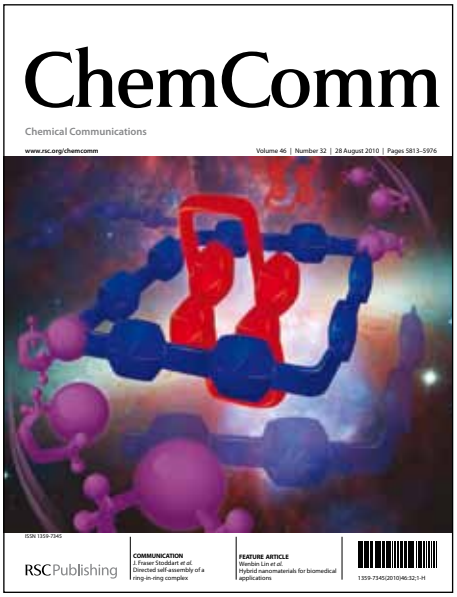
ChemComm

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the RSC Publishing peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, which is prior to technical editing, formatting and proof reading. This free service from RSC Publishing allows authors to make their results available to the community, in citable form, before publication of the edited article. This *Accepted Manuscript* will be replaced by the edited and formatted *Advance Article* as soon as this is available.

To cite this manuscript please use its permanent Digital Object Identifier (DOI®), which is identical for all formats of publication.

More information about *Accepted Manuscripts* can be found in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics contained in the manuscript submitted by the author(s) which may alter content, and that the standard **Terms & Conditions** and the **ethical guidelines** that apply to the journal are still applicable. In no event shall the RSC be held responsible for any errors or omissions in these *Accepted Manuscript* manuscripts or any consequences arising from the use of any information contained in them.

View Online

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

COMMUNICATION

Interactions of Non-Polar and "Click-able" Nucleotides in the Confines of a DNA Polymerase

Samra Obeid,*[a] Holger Bußkamp,[a] Wolfgang Welte,[b] Kay Diederichs[b] and Andreas Marx[a]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

Modified nucleotides play a paramount role in many cutting-edge biomolecular techniques. The present structural study highlights the plasticity and flexibility of the active site of a DNA polymerase while incorporating non-polar "Click-able" nucleotide analogs and emphasizing new insights into rational design guidelines for modified nucleotides.

Regarding biological core technologies such as next-generation sequencing[1], single molecule sequencing[2], microarray analysis based on labeled DNA amplificates[3], DNA conjugation[4], or the in vitro selection of ligands like aptamers by SELEX (systematic enrichment of ligands by exponential amplification)[5], functionalized 2'-deoxynucleoside triphosphates (dNTPs) play a fundamental role. One approach to improve the incorporation efficiency of modified substrates is to vary the linkage connecting nucleotide moiety and modification of choice. In most cases the modifications were introduced at the nucleobase, in particular, in case of pyrimidines at C5 position[1-6]. For rational linkage design the knowledge of the structure-function relationship of the enzyme is of great importance. Two recently solved structures showed *KlenTaq* (the large fragment of *Thermus aquaticus* DNA polymerase I) caught processing several modified dNTPs[7]. These studies focused on flexible and polar modifications that have distinct hydrogen-bonding capability. It turned out that the polar functionalities allow proper substrate-protein interaction and thereby might improve the substrate properties. However, little is known about rigid and nonpolar modification. So far, only one DNA polymerase structure capturing a rigid and nonpolar modification is available[7a]. This study suggested that the rigid and non-polar modification results in the distortion of the interaction between the enzyme and the substrate that leads to

