Structural basis for the selective incorporation of an artificial nucleotide opposite a DNA adduct by a DNA polymerase†

K. Betz, a A. Nilforoushan, b L. A. Wyss, b K. Diederichs, c S. J. Sturla d * b and A. Marx a * c

The possibility to sequence cytotoxic O6-alkylG DNA adducts would greatly benefit research. Recently we reported a benzimidazole-derived nucleotide that is selectively incorporated opposite the damaged site by a mutated DNA polymerase. Here we provide the structural basis for this reaction which may spur future developments in DNA damage sequencing.

There have been major recent advances addressing the distribution of epigenetically encoded DNA methylation in the genome due to the application of bisulfite sequencing,1 however, knowledge concerning the locations of chemically induced base alkylation lags far behind due to a lack of strategies for mapping it. An envisioned basis for DNA alkylation sequencing involves the use of artificial nucleotides incorporated opposite alkylated guanine by a DNA polymerase. Chemical alkylation of DNA by carcinogens or chemotherapeutics may occur, for example, at the O6 position of guanine resulting in O6-alkylG DNA adducts that induce toxicity and promote the misincorporation of T, leading to G-to-A mutations prevalent in human cancers.2–5

Recent DNA damage sequencing advances include versions of single molecule real time sequencing,6 and the use of DNA repair enzymes for excising oxidized or cross-linked bases7–10 and mapping these locations in the genome.11,12 For DNA polymerase-mediated selective amplification from O6-alkylG adducts,13 KlenTaq DNA polymerase (large fragment of Thermus aquaticus DNA polymerase I) has been shown to process an artificial benzimidazole-derived nucleoside triphosphate (Fig. 1, BenziTP) opposite O6-methylguanine (O6-MeG) more efficiently than opposite guanine.13,14 A mutated KlenTaq DNA polymerase...

† Electronic supplementary information (ESI) available: Experimental details, data collection statistics. See DOI: 10.1039/c7cc07173f
still lower compared to the natural incorporation of dCMP opposite dG (260 fold less efficient for M747K) The difference relies mainly on the higher $K_M$ of BenziTP while $k_{cat}$ is almost the same. Structural data would be highly beneficial for optimizing nucleotide design towards higher incorporation efficiency which is crucial for furthering the sequencing approach. Here, we provide structural insights into how the O6-MeG:Benzi pair is processed by KlenTaq M747K.

To investigate the preferential incorporation of BenziTP opposite O6-MeG on a structural level we crystallized KlenTaq M747K in a ternary complex with the substrate BenziTP paired opposite a templating O6-MeG in the active site (structure is termed M747K_{O6-MeG,Benzi}). For comparison we also solved the structure of KlenTaq M747K complexed with a natural dG:dCTP pair which is termed M747K_{G:C}. Analysis of the structures revealed that BenziTP forms a Watson–Crick-like base pair together with O6-MeG with some variations in positioning of the adduct methyl group and the substrate sugar pucker.

The structure M747K_{O6-MeG,Benzi} was obtained by crystallizing KlenTaq M747K in a binary complex with a primer/template duplex carrying O6-MeG at the templating position. The crystals were soaked with BenziTP together with up to 100 mM MnCl$_2$. Soaking without manganese ions did not result in a structure of a closed ternary complex. It was described that presence of Mn$^{2+}$ can lower the $K_M$ of nucleotides binding to a DNA polymerase 18 and adopts an anti conformation.14,22 It pairs with the O6-MeG templating nucleotide via two hydrogen bonds: one between the N1 of O6-MeG and the –NH donor on BenziTP (3.2 Å) and the second between the –NH$_2$ donor on O6-MeG and the carbonyl group of BenziTP (2.9 Å, Fig. 2). After superimposition of the nascent base-pairs in M747K_{O6-MeG,Benzi} and M747K_{G:C} based on the templating nucleotides the slightly larger propeller twist of the O6-MeG–Benzi pair becomes apparent (Fig. 2C). The simulated annealing omit map of both modified nucleotides is shown in Fig. 2A. For O6-MeG the density did not unambiguously define if the methyl group points towards or away from the BenziTP. We modelled both possibilities at the same time and refined their occupancies. Thereby the proximal O6-MeG refined to an occupancy of only 0.3, whereby the distal O6-MeG had an occupancy of 0.7. Therefore we decided to model the distal conformation keeping in mind that the alternative proximal conformation may also be populated in the crystal.

