Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1559–1562, March 1986
Biochemistry

Comparison of nucleotide interactions in water, proteins, and vacuum: Model for DNA polymerase fidelity

(replication fidelity/amplified base-pair discrimination/water exclusion)

JOHN PETRUSKA, LAWRENCE C. SOWERS, AND MYRON F. GOODMAN

Molecular Biology Section, ACB 406, Department of Biological Sciences, University of Southern California, University Park, Los Angeles, CA 90089-1481

Communicated by Sidney W. Benson, October 18, 1985

ABSTRACT **We propose a model for DNA polymerase fidelity in which free energy differences, $\Delta\Delta G$, between matched and mismatched nucleotides are magnified at the enzyme's active site. Both hydrogen bonding and stacking components of the interaction energy are amplified, with the most profound effect being on the magnitude of hydrogen-bonding interactions. Magnification in $\Delta\Delta G$ values follows from the exclusion of water around base pairs in the active site cleft of the enzyme. After showing that base-pair dissociation energies calculated from hydrogen-bonding and base-stacking interactions *in vacuo* are greatly reduced by water, it is proposed that water removal results in a proportional restoration of these contributions to base pairing. Assuming ≈40% exclusion of surrounding water, one predicts magnified values of $\Delta\Delta G$ sufficient to account for polymerase insertion and proofreading fidelity, thereby avoiding the need to postulate additional active site constraints in order to select or reject nucleotides.**

The high degree of replication fidelity exhibited by DNA polymerase has been interpreted to mean that these enzymes must be involved in selecting correct nucleotides for insertion into DNA (for a review, see ref. 1). Nucleotide selection could be imposed by change in enzyme conformation or active-site architecture, which greatly inhibits formation of phosphodiester bonds whenever a mismatched base is resident on the polymerase·DNA complex. The rationale for requiring polymerase base selection is that free energy differences between base mispairs observed in aqueous solution are between 1 and 3 kcal·mol^{-1}, a range that can only provide 5- to 150-fold discrimination against insertion of a wrong nucleotide (1). However, DNA polymerase insertion accuracies have been measured to be in the range of 10^{-3} to 10^{-5} *in vitro* (1), requiring free energy differences over twice as large as those measured in water.

We offer a physical explanation for the fidelity of DNA polymerases during the nucleotide insertion step, in which base-pairing free-energy differences are magnified several-fold in the environment of the active site as a result of water exclusion. There is no requirement to postulate that polymerases must undergo conformational changes to inhibit insertion of mismatched nucleotides. In this model, control of fidelity would then depend primarily on an enhanced magnitude of intrinsic hydrogen-bonding and base-stacking interactions between dNTP substrate and DNA primer-template molecules as water is excluded at the enzyme's active site.

MODEL

Consider the idea that the relative insertion (I) of two nucleotides, W (wrong) and R (right), competing at a template site bound to DNA polymerase, depends primarily on the difference between the corresponding free energies (ΔG_W and ΔG_R) evaluated at the polymerase surface, p.

$$\Delta\Delta G_p = (\Delta G_W - \Delta G_R)_p = RT \ln(I_R/I_W)_p \quad [1]$$

We intend to show that $\Delta\Delta G_p$ is significantly greater in magnitude than the corresponding free-energy difference in water, $\Delta\Delta G_{H_2O} = (\Delta G_W - \Delta G_R)_{H_2O}$, when the active-site region of the enzyme excludes water around base-paired nucleotides. Although we are focusing on polymerase insertion fidelity, the same type of base-pairing free-energy magnification applies to proofreading by 3′ exonuclease activities (2, 3). Proofreading exonucleases may excise previously inserted nucleotides that are melted out at primer termini. The relative stabilities of 3′ termini, as defined by the ratio of melted-to-annealed states, are proportional to $\Delta\Delta G_p$ values at the nuclease active site.

