

# A stable antiparallel cytosine-thymine base pair occurring only at the end of a duplex

**Citation for published version (APA):**

Vanhommerig, S. A. M., Genderen, van, M. H. P., & Buck, H. M. (1991). A stable antiparallel cytosine-thymine base pair occurring only at the end of a duplex. *Biopolymers*, 31(9), 1087-1094.  
<https://doi.org/10.1002/bip.360310908>

**DOI:**

[10.1002/bip.360310908](https://doi.org/10.1002/bip.360310908)

**Document status and date:**

Published: 01/01/1991

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# A Stable Antiparallel Cytosine–Thymine Base Pair Occurring Only at the End of a Duplex

SYLVIA A. M. VANHOMMERIG, MARCEL H. P. VAN GENDEREN,\* and HENK M. BUCK

Laboratory of Organic Chemistry, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands

## SYNOPSIS

Ultraviolet hyperchromicity experiments indicate that in DNA duplex formation, a C–T mismatch is destabilizing in the center of a duplex, but behaves as a stable base pair at the terminus of a duplex. The C–T base pair is thought to contain two hydrogen bonds, but has thermodynamic parameters ( $\Delta H^\circ$  and  $\Delta G^\circ$  of dissociation) that are similar to a G–C base pair. AMBER molecular mechanics calculations were performed to study the possible structural properties of DNA duplexes with central and terminal C–T combinations. These calculations also indicate that a central C–T pair destabilizes a duplex, while terminal C–T forms a stable base pair. Hydrogen bonding between cytosine and thymine occurs only in the energy-minimized structures when the helix diameter decreases and the propeller twist angle between the bases increases. These changes are found to occur only at the end of a duplex in the calculations, which may explain the experimental results.

## INTRODUCTION

The hybridization of DNA strands to form an antiparallel duplex usually follows the familiar Watson and Crick base pairing of adenine (A) with thymine (T), and guanine (G) with cytosine (C).<sup>1</sup> In this scheme, mismatched base pairs destabilize the duplex, or can even prohibit hybridization of the two DNA strands. However, several instances have been observed where noncanonical base pairs allow duplex formation. Among these are several wobble base pairs (e.g., G–T, A–C),<sup>1</sup> or base pairs formed after protonation of a ring nitrogen, thus changing it from a hydrogen-bond acceptor into a hydrogen-bond donor (e.g.,  $\text{CH}^+ - \text{C}$ ,<sup>2,3</sup>  $\text{AH}^+ - \text{A}^1$ ). Furthermore, in duplexes with parallel backbones, pyrimidine–pyrimidine base pairs have been found when the negative phosphate charges in the backbone are shielded by cationic peptide complexation<sup>4</sup> or methylation.<sup>5,6</sup> This allows the close proximity between the DNA strands that is necessary for such a hybridization.

In this paper, we wish to report uv hyperchromicity studies that point at the possibility of an an-

tiparallel cytosine–thymine base pair, which is formed only at the end of a duplex. AMBER molecular mechanics calculations are performed to offer a possible structural explanation for this behavior.

## MATERIALS AND METHODS

### Oligodeoxynucleotides

The DNA hexadecamers were all synthesized on an Applied Biosystems 381A DNA synthesizer with the standard  $\beta$ -cyanoethyl phosphoramidite protocol. After removal from the column and deprotection of the bases, the oligomers were purified by ethanol precipitation. DNA concentrations were determined in aqueous solution by uv absorption at 260 nm, assuming a single-strand concentration of 37  $\mu\text{g}/\text{mL}$  to give an  $A_{260}$  of 1 for a light path of 1 cm.<sup>7</sup>

### UV Measurements

All variable-temperature uv measurements were performed on a Perkin-Elmer 124 spectrophotometer in 10-mm quartz cuvettes at a wavelength of 260 nm. Total DNA single-strand concentrations ( $C_T$ ) of 2  $\mu\text{M}$  were used in a 2 mM Tris HCl/0.1

Biopolymers, Vol. 31, 1087–1094 (1991)

© 1991 John Wiley & Sons, Inc.

