ChemComm

Difluorotoluene, a thymine isostere, does not hydrogen bond after all

Xue Wang and K. N. Houk*

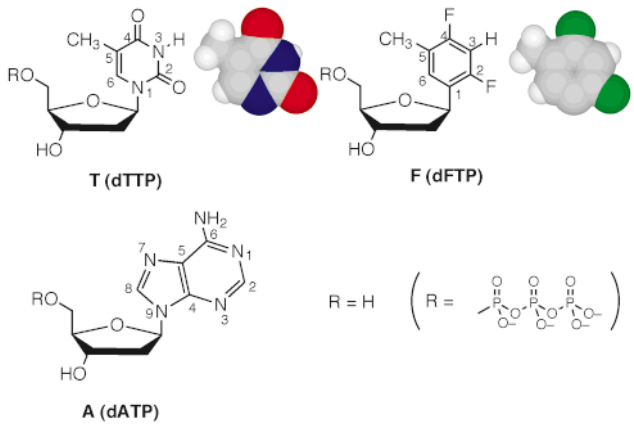
Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. E-mail: houk@chem.ucla.edu

Received (in Corvallis, OR, USA) 26th June 1998, Revised Manuscript Received 20th October 1998, Accepted 21st October 1998

The lack of significant hydrogen bonding by 2,4-difluorotoluene, an isostere of thymine, is confirmed by *ab initio* calculations and force field modelling of nucleic acids.

Watson–Crick base-pairing is the information code which directs replication, transcription and translation.[1] Base-pairing is produced by hydrogen-bonding which stabilizes the DNA double helical structure.[2] The free energy differences between matched and mismatched base pairs in aqueous solution are estimated to be in the range of 0.2 to 0.4 kcal mol^{-1} for terminal base pairs[3] and 1 to 3 kcal mol^{-1} for internal base pairs.[4] In 1988, Echols and Goodman noted that the energy difference between hydrogen bonding of matched and mismatched terminal pairs is insufficient to account for the remarkably high fidelity of Watson–Crick base pairing in DNA replication by the DNA polymerase (10^4 to 10^5).[5] It was proposed that the geometry of the complex between the template and substrate on DNA polymerase was more important than hydrogen bonding in controlling polymerase fidelity.

In 1994, Kool and co-workers proposed that there is no significant hydrogen bonding between adenine (**A**) and the



thymine (**T**) isostere, **F**.[6] Nevertheless, numerous studies now show that **F** can be readily incorporated in place of **T** by DNA polymerase.[7] In **F**, fluorines replace oxygens, and carbons replace nitrogens. Much evidence has been amassed in accord with the idea. ^{1}H NMR and X-ray data indicated that **F** does not H-bond with natural nucleotides in solution.[8] Single nucleotide insertion and 'running start' experiments demonstrated that **F** expressed similar selectivity to **T** during the replication of a DNA strand in the presence of Klenow fragment (KF) of *Escherichia coli* DNA polymerase.[7] The geometry of the reaction complex, not hydrogen bonding, seems of principal importance in the fidelity of the DNA polymerase.

However, Evans and Seddon proposed that the hydrogen bonding between **A** and **F** still plays an important role in DNA replication even with templates containing **F**.[9] In their communication, *ab initio* RHF/6-31G**, semi-empirical AM1 and PM3 calculations were presented which reveal large dipole moments of the methyl derivative of 2,4-difluorotoluene **1** (1.86 D) in comparison to H_2O (1.82 D). They suggested that the polar nature of 2,4-difluorotoluene is important for hydrogen bonding and hence, DNA replication. They reported the electrostatic potential surfaces of **1** and N^1-methylthymine **2** [Fig. 1(*a*)]. These are clearly similar, suggesting that **F** mimics the shape, the charge distribution and hydrogen-bonding patterns of **T** [Fig. 1(*b*)]. Evans and Seddon supported the 'status quo' concerning the major role of hydrogen-bonding on replication fidelity, even with **F**.

We have reinvestigated these issues with quantum mechanics and molecular mechanics, to test the ability of hydrogen-bonding by **F** and the potential role in base-pairing interactions in DNA.

