Factors Determining the Deriving Force of DNA Formation: Geometrical Differences of Base Pairs, Dehydration of Bases, and the Arginine Assisting

Lixiang Sun,[†] Robert I. Cukier,[‡] and Yuxiang Bu*[,†,‡]

Institute of Theoretical Chemistry, Shandong University, Jinan, 250100, P. R. China, and the Department of Chemistry, Michigan State University, East Lansing, Michigan 48824-1322

Received: June 12, 2006; In Final Form: December 12, 2006

The mechanism of the fidelity synthesis of DNA associated with the process of dGTP combination to the DNA template was explored. The exclusion of water molecules from the hydrated DNA bases can amplify the energy difference between the correct and incorrect base pairs, but the effect of the water molecules on the Gibbs free energy of formation is dependent on the binding sites for the water molecules. The water detachment from the incoming dNTP is not the only factor but the first step for the successful replication of DNA. The second step is the selection of the DNA polymerase on the DNA base pair through the comparison between the correct DNA base and the incorrect DNA base. The bonding of the Arg668 with the incoming dNTP can enlarge the Gibbs free energies of formation of the base pairs, especially the correct base pairs, thus increasing the driving force of DNA formation. When the DNA base of the primer terminus is correct, the extension of the guanine and the adenine is quicker than that of the cytosine and the thymine because of the hydrogen bonding fork formation of Arg668 with the minor groove of the primer terminus and the ring oxygen of the deoxyribose moiety of the incoming dNTP. Because of the geometry differences of the incorrect base pairs with the correct base pairs, the effect from the DNA polymerase is smaller on the incorrect base pair than on the correct base pair, and the extension of a mispair is slower than that of a correct base pair. This decreases the extension rate of the base pair and thus allows proofreading exonuclease activity to excise the incorrect base pair. Arg668 cannot prevent the extension of the GT mispair, as well as the GC correct base pair, and GA and GG mispairs. This may be attributed to the small geometry difference between the GT base pair and the correct AT base pair.

I. Introduction

The base pair mispairing has received considerable attention because of their connection with DNA mutation,[1−23] but only a few theoretical results about the high fidelity synthesis of the DNA have been published.[24−28] The fidelity of the successful replication of DNA is dependent on the proficiency of DNA polymerases in incorporating the correct dNTP into DNA. This high fidelity synthesis can be carried out despite the similarity in energy between correctly and incorrectly paired bases.[24,25] Three mechanisms have been proposed to explain how polymerases can achieve a selectivity of 1 in 10 000 with differences in $\Delta G°$ of less than 1 kcal/mol between correct and incorrect base pairs.[26] (1) Polymerases can exclude the water molecules from the active site of the base pair, thereby amplifying the energy difference between correct and incorrect base pairs.[27] (2) Polymerases can increase the $\Delta\Delta G°$ between the correct and incorrect base pairs by restricting the movement of the DNA.[24] (3) Polymerases can enhance the fidelity by selecting for base pairs of Watson−Crick geometry.[25,28] These three mechanisms are not mutually exclusive. By providing a check for Watson−Crick geometry, the polymerase would have to exclude H_2O from contact with the DNA and restrict the entropy of the DNA thereby increasing the $\Delta G°$ between correct and incorrect base pairs. All three theories involve interactions between the DNA and the polymerase.[29]

X-ray crystallographic studies with the closely related BF (*Bacillus stearothermophilus* DNA polymerase large fragment), Taq, and T7 DNA polymerases predict a structured minor groove with hydrogen bonds between the minor groove of the DNA and the water molecules and/or the amino acid side chains of KF^-.[30−32] The relevant interaction between the DNA polymerase and the DNA has been shown in Figure 1. This structure is based upon the ternary crystals of T7 DNA polymerases[32] in which the 2′-deoxyadenosine at the primer terminus was replaced by a 2′-dideoxyguanosine.[29]

Utilizing oligodeoxynucleotides containing 3-deazaguanine, McCain et al.[29] found evidence that Arg668 makes a critical hydrogen bond to the minor groove of the primer terminus as well as to the ring oxygen of the deoxyribose group of the incoming dNTP. This experiment supports a mechanism in which Arg668 forms a hydrogen bonding fork between the minor groove of the primer terminus and the ring oxygen of the deoxyribose moiety of the incoming dNTP to align the 3′-hydroxyl group with the α-phosphate of the dNTP. This is one mechanism by which the polymerase uses the geometry of the base pair to modulate the rates of the formation and the extension of the mispair. The results also show that the hydrogen bond between Arg668 and the N3-position of guanine exists and is important during the rate-limiting step of the extension past the GC base pair, as well as the GA and GG mispairs, but not the GT mispair.[29]

The main aim of this work is to explain the nature of the fidelity of the successful replication of DNA from the viewpoint

* Corresponding author: Yuxiang Bu, e-mail byx@sdu.edu.cn.
[†] Shandong University.
[‡] Michigan State University.

