Published in final edited form as:

ACS Synth Biol. 2017 October 20; 6(10): 1834–1840. doi:10.1021/acssynbio.7b00115.

Synthetic Biology Parts for the Storage of Increased Genetic Information in Cells

Sydney E. Morris, Aaron W. Feldman, and Floyd E. Romesberg

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Abstract

To bestow cells with novel forms and functions, the goal of synthetic biology, we have developed the unnatural nucleoside triphosphates dNaMTP and dTPT3TP, which form an unnatural base pair (UBP) and expand the genetic alphabet. While the UBP may be retained in the DNA of a living cell, its retention is sequence-dependent. We now report a steady-state kinetic characterization of the rate with which the Klenow fragment of *E. coli* DNA polymerase I synthesizes the UBP and its mispairs in a variety of sequence contexts. Correct UBP synthesis is as efficient as for a natural base pair, except in one sequence context, and *in vitro* performance is correlated with *in vivo* performance. The data elucidate the determinants of efficient UBP synthesis, show that the dNaM-dTPT3 UBP is the first generally recognized as a natural base pair, and importantly, demonstrate that dNaMTP and dTPT3TP are well optimized and standardized parts for the expansion of the genetic alphabet.

Graphical abstract

*Corresponding Author, Tel: +1 (858) 784-7290. Fax: +1 (858) 784-7472. floyd@scripps.edu.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/xxxxx.

Supporting methods, tables, and figures (PDF)

Author Contributions

SEM performed the kinetics assays; AWF synthesized and characterized dMF2, dMMS2, dMF2TP, and dMMS2TP. All authors contributed to the design of the experiments and the preparation of the manuscript.

The authors declare no competing financial interests.

Keywords

unnatural base pair; hydrophobic; DNA

Synthetic biology seeks to impart organisms with new forms and functions through the development of “parts” that may be mixed and matched within living cells to afford them with novel activities such as the ability to synthesize molecules not found in nature.^{1–4} Central to the approach is parts optimization and standardization; optimization for efficient performance, and standardization for efficient performance in different contexts without the need for additional optimization. The parts most commonly employed are borrowed from or inspired by nature.^{5,6} Although parts borrowed from nature can be repurposed to access a broad range of functions, their ability to create truly novel function is likely inherently limited by constraints acquired during evolution, for example, stop codons are recognized by release factors and rare codons are used to regulate transcription.⁷ An alternative strategy is to use parts that are synthetic in the truest sense in that they are not found in nature, but are instead the products of *de novo* design and synthesis. Such synthetic parts are free from the constraints acquired during evolution and free to draw upon forces unlike those employed by their natural counterparts, for example, the complementary hydrogen bonds might be replaced by hydrophobic and packing forces, potentially imparting increased orthogonality. However, such truly synthetic parts will likely require more optimization and standardization, as they do not benefit from eons of evolution for at least a similar function.

There is no more fundamental approach to the creation of cells with new forms and functions than the development of an unnatural base pair (UBP)^{8–10} that can be used to expand the genetic alphabet and to serve as the basis of semi-synthetic organisms (SSOs) that store and retrieve increased information. In this context, the parts are the unnatural triphosphates; their optimization is measured by DNA polymerase recognition; and their standardization is measured by their recognition in different sequence contexts. Although different metrics of recognition are possible, a particularly useful metric is the steady-state rate at which DNA polymerases synthesize the UBP (*e.g.* incorporation of an unnatural triphosphate opposite its cognate unnatural nucleotide in a template). While this is a complex series of steps that includes substrate binding, phosphoryl transfer, product release, and intervening conformational changes,¹¹ the observed rate under steady-state conditions is determined only by the slow step. This complexity is actually what makes the metric useful, because under steady-state conditions DNA polymerases recognize natural nucleotides in a binary fashion – the rates of correct pair synthesis are sufficiently fast that the observed rate corresponds to product dissociation, while the rate of mispair synthesis is sufficiently slow that the observed rate corresponds to the actual phosphoryl transfer.¹¹ Thus, an optimized UBP should be synthesized with a rate that is similar to a correct natural pair and the rate of mispairing should be variable and significantly slower; and a standardized UBP should be recognized in this binary fashion in different sequence contexts.

