Published in final edited form as:
J Am Chem Soc. 2013 December 11; 135(49): 18637–18643. doi:10.1021/ja409609j.

Structural Insights into DNA Replication Without Hydrogen-Bonds

Karin Betz[†,§], **Denis A. Malyshev**[‡,§], **Thomas Lavergne**[‡], **Wolfram Welte**[†], **Kay Diederichs**[†], **Floyd E. Romesberg**[‡,*], and **Andreas Marx**[†,*]

[†]Departments of Chemistry and Biology, Konstanz Research School Chemical Biology, Universität Konstanz, Universitätsstrasse 10, D-78464 Konstanz, Germany

[‡]Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California, 92037

Abstract

The genetic alphabet is comprised of two base pairs, and the development of a third, unnatural base pair would increase the genetic and chemical potential of DNA. d**5SICS**-d**NaM** is one of the most efficiently replicated unnatural base pairs identified to date, but its pairing is mediated by only hydrophobic and packing forces, and in free duplex DNA it forms a cross-strand intercalated structure that makes its efficient replication difficult to understand. Recent studies of the KlenTaq polymerase revealed that the insertion of d**5SICS**TP opposite d**NaM** proceeds via a mutually induced-fit mechanism, where the presence of the triphosphate induces the polymerase to form the catalytically competent closed structure, which in turn induces the pairing nucleotides of the developing unnatural base pair to adopt a planar Watson-Crick-like structure. To understand the remaining steps of replication, we now report the characterization of the pre-chemistry complexes corresponding to the insertion of d**NaM**TP opposite d**5SICS**, as well as multiple post-chemistry complexes in which the already formed unnatural base pair is positioned at the post-insertion site. Unlike with the insertion of d5**SICS**TP opposite d**NaM**, addition of d**NaM**TP does not fully induce the formation of the catalytically competent closed state. The data also reveal that once synthesized and translocated to the post-insertion position, the unnatural nucleobases again intercalate. Two modes of intercalation are observed, depending on the nature of the flanking nucleotides, and are each stabilized by different interactions with the polymerase, and each appear to reduce the affinity with which the next correct triphosphate binds. Thus, continued primer extension is limited by de-intercalation and rearrangements with the polymerase active site that are required to populate the catalytically active, triphosphate bound conformation.

INTRODUCTION

Successful development of a functional unnatural base pair that is orthogonally replicated in DNA is the first step toward creating a semi-synthetic organism with increased potential for information storage and retrieval, and would also expand the utility of nucleic acids for biological and biotechnological applications.[1–10] While a variety of unnatural base pair candidates have been reported,[11–15] only three have been shown to be efficiently replicated,[16–18] and only the pair formed between d**5SICS** and d**NaM** (d**5SICS**-d**NaM**; Fig.

[*]*Corresponding Author*. floyd@scripps.edu, andreas.marx@uni-konstanz.de.
[§]These authors contributed equally.

ASSOCIATED CONTENT

Supporting Information. Crystallization conditions and oligonucleotide sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

1) has been shown to be PCR amplified without sequence-bias[19] and efficiently transcribed in both directions.[20,21]

The efficient replication of d**5SICS**-d**NaM** is particularly interesting because it proceeds in the absence of complementary hydrogen-bonds (H-bonds) that underlie Watson-Crick-like pairing, and indeed, it forms an intercalated structure in duplex DNA.[22,23] This mode of pairing maximizes packing interactions, and is likely general for nucleotides with predominantly hydrophobic nucleobases,[24,25] but the resulting structure is reminiscent of a mispair between natural nucleotides[26–31] and is thus difficult to reconcile with efficient polymerase recognition. To investigate the structural basis for the efficient replication of DNA containing d**5SICS**-d**NaM**, we recently solved the crystal structure of KlenTaq DNA polymerase, the large fragment of the type I DNA polymerase from *Thermus aquaticus*, complexed with a templating d**NaM**, and with or without bound d**5SICS**TP ($KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$ and $KTQ_{d\mathbf{NaM}}$, respectively).[22] The structures of these pre-chemistry d**5SICS**TP incorporation complexes revealed that the pairing of d**5SICS**TP with d**NaM** drives the open-to-closed conformational change characteristic of a natural base pair[32–34] and interestingly, once in the closed environment, the pairing unnatural nucleotides adopt a planar, Watson-Crick-like geometry.[22] Thus we demonstrated that not only is the polymerase able to select for pairs that form a correct Watson-Crick structure, but at least with hydrophobic analogs, it is able to enforce the correct structure. This mutually induced fit mechanism highlights what might be a fundamental advantage of using hydrophobicity and packing forces to mediate replication, as they are sufficiently strong to mediate pairing, but also sufficiently plastic to adapt to the structure required by the polymerase. However, the insertion of d**5SICS**TP opposite d**NaM** is a particularly efficient step of replication,[16] and the mechanism by which d**NaM**TP is inserted opposite d**5SICS**, and the mechanism by which the primer containing either unnatural nucleotide is further elongated, which actually limits replication, remained unclear.

Here, to fully characterize the mechanism of unnatural base pair replication, we report the crystal structures of the pre-chemistry incorporation complexes leading to the insertion of d**NaM**TP opposite d**5SICS**, the binary complex of KlenTaq with a DNA template containing d**5SICS** at the templating position ($KTQ_{d\mathbf{5SICS}}$) and the corresponding ternary complex with d**NaM**TP bound ($KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$). We also report the structure of four post-incorporation complexes, the binary complex of KlenTaq and either a primer terminating with d**5SICS** paired opposite d**NaM** in a template ($KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}}$) in three different sequence contexts, or a primer terminating with d**NaM** paired opposite d**5SICS** ($KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}}$). Along with our previously reported structures, these structures provide key insights into the replication of the unnatural base pair and elucidate a mechanism that is based on a balance of intercalation and de-intercalation and structural rearrangements of the polymerase active site.

RESULTS

Pre-chemistry dNaMTP incorporation complexes

KlenTaq was first crystallized bound to a DNA primer/template with d**5SICS** at the templating position (position n). In the resulting complex ($KTQ_{d\mathbf{5SICS}}$; Fig. 2A), the d**5SICS** nucleoside adopts an extra-helical position that is similar to that observed for a natural dG in KTQ_{dG} (PDB ID 3SZ2).[22] However, compared to the previously described binary structures $KTQ_{d\mathbf{NaM}}$ (PDB ID 3SYZ) and KTQ_{dT} (PDB ID 3SV4) the single-stranded DNA of the template adopts a different arrangement (Fig. S1). It is likely that the single-stranded portion of the template is flexible in the binary structures, and that the

differences are not functionally relevant. Structural heterogeneity of the template is also implied by the absence of well defined electron density for the d**5SICS** nucleobase.