10.1021/jp063645f CCC: $37.00
Published on Web 02/01/2007

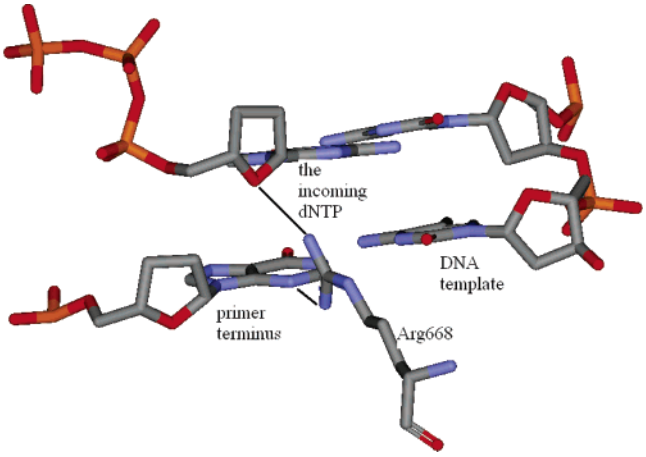
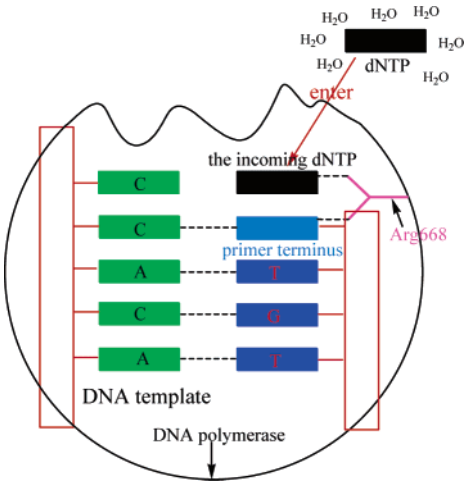


Figure 1. Representation of the hydrogen bonding fork between Arg668 and guanine at the 3-terminus and the deoxyribose ring oxygen of the incoming dGTP. This structure is based upon the ternary crystals of T7 DNA polymerases[32] in which the 2′,3′-dideoxyadenosine at the primer terminus was replaced by a 2′,3′-dideoxyguanosine. In the T7 DNA polymerase crystal structure, the distance between Arg429 (denoted as Arg668 here) and the N3-position of adenine at the primer terminus is 2.86 Å and the distance between Arg429 and the ring oxygen of the incoming 2′,3′-dideoxyguanosine 5′-triphosphate is 2.81 Å.

SCHEME 1: The Sketch Maps of the Formation and Extension of the DNA Base Pairs in the DNA Polymerase



of energetics and to explore how the water molecules and Arg668 modulate the relative preference between the pairing and the mispairing. We will probe into the mechanism of the fidelity synthesis of DNA associated with the process of dGTP combination to the DNA template step by step (shown in Scheme 1). First of all, we will clarify the role of excluding water molecules from the nucleoside if they can change the $\Delta\Delta G°$ between the correct and incorrect base-pairings. Second, the principle of the selection of the DNA polymerase on the incoming dNTP will be explained. Finally, we will show the effect of the extension rate of the DNA on the fidelity of the DNA replication. In addition, since the detachment of water molecules from the active site of the nucleoside and the interaction between the DNA and polymerase are important to the enhancement of the fidelity of the successful replication of DNA, many details regarding the mechanism need to be explored: (1) What is the dominating factor in the successful replication of DNA between the exclusion of the water molecules from the active site and the DNA−polymerase interaction? (2) Which sites of the nucleoside bound with water molecules have more effect on the replication? (3) The hydrogen bond between Arg668 and the N3-position of guanine exists and is important during the rate-limiting step of the extension past the GC base pair, as well as the GA and GG mispairs, but not the GT mispair. The reason is still unclear. Clearly, the resolution of these questions may provide some useful information to understanding the behavior of the successful replication of DNA.

TABLE 1: The ΔE (in kcal/mol) and $\Delta G_f°$ (in kcal/mol) of the Correct Base Pairs and Incorrect Base Pairs[a]

	GC	GT	GA-1	GA-2	GG	AT
ΔE	−24.6	−13.3	−12.9	−13.6	−12.1	−11.7
	−24.7	−13.5	−13.0	−13.8	−12.0	−11.9
$\Delta G_f°$	−11.9	−2.2	−0.3	−1.3	−1.5	0.4
	−12.6	−2.3	0.4	−0.7	−1.5	0.7

[a] The first row of data was obtained with the B3LYP/6-311+G* method, and the second row of data was obtained with the B3LYP/6-311++G** method.

II. Computational Methods

Molecular geometries and harmonic vibrational frequencies of the considered structures were obtained at the density functional theory level. The nonlocal hybrid three-parameter B3LYP density functional approach,[33−35] as implemented in the Gaussian 98 program,[36] was used in this work along with a 6-311+G* basis set because it shows good accuracy[37−40] with comparatively low cost in computer time. The geometric parameters and total energies of the complexes were given in Supporting Information. Because of the larger need of computer times for the systems bound with the Arg668, the optimizations and calculations of frequencies of the purine−purine base pairs bound with the Arg668 were performed at the B3LYP/6-31+G* level, and only the single-point calculations were carried out with the B3LYP/6-311+G* method employing the B3LYP/6-31+G*-optimized geometries. In addition, the Arg668 amino acid residue was modeled by the methylguanidinium in order to cut down the CPU time but is still denoted by Arg668 in the following analyses.

To further justify our results, we reoptimized six base pairs using the B3LYP/6-311++G** method. The ΔE and ΔG of these six base pairs were determined and the values were listed in Table 1. The B3LYP/6-311+G* method gave essentially identical results as the B3LYP/6-311++G** method.

