

# Unnatural Covalent DNA Base Pairing: Quantum Chemical Study

Eugene S. Kryachko\* and Minh Tho Nguyen

Department of Chemistry, University of Leuven, Celestijnenlaan 200 F, B-3001 Leuven, Belgium

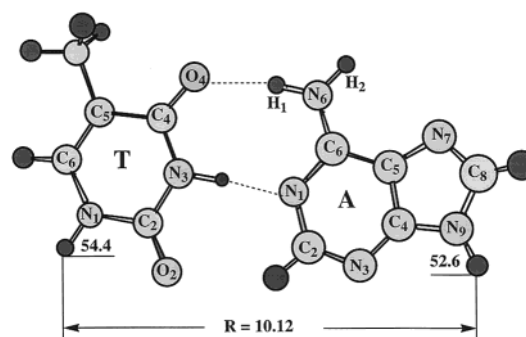
Received: March 27, 2002; In Final Form: June 17, 2002

The present work is aimed to perform a high-level ab initio computational study of the unnatural covalent bond base pairs, recently designed by Gao and Orgel (*Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14837) and Kishi et al. (*Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 928; *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 1471), for a better understanding of the underlying structural and energetic features of their base pair formations and for a demonstrating of how they actually accommodate the DNA double helix architecture. In addition, three new tautomeric forms of the Gao–Orgel covalent pair are found and their shape complementarity is analyzed.

## 1. Introduction

The Watson–Crick structure of the DNA comprises of the double helix. Two types of interaction, hydrogen bonding and stacking, maintain the DNA double helix structural motif where the former orchestrates the base pairing between adenine A and thymine T and between guanine G and cytosine C nucleic acids<sup>1</sup> (see also ref 2 for review and references therein), linking together two DNA strands. Stacking interaction primarily determines the helix architecture.<sup>3</sup> The nucleic acid base pairs A•T and G•C of DNA are usually referred to as the Watson–Crick (WC) complementary or canonical pairs (the former is pictured in Figure 1). The specific hydrogen-bonded patterns in the A•T and G•C base pairs predetermine the genetic code alphabet within the Watson–Crick model of DNA.<sup>1,2</sup> A case with altered hydrogen-bond patterns or, in other words, non-WC base pairings (mismatches or mispairings) might therefore cause an alternation of the genetic code leading to point mutations.<sup>1b,4,5</sup>

What is the key place of the hydrogen bonding in maintaining the DNA double helix architecture comparing with the inter-strand base-stacking and interstrand cross-stacking interactions? Whether the WC hydrogen bonds are required for a DNA polymerase to replicate a base pair? Is it possible to design such “unnatural” base pairs which will be not susceptible to effects of tautomerization? This fundamental paradigm,<sup>2f,6–13</sup> arisen since the discovery of the DNA structure in 1953, led in the mid 1980s to the concept of a “covalent base pair”, introduced by Devadas and Leonard.<sup>6</sup> During the last 2 decades, this concept has greatly inspired an appearance of a variety of covalently linked base pairs.<sup>6–10</sup> Nowadays, it has been reconsidered mainly thanks to the studies by Kool, Romesberg, and Schultz<sup>11,12,2f</sup> (see also ref 13). The current work in this area has particularly been focused on the base pairing between adenine and difluorotoluene and on a design of new covalently bonded base pairs by Gao and Orgel<sup>9</sup> and Kishi and co-workers.<sup>10</sup> The concept of a covalent 2 base pairing has recently been extended to other types of pairing of unnatural bases, particularly those designed by Schultz and Romesberg and co-workers on a basis of hydrophobic interactions.<sup>12</sup> An advantage of the hydrophobic bases consists first of all in that they are



