JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

The Energetic Difference between Synthesis of Correct and Incorrect Base Pairs Accounts for Highly Accurate DNA Replication

Andrew C. Olson,[†] Jennifer N. Patro,[†] Milan Urban,[†,‡] and Robert D. Kuchta*[,†]

[†]Department of Chemistry and Biochemistry, University of Colorado, UCB 215, Boulder, Colorado 80309, United States

[‡]Department of Organic Chemistry, Faculty of Science and Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic

Supporting Information

ABSTRACT: To better understand the energetics of accurate DNA replication, we directly measured ΔG^{o} for the incorporation of a nucleotide into elongating dsDNA in solution ($\Delta G^{o}_{incorporation}$). Direct measurements of the energetic difference between synthesis of correct and incorrect base pairs found it to be much larger than previously believed (average $\Delta\Delta G^{o}_{incorporation}$ = 5.2 ± 1.34 kcal mol^{-1}). Importantly, these direct measurements indicate that $\Delta\Delta G^{o}_{incorporation}$ alone can account for the energy required for highly accurate DNA replication. Evolutionarily, these results indicate that the earliest polymerases did not have to evolve sophisticated mechanisms to replicate nucleic acids; they may only have had to take advantage of the inherently more favorable ΔG^{o} for polymerization of correct nucleotides. These results also provide a basis for understanding how polymerases replicate DNA (or RNA) with high fidelity.

A hallmark of DNA replication is its low error frequency. Replicative DNA polymerases accurately copy the cell's genome, discriminating between four chemically similar substrates (dATP, dCTP, dTTP, and dGTP) during each polymerization event. In the absence of proofreading exonucleases, these enzymes typically make a mistake only once every 1,000 to 1,000,000 incorporation events.[1] While it has been well documented that different polymerases use different mechanisms to achieve their accuracy,[2−6] how they obtain the energy to so effectively differentiate between right and wrong nucleotides has remained unclear.[5,7−9] The prevailing hypothesis posits that the energy difference between correct and incorrect base pair formation is small and the polymerase must, therefore, greatly amplify this difference to attain high levels of fidelity.[1,7−12] However, this idea derives from studies that approximated the $\Delta\Delta G^{o}$ (~0.2−3 kcal mol^{-1}) between right and wrong base pairs using the melting profiles of duplex DNA.[10,13] We have now directly measured ΔG^{o} for the incorporation of a nucleotide ($\Delta G^{o}_{incorporation}$). These studies showed that the $\Delta\Delta G^{o}$ for forming correct versus incorrect base pairs is large ($\Delta\Delta G^{o}_{incorporation}$ ranges from 3.52 ± 0.80 to 6.98 ± 0.17 kcal mol^{-1} (mean = 5.2 ± 1.34 kcal mol^{-1})). Thus, the energetics of base pairing can account for an average misincorporation frequency of <10^{-3} per nucleotide polymerized without any amplification of $\Delta\Delta G^{o}_{incorporation}$, discrimination comparable to the level achieved by high-fidelity polymerases.

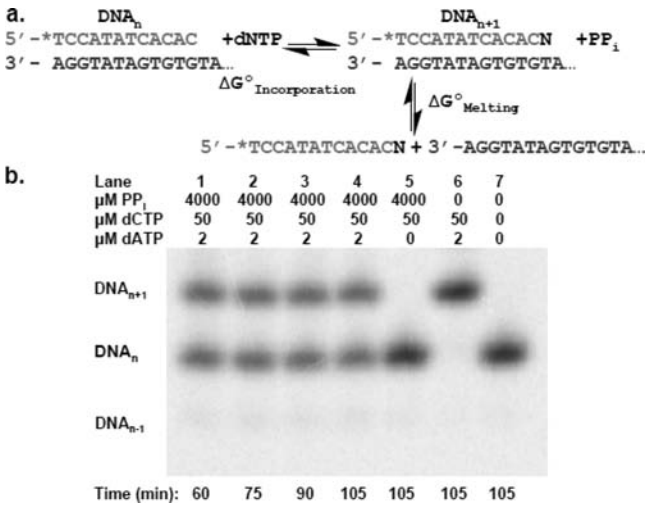

Figure 1. DNA$_n$ ↔ DNA$_{n+1}$ reaction. (a) Pictorial depiction of $\Delta G^{o}_{incorporation}$ versus $\Delta G^{o}_{melting}$. * indicates [^{32}P]-phosphate. (b) Gel of correct incorporation of dATP into Primer C/DNA$_t$; 50 μM dCTP prevents pyrophosphorolysis of DNA$_n$ to DNA$_{n-1}$.

