KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry

Karin Betz[1],[5], **Denis A. Malyshev**[2],[5], **Thomas Lavergne**[2], **Wolfram Welte**[1], **Kay Diederichs**[1], **Tammy J. Dwyer**[3], **Phillip Ordoukhanian**[4], **Floyd E. Romesberg**[2],[*], and **Andreas Marx**[1],[*]

[1]Departments of Chemistry and Biology, and Konstanz Research School Chemical Biology, Universität Konstanz, Konstanz, Germany.

[2]Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, U.S.A.

[3]Department of Chemistry and Biochemistry, University of San Diego, San Diego, CA 92110, U.S.A.

[4]The Center for Protein and Nucleic Acid Research, The Scripps Research Institute, La Jolla, California 92037, U.S.A.

Abstract

Many candidate unnatural DNA base pairs have been developed, but surprisingly, some of the best replicated adopt intercalated structures in free DNA that are difficult to reconcile with known mechanisms of polymerase recognition. Here we present crystal structures of KlenTaq DNA polymerase at different stages of replicating one of the more promising pairs, dNaM-d5SICS, and show that efficient replication results from the polymerase itself inducing the required natural-like structure.

The development of a third, unnatural DNA base pair, and an expanded genetic alphabet, is a central goal of synthetic and chemical biology and would increase the functional diversity of nucleic acids, provide tools for their site-specific labeling[1,2], increase the information potential of DNA[3], and lay the foundation of a semi-synthetic organism[4]. DNA replication is a complex process during which DNA polymerases undergo large substrate-induced conformational changes from an "open" complex to a catalytically competent "closed" complex with the cognate triphosphate forming complementary Watson-Crick hydrogen-bonds (H-bonds) with the templating nucleotide that positions it for incorporation into the

[*]Correspondence and requests for materials should be addressed to A.M. and F.E.R. floyd@scripps.edu or andreas.marx@uni-konstanz.de.
[5]These authors contributed equally to this work.

growing primer strand[5]. In contrast to these conformational changes in the polymerase, the structures of the two natural base pairs, both before and after covalent incorporation of the triphosphate, are virtually identical to those formed in duplex DNA, in the absence of a polymerase (Fig. 1a). Moreover, the structures of the two natural base pairs are virtually identical to each other, and the rigorous selection of this conserved structure by DNA polymerases is thought to be essential for high fidelity replication[6-9]. Given this exquisite structure-based substrate selectivity, efficient replication of an unnatural base pair would appear to require that it adopt a structure that closely mimics that of a natural base pair.

Surprisingly, several of the best replicated candidate unnatural base pairs bear little or no resemblance to their natural counterparts and rely not on complementary H-bonding for their pairing, but rather on complementary hydrophobic and packing forces[10,11], a strategy for unnatural base pair design first pursued in 1999[12]. In particular, some of the most promising candidates belong to a family of analogues exemplified by d**NaM**-d**5SICS** (Fig. 1b, **top**), for which the efficiency of every step of unnatural base pair synthesis is within an order-of-magnitude of that for a natural base pair[10,13]. Moreover, DNA containing d**NaM**-d**5SICS** may be amplified by PCR or transcribed into RNA with efficiencies and fidelities that approach those of fully natural DNA[14,15]. However, based on the solution structure of duplex DNA with d**5SICS** paired opposite a d**NaM** analog[16], and confirmed here for d**NaM**-d**5SICS** itself via solution state NOEs (Supplementary Results, Supplementary Fig. 1), this family of unnatural base pairs forms via an intercalative mode of pairing (Fig. 1b, **left**). Indeed, intercalation appears to be a general feature of predominantly hydrophobic base pairs[17-19], which lack the H-bonds that favor the Watson-Crick-like edge-to-edge mode of pairing. While intercalation maximizes the packing interactions between the predominantly hydrophobic nucleobase analogues, it results in a structure that is very different from that of a natural base pair, and in fact, its structure is more similar to that of a mispair. It is thus difficult to reconcile the replication of these unnatural base pairs with the accepted mechanism of polymerase recognition[5-9,20], especially in the case of d**NaM**-d**5SICS**, which is replicated with such high efficiency and fidelity. To address this apparent contradiction, we report the 1.9 to 2.2 Å resolution X-ray crystal structures (Supplementary Table 1 and Supplementary Fig. 2) of three binary complexes of the large fragment of *Taq* DNA polymerase I (KlenTaq) bound to templates with a natural nucleotide or d**NaM** at the first templating position, as well as two ternary complexes with cognate natural or unnatural triphosphates.

