3-17-2023

# 7-Deazaguanines in DNA: Functional and Structural elucidation of a DNA modification system

Samanthi Herath Gedara
*Portland State University*

Andrew Gustafson
*Portland State University*, andg@pdx.edu

Dirk Iwata-Reuyl
*Portland State University*, iwatard@pdx.edu

multiple additional authors

### Citation Details

Gedara, S. H., Wood, E., Gustafson, A., Liang, C., Hung, S. H., Savage, J., ... & Iwata-Reuyl, D. (2023). 7-Deazaguanines in DNA: functional and structural elucidation of a DNA modification system. Nucleic Acids Research, gkad141.

# 7-Deazaguanines in DNA: functional and structural elucidation of a DNA modification system

**Samanthi Herath Gedara[1],[†], Evan Wood [2],[†], Andrew Gustafson [1],[†], Cui Liang[3], Shr-Hau Hung [2],[6], Joshua Savage[2], Phuc Phan[2], Amit Luthra[2], Valérie de Crécy-Lagard [4], Peter Dedon [3],[5], Manal A. Swairjo [2],[6],* and Dirk Iwata-Reuyl [1],***

[1]Department of Chemistry, Portland State University, Portland, OR 97201, USA, [2]Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA, USA, [3]Singapore-MIT Alliance for Research and Technology, 138602, Singapore, [4]Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA, [5]Department of Biological Engineering, MIT, Cambridge, MA 02139, USA and [6]The Viral Information Institute, San Diego State University, San Diego, CA, USA

## ABSTRACT

The modified nucleosides 2′-deoxy-7-cyano- and 2′-deoxy-7-amido-7-deazaguanosine (dPreQ$_0$ and dADG, respectively) recently discovered in DNA are the products of the bacterial queuosine tRNA modification pathway and the *dpd* gene cluster, the latter of which encodes proteins that comprise the elaborate Dpd restriction–modification system present in diverse bacteria. Recent genetic studies implicated the *dpdA*, *dpdB* and *dpdC* genes as encoding proteins necessary for DNA modification, with *dpdD–dpdK* contributing to the restriction phenotype. Here we report the *in vitro* reconstitution of the Dpd modification machinery from *Salmonella enterica* serovar Montevideo, the elucidation of the roles of each protein and the X-ray crystal structure of DpdA supported by small-angle X-ray scattering analysis of DpdA and DpdB, the former bound to DNA. While the homology of DpdA with the tRNA-dependent tRNA-guanine transglycosylase enzymes (TGT) in the queuosine pathway suggested a similar transglycosylase activity responsible for the exchange of a guanine base in the DNA for 7-cyano-7-deazaguanine (preQ$_0$), we demonstrate an unexpected ATPase activity in DpdB necessary for insertion of preQ$_0$ into DNA, and identify several catalytically essential active site residues in DpdA involved in the transglycosylation reaction. Further, we identify a modification site for DpdA activity and demonstrate that DpdC functions independently of DpdA/B in converting preQ$_0$-modified DNA to ADG-modified DNA.

## GRAPHICAL ABSTRACT
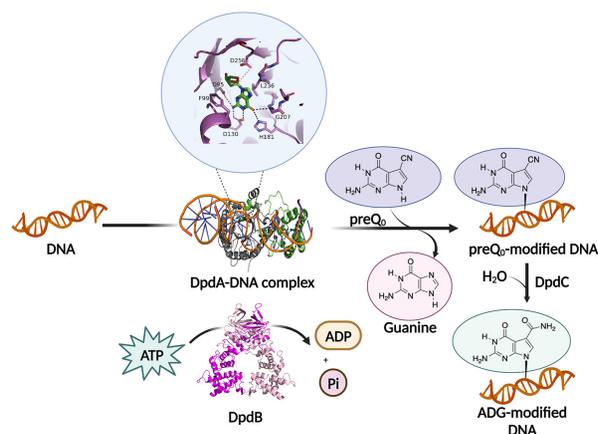


## INTRODUCTION

Nucleic acid processing comprises a diverse suite of inter-related systems essential to replication, genome maintenance and gene expression. Nucleotide modification, in which nucleobases undergo specific, enzymatically directed chemical changes, underpins all of these systems and is essential for their function. While both DNA and RNA are the targets of modification (1), RNA exhibits far more structural diversity and complexity in modification than DNA. Indeed, there are >150 distinct modifications known in RNA (2), many of which are structurally complex and require multistep biosynthetic pathways for their construction. In contrast, <25 modifications have been characterized in DNA (1,3), and in the main they
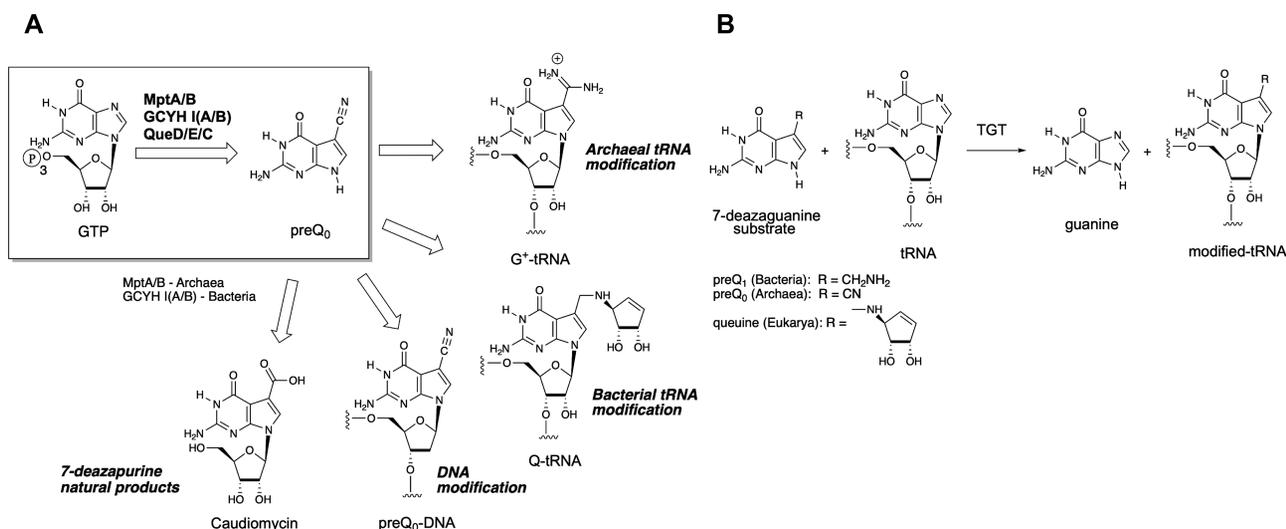
**Figure 1.** The biosynthesis of 7-deazapurine derivatives found in nucleic acids. (**A**) Overview of the *de novo* biosynthesis of 7-deazaguanosines. (**B**) The reactions catalyzed by the tRNA-dependent tRNA-guanine transglycosylases (TGTs).

exhibit relatively simple structural changes to the canonical bases.

In a remarkable example of the cross-talk between RNA and DNA processing, we recently discovered that the pathway to one of the most structurally complex class of modifications known to occur in RNA, that responsible for the 7-deazaguanine modifications queuosine (Q) and archaeosine ($G^+$) found in tRNA (Figure 1A), is also utilized by diverse organisms for the modification of DNA (4). Genomic modification with 2′-deoxy-7-cyano- and 2′-deoxy-7-amido-7-deazaguanosine (dPreQ$_0$ and dADG, respectively) has been demonstrated in bacteria (4), and these nucleosides, as well as dG$^+$ and 2′-deoxy-7-aminomethyl-7-deazaguanosine (dPreQ$_1$), have also been discovered in the DNA of phage (4–6) (Figure 1).

The biosynthetic pathway to the 7-deazaguanine-based tRNA modifications is elaborate (7,8), beginning with the multistep conversion of GTP to the common precursor base 7-cyano-7-deazaguanine (preQ$_0$; Figure 1A). Notably, the presence of a nucleic acid-independent component to the pathway is unique among modification pathways, and results in the formation of a precursor base that is subsequently inserted into the nucleic acid concomitant with the ejection of a genetically encoded guanine (Figure 1B). In the tRNA modification pathways, this reaction is catalyzed by the enzyme tRNA-guanine transglycosylase (TGT) (9,10), which has distinct subtypes (11) corresponding to their phylogenetic locations and specific functions in Archaea, Bacteria and Eukarya—the insertion of preQ$_0$ into position 15 of virtually all archaeal tRNA, or the insertion of 7-aminomethyl-7-deazaguanine (preQ$_1$) or queuine into position 34 (wobble position) of tRNA possessing a GUN anticodon in Bacteria and Eukarya, respectively (Figure 1B).

The first indication that 7-deazaguanines might also be present in DNA came from the observation that in many bacterial genomes, divergent *tgt*-like genes (originally annotated as *tgtA5* but now designated *dpdA*) and preQ$_0$ biosynthetic genes cluster with genes encoding putative DNA processing enzymes (12). We have confirmed the involvement of these systems in DNA modification in phylogenetically diverse bacteria (4) and phage (5,6), where the modifications appear to play a role in the protection of DNA—in the former by incoporating modification as part of a novel restriction–modification (RM) system (4,13), and in the latter by saturating the genome with the modified base to render the DNA resistant to host restriction nucleases (5,6), and potentially other DNA surveillance mechanisms such as the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) systems (5). Remarkably, while the stoichiometry of modification in the bacterial RM systems is of the order of 1 in $10^3$ nucleotides (4), in viral genomes the substitution of G by the 7-deazaguanine-modified base can be total (6), demonstrating that these highly modified bases not only protect the phage genome but are also the principal carriers of genetic information.

In bacteria, the 7-deazaguanine-based RM system is encoded by the *dpd* genes (4), which are typically located within a 20 kb genomic cluster. In *Salmonella enterica* serovar Montevideo (*S.* Montevideo), the Dpd system is comprised of the DpdA–DpdK proteins, and we have shown that *in vivo* DpdA/B/C are required for the formation of dPreQ$_0$ and dADG in the DNA (13), while DpdD–DpdK are responsible for the restriction phenotype, but specific roles for each protein are unknown. We report here the reconstitution of modification activity *in vitro*, demonstrate the surprising requirement for an ATPase activity in the transglycosylation reaction, identify a specific guanine that is a site for modification and describe the structural elucidation of DpdA and DpdB through X-ray crystallography and small angle X-ray scattering (SAXS), together with the identification of several DpdA residues important for catalysis.

## MATERIALS AND METHODS

### General

Buffers, salts and reagents (highest quality grade available) were purchased from Sigma (St. Louis, MO, USA). NTPs and dNTPs used in enzyme assays were from Fisher Scientific (Pittsburgh, PA, USA), and were at least 99% pure and certified free of contaminating nucleotide triphosphates. Synthetic DNA oligos (Table 1), were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). Dithiothreitol (DTT), isopropyl-β-D-thiogalactopyranoside (IPTG), kanamycin sulfate and ampicillin were purchased from RPI Corporation (Chicago, IL, USA). [8-$^{14}$C]Guanine was obtained from Moravek Inc. (Brea, CA, USA), and [α-$^{32}$P]ATP was obtained from Perkin Elmer (Waltham, MA, USA). Amicon Ultra 15 and 0.5 centrifugal filter units and NovaBlue Singles competent cells were acquired from EMD Millipore (Billerica, MA, USA). Nickel-nitrilotriacetic acid agarose (Ni$^{2+}$-NTA agarose), Whatman GF/C and polyvinyldifluoridene (PVDF) syringe filters were purchased from Fisher Scientific. GeneJet Plasmid Miniprep kits and PageRuler pre-stained protein ladder were purchased from Fermentas (Glen Burnie, MD, USA). Dialysis was carried out in Slide-A-Lyzer cassettes and SnakeSkin dialysis tubing from ThermoFisher (Waltham, MA, USA). All reagents for sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) were purchased from BioRad (Hercules, CA, USA). SDS–PAGE analysis was carried out using 12% gels and visualized with Coomassie Brilliant Blue. The recombinant DpdA, DpdB and DpdC proteins from *Salmonella enterica* subsp. enterica serotype Montevideo were overproduced and purified as previously described (13). PreQ$_0$ was synthesized as previously described (14), purified by reverse-phase high-performance liquid chromatography (HPLC) and stored at room temperature in dimethylsulfoxide (DMSO).

Ultraviolet-visible (UV-Vis) spectrophotometry was performed with a Varian Cary 100 spectrophotometer equipped with a thermostated multicell holder. Liquid scintillation counting was conducted with Hidex 300 SL or Beckman LS6500 liquid scintillation counters. Phosphorimaging was carried out with a Typhoon Trio phosphorimager (GE Healthcare). Chromatography-coupled mass spectrometry was performed with an Agilent 1290 HPLC coupled to Agilent 6490 triple-quadrupole mass spectrometer system.

### Initial screen for transglycosylation activity

Enzyme activity assays were carried out in 50 μl reactions consisting of tri-buffer (0.052 M MES, 0.052 M TAPSO and 0.1 M diethanolamine) at variable pH (6.3–8.8), 50 mM KCl, 10 mM MgCl$_2$, 5 mM DTT, 12.5 μg of pUC19 DNA, 5 μM DpdA, 5 μM DpdB, 20 μM [8-$^{14}$C]guanine (57 mCi/mmol) and various cofactors at 10 mM. Reactions were incubated at 37°C for 30–60 min, quenched with 400 μl of a 10% (w/v) solution of trichloroacetic acid (TCA), and the precipitated DNA was collected on a Whatman GF/C glass microfiber filter in a Büchner funnel attached to a vacuum flask. The reaction tubes were washed three times

with 450 μl of 10% TCA and the washes were added to the same glass microfiber filter. The samples were sandwiched between a second glass microfiber filter and covered with filter paper before washing with 300 ml of cold 95% ethanol. The glass microfiber filters were placed in liquid scintillation counting vials filled with Econo-Safe (RPI) liquid scintillation counting cocktail, and the samples were quantified by liquid scintillation counting.