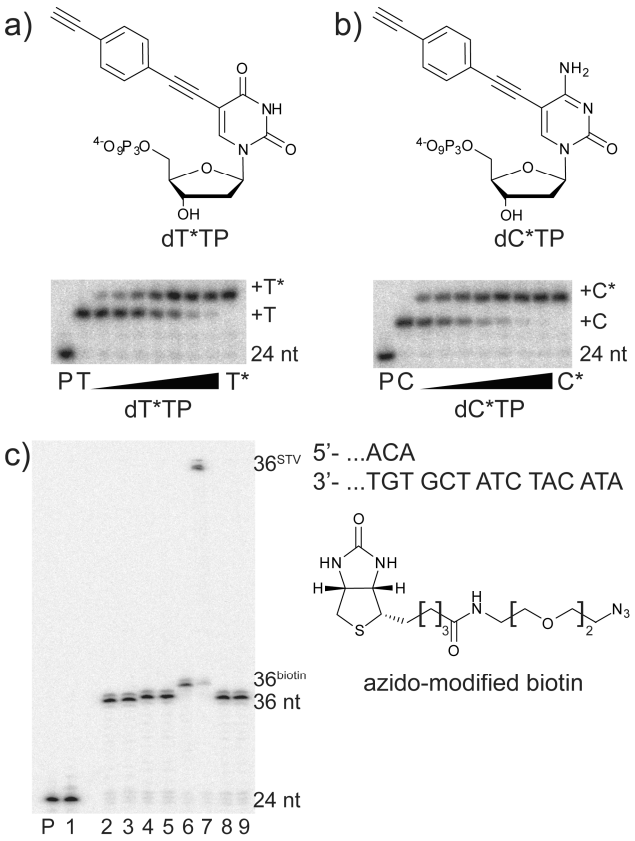


Fig. 1 a) Structure of dT*TP and primer extension. A representative PAGE analysis of the competition experiment dT*TP vs dTTP using *KlenTaq* is shown. The ratio of dT*TP/dTTP was varied from 1/1 to 100/1 (1/1, 2/1, 4/1, 10/1, 20/1, 50/1, 100/1). b) same as in a) for dC*TP. c) Primer extension using dC*TP followed by Click reaction. Partial sequence of the employed primer/template construct is indicated on the right. Lane P: primer only; lane 1: primer extension by *KlenTaq* in presence of dATP, dGTP and dTTP; lane 2: all four natural dNTPs; lane 3: same as lane 2 followed by incubation with streptavidin (STV); lane 4: same as in lane 1, but in the presence of dC*TP; lane 5: same as in lane 4 followed by incubation with STV; lane 6: same as in lane 4 followed by Click reaction with azido-modified biotin; lane 7: same as line 6 followed by incubation with STV; lane 8: same as lane 4 followed by incubation with azido-modified biotin (structure is shown) without BTTAA and $CuSO_4$; lane 9: same as in lane 8 followed by incubation with STV. For experimental detail see ESI †.

[a] *Department of Chemistry and Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany. Fax: (+49) 7531-88-5140; Tel: (+49) 7531-88-5139; E-mail: andreas.marx@uni-konstanz.de*

[b] *Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany.*

† Electronic Supplementary Information (ESI) available: See DOI: 10.1039/b000000x/

‡ Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4ELT, 4ELU and 4ELV).

the observed low incorporation efficiency. However, in some cases rigid linkers are needed to enable e.g. spectroscopic properties. Therefore, we extended the functional and structural studies with a new set of rigid C5 modified nucleotides (Fig. 1). The introduction of an aromatic ring in the rigid linkage should enable interactions to the protein. Further, these modified nucleotides serve as precursors, which can be post-synthetically addressed by Cu(I)-catalyzed azide-alkyne cycloaddition ("Click" chemistry), due to the terminal alkyne function[6l, 8]. We present crystal structures showing *KlenTaq* caught during the incorporation process of modified nucleotides. The structures reveal that the interaction pattern of the aromatic ring with the protein might account for the efficient processing of the analogs by *KlenTaq*.

Starting from 5-iodo 2'-deoxyuridine the modification was introduced by palladium-catalyzed cross-coupling[6n, 9] with 1,4-dienthynylbenzene. Next the modified thymidine analogs were transformed to the corresponding cytidine analog[10], followed by the conversion into the respective thymidine- or cytidine triphosphate analogs[11] (for details see ESI †).

Having the modified pyrimidine analogs in hand their acceptance and incorporation efficiency compared to the natural substrates by *KlenTaq* were investigated by competition experiments[3c, 7](Fig. 1a,b). In these single nucleotide incorporation experiments the modified substrates directly compete with the natural counterpart for incorporation. The objective was to determine the relative ratio of modified versus natural substrate that results in equal incorporation efficiencies. This ratio was assigned using 20% denaturing polyacrylamide gel electrophoresis exploiting the different retention times of the reaction products, due to the additional bulk of the modification[6n, 7a, 12]. The quantification of these experiments revealed that *KlenTaq* incorporates dT*MP and dC*MP with only 7- and 2-fold lower efficiency compared to the natural counterparts, respectively (Fig. S1).