We begin by noting that, for stacked nearest-neighbor doublets of base pairs in DNA, the following data are available: (*i*) empirical values of melting temperature (T_m) obtained from melting profiles of native DNA polymers in aqueous solution (4, 5); (*ii*) theoretical values of dissociation energy calculated from base-stacking and hydrogen-bonding interactions in vacuum (6, 7). It has already been pointed out that T_m correlates with the stacking component of dissociation energy (4). The strong correlation suggests that T_m values measured in water and base stacking enthalpies calculated in vacuum may be related as follows:

$$T_m = T^\circ_m + \alpha^\circ_s \Delta H^v_s, \quad [2]$$

where T_m is the melting temperature of each dinucleotide pair (within a polymer), ΔH^v_s is the corresponding base-stacking component of dissociation energy (in vacuum), α°_s is a factor representing the influence of the medium on intrinsic base-stacking interactions, and T°_m is a constant term (independent of base composition and sequence), which includes contributions of the medium and sugar–phosphate backbone interactions.

To assess the contribution to T_m from hydrogen bonding, we add another term to Eq. **2** as follows:

$$T_m = T^\circ_m + \alpha^\circ_s \Delta H^v_s + \alpha^\circ_h \Delta H^v_h, \quad [3]$$

where ΔH^v_h is the hydrogen-bonding component of dissociation energy (in vacuum) and α°_h represents the influence of the medium on the strength of this component. We then rearrange Eq. **3** into the form of a straight line:

$$(T_m - T^\circ_m)/\Delta H^v_h = \alpha^\circ_s(\Delta H^v_s/\Delta H^v_h) + \alpha^\circ_h, \quad [4]$$

and use literature values of T_m (4), ΔH^v_s (5, 6), and ΔH^v_h (7) to determine the value of T°_m that leads to the best correlation between $(T_m - T^\circ_m)/\Delta H^v_h$ and $\Delta H^v_s/\Delta H^v_h$. Trying various values of T°_m, we find that the correlation coefficient is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

maximum (0.982) near T°_m = 273 K (0°C) and falls off monotonically on either side. For T°_m = 273 K, the line obtained by a least-squares fit (Fig. 1) yields α°_s = 9.0 K kcal^{-1}·mol and α°_h = 0.0. Accordingly, for DNA in water, T°_m is very close to the melting temperature of ice and there is no apparent contribution to T_m from hydrogen bonding between bases.

Having evaluated the parameters in Eq. 3, we can now relate the dissociation energy in solution to the dissociation energy in vacuum. The dissociation energy in solution is

$$\Delta H = T_m \Delta S, \quad [5]$$

where ΔS is the average entropy increase per mol of base pair upon melting DNA. By combining Eqs. 3 and 5, we see that the energy required to dissociate DNA base pairs in aqueous medium is

$$\Delta H = T^\circ_m \Delta S + \beta^\circ_s \Delta H^v_s + \beta^\circ_h \Delta H^v_h, \quad [6]$$

where $\beta^\circ_s = \alpha^\circ_s \Delta S$ and $\beta^\circ_h = \alpha^\circ_h \Delta S$ are dimensionless constants for the medium. Taking ΔS to be 24 ± 2 cal·mol^{-1}·K^{-1} (4) and using the previously evaluated parameters (T°_m = 273 K, α°_s = 9.0 K kcal^{-1}·mol, α°_h = 0), we find β°_s = 0.22 and β°_h = 0, while $T^\circ_m \Delta S$ = 6.55 kcal·mol^{-1}. Thus, as shown in Fig. 2, the relationship between enthalpy changes in water and vacuum is

$$\Delta H_{H_2O} = 6.55 + 0.22\ \Delta H^v_s \text{ kcal·mol}^{-1}. \quad [7]$$

These results indicate that water reduces the intrinsic base-stacking component to 0.22 of its vacuum value and the hydrogen-bonding component to 0 (within experimental error). Accordingly, hydrogen bonding makes no net enthalpy contribution to helix-coil transitions in water, confirming the earlier predictions of DeVoe and Tinoco (8). The quantity 6.55 kcal·mol^{-1} is a contribution to the dissociation energy not found in vacuum. Being $T^\circ_m \Delta S$, it represents the energy required to melt the ice-like structure of surrounding water to allow dissociated base pairs their required number of degrees of freedom.* This is commonly referred to as the "hydrophobic" contribution to dissociation energy.