CCC 0006-3525/91/091087-08\$04.00

\* To whom correspondence should be addressed.

mM EDTA buffer solution (pH 7.0) with 0.1M NaCl. Samples of the duplexes were annealed for 2 min at 70°C, followed by slow cooling (45 min) to room temperature. Absorbance at 260 nm was then followed as a function of sample temperature, with a temperature gradient of 1°C/min. The uv hyperchromicity data were computer fitted to obtain melting temperatures ( $T_m$ ) and dissociation enthalpies ( $\Delta H^\circ$ ). For this, we used the following formula for the fraction  $f$  of single-stranded DNA<sup>4</sup>:

$$f = \frac{2}{1 + \{1 + 8 \cdot e((\Delta H^\circ/R)[(1/T) - (1/T_m)])\}^{1/2}}$$

The dissociation entropy ( $\Delta S^\circ$ ) and free energy ( $\Delta G^\circ$ ) were then determined with the formulae<sup>8</sup>:

$$\Delta S^\circ = \frac{\Delta H^\circ}{T_m} + R \ln \frac{C_T}{4}$$

$$\Delta G^\circ = \Delta H^\circ - T^\circ \Delta S^\circ \quad (T^\circ = 298.15 \text{ K})$$

$T_m$  values and  $\Delta H^\circ$  values were determined with an accuracy of 1°C and 0.5 kcal/mol, respectively. It then follows that the  $\Delta S^\circ$  values have an accuracy of 2 cal/mol·K, and the  $\Delta G^\circ$  values of 1 kcal/mol.

### AMBER Molecular Mechanics Calculations

Structural calculations on the DNA duplexes were performed with the AMBER program,<sup>9</sup> which uses a molecular mechanics force field to calculate the energy content of a molecule. This force field includes bond stretching, bond-angle deformation, torsional potentials, van der Waals and electrostatic interactions, and a specific hydrogen-bond energy term:

$$\begin{aligned} E_{\text{tot}} = & \sum_{\text{bonds}} K_R (R - R_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 \\ & + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ & + \sum_{i < j} \left( \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right) \\ & + \sum_{\text{Hbonds}} \left( \frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right) \end{aligned}$$

The energy parameters ( $K_R$ ,  $K_\theta$ ,  $V_n$ ,  $A_{ij}$ ,  $B_{ij}$ ,  $q_i$ ,  $C_{ij}$ , and  $D_{ij}$ ), and equilibrium values ( $R_{\text{eq}}$ ,  $\theta_{\text{eq}}$ , and  $\gamma$ ) were taken as described before.<sup>9</sup> The united-atom approach was used, and the dielectric constant was

chosen as  $\epsilon = R_{ij}$  to simulate the long-range shielding of electrostatic interactions in aqueous solution. All input structures were generated with standard B-DNA coordinates<sup>10</sup> and energy minimized until the gradient was smaller than 0.1 kcal/mol·Å. As we have found earlier,<sup>11</sup> the total interstrand energy ( $E_{\text{int}}$ ) is a better indicator of duplex stability than the total energy of the duplex ( $E_{\text{tot}}$ ), which varies with the number of atoms in the system.

## RESULTS AND DISCUSSION

### Duplex Stability

The parent duplex **1** consisted of the 16-mer d(CAT·GAA·TCC·TAG·CAG·T) and its complementary d(A·CTG·CTA·GGA·TTC·ATG). It was found to have a  $T_m$  of 47°C in 0.1M NaCl (see Table I), a  $\Delta H^\circ$  of 115 kcal/mol, and a  $\Delta G^\circ$  of 16.3 kcal/mol. This is in good agreement with the theoretical values as predicted with the method of Breslauer et al.<sup>12</sup> They describe the dissociation enthalpy and free energy as the sum of all dinucleotide contributions in the duplex, with an initiation free energy added. Using the Breslauer theory with the more recent numerical values of Quartin and Wetmur<sup>13</sup> for the dinucleotide contributions and initiation term, the following dissociation enthalpy and free energy are calculated for the completely complementary duplex **1** at 25°C, in a 1M NaCl solution of pH 7:  $\Delta H^\circ = 112.9$  kcal/mol,  $\Delta G^\circ = 19.2$  kcal/mol. Adjusting for the 0.1M NaCl salt concentration according to Quartin and Wetmur<sup>13</sup> involves an additional free energy term, given by