Towards the goal of expanding the genetic alphabet and the creation of SSOs with novel forms and functions, we have developed a UBP composed of the synthetic nucleotides dNaM and dTPT3 (dNaM-dTPT3; Figure 1A).⁶ In contrast to the natural nucleotides, the

pairing of these unnatural nucleotides relies on hydrophobic and packing forces as opposed to complementary hydrogen bonding. While the development of this UBP required years of optimization,⁸ the completely unnatural mode of pairing between the nucleobases of the UBP is expected to bestow them with an inherent orthogonality, with respect to their mispairing with the natural nucleotides and to their recognition by proteins that recognize the natural triphosphates (including the many proteins that interact with the triphosphates outside of information storage and retrieval). In 2014, with the aid of a nucleoside triphosphate transporter to import the unnatural triphosphates, we showed that a precursor of the dNaM-dTPT3 UBP may be retained within the DNA of growing and dividing *E. coli*,¹² and more recently, with the use of SpCas9 directed to eliminate sequences that have lost the UBP, we showed that dNaM-dTPT3 is retained in the DNA of the SSO with fidelity approaching that of a natural base pair.¹³ However, retention was observed to be sequence-dependent and the extent of optimization and standardization of dNaMTP and dTPT3TP remained to be determined. In fact, to this date no UBP has been reported to be recognized by a DNA polymerase in the same binary manner that a natural base pair is recognized, although a large amount of structure-activity relationship (SAR) data has determined that the substituent *ortho* to the glycosidic bond plays a particularly critical role. While the methoxy and thioamide moieties of dNaMTP and dTPT3TP appear to be suitable *ortho* substituents, whether they are optimal has also not been thoroughly examined.

We now report a steady-state kinetics characterization of the rate with which the Klenow fragment of *E. coli* DNA polymerase I (Kf) synthesizes the dNaM-dTPT3 UBP and its mispairs in primer-template substrates that correspond to the synthesis of the UBP, by either insertion of dNaMTP opposite dTPT3 or dTPT3TP opposite dNaM, in three different sequence contexts (Table 1 and 2). Two of the selected sequence contexts, contexts I and II with flanking dA-dT pairs are well replicated in our SSO, and one, with flanking dG-dC pairs, is replicated poorly (context III).¹³ In addition, we examine the contribution of the *ortho* methoxy substituent of dNaMTP. The data reveal that with five of the six substrates, the UBP is recognized by Kf in the same binary manner that the polymerase evolved to recognize a natural base pair, that it likely does so at least in part due to packing between the nucleobase of the incoming triphosphate and the primer terminal nucleobase, and that the *ortho* substituent of dNaMTP optimizes these interactions via specific packing and electrostatic interactions. Importantly, the single exception where UBP synthesis is less efficient than a natural counterpart is the insertion of dNaMTP opposite dTPT3 in sequence context III, which likely helps to explain why UBP retention in this sequence context is problematic in our SSO. Regardless of the underlying mechanism, we demonstrate that dNaM-dTPT3 is the first UBP synthesized more efficiently than any mispair, and notwithstanding the single exception, the natural-like recognition of the UBP reveals that, at least for this step of replication, dNaMTP and dTPT3TP are generally optimized and standardized parts for the expansion of the genetic alphabet.

Results and Discussion

To explore the extent to which dNaMTP and dTPT3TP are optimized for the expansion of the genetic alphabet, we characterized the efficiencies (second order rate constant, k_{cat}/K_M)

and fidelities ($(k_{\text{cat}}/K_M)_{\text{correct}}/(k_{\text{cat}}/K_M)_{\text{incorrect}}$) with which Kf inserts dNaMTP or dTPT3TP opposite its cognate unnatural nucleotide in a DNA template. To explore their standardization, we characterized the efficiency and fidelity of UBP synthesis in six primer-template substrates that correspond to three sequences in which the UBP is well replicated in our SSO (contexts I and II) or poorly replicated (context III). We first explored the insertion of dTPT3TP opposite dNaM (Table 1). We found this insertion to proceed with an efficiency of $1.0 \times 10^9 \text{ M}^{-1}\text{min}^{-1}$, $1.3 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, and $1.7 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, in contexts I, II, and III, respectively. The rates for dATP insertion opposite dT in the same three sequence contexts are $3.4 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, $8.4 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, and $4.9 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$. Thus, relative to the insertion of dATP opposite dT, the insertion of dTPT3TP opposite dNaM is only 3- to 7-fold less efficient in contexts II and III, but 3-fold more efficient in context I. While the insertion in context I is the first example of a UBP that is synthesized more efficiently than a natural base pair, all of the rates are sufficiently efficient and similar to a natural base pair that they are likely determined by product dissociation.