To determine whether the addition of d**NaM**TP to $KTQ_{d\mathbf{5SICS}}$ drives the same conformational change observed upon d**5SICS**TP binding to $KTQ_{d\mathbf{NaM}}$, we next determined the structure of $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ by soaking $KTQ_{d\mathbf{5SICS}}$ crystals with d**NaM**TP. The structure of $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ reveals that the unnatural triphosphate is bound to the O-helix (Fig. 2B), which is rotated and only partially closed. The position of the triphosphate appears to be stabilized by ionic interactions with Arg659 and Lys663 of the O-helix, as well as with His639 of the N-helix and Arg587 from the N-terminal end of the thumb domain K-helix. In addition, along with three water molecules, the triphosphate moiety coordinates a Mg^{2+} ion. The electron density for the sugar and the nucleobase moieties of d**NaM**TP are less well defined than that for the triphosphate moiety, suggesting an increased level of disorder and/or flexibility. The N- and O-helices of the fingers domain adopt a conformation intermediate between the open and closed states (Fig. 2C) (the root mean square deviation (rmsd) of residues 637 – 700 is 1.59 Å and 2.23 Å relative to $KTQ_{d\mathbf{5SICS}}$ and $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}TP}$ respectively). In addition, Tyr671 is slightly displaced from its open conformation position in the insertion site (Fig. S2), and the templating unnatural nucleobase moves from its extrahelical position toward the insertion site, again representative of a state intermediate between the open and closed conformations.

Post-chemistry extension complexes

We next sought to investigate the structures of the post-chemistry complexes, with the unnatural base pair positioned in the post-insertion site, where it is poised for continued primer elongation (*i.e.* extension of the unnatural base pair). We first characterized the structure of $KTQ_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}}$ with d**5SICS** at the primer terminus paired opposite d**NaM** at the n−1 position, with three different primer/templates (E1–E3, Table 1). In each binary complex, the polymerase adopts the expected open conformation, similar to that observed in $KTQ_{d\mathbf{NaM}}$, $KTQ_{d\mathbf{5SICS}}$, or KlenTaq bound to a fully natural primer/template.[22] However, the presence of the unnatural base pair has a significant effect on the structure of the primer/template. In the structure of $KTQ(E1)_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}}$, the template d**NaM** cross-strand intercalates into the primer strand between d**5SICS** and the 5’ dC (dC_{n-2}) (Fig. 3 and Fig. S3). To accommodate this intercalation, relative to their positions observed with natural substrates, the C1’ of the primer unnatural nucleotide moves 4.7 Å towards the template and the C1’ of the unnatural nucleotide in the template shifts 4.4 Å in the direction of translocation (Fig. 4A and C). The extent of intercalation is evident by the sugar C1’-C1’ distance of 8.5 Å, compared to the ~10.5 Å distance that is typical for a natural pair in the post-insertion site.[22,33] This degree of intercalation is even greater than in the free duplex, where the C1’-C1’ distance is 9.1 Å,[22] likely reflecting a decreased level of structural restraints when the unnatural base pair is positioned at the end of a duplex, as opposed to the middle. Intercalation also positions the templating nucleobase proximal to the primer terminus, and the N1 and C2 amino group of dG_n form H-bonds with the phosphate backbone of the primer (Fig. S4). Although perturbations are apparent with the n−2 and n−3 template nucleotides, they are smaller, and the remainder of the template is unperturbed, relative to its fully natural counterpart.[22] In contrast, at least minor distortions are apparent throughout the primer.

Examination of the polymerase reveals that, relative to its open form, the presence of the unnatural base pair at the post-insertion site induces the thumb domain to rotate, with helices H, I and K, which interact with the 3’ side of the primer, moving closer to the active site, and helices H1 and H2, which interact with the 5’ side of the primer, moving away (Fig. 4B). The position of the fingers domain is less perturbed. The intercalated state appears to be

accommodated by a network of protein residues of the finger domain, including Asn750, Tyr671, Gln754 and Glu615, which pack on the free 3' face of the d**5SICS** nucleobase (Fig. S3), and the primer terminus appears further stabilized by H-bonds between the 3' OH and phosphate backbone with His784 and Arg587, respectively. While the latter interaction is also observed with a natural substrate,[22] the former is not (by analogy to the homologous large fragment of polymerase I from *Bacillus stearothermophilus* (Bacillus fragment, BF), the 3' OH of a fully natural substrate forms an H-bond with Asp785[35]). Furthermore, the sulfur atom of d**5SICS** engages in a water mediated H-bond with Thr571. The phosphate of the template d**NaM** interacts with Arg746 (Fig. S3), and an analogous interaction is observed with a fully natural substrate.

In the structure of KTQ(E2)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$, the unnatural base pair again forms via intercalation, but in this case, by the primer d**5SICS** nucleobase inserting between the template d**NaM** and its 3' dG (dG$_{n-2}$) (Fig. 3). The extent of intercalation appears somewhat less than in KTQ(E1)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$, with a C1'-C1' distance of 9.2 Å (compared to 8.5 Å). The overall structure of the primer/template is similar to that observed with KTQ(E1)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$, but despite the decreased intercalation, it is somewhat more distorted, with the C1' of the primer unnatural nucleotide moving 5.7 Å towards the template and the C1' of the unnatural nucleotide in the template shifting 5.2 Å in the direction of translocation, relative to their positions observed with natural substrates (Fig. 4A and D). In addition, d**NaM** shields the templating nucleobase from interacting with the primer strand, and possibly as a result, neither the downstream nucleotides nor Arg587 are well resolved in the KTQ(E2)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ structure. While the overall structure of the polymerase is similar in KTQ(E2)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ and KTQ(E1)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$, there are significant differences in the interactions with the unnatural base pair. In KTQ(E2)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$, the O-helix residue Tyr671 stacks on the template d**NaM** nucleobase, while Gln754 and Glu615 form H-bonds with the sulfur of d**5SICS** and the primer 3'OH, respectively (Fig. S3). In contrast to KTQ(E1)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$, no specific interactions are observed between KlenTaq and the phosphates of the unnatural nucleotides in either the primer or template strands.

To examine the effect of strand context, we solved the structure of the KTQ(E2)$_{\text{d}\mathbf{5SICS}\text{-d}\mathbf{NaM}}$ binary complex, with d**NaM** at the primer terminus paired opposite d**5SICS** in the template (Table 1). Again, the unnatural nucleobases pair in an intercalated fashion, and in this case in a manner similar to KTQ(E2)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ (Fig. 3), although the C1'-C1' distance of 9.8 Å is somewhat longer. The protein-DNA interactions, including those involving the unnatural base pair are also conserved in the two structures, including the H-bond between the Gln754 H-bond donor and the H-bond acceptor of the unnatural nucleotide at the primer terminus, which with KTQ(E2)$_{\text{d}\mathbf{5SICS}\text{-d}\mathbf{NaM}}$ is the methoxy group of d**NaM** and with KTQ(E2)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ is the sulfur atom of d**5SICS** (Fig. S3).

