Commentary

Hydrogen bonding revisited: Geometric selection as a principal determinant of DNA replication fidelity

Myron F. Goodman

Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089-1340

That hydrogen bonds play a central role in forming Watson–Crick (W–C) A·T and G·C base pairs is a fundamental paradigm dating from the discovery of the structure of DNA in 1953. In addition to interstrand H bonding, intrastrand base-stacking and interstrand cross-stacking interactions are important in maintaining the bases in a stacked structure along the length of the DNA backbone. In general, H bonds between W–C base pairs are viewed as "informational," whereas the base-stacking interactions are regarded as "noninformational," merely stabilizing the double helix. Consequently, it is a common perception that the H bonds pairing A with T and G with C are primarily responsible for the ability of DNA polymerases to synthesize DNA with high fidelity.

Difluorotoluene, a nonpolar isosteric analog of thymine (T), contains fluorine atoms in place of oxygens on the pyrimidine ring and thus cannot form H bonds with A (1). Nevertheless, A·F base pairs are formed almost as well as A·T pairs by *Escherichia coli* proofreading-defective DNA polymerase I (KF exo⁻), as Moran *et al.* (2) report in this issue of the *Proceedings* [the chemical structures of F and T and space-filling models of each are shown in Moran *et al.* (2), figure 1]. The observation that KF exo⁻ fails to discriminate strongly against A·F pairs applies when F is present either as a template base on DNA (3) or as a dFTP substrate (2). The apparently inescapable conclusion is that H bonds are not absolutely required for polymerase to form W–C base pairs selectively.

These results provide an impetus to reconsider what role H bonds actually play in stabilizing DNA and enhancing DNA polymerase fidelity. Mismatched base pairs in a duplex DNA oligomer do cause marked reductions in DNA melting temperatures (4). The loss of H bonds upon replacement of T with F has this type of destabilizing effect (2). However, the notion that H bonds alone keep the two strands of a DNA double helix together, which is found in many textbooks, seems inadequate. When one considers that duplex alternating copolymers poly d(A,T) or poly d(G,C) have melting temperatures in aqueous solution that differ substantially from their respective homopolymer counterparts poly dA·poly dT or poly dG·poly dC, it becomes clear that base-stacking interactions have an important, perhaps dominant, sequence-dependent effect on duplex stability.

Furthermore, the free-energy differences ($\Delta\Delta G^0$) between matched and mismatched base pairs deduced from melting data are in a range of about 0.2–4.0 kcal/mol (4–6), depending on the identity of the mispair, the surrounding sequence context, and its location near the center or at the DNA terminus. These $\Delta\Delta G^0$ values, as measured in solution, are insufficient to account for the high nucleotide insertion fidelities of virtually all polymerases, including those that seem to be especially "error prone" such as eukaryotic Pol β (7) or HIV-1 reverse transcriptase (8–10). For example, $\Delta\Delta G^0 \approx 3.7$ kcal/mol measured for the natural base pairs A·T versus A·C (6) should result in an A·C misinsertion frequency of about $2 \times$ 10⁻³. However, dAMP·C and dCMP·A misinsertion frequencies are typically about one to two orders of magnitude lower than that (11).

The recognition that base-pairing free-energy differences are too small to completely account for polymerase insertion selectivities prompted H. Echols and me to propose a "geometric selection" mechanism as a key component of insertion specificity (12, 13). The idea is that geometrical and electrostatic properties of the polymerase active site are likely to have a profound influence on nucleotide-insertion specificities. This influence would strongly favor insertion of bases having an optimal geometry, such that the C1' distances and bond angles most closely approximate those of the Watson–Crick base pairs. For example, G·T, G·A, and C·A mispairs have markedly different bond angles than A·T and G·C pairs (Fig. 1) (14).

The observation by Moran *et al.* (2) that insertion of the base analog F opposite A is reduced by only 40-fold compared with its isosteric parent compound T opposite A suggests that the geometrical alignment of the substrate and template bases is a major determinant of polymerase fidelity. This result can be compared with earlier studies using the base analog 2-aminopurine (2AP), which forms 2AP·T base pairs with two H bonds in a proper W–C geometry and is reduced by 7-fold compared with A·T (15, 16). Thus, the absence of H bonds in the incorporation of F opposite A decreases selectivity only about 6-fold relative to incorporation of 2AP opposite T. Considering that mispairs assuming non-W–C geometries such as G·T (wobble), A·C (protonated wobble), G·A (*anti–syn*) (Fig. 1) are misinserted with frequencies on the order of 10⁻³–10⁻⁶ (11), geometric constraints imposed at the polymerase active site may improve selectivities by perhaps three orders of magnitude or more.