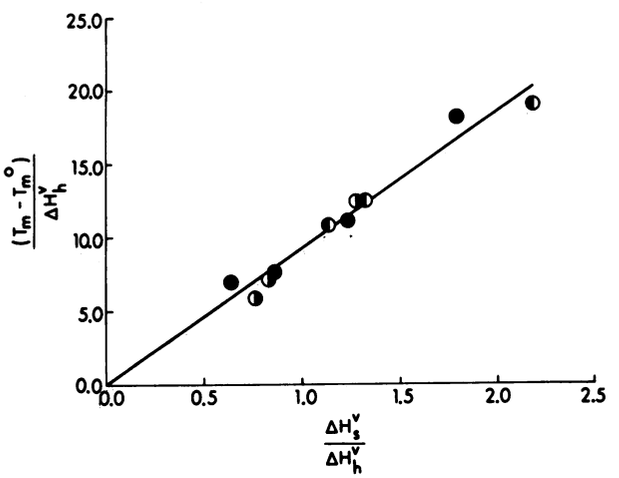

FIG. 1. Relationship between empirical melting temperatures for stacked DNA base pairs in water and theoretical dissociation energies in vacuum, plotted according to Eq. 4 with T°_m set equal to the melting temperature of water. T_m is the melting temperature for each doublet as determined from melting profiles of known polymers at 0.02 M ionic strength (4). ΔH^v_h is the theoretically computed hydrogen-bonding contribution to the dissociation energy per mol of base pair (7) and ΔH^v_s is the stacking component for the dinucleotide pair (6). Values are shown for each nearest-neighbor doublet of base pairs of type $\overrightarrow{RR}/\overleftarrow{YY}$ (●), $\overrightarrow{RY}/\overleftarrow{YR}$ (◐), and $\overrightarrow{YR}/\overleftarrow{RY}$ (◑); where R = purine, Y = pyrimidine, and arrows indicate the 5′ → 3′ direction. The line corresponds to Eq. 4 with intercept α°_h = 0 and slope α°_s = 9.0 K kcal·mol^{-1}.

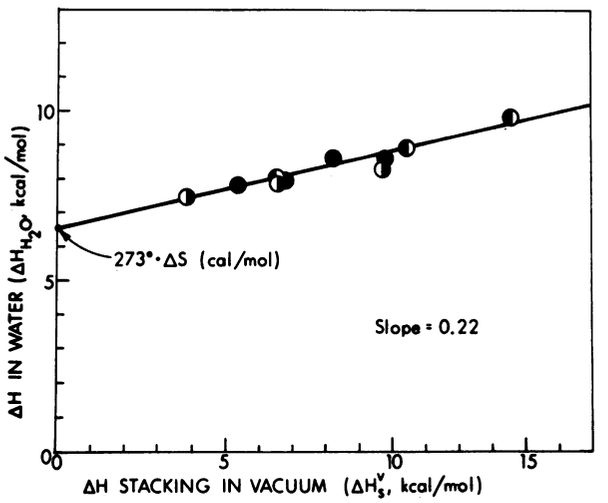

FIG. 2. Relationship between heats of dissociation (ΔH) for stacked DNA base pairs in water and in vacuum. Values of ΔH per mol of stacked base pair are shown for each nearest-neighbor doublet of base pairs of type $\overrightarrow{RR}/\overleftarrow{YY}$ (●), $\overrightarrow{RY}/\overleftarrow{YR}$ (◐), and $\overrightarrow{YR}/\overleftarrow{RY}$ (◑); where R = purine, Y = pyrimidine, and arrows indicate the 5′ → 3′ direction. The ΔH values in water are determined as equal to $T_m \Delta S$, where T_m is the melting temperature (in K) for each doublet, evaluated from melting profiles of known DNA polymers (4), and ΔS = 24 ± 2 cal·mol^{-1}·K^{-1}, the average entropy increase upon melting 1 mol of base pair. The ΔH values for stacking in vacuum (ΔH^v_s) are computed theoretically, using an optimized potential function for the calculation of nucleic acid interaction energies (6). The line corresponds to Eq. 7, with the intercept equal to (273 K)·(24 cal·mol^{-1}·K^{-1}) = 6.55 kcal·mol^{-1}, and slope = 0.22.