To demonstrate that the length of the single-stranded portion of the template does not affect the structure of the unnatural base pair at the primer terminus, we examined primer/template E3, which like E1 has the template sequence 3'-dG**NaM**G, but like E2, it has a 3-nt overhang (Table 1). With KTQ(E3)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$, the unnatural base pair is again intercalated, and in a manner similar to that observed with KTQ(E1)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ (Fig. S3). Furthermore, while the last single-stranded residue of KTQ(E3)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ (dT$_{n+2}$) is not resolved, the overall structures of KTQ(E3)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ and KTQ(E1)$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ are almost identical with an rmsd of 0.23 Å, and with the same residues interacting with the intercalated unnatural base pair.

Multiple attempts were made to solve the structures of the ternary complexes of either KTQ$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}}$ or KTQ$_{\text{d}\mathbf{5SICS}\text{-d}\mathbf{NaM}}$ and non-hydrolysable variants of the next correct natural triphosphate, dCTP or dGTP (*e.g.* NHdCTP or NHdGTP). In no case were we able to

detect electron density associated with the nucleoside triphosphate. Thus we conclude that the low affinity of the natural triphosphates for the primer terminus containing an intercalated unnatural base pair precludes their crystallization in a ternary complex.

DISCUSSION

Replication of natural DNA is mediated by the H-bonding and shape complementarity of the pairing nucleobases.[36–39] However, d**5SICS** and d**NaM** cannot form H-bonds and have shapes that are very different from the natural purines and pyrimidines. Nonetheless, during PCR amplification, d**5SICS**-d**NaM** is functionally equivalent to a natural base pair.[19] Our earliest efforts to understand this efficient replication focused predominantly on structure-activity relationships derived from kinetics assays. Later, we focused on free duplex DNA and showed that the pair forms via cross-strand intercalation,[22,23] raising more questions than we answered. This situation was at least partially clarified with our previous characterization of the pre-chemistry complexes leading to the insertion of d**5SICS**TP opposite d**NaM**. The structures of KTQ$_{\text{d}\mathbf{NaM}}$ and KTQ$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}\text{TP}}$ elucidated a mutual induced fit mechanism, wherein pairing of d**5SICS**-d**NaM** drives the open-to-closed conformational transition of the polymerase, and the closed conformation of the polymerase induces d**5SICS**-d**NaM** to adopt a Watson-Crick like structure. With these results, our attention turned to the mechanisms underlying the remaining steps of replication, including the insertion of d**NaM**TP opposite d**5SICS**, and the subsequent continued primer elongation after incorporation of either unnatural triphosphate.

Unlike with the addition of d**5SICS**TP to KTQ$_{\text{d}\mathbf{NaM}}$,[22] the addition of d**NaM**TP to KTQ$_{\text{d}\mathbf{5SICS}}$ did not induce the canonical open-to-closed conformational change observed during the synthesis of a natural base pair, but rather resulted in the formation of a structure wherein the DNA polymerase finger domain remains in a partially open conformation and the d**NaM**TP is bound via its triphosphate moiety to the O-helix. A similar conformation has been described by Beese *et al.* for BF polymerase with a dG-dTTP or a dG-ddTTP mismatch.[40] In this structure, the polymerase remains in an partially open conformation, referred to as “ajar,” and the templating nucleotide displaces the “gate keeping” residue Y714 (Y671 in KlenTaq) from the insertion site. It has been suggested that this ajar conformation allows the DNA polymerase to test for complementarity between the incoming and templating nucleotides before the enzyme transitions to the closed catalytically competent state. While Y761 remains in the templating position in KTQ$_{\text{d}\mathbf{5SICS}\text{-d}\mathbf{NaM}\text{TP}}$, both it and the nucleobase of d**5SICS** appear strained towards the same switch observed in the BF structure. A similar configuration has been observed with KlenTaq with an abasic site at the templating position.[41,42] Thus, the KTQ$_{\text{d}\mathbf{5SICS}\text{-d}\mathbf{NaM}\text{TP}}$ complex appears trapped in an intermediate state between the open binary complex and the ajar state observed with BF, similar to a partially closed state observed with the homologous *E. coli* polymerase I via biophysical studies.[43,44] Regardless, it is clear that incorporation of d**NaM**TP would require significant rearrangement of the polymerase to reach the catalytically competent closed state, while KTQ$_{\text{d}\mathbf{NaM}\text{-d}\mathbf{5SICS}\text{TP}}$ spontaneously forms the catalytically competent closed complex. This difference likely explains why the insertion of d**NaM**TP opposite d**5SICS** is often less efficient than the insertion of d**5SICS**TP opposite d**NaM**.[16]

In all four post-incorporation complexes characterized, the nucleobases pair in an intercalated manner, similar to their pairing in free duplex DNA.[23] However, two modes of intercalation are observed. With primer/templates E1 and E3, a common mode of intercalation is observed (Fig. 3), which demonstrates that the mode of pairing is unlikely to depend on the length of the single stranded template. In this mode of intercalation, the template d**NaM** inserts between its pairing d**5SICS** and the flanking dC$_{n-2}$, which allows for the template dG$_n$ to form stabilizing interactions with the primer terminus. In contrast, in

both complexes with primer/template E2, dC_n is unable to mediate such interactions, and the intercalated structure is formed by insertion of the primer d**5SICS** (KTQ(E2)$_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}}$) or d**NaM** (KTQ(E2)$_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}}$) between its pairing unnatural nucleobase and its flanking dG_{n-2} of the template, which likely optimizes packing interactions. Surprisingly, the mode of intercalation appears to depend most on sequence-specific interactions of the flanking nucleotides, with the specific packing interactions between the intercalating nucleobases being of secondary importance.