The interaction energy, ΔE, between the nucleic bases was evaluated as follows:

$$\Delta E = E_{compl} - E_A - E_B$$

where E_{compl} was the total energy of the optimized molecular complex, while E_A and E_B were the total energies of both the optimized subsystems, respectively. Finally, to obtain the true interaction energy, the basis set superposition error (BSSE) corrections were considered using the counterpoise procedure.[41]

The Gibbs free energy of formation, $\Delta G_f°$, of the base pair AB was determined according to the following formula:

$$\Delta G_f°(AB) = G°(AB) - (G°(A) + G°(B))$$

where the $G°(AB)$ was the Gibbs free energy of the base pair AB, while the $G°(A)$ and $G°(B)$ were the Gibbs free energies of the DNA base A and B, respectively.

The $G°$ was determined according to the formula:

$$G° = E_{tot} + \Delta E_{corr}$$

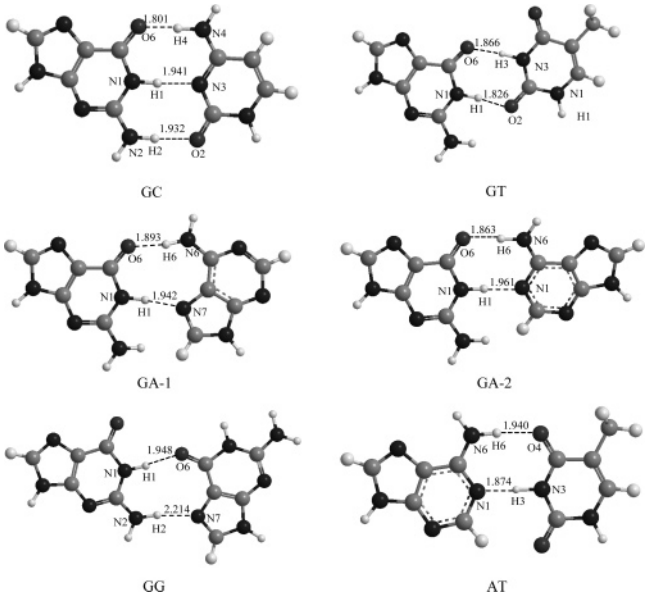


Figure 2. The geometries of the correct and incorrect base pairs.

TABLE 2: The ΔG_f° (in kcal/mol) of the Correct Base Pairs and Incorrect Base Pairs and Their Derivatives

	GC	GT	GA-1	GA-2	GG	AT
bare	−11.9	−2.2	−0.3	−1.3	−1.5	0.4
H_2O-1	−12.4	−2.2	−0.2	−1.3	−3.0	
H_2O-2	−11.0	−2.2	−0.2	−1.1	−1.8	
$2H_2O$	−11.6	−2.3	−0.1	−1.2	−2.6	
Arg−	−20.2	−5.6	−0.9	−3.8	−11.8	−3.2

where the E_{tot} denoted the total electronic energy of the molecule and the ΔE_{corr} was the thermal correction to the Gibbs free energy. The Gibbs free energies of the purine−purine base pairs linked by Arg668 were estimated by the total electronic energies obtained at the B3LYP/6-311+G* level and the thermal corrections to the Gibbs free energies obtained at the B3LYP/6-31+G* level.

III. Results and Discussion

A. The Driving Force of the Formation of DNA: Gibbs Free Energy. Six structures were obtained by optimizing the structures of the correct and incorrect base pairs including the guanine, and their corresponding geometries were drawn in Figure 2. The free energy differences between the correct and incorrect base pairs are less than 1 kcal/mol.[26] John Petruska et al. considered that polymerases can amplify the energy difference between the correct and incorrect base pairs by excluding water molecules from the active site.[27] We calculated the ΔG_f° of the correct and incorrect base pairs at the gas-phase condition (the data are listed in Table 2). Within the DNA stacked bases, the solvent is excluded and the DNA interior has a relatively low dielectric constant.[42,43] From Table 2, it can be seen that the ΔG_f° of GC is smaller than those of the mispairs by 9.7−11.6 kcal/mol. This difference between the correct and incorrect base pairs is enough for the DNA polymerase to select the correct base from the four DNA bases. If the nucleic base is the guanine or the cytosine in the template, the probability of the mispairing is very small. As far as we know, there are no examples for the mispair of the cytosine in the template. By comparison of only the ΔG_f° of the isolated base pairs, it can be deduced that the formations of the GA and GT mispairs are slightly easier than that of the correct AT base pair when the nucleic base in the template is the adenine or the thymine, and the ΔG_f° of the AT base pair is positive, which is in agreement with the results of Gutowski[44] and Laughton.[45] These are obviously different from the facts in the biological system. The nature for the decrease of the formation probability of the GA and GT incorrect base pairs and for the spontaneous formation of the AT base pair is the structural selection of the DNA polymerase on the DNA bases and the change of the AT base pair's ΔG_f° from positive to negative. The details will be discussed in the following section.