Central to our model, on going from water to a protein environment, β°_s and β°_h in Eq. 5 are replaced by new values (β_s and β_h) characteristic of the protein. Thus, the dissociation energy for DNA base pairs in a protein cleft is

$$\Delta H_p = 6.55 + \beta_s \Delta H^v_s + \beta_h \Delta H^v_h \text{ kcal·mol}^{-1}, \quad [8]$$

assuming that $T^\circ_m \Delta S$ is the same as in water, since dissociated base pairs are still released into water and require the same number of degrees of freedom. The important issue is to arrive at reasonable estimates of β_s and β_h at enzyme active sites. For conditions intermediate between water and vacuum, β_s lies between 0.22 and 1, while β_h lies between 0 and 1. For simplicity, we suggest that β_s and β_h are linearly related to the degree of exclusion of water around base pairs in the active site. Then, if f is the fraction of water excluded, β_s = 0.22 + 0.78f and $\beta_h = f$, approximately. Accordingly, Eq. 8 becomes

$$\Delta H_p = 6.55 + (0.22 + 0.78f)\Delta H^v_s + f\Delta H^v_h \text{ kcal·mol}^{-1}. \quad [9]$$

*In the backbone of DNA, each nucleotide has six single bonds about which rotation can occur (one C—C, two C—O, two O—P, and one C—N). Upon dissociation of base pairs, each single bond except the C—N glycosidic bond acquires 3 degrees of freedom (equivalent to the three staggered configurations in ethane). The C—N bond acquires 2 degrees of freedom corresponding to *anti* and *syn* conformations in the case of purine nucleotides only. Accordingly, for each dissociated base pair there are $2 \cdot (3^5)^2$ degrees of freedom. Assuming each associated base pair has only 1 degree of freedom for a fixed DNA conformation, the corresponding entropy change upon dissociation is $\Delta S = R \ln (2 \cdot 3^{10})$ = 23.2 cal·mol^{-1}·K^{-1}, in agreement with the assigned value 24 ± 2 cal·mol^{-1}·K^{-1}.

Let us now examine the free-energy difference between right and wrong nucleotides, $\Delta\Delta G_p$ defined by Eq. 1, in terms of ΔH_p given by Eq. 9. Since $\Delta G_p = \Delta H_p - T\Delta S$, then at constant T and ΔS,

$$\Delta\Delta G_p = \Delta\Delta H_p = (0.22 + 0.78f)\Delta\Delta H_s^v + f\Delta\Delta H_h^v. \quad [10]$$

The corresponding free-energy difference in water, equivalent to setting $f = 0$, is

$$\Delta\Delta G_{H_2O} = 0.22\Delta\Delta H_s^v. \quad [11]$$

Thus, on going from water to protein, the free energy difference between right and wrong nucleotides is predicted to be amplified by the factor,

$$\frac{\Delta\Delta G_p}{\Delta\Delta G_{H_2O}} = 1 + 3.5f + 4.5f\left(\frac{\Delta\Delta H_h^v}{\Delta\Delta H_s^v}\right). \quad [12]$$

Consider a protein cleft that excludes 40% of the water in contact with base pairs. In this case ($f = 0.4$), the amplification factor becomes

$$\frac{\Delta\Delta G_p}{\Delta\Delta G_{H_2O}} = 2.4 + 1.8\left(\frac{\Delta\Delta H_h^v}{\Delta\Delta H_s^v}\right). \quad [13]$$

Even if there is no difference in hydrogen bonding between right and wrong nucleotides, this factor will be 2.4. If hydrogen-bonding and stacking differences in vacuum are equal, the amplification factor will increase to 4.2.

DISCUSSION

The model presented here is based on correlations between empirical values of base-pair dissociation energy in water and theoretical values in vacuum (Fig. 2). The correlations indicate that the energies required to dissociate base-stacking and hydrogen-bonding interactions in vacuum are markedly diminished in the presence of water. The base-stacking component is reduced to a fraction (0.22) of its value in vacuum, while the hydrogen-bonding component approaches 0. On the other hand, water introduces a constant hydrophobic contribution of 6.55 kcal·mol^{-1} to the dissociation energy of base pairs.

