

## Building Better Enzymes: Molecular Basis of Improved Non-Natural Nucleobase Incorporation by an Evolved DNA Polymerase

Zahra Ouaray<sup>1</sup>, Isha Singh<sup>2,3</sup>, Millie M. Georgiadis<sup>2</sup> and Nigel G. J. Richards<sup>1,\*</sup>

<sup>1</sup> School of Chemistry, Cardiff University, Cardiff, CF10 3AT, United Kingdom

<sup>2</sup> Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, United States of America

<sup>3</sup> Present address: Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143, United States of America

\*Correspondence: Nigel G.J. Richards, School of Chemistry, Cardiff University, Park Place, Cardiff, CF10 3AT, United Kingdom. E-mail: [RichardsN14@cardiff.ac.uk](mailto:RichardsN14@cardiff.ac.uk).

RUNNING TITLE: Non-Natural Nucleobase Incorporation by an Evolved DNA Polymerase

Number of manuscript pages: 40

Supplementary material pages: 31

Tables: 0

Figures: 8

Supplementary material: Ouaray\_Manuscript\_Supp\_Information.pdf

Grant sponsors: BBSRC; Grant number: BB/P018017/1.  
NIH/NCI; Grant number: CA205166.

---

This is the author's manuscript of the article published in final edited form as:

Ouaray, Z., Singh, I., Georgiadis, M. M., & Richards, N. G. J. (2019). Building Better Enzymes: Molecular Basis of Improved Non-Natural Nucleobase Incorporation by an Evolved DNA Polymerase. Protein Science, n/a(n/a). <https://doi.org/10.1002/pro.3762>

**Abstract:** Obtaining semi-synthetic microorganisms that exploit the information density of “hachimoji” DNA requires access to engineered DNA polymerases. A KlenTaq variant has been reported that incorporates the “hachimoji” **P:Z** nucleobase pair with a similar efficiency to that seen for Watson-Crick nucleobase incorporation by the wild type (WT) KlenTaq DNA polymerase. The variant polymerase differs from WT KlenTaq by only four amino acid substitutions, none of which are located within the active site. We now report a series of molecular dynamics (MD) simulations on a series of binary complexes aimed at elucidating the contributions of the four amino acid substitutions to altered catalytic activity. These simulations suggest that WT KlenTaq is insufficiently flexible to be able to bind AEGIS DNA correctly, leading to the loss of key protein/DNA interactions needed to position the binary complex for efficient incorporation of the “hachimoji” **Z** nucleobase. In addition, we test literature hypotheses about the functional roles of each amino acid substitution and provide a molecular description of how individual residue changes contribute to the improved activity of the KlenTaq variant. We demonstrate that MD simulations have a clear role to play in systematically screening DNA polymerase variants capable of incorporating different types of non-natural nucleobases thereby limiting the number that need to be characterized by experiment.

**Keywords:** Expanded Genetic Alphabets; DNA Replication; Polymerase; Enzyme Engineering; Molecular Dynamics.

**Statement:** It is now possible to build DNA molecules containing non-natural nucleobase pairs in addition to A:T and G:C. Exploiting this development in synthetic biology requires engineered DNA polymerases that can replicate non-natural nucleobase pairs. Computational studies on a DNA polymerase variant reveal how amino acid substitutions outside of the active site yield an enzyme that replicates non-natural nucleobase pairs with high efficiency. This work will facilitate efforts to obtain bacteria possessing an expanded genetic alphabet.

**Abbreviations:** AEGIS, artificially expanded genetic information systems; **P**, 2-amino-8-(1-beta-D-2'-deoxyribofuranosyl)imidazo [1,2-a]-1,3,5-triazin-[8H]-4-one; **Z**, 6-amino-3-(2'-deoxyribofuranosyl)-5-nitro-1H-pyridin-2-one; DCCM, dynamic cross-correlation map; MD, molecular dynamics; PCA, principal component analysis; PDB, Protein Data Bank; RMSD, root mean square deviation; WT, wild type.

## Introduction

The development of “semi-synthetic” microorganisms possessing artificially expanded genetic information systems (AEGIS) will permit access to cells with novel phenotypes and biotechnological applications.<sup>1-4</sup> Non-natural nucleobase pairs that meet the size and/or hydrogen bonding complementarity rules of Watson-Crick base pairing have been described over the past two decades [Fig. S1(A)],<sup>5-8</sup> including the complementary 2-amino-8-(1-beta-D-2'-deoxyribofuranosyl)imidazo [1,2-a]-1,3,5-triazin-[8H]-4-one (trivially known as **P**) and 6-amino-3-(2'-deoxyribofuranosyl)-5-nitro-1H-pyridin-2-one (trivially known as **Z**) nucleobase pair that is present in “hachimoji” DNA (Fig. 1).<sup>1,5,9-12</sup> As is true of naturally occurring (Watson-Crick) DNA, AEGIS DNA duplexes containing **P:Z** pairs interconvert easily between A- and B-helical forms.<sup>13,14</sup> In addition, B-form DNA tolerates the inclusion of multiple consecutive **P:Z** nucleobase pairs with minimal structural impact on the double helix when compared to duplexes containing only A:T or G:C base pairs.<sup>9</sup>

Identifying DNA polymerases capable of catalyzing the incorporation of **P:Z** nucleobase pairs with efficiencies comparable to those that replicate Watson-Crick DNA is a necessary pre-requisite to realizing the promise of expanded genetic alphabets.<sup>15,16</sup> A variety of library generation and selection strategies have been developed to re-engineer the fidelity of DNA polymerases,<sup>17-19</sup> including the large (Klenow) fragment of *Thermus aquaticus* DNA polymerase I, which lacks the N-terminal 5'-3' exonuclease domain.<sup>20</sup> Thus, a compartmentalized self-replication strategy<sup>21,22</sup> was used to obtain a KlenTaq variant capable of incorporating dZTP

opposite a **P** nucleobase in the template strand with a greatly improved efficiency relative to the wild type (WT) precursor (Fig. 1).<sup>23</sup> The evolved KlenTaq variant contains four amino acid replacements (M444V, P527A, D551E and E832V; Fig. 1), all of which are distal to the active site. As a result, none of these residues interact directly with either the primer/template **P:Z** in the active site or with incoming nucleotide triphosphate (**dZTP**) in high resolution crystal structures of the pre- and post-incorporation complexes for this variant polymerase (PDB:5W6K and PDB:5W6Q, respectively).<sup>24</sup>