$$\Delta G_s^\circ = 0.26N \cdot RT^\circ \ln \left( \frac{1.3C_{\text{NaCl}}}{C_{\text{NaCl}} + 0.3} \right)$$

which amounts to -2.6 kcal/mol for  $C_{\text{NaCl}} = 0.1M$  and  $N = 15$  phosphates. Thus,  $\Delta G^\circ$  is expected to be 16.6 kcal/mol. Since  $\Delta S^\circ = (\Delta H^\circ - \Delta G^\circ)/T^\circ = 323.0$  cal/mol·K, it follows with a DNA concentration  $C_T = 2 \mu M$ , that  $T_m = \Delta H^\circ / (\Delta S^\circ - R \times \ln(C_T/4)) = 47.7^\circ C$ .

Duplex **2** consisted of the 16-mer d(CAT·GAA·TCC·TAG·CAG·T) with the mismatched counterpart d(A·CTG·CTA·GTA·TTC·ATG), which results in a C-T combination at position 8 (numbering from 5' to 3' in the first oligomer). This mismatch indeed led to a lower  $T_m$  value of 37°C, a decreased  $\Delta H^\circ$  of 109 kcal/mol, and a  $\Delta G^\circ$  of 12.4 kcal/mol (see Table I). This decrease in stability is quite large, when compared with the effect of

**Table I** Thermodynamic Parameters of the Duplex–Coil Equilibrium as Determined for the DNA Duplexes 1–3 by UV Hyperchromicity Measurements

System	Sequence <sup>a</sup>	$T_m$ (°C)	$\Delta H^\circ$ (kcal/mol)	$\Delta S^\circ$ (cal/mol · K)	$\Delta G^\circ$ (kcal/mol)
<b>1</b>	5'-d(CAT · GAA · TCC · TAG · CAG · T)-3' 3'-d(GTA · CTT · AGG · ATC · GTC · A)-5'	47	115	331	16.3
<b>2</b>	5'-d(CAT · GAA · TCC · TAG · CAG · T)-3' 3'-d(GTA · CTT · ATG · ATC · GTC · A)-5'	37	109	324	12.4
<b>3</b>	5'-d(CAT · GAA · TCC · TAG · CAG · T)-3' 3'-d(GTA · CTT · AGG · ATC · GTC · C)-5'	50	120	343	17.7

<sup>a</sup> Base-pair numbering is from left to right. Mismatches are indicated by bold letters.

changing the CG base pair at position 8 into an AT or TA base pair (see Table II).

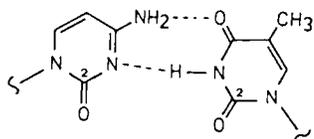
In such cases, the  $T_m$  is lowered only a few degrees and the  $\Delta G^\circ$  only by 1 kcal/mol. Therefore, we conclude that no base pairing is present for the C–T combination. The change in thermodynamic parameters can be evaluated with the formalism of dinucleotide contributions. The loss of the TC/AG and CC/GG dinucleotides around position 8 amounts to 16.6 kcal/mol in the enthalpy according to the Breslauer theory.<sup>13</sup> The overall decrease of only 6 kcal/mol in  $\Delta H^\circ$  indicates that the new TC/AT and CC/TG dinucleotides still yield an enthalpy contribution of 10.6 kcal/mol. However, most of this is probably due to hydrogen bonding in the canonical A–T and G–C base pairs in the two dinucleotides, and base stacking. For the free energy of dissociation, the loss of the TC/AG and CC/GG contributions should amount to 3.45 kcal/mol,<sup>13</sup> which almost equals the experimental decrease of 3.9 kcal/mol. The new TC/AT and CC/TG combinations have a slightly destabilizing effect on  $\Delta G^\circ$  of 0.45 kcal/mol, confirming the conclusion that no C–T base pair is present.