The M747K_{G:C} structure was solved by cocrystallization of KlenTaq M747K with a natural primer/template complex and dCTP in the presence of magnesium ions. It is very similar to the KlenTaq WT structure with the same primer/template and substrate complexed (PDB ID. 3RTV; 21 rmsd for C atoms: 0.224, Fig. S1A and B, ESI†). Just as 3RTV, the structure M747K_{G:C} is in a closed conformation and coordination of the two Mg$^{2+}$ ions and stabilization of the triphosphate in the active site is the same. (Fig. S1B, ESI†) The only significant difference is the Met747Lys mutation, which points towards the negatively charged template strand (Fig. S1C and D, ESI†). The electrostatic map in Fig. S1C and D (ESI†) visualizes the gained positive charge that may help to stabilize the template strand and promote selective lesion-bypass synthesis as was described before.16

The structure M747K_{O6-MeG,Benzi} also shows a closed enzyme conformation and superposes well with M747K_{G:C} (rmsd for C atoms: 0.182). The substrate BenziTP is bound in the active site and adopts an anti conformation.14,22 The structure M747K_{O6-MeG,Benzi} was obtained by crystallizing KlenTaq M747K in a ternary complex with a primer/template duplex carrying O6-MeG at the templating position. The crystals were soaked with BenziTP together with up to 100 mM MnCl$_2$. Soaking without manganese ions did not result in a structure of a closed ternary complex. It was described that presence of Mn$^{2+}$ can lower the $K_M$ of nucleotides binding to a DNA polymerase 18 and adopts an anti conformation.14,22 It pairs with the O6-MeG templating nucleotide via two hydrogen bonds: one between the N1 of O6-MeG and the –NH donor on BenziTP (3.2 Å) and the second between the –NH$_2$ donor on O6-MeG and the carbonyl group of BenziTP (2.9 Å, Fig. 2). After superimposition of the nascent base-pairs in M747K_{O6-MeG,Benzi} and M747K_{G:C} based on the templating nucleotides the slightly larger propeller twist of the O6-MeG–Benzi pair becomes apparent (Fig. 2C). The simulated annealing omit map of both modified nucleotides is shown in Fig. 2A. For O6-MeG the density did not unambiguously define if the methyl group points towards or away from the BenziTP. We modelled both possibilities at the same time and refined their occupancies. Thereby the proximal O6-MeG refined to an occupancy of only 0.3, whereby the distal O6-MeG had an occupancy of 0.7. Therefore we decided to model the distal conformation keeping in mind that the alternative proximal conformation may also be populated in the crystal.

The M747K_{G:C} structure was solved by cocrystallization of KlenTaq M747K with a natural primer/template complex and dCTP in the presence of magnesium ions. It is very similar to the KlenTaq WT structure with the same primer/template and substrate complexed (PDB ID. 3RTV; 21 rmsd for C atoms: 0.224, Fig. S1A and B, ESI†). Just as 3RTV, the structure M747K_{G:C} is in a closed conformation and coordination of the two Mg$^{2+}$ ions and stabilization of the triphosphate in the active site is the same. (Fig. S1B, ESI†) The only significant difference is the Met747Lys mutation, which points towards the negatively charged template strand (Fig. S1C and D, ESI†). The electrostatic map in Fig. S1C and D (ESI†) visualizes the gained positive charge that may help to stabilize the template strand and promote selective lesion-bypass synthesis as was described before.16