**Figure 1.** The Watson–Crick canonical base pair A•T and its shape complementarity characteristics.

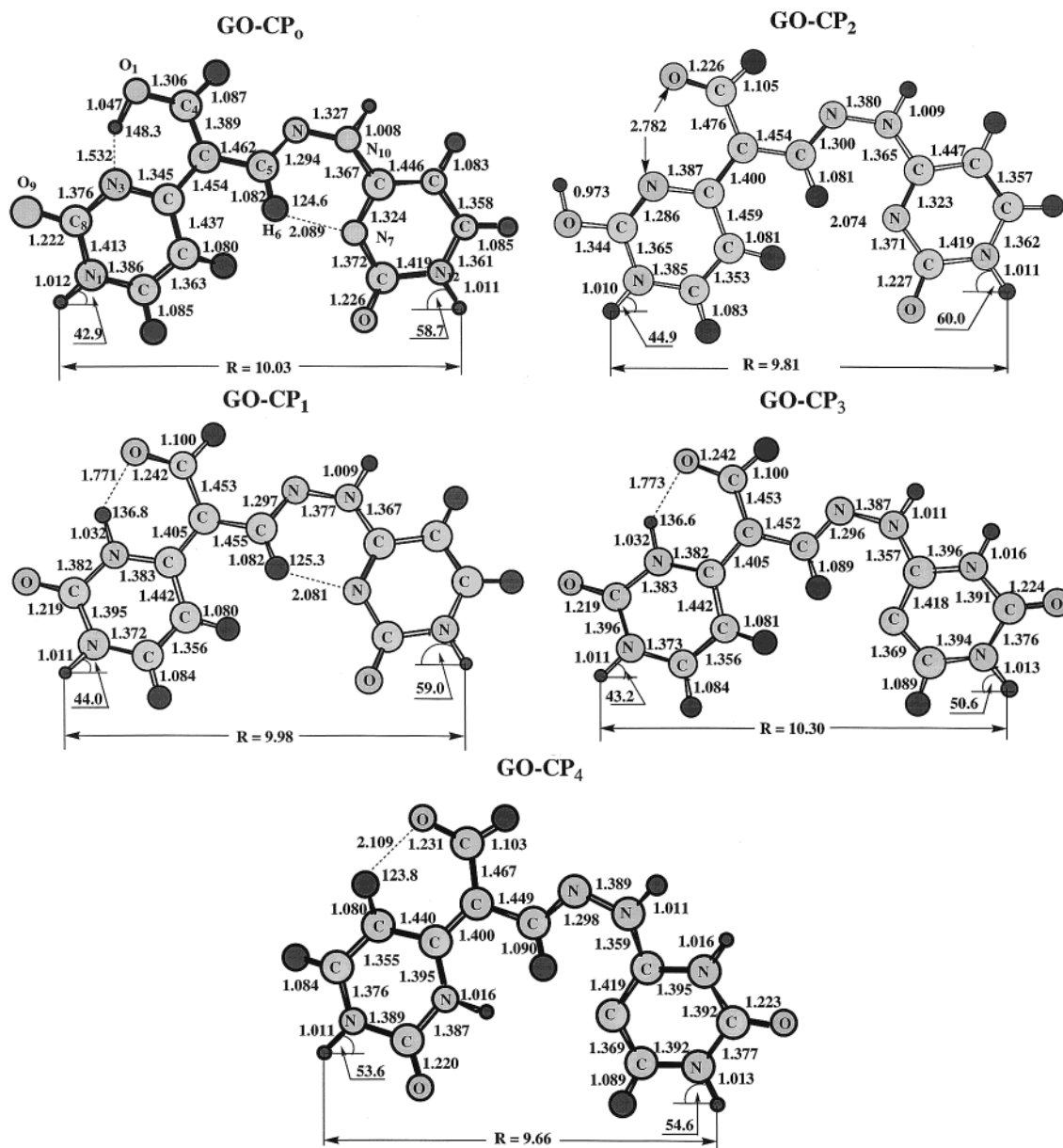
incapable of tautomerization and therefore, an insertion of unnatural base pairs into the DNA liaisons might significantly disfavor mispairing. One of the most important problems in the design of unnatural base pairs is how they geometrically accommodate the standard Watson–Crick architecture, without steric clashes.<sup>12,14</sup> Notice that such geometric constraint, or in other words, a shape complementarity, has long been thought to be the important factor in the DNA replication fidelity.<sup>2c–d,12,13a,15,16</sup>

The present paper reports a high-level performance theoretical study of the covalently bonded base pairs, focusing on the Gao–Orgel and Kishi models (computational notices are outlined in ref 17), with the emphasis on the analysis of their shape complementarity within the DNA double helix architecture. We thoroughly examine their optimized geometries and IR spectra and demonstrate some subtle features of these base pairs related particularly to their tautomerism, which might have an impact on a further development of the covalently bonded base-pair architecture. We also consider new covalent base pairs as analogues of Gao–Orgel model.

## 2. Gao–Orgel Covalent Base Pair and Its Tautomers

The Gao–Orgel covalently bonded base pair GO–CP<sub>0</sub> with the formyl-type hydrogen bond, shown in Figure 2, has recently been synthesized in the reaction of two nucleoside analogues, the dialdehyde and *N*<sup>4</sup>-aminocytidine. The semiempirical PM3 calculation, conducted by Gao and Orgel, demonstrates that, first, this base pair closely resembles the Watson–Crick base pair geometry; second, it is in fact planar, and third, it minimally distorts the double-helix structure motif (see Figure 1 for

\* Corresponding author. Tel: +32 (16) 32 73 84. FAX: +32 (16) 32 79 92. E-mail: eugene@bohr.quantchem.kuleuven.ac.be. On leave from Bogoliubov Institute for Theoretical Physics, Kiev, 03143 Ukraine.



**Figure 2.** The optimized geometries of the covalent base pairs GO-CP<sub>0-4</sub>. Distances in angstroms and angles in degrees.

the definition): its C' to C' distance  $R$  equal to 10.5 Å and glycosyl bond angles  $\alpha_1 = 48.5^\circ$  and  $\alpha_2 = 57.1^\circ$  are very close to those pertaining to the double-helix DNA ( $R = 10.8$  Å and both  $\alpha_1 = \alpha_2 = 52^\circ$ <sup>2b,c</sup> or  $\alpha_1 = \alpha_2 \approx 56^\circ$ <sup>16a</sup>  $R = 10.60 \pm 0.15$  and  $\alpha_1 = \alpha_2 = 68 \pm 2^\circ$ <sup>2</sup>  $R = 10.60 \pm 0.15$  Å;<sup>2</sup> see also refs 16b,c).

