rRNA<sub>p</sub>-*ogt* DNA; lane 2, nonincubated supercoiled pBS-rRNA<sub>p</sub>-*ogt* DNA; lane 3, linear pBS-rRNA<sub>p</sub>-*ogt* DNA incubated in an S30 mixture; and lane 4, supercoiled pBS-rRNA<sub>p</sub>-*ogt* DNA incubated in an S30 mixture.

when translation takes place at the same time as transcription the mRNA is stabilized and the ribosomes may bind more easily to the translation start sites.

To gain insight into other factors influencing the efficiency of SsOGT protein expression, we analysed the time course of the reaction with a fixed amount of the same construct. The highest expression level of the protein was observed after 60 min incubation, while at longer times (90 and 120 min) the efficiency decreased (Figures 4(c) and 4(d)), as observed in other *in vitro* expression systems [22]. This effect is probably due to the shortage of low molecular weight substrates (ATP, GTP, and amino acids) that are continuously used by the system with consequent drop of the reaction.

Furthermore, we tested whether the linearization of the construct could produce a transcriptional runoff at the end of the gene with a consequent increase of the product of our interest. This was not the case, however. Samples incubated with the linearized plasmid failed to yield a band corresponding to the expected size of the ORF OGT-6His (Figure 5(a)). Further analysis revealed that this was due to degradation of the linearized plasmid in the reaction mix (Figure 5(b)) similar to results obtained by other authors with different cell-free coupled transcription-translation systems [23].

**4.2. Characterization of SsOGT Activity.** To test whether the *in vitro*-produced SsOGT was functionally active, we incubated the construct pBS-rRNA<sub>p</sub>-*ogt* with the lysate at 70°C for 1 h in the presence of a fluorescein derivative of the O<sup>6</sup>-BG (SNAP-Vista Green™, New England BioLabs). As mentioned above, SsOGT catalyzes the formation of a covalent bond between the benzyl group of BG and a specific

cysteine residue in its active site; therefore, the successful completion of the reaction renders the protein fluorescent [21]. Indeed, we observed a fluorescent band corresponding to the expected size of the SsOGT in the reaction conditions adopted (Figure 6), demonstrating the active state of the expressed protein. The levels of *in vitro*-expressed SsOGT were assessed by comparing its fluorescence with that obtained with known amounts of recombinant protein. The outcome of the experiment also permitted excluding the possibility that *in vitro*-produced SsOGT was degraded after its translation and upon the irreversible transfer of the fluoresceinated-benzyl group to the active site, as previously demonstrated [24, 25]. In effect, incubation for 60 min at 70°C of the recombinant SsOGT in the *S. solfataricus* lysates in the presence of the SNAP-Vista Green™ did not affect the activity nor the fluorescent signal obtained (Figure 6, lane 3).

This analysis allowed us to estimate the amount of *in vitro*-translated SsOGT to an order of magnitude corresponding to about 10–20 ng of protein produced for μg of plasmid used, in 25 μl of reaction.

## 5. Discussion

The present study reports the development of a transcription/translation system for the synthesis of proteins at high temperature (70°C), based on an S30 extract from the thermophilic crenarchaeon *S. solfataricus*. The system makes use of an engineered classical pBS-SK plasmid, where efficient transcription is driven by a strong promoter, corresponding to the DNA region upstream from the 16S/23S rDNA gene, while translation is stimulated by the presence of a strong SD motif ahead of the start codon of the chosen gene. The

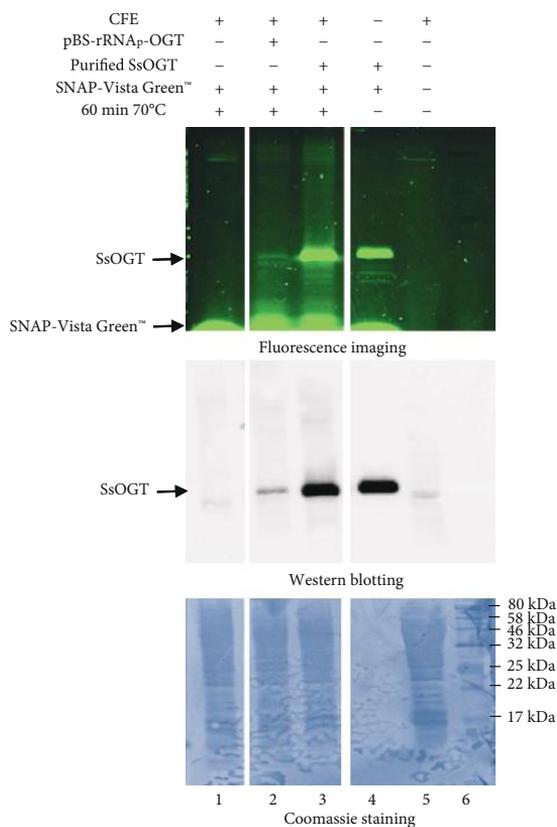


FIGURE 6: SsOGT labeling. SDS-PAGE of *in vitro*-expressed pBS-rRNA<sub>p</sub>-*ogt* plasmid and purified SsOGT protein both incubated with the BG-FL substrate (5 μM) for 60 min at 70°C. The gel was exposed for fluorescence imaging analysis, blotted, and stained with Coomassie blue. The filter was probed with the anti-OGT antibody (middle panel). Lane 1 contains 100 μg of *S. solfataricus* S30 fraction in the presence of the BG-FL substrate; lane 2 contains 8 μg of pBS-rRNA<sub>p</sub>-*ogt* plasmid in 100 μg of *S. solfataricus* S30 fraction and BG-FL substrate; lane 3 contains 200 ng of purified OGT protein with 100 μg of *S. solfataricus* S30 fraction and BG-FL substrate; lane 4 contains 200 ng of purified OGT protein with BG-FL substrate; lane 5 contains 100 μg of *S. solfataricus* S30 fraction; and lane 6 corresponds to the protein marker.

reaction works at the optimal temperature of 70°C, and maximal protein synthesis is achieved after 1 h of incubation.

We tested the system with two different genes, one encoding a ribosomal protein and another encoding SsOGT, an enzyme, whose activity was determined by using a fluorescent probe, as described above. The former gene had already shown to be efficiently translated *in vitro* from a pretranscribed mRNA [12], and served as a starting point to tune the system. Transcription/translation of the *ogt*-encoding gene allowed us to show that the protein product was active, thereby demonstrating that it was correctly folded/modified in the *in vitro* reaction. Moreover, the possibility to use fluorescent substrates of this enzyme is a clear advantage

for the quantification of the gene product, making this system flexible.

An important novelty of our system with respect to previous attempts described in the literature is that it requires only endogenous components present in the cell lysate. Indeed, the only described system for protein synthesis coupled with high-temperature translation makes use of a *Thermococcus kodakaraensis* lysate, but it requires an added thermostable T7 RNA polymerase to work [26]. Our assay is therefore an economically convenient choice, since extract preparation is simple and inexpensive.

While the present work describes a promising new technology mainly for the gene expression analysis, it is not yet usable as such for the *in vitro* scale-up production of recombinant proteins. To achieve this, further experiments and improvements are needed. For instance, one may envisage the division of the reaction into two compartments, one containing the modified extract and one containing a feeding solution that includes substrates such as amino acids, ATP, and GTP, and that is renewed by continuous flow, permitting substrate replenishment and byproduct removal.

Moreover, it should be observed that extant-coupled CFPS utilize DNA in three forms: linear PCR product, linearized plasmid, and circular plasmid. The use of linear PCR products has the distinct advantage of simplicity, since it eliminates the need for time-consuming cloning steps. However, circular DNA plasmids have typically been preferred to linearized plasmids or PCR products, due to the greater susceptibility of linear DNAs to nucleolytic cleavage. Indeed, in our case, samples incubated with the linearized plasmid failed to yield the expected protein product due to degradation of the linearized plasmid in the reaction mix. The removal of nucleases, and/or the utilization of overhang extensions to cyclize PCR products, could be adopted in the future for the optimization of the system.

In conclusion, we believe that the system described here has very good potential for use in fields such as protein display technologies, interactome analysis, and understanding of the molecular mechanisms governing coupled transcription-translation in archaea.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by grants to PL from the Istituto Pasteur-Fondazione Cenci Bolognetti project “Detecting and characterizing specialized ribosomes translating specific classes of mRNAs in Archaea” and by funds from “Sapienza” University of Rome to DB for the 2016 project “The

translational control of p53 executed by eIF5A”, protocol number: RP116154B8B63CB0.

## References

- [1] M. W. Nirenberg and J. H. Matthaei, “The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 47, no. 10, pp. 1588–1602, 1961.
- [2] M. R. Lamborg and P. C. Zamecnik, “Amino acid incorporation into protein by extracts of *E. coli*,” *Biochimica et Biophysica Acta*, vol. 42, no. 42, pp. 206–211, 1960.
- [3] G. Zubay, “In vitro synthesis of protein in microbial systems,” *Annual Review of Genetics*, vol. 7, no. 1, pp. 267–287, 1973.
- [4] Y. Shimizu, A. Inoue, Y. Tomari et al., “Cell-free translation reconstituted with purified components,” *Nature Biotechnology*, vol. 19, no. 8, pp. 751–755, 2001.
- [5] E. D. Carlson, R. Gan, C. E. Hodgman, and M. C. Jewett, “Cell-free protein synthesis: applications come of age,” *Biotechnology Advances*, vol. 30, no. 5, pp. 1185–1194, 2012.
- [6] R. J. Jackson and T. Hunt, “Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA,” *Methods in Enzymology*, vol. 96, pp. 50–74, 1983.
- [7] J. Josephsen and W. Gaastra, “DNA directed in vitro protein synthesis with *Escherichia coli* S-30 extracts,” *DNA Methods in Molecular Biology*, vol. 2, pp. 131–135, 1984.
- [8] H. R. B. Pelham and R. J. Jackson, “An efficient mRNA-dependent translation system from reticulocyte lysates,” *European Journal of Biochemistry*, vol. 67, no. 1, pp. 247–256, 1976.
- [9] T. Sawasaki, R. Morishita, M. D. Gouda, and Y. Endo, “Methods for high-throughput materialization of genetic information based on wheat germ cell-free expression system,” *Methods in Molecular Biology*, vol. 375, pp. 95–106, 2007.
- [10] T. Ezure, T. Suzuki, S. Higashide et al., “Cell-free protein synthesis system prepared from insect cells by freeze-thawing,” *Biotechnology Progress*, vol. 22, no. 6, pp. 1570–1577, 2006.
- [11] D. Ruggero, R. Creti, and P. Londei, “In vitro translation of archaeal natural mRNAs at high temperature,” *FEMS Microbiology Letters*, vol. 107, no. 1, pp. 89–94, 1993.
- [12] I. Condo, A. Ciammaruconi, D. Benelli, D. Ruggero, and P. Londei, “Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*,” *Molecular Microbiology*, vol. 34, no. 2, pp. 377–384, 1999.
- [13] M. Watanabe, K. Miyazono, M. Tanokura, T. Sawasaki, Y. Endo, and I. Kobayashi, “Cell-free protein synthesis for structure determination by X-ray crystallography,” *Methods in Molecular Biology*, vol. 607, pp. 149–160, 2010.
- [14] D. E. Ward, S. W. M. Kengen, J. van der Oost, and W. M. de Vos, “Purification and characterization of the alanine aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its role in alanine production,” *Journal of Bacteriology*, vol. 182, no. 9, pp. 2559–2566, 2000.
- [15] G. Andreotti, M. V. Cubellis, G. Nitti et al., “An extremely thermostable aromatic aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus*,” *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, vol. 1247, no. 1, pp. 90–96, 1995.
- [16] W. D. Reiter, U. Hüpdepohl, and W. Zillig, “Mutational analysis of an archaeobacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection in vitro,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 24, pp. 9509–9513, 1990.
- [17] D. Benelli and P. Londei, “In vitro studies of archaeal translational initiation,” *Methods in Enzymology*, vol. 430, pp. 79–109, 2007.
- [18] S. A. Qureshi, P. Baumann, T. Rowlands, B. Khoo, and S. P. Jackson, “Cloning and functional analysis of the TATA binding protein from *Sulfolobus shibatae*,” *Nucleic Acids Research*, vol. 23, no. 10, pp. 1775–1781, 1995.
- [19] K. H. Nierhaus, “Mg<sup>2+</sup>, K<sup>+</sup>, and the ribosome,” *Journal of Bacteriology*, vol. 196, no. 22, pp. 3817–3819, 2014.
- [20] G. Perugino, A. Vettone, G. Illiano et al., “Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage,” *The Journal of Biological Chemistry*, vol. 287, no. 6, pp. 4222–4231, 2012.
- [21] V. Visone, W. Han, G. Perugino et al., “In vivo and in vitro protein imaging in thermophilic archaea by exploiting a novel protein tag,” *PLoS One*, vol. 12, no. 10, article e0185791, 2017.
- [22] M. B. Iskakova, W. Szaflarski, M. Dreyfus, J. Remme, and K. H. Nierhaus, “Troubleshooting coupled in vitro transcription-translation system derived from *Escherichia coli* cells: synthesis of high-yield fully active proteins,” *Nucleic Acids Research*, vol. 34, no. 19, article e135, 2006.
- [23] H. L. Yang, L. Ivashkiv, H. Z. Chen, G. Zubay, and M. Cashel, “Cell-free coupled transcription-translation system for investigation of linear DNA segments,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 12, pp. 7029–7033, 1980.
- [24] G. Perugino, R. Miggiano, M. Serpe et al., “Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein,” *Nucleic Acids Research*, vol. 43, no. 18, pp. 8801–8816, 2015.
- [25] A. Vettone, M. Serpe, A. Hidalgo et al., “A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering,” *Extremophiles*, vol. 20, no. 1, pp. 1–13, 2016.
- [26] T. Endoh, T. Kanai, Y. T. Sato et al., “Cell-free protein synthesis at high temperatures using the lysate of a hyperthermophile,” *Journal of Biotechnology*, vol. 126, no. 2, pp. 186–195, 2006.

## The SNAP-tag technology revised: an effective chemo-enzymatic approach by using a universal azide-based substrate

Giuseppe Merlo<sup>a\*</sup>, Diego Caprioglio<sup>b\*</sup>, Michele Cillo<sup>c\*</sup>, Anna Valenti<sup>a</sup>, Rosanna Mattosovich<sup>a</sup>, Castrese Morrone<sup>b</sup>, Ildardo Massarotti<sup>b,d</sup>, Franca Rossi<sup>b</sup>, Riccardo Miggianno<sup>b,d</sup>, Antonio Leonardi<sup>c</sup>, Alberto Minassi<sup>b</sup> and Giuseppe Peruginò<sup>a</sup>

<sup>a</sup>Institute of Biosciences and BioResources, National Research Council of Italy, Naples, Italy; <sup>b</sup>Department of Pharmaceutical Sciences, University of Piemonte Orientale, Novara, Italy; <sup>c</sup>Department of Molecular Medicine and Medical Biotechnology, University of Naples “Federico II”, Naples, Italy; <sup>d</sup>IXTAL srl, Novara, Italy

### ABSTRACT

SNAP-tag<sup>®</sup> is a powerful technology for the labelling of protein/enzymes by using benzyl-guanine (BG) derivatives as substrates. Although commercially available or ad hoc produced, their synthesis and purification are necessary, increasing time and costs. To address this limitation, here we suggest a revision of this methodology, by performing a *chemo-enzymatic approach*, by using a BG-substrate containing an azide group appropriately distanced by a spacer from the benzyl ring. The SNAP-tag<sup>®</sup> and its relative thermostable version (*SsOGT-H<sup>2</sup>*) proved to be very active on this substrate. The stability of these tags upon enzymatic reaction makes possible the exposition to the solvent of the azide-moiety linked to the catalytic cysteine, compatible for the subsequent conjugation with DBCO-derivatives by azide-alkyne Huisgen cycloaddition. Our studies propose a strengthening and an improvement in terms of biotechnological applications for this self-labelling *protein-tag*.

### ARTICLE HISTORY

Received 16 September 2020  
Revised 5 October 2020  
Accepted 14 October 2020

### KEYWORDS

*Protein-tag*; protein labelling; enzymatic reaction; click chemistry; biotechnology

## 1. Introduction

The advent of the self-labelling *protein-tags* (SLPs) has led to a huge push in modern biotechnology, especially in the field of cell biology, where auto-fluorescent proteins (AFPs) for a long time dominated for their versatility in the localisation experiments of proteins, organelles, and membranes<sup>1</sup>. But the use of SLPs clearly goes beyond: they catalyse the covalent, highly specific and irreversible attachment of a part of their synthetic ligands upon reaction. This offers the opportunity to label them by conjugation to those ligands of an infinite number of chemical groups, such as fluorescent dyes, affinity molecules, or solid surfaces, expanding the application fields<sup>2</sup>. Among SLPs, of particular note are the *alotag*<sup>®</sup>, the *SpyTag*<sup>3</sup> the SNAP- and the CLIP-*tag*<sup>®</sup>. The Promega *alotag*<sup>®</sup> is a halo-alkane dehalogenase with a genetically modified active site, which reacts irreversibly with primary alkyl-halides<sup>4,5</sup>.

SNAP-*tag*<sup>®</sup> from New England Biolabs (NEB) is the engineered variant of the natural suicide human O<sup>6</sup>-methylguanine NA-methyltransferase protein (hMGMT). Alkylated DNA-alkyltransferases (AGTs, MGMTs or OGTs, E.C. 2.1.1.63) are ubiquitous and conserved proteins involved in the repair of the DNA alkylation damage, in particular, they remove alkyl adducts at the level of O<sup>6</sup>-position on guanine base<sup>6,7</sup>. The peculiar single-step mechanism are called “suicide enzymes,” in which the alkylated base is directly repaired by the irreversible transfer of the alkyl group

from the damaged guanine to the catalytic cysteine in the protein active site<sup>8</sup>. The protein is permanently inactivated upon the trans-alkylation reaction and susceptible to *in vivo* degradation by the proteasome.

In 2003, the group of Kai Johnsson developed a new strategy to exploit the hMGMT suicidal reaction in biotechnology, adopting a directed-evolution approach to engineer a variant to be used as an innovative *protein-tag*, that is, the SNAP-*tag*<sup>®</sup>. The rationale behind the SNAP-*tag* technology is the low substrate specificity of some AGT proteins, being able to efficiently recognise also the O<sup>6</sup>-benzyl-guanine (BG) nucleobase<sup>9</sup>. Likely, the reaction of these enzymes with BG-derivatives could happen: upon the irreversible transfer to the catalytic cysteine, they indeed demonstrated the specific labelling of the hMGMT with molecules, as fluorophore previously conjugated to the 4-position of the BG benzyl ring. Because of the small dimension of this protein, it was mutated to abolish any DNA binding activity and utilised as *protein tag* for the indirect labelling of proteins of interest genetically fused to it (Figure 1)<sup>9–13</sup>. Later, the same group further engineered the SNAP-*tag*<sup>®</sup> to obtain the CLIP-*tag*<sup>®</sup>, which specifically reacts with O<sup>2</sup>-benzyl-cytosine derivatives, expanding that technology for *in vivo* and *in vitro* multi-protein labelling<sup>14</sup>.

Apart from cell biology and fluorescence imaging, hundreds of papers are present in the literature showing many applications of SNAP-*tag*<sup>®</sup> in several fields, among which RNA-editing<sup>15</sup>, the

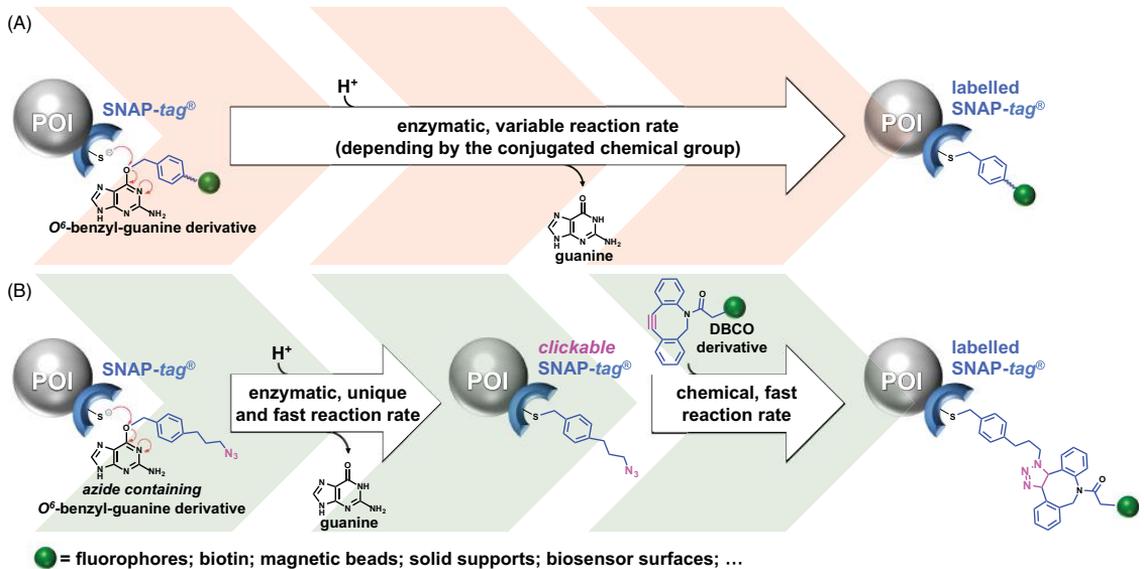
\*CONTACT Giuseppe Peruginò  giuseppe.peruginò@ibbr.cnr.it  Institute of Biosciences and BioResources, National Research Council of Italy, Via Pietro Castellino 1, Naples 80131, Italy; Alberto Minassi  alberto.minassi@uniupo.it  Department of Pharmaceutical Sciences, University of Piemonte Orientale, Largo Donega 3, Novara 28100, Italy

<sup>a</sup>These Authors contributed equally and joined as First Author.

 Supplemental data for this article can be accessed [here](#).

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Single-step reaction vs chemo-enzymatic approach. (A) The SNAP-tag<sup>®</sup> technology is based on BG-derivatives singularly synthesised and purified, and including that the conjugated chemical group (green sphere) could affect the enzymatic reaction rate. (B) The SNAP-tag<sup>®</sup> technology revised uses a unique and universal azide BG-derivative, converting SNAP-tag<sup>®</sup> in a clickable form, prone to perform a fast and efficient cycloaddition with DBCO-based chemical groups. POI, protein of interest genetically fused to the SNAP-tag<sup>®</sup>.

development of SNAP-based sensors for small molecules<sup>16–18</sup> and nucleic acids<sup>19,20</sup>, and protein-DNA complexes in “DNA Origami” structures<sup>21</sup>.

Following the same approach, Perugino and co-workers expanded this technology to extremophilic organisms and to all the applications which require harsh reaction conditions, not fully suitable for the employing of the mesophilic SNAP-tag<sup>®</sup>. To this aim, they developed a “thermo-SNAP-tag” by the production of a variant of the OGT from *Saccharolobus solfataricus* (previously *Jlfolobus solfataricus*, *SsOGT-H<sup>5</sup>*, hereinafter *H<sup>5</sup>*), an enzyme which revealed extremely resistant to high temperature, high ionic strength, proteases attack, and, in general, to common physical and chemical denaturants<sup>22,23</sup>. The intrinsic stability of *H<sup>5</sup>* made it incompatible with expression and utilisation *in vivo* as protein-tag in thermophilic organisms, as *Thermus thermophilus*<sup>24</sup> and *Sulfolobus islandicus*<sup>25</sup> as well as in an *in vitro* expression system using *Jlfolobus* lysates<sup>26</sup>. Recently, *H<sup>5</sup>* became a part of the new ASL<sup>tag</sup> system<sup>27</sup>, which was particularly useful for the *in vivo* immobilisation and contemporary labelling of proteins and enzymes of interest, stabilising them without any purification procedures needed<sup>28</sup>.

SNAP-tag<sup>®</sup> technology is essentially based on BG-substrates: though many of them are commercially available, the possibility of conjugation of infinite desired molecules to the 4-position on the guanine leads to the synthesis of *ad hoc* substrates. This is generally possible through the crosslinking reaction of the so-called “BG-building block” (such as the amine-reactive BG-NH<sub>2</sub>) with NHS-ester derivative compounds. The main disadvantage is the need to purify the final compounds before the reaction with the enzyme, increasing the times and costs of the experiments (Figure 1(A)). Furthermore, the presence of chemical groups conjugated to the benzyl moiety of the BG could affect the reaction efficiency of the SNAP-tag<sup>®</sup><sup>29–33</sup>, sometimes making this enzyme not fully applicable to particular requests.

In this work, we analysed and confirmed the catalytic dependence of SNAP-tag<sup>®</sup> and *H<sup>5</sup>* by several substrates having different

chemical groups conjugated to the O<sup>6</sup>-position of the guanine. To overcome these limitations, in the current study we suggest a further improvement of this technology with the application of *chemo-enzymatic approach*, by using a unique and universal azide decorated BG-derivative, to obtain the specific labelling of the *tc* (clickable-SNAP), that can be easily coupled with a potential infinite number of commercially available di-benzo-cyclo-oct (DBCO)-based molecules, through the copper-free azide-alkyne Huisgen cycloaddition (Figure 1(B)). This approach could mainly offer the advantage to take into account of a unique reaction rate for the enzyme (with the azide-based BG), saving costs and time for the linking to the tag of an infinite number of commercial available DBCO-molecules. Here, we successfully proved the labelling of the SNAP-tag<sup>®</sup> with several DBCO-based fluorophores and the covalent immobilisation of this protein on alkyne-coated surface sensors.

## 2. Materials and methods

### 2.1. Reagents

**BG** was from Activate Scientific GmbH (UK), whereas **MGPA** was gift of Prof D. Prosperi (University of Bicocca, Milan, Italy). **SNAL Vista<sup>®</sup> Green (SVG)**, **SNAP Cell<sup>®</sup> Block (SCB)**, **SNAP Cell<sup>®</sup> 4<sub>3</sub> (SC430)**, **BG-PEG-NH<sub>2</sub> (BGPA)**, **pSNAP-tag(m)** plasmid, DNA restriction endonucleases and DNA modification enzymes were purchased from New England Biolabs (USA). Molecular biology kits for plasmid preparations were from Macherey-Nagel GmbH (Germany). Oligonucleotides synthesis and DNA sequencing service were performed by Eurofins Genomics (Germany). **BDP F alkyne**, **BDP FL DBCO**, **Cy5 DBCO** were purchased from Lumiprobe GmbH (Germany). **DBCO-PEG<sub>4</sub>-Fluor 545**, **Tris(2-carboxyethyl)phosphine (TCEP)**, **Tris [(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA)** were from Sigma-Aldrich (St. Louis, MO, USA). **Pierce<sup>™</sup> Premium Grade Sulfo-NHS (N-hydroxy-sulfo-succinimide)** and **Pierce<sup>™</sup> Premium Grade 1-ethyl-3-(3-dimethyl-amino-propyl**

rbodiimide hydrochloride (EDC) were from Thermo Fisher Scientific (Carlsbad, CA).

## 2. Compounds synthesis: general procedures

<sup>1</sup>H (400 MHz) and <sup>13</sup>C (100 MHz) NMR spectra were measured on Bruker Avance Neo 400 MHz spectrometer. Chemical shifts were referenced to the residual solvent signal (CDCl<sub>3</sub>: δ<sub>H</sub> = 7.26, δ<sub>C</sub> = 7.0; DMSO: δ<sub>H</sub> = 2.50 δ<sub>C</sub> = 39.5). Low-resolution ESI-MS were obtained on an LTQ OrbitrapXL (Thermo Scientific) mass spectrometer. IR spectra were registered on Shimadzu DR 8001 spectrophotometer. Silica gel 60 (70–230 mesh) used for gravity column chromatography (CC) was purchased from Macherey–Nagel. Reactions were monitored by TLC on Merck 60 254 (0.25 mm) plates, visualised by staining with 5% H<sub>2</sub>SO<sub>4</sub> in ethanol or KMnO<sub>4</sub> and heating. Organic phases were dried with anhydrous Na<sub>2</sub>SO<sub>4</sub> before evaporation. Chemical reagents and solvents were from Aldrich, Alfa Aesar, and TCI and were used without any further purification unless stated otherwise.

## 3. Synthesis of BGN3

**BGN3** was synthesised according to the method of Zhang et al.<sup>34</sup>, whose experimental spectra were comparable. White solid. <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 7.84 (s, 1H), 7.53 (d, *J* = 7.8 Hz, 2H), 3.9 (d, *J* = 7.8 Hz, 2H), 6.28 (bs, 2H), 5.49 (s, 2H), 4.45 (s, 2H) (Figure S1(A)). <sup>13</sup>C NMR (100 MHz, DMSO-*d*<sub>6</sub>) δ 159.69, 136.80, 35.43, 128.82, 128.58, 66.44, 53.37. IR (KBr) cm<sup>-1</sup>: 3638, 3462, 322, 2799, 2132, 1424, 1257, 1163, 912, 790, 656, 514. ESI/MS: *m/z* [M + H<sup>+</sup>] 297 (Figure S1(B)).

## 4. Synthesis of BGSN3

**BGSN3** was synthesised by following the scheme in Figure S2.

### 4.1 Synthesis of 4-azido-N-(4-(hydroxymethyl) benzyl) butanamide (compound 3)

A stirred solution of compound **1** (see Figure S2; 1.176 g, 115 mmol, 1 eq/mol) was prepared according to the method by Wang et al.<sup>35</sup> in DCM (30 ml), compound **2** (1.500 g, 10.939 mmol, 2 eq/mol; prepared according to the method by Leng et al.<sup>36</sup>) and TEA (5.08 ml, 36.460 mmol, 4 eq/mol) were added. The mixture was stirred for 10 min at room temperature, then T3P (50% solution in EtOAc, 10.85 ml, 18.230 mmol, 2 eq/mol) was slowly added dropwise, and the stirred reaction was left overnight at room temperature until the complete conversion of the starting material (TLC: PE-EtOAc 4:6; R<sub>f</sub> 1 = 0.47; R<sub>f</sub> 3 = 0.16). The reaction was quenched by the addition of BRINE and extraction with DCM. After drying (Na<sub>2</sub>SO<sub>4</sub>) and evaporation, the residue was purified by gravity column chromatography on silica gel (gradient PE:EtOAc from 6:4 to 3:7) to afford compound **3** as a white solid (60 mg, 30%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.28 (d, *J* = 8.0 Hz, 2H), 2.0 (d, *J* = 7.9 Hz, 2H), 6.36 (t, *J* = 5.6 Hz, 1H), 4.63 (s, 2H), 4.35 (d, *J* = 5.7 Hz, 2H), 3.31 (t, *J* = 6.6 Hz, 2H), 2.27 (t, *J* = 7.3 Hz, 2H), 1.89 (s, *J* = 6.9 Hz, 2H) (Figure S3(A)). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 71.93, 140.30, 137.28, 127.69, 127.17, 64.53, 50.65, 43.19, 32.96, 4.69. IR (KBr) cm<sup>-1</sup>: 3276, 3055, 2921, 2880, 2103, 1635, 1540, 418, 1257, 1015, 827, 747, 677, 553. ESI/MS: *m/z* [M + H<sup>+</sup>] 249 (Figure S3(B)).

### 2.4.2. Synthesis of (N-(4-(((2-amino-9H-purin-6-yl)oxy)methyl)benzyl)-4-azidobutanamide) (BGSN3)

To a cooled solution (0 °C) of compound **3** (400 mg, 1.611 mmol, eq/mol) in dry DMF (10 ml) in a dry flask under N<sub>2</sub> atmosphere NaH (60% dispersion in mineral oil, 202 mg, 5.059 mmol, 3.14 eq/mol) was slowly added. The mixture was stirred at 0 °C for 10 min then DMAP (16 mg, 0.129 mmol, 0.08 eq/mol) and compound **4** (451 mg, 1.772 mmol, 1.1 eq/mol; prepared according to the method by Kindermann et al.<sup>37</sup>) were sequentially added. The reaction was then heated at room temperature and stirred for 4 h until the complete conversion of the starting material (TLC: DCM-MeOH 9:1; R<sub>f</sub> 4 = 0.70; R<sub>f</sub> BGSN3 = 0.55), then quenched by slow addition of BRINE and extraction with EtOAc. After drying (Na<sub>2</sub>SO<sub>4</sub>) and evaporation, the residue was purified by gravity column chromatography on silica gel (gradient DCM-MeOH from pure DCM 1:20:1) to afford **BGSN3** as a white solid (413 mg, 67%). <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 12.48 (bs, NH purine, 1H), 8.43 (t, *J* = 5.9 Hz, 1H), 7.85 (s, 1H), 7.49 (d, *J* = 7.7 Hz, 2H), 7.30 (d, *J* = 7.8 Hz, 2H), 6.31 (bs, NH2 purine 2H), 5.50 (s, 2H), 4.30 (d, *J* = 5.9 Hz, 2H), 3.3 (t, *J* = 6.8 Hz, 2H), 2.26 (t, *J* = 7.4 Hz, 2H), 1.81 (p, *J* = 7.1 Hz, 2H) (Figure S4(A)). <sup>13</sup>C NMR (100 MHz, DMSO-*d*<sub>6</sub>) δ 171.38, 159.9, 159.69, 155.23, 139.48, 137.90, 135.31, 128.60, 127.34, 113.5, 66.59, 50.36, 41.95, 32.23, 24.58. IR (KBr) cm<sup>-1</sup>: 3647, 3484, 337, 3282, 2794, 2100, 1580, 1403, 1282, 1163, 938, 835, 789, 635, 55. ESI/MS: *m/z* [M + H<sup>+</sup>] 382 (Figure S4(B)).

## 2.5 Plasmids and protein purification

The cloning procedures in the pQE31 expression vector (Qiagen Germany) were similar for both proteins. In particular, the pSNAI tag(m) Vector was used as a template to amplify the DNA fragment relative to the SNAP-tag<sup>®</sup> gene, by using QE\_SNAP-Fwc QE\_SNAP-Rev oligonucleotides pairs (5'-ATGGCAGGATCCA TGGACAAAGACTGCGAAATG-3'/5'-CTATCAAAGCTTAACCCAGCCCA GCTTGCCCA G-3'; BamH I and Hind III sites, respectively, are underlined). Afterwards, the resulting fragment and the pQE31 vector were digested with BamH I and Hind III restriction enzyme and ligated, leading to the final pQE-SNAP plasmid. The final SNAP-tag<sup>®</sup> protein was expressed with an extra N-terminal amino acid sequence, comprising a His<sub>6</sub>-tag (MRGSHHHHHHTDP-). The ligation mixture was used to transform *E. coli* KRX competent cells and positive colonies were confirmed by PCR and DNA sequence analyses.

*H*<sup>5</sup> was cloned as previously described<sup>22</sup>. SNAP-tag<sup>®</sup> and *t* proteins were expressed in *E. coli* ABLE C cells, grown at 37 °C in Luria–Bertani (LB) medium supplemented with 50 mg/l kanamycin and 100 mg/l ampicillin. The protein expression was induced with 1 mM isopropyl-thio-β-D-galactoside (IPTG) at an absorbance value of 0.5–0.6 A<sub>600</sub> nm. After overnight growth, cells were collected and resuspended 1:3 (w/v) in purification buffer (50 mM phosphate, 300 mM NaCl; pH 8.0) supplemented with 1% Triton X-100 and stored overnight at –20 °C. Subsequently, the biomass was treated in ice with lysozyme and DNase for 60 min and then sonicated as described (Perugino et al., 2012). After centrifugation (30 min at 60,000 × *g*), the cell extract was recovered and applied to a Protino Ni-NTA Column 1 ml (Macherey–Nagel) for His<sub>6</sub>-tag affinity chromatography. The eluted fractions containing the protein were collected and dialysed against phosphate-buffered saline (PBS 1 ×, 20 mM phosphate buffer, NaCl 150 mM, pH 7.3). Pool fractions were concentrated and protein purification was confirmed by SDS-PAGE analysis. Aliquots were finally stored at –20 °C.