Interestingly, the polymerase appears to be able to provide unique stabilizing interactions to the two types of intercalated structures at the primer terminus. In the KTQ(E1)$_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}}$ and KTQ(E3)$_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}}$ structures, the observed intercalated state leaves one face of the primer d**5SICS** unpacked by a flanking nucleobase, and its position is stabilized by packing interactions with Asn750, Tyr671, Gln754 and Glu615, an ionic interaction between its phosphate and Arg587, a water-mediated H-bond between its sulfur and Thr571, and by an H-bond between its 3'OH and His784. The position of d**NaM** in the template is stabilized by an ionic interaction between its phosphate and Arg746. In the KTQ(E2)$_{d\mathbf{NaM}\text{-}d\mathbf{5SICS}}$ and KTQ(E2)$_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}}$ structures, the intercalated state adopted leaves one face of the unnatural nucleobase in the template (d**NaM** and d**5SICS**, respectively) unpacked by a flanking nucleobase, and its position is stabilized by packing interactions with O-helix residue Tyr671. In this case, the position of the primer terminus is stabilized by an H-bond between its 3'OH and Gln615 and by an H-bond between Gln754 and the H-bond acceptor *ortho* to the glycosidic linkage (methoxy in d**NaM** and sulfur in d**5SICS**). Unlike with E1 and E3, neither the template nor primer strand in either complex with E2 is stabilized via interactions with their backbone phosphates. The rather different interactions by which the two intercalated structures are accommodated reveals that the polymerase is surprisingly plastic. Regardless, both structures of the post-insertion complexes require de-intercalation and significant remodeling of the polymerase active site for incorporation of the next dNTP, likely explaining why structures with the next correct natural triphosphate could not be obtained and also why extension of the unnatural base pair is less efficient than incorporation of the unnatural triphosphate.

Based on this and previously reported structural data,[22,23,32,36,39,40] we propose the following mechanism of replication (Fig. 5). The unnatural triphosphate initially binds to the O-helix, producing a flexible complex that samples different conformations, and when sufficiently stabilizing hydrophobic and packing interactions are made, the open-to-closed transition is induced, which induces the unnatural base pair to adopt a planar, Watson-Crick-like pairing, and incorporation of the triphosphate onto the growing primer terminus. With d**5SICS**TP incorporation, the intermediate states are populated only transiently, and the closed complex may only be captured by preventing incorporation with a dideoxy primer terminus. However, with d**NaM**TP incorporation, the series of conformational changes are halted at an ajar-like state with the unnatural triphosphate remaining bound to the O-helix, either due to the stability of this complex or the instability of the corresponding closed complex, and further progress towards the incorporation of d**NaM**TP requires thermal fluctuations to populate the closed state. After incorporation of either d**5SICS**TP or d**NaM**TP, the polymerase returns to the open conformation and pyrophosphate is released.[45] However, in this state, the unnatural base pair adopts a cross-strand intercalated structure, similar to the structure it adopts in free duplex DNA, and continued primer elongation requires thermal fluctuations to both de-intercalate the unnatural base pair and reorganize the polymerase active site. Because extension consistently limits the replication of DNA containing the unnatural base pair, the model predicts that further optimization of d**5SICS**-d**NaM** may be possible by making changes to the nucleobase analogs that decrease the stability of the intercalated structures. Efforts to test this hypothesis are currently underway.

MATERIALS AND METHODS

Oligonucleotide synthesis

Natural oligonucleotides were purchased from IDT (San Diego, CA). d**NaM** and d**5SICS** phosphoramidites and nucleosides were obtained from Berry & Associates Inc. (Dexter, MI), and the latter were phosphorylated using Ludwig & Eckstein conditions[46] as described.[16] Oligonucleotides containing an unnatural nucleotide were prepared using standard automated DNA synthesis methodology with ultra-mild DNA synthesis phosphoramidites on CPG ultramild supports (1 μmol, Glen Research; Sterling, VA) and an ABI Expedite 8905 synthesizer. After automated synthesis, the DMT-ON oligonucleotide was first purified by Glen-Pak cartridge (Glen Research) and then by 8 M urea 20% PAGE, followed by Synergi Fusion-RP HPLC (Phenomenex, Torrance, CA) to single-band purity (>98%) using a linear gradient of 100 mM triethyl ammonium bicarbonate buffer (pH 7.5) and acetonitrile (5–30% over 35 min). The fractions containing purified oligonucleotides were collected and dried by vacuum centrifugation and their identity was confirmed by MALDI-ToF with THAP matrix.

Protein production, crystallization, and structure determination

KlenTaq was prepared using an *E. coli* codonoptimized gene encoding amino acids 293–832 of Taq polymerase (purchased from GeneArt, Germany) cloned into the vector pGdR11 and expressed in *E. coli* strain BL21 (DE3) in LB medium for 4 h after induction with 1 mM IPTG. The harvested cell pellet was resuspended in lysis buffer (50 mM Tris HCl pH 8.5, 10 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$ 0.1% TritonX-100, 0.1% hydroxypolyethoxydodecane, and 1 mM PMSF) and lysed by the addition of 0.5 mg/mL lysozyme and incubation for 1 h at 37 °C. After lysis, a heat denaturation was performed (20 min, 80 °C) and the cell debris was pelleted by ultracentrifugation (1 h, 35,000 × *g*). Bacterial DNA in the supernatant was removed by PEI-precipitation and centrifugation. The resulting supernatant was purified by anion exchange chromatography (Q Sepharose) in 20 mM Tris HCl pH 8.5, 1 mM EDTA, 1 mM β-mercaptoethanol, eluting with a NaCl gradient. Fractions containing KlenTaq were pooled, concentrated and further purified by size-exclusion chromatography (Superdex 75) in 20 mM Tris HCl pH 7.5, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.15 M NaCl.

Purified KlenTaq was stored at 4 °C. Primers and templates were annealed prior to addition of protein and triphosphates. KlenTaq was mixed with triphosphate and/or primer/template and incubated for 30 min at 30 °C. The mixture was then filtered and crystallization conditions were screened using the sitting drop vapor diffusion method at 18 °C. Hits were reproduced using either the sitting or hanging drop vapor diffusion method. Prior to measurement, crystals were flash frozen in liquid nitrogen either with or without cryo protection (see Supporting Information).

Data was collected at the beamline PXIII (XO6DA) and PXI (XO6SA) at the Swiss Light Source of the Paul Scherrer Institute in Villigen, Switzerland. Data reduction was performed with the XDS package.[47] Statistics of data collection and refinement for all structures are given in Table S1. Data was used in refinement up to a resolution with a CC1/2 value[48] of around 50%. To facilitate comparison with other deposited structures, we also report resolution values at which $1/\sigma = 2$ (see Table S1). Data reduction of the $KTQ_{d\mathbf{5SICS}}$ and $KTQ_{d\mathbf{5SICS}\text{-}d\mathbf{NaM}TP}$ data was done in space group $P3_121$ (for cell dimensions see Table S1), and the structures were solved by rigid-body refinement using a previously published KlenTaq structure (PDB: 3M8S[49]) as a model. All binary elongation complexes (KTQ(E1), KTQ(E2) and KTQ(E3)) crystallized in space group $C222_1$ with similar cell dimensions (see Table S1). The KTQ(E1) complex was solved by molecular replacement using the binary KlenTaq structure 3SZ2[22] as a search model. The KTQ(E2) and KTQ(E3) structures were

solved by rigid body refinement against KTQ(E1). All structures were improved by altering refinement in PHENIX[50] and model building in COOT.[51] During refinement, structures were evaluated using the MolProbity server.[52] Restraint files of d**NaM**, d**5SICS** and d**NaM**TP for refinement were created using the Grade Web Server.[53] Figures were created with PyMOL.[54]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the beamline staff of the Swiss Light Source at the Paul Scherrer Institute for their assistance during data collection. We thank Dr. Phillip Ordoukhanian for assistance with oligonucleotide synthesis.

