Erschienen in: Angewandte Chemie International Edition ; 56 (2017), 39. - S. 12000-12003
https://dx.doi.org/10.1002/anie.201704190

Structural Basis for Expansion of the Genetic Alphabet with an Artificial Nucleobase Pair

Karin Betz, Michiko Kimoto, Kay Diederichs, Ichiro Hirao, and Andreas Marx**

Abstract: *Hydrophobic artificial nucleobase pairs without the ability to pair through hydrogen bonds are promising candidates to expand the genetic alphabet. The most successful nucleobase surrogates show little similarity to each other and their natural counterparts. It is thus puzzling how these unnatural molecules are processed by DNA polymerases that have evolved to efficiently work with the natural building blocks. Here, we report structural insight into the insertion of one of the most promising hydrophobic unnatural base pairs, the dDs–dPx pair, into a DNA strand by a DNA polymerase. We solved a crystal structure of KlenTaq DNA polymerase with a modified template/primer duplex bound to the unnatural triphosphate. The ternary complex shows that the artificial pair adopts a planar structure just like a natural nucleobase pair, and identifies features that might hint at the mechanisms accounting for the lower incorporation efficiency observed when processing the unnatural substrates.*

Expanding the genetic alphabet with an unnatural base pair increases the functional diversity of nucleic acids, and with this, the biological and biotechnological scope of potential applications. Besides nucleobase pairs with hydrogen-bonding patterns and structures different to the natural A–T and G–C pairs,[1] hydrophobic nucleobase surrogates that rely on hydrophobic and packing interactions have been shown to be able to selectively pair with each other and are therefore candidates to generate a third base pair.[2] Their main advantage is lower mispairing propensity with the four natural nucleotides since they lack appropriate H-bonding groups. Most prominent in this respect are the pairs between 2-methoxynaphthalene (NaM) and 6-methyl-2*H*-isoquinoline-1-thione (5SICS) or thieno[2,3-*c*]pyridine-7(6*H*)-thione (TPT3), as developed in the Romesberg group,[3] as well as the pair between 7-(2-thienyl)imidazo[4,5-*b*]pyridine (Ds) and 2-nitro-4-propynylpyrrole (Px) introduced by Hirao and coworkers[4–7] (Figure 1). The later analogues were used in the generation of high-affinity DNA aptamers containing natural and unnatural nucleotides by systematic evolution of ligands by exponential enrichment (SELEX).[8] The use of the unnatural base pair surrogate dDs–dPx increased the chemical and structural diversity of the DNA libraries used, resulting in improved selectivity and binding affinity. Further, a first post-transcriptional modification method for RNA transcripts using an expanded genetic alphabet was presented by using PCR-amplified DNA templates with the dDs–dPx pair.[9] Key for the success of these approaches is that the unnatural base pair is preserved beside the natural ones during enzymatic synthesis.

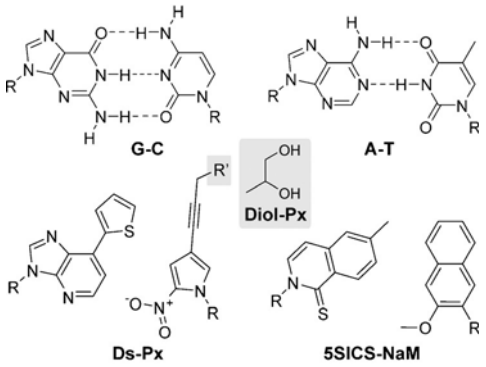

Figure 1. Structures of the natural base pairs and two successful hydrophobic artificial base pairs. *R* = 2′-deoxyribose, R′ can be different functional groups,for example, the diol functionality displayed in the gray box, which was used in this study.

The base pair surrogate dDs–dPx shows little similarity to either dNaM–d5SICS or the natural nucleobase pairs. It has been evolved over several screening rounds for efficiency and selectivity in replication.[4,7] The nitro group of dPx is crucial. First, it is believed to enable direct or water-mediated minor-groove interaction with the polymerase.[5,7] Second, it prevents mispairing with dA, thereby resulting in increase in selectivity.[5] One advantage of this base pair is that a variety of functionalities for site-specific labeling can be introduced at position 4 of the dPx base via a linker (gray box in Figure 1) without affecting the replication efficiency.[10,11] The best pairing partner for dDs is the dihydroxy derivative of dPx,[10] which might allow further derivatization through Schiff base formation. The obvious structural dissimilarity of the dDs–dPx pair to the natural nucleobase pairs raises the question of how these analogues are processed by DNA polymerases. Up to now, structural data has only been available for the dNaM–