We measured ΔG^{o} for polymerization of a correct dNTP ($\Delta G^{o}_{incorporation}$) for each correct incorporation event (Figure 1a, Table 1). Reactions containing 5′-[^{32}P]-DNA$_n$, the next correct dNTP needed for elongation of DNA$_n$ into DNA$_{n+1}$,

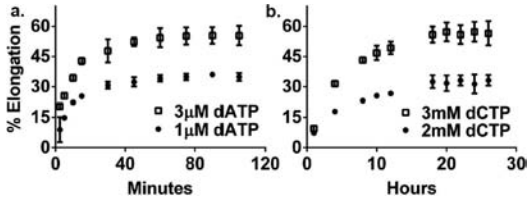

Figure 2. Time course of Primer C/DNA$_t$ elongation. (a) Correct incorporation of 1 μM and 3 μM dATP. (b) Misincorporation of 2 mM and 3 mM dCTP. All assays contained 4 mM pyrophosphate. Average results of two independent experiments are displayed with the estimated error (±standard deviation).

Received: October 31, 2012
Published: January 14, 2013

   dx.doi.org/10.1021/ja309866m | J. Am. Chem. Soc. 2013, 135, 1205−1208

Table 1. Primer−Template Sequences[a]

primer−template	incorp. event	ΔG° kcal/mol	$\Delta\Delta G^{\circ}$ kcal/mol
Primer T/DNA$_t$			
TCCATATCACAT	A → T	−4.68 ± 0.10	
AGGTATAGTGTA**T**GTCTTATCATCT	T → T	+0.52 ± 0.15	5.20 ± 0.18
Primer T/DNA$_t$ (BF)			
TCCATATCACAT	A → T	−4.88 ± 0.15	
AGGTATAGTGTA**T**GTCTTATCATCT	T → T	+0.96 ± 0.04	5.84 ± 0.16
Primer T/DNA$_t$ (KF)			
TCCATATCACAT	A → T	−4.97 ± 0.17	
AGGTATAGTGTA**T**GTCTTATCATCT	T → T	N/A	N/A
Primer C/DNA$_t$			
TCCATATCACAC	A → T	−4.64 ± 0.10	
AGGTATAGTGTG**T**ATCTTATCATCT	C → T	−0.11 ± 0.17	4.54 ± 0.20
Primer G/DNA$_t$			
TCCATATCACCG	A → T	−5.12 ± 0.16	
AGGTATAGTGGC**T**ATCTTATCATCT	G → T	−0.32 ± 0.59	4.81 ± 0.62
Primer T/DNA$_c$			
TCCATATCACAT	G → C	−5.08 ± 0.14	
AGGTATAGTGTA**C**TTCTTATCATCT	T → C	−0.55 ± 0.13	4.52 ± 0.19
Primer C/DNA$_c$			
TCCATATCACAC	G → C	−5.73 ± 0.11	
AGGTATAGTGTG**C**TTCTTATCATCT	C → C	+0.81 ± 0.12	6.54 ± 0.16
Primer A/DNA$_c$			
TCCATATCACGA	G → C	−6.20 ± 0.10	
AGGTATAGTGCT**C**AACTTATCATCT	A → C	−0.61 ± 0.12	5.58 ± 0.16
Primer T/DNA$_g$			
TCCATATCACAT	C → G	−5.09 ± 0.08	
AGGTATAGTGTA**G**TTCTTATCATCT	T → G	−1.57 ± 0.79	3.52 ± 0.80
Primer A/DNA$_g$			
TCCATATCACGA	C → G	−6.04 ± 0.04	
AGGTATAGTGCT**G**AACTTATCATCT	A → G	+0.95 ± 0.17	6.98 ± 0.17
Primer G/DNA$_g$			
TCCATATCACCG	C → G	−5.78 ± 0.20	
AGGTATAGTGGC**G**AACTTATCATCT	G → G	+0.18 ± 0.55	5.96 ± 0.58
Primer C/DNA$_a$			
TCCATATCACAC	T → A	−4.30 ± 0.06	
AGGTATAGTGTG**A**TTCTTATCATCT	C → A	+0.19 ± 0.29	4.49 ± 0.29
Primer A/DNA$_a$			
TCCATATCACGA	T → A	−4.86 ± 0.09	
AGGTATAGTGCT**A**GGCTTATCATCT	A → A	−1.14 ± 0.29	3.73 ± 0.31
Primer G/DNA$_a$			
TCCATATCACCG	T → A	−4.63 ± 0.08	
AGGTATAGTGGC**A**TTCTTATCATCT	G → A	+1.52 ± 0.27	6.15 ± 0.28