Encouraged by the excellent acceptance of the modified substrates, we next investigated the access of the terminal alkyne group via Click reaction (Fig. 1c, S2 and Scheme S2). First, we performed primer extension experiments using modified substrates to introduce a terminal alkyne group into DNA (Fig. 1c, lane 4). The mono alkyne functionalized DNA product was subsequently used for the Click reaction. An azido-modified biotin was clicked to the modified DNA. After 4h reaction time a quantitative conversion to the biotin labeled DNA product is observed, which shows a slight decrease in retention time, due to the additional bulk of the introduced modification (Fig. 1c, lane 6). Further, incubation with streptavidin and the resulting change in the electrophoretic mobility revealed that the Click reaction product indeed carried biotin (Fig. 1c, lane 7).

To gain structural insight into how the modified pyrimidine analogs are processed by *KlenTaq* we crystallized *KlenTaq* bound to a primer/template complex and the modified dT*TP and dC*TP positioned at the dNTP binding site (*KlenTaq*(dT*) or *KlenTaq*(dC*)). To stall the respective modified substrates in the position poised for catalysis the polymerization reaction was terminated by the incorporation of ddCMP or ddGMP at the 3'-primer terminus. The structures were solved by difference Fourier techniques and provide snapshots of the incorporation of the modified triphosphates at resolutions of 2.2 Å and 1.8 Å, respectively (Table S1 and Fig. 2, S3 and S4).

In *KlenTaq*(dT*), the polymerase adopts a similar overall structure as it is observed for *KlenTaq* in complex with DNA

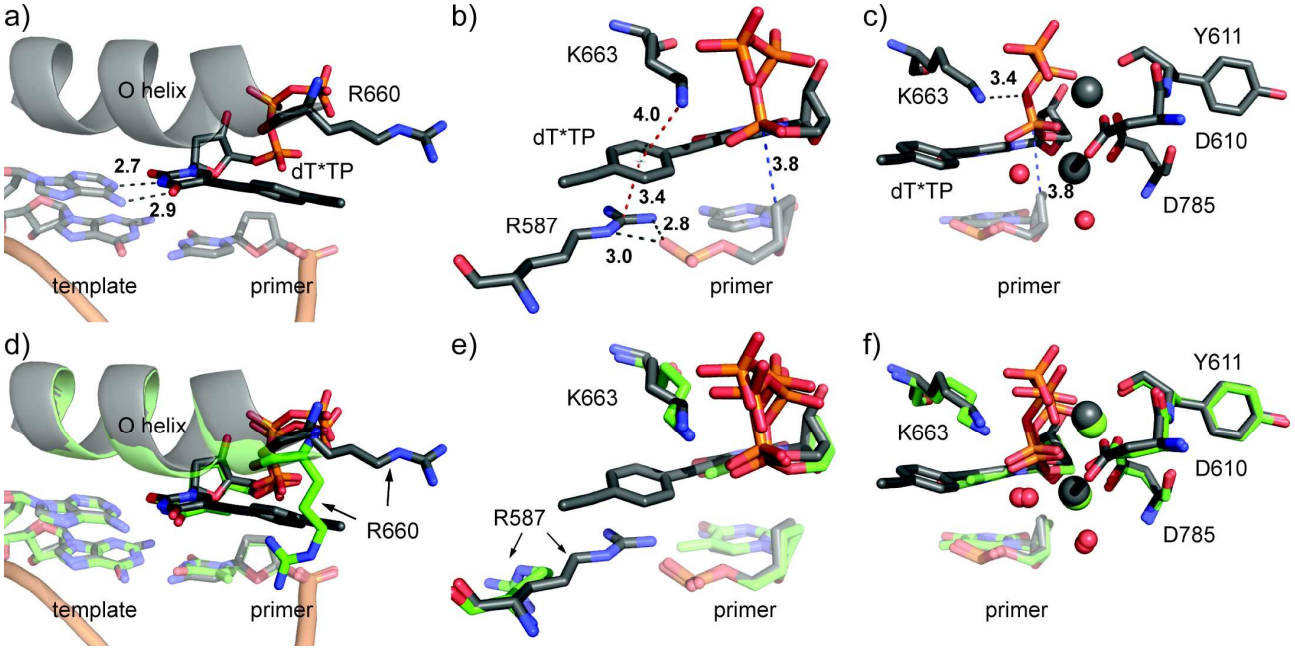


Fig. 2 *KlenTaq*(dT*) structure (gray). a) Zoom into the active site of *KlenTaq*(dT*). The incoming dT*TP is stabilized via Watson Crick base pairing. The O helix packs against the nascent base pair. R660 is indicated as stick. b) Interaction pattern of the modified dT*TP with the protein side chains R587 and K663 (red dashed lines indicate cation-π interactions). Distance of the α-phosphate to the 3'-primer terminus is shown in blue dashed line. Gray dashed lines indicate further interaction pattern. c) Close-up view of the complexation of Mg^{2+} ions by the incoming dT*TP and the amino acids responsible for catalysis. d) – f) Shown are the same cut-outs and orientations as in a) – c) for *KlenTaq*(dT*) (gray) superimposed with *KlenTaq*(1QTM) (green). All distances are in A□.