[*] Dr. K. Betz, Prof. Dr. K. Diederichs, Prof. Dr. A. Marx
Departments of Chemistry and Biology and Konstanz Research School Chemical Biology, Universität Konstanz
Universitätsstrasse 10, 78464 Konstanz (Germany)
E-mail: andreas.marx@uni-konstanz.de
Dr. M. Kimoto, Prof. Dr. I. Hirao
Institute of Bioengineering and Nanotechnology
31 Biopolis Way, The Nanos, #09-01, 138669 Singapore (Singapore)
and
RIKEN Center for Life Science Technologies
1–7 22 Suehiro-cho, Tsurumi-ku, Yokohama
Kanagawa 230-0045 (Japan)
E-mail: ichiro@ibn.a-star.edu.sg

d5SICS base pair surrogate in the confines of a DNA polymerase, specifically, the large fragment of the *Thermus aquaticus* DNA polymerase (KlenTaq).[12,13] These data indicate that the enzyme forces the unnatural base pair into a coplanar conformation when the triphosphate encounters the unnatural template, thereby resolving stacking interactions that are found in a free DNA duplex.[14] In this work, we present structural data for KlenTaq processing the unnatural base pair dDs–dPxTP (note: Diol-dPxTP was used in this study).

To obtain the structure, we followed a similar strategy as before.[12] For crystallization, KlenTaq was incubated with an annealed natural primer and a dDs-containing template, as well as ddCTP to terminate the primer strand. Crystals were grown and afterwards soaked with dPxTP, thereby resulting in the structure KlenTaq$_{\mathbf{dDs\text{-}dPxTP}}$. The structure was solved using data up to a resolution of 1.7 Å (see Methods and Table S1 in the Supporting Information).

The overall structure of KlenTaq$_{\mathbf{dDs\text{-}dPxTP}}$ with its four domains (N-terminal, palm, thumb, and finger domains; Figure 2A) is similar to the natural ternary complex KlenTaq$_{\mathrm{dG\text{-}dCTP}}$ (PDB ID: 3RTV[12]) and to KlenTaq$_{\mathbf{dNaM\text{-}d5SICSTP}}$ (PDB ID: 3SV3[12]), with root-mean-square deviation (rmsd) values for Cα atoms of only 0.185 Å (446 atoms aligned, 39 atoms rejected) and 0.173 Å (442 atoms aligned, 91 atoms rejected), respectively. All enzyme residues could be traced, but the region of the finger domain between residues 645 and 700 is less well resolved and shows higher flexibility than the rest of the enzyme (Figure S2A,D). The finger domain is in a closed conformation and the dDs–dPxTP pair is situated in the active site (Figure 2). The base pair could be unambiguously placed into the difference electron density map after initial refinement without it. All functionalities are well defined, and only the electron density at the diol moiety of the triphosphate is somehow extended. This can be explained by the rotational flexibility of the terminal hydroxy group on the one hand, but also by the presence of two epimers (*R* or *S* configuration at the diol) of dPxTP on the other. Both isomers were modeled at the same position and refined to occupancies of 0.40 and 0.46 for the *S* and *R* isomers, respectively, and the pyrrole, ribose, and triphosphate parts of the ligand superpose perfectly. The ribose moieties of the triphosphate and the templating dDs adopt C3′-*endo* conformations identical to the natural pair. Coordination of two magnesium ions by the triphosphate and Asp610, Asp785, and the backbone of Tyr611 characterizes an active closed complex (Figure 2C). The distance between the primer 3′ end (C3′ used for measuring) and the α-phosphate is almost identical to the natural case (3.8 Å vs. 3.9 Å). In addition to metal coordination, the triphosphate interacts with the side chains of Lys663, Arg659, and His639, and the backbone of Gln613, whereby it is well stabilized at its position. A water-mediated minor-groove interaction (in the natural case mediated by O2 atoms of pyrimidines or N3 atoms of purines) is formed between the nitro group of dPxTP and Asn750, Gln754, and Glu615. Finally, the dDs–dPxTP pair does not intercalate but pairs in a coplanar manner strikingly similar to the cognate Watson–Crick pairs (Figure 3). Even the propeller twist (relative rotation between bases within a base pair with respect to the base pairing axis) is similar between dDs–dPxTP and dG–dCTP (−10.5° versus −9.2°, Figure 3A and S3; determined by use of the 3DNA server[15]). This similarity was unexpected since the previously analyzed hydrophobic dNaM–d5SICSTP pair shows some differences in this regard (discussed below). Relative to the natural dG–dCTP pair, the unnatural dDs–dPxTP pair shows a larger overall base-pair width (C1′–C1′ distance between the pairing partners is 0.7 Å longer; Figure 3B). The shift takes place only on the side of the templating nucleotide, while the substrate triphosphate stays at its well-defined position just described. Amino acids interacting with the templating nucleotide (Arg677, Ser674, and Met673) are also slightly shifted (Figure 3D), following the movement of the template. Not only the width but also the height of the base pair is different to the cognate pairs. The thiophenyl as well as the propynyldiol moieties point towards the O-helix of the finger domain. This causes a small shift of the overall O-helix plus the connected helices and a different orientation of some amino acid side chains. The effect is pronounced for Thr664, which comes closest to the thiophenyl moiety and shifts upwards by 0.6 Å. The extent of the displacements of the Cα atoms along the O-helix is shown in detail in Figure S3D. Along with these shifts, the flexibility of the entire finger domain and also the newly formed base pair is higher than in the natural case, as indicated by elevated B-factors (Figure S2A–C). Arg660 makes space for the propynyldiol moiety and adopts a position where it may interact with one of the ligand OH groups through a hydrogen bond (Figure 3D). Its flexibility, however, indicates that the interaction cannot be strong in the present state. Similar shifts of Arg660 have already been found in other KlenTaq structures with modified substrates, and this displacement seems not to diminish incorporation efficiency.[16,17] All in all, the increased height of the pair only affects enzyme residues on the major-groove side of the pair, whereas the minor-groove side stays unperturbed.