[a]Average results of two independent experiments are displayed with the estimated error (±standard deviation). Within each experiment, ΔG° was determined at three different dNTP concentrations in quadruplicate. The underlined base is the templating position. BF denotes that *B. stearothermophilus* Large Fragment was used. KF denotes that Klenow Fragment (3′→5′ exo-) was used. Vent$_R$ (exo-) DNA Polymerase was used in all other cases.

pyrophosphate, and a trace amount of an exonuclease-deficient DNA polymerase were allowed to reach equilibrium (~60 min (Figure 2a). ΔG° values were always obtained at three different dNTP concentrations to ensure their accuracy and reproducibility. In contrast to previous studies that measured ΔG° when the DNA was bound to the polymerase,[4,14,15] we used a large excess of DNA such that the polymerase acted only as a catalyst, i.e. measured $\Delta G^{\circ}_{\text{incorporation}}$ for the reaction in solution. To avoid shortening of DNA$_n$ via pyrophosphorolysis, the reactions always contained ~50 μM of the dNTP present at the 3′ terminus (the nth position) of DNA$_n$. This concentration sufficed to prevent shortening of the DNA$_n$ but did not result in the misincorporation of this dNTP into the n+1 position (See below and Figure 1b, lane 5). The $\Delta G^{\circ}_{\text{incorporation}}$ for correct dNTP polymerization ranged from −4.3 ± 0.06 to −6.2 ± 0.10 kcal mol^{-1} and the average $\Delta G^{\circ}_{\text{incorporation}}$ was −5.2 ± 0.4 kcal mol^{-1}.

The ΔG° for the polymerization of an incorrect dNTP ($\Delta G^{\circ}_{\text{misincorporation}}$) was determined for all 12 possible misincorporation events (Table 1), and ranged from 1.52 ± 0.27 to −1.57 ± 0.79 kcal mol^{-1} with an average $\Delta G^{\circ}_{\text{misincorporation}}$ of 0.13 ± 1.28 kcal mol^{-1}. These reactions differed from those for correct incorporation in that they required ~18 h to attain equilibrium (Figure 2b) due to the slower rate of misincorporation, they contained higher concentrations of the incorrect dNTP and the template

dx.doi.org/10.1021/ja309866m | J. Am. Chem. Soc. 2013, 135, 1205−1208

sequences were constructed so as not to require a second dNTP to prevent shortening of the DNA_n via pyrophosphorolysis. In the absence of pyrophosphate, both correct and incorrect incorporation reactions were able to proceed to completion (full extension of DNA_n to DNA_{n+1}) over the time course of the experiment (Figure 1b, lane 6 and Figure S1, lane 2). With each misincorporation reaction, we observed that the percentage of DNA_n that was elongated to DNA_{n+1} did not change significantly after 18 h (example shown in Figure 2b), indicating that the $DNA_n \leftrightarrow DNA_{n+1}$ reaction had reached equilibrium. Additionally, after 18 h addition of the correct dNTP (1 mM) for conversion of any remaining DNA_n to DNA_{n+1} followed by a 1 h incubation period resulted in complete extension of any remaining DNA_n into DNA_{n+1}, indicating that the enzyme was still active (data not shown).