B. The Effect of the Water Molecules. Since the exclusion of water molecules from DNA bases can amplify the Gibbs free energy difference between the correct and incorrect base pairs, it is important to investigate which sites of the nucleoside bound with the water molecules have more effect on the replication of the DNA. Because the template DNA in the polymerase is an anhydrous surrounding, we optimized the base pairs on the model in which only the primer is bound with the water molecules. The sites on the primer bound with water molecules can be distinguished as two kinds. One kind of site bound with water molecules is on the exterior of the DNA, which are denoted as $1H_2O$-1 (the water molecule bonding to the O6 and N7 atoms of guanine), $1H_2O$-2 (the water molecule bonding to the N3 atom of guanine), and $2H_2O$, respectively (see Figure 3). Another kind of site is the hydrogen bond site between the DNA bases. We have optimized some multihydrated base pairs at the B3LYP/3-21G* level (see Supporting Information), but the structures of the DNA bases are not coplanar due to the effect of the water molecules. This does not accord with the fact in the biological system. Although we do not gain the ΔG_f° of the multihydrated base pairs, it can be known that the ΔG_f° between the correct and incorrect base pairs is less than 1 kcal/mol.[26]

Table 2 reveals that the ΔG_f° values of the monohydrated base pairs are slightly smaller than those of the corresponding bare base pairs when one water molecule binds to site 1, except for the GA-1 base pair. While one water molecule binds to site 2, the ΔG_f° values of the base pairs are slightly larger than those of the corresponding bare base pairs, except for the GG base pair. Similarly, when two water molecules bind to the primer molecule G, the ΔG_f° values of the hydrated base pairs are almost equal to those of the corresponding bare base pairs. These results indicate that the detachment of the water molecules from the base pairs on the exterior of the DNA has little effect on the ΔG_f°, and the ΔG_f° of the correct and incorrect base pairs are already different from each other significantly. Thus, it can be deduced that the exclusion of water molecules from the sites of the incoming dNTP hydrogen bonded with the base of the DNA template is the dominating factor for the amplification of the energy difference between the correct and incorrect base pairs, but it is not the only factor in the successful replication of DNA. If it is true, the GT and GA mispairs will exist with a large quantity in the biological system.

C. The Selectivity of the Polymerase on the DNA Base Geometries. As viewed from the ΔG_f°, the formation of the GA or GT mispairs is slightly easier than that of the correct AT base pair when the nucleic base pair in the template is adenine or thymine, but the polymerases can select for the correct base pairs on the basis of the geometric constraints because the four Watson−Crick base pairs have a similar topography.[46] Although the general shapes of the Watson−Crick base pairs are similar, the major groove displays differences for each base pair.[47] In contrast, the minor groove of the DNA may play an important role in this geometrical selection because the positions of two hydrogen bond acceptors (N3 of purine

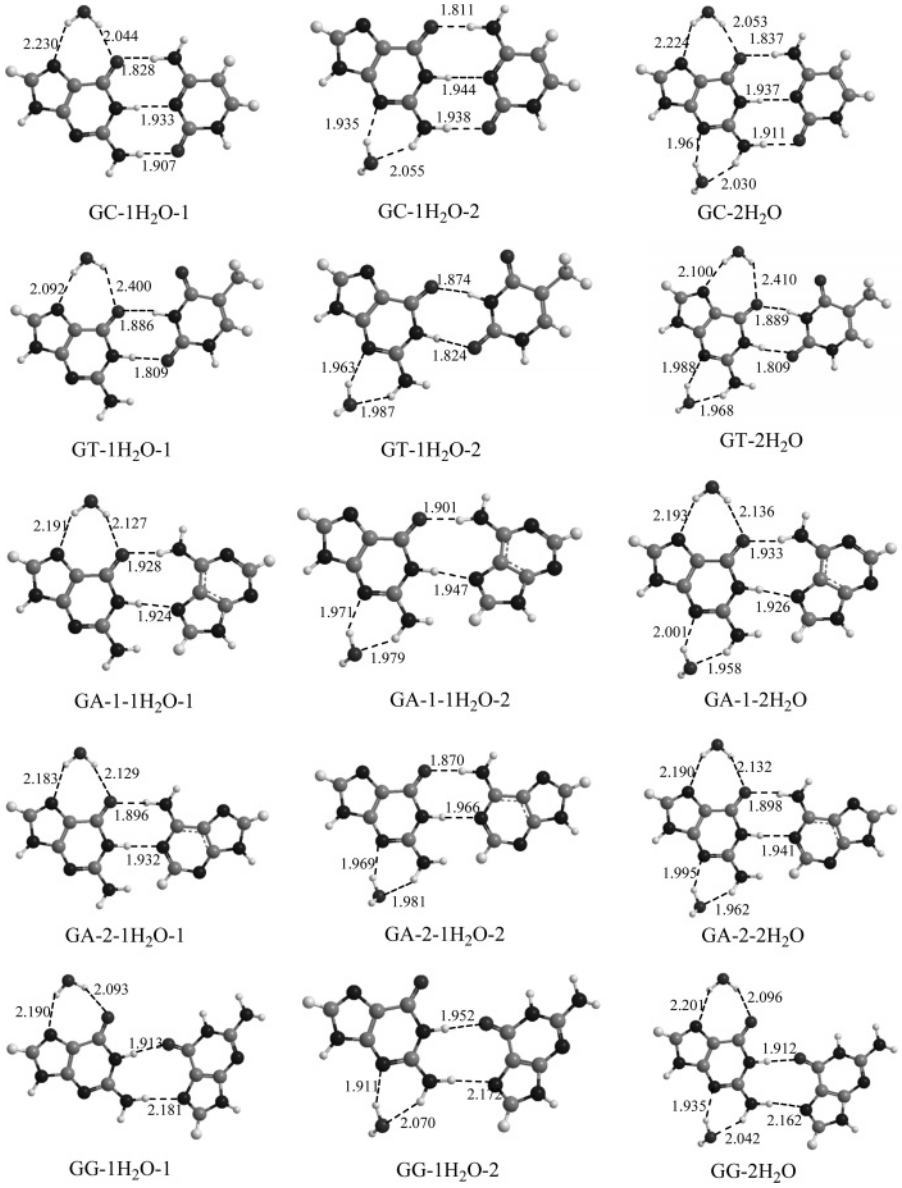


Figure 3. The geometries of the correct and incorrect base pairs bound to one or two water molecules.