and ddTTP (PDB: 1QTM; *KlenTaq*(1QTM)), resulting in a low r.m.s.d. for Cα of 0.417 A□. In analogy to *KlenTaq*(1QTM) the recognition of dT*TP relies on Watson–Crick base pairing and π-stacking interaction to the primer strand. Further all components required for efficient catalysis are properly arranged, except for two amino acids Arg660 and Arg587. Arg660 is positioned to make room for the rigid modification, thereby disabeling interactions with the 3'-primer terminus, which are observed in the unmodified case (Fig. 2a,d). Similar effects were observed earlier in case of another rigid non-polar modification that is processed about 2500-fold less effective by *KlenTaq*[7a]. However, in the present study the incoming dT*TP is further stabilized by Arg587. The guanidinium group of Arg587 is located above the aromatic ring of the modification allowing cation-π interaction (Fig. 2b). Further Arg587 stabilizes the phosphate backbone of the 3'-primer terminus. In addition, Lys663 is located on the other side of the aromatic ring also forming cation-π interactions. Thus the aromatic ring is sandwiched between two positively charged amino acid side chains (Arg587 and Lys663). While the C5 modification at the nucleobase does not induce a conformational change of Lys663, it clearly affects the conformation of Arg587 (Fig. 2e). The distance from the primer 3′-terminus to the α-phosphate of dT*TP is similar to the one observed in *KlenTaq*(1QTM) underpinning that dT*TP is properly aligned for catalysis (Fig. 2c,f). This interaction pattern may account for the proficient processing of dT*TP by *KlenTaq*.

The obtained structure of *KlenTaq* capturing dC*TP (*KlenTaq*(dC*)) shows very similar features as observed for *KlenTaq*(dT*) highlighted by the superimposed structures (Fig. S4). Once more, the incoming dC*TP is properly aligned and interacts via Watson-Crick base pairing with the templating G. In analogy to *KlenTaq*(dT*), dC*TP is additionally stabilized by cation-π interaction with Arg587 (Fig. S4b). The assembly of the catalytic residues with the required metal ions are very similar to *KlenTaq*(dT*) and consequently very alike *KlenTaq*(1QTM) indicating that dC*TP is properly positioned (Fig. S4c).

In summary, the obtained structures of *KlenTaq* capturing modified substrates in the active site illustrate the plasticity of the enzyme and highlights the importance of the flexible regions of the dNTP binding pocket defined in the model of "active site tightness"[13]. Attaching the modification at the C5 atom of the pyrimidines has two major advantages. Firstly, modifications at this position cause minor disruption of the DNA duplex, since it points to the major groove. Secondly, it also results in minor disruption of the enzyme's ternary complex, since *KlenTaq* possesses sufficient plasticity to adapt to the structure of the modified substrates. This fact underpins the idea that the enzyme is able to respond to the incoming nucleotide by interacting with the modified substrates. Previous structural studies of flexible and polar modified substrates revealed hydrogen-bonding capability might improve their substrate properties[7]. The present study focusing on rigid and non-polar modifications suggest that the introduction of an aromatic ring enables new interactions as cation-π interaction to positively charged amino acid side chains like arginine or lysine. This fact might explain the efficient processing of dT*TP and dC*TP. For instance, the earlier employed thymidine anlog[7a] that bears a rigid non-polar modification lacking an aromatic ring is less effectively processed as dT*TP (i.e. 7- versus 2500-fold reduction of incorporation efficiency compared to dTTP). In line with this findings, Burgess and coworkers demonstrated that dyes attached to the nucleobase via a rigid conjugate linkage show enhanced spectroscopic properties and their substrate properities increase with the linker bearing an aromatic ring[14]. These insights and the present study suggest that implementing an aromatic ring at the discussed position in modified dNTPs may improve their substrate properties. This enhancement of design guidelines for the development of new modified dNTPs in combination with directed evolution of DNA polymerases[15] will stimulate the development of future applications.

