First publ. in: Nature Chemical Biology ; 8 (2012), 7. - pp. 612-614

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

Karin Betz[1,2,7], Denis A Malyshev[3,7], Thomas Lavergne[3], Wolfram Welte[2,4], Kay Diederichs[2,4], Tammy J Dwyer[5], Phillip Ordoukhanian[6], Floyd E Romesberg[3]* & Andreas Marx[1,2]*

Many candidate unnatural DNA base pairs have been developed, but some of the best-replicated pairs 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 replication for one such pair, 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 for a semisynthetic organism[4]. DNA replication is a complex process during which DNA polymerases undergo substantial substrate-induced conformational changes from an 'open' complex to a catalytically competent 'closed' complex, with the cognate triphosphate forming complementary Watson-Crick hydrogen bonds with the templating nucleotide that positions it for incorporation into the 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 seem to require that it adopts a structure that closely mimics that of a natural base pair.

Several of the best replicated candidate unnatural base pairs bear little or no resemblance to their natural counterparts and rely not on complementary hydrogen 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 (ref. 12). In particular, some of the most promising candidates belong to a family of analogs exemplified by **dNaM**-**d5SICS** (**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 **dNaM**-**d5SICS** may be amplified by PCR or transcribed into RNA with efficiencies and fidelities that approach those of fully natural DNA[14,15]. However, as observed in the solution structure of duplex DNA with **d5SICS** paired opposite a **dNaM** analog[16] and as confirmed here for **dNaM**-**d5SICS** 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 seems to be a general feature of predominantly hydrophobic base pairs[17–19], which lack the hydrogen bonds that favor the Watson-Crick–like edge-to-edge mode of pairing. Although intercalation maximizes the packing interactions between the predominantly hydrophobic nucleobase analogs, 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 **dNaM**-**d5SICS**, which is replicated with such high efficiency and fidelity. To address this apparent contradiction, we determined 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 **dNaM** at the first templating position as well as those of 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 **dNaM** at the first templating position (KTQ_{dNaM}; **Fig. 2a** and **Supplementary Methods**). 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}, respectively; **Supplementary Fig. 3**). In KTQ_{dNaM}, the polymerase adopts an overall structure that is similar to that observed in KTQ_{dG} and KTQ_{dT} (r.m.s. deviation for Cα atoms: 0.83 Å and 0.35 Å, respectively). Moreover, the bound template of KTQ_{dNaM} 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 **dNaM** 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 and Thr664 and the template nucleotides dA_{T3} and dA_{T2}. The methyl group of **dNaM** 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 (Protein Data Bank (PDB) code 4KTQ)[21]. In the these structures, the

[1]Department of Chemistry, Universität Konstanz, Konstanz, Germany. [2]Konstanz Research School-Chemical Biology, Universität Konstanz, Konstanz, Germany. [3]Department of Chemistry, The Scripps Research Institute, La Jolla, California, USA. [4]Department of Biology, Universität Konstanz, Konstanz, Germany. [5]Department of Chemistry and Biochemistry, University of San Diego, San Diego, California, USA. [6]The Center for Protein and Nucleic Acid Research, The Scripps Research Institute, La Jolla, California, USA. [7]These authors contributed equally to this work. *e-mail: floyd@scripps.edu or andreas.marx@uni-konstanz.de

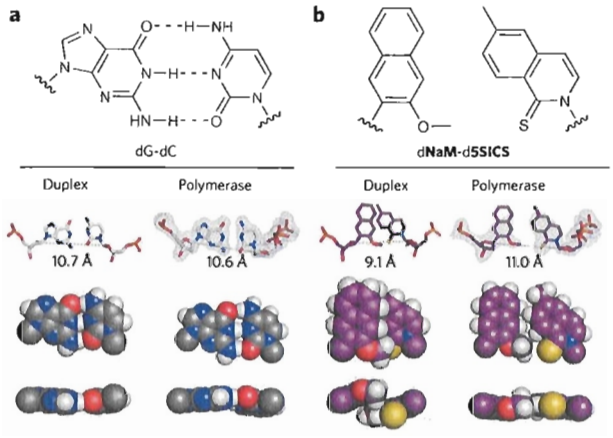

Figure 1 | KlenTaq polymerase induces the dNaM-d5SICS unnatural base pair to adopt a natural, Watson-Crick-like structure. (**a**,**b**) Structure of a natural dG-dC base pair (**a**) and **dNaM**-d**5SICS** (**b**). Chemical structures are shown at the top of each panel, with a comparison of the structure 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 space-filling models are viewed from above and from the minor groove.