and O2 of pyrimidine) are similar in the four Watson−Crick base pairs but different in the mispairs.[47,48]

Comparison of the geometries between the correct and incorrect base pairs displayed in Figure 2 indicates that the correct and incorrect base pairs have many significant geometric differences. Although the N···N distance of the incorrect base pairs GT and GA-1 is close to that of the correct base pair, there are a lot of other differences between the incorrect and correct base pairs. (1) The two hydrogen bond positions of the thymine in the incorrect GT base pair are not the same as those in the correct AT base pair. The two hydrogen bonds form on the sites of the H3(N3) and O2 atoms of the thymine in the incorrect GT base pair, but those form on the sites of the H3-(N3) and O4 atoms of the thymine in the correct AT base pair. (2) The purine can form two hydrogen bonds at the position toward the major groove in the correct and incorrect base pairs. In the two hydrogen bonds, adenine acts as the proton acceptor in one hydrogen bond but acts as the proton donor in the other hydrogen bond, while in the incorrect GT base pair, the guanine acts as the proton acceptor in the two hydrogen bonds. The position of the N3 atom of the guanine relative to the thymine in the incorrect GT base pair is not the same as that of the adenine in the correct AT base pair. It is obvious that the geometries of the incorrect GA-1 and GG base pairs are different from those of the correct TA and CG base pairs because the tropism of the DNA base in the incorrect base pair is reverse to that in the correct base pair. Although the tropism of the DNA base in the GA-2 is the same as that in the correct base pair, the geometry of the GA-2 base pair is significantly different from that of the TA base pair. The position of the N atom linked with the deoxyribose moiety on the thymine in the TA base pair corresponds to that of the N3 atom of the guanine in the incorrect GA-2 base pair. All the structures of the incorrect base pairs are nonplanar except for the GT base pair. The side view of the base pairs are drawn in Figure 4. The other mispairs including the cytosine, adenine, or thymine in the primer are also different from the correct base pairs. These differences between the correct and incorrect base pairs can be distinguished by the DNA polymerase, so the incorrect DNA bases may be put away before they form hydrogen bonds with the template base.

D. The Effect of the Amino Acid Arg668 of the Polymerase. McCain et al. proved that the interaction between the minor groove of the primer and Arg668 of DNA polymerase I

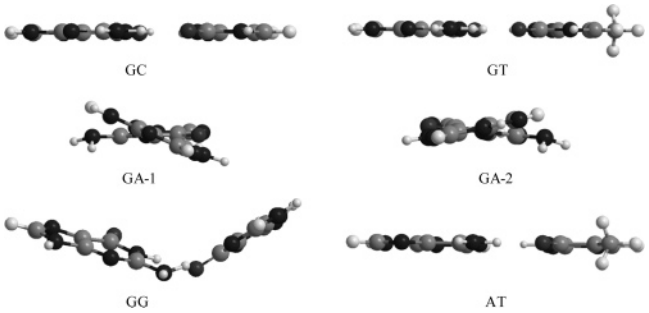


Figure 4. The side views of the correct and incorrect base pairs.

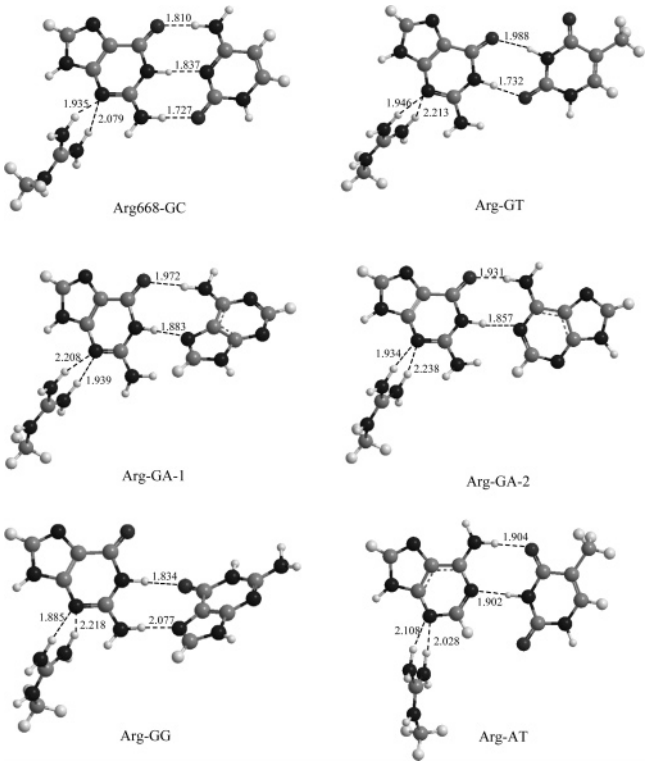


Figure 5. The geometries of the correct and incorrect base pairs bound with Arg668.

of *Escherichia coli* is important for the fidelity of both the formation and the extension of mispairs.[29] On the basis of this experimental observation, the structures of several base pairs bound with Arg668 were designed and optimized. Because of the representation of the hydrogen bonding fork between Arg668 and guanine at the 3′-terminus and between Arg668 and the deoxyribose ring oxygen of the incoming dNTP, the arginine must be protonated as it exists naturally in biological systems. In these structures of the base pairs bound with the Arg668 (as drawn in Figure 5), the Arg668 binds to the N3 atom of the primer guanine through two hydrogen bonds. The two hydrogen bond distances are slightly different from those of the ternary crystals of T7 DNA polymerases.[32] In the T7 DNA polymerase crystal structure, the distances between Arg429 and the N3-position of adenine at the primer terminus are 2.86 and 3.64 Å, respectively.