Duplex **3** combined the 16-mer d(CAT · GAA · TCC · TAG · CAG · T) with the mismatched

counterpart d(C · CTG · CTA · GGA · TTC · ATG), giving a terminal C–T combination at position 16. Now, a  $T_m$  value of 50°C was found, which is even slightly higher than the parent duplex **1** under identical conditions (see Table I). In addition, the  $\Delta H^\circ$  value was 120 kcal/mol, and the  $\Delta G^\circ$  17.7 kcal/mol, indicating a high intrinsic duplex stability. In comparison, a CG or GC base pair at position 16 would also have given a  $T_m$  of 50 or 51°C, and thermodynamic parameters in the same range as for **3** (see Table II). Complete removal of the terminal base pair results in a decreased duplex stability ( $T_m = 44.9^\circ\text{C}$ , see Table II), despite the higher GC content of the remaining 15-mer duplex. However, even the presence of unpaired dangling bases at the end of a duplex can raise the stability through stacking interactions. The work of Sugimoto et al.,<sup>14</sup> although it concerns RNA duplexes, allows one to estimate the free energy term involved. Corrected for a temperature effect (reference temperature of 37°C for the RNA systems instead of 25°C for DNA systems), the effect of a 5'-dangling C is negligible, while a 3'-dangling U contributes ca. 0.7 kcal/mol to the  $\Delta G^\circ$ . Obviously, this cannot account for the 2.5 kcal/mol difference between the experimental  $\Delta G^\circ$  of **3**, and the calculated  $\Delta G^\circ$  of **4** where no

**Table II** Calculated Thermodynamic Parameters for the Duplex–Coil Equilibrium of Reference Duplexes

System	Change from <b>1</b>	$T_m$ (°C)	$\Delta H^\circ$ (kcal/mol)	$\Delta S^\circ$ (cal/mol · K)	$\Delta G^\circ$ (kcal/mol)
<b>2a</b>	Central CG → AT	44.9	108.8	313.3	15.4
<b>2b</b>	Central CG → TA	45.9	111.0	319.1	15.9
<b>3a</b>	Terminal TA → GC	49.6	117.4	355.1	17.5
<b>3b</b>	Terminal TA → CG	50.8	117.5	334.1	17.9
<b>4</b>	Terminal TA deleted	44.9	106.4	305.9	15.2



**Figure 1.** Hydrogen-bonding scheme in an antiparallel C-T base pair. Note that the carbonyl oxygens  $O_2$  must come into close contact when the base pair is formed.

effect of the terminal C-T combination is present. Sugimoto et al. have also examined the effect of terminal mismatches in RNA duplexes.<sup>15</sup> Again corrected for the reference temperature, a C-U mismatch contributes 0.6 kcal/mol to the  $\Delta G^\circ$ . It can be seen from this that even two dangling bases cannot explain the strong increase in duplex stability for our DNA system **3**. The remaining free energy difference (ca. 2 kcal/mol) when the dangling bases are accounted for would fit nicely with two hydrogen bonds according to Sugimoto et al.,<sup>15</sup> indicating that C-T base pairing indeed occurs in system **3**. In a Breslauer-type analysis of the thermodynamic parameters, a GT/CA dinucleotide is replaced by a GT/CC dinucleotide. For the  $\Delta H^\circ$  this should entail a loss of 6.5 kcal/mol,<sup>13</sup> so the net increase from 115 to 120 kcal/mol indicates an enthalpy contribution of 11.5 kcal/mol for the new dinucleotide. The  $\Delta G^\circ$  increased from 13.6 to 17.7 kcal/mol, which includes a loss of 1.4 kcal/mol for the GT/CA dimer.<sup>13</sup> Therefore, the new GT/CC pair contributes 2.8 kcal/mol to the free energy of the duplex. Both the  $\Delta G^\circ$  and the  $\Delta H^\circ$  are in the same range as for a GC/CG dinucleotide. It must be concluded that the terminal C-T combination significantly increases the duplex stability, and can be seen