We first solved the structure of the binary complex of KlenTaq bound to a primer-template with d**NaM** at the first templating position (KTQ$_{d\text{NaM}}$) (Supplementary Methods, Fig. 2a). For comparison we also solved the structure of the binary complex between KlenTaq and a fully natural primer-template containing dG or dT at the same position (KTQ$_{dG}$ and KTQ$_{dT}$) (Supplementary Fig. 3). In KTQ$_{d\text{NaM}}$, the polymerase adopts an overall structure that is similar to that observed in KTQ$_{dG}$ and KTQ$_{dT}$ (rmsd for C$_\alpha$ atoms: 0.83 Å and 0.35 Å, respectively). Moreover, the bound template of KTQ$_{d\text{NaM}}$ is virtually superimposable with that of KTQ$_{dT}$ (Supplementary Fig. 4), with the templating nucleotides flipped away from the developing duplex, and the two downstream nucleotides, dA$_{T3}$ and dA$_{T2}$ stacked on the exposed nascent base pair (Supplementary Fig. 3b,c). The hydrophobic nucleobase of d**NaM**

is positioned in the same pocket as the templating dT of KTQ_{dT}, where it engages in hydrophobic packing interactions with O helix residues Met673, Tyr671, Phe667, Tyr664, and the template nucleotides dA_{T3} and dA_{T2}. The methyl group of d**NaM** does not appear to make any specific contacts with the polymerase and the relatively weak signal in the electron density map indicates that it is nearly freely rotating. In contrast, the bound template of KTQ_{dG} adopts a conformation similar to that observed in the previously reported open structure of KlenTaq bound to a natural primer-template (PDB ID: 4KTQ)[21]. In these structures, the single-stranded template again kinks at its junction with the duplex portion of the primer-template, but instead of being packed by downstream nucleotides, the nascent base pair is packed by Tyr671 (Supplementary Fig. 5). The different binary complexes reveal that their structures are sequence-dependent, and the relatively large B factors (especially in fingers domain and the region proximal to the primer terminus, including Tyr671) (Supplementary Fig. 6), suggest that they are relatively dynamic. Nonetheless, the structural similarity of $KTQ_{d\mathbf{NaM}}$ and KTQ_{dT} suggests that the unnatural nucleotide is not abnormally perturbative.

To investigate whether the formation of d**NaM**-d**5SICS** is able to induce conformational changes similar to those induced by the formation of a natural base pair[5,21,22], we next solved the structure of the corresponding ternary complex ($KTQ_{d\mathbf{NaM}\text{-}d}\mathbf{5SICS}TP$) (Fig. 2b). The structure of $KTQ_{d\mathbf{NaM}\text{-}d}\mathbf{5SICS}TP$ reveals that d**5SICS**TP is bound in the active site, and as with natural substrates, its binding does indeed induce the closure of the fingers domain over the active site and a dramatic conformational change in the single-stranded portion of the template, with the phosphate backbone moved significantly and d**NaM** flipped back along the axis of the developing duplex where the two unnatural nucleotides pair (Fig. 2c -e). Interestingly, similar stabilization of the catalytically active complex is apparently not afforded by mispairing between d**NaM** and a natural triphosphate, as repeated attempts to soak crystals of $KTQ_{d\mathbf{NaM}}$ with natural triphosphates failed to produce a stable ternary complex. For a more detailed comparison of the conformational changes induced by correct natural or unnatural triphosphate binding, we also solved the structure of the analogous fully natural complex ($KTQ_{dG\text{-}dCTP}$). The structures of $KTQ_{dG\text{-}dCTP}$ and $KTQ_{d\mathbf{NaM}\text{-}d}\mathbf{5SICS}TP$ are similar to each other (Fig. 2f; rmsd for C_α atoms: 0.43 Å), and to the fully natural ternary complex of KlenTaq reported earlier (PDB ID: 3KTQ)[21] (rmsd for C_α atoms between $KTQ_{dG\text{-}dCTP}$ and 3KTQ: 0.30 Å). Relative to $KTQ_{dG\text{-}dCTP}$, the active site of $KTQ_{d\mathbf{NaM}\text{-}d}\mathbf{5SICS}TP$ is slightly enlarged to accommodate the unnatural base pair, and the relatively larger B-factors suggest that the fingers domain is somewhat more flexible (Supplementary Fig. 6).