Figure 2. The artificial base pair dDs–dPxTP pairs in an edge-to-edge manner in the active site of KlenTaq and shows the same interactions with the enzyme and catalytic ions as a natural base pair. A) Overall structure of KlenTaq (domains colored as indicated in the Figure) with bound DNA (dark gray) and substrate. B) dDs–dPxTP pair surrounded by its simulated annealing mFo-DFc omit map contoured at 3σ. The artificial templating dDs is shown in dark green, and the *R* and *S* isomers of dPxTP are shown in blue-cyan and green-cyan, respectively C) Interactions of the triphosphate dPxTP with the enzyme. Magnesium ions are shown in green and water molecules are shown in white.

Figure 3. Comparison of dDs–dPxTP with the natural dG-dCTP pair and the unnatural dNaM–d5SICSTP pair in KlenTaq. A) Base-pair propeller twist visualized with the templating nucleobase aligned orthogonal to the paper plane (ribose and triphosphate moieties are not shown). B) Relative position of the base pairs when the whole enzyme complex is superposed. C1′–C1′ distances between the pairing partners are indicated C) Shifts of the residues surrounding the artificial pair are visualized. The finger domain O-helix is shown as a ribbon diagram.

The previously analyzed hydrophobic dNaM–d5SICSTP pair also pairs edge-to-edge in the active site of KlenTaq as mentioned above. However, in contrast to the pairs dG–dCTP and dDs–dPxTP, the propeller twist is significantly smaller (2.4°, Figure 3 A and Figure S3), and the pair exhibits a larger relative shift of the bases along the *z* axis (stagger = −1.2 Å) which is zero for dG–dCTP and only −0.12 Å for dDs–dPxTP (Figure S3). These two base-pair parameters have not been discussed previously when comparing the natural pair dG–dCTP with the hydrophobic artificial base pair dNaM–d5SICSTP. With the new structure KlenTaq$_{\mathbf{dDs\text{-}dPxTP}}$, this unapparent difference between the two unnatural base pair surrogates in relation to the natural pair was identified.