We next characterized the insertion of dNaMTP opposite dTPT3 in sequence contexts I–III (Table 2), which we found to proceed with an efficiency of $1.1 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, $1.3 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, and $9.6 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$, respectively. The rates for dATP insertion opposite dT in the same three sequence contexts are $2.0 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, $9.4 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, and $2.6 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$. Thus, insertion of dNaMTP opposite dTPT3 is 2- to 8-fold less efficient than insertion of dATP opposite dT in sequence contexts I and II, and 27-fold less efficient in context III. Thus, insertion of dNaMTP is likely limited by product dissociation in contexts I and II, but it may be limited by phosphoryl transfer in context III.

Synthesis of the mispair between unnatural nucleotides resulting from the insertion of dNaMTP opposite dNaM, was found to proceed with an efficiency of $4.5 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$, $4.0 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$, or $<10^3 \text{ M}^{-1}\text{min}^{-1}$, in sequence contexts I–III, respectively. Of the twelve possible mispairs resulting from the insertion of a natural triphosphate in the three sequence contexts with dNaM in the template, ten are synthesized with a rate too low to detect ($<10^3 \text{ M}^{-1}\text{min}^{-1}$). Two mispairs are synthesized with detectable rates, both involving the insertion of dATP, which proceeded in context I with an efficiency of $8.0 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$, and in context II, with an efficiency of $3.9 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$.

Synthesis of the mispair between unnatural nucleotides resulting from the insertion of dTPT3TP opposite dTPT3, was found to proceed with an efficiency of $1.2 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$ in sequence context I and $<10^3 \text{ M}^{-1}\text{min}^{-1}$ in contexts II and III. Of the twelve possible mispairs resulting from the insertion of a natural triphosphate in the three sequence contexts opposite dTPT3, all were undetectable ($<10^3 \text{ M}^{-1}\text{min}^{-1}$) except for insertion of dGTP and dTTP in context I ($1.6 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$ and $4.5 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$, respectively), and dGTP and dTTP in context II ($2.6 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$ and $1.9 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$, respectively).

Thus, of the six possible unnatural mispairs, two are synthesized ~30-fold less efficiently than the correct UBP, one is synthesized 94-fold less efficiently, and three are synthesized greater than 10^5 -fold less efficiently. Of the 24 possible mispairs with a natural nucleotide, 18 are synthesized more than 10^5 -fold less efficiently than the UBP, five are synthesized 240- to 700-fold less efficiently, and one is synthesized an order of magnitude less efficiently

than the UBP. This is the first example of a UBP whose synthesis is more efficient than any possible mispair, and the conservation of this selective recognition in three different sequence contexts suggests that it is general.

In five of the six primer-templates examined, the similar efficiencies observed for natural and UBP synthesis and the generally much reduced and more variable efficiencies with which the mispairs are synthesized suggest that Kf recognizes the UBP in the same binary fashion that it recognizes a natural base pair. The single exception is for the insertion of dNaMTP opposite dTPT3 in sequence context III. While the UBP is still synthesized in this context more than four orders of magnitude more efficiently than any mispair, it is synthesized significantly less efficiently (27-fold) than the natural base pair. The reduced efficiency suggests that the rate may result from slow phosphoryl transfer, and thus may be interpreted in terms of this step. Structural studies have suggested that during incorporation, the nucleobase of the incoming unnatural triphosphate forms a Watson-Crick-like structure with the templating unnatural nucleobase, with significant packing interactions provided by the nucleobase at the primer terminus (Figure 2A).^{14,15} In general, the absence of heteroatoms in the aromatic ring of dNaMTP that might interact via favorable dispersive interactions with specific regions of high electron density of the primer terminal nucleobase, may make insertion efficiency more sensitive to specific structural details,⁸ and the decreased surface area of the pyrimidine that is uniquely present in context III, may exacerbate this effect. Regardless, replication of the UBP in sequence context III is the least efficient of the three contexts examined in the SSO, and thus the data suggest that this is, at least in part, caused by the reduced efficiency of insertion of dNaMTP opposite dTPT3.