Funding Sources

This work was supported by the National Institutes of Health (GM 060005 to F.E.R.) and by the Konstanz Research School Chemical Biology.

REFERENCES

1. Collins ML, Irvine B, Tyner D, Fine E, Zayati C, Chang C, Horn T, Ahle D, Detmer J, Shen LP, Kolberg J, Bushnell S, Urdea MS, Ho DD. Nucleic Acids Res. 1997; 25:2979–2984. [PubMed: 9224596]
2. Johnson SC, Marshall DJ, Harms G, Miller CM, Sherrill CB, Beaty EL, Lederer SA, Roesch EB, Madsen G, Hoffman GL, Laessig RH, Kopish GJ, Baker MW, Benner SA, Farrell PM, Prudent JR. Clin. Chem. 2004; 50:2019–2027. [PubMed: 15319316]
3. Arens MQ, Buller RS, Rankin A, Mason S, Whetsell A, Agapov E, Lee WM, Storch GA. J. Clin. Microbiol. 2010; 48:2387–2395. [PubMed: 20484608]
4. Lee WM, Grindle K, Pappas T, Marshall DJ, Moser MJ, Beaty EL, Shult PA, Prudent JR, Gern JE. J. Clin. Microbiol. 2007; 45:2626–2634. [PubMed: 17537928]
5. Kimoto M, Yamashige R, Matsunaga K, Yokoyama S, Hirao I. Nat. Biotechnol. 2013; 31:453–457. [PubMed: 23563318]
6. Hollenstein M, Hipolito CJ, Lam CH, Perrin DM. Nucleic Acids Res. 2009; 37:1638–1649. [PubMed: 19153138]
7. Keefe AD, Cload ST. Curr. Opin. Chem. Biol. 2008; 12:448–456. [PubMed: 18644461]
8. Seeman NC. Ann. Rev. Biochem. 2010; 79:65–87. [PubMed: 20222824]
9. Wang H, Yang R, Yang L, Tan W. ACS Nano. 2009; 3:2451–2460. [PubMed: 19658387]
10. Chen T, Shukoor MI, Chen Y, Yuan Q, Zhu Z, Zhao Z, Gulbakan B, Tan W. Nanoscale. 2011; 3:546–556. [PubMed: 21109879]
11. Chelliserrykattil J, Lu H, Lee AH, Kool ET. ChemBio-Chem. 2008; 9:2976–2980.
12. Brotschi C, Mathis G, Leumann CJ. Chem. Eur. J. 2005; 11:1911–1923. [PubMed: 15685710]
13. Kaul C, Muller M, Wagner M, Schneider S, Carell T. Nat. Chem. 2011; 3:794–800. [PubMed: 21941252]
14. Meggers E, Holland PL, Tolman WB, Romesberg FE, Schultz PG. J. Am. Chem. Soc. 2000; 122:10714–10715.
15. Minakawa N, Ogata S, Takahashi M, Matsuda A. J. Am. Chem. Soc. 2009; 131:1644–1645. [PubMed: 19146369]
16. Lavergne T, Malyshev DA, Romesberg FE. Chem. Eur. J. 2012; 18:1231–1239. [PubMed: 22190386]
17. Yamashige R, Kimoto M, Takezawa Y, Sato A, Mitsui T, Yokoyama S, Hirao I. Nucleic Acids Res. 2012; 40:2793–2806. [PubMed: 22121213]

18. Yang Z, Chen F, Alvarado JB, Benner SA. J. Am. Chem. Soc. 2011; 133:15105–15112. [PubMed: 21842904]
19. Malyshev DA, Dhami K, Quach HT, Lavergne T, Ordoukhanian P, Torkamani A, Romesberg FE. Proc. Natl. Acad. Sci. U.S.A. 2012; 109:12005–12010. [PubMed: 22773812]
20. Seo YJ, Matsuda S, Romesberg FE. J. Am. Chem. Soc. 2009; 131:5046–5047. [PubMed: 19351201]
21. Seo YJ, Malyshev DA, Lavergne T, Ordoukhanian P, Romesberg FE. J. Am. Chem. Soc. 2011; 133:19878–19888. [PubMed: 21981600]
22. Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE, Marx A. Nat. Chem. Biol. 2012; 8:612–614. [PubMed: 22660438]
23. Malyshev DA, Pfaff DA, Ippoliti SI, Hwang GT, Dwyer TJ, Romesberg FE. Chem. Eur. J. 2010; 16:12650–12659. [PubMed: 20859962]
24. Johar Z, Zahn A, Leumann CJ, Jaun B. Chem. Eur. J. 2008; 14:1080–1086. [PubMed: 18038386]
25. Matsuda S, Fillo JD, Henry AA, Rai P, Wilkens SJ, Dwyer TJ, Geierstanger BH, Wemmer DE, Schultz PG, Spraggon G, Romesberg FE. J. Am. Chem. Soc. 2007; 129:10466–10473. [PubMed: 17685517]
26. Krahn JM, Beard WA, Wilson SH. Structure. 2004; 12:1823–1832. [PubMed: 15458631]
27. Chou SH, Chin KH, Wang AH. Nucleic Acids Res. 2003; 31:2461–2474. [PubMed: 12736295]
28. Chou SH, Zhu L, Reid BR. J. Mol. Biol. 1994; 244:259–268. [PubMed: 7966337]
29. Shepard W, Cruse WB, Fourme R, de la Fortelle E, Prange T. Structure. 1998; 6:849–861. [PubMed: 9687367]
30. Špacková, Na; Berger, I.; Šponer, J. J. Am. Chem. Soc. 2000; 122:7564–7572.
31. Sunami T, Kondo J, Hirao I, Watanabe K, Miura KI, Takenaka A. Acta Crystallogr. D Biol. Crystallogr. 2004; 60:90–96. [PubMed: 14684897]
32. Rothwell PJ, Waksman G. Adv. Protein Chem. 2005; 71:401–440. [PubMed: 16230118]
33. Li Y, Korolev S, Waksman G. EMBO J. 1998; 17:7514–7525. [PubMed: 9857206]
34. Doublié S, Tabor S, Long AM, Richardson CC, Ellenberger T. Nature. 1998; 391:251–258. [PubMed: 9440688]
35. Kiefer JR, Mao C, Braman JC, Beese LS. Nature. 1998; 391:304–307. [PubMed: 9440698]
36. Echols H, Goodman MF. Annu. Rev. Biochem. 1991; 60:477–511. [PubMed: 1883202]
37. Kool ET. Annu. Rev. Biochem. 2002; 71:191–219. [PubMed: 12045095]
38. Goodman MF. Proc. Natl. Acad. Sci. U.S.A. 1997; 94:10493–10495. [PubMed: 9380666]
39. Kunkel TA. J. Biol. Chem. 2004; 279:16895–16898. [PubMed: 14988392]
40. Wu EY, Beese LS. J. Biol. Chem. 2011; 286:19758–19767. [PubMed: 21454515]
41. Obeid S, Welte W, Diederichs K, Marx A. J. Biol. Chem. 2012; 287:14099–14108. [PubMed: 22318723]
42. Obeid S, Blattner N, Kranaster R, Schnur A, Diederichs K, Welte W, Marx A. EMBO J. 2010; 29:1738–1747. [PubMed: 20400942]
43. Hohlbein J, Aigrain L, Craggs TD, Bermek O, Potapova O, Shoolizadeh P, Grindley NDF, Joyce CM, Kapanidis AN. Nat. Commun. 2013; 4:2131. [online]. [PubMed: 23831915]
44. Berezhna SY, Gill JP, Lamichhane R, Millar DP. J. Am. Chem. Soc. 2012; 134:11261–11268. [PubMed: 22650319]
45. Golosov AA, Warren JJ, Beese LS, Karplus M. Structure. 2010; 18:83–93. [PubMed: 20152155]
46. Ludwig J, Eckstein F. J. Org. Chem. 1989; 54:631–635.
47. Kabsch W. Acta Crystallgr. D Biol. Cyrstallogr. 2010; 66:125–132.
48. Karplus PA, Diederichs K. Science. 2012; 336:1030–1033. [PubMed: 22628654]
49. Betz K, Streckenbach F, Schnur A, Exner T, Welte W, Diederichs K, Marx A. Angew. Chem. Int. Ed. 2010; 49:5181–5184.
50. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. Acta Crystallogr. D Biol. Crystallogr. 2010; 66:213–221. [PubMed: 20124702]