The data listed in Table 2 indicate that ΔG_f° of the base pairs bound with Arg668 is smaller than that of the corresponding isolated base pairs, implying that the interaction between the DNA base and the DNA polymerase can increase the driving force of the combination of the DNA bases, especially for the GC base pair and the GG base pair. This is also proven by the ΔG_f° changes of the AT base pair. The formation of the AT base pair cannot occur spontaneously in the gas phase because the ΔG_f° is positive, but in the DNA polymerase this process becomes spontaneous. Although the increase of the driving force of the combination of the GG base pair is larger than that of the GC base pair by 2.0 kcal/mol, the overall driving force of the combination of the GC base pair is larger than that of the GG base pair by 8.4 kcal/mol. This is in accordance to the above analyses regarding the isolated base pairs in which the nucleic base is the guanine or the cytosine in the template, and the probability of the mispairing is very small. Although the Arg668 links with the ring oxygen of the deoxyribose moiety of the incoming dNTP through the hydrogen bond in the DNA polymerase, it has the same effect on decreasing the ΔG_f° when the Arg668 links with the N3 atom of the guanine through the hydrogen bond. The decrease of the ΔG_f° is because the Arg668 has one positive charge. The one-electron oxidation and the coupling of Li^+ to the guanine−cytosine base pair can strengthen the interaction between guanine and cytosine[49] and decrease the ΔG_f° of the GC base pair. These analyses indicate that the interaction between the DNA polymerase and the incoming dNTP can increase the driving force of the formation of the DNA; thus the rate of the formation of the DNA may increase although the increase of the rate results primarily from the decrease of the activation energy.

Although the interaction between the Arg668 of the DNA polymerase and the DNA can increase the driving force of the combination of the DNA bases, the disadvantageous geometries of the incorrect base pairs decrease the interaction between the DNA polymerases and the DNA base pair of the primer terminus.

The presence of the hydrogen bond in the GA-1 base pair would require the guanine on the primer terminus to be in the anti conformation (see Figure 6). Potential GA complexes have been observed in the oligodeoxynucleotides.[50−54] The structure GA-1 has the template dA to be in the syn conformation with Hoogsteen hydrogen bonding with the G on the primer strand.[47−49] In GA-2, the adenine on the template is in the anti conformation and would cause the distance between the deoxyribose moieties to lengthen.[53,54] Neither of the GA complexes appears to be optimal for DNA replication.

The GG-Arg668 complex with base pairs has been observed in oligodeoxynucleotides.[55] The presence of a hydrogen bond between the primer guanine and Arg668 requires the guanine at the primer terminus to be in the anti conformation, while the template G is in the syn conformation. The structure of the GG complex found in the crystallography experiments with BF is that the primer G is in the syn conformation while the template G is in the anti conformation.[56] The GG mispair is formed with the incoming dNTP in the syn orientation and the template in the anti conformation, but that extension occurs with the primer G in the anti conformation and the template G in the syn conformation.[29]

The hydrogen bond between Arg668 and the N3-position of guanine exists and is important during the rate-limiting step of the extension past the GC base pair as well as the GA and GG mispairs but not the GT mispair.[29] The loss of selectivity to prevent extension of GT mispair is consistent with the results that Arg668 is not as good as the wild-type in preventing the extension of GT mispair.[57] From the geometries of the correct and incorrect base pairs of the primer terminus (drawn in Figure 2), it can be seen that the position of the N atom linked with the deoxyribose moiety on the guanine and the tropism of the N−H bond, viz. the position of the deoxyribose moiety in the incorrect base pairs, does not lie in the position that they should

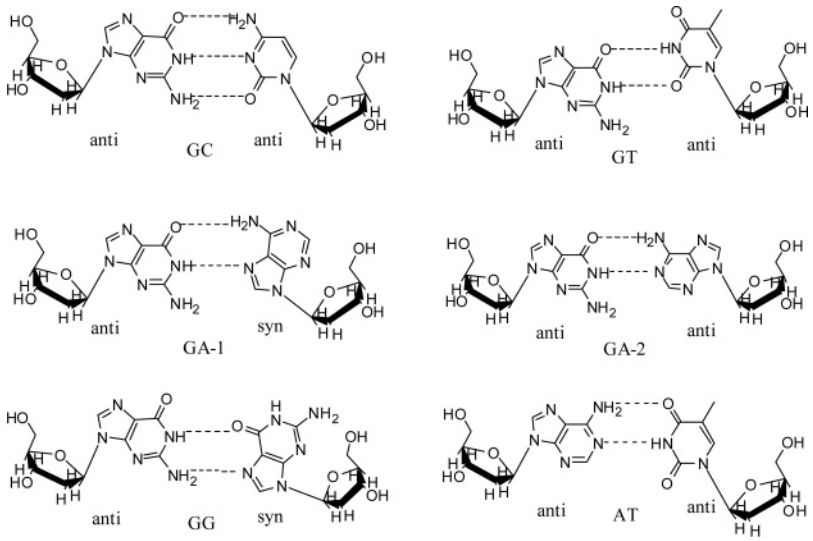


Figure 6. The sketch map of the anti and syn structures of the correct and incorrect base pairs.