### *In vitro* dPreQ$_0$ and dADG formation

To prepare preQ$_0$-modified DNA, a 400 μl reaction with tri-buffer (pH 7.1), 50 mM KCl, 10 mM MgCl$_2$, 5 mM DTT, 400 μg of pUC19 DNA, 5 μM DpdA, 5 μM DpdB, 20 μM preQ$_0$ and 10 mM ATP was incubated at 37°C for 60 min. The preQ$_0$-modified DNA was isolated from the reaction with a phenol–chloroform extraction followed by ethanol precipitation. The preparation of ADG-modified DNA was carried out in two ways; in the first, DpdC (5 μM) was included in a reaction as described above, while in the second it was prepared in a 100 μl reaction with tri-buffer (pH 7.1), 50 mM KCl, 10 mM MgCl$_2$, 2 mM DTT, 5 μM DpdC and 85 μg of preQ$_0$-modified pUC19 DNA incubated at 37°C for 60 min. The DNA was isolated as before and samples of both reactions were lyophilized in preparation for mass spectrometry.

### Nucleoside analysis by LC-MS

DNA analysis was performed as previously described (13) but with several modifications. Purified DNA (20 μg) was hydrolyzed in 10 mM Tris–HCl (pH 7.9) with 1 mM MgCl$_2$ with Benzonase (20 U), DNase I (4 U), calf intestine phosphatase (17 U) and phosphodiesterase (0.2 U) for 16 h at ambient temperature. Following passage through a 10 kDa filter to remove proteins, the filtrate was lyophilized and resuspended to a final concentration of 0.2 μg/μl (based on initial DNA quantity).

Quantification of dADG and dPreQ$_0$ was achieved by liquid chromatography-coupled triple quadrupole mass spectrometry (LC-MS/MS), with the four canonical 2′-deoxyribonucleosides (dA, dT, dG and dC) quantified with an in-line diode array detector (LC-DAD). Aliquots of hydrolyzed DNA were injected onto a Phenomenex Luna Omega Polar C18 column (2.1 x 100 mm, 1.6 μm particle size) equilibrated with 98% solvent A (0.1% v/v formic acid in water) and 2% solvent B (0.1% v/v formic acid in acetonitrile) at a flow rate of 0.3 ml/min and eluted with the following solvent gradient: 12% B for 10 min, 1 min ramp to 100% B for 10 min, 1 min ramp to 2% B for 10 min. The HPLC column was coupled to an Agilent 1290 Infinity DAD and an Agilent 6490 triple quadruple mass spectrometer (Agilent, Santa Clara, CA, USA). The column was kept at 40°C and the auto-sampler was cooled at 4°C. The UV wavelength of the DAD was set at 260 nm and the electrospray ionization of the mass spectrometer was performed in positive ion mode with the following source parameters: drying gas temperature 200°C with a flow of 14 l/min; nebulizer gas pressure 30 psi; sheath gas temperature 400°C with a flow of 11 l/min; capillary voltage 3000 V; and nozzle voltage 800 V. Compounds were quantified in multiple reaction monitoring (MRM) mode with the following

**Table 1.** DNA oligomers used to construct duplexes for guanine exchange assays and binding experiments

| Name | Oligomers used | Sequence |
|---|---|---|
| 80a | pUC19 169–248 F | 5′-TAC TGA GAG TGC ACC ATA TGC GGT GTG AAA TAC CGC ACA GAT GCG TAA GG A GAA AAT ACC GCA TCA GGC GCC ATT C GC CA-3′ |
| | pUC19 169–248 R | 5′-TGG CGA ATG GCG CCT GAT GCG GTA T TT TCT CCT TAC GCA TCT GTG CGG TAT TT C ACA CCG CAT ATG GTG CAC TCT CAG TA-3′ |
| 80b | pUC19 249–328 F | 5′-TTC AGG CTG CGC AAC TGT TGG GAA GGG CGA TCG GTG CGG GCC TCT TCG CT A TTA CGC CAG CTG GCG AAA GGG GG A TGT GC-3′ |
| | pUC19 249–328 R | 5′-GCA CAT CCC CCT TTC GCC AGC TGG CGT AAT AGC GAA GAG GCC CGC ACC GA T CGC CCT TCC CAA CAG TTG CGC AGC C TG AA-3′ |
| 81c | pUC19 329–409 F | 5′-TGC AAG GCG ATT AAG TTG GGT AAC GCC AGG GTT TTC CCA GTC ACG ACG TT G TAA AAC GAC GGC CAG TGA ATT CG A GCT CGG-3′ |
| | pUC19 329–409 R | 5′-CCG AGC TCG AAT TCA CTG GCC GTC GTT TTA CAA CGT CGT GAC TGG GAA AA C CCT GGC GTT ACC CAA CTT AAT CGC C TT GCA-3′ |
| 40a | pUC19 249–288 F | 5′-TTC AGG CTG CGC AAC TGT TGG GAA GGG CGA TCG GTG CGG G-3′ |
| | pUC19 249–288 R | 5′-CCC GCA CCG ATC GCC CTT CCC AAC AGT TGC GCA GCC TGA A-3′ |
| 40b | pUC19 289–328 F | 5′-CCT CTT CGC TAT TAC GCC AGC TGG CGA AAA GGG GAT GTG C-3′ |
| | pUC19 289–328 R | 5′-GCA CAT CCC CTT TTC GCC AGC TGG CGT AAT AGC GAA GAG G-3′ |
| 18T | 40bF18T | 5′-CCT CTT CGC TAT TAC GCT AGC TGG CGA AAA GGG GAT GTG C-3′ |
| | 40bF18T comp | 5′-GCA CAT CCC CTT TTC GCC AGC TAG CGT AAT AGC GAA GAG G-3′ |
| 20A | 40bF20A | 5′-CCT CTT CGC TAT TAC GCC AAC TGG CGA AAA GGG GAT GTG C-3′ |
| | 40bF20A comp | 5′-GCA CAT CCC CTT TTC GCC AGT TGG CGT AAT AGC GAA GAG G-3′ |
| 21T | 40bF21T | 5′-CCT CTT CGC TAT TAC GCC AGT TGG CGA AAA GGG GAT GTG C-3′ |
| | 40bF21T comp | 5′-GCA CAT CCC CTT TTC GCC AAC TGG CGT AAT AGC GAA GAG G-3′ |
| 23A | 40bF23A | 5′-CCT CTT CGC TAT TAC GCC AGC TAG CGA AAA GGG GAT GTG C-3′ |
| | 40bF23A comp | 5′-GCA CAT CCC CTT TTC GCT AGC TGG CGT AAT AGC GAA GAG G-3′ |
| 28bp | 28bpDuplexF | 5′-GGG GTT TTC GCC AGC TGG CGT TTT GGG G-3′ |
| | 28bpDuplexR | 5′-CCC CAA AAC GCC AGC TGG CGA AAA CCC C-3′ |

$m/z$ transitions defined using synthetic standards: 310.1 → 194.1, 310.1 → 177.1, 310.1 → 293.1 for dADG, and 292.1 → 176.1, 176.1 → 159.1, 176.1 → 52.1 for dPreQ$_0$. External calibration curves were used for the quantification of the modified canonical 2′-deoxynucleosides, with replicate measurements of eight concentrations of each standard. A linear regression with $r^2 > 0.995$ was obtained in all relevant ranges. The limit of detection (LOD), defined by a signal-to-noise ratio (S/N) ≥3, ranged from 0.1 to 1 fmol for the modified 2′-deoxynucleosides. Data acquisition and processing were performed using MassHunter software (Agilent).

### Preparation of pUC19 fragments and synthetic DNA oligos as substrates for the transglycosylase reaction

Approximately 1 mg of pUC19 was subjected to restriction digestion with CviQI (New England Biolabs) in a 0.5 ml reaction according to the manufacturer's instructions. The resulting fragments were purified via continuous elution PAGE using a 491 Prep Cell (BioRad) with a 3.5% gel as described previously (15,16). Tris/borate/EDTA (TBE) was used as the elution buffer with a flow rate of 0.5 ml/min, and 2 ml fractions were collected. Fractions containing each of the relevant pUC19 fragments were pooled and the DNA precipitated with ethanol. After washing the pellets with cold 70% ethanol and drying, the DNA was resuspended in Tris/EDTA (TE) buffer and stored at –20°C.

Synthetic DNA duplexes were prepared by mixing the forward and reverse oligos (Table 1) in a 1:1 molar ratio in a buffer solution containing 10 mM Tris, pH 7.5, 1 mM EDTA and 50 mM NaCl. The mixture was then heated at 94°C for 2 min, cooled to 25°C over 30 min and stored at –20°C.

### Radiochemical enzyme activity assays

Enzyme activity assays were carried out in 50 µl reactions consisting of tri-buffer (pH 7.1), 50 mM KCl, 10 mM MgCl$_2$, 5 mM DTT, variable amounts of DNA, 1–5 µM DpdA, 1–5 µM DpdB, 20 µM [8-$^{14}$C]guanine (57 mCi/mmol) and 10 mM ATP, and were incubated at 37°C for variable times. Reactions were quenched and the DNA was precipitated with 600 µl of a 10% (w/v) solution of TCA and added to Whatman GF/C glass microfiber filters in a Büchner funnel attached to a vacuum flask. The reaction tubes were washed three times with 1000 µl of 10% TCA and the washes were added to the same glass microfiber filter. The samples were sandwiched between a second glass microfiber filter and covered with filter paper before washing with 300 ml of cold 95% ethanol. The glass microfiber filters were placed in liquid scintillation counting vials filled with Econo-Safe (RPI) liquid scintillation counting cocktail and quantified by liquid scintillation counting.

### ATPase assay

To investigate ATP consumption during modification, 10 µl reactions containing tri-buffer (pH 7.1), 50 mM KCl, 10 mM MgCl$_2$, 5 mM DTT, 20 µM guanine, 100 µM [α-$^{32}$P]ATP (Perkin-Elmer) and either 5 µM DpdA or 5 µM DpdB, or both, with or without 144 nM pUC19 plasmid were incubated at 37°C for 30 min. The reactions were quenched with 2 µl of 0.1 M EDTA (pH 8.0) prior to 2 µl of each reaction being spotted on a PEI cellulose thin-layer chromatography (TLC) plate (Millipore) and developed with KH$_2$PO$_4$ (pH 3.5). The plate was exposed to a phosphor storage screen then imaged on a Typhoon Trio (GE Healthcare). The Rf values for ATP, ADP and AMP were determined with authentic standards utilizing the fluorescence indicator in the TLC plates. To quantify the ATPase activity, reaction assays were carried out as above but terminated at various time points prior to TLC analysis and phosphorimaging, and the data were used to calculate initial velocities.

### Subcloning and site-directed mutagenesis of dpdA

The initial cloning of the *dpdA* gene from *Salmonella enterica* subsp. enterica serotype Montevideo (NCBI accession number AHW12286.1) into the pET30-Xa vector was

described previously ([13](#)). For producing protein for X-ray analysis, the gene was subcloned from the initial pET30-Xa vector into the NheI (5′ end) and Xho1 (3′ end) sites of the pET28a(+) expression vector (Novagen, San Diego, CA, USA) carrying an N-terminal His$_6$ sequence followed by a thrombin cleavage site. The nucleotide sequence of the resulting construct (pET28a-*Sm*DpdA$^{WT}$) was confirmed by sequencing (Genewiz, Inc., South Plainfield, NJ, USA) and used for expression. Point mutagenesis was performed on pET28a-*Sm*DpdA$^{WT}$ using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs). All constructs were verified by sequencing. Primers used for subcloning and mutagenesis are listed in Supplementary Table S1.

### DpdA overexpression and purification for X-ray crystallography

The wild-type *S.* Montevideo DpdA protein (M$_r$ 48100 Da) and mutants were overexpressed in *Escherichia coli* C41 (DE3) cells (Lucigen Corporation, Middleton, WI, USA). A 1 liter culture was grown at 37°C in LB-Miller broth containing kanamycin (50 µg/ml) with vigorous shaking (120 rpm) to an optical $A_{600}$ of ∼0.6 before expression was induced by addition of IPTG to a final concentration of 0.5 mM. After further growth for 8–9 h, cells were harvested by centrifugation at $6000 \times g$ for 20 min, and lysed in 40 ml of Buffer A [50 mM Tris (pH 7.5), 10 mM imidazole, 300 mM NaCl, 5% glycerol, 1 mM DTT, 1 U of DNase I and a tablet of cOmplete™ ULTRA protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IL, USA)]. The cell lysate was cleared by centrifugation at $20\ 000 \times g$ for 30 min at 4°C, and the supernatant was filtered using a 0.22 µm filter and loaded onto a 5 ml Ni-NTA column (Qiagen, Valencia, CA, USA) pre-equilibrated in Buffer A. Resin was then washed with 100 ml of Buffer A, 100 ml of 50 mM Tris (pH 7.5), 20 mM imidazole, 500 mM NaCl, 5% glycerol and 1 mM DTT (Buffer B), followed by 100 ml of 50 mM Tris (pH 7.5), 20 mM imidazole, 200 mM NaCl, 5% glycerol and 1 mM DTT (Buffer C) before the protein was eluted with Buffer C supplemented with 400 mM imidazole. The Ni-NTA eluate was further purified by gel filtration chromatography using a Sephacryl S-200 HR 16/60 column (GE Healthcare Life Sciences, Pittsburgh, PA, USA) pre-equilibrated in buffer containing 50 mM Tris (pH 7.5), 100 mM KCl and 1 mM DTT. For SAXS samples, an Enrich™ SEC 70 column (Biorad, Hercules, CA, USA) was used instead. Protein purity was confirmed to be >95% by SDS–PAGE. Attempts to cleave the N-terminal His$_6$ tag with thrombin resulted in protein precipitation and loss; therefore, the His$_6$-tagged protein was directly used in crystallization, electrophoretic mobility shift assay (EMSA) and SAXS experiments.