**Table 1.** Substrate specificity of SNAP-tag<sup>®</sup> and H<sup>5</sup> by competitive inhibition method (IC<sub>50</sub>) by using SVG as substrate, and second order rate constant of the enzymic reaction of these protein-tags only on the BGSN3 substrate.

Structure	Name	SNAP-tag <sup>®</sup>		SsOGT-H <sup>5</sup>		Note
		IC <sub>50</sub> (μM)	k <sup>a</sup> (s <sup>-1</sup> M <sup>-1</sup> )	IC <sub>50</sub> (μM)	k (s <sup>-1</sup> M <sup>-1</sup> )	
	SVG	–	2.8 × 10 <sup>4</sup> <sup>b</sup>	–	1.6 × 10 <sup>4</sup>	[14,24]
	BG	36.8 ± 5.6	–	10.1 ± 1.0	–	This work
	SCB	2.1 ± 0.5	–	4.4 ± 0.8	–	This work
	BGN3	15.6 ± 0.3	–	23.5 ± 1.0	–	This work
	BG430	ND <sup>c</sup>	–	ND	–	This work
	BGPA	86.0 ± 6.7	–	14.3 ± 1.9	–	This work
	MGPA <sup>d</sup>	–	–	268.9 ± 19.1 <sup>e</sup>	–	This work
	BGSN3	17.8 ± 1.1	4.64 ± 1.04 × 10 <sup>5</sup>	10.0 ± 0.7	1.40 ± 0.47 × 10 <sup>4</sup>	This work

For each compound, the guanine moiety is drawn in black and the chemical group conjugated to the benzyl ring in blue. The fluorescein moiety of the SVG is green, whereas SCB differs from the other derivatives by the presence of a benzylic ring (in red). Azide group is conventionally coloured in magenta. Reaction rates at 25 °C; <sup>b</sup>this value was obtained by using a BG-fluorescein substrate (BG-FL) very similar to SVG; <sup>c</sup>not determined; <sup>d</sup>this molecule is a O<sup>6</sup>-methylguanine derivative; <sup>e</sup>competitive assay for H<sup>5</sup> was performed at 65 °C.

### 6 AGTs' substrate assay by competitive inhibition method

Competitive inhibition assay was performed as described<sup>23,38</sup>. Briefly, by using a fixed concentration of the fluorescent SVG (1 μM) and enzymes (5 μM), an increasing amount of guanine derivatives (0–2 mM) was added to the mixtures. The reactions were incubated for 30 min at 25 °C and 50 °C for SNAP-tag<sup>®</sup> and H<sup>5</sup> respectively, and loaded on SDS-PAGE. Subsequently, fluorescent bands were measured by gel-imaging on a VersaDoc 4000<sup>TM</sup> system (Bio-Rad), by applying a blue LED/530 bandpass filter. Obtained data were finally plotted by Equation (1),

$$RF = \frac{100\%}{1 + \left(\frac{[I]}{IC_{50}}\right)[S]} \quad (1)$$

where RF is the obtained Relative Fluorescence, [I] and [S] are the concentration of the inhibitor and the substrate, respectively, and

finally the IC<sub>50</sub> is the concentration needed to reduce by 50% the fluorescence intensity of the protein band.

We evaluated the activity of SNAP-tag<sup>®</sup> and H<sup>5</sup> enzymes on BGN3 and BGSN3 by the afore-mentioned IC<sub>50</sub> method (Figure S5(A,B)) and other O<sup>6</sup>-guanine-derivatives (Table 1).

### 2.7. In vitro Huisgen Cu(I)-catalysed cycloaddition reaction

The Huisgen chemical reaction was evaluated on SNAP-tag<sup>®</sup> and H<sup>5</sup> previously incubated with BGN3 and BGSN3. An opportune amount of purified proteins was incubated within an equimolar ratio of these substrates for 60–120 min at 25 °C and 37 °C respectively, to ensure the complete enzymatic labelling reaction. Later, we performed the subsequent cycloaddition using 5 μM of an alkyne-derivative of the fluorescein (BDP FL alkyne), in the

presence of copper (1 mM), TCEP (1 mM), TBTA (0.1 mM) and, here indicated, of SDS (0.05%). Finally, mixtures were loaded on SDS-PAGE and analysed as described in Section 4 (Figure S5(C,D)).

### 8. Molecular modelling

All molecular modelling studies were performed on a Tesla workstation equipped with two Intel Xeon X5650 2.67 GHz processors and Ubuntu 14.04 (<http://www.ubuntu.com>). The protein structures and 3D chemical structures were generated in PyMOL (The PyMOL Molecular Graphics System, version 2.2.3, Schrödinger, Inc., 2019).

### 9. Molecular dynamics (MD) simulation

The MD simulations were carried out using the Desmond simulation package of Schrödinger LLC (Schrödinger Release 2019-1: Desmond Molecular Dynamics System; D. E. Shaw Research: New York, NY, 2019; Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2019). The X-ray structure of the  $H^5$  covalently bound to **SVG** was used in this study, entry code 5A03<sup>9</sup>, water molecules were removed, and all hydrogen atoms and charges were added. The NPT ensemble with the temperature of 300 K and a pressure 1 bar was applied in all runs. The simulation length was 100 ns with relaxation time 1 ps. The OPLS\_2005 force field parameters were used in all simulations<sup>40</sup>. The long-range electrostatic interactions were calculated using the particle mesh Ewald method<sup>41</sup>. The cut-off radius in Coulomb interactions was 9.0 Å. The water molecules were explicitly described using the simple point charge model<sup>42</sup>. The Martyna-Tuckerman-Klein chain coupling scheme<sup>43</sup> with a coupling constant of 2.0 ps was used for the pressure control and the Nosé-Hoover chain coupling scheme<sup>44</sup> for the temperature control. Non-bonded forces were calculated using an r-RESPA integrator where the short-range forces were updated every step and the long-range forces were updated every three steps. The trajectory sampling was done at an interval of 1.0 ps. The behaviour and interactions between the ligands and protein were analysed using the Simulation Interaction Diagram tool implemented in the Desmond MD package. The stability of MD simulations was monitored by looking at the RMSD of the ligand and protein atom positions in time.

### 10. Determination of the rate constants of the chemo-enzymatic labelling approach

Rate constants of the enzymatic reactions with the only **BGSN3** were determined by the method of Gautier et al.<sup>14</sup>. In this case, purified proteins (5  $\mu$ M) were incubated with the substrate (5  $\mu$ M) in PBS 1 $\times$  buffer at 25 °C. Aliquots were taken at different times, the reactions were immediately stopped in Leammli Buffer 1 $\times$  in addition with 10  $\mu$ M of **Cy5 DBCO** fluorophore and placing tubes on ice.

Rate constants for the chemical reaction needed for the preliminary achievement of the *clickable*-SNAP and *clickable*- $H^5$  with **BGSN3**, which was obtained by the afore-described protocol, in order to get the complete labelling. Then, to each aliquot of 5  $\mu$ M of *clickable* proteins, 20  $\mu$ M of **DBCO-PEG<sub>4</sub>-Fluor 545** fluorophore was added. At different times, an excess of sodium azide (NaN<sub>3</sub>, 30 mM) was immediately added to each aliquot and then placing tubes on ice, in order to stop the click reaction between the azide group on the **BGSN3** and the **DBCO-PEG<sub>4</sub>-Fluor 545** molecule.

Finally, for both the experiments, all aliquots were boiled in a SDS buffer for 5 min, and immediately loaded on a SDS-PAGE, for the *gel-imaging* and *coomassie staining* analyses, as previously described. Data were fitted to a pseudo-first-order reaction model using the GraFit 5.0 software package (Erithacus Software Ltd Second-order rate constants  $k$  (in  $s^{-1} M^{-1}$ ) were then obtained by dividing the pseudo-first-order constant by the concentration of the substrate (Figure 2 and Table 1). Values given are an average of at least three independent measurements.

### 2.11. In vitro Huisgen copper-free cycloaddition reaction with different DBCO-fluorophores

For the copper-free click reaction, aliquots of 5  $\mu$ M of each *clickable*-protein were incubated for 60 min at room temperature in the dark with 5  $\mu$ M of fluorescent DBCO-derivative substrate (**BDP FL DBCO**, **Cy5 DBCO**, and **DBCO-PEG<sub>4</sub>-Fluor 545**) in a total volume of 10  $\mu$ l of PBS 1 $\times$  buffer (Figure 4 and Figure S6). The reactions were finally stopped in Leammli Buffer 3 $\times$ , loaded on SDS-PAGE, and analysed as described in Section 4, by applying blue LED/530 bandpass filter, red LED/695 bandpass filter and green LED/605 bandpass filter as excitation/emission parameters for each DBCO-fluorophore, respectively. The click reaction was also performed on 5  $\mu$ M of both the enzymes, but in the presence of an *EcCFE* diluted in PBS 1 $\times$  buffer.

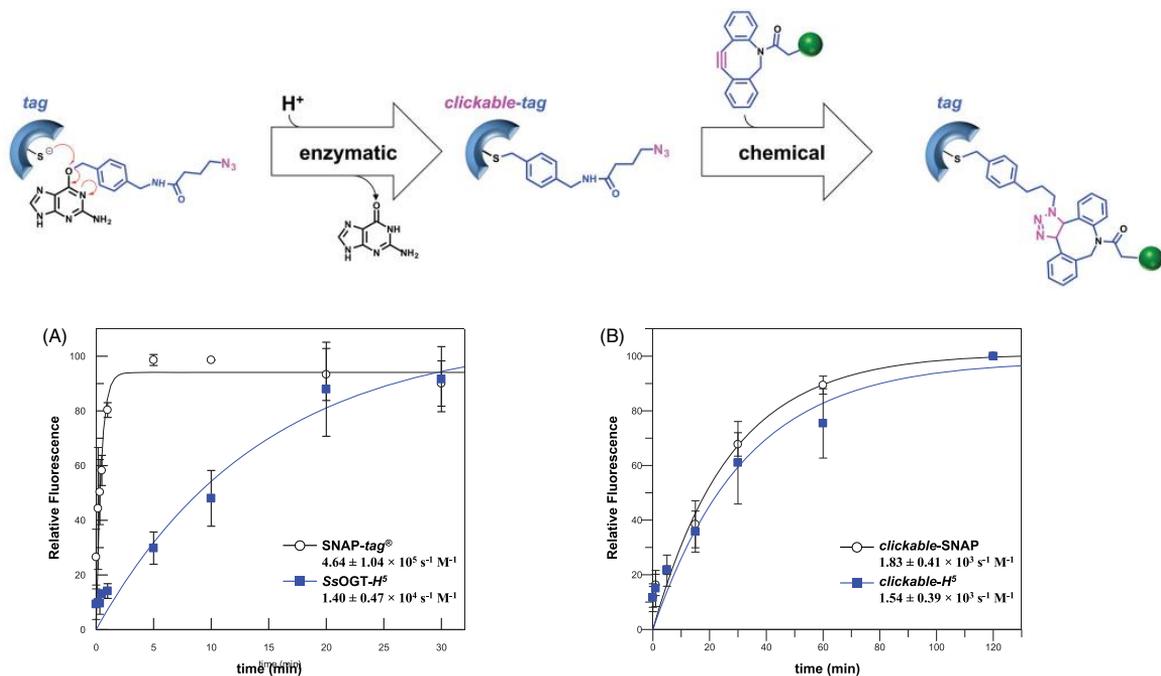
### 2.12. Procedure for protein immobilisation on bio layer interferometry (BLI), by following the chemo-enzymatic approach

OctetRED96<sup>TM</sup> (ForteBio, Fremont, CA) was used to immobilise specifically SNAP-tag<sup>®</sup> and  $H^5$  with the *chemo-enzymatic approach* (Figure 5(A,B)). Samples and reaction buffers were located in black 96-well plates (OptiPlate-96 Black, Black Opaque 96-well Microplate, PerkinElmer, Billerica, MA) in a maximum reaction volume of 300  $\mu$ l per well with 800 rpm shaking for each step. For the immobilisation procedure, AR2G sensors were first wetted in 200  $\mu$ l of pure water for at least 15 min, followed by an equilibration step (3 min) in acetate buffer 0.1 M, pH 5.0. Afterwards, they were activated with 20 mM 1-ethyl-3-(3-dimethyl-amino-propyl) carbodiimide hydrochloride (EDC)/20 mM N-hydroxy-sulfo-succinimide (sulfo-NHS) mixture in acetate buffer (60 min) and covered with 2 mM **propargyl-PEG<sub>3</sub>-amine** bifunctional linker (BroadPharm, San Diego, CA) in Loading step (20 min). To avoid the presence of any free amine groups on the biosensors, Blocking step with Ethanolamine 1 M (30 min) was performed. Subsequently, a Washing step (15 min) with water and an Equilibration step in click-reaction buffer (15 min) are followed.

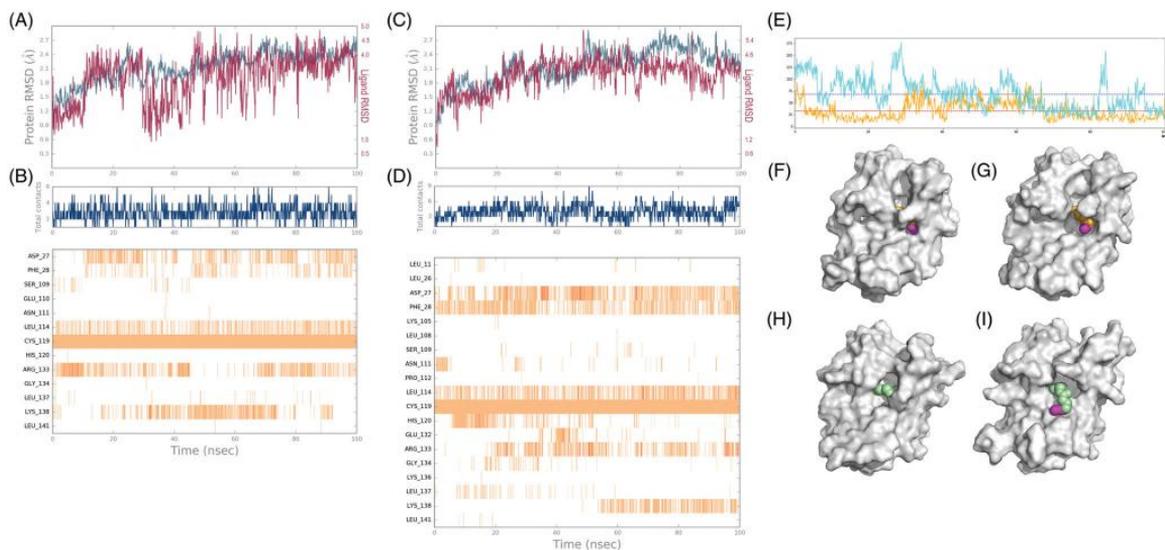
During the afore-described procedure, proteins were labelled with **BGSN3**. Finally, the immobilisation step for each sample via Huisgen reaction was carried out at 30 °C for 80 min, followed by a Washing step (20 min), in order to remove all the unbound molecules. This procedure was the same in the presence of the *EcCFE*. All measurements were performed in triplicates.

### 2.13. Permeability of eukaryotic and prokaryotic cells to BGSN3:

HEK293T cells were maintained at 37 °C with 5% CO<sub>2</sub> in Dulbecco's Modified Essential Medium (Invitrogen, Carlsbad, CA) supplemented with 10% Fetal Bovine Serum (FBS) (Invitrogen) and 100 U/ml Penicillin/Streptomycin (Roche, Switzerland). HEK293T cells were transfected with SNAP-tag<sup>®</sup> plasmid by using



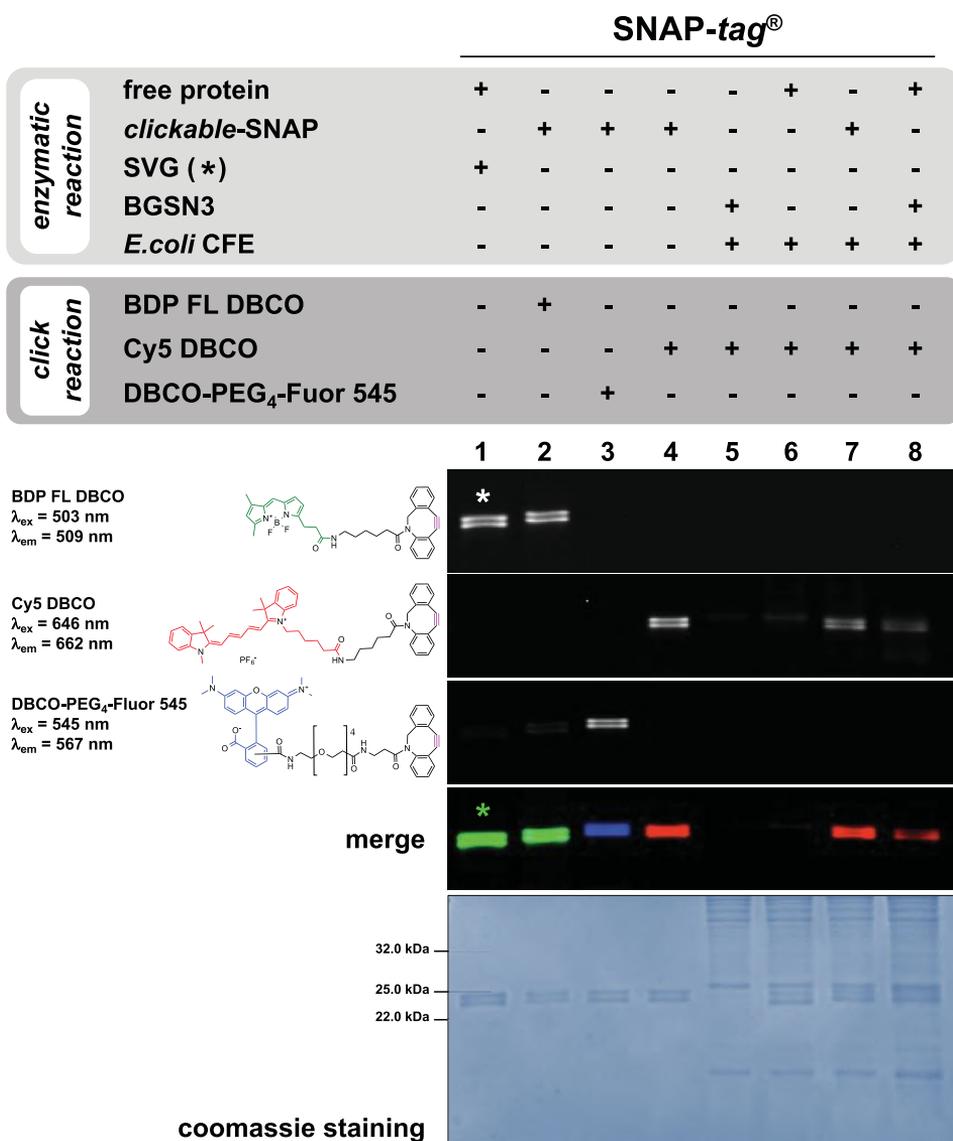
**Figure 2.** Reaction rates of the chemo-enzymatic approach. Pseudo-first-order reaction of *protein-tags* for (A) the enzymatic reaction with BGSN3 (see *k* values also in Table 1), and of *clickable-tags* for (B) Huisgen reaction with DBCO-PEG<sub>2</sub>-Fluor 545 (see values in the main text). Values given are an average of three independent measurements. The reaction scheme was an exemplification of Figure 1(B) in the main text. Data are represented as mean  $\pm$  SEM.



**Figure 3.** Molecular modelling on H<sup>5</sup> with BG-azides. (A) RMSD of the atomic positions for the compound BGN3 (Lig fit Prot, in red) and the protein H<sup>5</sup> (C $\alpha$  position in blue) of the 100 ns molecular dynamics simulations using Desmond package. (B) A timeline representation of the interactions and contacts (H-bonds, Hydrophobic, Water bridges). (C) RMSD of the atomic positions for the compound BGSN3 (Lig fit Prot, in red) and the protein H<sup>5</sup> (C $\alpha$  positions, in blue) of the 100 ns molecular dynamics simulations using Desmond package. (D) A timeline representation of the interactions and contacts (H-bonds, Hydrophobic, Ionic, Water bridges). (E) Solvent Accessible Surface Area (SASA) of BGN3/H<sup>5</sup> (in orange) and BGSN3/H<sup>5</sup> (in cyan) complexes over the MD simulation time (mean values are depicted as dot lines). (F-I) 3D surface models of H<sup>5</sup>-probe complexes with lower (F, H) and higher (G, I) SASA value for BGN3 (F, G) and BGSN3 (H, I), respectively.

pofectamine 2000 (Invitrogen) following manufacturer's protocol. The treatment with BGSN3 were performed, at the concentration and time indicated for each experiment. Twenty-four hours after transfection, we treated cells with BGSN3 for 2 h at different

concentrations ranging (from 1 to 25  $\mu\text{M}$ ), directly dissolving the compound in complete culture medium. Then cells were harvested, washed with PBS 1 $\times$  buffer and lysed with 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5 mM EDTA, 0.1% Triton X-100



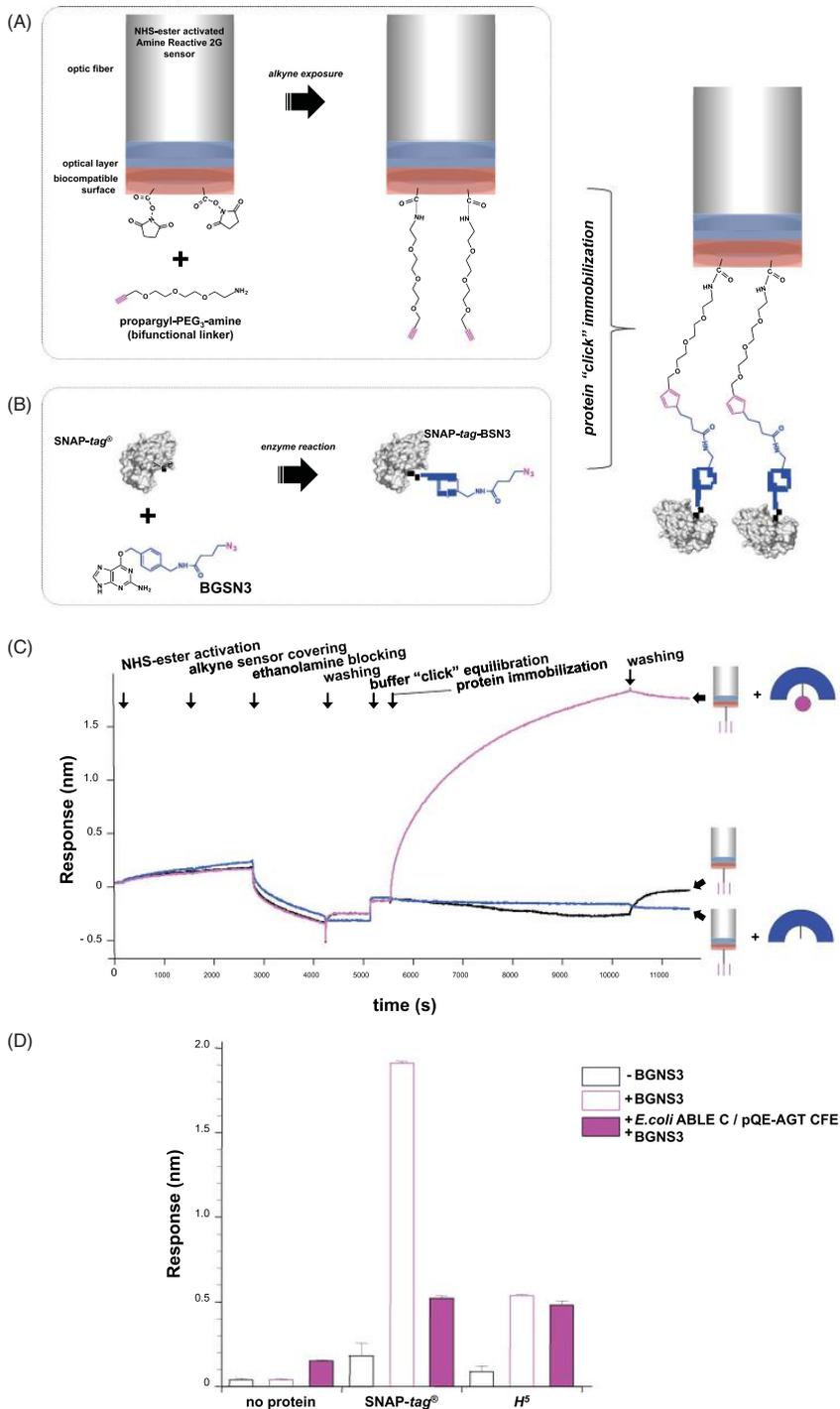
**Figure 4.** Specificity of the Huisgen reaction. *Gel-imaging* analysis of SNAP-tag<sup>®</sup> labelling by a chemo-enzymatic approach with BGSN3 and three different DBCO-derivative fluorophores. Protein (5 μM) was incubated with 5 μM of the azide-based BG for 60 min at 25 °C; then, an equimolar amount of DBCO-based substrate was added for the chemical click reaction, keeping the same time and temperature conditions. As control, SNAP-tag<sup>®</sup> was incubated only with SVG (lane 1, signal marked with an asterisk).

plemented with complete protease (Roche, Switzerland) and phosphatase (SERVA Electrophoresis, Germany) inhibitors. Afterwards, transfected cells were treated with a fixed concentration of BGSN3 (10 μM) at different time points (from 30 to 20 min). Again, HEK293T cells were washed and lysed as described before. To confirm the reaction with BGSN3, the same amount of protein extract (0.91 μg/μL for each sample) was incubated for 30 min at 25 °C with SVG. Subsequently, proteins were loaded on SDS-PAGE and analysed by *gel-imaging* on a VersaDoc 300™ system (Bio-Rad), by applying a blue LED/530 bandpass filter (Figure 6).

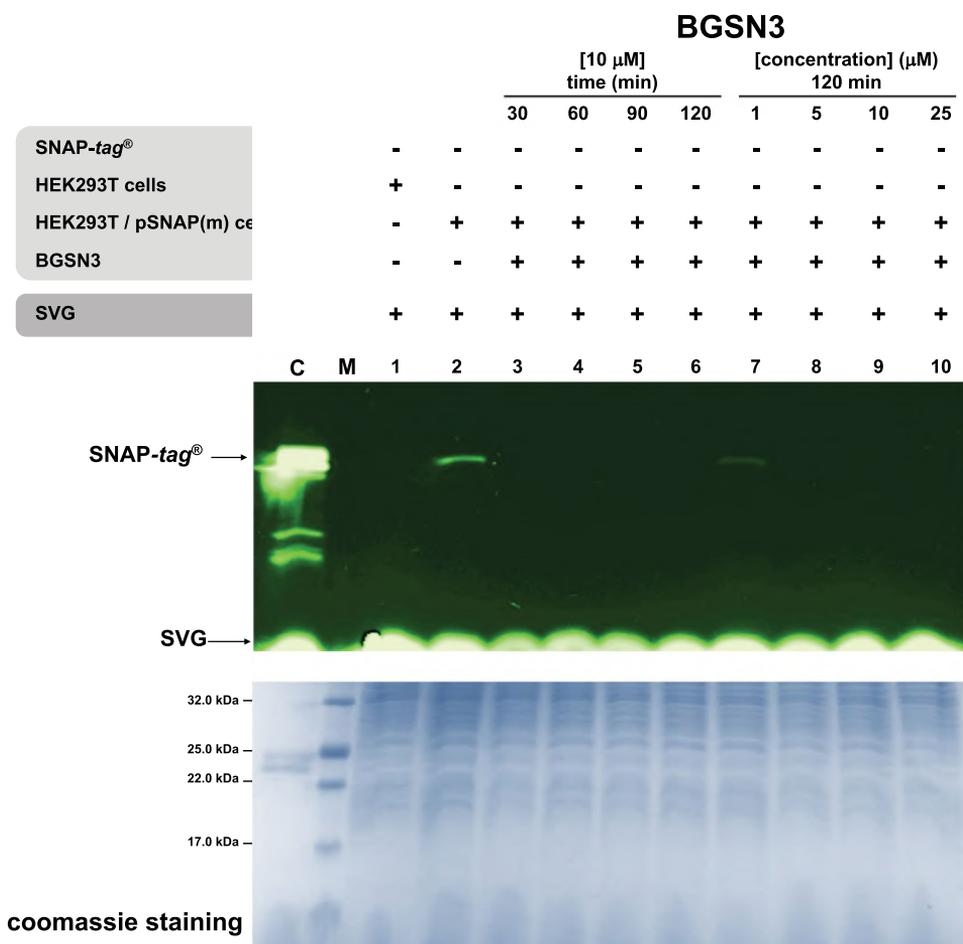
For flow cytometry analysis, HeLa cells were seeded in 24-well plates and transfected with SNAP-tag<sup>®</sup> plasmid by using

Lipofectamine 2000 (Invitrogen, USA) following manufacturer protocol. Twenty-four hours after the transfection, cells were treated with 25 μM BGSN3 for 1 h, and the excess of the substrate was washed out by 2 × 15 min, followed by 1 × 30 min washes. Cells were then treated with 2.5 μM BDP FL DBCO for 30 min and unbound fluorophore was removed by following the same procedure performed for the BGSN3. All treatments and washes were performed at 37 °C in a complete culture medium. Last cells were harvested by trypsinization, and fluorescence was measured using FACS CANTO II instrument. The analysis was performed on live singlet cells using FlowJo software (Figure S7(A)).

*E. coli* ABLE C strain was transformed with SNAP-tag<sup>®</sup> plasmid and protein expressed as previously described. After overnight



**Figure 5.** Covalent immobilisation of clickable-tags on the BLI sensor. (A) Covering of the BLI sensor with a bi-functional linker, exposing alkyne groups for the Huisgen 1,3-dipolar cycloaddition reaction; (B) reaction of the SNAP-tag<sup>®</sup> with BGSN3; (C) chemo-enzymatic SNAP-tag<sup>®</sup> immobilisation on BLI. The alkyne-covered sensor (*silver cylinder*) was immersed in wells containing the buffer (*in black*), the free SNAP-tag<sup>®</sup> (*in blue*) and the clickable-SNAP (*in magenta*); (D) column chart relative to the BLI immobilisation of purified *protein-tags* alone (*black-bordered bars*) or in the presence of BGSN3 (*magenta-bordered bars*). Filled magenta bars represent the BLI immobilisation using the EcCFE upon heterologous expression of *protein-tags*. Standard deviations were obtained from three independent experiments. Data are represented as mean  $\pm$  SEM.



**Figure 6.** Eukaryotic permeability to BGSN3. SDS-PAGE analysis by *gel-imaging* and *coomassie staining* of HEK293T cell lysates. After BGSN3 in medium treatment, lysates were incubated with SVG.

rowth, samples of 2 ml were treated with 100  $\mu$ M of **BGSN3** for 1 h at 25 °C and then collected by centrifugation at 2000  $\times g$ . Cell pellets of 0.05 g were resuspended 1:3 (w/v) in PBS 1 $\times$  supplemented with 1% Triton X-100 and subjected to cell lysis, by applying 5 cycles of freeze-thawing. After a centrifugation at 13,000  $\times g$ , the supernatants containing the protein extract were incubated 30 min at 25 °C with **SVG**, and proteins were loaded on SDS-PAGE. Finally, fluorescent bands were analysed by *gel-imaging* techniques (Figure S7(B)).

## 2. Result and discussion

### 2.1. Substrate specificity of AGTs on BG-based substrates

Following the irreversible reaction shown in Figure 1, we evaluated the activity of two enzymes in our possession on several <sup>6</sup>-guanine-derivatives (Table 1). Because most of them are non-fluorescent compounds, we performed an AGTs' competitive inhibition assay by using the fluorescein-derivative SNAP-Vista<sup>®</sup> as substrate (**SVG**), as previously described<sup>22–24,45</sup>. Briefly, the reaction of an AGT with **SVG** led to a fluoresceinated protein, which can be visualised as a fluorescent band in *gel-imaging* analysis after SDS-PAGE. The presence of increasing amounts of a

non-fluorescent competitor in the reaction causes a decrease of the fluorescent signals, which can be measured and plotted for the IC<sub>50</sub> values determination<sup>23,46</sup>. As shown in Table 1, SNAP-tag and H<sup>5</sup> displayed different behaviours versus these competitor without any rationale for the dimension and/or polarity of the conjugated chemical groups. While SNAP-Cell<sup>®</sup> 430 (**SC430**) completely lost the competition with **SVG**, both the enzymes are extremely active on the SNAP Cell<sup>®</sup> Block (**SCB**), displaying the lowest IC<sub>50</sub> value measured. This result was expected, because **SCB** has a structure very similar to the Lomeguatrib, one of the most efficient inhibitors of the hMGMT protein, employed in the cancer treatment in combination with alkylating agents-based chemotherapeutics<sup>47</sup>.

In general, all commercially available products used (**SVG**, **SCB**, **BG430**, and **BG-PEG-NH2**, **BGPA**) are good substrates for the SNAP-tag<sup>®</sup> and H<sup>5</sup> enzymes, completing their labelling reaction in a few hours (data not shown). However, based on our results, the choice of the chemical group to be conjugated to the O<sup>6</sup>-guanine for customized substrates may present risks, with consequent decreases in the reaction rate for these *protein-tags*. This was the case of methyl-guanine-PEG-NH2 (**MGPA**), which is an O<sup>6</sup>-methyl-guanine derivative, used for the immobilisation of SNAP-tag<sup>®</sup> on nanoparticles<sup>48</sup>. The latter is not a preferred substrate, probab

because of the absence of the benzyl ring, which leads to complete labelling of the SNAP-tag<sup>®</sup> and  $H^5$  after over-night incubation at 4 °C<sup>48</sup> and 65 °C (data not shown), respectively.

## 2. In vitro enzymatic reaction of engineered AGTs with BG-zide substrates

Recent studies were focussed on the synthesis of alternative “BG-building blocks,” which offer the opportunity to produce SNAP-substrates by following easier and faster protocols: an alkyne substituted  $O^6$ -BG was employed in the synthesis of compounds by the Huisgen cycloaddition with azide-based fluorescent probes<sup>49</sup> or, inversely, by using the  $O^6$ -BG- $N_3$  (**BGN3**, Figure S1) for the conjugation with alkyne-based chemical groups<sup>34</sup>. We evaluated the enzymatic reaction of the  $H^5$  and the SNAP-tag<sup>®</sup> directly on **BGN3** and a synthesised BG-derivative containing a benzyl ring oppositely spaced from the azide group (**BGSN3**, Figure S4): after the reaction, no fluorescent signal was obtained on SDS-PAGE gel-imaging upon the addition of **SVG** (Figure S5(A,B)). This indicates that the catalytic cysteine was completely blocked by the benzyl-zide moiety, impeding the access of the fluorescent substrate to the active site. Compared to the classical BG-derivatives, these *pro-in-tags* showed a reasonable activity on both these BG-azides, as shown by the calculated  $IC_{50}$  (Table 1 and Figure S5(A,B)).

After the enzymatic reaction of  $H^5$  with **BGN3** and **BGSN3**, we performed the subsequent cycloaddition using an alkyne-derivative of the fluorescein (**BDP FL alkyne**): however, the chemical reaction was less efficient using the former substrate (Figure S5(C), lane 2). In this case, the complete fluorescein labelling of the protein was achieved only in the presence of a small amount of SDS during the cycloaddition step (lane 3), suggesting that the protein is still folded after the enzymatic reaction and the azide is hidden in the active site core. The addition of the denaturant could have slightly opened the protein structure, favouring a better exposure of the azide group to the solvent, and allowing the click reaction to occur.