The B3LYP/6-31+G(d,p) optimized structure of the Gao-Orgel covalently bonded base pair GO-CP<sub>0</sub>, obtained in the present work, is displayed in Figure 2. It is perfectly planar and strongly polar compared to the A·U base pair (see Table 1). Its geometrical parameters  $R^{\text{GO-CP}_0}$ ,  $\alpha_1^{\text{GO-CP}_0}$ , and  $\alpha_2^{\text{GO-CP}_0}$  are equal to 10.03 Å, 42.9°, and 58.7°, respectively. Their comparison with, for instance, the corresponding values of  $R^{\text{AU}}$ ,  $\alpha_1^{\text{AU}}$ , and  $\alpha_2^{\text{AU}}$  of the A·U base pair, obtained at the same computational level,<sup>19</sup> shows that  $R^{\text{GO-CP}_0}$  is smaller than  $R^{\text{AU}}$  by 0.09 Å and  $\alpha_1^{\text{GO-CP}_0}$  by 11.5° than  $\alpha_1^{\text{AU}}$ , whereas  $\alpha_2^{\text{GO-CP}_0}$  slightly exceeds (by 3.2°)  $\alpha_2^{\text{AU}}$ . The key properties of the base pair GO-CP<sub>0</sub>, including the rotational constants, mean dipole polarizability and mean quadrupole moment, are gathered in

Table 1. It is worth mentioning that a large dipole polarizability of GO-CP<sub>0</sub> implies its higher interaction capability with polar solvents.

The base pair GO-CP<sub>0</sub> is characterized by two intramolecular hydrogen bonds: the O<sub>1</sub>-H<sub>2</sub>···N<sub>3</sub> bond with the bond length of 1.532 Å and bond angle of 148.3° between the formyl group C<sub>4</sub>=O<sub>1</sub> and N<sub>3</sub> of the dialdehyde ring and the C<sub>5</sub>-H<sub>6</sub>···N<sub>7</sub> one whose bond length is equal to 2.089 Å and bond angle to 124.6°. The former refers to the formyl-type hydrogen bond which is characterized by the highly IR active O<sub>1</sub>-H<sub>2</sub> stretching vibrational mode at 2412 cm<sup>-1</sup> (Table 2). A proton transfer along this hydrogen bond results in the amide tautomeric form GO-CP<sub>1</sub> of GO-CP<sub>0</sub>, which appears to be more stable by 5.9 kcal/mol (see Figure 2). Due to the entropy excess, equal to 6.4 cal/mol·T of the GO-CP<sub>1</sub> base pair over the original GO-CP<sub>0</sub> one, its stability increases to 7.1 kcal/mol at ambient temperatures.

The tautomerization process GO-CP<sub>1</sub> ⇌ GO-CP<sub>0</sub> is accompanied by substantial geometrical changes in the vicinity of the N<sub>3</sub>···H<sub>2</sub>···O<sub>1</sub> proton-share hydrogen bond. As shown in

**TABLE 1: B3LYP/6-31+G(d,p) Relative Energies (in kcal/mol), Rotational Constants (in MHz), Total Dipole Moments (in debye), Mean Dipole Polarizabilities  $\alpha = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$  (in au), Quadrupole Moments  $Q = (Q_{xx} + Q_{yy} + Q_{zz})/3$  (in D·Å), and the Base Geometrical Parameters of the normal Amide Covalent Base Pair GO-CP<sub>1</sub>, Its Formyl- and Enol-Tautomers, GO-CP<sub>3</sub>, GO-CP<sub>4</sub>, and the WC A·U**

form	energy	rotational constants			dipole	$\alpha$	$Q$	geometrical parameters		
		$A$	$B$	$C$				$R$ (Å)	$a_1$ (deg)	$a_2$ (deg)
amide	0.0	679.28 (690.56) <sup>a</sup>	160.61 (162.08)	129.92 (131.27)	7.9 (8.5)	221.4	111.3 (111.9)	9.98 (9.94)	44.0 (44.3)	59.0 (58.4)
formyl	5.9	691.08	159.75	129.76	8.5	214.2	111.4	10.03	42.9	58.7
enol	19.0	653.53	164.19	131.22	4.8	226.2	101.9	9.81	44.9	60.0
GO-CP <sub>3</sub>		822.77 (840.93)	134.74 (135.33)	115.78 (116.57)	5.8		177.1	10.30 (10.35)	43.2 (42.8)	50.6 (49.7)
GO-CP <sub>4</sub>		598.93 (624.09)	163.74 (160.07)	128.58 (127.39)	2.5		116.2	9.66 (9.91)	53.6 (51.1)	54.6 (52.6)
A·U		200.32	169.17	108.79	1.8	163.9	89.3	10.12	54.41	55.54

<sup>a</sup> The corresponding HF/6-31+G(d,p) values are given in parentheses. The B3LYP/6-31+D(d,p) electronic energy of the base pair GO-CP<sub>0</sub> is equal to  $-979.42967$  hartree.

**TABLE 2: Stretching Vibrational Modes of the normal Covalent Base Pair GO-CP<sub>1</sub> and Its Formyl- and Enol-Tautomeric Forms**

GO-CP <sub>1</sub>	frequency		assignment
	GO-CP <sub>0</sub>	GO-CP <sub>2</sub>	
3266 (179) <sup>a</sup>			$\nu(\text{N}_3\text{-H})$
3631 (121)	3630 (121)	3633 (121)	$\nu(\text{N}_{10}\text{-H})$
3644 (181)	3624 (127)	3644 (217)	$\nu(\text{N}_{11}\text{-H})$
3654 (61)	3655 (59)	3652 (59)	$\nu(\text{N}_{12}\text{-H})$
		3753 (123)	$\nu(\text{O}_9\text{-H})$

<sup>a</sup> Frequencies are given in  $\text{cm}^{-1}$ ; IR intensities in  $\text{km/mol}$  are shown in parentheses.