The reduced and variable rates at which the mispairs are synthesized suggest that they are also limited by phosphoryl transfer, and may again be interpreted in terms of this step. The most efficiently synthesized mispairs result from the insertion of dATP or dNaM opposite dNaM in context I and II. Unlike with correct UBP synthesis, mispairs with natural nucleobases and dNaM are unlikely to form a Watson-Crick-like structure, as the natural nucleobases are likely solvated, and this mode of edge-to-edge pairing would thus force an energetically disfavorable desolvation. In duplex DNA, the UBP, as well as other pairs of unnatural nucleotides with large aromatic nucleobases,^{8,16} form cross-strand intercalated structures that optimize packing interactions in the absence of edge-to-edge complementary hydrogen bonding (Figure 2B). The absence of complementary hydrogen bonding and the increased hydrophobicity and surface area with the adenine and naphthyl moieties, suggest that a similar and stabilizing intercalative interaction might underlie the favorable synthesis of these mispairs. In addition, contexts I and II are the only ones in which the UBP is flanked by dA-dT pairs, which are perhaps better able to structurally adjust to optimize the packing interactions within the cross-strand intercalated structure. Indeed, no mispairs are efficiently synthesized in sequence context III, where the UBP is flanked by dG-dC pairs.

To explore the contribution of the *ortho* substituent of dNaMTP to polymerase recognition, we synthesized and characterized two new analogs, dMMS2TP and dMF2TP, and also characterized the previously reported analogs, dMMO2TP and dDM5TP^{17,18} (Figure 1B and Supporting Information). The *ortho* substituent of these single ring analogs of dNaMTP vary systematically in hydrophobicity, hydrogen bonding potential, and polarizability. In the

three sequence contexts, dMMO2TP and dMMS2TP are inserted opposite dTPT3 similarly and slightly more efficiently than dMF2TP and dDM5TP (Table 2). However, each of these analogs is inserted at least 40-, 20-, and 8-fold less efficiently than dNaMTP in sequence contexts I, II, and III, respectively. The reduced efficiency with which dMMO2TP is inserted relative to dNaMTP clearly demonstrates the importance of the extended aromatic surface area of dNaMTP for polymerase-mediated UBP synthesis.

The variable and reduced efficiency with which each of these analogs is inserted again suggests that their rate of insertion, unlike dNaMTP, is limited by phosphoryl transfer, and as with the mispairs discussed above, the efficiencies may be interpreted in terms of this step. The most consistent difference between the more efficiently inserted analogs, dMMO2TP or dMMS2TP, and the less efficiently inserted analogs, dMF2TP or dDM5TP, is a higher K_M for the latter, which is most pronounced in context III. In addition, the k_{cat} for insertion of dMF2TP or dDM5TP is somewhat reduced in context I, and similarly reduced for dDM5TP in context III, which along with the elevated K_M , results in the particularly inefficient insertion in this sequence context. The general exaggeration of the effects observed for context III again suggests that a dC at the primer terminus is less suitable for UBP insertion than a dA (see above). This is also consistent with structural studies that reveal significant packing interactions between the *ortho* substituent of the incoming nucleobase and the template nucleobase of the primer-terminal base pair (Figure 2A).¹⁴ Regardless of mechanism, the data clearly indicates that a methoxy or thioether moiety is a more optimal *ortho* substituent than a methyl or fluoro moiety.

SAR studies involving many previously reported analogs led to the conclusion that the hydrophobicity of the *ortho* substituent is critical for UBP synthesis.⁸ However, the current results demonstrate a more nuanced SAR, as the methyl substituent of dDM5TP is the most hydrophobic and the fluoro substituent of dMF2 is the least polarizable. The current data suggests that polarization and dispersive interactions, which are more sensitive to the structure than the hydrophobic effects, are critical, and perhaps more critical than simple hydrophobicity. An important role for polarization and dispersive interactions is consistent with theoretical studies of the packing between natural nucleobases.⁸