A more detailed comparison of the active sites of $KTQ_{dG\text{-}dCTP}$ and $KTQ_{d\mathbf{NaM}\text{-}d}\mathbf{5SICS}TP$ further reveal their similarity (Fig. 2g and h). Just as in the fully natural complex, the orientation of the unnatural triphosphate is stabilized by interactions between its phosphate groups and the side chains of His639, Arg659, Lys663, and the amide backbone of Gln613. Also as with the natural triphosphates, the sugar rings of d**NaM** and d**5SICS**TP adopt the C3′-endo conformation, and the phosphate groups of the incoming triphosphate coordinate the two catalytically essential magnesium ions, which also coordinate polymerase residues Asp785, Asp610 and Tyr611 (Fig. 2h). The *ortho* substituents of both unnatural

nucleobases, which structure-activity-relationship data reveal are essential for replication[23], are oriented into the developing minor groove, in a fashion analogous to that of the H-bond acceptors of the natural nucleobases[24]. The d**5SICS**TP sulfur atom participates in a water-mediated H-bonding network with Glu615, Gln754 and Asn750, and the d**NaM** methoxy group, unlike in the binary structure, is well ordered and packed with the sulfur atom of d**5SICS**TP, the backbone of Gly668, and the carbonyl group of Phe667 on one side, and the guanine of the 3′ template nucleotide on the other. Lastly, the distance between the sugar C3′ of d**5SICS**TP and the α-phosphorus atom in $KTQ_{dNaM-d5SICSTP}$ is virtually identical to that observed for the natural triphosphate in $KTQ_{dG-dCTP}$ (3.9 and 3.8 Å, respectively).

Most remarkably, unlike the intercalated structure formed in a free duplex, the nucleobases of d**NaM** and d**5SICS**TP adopt a co-planar structure with nearly optimal edge-to-edge packing (average distance of 4.2 Å between the hydrophobic edges of the nucleobases), and a C1′-C1′ internucleotide distance that is roughly the same as that of a natural base pair (11.0 Å versus 10.6 Å, respectively, compared to 9.1 Å for d**NaM**-d**5SICS** in a free duplex). Thus, despite the absence of Watson-Crick-like H-bonds, and unlike in duplex DNA, the structure of d**NaM**-d**5SICS** in the polymerase active site is similar to that of a natural base pair (Fig. 1). The reduced level of intercalation is likely due in part to the A-form structure of the primer terminus, which is wider than the B-form structure of the free duplex. However, given that the triphosphate is only constrained by non-covalent interactions, greater intercalation should be possible, and the fact that it is not observed suggests that the sum of the interactions between the developing base pair and the polymerase active site favors a planar, Watson-Crick-like geometry.

The data provide an explanation for the empirical observation that complementary H-bond formation is not required for the efficient and selective replication of DNA[10-12,25]. d**NaM**-d**5SICS** is efficiently and selectively replicated because its formation provides a suitably strong driving force to induce the required structural transitions in the polymerase, and because it also possesses sufficient plasticity to adapt to the structure it induces in the polymerase. Correspondingly, the data reveal that the polymerase active site is not only capable of selecting for a correct structure among the pairing nucleotides, but also at least in some cases it is capable of enforcing it. Moreover, the efficient replication of d**NaM**-d**5SICS** by a variety of other polymerases, including polymerases from different families[13,14], suggests that these observations with KlenTaq may be general. It is interesting to speculate that polymerases may have evolved to favor a co-planar geometry to prevent natural nucleotide mispairing via cross-strand intercalation and instead allow only the more specific, edge-to-edge H-bonding interactions. Finally, the data reveal that structural mimicry of a natural base pair is not required for unnatural base pair design and that, as is the case with protein structure and folding, the strong but relatively plastic nature of hydrophobic and packing forces makes them particularly well suited to underlie an expanded genetic alphabet. Further studies aimed at elucidating the factors underlying the efficient continued DNA synthesis after d**NaM**-d**5SICS** synthesis are currently underway.