### Crystallization and crystal structure determination of *S.* Montevideo DpdA

A search for crystallization conditions for *S.* Montevideo DpdA was carried out by high-throughput screening using the vapor diffusion method. Using a Mosquito robot (TTP Labteck, Melbourn, UK), sitting drops were set up in 96-well microplates, incubated at 20°C and monitored

using a Rock Imager system (Formulatrix, Bedford, MA, USA). Initial crystallization conditions were identified in the JCSG+ crystallization screen (Qiagen Inc., Valencia, CA, USA), and were further subjected to optimization using the Hampton Research Additive Screen (Hampton Research Inc., Carlsbad, CA, USA). Optimal single crystals were grown at 20°C in sitting drops by mixing 2 µl of sample containing 4.0 mg/ml protein (81.5 µM), 50 mM Tris (pH 7.5), 100 mM NaCl and 1 mM DTT with 3 µl of reservoir solution containing 10% (w/v) polyethylene glycol 8000, 8% (v/v) ethylene glycol, 100 mM HEPES (pH 7.5), 0.1 mM ZnCl$_2$, 3 mM reduced glutathione and 3 mM oxidized glutathione. Crystals were cryoprotected by soaking for 10 min in a 5 µl drop containing reservoir solution without ZnCl$_2$ plus 20% (v/v) glycerol, followed by flash cooling in liquid nitrogen. X-ray diffraction data were collected on beamline BL14-1 at the Stanford Synchrotron Radiation Lighthouse (SSRL; Menlo Park, CA, USA). The presence of the Zn anomalous scatterer was verified by X-ray fluorescence using an excitation wavelength near the Zn edge. A three-wavelength anomalous dispersion dataset was collected at the peak and inflection energies of the Zn edge, and at a remote energy above the Zn edge. All data were processed in the XDS program suite ([17](#)) and data quality was checked in Xtriage ([18](#)). Data collection parameters and processing statistics are summarized in Table [3](#). The crystal structure was determined using the multiwavelength anomalous dispersion method, the endogenous Zn as the anomalous scatterer (Zn-MAD) and the PHENIX suite of programs ([18](#)). Solvent content analysis indicated a single protein molecule in the asymmetric unit. Autotracing of the density-modified FoFOM electron density map in PHENIX yielded an initial refined model containing 321 of the 416 residues in the protein sequence (77% of the structure) and exhibiting a crystallographic- and free-R factor of 0.33 and 0.39, respectively. In several rounds of model building, solvent fitting and refinement in CCP4 Refmac5 ([19](#)) and Coot ([20](#)), the remaining 96 residues in the structure were successfully modeled. Final structure refinement statistics are listed in Table [3](#).

### Analytical size exclusion chromatography (SEC) of *S.* Montevideo DpdA and DpdB

Analytical SEC of DpdA and DpdB was carried out after His$_6$-tag removal. For each experiment, 500 µl of a 0.5 mg/ml protein sample were loaded on a Bio-Rad SEC 650 $10 \times 300$ mm column attached to an NGC chromatography system (Bio-Rad) and pre-equilibrated in buffer containing 100 mM Tris–HCl (pH 7.5), 100 mM KCl and 1 mM DTT. Samples were eluted using the same buffer at a flow rate of 0.25 ml/min, with absorption detection at 280 nm. The column was calibrated using gel filtration standards (Bio-Rad, cat no. 151–1901) that included thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa), and a standard curve relating elution volume to apparent molecular weight was generated by linear regression of the partition coefficient ($K_{av}$) as a function of the log molecular weight using the equation: $K_{av} = V_e - V_0 / V_g - V_0$, where $V_e$ is the protein elution volume, $V_0$ is the void volume determined by elu-

tion of Blue Dextran (2000 kDa), and $V_g$ is the geometric column volume.

## Preparation of a 28 bp DNA duplex for EMSA and SAXS experiments

Synthetic DNA oligos (28bpDuplexF and 28bpDuplexR, Table 1) were purchased from Integrated DNA Technologies (IDT) and suspended in the LabReady formulation IDTE (10 mM Tris, 0.1 mM EDTA, pH 8.0). The DNA duplexes were prepared by mixing the two oligos in a 1:1 molar ratio to a final concentration of 48 μM in buffer solution containing 5 mM $MgCl_2$, 50 mM NaCl, 15 mM HEPES (pH 7.8) and 0.5 mM DTT. Using a thermocycler, the mixture was then heated to 95°C for 5 min and cooled to 25°C at a rate of –1°C/min, and then stored at –20°C. Before use in binding reactions, duplexes were buffer-exchanged into buffer containing 20 mM NaCl, 20 mM Tris (pH 7.5) and 2 mM β-mercaptoethanol, using a 10 kDa Amicon Ultra centrifugal device.

## EMSA for *S*. Montevideo DpdA

Binding reactions were carried out for 5 min at 37°C in 10 μl solutions containing 0–15 μM wild-type or mutant *S*. Montevideo DpdA, 1 μM DNA duplex, 20 mM Tris (pH 7.5), 20 mM NaCl, 2 mM β-mercaptoethanol and 2.5% (w/v) Ficoll 400. Reactions (1 μl) were loaded onto a 5% nondenaturing polyacrylamide gel cast in 50 mM Tris and 50 mM borate (pH 7.6) and pre-run at 80 V for 2 h immediately before use, and electrophoresed at 4°C for 15 min at 200 V in running buffer containing 50 mM Tris and 50 mM borate (pH 8.3). DNA migration was visualized by SYBR-Gold staining. For quantitative EMSA experiments, the binding reactions were conducted using a DNA duplex prepared by spiking the oligo mixture with 1% 6-FAM-3′-labeled 28bpDuplexR before annealing, and the gel was imaged using a Typhoon Fla 9600 instrument ($λ_{ex}$ = 473 nm, $λ_{em}$ = 520 nm). Band densities were measured using Image Studio™ Lite 5.2.5 (LI-COR Biotechnology, Lincoln, NE, USA), and the density data were fit using the non-linear least squares method in GraphPad Prism version 9.1.2 (GraphPad Software, La Jolla, CA, USA, www.graphpad.com).

## SEC-SAXS data collection and analysis

SEC-SAXS data were collected on the SIBYLS beamline at the Lawrence Berkeley National Laboratory's Advanced Light Source (ALS), using an incident beam wavelength of 1.03 Å and a sample to detector distance of 1.5 m. The scattering vector q for all data ranged from 0.013/Å to 0.5/Å, where q = $4π\sin(θ)/λ$ and 2θ is the measured scattering angle. Multiangle light scattering (MALS), quasi-elastic light scattering (QELS), UV absorption and refractive index data were also collected and analyzed using the Wyatt Astra 6 software (21). Samples of free DNA (1 mg/ml), free *S*. Montevideo DpdA$^{D95A}$ (3 mg/ml) and DpdB (3 mg/ml) were prepared in buffer containing 100 mM KCl, 50 mM Tris (pH 7.0) and 1 mM DTT. The DpdA$^{D95A}$–DNA complex

(4 mg/ml) was prepared in the same buffer by mixing protein and DNA duplex in equimolar concentrations (27 μM) and further purification by gel filtration on an Enrich SEC 70 column (Biorad Inc.), followed by concentration to 75.5 μM (3.8 mg/ml) in an Amicon filtration device. Samples were loaded on a Shodex PROTEIN KW-802.5 size exclusion column (8.0 mm ID × 300 mm) pre-equilibrated in the same buffer above, and eluted at a flow rate of 0.5 ml/min in line with the X-ray beam. Scattering images were collected with 3 s exposures per frame over the course of 40 min. Frames were averaged and processed using standard procedures in ScÅtter IV (https://www.bioisis.net/tutorials/9) (22). Data analysis was performed using ATSAS package tools (23) integrated in BioXTAS RAW 2.1.1 (24), except for Porod volumes which were calculated using PRIMUS in the ATSAS package. Rg values were calculated using GNOM in BioXTAS RAW. Molecular masses were determined using the methods Shape & Size (25) and SAXS-MoW 2.0 (26). P(r) functions were calculated using GNOM (27). Theoretical scattering curves were computed from different structural models and compared with experimental scattering curves using the program FoXS (28) and, in the case of *Sm*DpdB, followed by MultiFoXs (29) for multistate models.

## Generation of structural models and fitting to the SAXS data

A 13 bp 90° bent DNA model extracted from the crystal structure of *E. coli* endonuclease IV (Endo IV) in complex with DNA [PDB ID 1QUM (30)] was extended to 28 bp and its sequence modified in the program Coot (20) to represent the sequence of the 28 bp DNA duplex in the SAXS sample. The resulting DNA duplex model was docked onto the *S*. Montevideo DpdA crystal structure using the HADDOCK server version 2.2 (31). In the docking protocol, active site residues Asp95, Asp130 and His181 and the substrate guanine residue of DNA were designated as active residues to apply distance restraints. The His$_6$ tag (22 residues including a linker) and the D95A mutation were added to the model which was then energy minimized using the AMBER99SB-ILDN forcefield (32) in the GROMACS package version 2021.4-Homebrew (33) before use for fitting to the experimental SAXS data. The same approach was used to build a structural model of the DpdA–DNA complex with straight DNA. For this, first the 28 bp DNA model was built in Coot (34) and subjected to molecular dynamics simulations in GROMACS, and the conformation that best fit the experimental scattering data from the free DNA sample (Supplementary Figure S9E, F) was selected for docking in HADDOCK. Two additional complex models were generated using the bent or straight DNA models by maximally superposing the protein and DNA models with the corresponding components in the crystal structure of the *E. coli* Endo IV–DNA complex [PDB ID 1QUM (30)], followed by energy minimization in GROMACS (33). The four structural models of the DpdA/DNA complex, free DNA and free DpdA were used for fitting attempts to the experimental data using FoXS (28). For *S*. Montevideo DpdB, a structural model representing the closed state of the protein homodimer was generated using AlphaFold

(35) and subjected to conformational sampling using the BILBOMD webserver (https://bl1231.als.lbl.gov/bilbomd), which yielded a second model representing an open state of the protein. *Ab initio* model reconstruction of DpdB was computed and validated using DENSS integrated in BioX-TAS RAW. All structural figures were made in the PyMOL Molecular Graphics program version 2.4.1 (Schrödinger, LLC).

## RESULTS

### Reconstitution of *in vitro* activity for DpdAB

In considering possible routes to $dPreQ_0$ incorporation into DNA, the precedence of transglycosylation in tRNA modification, as well as mechanistic economy, made this route more likely in our view than alternative routes requiring incorporation at the nucleotide level. Notably, the ability of the tRNA-dependent TGT enzymes to catalyze guanine exchange in tRNA forms the basis of a quantitative activity assay in which the TGT-catalyzed incorporation of [$^{14}$C]guanine into tRNA can be measured after precipitation and collection of the labeled tRNA followed by liquid scintillation counting (36,37). Given the predicted homology between the cores of the DpdA protein and the TGT proteins, which structural modeling revealed included extensive conservation of the active site (4), we designed a potential activity assay based on the measurement of [$^{14}$C]guanine incorporation into DNA to test the putative transglycosylation activity of DpdA. Because plasmid DNA was effective *in vivo* as a substrate (13), we used pUC19 as the substrate for *in vitro* assays of enzymatic activity.

While the bacterial and archaeal TGT enzymes are functional monomers (38,39), the observation that *in vivo* modification of DNA with $preQ_0$ required the expression of both DpdA and DpdB was reminiscent of the eukaryotic TGT system, which functions as a heterodimer comprised of the homologous QTRT1 and QTRT2 proteins (40–42). However, based on sequence analysis, there is no predicted structural homology between DpdA and DpdB, suggesting that the functional relationship between these proteins might be different in the DNA modification system. Consistent with the *in vivo* experiments (13), modification activity was not observed with *in vitro* assays of DpdA (or DpdB) alone (Figure 2A), but surprisingly DpdA and DpdB together also failed to support enzymatic activity (Figure 2A).

A screen of possible cofactors in the reaction identified NTPs, in particular ATP (Figure 2A; Supplementary Figure S1A), as a necessary component of the reaction, while other activated phosphate compounds failed to support activity (Supplementary Figure S1A). Optimal enzyme activity was observed at a pH of ∼7.1 (Supplementary Figure S1B), and both supercoiled and linear pUC19 served as effective substrates (Figure 2A). Time-course assays of the reaction (Supplementary Figure S1C) revealed that the stoichiometry for [$^{14}$C]guanine incorporation into pUC19 was in excess of 15 per plasmid molecule, or >1 in 200 nucleotides. Despite the functional coupling of DpdA and DpdB, there does not appear to be any physical coupling as analytical SEC did not reveal the presence of a DpdA–DpdB complex (Supplementary Figure S2), in either the presence of absence of ATP.

### Source of ATP dependence

To probe the ATP dependence of the modification reaction, we investigated the reaction using [α-$^{32}$P]ATP and unlabeled guanine, and analyzed the fate of the ATP by TLC. As shown in Figure 2B, ATP is converted to ADP and $P_i$ in the reaction, indicating that guanine exchange is dependent on net ATP hydrolysis. Furthermore, DpdB is clearly responsible for ATP hydrolysis (Figure 2B, lanes 2 and 3) and, while DpdB is a robust ATPase in the absence of DNA, its activity is enhanced in the presence of DNA (Figure 2B, lane 6) and DNA–DpdA (Figure 2B, lane 7, and Figure 2C).

### Nucleoside analysis of pUC19 modified by DpdAB and DpdC

To confirm that the activity measured with the radiochemical assays accurately reflected the native activity of $preQ_0$ incorporation, and to investigate the activity of DpdC, we carried out analogous assays with authentic $preQ_0$ in place of [$^{14}$C]guanine. Subsequent isolation of the pUC19 DNA, nuclease digestion and dephosphorylation to give the 2′-deoxynucleosides, followed by LC-MS analysis, clearly showed the presence of $dPreQ_0$ (Figure 3), and at levels consistent with the insertion of [$^{14}$C]guanine (Table 2). Furthermore, inclusion of DpdC in the assays, either together with DpdA/B or in a separate assay lacking DpdA/B but containing $preQ_0$-modified pUC19, resulted in the formation of dADG-modified DNA (Figure 3; Table 2), demonstrating that our conditions reconstituted native DpdA/B/C activity, and that insertion of $preQ_0$ into DNA by DpdA/B and the subsequent conversion of $preQ_0$-modified DNA to ADG-modified DNA by DpdC are discrete events.