Figure 2, as far as the proton migrates along this bond, it undergoes a contraction by  $\sim 0.14$  Å. This also causes a contraction of the  $\text{C}_9=\text{O}_{10}$  bond by  $0.06$  Å which in turn results in a small blue shift ( $\approx 19$   $\text{cm}^{-1}$ ) of the  $\text{C}_4=\text{O}_1$  stretching vibration and in an enhancement of IR intensity by a factor of ca. 5. In contrast, the intra-ring bonds in the vicinity of the  $\text{N}_3$  atom undergo smaller changes (ca.  $\pm 0.01\text{--}0.04$  Å). The net effect of the proton transfer on the  $\text{N}_4$ -aminocytidine geometry is negligible although it slightly changes the  $C'$  to  $C'$  distance and glycosyl bond angles to  $9.98$  Å, and  $44.0^\circ$  and  $59.0^\circ$ , respectively. The key properties of the covalent base pair GO-CP<sub>1</sub> are summarized in Table 1. Notice that its amide-type hydrogen bond  $\text{N}_3\text{-H}_2\cdots\text{O}_1$  is characterized by the bond length of  $1.771$  Å and the bond angle of  $136.8^\circ$ . The N-H stretching vibration is predicted at  $3266$   $\text{cm}^{-1}$  (see Table 2).

A further proton transfer from  $\text{N}_3$  to the carbonyl oxygen atom  $\text{O}_7$  leads to the enol-type tautomer GO-CP<sub>2</sub> displayed in Figure 2. It is less stable by  $18.8$  kcal/mol and less polar by  $3.7$  D than the covalent base pair GO-CP<sub>0</sub>. GO-CP<sub>2</sub> has a slightly smaller  $C'$  to  $C'$  distance of  $9.81$  Å. A comparison of the geometries of the dialdehyde ring in the GO-CP<sub>1</sub> and GO-CP<sub>2</sub> base pairs demonstrates that the keto-enol tautomerism leads to the substantial elongation of the  $\text{C}_4\text{-O}_1$  bond by  $0.12$  Å and contraction of the  $\text{N}_3\text{-C}_8$  one by ca.  $0.1$  Å.

Let us now estimate the occurrence of the GO-CP<sub>2</sub> base pair using the well-known expression for the equilibrium constant  $k_{\text{GO-CP}_1 \rightleftharpoons \text{GO-CP}_2}$  of the tautomerization process  $\text{GO-CP}_1 \rightleftharpoons \text{GO-CP}_2$ :  $k_{\text{GO-CP}_1 \rightleftharpoons \text{GO-CP}_2} = \exp[-(\Delta H - T\Delta S)/k_B T]$  where  $\Delta H$  is the difference in enthalpies of GO-CP<sub>1</sub> and GO-CP<sub>2</sub>,  $\Delta S$  is their corresponding entropy difference, and  $k_B$  is Boltzmann constant equal to  $198.72156 \times 10^{-2}$  cal/mol T. We then obtain that  $k_{\text{GO-CP}_1 \rightleftharpoons \text{GO-CP}_2} = 6.0 \times 10^{-19}$  which is much lower, by eleven orders, than the estimated misinsertion frequency of the spontaneous point mutations in the conventional DNA base pairs.<sup>20</sup> We therefore conclude that the covalently bonded base

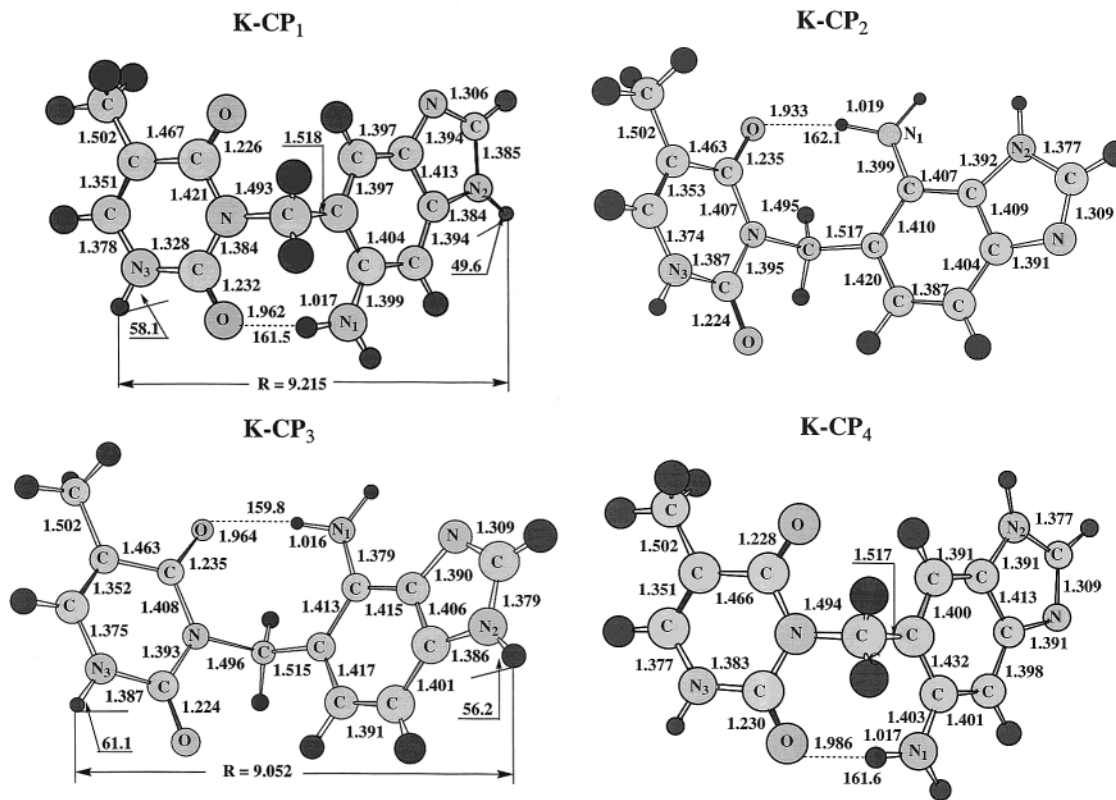
pair GO-CP<sub>1</sub> does not substantially suffer from possible GO-CP<sub>2</sub>-type mismatches. Although, on the other hand, the equilibrium constant  $k_{\text{GO-CP}_1 \rightleftharpoons \text{GO-CP}_0}$  of the tautomerization  $\text{GO-CP}_1 \rightleftharpoons \text{GO-CP}_0$ , equal to  $6.0 \times 10^{-6}$ , demonstrates that the occurrence of the GO-CP<sub>0</sub> mismatches is of the same order as that in the DNA base pairs. However, we should notice that such type of mismatches is not precisely identical to that occurred in DNA<sup>21</sup>, and comparing the geometries of GO-CP<sub>1</sub> and GO-CP<sub>0</sub>, we suggest that such tautomerism may only cause negligibly minor perturbations with respect to the overall double helix, without, likely, any extension beyond the mismatched covalent base pair.