In considering how this evolved KlenTaq variant might differ from WT KlenTaq, Singh et al. analyzed differences in relative domain motions in the unnatural complexes as compared to the natural complexes.<sup>24</sup> There were no significant differences between the relative domain positions in the crystal structures of pre-incorporation complexes for the Watson-Crick (WT KlenTaq within incoming dNTP) and AEGIS (KlenTaq variant with incoming **dZTP** paired to template **P**) systems. In contrast, there were significant differences in the post-incorporation Watson-Crick (WT KlenTaq with G:C bound in the active site) *versus* AEGIS (evolved KlenTaq with **P:Z** bound in the active site) complexes. In forming the ternary pre-incorporation complex, there is a large motion of the fingers domain as it closes down on the incoming dNTP when compared to the binary post-incorporation complex.<sup>24</sup> Thus, the average rotation angle calculated from a comparison of the pre- and post-incorporation AEGIS complexes was 63.6° while that for the Watson-Crick complexes was decreased by 4-5° despite the fact that the Watson-Crick and AEGIS

Accepted Article

pre-incorporation complexes ultimately reached a similar structural state. Modeling the AEGIS template/primer bound to WT KlenTaq in a post-incorporation complex resulted in two significant “clashes”, suggesting that the structure of the AEGIS template-primer differed significantly from that of its Watson-Crick counterpart. Collectively, the crystallographic evidence supported the hypothesis that the four amino acid substitutions in the KlenTaq variant led to increased flexibility in the enzyme and that this property was required for efficient incorporation of dZTP opposite template **P**.<sup>24</sup> In addition, this study showed that the post-incorporation complex was the most affected by the residue substitutions. We now report a series of molecular dynamics (MD) simulations on eight binary complexes that not only test the idea that amino acid substitutions in the KlenTaq variant give rise to increased flexibility but also reveal how altered dynamical motions might contribute to the increased efficiency of **P:Z** nucleobase incorporation by the KlenTaq variant. These calculations set the scene for obtaining novel KlenTaq polymerase variants that can incorporate multiple types of non-natural nucleobase pairs using MD simulations as part of rational, structure-based strategies.

## Results and Discussion

### *WT KlenTaq and the evolved KlenTaq polymerase have similar dynamical properties in the absence of DNA*

MD simulations of WT KlenTaq and the KlenTaq variant in water show that there is little difference in the dynamics of the two enzymes in the absence of bound

DNA (Figs. S2-S4). The similarity of motions in the “thumb”, “palm” and “fingers” domains for both polymerases is evident from principal component analysis (PCA).<sup>25</sup> Thus, projecting the simulation snapshots along the largest components (PC1 and PC2) shows that WT KlenTaq and the KlenTaq variant explore the same phase space (Fig. S4).

### ***WT KlenTaq polymerase binds differently to Watson-Crick and AEGIS DNA template/primer duplexes***

We tested the hypothesis that the four amino acid substitutions in the KlenTaq variant exert their effects by modifying dynamical motions in the binary complex<sup>21</sup> using four MD simulations: WT KlenTaq bound to Watson-Crick DNA (template: 5'-AAAGGGCGCCGTGGTC-3'/primer: 5'-GACCACGGCGCC-3') and **P:Z**-containing DNA (template: 5'-AAAG**P**GCGCCGTGGTC-3'/primer: 5'-GACCACGGCGC**Z**-3') duplexes, and the KlenTaq variant bound to the same Watson-Crick and AEGIS DNA duplexes. Template and primer strands were positioned in these models based on the X-ray crystal structure of the variant KlenTaq/AEGIS DNA binary complex.<sup>24</sup> All four models were stable throughout the MD simulations based on RMSD values (Fig. S2). Most of the mobile residues were located in the fingers and thumb domains. The thumb domains in both polymerases when in the binary complexes, however, were less flexible when compared the cognate, uncomplexed enzymes (Fig. S3).

The dynamical motions of both the protein and the DNA in the four trajectories were analyzed to understand how the altered residues (M444V, P527A, D551E and E832V) in the KlenTaq variant might impact incorporation of both Watson-Crick and AEGIS nucleobases. The trajectories of the binary complexes in which WT KlenTaq was bound to either Watson-Crick or AEGIS DNA gave two main principal components, which captured 39.6% of the coordinate variance [Fig. 2(A)]. The largest of these, PC1, is associated with a motion in which the thumb and fingers domains move away from each other in the binary complex. The second major component, PC2, describes local motions in the thumb domain. Projecting structures sampled in the two MD simulations along PC1 and PC2 [Fig. 2(A)] shows that the two binary complexes separate along PC1, although they can sample common regions of phase space.

Domain motions (DynDom)<sup>26</sup> computed from the MD trajectories highlight these differences in the dynamics of WT polymerase when complexed with either Watson-Crick or AEGIS DNA duplexes (Fig. 3). In the WT KlenTaq/Watson-Crick DNA binary complex, the inactive exonuclease domain (residues 290-423) moves away from the polymerase domain (424-832) due to bending of residues in the palm (424-449, 553-614 and 774-832) [Fig. 3(A), Table S1]. This motion is consistent with the need for the polymerase to change its conformation in order to bind the DNA duplex. Additional evidence is provided from the dynamic cross-correlation map (DCCM)<sup>27</sup> of this binary complex, which shows the motions of the fingers and thumb domains to be anti-correlated [red box in Fig. 4 (left/top)]. An anti-correlated motion

of residues in the  $\alpha J/K$  helix of the palm domain (Fig. 1) and fingers domain [blue box in Fig. 4 (left/top)], which is not seen in the MD trajectory of the uncomplexed WT polymerase [Fig. S4(C)], also supports the DynDom analysis.