It is well known that fluorinated olefins and aromatic compounds are poor hydrogen bond acceptors. Fluorine has a greater electronegativity and lower polarizability than oxygen, which renders it a much poorer hydrogen bond acceptor. At the MP2/TZV++(3d,1f,1p) level of theory, vinyl C(sp^2)–F forms hydrogen bonds with water with a strength of only 1.5 kcal mol^{-1}, significantly lower than typical hydrogen bonds of 5 to 10 kcal mol^{-1}.[10] **F** contains two aromatic C(sp^2)–F moieties. The hydrogen bonding strength of **F** with water should be even less.

We recomputed **1** and **2** with semi-empirical PM3, *ab initio* RHF/3-21G and RHF/6-31G** calculations.[11] At the RHF/6-31G** level, the electrostatic charges of fluorines in **1** are both 0.21, and the dipole moment of **1** is 1.90 D. The electrostatic charges of the oxygens in **2** are both 0.61, and the dipole moment of **2** is 4.83 D. The dipole moment of the water molecule has also been calculated using the RHF/6-31G** method for comparison, giving a value of 2.15 D. In spite of the representation displayed in Fig. 1(*a*), the fluorines of **F** and oxygens of **T** have significantly different electrostatic charges. The electrostatic potential range used in the published pictorial representation for **1** is from −19 to +27 electron per atomic unit, while it is from −46 to +51 electron per atomic unit for **2**.

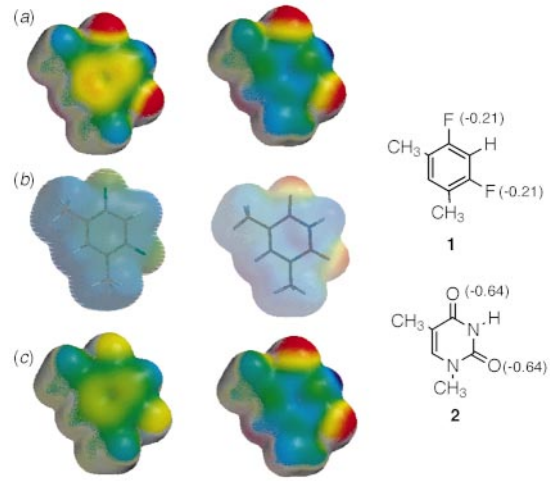


Fig. 1 The electrostatic potential surfaces of **1** and **2**: (*a*) as presented in ref. 9 [the range of electrostatic potential shown in thymine is from red (−46.0) to blue (+50.5) and in difluorotoluene is from red (−19.0) to blue (+27.3); the units are electrons per atomic unit), (*b*) PM3 model and (*c*) RHF/6-31G** model [both (*b*) and (*c*) using the same potential range (−46.1 to +50.5 from red to blue for RHF) for both structures from our calculations; the units are electrons per atomic unit and charges are given in parenthesis].

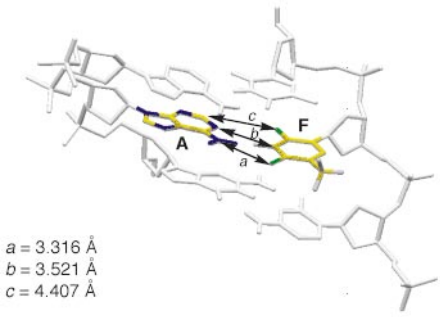


Fig. 2 Molecular modelling shows no obvious hydrogen bonding of **F** with **A**.

Because of this, the color pictures in Fig. 1(*a*) provide the incorrect impression that the electrostatic potentials are very similar. When plotted on the same color scale [Fig. 1(*b*),(*c*)], it is obvious that the electrostatic potential range of **1** is much more limited than that of **2**. Hydrogen bonds involving **F** will be very weak, perhaps impossible to detect in solution. Indeed, this is the conclusion of Kool *et al*. based upon experimental data.[8] The dipole moment of **1** is also significantly smaller than **2**, even smaller than H_2O. Since replication experiments are performed in aqueous solution, DNA strands are solvated. Hydrogen bonding between **F** and **A** is not competitive with hydrogen bonding between solvent water molecules and **A**, even if **F** can weakly complex with **A** in the gas phase.