The structure M747K_{O6-MeG,Benzi} also shows a closed enzyme conformation and superposes well with M747K_{G:C} (rmsd for C atoms: 0.182). The substrate BenziTP is bound in the active site and adopts an anti conformation.14,22 The structure M747K_{O6-MeG,Benzi} was obtained by crystallizing KlenTaq M747K in a ternary complex with a primer/template duplex carrying O6-MeG at the templating position. The crystals were soaked with BenziTP together with up to 100 mM MnCl$_2$. Soaking without manganese ions did not result in a structure of a closed ternary complex. It was described that presence of Mn$^{2+}$ can lower the $K_M$ of nucleotides binding to a DNA polymerase 18 and adopts an anti conformation.14,22 It pairs with the O6-MeG templating nucleotide via two hydrogen bonds: one between the N1 of O6-MeG and the –NH donor on BenziTP (3.2 Å) and the second between the –NH$_2$ donor on O6-MeG and the carbonyl group of BenziTP (2.9 Å, Fig. 2). After superimposition of the nascent base-pairs in M747K_{O6-MeG,Benzi} and M747K_{G:C} based on the templating nucleotides the slightly larger propeller twist of the O6-MeG–Benzi pair becomes apparent (Fig. 2C). The simulated annealing omit map of both modified nucleotides is shown in Fig. 2A. For O6-MeG the density did not unambiguously define if the methyl group points towards or away from the BenziTP. We modelled both possibilities at the same time and refined their occupancies. Thereby the proximal O6-MeG refined to an occupancy of only 0.3, whereby the distal O6-MeG had an occupancy of 0.7. Therefore we decided to model the distal conformation keeping in mind that the alternative proximal conformation may also be populated in the crystal.
Depending on the position of the methyl group, the BenziTP is slightly shifted. As the distal methyl group comes closer to the phenyl ring of BenziTP (closest distance 3.4 Å) BenziTP is shifted further away from the pairing partner resulting in a little enlarged base-pair width compared to the natural case (C1′-C1′-distance 11.0 Å vs. 10.6 Å, Fig. 2B).

The electron density profiles for the ribose moiety suggest that several sugar conformations may be present at the same time in the crystal. We placed two different conformations (C2′-endo and C3′-endo), performed one round of occupancy refinement for the two BenziTP ligands and found a 1/3 to 2/3 occupancy ratio for C3′-endo versus C2′-endo. For natural substrates the common sugar mostly observed is C3′-endo in the KlenTaq DNA polymerase active site.23 The triphosphate moiety is very well defined in the electron density and interactions with polar residues of the O-helix allow the same stabilization as in the natural case (Fig. 3A and B). Additional stabilization may be possible with Asn750 and Gln754 and the carbonyl group of Benzi via a water molecule. The distance to the water molecule, however, is larger than in the natural case (3.4 Å versus 2.8 Å).

Characteristic for an active complex prior to catalysis is that the finger domain is in a closed conformation and the triphosphate and amino acid side chains within the palm domain coordinate two divalent metal ions. The residues involved in coordination two divalent metal ions. The residues involved in phosphate and amino acid side chains within the palm domain are the same as with a natural base-pair but the coordination shows a slightly different geometry (Fig. 3). The overlay in Fig. 3C shows the metal coordination in M747K_O6MeG:Benzi and M747K_G:C. Metal ion B which is coordinated (Fig. 3). The overlay in Fig. 3C shows the metal coordination in M747K_O6MeG:Benzi and M747K_G:C. Metal ion B which is coordinated (Fig. S3C, black arrow, ESI†). Metal ion A is postulated to stabilize the pentacovalent negative charge of the leaving pyrophosphate.24 The second ion would not clash with the primer terminus. The Arg660 side chain is rotated compared to its position in the natural dGdCTP case and loses its interaction with the primer strand (Fig. S3, ESI†). Instead, Arg587 is rotated towards the primer 3′-end, taking over the lost interaction. The size of the BenziTP nucleobase however does not explain, why Arg660 is shifted as it would not clash with the side chain in a position observed in M747K_G:C (Fig. S3C, black arrow, ESI†). Concerning the position of Arg660 and Arg387 M747K_{O6MeG:Benzi} seems to be similar to a KlenTaq complex with a bound dGTP (PDB ID:1QSS,20 Fig. S3D, ESI†) it shows only four coordination partners. In both complexes an additional coordination of metal ion A would be mediated by the 3′-OH group of the primer end which is not present in our structures since we are using a dideoxy-terminated primer as a common strategy to capture the ternary complex in crystallization. The assigned role of metal ion A is to lower the pKa of that terminal OH group for the nucleophilic attack on the alpha phosphate.25,26 Why this different ion binding (along with the different sugar conformation) occurs and if it still displays an active complex from which the catalytic reaction would occur or if we trapped the structure in a pre-catalytic state from which catalysis can take place only after an additional conformational change is not clear.