The sequences of the primer templates used to measure misincorporation were designed to prevent net pyrophosphorolysis of the primer strand (DNA_n) during the long incubations required to allow the reactions to reach equilibrium. Both the misincorporated nucleotide (i.e., at the n+1 position) and the nth nucleotide of primer strand were identical.[a] Thus, if pyrophosphorolysis of the nucleotide at the primer terminus occurred, a relatively high concentration of this just-removed dNTP was present, thereby allowing the polymerase to immediately replace the terminal nucleotide. Ultimately, this approach succeeds because ΔG° for a correct incorporation reaction is much more negative than ΔG° for a misincorporation reaction.

We used three different exonuclease-deficient polymerases from two different evolutionary families to demonstrate that the enzyme acts only as a catalyst and does not affect $\Delta G^\circ_{incorporation}$. *Bacillus stearothermophilus* Large Fragment (BF, an A family enzyme), $Vent_R$ (exo^-) DNA Polymerase (a B family enzyme), and Klenow Fragment (KF (exo^-), an A family enzyme) were compared using Primer T/DNA_t. All three enzymes gave similar $\Delta G^\circ_{incorporation}$ values for correct incorporation of dATP opposite a templating T (Table 1). Only the thermostable enzymes, BF and $Vent_R$, could be compared for misincorporation of dTTP opposite the templating T due to the 18 h incubation required to achieve equilibrium at 37 °C. Again, similar $\Delta G^\circ_{misincorporation}$ values were measured with both polymerases (Table 1). Together, these data indicate that $\Delta G^\circ_{incorporation}$ is polymerase independent, as one would predict for the polymerase acting as a catalyst.

The $\Delta\Delta G^\circ_{incorporation}$ between right and wrong dNTPs varied from 3.52 ± 0.80 to 6.98 ± 0.17 kcal mol^{-1} with an average $\Delta\Delta G^\circ_{incorporation}$ of 5.2 ± 1.34 kcal mol^{-1}, enough energy on average to account for misincorporation frequencies $<10^{-3}$ per nucleotide polymerized and close to those observed with high fidelity polymerases.[10] Thus, DNA polymerases could achieve high fidelity with little, if any, amplification of $\Delta\Delta G^\circ_{incorporation}$.

To determine if the large $\Delta\Delta G^\circ_{incorporation}$ is independent of primer–template length, we compared polymerization of a correct (dTTP) and incorrect (dCTP) nucleotide using two DNAs with different duplex lengths but identical sequences around the polymerization site, Primer C_{long}/DNA_a (a 27-base pair duplex) and Primer C/DNA_a, (a 12-base pair duplex). The incorporation of dTTP and dCTP opposite a template A yielded a $\Delta\Delta G^\circ_{incorporation}$ of 4.37 ± 0.13 kcal mol^{-1} on Primer C_{long}/DNA_a, very similar to the $\Delta\Delta G^\circ_{incorporation}$ of 4.49 ± 0.29 kcal mol^{-1} on Primer C/DNA_a (Table 1 and Table S1 in Supporting Information [SI]). Thus, the large $\Delta\Delta G^\circ_{incorporation}$ is independent of template length for identical sequence contexts.

We measured the correct incorporation events within the context of three different sequences to ask if the sequence could affect $\Delta G^\circ_{incorporation}$. Comparing these values showed that, while sequence affected $\Delta G^\circ_{incorporation}$ by up to 1.1 kcal mol^{-1}, the values were always highly negative, and a large $\Delta\Delta G^\circ_{incorporation}$ between right and wrong dNTPs was always observed (Table 1). Elucidating the cause of this sequence dependence of $\Delta G^\circ_{incorporation}$ will, however, require a much more extensive investigation.