### Identification of a modification site

To gain insight into the identity of modification sites, we subjected $preQ_0$- and ADG-modified pUC19 to restriction enzyme digestion with several restriction endonucleases (PvuI, PvuII, AcuI, Ac1I and BciVI) that possess G in the recognition motif in a variety of sequence contexts. Only in the case of PvuII did we observe a disruption in restriction of pUC19 (Figure 4A; Supplementary Figure S3), suggesting that $preQ_0$ or ADG occupied at least one of the G positions in the recognition site CAGCTG. pUC19 possesses two PvuII sites and, under the conditions of the experiment, restriction of unmodified plasmid is complete at both sites, generating two fragments, one of 2364 bp and the second of 322 bp (Figure 4A). When the plasmid has been modified to contain either $preQ_0$ or ADG, PvuII restriction is disrupted, with the production of nicked (open circle) plasmid, what appears to be a linear plasmid and a small amount of fully restricted plasmid as evidenced by the presence of the 2364 and 322 bp fragments. Note that while disruption of restriction by PvuII is consistent with modification of the recognition sequence, the observation of normal restriction in the case of the other restriction endonucleases (Supplementary Figure S3) does not unequivocally rule out modification at those sites.
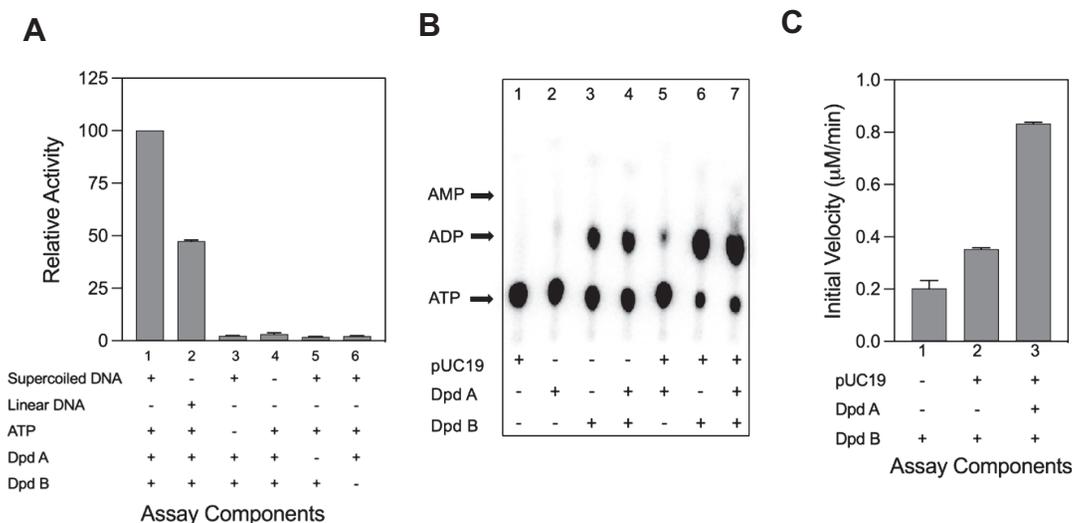
**Figure 2.** Enzymatic activities of the DpdA/DpdB system. (**A**) Relative [$^{14}$C]guanine exchange activity in the presence (+) or absence (–) of the indicated assay components. Error bars represent the standard error from two independent assays, each carried out in triplicate. (**B**) TLC analysis of ATPase assays with [$\alpha$-$^{32}$P]ATP in the presence (+) or absence (–) of the components listed. The Rf values for ATP, ADP and AMP were determined with authentic standards. (**C**) Rates of DpdB ATPase activity (initial velocity) in the presence (+) or absence (–) of DNA and DpdA. Error bars represent the standard error from two independent time-course assays.

**Table 2.** Quantification of 7-deazaguanine modifications in pUC19 DNA treated with DpdABC under varying conditions

| # | Conditions | dADG per $10^6$ nt | dPreQ$_0$ per $10^6$ nt |
|---|-----------|-------------------|------------------------|
| 1 | DpdAB | nd[a] | $3593 \pm 212$[b] |
| 2 | DpdABC | $7844 \pm 456$ | $31 \pm 3$ |
| 3 | DpdC with preQ$_0$-modified DNA | $9987 \pm 536$ | $238 \pm 14$ |
| 4 | DNA control (no enzyme) | nd | nd |

[a]nd, not detected; [b]modifications per $10^6$ nucleotides; values represent the mean ± deviation about the mean for two technical replicates of a single biological sample.

### Determination of a minimal DNA substrate and the site of guanine exchange

As a complement to the restriction analysis described above, we also sought to test smaller DNA duplexes as substrates, and identify which guanine(s) in the PvuII sequence was subject to exchange with preQ$_0$. To this end, pUC19 was restricted with the endonuclease CviQI to generate three fragments; a large fragment (1769 bp) possessing one of the PvuII sites, a medium fragment (676 bp), and a small fragment (241 bp) possessing the other PvuII site, and each fragment was isolated and purified. All three fragments were observed to serve as effective substrates in the guanine exchange assay (data not shown), and the small fragment was further subdivided into three roughly equal size duplexes and each prepared synthetically (Table 1). These were 80a (corresponding to nucleotides 169–248 in pUC19), 80b (corresponding to nucleotides 249–328 and possessing the PvuII site) and 81c (corresponding to nucleotides 329–409). While 80b exhibited robust activity (Figure 4B) as the substrate in the guanine exchange assay, 80a was a far poorer substrate and we failed to detect any activity with 81c as substrate (Figure 4B).

Two 40 bp duplexes corresponding to the two halves of 80b were subsequently synthesized and denoted 40a and 40b (Table 1). Of these, only 40b, which harbors the PvuII site, served as an efficient substrate for the guanine exchange reaction (Figure 4B). The sequence of 40b shows the PvuII site in the center of the duplex, nested within a larger 12 bp palindrome. To identify the guanine targeted by DpdA/B in the transglycosylation reaction, we designed mutated duplexes in which each GC base pair in the PvuII sequence in 40b was individually changed to an AT base pair, and these mutated duplexes were then assayed in the guanine exchange reaction. Duplexes 20A and 21T, corresponding to replacement of the GC base pair in positions 3 and 4 of the PvuII recognition sequence (CA**G**/C**TG**) and at positions 20 and 21 in the sequence of the sense strand of 40b, respectively, were ineffective as substrates (Figure 4C), while duplexes 18T and 23A, corresponding to replacement of the GC base pair in positions 1 and 6 of the recognition sequence (**C**AG/CT**G**), respectively, exhibited low but measurable activity as substrates in the guanine exchange reaction (Figure 4C).

### X-ray crystal structure of DpdA

The crystal structure of *S.* Montevideo DpdA was determined in space group P2$_1$2$_1$2 using the multiwavelength anomalous dispersion method and an endogenous Zn$^{2+}$ ion as the anomalous scatterer. The structure was refined to a resolution of 2.51 Å (Figure 5A; Table 3; Supplementary Figure S4A), and reveals one protein molecule in the asymmetric unit. Analysis of the molecular interfaces in PISA (43) revealed no significant multimerization interfaces with symmetry mates, consistent with a monomeric protein as
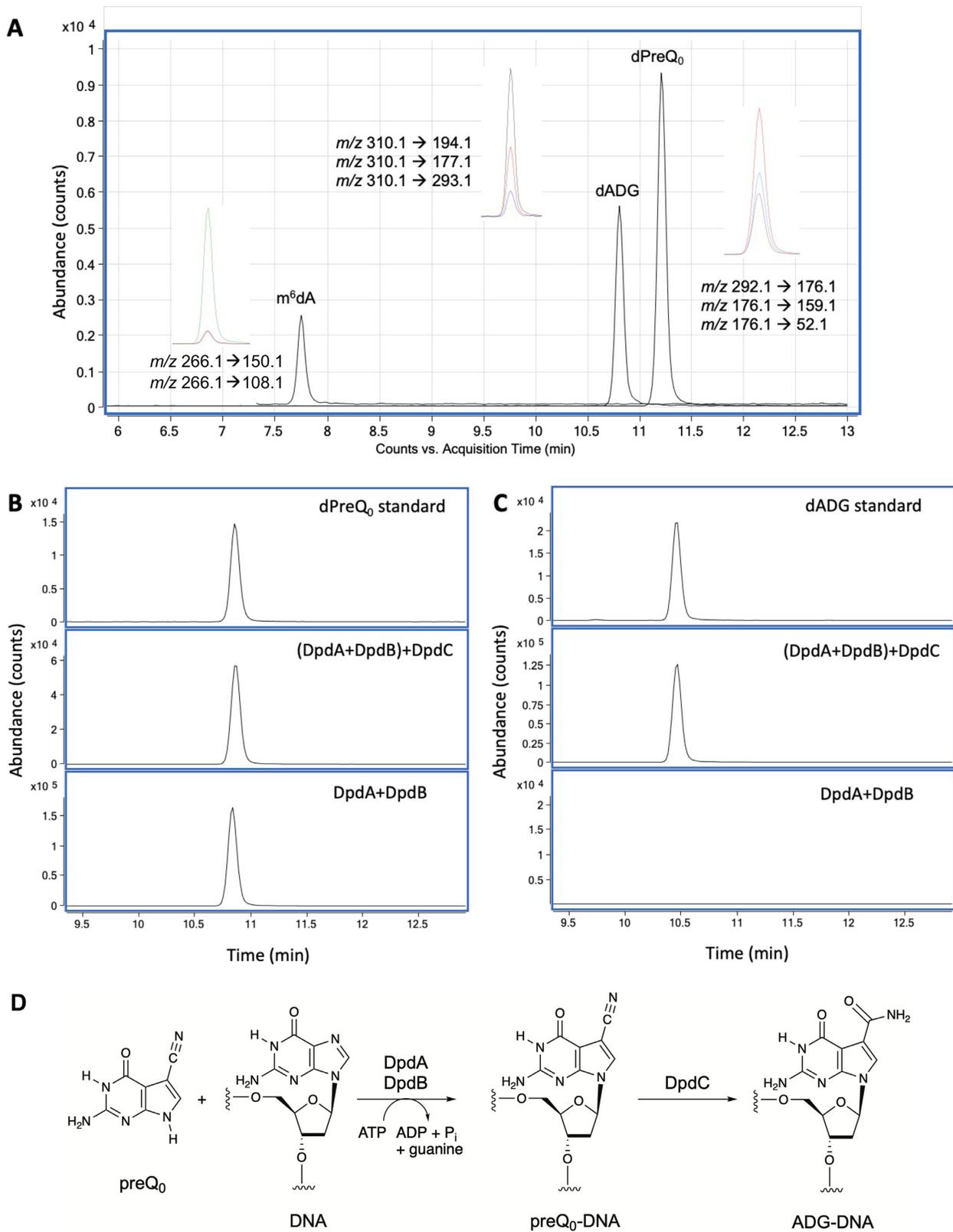
**Figure 3.** LC-MS analysis of 2′-deoxyribonucleosides produced during *in vitro* reactions of pUC19 with Dpd proteins. (**A**) Extracted ion chromatograms showing the three MS collision-induced dissociation (CID) transitions used to identify and quantify dADG and dPreQ$_0$, the only 2′-deoxyribonucleosides detected in the pUC19 DNA; the m$^6$dA is shown for reference and was not detected in the samples. Inset chromatograms show the main CID transition for loss of 2′-deoxyribose (116 Da; upper) and other CID transitions characteristic of each 2′-deoxyribonucleoside (middle and lower). (**B and C**) Extracted ion chromatograms for dPreQ$_0$ (**B**) and dADG (**C**) reveal that DpdA and DpdB are required for the formation of dPreQ$_0$ (middle panels) and the presence of DpdC is required to convert dPreQ$_0$ to dADG (lower panels). Chemical standards are shown in the upper panel for reference. (**D**) Reactions catalyzed by the DpdABC proteins.
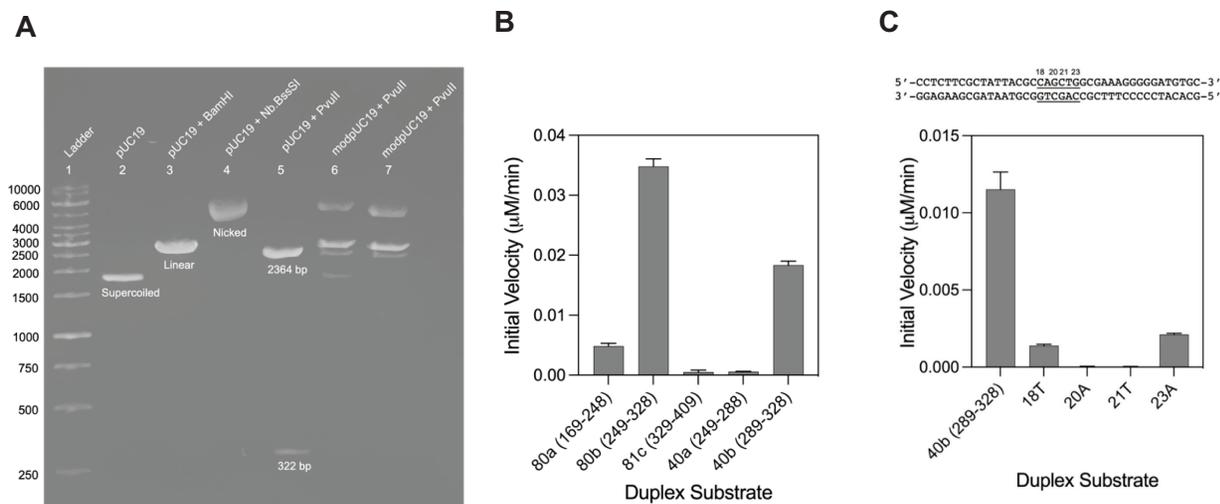
**Figure 4.** Identification of a modification site. (**A**) Restriction analysis of unmodified and preQ$_0$- and ADG-modified pUC19. Lane 1, DNA kb ladder; lane 2, supercoiled pUC19; lane 3, linear pUC19 (restricted with BamHI); lane 4, open circle pUC19 (nicked with BssSI); lanes 5–7, unmodified, preQ$_0$-modified and dADG-modified pUC19, respectively, restricted with PvuII. (**B**) [$^{14}$C]guanine exchange activity catalyzed by DpdA/B in the presence of various DNA duplexes derived from the small fragment of CviQI-restricted pUC19. The first number in the name corresponds to the duplex length, while the numbers in parentheses correspond to the region of pUC19 from which the fragment is derived in standard pUC19 numbering. (**C**) [$^{14}$C]guanine exchange activity catalyzed by DpdA/B in the presence of the duplex 40b and several 40b mutants. The sequence of 40b is shown with the PvuII site underlined and the mutation sites numbered according to their position in the sequence of the sense strand of 40b. For all plots, error bars represent standard error from at least two independent assays, each carried out in duplicate or triplicate.

seen in solution by SEC (Supplementary Figure S2). DpdA is an αβ protein (70 Å × 54 Å × 41 Å) built around the (β/α)$_8$ fold characteristic of the triosephosphate isomerase (TIM) barrel structural superfamily, except that the first helix in the TIM barrel is replaced by a one-turn 3$_{10}$-helix (η1 in Figure 5A). When compared with the basic (β/α)$_8$ TIM barrel proteins, DpdA differs by insertions in the β1/α1, β2/α2, β3/α3 and β8/α8 loops (Ins1, Ins2, Ins3 and Ins8, respectively, Figure 5B). Ins1 (residues Asp16–Asp36) is a 21-residue extended loop, Ins2 (residues Lys54–Thr68) is a 1.5-turn α-helix followed by a 10-residue glycine-rich loop and Ins3 (residues Gly97–Tyr109) is a single-turn 3$_{10}$-helix followed by an eight-residue loop. Ins8 (residues Ser256–Phe383) is a large insertion and represents 30% of the protein structure, and contains a two-turn α-helix (αB) laying roughly parallel to the TIM barrel helices, a short β-hairpin (βA and βB) followed by a protruding α-helical domain (αC–αF) and three orthogonal one-turn helices (αG, αH and ηB). Ins8 is pinned against the TIM barrel by a structural Zn site with a CXCX$_2$CX$_{22}$H motif, where the Zn$^{2+}$ ion is coordinated by Cys368, Cys370 and Cys373 from Ins8 and His395 from the terminal helix (α8) of the TIM barrel (Figure 5A; Supplementary Figure S4B). All insertions and the Zn site are conserved in the DpdA family (Supplementary Figure S5). The active site is located at the wide (C-terminal) face of the barrel (see below).