Conclusion

We have characterized the steady-state rate at which Kf synthesizes the UBP in six different primer-templates, which correspond to the replication of DNA containing the UBP in three different sequence contexts that we have used for *in vivo* replication of the UBP. With five of the six primer-templates, the UBP is synthesized with an efficiency similar to that of a natural base pair. Similar small differences in efficiency are also observed with natural base pairs in different sequence contexts, which are all thought to be limited by product dissociation; this suggests that the differences result from effects of the UBP on product dissociation. In addition, 18 of the 24 possible mispairs are synthesized more than 10⁵-fold less efficiently than the dNaM-dTPT3 UBP, and of the remaining six possible mispairs, five are synthesized 240- to 700-fold less efficiently, and one is synthesized 10-fold less efficiently. Thus, Kf appears to recognize the UBP in generally the same binary fashion that it recognizes a natural base pair. While the hydrophobic effect undoubtedly contributes, the

mechanism underlying this binary recognition appears to be based on polarization and dispersive interactions between the nucleobase analog and the natural nucleobase at the primer terminus. Moreover, our results suggest that these interactions are at least in part mediated by the *ortho* substituent of the unnatural nucleotide, and that these are at least fairly well optimized with dNaMTP and dTPT3TP. The only case where these interactions may not be sufficiently favorable to impart the UBP with natural-like synthesis is the insertion of dNaMTP opposite dTPT3 in sequence context III, where the observation of a 27-fold reduced rate suggests that insertion may be limited by phosphoryl transfer. Importantly, the reduced efficiency of this specific step of UBP replication likely contributes to its reduced *in vivo* retention in sequence context III. While the synthesis of mispairs is likely facilitated by similar interactions, they appear to proceed via an intercalative mechanism that requires greater distortion of the primer terminal base pair.

Regardless of the underlying mechanisms of correct UBP and mispair synthesis, this is the first demonstration of a UBP whose synthesis is consistently more efficient than any possible mispair. Furthermore, with the single exception of insertion of dNaMTP opposite dTPT3 in sequence context III, this is to our knowledge the first truly synthetic part that is recognized by a biological system as natural-like. However, replication of DNA containing the UBP involves both its synthesis and continued primer extension, and the overall fidelity is the product of the fidelity of both steps. Previous studies with analogs of dNaM and dTPT3 have consistently shown that the extension step is more efficient for a UBP than a natural mispair.¹⁹ Thus, while a similar detailed characterization of UBP extension is still in progress, the overall fidelity of UBP replication is likely to be significantly higher than that reported here for synthesis alone, and indeed likely to approach that of a natural base pair. To retrieve the increased genetic information made possible by the dNaM-dTPT3 UBP, corresponding ribotriphosphate “parts” will need to be optimized and standardized for efficient and sequence-independent transcription and translation, which is currently in progress.

Methods

Synthesis of Unnatural Triphosphates

dNaMTP and dTPT3TP were a generous gift from Synthorx and dNaM analogs, dMMO2TP and dDM5TP, were synthesized as previously described.^{17,18} dMMS2TP and dMF2TP, were synthesized from commercially available aryl bromides according to literature methods. Briefly, lithiation of the corresponding aryl halide was followed by coupling to disiloxane-protected 2-deoxyribonolactone and the resulting hemiacetal intermediate was reduced with triethylsilane and boron trifluoride etherate to afford the corresponding protected nucleosides.²⁰ Nucleosides were then deprotected with tetra-*n*-butylammonium fluoride and phosphorylated using standard conditions²¹ to provide the desired nucleotide triphosphates. Compounds were assayed for purity by NMR and HPLC (see Figure S2 and Supporting Methods for details).

Steady-state Kinetics Assay

Steady-state kinetic reactions were performed as previously described with template and 5'-[γ - ^{32}P] labeled primer oligonucleotides (See Supporting Methods and Tables S1 and S2).²² Reactions were analyzed by denaturing polyacrylamide gel electrophoresis followed by overnight exposure to a phosphor screen and imaging. Image Studio Lite 5.2.5 software was used to quantify primer and extended primer gel band intensities and determine the fraction of extended primer or velocity at each triphosphate concentration. GraphPad Prism 7 was used to plot measured velocities versus the eight triphosphate concentrations examined and to fit the Michaelis-Menten equation to determine k_{cat} and K_M ; values reported are the averages and standard deviations of three, independent experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (GM060005 and GM118178 to F.E.R.) and by National Science Foundation Graduate Research Fellowships (S.E.M. and A.W.F.; NSF/DGE-1346837).