There are at least three possible check points for proper geometric alignment during base insertion by polymerases: initial dNTP binding (16, 17), postbinding selection for the correct geometry (12, 18) by an induced-fit mechanism (19–21), and the chemical step of phosphodiester formation. Previous data suggest there are significant differences in the extent to which different polymerases use each of the check points. We have suggested that the remarkable base-insertion fidelity of DNA polymerases derives from the sequential application of each check point to provide exquisite sensitivity to Watson–Crick geometry at the transition state for phosphodiester formation (13).

Taking the geometrical constraints imposed by the polymerase active site into consideration in conjunction with the active site electrostatic environment, it may be possible to relate the polymerase-insertion selectivity to solution free-energy differences between matched and mismatched base pairs. Measurements of $\Delta G^0 = \Delta H^0 - T\Delta S^0$ indicate relatively small differences between right and wrong base pairs at 37°C; $\Delta\Delta G^0$ is in a range of 0.2–4 kcal/mol, as mentioned above. The differences are small because ΔS^0 correlates with ΔH^0 (5, 22), a phenomenon called enthalpy–entropy compensation, which is observed in aqueous solution (23). That is, it takes more energy to melt highly stable, rigidly constrained base pairs than it does to melt less stable, weakly constrained base pairs.
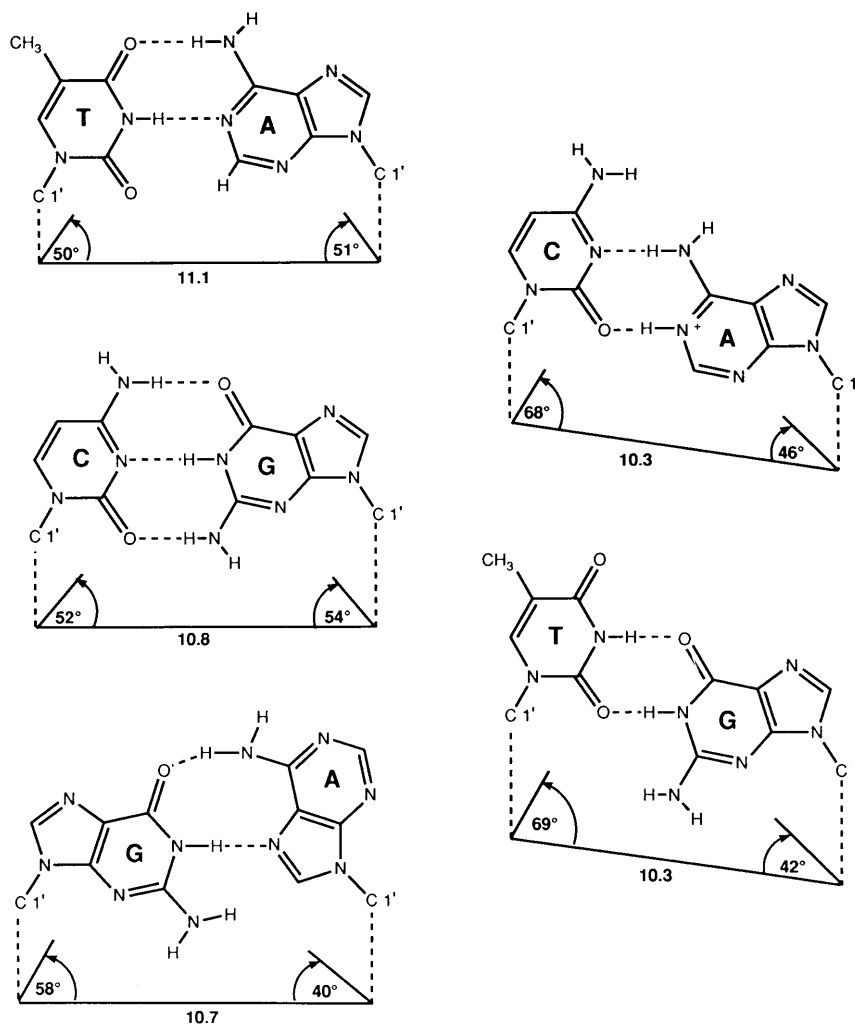
FIG. 1.   Geometric properties of Watson–Crick and mismatched base pairs. This figure is based on x-ray crystallography of duplex B-DNA oligonucleotides. The striking geometric identity of the Watson–Crick A·T and G·C base pairs is not matched by the A·C protonated wobble and G·T wobble base mispairs or by the G(*anti*)·A(*syn*) base mispair. [Reproduced with permission from ref. 14 (copyright 1987, Springer, Heidelberg).]