51. Emsley P, Lohkamp B, Scott WG, Cowtan K. Acta Crystallogr. D Biol. Crystallogr. 2010; 66:486–501. [PubMed: 20383002]

52. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. Acta Crystallogr. D Biol. Crystallogr. 2010; 66:12–21. [PubMed: 20057044]

53. Smart, OS.; Womack, TO.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. grade, version 1.2.2. Cambridge, U.K.: 2011. http://www.globalphasing.com

54. DeLano, WL. The PyMOL Molecular Graphics System. San Carlos, CA: 2002. http://www.pymol.org

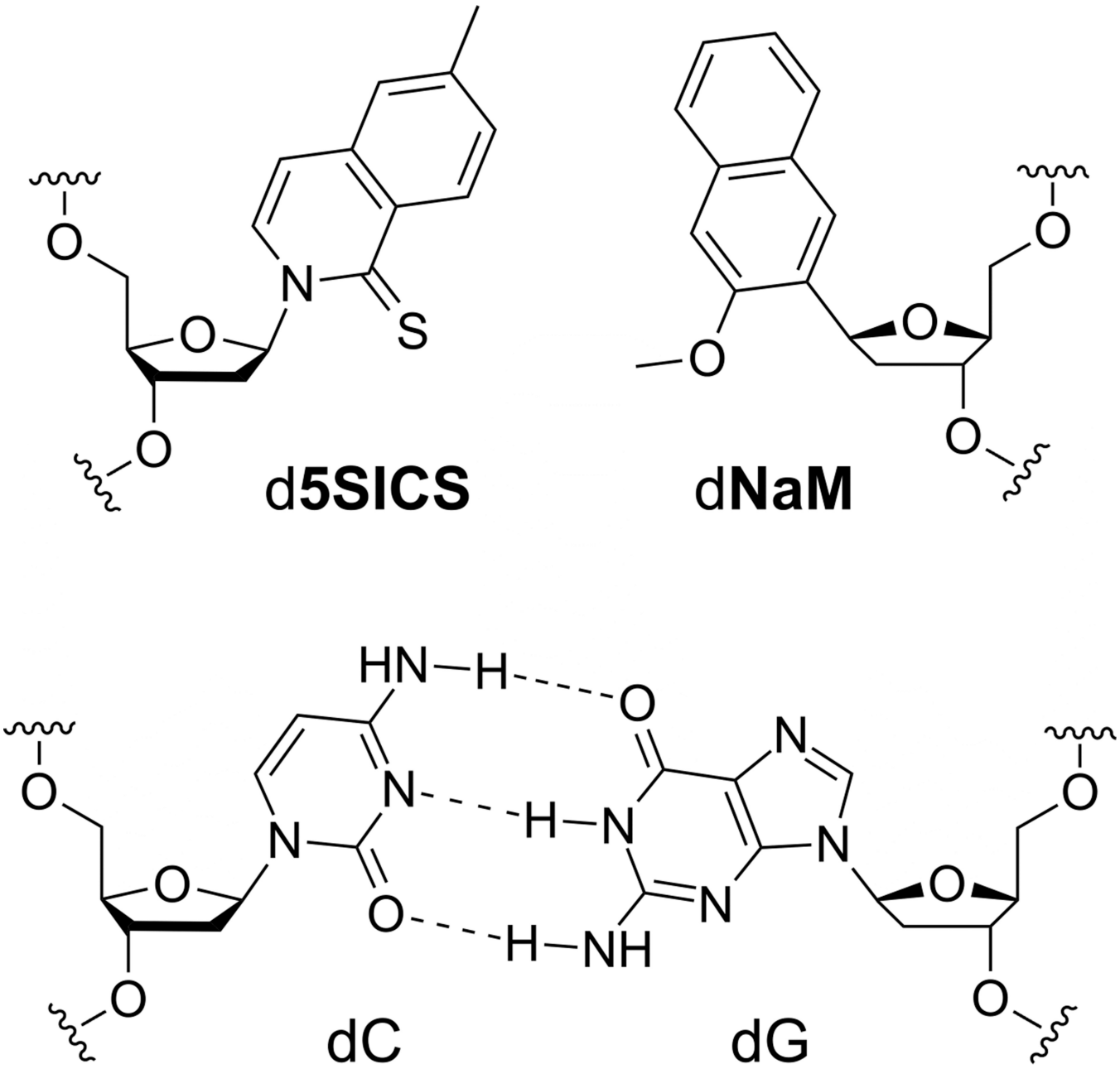

Figure 1.
The d**5SICS**-d**NaM** unnatural base pair, with a natural Watson-Crick base pair shown for comparison.