References

1. Arkin AP, Fletcher DA. Fast, cheap and somewhat in control. *Genome Biol.* 2006; 7:114. [PubMed: 16942631]
2. Wang YH, Wei KY, Smolke CD. Synthetic biology: advancing the design of diverse genetic systems. *Annu. Rev. Chem. Biomol. Eng.* 2013; 4:69–102. [PubMed: 23413816]
3. Purnick PE, Weiss R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* 2009; 10:410–422. [PubMed: 19461664]
4. Way JC, Collins JJ, Keasling JD, Silver PA. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell.* 2014; 157:151–161. [PubMed: 24679533]
5. Nandagopal N, Elowitz MB. Synthetic biology: integrated gene circuits. *Science.* 2011; 333:1244–1248. [PubMed: 21885772]
6. Cameron DE, Bashor CJ, Collins JJ. A brief history of synthetic biology. *Nat. Rev. Microbiol.* 2014; 12:381–390. [PubMed: 24686414]
7. Short GF 3rd, Golovine SY, Hecht SM. Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry.* 1999; 38:8808–8819. [PubMed: 10393557]
8. Malyshev DA, Romesberg FE. The expanded genetic alphabet. *Angew. Chem. Int. Ed.* 2015; 54:11930–11944.
9. Zhang L, Yang Z, Sefah K, Bradley KM, Hoshika S, Kim MJ, Kim HJ, Zhu G, Jimenez E, Cansiz S, Teng IT, Champanhac C, McLendon C, Liu C, Zhang W, Gerloff DL, Huang Z, Tan W, Benner SA. Evolution of functional six-nucleotide DNA. *J. Am. Chem. Soc.* 2015; 137:6734–6737. [PubMed: 25966323]
10. Kimoto, M., Hirao, I. Creation of unnatural base pair systems toward new DNA/RNA biotechnologies. In: Erdmann, AV, Markiewicz, TW., Barciszewski, J., editors. *Chemical Biology of Nucleic Acids: Fundamentals and Clinical Applications*. Springer Berlin Heidelberg; Berlin, Heidelberg: 2014. p. 131-148.
11. Kuchta RD, Benkovic P, Benkovic SJ. Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry.* 1988; 27:6716–6725. [PubMed: 3058205]
12. Malyshev DA, Dhama K, Quach HT, Lavergne T, Ordoukhanian P, Torkamani A, Romesberg FE. Efficient and sequence-independent replication of DNA containing a third base pair establishes a

- functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. USA*. 2012; 109:12005–12010. [PubMed: 22773812]
13. Zhang Y, Lamb B, Feldman AW, Zhou AX, Lavergne T, Li L, Romesberg FE. A semi-synthetic organism engineered for the stable expansion of the genetic alphabet. *Proc. Natl. Acad. Sci. USA*. 2017; 114:1317–1322. [PubMed: 28115716]
14. Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE, Marx A. KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat. Chem. Biol.* 2012; 8:612–614. [PubMed: 22660438]
15. Betz K, Malyshev D, Lavergne T, Welte W, Diederichs K, Romesberg FE, Marx A. Structural insights into DNA replication without hydrogen bonds. *J. Am. Chem. Soc.* 2013; 135:18637–18643. [PubMed: 24283923]
16. Malyshev DA, Pfaff DA, Ippoliti SI, Hwang GT, Dwyer TJ, Romesberg FE. Solution structure, mechanism of replication, and optimization of an unnatural base pair. *Chem. Eur. J.* 2010; 16:12650–12659. [PubMed: 20859962]
17. Matsuda S, Henry AA, Romesberg FE. Optimization of unnatural base pair packing for polymerase recognition. *J. Am. Chem. Soc.* 2006; 128:6369–6375. [PubMed: 16683801]
18. Matsuda S, Leconte AM, Romesberg FE. Minor groove hydrogen bonds and the replication of unnatural base pairs. *J. Am. Chem. Soc.* 2007; 129:5551–5557. [PubMed: 17411040]
19. Seo YJ, Hwang GT, Ordoukhanian P, Romesberg FE. Optimization of an unnatural base pair toward natural-like replication. *J. Am. Chem. Soc.* 2009; 131:3246–3252. [PubMed: 19256568]
20. Wichai U, Woski SA. Disiloxane-protected 2-deoxyribonolactone as an efficient precursor to 1,2-dideoxy-1- β -aryl-d-ribofuranoses. *Org. Lett.* 1999; 1:1173–1175. [PubMed: 10825968]
21. Ludwig J, Eckstein F. Rapid and efficient synthesis of nucleoside 5'-*O*-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* 1989; 54:631–635.
22. Romesberg FE, Yu C, Matsuda S, Henry AA. Development of a universal nucleobase and modified nucleobases for expanding the genetic code. *Curr. Protoc. Nucleic Acid Chem.* Chapter 1. 2002 Unit 1.5.

