Published in final edited form as:

Nat Rev Chem. 2017 September ; 1(9): . doi:10.1038/s41570-017-0074.

Relating DNA base-pairing in aqueous media to DNA polymerase fidelity

John Petruska¹ and Myron F. Goodman^{1,2}

¹Department of Biological Sciences, University of Southern California

²Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA

Abstract

Controversy surrounds the perceived absence of a relationship between DNA polymerase fidelity (kinetic discrimination) and free energy changes determined from DNA melting studies (thermodynamic discrimination). Thermodynamic discrimination together with aqueous solvent effects can account for kinetic fidelities on the order of those observed experimentally.

DNA polymerases are enzymes that are responsible for rapidly and accurately copying the genome. When immersed in an aqueous sea of A, T, G and C deoxyribonucleoside triphosphates (dNTPs), polymerases catalyse template-directed synthesis of Watson–Crick (W–C) base pairs in preference to mispairs. For many polymerases, this proceeds with very high selectivity, which is remarkable given that the choice to accept or reject a dNTP substrate must be made in about 10^{-3} s. Identifying the source of free energy that enables high-fidelity DNA replication would further our fundamental understanding of biology.

The thermal stabilities of AT and GC base-pair matches are greater than those of mismatches, such as AC and GT, on account of stronger hydrogen-bonding and base-stacking interactions in double-stranded DNA (dsDNA) comprising matched pairs¹. How can DNA polymerases take advantage of the differences in standard free energy change ($\Delta G^\circ = \Delta G^\circ(\text{mismatch}) - \Delta G^\circ(\text{match})$) between matched and mismatched base pairs to copy DNA accurately? This is a fundamental unresolved question. Replicative polymerases incorporate mismatches very rarely ($f = e^{-\Delta G^\circ/RT}$; where f is the error frequency, T is temperature and R is the universal gas constant), with errors occurring only once every 10^3 – 10^6 base pairs. Under physiological conditions, this low error rate of high-fidelity polymerases necessitates ΔG° to be 4–8 kcal mol⁻¹ (REF. 1). Determining the melting temperatures of DNA enables standard ΔG° values for individual stacked base pairs to be calculated; results are obtained by comparing matched versus mismatched pairs that are located internally in dsDNA² or at primer–template DNA termini³. However, these data yield ΔG° values of about 2 kcal mol⁻¹ or less, which alone would afford polymerase error frequencies of greater than 10^{-2} . We have suggested that polymerase active sites may need additional sources of ΔG° — a supplement to the apparently small differences in ΔG°

Correspondence to M.F.G. mgoodman@usc.edu.

Competing interests statement

The authors declare no competing interests.

indicated by thermodynamic studies of DNA strand association² or dissociation⁴ — to favour incorporation of matched pairs over mismatched ones. This Comment explores how and to what extent DNA thermodynamics in aqueous media can serve as a source of the G° required for the high fidelities exhibited by certain DNA polymerases.

Stalling with pyrophosphate

The impetus for us to investigate anew the relationship between the thermodynamic stabilities of base pairs and polymerase kinetic base-pair selection came from an insightful recent study on DNA polymerase fidelity⁵. In that paper, Olson *et al.* questioned whether variations in the fidelity of DNA polymerases could arise from differences between the thermodynamic stabilities of base pairs and mispairs alone, or instead whether amplification of these differences in G° are required in the form of specific discriminatory interactions in the active site. It was proposed that by examining match and mismatch incorporation rates in the presence of high concentrations of pyrophosphate (PP_i) — used to equalize the rates of dNTP insertion and release — it becomes possible to calculate the equilibrium constant (K_{eq}) for both match and mismatch incorporation; thus, the corresponding standard free energy changes, using $G^\circ = -RT \ln K_{eq}$, can be obtained.