On the contrary, using **BGSN3** as substrate, the labelling of both the enzymes was comparable to the classical reaction with **VG** without any denaturing agent, likely the longer spacer of **BGN3** could sufficiently move away from the azide group from the protein surface for the Huisgen reaction (Figure S5(D), lanes 2 and 4). From now on, experiments were only performed by using the longer BG-azide. We first calculated the rate of the enzymatic reaction, demonstrating that both *protein-tags* show a high catalytic activity comparable to the commercial BG-derivatives currently used (Figure 2(A) and Table 1), also indicating that the complete protein labelling in less of an hour can be performed<sup>13,14,24</sup>.

## 3. Molecular modelling on the $H^5$ with BG-azides

**BGN3** and **BGSN3** differ in length since the chemical spacer between the benzyl ring and the active azide makes the latter potentially more prone to the labelling reaction. It could be assumed that this aspect alone influences the availability of the azide moiety to react. However, proteins are not a static system, the amino acids side-chain movements could mask the azide and prevent the “click” chemistry reaction. The covalent complexes of these compounds with  $H^5$  were analysed with Molecular Dynamics (MD) simulations using the Desmond package (see Experimental section). The complexes were simulated for 100 ns at 300 K using standard protocol. The protein structure has been stabilised, as shown in the RMSDs for both the IDO1  $C\alpha$  and the ligand (Figure

3(A,C)). The MD results were analysed in terms of Solvent Accessible Surface Area (SASA) of the compounds: more time the compounds are exposed to the solvent, the higher is the possibility to react<sup>50</sup>. In Figure 3 is reported the fluctuation of the SASA values over the simulation time together with the structure model of the  $H^5$  protein in complex with **BGN3** and **BGSN3**, respectively. The former is less exposed to the solvent with a SASA value of  $32.967 \pm 18.573 \text{ \AA}^2$  compared to **BGSN3**, which shows a high SASA value  $68.302 \pm 32.455 \text{ \AA}^2$ . This simulation confirmed our biochemical data, proposing the BG-derivative with the spacer as a better substrate for our *chemo-enzymatic approach*.

## 3.4. Specificity and versatility of the chemo-enzymatic reaction

The  $O^6$ -BG-based **BGSN3** is a good substrate for the two *protein tags* used (Table 1 and Figure 2(A)) and offering the advantage to sufficiently expose the azide group for the Huisgen reaction. This was the starting point to examine: (i) the labelling efficiency of the *clickable*-SNAP and *clickable*- $H^5$  by using different DBCO-based fluorophores; (ii) the specificity of the “click” reaction.

Upon the reaction with **BGSN3**, all cycloaddition reactions with three different DBCO-based fluorophores were complete in c. 30–45 min in PBS 1× buffer (Figure 4, lanes 2–4), with a protein labelling as efficient as the enzymatic reaction using the sole **SV** (lane 1). We quantitatively evaluated the rate ( $k$ ) of the click reaction by using the **DBCO-PEG4-Fluor 545** fluorophore: as expected, both the *clickable-tags* were labelled with the same efficiency ( $1.83 \pm 0.41 \times 10^3 \text{ s}^{-1} \text{ M}^{-1}$  for SNAP-tag<sup>®</sup>;  $1.54 \pm 0.39 \times 10^3 \text{ s}^{-1} \text{ M}^{-1}$  for  $H^5$ ), demonstrating that the chemical reaction is sufficiently fast and independent from the *tags* (Figure 2(B)).

Concerning the specificity, we added a crude protein extract from *Escherichia coli* ABLE C (EcCFE), without any AGT activity: the *gel-imaging* analysis (Figure 4, lane 5). In this context, the presence of the free *protein-tag* and the DBCO-fluorophore also did not result in any fluorescent signal (lanes 6), whereas the previously purified *clickable*-SNAP (lane 7), as well as its free form in the presence of **BGSN3** (lane 8), was specifically able to complete the chemo-enzymatic reaction, giving an evident fluorescent signal. The high specificity of our approach was also confirmed by using the  $H^5$  enzyme, which displays a better labelling reactivity than the mesophilic SNAP-tag<sup>®</sup> (Figure S6). Probably, something in the extract might impede SNAP-tag<sup>®</sup> activity. These results clearly demonstrated the high efficiency of our *chemo-enzymatic approach* for the labelling of both the *protein-tags* used.

## 3.5. Application to the bio layer interferometry

The possibility to apply the SNAP-tag<sup>®</sup> technology to the Surface Plasmon Resonance (SPR) for the covalent immobilisation of a protein of interest was first explored by the group of Kai Johnsson<sup>3</sup> followed by other groups with the same substrate<sup>51</sup> or a biotinylated BG-derivative<sup>52</sup>. Their approaches, again, required preliminarily the synthesis and the purification of a compatible substrate to cover the sensor chip surface. We used, instead, the **BGSN3** substrate for the immobilisation of the SNAP-tag<sup>®</sup> directly on an alkyne-derived sensor chip of the bio layer interferometry (BLI) equipment, as shown in Figure 5. This technique is more advantageous with respect to the SPR because: (i) it needs a smaller amount of sample, making it more compatible to higher throughput (throughput capacity of running up to 96 samples in parallel); (ii) the possibility to reuse samples, and (iii) of the total independency from any microfluidic issues.

Given the lack of any available BLI alkyne-derived sensors, we first activated the AR2G type by a bi-functional linker (**propargyl-EG 3-amine**) in order to expose an alkyne group on the surface (Figure 5(A)). This modified protocol provides first the coating of the sensor tips with alkyne groups (approx. 80 min), during that reaction between the *protein-tag* and **BGSN3** inside the 96-wells rack takes place (Figure 5(B)). Only the contemporary presence of the *clickable*-SNAP and the alkyne-coated sensor led to a measurable response (Figure 5(C)). After washing procedures, the signal did not significantly drop-down, given the covalent reaction between the protein and the sensor. We successfully achieved results with both the enzymes, although temperature and times of the enzymatic reaction on BLI (30 °C) favoured the SNAP-*tag*<sup>®</sup> respect to the thermophilic *H*<sup>51,3,24</sup>. Furthermore, in *EcCFEs* where both the enzymes were expressed, a specific and efficient immobilisation on BLI sensor tips occurred (Figure 5(D)), although the VAP-*tag*<sup>®</sup> displayed a lower labelling efficiency in the *EcCFE*, as expected (compare lane 8 in Figure 4 and Figure S6). As for other techniques, this specific surface immobilisation of SNAP-*tag*<sup>®</sup> gives the opportunity to perform a directly *on-chip purification* of a tagged-POI from a crude lysate, without any purification step, in an indirect manner, which favours a better orientation of the POI in its biological activities.

## 6. Permeability of eukaryotic and prokaryotic cells to G-azides

One of the major applications of the SNAP-*tag*<sup>®</sup> technology concerns the field of cell biology, where detecting fluorescently-tagged-POIs in living cells represents an important tool to study protein functions and locations<sup>53</sup>. To test our *chemo-enzymatic approach*, we first investigated the permeability of **BGSN3**. Lysates of HEK293T cells pre-treated with **BGSN3** were then incubated with the **SVG** substrate: the absence of any fluorescent signal by *el-imaging* only in BG-azide treated lysates demonstrated that the internalisation of **BGSN3** was fast (ca. 30 min; Figure 6, lane 3) and at concentrations comparable with commercial cell biology G-substrates (in the range of <5 μM; Figure 6, lane 8). Preliminary experiments by FACS analysis confirmed that the *in vivo* cycloaddition between **BGSN3** and the **BDP-FL DBCO** occurred (Figure S7(A)). This was also confirmed for *E. coli* bacterial cells (Figure S7(B), lane 2).

## 7. Conclusions and perspectives

We developed an innovative modification of the SNAP-*tag*<sup>®</sup> technology, in order to overcome times and costs relative to the production and the utilisation of commercial or purified customised G-derivatives. Although they are compatible in terms of catalytic activity as for the SNAP-*tag*<sup>®</sup>, as well as for the others *GT*<sub>S</sub><sup>22–24,37,46,54</sup> the risk of lowering the catalytic activity of these tags with customised BG-derivatives should not be underestimated (Table 1). We started by the knowledge that: (i) self-labelling *protein-tags* are still folded and enough stability in their enzymated form after the enzymatic reaction<sup>13,24</sup>; (ii) the Huisgen cycloaddition is extremely versatile, fast and specific. Recently, the latter was used for the entrapment of catalytic activities by azide-activated pseudo-substrates in a well-known powerful method, the *in vivo* activity-based protein profiling (ABPP)<sup>55</sup>. For these reasons, the *chemo-enzymatic approach* (Figure 1(B)) with an opportunely selected azide-based BG-substrate (**BGSN3**) was set up: the efficient exposition of the azide outside the protein surface allows the reaction with a huge number of commercially DBCO-based

molecules, more than those BG-derivatives, keeping high the specificity in the presence of *in vitro* “perturbing” proteins (like in cell lysates) and the *in vivo* labelling of expressed SNAP-*tag*<sup>®</sup> in eukaryotic cells. Finally, **BGSN3** proved to be a good substrate for the direct immobilisation of these tags on solid surfaces. We demonstrated that splitting the SNAP-*tag*<sup>®</sup> reaction into two fast steps as experimentally measured (Figure 2(A,B)), does not affect the overall rate and efficiency of the protein labelling<sup>13,24</sup>, thus opening new perspectives and widening the applications of this powerful biotechnology.

## Authors’ contributions

Conceptualisation, G.P. and A.Mi.; Methodology, R.Me., G.F. Investigation, R.Me., D.C. and M.C.; Formal Analysis, C.M., R.Mi. an A.Ma.; Validation, R.Me., R.Ma., A.Mi. and G.P.; Writing – Origin Draft, R.Me. and G.P.; Writing – Review & Editing, C.M., F.R. an A.V. and G.P.; Funding Acquisition, A.V. and G.P.; Resources, R.M. R.Ma. and D.C.; Supervision, A.L., A.Mi. and G.P.

## Acknowledgement

We are grateful to prof. Davide Proserpi (University of Milano Bicocca) for the kind gift of the **MGPA** substrate.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was funded by Ministero dell’Istruzione Università Ricerca (MIUR) National Operational Program (PON) Research an Innovation 2014–2020 [CCI 2014IT16M2OP005], European Soci Fund, Action I0.1 “Innovative Doctorates with Industrial characterization.” This research was also founded by Università d Piemonte Orientale (project: RIVmiggianorFAR\_2019).

## References

1. Chudakov DM, Matz MV, Lukyanov S, Lukyanov K. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* 2010;90:1103–63.
2. Liss V, Barlag B, Nietschke M, Hensel M. Self-labelling enzymes as universal tags for fluorescence microscopy super-resolution microscopy and electron microscopy. *S Rep* 2015;5:17740–52.
3. Hatlem D, Trunk T, Linke D, Leo JC. Catching a SPY: using the SpyCatcher-SpyTag and related systems for labeling an localizing bacterial proteins. *Int J Mol Sci* 2019;20:2129–39.
4. Los GV, Encell LP, McDougall MG, et al. HaloTag: a novel protein labeling technology for cell imaging and proteomic analysis. *ACS Chem Biol* 2008;3:373–82.
5. England CG, Luo H, Cai W. HaloTag technology: a versatile platform for biomedical applications. *Bioconjug Chem* 2012;23:975–86.
6. Pegg AE. Repair of O(6)-alkylguanine by alkyltransferase. *Mutat Res* 2000;462:83–100.
7. Daniels DS, Mol CD, Arvai AS, et al. Active and alkylate human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *Embo J* 2000;19:1719–30.

3. Daniels DS, Woo TT, Luu KX, et al. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat Struct Mol Biol* 2004;11:714–20.
2. Keppler A, Gendreizig S, Gronemeyer T, et al. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat Biotechnol* 2003;21:86–9.
2. Keppler A, Pick H, Arrivoli C, et al. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci USA* 2004;10:9955–9.
1. Juillerat A, Gronemeyer T, Keppler A, et al. Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules *in vivo*. *Chem Biol* 2003;10:313–7.
2. Gronemeyer T, Chidley C, Juillerat A, et al. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng Des Sel* 2006;19:309–16.
3. Mollwitz B, Brunk E, Schmitt S, et al. Directed evolution of the suicide protein O6-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* 2012;51:986–94.
4. Gautier A, Juillerat A, Heinis C, et al. An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* 2008;15:128–36.
5. Vogel P, Moschref Merkle MM, Li Q, et al. Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat Methods* 2018;15:535–8.
5. Srikun D, Albers AE, Nam CI, et al. Organelle-targetable fluorescent probes for imaging hydrogen peroxide in living cells via SNAP-Tag protein labeling. *J Am Chem Soc* 2010;132:4455–65.
7. Abo M, Minakami R, Miyano K, et al. Visualization of phagosomal hydrogen peroxide production by a novel fluorescent probe that is localized via SNAP-tag labeling. *Anal Chem* 2014;86:5983–90.
3. Griss R, Schena A, Reymond L, et al. Bioluminescent sensor proteins for point-of-care therapeutic drug monitoring. *Nat Chem Biol* 2014;10:598–603.
2. Tomat E, Nolan EM, Jaworski J, Lippard SJ. Organelle-specific zinc detection using zinpyr-labeled fusion proteins in live cells. *J Am Chem Soc* 2008;130:15776–7.
2. Kamiya M, Johnsson K. Localizable and highly sensitive calcium indicator based on a BODIPY fluorophore. *Anal Chem* 2010;82:6472–9.
1. Saccà B, Meyer R, Erkelenz M, et al. Orthogonal protein decoration of DNA origami. *Angew Chem Int Ed Engl* 2010;49:9378–83.
2. Perugino G, Vettone A, Illiano G, et al. Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J Biol Chem* 2012;287:4222–31.
3. Perugino G, Miggiano R, Serpe M, et al. Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic Acids Res* 2015;43:8801–16.
4. Vettone A, Serpe M, Hidalgo A, et al. A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering. *Extremophiles* 2016;20:1–13.
5. Visone V, Han W, Perugino G, et al. *In vivo* and *in vitro* protein imaging in thermophilic archaea by exploiting a novel protein tag. *PLoS One* 2017;12:e0185791.
5. Lo Gullo G, Mattosovich R, Perugino G, et al. Optimization of an *in vitro* transcription/translation system based on *Sulfolobus solfataricus* cell lysate. *Archaea* 2019;2019:848253.
27. Merlo R, Del Prete S, Valenti A, et al. An AGT-based protein tag system for the labelling and surface immobilization of enzymes on *E. coli* outer membrane. *J Enzyme Inhib Med Chem* 2019;34:490–9.
28. Del Prete S, Merlo R, Valenti A, et al. Thermostability enhancement of the  $\alpha$ -carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* by using the anchoring and self-labelling-protein-tag system (ASL<sup>tag</sup>). *J Enzyme Inhib Med Chem* 2019;34:946–54.
29. Moschel RC, McDougall MG, Dolan ME, et al. Structural features of substituted purine derivatives compatible with depletion of human O6-alkylguanine-DNA alkyltransferase. *Med Chem* 1992;35:4486–91.
30. Chae MY, McDougall MG, Dolan ME, et al. Substituted O6-benzylguanine derivatives and their inactivation of human O6-alkylguanine-DNA alkyltransferase. *J Med Chem* 1994;37:342–7.
31. Terashima I, Kawate H, Sakumi K, et al. Substrate specificity of human O6-methylguanine-DNA methyltransferase for O6-benzylguanine derivatives in oligodeoxynucleotides. *Chem Res Toxicol* 1997;10:1234–9.
32. McElhinney RS, Donnelly DJ, McCormick JE, et al. Inactivation of O6-alkylguanine-DNA alkyltransferase. Novel O6-(hetaryl)methyl)guanines having basic rings in the side chain. *J Med Chem* 1998;41:5265–71.
33. Sun G, Fan T, Zhang N, et al. Identification of the structural features of guanine derivatives as MGMT inhibitors using 3D-QSAR modeling combined with molecular docking. *Molecules* 2016;21:823–44.
34. Zhang CJ, Li L, Chen GYJ, et al. One- and two-photon live cell imaging using a mutant SNAP-Tag protein and its FRE substrate pairs. *Org Lett* 2011;13:4160–3.
35. Huang M, Deng Z, Tian J, Liu T. Synthesis and biological evaluation of salinomycin triazole analogues as anticancer agents. *Eur J Med Chem* 2017;127:900–8.
36. Leng S, Qiao Q, Miao L, et al. A wash-free SNAP-tag fluorescent probe based on the additive effects of quench release and environmental sensitivity. *Chem Commun* 2013;6448–51.
37. Kindermann M, George N, Johnsson N, Johnsson K. Covalent and selective immobilization of fusion proteins. *J Am Chem Soc* 2003;125:7810–1.
38. Mattosovich R, Merlo R, Fontana A, et al. A journey down to hell: new thermostable protein-tags for biotechnology at high temperatures. *Extremophiles* 2020;24:81–91.
39. Rossi F, Morrone C, Massarotti A, et al. Crystal structure of thermophilic O6-alkylguanine-DNA alkyltransferase-derived self-labeling protein-tag in covalent complex with a fluorescent probe. *Biochem Biophys Res Commun* 2018;50:698–703.
40. Banks JL, Beard HS, Cao Y, et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* 2005;26:1752–80.
41. Toukmaji AY, Board JA. Jr., Ewald summation techniques in perspective: a survey. *Comput Phys Commun* 1996;95:73–9.
42. Zielkiewicz J. Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. *J Chem Phys* 2005;123:104501.
43. Martyna GJ, Klein ML, Tuckerman M. Reversible multiple time scale molecular dynamics. *J Chem Phys* 1992;97:2635–43.

4. Evans DJ, Holian BL. The Nose–Hoover thermostat. *J Chem Phys* 1985;3:4069.
5. Miggiano R, Perugino G, Ciaramella M, et al. Crystal structure of *Mycobacterium tuberculosis* O6-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA. *Biochem J* 2016;473:123–33.
5. Morrone C, Miggiano R, Serpe M, et al. Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase. *Biochim Biophys Acta Gen Subj* 2017;1861:86–96.
7. Ranson M, Middleton MR, Bridgewater J, et al. Lomeguatrib, a potent inhibitor of O6-alkylguanine-DNA-alkyltransferase: phase I safety, pharmacodynamic, and pharmacokinetic trial and evaluation in combination with temozolomide in patients with advanced solid tumors. *Clin Cancer Res* 2006;12:1577–84.
3. Colombo M, Mazzucchelli S, Montenegro JM, et al. Protein oriented ligation on nanoparticles exploiting O6-alkylguanine-DNA transferase (SNAP) genetically encoded fusion. *Small* 2012;8:1492–7.
3. Song X, Wang C, Han Z, et al. Terminal alkyne substituted O6-benzylguanine for versatile and effective syntheses of fluorescent labels to genetically encoded SNAP-tags. *RS Adv* 2015;5:23646–9.
50. Ferraris DM, Spallek R, Oehlmann W, et al. Structure of *Thermococcus litoralis*  $\Delta^1$ -pyrroline-2-carboxylate reductase in complex with NADH and L-proline. *Proteins* 2014;82:2268–74.
51. Huber W, Perspicace S, Kohler J, et al. SPR-based interaction studies with small molecular weight ligands using hAC fusion proteins. *Anal Biochem* 2004;333:280–8.
52. Niesen J, Sack M, Seidel M, et al. SNAP-Tag technology: useful tool to determine affinity constants and other functional parameters of novel antibody fragments. *Bioconjug Chem* 2016;27:1931–41.
53. Testa I, Wurm CA, Medda R, et al. Multicolor fluorescence nanoscopy in fixed and living cells by exciting conventional fluorophores with a single wavelength. *Biophys J* 2010;98:2686–94.
54. Miggiano R, Valenti A, Rossi F, et al. Every OGT is illuminated ... by fluorescent and synchrotron lights. *Int J Mol Sci* 2017;18:2613–31.
55. Zweerink S, Kallnik V, Ninck S, et al. Activity-based proteome profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nat Commun* 2017;8:15352.

RESEARCH PAPER



## An AGT-based *protein-tag* system for the labelling and surface immobilization of enzymes on *E. coli* outer membrane

Luca Merlo<sup>a\*</sup>, Sonia Del Prete<sup>a\*</sup>, Anna Valenti<sup>a</sup>, Rosanna Mattosovich<sup>a</sup>, Vincenzo Carginale<sup>a</sup>, Claudio T. Supuran<sup>b</sup> , Clemente Capasso<sup>a</sup>  and Giuseppe Perugino<sup>a</sup>

<sup>a</sup>Department of Biology Agriculture and Food Sciences, Institute of Bioscience and BioResources – National Research Council of Italy, Naples, Italy; <sup>b</sup>Neurofarba Department, University of Florence, Polo Scientifico, Sesto Fiorentino Firenze, Italy

### ABSTRACT

The use of natural systems, such as outer membrane protein A (OmpA), phosphoprotein E (PhoE), ice nucleation protein (INP), etc., has been proved very useful for the surface exposure of proteins on the outer membrane of Gram-negative bacteria. These strategies have the clear advantage of unifying in a one-step production, the purification and the *in vivo* immobilisation of proteins/biocatalysts onto a specific biological support. Here, we introduce the novel *Anchoring-and-Self-Labeling-protein-tag* (ASL<sup>tag</sup>), which allows the *in vivo* immobilisation of enzymes on *E. coli* surface and the labelling of the neosynthesised proteins with the engineered alkylguanine-DNA-alkyl-transferase (H<sup>5</sup>) from *Sulfolobus solfataricus*. Our results demonstrated that this *tag* enhanced the overexpression of thermostable enzymes, such as the carbonic anhydrase (SspCA) from *Sulfurihydrogenibium yellowstonense* and the  $\beta$ -glycoside hydrolase (S $\beta$ Gly) from *S. solfataricus*, without affecting their folding and catalytic activity, proposing a new tool for the improvement in the utilisation of biocatalysts of biotechnological interest.

### ARTICLE HISTORY

Received 30 October 2018  
Revised 28 November 2018  
Accepted 4 December 2018

### KEYWORDS

Carbonic anhydrase;  
 $\beta$ -glycoside hydrolase;  
thermostable *protein-tag*;  
ice nucleation protein;  
enzyme immobilisation

### 1. Introduction

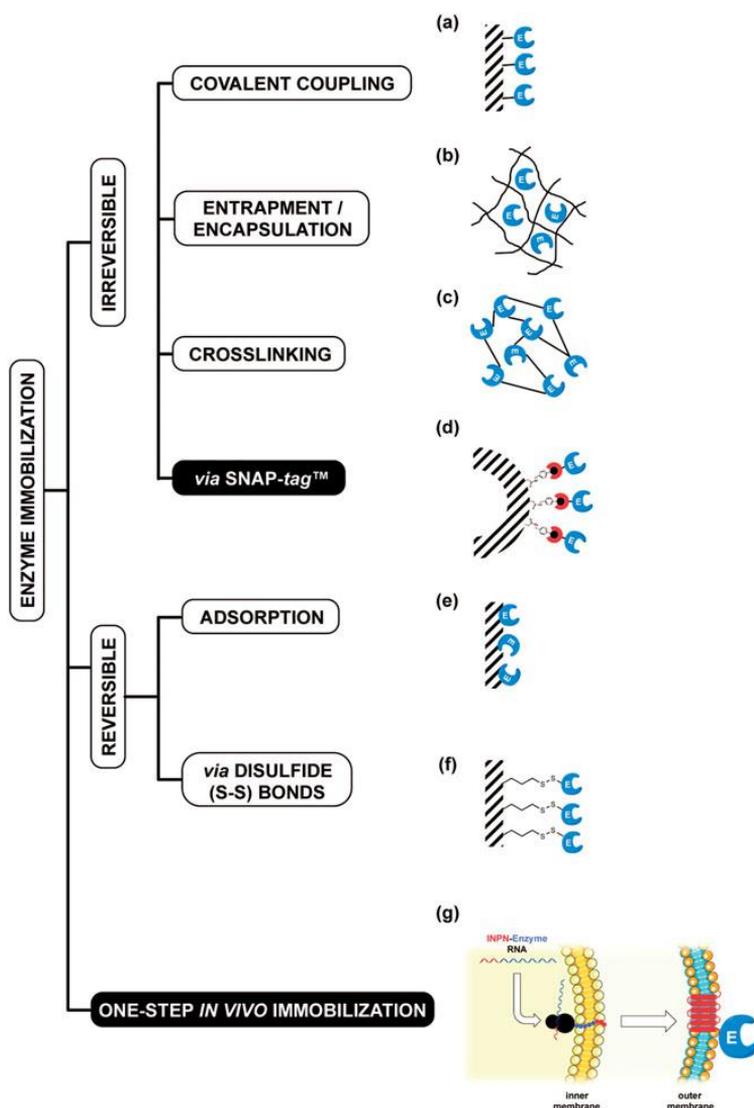
The term *immobilised enzymes* refers to 'enzymes physically confined or localised in a certain defined region of space with retention of their catalytic activities, and which can be used repeatedly and continuously'<sup>1</sup>. The immobilisation of enzymes on solid supports is historically very important for overcoming their general instability in harsh operational conditions and their low shelf-life, as well as the need for their recycling more times<sup>2</sup>. Furthermore, the physical separation of the biocatalyst from the reaction mixture avoids the protein contamination of the products. Although a reduction in reaction rates sometimes occurs, because the enzyme cannot mix freely with the substrate or a particular conformational change is needed for the biocatalyst efficiency, there are many examples of increased activity and stability of immobilised enzymes<sup>3</sup>. Many chemical or physical methods for the enzyme immobilisation are currently available, from the physical adsorption to the covalent coupling on supports (Figure 1(a–c,e–f))<sup>3–6</sup>. Recently, the use of *protein-tags* based on an engineered version of the human O<sup>6</sup>-alkyl-guanine-DNA-alkyl-transferases (hAGT) is an effective alternative for the covalent immobilisation of proteins and enzymes (Figure 1(d))<sup>7–9</sup>. AGTs (or AGTs, MGMTs; E.C.: 2.1.1.63) are DNA repair enzymes, which irreversibly transfer the alkyl group from the damaged DNA containing O<sup>6</sup>-alkyl-guanines to their cysteine residue in the active site<sup>10–13</sup>. In 2003, Johnsson and his group demonstrated that most enzymes of this class display relatively low substrate specificity, making them reactive also with free O<sup>6</sup>-benzyl-guanines (O<sup>6</sup>-BG)

nucleobases<sup>14</sup>. This led to the development of the so-called SNAP *tag*<sup>TM</sup> technology, which uses derivatives of O<sup>6</sup>-BG potentially conjugated with an unlimited number of chemical groups<sup>15–18</sup>. This system allows the immobilisation on O<sup>6</sup>-BG-derivatised surface of the protein expressed in fusion with the SNAP-*tag*<sup>18</sup> (Figure 1(d)). However, all these approaches mainly depend on the high cost due to the isolation and purification of the biocatalysts. This limitation can be easily overcome by the heterologous expression of enzymes and their *in vivo* direct immobilisation on the surface of bacterial hosts, by the utilisation of transmembrane protein domains, as the ice nucleation protein (INP) of the Gram-negative bacterium *Pseudomonas syringae* (Figure 1(g))<sup>19,20</sup>. This protein is composed of an N-terminal domain (N, 175 residues) structurally separated from a C-terminal domain (C, 49 residues) by a repetitive central domain<sup>21</sup>. Both domains play a role in the anchoring of proteins to the outer membrane<sup>21</sup>. The use of INP as *anchoring carrier* is considered of great interest in biotechnological applications, ranging from the development of bacterial cell surface display systems for vaccine delivery to the fabrication of whole-cell biocatalysts and biosensors<sup>22–24</sup>. The N-terminal domain of INP (INPN) was recently and successfully used for the one-step procedure immobilisation (Figure 1(g))<sup>22–29</sup>. Moreover, Capasso et al. demonstrated that the amount of a thermostable carbonic anhydrase<sup>30–33</sup> fused to the INPN domain and expressed on the bacterial cell surface had a hydratase activity similar to that of the enzyme covalently immobilised onto magnetic nanoparticles<sup>30,3</sup>. Here, we introduce a novel *protein-tag* system, (hereinafter

**CONTACT** Clemente Capasso  clemente.capasso@ibbr.cnr.it; Giuseppe Perugino  giuseppe.perugino@ibbr.cnr.it  Department of Biology Agriculture and Food Sciences, Institute of Bioscience and BioResources – National Research Council of Italy, Via Pietro Castellino 111, Naples 80131, Italy  
\*These authors equally contributed to this work.

 Supplemental data for this article can be accessed on the publisher's website.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Examples of enzyme immobilisation methods. Among the traditional methods for the irreversible immobilisation (a–c, e–f) of an enzyme (E, in blue), the recent introduction of the SNAP-tag™ technology (d, red semicircle) allowed an indirect immobilisation of protein of interest. The *one step in vivo immobilisation* of an enzyme (g) is possible when it is expressed as fusion protein with the N-terminal domain of the ice nucleation protein (INPN, in red). Because of the presence of a peptide leader upstream the coding sequence, the nascent polypeptide is translocated to the cell OM by the anchoring transmembrane INPN domain, leading to the immobilisation and exposition of the biocatalyst outside the cell.

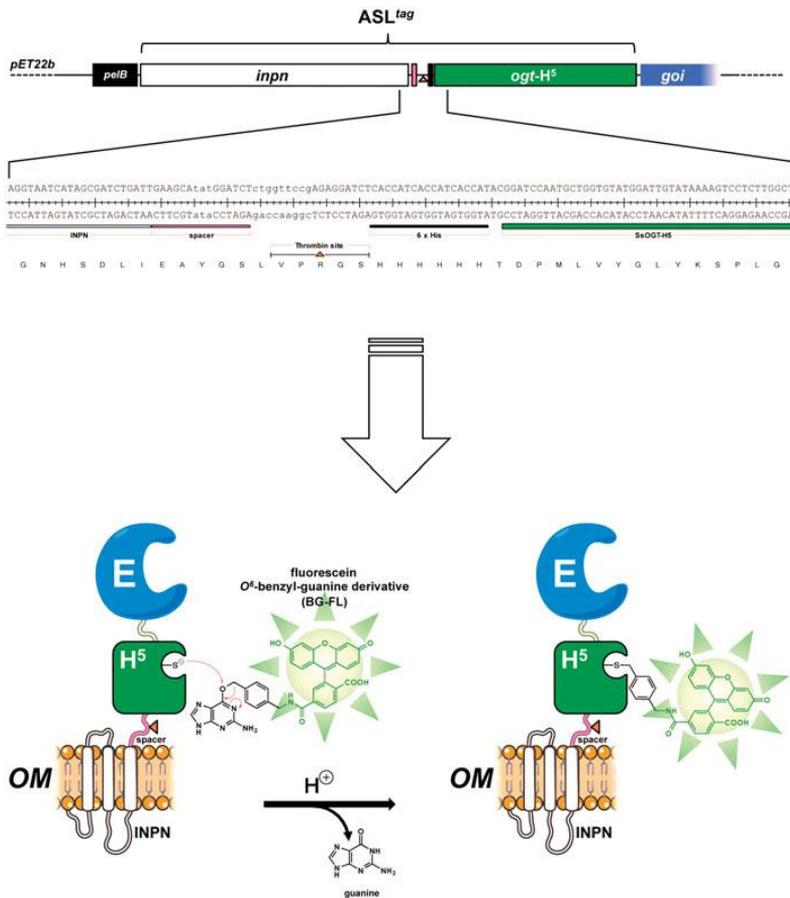
anchoring-and-Self-Labeling-protein-tag or ASL<sup>tag</sup>), which simultaneously allows the *in vivo* immobilisation of the enzyme of interest on the *E. coli* surface and its quantitative determination (Figure 2). The ASL<sup>tag</sup> system is formed by the INPN domain fused to an engineered and thermostable variant of the alkylguanine-DNA-alkyl-transferase (H<sup>5</sup>) from the hyperthermophilic archaeon *Halofolobus solfataricus*<sup>35,36</sup>. This enzyme was extensively characterised, suggesting its biotechnological role as thermostable alternative to the commercial SNAP-tag™ and its utilisation as protein-tag for heterologous expression of proteins of interest in *E. coli* and, for the first time, in thermophilic organisms as *Thermus thermophilus* and *Sulfolobus islandicus*<sup>35–37</sup>. Thus, using the substrate of H<sup>5</sup>, a fluorescein derivative of the O<sup>6</sup>-BG (Figure 2), we successfully estimated the expression of the ASL<sup>tag</sup> in *E. coli* cells, *in vitro gel-imaging* techniques, as well as by *in vivo* fluorescent

microscopy. Furthermore, we demonstrated that the activity and the stability of the enzymes of interest (SspCA, the  $\alpha$ -carbon anhydrase from *Sulfurihydrogenibium yellowstonense*; and Ss $\beta$ GI the  $\beta$ -glycoside hydrolase from *S. solfataricus*) fused to the ASL<sup>tag</sup> and exposed on the surface of *E. coli* cells were not affected in the presence of this novel protein-tag.

## 2. Materials and methods

### 2.1. Reagents

All DNA restriction and modification enzymes and the fluorescent substrate for the OGT activity (SNAP-Vista Green™, hereinafter BC FL) were purchased from New England Biolabs (Ipswich, MA). Molecular biology kits for the plasmid preparations and DNA g



**Figure 2.** The ASL<sup>tag</sup> protein. The ASL<sup>tag</sup> gene is composed by the *inpn* ORF (in white) in frame fused to the *ogtH*<sup>5</sup> gene (in green) in the pET22b expression vector. The *tag* can be further fused to a gene of interest (*goi*, in blue), for a one step procedure of the expression and immobilisation of an enzyme (E). The presence of the *inpn* moiety allows the quantitative estimation of the yield of E by the irreversible alkyl-transferase assay using a fluorescent O<sup>6</sup>-benzyl-guanine derivative (BG-FL) between *inpn* and *ogtH*<sup>5</sup>, a spacer (in pink), a thrombin cleavage site (shown as an orange triangle) and a 6 × His-tag (in black) were inserted, for the easy separation and purification of any H<sup>5</sup>-E fusion protein.

extractions (NucleoSpin<sup>®</sup> Gel and PCR Clean-up<sup>®</sup>) were from Macherey-Nagel GmbH (Germany); Lyophilised Thrombin Protease from GE Healthcare (Illinois, US). Eurofins Genomics (Germany) performed the oligonucleotides synthesis and the DNA sequencing service.

## 2. DNA constructs

To obtain the pET-ASL<sup>tag</sup> construct, we replaced the  $\alpha$ -carbonic dehydratase (*SspCA*) gene with the *ogtH*<sup>5</sup> gene in the previously described vector pET-22b/INPN-*SspCA*<sup>30</sup>. By the latter and the pQE-*ogtH*<sup>5</sup> plasmid as template, the DNA fragments relative to the INPN domain and H<sup>5</sup> were respectively amplified with the INPN- and H<sup>5</sup>-Fwd/Rev oligonucleotides pairs (listed in Table 1), under the following conditions: an initial denaturation at 95.0 °C for 5 min, 30 cycles of 30 s at 95.0 °C, 30 s at 50.0 °C and 30 s at 72.0 °C, followed by a final extension of 5 min at 72.0 °C. DNA products were fused to each other in a further PCR amplification, taking advantage of the total complementarity of the INPN-Rev to the H<sup>5</sup>-Fwd oligonucleotide, obtaining the final ASL<sup>tag</sup> DNA fragment. Subsequently, this fragment and the pET-22b/INPN-*SspCA* vector were digested with Hind III and Xba I restriction endonucleases, gel-purified, and ligated. The ligation mixture was used to

transform the *E. coli* DH5 $\alpha$  strain, and positive colonies were confirmed by colony PCR and DNA sequencing.

The same cloning strategy was used to achieve the pET-ASL<sup>tag</sup>-*SspCA* construct: the ASL<sup>tag</sup> DNA fragment obtained this time using the H<sup>5</sup>-Rev2 oligonucleotide was used as template to further fuse to the *SspCA* gene (obtained by amplification with *SspC* Fwd/Rev oligonucleotide pairs, Table 1). Again, positive colonies after ligation and transformation were confirmed as above.