single-stranded template 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**). Overall, we found that the structures of the binary complexes are sequence dependent, and the relatively large B-factors (especially in the 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_{dNaM} 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 ($KTQ_{dNaM\text{-}d5SICSTP}$) (**Fig. 2b**). The structure of $KTQ_{dNaM\text{-}d5SICSTP}$ reveals that d**5SICSTP** 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, including movement of the phosphate backbone and flipping of d**NaM** back along the axis of the developing duplex, where the two unnatural nucleotides pair (**Fig. 2c–e**). Remarkably, 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_{dNaM} 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 structure of $KTQ_{dG\text{-}dCTP}$ is similar to that of $KTQ_{dNaM\text{-}d5SICSTP}$ (**Fig. 2f**; r.m.s. deviation for Cα atoms: 0.43 Å) and to that of the fully natural ternary complex of KlenTaq reported earlier (PDB code 3KTQ[21]; r.m.s. deviation for Cα atoms between $KTQ_{dG\text{-}dCTP}$ and 3KTQ: 0.30 Å). Relative to $KTQ_{dG\text{-}dCTP}$, the active site of $KTQ_{dNaM\text{-}d5SICSTP}$ 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 $KTQ_{dNaM\text{-}d5SICSTP}$ further reveal their similarity (**Fig. 2g,h**).

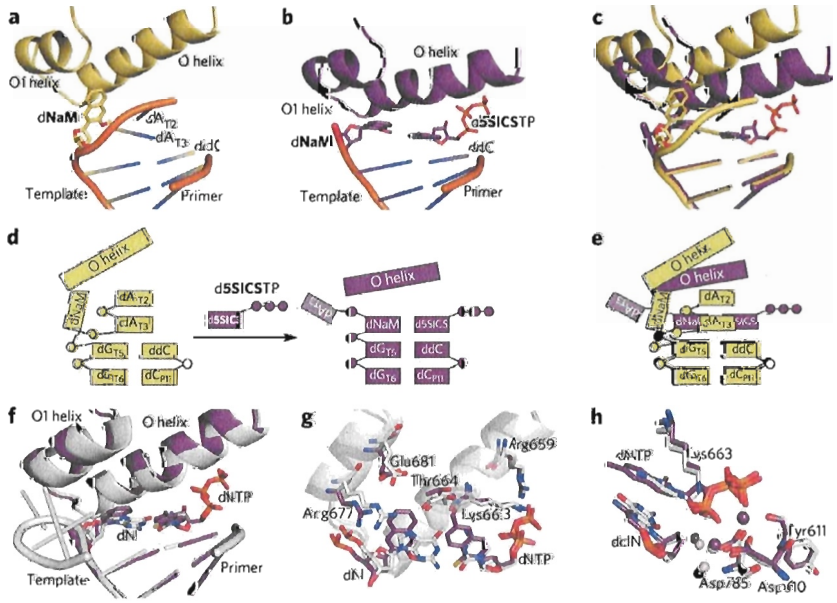

Figure 2 | Unnatural base pair formation induces conformational transitions of KlenTaq and the formation of a natural-like ternary complex. (**a–c**) Structure of complexes showing helices O and O1, primer-template and d**5SICS**TP (if present) in KTQ_{dNaM} binary complex (yellow) (**a**); $KTQ_{dNaM\text{-}d5SICSTP}$ ternary complex (purple) (**b**); and a superposition of the two, highlighting the structural transition induced by d**5SICS**TP binding (**c**). (**d**,**e**) Schematic illustration of conformational transition induced by d**5SICS**TP binding (**d**) and superposition of binary and ternary complexes (**e**). (**f–h**) Superposition of $KTQ_{dNaM\text{-}d5SICSTP}$ (purple) and $KTQ_{dG\text{-}dCTP}$ (gray), illustrating the similarities of helices O and O1 and primer-template (**f**); the active site (**g**); and the catalytically critical network of side chains, water molecules and Mg^{2+} ions (**h**) (water molecules and magnesium ions are shown as light pink and purple spheres, respectively, for $KTQ_{dNaM\text{-}d5SICSTP}$, and as dark gray and light gray spheres, respectively, for $KTQ_{dG\text{-}dCTP}$). dNTP, incoming triphosphate; dN, templating nucleotide; ddN, dideoxynucleotide at the primer terminus.

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 and Lys663 and the amide backbone of Gln613. 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 hydrogen-bond acceptors of the natural nucleobases[24]. The d**5SICS**TP sulfur atom participates in a water-mediated hydrogen-bonding network with Glu615, Gln754 and Asn750. 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 $KTQ_{dNaM\text{-}d5SICSTP}$ is virtually identical to that observed for the natural triphosphate in $KTQ_{dG\text{-}dCTP}$ (3.8 Å and 3.9 Å, respectively).