To provide insights into the importance of Watson–Crick hydrogen bonding during dNTP polymerization, we examined Primer $C/DNA_{abasic1}$. This DNA is identical to Primer C/DNA_t except the T in the templating position has been replaced by an abasic site (Table S1 in SI). Unlike the generation of a correct base pair, only phosphodiester bond formation and stacking of the base from the incoming dNTP can drive incorporation. Polymerization of purine dNTPs was significantly more favorable than polymerization of pyrimidine dNTPs (by ~1.8 kcal mol^{-1}) consistent with stacking of purines being more favorable than stacking of pyrimidines (Table S1 in SI) and as predicted by the differing stacking potentials of the bases.[16] Similar results were obtained with a DNA that contained four consecutive abasic sites, indicating that the identity of the templating nucleotide at the n+2 position does not affect dNTP incorporation opposite an abasic site at the n+1 position (Table S1 in SI, Primer $C/DNA_{abasic4}$). The lack of a templating base resulted in a much less favorable $\Delta G^\circ_{incorporation}$ than when the correct templating base was present. Potentially, this could result either from the lack of Watson–Crick hydrogen bonds and/or altered stacking interactions of the template base at the n+1 and/or n+2 position upon dNTP incorporation. We suspect that hydrogen bonding and base stacking are intrinsically linked; if a base pair can form Watson–Crick hydrogen bonds, it will help position the bases for optimum base stacking, and the stacking of bases will likewise favorably align the base pair for hydrogen bonding.

These data show that the $\Delta\Delta G^\circ$ between right and wrong base pair formation in DNA is much larger than previously believed and is sufficient to account for most, but not quite all, of the discrimination exhibited by high fidelity polymerases. This contrasts with current dogma, which postulates that polymerases must greatly amplify $\Delta\Delta G^\circ_{incorporation}$ to achieve high fidelity.[1,7−12] However, this model is based upon melting profiles of dsDNA containing matched or mismatched basepairs at the 3′-terminus of a primer–template[10,13] (i.e., $\Delta G^\circ_{melting}$ (Figure 1a) rather than from direct measurements of $\Delta G^\circ_{incorporation}$. Why, however, should these melting studies give such different results than direct measurement of $\Delta\Delta G^\circ_{incorporation}$? DNA melting is a highly cooperative process, and previous studies have shown that the effect of a mismatch is very position dependent.[17,18] If the mismatch is placed in the middle of a DNA duplex as opposed to near one end, $\Delta G^\circ_{melting}$ is greatly altered, raising the question of whether melting profiles are the best way to determine the energetics for the generation of new base pairs, as occurs during DNA synthesis (i.e., $\Delta G^\circ_{incorporation}$ (Figure 1a).[10,19] The smaller effect of a mismatch at the primer terminus likely results from the mismatch at the primer terminus only disrupting one neighboring stacking interaction, whereas an internal mismatch disrupts two stacking interactions (one on either side of the mismatch). Previous studies have shown that stacking

dx.doi.org/10.1021/ja309866m | J. Am. Chem. Soc. 2013, 135, 1205−1208

interactions, even in the absence of Watson−Crick hydrogen bonding, have a significant impact on DNA melting thermodynamics.[20]

Evolutionarily, this large $\Delta\Delta G^{o}$ may have simplified the fidelity problem for the first nucleotide polymerases. Rather than having to develop sophisticated mechanisms to accurately replicate nucleic acids, they could have taken advantage of the much greater stability of correct base pairs. The more favorable binding of a correct dNTP to a templating base would favor the synthesis of correct base pairs opposite a nucleic acid template.

However, in terms of today's enzymes and thinking about how polymerases obtain fidelity, several issues must be considered. First, DNA synthesis inside of a cell operates under nonequilibrium conditions since one of the products, PP_i, is rapidly destroyed by pyrophosphatase.[21] Second, polymerases generally synthesize DNA quite rapidly (>1000 nucleotides s^{-1} in some cases[9]), and it is unlikely that allowing a reaction to reach equilibrium on an enzyme could accommodate rapid DNA synthesis. Assuming the enzyme can "harvest" this $\Delta\Delta G^{o}$, it could be expressed at any stage of the reaction cycle (dNTP binding, chemistry, etc.) and this could vary for different enzymes, as one observes when comparing how different polymerases discriminate against wrong dNTPs.[2−6,12,22−24] Recent simulations of $\Delta\Delta G^{o}$ of transition-state binding between correct and incorrect bases within the DNA polymerase β-active site are within the range of our $\Delta\Delta G^{o}_{incorporation}$ observations (~5 kcal/mol).[25] In light of these constraints, polymerases may well have developed catalytic strategies to amplify the $\Delta\Delta G^{o}$ between right and wrong base pairs. Finally, it remains to be seen if the different structures of DNA/RNA and RNA/RNA duplexes provide different base-pairing energetics, thus requiring polymerases that generate these duplexes to adopt different catalytic strategies.