Finally, the pET-ASL<sup>tag</sup>-*lacS* plasmid preparation started with the achievement of the DNA fragment relative to the *ogtH*<sup>5</sup> gene fused to the  $\beta$ -glycoside hydrolase from the thermophilic archae *Sulfolobus solfataricus* (*lacS*), obtained by the Hind III/BamH I digestion from the pQE-*ogtH*<sup>5</sup>-*lacS* plasmid<sup>36</sup>. The pET-ASL<sup>tag</sup> plasmid was similarly digested to obtain the pET22 recipient for the first ligation/transformation round. Positive blue colonies were selected by the hydrolase activity of the *lacS* gene product (S $\beta$ Gly) on the Ampicillin selective Luria-Bertani (LB) Agar plate supplemented with 5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucopyranoside (X-Glc), and the insertion confirmed using PCR analysis. The intermediate vector was further digested with BamH I, treated with Alkaline Phosphatase Calf Intestinal (CIP) and ligated to the BamH I/BamH I INPN DNA fragment, derived from the digestion of the pET-ASL<sup>tag</sup> plasmid. In this case, positive blue colonies were

**Table 1.** Oligonucleotides used in this study.

Oligonucleotide	Sequence
INPN-Fwd	5'-TAATACGACTCACTATAGGG-3'
INPN-Rev	5'-GGTGATGGTGAGATCCTCTCGGAACCAAGAGATCCATAGGCTTCAATCAGATCGC-3'
H <sup>5</sup> -Fwd	5'-GCCGATCTGATTGAAGCCTATGGATCTCTGGTCCGAGAGGATCTCACATCACC-3'
H <sup>5</sup> -Rev	5'-TCACCTTCATATGACCATTATGTTCTCGCTACCATAAGCTTCTGTCGACGGTACCTCGAGTTCTGG-3'
H <sup>5</sup> -Rev2	5'-ATTGAGCAACTGACTGAAATGCC-3'
SspCA-Fwd	5'-CCAGAACTCGAGTACCCTCGACAGAGGCTTATGGTAGCGAACTGAATGGTCATATGAAGGTGA-3'
SspCA-Rev	5'-CTAGTTATTGCTACGGGT-3'

3-Amp-X-Glc agar plates were confirmed by Pst I restriction enzyme digestion analysis.

### 3. Determination of the protein expression by a fluorescent assay based on the H<sup>5</sup> activity

All constructs were used to transform *E. coli* BL21(DE3) cells. Cultures were grown at 37.0 °C in LB selective medium supplemented with 100.0 mg/L ampicillin and 30.0 mg/L chloramphenicol; expression was induced with 1.0 mM isopropyl-thio- $\beta$ -D-lactoside (IPTG) when an absorbance value of 0.5–0.6 A<sub>600 nm</sub> was reached, or in the ZY auto-induction medium (AI)<sup>38</sup>, supplemented with the same selective antibiotics. After an overnight incubation, whole cells were collected and assayed by using the BG-FL fluorescent H<sup>5</sup> substrate previously described<sup>35–37,39</sup> for a qualitative measurement of the protein expression, an aliquot of 0 mL of cells was centrifuged at 4000  $\times$  g and the pellet was resuspended in 50.0  $\mu$ L of 5.0  $\mu$ M BG-FL in phosphate-buffered saline (PBS 1 $\times$ , 20.0 mM phosphate buffer, NaCl 150.0 mM, pH 7.3). After an incubation for 2.0 h at 37.0 °C, reactions were stopped by denaturing the samples in Laemmli Buffer 1 $\times$  and directly loaded on SDS-PAGE, followed by *gel-imaging* on a VersaDoc 300™ system (Bio-Rad), by applying a blue LED/530 bandpass filter as excitation/emission parameters, respectively. Finally, the fluorescence intensity of each band was normalised to the intensity of the signal obtained from the Coomassie Blue staining analysis.

For a quantitative determination of the expression, whole cells were opportunistically diluted to achieve an OD<sub>600 nm</sub> of 1.0. By following the same above-mentioned assay, three different volumes of whole cells were loaded *in the same gel* with defined amounts (in the range of 0.8–50.0 pmols) of purified free H<sup>5</sup> protein, after the reaction on BG-FL in the same conditions. The obtained values of the relative fluorescence as a function of the purified loaded H<sup>5</sup> were fitted by a linear equation, whose parameters were then used for the estimation of the amount of expressed H<sup>5</sup>-derived fusion proteins, assuming that the activity of the H<sup>5</sup> moiety in the fusions is not affected by the presence of the other protein partner(s). Given that the concentration of *E. coli* cell cultures of 1.0 OD<sub>600 nm</sub> is ca. 8.0  $\times$  10<sup>8</sup> cells/mL, and the amount of wet cells is 7 g/L, it is possible to calculate the yield of expressed proteins in terms of pmol/mg of cells<sup>40</sup>.

### 4. Membranes fractionation

The *E. coli* outer (OM) and inner (IM) membranes were purified following a procedure previously described<sup>27</sup>. Briefly, harvested bacterial cells were resuspended 1:20 (g/mL) in 25.0 mM Tris/HCl buffer, pH 8.0 and disrupted by sonication on ice (10 cycles of 30 s on/off treatment). Cell extract was centrifuged at 20,000  $\times$  g for 1.0 h, and the supernatant containing the cytoplasmic fraction was discarded. Both IM and OM fractions were recovered in the pellet and resuspended in 20.0 mL of PBS 1 $\times$ ,

containing 0.01 mM MgCl<sub>2</sub> and 2.0% Triton X-100. After incubation at room temperature for 30.0 min, the solution was centrifuged as described above. Then, a defined amount of the OM fraction obtained was assayed for the H<sup>5</sup> activity, leading to the quantitative determination of the total amount of fusion protein, as previously described.

### 2.5. Thrombin assay

The ASL<sup>tag</sup> on the OM was cleaved with the Thrombin Protease, in order to separate the H<sup>5</sup> moiety from the INPN transmembrane domain. A suspension of *E. coli* whole cells was gently centrifuged at 3000  $\times$  g for 10.0 min at 4.0 °C and then resuspended in a 10-fold volume of PBS 1 $\times$ . The sample was incubated with 30.0 U of Thrombin Protease at 25.0 °C overnight under gentle agitation in the presence of 5.0  $\mu$ M of BG-FL, followed by the same centrifugation. Cells and supernatants were separately loaded on SDS-PAGE and analyzed by *gel-imaging* and Coomassie staining as described.

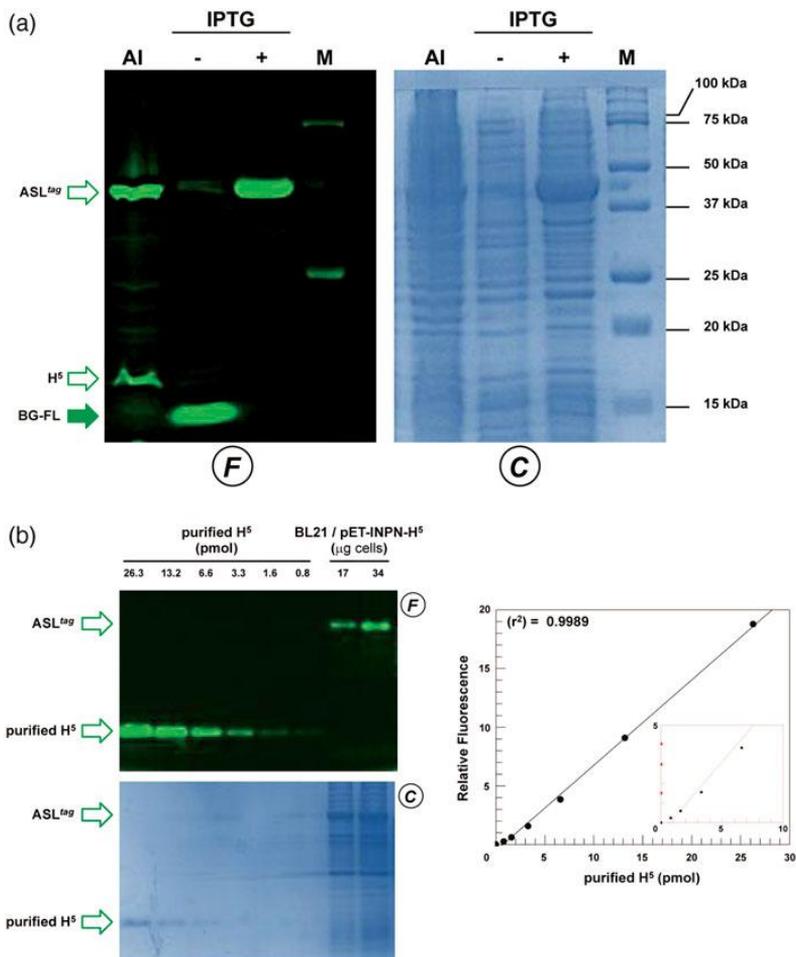
### 2.6. Microscopy analysis of *E. coli* cells

For *in vivo* imaging, *E. coli* BL21(DE3) cells transformed with pE22b/INPN-SspCA or pET-ASL<sup>tag</sup> plasmids were IPTG-induced and grown overnight at 37.0 °C and finally diluted until OD<sub>600 nm</sub> of 1.0. An amount of cells corresponding to a volume of 1.0 mL of the culture was washed twice in PBS 1 $\times$  and finally resuspended in 50.0  $\mu$ L of the same buffer supplemented with 5.0  $\mu$ M of the BG-FL substrate. After an incubation at 37.0 °C for 30.0 min, cells were washed twice, resuspended and again incubated for 30.0 min at 37.0 °C, to allow the external diffusion of the unreacted substrate. Labelling was first verified by fluorescent *gel-imaging* on SDS-PAGE and then spotted on poly-L-lysine coated slides for microscopy analysis.

Images were collected using a DM6 fluorescence microscope and a Hamamatsu camera under the control of Leica LAS AF 6000 software. Excitation and emission wavelengths used suitably for AlexaFluor488 dye were  $\lambda_{ex}$  = 490 nm;  $\lambda_{em}$  = 525 nm, respectively.

### 2.7. $\beta$ -glycoside hydrolase assay

The  $\beta$ -glycoside hydrolase assay was performed as previously described<sup>41</sup> at different temperatures in 50 mM sodium phosphate buffer at pH 6.5, in the presence of 2 Np- and 4 Np-Glc substrate at 5.0 mM final concentration. OM fractions containing ASL<sup>tag</sup> and relative fusion proteins amounts ranging from 1.0 to 5.0  $\mu$ g were used in each assay. For the correction of the spontaneous hydrolysis of the substrates, all the reactants except the enzyme (clarified mixture), was taken into account. The enzymatic activity was calculated on the basis of the molar extinction coefficient ( $\epsilon_{M}$ ) value of 2- and 4-nitrophenol in 50 mM sodium phosphate buffer pH 6 at different temperatures, as previously reported<sup>41</sup>. We defined as one unit of enzyme activity the amount of enzyme hydrolyzing 1.0  $\mu$ mol of substrate in 1.0 min, under the above described conditions.



**Figure 3.** The ASL<sup>tag</sup> expression in *E. coli*. (a) *E. coli* BL21(DE3) strain transformed with the pET-ASL<sup>tag</sup> plasmid was grown in IPTG-induced or in auto-inductive medium (AI). After the *in vivo* OGT assay, a defined amount in micrograms of whole cells at OD<sub>600nm</sub> of 1.0<sup>40</sup> was directly loaded on SDS-PAGE, followed by *gel-ima* *g* fluorescence (F) and Coomassie staining (C) analyses. Open and closed green arrows indicate fluorescent signals of free H<sup>5</sup> or H<sup>2</sup>-based fusion proteins, and the BG-FL substrate, respectively. M: molecular weight marker. (b) Quantitative estimation of the ASL<sup>tag</sup> expression: defined amount of cells and purified H<sup>5</sup> protein (pmols) were loaded and analyzed on a SDS-PAGE (on the left). Fluorescent values obtained from H<sup>5</sup> were fitted in a linear plot (on the right), as described in Materials and Methods. Obtained parameters allowed the quantitative determination of the amount of ASL<sup>tag</sup> in *E. coli* cells and shown in Table 2.

**Table 2.** Quantitative estimation of the heterologous expression of ASL<sup>tag</sup> and native fusion proteins in *E. coli* BL21(DE3) whole wet cells.

	MW <sup>a</sup> (kDa)	ASL <sup>tag</sup> -E ratio	Fusion yield <sup>b</sup> (pmol H <sup>5</sup> /mg)	E yield <sup>c</sup> (μg/mg)	(r <sup>2</sup> ) <sup>d</sup>
iL <sup>tag</sup>	42.0	–	152.5 ± 15.7	–	0.9977
iL <sup>tag</sup> -SspCA	70.2	1:1	58.4 ± 11.6	1.54 ± 0.30	0.9986
iL <sup>tag</sup> -SsβGly	98.8	4:1	30.6 ± 13.2	7.2 ± 3.1 <sup>e</sup>	0.9980

<sup>a</sup>calculated from the primary structure.

<sup>b</sup>On the basis of the H<sup>5</sup> activity on fluorescent *gel-imaging* analysis (see Materials and Methods).

<sup>c</sup>the amount of the immobilized enzyme without the ASL<sup>tag</sup>.

<sup>d</sup>Correlation coefficient of the linear curve obtained from the H<sup>5</sup> values of fluorescence as a function of the amount of the loaded protein.

<sup>e</sup>the tetrameric form of the catalytically active SsβGly enzyme.

## 8. Carbonic anhydrase assay

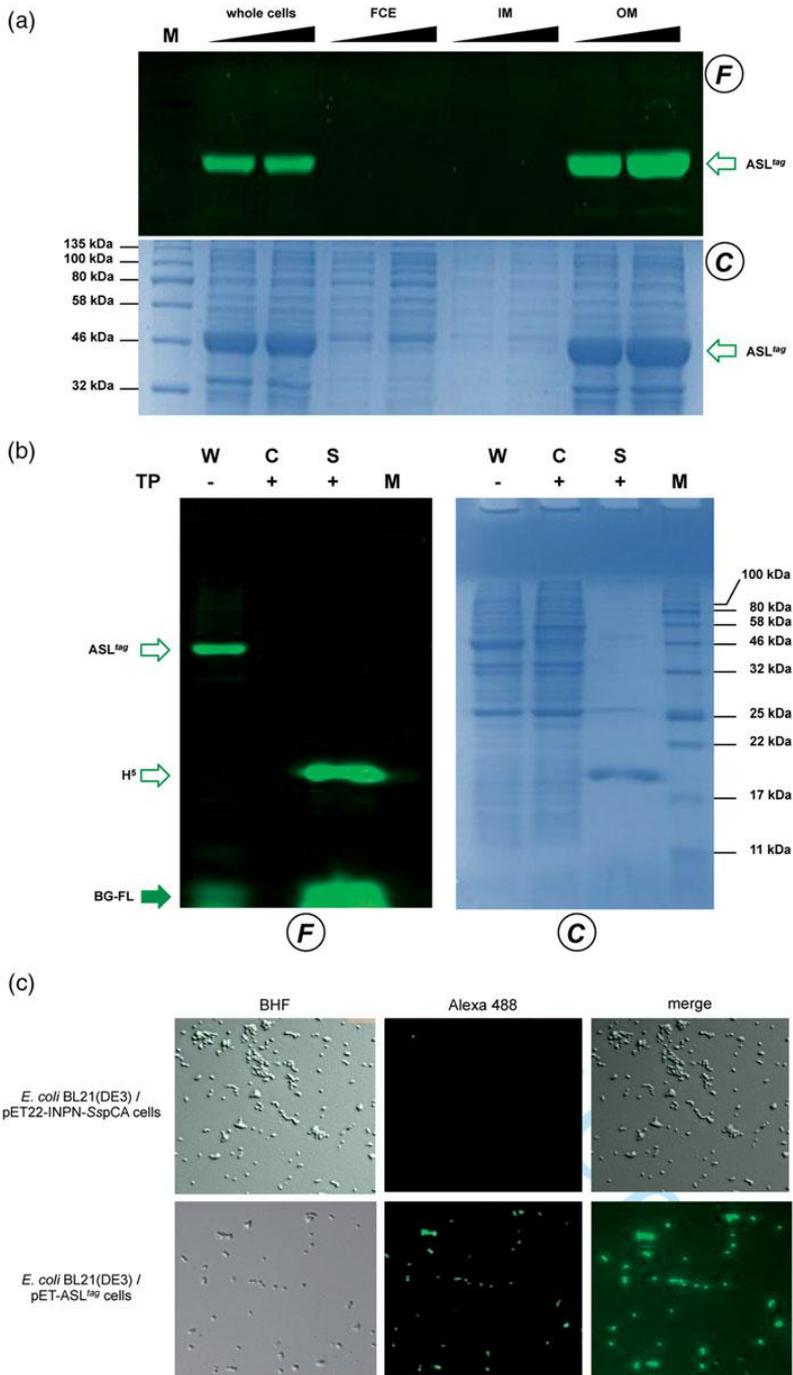
A activity assay was a modification of the procedure described by Capasso *et al.*<sup>33</sup>. Briefly, the hydratase assay was performed at 37°C using CO<sub>2</sub> as substrate following the pH variation due to the catalyzed conversion of CO<sub>2</sub> to bicarbonate. Bromothymol blue was used as pH indicator. The production of hydrogen ions during

the CO<sub>2</sub> hydration reaction lowers the pH of the solution leading to a colour transition of the dye. The time required for the colour change is inversely proportional to the amount of CA present in the sample. The Wilbur–Anderson units (WAU) were calculated according to the following definition: one WAU of CA activity defined as the ratio  $(T_0 - T)/T$ , where  $T_0$  (the time needed for the pH indicator color change for the uncatalyzed reaction) and  $T$  (the time needed for the pH indicator color change for the catalyzed reaction) are recorded as the time (in seconds) required for the pH to drop from 8.3 to the transition point of the dye (pH 6.8) in a control buffer and the presence of enzyme, respectively.

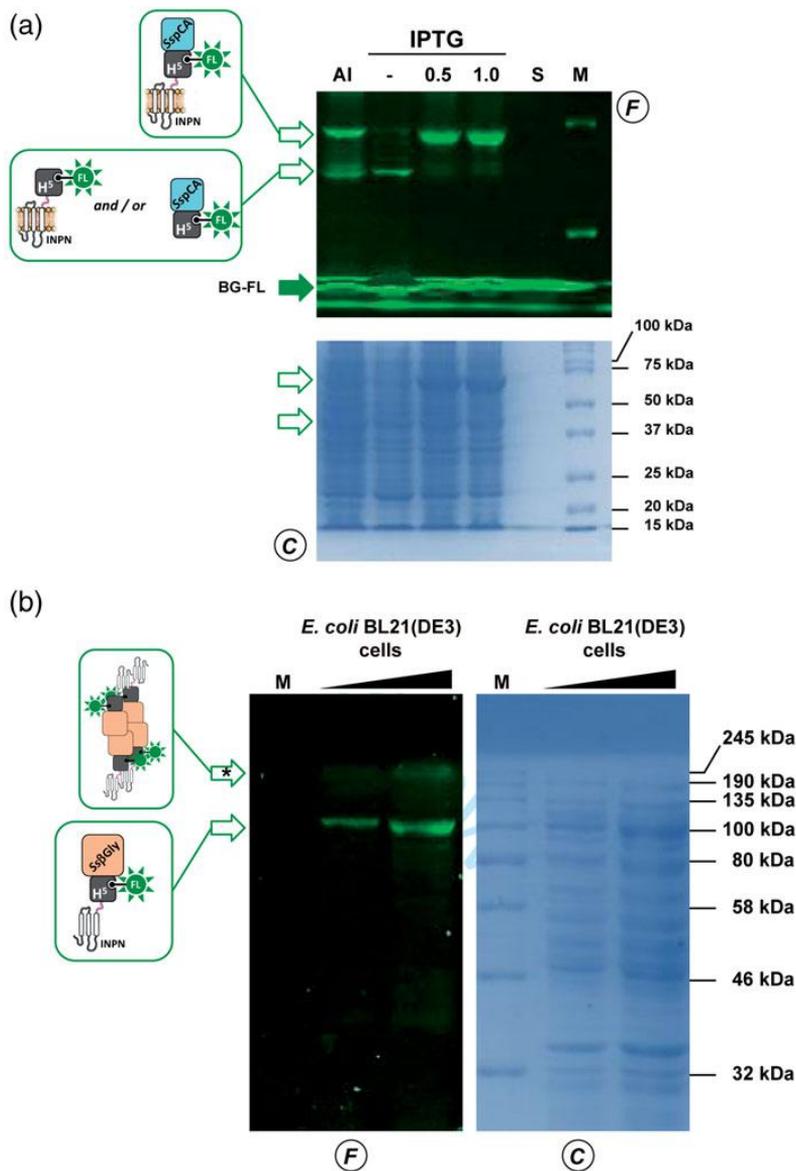
## 3. Results and discussion

### 3.1. Expression analysis and localisation of ASL<sup>tag</sup> in *E. coli*

The *in vivo* alkyl-transferase fluorescent assay of H<sup>5</sup> on whole bacterial cells is a useful method to determine the heterologous expression of this protein and/or relative fusion proteins in the mesophilic *E. coli* and in the thermophilic species *T. thermophil*



**Figure 4.** Localisation of ASL<sup>tag</sup> in *E. coli*. (a) gel-imaging and Coomassie staining analyses after SDS-PAGE of different loaded amount of the whole cells, the relative topoplasmic fraction (FCE), the inner (IM) and the outer membrane (OM) fractions. (b) Cleavage of ASL<sup>tag</sup> by the Thrombin protease (T) on whole cells (W). After the reaction during the protease treatment, cells were centrifuged, separating the supernatant (S) from the intact cells (C). (c) *E. coli* BL21(DE3) cells transformed with pET22b/INPN-SspCA (Top) or with pET-ASL<sup>tag</sup> (bottom) were incubated with BG-FL and then analyzed at fluorescence microscopy. Images show bright field (BHF), AlexaFluor488 (green) and merged images. All used symbols are described in Figure 3(a).



**Figure 5.** Heterologous expression of thermostable enzymes fused to ASL<sup>tag</sup>. SDS-PAGE analysis of the ASL<sup>tag</sup>-SspCA (a) and ASL<sup>tag</sup>-SsβGly (b) expression. (F) Fluorescent signals on the F panel correspond to bands whose molecular weights are attributable to the protein fusions shown in the schemes. (b) 1.0 mM IPTG induced cells gave a very high fluorescent band (marked by an asterisk), presumably corresponding to the tetrameric form of SsβGly linked to 4 × ASL<sup>tag</sup> units. All used symbols are described in Figure 3(a).

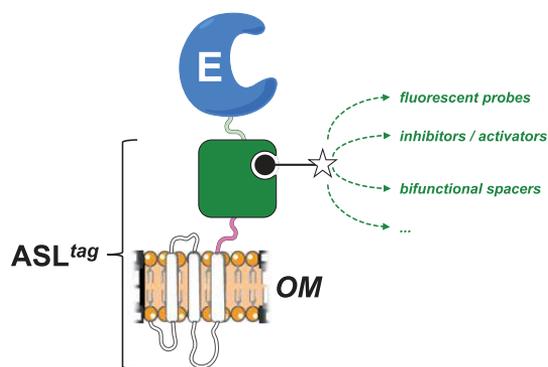
and *S. islandicus*, without any purification procedure<sup>35–37</sup>. The *in vivo* assay is possible since the bacterial cells are permeable to GT fluorescent substrates (as the commercially available SNAP-ista Green<sup>TM</sup> or SNAP-Cell TMR<sup>TM</sup>); and the catalytic activity at mesophilic temperatures (25–37 °C) of H<sup>5</sup> mutant is one order of magnitude higher than the SsOGT wild-type counterpart and comparable to the commercial SNAP-tag<sup>TM36</sup>. Despite GFPs utilisation, the covalent conjugation of H<sup>5</sup> with the benzyl-fluorophore moiety of the substrate (Figure 2) allows the denaturation of the whole cells and the direct loading of the samples on SDS-PAGE or the *gel-imaging* analysis, as described in the Materials and Methods<sup>35–37,39</sup>.

To evaluate the expression of the ASL<sup>tag</sup>, *E. coli* BL21(DE3)/pE ASL<sup>tag</sup> cells were grown in LB medium and induced with IPTG in the AI. In the latter case, although the expression level of the fusion protein was satisfactory, the presence of only the H<sup>5</sup> signal (ca. 20% of the total fluorescence intensity) in the *gel-imaging* analysis is difficult to rationalise (Figure 3(a)). Probably, during the expression of the protein (e.g., in the advanced stage of the growth) could occur interruptions or failures in the translocation process of the ASL<sup>tag</sup> on the outer membrane of *E. coli* with following breaking/cleavage events, especially in the spacer region between INPN and H<sup>5</sup> (Figure 2). On the other hand, after the IPTG induction, a strong and clear signal at the expected

molecular weight is visible, without any fragmentation (Figure 4a). This is an important result since, to date, the heterologous expression of SsOGT wild-type and relative variants in *E. coli* has been generally performed in the ABLE C strain, which keeps low the number of copies of the *ogt*-containing plasmids<sup>35,36</sup>. As proof, transformed *E. coli* BL21(DE3) strain with the pQE-gtH<sup>5</sup> plasmid<sup>36</sup> showed a very low expression of free H<sup>5</sup> (supplementary Figure S1), whereas the fusion with the INPN domain and the consequent translocation on the outer cell membrane made the H<sup>5</sup> expression possible in this strain. The comparative fluorescent analysis with a defined amount of free H<sup>5</sup> enzyme, allowed us to quantitatively measure the heterologous expression of ASL<sup>tag</sup> as pmol/mg of the whole wet cells (Figure 3(b) and Table 2). The assay on H<sup>5</sup> confirmed the anchoring function of the INPN trans-membrane protein: only the whole cells and the fraction containing the outer membrane displayed a fluorescent band corresponding to the ASL<sup>tag</sup> fusion protein, whereas the signal was missed in the lanes relative to the cytoplasmic and the inner membrane fractions (Figure 4(a)). Besides, the evidence of the INPN on the bacterial outer membrane was confirmed by treating the whole bacterial cells with the thrombin, too (Figure 2). A cleavage site for this protease was localised between the INPN domain and the H<sup>5</sup> moiety (Figure 2). The fluorescent signal corresponding to the MW of the H<sup>5</sup> protein was present only in the supernatant fraction when the protease was added at the same time with the G-FL substrate (Figure 4(b)). Finally, we checked the OGT activity of ASL<sup>tag</sup> in living cells by microscope analysis, upon the labelling with the fluorescent substrate. The obtained images of living *E. coli* BL21(DE3) cells showed a strong and specific fluorescent signal only in those transformed with the ASL<sup>tag</sup>-containing plasmid (Figure 4(c)), indicating that the fusion protein is stable and proficient to labelling. This data suggest that ASL<sup>tag</sup> is suitable for localisation and analysis of membrane proteins, and provide an opportunity for further *in vivo* analyses of SL<sup>tag</sup>-tagged proteins of interest under physiological conditions.

## 2. In vivo immobilisation of thermostable enzymes by ASL<sup>tag</sup>

As described in the literature, it has been demonstrated that the monomeric  $\alpha$ -carbonic anhydrase (SspCA) from the thermophilic bacterium *S. yellowstonense* can be actively expressed onto the outer membrane of *E. coli*<sup>30</sup>. Following this strategy, we realised a plasmid expressing the ASL<sup>tag</sup>-SspCA construct by inserting the *ogtH<sup>5</sup>* gene between the INPN and SspCA. The expression profile analyzed by following the H<sup>5</sup> activity on G-FL confirmed that in the AI multiple fluorescent signals are present (Figure 5(a)), mainly represented by the full-length SL<sup>tag</sup>-SspCA (70.2 kDa; ca. 65% of the total fluorescence intensity) and lower band (ca. 35%) closer to 37 kDa than 50 kDa (Figure 5(a)). This signal is compatible to the ASL<sup>tag</sup> (42.0 kDa) as well as the H<sup>5</sup>-SspCA moiety (46.0 kDa), suggesting that both translation interruption and translocation failure events in this growth conditions can be not excluded. Again, we detected an optimal achievement of the full-length fusion protein under IPTG-based expression in LB medium (ca. 95%; Figure 5(a)). In this condition and considering the INPN:H<sup>5</sup>:SspCA ratio as 1:1:1, the amount of the whole fusion protein expressed was estimated as ca. 60.0 pmol/g cells, which corresponds to ca. 1.5  $\mu$ g of the sole immobilised carbonic anhydrase per mg of cells (Table 2). Preliminary assays on ASL<sup>tag</sup>-SspCA indicated that the presence of H<sup>5</sup> does not hamper the hydratase activity of the



**Figure 6.** The biotechnological potential of the ASL<sup>tag</sup>. Any chemical group of interest (open star) conjugated to the BG-substrate (black closed circle) could be covalently bound to the H<sup>5</sup> moiety (in green) of the ASL<sup>tag</sup>. This enhances the potential use of an immobilised enzyme (E) on the *E. coli* surface (OM), making available to it a series of molecules, e.g., fluorescent probes and enzymatic activity modulators, or bi-functional groups for cascade reactions with other biocatalysts.

SspCA, if compared with that of the previously expressed INPN anchored enzyme<sup>30</sup> (data not shown).

ASL<sup>tag</sup> system was also tested with another thermostable enzyme, the  $\beta$ -glycosidase (Ss $\beta$ Gly) from the thermophilic archaea *S. solfataricus*<sup>42</sup>. We previously demonstrated that the cytoplasmic H<sup>5</sup>-Ss $\beta$ Gly fusion protein is stable and active for both the OGT and the  $\beta$ -glycosidase assays, suggesting that the presence of one enzyme does not interfere with the folding and activity of the other<sup>36</sup>. Interesting to note that anchoring this protein fusion to the bacterial outer membrane by the INPN domain is particularly challenging because Ss $\beta$ Gly is active only in its tetrameric form<sup>43,44</sup>. The presence of blue colonies on LB agar plate in the presence of a glucoside chromogen derivative (X-Glc), which is a preferred substrate of Ss $\beta$ Gly<sup>41</sup> but not of the *E. coli* LacZ ( $\beta$ -galactosidase enzyme) was a first and convincing indication of the oligomerisation of this thermostable enzyme. Although in this case the amount of the expressed fusion protein was lower than the above example (Figure 5(b) and Table 2), IPTG-induced *E. coli* BL21(DE3)/pE-ASL<sup>tag</sup>-Ss $\beta$ Gly cells displayed a fluorescent signal of expected molecular weight (98.8 kDa), corresponding to one monomer of Ss $\beta$ Gly fused to one ASL<sup>tag</sup> unit. However, a higher band clearly visible in the fluorescence analysis (marked with an asterisk in Figure 5(b)), out of the molecular weight marker range used: as for the cytoplasmic H<sup>5</sup>-Ss $\beta$ Gly fusion, it could suggest that it corresponds to a partially denatured part of the tetrameric form of ASL<sup>tag</sup>-Ss $\beta$ Gly (ca. 400.0 kDa), which is particularly resistant to thermal denaturation<sup>36,42,43,45</sup>. Finally, an amount of 0.24  $\mu$ g/mg of immobilised Ss $\beta$ Gly on the OM fraction (on the basis of the calculated H<sup>5</sup> pmol) was assayed on 2 Np- and 4 Np-Glc at three different temperatures. The result shows an activity of 12.6  $\pm$  0.7 and 8.8  $\pm$  0.4 (50  $^{\circ}$ C), 30.3  $\pm$  0.4 and 20.3  $\pm$  0.9 (60  $^{\circ}$ C), 51.5  $\pm$  0.9 and 30.8  $\pm$  1.5 U/mg (70  $^{\circ}$ C), respectively, whereas OM fraction containing the sole ASL<sup>tag</sup> did not result in any  $\beta$ -glucosidase activity, as expected (data not shown). These values are correctly related to the activity of Ss $\beta$ Gly<sup>45</sup>, clearly indicating that the formation of the quaternary structure on the *E. coli* OM occurs. However, since the 3D structure of this thermostable glycoside hydrolase showed that it is not laying on a surface<sup>43</sup>, we hypothesised an invagination of the *E. coli* outer membrane to allow the assembly of four units of the ASL<sup>tag</sup>-Ss $\beta$ Gly (Supplementary Figure S2).

## Conclusions and perspectives

In the present work, we introduced and demonstrated the utility of a novel *protein-tag*, composed by the N-terminal domain of the IP protein fused to a DNA repair enzyme. From our results, it is readily apparent that ASL<sup>tag</sup> offers: (i) an easy expression and *in vivo one-step procedure* of enzyme immobilisation on biological supports (e.g., *E. coli* outer membrane); (ii) significantly reduces the costs of the enzyme purification and those of the immobilisation support, allowing a direct exposition of the enzyme to the solvent<sup>20,30</sup>; (iii) an indirect labeling, by the reaction of a thermostable variant of the SNAP-tag<sup>TM</sup> (H<sup>5</sup>)<sup>46,47</sup>, which covalently links desired chemical groups conjugated to its benzyl-guanine substrate<sup>14,35</sup>. ASL<sup>tag</sup> favoured the expression of a monomeric protein (e.g., the thermostable SspCA) and an enzyme having a complex quaternary structure (e.g., the thermophilic SsβGly), without compromising their overall folding and enzymatic activity. Moreover, we showed that the utilisation of a fluorescein-derivative of the SNAP-tag led to the localisation and the quantitative determination of the yield of the expressed ASL<sup>tag</sup> and the relative fusion proteins (Table 2). On the other hand, despite the GFPs utilisation limited only to all fluorescence-based applications, the possibility to conjugate different groups to the BG- for the H<sup>5</sup> reaction<sup>36</sup> dramatically expands the biotechnological potential of this novel *pro-in-tag*. For example, it will be possible to modulate the activity of biocatalysts (by introducing inhibitors/activators), or connecting them with other proteins for the improvement of cascade reactions in the presence of bi-functional chemical groups (Figure 6).

## Acknowledgements

We are grateful to Maria Ciaramella for the constant and fruitful discussions during the execution of the experiments and the preparation of the manuscript. We are also grateful to Giovanni Del Lonaco for technical assistance.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work is supported by FIRB-Futuro in Ricerca 3FR12001G\_002 “Nematic” and by the grant “SMART GENERATION-Sistemi e tecnologie sostenibili per la generazione di energia-PON03PE\_00157\_1, OR3-Bio-sistemi di cattura ed utilizzazione della CO<sub>2</sub>”.