### Similarity of DpdA to other DNA and RNA modification enzymes (DNA repair and tRNA-dependent transglycosylses)

A search using the DALI server (44) yielded all the bacterial, archaeal and eukaryotic TGTs as the most similar structures to DpdA (Figure 5C–E). These include *Zy-momonous mobilis* bTGT [PDB ID 1P0B (45), root mean square deviation (rmsd) 2.8 Å over 291 C$_\alpha$ atoms, Z-score 26.2], *Pyrococcus horikoshii* aTGT [PDB ID 1IT8 (38), rmsd 2.8 Å over 283 C$_\alpha$ atoms, Z-score 25.6] and the human eTGT catalytic subunit [PDB ID 6H42 (40), rmsd 2.7 Å over 289 C$_\alpha$ atoms, Z-score 23.9]. 3D superpositions of *S.* Montevideo DpdA with these structures are shown in Supplementary Figure S6. A structure-based multisequence alignment of DpdA with the various TGTs of known structure shows 22–26% sequence similarity and 9–13% identity, and conservation of the catalytic residues and structural Zn$^{2+}$ site (Supplementary Figure S7). However, key differences are observed. In addition to the absence of the N-terminal β-sheet lid seen on the narrow face of the TIM barrel in the TGTs, the four insertions on the wide face of the TIM barrel of DpdA distinguish it from the TGTs (Figure 5B–E). Ins1 and Ins2 are DpdA specific, while Ins3 carries a different secondary structure (3$_{10}$-helix–loop in DpdA versus a β-sheet or α-helix in TGTs). Further, Ins8 is larger in DpdA by a 70-residue four-helix bundle (αC–αF). These TIM barrel insertions carry functional features specific to each family. For example, in *Z. mobilis* TGT, Ins3 and Ins8 flank the anticodon stem–loop of substrate tRNA and interact with the anticodon loop in a sequence-specific manner, providing needed specificity for tRNA substrates containing GUN anticodons, and Ins8 also provides numerous interactions with the anticodon stem (46). In DpdA, Ins1 and Ins8 extend the TIM barrel core, creating a crescent-shaped, positively charged surface that can accommodate double-stranded DNA (dsDNA; see below), while Ins2 and Ins3 flank the active site and thus may contribute to sequence-specific DNA recognition.
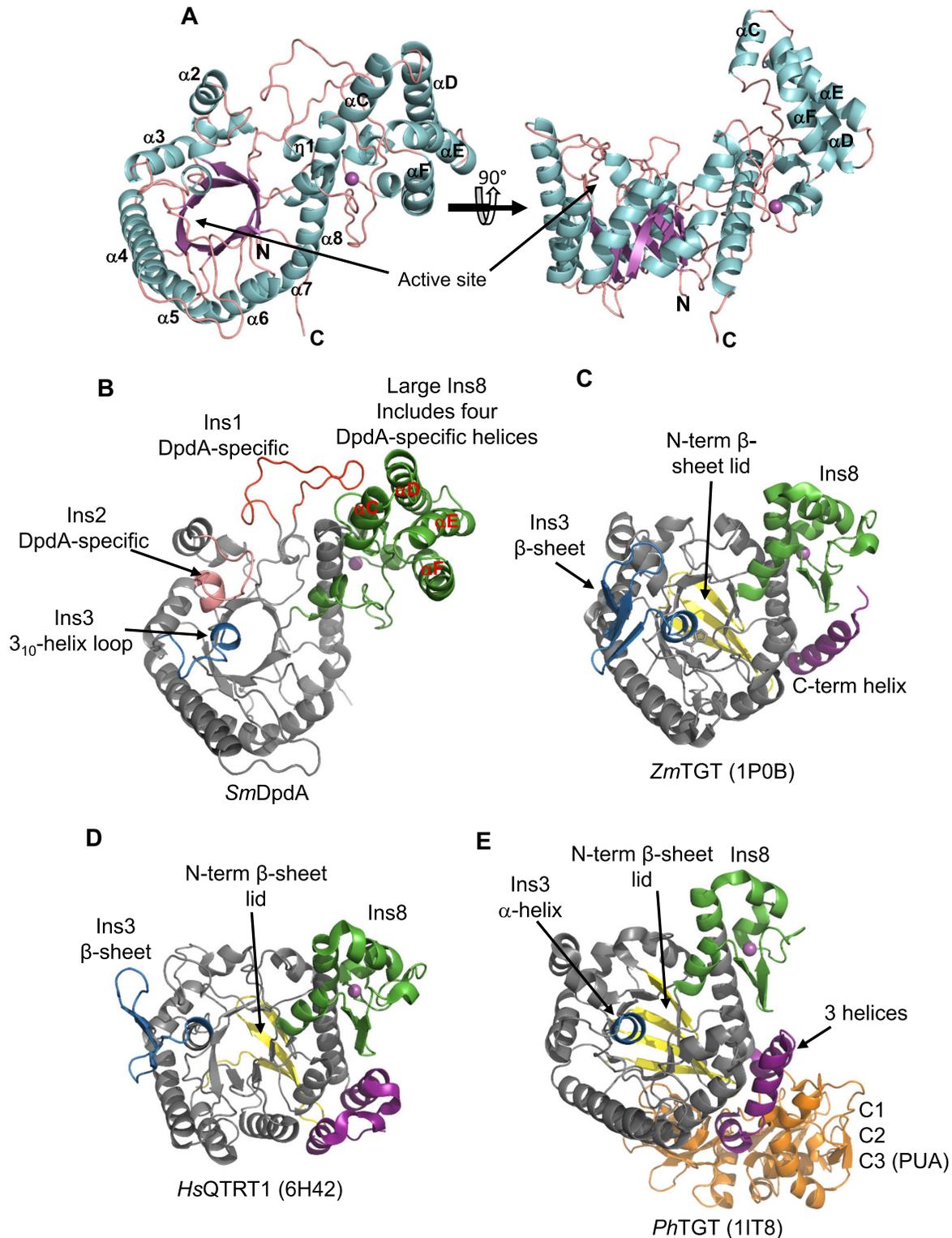
**Figure 5.** Crystal structure of *S.* Montevideo DpdA and comparison with the tRNA-guanine transglycosylases. (**A**) Orthogonal views of the overall structure with β-strands and helices colored in magenta and cyan, respectively. The active site is located within a deep pocket at the wide end of the β-barrel. For clarity, only the helices are labeled. The Zn ion is shown as a magenta ball. (**B–E**) Comparison of TIM barrel insertions of DpdA (B) with those of bTGT (C), QTRT1 (D) and aTGT (E). Shown are crystal structures of representative members of the families (*Sm*, *Salmonella enterica* serovar Montevideo; *Zm*, *Zymomonus mobilis*; *Hs*, *Homo sapiens*; *Ph*, *Pyrococcus horikoshii*). In all structures, the TIM barrel is colored in gray. Insertions Ins1, Ins2, Ins3 and Ins8 are colored in red, salmon pink, blue and green, respectively. The N- and C-terminal extensions are colored in yellow and purple, respectively. The Zn ions are shown as magenta balls. The C-terminal domains C1–C3 of aTGT are colored in orange. Insertions and some family-specific secondary structure features are labeled.

**Table 3.** X-ray data collection and structure refinement statistics for *S.* Montevideo DpdA

| Data collection | | | |
|---|---|---|---|
| Space group | | P2$_1$2$_1$2 | |
| Matthew's coefficient (Å$^3$/Da) | | 2.77 | |
| Solvent content (%) | | 55.6 | |
| Monomers/AU | | 1 | |
| Dataset | Peak | Remote | Inflection |
| Unit cell a, b, c (Å) | 86.10, 141.77, 42.50 | 86.12, 141.83, 42.50 | 86.13, 141.88, 42.51 |
| Wavelength (Å) | 1.28149 | 0.98397 | 1.28268 |
| Resolution (Å)[a] | 73.59–2.51 (2.65–2.51) | 73.61–2.65 (2.79–2.65) | 73.62–2.59 (2.73–2.59) |
| Measured reflections | 207 845 (11 600) | 225 828 (32 623) | 201 380 (13 261) |
| Unique reflections | 16 179 (1324) | 15 829 (2245) | 15 462 (1439) |
| Completeness (%) | 87.1 (50.4) | 100 (100) | 90.8 (59.6) |
| Multiplicity | 12.8 (8.8) | 14.3 (14.5) | 13.0 (9.2) |
| R-merge (%) | 0.103 (0.964) | 0.197 (1.488) | 0.105 (0.873) |
| R-meas (%) | 0.107 (1.025) | 0.204 (1.542) | 0.109 (0.925) |
| R-pim (%) | 0.029 (0.341) | 0.054 (0.400) | 0.040 (0.412) |
| CC$_{1/2}$ | 0.99 (0.92) | 1.0 (0.88) | 1.0 (0.93) |
| $<I/\sigma(I)>$ | 21.8 (2.3) | 15.1 (2.5) | 22.4 (2.8) |
| FOM after density modification 0.68 | | | |
| No. of Zn sites 1 | | | |
| **Structure refinement** | | | |
| Resolution range (Å) | | 54.78–2.51 (2.57–2.51) | |
| No. of reflections | | 15 348 (571) | |
| No. of atoms in the asymmetric unit | | | |
| Protein/water/Zn$^{2+}$ | | 3368/91/1 | |
| R$_{cryst}$/R$_{free}$[b] | | 0.132/0.179 | |
| Deviation from ideality | | | |
| Bond length (Å) | | 0.006 | |
| Bond angles (°) | | 1.078 | |
| Ramachandran plot—residues in: | | | |
| Favored (%) | | 96.6 | |
| Allowed (%) | | 2.9 | |
| Outliers (%)[c] | | 0.5 | |
| Mean B factor (Å$^2$) | | 55.9 | |
| Estimated coordinate error (Å)[d] | | 0.23 | |

[a]Highest resolution shell information in parentheses; [b]R$_{free}$ was monitored with 5% of the data excluded from the refinement; [c]two outliers, Gln33 and Arg384, both ordered residues and engaged in hydrophilic interactions with their envirnments; [d]value calculated based on R$_{free}$.

In addition to the TGTs, the DALI search yielded the crystal structure of the *E. coli* DNA repair enzyme Endo IV, also a TIM barrel enzyme, in complex with dsDNA (30) [PDB ID 1QUM, rmsd 4.1 Å over 195 C$_\alpha$ atoms, Z-score 11.9, Supplementary Figure S8], despite low sequence homology (11% similarity and 3.6% identity). Endo IV primes DNA repair synthesis by cleaving the DNA backbone immediately 5′ of apurinic/apyrimidinic sites in damaged dsDNA. Endo IV recognizes abasic sites by flipping both the abasic ribose and its orphan partner nucleotide out of duplex DNA and bending the DNA ~90° at the flipped-out nucleotides. To our knowledge, DpdA is only the third known example of the utilization of the TIM-fold in DNA metabolism, after Endo IV and the less characterized 3′→5′ single-stranded DNA (ssDNA)/RNA exonuclease TatD [PDB ID 4PE8 (47)].

**The DpdA active site suggests a double-displacement mechanism with a covalent intermediate**

The structural similarity of TGTs and Endo IV includes a conserved deep pocket at the center of the TIM barrel of DpdA that is the putative active site. Placing a guanine nucleoside in this pocket based on its maximum superposition with the active sites of the archaeal, bacterial or eukaryotic TGTs (46,48,49) reveals the same conserved enzyme–substrate interactions seen in TGTs (Figure 6A), and suggests a similar mechanism of binding the substrate nucleotide, consistent with the reaction catalyzed. In this model, Phe99 and Leu236, both conserved hydrophobic residues in the TGT and DpdA families, are positioned to sandwich the substrate base, Asp130 and His181 are positioned to hydrogen-bond to the Watson–Crick edge of the base, and Asp256 and Asp95 are positioned to act as a nucleophile and general acid/base, respectively, in the transglycosylation reaction. All three active site aspartate residues are invariant in both enzyme families (Supplementary Figure S7). These conserved residues provide a hydrogen bond network buried in a hydrophobic environment which facilitates specific recognition of the guanine and 7-deazaguanine bases, with the Phe99 side chain potentially acting as a gate that flips to open the active site and facilitate exchange of the leaving guanine base with the incoming 7-deazaguanine, as has been proposed for TGTs (46).