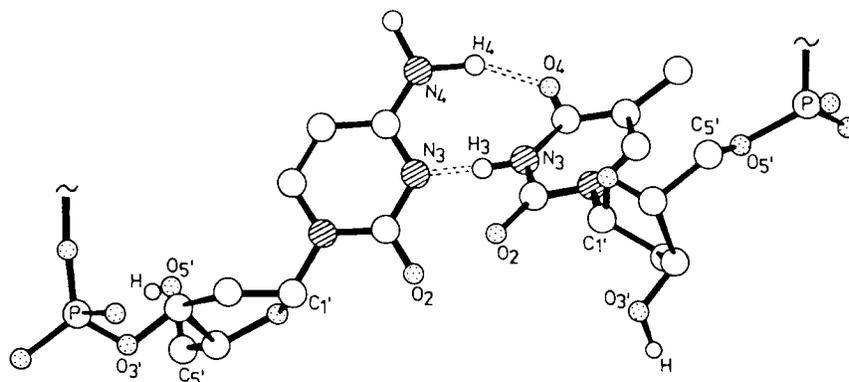
as a C-T base pair. In nonempirical calculations of base-pair formation, the antiparallel C-T base pair was not found to be very stable.<sup>16</sup> The interaction energy between a cytosine and a thymine base was 14.3 kcal/mol, as compared to an A-T base pair (16 kcal/mol) or a G-C base pair (26.5 kcal/mol). However, these calculations did not include base-stacking interactions, which we have found to be very important for pyrimidine-pyrimidine base pairs.<sup>6</sup> Also, the interaction energies cannot be compared directly to the measured  $\Delta H^\circ$  values, since phosphate-phosphate repulsions were not included.

A stable antiparallel C-T base pair seems only possible as depicted in Figure 1, with an  $N_3 \cdots H-N_3$  hydrogen bond and an  $N_4-H \cdots O_4$  hydrogen bond,<sup>16</sup> both of which are well known in base-pair interactions.

Two obstacles would appear to be present, however, for the formation of such a C-T base pair. First, the bases are both pyrimidines, which must approach each other more closely (e.g., in terms of  $C_1'-C_1'$  distance) than a standard purine-pyrimidine base pair, to form stable hydrogen bonds. Second, even when the hydrogen bonds are formed, the carbonyl oxygens  $O_2$  come into very close steric and electronic contact, thus disfavoring the base-pair formation. In order to investigate the possibility of C-T base-pair formation in the light of these problems, we performed structural calculations on the duplexes **1-3**.

### Structural Calculations

Molecular mechanics calculations with the AMBER program were performed to get an indication of the relative stability of the systems, and to examine



**Figure 2.** Structure of the terminal C-T base pair in duplex **3**, as obtained by energy-minimization with the AMBER program. Nitrogen atoms are shaded, and oxygen atoms are dotted.

**Table III Energy Contributions for the Minimized Structures of the Duplexes 1–3 as Determined with the AMBER Program**

Energy Term <sup>a</sup>	1 (No Mismatch)	2 (Central C-T)	3 (Terminal C-T)
$E_{\text{tot}}$	-1583.3	-1590.1	-1578.1
$E_{\text{bond}}$	8.2	8.8	8.3
$E_{\text{angle}}$	139.7	138.3	145.5
$E_{\text{dihed}}$	458.3	457.1	450.9
$E_{\text{vdW,nb}}$	-559.9	-557.3	-558.7
$E_{\text{el,nb}}$	1129.9	1147.0	1151.3
$E_{\text{vdW,14}}$	149.6	153.0	148.8
$E_{\text{el,14}}$	-2904.9	-2933.1	-2919.4
$E_{\text{Hbond}}$	-4.2	-3.9	-4.8
$E_{\text{int}}$	-230.1	-212.0	-229.3

<sup>a</sup> Energy terms are in kcal/mol, and denote, respectively, total energy, bond stretching energy, angle deformation energy, torsional energy, nonbonded van der Waals energy, nonbonded electrostatic energy, vicinal van der Waals energy, vicinal electrostatic energy, hydrogen-bond energy, and interstrand energy.

possible conformational explanations for the different behavior of central and terminal C–T combinations. The energy-refined structures did not differ strongly in total energy, and also had a similar decomposition of the total energy into the several energy terms involved (see Table III).