## ORCID

Roberto T. Supuran  <http://orcid.org/0000-0003-4262-0323>  
 Eleonora Capasso  <http://orcid.org/0000-0003-3314-2411>

## References

- Tosa T, Mori T, Fuse N, Chibata I. Studies on continuous enzyme reactions. I. Screening of carriers for preparation of water-insoluble aminoacylase. *Enzymologia* 1966;31:214–24.
- Zhou Z, Hartmann M. Progress in enzyme immobilization in ordered mesoporous materials and related applications. *Chem Soc Rev* 2013;42:3894–912.
- Mohamad NR, Che Marzuki NH, Buang NA, et al. An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. *Biotec Biotech Equip* 2015;29:205–20.
- Nguyen HH, Kima M. An overview of techniques in enzyme immobilization. *App Sci Conv Tech* 2017;26:157–63.
- Jakub ZJ, Meyer AS, Jesionowski T, Pinelo M. A general overview of support materials for enzyme immobilization: characteristics, properties, practical utility. *Catalysts* 2018;8:92.
- Sirisha VL, Jain A, Chapter JA. Nine – enzyme immobilization: an overview on methods, support material, and applications of immobilized enzymes. *Adv Food Nut Res* 2016;7:179–211.
- Engin S, Trouillet V, Franz CM. Benzylguanine thiol self-assembled monolayers for the immobilization of SNAP-tag proteins on microcontact-printed surface structure. *Langmuir* 2010;26:6097–101.
- Engin S, Fichtner D, Wedlich D, Fruk L. SNAP-tag as a tool for surface immobilization. *Curr Pharm Des* 2013;19:5443–8.
- Meldal M, Schoffelen S. Recent advances in covalent, site-specific protein immobilization. *F1000Res*. 2016;5. pii: F100 Faculty Rev-2303. eCollection 2016.
- Pegg AE. Repair of O6-alkylguanine by alkyltransferase. *Mutat Res* 2000;462:83–100.
- Daniels DS, Mol CD, Arvai AS, et al. Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *Embo J* 2000;19:1719–30.
- Daniels DS, Woo TT, Luu KX, et al. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nucl Acids Res* 2004;32:1714–20.
- Tubbs JL, Pegg AE, Tainer JA. 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* 2007;6:1100–15.
- Keppeler A, Gendrezig S, Gronemeyer T, et al. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotech* 2003;21:86–9.
- Gautier A, Juillerat A, Heinis C, et al. An engineered protein tag for multi-protein labeling in living cells. *Chem Biol* 2005;13:128–36.
- Gronemeyer T, Chidley C, Juillerat A, et al. Directed evolution of O6-alkylguanine-DNA alkyltransferase for application in protein labeling. *Prot Eng Des Sel* 2006;19:309–16.
- Yu Q, Griss R, Schena A, Johnsson K. Highly modular bioluminescent sensors for small molecules and proteins. *Mol Enz* 2017;589:365–82.
- Huber W, Perspicace S, Kohler J, et al. SPR-based interaction studies with small molecular weight ligands using hACh fusion proteins. *Anal Biochem* 2004;333:280–8.
- Cochet N, Widehem P. Ice crystallization by *Pseudomonas syringae*. *Appl Microbiol Biotechnol* 2000;54:153–61.
- Gurian-Sherman D, Lindow SE. Bacterial ice nucleation: significance and molecular basis. *FASEB J* 1993;7:1338–43.
- Graether SP, Jia Z. Modeling *Pseudomonas syringae* ice nucleation protein as a beta-helical protein. *Biophys J* 2000;78:1169–73.
- Samuelson P, Gunneriusson E, Nygren PA, Ståhl S. Display of proteins on bacteria. *J Biotechnol* 2002;96:129–54.
- Lee SY, Choi JH, Xu Z. Microbial cell-surface display. *Trends Biotechnol* 2003;21:45–52.
- Daugherty PS. Protein engineering with bacterial display. *Curr Opin Struct Biol* 2007;17:474–80.

5. Jung H-C, Park J-H, Park S-H, et al. Expression of carboxymethylcellulase on the surface of *Escherichia coli* using *Pseudomonas syringae* ice nucleation protein. *Enzyme Microb Technol* 1998;22:348–54.
5. Li L, Kang DG, Cha HJ. Functional display of foreign protein on surface of *Escherichia coli* using N-terminal domain of ice nucleation protein. *Biotechnol Bioeng* 2004;85:214–21.
7. Li Q, Yu Z, Shao X, et al. Improved phosphate biosorption by bacterial surface display of phosphate-binding protein utilizing ice nucleation protein. *FEMS Microbiol Lett* 2009; 299:44–52.
3. Li Q, Yan Q, Chen J, et al. Molecular characterization of an ice nucleation protein variant (inaQ) from *Pseudomonas syringae* and the analysis of its transmembrane transport activity in *Escherichia coli*. *Int J Biol Sci* 2012;8:1097–108.
9. Xu Y, Liu Q, Zhou L, et al. Surface display of GFP by *Pseudomonas syringae* truncated ice nucleation protein in attenuated *Vibro anguillarum* strain. *Mar Biotechnol* 2008;10: 701–8.
2. Del Prete S, Perfetto R, Rossi M, et al. A one-step procedure for immobilising the thermostable carbonic anhydrase (SspCA) on the surface membrane of *Escherichia coli*. *J Enz Inhib Med Chem* 2017;32:1120–8.
1. Nakagawa S, Shtaih Z, Banta A, et al. *Sulfurihydrogenibium yellowstonense* sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium*, *Sulfurihydrogenibium subterraneum* and *Sulfurihydrogenibium azorense*. *Int J Syst Evol Microbiol* 2005;55:2263–8.
2. Akdemir A, Vullo D, De Luca V, et al. The extreme-alpha-carbonic anhydrase (CA) from *Sulfurihydrogenibium azorense*, the fastest CA known, is highly activated by amino acids and amines. *Bioorg Med Chem Lett* 2013;23:1087–90.
3. Capasso C, De Luca V, Carginale V, et al. Biochemical properties of a novel and highly thermostable bacterial alpha-carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* YO3AOP1. *J Enzyme Inhib Med Chem* 2012;27: 892–789.
4. Perfetto R, Del Prete S, Vullo D, et al. Production and covalent immobilisation of the recombinant bacterial carbonic anhydrase (SspCA) onto magnetic nanoparticles. *J Enzyme Inhib Med Chem* 2017;32:759–66.
5. Perugino G, Vettone A, Illiano G, et al. Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J Biol Chem* 2012;287:4222–31.
36. Vettone A, Serpe M, Hidalgo A, et al. A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA alkyl-transferase by protein engineering. *Extremophiles* 2016;20:13.
37. Visone V, Han W, Perugino G, et al. In vivo and in vitro protein imaging in thermophilic archaea by exploiting a novel protein tag. *PLoS One* 2017;12:e0185791.
38. Studier FW. Protein production by auto-induction in high density shaking cultures. *Prot Expr Purif* 2005;41:207–34.
39. Miggiano R, Valenti A, Rossi F, et al. Every OGT illuminated... by fluorescent and synchrotron lights. *Int Mol Sci* 2017;18:2613–31.
40. Glazyrina J, Materne EM, Dreher T, et al. High cell density cultivation and recombinant protein production with *Escherichia coli* in a rocking-motion-type bioreactor. *Micro Cell Fact* 2010;9:42–52.
41. Perugino G, Trincone A, Giordano A, et al. Activity of hyperthermophilic glycosynthases is significantly enhanced at acidic pH. *Biochemistry* 2003;42:8484–93.
42. Moracci M, Capalbo L, Ciaramella M, Rossi M. Identification of two glutamic acid residues essential for catalysis in the beta-glycosidase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Protein Eng* 1996;9:1191–5.
43. Aguilar C, Sanderson I, Moracci M, et al. Crystal structure of the beta-glycosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: resilience as a key factor in thermal stability. *J Mol Biol* 1997;271:789–802.
44. Pouwels J, Moracci M, Cobucci-Ponzano B, et al. Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal-glycosidases. *Extremophiles* 2000; 4:157–64.
45. Moracci M, Trincone A, Perugino G, et al. Restoration of the activity of active-site mutants of the hyperthermophilic-glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophile. *Biochemistry* 1998; 37:17262–70.
46. Perugino G, Miggiano R, Serpe M, et al. Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nuc Ac Res* 2011;39:8801–16.
47. Rossi F, Morrone C, Massarotti A, et al. Crystal structure of thermophilic O6-alkylguanine-DNA alkyltransferase-derived self-labeling protein-tag in covalent complex with a fluorescent probe. *Bioch Biophys Res Comm* 2018;500:698–703.

## Thermostability enhancement of the $\alpha$ -carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* by using the anchoring-and-self-labelling-protein-tag system (ASL<sup>tag</sup>)

Concetta Del Prete<sup>a\*</sup>, Rosa Merlo<sup>a\*</sup>, Anna Valenti<sup>a</sup>, Rosanna Mattossovich<sup>a</sup>, Mosè Rossi<sup>a</sup>, Vincenzo Carginale<sup>a</sup>,  
Claudio T. Supuran<sup>b</sup> , Giuseppe Peruginò<sup>a</sup> and Clemente Capasso<sup>a</sup> 

<sup>a</sup>Department of Biology Agriculture and Food Sciences, Institute of Bioscience and BioResources – National Research Council of Italy, Naples, Italy; <sup>b</sup>Neurofarba Department, University of Florence, Polo Scientifico, Sesto Fiorentino Firenze, Italy

### ABSTRACT

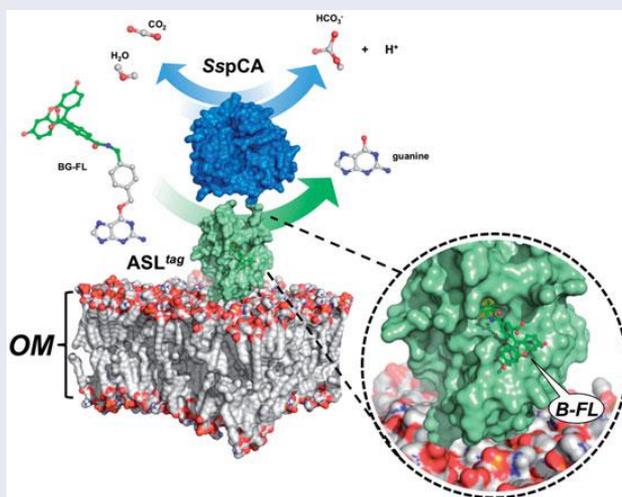
Carbonic anhydrases (CAs, EC 4.2.1.1) are a superfamily of ubiquitous metalloenzymes present in all living organisms on the planet. They are classified into seven genetically distinct families and catalyse the hydration reaction of carbon dioxide to bicarbonate and protons, as well as the opposite reaction. CAs were proposed to be used for biotechnological applications, such as the post-combustion carbon capture processes. In this context, there is a great interest in searching CAs with robust chemical and physical properties. Here, we describe the enhancement of thermostability of the  $\alpha$ -CA from *Sulfurihydrogenibium yellowstonense* (SspCA) by using the anchoring-and-self-labelling-protein-tag system (ASL<sup>tag</sup>). The anchored chimeric H<sup>2</sup>-SspCA was active for the CO<sub>2</sub> hydration reaction and its thermostability increased when the cells were heated for a prolonged period at high temperatures (e.g. 70 °C). The ASL<sup>tag</sup> can be considered as a useful method for enhancing the thermostability of a protein useful for biotechnological applications, which often need harsh operating conditions.

### ARTICLE HISTORY

Received 25 February 2019  
Revised 3 April 2019  
Accepted 5 April 2019

### KEYWORDS

Carbonic anhydrase;  
thermostability;  
*Sulfurihydrogenibium yellowstonense*;  
alkylguanine-DNA-alkyl-transferase; ASL<sup>tag</sup>



### 1. Introduction

The hydration/dehydration reaction involving carbon dioxide, water, bicarbonate, and protons ( $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+$ ) is a fundamentally important process for the planet and all its associated forms of life<sup>1–9</sup>. The dissolution of CO<sub>2</sub> in the aqueous phase develops carbonic acid (H<sub>2</sub>CO<sub>2</sub>), which is subject to an ionisation

reaction producing bicarbonate (HCO<sub>3</sub><sup>−</sup>), whereas this last species then generates carbonate (CO<sub>2</sub>) through a second dissociative reaction. These species are disseminated in the fluids of the all living organisms and are involved in a large number of physiological processes, such as some biosynthetic pathways, photosynthesis, respiration, pH homeostasis, secretion of electrolytes, etc.<sup>9–11</sup>.

**CONTACT** Clemente Capasso  [clemente.capasso@ibbr.cnr.it](mailto:clemente.capasso@ibbr.cnr.it); Giuseppe Peruginò  [giuseppe.peruginò@ibbr.cnr.it](mailto:giuseppe.peruginò@ibbr.cnr.it)  Department of Biology Agriculture and Food Sciences, Institute of Bioscience and BioResources – National Research Council of Italy, Via Pietro Castellino 111, Naples, 80131, Italy  
These authors equally contributed to this work.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

physiological pH, the naturally uncatalysed CO<sub>2</sub> hydration reaction as a catalytic constant ( $k_{\text{cat}}$ ) of 0.15 s<sup>-1</sup>, whereas the uncatalysed hydration shows a  $k_{\text{cat}}$  of 50.0 s<sup>-1</sup><sup>12,13</sup>. These values are typical of slow reactions and are not sufficient for accomplishing fast cellular physiological processes which support metabolic activities dependent on the dissolved inorganic carbon species (CO<sub>2</sub>, CO<sub>3</sub><sup>-</sup>, CO<sub>3</sub><sup>2-</sup>)<sup>9</sup>. Probably this is the reason why living organisms evolved a superfamily of ubiquitous metalloenzymes, the carbonic anhydrases (CAs, EC 4.2.1.1), which catalyse, and highly accelerate, the above-mentioned reactions, at a very high rate with respect to the non-catalysed reaction. CAs show kinetic constants  $k_{\text{cat}}$  varying from 10<sup>4</sup> to 10<sup>6</sup> s<sup>-1</sup> for the hydration reaction<sup>12,13</sup>.

Up to date, CA superfamily contains seven genetically distinct families (or classes), named  $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ -,  $\zeta$ -,  $\eta$ -, and  $\theta$ -CAs<sup>9,14,15</sup>, characterised by multiple transcript variants and protein isoforms, with different biochemical properties and specific tissue/organ and sub-cellular localisations<sup>7,9,12,16–19</sup>. Generally, only  $\alpha$ -class enzymes are present in the animals<sup>20,21</sup>, whereas  $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ - and  $\zeta$ -classes are found in plants and algae;  $\alpha$ - and  $\beta$ -CAs in fungi;  $\alpha$ -,  $\gamma$ -, and/or  $\eta$ -CAs in protozoans;  $\alpha$ -,  $\beta$ -, and  $\gamma$ -CA classes in bacteria<sup>7,19,15,22–25</sup>.

Studies carried out on the bacterial CAs concern two main aspects. They are considered an attractive and rather new drug target, because their inhibition affects the growth or virulence of many pathogens<sup>4,7,26–28</sup>. Furthermore, they are biocatalysts often used in biotechnological applications<sup>29,30</sup>, such as the post-combustion carbon capture process, artificial lungs, and biosensors<sup>31,32</sup>. Many such processes are characterised by conditions, which may be deleterious to an enzyme belonging to the mesophilic organisms<sup>25,33–45</sup>. In the field of biotechnology, there is a great interest in searching proteins with robust chemical and physical properties, which resist the hard conditions of industrial processes. In this context, our groups identified in the genome of the extreme thermophiles *Sulfurihydrogenibium yellowstonense* and *Sulfurihydrogenibium azorense* two CAs, indicated with the synonyms SspCA and SazCA, respectively. It has been demonstrated that these two CAs belong to the  $\alpha$ -CA class and showed an excellent activity as catalysts for the CO<sub>2</sub> hydratase reaction ( $k_{\text{cat}} = 10^5$ – $10^6$  s<sup>-1</sup>)<sup>30,46–52</sup>. Interestingly, the two extreme enzymes resulted to be highly thermostable, retaining an excellent catalytic activity when heated for a prolonged period at a temperature higher than 80 °C<sup>30,46–52</sup>. The X-ray tridimensional structures of the two proteins demonstrated that the high compactness of the dimeric structure, the higher content of secondary-structural elements, the increased number of charged residues on the protein surface, and the vast number of ionic networks with respect to the mesophilic counterparts, are the main structural elements responsible for the protein thermostability<sup>29,30</sup>. Moreover, Russo et al. reported the use of free SspCA in experiment of CO<sub>2</sub> absorption<sup>53</sup> demonstrating that it is an excellent candidate for the biomimetic capture of CO<sub>2</sub>. Subsequently, the necessity to use this biocatalyst repeatedly and continuously, led to the immobilisation of the recombinant SspCA on polyurethane foam (PU), a pre-polymer of polyethylene glycol<sup>54</sup>; onto supported ionic liquid membranes (SMLs), in order to realise a system able to selectively separate and transform CO<sub>2</sub><sup>55</sup>. Furthermore, the immobilisation onto magnetic support for recovering the biocatalyst from the reactor effortlessly and practically, for example through the use of a magnet, was also proposed for these thermostable CAs<sup>56</sup>. Unfortunately, these strategies may discourage the wide utilisation of enzymes in industrial applications because of the high costs connected to the biocatalyst production and purification, and the expenses for the preparation of the immobilisation support.

Thus, to overcome this limitation, a one-step immobilisation procedure has been proposed, which consists in the overexpression of SspCA directly onto the surface of bacterial hosts, by using the ice nucleation protein (INP) from the Gram-negative bacterium *Pseudomonas syringae*<sup>57</sup>.

In this article, we describe the improvement of the thermal stability of SspCA by using a novel protein-tag system, the ASL<sup>tag</sup><sup>58</sup>. The anchored SspCA was fused to the thermostable variant of the alkylguanine-DNA-alkyl-transferase (H<sup>5</sup>) from the hyperthermophilic archaeon *Sulfolobus solfataricus*. The chimeric H<sup>5</sup>-SspCA was efficiently overexpressed on the bacterial surface of *Escherichia coli*. The protonography technique showed that the neosynthesised H<sup>5</sup>-SspCA was active for the CO<sub>2</sub> hydration reaction. Even more intriguing, the chimeric H<sup>5</sup>-SspCA expression onto the bacterial surface resulted to be more stable with respect to the non-chimeric SspCA, when treated at high temperatures (50.0 and 70.0 °C) for a prolonged time. The ASL<sup>tag</sup> system may thus be considered as a brilliant strategy to further increase the thermostability of proteins to be used in biotechnological applications, in which a highly effective and thermostable catalyst is needed.

## 2. Materials and methods

### 2.1 Construction of vectors for surface fusion and H<sup>5</sup>-SspCA overexpression

The vector pET-22b/INPN-SspCA was used to produce the pE<sup>-</sup>ASL<sup>tag</sup>-SspCA vector, which overexpressed onto the bacterial surface the chimeric H<sup>5</sup>-SspCA. The pET-22b/INPN-SspCA and pE<sup>-</sup>ASL<sup>tag</sup>-SspCA vectors were prepared as described previously<sup>57,59</sup>. For overexpressing the chimeric H<sup>5</sup>-SspCA or SspCA on the bacterial cell surface, competent *E. coli* BL21 (DE3) cells were transformed with the above-mentioned constructs. They were grown at 37.0 °C and induced with 1.0 mM isopropyl-thio- $\beta$ -D-galactoside (IPTG) and 0.5 mM ZnSO<sub>4</sub> at an OD<sub>600</sub> of 0.6–0.7. After additional growth for 6 h, the cells were harvested by centrifugation and washed three times with PBS 1×. Aliquots of cells were resuspended in 25 mM Tris/HCl and used to determine the enzyme activity and for the preparation of the outer membrane fraction.

### 2.2 Carbonic anhydrase assay and SDS-PAGE

CA activity assay was a modification of the procedure described by Capasso et al.<sup>59</sup>. Briefly, the assay was performed at 0 °C using CO<sub>2</sub> as substrate and following the pH variation due to the catalysed conversion of CO<sub>2</sub> to bicarbonate. Bromothymol blue was used as the indicator of pH variation. The production of hydrogen ions during the CO<sub>2</sub> hydration reaction lowers the pH of the solution until the colour transition point of the dye is reached. The time required for the colour change is inversely related to the quantity of CA present in the sample. Wilbur-Anderson units (WAU) were calculated according to the following definition: or WAU of activity is defined as (T<sub>0</sub>–T)/T, where T<sub>0</sub> (uncatalysed reaction) and T (catalysed reaction) are recorded as the time (i seconds) required for the pH to drop from 8.3 to the transition point of the dye in a control buffer and in the presence of enzyme, respectively. Assay of the membrane-bound enzyme (H<sup>5</sup>-SspCA or SspCA) was carried out using an amount of whole cells or outer membranes ranging from 1.0 to 5.0 mg. Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE) was performed as described by Laemmli using 12%

els.<sup>60</sup> Samples were dissolved in buffer with 5%  $\beta$ -mercaptoethanol. The gel was stained with Coomassie blue and protein concentration was determined by Bio-Rad assay kit (Bio-Rad, Hercules, CA).

### 3 Protonography and his-tag Western blotting

To perform the protonography, wells of 12% SDS-gel were loaded with solubilised outer membranes having on their surface H<sup>5</sup>-SspCA or SspCA, and a solution of free SspCA (enzyme overexpressed as cytoplasmic protein and purified as described previously<sup>59</sup>). Samples were mixed with loading buffer without 2-mercaptoethanol and without boiling the samples, to solubilise dyes and avoid protein denaturation. The gel was run at 180 V until the dye front moved off the gel. Following the electrophoresis, the 12% SDS-gel was subject to protonography to detect the cytoplasmic SspCA, the surface-SspCA, and surface-H<sup>5</sup>-SspCA hydrolase activity on the gel as described by Del Prete et al.<sup>61,62</sup> and De Luca et al.<sup>63</sup>. To perform the Western-Blot, after a 12% (v/v) SDS-PAGE, the overexpressed cytoplasmic SspCA and the membrane-bound enzymes (SspCA and H<sup>5</sup>-SspCA) were also electrophoretically transferred to a PVDF membrane with transfer buffer (25 mM Tris, solubilised whole cells 192 mM glycine, 20% methanol) by using Trans-Plot SD Cell (Bio-Rad, Hercules, CA). His-tag Western blot was carried out using the Pierce Fast Western Blot Kit (Thermo Scientific, Waltham, MA). The blotted membrane has been placed in the wash blot solution Fast Western 1× Wash Buffer to remove transfer buffer. Primary antibody Working Dilution was added to the blot and incubated for 30.0 min at room temperature (RT) with shaking. After, the blot was removed from the primary antibody solution and incubated for 10.0 min with the Fast Western Optimized HRP Reagent Working Dilution. Subsequently, the membrane was washed two times in about 20 ml of Fast Western 1× Wash Buffer. Finally, the membrane was incubated with the Detection Reagent Working Solution and incubated for 3.0 min at RT and then developed with X-ray film.

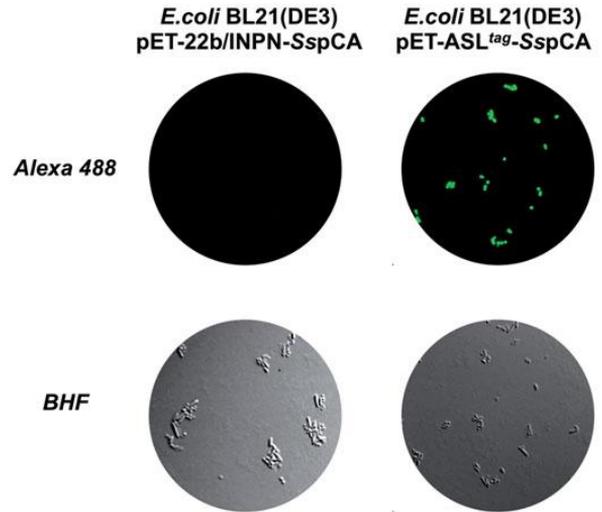
### 4 Determination of the H<sup>5</sup> activity by an *in vitro* and *in vivo* fluorescent assay

Whole overnight induced *E. coli* BL21(DE3) cells were collected and the expression of the H<sup>5</sup>-derived fusion proteins was analysed by an *in vitro* assay with the fluorescent SNAP-Vista Green<sup>TM</sup> substrate (New England Biolabs, Ipswich, MA; hereinafter BG-FL), as previously described<sup>58,64</sup>. The *in vivo* imaging was carried out as described by Merlo et al.<sup>58</sup>. Briefly, bacterial cells expressing the H<sup>5</sup>-SspCA onto cell surface were washed twice in PBS 1× and resuspended in 50.0  $\mu$ l of the same buffer supplemented with 0  $\mu$ M of the BG-FL. After incubation at 37.0 °C for 30.0 min, cells were washed twice, resuspended, and again incubated for 30.0 min at 37.0 °C, to allow the external diffusion of the reacted substrate. Images were collected using a DM6 fluorescence microscope and Hamamatsu camera under the control of Zeiss LAS AF 6000 software; excitation and emission wavelengths used suitably for AlexaFluor488 dye were  $\lambda_{ex}$  = 490 nm;  $\lambda_{em}$  = 525 nm, respectively.

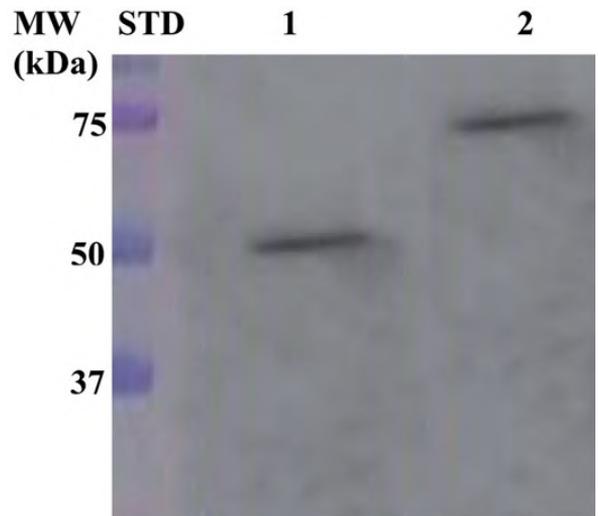
### 5 Outer membrane preparation

The bacterial outer membranes were fractioned by inner membranes as described previously by Del Prete et al.<sup>57</sup>. Briefly, 2.0 g

of harvested bacterial cells were resuspended and disrupted by sonication on the ice. Cell extract was ultracentrifuged to recover the total membrane fraction. The outer membrane fraction was purified resuspending the pellet in phosphate-buffered saline (PBS 1×) containing 0.01 mM MgCl<sub>2</sub> and 2% Triton X-100 and incubated at RT for 30.0 min to solubilise the inner membrane. The outer membrane fraction was then pelleted by ultracentrifugation at 120,000×g and used for further experiments.



**Figure 1.** Fluorescence microscopy of *E. coli* BL21(DE3) cells transformed with pET-22b/INPN-SspCA (left) or with pET-ASL<sup>tag</sup>-SspCA (right). The cells were incubated with BG-FL and then analysed by fluorescence microscopy. Images show bright field (BHF) and AlexaFluor488 (green). As expected, the fluorescence is only evidenced for the bacterial cell transformed with the ASL<sup>tag</sup> system.



**Figure 2.** Western Blot performed on the outer membrane purified from the whole bacterial cells. The anti His-tag antibody was raised against the C-terminus of His-tagged SspCA. Legend: Lane Std, molecular markers, M.W. starting from the top: 75.0, 50.0, and 37.0 kDa; Lane 1, anchored SspCA; Lane 2, anchored H<sup>5</sup>-SspCA.

## 6 Temperature stability studies

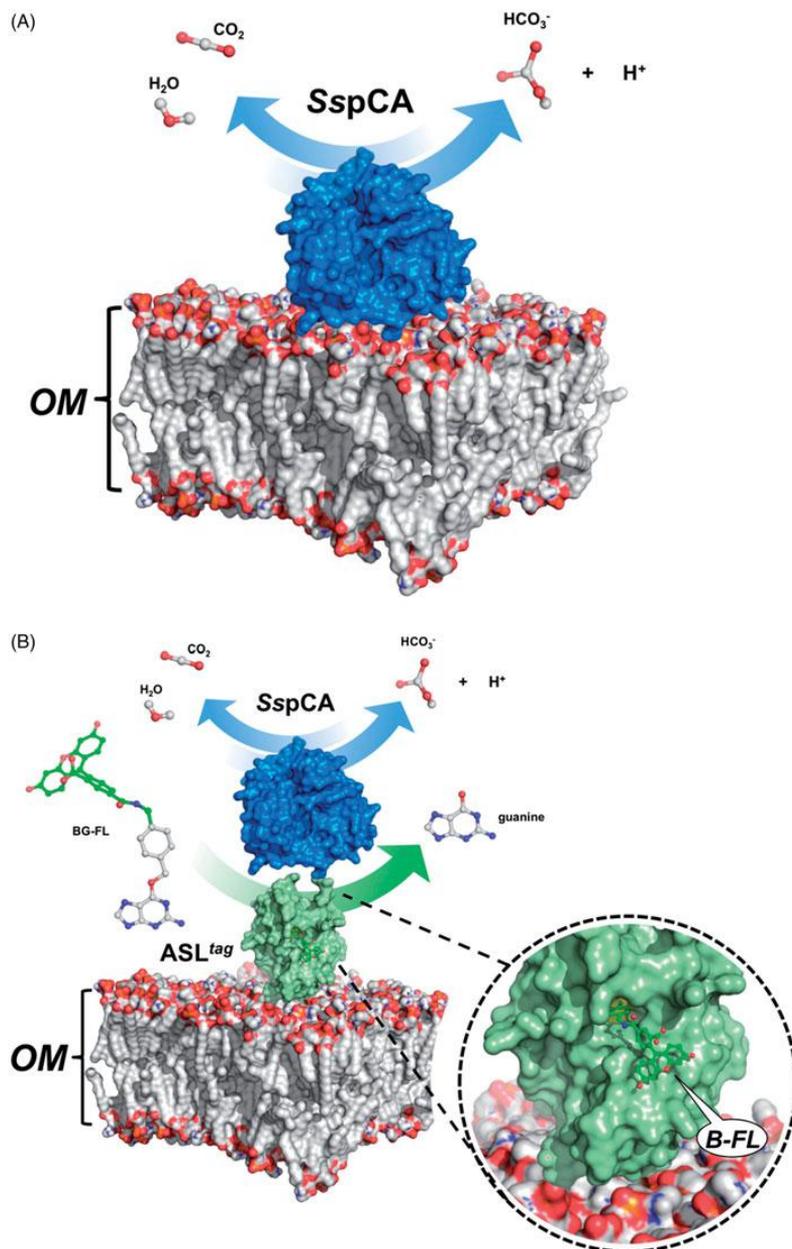
### 6.1 Thermostability

arterial cells (2.0g/20 ml) were incubated at 25.0, 50.0, and 70.0 °C for different time up to 24 h to compare the stability of the membrane-bound enzymes (SspCA and H<sup>5</sup>-SspCA) at the above-indicated temperatures. Cell membrane-bound enzyme aliquots were withdrawn at appropriate times and the residual activity was measured using CO<sub>2</sub> as the substrate. All data have been analysed using GraphPad Prism version 5.0 software (GraphPad

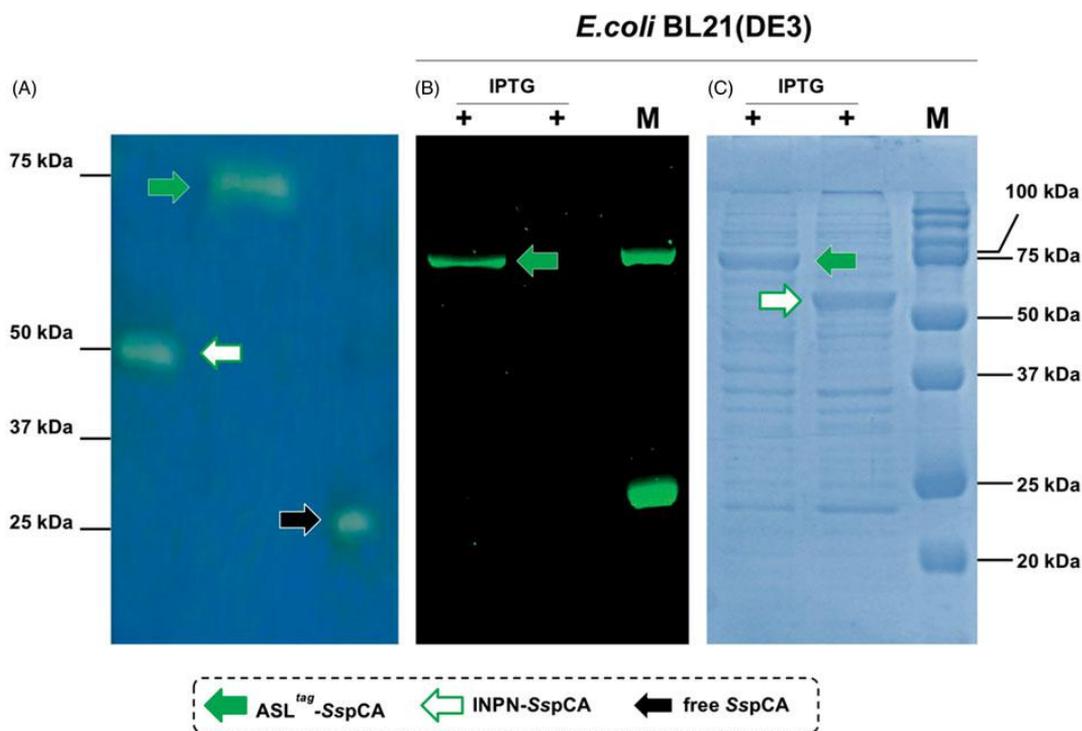
Software, San Diego, CA). Curves were obtained by the mean of three independent determinations.

### 2.6.2 Long-term stability

Membrane-bound SspCA or H<sup>5</sup>-SspCA were investigated for the long-term stability at different temperatures (25.0, 50.0, and 70.0 °C) by assaying their hydratase residual activities using CO<sub>2</sub> as substrate and withdrawing aliquots of cell surface SspCA



**Figure 3.** Model representation of an outer membrane fraction (OM; pdb from Tieleman and Berendsen<sup>65</sup>) describing the *in vivo* immobilisation of SspCA (*in blue*; PDB ID: 4G7A; panel A) and in fusion with H<sup>5</sup> (*in green*; PDB ID: 6GA0; panel B). The INPN domain is omitted because inserted in the OM. The catalytic reaction of SspCA (the hydration/dehydration of CO<sub>2</sub>) and H<sup>2</sup> (the conversion of BG-FL in the free guanine and the fluorescent benzyl-guanine derivative, B-FL, covalently linked to the active site of H<sup>5</sup>) are also shown.



**Figure 4.** Protonography (Panel A), fluorescence gel-imaging (Panel B) and Coomassie staining (Panel C) of SspCA and H<sup>5</sup>-SspCA carried out with different amounts of whole *E. coli* cells (see Materials and Methods). Filled green, white and black arrows represent the ASL<sup>tag</sup>-SspCA, INPN-SspCA and the free SspCA, respectively.

H<sup>5</sup>-SspCA at appropriate times. All the buffers were sterilised by using a sterile 0.22  $\mu$ m filter, while samples containing the membrane-bound enzymes were treated with a diluted solution of aN<sub>3</sub> to avoid contamination. All data were obtained by the mean of 3 independent determinations.

