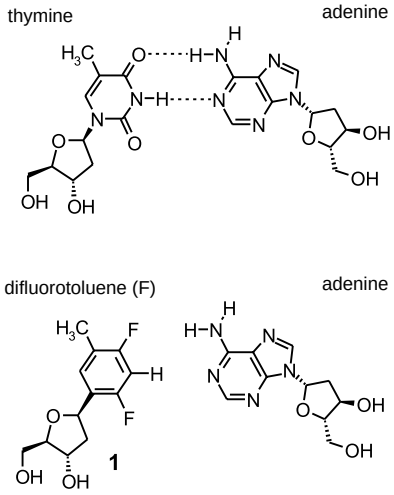
Selectivity of DNA Replication: The Importance of Base-Pair Geometry over Hydrogen Bonding

Ulf Diederichsen*

Polymerase is the central enzyme of the many proteins involved in DNA replication. It is responsible for the selection and bonding of nucleotides complementary to a template strand as well as the correction of faulty base pairings. A high degree of selectivity is prerequisite for the integrity of the genom over generations.[1] The hydrogen-bonding recognition between the decoding nucleobase and its complementary nucleotide is an information-transferring interaction and appears initially to be responsible for selectivity. Considering the many combination possibilities in which a nucleobase can produce two or three hydrogen bonds and a mistake percentage in the duplication of a *Escherichia coli* genom of only $10^{-9}-10^{-10}$ per base, the selectivity in base pairing alone cannot be the reason for this selectivity. The exonuclease as well as the post-replicative mismatch repair procedures have to be taken into consideration along with the aforementioned transfer fidelity of the applicable assortment of nucleotides. The initial insertion of nucleotides is achieved with only one mutation per 10^3 to 10^5 incorporated nucleotides.[2] In addition, the following factors must be considered as reasons for high selectivity along with the base pair identification through hydrogen-bonding interactions: 1. Base stacking; 2. polarity of the base and the resulting solvation; 3. geometry of the recognition (pairing mode, the shape and size of the bases); 4. DNA–protein and protein–substrate interactions (dNTP–substrate binding, geometrical selectivity, and phosphate diester bonding structures); 5. enzyme kinetics; 6. sequential factors. The importance of individual factors must be considered, because the mechanism of the selection and incorporation of nucleotides has not yet been established. Publications from E. Kool's group describe in vitro replication experiments in which instead of thymine the isoster nucleobase analogue difluorotoluene, which has a much lower hydrogen-bonding capability, was used.[3–5] The replication fidelity when using the thymine mimetic remains largely intact, which appears to have led to an underestimation of shape recognition, stacking, and solvation as factors for nucleotide selectivity.

[*] Dr. U. Diederichsen
Institut für Organische Chemie und Biochemie
Technische Universität München
Lichtenbergstrasse 4, D-85747 Garching (Germany)
Fax: (+49) 89 2891 3210
E-mail: ud@linda.org.chemie.tu-muenchen.de

The experimental approach is based on nucleoside **1** which contains the pseudobase difluorotoluene (F) instead of thymine.[6] In peptide chemistry, the use of a fluoroethylene unit as a peptidomimetic and conformational control unit to replace the carboxamide functionality has been established.[7] Through X-ray crystal structure analysis and NMR spectroscopy, the geometrical and conformational high unanimity between nucleoside **1** and thymidine could actually be determined (Scheme 1).[6a] The $C-F\cdots H-N$ hydrogen bonding is weaker and a $C-H\cdots N$ interaction can be rated