Since the Evans–Seddon paper, others have tested the ideas computationally.[12,13] Meyer and Sühnel did calculations on **F**–**A** and **T**–**A** complexes using the RHF, B3LYP and MP2 methods and the 6-31G** basis set.[12] Without zero point correction, the complexation energy for **T**–**A** is 11.7 kcal mol^{-1}, significantly larger than the complexation energy for **F**–**A** (−3.8 kcal mol^{-1}).

However, in aqueous solvent, the loosely complexed **F**–**A** is readily dissociated. Santhosh and Mishra did calculations on **F**–**A** and **T**–**A** systems in a solvent cavity model (PCM) at the 4-31G level with AM1 geometries.[13] The complexation energy of **F**–**A** becomes repulsive (3.5 kcal mol^{-1}) in aqueous media, while it is attractive for the **T**–**A** complex (−11.1 kcal mol^{-1}). The best distance for fluorine of **F** and nitrogen of **A** is 3.533 Å (calculated from their graph), significantly longer than that of **T**–**A** (3.0 Å). We performed RHF/6-31G* optimizations and SCI-PCM solvent cavity model calculations on **F**–**A** and **T**–**A** as well. With RHF/6-31G* optimized geometries, and SCI-PCM solvation calculations at the RHF/6-31G* level, the energy difference between **F**–**A** and uncomplexed **A** and **F** is −1.4 kcal mol^{-1}. This means that the **F**–**A** complex becomes repulsive in aqueous solution, as Santhosh and Mishra showed.[13] The complexation energy for **T**–**A** in water is computed to be −5.2 kcal mol^{-1}. **F** does not hydrogen bond with **A** in aqueous solution.

We have also modelled the polydeoxynucleotides containing **F**–**A** or **T**–**A** complexes. The AMBER* force field was employed in these calculations.[14] The charges on the oxygens in **T** are −0.47 and −0.53, respectively, and the charges on the fluorines in **F** are both −0.25. A strand of 3-nt DNA is shown in Fig. 2. Whereas the distances between O^4 of **T** and NH_2 of **A**, and between N^3 of **T** and N^1 of **A**, are computed to be 2.832 and 2.851 Å respectively, the distance between F^4 of **F** and NH_2 of **A** is 3.316 Å, and the distance between C^3 of **F** and N^1 of **A** is 3.521 Å. These are 0.484–0.670 Å longer, a clear indication of the absence of hydrogen-bonding in the small DNA oligomer. Importantly, the distance between N^3 of **T** and N^1 of **A** calculated with the AMBER* force field (2.851 Å) is close to those measured from high resolution DNA X-ray structures (2.82 ± 0.07 Å),[12] and the distance between F^4 of **F** and N^1 of **A** calculated with the AMBER* force field (3.521 Å) is close to the value obtained from RHF/4-31G//AM1 calculations by Santhosh and Mishra (3.533 Å).[13]

Force field calculations of a 12-nt DNA fragment with the GB/SA* solvation model[15] also show that there is no hydrogen-bonding between **F** and **A** (Fig. 3). In DNA double helices, the

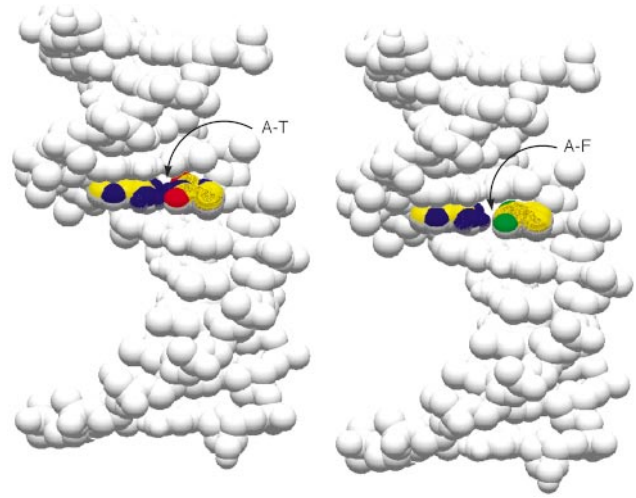


Fig. 3 The structure of 12-nt DNA with **A**–**T** or **A**–**F** base pairs.