Accession codes. Protein Data Bank: The atomic coordinates and structure factors for the reported crystal structures are deposited under accession 3SZ2 (KTQ_{dG}), 3SV4 (KTQ_{dT}), 3SYZ (KTQ_{dNaM}), 3RTV ($KTQ_{dG-dCTP}$), 3SV3 ($KTQ_{dNaM-d5SICSTP}$).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Hollenstein M, Hipolito CJ, Lam CH, Perrin DM. Nucleic Acids Res. 2009; 37:1638–1649. [PubMed: 19153138]

2. Seeman NC. Ann. Rev. Biochem. 2010; 79:65–87. [PubMed: 20222824]

3. Piccirilli JA, Krauch T, Moroney SE, Benner SA. Nature. 1990; 343:33–37. [PubMed: 1688644]

4. Xie JM, Schultz PG. Nat. Rev. Mol. Cell Biol. 2006; 7:775–782. [PubMed: 16926858]

5. Rothwell PJ, Waksman G. Adv. Prot. Chem. 2005; 71:401–440.

6. Echols H, Goodman MF. Annu. Rev. Biochem. 1991; 60:477–511. [PubMed: 1883202]

7. Goodman MF. Proc. Natl. Acad. Sci. USA. 1997; 94:10493–10495. [PubMed: 9380666]

8. Kool ET. Ann. Rev. Biochem. 2002; 71:191–219. [PubMed: 12045095]

9. Kunkel TA. J. Biol. Chem. 2004; 279:16895–16898. [PubMed: 14988392]

10. Seo YJ, Hwang GT, Ordoukhanian P, Romesberg FE. J. Am. Chem. Soc. 2009; 131:3246–3252. [PubMed: 19256568]

11. Hirao I, et al. Nat. Methods. 2006; 3:729–735. [PubMed: 16929319]

12. McMinn DL, et al. J. Am. Chem. Soc. 1999; 121:11585–11586.

13. Lavergne T, Malyshev DA, Romesberg FE. Chem. Eur. J. 2012; 18:1231–1239. [PubMed: 22190386]

14. Malyshev DA, Seo YJ, Ordoukhanian P, Romesberg FE. J. Am. Chem. Soc. 2009; 131:14620–14621. [PubMed: 19788296]

15. Seo YJ, Matsuda S, Romesberg FE. J. Am. Chem. Soc. 2009; 131:5046–5047. [PubMed: 19351201]

16. Malyshev DA, et al. Chem. Eur. J. 2010; 16:12650–12659. [PubMed: 20859962]

17. Matsuda S, et al. J. Am. Chem. Soc. 2007; 129:10466–10473. [PubMed: 17685517]

18. Wojciechowski F, Leumann CJ. Chem. Soc. Rev. 2011; 40:5669–5679. [PubMed: 21431120]

19. Brotschi C, Haberli A, Leumann CJ. Angew. Chem. Int. Ed. 2001; 40:3012–3014.

20. Wu EY, Beese LS. J. Biol. Chem. 2011; 286:19758–19767. [PubMed: 21454515]

21. Li Y, Korolev S, Waksman G. EMBO J. 1998; 17:7514–7525. [PubMed: 9857206]

22. Doublié S, Tabor S, Long AM, Richardson CC, Ellenberger T. Nature. 1998; 391:251–258. [PubMed: 9440688]

23. Leconte AM, et al. J. Am. Chem. Soc. 2008; 130:2336–2343. [PubMed: 18217762]

24. Meyer AS, Blandino M, Spratt TE. J. Biol. Chem. 2004; 279:33043–33046. [PubMed: 15210707]

25. Moran S, Ren RXF, Rumney S, Kool ET. J. Am. Chem. Soc. 1997; 119:2056–2057. [PubMed: 20737028]
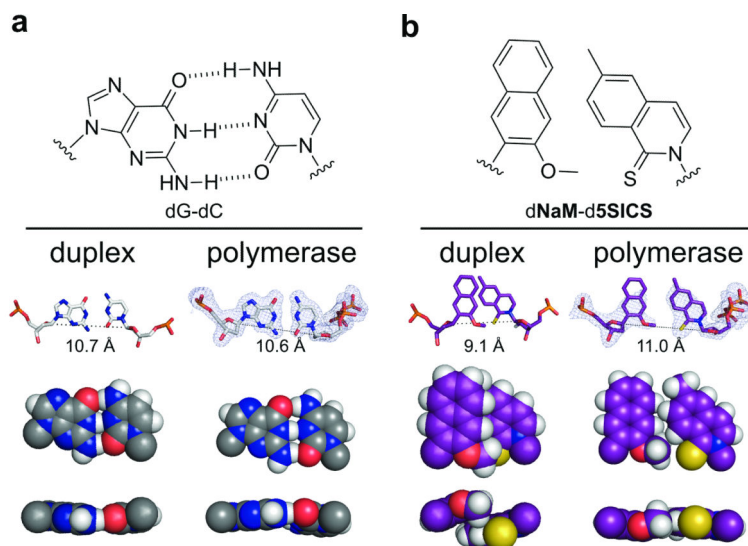
Figure 1. KlenTaq polymerase induces the dNaM-d5SICS unnatural base pair to adopt a natural, Watson-Crick-like structure