## Results and discussion

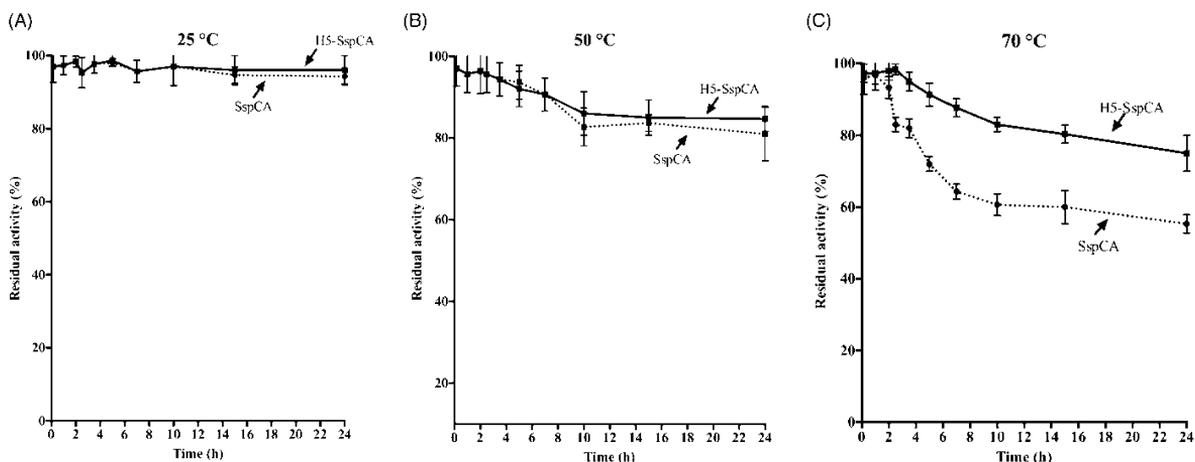
### 1 Expression and surface localisation of SspCA and H<sup>5</sup>-SspCA

Expression of the SspCA and H<sup>5</sup>-SspCA was realised through the one-step procedure, by transforming the *E. coli* cells with the construct expressing a gene composed of a signal peptide (necessary for the periplasmic translocation of the protein), the *P. syringae* INPN domain (fundamental for displaying the overexpressed protein onto the bacterial surface), and the protein of interest (SspCA or H<sup>5</sup>-SspCA). This strategy has the advantage to overexpress and directly immobilise *in vivo* the  $\alpha$ -CA or other proteins on the bacterial cell surface. Besides, the system expressing the H<sup>5</sup>-SspCA, named anchoring-and-self-labelling-protein-tag (ASL<sup>tag</sup>), allowed the labelling of the neosynthesised protein fused to H<sup>5</sup> through the use of the fluorescein derivative of the O<sup>6</sup>-BG (BG-FL), which is the substrate of H<sup>5</sup>. As reported in Figure 1, the expression of the chimeric H<sup>5</sup>-SspCA on the bacterial surface has been confirmed using the H<sup>5</sup> substrate and analysing the whole bacterial cells with fluorescent microscopy. The irreversible reaction of the SL<sup>tag</sup> system with a fluorescent substrate allowed the quantitative determination of the immobilised bacterial  $\alpha$ -CA or of other proteins fused to H<sup>5</sup>, by *in vitro* gel-imaging techniques as described by Del Prete et al.<sup>57</sup> and Merlo et al.<sup>58</sup>. Differently from H<sup>5</sup>-SspCA, the expression of the anchored His-tagged SspCA

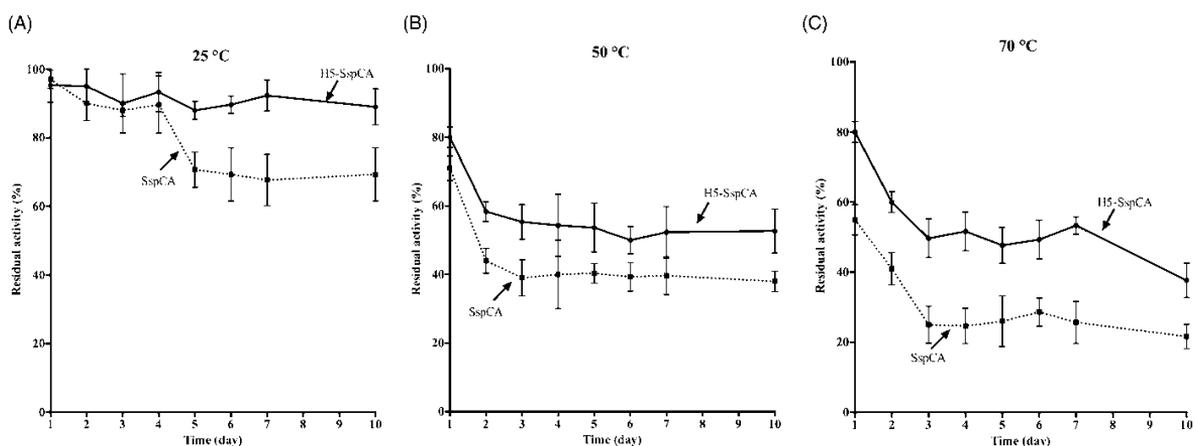
(without the H<sup>5</sup>) has been confirmed only by the Western Blot analysis using an anti-His-tag antibody (Figure 2), indicating an expected molecular weight of 50.0 kDa (the sum of the INPN and SspCA polypeptide chains produced with the construct pET-22L INPN-SspCA; see the experimental section). Anchored His-tagged H<sup>5</sup>-SspCA showed a higher molecular weight (70.0 kDa) with respect to the non-chimeric protein because of the presence of the H<sup>5</sup> protein (158 amino acid residues). The H<sup>5</sup>-SspCA Western Blot fully supports the fluorescence microscopy results. Thus, using this one-step procedure, the thermostable proteins  $\alpha$ -C (SspCA) and the chimeric ASL<sup>tag</sup>-SspCA<sup>56</sup> were efficiently expressed on the external side of the bacterial outer membrane. Figure 3 reports a model representing the *in vivo* immobilisation of SspCA (Panel A) and chimeric H<sup>5</sup>-SspCA (Panel B) on the bacterial external cell surface. Moreover, Figure 3 shows the reaction catalysed by both the biocatalysts (the hydration/dehydration of CO<sub>2</sub> and the conversion of BG-FL in the free guanine and the fluorescent benzyl-guanine derivative covalently linked to the active site of H<sup>5</sup>).

### 3.2 Hydratase activity of the membrane anchored SspCA and H<sup>5</sup>-SspCA

Using CO<sub>2</sub> as the substrate, the hydratase activity of all the forms of SspCA has been investigated in solution<sup>57</sup>. The results show that the membrane-bound SspCA with and without H<sup>5</sup> was an active enzyme, when immobilised on the bacterial surface. The CO<sub>2</sub> hydratase activity of SspCA and H<sup>5</sup>-SspCA did not show any differences. The results also evidenced that 1.0  $\mu$ g of bacterial cells had a CO<sub>2</sub> hydratase activity corresponding to that of 10.0 ng of



**Figure 5.** The thermostability of immobilised SspCA and H<sup>5</sup>-SspCA on the bacterial surface. Measures were carried out at indicated temperatures, by using aliquots of whole cells incubated up to 24 h. Legend: continuous line, membrane-bound H<sup>5</sup>-SspCA; dashed line, membrane-bound SspCA. Each point is the mean of three independent determinations.



**Figure 6.** The long-term stability of immobilised SspCA and H<sup>5</sup>-SspCA on the bacterial surface. Measures were carried out at indicated temperatures up to 10 d, using aliquots of whole bacterial cells. Continuous line: free SspCA; Dashed line: membrane-bound SspCA. Each point is the mean of three independent determinations.

ee SspCA. Probably, anchored SspCA or H<sup>5</sup>-SspCA is subjected to various phenomena, which influence the enzymatic reaction, e.g. a reduction of the structural conformational changes (this is typical of an immobilised enzyme); a different substrate access to the active site with respect to the free biocatalyst due to the bacterial cell surface microenvironment, and, finally, an aggregation of the cells or outer membranes used in the assay. Otherwise, the activity of SspCAs was compared by using the protonography, which is a technique able to reveal the hydrogen ions produced by the CO<sub>2</sub> hydratase activity reaction as a yellow band on the SDS-PAGE. The protonography results showed that all the forms of SspCA (the two membrane-bound ones and the free enzyme) had a comparable enzyme activity and a different molecular weight on SDS-PAGE (Figure 4, panel A and C). Protonography corroborated the results obtained with the fluorescent microscopy (Figure 1) and Western Blot (Figure 2). Interestingly, H<sup>5</sup>-SspCA fluorescent band had a molecular weight of about 70.0 kDa (Figure 4, panel B)

indicated that the presence of SspCA does not affect the activity of the thermostable H<sup>5</sup> enzyme on the BG-FL substrate.

### 3.3 Stability of SspCA and H<sup>5</sup>-SspCA linked to the bacterial cell surface

Using the whole bacterial cells expressing on the external surface SspCA or H<sup>5</sup>-SspCA, the effect of the CO<sub>2</sub> hydratase reaction as a function of temperature has been investigated. In Figure 5, the residual activity of the SspCA and H<sup>5</sup>-SspCA remained almost constant at 25.0 and 50.0 °C, retaining their residual activity at 100% up to 24 h (panel A) and at 70% up to 6 h of incubation (panel B) respectively. In contrast, it is readily apparent that at higher temperatures (70.0 °C) SspCA and H<sup>5</sup>-SspCA behave differently (Figure 5, panel C): the residual activity of SspCA started to decline rapidly after 2 h, getting a value of about 60% after 14 h of incubation

whereas the stabilising effect of H<sup>5</sup> on the SspCA showed a residual activity of about 85% and remained almost constant for the rest of the time indicated in the figure (panel C). These results demonstrated that the anchoring ASL<sup>tag</sup> system, enhanced the SspCA stability of about 20%. On the other hand, it is important to highlight that both anchored enzymes continued to work for several hours at temperatures considered prohibitive for the free enzymes, as SspCA, which Russo et al. demonstrated to show a residual activity of 20% when heated at 70.0 °C for 15 min<sup>57</sup>. This aspect is crucial in the context of the post-combustion carbon capture process, which requires temperatures ranging from 40.0 and 60.0 °C<sup>53</sup>. Figure 6 shows the residual activity for the CO<sub>2</sub> hydration reaction for SspCA and H<sup>5</sup>-SspCA when the whole cells were treated at different temperatures for a very long period (up to 30 d). At 25.0 °C, the SspCA residual activity started to decrease after 4 d and reached a value of about 70% after 10 d, while H<sup>5</sup>-SspCA remained almost constant (panel A). At 50.0 and 70.0 °C, the residual activity of SspCA decreased up to 40 and 20%, respectively (panel B and C), whereas H<sup>5</sup>-SspCA showed a residual activity of about 60 and 40%, respectively (panel B and C). All these data confirmed that the presence of a thermostable *protein-tag* between the IPN anchoring domain and the SspCA significantly improved the long-term stability and the storage of this CA.

## Conclusions

The ASL<sup>tag</sup> system efficiently overexpressed the chimeric H<sup>5</sup>-SspCA onto the bacterial cell surface, as demonstrated by fluorescence microscopy and Western-Blot. As expected, the CO<sub>2</sub> hydratase assay and the protonography showed that SspCA was still very active, even linked on the bacterial surface and the H<sup>5</sup> moiety, showing a CO<sub>2</sub> hydratase activity similar to that of its anchored counterpart without H<sup>5</sup>. Furthermore, by investigating the behaviour of the whole bacterial cells expressing on the external surface SspCA or H<sup>5</sup>-SspCA at different temperatures, we demonstrated an enhancement in terms of thermal stability of the chimeric protein. In conclusion, the H<sup>5</sup>-SspCA obtained by the ASL<sup>tag</sup> system constitutes a valid strategy for further increasing the thermostability of proteins, for processes in which a highly effective, thermostable catalyst is needed.

## Acknowledgements

We are grateful to Giovanni Del Monaco for technical assistance.

## Disclosure statement

The authors state no conflict of interests.

## Funding

This work was supported by FIRB-Futuro in Ricerca RBF12001G\_22 “Nematic” and by the grant “SMART GENERATION – Sistemi e tecnologie sostenibili per la generazione di energia-PON03PE\_20157\_1, OR3-Bio-sistemi di cattura ed utilizzazione della CO<sub>2</sub>”.

## ORCID

Laudiu T. Supuran  <http://orcid.org/0000-0003-4262-0323>  
 Capasso  <http://orcid.org/0000-0003-3314-2411>

## References

1. Capasso C, Supuran CT. Inhibition of bacterial carbon anhydrases as a novel approach to escape drug resistant *Curr Top Med Chem* 2017;17:1237–48.
2. Del Prete S, Vullo D, De Luca V, et al. Cloning, expression purification and sulfonamide inhibition profile of the complete domain of the eta-carbonic anhydrase from *Plasmodium falciparum*. *Bioorg Med Chem Lett* 2016;24:4184–90.
3. Del Prete S, Vullo D, De Luca V, et al. Anion inhibition profiles of the complete domain of the eta-carbonic anhydrase from *Plasmodium falciparum*. *Bioorg Med Chem* 2016;24:4410–4.
4. Annunziato G, Angeli A, D’Alba F, et al. Discovery of new potential anti-infective compounds based on carbonic anhydrase inhibitors by rational target-focused repurposing approaches. *Chem Med Chem* 2016;11:1904–14.
5. Del Prete S, Vullo D, De Luca V, et al. Anion inhibition profiles of alpha-, beta- and gamma-carbonic anhydrases from the pathogenic bacterium *Vibrio cholerae*. *Bioorg Med Chem* 2016;24:3413–7.
6. Abdel Gawad NM, Amin NH, Elsaadi MT, et al. Synthesis of 4-(thiazol-2-ylamino)-benzenesulfonamides with carbonic anhydrase I, II and IX inhibitory activity and cytotoxic effect against breast cancer cell lines. *Bioorg Med Chem* 2016;24:3043–51.
7. Capasso C, Supuran CT. An overview of the carbonic anhydrases from two pathogens of the oral cavity: streptococcus mutans and porphyromonas gingivalis. *Curr Top Med Chem* 2016;16:2359–68.
8. Del Prete S, Vullo D, De Luca V, et al. Comparison of the sulfonamide inhibition profiles of the alpha-, beta- and gamma-carbonic anhydrases from the pathogenic bacterium *Vibrio cholerae*. *Bioorg Med Chem Lett* 2016;26:1941–6.
9. Supuran CT, Capasso C. Biomedical applications of prokaryotic carbonic anhydrases. *Expert Opin Ther Pat* 2018;28:745–54.
10. Nishimori I, Onishi S, Takeuchi H, Supuran CT. The alpha and beta classes carbonic anhydrases from *Helicobacter pylori*: novel drug targets. *Curr Pharm Des* 2008;14:622–30.
11. Morishita S, Nishimori I, Minakuchi T, et al. Cloning, polymorphism, and inhibition of beta-carbonic anhydrase from *Helicobacter pylori*. *J Gastroenterol* 2008;43:849–57.
12. Supuran CT. Advances in structure-based drug discovery of carbonic anhydrase inhibitors. *Expert Opin Drug Discov* 2017;12:61–88.
13. Supuran CT. Structure and function of carbonic anhydrase. *Biochem J* 2016;473:2023–32.
14. Supuran CT, Capasso C. An overview of the bacterial carbonic anhydrases. *Metabolites* 2017;7:pii: E56.
15. Capasso C, Supuran CT. An overview of the alpha-, beta- and gamma-carbonic anhydrases from bacteria: can bacterial carbonic anhydrases shed new light on evolution of bacteria? *J Enzyme Inhib Med Chem* 2015;30:325–32.
16. Supuran CT. Carbonic anhydrase activators. *Future Med Chem* 2018;10:561–73.
17. Supuran CT, Capasso C. An overview of the bacterial carbonic anhydrases. *Metabolites* 2017;7:56–74.

3. Supuran CT, Capasso C. Carbonic anhydrase from *Porphyromonas gingivalis* as a drug target. *Pathogens* 2017; 6:30–42.
3. Capasso C, Supuran CT. Bacterial, fungal and protozoan carbonic anhydrases as drug targets. *Expert Opin Ther Targets* 2015;19:1689–704.
2. Aspatwar A, Tolvanen ME, Ortutay C, Parkkila S. Carbonic anhydrase related proteins: molecular biology and evolution. *Subcell Biochem* 2014;75:135–56.
1. Supuran CT. Carbonic anhydrases as drug targets-an overview. *Curr Top Med Chem* 2007;7:825–33.
2. Supuran CT, Capasso C. The  $\eta$ -class carbonic anhydrases as drug targets for antimalarial agents. *Expert Opin Ther Targets* 2015;19:551–63.
3. Capasso C, Supuran CT. An overview of the selectivity and efficiency of the bacterial carbonic anhydrase inhibitors. *Curr Med Chem* 2015;22:2130–9.
4. Capasso C, Supuran CT. Sulfa and trimethoprim-like drugs - antimetabolites acting as carbonic anhydrase, dihydropteroate synthase and dihydrofolate reductase inhibitors. *J Enzyme Inhib Med Chem* 2014;29:379–87.
5. Capasso C, Supuran CT. Anti-infective carbonic anhydrase inhibitors: a patent and literature review. *Expert Opin Ther Pat* 2013;23:693–704.
5. Ozensoy Guler O, Capasso C, Supuran CT. A magnificent enzyme superfamily: carbonic anhydrases, their purification and characterization. *J Enzyme Inhib Med Chem* 2016;31: 689–94.
7. Del Prete S, Vullo D, De Luca V, et al. Sulfonamide inhibition studies of the beta-carbonic anhydrase from the pathogenic bacterium *Vibrio cholerae*. *Bioorg Med Chem* 2016;24: 1115–20.
3. Del Prete S, De Luca V, De Simone G, et al. Cloning, expression and purification of the complete domain of the eta-carbonic anhydrase from *Plasmodium falciparum*. *J Enzyme Inhib Med Chem* 2016;31:54–59.
3. De Simone G, Monti SM, Alterio V, et al. Crystal structure of the most catalytically effective carbonic anhydrase enzyme known, SazCA from the thermophilic bacterium *Sulfurihydrogenibium azorense*. *Bioorg Med Chem Lett* 2015; 25:2002–6.
2. Di Fiore A, Capasso C, De Luca V, et al. X-ray structure of the first 'extremo-alpha-carbonic anhydrase', a dimeric enzyme from the thermophilic bacterium *Sulfurihydrogenibium yellowstonense* YO3AOP1. *Acta Crystallogr D Biol Crystallogr* 2013; 69:1150–9.
1. Alterio V, Langella E, De Simone G, Monti SM. Cadmium-containing carbonic anhydrase CDCA1 in marine diatom *Thalassiosira weissflogii*. *Mar Drugs* 2015;13:1688–97.
2. Di Fiore A, Alterio V, Monti SM, et al. Thermostable carbonic anhydrases in biotechnological applications. *Int J Mol Sci* 2015;16:15456–80.
3. Supuran CT. CA IX stratification based on cancer treatment: a patent evaluation of US2016/0002350. *Expert Opin Ther Pat* 2016;26:1105–9.
4. Lomelino C, McKenna R. Carbonic anhydrase inhibitors: a review on the progress of patent literature (2011–2016). *Expert Opin Ther Pat* 2016;26:947–56.
5. Monti SM, Supuran CT, De Simone G. Anticancer carbonic anhydrase inhibitors: a patent review (2008–2013). *Expert Opin Ther Pat* 2013;23:737–49.
36. Masini E, Carta F, Scozzafava A, Supuran CT. Antiglucom carbonic anhydrase inhibitors: a patent review. *Expert Opin Ther Pat* 2013;23:705–16.
37. Scozzafava A, Supuran CT, Carta F. Antiobesity carbon anhydrase inhibitors: a literature and patent review. *Expert Opin Ther Pat* 2013;23:725–35.
38. Aggarwal M, Kondeti B, McKenna R. Anticonvulsant/antiepileptic carbonic anhydrase inhibitors: a patent review. *Expert Opin Ther Pat* 2013;23:717–24.
39. Carta F, Supuran CT. Diuretics with carbonic anhydrase inhibitory action: a patent and literature review (2005–2013). *Expert Opin Ther Pat* 2013;23:681–91.
40. Winum JY, Capasso C. Novel antibody to a carbonic anhydrase: patent evaluation of WO2011138279A1. *Expert Opin Ther Pat* 2013;23:757–60.
41. Aggarwal M, McKenna R. Update on carbonic anhydrase inhibitors: a patent review (2008–2011). *Expert Opin Ther Pat* 2012;22:903–15.
42. Carta F, Scozzafava A, Supuran CT. Sulfonamides: a patent review (2008–2012). *Expert Opin Ther Pat* 2012;22:747–58.
43. Carta F, Supuran CT, Scozzafava A. Novel therapies for glaucoma: a patent review 2007–2011. *Expert Opin Ther Pat* 2012;22:79–88.
44. Poulsen SA. Carbonic anhydrase inhibition as a cancer therapy: a review of patent literature, 2007–2009. *Expert Opin Ther Pat* 2010;20:795–806.
45. Boone CD, Habibzadegan A, Gill S, McKenna R. Carbonic anhydrases and their biotechnological application. *Biomolecules* 2013;3:553–62.
46. Alafeefy AM, Abdel-Aziz HA, Vullo D, et al. Inhibition of carbonic anhydrases from the extremophilic bacterium *Sulfurihydrogenibium yellowstonense* (SspCA) and *Sulfurihydrogenibium azorense* (SazCA) with a new series of sulfonamides incorporating aroylhydrazone-, [1,2,4]triazolo[3,4-b][1,3,4]thiadiazinyl- or 2-(cyanophenylmethylene 1,3,4-thiadiazol-3(2H)-yl) moieties. *Bioorg Med Chem* 2014;24: 141–7.
47. Vullo D, De Luca V, Scozzafava A, et al. The extremo-alpha carbonic anhydrase from the thermophilic bacterium *Sulfurihydrogenibium azorense* is highly inhibited by sulfonamides. *Bioorg Med Chem* 2013;21:4521–5.
48. Vullo D, Luca VD, Scozzafava A, et al. The alpha-carbonic anhydrase from the thermophilic bacterium *Sulfurihydrogenibium yellowstonense* YO3AOP1 is highly susceptible to inhibition by sulfonamides. *Bioorg Med Chem* 2013;21:1534–8.
49. Akdemir A, Vullo D, De Luca V, et al. The extremo-alpha carbonic anhydrase (CA) from *Sulfurihydrogenibium azorense* the fastest CA known, is highly activated by amino acids and amines. *Bioorg Med Chem Lett* 2013;23:1087–90.
50. De Luca V, Vullo D, Scozzafava A, et al. Anion inhibitory studies of an alpha-carbonic anhydrase from the thermophilic bacterium *Sulfurihydrogenibium yellowstonense* YO3AOP1. *Bioorg Med Chem Lett* 2012;22:5630–4.
51. Vullo D, De Luca V, Scozzafava A, et al. Anion inhibitory studies of the fastest carbonic anhydrase (CA) known, the extremo-CA from the bacterium *Sulfurihydrogenibium azorense*. *Bioorg Med Chem Lett* 2012;22:7142–5.
52. Vullo D, De Luca V, Scozzafava A, et al. The first activatable study of a bacterial carbonic anhydrase (CA). The thermally stable alpha-CA from *Sulfurihydrogenibium yellowstonense* YO3AOP1 is highly activated by amino acids and amines. *Bioorg Med Chem Lett* 2012;22:6324–7.

3. Russo ME, Olivieri G, Capasso C, et al. Kinetic study of a novel thermo-stable alpha-carbonic anhydrase for biomimetic CO<sub>2</sub> capture. *Enzyme Microb Technol* 2013;53:271–7.
4. Migliardini F, De Luca V, Carginale V, et al. Biomimetic CO<sub>2</sub> capture using a highly thermostable bacterial alpha-carbonic anhydrase immobilized on a polyurethane foam. *J Enzyme Inhib Med Chem* 2014;29:146–50.
5. Abdelrahim MYM, Martins CF, Neves LA, et al. Supported ionic liquid membranes immobilized with carbonic anhydrases for CO<sub>2</sub> transport at high temperatures. *J Membr Sci* 2017;528:225–30.
5. Perfetto R, Del Prete S, Vullo D, et al. Production and covalent immobilisation of the recombinant bacterial carbonic anhydrase (SspCA) onto magnetic nanoparticles. *J Enzyme Inhib Med Chem* 2017;32:759–66.
7. Del Prete S, Perfetto R, Rossi M, et al. A one-step procedure for immobilising the thermostable carbonic anhydrase (SspCA) on the surface membrane of *Escherichia coli*. *J Enzyme Inhib Med Chem* 2017;32:1120–8.
3. Merlo R, Del Prete S, Valenti A, et al. An AGT-based protein-tag system for the labelling and surface immobilization of enzymes on *E. coli* outer membrane. *J Enzyme Inhib Med Chem* 2019;34:490–9.
3. Capasso C, De Luca V, Carginale V, et al. Biochemical properties of a novel and highly thermostable bacterial alpha-carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* YO3AOP1. *J Enzyme Inhib Med Chem* 2012;27:892–7.
60. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970;227:680–5.
61. Del Prete S, De Luca V, Supuran CT, Capasso C. Protonography, a technique applicable for the analysis of alpha-carbonic anhydrase activity. *J Enzyme Inhib Med Chem* 2015;30:920–4.
62. Del Prete S, De Luca V, Landolo E, et al. Protonography, a powerful tool for analyzing the activity and the oligomeric state of the gamma-carbonic anhydrase identified in the genome of *Porphyromonas gingivalis*. *Bioorg Med Chem* 2015;23:3747–50.
63. De Luca V, Del Prete S, Supuran CT, Capasso C. Protonography, a new technique for the analysis of carbonic anhydrase activity. *J Enzyme Inhib Med Chem* 2015;30:277–82.
64. Vettone A, Serpe M, Hidalgo A, et al. A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNAalkyl-transferase by protein engineering. *Extremophiles* 2016;20:13.
65. Tieleman DP, Berendsen HJ. A molecular dynamics study of the pores formed by *Escherichia coli* OmpF porin in a fully hydrated palmitoyl-oleoyl-phosphatidylcholine bilayer. *Bioophys J* 1998;74:2786–801.

Review

# $O^6$ -alkylguanine-DNA Alkyltransferases in Microbes Living on the Edge: From Stability to Applicability

Rosanna Mattossovich <sup>1,†</sup>, Rosa Merlo <sup>1,†</sup>, Riccardo Miggiano <sup>2</sup>, Anna Valenti <sup>1,\*</sup>  
and Giuseppe Perugino <sup>1,\*</sup>

<sup>1</sup> Institute of Bioscience and BioResources, National Research Council of Italy, Via Pietro Castellino 111, 80131 Naples, Italy; rosanna.mattossovich@ibbr.cnr.it (R.M.); rosa.merlo@ibbr.cnr.it (R.M.)

<sup>2</sup> Department of Pharmaceutical Sciences, University of Piemonte Orientale, Via Bovio 6, 28100 Novara, Italy; riccardo.miggiano@uniupo.it

\* Correspondence: anna.valenti@ibbr.cnr.it (A.V.); giuseppe.perugino@ibbr.cnr.it (G.P.); Tel.: +39-081-6132-247 (A.V.); +39-081-6132-496 (G.P.); Fax: +39-081-6132-646 (A.V. & G.P.)

† These authors contribute equally to this work.

Received: 26 March 2020; Accepted: 16 April 2020; Published: 20 April 2020



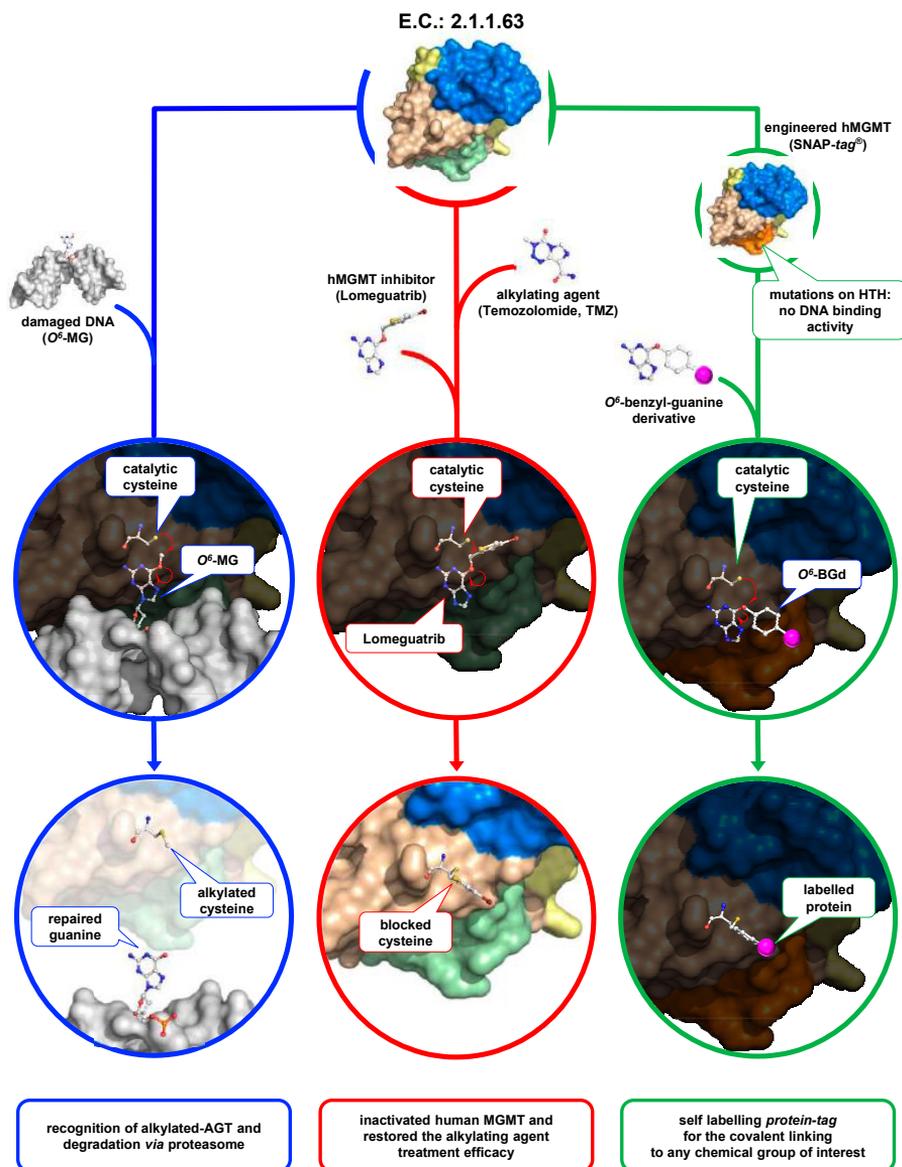
**Abstract:** The genome of living cells is continuously exposed to endogenous and exogenous attacks, and this is particularly amplified at high temperatures. Alkylating agents cause DNA damage, leading to mutations and cell death; for this reason, they also play a central role in chemotherapy treatments. A class of enzymes known as AGTs (alkylguanine-DNA-alkyltransferases) protects the DNA from mutations caused by alkylating agents, in particular in the recognition and repair of alkylated guanines in  $O^6$ -position. The peculiar irreversible self-alkylation reaction of these enzymes triggered numerous studies, especially on the human homologue, in order to identify effective inhibitors in the fight against cancer. In modern biotechnology, engineered variants of AGTs are developed to be used as *protein tags* for the attachment of chemical ligands. In the last decade, research on AGTs from (hyper)thermophilic sources proved useful as a model system to clarify numerous phenomena, also common for mesophilic enzymes. This review traces recent progress in this class of *thermozymes*, emphasizing their usefulness in basic research and their consequent advantages for in vivo and in vitro biotechnological applications.

**Keywords:** thermophilic sources; DNA repair; biotechnological tools; alkylation damage; AGT

## 1. Introduction

Monofunctional alkylating agents, a class of mutagenic and carcinogenic agents present in the environment, induce DNA alkylation in several positions including guanine at  $O^6$  ( $O^6$ -MG; 6% of adducts formed), the  $N^7$  of guanine ( $N^7$ -MG; 70%), and the  $N^3$  of adenine ( $N^3$ -MA; 9%) [1]. Alkylation of guanine ( $O^6$ -AG) is a cytotoxic lesion, although the specific mechanism of this cytotoxicity is not yet fully understood. It was proposed that the toxic effect occurs after DNA replication, because the  $O^6$ -AG incorrectly base-pairs with thymine generating a transition from G:C to A:T [2]. The mutations caused by  $O^6$ -MG that occur at the time of replication are recognized by the post-replication mismatch repair system with potential harmful implications for cell viability. Apart from conventional DNA repair pathways as Mismatch Excision Repair (MMR), Nucleotide Excision Repair (NER), Base Excision Repair (BER), alkylated-DNA protein alkyl-transferases (called  $O^6$ -alkyl-guanine-DNA-alkyl-transferase (AGT or OGT) or  $O^6$ -methyl-guanine- DNA-alkyl-transferase (MGMT); EC: 2.1.1.63) perform the direct repair of alkylation damage in DNA [3,4]. They represent the major factor in counteracting the effects of alkylating agents that form such adducts [4]. These are small enzymes (17–22 kDa) that are widely present in organisms of the three kingdoms (bacteria, archaea, eukaryotes) but apparently absent from

plants, *Schizosaccharomyces pombe*, *Thermus thermophilus*, and *Deinococcus radiodurans*. The reaction mechanism of AGTs is based on the recognition of the damaged nucleobase on DNA [5], followed by a one-step  $SN_2$ -like mechanism, in which the alkyl group of the damaged guanine is irreversibly transferred to a cysteine residue in its active site [5–8] (Figure 1, blue path).



**Figure 1.** The O<sup>6</sup>-alkyl-guanine-DNA-alkyl-transferase (AGT)s’ world. AGTs are small enzymes composed by a N-terminal domain (in sky blue), a C-terminal domain (in light brown) connected by a loop (in yellow). In the C-terminal domain, a helix-turn-helix motif (in light green, or in orange in the SNAP-tag) is responsible of the DNA-binding activity. The peculiar irreversible reaction mechanism of these enzymes plays a pivotal role in the physiological DNA repair (blue path), and it has important repercussions in cancer cell treatment (red path) and biotechnological applications (green path). Atoms are coloured by the CPK colour convention.

For these reasons, they are also called *suicide* or *kamikaze* proteins, showing a 1:1 stoichiometry of their reaction with the natural substrate. The disadvantage of this elegant catalysis is that, upon alkylation, the protein is self-inactivated and destabilized, triggering its recognition by cellular systems to be degraded by the proteasome [8,9].

### 1.1. AGTs as Targets in Cancer Therapy

Alkylation damage to DNA occurs in various living conditions, and for this reason the widespread presence of AGT protects cells from killing by alkylating agents. However, human AGT (hMGMT) is a *double-edged sword*—on the one hand, it protects healthy cells from these genotoxic and carcinogenic effects, but also counteracts alkylating agents-based chemotherapy by protecting cancer cells from the killing effect of these drugs [10,11]. Consequently, hMGMT has emerged as a crucial factor in anticancer therapies [12]—an inverse relationship has been discovered between the presence of hMGMT and the sensitivity of cells to the cytotoxic effects of alkylating agents, such as temozolomide (TMZ), in different types of cancer cells, including prostate, breast, colon, and lung cancer cells [13].

The resistance to chemotherapy may be reduced by inhibition of these enzymes; as described before, after removing the lesion, the alkylated form of the protein is inactivated and enters the intracellular degradation pathways. Hence, in order to counteract the action of hMGMT in chemotherapy regimens, a large number of studies aimed to develop hMGMT inhibitors to be used in combination with alkylating agents. In view of this therapeutic relevance, much success has been obtained through the design of hMGMT pseudo-substrates, namely, the  $O^6$ -benzylguanine ( $O^6$ -BG) and the strong inactivator  $O^6$ -[4-bromophenyl]-guanine ( $O^6$ -BTG, Lomeguatrib) [13,14]. These compounds mimic damaged guanine on DNA and react with the protein by the covalent transfer of the alkyl adduct to the active site cysteine residue, thus irreversibly inactivating the enzyme (Figure 1, red path). Therapeutically,  $O^6$ -BG is not toxic on its own, but renders cancer cells 2 to 14 times more sensitive to alkylating agents' effects. The oligonucleotides containing several  $O^6$ -BG are potent inhibitors and represent a valid alternative to the use of free modified guanines, thereby improving the activity of the alkylating chemotherapy drug in the treatment of some tumours [15–17].

### 1.2. AGTs and Biotechnology

The specific labelling of proteins with synthetic probes is an important advance for the study of protein function. To achieve this, the protein of interest is expressed in a fusion with additional genetically encoded polypeptides, called *tags*, which mediate the labelling. The first example of an *autofluorescent tag* was the *Aequorea victoria* green fluorescent protein (GFP) allowing the *in vivo* localization of fusion proteins in cellular and molecular biology [18,19]. Among *affinity tags*, of particular importance are the poly(His)-*tag*, the chitin-binding protein, the maltose-binding protein [20], the Strep-*tag* [21], and the glutathione-S-transferase (GST-*tag*) [22], which allow fast and specific purification of proteins of interest from their crude biological source using affinity techniques. *Solubilization tags* are especially useful to assist the proper folding of recombinant proteins expressed in chaperone-deficient species such as *Escherichia coli*, avoiding protein precipitation and the use of alternative expression protocols [23,24]—these include thioredoxin [25] and poly(NANP).

However, all the *tags* listed above are limited by the fact that each of them can be used for one or only a few applications. The need therefore emerged to develop a *universal tag* that could widely cover several applications.