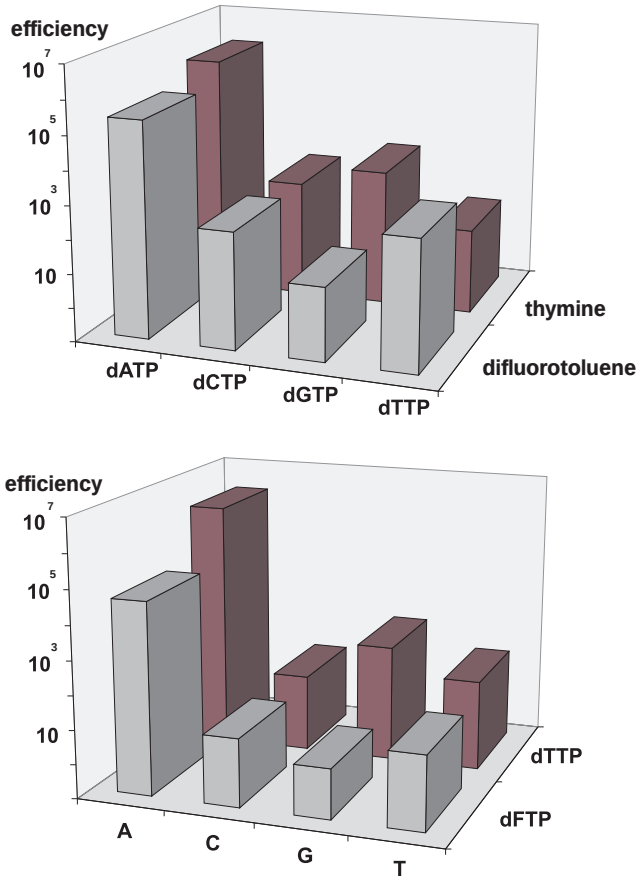
Scheme 1. Difluorotoluene (F) can be substituted for a pyrimidine as a thymine–isoster in Watson–Crick pairing. The ability to form hydrogen bonds is considerably lower for F.

distinctly weaker than the corresponding $C=O\cdots H-N$ and $N-H\cdots N$ hydrogen bonding.[8] An interaction of difluorotoluene with adenine was not even observed in chloroform.[4] An A–F mismatch in the center of a dodecamer double strand resulted in destabilization by 18 K relative to the A–T pairing,[9] which indicates negligible hydrogen bond formation with difluorotoluene. It has to be borne in mind that according to ab initio calculations (6-31G**) the charge distribution and electronic density of the aromatic isoster **1** is similar to that of thymine.[10] Hydrogen bonding, especially in an enzymatic environment in which the orientation of the proton donors and acceptors can be geometrically compelled, cannot be

  1433-7851/98/3712-1655 $ 17.50+.50/0

excluded. In this instance it is assumed that the contribution of an F···H–X interaction gives almost half of that of an O···H–X bond.[8c]

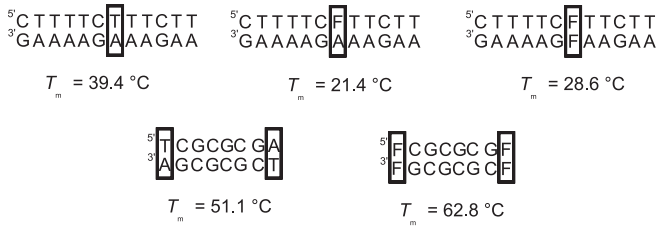
The importance of the hydrogen bond formation with the complementary nucleotide triphosphate was investigated by comparing the replication fidelity of a DNA template that contains the pseudo base difluorotoluene instead of thymine.[3] The replication started with the template-bound primer sequence with the assistance of the Klenow fragment from DNA polymerase I, and the selectivity with regard to insertion of dATP, dGTP, dTTP, or dCTP was studied.[4] A comparably good selectivity for difluorotoluene and thymine with regard to insertion of dATP was demonstrated (Scheme 2 (top)). In both instances the adenosine nucleotide



Scheme 2. Top: Efficiency (v_{max}/K_m) of the incorporation of canonical nucleotide triphosphates as complementary for thymine or difluorotoluene.[4] Bottom: Comparison of the efficiency of incorporation of dTTP and dFTP complementary to A, C, G, or T template bases.[5]

was inserted as complementary with a more than 1000 times greater efficiency than the other nucleotides. A very similar result is displayed for the selectivity in which the nucleotide triphosphate dTTP and mimetic dFTP are identified and inserted as the template complementary (Scheme 2 (bottom)).[5] Although difluorotoluene has a much smaller tendency for hydrogen bonding than thymine, the selectivity is surprisingly similar which points to an insignificant contribution of hydrogen bonding for the recognition of the complementary unit. In both instances the aryl units are of very similar shape and size and a similar preferred conformation and charge distribution appear to be the important selection criteria for the polymerase. These findings are in accord with the suggested "geometrical selection" of Echols and Goodman as the important parameter for insertion specificity.[11] Hence, the selection criterium is the base pair geometry, which is supplied through the pairing mode and which is reflected in the distances between the anomeric centers and the bonding angles.