The motional properties of the binary complex in which WT KlenTaq binds AEGIS DNA differ in a number of important aspects from that containing Watson-Crick DNA. For example, the domain motions are more complicated even though the exonuclease and polymerase domains still move apart [Fig. 3(B)]. Not only do the fingers and thumb move away from each other but the tip of the thumb about the AEGIS DNA becomes repositioned due to residues 477-543 moving away from the fingers [Fig. 3(B)] with a rotation of  $6.4^\circ$  and a translation of  $0.2 \text{ \AA}$ . In addition, the base of the fingers domain moves toward the thumb with a rotation of  $7.5^\circ$  and a translation of  $0.2 \text{ \AA}$  (Table S1) due to bending of residues located in the palm, thereby also causing the tip of the fingers domain to move away from the thumb [Fig. 3(B)]. In agreement with these observations, the DCCM for the WT KlenTaq/AEGIS DNA binary complex shows anti-correlated motions of (i) the fingers and thumb [red box in Fig. 4 (left/bottom)] and (ii) residues in the  $\alpha J/K$  helix of the palm domain with the fingers [blue box in Fig. 4 (left/bottom)]. Neither of these motions are seen for the uncomplexed WT polymerase [Fig. S4(C)]. The presence of the **Z:P** nucleobase pair also leads to anti-correlated residue motions in the  $\alpha G$  helix and the fingers domain [green box in Fig. 4 (left/bottom)].

***The evolved KlenTaq polymerase accommodates both Watson-Crick and AEGIS DNA template/primer duplexes albeit through distinct dynamical motions***

While it is important for the evolved polymerase to recognize the unnatural AEGIS substrate, it must also recognize Watson-Crick nucleobases to be useful for the efficient replication of hachimoji DNA. Two major principal components capture 53.4% of the coordinate variance in the MD trajectories of the binary complexes of the KlenTaq variant bound to either Watson-Crick or AEGIS DNA. Projecting structures sampled in the two MD simulations along PC1 and PC2 shows a clear separation along both components, meaning that the two complexes sample different regions of phase space [Fig. 2(B)]. Comparing this projection for the KlenTaq variant [Fig. 2(B)] with that for the binary complexes of WT enzyme [Fig. 2(A)] shows an increased amplitude of the motions along PC1, supporting the idea that the residue substitutions confer increased flexibility onto the KlenTaq variant.

DynDom analysis suggests that the fingers domain of the variant polymerase exhibits different motions when binding Watson-Crick rather than AEGIS DNA. In addition, the tip of the fingers domain is tilted towards the base with a 14.6° rotation and a 0.5 Å translation in the KlenTaq variant/Watson-Crick DNA binary complex (Table S1), and the domain motions differ from those present in the KlenTaq variant/AEGIS DNA complex [Fig. 3(C), 3(D)]. These findings are consistent with inferences based on comparing the X-ray crystal structures of the KlenTaq

variant/AEGIS DNA and the WT KlenTaq/Watson-Crick DNA binary complexes.<sup>24</sup> The rotation of the thumb domain about residues 468-534 in the MD trajectory of the KlenTaq variant/AEGIS DNA binary complex is larger than that seen in the X-ray structure (Table S1), perhaps as a consequence of replacing Pro-527 by alanine. The exonuclease domain also moves away from the polymerase domain in the KlenTaq variant/AEGIS DNA binary complex, which is not seen when the variant polymerase binds to Watson-Crick DNA [Fig. 3(C), 3(D)]. Finally, an anti-correlated motion between the  $\alpha J/K$  helix in the palm and the fingers is present when the KlenTaq variant binds AEGIS DNA [blue box in Fig. 4 (right/bottom)] that is absent in the free polymerase. In contrast, this anti-correlated motion is decreased in the KlenTaq variant/Watson-Crick DNA binary complex. We conclude that the KlenTaq variant exhibits altered dynamical motions in binding the Watson-Crick DNA duplex, as compared to AEGIS DNA, even though it still readily accommodates this substrate.

***Z:P-containing DNA in the binary complex is sub-optimally recognized by the WT KlenTaq polymerase***

Interactions of the terminal nucleobase pair (C112:G205) of Watson-Crick DNA with the WT polymerase seen in the MD-derived trajectory are consistent with experimental findings.<sup>28-30</sup> Three intermolecular interactions between the C112:G205 nucleobase pair and protein residues Arg-573, Arg-746 and Asp-785 are highly populated in the simulation, two of which (Arg-746/G205 and Asp-785/C112) involve

the sugar-phosphate backbone [Fig. S5(A)]. The third interaction is a hydrogen bond between the Arg-573 side chain and the newly added nucleobase C112 (in the N<sup>th</sup> position at the 3'-end of the primer) [Fig. S5(A), Table S4]. In addition, the side chain of Arg-587, which is located in the  $\alpha JK$  helical region of the palm domain, forms three highly populated hydrogen bonds to C111, located in the [N-1]<sup>th</sup> position, in the MD trajectory [orientation 1 of Fig. 5(A)]. This finding is consistent with prior proposals that Arg-587 correctly positions the template strand through interactions with the [N-1]<sup>th</sup> nucleotide thereby allowing the incoming dNTP to bind appropriately for reaction with the 3'-OH in the ternary complex prior to primer extension.<sup>29</sup>

The interaction of Arg-587 with the [N-1]<sup>th</sup> nucleotide in the binary complex is also correlated with the presence of a salt bridge between Arg-596 and Glu-832 in the simulation trajectory of the WT KlenTaq/Watson-Crick DNA binary complex [Fig. 5(A), S6(A)]. Two “coupled” arrangements of the Arg-587 and Arg-596 side chains located at the ends of the  $\alpha JK$  helical region are also seen during the MD trajectory [Fig. S6(A)]. In the first arrangement (orientation 1), which is present in a majority of sampled structures, Arg-587 interacts with the [N-1]<sup>th</sup> nucleotide (C111) and Arg-596 is positioned to interact with the Glu-826 and Glu-832 side chains (Fig. 5). In the second arrangement (orientation 2) Arg-587 interacts with the N<sup>th</sup> nucleotide (C112) and Arg-596 reorients to interact with the  $\alpha H$  helix in the thumb domain (Fig. 5).

These interactions at the growing end of the DNA duplex are altered in the MD simulations of the WT KlenTaq/AEGIS DNA binary complex (Table S5). For example, only two interactions with the N<sup>th</sup> nucleotide (Z112) are observed, in which

Arg-573 and Asp-785 continue to form populated interactions to the nucleobase and sugar-phosphate backbone of **Z112** [Fig. S5(B)]. Moreover, the AEGIS nucleobase **P205**, located in the template strand, becomes hydrogen bonded to the Gln-754 side chain, a protein/DNA interaction that is absent in the trajectory for the WT KlenTaq/Watson-Crick DNA binary complex [Fig. S5(A)]. Importantly, because of its functional role in primer extension, the Arg-587 side chain interacts with **Z112** rather adopting its optimal location in which it can interact with C111, the [N-1]<sup>th</sup> nucleotide (Fig. 5, Table S5). Arg-596 is also oriented so as to interact with Asp-452 in the ***αH*** helix at the base of the thumb domain presumably as a consequence of the altered conformational preference of the Arg-587 side chain [Fig 5(B)]. Finally, the salt bridge between Arg-596 and Glu-832 is absent in the MD trajectory computed for the WT KlenTaq/AEGIS DNA binary complex.