Closing this section, we consider two new covalent base pairs GO-CP<sub>3</sub> and GO-CP<sub>4</sub>. Their B3LYP/6-31+G(d,p) optimized and perfectly planar geometries are shown in Figure 2 and their properties are gathered in Table 1. It is worth noticing that GO-CP<sub>3</sub> is characterized by the  $C'$  to  $C'$  interglycosidic distance  $R = 10.30$  Å and by the glycosyl bond angles  $\alpha_1 = 43.2^\circ$  and  $\alpha_2 = 50.6^\circ$ . It is then clear that  $R$  certainly lies in a better agreement with the experimental range of  $R$  inherent for the DNA base pairs than that of GO-CP<sub>0</sub> and GO-CP<sub>1</sub>. Furthermore, the bond angle  $\alpha_1$  is nearly the same as  $\alpha_1^{\text{GO-CP}_0}$  and  $\alpha_1^{\text{GO-CP}_1}$ , while  $\alpha_2$  is closer to the experimental value of  $52^\circ$ .<sup>2b,c</sup> Altogether, this implies that GO-CP<sub>3</sub> has much more in common with the DNA double helix motif than the base pairs GO-CP<sub>0</sub> and GO-CP<sub>1</sub>. Interestingly, the base pairs GO-CP<sub>1</sub> and GO-CP<sub>3</sub> have the amide-type hydrogen bond with rather similar geometries.

The covalent base pair GO-CP<sub>4</sub> is characterized by a smaller  $C'$  to  $C'$  interglycosidic distance  $R = 9.66$  Å. Its glycosyl bond angles  $\alpha_1 = 53.6^\circ$  and  $\alpha_2 = 54.6^\circ$  perfectly match the DNA experimental range (cf. with the values of the A·U given in Table 1), and thus, such covalent pair excellently accommodates the DNA double helix among all covalent base pairs of the Gao-Orgel type treated in the present work. It also possesses a lower total dipole moment of  $2.5$  D, comparable to that of the conventional WC base pairs.

### 3. Kishi Covalently Cross-Linked Base Pair

Recently, Kishi and co-workers<sup>10</sup> have designed the  $\text{CH}_2$ -bridged base pairs and reported their synthesis. Comparing them, for instance, with the WC A·T base pair, we notice that they possess the single hydrogen bond of the  $\text{N-H}\cdots\text{O}$  type which, as known, is characterized by the proton-transfer barrier of more than  $12$  kcal/mol, and therefore, the resulting tautomeric mispair is unlikely to occur. Following the original order,<sup>10</sup> Figure 3 displays the B3LYP/6-31+G(d,p) optimized geometries of four  $\text{CH}_2$ -bridged base pairs. The former two belong to class I



**Figure 3.** The optimized geometries of Kishi covalent base pairs (distances in angstroms and angles in degrees).

**TABLE 3: Key Properties of Kishi Covalent Base Pairs Including the B3LYP/6-31+G(d,p) and B3LYP/6-31G Electronic Energies (in hartree), ZPVE (in kcal/mol), and N-H Stretching Vibrational Modes Calculated at the Latter Computational Level**

properties	K-CP1	K-CP2	K-CP3	K-CP4
$-\text{energy}_{\text{B3LYP/6-31+G(d,p)}}$	0.53454	0.53471	0.54175	0.53460
+ 927				
$-\text{energy}_{\text{B3LYP/6-31G}}$	0.19947	0.19838	0.20805	0.19906
+ 927				
ZPVE + 164	0.55	0.37	1.04	0.56
rotational constants				
A	617.21	692.52	619.30	696.62
B	218.28	209.04	218.32	207.45
C	176.63	178.09	176.46	177.32
dipole moment	5.6	6.4	5.0	6.6
Q	105.2	105.1	110.8	110.7
a	167.0	165.5	166.6	165.1
$\nu_{\text{N}_1-\text{H}^{\cdots}\text{O}}$	3516 (375)	3502 (396)	3525 (297)	3534 (290)
$\nu_{\text{N}_2-\text{H}}$	3691 (46)	3691 (35)	3690 (56)	3692 (49)
$\nu_{\text{N}_3-\text{H}}$	3654 (92)	3651 (94)	3652 (91)	3653 (93)

<sup>a</sup>Frequencies in  $\text{cm}^{-1}$ ; IR intensities in  $\text{km/mol}$  are given in parentheses. Rotational constants are indicated in MHz, total dipole moment in debye, mean dipole polarizability  $\alpha = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$  in atomic units, mean quadrupole moment  $Q = (Q_{xx} + Q_{yy} + Q_{zz})/3$  in  $\text{D}\cdot\text{\AA}$ . Numbering of atoms is shown in Figure 3.