In 2003, the group headed by Kai Johnsson pioneered the use of an engineered hMGMT variant as a fusion protein for *in vitro* and *in vivo* biotechnology applications, which led then to its commercialization, namely, the SNAP-*tag* (New England Biolabs) [26–29]. They started from the knowledge that hMGMT tolerates the presence of groups conjugated to the pseudo-substrate  $O^6$ -BG ( $O^6$ -BG derivatives)—the unusual covalent bond with the benzyl moiety can therefore be exploited for “biotech” purposes (Figure 1, green path). Thanks to its small size, the engineered hMGMT (SNAP-*tag*) can be fused with other proteins of interest. The expression of the fusion protein inside the cells

followed by incubation with opportune fluorescent derivatives leads to in vivo labelling of fusion proteins with the probe, which can be used for localization studies [26]. The same principle has also been used for the immobilization of tagged fusion proteins in vitro [30]. This offers a delicate condition for fixing and disposing in a better orientation of a wide range of proteins/enzymes on a surface. The SNAP-tag technology was successfully applied to surface plasmon resonance (SPR) for the covalent immobilization of proteins of interest [31]. Another interesting application of this protein-tag is the possibility to produce new antibody fragments (scFv-SNAP) to be employed in the SPR analysis [32].

Despite the need to use a specific substrate, SNAP-tag offers endless applications—the possibility to covalently link a desired chemical group (conjugated to the  $O^6$ -BG) to a protein of interest (genetically fused to it) makes it decidedly advantageous, if compared to traditional *protein tags* currently in use. Table 1 shows a brief comparison between some examples of *protein tags* and the SNAP-tag in several application fields.

**Table 1.** The use of *protein tags* in some applicative examples.

Applications	FPs	Affinity Tag	SNAP-Tag	Notes
In vivo imaging	+ <sup>a</sup>		+	
Substrate utilization	+	–	–	FPs do not need of any substrate for their fluorescence
Emission spectra	±	–	+	FPs are in a limited number with respect to chemical probes
Time-resolved fluorescence	±	–	+	
Multi-colour fluorescence	±	–	+	For FPs, multi-cloning and expression is necessary
In vitro applications	±	±	+	
Variety of chemical group labelling	–	–	+	
Pulse-chase analysis	–	–	+	Fresh synthesized FPs cannot be efficiently quenched
Anaerobic conditions	–	+	+	FPs' fluorophore formation requires oxygen
Protein purification	+	+	+	Utilization of the GFP-trap matrix
Protein immobilization	+	+	+	Utilization of the GFP-trap matrix
Pull-down experiments	+	+	+	Utilization of the GFP-trap matrix

<sup>a</sup> +, fully applicable or advantageous; ±, limited applicability; –, not applicable or disadvantageous; FPs = fluorescent proteins.

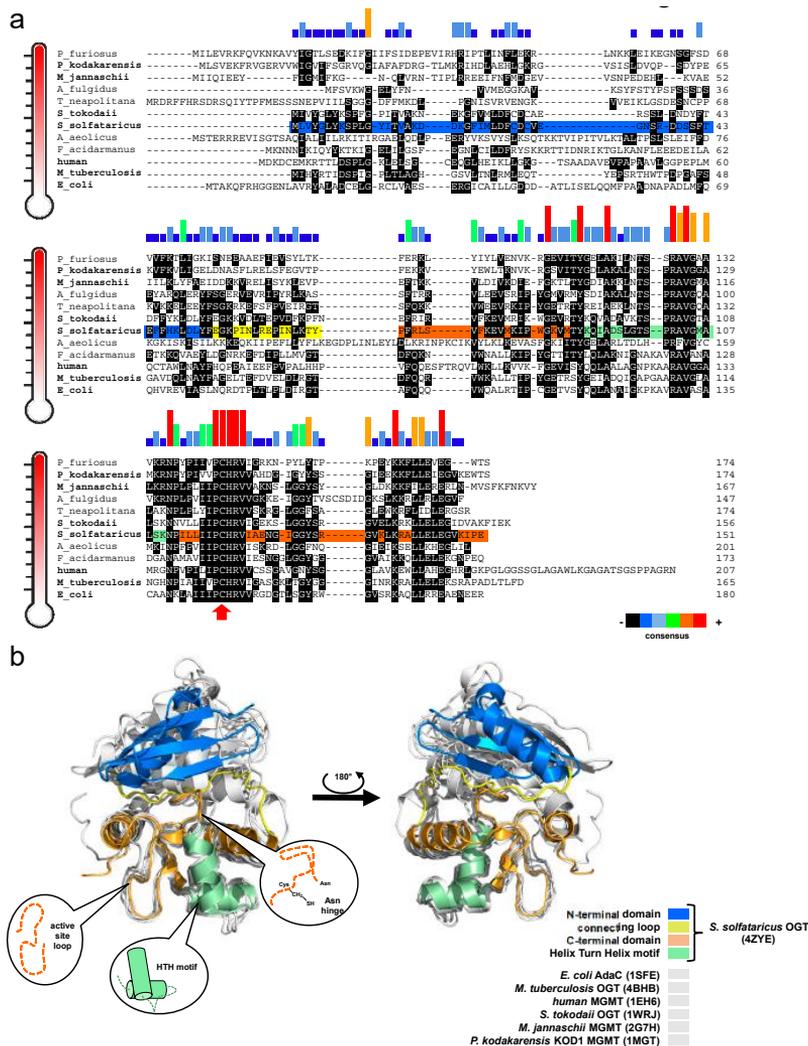
## 2. Thermophilic and Thermostable AGTs

As for organisms living under mesophilic conditions, environmental and endogenous alkylating agents also attack the genome of thermophilic and hyper-thermophilic organisms. Additionally, high temperatures accelerate the process of alkylation, leading to DNA breaks [33], because alkylating agents are chemically unstable at the physiological conditions of these organisms, and their collateral decomposition may worsen the formation of DNA alkylation products [34]. Thus, the presence of AGTs and methylpurine glycosylases in hyperthermophilic organisms implies that they are naturally exposed to endogenous methylating agents [34], thus supporting the crucial role of AGTs [35,36].

Apart from some studies on Archaea using cell-free extracts, a few examples of biochemical studies of AGTs from thermophilic sources include the enzymes from *Pyrococcus* sp. KOD1 [35] conducted by Imanaka and co-workers from *Aquifex aeolicus* and *Archaeoglobus fulgidus* performed by the group of Prof. Pegg in 2003 (Figure 2) [34]. Intriguingly, *A. aeolicus* AGT, whose organism was identified as the most primitive bacterium, is closer to the mammalian AGTs than other bacterial homologues in terms of  $O^6$ -BG sensitivity [34].

The Common Themes in AGTs' Tertiary Structure and the Intrinsic Factors of Stability

Despite the different primary structures (Figure 2a), thermophilic enzymes show a typical AGT protein architecture, consisting of two domains [37]: a highly conserved C-terminal domain (CTD), surprisingly superimposable for all available AGT structures (Figure 2b), and a N-terminal domain (NTD), which is very different among AGTs and whose function is not well understood (likely involved in regulation, cooperative binding, and stability [6,38,39]). The CTD contains the DNA binding *helix-turn-helix* motif (HTH); the *Asn hinge*, which precedes the -V/IPCHR<sup>W</sup>V/I- amino acid sequence containing the conserved catalytic cysteine (except the *Caenorhabditis elegans* AGT-2 that has the -PCHP- sequence [40,41]); and the *active site loop*, responsible for the substrate specificity.



**Figure 2.** (a) Alignment of biochemically and structurally (in bold) characterized AGTs. DNA sequences are listed in decreasing order of temperature. The histograms in different colours show the sequence consensus, and the red arrow indicates the highly conserved catalytic cysteine. (b) Superimposition of all known AGT structures in their free form (in grey). All common domains and elements are coloured only for the *O*<sup>6</sup>-alkyl-guanine-DNA-alkyl-transferase protein from the archaeon *Saccharolobus solfataricus* (hereinafter *Ss*OGT) enzyme. Coloured bars behind the *Ss*OGT sequence in (a) recall the enzyme domains highlighted in the structure and in the legend in (b).

A comparative structural analysis performed on AGT proteins whose structures are in the Protein Data Bank revealed significant differences of the intrinsic structural features that have been considered to be relevant for thermostability, such as helix capping, intramolecular contacts (hydrogen bonds, ion-pairs), and solvent-accessible surface areas. Helix capping plays a central role in the stability of  $\alpha$ -helices, due to lack of intra-helical hydrogen bonds in the first and last turn [42,43], and its effect results in an overall structural stabilization of protein folding [44]. By inspecting the crystal structure of SsOGT (PDB ID: 4ZYE), considered here as the thermophilic reference AGT protein, we verified that the five  $\alpha$ -helices of the composing the protein tertiary structure are characterized by the presence of helix capping, this possibly increasing the thermal stability. In particular, the helix *H1* at the NTD is stabilised by a peculiar double serine sequence (S40–S41) and a glutamic acid (E54) at its CTD, the latter is strictly conserved in all AGTs from thermophilic organisms (see Figure 2a). The HTH motif, built on helices *H3* and *H4*, is stabilized at the level of *H3* by a highly conserved threonine residue (T89) as N-cap and a serine (S96), distinctive of SsOGT, as C-cap. Furthermore, helix *H4* contains two serine-based capping among which the one placed at NTD (S100) is strictly conserved in all thermophilic AGTs and is followed by a proline (P101) that fits well in the first turn of the helix thanks to its own backbone conformation. Finally, the helix *H5* is protected by glutamic acid capping that is present in all the AGTs from different species. Another feature contributing to thermal stability is the solvent-accessible surface area (SASA). Indeed, the decrease of SASA and the increase of hydrophobic residues that are buried from the solvent have stabilizing principles for thermostable protein [45]. As described in Table 2, SsOGT shows the smaller total SASA value, in line with its exceptional stability. On the contrary, OGT from *Mycobacterium tuberculosis* [38,46] has a higher value due to the peculiar conformation of both the active site loop and the C-terminal tail that are exposed to the bulk solvent and are less heat stable (Table 2).

Finally, by comparing hyperthermophilic AGTs with the orthologs from mesophilic organisms, in terms of atomic contacts between charged residues as well as intramolecular hydrogen bonds (Table 2), significant differences emerged in the number of charged residues contacts. As expected for thermostable proteins [47], SsOGT, as well as the proteins from *Sulfurisphaera tokodaii* and *Pyrococcus kodakaraensis*, shows a larger number of electrostatic contacts, characterized by higher bond-dissociation energy, with respect to hydrogen bonds for which we did not detect significant differences among the analysed structures, apart from MGMT of *P. kodakaraensis* (*Pk*-MGMT) [48].

**Table 2.** Comparison of solvent-accessible surface area and intramolecule contacts.

T <sub>opt</sub> (°C)	Enzyme (PDB ID)	Total SASA (Å <sup>2</sup> )	Charged Residues Contacts	Intramolecule H-bonds <sup>a</sup>	References
37	<i>Escherichia coli</i> Ada-C (1SFE)	8421.8	74	141	[49]
37	<i>Mycobacterium tuberculosis</i> OGT (4BHB)	9535.2	56	143	[38]
37	<i>Homo sapiens</i> MGMT (1EH6)	8764.3	71	127	[6]
80	<i>Saccharolobus solfataricus</i> OGT (4ZYE)	8054.1	94	137	[39]
80	<i>Sulfurisphaera tokodaii</i> OGT (1WRJ)	8049.5	124	134	PDB <sup>b</sup>
80	<i>Methanocaldococcus jannaschii</i> MGMT (2G7H)	17,770.8 <sup>c</sup>	N.D.	N.D.	[50]
85	<i>Pyrococcus kodakaraensis</i> MGMT (1MGT)	8302.8	111	157	[48]

<sup>a</sup> Excluding intra-residues H-bonds. <sup>b</sup> <https://www.rcsb.org/structure/1wrj>. <sup>c</sup> The structure has been solved by means of NMR explaining the high SASA value.

Although the number of H-bonds is approximately similar across the AGTs from different organisms, there should be differences in the position-related role of such bonds, supporting overall stability of thermophilic variants. With reference to *Pk*-MGMT, Hashimoto and co-workers detected

the same number of ion-pairs between the extremophilic protein and *E. coli* Ada-C [49]; however, more intra- and inter-helix ion pairs were found in *Pk*-MGMT. Although the absence of a correlation between ion pairs' position and stabilization in Ada-C exists, the intra-helix ion pairs act in the secondary structure of *Pk*-MGMT, stabilizing helices, and the inter-helix ion pairs consolidate the inter-domain interactions, enhancing the stability of the tertiary structure packing.

### 3. The $O^6$ -Alkylguanine-DNA-Alkyltransferase from *Saccharolobus solfataricus*

In the last decade, SsOGT has been characterized through detailed physiological, biochemical, and structural analysis. Due to its intrinsic stability, the SsOGT protein has proven to be an outstanding model for clarifying the relationships between function and structural characteristics.

*Saccharolobus solfataricus* (previously known as *Sulfolobus solfataricus*) is a microorganism first isolated and discovered in 1980 in the Solfatara volcano (Pisciarelli-Naples, Italy) [51], which thrives in volcanic hot springs at 80 °C and a pH 2.0–4.0 range. In order to protect its genome in these harsh conditions, *S. solfataricus* evolved several efficient protection and repair systems [33,52]. *S. solfataricus* is highly sensitive to the alkylating agent methyl methane sulfonate (MMS), showing a transient growth arrest when treated with MMS concentrations in the range of > 0.25 mM to 0.7 mM [33,52]. Interestingly, although the *ogt* RNA level increases after MMS treatment, the relative enzyme concentration decreases, suggesting its degradation in cells in response to the alkylating agent and, in general, to a cellular stress [52]. Under these treatment conditions, however, the protein level rises after few hours, and, in parallel, the growth of *Saccharolobus* starts again [52], indicating a role of SsOGT in efficient DNA repair by alkylation damage.

#### 3.1. Innovative OGT Assays

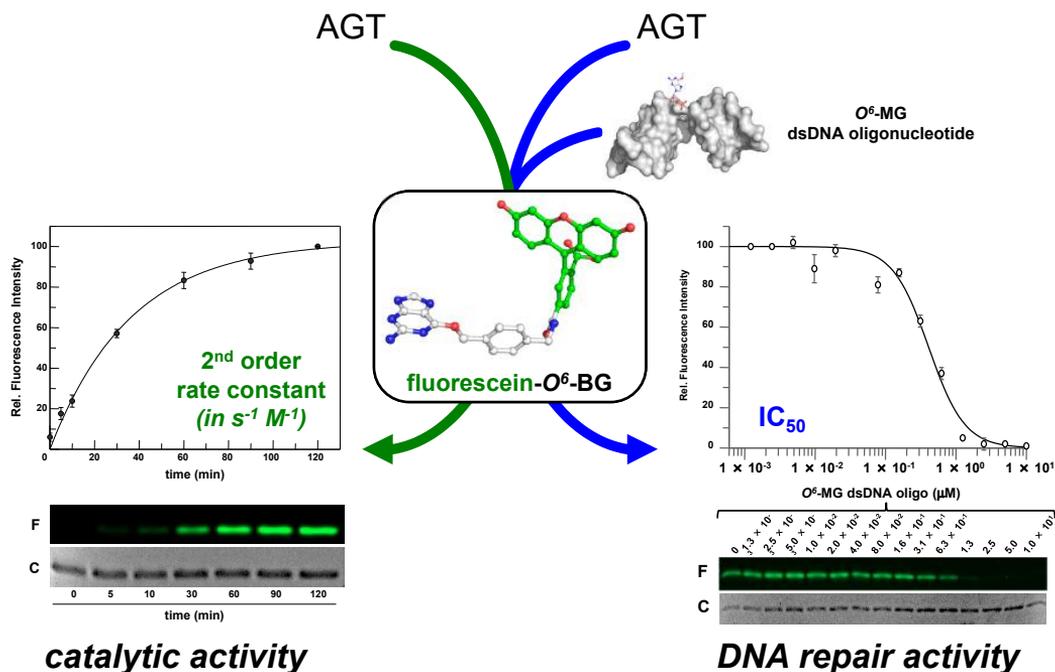
Various assays to measure AGT activity are reported in the literature. The first methods were based on the use of oligonucleotides carrying radioactive ( $^3\text{H}$  or  $^{14}\text{C}$ )  $O^6$ -alkylguanine groups. Proteinase K digestion was then carried out to measure the levels of marked S-methyl-cysteine in the lysate in an automatic amino acid analyser [53]. A very similar, but simpler and faster radioactive assay was used in another procedure with a  $^{32}\text{P}$ -terminal labelled oligonucleotide containing a modified guanine in a methylation-sensitive restriction enzyme sequence (as *Mbo* I). The AGT DNA repair activity thereby allowed the restriction enzyme to cut [54]. This procedure was also used by Ciaramella's group to identify for the first time the activity of SsOGT [52]. This test has the advantage of analysing the digested fragment directly by electrophoresis on a polyacrylamide gel [55].

It was therefore improved in terms of precision by the subsequent separation of the digested oligonucleotides by HPLC. The chromatographic separation allowed the calculation of the concentration of active AGT after measuring the radioactivity of the peak corresponding to the digested fragment [55]. Similarly, Moschel's group developed the analysis of hMGMT reaction products based on HPLC separation in 2002. This test investigated the degree of inhibition of oligonucleotides with  $O^6$ -MG or  $O^6$ -BG in different positions that varied from the 3' to the 5' end and whether they could be used as chemotherapy agents.  $\text{IC}_{50}$  values were obtained by quantifying the remaining active protein after the radioactive DNA reaction [56].

Although the assay measures the protein activity, the use of radioactive materials and chromatographic separations made these assays long, tedious, and unsafe.

An alternative approach was proposed in 2010 by the group of Carme Fàbrega, who set up an assay based on the thrombin DNA aptamer (TBA), a single-stranded 15 mer DNA oligonucleotide identified via Systematic Evolution of Ligands by EXponential enrichment (SELEX), which in its quadruplex form binds thrombin protease with high specificity and affinity [57]. In this assay, they put a fluorophore and a quencher to the TBA—the quadruplex structure of this oligonucleotide is compromised if a central  $O^6$ -MG is present, preventing the two probes to stay closer. An AGT's repair activity on the oligonucleotide allows the folding of the quadruplex structure and the Förster Resonance Energy Transfer (FRET) energy transfer takes place, resulting in a decrease of the fluorescence intensity [58].

Recently, the introduction of fluorescent derivatives of the  $O^6$ -BG (as SNAP Vista Green, New England Biolabs) made possible the development of a novel DNA alkyl-transferase assay. Because AGT covalently binds a benzyl-fluorescein moiety of its substrate after reaction, it is possible to immediately load the protein product on a SDS-PAGE—the *gel-imaging* analysis of the fluorescence intensity gives a direct measure of the protein activity because of the 1:1 stoichiometry of protein/substrate (Figure 3). Signals of fluorescent protein (corrected by the amount of loaded protein by Coomassie staining analysis) obtained at different times are plotted, and a second order reaction rate is determined [38,39,46,52,59,60]. This method can be applied to all AGTs that bind  $O^6$ -BG, with the exception of the *E. coli* Ada-C [61,62].



**Figure 3.** Innovative fluorescent AGT assay. The substrate could be used alone for the determination of the AGT catalytic activity, or in combination with a competitive non-fluorescent substrate (alkylated-DNA). In the latter case, an indirect measure of the DNA repair activity on natural substrates is determined (adapted from [63]).

Furthermore, an alkylated double strand DNA (dsDNA) oligonucleotide can be included in a competition assay with the fluorescein substrate. This non-fluorescent substrate lowers the final fluorescent signal on *gel imaging* analysis, depending on its concentration. In this way, it is possible to measure the activity of AGTs for their natural substrate, giving an indirect measure of methylation repair efficiency (Figure 3) [38,39,46,52,59,60]. By using this methodology, it was even possible to discriminate the SsOGT activity regarding the position of the  $O^6$ -MG on DNA (see below; [39]), in line with previous data on hMGMT [64].

### 3.2. Biochemical Properties of *S. solfataricus* OGT

The recombinant SsOGT protein, heterologously expressed in *E. coli*, has been fully characterized using the fluorescent assay described and summarized in Section 3.1, and some results are compiled in Table 3. In agreement with its origin, the protein showed optimal catalytic activity at 80 °C, although retaining a residual activity at lower temperatures (Table 3), and in a pH range between 5.0 and 8.0. As for the most part of many thermophilic enzymes, SsOGT is resistant over a wide range of reaction

conditions, such as ionic strength, organic solvents, common denaturing agents, and proteases [52,59]. Interestingly, chelating agents do not affect the activity of this enzyme. Crystallographic data clarified this observation, as the archaeal enzyme lacks a zinc ion in the structure [39], whereas this ion is important for correct folding of hMGMT [6].

### 3.3. Crystal Structure of SsOGT

All catalytic steps of the AGTs' activity (alkylated DNA recognition, DNA repair, irreversible trans-alkylation of the catalytic cysteine, recognition, and degradation of the alkylated protein) have been structurally characterized. Most information comes from the classic studies on hMGMT, as well as the Ada-C and OGT from *Escherichia coli* [5–8,49]. Other AGTs' structures are also available in the Protein Data Bank site (Figure 2a) (<http://www.rcsb.org/pdb/results/results.do?tabtoshow=Current&qrid=D3B02F3B>).

**Table 3.** Biochemical properties comparison among SNAP-tag®, SsOGT and the relative H<sup>5</sup> mutant.

Features		SNAP-tag®	SsOGT	SsOGT-H <sup>5</sup>
Molecular weight (kDa) <sup>a</sup>		23.0	17.0	17.0
T <sub>opt</sub> (°C)		37.0	80.0	75.0
Relative activity	at 25.0°C	80%	25%	<b>50%</b> <sup>b</sup>
	at 37.0°C	100%	45%	<b>65%</b>
	at 80°C	-	100%	95%
Catalytic activity at 37 °C (M <sup>-1</sup> s <sup>-1</sup> )		2.8 × 10 <sup>4</sup>	2.8 × 10 <sup>3</sup>	<b>1.6 × 10<sup>4</sup></b>
pH <sub>opt</sub>		6.0	7.5	6.0
Thermal stability T <sub>1/2</sub> (°C)		6 h (37)	3 h (70)	3 h (70)
Thermal stability T <sub>1/2</sub> at 37 °C (h)		6	>24	>24
Additives	NaCl	<0.3 M	>1.0 M	>1.0 M
	EDTA	no	yes	yes
	sarcosyl	no	>0.5%	>0.5%
	DDT	yes	no	no

<sup>a</sup> Data from [50,57,64]. <sup>b</sup> Enhancement with respect to the SsOGT (in bold).

As shown in Figure 1, all AGTs are inactivated after the reaction and degraded via proteasome, whereas in higher organisms, the degradation is preceded by protein ubiquitination [9]. It is a common view that the recognition of alkylated-AGTs is due by a conformational change; however, data on structure and properties of alkylated AGTs are limited because alkylation greatly destabilizes their folding [39]. The methylated-hMGMT and benzylated-hMGMT 3D structures were only obtained by flash-frozen crystals, showing that alkylation of the catalytic cysteine (C145) induces subtle conformational changes [6,7,65]. Consequently, these structures might not reflect the physiological conformation of the alkylated hMGMT [39].

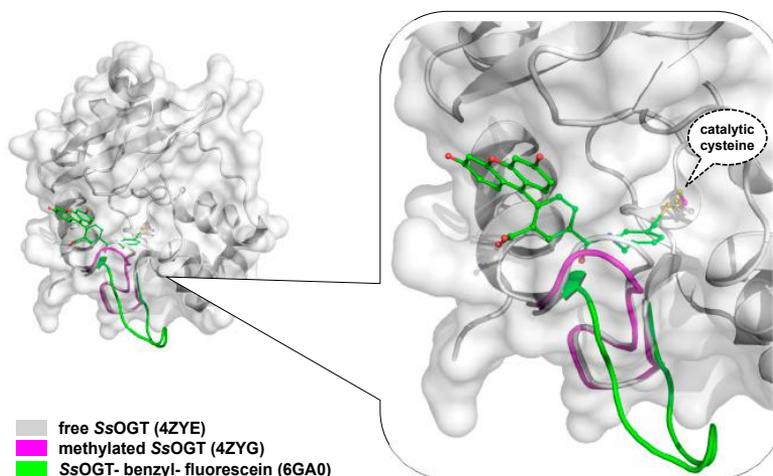
Concerning the interactions with the DNA, SsOGT binds methylated oligonucleotides. However, the repair activity depends on the position of the alkyl-group [39]. To efficiently repair the alkylated base on DNA double helix, the protein requires at least three bases from either the 5' or the 3' end. This is due to the necessary interactions formed with the double helix. Structural analysis confirmed these data [39].

To overcome the serious limitation to obtain structural data from mesophilic AGTs after reaction, studies have moved to thermostable homologues, based also on the knowledge that all AGTs share a common CTD domain structure (Figure 2b). In contrast to the human counterpart, alkylated SsOGT was soluble and relatively stable, thus allowing *in-deep* analysis of the protein in its post-reaction form [39]. Structural and biochemical analysis of the archaeal OGT, as well as after the reaction with a bulkier adduct in the active site (benzyl-fluorescein; [66]), suggested a possible mechanism of alkylation-induced SsOGT unfolding and degradation (Figure 4).

On the basis of their data, Perugino and co-workers suggested a general model for the mechanism of post-reaction AGT destabilization—the so called *active-site loop* moves towards the bulk solvent as a result of the covalent binding of alkyl adduct on the catalytic cysteine and the extent of the loop movement and dynamic correlates with the steric hindrance of the adduct [39,66] (Figure 4). The destabilization of this protein region triggers then the recognition of the alkylated protein by degradation pathway.

### 3.4. Biotechnological Applications of an Engineered SsOGT—the H<sup>5</sup> Mutant

As described in Section 1.2, the introduction of the SNAP-tag technology enabled a wide in vivo and in vitro labelling variety for biological studies by fusing any protein of interest (POI) to this *protein tag* [67]. However, being originated from hMGMT, the extension to extremophilic organisms and/or harsh reaction conditions is seriously limited.



**Figure 4.** Conformational changes of the SsOGT *active-site loop* after reaction with an O<sup>6</sup>-MG dsDNA oligonucleotide (in magenta; [39]), or with SNAP Vista Green substrate (in green; [66]).

By following the same approach used for the hMGMT as Kai Johnsson [26–30], an engineered version of SsOGT was produced [52,59]. This protein, called SsOGT-H<sup>5</sup>, contains five mutations in the helix-turn-helix domain, abolishing any DNA-binding activity [52]. In addition, a sixth mutation was made—in the *active site loop*, where serine residue was replaced by a glutamic acid at position 132 (S132E). This modification increased the catalytic activity of SsOGT [52,59], as it was observed in the engineered version of the hMGMT during the SNAP-tag development [26]. SsOGT-H<sup>5</sup> shows slightly lower heat stability in respect to the wild-type protein (Table 3), whereas the resistance to other denaturing agents is maintained. Moreover, SsOGT-H<sup>5</sup> is characterised by a surprisingly high catalytic activity at lower temperatures, keeping the rate of reaction to the physiological ones (Table 3) [52,59]. These characteristics make this mutant a potential alternative to SNAP-tag for in vivo and in vitro biotechnological applications. The stability against thermal denaturation allowed Miggiano and co-workers to obtain the structure of the protein after the reaction with the fluorescent substrate SNAP-Vista Green, revealing the peculiar destabilization of the *active site loop* after the alkylation of the active cysteine [66].

#### 3.4.1. In vitro Thermostable H<sup>5</sup>-Based Chimeras

The *Saccharolobus* OGT mutant has been firstly tested as *protein tag* fused to two thermostable *S. solfataricus* proteins heterologously expressed in *E. coli*. The chimeric proteins were correctly folded, and the tag did not interfere with the enzymatic activity of the tetrameric *S. solfataricus*  $\beta$ -glycosidase

(Ss $\beta$ gly) [59], nor with the hyperthermophile-specific DNA topoisomerase reverse gyrase [68–72]. Furthermore, the stability of H<sup>5</sup> made possible a heat treatment of the cell-free extract to remove most of the *E. coli* proteins and performing the  $\beta$ -glycosidase assay at high temperatures without the need of removing the tag [60].

### 3.4.2. Expression in Thermophilic Organisms Models

As the applicability of the thermostable tag under in vivo conditions is very important, the SsOGT-H<sup>5</sup> was also expressed in thermophilic organisms. The fluorescent AGT assay allows for the detection of the presence of SsOGT-H<sup>5</sup> both in living cells as well as in vitro in cell-free extracts [59,72]. To assay the activity to SsOGT-H<sup>5</sup>, it was necessary to choose models in which the endogenous AGT activity is suppressed. *Thermus thermophilus* is an *ogt*-species, showing only one *agt* homologue (TTHA1564), whose annotation corresponds to an alkyltransferase-like protein (ATL) [73]. ATLs are a class of proteins present in prokaryotes and lower eukaryotes [74], presenting aminoacidic motifs similar to those of AGTs' CTD, in which a tryptophan residue replaces the cysteine in the active site [75]. Like AGTs, ATLs use a helix-turn-helix motif to bind the minor groove of the DNA, but they do not repair it as they only recruit and interact with proteins involved in the nucleotide excision repair system [76,77].

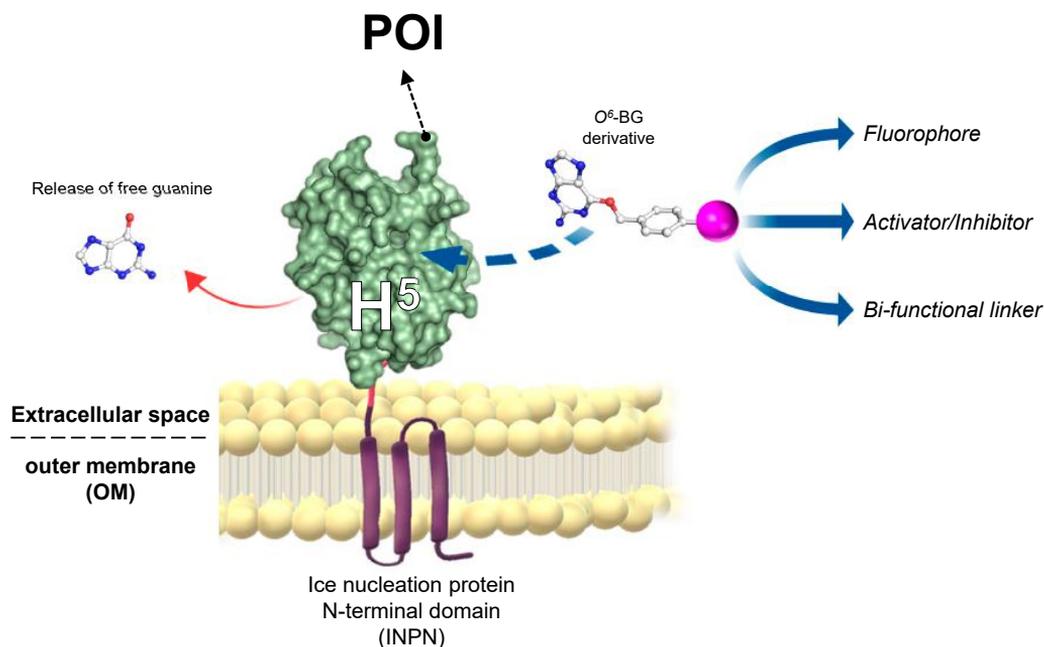
Although *T. thermophilus* is a natural *ogt* knockout organism, *Sulfolobus islandicus* possesses an *ogt* gene very similar to that of *S. solfataricus*, which was silenced by a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-based technique and then used as a host organism [72].

The fluorescent signal obtained by SDS-PAGE gel imaging revealed that SsOGT-H<sup>5</sup> not only is efficiently expressed in these thermophilic microorganisms, but it also showed that this tag was correctly folded and active, demonstrating the fact that SsOGT-H<sup>5</sup> might be used as an in vivo protein tag at high temperatures [59,72]. As is the case with SNAP-tag in human cells, the utilization of SsOGT-H<sup>5</sup> with different fluorescent substrates gives the opportunity to perform a multi-colour fluorescence study (see Table 1), by following a POI inside living “thermo cells” at different stages and localization.

### 3.4.3. The ASL<sup>tag</sup> System

As most biotechnological processes require harsh operational conditions, the immobilization of very robust enzymes on solid supports is often essential [78]. By definition, an immobilized enzyme is a “physically confined biocatalyst, which retains its catalytic activity and can be used repeatedly” [79]. Protein immobilisation offers several advantages, such as the catalysts' recovery and reuse, as well as the physical separation of the enzymes from the reaction mixture. Currently, different immobilisation strategies are available, from physical adsorption to covalent coupling [80–83]. However, all these procedures require purified biocatalysts and suffer from problems related to steric hindrance between the catalyst, the substrate, and the solid support, with increasing of costs and time for the production processes.

The introduction of “cell-based” immobilisation systems resulted in a significant improvement and reduces both time and costs of the process. One of the most widely used display strategies is the simultaneous heterologous expression of enzymes and their in vivo immobilisation on the external surface of Gram-negative bacteria cells, by the utilisation of the ice nucleation protein (INP) from *Pseudomonas syringae* [84,85]. Most recently, the N-terminal domain of INP (INPN) was used to produce a novel anchoring and self-labelling protein tag (hereinafter ASL<sup>tag</sup>). The ASL<sup>tag</sup> consists of two moieties, the INPN and the engineered and SsOGT-H<sup>5</sup> mutant (Figure 5) [86].



**Figure 5.** The novel anchoring and self-labelling protein tag (ASL<sup>tag</sup>) system. A protein of interest (POI) is genetically encoded with the tag, which in turn makes it anchored in the outer membrane and accessible for the covalent linkage to a desired chemical group (magenta sphere) by the activity of SsOGT-H<sup>5</sup> (adapted from [87]).

The INPN allows an *in vivo* immobilisation on *E. coli* outer membrane of enzymes of interest and their exposition to the solvent. The significant reduction of the costs related to the purification and immobilization is added to the overcoming of problems related to the recovery of enzymes by simple filtration or centrifugation methods [88]. SsOGT-H<sup>5</sup>, in turn, gives the unique opportunity to label immobilized enzymes with any desired chemical groups (opportunistically conjugated to the benzyl-guanine; in magenta in Figure 5) [27,59], dramatically expanding biotechnological applications of this new tool. Depending of the chemical group of choice, modulating the activity of enzymes fused with the ASL<sup>tag</sup> can be possible by introducing activator or inhibitor molecules (Figure 5). The ASL<sup>tag</sup> system was successfully employed for the expression and immobilization of monomeric biocatalysts, such as the thermostable carbonic anhydrase from *Saccharolobus solfataricus* (SspCA), as well as the tetrameric Ssβgly, without affecting their folding and catalytic activity [86]. Moreover, SspCA fused to the ASL<sup>tag</sup> showed an increase in residual activity of up to 30 % for a period of 10 days at 70 °C [87], representing a huge advantage in pushing beyond reactions in bioreactors and in the reutilization of biocatalysts.

#### 4. *Pyrococcus furiosus* and *Thermotoga neapolitana* OGT

To extend the SNAP-tag technology to hyperthermophilic microorganisms for *in vivo* studies, an O<sup>6</sup>-alkylguanine-DNA alkyltransferase has been recently characterized from the archaeon *Pyrococcus furiosus* [89]. This extremophilic microorganism was originally isolated from hot marine sediments in Vulcano Island (Italy) [90], with an optimum growth temperature around 100 °C, thus thriving under extremely harsh conditions. Like those of other thermophilic Archaea, its enzymes are extremely thermostable and can be used in various biotechnological applications. For example, DNA polymerase I, also known as *Pfu*DNA polymerase, is one of the most famous and frequently used enzymes from *P. furiosus* because of its high activity, thermostability, and strong 3'-5' proof-reading activity [91]. The first demonstration of an OGT activity in *P. furiosus* was in 1998, when Margison and co-workers identified a protein of 22 kDa, whose catalytic activity was abolished by the O<sup>6</sup>-BG pseudo-substrate.