base pairs adjacent to **F**–**A** are strongly hydrogen bonded. This, along with π-stacking to **F** and **A**, keeps the **F**–**A** base pair from separating. Aqueous solvation also keeps DNA strands tightly coiled. Even with the assistance from neighboring base pairs, the **F**–**A** hydrogen bonding distances are still 0.6–0.7 Å longer than those of **T**–**A**. All evidence suggests that there is no hydrogen bonding between **F** and **A**. However, the π-stacking and the shape of the whole DNA strands are not distorted significantly with **F** in place of **T**. As a nearly perfect isostere, **F** should be able to fit into the DNA template-polymerase-substrate complex without disturbing the structure.

The experimental and theoretical studies support the conclusion of Kool *et al*. that the hydrogen bonding of **F**–**A** is very unlikely to play an important role in DNA replication. Geometrical effects must account for the high fidelity of DNA replication in the presence of polymerase.

Notes and references

1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737.
2 G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, Freeman, San Francisco, 1960; G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer-Verlag, New York, 1991; G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford, New York, 1997.
3 J. Petruska, M. F. Goodman, M. S. Boosalis, L. C. Sowers and C. Cheong, *Proc. Natl. Acad. Sci. U.S.A.*, 1988, **85**, 6252.
4 L. A. Loeb and T. A. Kunkel, *Annu. Rev. Biochem.*, 1982, **52**, 429.
5 H. Echols and M. F. Goodman, *Annu. Rev. Biochem.*, 1991, **60**, 477; M. F. Goodman, *Proc. Natl. Acad. Sci. U.S.A.*, 1997, **94**, 10 493.
6 B. A. Schweitzer and E. T. Kool, *J. Org. Chem.*, 1994, **59**, 7238.
7 R. X.-F. Ren, B. A. Schweitzer, C. J. Sheils and E. T. Kool, *Angew. Chem., Int. Ed. Engl.*, 1996, **35**, 743; S. Moran, R. X.-F. Ren, C. J. Sheils, S. Rumney and E. T. Kool, *Nucleic Acid Res.*, 1996, **24**, 2044; R. X.-F. Ren, N. C. Chaudhuri, P. L. Paris, S. Rumney and E. T. Kool, *J. Am. Chem. Soc.*, 1996, **118**, 7671; K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, P. L. Paris, D. C. Tahmassebi and E. T. Kool, *J. Am. Chem. Soc.*, 1996, **118**, 8182; S. Moran, R. X.-F. Ren, S. Rumney IV and E. T. Kool, *J. Am. Chem. Soc.*, 1997, **119**, 2056; S. Moran, R. X.-F. Ren and E. T. Kool, *Proc. Natl. Acad. Sci. U.S.A.*, 1997, **94**, 10 506; D. Liu, S. Moran and E. T. Kool, *Chem. Biochem.*, 1997, **4**, 919.
8 B. A. Schweitzer and E. T. Kool, *J. Am. Chem. Soc.*, 1995, **117**, 1863; K. M. Guckian and E. T. Kool, *Angew. Chem., Int. Ed.*, 1998, **36**, 2825.
9 T. A. Evans and K. R. Seddon, *Chem. Commun.*, 1997, 2023.
10 J. A. K. Howard, V. J. Hoy, D. O'Hagan and G. T. Smith, *Tetrahedron*, 1996, **52**, 12 613.
11 SPARTAN Version 4.0, Wavefunction, Inc., 18401 Von Karman Ave., #370, Irvine, CA 92715, USA.
12 M. Meyer and J. Sühnel, *J. Biomol. Struct. Dynamics*, 1997, **15**, 619.
13 C. Santhosh and P. C. Mishra, *Int. J. Quantum Chem.*, 1998, **68**, 351.
14 F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Canfield, G. Chang, T. Hendrickson and W. C. Still, *J. Comput. Chem.*, 1990, **11**, 440.
15 D. Qiu, P. S. Shenkin, F. P. Hollinger and W. C. Still, *J. Phys. Chem.*, 1997, **101**, 3005.

Communication 8/07425I