***Both C:G and Z:P are accommodated in the active site of the KlenTaq variant***

In contrast to what is observed for the WT polymerase, the N<sup>th</sup> nucleotide pair in both binary complexes of the KlenTaq variant is stabilized by interactions between the phosphate backbone and Arg-746 and Arg-785 [Fig. S5(B), Table S6, S7]. A direct interaction between G206 and both Gln-754 and Arg-573 is also present in the KlenTaq variant/Watson-Crick DNA binary complex. Arg-573 is also seen to interact with C207. When AEGIS DNA is bound by the KlenTaq variant, however, the Gln-754/**P206** interaction is maintained but Arg-573 binds to **Z112** (Table S6, S7).

Importantly, and unlike what is observed for the WT polymerase, Arg-587 preferentially binds to the [N-1]<sup>th</sup> nucleotide in the primer (C111 or Z112) in *both* binary complexes involving the KlenTaq variant [Fig. S7(A), Table S6, S7]. This finding suggests that the salt bridge between the side chains of Arg-596 and Glu-832 in the KlenTaq variant is not necessarily required to maintain the Arg-587/[N-1]<sup>th</sup> nucleobase interaction [Fig. S7(B)].

Although the DNA duplexes maintain their helical structure and bending angle in all MD simulations, the distribution of the slide and twist angles<sup>31</sup> in the CC dinucleotide step in the Watson-Crick primer (slide: -1.3 and -1.9 Å, twist: 25°) differ from those of the cognate (CZ) step in the AEGIS primer (slide: -1.9 Å, twist: 30°) [Fig. 6(A), 6(B)]. This finding again supports the idea that WT KlenTaq binds DNA differently when the Z:P nucleobase pair replaces C:G whereas both Watson-Crick and AEGIS DNA are bound in a similar manner by the KlenTaq [variant Fig. 6(C), 6(D)]. The molecular origin of these differences in the slide and twist angle distributions are difficult to assign to specific protein/DNA interactions or individual amino acid substitutions in the KlenTaq variant, although they may be associated with the altered electrostatic and dispersion properties of the Z:P nucleobase pair.<sup>13</sup> As far as we can tell from our MD trajectories of the four binary complexes, the 3'-OH in either of the terminal C or Z nucleotides adopts a similar location and participates in the same hydrogen bonding interactions (Fig. S8). We cannot rule out the possibility, however, that longer MD simulations might exhibit altered positioning of the 3'-OH in the Watson-Crick and/or AEGIS binary complexes.

## ***M444V and D551E substitutions affect the dynamics and DNA binding of the evolved KlenTaq variant***

Throughout the MD trajectory of the WT KlenTaq/Watson-Crick DNA binary complex, Met-444 is held within a hydrophobic pocket defined by the side chains of residues Phe-564 and Met-779, which are located in the  $\beta 6$  and  $\beta 12$  strands, respectively [Fig. 7, S9(A)]. The  $\beta 6$  strand participates in an anti-parallel sheet linking the  $\alpha I$  helix of the thumb with the  $\alpha J/K$  region, and  $\beta 12$  in the palm domain is directly connected to the  $\alpha Q$  helix in the fingers. When WT KlenTaq is bound to Watson-Crick DNA, the Met-444 side chain makes an additional interaction with Met-765 that can be replaced by Gln-782, when the wild-type polymerase is bound to AEGIS DNA [Fig. 7(A), 7(B)]. The introduction of valine in this position (Val-444), however, means that the smaller side chain only forms populated interactions with Phe-564, Met-779 and Leu-780 in the simulations of both KlenTaq variant/Watson-Crick and KlenTaq variant/AEGIS DNA binary complexes [Fig. 7(C), 7(D)]. The consequence of this reduced set of interactions is that the  $\alpha G$  helix becomes more flexible in the variant compared to the WT polymerase [Fig. S9(B)]. Our calculated result is again in agreement with inferences about the elasticity of the  $\alpha G$  helix based on the X-ray crystal structure of the KlenTaq variant /AEGIS DNA binary complex.<sup>24</sup> Replacing Asp-551 (located in the  $\alpha I$  helix close to the  $\alpha G$  helix) by glutamate also impacts the flexibility of the  $\alpha G$  helix based on the MD simulations of the four binary complexes. Thus, the presence of the longer side chain at residue 551 permits a shorter (and stronger) salt bridge to the Arg-547 side chain located on the  $\alpha H$  helix

than is present in WT KlenTaq [Fig. S10(B)]. When AEGIS DNA is complexed within the KlenTaq variant, the interaction between Asp-551 and Arg-547 has the effect of stiffening the thumb thereby moving the  $\alpha G$  helix towards the fingers and giving rise to an anti-correlated motion of the  $\alpha G$ ,  $\alpha P$ ,  $\alpha Q$  and  $\alpha R$  helices with the fingers domain [green and orange boxes in Fig. 4 (right/bottom)].

### ***Single amino acid substitutions fail to capture the properties of the evolved polymerase-DNA complexes***

Additional MD simulations (only one repeat per variant) of the binary complexes for four single-point KlenTaq variants were performed to disentangle the contributions of the individual residue replacements to altering the properties of the evolved enzyme. All of these binary complexes were stable under our simulation conditions (Fig. S11) and the resulting trajectories were compared to that obtained for the WT KlenTaq/AEGIS DNA binary complex. In particular, we wanted to ascertain which, if any, of the residue substitutions corrected mis-positioning of arginine side chains about the **P:Z** nucleobase pair in the AEGIS DNA duplex. No change was observed for the polymerase/**Z112** interactions in the simulations of the P527A and E832V KlenTaq variants compared to those discussed above for the WT polymerase/AEGIS DNA complex (Fig. S12). This was not the case, however, in the MD simulation of the M444V variant/AEGIS DNA binary complex in which only the Arg-573 side chain interacted with **Z112** [Fig. 8(A)]. Altered polymerase/**Z112** interactions were also observed for the D551E variant/AEGIS DNA binary complex in