The PF1878 ORF is relative to a protein of 20.1 kDa. From its primary structure, the relative polypeptide seems to be closely related to the MGMT from *Pyrococcus kodakarensis* KOD1 (*Pk*-MGMT) [48,89,92]. The extreme thermostability was confirmed by in vitro biochemical studies on the heterologous expressed and characterized OGT protein from *P. furiosus* *Pfu*OGT. This enzyme was active on BG-fluorescent substrates, thus allowing the competitive assay with methylated dsDNA. However, the experiments were performed at 65 °C instead of the standard procedure at 50 °C, as described for *Ss*OGT [39,59,60], due to the strong thermophilicity of this enzyme. This behaviour was effectively confirmed by differential scan fluorimetry analysis where the temperature melting ( $T_m$ ) of *Pfu*OGT was found to be 80 °C, much higher than that of *Ss*OGT (68 °C) [89]. It is worth noting that, in order to obtain a the sigmoidal melting curve for *Pfu*OGT, a slower heating rate (10 min/°C × cycle) was set up, whereas the  $T_m$  value measurement is usually performed at 1 min/°C × cycle [93].

*Thermotoga neapolitana* is a hyperthermophilic Gram-negative bacterium of the order of Thermotogales [94–96], which are excellent models for genetic engineering and biotechnological applications [97–100]. The CTN1690 ORF shows a clear homology of the  $O^6$ -alkylguanine-DNA-alkyl-transferase. SDS-PAGE gel imaging analysis on lyophilized *T. neapolitana* cells incubated with the AGT fluorescent substrate showed a strong fluorescent signal with a molecular weight close to that of *Ss*OGT. The observed molecular weight and, above all, the sensitivity to the  $O^6$ -BG derivative, led to the cloning and heterologous expression of the *Thermotoga neapolitana* OGT protein (*Tn*OGT) in *E. coli* [89]. This protein, like most AGTs, has a role in DNA repair, as confirmed by competitive fluorescent assay in the presence of methylated dsDNA. As shown in Figure 3, the  $IC_{50}$  value was similar to that obtained for *Ss*OGT. Surprisingly, the enzyme from *T. neapolitana* exhibited a very high activity at low temperatures [89], similar to that possessed by the mutant *Ss*OGT-H<sup>5</sup> (Table 3) [52,59]. Superimposition analysis between a *Tn*OGT 3D model and the free form of *Ss*OGT (ID PDB: 4ZYE) revealed in both structures the presence of a serine residue in the active site loop (S132 in *Ss*OGT, see Figure 2a), which was replaced in *Ss*OGT-H<sup>5</sup> by a glutamic acid to improve its activity at lower temperatures. Interestingly, some residues are missing in *Tn*OGT that play an important role in stabilizing *Ss*OGT. In particular, the ionic interactions that play a crucial role in the stability of the *Saccharolobus* enzyme at high temperatures, such as the pair R133-D27 [39] and the *K-48 network* [60], are largely replaced by hydrophobic residues in the *Thermotoga* homolog. Evidently, different residues and mechanisms of stabilization may contribute to its exceptional catalytic activity at moderate temperatures and the high thermal stability.

## 5. Future Perspectives

The interest shown from the important insights of this class of small proteins led to novel biotechnological applications [101]. Studies on thermophilic AGTs represent a unique opportunity for structural analysis and, in the case of the *S. solfataricus* protein, for the identification of conformational changes after the trans-alkylation reaction, which are detectable with mesophilic AGTs, as the alkylated form are rapidly destabilized [6]. These results could have a wide impact, especially in medical fields for the design of novel hMGMT inhibitors to be used in cancer therapy [102]. Furthermore, given their small size, thermophilic enzymes are very useful for studying general stabilization mechanisms at high temperatures (as for *Pk*-MGMT and *Ss*OGT), which can then be applied to mesophilic enzymes. Searching for alternative *Ss*OGT homologues was clearly useful, leading to the identification of AGTs that are more resistant to thermal denaturation (*Pfu*OGT) or to enzymes with a higher reaction rate at all tested temperatures (*Tn*OGT).

Concerning biotechnology, the use of a modified hMGMT as *protein tag* opened the possibility to generalise this method—a targeted mutagenesis on a thermostable OGT by following a *rational approach* led to the characterization of *Ss*OGT-H<sup>5</sup>, applicable to in vitro harsh reaction conditions and to in vivo (hyper)thermophilic model organisms. On the other hand, by an *irrational approach* (random mutagenesis) it is also possible to enhance their catalytic activity [103], or modify the substrate

specificity of these enzymes, making them active on benzyl-cytosine ( $O^2$ -BC) derivatives, such as that which happened for the production of the CLIP-tag [104].

This knowledge could be the starting point of developing a new engineered *thermo*-SNAP-tag to be employed in particular biotechnological fields, from in vivo studies in (hyper)thermophilic microorganisms (such as the in vivo CRISPR-Cas immune system in *P. furiosus* [105,106]) to industrial processes that require high temperatures or, in general, harsh reaction conditions.

**Author Contributions:** R.M. (Rosanna Mattosovich) and R.M. (Rosa Merlo) equally contributed to the present review article by writing the manuscript and joined as First Author. R.M. (Riccardo Miggiano) analysed the 3D structures, prepared Table 2, and wrote the relative text. A.V. contributed to the editing of the manuscript and to the collecting part of the references. G.P. contributed to the editing of the texts and the drawing of figures. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Ministero dell’Istruzione Università e Ricerca (MIUR) National Operational Program (PON) Research and Innovation 2014-2020 (CCI 2014IT16M2OP005), European Social Fund, Action I.1 “Innovative Doctorates with Industrial characterization”.

**Acknowledgments:** G.P. would like to thank all the authors, Elena and Elisa Perugino for their efforts in writing this work and in their technical assistance, but mainly for their human support during the difficult and delicate period of staying at home following the COVID-19 outbreak. All authors strongly thank the reviewers for their important corrections and changes, increasing the scientific level of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

AGT	$O^6$ -alkyl-guanine-DNA-alkyl-transferase
CLIP-tag	engineered version of SNAP-tag active on $O^2$ -BC
FP	auto-fluorescent protein
hMGMT	human $O^6$ -methyl-guanine-DNA-alkyl-transferase
MGMT	$O^6$ -methyl-guanine-DNA-alkyl-transferase
$O^2$ -BC	$O^2$ -benzyl-cytosine
$O^6$ -AG	$O^6$ -alkyl-guanine
$O^6$ -BG	$O^6$ -benzyl-guanine
$O^6$ -MG	$O^6$ -methyl-guanine
OGT	$O^6$ -alkyl-guanine-DNA-alkyl-transferase
SNAP-tag	engineered version of hMGMT for biotech purposes

## References

- Liu, L.; Gerson, S.L. Targeted modulation of MGMT: Clinical implications. *Clin. Cancer Res.* **2006**, *12*, 328–331. [[CrossRef](#)]
- Leonard, G.A.; Thomson, J.; Watson, W.P.; Brown, T. High-resolution structure of a mutagenic lesion in DNA. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9573–9576. [[CrossRef](#)] [[PubMed](#)]
- Drablos, F.; Feyzi, E.; Aas, P.A.; Vaagbø, C.B.; Kavli, B.; Bratlie, M.S.; Peña-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H.E. Alkylation damage in DNA and RNA-repair mechanisms and medical significance. *DNA Repair* **2004**, *11*, 1389–1407. [[CrossRef](#)] [[PubMed](#)]
- Pegg, A.E. Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. *Chem. Res. Toxicol.* **2011**, *24*, 618–639. [[CrossRef](#)] [[PubMed](#)]
- Duguid, E.M.; Rice, P.A.; He, C. The structure of the human AGT protein bound to DNA and its implications for damage detection. *J. Mol. Biol.* **2005**, *350*, 657–666. [[CrossRef](#)] [[PubMed](#)]
- Daniels, D.S.; Mol, C.D.; Arvai, A.S.; Kanugula, S.; Pegg, A.E.; Tainer, J.A. Active and alkylated human AGT structures: A novel zinc site, inhibitor and extrahelical base binding. *EMBO J.* **2000**, *19*, 1719–1730. [[CrossRef](#)] [[PubMed](#)]
- Daniels, D.S.; Woo, T.T.; Luu, K.X.; Noll, D.M.; Clarke, N.D.; Pegg, A.E.; Tainer, J.A. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* **2004**, *11*, 714–720. [[CrossRef](#)]

8. Tubbs, J.L.; Pegg, A.E.; Tainer, J.A. DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O<sup>6</sup>-alkylguanine-DNA-alkyltransferase and its implications for cancer chemotherapy. *DNA Repair* **2007**, *6*, 1100–1115. [[CrossRef](#)]
9. Xu-Welliver, M.; Pegg, A.E. Degradation of the alkylated form of the DNA repair protein, O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *Carcinogenesis* **2002**, *23*, 823–830. [[CrossRef](#)]
10. Gerson, S.L. MGMT: Its role in cancer aetiology and cancer therapeutics. *Nat. Rev. Cancer* **2004**, *4*, 296–307. [[CrossRef](#)]
11. Sabharwal, A.; Middleton, M.R. Exploiting the role of O<sup>6</sup>-methylguanine-DNA-methyltransferase (MGMT) in cancer therapy. *Curr. Opin. Pharmacol.* **2006**, *6*, 355–363. [[CrossRef](#)] [[PubMed](#)]
12. Zhong, Y.; Huang, Y.; Huang, Y.; Zhang, T.; Ma, C.; Zhang, S.; Fan, W.; Chen, H.; Qian, J.; Lu, D. Effects of O<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT) polymorphisms on cancer: A meta-analysis. *Mutagenesis* **2010**, *25*, 83–95. [[CrossRef](#)] [[PubMed](#)]
13. Kaina, B.; Christmann, M. DNA repair in personalized brain cancer therapy with temozolomide and nitrosoureas. *DNA Repair* **2019**, *78*, 128–141. [[CrossRef](#)] [[PubMed](#)]
14. Khan, O.; Middleton, M.R. The therapeutic potential of O<sup>6</sup>-alkylguanine DNA alkyltransferase inhibitors. *Expert Opin. Investig. Drugs* **2007**, *10*, 1573–1584. [[CrossRef](#)] [[PubMed](#)]
15. Paranjpe, A.; Zhang, R.; Ali-Osman, F.; Bobustuc, G.C.; Srivenugopal, K.S. Disulfiram is a direct and potent inhibitor of human O<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT) in brain tumor cells and mouse brain and markedly increases the alkylating DNA damage. *Carcinogenesis* **2014**, *35*, 692–702. [[CrossRef](#)] [[PubMed](#)]
16. Rabik, C.A.; Dolan, M.E. Molecular mechanisms of resistance and toxicity associated with platinating agents. *Cancer Treat. Rev.* **2007**, *33*, 9–23. [[CrossRef](#)]
17. Kaina, B.; Margison, G.P.; Christmann, M. Targeting O<sup>6</sup>-methylguanine-DNA methyltransferase with specific inhibitors as a strategy in cancer therapy. *Cell Mol. Life Sci.* **2010**, *67*, 3663–3681. [[CrossRef](#)]
18. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W.W.; Prasher, D.C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802–805. [[CrossRef](#)]
19. Tsien, R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **1998**, *67*, 509–554. [[CrossRef](#)]
20. di Guan, C.; Li, P.; Riggs, P.D.; Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **1988**, *67*, 21–30. [[CrossRef](#)]
21. Schmidt, T.G.M.; Koepke, J.; Frank, R.; Skerra, A. Molecular Interaction Between the Strep-tag Affinity Peptide and its Cognate Target, Streptavidin. *J. Mol. Biol.* **1996**, *255*, 753–766. [[CrossRef](#)] [[PubMed](#)]
22. Ren, L.; Chang, E.; Makky, K.; Haas, A.L.; Kaboord, B.; Walid Qoronfle, M. Glutathione S-transferase pull-down assays using dehydrated immobilized glutathione resin. *Anal. Biochem.* **2003**, *322*, 164–169. [[CrossRef](#)] [[PubMed](#)]
23. Donini, S.; Ferraris, D.M.; Miggiano, R.; Massarotti, A.; Rizzi, M. Structural investigations on orotate phosphoribosyltransferase from *Mycobacterium tuberculosis*, a key enzyme of the *de novo* pyrimidine biosynthesis. *Sci. Rep.* **2017**, *7*, 1180. [[CrossRef](#)] [[PubMed](#)]
24. Donini, S.; Garavaglia, S.; Ferraris, D.M.; Miggiano, R.; Mori, S.; Shibayama, K.; Rizzi, M. Biochemical and structural investigations on phosphoribosylpyrophosphate synthetase from *Mycobacterium smegmatis*. *PLoS ONE* **2017**, *12*, e0175815. [[CrossRef](#)] [[PubMed](#)]
25. LaVallie, E.R.; DiBlasio, E.A.; Kovacic, S.; Grant, K.L.; Schendel, P.F.; McCoy, J.M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio Technol.* **1993**, *11*, 187–193. [[CrossRef](#)]
26. Juillerat, A.; Gronemeyer, T.; Keppler, A.; Gendreizig, S.; Pick, H.; Vogel, H.; Johnsson, K. Directed evolution of O<sup>6</sup>-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem. Biol.* **2003**, *10*, 313–317. [[CrossRef](#)]
27. Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **2003**, *21*, 86–89. [[CrossRef](#)]
28. Kindermann, M.; George, N.; Johnsson, N.; Johnsson, K. Covalent and selective immobilization of fusion proteins. *J. Am. Chem. Soc.* **2003**, *125*, 7810–7811. [[CrossRef](#)]
29. Gronemeyer, T.; Chidley, C.; Juillerat, A.; Heinis, C.; Johnsson, K. Directed evolution of O<sup>6</sup>-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng. Des. Sel.* **2006**, *19*, 309–316. [[CrossRef](#)]
30. Hinner, M.J.; Johnsson, K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **2010**, *21*, 766–776. [[CrossRef](#)]

31. Huber, W.; Perspicace, S.; Kohler, J.; Müller, F.; Schlatter, D. SPR-based interaction studies with small molecular weight ligands using hAGT fusion proteins. *Anal. Biochem.* **2004**, *333*, 280–288. [[CrossRef](#)] [[PubMed](#)]
32. Niesen, J.; Sack, M.; Seidel, M.; Fendel, R.; Barth, S.; Fischer, R.; Stein, C. SNAP-tag technology: A useful tool to determine affinity constants and other functional parameters of novel anti-body fragments. *Bioconjug. Chem.* **2016**, *27*, 1931–1941. [[CrossRef](#)] [[PubMed](#)]
33. Valenti, A.; Napoli, A.; Ferrara, M.C.; Nadal, M.; Rossi, M.; Ciaramella, M. Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* **2006**, *34*, 2098–2108. [[CrossRef](#)] [[PubMed](#)]
34. Kanugula, S.; Pegg, A.E. Alkylation damage repair protein O<sup>6</sup>-alkyl-guanine-DNA alkyltransferase from the hyperthermophiles *Aquifex aeolicus* and *Archaeoglobus fulgidus*. *Biochem. J.* **2003**, *375*, 449–455. [[CrossRef](#)]
35. Leclere, M.M.; Nishioka, M.; Yuasa, T.; Fujiwara, S.; Takagi, M.; Imanaka, T. The O<sup>6</sup>-methylguanine-DNA methyltransferase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1: A thermostable repair enzyme. *Mol. Gen. Genet.* **1998**, *258*, 69–77. [[CrossRef](#)]
36. Skorvaga, M.; Raven, N.D.; Margison, G.P. Thermostable archaeal O<sup>6</sup>-alkylguanine-DNA alkyltransferases. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6711–6715. [[CrossRef](#)]
37. Fang, Q.; Kanugula, S.; Pegg, A.E. Function of domains of human O<sup>6</sup>-alkyl-guanine-DNA alkyltransferase. *Biochemistry* **2005**, *44*, 15396–15405. [[CrossRef](#)]
38. Miggiano, R.; Casazza, V.; Garavaglia, S.; Ciaramella, M.; Perugino, G.; Rizzi, M.; Rossi, F. Biochemical and structural studies of the *Mycobacterium tuberculosis* O<sup>6</sup>-methylguanine methyltransferase and mutated variants. *J. Bacteriol.* **2013**, *195*, 2728–2736. [[CrossRef](#)]
39. Perugino, G.; Miggiano, R.; Serpe, M.; Vettone, A.; Valenti, A.; Lahiri, S.; Rossi, F.; Rossi, M.; Rizzi, M.; Ciaramella, M. Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic Acids Res.* **2015**, *43*, 8801–8816. [[CrossRef](#)]
40. Kanugula, S.; Pegg, A.E. Novel DNA Repair Alkyltransferase from *Caenorhabditis elegans*. *Env. Mol. Mut.* **2001**, *38*, 235–243. [[CrossRef](#)]
41. Serpe, M.; Forenza, C.; Adamo, A.; Russo, N.; Perugino, G.; Ciaramella, M.; Valenti, A. The DNA alkylguanine DNA alkyltransferase-2 (AGT-2) of *Caenorhabditis elegans* is involved in meiosis and early development under physiological conditions. *Sci. Rep.* **2019**, *9*, 6889. [[CrossRef](#)] [[PubMed](#)]
42. Presta, L.G.; Rose, G.D. Helix signals in proteins. *Science* **1988**, *240*, 1632–1641. [[CrossRef](#)] [[PubMed](#)]
43. Richardson, J.S.; Richardson, D.C. Amino acid preferences for specific locations at the ends of  $\alpha$ -helices. *Science* **1988**, *240*, 1648–1652. [[CrossRef](#)] [[PubMed](#)]
44. Koscielska-Kasprzak, K.; Cierpicki, T.; Otlewski, J. Importance of alpha-helix N-capping motif in stabilization of betabetaalpha fold. *Protein Sci.* **2003**, *12*, 1283–1289. [[CrossRef](#)] [[PubMed](#)]
45. Chan, M.K.; Mukund, S.; Kletzin, A.; Adams, M.W.W.; Rees, D.C. Structure of a hyperthermo-philic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **1995**, *267*, 1463–1469. [[CrossRef](#)]
46. Miggiano, R.; Perugino, G.; Ciaramella, M.; Serpe, M.; Rejman, D.; Páv, O.; Pohl, R.; Garavaglia, S.; Lahiri, S.; Rizzi, M. Crystal structure of *Mycobacterium tuberculosis* O<sup>6</sup>-methylguanine-DNA methyltransferase protein clusters assembled on to damaged DNA. *Biochem. J.* **2016**, *473*, 123–133. [[CrossRef](#)]
47. Rice, D.W.; Yip, K.S.; Stillman, T.J.; Britton, K.L.; Fuertes, A.; Connerton, J.; Pasquo, A.; Scandura, R.; Engel, P.C. Insight into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase. *FEMS Microbiol. Rev.* **1996**, *18*, 105–117. [[CrossRef](#)]
48. Hashimoto, H.; Inoue, T.; Nishioka, M.; Fujiwara, S.; Takagi, M.; Imanaka, T.; Kai, Y. Hyperthermostable protein structure maintained by intra and inter-helix ion-pairs in archaeal O<sup>6</sup>-methylguanine-DNA methyltransferase. *J. Mol. Biol.* **1999**, *292*, 707–716. [[CrossRef](#)]
49. Moore, M.H.; Gulbis, J.M.; Dodson, E.J.; Demple, B.; Moody, P.C. Crystal structure of a suicidal DNA repair protein: The Ada O<sup>6</sup>-methylguanineDNA methyltransferase from *E. coli*. *EMBO J.* **1994**, *13*, 1495–1501. [[CrossRef](#)]
50. Roberts, A.; Pelton, J.G.; Wemmer, D.E. Structural studies of MJ1529, an O<sup>6</sup>-methylguanine-DNA methyltransferase. *Magn. Reson. Chem.* **2006**, *44*, S71–S82. [[CrossRef](#)]
51. Zillig, W.; Stetter, K.O.; Wunderl, S.; Schulz, W.; Priess, H.; Scholz, I. The *Sulfolobus*-“*Caldariella*” group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.* **1980**, *125*, 259–269. [[CrossRef](#)]

52. Perugino, G.; Vettone, A.; Illiano, G.; Valenti, A.; Ferrara, M.C.; Rossi, M.; Ciaramella, M. Activity and regulation of archaeal DNA alkyltransferase: Conserved protein involved in repair of DNA alkylation damage. *J. Biol. Chem.* **2012**, *287*, 4222–4231. [[CrossRef](#)] [[PubMed](#)]
53. Olsson, M.; Lindahl, T. Repair of alkylated DNA in *Escherichia coli*. Methyl group transfer from O<sup>6</sup>-methylguanine to a protein cysteine residue. *J. Biol. Chem.* **1980**, *255*, 10569–10571. [[PubMed](#)]
54. Wu, R.S.; Hurst-Calderone, S.; Kohn, K.W. Measurement of O<sup>6</sup>-alkylguanine-DNA-alkyltransferase activity in human cells and tumor tissues by restriction endonuclease inhibition. *Cancer Res.* **1987**, *47*, 6229–6235. [[PubMed](#)]
55. Klein, S.; Oesch, F. Assay for O<sup>6</sup>-alkylguanine-DNA-alkyltransferase using oligonucleotides containing O<sup>6</sup>-methylguanine in a BamHI recognition site as substrate. *Anal. Biochem.* **1992**, *205*, 294–299. [[CrossRef](#)]
56. Luu, K.X.; Kanugula, S.; Pegg, A.E.; Pauly, G.T.; Moschel, R.C. Repair of oligodeoxyribonucleotides by O(6)-alkylguanine-DNA alkyltransferase. *Biochemistry* **2002**, *41*, 8689–8697. [[CrossRef](#)]
57. Bock, L.C.; Griffin, L.C.; Latham, J.A.; Vermaas, E.H.; Toole, J.J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **1992**, *355*, 564–566. [[CrossRef](#)]
58. Tintoré, M.; Aviñó, A.; Ruiz, F.M.; Eritja, R.; Fábrega, C. Development of a novel fluorescence assay based on the use of the thrombin binding aptamer for the detection of O<sup>6</sup>-alkyl-guanine-DNA alkyltransferase activity. *J. Nuc. Ac.* **2010**, *2010*, 632041.
59. Vettone, A.; Serpe, M.; Hidalgo, A.; Berenguer, J.; del Monaco, G.; Valenti, A.; Ciaramella, M.; Perugino, G. A novel thermostable protein-tag: Optimization of the *Sulfolobus solfataricus* DNA-alkyl-transferase by protein engineering. *Extremophiles* **2016**, *20*, 1–13. [[CrossRef](#)]
60. Morrone, C.; Miggiano, R.; Serpe, M.; Massarotti, A.; Valenti, A.; Del Monaco, G.; Rossi, M.; Rossi, F.; Rizzi, M.; Perugino, G. Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase. *Biochim. Biophys. Acta* **2017**, *1861*, 86–96. [[CrossRef](#)]
61. Elder, R.H.; Margison, G.P.; Rafferty, J.A. Differential inactivation of mammalian and *Escherichia coli* O<sup>6</sup>-alkylguanine-DNA alkyltransferases by O<sup>6</sup>-benzylguanine. *Biochem. J.* **1994**, *298*, 231–235. [[CrossRef](#)] [[PubMed](#)]
62. Goodtzova, K.; Kanugula, S.; Edara, S.; Pauly, G.T.; Moschel, R.C.; Pegg, A.E. Repair of O<sup>6</sup>-benzylguanine by the *Escherichia coli* Ada and Ogt and the human O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *J. Biol. Chem.* **1997**, *272*, 8332–8339. [[CrossRef](#)] [[PubMed](#)]
63. Miggiano, R.; Valenti, A.; Rossi, F.; Rizzi, M.; Perugino, G.; Ciaramella, M. Every OGT Is Illuminated... by Fluorescent and Synchrotron Lights. *Int. J. Mol. Sci.* **2017**, *18*, 2613. [[CrossRef](#)] [[PubMed](#)]
64. Melikishvili, M.; Rasimas, J.J.; Pegg, A.E.; Fried, M.G. Interactions of human O(6)-alkylguanine-DNA alkyltransferase (AGT) with short double-stranded DNAs. *Biochemistry* **2008**, *47*, 13754–13763. [[CrossRef](#)]
65. Brunk, E.; Mollwitz, B.; Rothlisberger, U. Mechanism to Trigger Unfolding in O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase. *ChemBioChem* **2013**, *14*, 703–710. [[CrossRef](#)]
66. Rossi, F.; Morrone, C.; Massarotti, A.; Ferraris, D.M.; Valenti, A.; Perugino, G.; Miggiano, R. Crystal structure of a thermophilic O<sup>6</sup>-alkylguanine-DNA alkyltransferase-derived self-labeling protein-tag in covalent complex with a fluorescent probe. *Biochem. Biophys. Res. Comm.* **2018**, *500*, 698–703. [[CrossRef](#)]
67. Keppler, A.; Pick, H.; Arrivoli, C.; Vogel, H.; Johnsson, K. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc. Natl. Acad. Sci. USA* **2004**, *10*, 9955–9959. [[CrossRef](#)]
68. Valenti, A.; Perugino, G.; D’Amaro, A.; Cacace, A.; Napoli, A.; Rossi, M.; Ciaramella, M. Dissection of reverse gyrase activities: Insight into the evolution of a thermostable molecular machine. *Nucleic Acids Res.* **2008**, *36*, 4587–4597. [[CrossRef](#)]
69. Valenti, A.; Perugino, G.; Nohmi, T.; Rossi, M.; Ciaramella, M. Inhibition of translesion DNA polymerase by archaeal reverse gyrase. *Nucleic Acids Res.* **2009**, *37*, 4287–4295. [[CrossRef](#)]
70. Perugino, G.; Valenti, A.; D’Amaro, A.; Rossi, M.; Ciaramella, M. Reverse gyrase and genome stability in hyperthermophilic organisms. *Biochem. Soc. Trans.* **2009**, *37*, 69–73. [[CrossRef](#)]
71. Valenti, A.; Perugino, G.; Rossi, M.; Ciaramella, M. Positive supercoiling in thermophiles and mesophiles: Of the good and evil. *Biochem. Soc. Trans.* **2011**, *39*, 58–63. [[CrossRef](#)] [[PubMed](#)]
72. Visone, V.; Han, W.; Perugino, G.; Del Monaco, G.; She, Q.; Rossi, M.; Valenti, A.; Ciaramella, M. In vivo and in vitro protein imaging in thermophilic archaea by exploiting a novel protein tag. *PLoS ONE* **2017**, *10*, e0185791. [[CrossRef](#)] [[PubMed](#)]

73. Morita, R.; Nakagawa, N.; Kuramitsu, S.; Masui, R. An O<sup>6</sup>-methylguanine-DNA methyltransferase-like protein from *Thermus thermophilus* interacts with a nucleotide excision repair protein. *J. Biochem.* **2008**, *144*, 267–277. [[CrossRef](#)]
74. Schärer, O.D. Alkyltransferase-like Proteins: Brokers Dealing with Alkylated DNA Bases. *Mol. Cell* **2012**, *47*, 3–4. [[CrossRef](#)] [[PubMed](#)]
75. Tubbs, J.L.; Latypov, V.; Kanugula, S.; Butt, A.; Melikishvili, M.; Kraehenbuehl, R.; Fleck, O.; Marriott, A.; Watson, A.J.; Verbeek, B.; et al. Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **2009**, *459*, 808–813. [[CrossRef](#)] [[PubMed](#)]
76. Latypov, V.F.; Tubbs, J.L.; Watson, A.J.; Marriott, A.S.; McGown, G.; Thorncroft, M.; Wilkinson, O.J.; Senthong, P.; Butt, A.; Arvai, A.S.; et al. AtI1 regulates choice between global genome and transcription-coupled repair of O(6)-alkylguanines. *Mol. Cell.* **2012**, *47*, 50–60. [[CrossRef](#)] [[PubMed](#)]
77. Lahiri, S.; Rizzi, M.; Rossi, F.; Miggiano, R. *Mycobacterium tuberculosis* UvrB forms dimers in solution and interacts with UvrA in the absence of ligands. *Proteins* **2018**, *86*, 98–109. [[CrossRef](#)]
78. Zhou, Z.; Hartmann, M. Progress in enzyme immobilization in ordered mesoporous materials and related applications. *Chem. Soc. Rev.* **2013**, *42*, 3894–3912. [[CrossRef](#)]
79. Tosa, T.; Mori, T.; Fuse, N.; Chibata, I. Studies on continuous enzyme reactions. I. Screening of carriers for preparation of water-insoluble aminoacylase. *Enzymologia* **1966**, *31*, 214–224.
80. Mohamad, N.R.; Che Marzuki, N.H.; Buang, N.A.; Huyop, F.; Wahab, R.A. An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. *Biotech. Biotech. Equip.* **2015**, *29*, 205–220. [[CrossRef](#)]
81. Nguyen, H.H.; Kima, M. An Overview of Techniques in Enzyme Immobilization. *App. Sci. Conv. Tech.* **2017**, *26*, 157–163. [[CrossRef](#)]
82. Jakub Zdarta, J.; Meyer, A.S.; Jesionowski, T.; Pinelo, M. A General Overview of Support Materials for Enzyme Immobilization: Characteristics, Properties, Practical Utility. *Catalysts* **2018**, *8*, 92. [[CrossRef](#)]
83. Sirisha, V.L.; Jain, A.; Jain, A. Chapter Nine—Enzyme Immobilization: An Overview on Methods, Support Material, and Applications of Immobilized Enzymes. *Adv. Food Nut. Res.* **2016**, *79*, 179–211.
84. Gurian-Sherman, D.; Lindow, S.E. Bacterial ice nucleation: Significance and molecular basis. *FASEB J.* **1993**, *7*, 1338–1343. [[CrossRef](#)] [[PubMed](#)]
85. Cochet, N.; Widehem, P. Ice crystallization by *Pseudomonas syringae*. *Appl. Microbiol. Biotechnol.* **2000**, *54*, 153–161. [[CrossRef](#)]
86. Merlo, R.; Del Prete, S.; Valenti, A.; Mattosovich, R.; Carginale, V.; Supuran, C.T.; Capasso, C.; Perugino, P. An AGT-based protein-tag system for the labelling and surface immobilization of enzymes on *E. coli* outer membrane. *J. Enz. Inhib. Med. Chem.* **2019**, *34*, 490–499. [[CrossRef](#)]
87. Del Prete, S.; Merlo, R.; Valenti, A.; Mattosovich, R.; Rossi, M.; Carginale, M.; Supuran, C.T.; Perugino, G.; Capasso, C. Thermostability enhancement of the  $\alpha$ -carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* by using the anchoring-and-selflabelling-protein-tag system (ASL<sup>tag</sup>). *J. Enz. Inhib. Med. Chem.* **2019**, *34*, 946–954. [[CrossRef](#)]
88. Del Prete, S.; Perfetto, R.; Rossi, M.; Alasmay, F.A.S.; Osman, S.M.; AlOthman, Z.; Supuran, C.T.; Capasso, C. A one-step procedure for immobilising the thermostable carbonic anhydrase (SspCA) on the surface membrane of *Escherichia coli*. *J. Enz. Inhib. Med. Chem.* **2017**, *32*, 1120–1128. [[CrossRef](#)]
89. Mattosovich, R.; Merlo, R.; Fontana, A.; d’Ippolito, G.; Terns, M.P.; Watts, E.A.; Valenti, A.; Perugino, P. A journey down to hell: New thermostable protein-tags for biotechnology at high temperatures. *Extremophiles* **2020**, *24*, 81–91. [[CrossRef](#)]
90. Fiala, G.; Stetter, K.O. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100 °C. *Arch. Microb.* **1986**, *145*, 56–61. [[CrossRef](#)]
91. Lundberg, K.S.; Shoemaker, D.D.; Adams, M.W.W.; Short, J.M.; Sorge, J.A.; Mathur, E.J. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **1991**, *108*, 1–6. [[CrossRef](#)]
92. Nishikori, S.; Shiraki, K.; Yokota, K.; Izumikawa, N.; Fujiwara, S.; Hashimoto, H.; Imanaka, T.; Takagi, M. Mutational effects on O<sup>6</sup>-methylguanine-DNA methyltransferase from hyperthermophile: Contribution of ion-pair network to protein thermostability. *J. Biochem.* **2004**, *135*, 525–532. [[CrossRef](#)] [[PubMed](#)]
93. Niesen, F.H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc.* **2007**, *2*, 2212–2221. [[CrossRef](#)] [[PubMed](#)]

94. Belkin, S.; Wirsén, C.O.; Jannasch, H.W. A new sulfur-reducing, extremely thermophilic eubacterium from a submarine thermal vent. *App. Environ. Microbiol.* **1986**, *51*, 1180–1185. [CrossRef]
95. Jannasch, H.W.; Huber, R.; Belkin, S.; Stetter, K.O. *Thermotoga neapolitana* sp. nov. of the extremely thermophilic, eubacterial genus *Thermotoga*. *Arch. Microbiol.* **1988**, *150*, 103–104. [CrossRef]
96. Connors, S.B.; Mongodin, E.F.; Johnson, M.R.; Montero, C.I.; Nelson, K.E.; Kelly, R.M. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS Microb. Rev.* **2006**, *30*, 872–905. [CrossRef] [PubMed]
97. Zhang, J.; Shi, H.; Xu, L.; Zhu, X.; Li, X. Site-directed mutagenesis of a hyperthermophilic endoglucanase Cel12B from *Thermotoga maritima* based on rational design. *PLoS ONE* **2015**, *10*, e0141937. [CrossRef]
98. Fink, M.; Trunk, S.; Hall, M.; Schwab, H.; Steiner, K. Engineering of TM1459 from *Thermotoga maritima* for increased oxidative alkene cleavage activity. *Front. Microb.* **2016**, *7*, 1–9. [CrossRef]
99. Donaldson, T.; Iozzino, L.; Deacon, L.J.; Billones, H.; Ausili, A.; D'Auria, S.; Dattelbaum, J.D. Engineering a switch-based biosensor for arginine using a *Thermotoga maritima* periplasmic binding protein. *An. Biochem.* **2017**, *525*, 60–66. [CrossRef]
100. Han, D.; Xu, Z. Development of a pyrE-based selective system for *Thermotoga* sp. strain RQ7. *Extremophiles* **2017**, *21*, 297–306. [CrossRef]
101. Pegg, A.E. Repair of O<sup>6</sup>-alkylguanine by alkyltransferases. *Mutat. Res.* **2000**, *462*, 83–100. [CrossRef]
102. Margison, G.P.; Povey, A.C.; Kaina, B.; Santibanez Koref, M.F. Variability and regulation of O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *Carcinogenesis* **2003**, *24*, 625–635. [CrossRef] [PubMed]
103. Mollwitz, B.; Brunk, E.; Schmitt, S.; Pojer, F.; Bannwarth, M.; Schiltz, M.; Rothlisberger, U.; Johnsson, K. Directed evolution of the suicide protein O<sup>6</sup>-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* **2012**, *51*, 986–994. [CrossRef] [PubMed]
104. Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa, I.R., Jr.; Kindermann, M.; Beaufils, F.; Johnsson, K. An engineered protein-tag for multi-protein labeling in living cells. *Chem. Biol.* **2008**, *15*, 128–136. [CrossRef]
105. Hale, C.R.; Zhao, P.; Olson, S.; Duff, M.O.; Graveley, B.R.; Wells, L.; Terns, R.M.; Terns, M.P. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **2009**, *139*, 945–956. [CrossRef]
106. Terns, R.M.; Terns, M.P. The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochem. Soc. Trans.* **2013**, *41*, 1416–1421. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

La borsa di dottorato è stata cofinanziata con risorse del  
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),  
Fondo Sociale Europeo, Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale”



UNIONE EUROPEA  
Fondo Sociale Europeo



*Ministero dell'Istruzione,  
dell'Università e della Ricerca*



PON  
RICERCA  
E INNOVAZIONE  
2014 - 2020