The Vent and Pfu polymerases — thermally stable enzymes from *Thermococcus litoralis* and *Pyrococcus furiosus*, respectively — have proven particularly useful in studies of DNA_n elongation to DNA_{n+1}. For example, Vent polymerase has been applied to the incorporation of dNTP (N = A or G), present in varying amounts, opposite to a T template. At high $[PP_i]$, the net conversion rate of DNA_n to DNA_{n+1} approaches zero over time, such that a constant concentration quotient $[DNA_{n+1}]/[DNA_n]$, interpreted as equilibrium, was attained after about 1 h in the case of the match reaction (A opposite to T) and about 20 h for the mismatch (G opposite to T)⁵. The equilibrium constant $K_{eq} = [DNA_{n+1}][PP_i]/[DNA_n][dNTP]$ was used to compute G° in each case, and the resulting G° values of 3.5–7 kcal mol⁻¹ suggest that the imposition of additional kinetic constraints on the active site would be unnecessary. This conclusion was derived solely from match and mismatch incorporation data; pyrophosphorolysis — necessary for dNTP release from DNA_{n+1} — was inferred but not measured.

The observation of constant $[DNA_{n+1}]/[DNA_n]$ values in the case of both match and mismatch incorporation reactions is a necessary but insufficient equilibrium condition. To be sure that a system has reached equilibrium, one must obtain the same steady-state ratios for $[DNA_{n+1}]/[DNA_n]$ when studying the reverse of both incorporation reactions by initiating them from the products. Therefore, we measured the rates of both the forward and reverse reactions³. The key result was that the reverse reaction was undetectable for mismatches; thus, pyrophosphorolysis of G at the primer-3' end mispaired opposite to T did not occur³. Mismatch incorporation at high $[PP_i]$ is far from equilibrium despite showing an approximate levelling off³. Thus, the data of Olson *et al.*⁵ do not support the proposal that large G° values obtained from polymerase-catalysed steady-state incorporation of matches and mismatches provide a direct measure of polymerase fidelity without having to consider active site constraints.

Although pyrophosphorolysis was not detected for the mismatch (G opposite to T), it was readily observed for the match (A opposite to T). The removal of A opposite to T reached steady state after ~30–60 min — the same timescale observed for incorporation of A opposite to T (REF. 3). The ΔG° values observed for the incorporation reaction ($\Delta G^\circ_{\text{inc}} = -5.22 \pm 0.05 \text{ kcal mol}^{-1}$) and reverse ($\Delta G^\circ_{\text{pyro}} = 5.51 \pm 0.09 \text{ kcal mol}^{-1}$) were opposite in sign and of almost identical magnitude, which suggests that match incorporation at high $[\text{PP}_i]$ was indeed at or near equilibrium. Overall, it is noted that polymerase-catalysed match incorporation reaches steady state close to equilibrium when subjected to reversal with PP_i , whereas Vent polymerase-catalysed mismatch incorporation seems to be irreversible even upon treatment with excess PP_i . These results prompted us to re-examine here what DNA melting studies for matched and mismatched base pairs in aqueous solution can reveal about DNA polymerase fidelity when the environmental influence of H_2O hydrogen bonding is considered.

The influence of solvent on ΔG°

The influence of the solvent environment on DNA stability can be determined by plotting ΔS° versus ΔH° for melting of each kind of DNA doublet (d) (FIG. 1), with values obtained by measuring the melting temperature (T_m) of dsDNA in 1 M aqueous NaCl (REF. 2). The data in the upper range of ΔS°_d and ΔH°_d correspond to the ten W–C nearest-neighbour doublets, whereas those in the lower range are for a subset of doublets that contain a single mispair². The GA–CT doublet is of the type used in our polymerase fidelity experiment, in which correct incorporation of A opposite to T takes place next to 5′-nearest-neighbour G (on primer) bound to C (on template)³. Match incorporation reaches equilibrium with respect to pyrophosphorolysis at high $[\text{PP}_i]$, yielding $\Delta G^\circ_{\text{inc}}(\text{AT}) = -5.2 \text{ kcal mol}^{-1}$ (REF. 3), a value much larger in magnitude than $\Delta G^\circ_{\text{GA-CT}}(1.3 \text{ kcal mol}^{-1})$ calculated using ΔH°_d and ΔS°_d for melting of doublet GA–CT². The value of $\Delta G^\circ_{\text{GA-CT}}$ is small because the large and positive value of $\Delta H^\circ_{\text{GA-CT}}(8.2 \text{ kcal mol}^{-1})$ is almost cancelled by the entropic contribution $-T\Delta S^\circ_{\text{GA-CT}}(-310 \text{ K} \times 22.2 \text{ cal mol}^{-1}\text{K}^{-1} = -6.9 \text{ kcal mol}^{-1})$. This approximate cancellation has been referred to as “enthalpy–entropy compensation” (REF. 4). For W–C doublets at 37 °C, ΔG°_d is largest in the case of CG–GC ($2.24 \text{ kcal mol}^{-1}$) and smallest for TA–AT ($0.56 \text{ kcal mol}^{-1}$).