which only the Arg-587 side chain can interact with the non-natural nucleotide, albeit via an electrostatic interaction with the phosphate group in the oligonucleotide backbone [Fig. 8(A)]. As importantly, the motions of the M444V KlenTaq variant do permit the side chain of Arg-587 to hydrogen bond to C111, the [N-1]<sup>th</sup> nucleotide (70-73% occurrence). In addition, the DCCMs for the M444V and D551E KlenTaq variants [Fig. 8(B)] reveal an anti-correlated motion between the palm and thumb domains, which is absent in the P527A and E832V KlenTaq variants [Fig. S13(B)]. Turning the effects of individual residue substitutions on the structural properties of the template/primer duplex about the non-natural **P:Z** nucleobase pair in the AEGIS DNA, we find that the **CZ** dinucleotide slide and twist distributions observed for the M444V, P527A, and E832V KlenTaq variants are essentially identical to those seen for the WT polymerase. On the other hand, substituting Asp-551 by Glu results in the **CZ** dinucleotide slide and twist distributions becoming more like that observed for the **CC** dinucleotide when Watson-Crick DNA is bound by the WT polymerase [Fig. 6(A), 6(B), 8(C)]. In addition, we note that the interaction of Asp-785 with **Z112** is absent in both MD trajectories computed for the M444V/AEGIS DNA and D551E/AEGIS DNA complexes, suggesting that the addition of dZTP to the primer will be slowed in both of these KlenTaq single substitution variants.

## Conclusions

These MD simulations systematically explore the role of dynamical motions within the WT and evolved KlenTaq polymerases involved in positioning of Watson-Crick or AEGIS DNA template/primer substrates. In the absence of substrate, the amino acid substitutions within the evolved polymerase have no effect on the overall dynamical properties of the enzyme. In the presence of either Watson-Crick or AEGIS template/primer, however, the dynamical motions of the evolved KlenTaq are distinct from those of the WT polymerase. Specifically, WT KlenTaq is unable to maintain sets of correlated motions and specific interactions of amino acid residues required to appropriately position **P:Z** in the active site. The four amino acid substitutions in the KlenTaq variant resolve this problem by altering the flexibility of key segments of the enzyme with the consequence that the enzyme can bind to Watson-Crick and AEGIS DNA in an equivalent fashion.

Our MD simulations also confirm the proposal<sup>24</sup> that the KlenTaq variant has increased flexibility compared with the WT polymerase, which facilitates positioning of the Arg-587 side chain to interact with the [N-1]<sup>th</sup> nucleotide in the growing strand of both Watson-Crick and AEGIS DNA duplexes. Replacing Met-444 in the  $\alpha$ G helix and Asp-551 at the base of the thumb domain by valine and glutamate, respectively, allows the fingers, palm and thumb domains in the KlenTaq variant to bind AEGIS DNA in an optimal conformation. All three mutations are synergistic. Changing Met-444 to valine increases the flexibility of the  $\alpha$ G helix thereby ensuring an optimal interaction of Arg-587 with both Watson-Crick and AEGIS DNA duplexes. Replacing

Asp-551 by glutamate rigidifies the base of the thumb and also increases the flexibility of the  $\alpha$ G helix. Our calculations also confirm that replacing Pro-527 by alanine permits the tip of the thumb to be more flexible,<sup>24</sup> allowing the variant to bind both Watson-Crick and AEGIS DNA and facilitating the interaction of Asp-785 with the newly added nucleotide (C112 or Z112) in the binary complex. In complexes of WT KlenTaq, salt bridge formation between Glu-832 and Arg-596 is correlated with positioning of the Arg-587 side chain. This is no longer possible for the KlenTaq variant because Glu-832 is substituted by valine. No obvious dynamical changes were observed, however, for the E832V KlenTaq variant.

Both MD simulation and X-ray crystallography indicate that increasing domain flexibility will be required for polymerases capable of replicating DNA containing non-natural nucleobases with high efficiency. This appears to be the case even for DNA duplexes that look very similar to those containing only Watson-Crick nucleobase pairs. In identifying correlated residue/residue and residue/nucleotide interaction networks, hinge regions, and domain movements required for productive positioning of both natural and unnatural template/primer substrates in the active site of the KlenTaq variant, our calculations lay a foundation for screening altered KlenTaq polymerases with potential use in not only replicating hachimoji DNA but also the shape complementary nucleobases developed by Romesberg<sup>2</sup> and Hirao.<sup>7</sup> Further, our comparative analyses of MD simulations for all possible combinations of WT and evolved polymerases complexed to Watson-Crick and AEGIS DNA provides a

general, structure-based strategy for obtaining polymerases capable of replicating expanded genetic alphabets.

Accepted Article

## Methods

### *Parameterization of the non-natural nucleobases*

N9-methylated forms of the **Z** and **P** nucleobases were geometry optimized at the HF/6-31G\* level of theory using the NWChem software package<sup>32</sup> prior to obtaining RESP charges<sup>33</sup> subject to the constraint of zero net charge. This procedure is consistent with the GAFF2 parameterization of small molecules.<sup>34</sup> Partial charges on the sugar-phosphate atoms in the corresponding nucleotide fragments were taken from the PARMBSC1 force field used to model nucleic acids,<sup>35</sup> with the desired overall charge being obtained by adjusting the initial partial charges on C1' and H1' (Table S2).<sup>36</sup> Force field parameters for the bonding interactions were obtained from PARMBSC1 and GAFF2. The resulting set of parameters for the two non-natural nucleobases are provided elsewhere (Data S1).