■ ASSOCIATED CONTENT

Supporting Information

Abbreviations used, the experimental details, and the data for polymerization of dNTPs opposite an abasic site and on longer templates. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

kuchta@colorado.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by NIH Grant AI59764.

■ ADDITIONAL NOTE

[a]The conditions required to measure $\Delta G^{o}_{misincorporation}$ (high PP_i concentrations and long incubation times) could result in substantial pyrophosphorolysis of DNA_n. Avoiding this problem required that the nucleotide at the primer 3′ terminus (the nth position) of DNA_n be the same as the nucleotide for which we measured misincorporation at the n+1 position For example, misincorporation of only dCTP could be measured with Primer C/DNA_t (Table 1). If the nth nucleotide were removed via pyrophosphorolysis, the high level of dCTP in combination with the favorable ΔG^{o} of correct dNTP polymerization ensured it was rapidly replaced. If the nth nucleotide were different (for example A) from the misincorporated nucleotide (dCTP), the high levels of dATP needed to replace an A removed via pyrophosphorolysis would have competed with dCTP during misincorporation.

■ REFERENCES

(1) Kunkel, T. A.; Bebenek, K. *Annu. Rev. Biochem.* **2000**, *69*, 497.
(2) Capson, T. L.; Peliska, J. A.; Kaboord, B. F.; Frey, M. W.; Lively, C.; Dahlberg, M.; Benkovic, S. J. *Biochemistry* **1992**, *31*, 10984.
(3) Kuchta, R. D.; Benkovic, P.; Benkovic, S. J. *Biochemistry* **1988**, *27*, 6716.
(4) Kuchta, R. D.; Mizrahi, V.; Benkovic, P. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 8410.
(5) McCulloch, S. D.; Kunkel, T. A. *Cell Res.* **2008**, *18*, 148.
(6) Wong, I.; Patel, S. S.; Johnson, K. A. *Biochemistry* **1991**, *30*, 526.
(7) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.
(8) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83.
(9) Berdis, A. J. *Chem. rev.* **2009**, *109*, 2862.
(10) Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C.; Cheong, C.; Tinoco, I., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252.
(11) Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493.
(12) Joyce, C. M.; Benkovic, S. J. *Biochemistry* **2004**, *43*, 14317.
(13) Loeb, L. A.; Kunkel, T. A. *Annu. Rev. Biochem.* **1982**, *51*, 429.
(14) Dahlberg, M. E.; Benkovic, S. J. *Biochemistry* **1991**, *30*, 4835.
(15) Eger, B. T.; Benkovic, S. J. *Biochemistry* **1992**, *31*, 9227.
(16) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213.
(17) Guo, Z.; Liu, Q.; Smith, L. M. *Nat. Biotechnol.* **1997**, *15*, 331.
(18) Piao, X.; Sun, L.; Zhang, T.; Gan, Y.; Guan, Y. *Acta Biochim. Pol.* **2008**, *55*, 713.
(19) Aboul-ela, F.; Koh, D.; Tinoco, I., Jr.; Martin, F. H. *Nucleic acids rese.* **1985**, *13*, 4811.
(20) Bommarito, S.; Peyret, N.; SantaLucia, J., Jr. *Nucleic Acids Res.* **2000**, *28*, 1929.
(21) Kornberg, A.; Rao, N. N.; Ault-Riche, D. *Annu. Rev. Biochem.* **1999**, *68*, 89.
(22) Frey, M. W.; Sowers, L. C.; Millar, D. P.; Benkovic, S. J. *Biochemistry* **1995**, *34*, 9185.
(23) Johnson, K. A. *Annu. Rev. Biochem.* **1993**, *62*, 685.
(24) Purohit, V.; Grindley, N. D.; Joyce, C. M. *Biochemistry* **2003**, *42*, 10200.
(25) Rucker, R.; Oelschlaeger, P.; Warshel, A. *Proteins* **2010**, *78*, 671.