Most remarkably, unlike the intercalated structure formed in a free duplex, the structure adopted by the nucleobases of d**NaM** and d**5SICS**TP is co-planar 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 hydrogen 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 intercalation is most 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 noncovalent 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 hydrogen 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 has 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, 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 generalized to other polymerases. Perhaps polymerases have evolved to favor a coplanar geometry to prevent natural nucleotide mispairing via cross-strand intercalation and instead allow only the more specific, edge-to-edge hydrogen-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 under way.

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

Received 11 October 2011; accepted 3 April 2012;
published online 3 June 2012; corrected online 6 June 2012

References

1. Hollenstein, M., Hipolito, C.J., Lam, C.H. & Perrin, D.M. *Nucleic Acids Res.* **37**, 1638–1649 (2009).
2. Seeman, N.C. *Annu. Rev. Biochem.* **79**, 65–87 (2010).
3. Piccirilli, J.A., Krauch, T., Moroney, S.E. & Benner, S.A. *Nature* **343**, 33–37 (1990).
4. Xie, J. & Schultz, P.G. *Nat. Rev. Mol. Cell Biol.* **7**, 775–782 (2006).
5. Rothwell, P.J. & Waksman, G. *Adv. Protein Chem.* **71**, 401–440 (2005).
6. Echols, H. & Goodman, M.F. *Annu. Rev. Biochem.* **60**, 477–511 (1991).
7. Goodman, M.F. *Proc. Natl. Acad. Sci. USA* **94**, 10493–10495 (1997).
8. Kool, E.T. *Annu. Rev. Biochem.* **71**, 191–219 (2002).
9. Kunkel, T.A. *J. Biol. Chem.* **279**, 16895–16898 (2004).
10. Seo, Y.J., Hwang, G.T., Ordoukhanian, P. & Romesberg, F.E. *J. Am. Chem. Soc.* **131**, 3246–3252 (2009).
11. Hirao, I. *et al. Nat. Methods* **3**, 729–735 (2006).
12. McMinn, D.L. *et al. J. Am. Chem. Soc.* **121**, 11585–11586 (1999).
13. Lavergne, T., Malyshev, D.A. & Romesberg, F.E. *Chemistry* **18**, 1231–1239 (2012).
14. Malyshev, D.A., Seo, Y.J., Ordoukhanian, P. & Romesberg, F.E. *J. Am. Chem. Soc.* **131**, 14620–14621 (2009).
15. Seo, Y.J., Matsuda, S. & Romesberg, F.E. *J. Am. Chem. Soc.* **131**, 5046–5047 (2009).
16. Malyshev, D.A. *et al. Chemistry* **16**, 12650–12659 (2010).
17. Matsuda, S. *et al. J. Am. Chem. Soc.* **129**, 10466–10473 (2007).
18. Wojciechowski, F. & Leumann, C.J. *Chem. Soc. Rev.* **40**, 5669–5679 (2011).
19. Brotschi, C., Haberli, A. & Leumann, C.J. *Angew. Chem. Int. Ed. Engl.* **40**, 3012–3014 (2001).
20. Wu, E.Y. & Beese, L.S. *J. Biol. Chem.* **286**, 19758–19767 (2011).
21. Li, Y., Korolev, S. & Waksman, G. *EMBO J.* **17**, 7514–7525 (1998).
22. Doublié, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. *Nature* **391**, 251–258 (1998).
23. Leconte, A.M. *et al. J. Am. Chem. Soc.* **130**, 2336–2343 (2008).
24. Meyer, A.S., Blandino, M. & Spratt, T.E. *J. Biol. Chem.* **279**, 33043–33046 (2004).
25. Moran, S., Ren, R.X.F., Rumney, S. & Kool, E.T. *J. Am. Chem. Soc.* **119**, 2056–2057 (1997).

Acknowledgments

We thank the beamline staff of the Swiss Light Source at the Paul Scherrer Institute for their assistance during data collection. This work was supported by the Konstanz Research School Chemical Biology (to K.B.) and the US National Institutes of Health (GM060005 to F.E.R.).

Author contributions

K.B., D.A.M., T.L., F.E.R. and A.M. conceived of the project, designed the experiments and analyzed the data. K.B., D.A.M., T.L. and P.O. performed chemical synthesis. K.B., W.W. and K.D. performed crystallography studies. T.J.D. performed the NMR experiments, and D.A.M. and T.J.D. performed modeling studies. K.B., D.A.M., A.M. and F.E.R. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to A.M. or F.E.R.