To test this model we conducted site-directed mutagenesis of Asp265 and Asp95 and investigated the ability of the mutant proteins to catalyze the guanine exchange reaction (in the presence of DpdB and ATP) in pUC19 DNA. Substitution of either residue with alanine abolished the guanine exchange activity (Figure 6B). We therefore propose that DpdA utilizes a mechanistic strategy similar to that of the tRNA-dependent TGTs for the exchange of the ge-
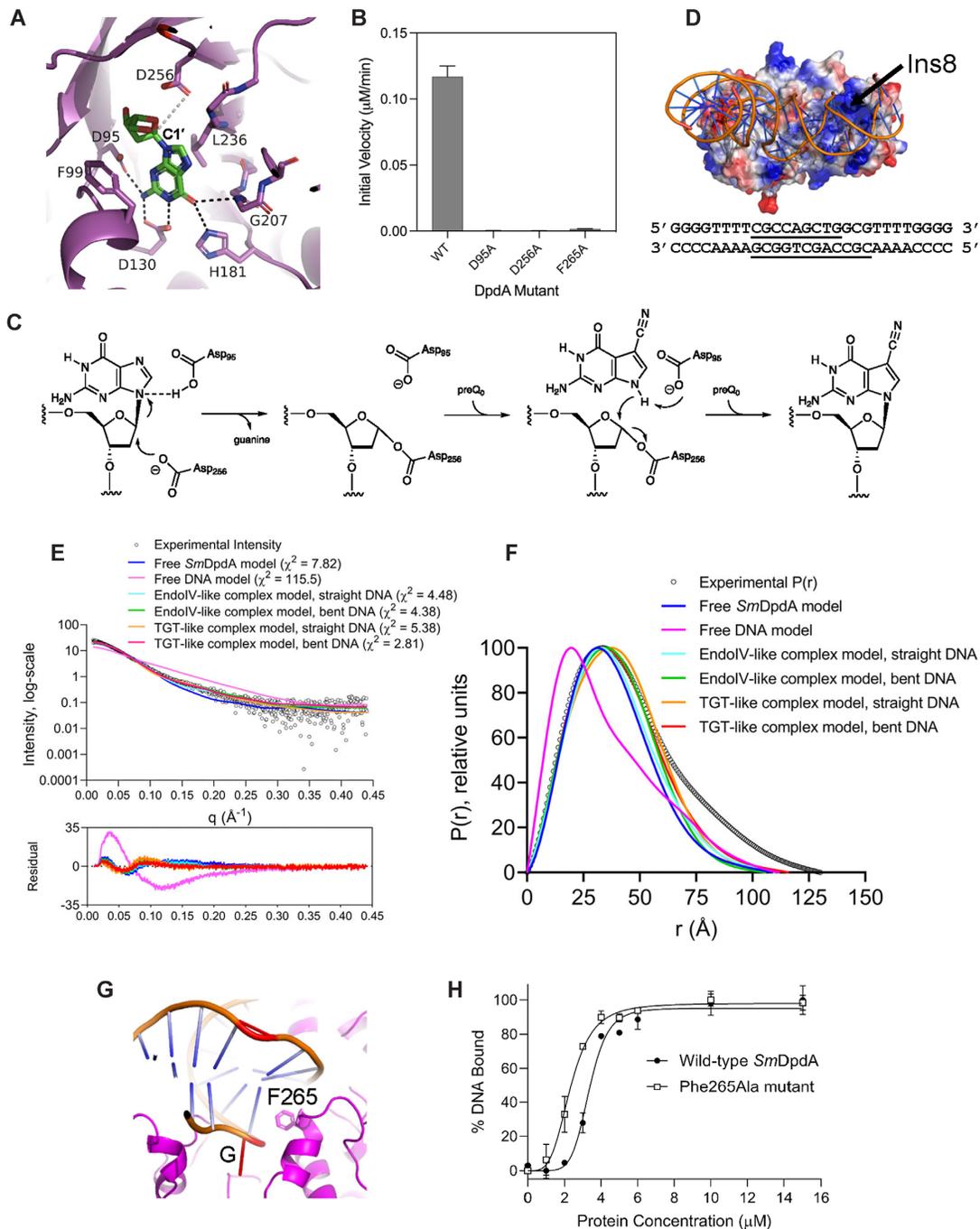
**Figure 6.** Insights into DpdA mechanism and DNA recognition. (A–C) The active site of DpdA suggests a similar base exchange mechanism to tRNA-dependent translgycosylases. (**A**) Close-up view of a docked guanosine nucleoside (green) in the active site of *S.* Montevideo DpdA showing putative coordination of the base and ribose by conserved active site residues (magenta). Putative interactions are indicated as dashed lines. (**B**) Initial velocities of [$^{14}$C]guanine exchange activities of the wild-type *S.* Montevideo DpdA/B, the Asp95Ala and Asp256Ala active site mutants and the Phe265Ala mutant using pUC19 DNA as substrate. Error bars represent the standard error. (**C**) Proposed minimal catalytic mechanism of DpdA showing the key roles of Asp95 and Asp256. (**D-G**) SAXS analysis of *S.* Montevideo DpdA in complex with 28 bp DNA duplex carrying the modification target sequence. (**D**) Model of the DpdA–DNA complex that yielded the best fit to the SAXS data, with the DNA bent and docked in a TGT-like orientation. The electrostatic surface potential of the protein is shown and the DpdA-specific insertion Ins8 is indicated. The sequence of the 28 bp DNA duplex used in SAXS experiments and model building is shown on the bottom, with the modification target sequence underlined. (**E**) Comparison of the experimental scattering curve (black circles) with the theoretical curves of the various models (colored lines). The SAXS model in (D) and other models in which DNA is not bent or docked in an Endo IV-like orientation are compared. The $\chi^2$ values for each fit and the associated residual plots are shown. (**F**) Corresponding P(r) functions calculated from the experimental curve (black circles) and theoretical curves (colored lines). (**G**) Close up view of the DNA-binding surface of *S.* Montevideo DpdA near the DNA bend, based on the SAXS model of the DpdA–DNA complex, showing Phe265 putatively positioned in proximity to the modification site in the minor groove of DNA. (**H**) Binding of wild-type DpdA (black circles, apparent $K_d = 3.4 \pm 0.2\,\mu$M) and Phe265Ala mutant (open squares, apparent $K_d = 2.3 \pm 0.1\,\mu$M) to the 28 bp DNA duplex, as measured by EMSA (see EMSA gels in Supplementary Figure S11), $n = 3$. *Sm*, *Salmonella enterica* serovar Montevideo.

**Table 4.** Parameters calculated from SAXS and MALS data of *S.* Montevideo DpdA in complex with 28 bp DNA duplex, and of samples containing DpdA alone and DNA alone

|  | DNA | DpdA | DpdA–DNA |
| --- | --- | --- | --- |
| **Structural parameters** | | | |
| $R_g$ (Å) from Guinier plot | $26.2 \pm 0.5$ | $27.0 \pm 0.3$ | $34.1 \pm 0.3$ |
| $R_g$ (Å) from P(r) | $28.0 \pm 0.4$ | $27.2 \pm 0.3$ | $36.3 \pm 0.2$ |
| $D_{max}$ (Å) P(r) GNOM | 102 | 100 | 130 |
| Porod volume estimate, $V_p$ (Å$^3$) | 24 188 | 67 005 | 103 690 |
| **Molecular mass (kDa)** | | | |
| Shape and size | 23 | 47 | 68 |
| SAXSMoW 2.0 | 20 | 51 | 69 |
| Theoretical from sequence | 17 | 50.4 | 67 |
| MALS[a] | 53 | 56 | 64 |

[a]Multiangle light scattering.

netically encoded guanine for preQ$_0$ in DNA (46,49) (Figure 6C). In this proposal, Asp95 functions as a general acid to activate the guanine for elimination, and Asp256 functions as an enzyme nucleophile to displace the guanine and form a covalent enzyme–DNA intermediate. With the subsequent binding of preQ$_0$, the deprotonated Asp95 serves as a general base, facilitating the nucleophilic attack of N9 of preQ$_0$ at the glycosidic carbon to break down the covalent intermediate and form preQ$_0$-modified DNA.

## SAXS analysis of DpdA–DNA complex suggests a novel mode of DNA binding by the TIM fold

To gain insight into how DpdA binds DNA, we performed SEC-SAXS experiments on the *S.* Montevideo DpdA$^{D95A}$ mutant in complex with a 28 bp DNA duplex that possesses a modification site (sequence shown in Figure 6D), as well as on the free protein and free DNA (Supplementary Figure S9C–F). The catalytically inactive mutant protein was chosen for these experiments to prevent product formation and trap the pre-exchange complex. Structural parameters derived from the SAXS data are listed in Table 4. The protein and protein–DNA complex exhibit globular compact structures as indicated by their dimensionless Kratky plots (Supplementary Figure S9A). Guided by the structural similarity of DpdA to Endo IV which suggested that, similarly to Endo IV, DpdA may bend the DNA, we generated a model of a bent 28 bp DNA duplex using the bent DNA from the crystal structure of the Endo IV–DNA complex (30) as a starting template, and docked the constructed 28 bp bent DNA model on the DpdA crystal structure using HADDOCK (31) as described in the Materials and Methods. In the resulting energy-minimized model, the DNA is bound in a bent conformation along the long axis of the positively charged concave surface on the open face of the TIM barrel of DpdA, and contacting the helical bundle of Ins8 (Figure 6D). This model was used to calculate theoretical scattering curves and intra-particle distance distribution functions [P(r)] (Figure 6E, F) and provided a good fit to the experimental scattering data ($\chi^2 = 2.81$), with an experimental molecular weight of 68–69 kDa, close to the theoretical molecular weight of 67 kDa (Table 2).

When compared with the crystal structure of *E. coli* Endo IV bound to DNA [PDB ID 1QUM (30)], the SAXS model suggests that the TIM barrel of DpdA binds DNA in a different orientation, ~90° away from its general orientation in the Endo IV–DNA structure (compare Supplementary Figures S10A and B) in a fashion similar to TGT enzymes as seen in their crystal structures with bound RNA substrates (specifically *Z. mobilis* TGT [PDB ID 1Q2R (46)]) or human TGT in complex with ASL RNA [PDB ID 7NQ4 (49)]. In these structures, the helical stem of RNA lies parallel to the long axis of the protein molecule, extending from the center of the TIM barrel to Ins8 (compare Supplementary Figures S10A and C–E). Indeed, models of the complex in which DNA is not bent and/or docked onto the protein in an Endo IV-like orientation provided a worse fit to the experimental scattering data, with $\chi^2$ values >4 for models in which DNA is bound in an Endo IV-like orientation (with or without bending), and $\chi^2 = 5.38$ for the model in which DNA is bound in a TGT-like orientation without bending (Figure 6E, F).

Based on the structural homology between DpdA and Endo IV and the SAXS model, we hypothesized that, similarly to Endo IV, DpdA might flip the base pair at the modification site and stabilize the flipped base pair by filling the hole left by the flipped pair with an aromatic side chain. We identified Phe265 as the aromatic residue that could play that role in *S.* Montevideo DpdA (Figure 6G). Phe265 is the only conserved, solvent-exposed aromatic side chain on the DNA-binding face of the TIM barrel, and it is located in Ins8 and points into the major groove of bound DNA in the vicinity of the active site. Substitution of Phe265 with alanine resulted in complete loss of guanine exchange activity (Figure 6B), even though the mutant binds to the 28 bp DNA duplex containing a modification site with affinity comparable with the wild-type enzyme as measured by EMSA (Figure 6H; Supplementary Figure S11). These data are consistent with Phe265 playing a role in a step subsequent to DNA binding, such as stabilizing a flipped base pair and bent DNA conformation at the target site.

## SAXS solution structure of DpdB

DpdB is annotated in databases as a member of the ParB-like protein structural superfamily. Members of this superfamily have been identified in biologically diverse contexts and include the prokaryotic chromosome segregation protein and CTPase ParB (50), the DNA phosphorothioation proteins DndB (51) and SspE (52), and the functionally distant eukaryotic sulfiredoxin Srx which catalyzes the repair of hyperoxidized cysteines in peroxiredoxin (53). To gain further insight into the function of DpdB proteins, we performed SEC-SAXS and MALS analyses of *S.* Montevideo DpdB in solution. Structural parameters derived from the SAXS data are listed in Table 5. As expected, the molecular mass estimated from the SAXS and MALS data (80–95 kDa) is double the monomer mass based on sequence, consistent with a homodimeric structure and with SEC analysis (Supplementary Figure S2). To fit the SAXS data, we used a 3D model generated by the artificial intelligence structure prediction program AlphaFold (54) which predicted a pyramid-shaped homodimer consisting of an N-

**Table 5.** Parameters calculated from SAXS and MALS data for *S.* Montevideo DpdB

| Structural parameters | DpdB |
| --- | --- |
| $R_g$ (Å) from Guinier plot | $34.0 \pm 0.4$ |
| $R_g$ (Å) from P(r) | $36.8 \pm 1.1$ |
| $D_{max}$ (Å) from GNOM | 164 |
| Porod volume estimate, Vp ($\text{Å}^3$) | 125 967 |
| **Molecular mass (kDa)** | |
| Shape a size | 90 |
| SAXSMoW 2.0 | 95 |
| Theoretical from sequence[a] | 44.2 |
| MALS[b] | 80 |
| DENSS RSC[c] | $0.84 \pm 0.03$ |
| FSC[d] | 0.97 |

[a]Monomer mass including the $His_6$ tag; [b]multiangle light scattering; [c]RSC, real space correlation averaged over 20 envelopes; [d]FSC, Fourier shell correlation averaged over 20 envelopes.

terminal ParB-fold domain and an α-helical C-terminal domain (Figure 7). The theoretical scattering curve and intra-particle distance distribution function [P(r)] calculated from this model provided a good fit to the experimental scattering data ($\chi^2 = 1.73$, Figure 7A, B). However, the SAXS-derived experimental maximum particle size ($D_{max}$) was significantly larger than that presented by the AlphaFold model (164 Å versus 92 Å), suggesting multiple conformations in solution. Indeed, conformational sampling followed by multistate modeling resulted in a two-state model constituted of two conformations with equal distribution, and this model provides the best fit to the SAXS data ($\chi^2 = 0.94$, Figure 7A). In this model, DpdB undergoes a conformational change around a five-residue proline-rich linker ($T_{156}KPLP_{160}$) connecting the two domains within the monomer. The two states represent a closed conformation similar to the AlphaFold prediction with a 40 Å distance between the centers of mass of the two C-terminal domains, and an open conformation in which the C-terminal domains are apart by 86 Å (Figure 7C). This model shows an excellent fit to the *ab initio* shape reconstructed from the SAXS data without applying 2-fold symmetry (Figure 7D). To validate the fit of the two-state model to the *ab initio* density, we aligned the theoretical envelopes calculated from the closed, open and two-state models to the *ab initio* envelope, and calculated the Fourier shell correlation (FSC) of the aligned envelopes (theoretical versus experimental, Supplementary Figure S12). The results show that the theoretical envelope calculated from the two-state model shows the best fit to the *ab initio* envelope with overall FSC of 0.86, while the closed model and open model each provides a less fitting theoretical envelope, with overall FSC values of 0.74 and 0.75, respectively.