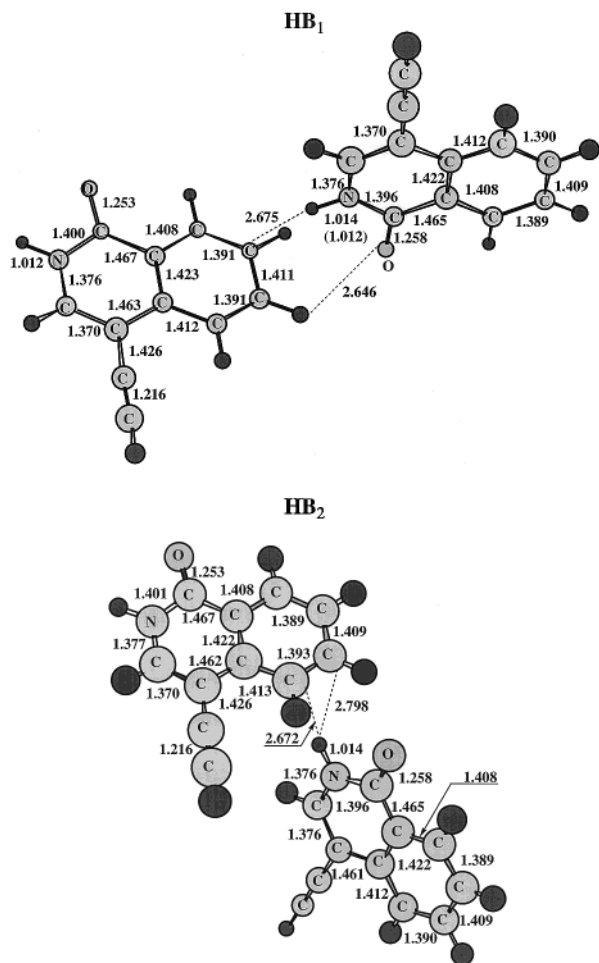
consisting of K-CP1 of the WC type and K-CP2 of the reverse WC type, and the other two pairs of class II are K-CP3 of the WC type and K-CP4 of the reverse WC type. Table 3 collects their theoretical data, including the NH stretching frequencies. First of all, it is worth noticing that all these pairs are nonplanar, as already mentioned in ref 10a. They are nearly iso-energetic, although K-CP3 is slightly favored over the others: by 4.5 (4.0) kcal/mol over K-CP1 and K-CP4 and 4.4 (3.8) kcal/mol over K-CP2 (after ZPVE correction evaluated at the B3LYP/6-31G computational level). K-CP3 has the C' to C' distance R of 9.052 Å and glycosyl bond angles  $\alpha_1 = 56.2^\circ$  and  $\alpha_2 = 61.1^\circ$ . This implies that the distortion of the DNA double helix motif,

caused by its insertion, is not so substantial. However, K-CP1, which is less energetically favorable covalent base pair, better accommodates the DNA double helix because  $R = 9.215 \text{ \AA}$ ,  $\alpha_1 = 49.6^\circ$  and  $\alpha_2 = 58.1^\circ$ . This value of R rather well correlates with that provided by the X-ray structural analysis.<sup>10a</sup>

Kishi-type covalent base pairs K-CP1 and K-CP2 have a quite interesting feature distinguishing them from Gao-Orgel pair. This is actually their ability to form the reverse WC base pairing which is energetically equivalent to the normal one. One of these reverse pairs, namely, K-CP2, has the smallest hydrogen bond length of 1.933 Å among all pairs K-CP<sub>n</sub> ( $n = 1 - 4$ ) and correspondingly, the lowest stretching mode  $\nu_{\text{N-H}\cdots\text{O}}$  stretch peaked at  $3502 \text{ cm}^{-1}$  that, in overall, may lead to a feasible formation of its tautomeric mispairing counterpart.

#### 4. Summary

Summarizing the presented theoretical results, we conclude that the amide covalently bonded base pair GO-CP1 appears to be more stable compared to the formyl one. However, due to a larger dipole moment of GO-CP1, polar solvents like, e.g., water, favor it by a stronger stabilization. We also reveal that GO-CP1 is much more strongly stabilized by a polar environment and likely favors a stacking due to a larger mean polarizability and total dipole moment, compared, for instance, to the A·U base pair. Its C' to C' interglycosidic distance R of 9.98 Å and glycosyl bond angle  $\alpha_2 = 59.0^\circ$  almost coincide with the corresponding calculated values for the A·U pair. An agreement in glycosyl bond angle  $\alpha_1 = 44.0^\circ$ , indicating a slight shift of the dealdehyde toward the minor groove, is less satisfactory. However, all these values are within the range of the polymerase clamp<sup>2,12,14</sup> and, therefore, GO-CP1 may be incorporated into the double helix without causing major perturbations. Also, the calculated vibrational frequencies of the covalently bonded base pairs can be used as their "fingerprints" in further experimental studies. Altogether, the present work indeed demonstrates a use-



**Figure 4.** The B3LYP/6-31G hydrophobic base pairs. Distances are given in angstroms. The monomeric N–H bond length is indicated in parentheses.

ful consistency of Gao–Orgel and Kishi et al. experiments and high-level computational approach toward a further exploring of the concept of covalently bonded base pairing in the DNA double helix.