### *Building the initial models of WT KlenTaq, the evolved KlenTaq variant and the binary complexes*

An initial model of the evolved KlenTaq polymerase variant bound to template (5'-AAAGPGCGCCGTGGTC-3') and primer (5'-GACCACGGCGCZ-3') DNA [Fig. S1(B)] was prepared from the corresponding X-ray crystal structure (PDB: 5W6Q).<sup>24</sup> Unobserved residues (291-293, 642-665 and 832) in the original X-ray crystal structure of the evolved KlenTaq variant were added based on their positions in other structures of WT KlenTaq (PDB:4KTQ<sup>30</sup> and PDB:1BGX<sup>37</sup>). Energy

minimization was performed to remove steric clashes in the initial model using the ff14SB<sup>38</sup> and PARMBSC1<sup>35</sup> force field parameters for the protein and DNA components, respectively, and our “in-house” parameters for the **Z:P** nucleobase pair. These calculations together with subsequent equilibration and MD simulations employed the PMEMD module implemented in the Amber2016 software suite.<sup>39,40</sup> The resulting model of the KlenTaq variant/AEGIS DNA binary complex was then used to construct models of the apo-forms of WT KlenTaq and the evolved KlenTaq variant, and the KlenTaq variant/Watson-Crick, the WT KlenTaq/Watson-Crick and the WT KlenTaq/AEGIS DNA binary complexes. Similarly, the last frame of WT KlenTaq/AEGIS DNA binary complex was used to construct the initial models of the binary complexes in which AEGIS DNA was bound to the four single-point KlenTaq variants. Each of these models was placed in a box of explicit TIP3P<sup>41</sup> water molecules containing K<sup>+</sup> and Cl<sup>-</sup> ions to yield an ionic strength of 10 mM (Table S3). These counterions were chosen in order to mimic the conditions under which the KlenTaq variant was obtained by directed evolution.<sup>23</sup> Each system was energy minimized to remove bad contacts, heated, and then equilibrated for 100 ns to allow the quasi-immobile ions sufficient time to equilibrate prior to the production phase.<sup>42</sup> Three independent MD simulations of 110 ns were then performed on each of the six systems in the NPT ensemble (T = 323 K and P = 100,000 Pa). Models of the binary complexes of the M444V, P527A, D551E and E832V single-point KlenTaq variants bound to AEGIS DNA were built from the last frame of the WT KlenTaq/**Z:P**-containing DNA complex. These model structures were then solvated and

Accepted Article

equilibrated following the procedures outlined above, and MD trajectories (110 ns) determined using identical simulation conditions to those outlined above. Periodic boundary conditions were used in all MD simulations, with an 8 Å cutoff being used for non-bonded interactions, and particle-mesh Ewald methods were used to describe long-range electrostatics.<sup>43-45</sup> The temperature and pressure of each system was maintained using Langevin constraints,<sup>46</sup> and the SHAKE algorithm was used to constrain bonds involving hydrogen atoms,<sup>47</sup> which allowed the use of 2.0 fs time steps.

### ***Analysis of the MD trajectories***

Structures were sampled at 100 ps intervals in each of the 23 MD trajectories to give a combined 3300 “snapshots” for each of the six WT and variant Klentaq systems, and 1100 “snapshots” for the four single variants. These trajectories that were analyzed using the DynDom,<sup>26</sup> 3DNA,<sup>31</sup> Bio3D,<sup>27</sup> CPPTRAJ<sup>48</sup> and VMD<sup>59</sup> software packages. Thus, DCCMs were computed based on the motions of the C $\alpha$  carbons of the protein using algorithms implemented in the Bio3D library in “R”.<sup>27</sup> These maps reveal the extent to which the motions of two C $\alpha$  carbons are correlated during the MD trajectory, with the correlation coefficient for the residue pair varying over a range of -1 to +1. When the C $\alpha$  carbons have completely correlated motions (identical phase space, period and direction), the correlation coefficient has a value of +1. For “anti-correlated” motions, in which the correlated motions take place in opposite directions, the value of the correlation coefficient is -1. For completely

uncorrelated motions, the correlation coefficient has a value of zero. The Bio3D library was also used to perform principal component analysis (PCA) for all snapshots obtained in the combined MD trajectories for each of the six systems.

Reference structures needed for the DynDom analysis were obtained by averaging the coordinates of the last snapshot in each of the three independent MD simulations for WT KlenTaq, the KlenTaq variant and four binary complexes (WT KlenTaq/Watson-Crick DNA; WT KlenTaq/AEGIS DNA; KlenTaq variant/Watson-Crick DNA; KlenTaq variant/AEGIS DNA). The domain motions of each reference structure compared to the X-ray crystal structure of WT KlenTaq in a binary complex (PDB: 3SZ2)<sup>30</sup> were then obtained using the DynDom package<sup>26</sup> by aligning the C $\alpha$  carbons of the two proteins. Helical parameters for the DNA molecules in each of the four complexes throughout the MD trajectories were calculated using the 3DNA<sup>31</sup> and do\_x3DNA programs.<sup>50</sup> Hydrogen bonds were identified in the simulations assuming that the heavy-atom donor (D)/acceptor (A) distance cutoff was 3.0 Å and that the D-H-A angle was in the range of 135-180°. Populated hydrogen bonds were defined to be present in at least 10% of the sampled structures in each MD trajectory with the exception of those between the Arg-587 side chain and the nitro group of the **Z** nucleobase.

## References

1. Hoshika S, Leal NA, Kim M-H, Kim M-S, Karalkar NB, Kim H-I, Bates AM, Watkins NE Jr, Santa Lucia HA, Meyer AJ, Das Gupta S, Piccirilli JA, Ellington AD, Santa Lucia J Jr, Georgiadis MM, Benner SA (2019) Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* 363:884-887.
2. Morris SE, Feldman AW, Romesberg FE (2017) Synthetic biology parts for the storage of increased genetic information in cells. *ACS Synth Biol* 6:1834–1840.
3. Wang Y-H, Wei KY, Smolke CD (2013) Synthetic Biology: Advancing the design of diverse genetic systems. *Annu Rev Chem Biomol Eng* 4:69–102.
4. Herdewijn P, Marlière P (2009) Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem Biodivers* 6:791–808.
5. Benner SA (2004) Understanding nucleic acids using synthetic chemistry. *Acc Chem Res* 37:784–797.
6. Feldman AW, Romesberg FE (2018) Expansion of the genetic alphabet: A chemist's approach to synthetic biology. *Acc Chem Res* 51:394–403.
7. Hirao I, Kimoto M, Yamashige R (2012) Natural versus artificial creation of base pairs in DNA: Origin of nucleobases from the perspectives of unnatural base pair studies. *Acc Chem Res* 45:2055–2065.