The apparent suppression of intrinsic base–base interactions in water may be the result of strong water–base interactions. Water may interact with bases through its strong dipole character and associated hydrogen-bonding capability. Since hydrogen bonding is mainly a dipole–dipole interaction (7), while intrinsic base stacking is mainly an induced-dipole–induced-dipole interaction (6), it is reasonable for water to diminish base–base hydrogen bonding more than stacking.

We suggest that in the active site of polymerase, intrinsic interactions between base pairs are partially restored by the exclusion of water. This could happen if the protein does not interact with DNA bases in the same way as water does. For example, the protein might interact only with the sugar–phosphate backbone on the narrow-groove side, so as to pull the DNA into a cleft and partially free the bases from water. If the protein does not interact significantly with the bases themselves, base–base interactions may be restored to values intermediate between vacuum and water. The implication of this model is that stacking and hydrogen-bonding interactions are in essence amplified in the active site of the enzyme, as a consequence of water expulsion, and that the resultant amplification of free-energy differences is sufficient to account for the fidelity of DNA replication.

The model is applicable to studies on the relative incorporation of base analogues by DNA polymerase. Eq. 13 predicts that, even if the difference in hydrogen bonding between nucleotides competing for insertion is 0, free-energy differences observed in water will be amplified by a factor of 2.4. This situation may reasonably apply to the incorporation of 5-substituted pyrimidines. The difference in hydrogen bonding between uracil and thymine with adenine is probably near 0 as is the difference between cytosine and 5-methylcytosine with guanine. The value of $\Delta\Delta G_{H_2O}$ for these homologous pairs can be determined from melting studies of polymers containing the homologs (9–11). Polymers containing the methylated pyrimidine (thymine, 5-methylcytosine) in place of the unmethylated derivative (uracil, cytosine) invariably melt at higher temperatures because of higher intrinsic stacking energy. $\Delta\Delta G_{H_2O}$ values can be determined from the averages of the corresponding homopolymer and alternating copolymer melting temperatures. The resultant $\Delta\Delta G_{H_2O}$ is 0.24 kcal·mol^{-1} between thymine and uracil and 0.186 kcal·mol^{-1} between 5-methylcytosine and cytosine. According to Eq. 13, these values are amplified by 2.4, so that $\Delta\Delta G_p$ = 0.58 kcal·mol^{-1} and 0.45 kcal·mol^{-1}, respectively. Substituting in Eq. 1, we predict the ratio of incorporation of uracil/thymine to be 0.4 and of 5-methylcytosine/cytosine to be 2.1, in fairly good agreement with the values (0.54 and 1.85, respectively) observed by Kornberg (12) and coworkers in a DNA synthesis assay *in vitro*.

The model predicts an amplification factor >2.4 when there is a difference in hydrogen bonding *in vacuo*. Base-pairing properties of 2-aminopurine, a mutagenic base analogue of adenine, have been analyzed extensively from the points of view of (*i*) melting temperature differences for 2-aminopurine·thymine vs. A·T base pairs of DNA polymers (13), and (*ii*) measurements of relative insertion frequencies of 2-aminopurine, in competition with adenine, opposite template thymine sites by DNA polymerases (13, 14). Table 1 shows a comparison of the free energy differences between 2-aminopurine·thymine vs. A·T base pairs in water (column 2) and in the presence of DNA polymerase (column 3) for each of the possible nearest-neighbor base-stacking partners. The ratios (column 4) indicate that free-energy differences have been amplified in the active site by a factor of 3.7 ± 1.1, in agreement with our predicted range of 2.4–4.2.