An additional difference to the natural complex is the fact that the enzyme is not as tightly closed, which is indicated by higher flexibility of the whole finger domain (high B-factors and weak electron density in that region, Fig. S2A, ESI†). This might be due to the larger size of the nucleobase-pair where the methyl group attached to dG and the phenyl ring of Benzi both extend beyond the consensus pocket of natural base-pairs (Fig. S2B, ESI†). Similar flexibility has already been observed in KlenTaq DNA polymerase complexes with artificial base-pairs and non cognate substrates.21,28,29 An anticipated outcome may be reduced capacity to stabilize the substrate/enzyme complex, thus allowing some deviation from the optimal geometry for catalysis. Finally this conformational heterogeneity might account for the observed impaired enzymatic efficiency of the DNA polymerase when processing modified substrates in comparison to the cognate substrates.

Another difference in M747K_{O6MeG:Benzi} compared to M747K_G:C is the interaction of Arg600 with the 3′-primer terminus. The Arg660 side chain is rotated compared to its position in the natural dGdCTP case and loses its interaction with the primer strand (Fig. S3, ESI†). Instead, Arg587 is rotated towards the primer 3′-end, taking over the lost interaction. The size of the BenziTP nucleobase however does not explain, why Arg660 is shifted as it would not clash with the side chain in a position observed in M747K_G:C (Fig. S3C, black arrow, ESI†). Concerning the position of Arg660 and Arg387 M747K_{O6MeG:Benzi} seems to be similar to a KlenTaq complex with a bound dGTP (PDB ID:1QSS,20 Fig. S3D, ESI†)
as this is the only nucleobase that was found to cause a different positioning of these residues in the KlenTaq DNA polymerase structures elucidated by Waksman et al.30,31. Based on the position of Arg660 in 1QSS the different shape of BenziTP would indeed cause a clash with Arg660 (Fig. S3D, ESI†) and therefore explain why it is rotated away.

In summary we could show why the synthetic adduct-directed nucleotide BenziTP is specifically incorporated opposite the alkylation-induced DNA adduct O6-MeG. The structural composition of both components allow formation of a Watson–Crick-like base-pair mediated by two hydrogen bonds with a size and geometry similar to the cognate base-pairs to be readily accepted by KlenTaq DNA polymerase and its M747K mutant. Neither should the other four natural nucleotides be able to form an equally similar base-pair with O6-MeG nor should Benzi pair in the same way with natural G. The study shows that KlenTaq DNA polymerase forms an active site that is able to accommodate the unnatural base-pair. While the mutated amino acid seems to be not directly involved in selecting the nucleotide opposite O6-MeG e.g., by forming hydrogen bonds to BenziTP, it might contribute nonetheless to stabilizing the conformations at the active site that favour selective incorporation by increasing the positive electrostatic potential at the negatively charged template phosphate backbone site (Fig. S1C and D, ESI†). The fact that well-diffracting crystals could only be obtained in the presence of manganese ions, and the higher flexibility of the substrates and the finger domain in the complex might be indicative for the lower incorporation efficiency of BenziMP opposite O6-MeG compared to a canonical base pair. The structural data provided here might guide optimization of the processes e.g., by altering the chemical structure of the triphosphate or further mutating the enzyme.

We thank the beamline staff of the Swiss Light Source at the Paul Scherrer Institute for access and help at the beamline. We also acknowledge the European Research Council (Grant 260341) and the Swiss National Science Foundation (Grant 156280) (S. J. S.) for funding.

Conflicts of interest

There are no conflicts to declare.

Notes and references

† Atomic coordinates and structure factors have been deposited in the Protein Data Bank with PDB codes 5OT7 for M747K_Gc: and 5OXJ for M747K_MeG:Benzi+.