A search for similar structures to the SAXS closed model using the DALI server (44) identified a number of ParB-fold proteins as the top hits, consistent with initial predictions. The top five hits are *Myxococcus xanthus* PadC [PDB ID 4RYK (52), rmsd 3.2 Å over 119 $C_\alpha$ atoms, Z-score 7.9, 12% sequence identity], *Geobacillus thermoleovorans* nucleoid occlusion protein Noc [PDB ID 7NFU (55), rmsd 7.6 Å over 116 $C_\alpha$ atoms, Z-score 7.6, 16% sequence identity], *Saccharolobus solfataricus* pNOB8-like ParB [PDB ID

5K5A (56), rmsd 10.5 Å over 125 $C_\alpha$ atoms, Z-score 7.5, 8% identity], *Agrobacterium tumefaciens* hypothetical protein Atu1540 [PDB ID 2HWJ, rmsd 3.9 Å over 97 $C_\alpha$ atoms, Z-score 5.7, 9% identity] and *Arabidopsis thaliana* sulfiredoxin [PDB ID 6KY4 (57), rmsd 2.9 Å over 84 $C_\alpha$ atoms, Z-score 7.2, 12% identity]. Additionally, one of the top 10 hits is the recent crystal structure of the NTPase and endonuclease SspE from *Streptomyces yokosukanensis* [PDB ID 6JIV (52), rmsd 4.4 Å over 165 $C_\alpha$ atoms, Z-score 4.4, 12% sequence identity and 26% similarity] (Supplementary Figure S13). SspE is a component of the SspABCD–SspE RM system involved in anti-phage defense based on single-stranded phosphorothioation of the DNA sugar–phosphate backbone (52). SspE acts as a phosphorothioate-sensing NTPase and endonuclease that nicks and degrades invasive phage DNA.

The N-terminal domain of DpdB harbors a conserved DGQQR motif that was shown in SspE to be required for the NTPase activity (52) (Supplementary Figure S13), and that is analogous to the GxxRxxA motif directly involved in nucleotide binding in the CTPase domain of the chromosome segregation protein ParB (50). The C-terminal α-helical domain of DpdB is homologous to the endonuclease domain of SspE except that it lacks the SspE-specific motif EHxxP required for the nicking activity, consistent with the absence of endonuclease activity for DpdB. Further, superposition onto the CTPase domain of the homodimeric *B. subtilis* ParB [PDB ID 6SDK (50)] or *M. xanthus* PadC [PDB ID 4RYK (52)] with a CDP nucleotide bound in the active site provides insight into the putative ATP-binding site of DpdB (Figure 7E). The model suggests that, like the ParB CTPase domain which mediates CTP-dependent homodimerization and assembly on DNA, the ATPase domain of DpdB may provide a homodimer interface that is regulated by ATP binding and hydrolysis. Consistent with this, the C-terminal domains in the closed state (or AlphaFold) model encircle a positively charged central tunnel with an average width of 20 Å, wide enough to accommodate dsDNA (Figure 7F).

## DISCUSSION

The Dpd system represents the largest and most complex RM system yet discovered, but the rationale for this complexity is not yet understood. Based on prior genetic data (13) and now *in vitro* biochemical investigations, the modification components of this system clearly comprise the proteins DpdA, DpdB and DpdC, with DpdA and DpdB responsible for incorporation of $preQ_0$ into DNA, and DpdC responsible for the subsequent conversion of $preQ_0$-modified DNA to dADG-modified DNA in a reaction independent of DpdA and DpdB.

The stoichiometry of *in vitro* pUC19 modification was consistent in both the radiochemical and $preQ_0$ modification assays, but was an order of magnitude higher than the level of *in vivo* genomic modification (13), suggesting that *in vivo* modification is incomplete, or that secondary, non-cognate sites are being modified under the conditions of the *in vitro* assays. Insight into the identity of the modification site(s) was provided by restriction endonuclease analysis, which demonstrated that of the endonucleases pos-
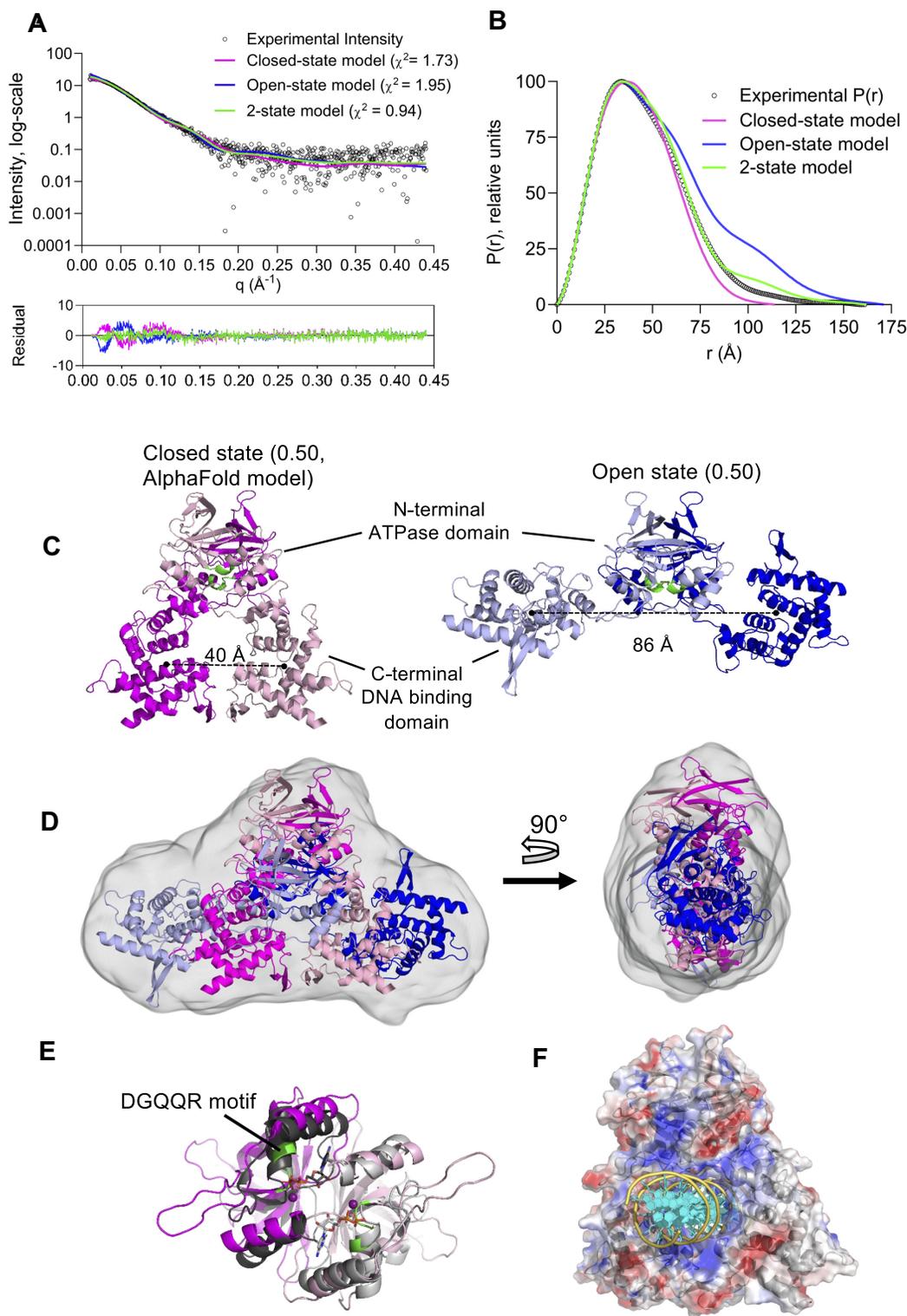
**Figure 7.** SAXS solution structure of *S*. Montevideo DpdB. (**A**) Comparison of the experimental scattering curve (black circles) with the theoretical curves of the various models (colored lines), the $\chi^2$ values for each fit and the associated residual plots. (**B**) P(r) functions calculated from the experimental curve (black circles) and theoretical curves (colored lines). (**C**) Ribbon diagrams of the closed and open states of the DpdB homodimer in the two-state model, with relative abundances shown in parentheses. Monomers are colored in different shades, the domains are labeled and distances between centers of mass of the C-terminal domains across the dimer are shown. (**D**) *Ab initio* reconstruction of the molecular envelope (gray surface) calculated from the SAXS data and overlaid on the closed and open models in magenta and blue, respectively. (**E**) The ATPase domain of DpdB extracted from the SAXS model and superposed on the homodimeric CTPase domain of ParB [PDB ID 6SDK]. The DGQQR motif in DpdB is highlighted in green, and the CDP molecules and $Ca^{2+}$ ions bound to ParB are shown as a stick model and purple balls, respectively. (**F**) Docking model of dsDNA onto the DpdB closed model shown in electrostatic surface potential representation. The view is rotated 15° counter-clockwise around the vertical axis relative to (**C**).

sessing G in various recognition sequence contexts investigated here (Figure 4; Supplementary Figure S3) and previously (13), only PvuII activity was disrupted by modification (Figure 4), consistent with modification within the PvuII recognition/cleavage sequence (CAG/CTG). However, the fact that only three PvuII sites exist within pUC19 while modification occurs at >15 sites (Table 2; Supplementary Figure S1C) indicates that the DpdA recognition site is not identical to PvuII. Experiments with smaller duplexes derived from pUC19 in which the PvuII site was disrupted (Figure 4C) are consistent with guanine exchange at the GC base pair in positions 3 and 4 of the PvuII sequence (CA**G**/**C**TG), as duplexes 20A and 21T exhibited no detectable activity as substates in the $[^{14}C]$guanine exchange reaction (Figure 4C). The failure of 20A and 21T to act as substrates demonstrates that 40b possesses only a single recognition site, and that a C following the target G is required for recognition by DpdA. Furthermore, the low but detectable activity exhibited with 18T and 23A suggests that these positions are also part of the DpdA recognition motif, consistent with CxGCxG comprising at least a portion of the recognition sequence, where the identity of the 'x' nucleotides is still undetermined. While the unimpeded restriction of $preQ_0$- or dADG-modified pUC19 by the remaining restriction endonucleases is consistent with this minimal motif, elucidation of the complete recognition motif awaits more rigorous, systematic experiments directly probing the sequence requirements for recognition by DpdA, and that work is in progress.

Given that both $preQ_0$ and $preQ_1$ are present as stable metabolites in bacterial cells, and that $dPreQ_1$ is not detected in bacterial DNA, the bacterial DpdA must recognize $preQ_0$ with high selectively. To gain insight into how bacterial DpdA enzymes recognize $preQ_0$ and discriminate against $preQ_1$, we placed $preQ_0$ and $preQ_1$ in the active site of *S.* Montevideo DpdA based on the positions of these bases in the active sites of *P. horikoshii* TGT [PDB ID 1IT8 (38)] and *Z. mobilis* TGT [PDB ID 1P0E (45)], respectively, as seen in their crystal structures in complex with substrates (38,40,45). Placement was done manually by maximally superposing the conserved active site residues D95, D130, D256, H181, G182, G206 and G207 with their counterparts in the respective TGT. TGTs coordinate the seven-substituent moiety of their substrates through backbone interactions with residues in a conserved Gly-rich loop connecting the sixth β-strand and α-helix of the TIM barrel, while structural features in the same loop preclude binding of 'non-cognate' substrates (Supplementary Figure S14). For example, *P. horikoshii* TGT coordinates the 7-cyano moiety of $preQ_0$ via a hydrogen bond donated by the backbone NH of Val198 (in *P. horikoshii* TGT residue numbers), while the side chain of that valine protrudes into and restricts the substrate-binding pocket, preventing accommodation of the bulky aminomethyl moiety of $preQ_1$ (Supplementary Figure S14A) (38,58). In *Z. mobilis* TGT, the analogous valine residue (Val233 in *Z. mobilis* TGT numbers) points away from the pocket through a backbone flip, allowing space for accommodation of the 7-aminomethyl moiety of $preQ_1$ (Supplementary Figure S14B). In the catalytic subunit of eukaryotic TGT [e.g. *H. sapiens* QTRT1, PDB ID 6H45 (40)], the absence of a valine side chain creates space

for accommodation of the bulky cyclopentenediol ring of queuine (Supplementary Figure S14C).

The $preQ_0$-docked *S.* Montevideo DpdA models show that the cyano group of $preQ_0$ can be accommodated through hydrogen bonding with the backbone NH of Met208 (Supplementary Figure S14D), similar to the interaction seen in aTGT, while the conserved Val209 side chain protruding into the substrate-binding pocket clashes with and thus obstructs accommodation of the bulky aminomethyl moiety of $preQ_1$ (Supplementary Figure S14E). Further, Val209 is followed by two conserved bulky side chains (Pro210–Leu211) that would hamper its backbone flipping and associated movement of its side chain to make room for the aminomethyl of $preQ_1$. These two residues are also bulky in the aTGT family (P199–Leu200 in *P. horikoshii* TGT), consistent with its $preQ_0$ specificity, and are glycines or other small amino acids (GlyGly/Ser/Ala) in bTGT, QTRT1 and the eurkaryotic-like *C. trachomatis* TGT, which would allow flexibility in the polypeptide backbone, providing space for accommodation of a bulky seven-substituent group. In aTGT and DpdA, the presence of a proline residue (Pro197 in *P. horikoshii* TGT and Pro210 in *S.* Montevideo DpdA) after Val198 (Val209 in *S.* Montevideo DpdA) hampers the mobility of Val198, while in bTGT and eukaryotic QTRT1 that proline residue is replaced by Gly and/or is followed by Gly. This suggests the consensus sequence motif GGhsbb (h = hydrophobic, s = small, b = bulky amino acid) for the $preQ_0$-targeting enzymes aTGT and bacterial DpdA, and GGhsG/sG (h = hydrophobic, s = small amino acid) for $preQ_1$- and queuine-targeting TGTs (Supplementary Figure S14).