In summary, we suggest that when nucleotides enter the active site of DNA polymerase, water is displaced. The removal of water around bases restores intrinsic hydrogen-bonding and stacking interactions between bases, thereby amplifying free-energy differences between right and wrong base pairs. In this manner, a high degree of nucleotide insertion fidelity can be achieved without nucleotide selection or rejection either by protein conformational changes or "kinetic proofreading" (15, 16). This concept generally

Table 1. Comparison of free-energy differences for 2-aminopurine·thymine vs. A·T base pairs in water ($\Delta\Delta G_{H_2O}$) and in the active site of DNA polymerase ($\Delta\Delta G_p$) for various 5′ nearest neighbors

5′ nearest neighbor*	Free-energy difference, kcal·mol^{-1}		$\Delta\Delta G_p/\Delta\Delta G_{H_2O}$
	$\Delta\Delta G_{H_2O}$†	$\Delta\Delta G_p$‡	
T	0.20 ± 0.05	0.37 ± 0.04	1.9 ± 0.7
C	0.20 ± 0.05	0.98 ± 0.10	4.9 ± 1.9
G	0.40 ± 0.05	1.44 ± 0.15	3.6 ± 0.8
A	0.35 ± 0.05	1.55 ± 0.15	4.4 ± 1.1
Average	0.29 ± 0.05	1.08 ± 0.11	3.7 ± 1.1

*Base preceding 2-aminopurine or adenine on primer strand and hydrogen-bonded to complementary base on template strand.
†$\Delta\Delta G_{H_2O} = (\Delta G_{AP\cdot T} - \Delta G_{A\cdot T})_{H_2O}$ as computed from DNA polymer melting data (T_m values are in ref. 13). AP, 2-aminopurine.
‡$\Delta\Delta G_p = (\Delta G_{AP\cdot T} - \Delta G_{A\cdot T})_p$ as computed from DNA polymerase catalyzed insertion frequencies (I_{max} values in ref. 13).

agrees with the recent proposal of Dewar and Storch (17) in which removal of water from the active site is responsible for both high substrate selectivity and turnover number in enzymes. However, it should be pointed out that our approach is basically thermodynamic rather than kinetic. While defining the free-energy differences between states relevant to fidelity, we have not defined the pathway between the states. To describe the pathway, one needs to know the structure of the enzyme cleft and the manner in which nucleotides are bound so as to exclude water around bases.

We thank Sarah Wright, Maria McElwee, and Sheau-Fang Oblonsky for preparation of this manuscript. Work reported here was supported by National Institutes of Health Grants GM21422 and GM33863 and by National Institutes of Health Post-Doctoral Training Grant 2T32CA09320 to L.S.

1. Loeb, L. A. & Kunkel, T. A. (1982) *Annu. Rev. Biochem.* **52,** 429–457.
2. Brutlag, D. & Kornberg, A. (1972) *J. Biol. Chem.* **247,** 241–248.
3. Muzyczka, N., Poland, R. L. & Bessman, M. J. (1972) *J. Biol. Chem.* **247,** 7122–7166.
4. Gotoh, O. & Takashira, Y. (1981) *Biopolymers* **20,** 1033–1042.
5. Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, ed. Cantor, C. R. (Springer, New York), pp. 139–149.
6. Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. (1978) *Biopolymers* **17,** 2341–2360.
7. Jordan, F. & Sostman, H. D. (1973) *J. Am. Chem. Soc.* **95,** 6544–6554.
8. DeVoe, H. & Tinoco, I. (1962) *J. Mol. Biol.* **4,** 500–517.
9. Gill, J. E., Mazrimas, J. A. & Bishop, C. C. (1974) *Biochem. Biophys. Acta* **335,** 330–348.
10. Sober, E. K. (1970) in *Handbook of Biochemistry*, ed. Sober, H. A. (Chemical Rubber, Cleveland, OH), 2nd Ed., pp. 21–22.
11. Vilpo, J. A. & Ridell, J. (1983) *Nucleic Acids Res.* **11,** 3753–3765.
12. Kornberg, A. (1980) *DNA Replication* (Freeman, San Francisco), p. 119.
13. Petruska, J. & Goodman, M. F. (1985) *J. Biol. Chem.* **260,** 7533–7539.
14. Pless, R. C. & Bessman, M. J. (1983) *Biochemistry* **22,** 4905–4915.
15. Hopfield, J. J. (1974) *Proc. Natl. Acad. Sci. USA* **72,** 4135–4139.
16. Ninio, J. (1975) *Biochimie* **57,** 587–595.
17. Dewar, M. J. S. & Storch, D. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2225–2229.