The above method for determining the ΔG°_d of melting for an individual doublet does not take into account the environmental (e) influence of H_2O , with doublet melting being accompanied by significant rearrangement of the hydrogen bonding in the solvation shell. The contribution of the solvent to ΔG°_d can be obtained by deriving an analytical curve fit for the observed relationship between ΔS°_d and ΔH°_d (FIG. 1). Clearly, ΔS°_d is not constant, and $T_m = \Delta H^\circ_d / \Delta S^\circ_d$ is not directly proportional to ΔH°_d as has previously been suggested⁶. Instead, we have proposed that T_m is linearly related to ΔH°_d according to the relationship $T_m = T_0 + \Delta H^\circ_d / a$, where T_0 is a constant corresponding to the value of T_m when $\Delta H^\circ_d = 0$,

and a is an entropy constant with a value greater than any ΔS_d° (REF. 4). Upon substituting $\Delta H_d^\circ/\Delta S_d^\circ$ in place of T_m , we obtain the hyperbolic function $\Delta S_d^\circ = a\Delta H_d^\circ/(\Delta H_d^\circ + aT_0)$ (REF. 4). This hyperbolic equation for melting of nearest-neighbour doublets in dsDNA has been confirmed by a statistical thermodynamic treatment of dsDNA melting⁷. The constants a and T_0 define the shape of the curve and were evaluated by a least-squares fit to the data for the ten W–C doublets (FIG. 1), which gives $a = 80 \pm 2 \text{ cal mol}^{-1} \text{ K}^{-1}$. When $\Delta H_d^\circ = 0$, the slope of the curve is $1/T_0$, and T_m for the doublet becomes $T_0 = 273 \pm 5 \text{ K}$. This value of T_0 is close to the melting temperature of H₂O (REF. 4), which implies that hydrogen bonding between H₂O molecules within the solvation shell contributes to the stability of each doublet. The aqueous solvent effects on doublet melting can now be described as $\Delta G_{d+e}^\circ = \Delta H_{d+e}^\circ - T\Delta S_{d+e}^\circ$ without changing the T_m value for each doublet, such that $T_m = \Delta H_d^\circ/\Delta S_d^\circ = \Delta H_{d+e}^\circ/\Delta S_{d+e}^\circ = (aT_0 + \Delta H_d^\circ)/a$. Accordingly, the environmentally adjusted ΔH_{d+e}° is the sum of ΔH_d° and the constant aT_0 (21.8 kcal mol^{−1}), whereas ΔS_{d+e}° remains constant at the value of $a = 80 \text{ cal mol}^{-1} \text{ K}^{-1}$.

When the contribution from an aqueous medium is described in this manner for melting of doublet GA–CT, then $\Delta G_{d+e}^\circ = (8.2 + 21.8) - 24.8 \text{ kcal mol}^{-1} = 5.2 \text{ kcal mol}^{-1}$, which agrees in magnitude with the polymerase result, $\Delta G_{\text{inc}}^\circ(\text{AT}) = -5.22 \pm 0.05 \text{ kcal mol}^{-1}$, determined from K_{eq} between polymerization and phosphorolysis³. Although the entropic parameter a is surprisingly large, the ‘proof of the pudding is in the eating’, as shown by the quantitative accord between $\Delta G_{\text{inc}}^\circ$ determined for the polymerase and the thermodynamic ΔG_{d+e}° value for the same nearest-neighbour stacking partners. We suggest that the thermodynamic stability of matched dsDNA, reinforced by hydrogen bonding in the immediate aqueous environment, provides a sufficiently negative G° value for DNA polymerases to preferentially mediate formation of matched W–C base pairs.