What importance does solvation and stacking have for the suitability of nucleoside **1** as an isoster mimetic? The comparison of the stability of regular double strands with duplexes that contain A–F or F–F mismatch positions proves the fundamental influence of both factors (Scheme 3):[9] The



Scheme 3. Result of difluorotoluene mismatch pairing on the double strand stability of DNA dodecamers (1.5 μM, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Na-PIPES, pH 7) and octamers (6 μM, 1 M NaCl, 10 mM Na-phosphate, pH 7).[9]

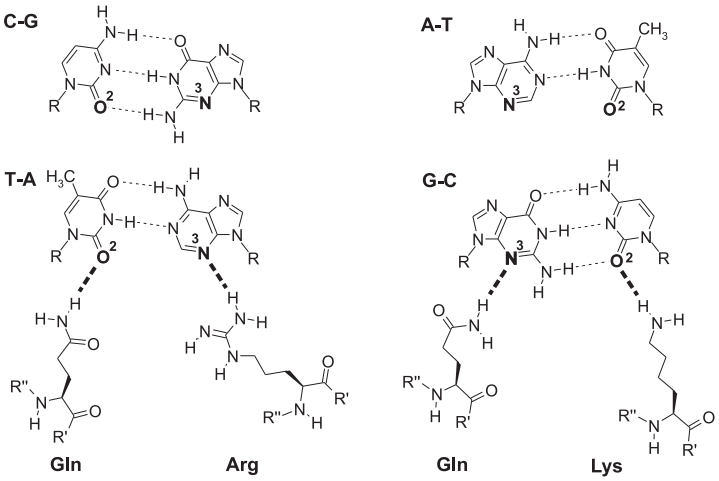
substitution of an A–F mismatch for an A–T base pair severely destabilizes the double strand (the melting point (T_m) is lowered by 18 K) whereas for an F–F mismatch T_m is only 11 K lower than a regular pairing duplex. This tendency is likely to depend on the solvation and stacking contributions. In aqueous solution difluorotoluene is not solvated distinctly, neither in a single nor a double strand. In an A–T base pair the solvation is compensated by the base pairing, whereas adenine in a double strand with A–F mismatch loses its hydrogen bonding without any adequate compensation. The significance of the stacking is indicated by the modifications in the terminal positions of the double strand.[9] The octamer with two F–F mismatch positions is more stable (T_m increases by 12 K) than the oligomer with terminal A–T pairing: This ordering indicates impressively that difluorotoluene provides a higher stacking contribution than adenine.[12]

Polymerases appear to rely to a large extent upon shape, polarity of the bases and the Watson–Crick geometry to accomplish selectivity. These results are in full agreement with two recent publications by Doublié et al.[13] and Kiefer et al.,[14] who described the X-ray structure of the cocrystallization of DNA polymerase with a DNA-template primer double strand.[15] On the one hand, the ternary complex could be stabilized in the crystal with a nucleotide triphosphate when the primer is lacking the terminal 3′-hydroxyl group,[13] while on the other hand, even in the crystal the polymerase–template primer complex retains its catalytic activity and selectivity with respect to the base complementarity.[14] Without detailing the X-ray structures and the mechanism of the phosphate diester bonding formation, three characteristics which have an impact on selectivity must be emphasized:

  1433-7851/98/3712-1656 $ 17.50+.50/0

1. The entering nucleotide triphosphate will be held in a tight pocket by stringently conserved amino acids, so that it arrives in a Watson – Crick orientation facing the complementary nucleobase. The template base is fixed through stacking interactions with the adjacent nucleobase and on the other side with a tyrosine residue or through a van der Waals contact with glycine.

2. Apart from the nonspecific recognition of the polyanionic backbone of the DNA double strand another sequence-independent hydrogen-bonding recognition in the minor groove is ascribed to the first four or five base pairs of the active center. Amino acids with proton donor properties recognize the acceptor position N3 in purine and O2 in pyrimidine (Scheme 4). The A – T and G – C pairs both have



Scheme 4. O2 in pyrimidines and N3 in purines are proton acceptors and point to the minor groove in the DNA double helix. They are hardly any different in their hydrogen-bonding ability and are aligned quasi-symmetrically in the base pair. The recognition of O2 and N3 by Asn, Gln, Arg, His, or Lys is selective for Watson – Crick pairing.

this double-acceptor pattern, irrespective of whether a purine or pyrimidine is the template.[16] This interaction can distinguish between the Watson – Crick pairing mode and all other pairing geometries. Analogous hydrogen-bonding interactions of the polymerase in the major groove in which each base pair has its individual donor/acceptor pattern are nonexistent. The nonspecific recognition is a prerequisite for the elongation of the double strand during continuous synthesis in the enzyme pocket.