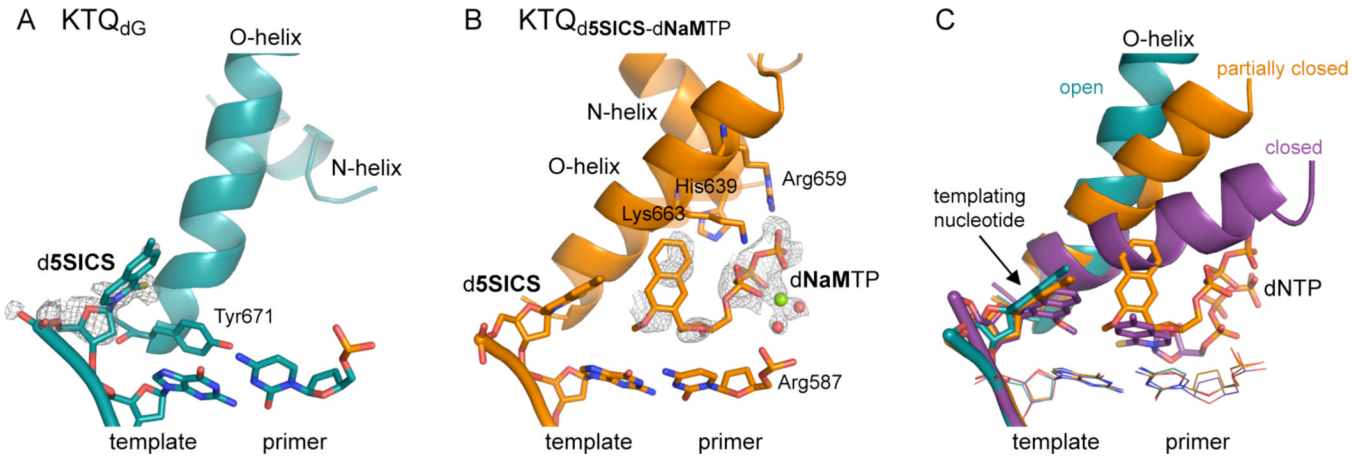

Figure 2.

Open binary and pre-catalytic ternary complex of KTQ_{d5SICS} and $KTQ_{d5SICS\text{-}dNaMTP}$, respectively. (A) The natural base pair at the post-insertion site, the templating d**5SICS**, and Tyr671 are shown as sticks and the O- and N-helices are shown as cartoon. Simulated annealing mFo-DFc omit map around d**5SICS** is shown, contoured at 3σ. (B) Same arrangement as in (A) but for the ternary complex $KTQ_{d5SICS\text{-}dNaMTP}$. Simulated annealing mFo-DFc omit map around the bound d**NaM**TP and the coordinated Mg^{2+} ion (green sphere) and associated water molecules (red spheres) is shown, contoured at 3σ. (C) Superposition of KTQ_{d5SICS} (cyan), $KTQ_{d5SICS\text{-}dNaMTP}$ (orange) and $KTQ_{dNaM\text{-}d5SICSTP}$ (PDB ID 3SZ2, purple) shows the open, partially closed and closed state of the enzyme.

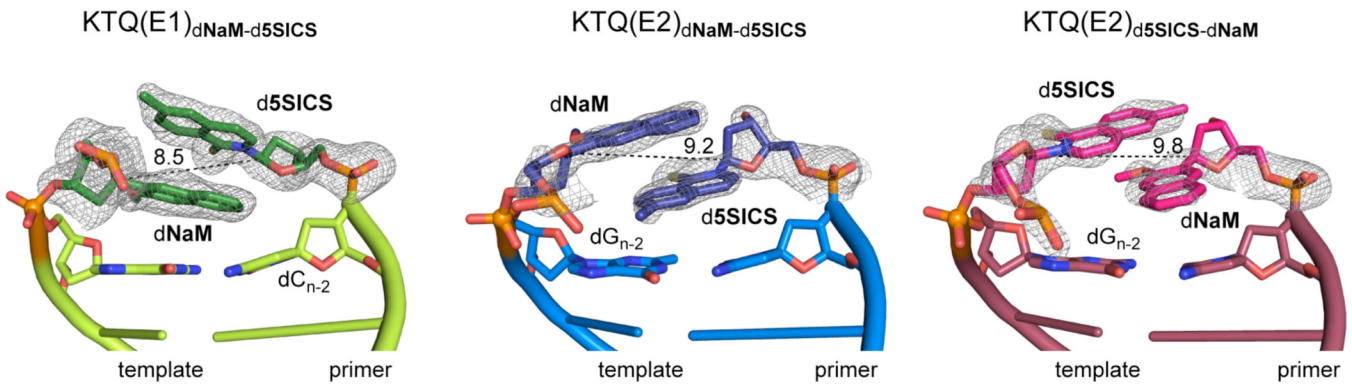

Figure 3.

Primer termini of open binary complexes with d**NaM**-d**5SICS** in the post-insertion site. KTQ(E1)$_{\text{dNaM-d5SICS}}$, KTQ(E2)$_{\text{dNaM-d5SICS}}$, and KTQ(E2)$_{\text{d5SICS-dNaM}}$ are labeled and shown in green, blue, and red, respectively. The intercalated unnatural base pair is shown in dark green, dark blue and pink, respectively, surrounded by their simulated annealing mFo-DFc omit maps contoured at 3σ. C1'-C1' distances (Å) within each unnatural pair are shown.