“the thermodynamic stability of matched dsDNA, reinforced by hydrogen bonding in the immediate aqueous environment, provides a sufficiently negative G° value”

Relating $\Delta\Delta G_{d+e}^\circ$ to DNA polymerase fidelity

When cognate and non-cognate dNTPs are present in equimolar concentrations, these substrates compete for incorporation by polymerases. As noted earlier, under conditions that promote optimal forward rates for DNA synthesis, G° is in the range 4–8 kcal mol^{−1} (REF. 1). With respect to the influence of solvent on the G° between doublet formation, ΔS_{d+e}° is predicted to be constant for all doublets — that is, matches and mismatches — such that $\Delta\Delta S_{d+e}^\circ = 0$. In addition, $\Delta H_{d+e}^\circ = \Delta H_d^\circ + aT_0$, and as aT_0 is the same for all doublets, $\Delta\Delta H_{d+e}^\circ = \Delta\Delta H_d^\circ$. Consequently, $\Delta\Delta G_{d+e}^\circ = \Delta\Delta H_d^\circ$.

For the ten W–C doublets in 1 M aqueous NaCl (FIG. 1), ΔH_d° ranges from 11.0 (for CG–GC) to 6.6 kcal mol^{−1} (for TA–AT). For doublets with single mismatches, ΔH_d° is between

4.1 and $-3.0 \text{ kcal mol}^{-1}$. After the inclusion of solvent effects, $\Delta\Delta G_{d+e}^{\circ} = \Delta\Delta H_d^{\circ}$ ranges between about 7 and 13 kcal mol^{-1} , so the magnitudes of $\Delta\Delta G_{d+e}^{\circ}$ are either roughly coincident with or far exceed the G° values needed to account for DNA polymerase fidelity.

Linking solvent, geometry and fidelity

The reversibility and irreversibility with which Vent polymerase incorporates matches and mismatches, respectively, offers fresh insight into the geometric constraints of DNA polymerase active sites in relation to fidelity³. Several kinetic studies have delineated individual steps in the polymerase catalytic cycle: from the binding of cognate and non-cognate substrates, to polymerase conformational changes and phosphodiester bond chemistry⁸. Each step may serve as a kinetic fidelity ‘checkpoint’ that discriminates between matches and mismatches in a polymerase-specific manner⁹.

Each step in the match and mismatch incorporation pathways must adhere to microscopic reversibility. Given the many steps involved, it might seem surprising that match incorporation is overall reversible. Yet, when Vent and Pfu polymerases process AT pairs, the similar magnitudes of $-\Delta G_{inc}^{\circ}$ and $+\Delta G_{pyro}^{\circ}$ imply that the substrates and products remain properly aligned throughout the forward and reverse pathways³. The overriding importance of maintaining optimal geometric alignment for catalysis in the active site of DNA polymerases is discussed in a concurrent Review article in *Nature Reviews Chemistry*⁹. Although binding of Vent or Pfu to a terminal G (primer)···T (template) mismatch is strong enough to allow 100% extension by incorporation of a next correct (TA) base pair, misalignments of the GT mismatch in subsequent reverse checkpoint steps are apparently too severe to allow nucleophilic attack by PP_i to yield and release detectable dGTP after a two-day incubation with Vent polymerase³. In the case of Pfu polymerase, some dGTP release is observed (<2%), although the system nevertheless remains far from equilibrium³.

We have shown here that DNA hydrogen-bonding and base-stacking interactions can provide an ample source of G° to account for high-fidelity DNA synthesis as long as the stabilizing hydrogen-bonding effects of the solvent are also accounted for. The relevance of the $\Delta\Delta G_{d+e}^{\circ}$ values (obtained from doublet melting thermodynamics) to the $\Delta\Delta G_{inc}^{\circ}$ values (derived from polymerase kinetics) originates from the retention of hydrogen-bonding and stacking primer–template interactions within the DNA polymerase active site. Geometric selection¹ can be imposed on the active site through induced fit mechanisms that properly align cognate dNTP, which suppress entropic losses associated with the necessary conformational changes and chemical step. In this way, the enzymes take full advantage of the different H° values associated with installing matches and mismatches.