Finally, it would be worth mentioning the hydrophobic analogue of the base pair which has recently been proposed by Schultz, Romesberg, and co-workers.<sup>12</sup> Figure 4 displays two B3LYP/6-31G optimized structures, HB<sub>1</sub> and HB<sub>2</sub>, of the 7-ethynyl iscarbostyryl base pair (a lower analogue of the 7-propynyl iscarbostyryl base pair treated in ref 12), excluding a very weak and planar van der Waals pair, characterized by the large distance  $R = 11.586 \text{ \AA}$  and by a head-to-head arrangement of the C–C≡C–H bonds. The energies of the pair formation of HB<sub>1</sub> and HB<sub>2</sub> amount to 4.1 and 4.0 kcal/mol, respectively, after ZPVE correction, likely attributable to the  $\pi$  hydrogen bonding N–H $\cdots\pi$  and two weak C–H $\cdots$ O bonds (such a large energy of complexation is not consistent with the suggested hydrophobic character). The formation of this complexation results in a red shift by  $38 \text{ cm}^{-1}$  of the corresponding  $\nu_{\text{NH}}$  stretching mode. Other theoretical properties of the reported hydrophobic base pairs are listed in Table 4. These structures are actually inconsistent with the experimentally designed pair.<sup>12</sup> We suggest that such an inconsistency in the retention of the typical DNA base pair configuration may arise due to a neglect of solvent effects, in particular with an aqueous environment, inherent for the DNA double helix, and resolving this inconsistency would indeed be an interesting topic for further study, on both experimental and theoretical sides.

**TABLE 4: Key Properties of the Hydrophobic Base Pairs Including the B3LYP/6–31G Electronic Energies (in hartree), ZPVE (in kcal/mol), and N–H Stretching Vibrational Modes**

properties	HB <sub>1</sub>	HB <sub>2</sub>
-energy + 1106	−0.36549	−0.36550
ZPVE + 191	0.20	0.25
rotational constants		
A	387.7332	388.4257
B	77.5347	88.6447
C	75.4575	84.9759
dipole moment	4.2	4.7
$\alpha$	247.4	245.2 19

Frequencies in  $\text{cm}^{-1}$ ; IR intensities in  $\text{km/mol}$  are given in parentheses. Rotational constants are indicated in MHz, total dipole moment in D, mean dipole polarizability  $\alpha = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$  in atomic units.