8. Kool ET (2002) Replacing the nucleobases in DNA with designer molecules. *Acc Chem Res* 35:936–943.
9. Georgiadis MM, Singh I, Kellett WF, Hoshika S, Benner SA, Richards NGJ (2015) Structural basis for a six nucleotide genetic alphabet. *J Am Chem Soc* 137:6947–6955.
10. Chawla M, Credendino R, Chermak E, Oliva R, Cavallo L (2016) Theoretical characterization of the H-bonding and stacking potential of two nonstandard nucleobases expanding the genetic alphabet. *J Phys Chem B* 120:2216–2224.
11. Wang W, Sheng X, Zhang S, Huang F, Sun C, Liu J, Chen D (2016) Theoretical characterization of the conformational features of unnatural oligonucleotides containing a six nucleotide genetic alphabet. *Phys Chem Chem Phys* 18:28492–28501.
12. Reichenbach LF, Sobri AA, Zaccai NR, Agnew C, Burton N, Eperon LP, de Ornellas S, Eperon IC, Brady RL, Burley GA (2016) Structural basis of the mispairing of an artificially expanded genetic information system. *Chem* 1:946–958.
13. Molt RW Jr, Georgiadis MM, Richards NGJ (2017) Consecutive non-natural PZ nucleobase pairs in DNA impact helical structure as seen in 50  $\mu$ s molecular dynamics simulations. *Nucleic Acids Res* 45: 3643–3653.

14. Richards NGJ, Georgiadis MM (2017) Toward an expanded genome: Structural and computational characterization of an artificially expanded genetic information system. *Acc Chem Res* 50:1375–1382.
15. Hottin A, Marx A (2016) Structural insights into the processing of nucleobase-modified nucleotides by DNA polymerases. *Acc Chem Res* 49:418–427.
16. Yang Z, Chen F, Alvarado JB, Benner SA (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J Am Chem Soc* 15105–15112.
17. Galloway DA, Laimins LA, Division B, Hutchinson F (2016) Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat Chem* 8:87–92.
18. Wynne SA, Pinheiro VB, Holliger P, Leslie AGW (2013) Structures of an apo and a binary complex of an evolved archeal B family DNA polymerase capable of synthesising highly Cy-dye labelled DNA. *PLoS One* 8:e70892.
19. Coulther TA, Stern HR, Beuning PJ (2019) Engineering polymerases for new functions. *Trends Biotechnol* 37:1091-1103.
20. Kim Y, Eom SH, Wang J, Lee D.-S., Suht SE, Steitz TA (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* 376:612-616.

21. Ghadessy FJ, Ong JL, Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci USA* 98:4552–4557.
22. Houlihan G, Arangundy-Franklin S, Holliger P (2017) Exploring the chemistry of genetic storage and propagation through polymerase engineering. *Acc Chem Res* 50:1079-1087.
23. Laos R, Shaw R, Leal NA, Gaucher E, Benner SA (2013) Directed evolution of polymerases to accept nucleotides with nonstandard hydrogen bond patterns. *Biochemistry* 52:5288–5294.
24. Singh I, Laos R, Hoshika S., Benner SA, Georgiadis MM (2018) Snapshots of an evolved DNA polymerase pre- and post-incorporation of an unnatural nucleotide. *Nucl Acids Res* 46:7977–7988.
25. Skjaerven L, Martinez A, Reuter N (2011) Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins: Struct Funct Bioinform* 79:232–243.
26. Taylor D, Cawley G, Hayward S (2014) Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics* 30:3189–3196.

27. Grant BJ, Rodrigues APC, El Sawy KM, McCammon JA, Caves LSD (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22:2695–2696.
28. Eom SH, Wang J, Steitz TA (1996) Structure of Taq polymerase with DNA at the polymerase active site. *Nature* 382:278–281.
29. Li Y, Kong Y, Sergey K, Waksman G (1998) Crystal structures of the Klenow fragment of *Thermus aquaticus* DNA polymerase I complexed with deoxyribonucleoside triphosphates. *Protein Sci* 7:1116–1123.
30. Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE, Marx A (2012) KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat Chem Biol* 8:612-614.
31. Lu X-J, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108-5121
32. Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, Van Dam HJJ, Wang D, Nieplocha J, Apra E, Windus TL, De Jong WA (2010) NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput Phys Commun* 181:1477–1489.

33. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J Phys Chem* 97:10269–10280.
34. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general Amber force field. *J Comput Chem* 25:1157-1174.
35. Ivani I, Dans PD, Noy A, Pérez A, Faustino I, Hospital A, Walther J, Andrio P, Goñi R, Balaceanu A, Portella G, Battistini F, Gelpi JL, González C, Vendruscolo M, Laughton CA, Harris SA, Case DA, Orozco M (2016) PARMBSC1 : A refined force-field for DNA simulations. *Nat Methods* 13:55-58.
36. Cieplak P, Cornell WD, Bayly C, Kollman PA (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J Comput Chem* 16:1357-1377.

37. Murali R, Sharkey DJ, Daiss JL, Murthy HM (1998) Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: the Fab is directed against an intermediate in the helix-coil dynamics of the enzyme. *Proc Natl Acad Sci USA* 95:12562–12567.
38. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11:3696-3713.
39. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668-1688.
40. Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv. Protein Chem* 66:27-85.
41. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926-935.
42. Pasi M, Maddocks JH, Beveridge D, Bishop TC, Case DA, Cheatham T, Dans PD, Jayaram B, Lankas F, Laughton C, Mitchell J, Osman R, Orozco M, Pérez A, Petkevičiūtė D, Spackova N, Sponer J, Zakrzewski K, Lavery R (2014)  $\mu$ ABC: A

systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res* 42:12272-12283.

43. Cerutti DS, Case DA (2010) Multi-Level Ewald: A hybrid multigrid/Fast Fourier Transform approach to the electrostatic particle-mesh problem. *J Chem Theory Comput* 6:443-458.
44. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An  $N\log(N)$  method for Ewald sums in large systems. *J Chem Phys* 98:10089-10092.
45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995). A smooth Particle mesh Ewald method. *J Chem Phys* 103:8577-8593.
46. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684-3690.
47. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J Comput Phys* 23:327-341.
48. Roe DR, Cheatham TE (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 9:3084-3095.

49. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual Molecular Dynamics. J Mol Graph 14:33-38.

50. Kumar R, Grubmüller H (2015) do-x3dna: A tool to analyze structural fluctuations of dsDNA or dsRNA from molecular dynamics simulations. Bioinformatics 31:2583-2585

## FIGURE LEGENDS

Figure 1. (A) Cartoon representation of the X-ray crystal structure of the “evolved” variant KlenTaq polymerase in its binary complex (PDB: 5W6Q),<sup>24</sup> showing the location of the **P:Z** nucleobase pair in the polymerase domain. Helices and strands are colored purple and yellow, respectively. (B) Chemical structure of the **P:Z** nucleobase pair. R indicates the location of the 2'-deoxyribose substituent and hydrogen bonds are indicated by dashed lines. (C) Close-up of the polymerase domain showing the palm, fingers and thumb domains, the helical regions  **$\alpha$ G- $\alpha$ K**,  **$\alpha$ P** and  **$\alpha$ Q**, and the side chains of the mutated residues, Val444, Ala527, Glu551 and Val832, in the variant KlenTaq polymerase.