be in the correct base pairs, especially in the incorrect GA-1, GA-2, and GG base pairs, while the differences between the incorrect GT base pair and the correct AT base pair are much smaller than those between the incorrect GA-1, GA-2, and the TA base pairs and between the incorrect GG and CG base pairs. Because of the flexibility of the DNA polymerase, the hydrogen bonds between the Arg668 and the minor groove of the primer terminus and between the Arg668 and the ring oxygen of the deoxyribose moiety of the incoming dNTP can form when the incorrect base pair is the GT base pair, since the position of the guanine in the incorrect GT base pair is slightly different from that of the adenine in the correct AT base pair. The incorrect base pair, except for GT, cannot form the phosphodiester with the incoming dNTP quickly, even though the DNA polymerase can distort in some measure, because the geometries of these incorrect base pairs have larger differences from those of the correct base pairs.

The loss of the hydrogen bond between the minor groove of the primer terminus and the Arg668 decreases the rate of the extension past the GC base pair, as well as the GA and GG mispairs, but not the GT mispair.[29] R668A is not as good as the wild-type in preventing the extension of GT mispair.[56] Clearly, the above analyses indicate that the geometries have some differences between the incorrect GT base pair and the correct base pairs and between the incorrect GT base pair and the other incorrect base pairs. We presume that the geometries of the DNA polymerase only have a little change when the hydrogen bond between the minor groove of the primer terminus and the Arg668 of the DNA polymerase is lost. It is possible that these changes make the incoming dNTP lay in the position that suits it to form the phosphodiester bond between the guanine of the incorrect GT base pair of the primer terminus and the incoming dNTP quickly but not between the guanine of the correct GC base pair on the primer terminus and the incoming dNTP as quickly as that of the GT incorrect base pair.

When the DNA base on the primer terminus is adenine, the interaction between the minor groove of the primer terminus and the Arg668 is similar to that of the guanine, and the effect of Arg668 on the extension of DNA is also similar to that of the guanine. When the adenine is the correct base, the extension of DNA is very fast, but if the adenine is the incorrect base, the extension of DNA is very slow. The interaction between Arg668 and the minor groove of the primer terminus is lost when the DNA base on the primer terminus is cytosine or thymine, so the rate of the extension of DNA is smaller than that of the guanine and the adenine on the DNA primer terminus. The decrease of the extension rate of the base pair allows proofreading exonuclease activity to excise the incorrect base pair.

IV. Conclusions

The nature of the fidelity of the successful replication of DNA associated with the process of the dGTP combination to the DNA template is explicated. The analyses of the Gibbs free energies of formation reveal that the water detachment from the DNA base can amplify the energy difference between the correct and incorrect base pairs, but the effect of the water molecules on the Gibbs free energy of formation depends on the binding sites for the water molecules. The effect of the exclusion of the water molecules is not the only factor but the first step for the successful replication of DNA. The second step is the selection of the DNA polymerase on the DNA base pairs through the geometric comparison between the correct DNA bases and the incorrect DNA bases. By this effect, the DNA polymerase will throw away the incorrect DNA bases, thus to further decrease the probability of the mispairing formation, but some mispairs will still form. The interaction between the DNA polymerase and the incoming dNTP can increase the driving force for the formation of the DNA. When the DNA base of the primer terminus is correct, the extension of the guanine and the adenine is quicker than that of the cytosine and the thymine because Arg668 forms a hydrogen bonding fork between the minor groove of the primer terminus and the ring oxygen of the deoxyribose moiety of the incoming dNTP.[29] Because of the geometry differences of the incorrect base pairs and the correct base pairs, the effect from the DNA polymerase is smaller on the incorrect base pair than on the correct base pair, and the extension of a mispair is slower than that of a correct base pair. This slower extension of a mispair would allow proofreading exonuclease activity to excise the incorrect base pair.[29] Arg668 cannot prevent the extension of the GT mispair as well as the GC, GA, and GG mispairs. This may be because the geometry of the GT base pair is only slightly different from that of the correct base pair and the DNA polymerase can flex in some measure.

Acknowledgment. This work is supported by NSFC (Grant Nos. 20573063 and 20633060), NIH (Grant No. GM62790),

NCET and the NSF of Shandong Province (Grant No. Z2003B01), and SCF for ROCS-SEM. Support from the Virtual Laboratory for Computational Chemistry of CNIC and the Supercomputing Center of CNIC-CAS is also acknowledged. A part of the calculations were performed on the Center for Biological Modeling and the Michigan Center for Biological Information Linux clusters at Michigan State University and the High-Performance Computational Center in Shandong University.