A similarity between the two hydrophobic base pairs is increased distance between the pairing partners (Figure 3 B). Additionally, for dNaM–d5CISTP, the C1′–C1′ distance is larger (11.0 Å vs. 10.6 Å in the dG-dCTP case; Figure S1A). The residues around the templating nucleotide are even shifted a bit more compared to KlenTaq$_{\mathbf{dDs\text{-}dPxTP}}$ owing to the different nucleobase structure (Figure S1B). It seems that if an artificial pair with elevated size forms in the active site of KlenTaq, the enzyme allows extension to one side of the binding pocket (the template side) but not the other. This behavior might assist the enzyme in aligning the triphosphate substrate perfectly for proper attack by the 3′-OH group of the primer end, although the pair shows some alterations from the natural consensus structure.

Along with the shift on the template side, an O-helix shift similar to that described for KlenTaq$_{\mathbf{dDs\text{-}dPxTP}}$ was found in the KlenTaq$_{\mathbf{dNaM\text{-}d5SICSTP}}$ structure. In both cases, the O-helix is closed far enough that the substrate and other components are well aligned for catalysis.

A recent report featuring a modelled dDs–dPx base pair in the active state of Deep Vent DNA polymerase, which is among the most proficient polymerases in recognizing the dDs–dPx pair with high fidelity in replication,[10] proposes accommodation of the unnatural base pair in the polymerase.[18] In contrast, modelling the artificial pair into the active site of a KlenTaq ternary structure suggested that the side-chain oxygen atom of Thr664 clashes the templating dDs base.[18] Indeed, this residue in the O-helix of the finger domain is the one that comes closest to the unnatural pair, but our structural data indicates that the enzyme is flexible enough to accommodate the base pair and adapt to its structure. The described shift and elevated B-factors of the O-helix and other parts of the finger domain (see Figure S2), however, indicate that the O-helix does not close as tightly as with a cognate pair. This small difference could provide an explanation for why the two most successful hydrophobic artificial pairs are still formed with somewhat diminished efficiency compared to the natural counterparts. The perfect arrangement of components taking part in catalysis might be slightly changed in one or several steps during insertion. Modulating the size of the pairs towards the O-helix or mutating the enzyme in this region might provide opportunities for further optimization.

One of the features of the dPx nucleobase is the fact that it can be modified with different functional groups. The linker moiety of dPxTP is placed in such a way that the modification can leave the enzyme through two possible channels, as was shown before for C5-modified deoxyuridines and C7-modified 7-deaza-adenosines in KlenTaq.[16,17,19] (The data in Table 2 of Ref. [10] show the side-chain dependency for the replication efficiency, in which large side chains reduced the efficiency).

In conclusion, we provide herein the structural basis for expansion of the genetic alphabet with the dDs–dPxTP pair. The data show how the artificial base pair is well accepted in the active site of KlenTaq by pairing edge-to-edge just like the natural pairs. This is consistent with previous observations with the artificial base pair dNaM–d5SICSTP. Compared to dNaM–d5SICSTP, which pairs in an almost coplanar manner, the dDs–dPxTP pair adopts a similar propeller twist to the natural pair. Both unnatural pairs exhibit increased base pair width and a different height compared to the natural counterparts. KlenTaq has some flexibility to adapt to their slightly increased height and width. While the triphosphate and the minor-groove side of the pair are perfectly aligned in the active site, changes in width and height are transferred to the side of the templating nucleotide and the O-helix. With

both pairs, the O-helix is not as tightly closed as in the natural case, and the resulting small changes in positions may explain the still lower incorporation efficiency of the artificial substrates. Therefore, the size of the pairs in this direction, as well as residues in the O-helix, are suitable targets for further modification and improvement.

Our results again confirm that hydrophobic base pairs are well suited to expand the genetic alphabet as long as they can adopt a similar shape to the cognate pairs and match the constraints of the active sites of the respective enzymes. Furthermore, it seems that base-pair parameters (like propeller twist) can still vary between different hydrophobic base-pair candidates. We assume that the exact pairing behavior in the active site of a DNA polymerase is difficult to predict directly from the base-pair architecture, thus rendering structural studies indispensable to further characterize future artificial base pairs for the expansion of the genomic alphabet.

Acknowledgements

We thank the beamline staff of the Swiss Light Source at the Paul Scherrer Institute for access and help at the beamline. We also thank the Konstanz Research School Chemical Biology (K.B.) and the Japan Science and Technology Agency, Precursory Research for Embryonic Science and Technology (JPMJPR13K9) (M.K.) for financial support. A part of this work was funded by the Institute of Bioengineering and Nanotechnology (Biomedical Research Council, Agency for Science, Technology and Research, Singapore).