From a ‘DNA-centric’ perspective, G° values for match and mismatch pairings are insufficient to account for high polymerase fidelity. Our Comment shows that a ‘solvent + DNA’ perspective affords much larger differences in G° , the magnitudes of which are easily sufficient to account for high polymerase fidelities. Different DNA polymerases impose active site constraints that can strongly modify incorporation specificities. One

striking example is DNA polymerase η , the X-ray structure of which reveals an active site that favours a wobble GT base mispair at 3'-AT template motifs — a hallmark of somatic hypermutation of immunoglobulin variable DNA regions that can enable an immune system to adapt to new threats¹⁰. Although the difference in G° from solvent reinforced hydrogen bonding and base stacking is sufficient for discrimination between substrates, nuances in the polymerase active site structures may regulate selection. As such, the linkage between solvent, geometry and fidelity remains in the hands of the polymerase.

Acknowledgments

The authors thank C. H. Mak and K. Oertell for insightful discussions, and J. G. Bertram for collating the data. This work was supported by the US National Institutes of Health (Grant Nos ES012259, U19CA177547 and GM214422 to M.F.G.).

References

1. Echols H, Goodman MF. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* 1991; 60:477–511. [PubMed: 1883202]
2. SantaLucia J, Hicks D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 2004; 33:415–440. [PubMed: 15139820]
3. Oertell K, et al. Kinetic selection versus free energy of DNA base pairing in control of polymerase fidelity. *Proc. Natl Acad. Sci. USA.* 2016; 113:E2277–E2285. [PubMed: 27044101]
4. Petruska J, Goodman MF. Enthalpy–entropy compensation in DNA melting thermodynamics. *J. Biol. Chem.* 1995; 270:746–750. [PubMed: 7822305]
5. Olson AC, Patro JN, Urban M, Kuchta RD. The energetic difference between synthesis of correct and incorrect base pairs accounts for highly accurate DNA replication. *J. Am. Chem. Soc.* 2013; 135:1205–1208. [PubMed: 23316816]
6. Gotoh O, Tagashira Y. Locations of frequently opening regions on natural DNAs and their relation to functional loci. *Biopolymers.* 1981; 20:1043–1058. [PubMed: 7225530]
7. Starikov EB, Nordén B. Physical rationale behind the nonlinear enthalpy–entropy compensation in DNA duplex stability. *J. Phys. Chem. B.* 2009; 113:4698–4707. [PubMed: 19278234]
8. Joyce CM, Benkovic SJ. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry.* 2004; 43:14317–14324. [PubMed: 15533035]
9. Wu W-J, Yang W, Tsai M-D. How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nat. Rev. Chem.* 2017; 1:0068.
10. Zhao Y, et al. Mechanism of somatic hypermutation at the WA motif by human DNA polymerase η . *Proc. Natl Acad. Sci. USA.* 2013; 110:8146–8151. [PubMed: 23630267]

Figure 1. Correlation between ΔS° and ΔH° associated with melting nearest-neighbour doublets in DNA

For the ten Watson–Crick (W–C) doublets, the values of the standard enthalpy change (ΔH°_d) and the standard entropy change (ΔS°_d) represent the average of measurements reported in the literature². The ΔH°_d values (± 0.3 kcal mol⁻¹) range from 11.0 (for CG–GC) to 6.6 (for TA–AT). ΔS°_d versus ΔH°_d is fit to a rectangular hyperbola. Also shown are ΔH°_d values (± 0.6 kcal mol⁻¹) for doublets with single mismatches, with the values for CG–GT, CC–GT, CA–GC and TG–AA, ranging from 4.1 to -3.0 kcal mol⁻¹ (REF. 2). All data were acquired in aqueous NaCl (1 M) at pH 7.