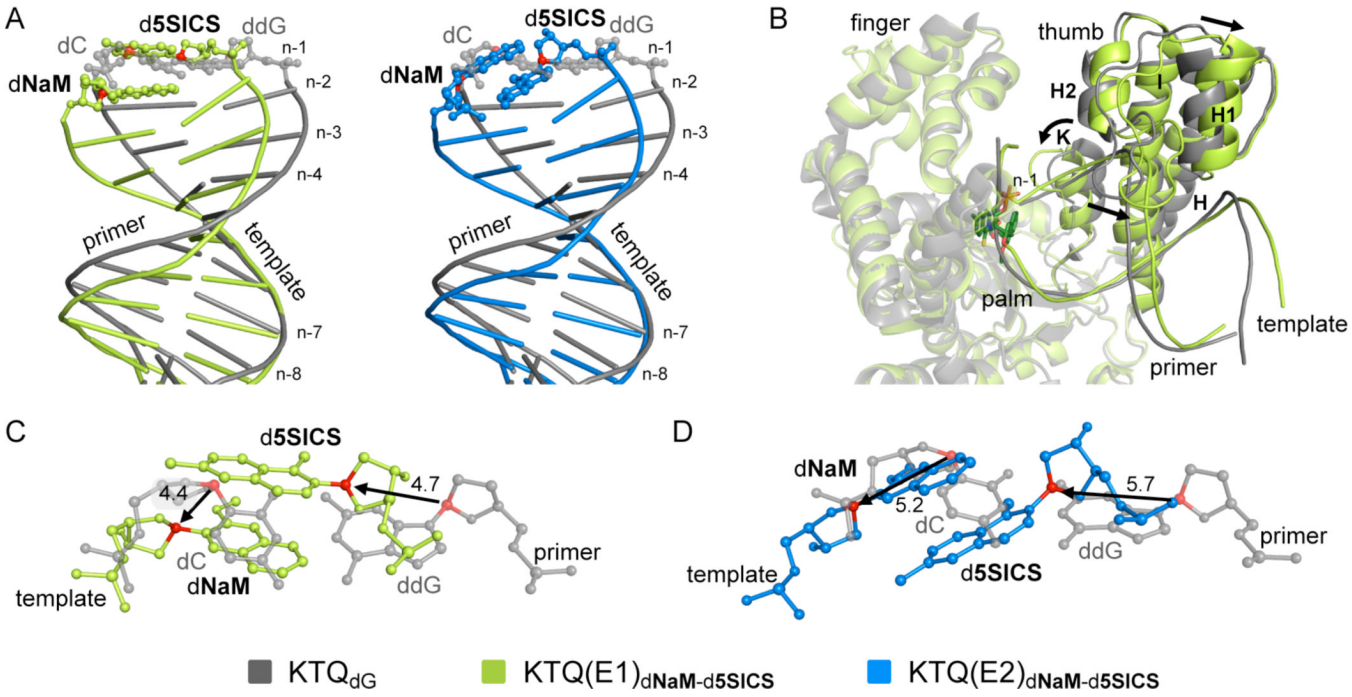

Figure 4.

Comparison of KTQ(E1)$_{\text{dNaM-d5SICS}}$ (green) and KTQ(E2)$_{\text{dNaM-d5SICS}}$ (blue) with KTQ$_{\text{dG}}$ with a natural dC-ddG base pair at the post-insertion site (grey). (A) Superposition of duplex portion of primer/template of KTQ(E1)$_{\text{dNaM-d5SICS}}$ and KTQ(E2)$_{\text{dNaM-d5SICS}}$ with KTQ$_{\text{dG}}$. The unnatural base pair in the post-insertion site is shown in ball and stick representation. (B) Superposition of KTQ(E1)$_{\text{dNaM-d5SICS}}$ and KTQ$_{\text{dG}}$, shown as cartoon. The finger and palm domains which are only slightly affected by the unnatural base pair are transparent. The larger movement of the thumb domain is indicated with black arrows. (C and D) Superposition of unnatural and natural base pair (from KTQ$_{\text{dG}}$) at post-insertion site with distance between C1' atoms indicated in Å, (C) KTQ(E1)$_{\text{dNaM-d5SICS}}$ and (D) KTQ(E2)$_{\text{dNaM-d5SICS}}$.

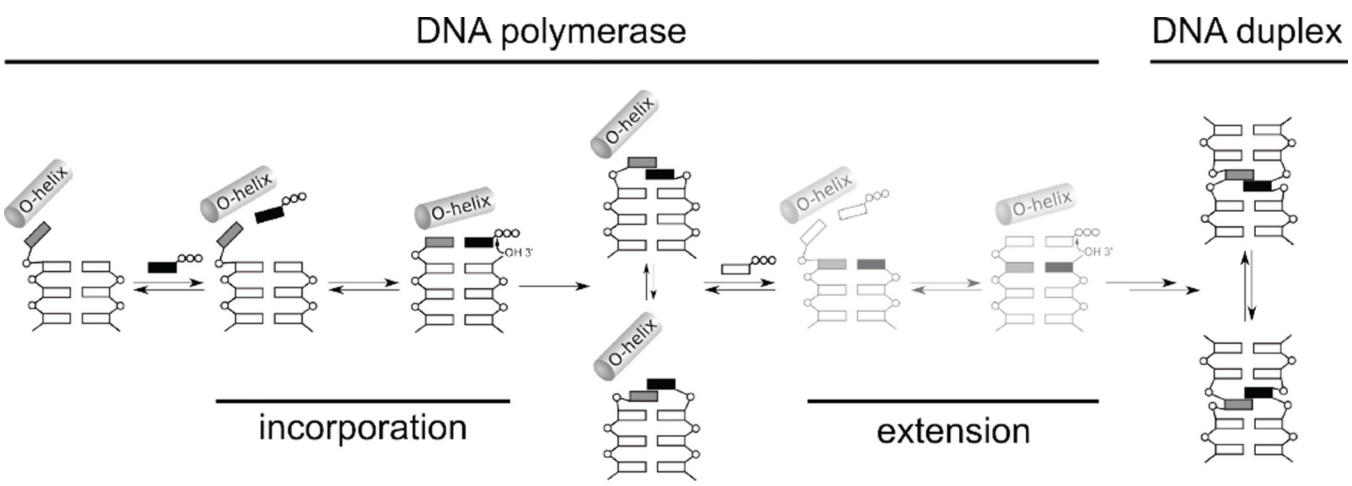

Figure 5.
Proposed mechanism of replication. Intermediates not yet validated by structural studies (*i.e.* extension complexes) are shown in lighter color. The steps corresponding to incorporation of the unnatural triphosphate and subsequent extension of the nascent unnatural base pair are indicated. The O-helix of the protein is shown, phosphates are indicated with open circles, natural nucleotides are indicated with open rectangles, and the unnatural nucleotides are indicated with grey and black rectangles.

Table 1

Primer/template sequences of post-incorporation complexes characterized.

Post-Incorporation Complex	Primer/Template Sequence	PDB ID
KTQ(E1)$_{\text{dNaM-d5SICS}}$	5‘-ACC ACG GCG C **5SICS** 3‘-TGG TGC CGC G **NaM** GA	4C8L
KTQ(E2)$_{\text{dNaM-d5SICS}}$	5‘-GCC ACG GCG C **5SICS** 3‘-CGG TGC CGC G **NaM** CTT	4C8O
KTQ(E2)$_{\text{d5SICS-dNaM}}$	5‘-GCC ACG GCG C **NaM** 3‘-CGG TGC CGC G **5SICS** CTT	4C8M
KTQ(E3)$_{\text{dNaM-d5SICS}}$	5‘-ACC ACG GCG C **5SICS** 3‘-TGG TGC CGC G **NaM** GTT	4C8N