A **B** **Figure 1.**

Unnatural base pairs (UBPs). (A) The **dNaM-dTPT3** UBP. (B) The **dNaM** analogs with different *ortho* substituents explored in this study.

Figure 2.

Structural data.^{14,16} (A) Structure of a previously characterized analog of dTPT3TP (d5SICSTP) pairing with a template dNaM within the active site of KlenTaq DNA polymerase. For simplicity, only the nascent UBP and the natural primer-terminal dA-dT base pair are shown.¹⁴ (B) Structure of dNaM paired with a previously characterized analog of dTPT3 (d5SICS) in duplex DNA.¹⁶ For simplicity, only the UBP and the flanking natural base pairs are shown.

Table 1

Steady-state Kinetic Constants for Kf-Mediated Insertion of Triphosphate (dYTP) Opposite dNaM (X) in the Template

X	dYTP	k_{cat} (min ⁻¹)	K_M (μM)	k_{cat}/K_M (M ⁻¹ min ⁻¹)
<i>Sequence Context I</i>				
5' -d (TCCGGCTCGTATGTTGTGTGGAA)				
3' -d (AGGCCGAGCATACAACACACCTT X ACACTCGCCTATTGTTAAAGTGTGTCCTTTGTC)				
NaM	TPT3	24.31 ± 0.94	0.02 ± 0.005	1.0 × 10 ⁹
	NaM	9.15 ± 0.18	0.20 ± 0.02	4.5 × 10 ⁷
	A	4.60 ± 0.12	0.06 ± 0.01	8.0 × 10 ⁷
	G	n.d. ^a	n.d. ^a	< 1 × 10 ³
	C	n.d. ^a	n.d. ^a	< 1 × 10 ³
	T	n.d. ^a	n.d. ^a	< 1 × 10 ³
T	A	4.55 ± 0.10	0.01 ± 0.001	3.4 × 10 ⁸
<i>Sequence Context II</i>				
5' -d (CTTTTGTTTGTCTGCCGTGATCT)				
3' -d (GAAAACAAACAGACGGCACTAG X ATGTAACACACTCAATTTCAACATGAGCTC)				
NaM	TPT3	14.30 ± 0.54	0.11 ± 0.02	1.3 × 10 ⁸
	NaM	9.15 ± 0.44	2.32 ± 0.33	4.0 × 10 ⁶
	A	6.93 ± 0.17	17.64 ± 1.17	3.9 × 10 ⁵
	G	n.d. ^a	n.d. ^a	< 1 × 10 ³
	C	n.d. ^a	n.d. ^a	< 1 × 10 ³
	T	n.d. ^a	n.d. ^a	< 1 × 10 ³
T	A	19.99 ± 0.85	0.02 ± 0.002	8.4 × 10 ⁸
<i>Sequence Context III</i>				
5' -d (TCTTTTGTGTTGTCTGCCGTGATC)				
3' -d (AGAAAACAAACAGACGGCACTAG X CATGTAACACACTCAATTTCAACATGAGCTC)				
NaM	TPT3	18.52 ± 0.34	0.11 ± 0.01	1.7 × 10 ⁸
	NaM	n.d. ^a	n.d. ^a	< 1 × 10 ³
	A	n.d. ^a	n.d. ^a	< 1 × 10 ³
	G	n.d. ^a	n.d. ^a	< 1 × 10 ³
	C	n.d. ^a	n.d. ^a	< 1 × 10 ³
	T	n.d. ^a	n.d. ^a	< 1 × 10 ³
T	A	13.15 ± 0.42	0.03 ± 0.003	4.9 × 10 ⁸

^aRates too slow for determination of k_{cat} and K_M independently.