Keywords: artificial base pairs · DNA polymerase · enzyme catalysis · nucleobases · structure determination

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 12000–12003
Angew. Chem. **2017**, *129*, 12162–12166

[1] a) C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323; b) J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37; c) M. Ishikawa, I. Hirao, S. Yokoyama, *Tetrahedron Lett.* **2000**, *41*, 3931–3934; d) T. Fujiwara, M. Kimoto, H. Sugiyama, I. Hirao, S. Yokoyama, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2221–2223; e) I. Hirao, Y. Harada, M. Kimoto, T. Mitsui, T. Fujiwara, S. Yokoyama, *J. Am. Chem. Soc.* **2004**, *126*, 13298–13305; f) Z. Yang, D. Hutter, P. Sheng, A. M. Sismour, S. A. Benner, *Nucleic Acids Res.* **2006**, *34*, 6095–6101; g) N. Tarashima, Y. Komatsu, K. Furukawa, N. Minakawa, *Chem. Eur. J.* **2015**, *21*, 10688–10695.

[2] a) D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586; b) E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191–219; c) T. Mitsui, A. Kitamura, M. Kimoto, T. To, A. Sato, I. Hirao, S. Yokoyama, *J. Am. Chem. Soc.* **2003**, *125*, 5298–5307.

[3] a) Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 3246–3252; b) T. Lavergne, D. A. Malyshev, F. E. Romesberg, *Chem. Eur. J.* **2012**, *18*, 1231–1239; c) L. Li, M. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2014**, *136*, 826–829; d) Y. Zhang, B. M. Lamb, A. W. Feldman, A. X. Zhou, T. Lavergne, L. Li, F. E. Romesberg, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 1317–1322..

[4] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*(9), 729–735.

[5] I. Hirao, T. Mitsui, M. Kimoto, S. Yokoyama, *J. Am. Chem. Soc.* **2007**, *129*, 15549–15555.

[6] M. Kimoto, R. Kawai, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2009**, *37*, e14.

[7] I. Hirao, M. Kimoto, R. Yamashige, *Acc. Chem. Res.* **2012**, *45*, 2055–2065.

[8] a) M. Kimoto, R. Yamashige, K. Matsunaga, S. Yokoyama, I. Hirao, *Nat. Biotechnol.* **2013**, *31*, 453; b) K. Matsunaga, M. Kimoto, C. Hanson, M. Sanford, H. A. Young, I. Hirao, *Sci. Rep.* **2015**, *5*, 18478; c) M. Kimoto, M. Nakamura, I. Hirao, *Nucleic Acids Res.* **2016**, *44*, 7487–7494; d) K. Matsunaga, M. Kimoto, I. Hirao, *J. Am. Chem. Soc.* **2017**, *139*, 324–334.

[9] T. Someya, A. Ando, M. Kimoto, I. Hirao, *Nucleic Acids Res.* **2015**, *43*, 6665–6676.

[10] R. Yamashige, M. Kimoto, Y. Takezawa, A. Sato, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2012**, *40*, 2793–2806.

[11] I. Okamoto, Y. Miyatake, M. Kimoto, I. Hirao, *ACS Synth. Biol.* **2016**, *5*, 1220–1230.

[12] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, P. Ordoukhanian, F. E. Romesberg, A. Marx, *Nat. Chem. Biol.* **2012**, *8*, 612–614.

[13] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, F. E. Romesberg, A. Marx, *J. Am. Chem. Soc.* **2013**, *135*, 18637–18643.

[14] D. A. Malyshev, D. A. Pfaff, S. I. Ippoliti, G. T. Hwang, T. J. Dwyer, F. E. Romesberg, *Chem. Eur. J.* **2010**, *16*, 12650–12659.

[15] X.-J. Lu, W. K. Olson, *Nat. Protoc.* **2008**, *3*, 1213–1227.

[16] K. Bergen, A.-L. Steck, S. Strütt, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *J. Am. Chem. Soc.* **2012**, *134*, 11840–11843.

[17] A. Hottin, K. Betz, K. Diederichs, A. Marx, *Chem. Eur. J.* **2017**, *23*, 2109–2118.

[18] Y. Hikida, M. Kimoto, I. Hirao, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2017**, *483*, 52–57.

[19] A. Hottin, A. Marx, *Acc. Chem. Res.* **2016**, *49*, 418–427.