3. That functional efficiency and selectivity of the polymerase are preserved in the crystal emphasizes how participating molecules have to be aligned.

As a result the requirement for pseudobases to be incorporated instead of a canonical nucleobase with comparable selectivity is that its corresponding nucleotide phosphate triester analogues are isosters of dNTP with rather similar conformations. Furthermore, it should have an aryl group that shows good stacking interactions, as well as acceptor functionalities in the pseudo N3 and O2 positions, respectively. The difluorotoluene nucleotide **1** fulfills many of these requirements, and it is consequently recognized and incorporated as a Watson – Crick mimetic with high selectivity.

German version: *Angew. Chem.* **1998**, *110*, 1745 – 1747

Keywords: base stacking · DNA replication · geometrical selection · hydrogen bonds · isosteric nucleobases

[1] a) D. Voet, J. G. Voet, *Biochemistry*, 2nd ed, Wiley, New York, **1992**; b) P. Strazewski, C. Tamm, *Angew. Chem*. **1990**, *102*, 37 – 59; *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 36 – 67.
[2] A. L. Loeb, T. A. Kunkel, *Annu. Rev. Biochem*. **1982**, *52*, 429 – 457.
[3] D. Liu, S. Moran, E. T. Kool, *Chem. Biol*. **1997**, *4*, 919 – 926.
[4] S. Moran, R. X.-F. Ren, S. Rumney IV, E. T. Kool, *J. Am. Chem. Soc*. **1997**, *119*, 2056 – 2057.
[5] S. Moran, R. X.-F. Ren, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10506 – 10511.
[6] a) K. M. Guckian, E. T. Kool, *Angew. Chem*. **1997**, *109*, 2942 – 2959; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2825 – 2828; b) X.-F. Ren, B. A. Schweitzer, C. J. Sheils, E. T. Kool, *ibid*. **1996**, *108*, 834 – 837 and **1996**, *35*, 743 – 746.
[7] a) T. Allmendinger, E. Felder, E. Hungerbühler, *Tetrahedron Lett*. **1990**, *31*, 7301 – 7304; b) L. G. Boros, B. De Corte, R. H. Gimi, J. T. Welch, Y. Wu, R. E. Handschumacher, *ibid*. **1994**, *35*, 6033 – 6036; c) P. A. Bartlett, A. Otake, *J. Org. Chem*. **1995**, *60*, 3107 – 3111.
[8] a) J. A. K. Howard, V. J. Hoy, D. O'Hagan, G. T. Smith, *Tetrahedron* **1996**, *52*, 12613 – 12622; b) G. R. Desiraju, *Angew. Chem*. **1995**, *107*, 2541 – 2558; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2311 – 2327; c) W. L. Jorgensen, D. L. Severance, *J. Am. Chem. Soc.* **1990**, *112*, 4768 – 4774.
[9] B. A. Schweitzer, E. T. Kool, *J. Am. Chem. Soc*. **1995**, *117*, 1863 – 1872.
[10] T. A. Evans, K. R. Seddon, *Chem. Commun*. **1997**, 2023 – 2024.
[11] a) M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10493 – 10495; b) H. Echols, M. F. Goodman, *Annu. Rev. Biochem*. **1991**, *60*, 477 – 511.
[12] For the arene stacking, see K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, P. L. Paris, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc*. **1996**, *118*, 8182 – 8183.
[13] S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **1998**, *291*, 251 – 258.
[14] J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *291*, 304 – 307.
[15] For further X-ray structure analysis of a ternary complex of DNA polymerase, DNA template primer and dCTP, see H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, J. Kraut, *Science* **1994**, *264*, 1891 – 1903. See also the commentary of T. A. Steitz, *Nature* **1998**, *291*, 231 – 232.
[16] N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 804 – 808.

  1433-7851/98/3712-1657 $ 17.50+.50/0