The interstrand interaction energies on the other hand, were strikingly different, and indicate a far lower duplex stability for duplex **2** with the central C–T mismatch. Systems **1** and **3** were equally stable, which is in good agreement with the experimentally determined duplex stabilities. It therefore seems that

**Table IV Average Conformations of the Energy-Minimized Duplexes 1–3**

Conformation <sup>a</sup>	1 (No Mismatch)	2 (Central C-T)	3 (Terminal C-T)
$\alpha$ (P—O <sub>5'</sub> )	-67.2°	-67.0°	-67.7°
$\beta$ (C <sub>5'</sub> —O <sub>5'</sub> )	177.6°	178.2°	177.4°
$\gamma$ (C <sub>4'</sub> —C <sub>5'</sub> )	58.7°	58.5°	59.0°
$\delta$ (C <sub>3'</sub> —C <sub>4'</sub> )	129.3°	130.4° <sup>b</sup>	130.1° <sup>c</sup>
$\epsilon$ (C <sub>3'</sub> —O <sub>3'</sub> )	-177.6°	-178.4°	-177.2°
$\zeta$ (P—O <sub>3'</sub> )	-99.8°	-99.3°	-100.0°
$\chi$ (C <sub>1'</sub> —N)	-123.2°	-123.2°	-123.4°
P	144.3°	146.3° <sup>b</sup>	145.2° <sup>c</sup>
$\nu_{\text{max}}$	32.8°	32.8°	33.8°
Twist	7.0°	8.0°	8.4°
hra	36.5°	36.4°	36.8°
Diameter <sup>d</sup>	18.1 Å	18.0 Å <sup>e</sup>	18.1 Å <sup>f</sup>

<sup>a</sup> Nomenclature according to IUPAC-IUB recommendations,<sup>17</sup> and Altona and Sundaralingam.<sup>18</sup> The hra and twist denote the helical repeat angle and the propeller twist angle, respectively.

<sup>b</sup> Excluding C<sub>8</sub>.

<sup>c</sup> Excluding C<sub>16</sub>.

<sup>d</sup> Diametrical phosphorus–phosphorus distance within a base pair.

<sup>e</sup> Excluding base pair 8.

<sup>f</sup> Excluding base pair 16.

**Table V Conformational Characteristics for Base Pair 8 in the Energy-Minimized Duplexes 1–3**

Conformation <sup>a</sup>	<b>1</b> (No Mismatch)		<b>2</b> (Central C-T)		<b>3</b> (Terminal C-T)	
	C	G	C	T	C	G
$\alpha$	-66.0°	-66.7°	-70.2°	-66.6°	-67.7°	-67.9°
$\beta$	178.1°	179.3°	169.8°	179.1°	-178.3°	178.1°
$\gamma$	58.9°	58.1°	57.1°	57.8°	59.8°	59.3°
$\delta$	136.4°	136.7°	79.7°	117.0°	140.4°	138.7°
$\epsilon$	-178.3°	-177.9°	-177.5°	177.9°	-178.0°	-176.8°
$\zeta$	-111.6°	-111.6°	-78.8°	-100.3°	-117.1°	-113.6°
$\chi$	-117.9°	-121.8°	-148.6°	-134.9°	-116.8°	-122.9°
P	153.7°	154.3°	32.2°	124.1°	158.8°	156.7°
$\nu_{\max}$	33.8°	34.9°	39.0°	31.8°	34.9°	34.9°
Twist		6.3°		11.0°		4.2°
hra		38.5°		37.8°		40.1°
Diameter		17.8 Å		17.4 Å		17.7 Å

<sup>a</sup> See notes for Table IV.

the AMBER calculations give a reliable picture of the duplexes 1–3. It should be noted that any intrastrand stacking interactions from the C–T combination are not included in  $E_{\text{int}}$ . However, we have seen that this contribution is minor (0.6 kcal/mol) compared to the hydrogen bonding (ca. 2 kcal/mol).