The ability of the tRNA-dependent TGT enzymes (59) to catalyze the base exchange reaction without an ATP dependence, as well as DNA-modifying enzymes to catalyze their reactions in an ATP-independent process (60), renders the ATP dependence (or more generally NTP) in the bacterial Dpd system somewhat of an enigma. Indeed, as already noted, the structure of *S.* Montevideo DpdA exhibits significant homology to the TGT enzymes, and SEC-SAXS experiments with a bound 28 bp DNA duplex suggest that DpdA binds DNA in a similar manner to TGT binding of tRNA. Furthermore, neither the crystal structure nor the SAXS data reveals any structural impediment to catalysis in the absence of ATP hydrolysis by DpdB. While the role of the NTPase activity of the homologous SspE is not clear, it has been proposed that the phosphorothioate-enhanced NTP hydrolysis by SspE provides the energy needed to translocate SspE (perhaps via hydrolysis-driven conformational changes in the protein) along the DNA substrate, allowing processive nicking along the DNA (52). A similar role was proposed for the CTPase activity of the DNA segregation protein ParB (50), where CTP binding promotes the assembly of ParB homodimers on DNA while CTP hydrolysis serves to release ParB from DNA, thereby enabling its recycling to bind to another cognate site. The structural homology of DpdB to the ParB family NTPases SspE and ParB suggests that the ATPase activity of DpdB may serve a similar function or functions, and the observation that the ATPase activity of DpdB is enhanced in the presence of DpdA–DNA (Figure 2C) is consistent with such a role. Alternatively, ATP hydrolysis may be associated with a

helicase-like activity necessary to generate local regions of ssDNA that may be required for the formation of a catalytically competent complex of DpdA and DNA. These and related questions are the focus of ongoing investigation.

## DATA AVAILABILITY

Atomic coordinates and structure factors for *S. Montevideo* DpdA have been deposited in the Protein Data Bank under accession number 7UI4. The SAXS data have been deposited in the Small Angle Scattering Biological Data Bank under accession numbers SASDQ59, SASDQ89, SASDQ79 and SASDQ69 for DpdA$^{D95A}$, DpdB, the DpdA$^{D95A}$–DNA complex and 28 bp DNA, respectively.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Dr Michal Hammel for help with SAXS data processing, Susan Bayooz for help with initial protein production in the Swairjo lab, and Eric Hellie for help with generation of DNA models. The graphical abstract was created with BioRender.com.

## REFERENCES

1. Grosjean,H. (2009) Nucleic acids are not boring long polymers of only four types of nucleosides: A guided tour. In: Grosjean,H. (ed.) *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin, pp. 1–18.
2. Boccaletto,P., Machnicka,M.A., Purta,E., Piatkowski,P., Baginski,B., Wirecki,T.K., de Crecy-Lagard,V., Ross,R., Limbach,P.A., Kotter,A. *et al.* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, **46**, D303–D307.
3. Weigele,P. and Raleigh,E.A. (2016) Biosynthesis and function of modified bases in bacteria and their viruses. *Chem. Rev.*, **116**, 12655–12687.
4. Thiaville,J.J., Kellner,S.M., Yuan,Y., Hutinet,G., Thiaville,P.C., Jumpathong,W., Mohapatra,S., Brochier-Armanet,C., Letarov,A.V., Hillebrand,R. *et al.* (2016) Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc. Natl Acad. Sci. USA*, **113**, E1452–E1459.
5. Hutinet,G., Kot,W., Cui,L., Hillebrand,R., Balamkundu,S., Gnanakalai,S., Neelakandan,R., Carstens,A.B., Fa Lui,C., Tremblay,D. *et al.* (2019) 7-Deazaguanine modifications protect phage DNA from host restriction systems. *Nat. Commun.*, **10**, 5442.
6. Crippen,C.S., Lee,Y.J., Hutinet,G., Shajahan,A., Sacher,J.C., Azadi,P., de Crecy-Lagard,V., Weigele,P.R. and Szymanski,C.M. (2019) Deoxyinosine and 7-deaza-2-deoxyguanosine as carriers of genetic information in the DNA of *Campylobacter* viruses. *J. Virol.*, **93**, e01111-19.
7. Hutinet,G., Swairjo,M.A. and de Crecy-Lagard,V. (2017) Deazaguanine derivatives, examples of crosstalk between RNA and DNA modification pathways. *RNA Biol.*, **14**, 1175–1184.
8. Iwata-Reuyl,D. and de Crecy-Largard,V. (2009) Enzymatic formation of the 7-deazaguanosine hypermodified nucleosides of tRNA. In: Grosjean,H. (ed.) *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin, pp. 377–391.
9. Okada,N., Noguchi,S., Kasai,H., Shindo-Okada,N., Ohgi,T., Goto,T. and Nishimura,S. (1979) Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction. *J. Biol. Chem.*, **254**, 3067–3073.
10. Okada,N., Harada,F. and Nishimura,S. (1976) Specific replacement of Q base in the anticodon of tRNA by guanine catalyzed by a cell-free extract of rabbit reticulocytes. *Nucleic Acids Res.*, **3**, 2593–2603.
11. Stengl,B., Reuter,K. and Klebe,G. (2005) Mechanism and substrate specificity of tRNA-guanine transglycosylases (TGTs): tRNA-modifying enzymes from the three different kingdoms of life share a common catalytic mechanism. *Chembiochem*, **6**, 1926–1939.
12. Iyer,L.M., Zhang,D., Burroughs,A.M. and Aravind,L. (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res.*, **41**, 7635–7655.
13. Yuan,Y., Hutinet,G., Valera,J.G., Hu,J., Hillebrand,R., Gustafson,A., Iwata-Reuyl,D., Dedon,P.C. and de Crecy-Lagard,V. (2018) Identification of the minimal bacterial 2′-deoxy-7-amido-7-deazaguanine synthesis machinery. *Mol. Microbiol.*, **110**, 469–483.
14. Migawa,M.T., Hinkley,J.M., Hoops,G.C. and Townsend,L.B. (1996) A two step synthesis of the nucleoside Q precursor 2-amino-5-cyanopyrrolo(2,3-*d*)pyrimidin-4-one (PreQ$_0$). *Synth. Comm.*, **26**, 3317–3322.
15. Van Lanen,S.G., Kinzie,S.D., Matthieu,S., Link,T., Culp,J. and Iwata-Reuyl,D. (2003) tRNA modification by S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Assay development and characterization of the recombinant enzyme. *J. Biol. Chem.*, **278**, 10491–10499.
16. Hediger,M.A. (1986) High resolution preparative gel electrophoresis of DNA fragments and plasmid DNA using a continuous elution apparatus. *Anal. Biochem.*, **159**, 280–286.
17. Kabsch,W. (2010) XDS. *Acta Crystallogr D*, **66**, 125–132.
18. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
19. Murshudov,G.N., Skubak,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
20. Emsley,P. and Cowtan,K. (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
21. Wyatt,P.J. (1993) Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta*, **272**, 1–40.
22. Tully,M.D., Tarbouriech,N., Rambo,R.P. and Hutin,S. (2021) Analysis of SEC-SAXS data via EFA deconvolution and Scatter. *J. Vis. Exp.*, https://doi.org/10.3791/61578.
23. Manalastas-Cantos,K., Konarev,P.V., Hajizadeh,N.R., Kikhney,A.G., Petoukhov,M.V., Molodenskiy,D.S., Panjkovich,A., Mertens,H.D.T., Gruzinov,A., Borges,C. *et al.* (2021) ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **54**, 343–355.
24. Hopkins,J.B., Gillilan,R.E. and Skou,S. (2017) BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.*, **50**, 1545–1553.

25. Franke,D., Jeffries,C.M. and Svergun,D.I. (2018) Machine learning methods for X-ray scattering data analysis from biomacromolecular solutions. *Biophys. J.*, **114**, 2485–2492.

26. Piiadov,V., Ares de Araújo,E., Oliveira Neto,M., Craievich,A.F. and Polikarpov,I. (2019) SAXSMoW 2.0: online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Sci.*, **28**, 454–463.

27. Svergun,D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.

28. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.*, **105**, 962–974.

29. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2016) FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.*, **44**, W424–W429.

30. Hosfield,D.J., Guan,Y., Haas,B.J., Cunningham,R.P. and Tainer,J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397–408.

31. van Zundert,G.C., Rodrigues,J.P., Trellet,M., Schmitz,C., Kastritis,P.L., Karaca,E., Melquiond,A.S., van Dijk,M., de Vries,S.J. and Bonvin,A.M. (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **428**, 720–725.

32. Lindorff-Larsen,K., Piana,S., Palmo,K., Maragakis,P., Klepeis,J.L., Dror,R.O. and Shaw,D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.

33. Abraham,M.J., Murtola,T., Schulz,R., Páll,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1-2**, 19–25.

34. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

35. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

36. Okada,N. and Nishimura,S. (1979) Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase, from *Escherichia coli*. *J. Biol. Chem.*, **254**, 3061–3066.

37. Bai,Y., Fox,D.T., Lacy,J.A., Van Lanen,S.G. and Iwata-Reuyl,D. (2000) Hypermodification of tRNA in *Thermophilic archaea*. Cloning, overexpression, and characterization of tRNA-guanine transglycosylase from *Methanococcus jannaschii*. *J. Biol. Chem.*, **275**, 28731–28738.

38. Ishitani,R., Nureki,O., Fukai,S., Kijimoto,T., Nameki,N., Watanabe,M., Kondo,H., Sekine,M., Okada,N., Nishimura,S. *et al.* (2002) Crystal structure of archaeosine tRNA-guanine transglycosylase. *J. Mol. Biol.*, **318**, 665–677.

39. Romier,C., Reuter,K., Suck,D. and Ficner,R. (1996) Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. *EMBO J.*, **15**, 2850–2857.

40. Johannsson,S., Neumann,P. and Ficner,R. (2018) Crystal structure of the human tRNA guanine transglycosylase catalytic subunit QTRT1. *Biomolecules*, **8**, 81.

41. Behrens,C., Biela,I., Petiot-Becard,S., Botzanowski,T., Cianferani,S., Sager,C.P., Klebe,G., Heine,A. and Reuter,K. (2018) Homodimer architecture of QTRT2, the noncatalytic subunit of the eukaryotic tRNA-guanine transglycosylase. *Biochemistry*, **57**, 3953–3965.

42. Slany,R.K. and Muller,S.O. (1995) tRNA-guanine transglycosylase from bovine liver. Purification of the enzyme to homogeneity and biochemical characterization. *Eur. J. Biochem.*, **230**, 221–228.

43. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.

44. Holm,L. (2020) Using Dali for protein structure comparison. *Methods Mol. Biol.*, **2112**, 29–42.

45. Brenk,R., Stubbs,M.T., Heine,A., Reuter,K. and Klebe,G. (2003) Flexible adaptations in the structure of the tRNA-modifying enzyme tRNA-guanine transglycosylase and their implications for substrate selectivity, reaction mechanism and structure-based drug design. *Chembiochem*, **4**, 1066–1077.

46. Xie,W., Liu,X. and Huang,R.H. (2003) Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nat. Struct. Biol.*, **10**, 781–788.

47. Chen,Y.C., Li,C.L., Hsiao,Y.Y., Duh,Y. and Yuan,H.S. (2014) Structure and function of TatD exonuclease in DNA repair. *Nucleic Acids Res.*, **42**, 10776–10785.

48. Ishitani,R., Nureki,O., Nameki,N., Okada,N., Nishimura,S. and Yokoyama,S. (2003) Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell*, **113**, 383–394.

49. Sievers,K., Welp,L., Urlaub,H. and Ficner,R. (2021) Structural and functional insights into human tRNA guanine transgylcosylase. *RNA Biol.*, **18**, 382–396.

50. Soh,Y.M., Davidson,I.F., Zamuner,S., Basquin,J., Bock,F.P., Taschner,M., Veening,J.W., De Los Rios,P., Peters,J.M. and Gruber,S. (2019) Self-organization of parS centromeres by the ParB CTP hydrolase. *Science*, **366**, 1129–1133.

51. Xia,S., Chen,J., Liu,L., Wei,Y., Deng,Z., Wang,L. and Chen,S. (2019) Tight control of genomic phosphorothioate modification by the ATP-modulated autoregulation and reusability of DndB. *Mol. Microbiol.*, **111**, 938–950.

52. Xiong,X., Wu,G., Wei,Y., Liu,L., Zhang,Y., Su,R., Jiang,X., Li,M., Gao,H., Tian,X. *et al.* (2020) SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities. *Nat. Microbiol.*, **5**, 917–928.

53. Basu,M.K. and Koonin,E.V. (2005) Evolution of eukaryotic cysteine sulfinic acid reductase, sulfiredoxin (Srx), from bacterial chromosome partitioning protein ParB. *Cell Cycle*, **4**, 947–952.

54. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

55. Jalal,A.S.B., Tran,N.T., Wu,L.J., Ramakrishnan,K., Rejzek,M., Gobbato,G., Stevenson,C.E.M., Lawson,D.M., Errington,J. and Le,T.B.K. (2021) CTP regulates membrane-binding activity of the nucleoid occlusion protein Noc. *Mol. Cell*, **81**, 3623–3636.

56. Schumacher,M.A., Tonthat,N.K., Lee,J., Rodriguez-Castañeda,F.A., Chinnam,N.B., Kalliomaa-Sanford,A.K., Ng,I.W., Barge,M.T., Shaw,P.L. and Barillà,D. (2015) Structures of archaeal DNA segregation machinery reveal bacterial and eukaryotic linkages. *Science*, **349**, 1120–1124.

57. Liu,M., Wang,J., Li,X., Sylvanno,M.J., Li,M., Zhang,M. and Wang,M. (2019) The crystal structure of sulfiredoxin from *Arabidopsis thaliana* revealed a more robust antioxidant mechanism in plants. *Biochem. Biophys. Res. Commun.*, **520**, 347–352.

58. Watanabe,M., Matsuo,M., Tanaka,S., Akimoto,H., Asahi,S., Nishimura,S., Katz,J.R., Hashizume,T., Crain,P.F., McCloskey,J.A. *et al.* (1997) Biosynthesis of archaeosine, a novel derivative of 7-deazaguanosine specific to archaeal tRNA, proceeds via a pathway involving base replacement of the tRNA polynucleotide chain. *J. Biol. Chem.*, **272**, 20146–20151.

59. Romier,C., Meyer,J.E. and Suck,D. (1997) Slight sequence variations of a common fold explain the substrate specificities of tRNA-guanine transglycosylases from the three kingdoms. *FEBS Lett.*, **416**, 93–98.

60. Hong,S. and Cheng,X. (2016) DNA base flipping: a general mechanism for writing, reading, and erasing DNA modifications. *Adv. Exp. Med. Biol.*, **945**, 321–341.