Supporting Information Available: Molecular geometry coordinates and total energies of all base pairs considered in this paper. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Duncan, B. K.; Miller, J. H. *Nature* **1980**, *287*, 560.
(2) Wiebauer, K.; Jiricny, J. *Nature* **1989**, *339*, 234.
(3) Wiebauer, K.; Jiricny, J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 5842.
(4) Rideout, W. M., III; Coetzee, G. A.; Olumi, A. F.; Jones, P. A. *Science* **1990**, *249*, 1288.
(5) Coulondre, C.; Miller, J. H.; Farabaugh, P. J.; Gilbert, W. *Nature* **1978**, *274*, 775.
(6) Jones, P. A. *Cancer Res.* **1996**, *56*, 2463.
(7) Kan, L. S.; Chandrasegaran, S.; Pulford, S. M.; Miller, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4263.
(8) Patel, D. J.; Kozlowski, S. A.; Ikuta, S.; Itakura, K. *Biochemistry* **1984**, *23*, 3207.
(9) Waters, S. B.; Akman, S. A. *Mutat. Res.* **2001**, *487*, 109.
(10) Lari, S.-U.; Al-Khodairy, F.; Paterson, M. C. *Biochemistry* **2002**, *41*, 9248.
(11) Hunter, W. N.; Brown, T.; Kennard, O. *J. Biomol. Struct. Dyn.* **1986**, *4*, 173.
(12) Keepers, J. W.; Schmidt, P.; James, T. L.; Jollman, P. A. *Biopolymers* **1984**, *23*, 2901.
(13) Chuprina, V. P.; Poltev, V. I. *Nucleic Acids Res.* **1983**, *11*, 5205.
(14) Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L. S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498.
(15) Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L. S.; Kopka, M. L.; Dickerson, R. E. *Struct. Expression* **1988**, *2*, 27.
(16) Nikonowicz, E. P.; Gorenstein, D. G. *Biochemistry* **1990**, *29*, 8845.
(17) Nikonowicz, E. P.; Meadows, R. P.; Fagan, P.; Gorenstein, D. G. *Biochemistry* **1991**, *30*, 1323.
(18) Li, Y.; Zon, G.; Wilson, W. D. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 26.
(19) Greene, K. L.; Jones, R. L.; Li, Y.; Robinson, H.; Wang, A. H.; Zon, G.; Wilson, W. D. *Biochemistry* **1994**, *33*, 1053.
(20) Maskos, K.; Gunn, B. M.; LeBlanc, D. A.; Morden, K. M. *Biochemistry* **1993**, *32*, 3583.
(21) Gao, X. L.; Patel, D. J. *J. Am. Chem. Soc.* **1988**, *110*, 5178.
(22) Zharkov, D. O.; Gilboa, R.; Yagil, I.; Kycia, J. H.; Gerchman, S. E.; Shoham, G.; Grollman, A. P. *Biochemistry* **2000**, *39*, 14768.
(23) Scholdberg, T. A.; Nechev, L. V.; Merritt, W. K.; Harris, T. M.; Harris, C. M.; Lloyd, R. S.; Stone, M. P. *Chem. Res. Toxicol.* **2005**, *18*, 145.
(24) 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.
(25) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.
(26) Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493.
(27) Petruska, J.; Sowers, L. C.; Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1559.
(28) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191.
(29) McCain, M. D.; Meyer, A. S.; Schultz, S. S.; Glekas, A.; Epratt, T. E. *Biochemistry* **2005**, *44*, 5647.
(30) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391*, 304.
(31) Li, Y.; Korolev, S.; Waksman, G. *EMBO J.* **1998**, *17*, 7514.
(32) Doublie, S.; Tabor, S.; Long, A. M.; Richardson, C. C.; Ellenberger, T. *Nature* **1998**, *391*, 251.
(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(34) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.
(35) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
(36) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
(37) Galbraith, J. M.; Schaefer, H. F., III. *J. Chem. Phys.* **1996**, *105*, 862.
(38) Roesch, N.; Trickey, S. B. *J. Chem. Phys.* **1997**, *106*, 8940.
(39) Wesolowski, S. S.; Leininger, M. L.; Pentchew, P. N.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **2001**, *123*, 4023.
(40) Rienstra-Kiracofe, J. C.; Barden, C. J.; Brown, S. T.; Schaefer, H. F., III. *J. Phys. Chem. A* **2001**, *105*, 524.
(41) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(42) Colson, A. O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1992**, *96*, 9787.
(43) Colson, A. O.; Besler, B.; Close, D. M.; Sevilla, M. D. *J. Phys. Chem.* **1992**, *96*, 661.
(44) Radisic, D.; Bowen, K. H., Jr.; Dabkowska, I.; Storoniak, P.; Zak, J.; Gutowski, M. *J. Am. Chem. Soc.* **2005**, *127*, 6443.
(45) Cubero, M.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 8653.
(46) Kennard, O.; Salisbury, S. A. *J. Biol. Chem.* **1993**, *268*, 10701−10704.
(47) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804.
(48) Brown, T.; Kennard, O. *Curr. Opin. Struct. Biol.* **1992**, *2*, 354.
(49) Sun, L. X.; Bu, Y. X. *J. Phys. Chem. B* **2005**, *109*, 593.
(50) Kan, L. S.; Chandrasegaran, S.; Pulford, S. M.; Miller, P. S. *Proc. Natl. Acad. Sci U.S.A.* **1983**, *80*, 4263.
(51) Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L. S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498.
(52) Webster, G. D.; Sanderson, M. R.; Skelley, J. V.; Neidle, S.; Swann, P. F.; Li, B. F. L.; Tickle, I. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6693.
(53) Hunter, W. N.; Brown, T.; Kennard, O. *J. Biomol. Struct. Dyn.* **1986**, *4*, 173.
(54) Brown, T.; Hunter, W. N.; Kneale, G.; Kennard, O. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2402.
(55) Skelly, J. V.; Edwards, K. J.; Jenkins, T. C.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 804.
(56) Johnson, S. J.; Beese, L. S. *Cell* **2004**, *116*, 803.
(57) Minnick, D. T.; Bebenek, K.; Osheroff, W. P.; Turner, R. M., Jr.; Astatke, M.; Liu, L.; Kunkel, T. A.; Joyce, C. M. *J. Biol. Chem.* **1999**, *274*, 3067.