Notes and references

1 (a) D. R. Bentley, et al., *Nature* 2008, **456**, 53; (b) J. Bowers, et al., *Nat. Methods* 2009, **6**, 593.

2 (a) H. Ruparel, L. Bi, Z. Li, X. Bai, D. H. Kim, N. J. Turro, J. Ju, *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**, 5932; (b) T. D. Harris, et al., *Science* 2008, **320**, 106.

3 (a) J. Guo, et al., *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**, 9145; (b) J. R. Pollack, C. M. Perou, A. A. Alizadeh, M. B. Eisen, A. Pergamenschikov, C. F. Williams, S. S. Jeffrey, D. Botstein, P. O. Brown, *Nat. Genet.* 1999, **23**, 41; (c) A. Baccaro, A.-L. Steck, A. Marx, *Angew. Chem. Int. Ed.* 2012, **51**, 254.

4 C. M. Perou, et al., *Nature* 2000, **406**, 747.

5 T. S. Seo, X. Bai, D. H. Kim, Q. Meng, S. Shi, H. Ruparel, Z. Li, N. J. Turro, J. Ju, *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**, 5926.

6 (a) S. H. Weisbrod, A. Marx, *Chem. Commun.* 2007, 1828; (b) G. Mayer, *Angew. Chem., Int. Ed. Engl.* 2009, **48**, 2672; (c) O. Thum, S. Jäger, M. Famulok, *Angew. Chem. Int. Ed.* 2001, **40**, 3990; (d) H. Weizman, Y. Tor, *J. Am. Chem. Soc.* 2002, **124**, 1568; (e) G. Giller, T. Tasara, B. Angerer, K. Mühlegger, M. Amacker, H. Winter, *Nucleic Acids Res.* 2003, **31**, 2630; (f) S. Jäger, M. Famulok, *Angew. Chem. Int. Ed.* 2004, **43**, 3337; (g) J. P. Anderson, B. Angerer, L. A. Loeb, *BioTechniques* 2005, **38**, 257; (h) T. Ohbayashi, M. Kuwahara, M. Hasegawa, T. Kasamatsu, T. Tamura, H. Sawai, *Org. Biomol. Chem.* 2005, **3**, 2463; (i) S. Jäger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum, M. Famulok, *J. Am. Chem. Soc.* 2005, **127**, 15071; (j) S. G. Srivatsan, Y. Tor, *J. Am. Chem. Soc.* 2007, **129**, 2044; (k) P. Čapek, H. Cahová, R. Pohl, M. Hocek, C. Gloeckner, A. Marx, *Chem.--Eur. J.* 2007, **13**, 6196; (l) P. M. E. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem., Int. Ed.* 2008, **47**, 8350; (m) M. Hocek, M. Fojta, *Org. Biomol. Chem.* 2008, **6**, 2233; (n) S. Obeid, M. Yulikov, G. Jeschke, A. Marx, *Angew. Chem. Int. Ed.* 2008, **47**, 6782.

7 (a) S. Obeid, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *Proc. Natl. Acad. Sci. U. S. A.* 2010, **107**, 21327; (b) K. Bergen, A. L. Steck, S. Strutt, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *J. Am. Chem. Soc.* 2012 doi: 10.1021/ja3017889 (available online).

8 A. H. El-Sagheer, T. Brown, *Chem Soc Rev* 2010, **39**, 1388.

9 R. Chinchilla, C. Najera, *Chem. Rev.* 2007, **107**, 874.

10 J. Gaster, A. Marx, *Chemistry* 2005, **11**, 1861.

11 T. Kovács, L. Ötvös, *Tetrahedron Lett.* 1988, **29**, 4525.

12 A. Baccaro, A. Marx, *Chem.--Eur. J.* 2010, **16**, 218.

13 E. T. Kool, *Annu. Rev. Biochem.* 2002, **71**, 191.

14 L. H. Thoresen, G.-S. Jiao, W. C. Haaland, M. L. Metzker, K. Burgess, *Chem.--Eur. J.* 2003, **9**, 4603.

15 (a) A. M. Leconte, et al., *Angew. Chem. Int. Ed.* 2010, **49**, 5921; (b) N. Ramsay, A.-S. Jemth, A. Brown, N. Crampton, P. Dear, P. Holliger, *J. Am. Chem. Soc.* 2010, **132**, 5096; (c) N. Staiger, A. Marx, *ChemBioChem* 2010, **11**, 1963.