Figure 2. (A) Cartoon representation of the motions associated with the two principal components PC1 (left) and PC2 (center) describing the variance of the WT polymerase/Watson-Crick DNA and WT polymerase/AEGIS DNA binary complexes at early (red), intermediate (white) and late (blue) stages of the MD simulations; (right) projection of trajectory snapshots along PC1 and PC2 for the WT polymerase/Watson-Crick DNA (blue) and WT polymerase/AEGIS DNA (red) binary

complexes. (B) Cartoon representation of the motions associated with the two principal components PC1 (left) and PC2 (center) describing the variance of the variant polymerase/Watson-Crick DNA and variant polymerase/AEGIS DNA binary complexes at early (red), intermediate (white) and late (blue) stages of the MD simulations; (right) Projection of trajectory snapshots along PC1 and PC2 for the variant polymerase/Watson-Crick DNA (blue) and variant polymerase/AEGIS DNA (red) binary complexes.

Figure 3. DynDom<sup>26</sup> analysis of the four polymerase/DNA complexes. For reference, WT polymerase/Watson-Crick DNA binary complex (PDB: 3SZ2)<sup>30</sup> is shown in grey in all of the panels. The axis of motion in each panel is indicated by an arrow, the head of which is colored to represent the moving domain (red, yellow, or purple). The fixed domain (blue) and regions that bend during the motion are shown in green. (A) C $\alpha$  trace of the WT KlenTaq/Watson-Crick DNA binary complex. The exonuclease domain (red) moves away from the fingers, palm, and thumb domains. (B) C $\alpha$  trace of the WT KlenTaq/AEGIS DNA binary complex. In this case, the fingers domain (purple) moves back into the plane of this image (purple headed arrow), the thumb domain (yellow) moves away from the fingers domain to the right, and the exonuclease domain (red) moves away from fingers, palm, and thumb domains as for the WT KlenTaq/Watson-Crick DNA binary complex. (C) C $\alpha$  trace of the variant polymerase/Watson-Crick DNA binary complex. The tip of the fingers domain (yellow) moves to the left, away from the thumb domain. The exonuclease domain exhibits minimal movement in contrast to the other complexes. (D) C $\alpha$  trace of the

variant polymerase/AEGIS DNA binary complex. The thumb domain (yellow) moves away from the fingers domain; the exonuclease domain (red) moves away from the other domains.

Figure 4. Dynamical cross correlation maps (DCCMs) computed for the four polymerase/DNA binary complexes; (left/top triangle) WT/Watson-Crick DNA binary complex, (left/bottom triangle) WT/AEGIS DNA binary complex, (right/top triangle) variant/Watson-Crick DNA binary complex, (right/bottom triangle) variant/AEGIS DNA binary complex. Correlated (range: 0.25 to 1) and anti-correlated (range: -0.25 to -1) motions are colored from light to dark blue and pink, respectively. Areas rendered in white correspond to non-correlated motions (range: -0.25 to 0.25). Secondary structural elements are also included on the map. See text for details.

Figure 5. (A) Orientational preferences of Arg-587 in the trajectories of the binary complexes between WT KlenTaq and Watson-Crick (left) and AEGIS (right) DNA. (B) Orientational preferences of the Arg-596 side chain in trajectories of the WT KlenTaq/Watson-Crick (left/middle) and variant KlenTaq/AEGIS (right) binary complexes. The  $\alpha H$  and  $\alpha J/K$  helices and the C-terminal (Cter) region are shown in a cartoon representation (purple). Atoms in named residues are colored according to the scheme: C, cyan; H, white; N, blue; O, red. Hydrogen bonds seen in the trajectories are represented by a black dashed line. Percentage occupancies of selected interactions are computed based on the number of structures in the trajectory in which the hydrogen bond is present.

Figure 6. Representative histograms showing the distribution of slide and twist values for the dinucleotide step involving the (N-1)th and Nth nucleobase pairs throughout the MD trajectories. (A) Slide and (B) twist values in the MD simulations of the WT polymerase binary complexes. (C) Slide and (D) twist values in the MD simulations of the variant polymerase binary complexes. In all figures CC and CZ refer to C111/C112 and C111/Z112 in the Watson-Crick and AEGIS DNA primer, respectively. The insets showing the molecular definitions of slide and twist for dinucleotide base pairs are taken from Lu and Olson.<sup>31</sup>

Figure 7. Interactions between residue 444 and surrounding residues in the MD trajectories of (A) the WT polymerase/Watson-Crick DNA binary complex, (B) the WT polymerase/AEGIS DNA binary complex, (C) the variant polymerase/Watson-Crick DNA binary complex, and (D) the variant polymerase/AEGIS DNA binary complex. The percentage occurrence (blue) of a specific interaction throughout the MD trajectory is also shown.

Figure 8. (A) Hydrogen bonds (dashed lines) between the **P:Z** nucleobase pair and protein residues, which are seen in the trajectories with greater than 30% occupancy in the MD trajectories computed for the M444V KlenTaq variant/AEGIS DNA (top) and D551E KlenTaq variant/AEGIS DNA (bottom) binary complexes. Percentage occupancies of selected interactions are computed based on the number of structures in the trajectory in which the hydrogen bond is present. Atoms in named residues are colored according to the scheme: C, cyan; H, white; N, blue; O, red; P,

orange. (B) Dynamical cross correlation maps computed for the M444V KlenTaq variant/AEGIS DNA (top triangle) and D551E KlenTaq variant/AEGIS DNA (bottom triangle) binary complexes. Correlated (range: 0.25 to 1) and anti-correlated (range: -0.25 to -1) motions are colored from light to dark blue and pink, respectively. Areas rendered in white correspond to non-correlated motions (range: -0.25 to 0.25). Secondary structural elements are also included on the map. See text for details. (C) Slide (left) and twist (right) values for the C1111/Z112 dinucleotide step of the AEGIS DNA primer in the MD simulations of the binary complexes of AEGIS DNA bound to WT KlenTaq and the four single-point (M444V, P527A, D551E and E832V) KlenTaq variants.