Structure of (**a**) a natural dG-dC base pair and (**b**) d**NaM**-d**5SICS**. Chemical structures are shown at the top of each panel, with a comparison of the structure formed between two nucleotides in duplex DNA (left) and between the templating nucleotide and the incoming triphosphate in the active site of KlenTaq polymerase (right). C1′-C1′ distances are indicated, and CPK renderings are viewed from above and from the minor groove. Only nucleobases are shown with sugar and phosphates omitted for clarity.
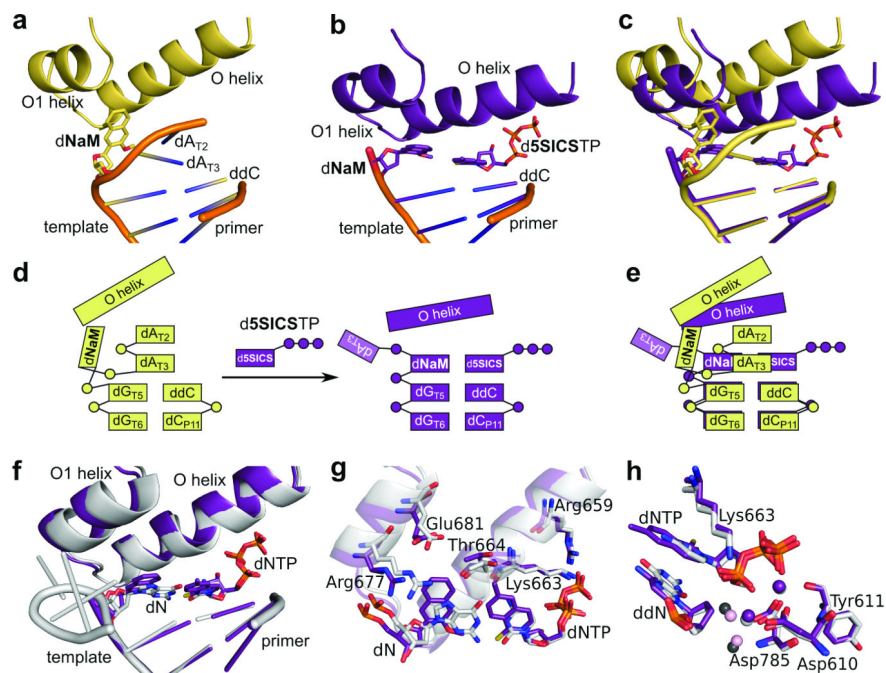
Fig. 2. Unnatural base pair formation induces conformational transitions of KlenTaq and the formation of a natural-like ternary complex

Structure of complexes showing helices O and O1, primer-template, and d**5SICS**TP (if present) in: (**a**) KTQ$_{dNaM}$ binary complex (yellow); (**b**) KTQ$_{dNaM-d}$**5SICS**TP ternary complex (purple); and (**c**) their superposition highlighting the structural transition induced by d**5SICS**TP binding. Schematic illustration of conformational transition induced by d**5SICS**TP binding (**d**) and superposition of binary and ternary complexes (**e**). Superposition of KTQ$_{dNaM-d5SICS}$TP (purple) and KTQ$_{dG-dCTP}$ (grey) illustrating the similarities of: (**f**) helices O and O1, and primer-template; (**g**) active site; and (**h**) catalytically critical network of side chains, water molecules, and Mg^{2+} ions (water molecules and magnesium ions are shown as light pink and purple spheres, respectively, for KTQ$_{dNaM-d}$**5SICS**TP and as dark grey and light grey spheres, respectively, for KTQ$_{dG-dCTP}$; incoming triphosphate is labeled dNTP, templating nucleotide is labeled with dN and dideoxynucleotide at the primer terminus is labeled with ddN, as appropriate).