However, rigid base pairs having fewer degrees of freedom in the double helix will gain more degrees of freedom upon melting, whereas the opposite is true for less stable base pairs. As long as entropy and enthalpy changes are proportional, $\Delta\Delta H^0$ is reduced by $T\Delta\Delta S^0$, resulting in a small $\Delta\Delta G^0$ (5).

How might polymerases increase free-energy differences to achieve high discrimination? Perhaps the geometric constraints imposed on the substrate and template bases in the polymerase active cleft can suppress $\Delta\Delta S$ enough to bring $\Delta\Delta G$ much closer in magnitude to $\Delta\Delta H$. Typical values of $\Delta\Delta H^0$ measured in aqueous solution are, by themselves, almost large enough to accommodate polymerase-insertion fidelities (5). Then, to the extent that dNTP and template bases confront each other in a lower dielectric medium that acts to partially exclude water, $\Delta\Delta H$ values may be even larger than in water (24). Thus, a polymerase active site that snugly accommodates correct base pairs by geometric selection and also reduces water in the vicinity of the base pair may amplify base pair free-energy differences by reducing entropy differences and increasing enthalpy differences by amounts sufficient to account for nucleotide-insertion fidelity.

In addition to discrimination during nucleotide insertion, fidelity in DNA replication is often enhanced by an associated exonuclease activity. Proofreading exonucleases increase fidelity by approximately 40- to 200-fold (25), displaying significantly less selectivity than polymerases. It is generally believed that exonuclease relies on the "melting capacity" of the 3′

terminus to distinguish between correct and incorrect insertions (26, 27). It is reasoned that polymerization and proofreading are competing reactions at a primer-3′ terminus, requiring an annealed or melted terminus, respectively (16, 17). Discrimination arises because the 3′ terminus is more likely to be annealed following correct insertions, favoring polymerization, but much more likely to be melted out following incorrect insertions, favoring excision (16, 28). It is tempting to ask if a geometric selection mechanism might be occurring in the exonuclease active site, enhancing the excision of non-Watson–Crick base pairs beyond what would be expected based solely on the relative stabilities of base pairs in solution.

Qualitatively, differences in base pair stabilities appear to be sufficient. Evidence comes from presteady state kinetics measurements on the excision of 2AP paired opposite T, C, A, and G using bacteriophage T4 DNA polymerase (29). The rate of excision of 2AP from a primer-3′ terminus is inversely correlated with the melting temperature of 2AP·N base pairs imbedded in an oligomer DNA duplex. For example, when present in the same sequence context, a "stable" 2AP·T Watson–Crick base pair is hydrolyzed much more slowly than an unstable 2AP·C wobble mispair. However, a terminal 2AP·T in a A-T rich environment is hydrolyzed more rapidly than 2AP·C in a G-C rich environment. Thus, exonuclease specificity appears to be more strongly tied to DNA stability than to terminal base pair geometry.

Nevertheless, it would be of great interest to use proofreading-proficient polymerases to measure the excision of difluorotoluene. Based on thermal denaturation measurements, Moran et al. (2) demonstrate that F pairs poorly with all of the natural bases. Significantly, the magnitude of the free energy difference between F·A and T·A base pairs is 3.6 kcal/mol, similar to that found for C·A versus T·A base pairs (6). If proofreading activities are governed primarily by primer stability and not geometric selection, then one would expect F to be excised much more rapidly than T opposite A, and, furthermore, the rate of excision of F might be the same whatever natural bases with which it paired.