Table 2

Steady-state Kinetic Constants for Kf-Mediated Insertion of Triphosphate (dXTP) Opposite dTPT3 (Y) in the Template

Y	dXTP	k_{cat} (min ⁻¹)	K_M (μM)	k_{cat}/K_M (M ⁻¹ min ⁻¹)
<i>Sequence Context I</i>				
5' -d (TGTGAAATTGTTATCCGCTCACA)				
3' -d (ACACCTTTAACAATAGGCGAGTGT X AAGGTGTGTTGTATGCTCGGCCTTCGTATTTTCACATTTTCGG)				
TPT3	NaM	6.76 ± 0.24	0.06 ± 0.01	1.1 × 10 ⁸
	TPT3	3.01 ± 0.10	2.58 ± 0.43	1.2 × 10 ⁶
	MMO2	5.68 ± 0.18	2.07 ± 0.18	2.7 × 10 ⁷
	DM5	2.96 ± 0.09	7.18 ± 0.61	4.1 × 10 ⁷
	MMS2	9.27 ± 0.30	3.53 ± 0.26	2.6 × 10 ⁷
	MF2	3.13 ± 0.08	5.86 ± 0.46	5.3 × 10 ⁷
	A	n.d. ^a	n.d. ^a	< 1 × 10 ³
T	G	3.15 ± 0.14	20.26 ± 2.31	1.6 × 10 ⁵
	C	n.d. ^a	n.d. ^a	< 1 × 10 ³
	T	12.85 ± 0.46	28.28 ± 2.96	4.5 × 10 ⁵
	A	8.66 ± 0.50	0.04 ± 0.01	2.0 × 10 ⁸
<i>Sequence Context II</i>				
5' -d (ACAACTTTAACTCACACAATGTA)				
3' -d (TGTGAAATTGAGTGTGTTACAT X TCTAGTGCCGCTCTGTTTGTGTTTCTTACCTTAG)				
TPT3	NaM	10.71 ± 0.44	0.09 ± 0.01	1.3 × 10 ⁸
	TPT3	n.d. ^a	n.d. ^a	< 1 × 10 ³
	MMO2	9.77 ± 0.28	1.61 ± 0.21	6.1 × 10 ⁶
	DM5	4.49 ± 0.08	3.71 ± 0.29	1.2 × 10 ⁶
	MMS2	6.34 ± 0.22	0.94 ± 0.09	6.8 × 10 ⁶
	MF2	9.59 ± 0.24	4.16 ± 0.37	2.3 × 10 ⁶
	A	n.d. ^a	n.d. ^a	< 1 × 10 ³
T	G	3.33 ± 0.09	12.93 ± 1.08	2.6 × 10 ⁵
	C	n.d. ^a	n.d. ^a	< 1 × 10 ³
	T	5.82 ± 0.13	30.57 ± 1.95	1.9 × 10 ⁵
	A	19.07 ± 0.41	0.02 ± 0.002	9.4 × 10 ⁸
<i>Sequence Context III</i>				
5' -d (CAACCTTTAACTCACACAATGTAC)				
3' -d (GTTGAAATTGAGTGTGTTACATG X CTAGTGCCGCTCTGTTTGTGTTTCTTACCTTAG)				
TPT3	NaM	3.03 ± 0.12	0.31 ± 0.04	9.6 × 10 ⁶
	TPT3	n.d. ^a	n.d. ^a	< 1 × 10 ³
	MMO2	4.52 ± 0.11	6.90 ± 0.49	6.6 × 10 ⁵
	DM5	1.15 ± 0.02	28.02 ± 1.41	4.1 × 10 ⁴

Y	dXTP	k_{cat} (min^{-1})	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ ($\text{M}^{-1}\text{min}^{-1}$)
	MMS2	4.88 ± 0.09	4.01 ± 0.20	1.2×10^6
	MF2	5.17 ± 0.11	25.33 ± 1.26	2.0×10^5
	A	n.d. ^a	n.d. ^a	$< 1 \times 10^3$
	G	n.d. ^a	n.d. ^a	$< 1 \times 10^3$
	C	n.d. ^a	n.d. ^a	$< 1 \times 10^3$
	T	n.d. ^a	n.d. ^a	$< 1 \times 10^3$
T	A	16.99 ± 0.76	0.06 ± 0.01	2.6×10^8

^aRates too slow for determination of k_{cat} and K_{M} independently.