**Acknowledgment.** The authors gratefully thank the Research Council of the University of Leuven (GOA program).

## References and Notes

- (1) (a) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737 (reprinted in *Ann. N. Y. Acad. Sci.* **1995**, *758*, 737). (b) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 964. (c) Crick, F. H. C.; Watson, J. D. *Proc. R. Soc. A* **1954**, *223*, 80. (d) Watson, J. D. *The Double Helix*; Penguin Books: New York, 1968. (e) Crick, F. H. C. *What Mad Pursuit*; Basic Books: New York, 1988. (f) Watson, J. D.; Crick, F. H. C. *J. Am. Med. Assoc.* **1993**, *269*, 1966, 1967.
- (2) (a) Saenger, W. *Principles of Nucleic Acid Structure*; Springer: Berlin, 1984. (b) Sinden, R. R. *DNA Structure and Function*; Academic: San Diego, 1994. (c) *Nucleic Acids in Chemistry and Biology*, Blackburn, G. M., Gait, M. J., Eds.; Oxford University Press: Oxford, 1996; p 21. (d) Kornberg, A.; Baker, T. A. *DNA Replication*; Freeman: New York, 1992. (e) Leontis, N. B.; Westhof, E. *Q. Rev. Biophys.* **1998**, *31*, 399. (f) Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1.
- (3) Schuster, P.; Wolschann, P. *Monatsh. Chem.* **1999**, *130*, 947.
- (4) (a) Drake, J. W. *The Molecular Basis of Mutation*; Holden-Day: San Francisco, 1970. (b) Auerbach, C. *Mutations Research. Problems, Results and Perspectives*; Chapman and Hall: London, 1976. (c) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.
- (5) (a) Lawley, P. D.; Brookes, P. *Nature* **1961**, *192*, 1081. (b) Lawley, P. D.; Brookes, P. *J. Mol. Biol.* **1962**, *4*, 216. (c) Sowers, L. C.; Shaw, B. R.; Veigl, M. L.; Sedwick, W. D. *Mutat. Res.* **1987**, *177*, 201. (d) Topal, M. D.; Fresco, J. R. *Nature* **1986**, *263*, 285. (e) Topal, M. D.; Fresco, J. R. *Nature* **1986**, *263*, 289. (f) Danilov, V. I.; Kventsel, G. F. *Electronic Representation in the Point Mutation Theory*; Naukova Dumka: Kiev, 1971. (g) Rein, R.; Shibata, M.; Garduno-Juarez, R.; Kieber-Emmons, T. In *Structure and Dynamics: Nucleic Acids and Proteins*; Clementi, E., Sarma, R. H., Eds.; Adenine Press: New York, 1983; p 269. (h) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149.
- (6) (a) Devadas, B.; Leonard, N. J. *J. Am. Chem. Soc.* **1986**, *108*, 5012. (b) Devadas, B.; Leonard, N. J. *J. Am. Chem. Soc.* **1990**, *112*, 3125. (c) See also Bhat, B.; Leonard, N. J.; Robinson, H.; Wang, A. H.-J. *J. Am. Chem. Soc.* **1996**, *118*, 10744.
- (7) (a) Webb, T. R.; Matteucci, M. D. *J. Am. Chem. Soc.* **1986**, *108*, 2764. (b) Weber, A. L. *Origins of Life and the Evolution of the Biosphere* **1989**, *19*, 179. (c) Ferentz, A. E.; Verdine, G. L. *J. Am. Chem. Soc.* **1991**, *113*, 4000. (d) Cowart, M.; Benkovic, S. J. *Biochemistry* **1991**, *30*, 788. (e) Mitchell, M. A.; Kelly, R. C.; Wicniewski, N. A.; Hatzenbuehler, N. T.; Williams, M. G.; Petzold, G. L.; Slightom, J. L.; Siemieniak, D. R. *J. Am. Chem. Soc.* **1991**, *113*, 8994. (f) Dieter-Wurm, I.; Sabat, M.; Lippert, B. *J. Am. Chem. Soc.* **1992**, *114*, 357. (g) Armstrong, R. W.; Salvati, M. E.; Nguyen, M. *J. Am. Chem. Soc.* **1992**, *114*, 3144. (h) Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I. *Yokoyama, S. Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4922.
- (8) (a) Kirchner, J. J.; Sigurdsson, S. T.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4021. (b) Huang, H.; Solomon, M. S.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 9240. (c) Seaman, F. C.; Hurley, L. *Biochemistry* **1993**, *32*, 12577. (d) Osborne, S. E.; Cain, R. J.; Glick, G. D. *J. Am. Chem. Soc.* **1997**, *119*, 1172.
- (9) Gao, K.; Orgel, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14837.
- (10) (a) Qiao, X.; Kishi, Y. *Angew. Chem.* **1999**, *111*, 977; *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 928. (b) Qiu, Y.-L.; Li, H.-Y.; Topalov, G.; Kishi, Y. *Tetrahedron Lett.* **2000**, *41*, 9425. (c) Li, H.-Y.; Qiu, Y.-L.; Moyroud, E.; Kishi, Y. *Angew. Chem., Int. Ed.* **2001**, *40*, 1471.
- (11) (a) Moran, S.; Rex, X.-F. Ren; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056. (b) Morales, J. C.; Kool, E. T. *Nature Struct.*

*Biol.* **1998**, *5*, 950. (c) Guckian, K. M.; Krugh, T. K.; Kool, E. T. *Nature Struct. Biol.* **1998**, *5*, 954. (d) Barsky, D.; Kool, E. T.; Colvin, M. E. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1119. (e) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 990. (f) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 6841 and references therein. Also see: Meyer, M.; Silhnel, J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 619.

(12) (a) McMin, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. C.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585. (b) Ogawa, A. K.; Wu, Y.; McMin, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274.

(13) (a) Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493. (b) Rawls, R. *Chem. Eng. News* **1997**, March 3, 9. (c) Diederichsen, U. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 1655.

(14) (a) Lutz, M. J.; Held, H. A.; Hottinger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308. (b) Kool, E. T. *Biopolymers* **1998**, *48*, 3.

(15) Watson, J. D.; Hopkins, N. H.; Roberts, J. W.; Steitz, J. A.; Weiner, A. M. *Molecular Biology of the Gene*; Benjamin/Cummings: Menlo Park, 1987.

(16) (a) Dickerson, R. E. *Methods Enzymol.* **1992**, *211*, 67. (b) Voet, A.; Rich, A. *Prog. Nucleic Acids Res. Mol. Biol.* **1970**, *10*, 183. (c) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145.

(17) All computations of the Gao–Orgel and Kishi covalent base pairs were performed at the density-functional exchange-correlation B3LYP computational level with 6-31+G(d, p) basis set using GAUSSIAN 98 suit of packages.<sup>18</sup> The geometries of all covalent base pairs considered in the present work were fully optimized without any constraint on a possible

planarity. The harmonic vibrations were also calculated in order to locate their minimum energy structure and to analyze further their vibrational spectra and thermodynamic properties. The hydrophobic base pairs were studied via the B3LYP/6-31G method.

(18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(19) Kryachko, E. S.; Nguyen, M. T. *Adv. Quantum Chem.* **2001**, *40*, 79.

(20) (a) Mendelman, L. V.; Boosalis, M. S.; Petruska, J.; Goodman, M. F. *J. Biol. Chem.* **1989**, *264*, 14415.

(21) (a) Lomant, A. J.; Fresco, J. R. In *Progress in Nucleic Acid Research Molecular Biology*; Lomant, A. J., Fresco, J. R., Eds.; Academic: New York, 1975; Vol. 15, p 185. (b) Lindhal, T. *Annu. Rev. Biochem.* **1982**, *51*, 61. (c) Sloane, D. L.; Goodman, M. F.; Echols, H. *Nucleic Acids Res.* **1998**, *26*, 6465. (d) Moran, S.; Ren, R. X.; Sheils, C. J.; Rummey, S.; Kool, E. T. *Nucleic Acids Res.* **1996**, *24*, 2044. (e) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477.