The surface has barely been scratched in terms of understanding the interactions between polymerases and DNA that determine replication fidelity. The magnitude and location of mutations depend on a complex interplay between polymerases, proofreading exonucleases, processivity factors and the properties of the DNA primer-template sequences. Although models have been proposed to include polymerase steady-state kinetic parameters along with base stacking and sequence context to explain fidelity (11), precise molecular mechanisms governing mutagenic hot and cold spots remain obscure. Different polymerases copying the same primer-template DNA can exhibit markedly different mutation frequencies and spectra. The ability to separate the effects of H bonding from base stacking holds the promise of new progress in these directions. The future use of the difluorotoluene T analog along with an anticipated group of other non-H bonding base analogs should enable a precise determination of the effects of nearest-neighbor base stacking on misinsertion frequencies and proofreading efficiencies, and shed light on how different polymerase and exonuclease active sites sense the presence of nearby primer and template bases.

1. Schweitzer, B. A. & Kool, E. T. (1994) *J. Org. Chem.* **59,** 7238–7242.
2. Moran, S., Ren, R. X.-F. & Kool, E. T. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10506–10511.
3. Moran, S., Ren, R. X.-F., Rumney, S. & Kool, E. T. (1997) *J. Am. Chem. Soc.* **119,** 2056–2057.
4. Aboul-ela, F., Koh, D., Tinoco, I. J. & Martin, F. H. (1985) *Nucleic Acids Res.* **13,** 4811–4825.
5. Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. C., Cheong, C. & Tinoco, I., Jr. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6252–6256.
6. Law, S. M., Eritja, R., Goodman, M. F. & Breslauer, K. J. (1996) *Biochemistry* **35,** 12329–12337.
7. Kunkel, T. A. (1985) *J. Biol. Chem.* **260,** 5787–5796.
8. Preston, B. D., Poiesz, B. J. & Loeb, L. A. (1988) *Science* **242,** 1168–1171.
9. Roberts, J. D., Bebenek, K. & Kunkel, T. A. (1988) *Science* **242,** 1171–1173.
10. Yu, H. & Goodman, M. F. (1992) *J. Biol. Chem.* **267,** 10888–10896.
11. Mendelman, L. V., Boosalis, M. S., Petruska, J. & Goodman, M. F. (1989) *J. Biol. Chem.* **264,** 14415–14423.
12. Sloane, D. L., Goodman, M. F. & Echols, H. (1988) *Nucleic Acids Res.* **16,** 6465–6475.
13. Echols, H. & Goodman, M. F. (1991) *Annu. Rev. Biochem.* **60,** 477–511.
14. Kennard, O. (1987) *Nucleic Acids and Molecular Biology*, eds. Eckstein, F. & Lilley, D. M. J. (Springer, Heidelberg), pp. 25–52.
15. Bessman, M. J., Muzyczka, N., Goodman, M. F. & Schnaar, R. L. (1974) *J. Mol. Biol.* **88,** 409–421.
16. Clayton, L. K., Goodman, M. F., Branscomb, E. W. & Galas, D. J. (1979) *J. Biol. Chem.* **254,** 1902–1912.
17. Galas, D. J. & Branscomb, E. W. (1978) *J. Mol. Biol.* **88,** 653–687.
18. Echols, H. (1982) *Biochimie* **64,** 571–575.
19. Kuchta, R. D., Mizrahi, V., Benkovic, P. A., Johnson, K. A. & Benkovic, S. J. (1987) *Biochemistry* **26,** 8410–8417.
20. Kuchta, R. D., Benkovic, P. & Benkovic, S. J. (1988) *Biochemistry* **27,** 6716–6725.
21. Wong, I., Patel, S. S. & Johnson, K. A. (1991) *Biochemistry* **30,** 526–537.
22. Petruska, J. & Goodman, M. F. (1995) *J. Biol. Chem.* **270,** 746–750.
23. Lumry, R. & Rajender, S. (1970) *Biopolymers* **9,** 1125–1227.
24. Petruska, J., Sowers, L. C. & Goodman, M. F. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1559–1562.
25. Schaaper, R. M. (1993) *J. Biol. Chem.* **268,** 23762–23765.
26. Brutlag, D. & Kornberg, A. (1972) *J. Biol. Chem.* **247,** 241–248.
27. Muzyczka, N., Poland, R. L. & Bessman, M. J. (1972) *J. Biol. Chem.* **247,** 7116–7122.
28. Bessman, M. J. & Reha-Krantz, L. J. (1977) *J. Mol. Biol.* **116,** 115–123.
29. Bloom, L. B., Otto, M. R., Eritja, R., Reha-Krantz, L. J., Goodman, M. F. & Beechem, J. M. (1994) *Biochemistry* **33,** 7576–7586.