The average conformational features of the three energy-minimized duplexes are given in Table IV. All systems showed standard B-DNA conformations of the backbone, sugar ring, and base (excepting the mismatched cytosines, *vide infra*) and resembled each other strongly.

A more detailed look at the base pairs 8 and 16 in the three duplexes is presented in Tables V and VI. As can be seen, the unchanged base pairs (16 in duplex 2 and 8 in duplex 3) did not markedly react to the introduced mismatches (8 in duplex 2 and 16 in duplex 3).

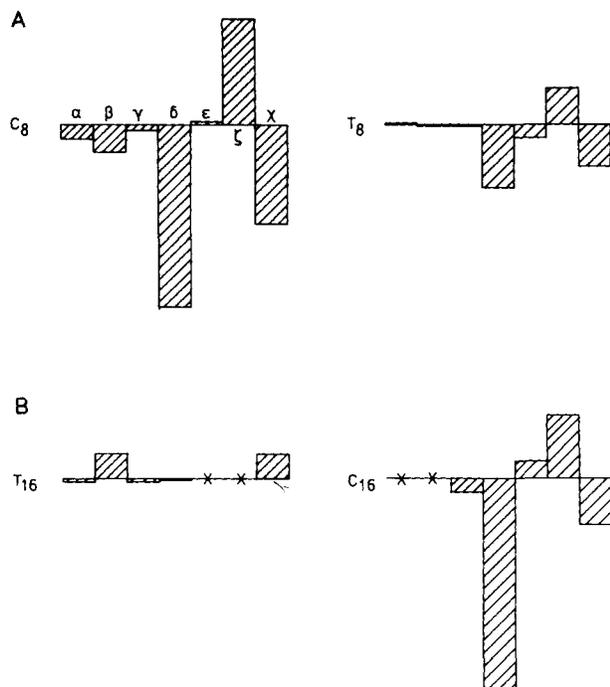
In the C–T combinations (see Figure 3), changes were seen in these calculations primarily for the sugar pucker (torsion angle  $\delta$ ), the P–O<sub>5'</sub> torsion angle  $\zeta$ , and the base orientation (torsion angle  $\chi$ ).

These changes were most evident for the cytidine nucleotide in both duplexes. The cytidine sugar rings

**Table VI Conformational Characteristics for Base Pair 16 in the Energy-Minimized Duplexes 1–3**

Conformation <sup>a</sup>	<b>1</b> (No Mismatch)		<b>2</b> (Central C-T)		<b>3</b> (Terminal C-T)	
	T	A	T	A	T	C
$\alpha$	-67.2°	—	-66.6°	—	-68.3°	—
$\beta$	-179.1°	—	-178.3°	—	-171.4°	—
$\gamma$	58.7°	59.6°	58.4°	59.6°	57.7°	55.4°
$\delta$	148.1°	145.0°	147.3°	144.8°	147.7°	79.3°
$\epsilon$	—	178.8°	—	178.6°	—	-176.1°
$\zeta$	—	-96.8°	—	-94.4°	—	-77.4°
$\chi$	-109.5°	-104.6°	-110.5°	-104.8°	-117.2°	-118.8°
P	164.4°	173.1°	163.5°	174.3°	164.3°	46.8°
$\nu_{\max}$	37.9°	32.8°	37.9°	31.8°	40.0°	40.0°
Twist		15.3°		15.0°		35.6°
hra		35.3°		34.8°		39.0°
Diameter		17.9 Å		18.0 Å		16.6 Å

<sup>a</sup> See notes for Table IV.



**Figure 3.** Calculated deviations of the torsion angles in the central C-T combination of duplex **2** (A), and the terminal C-T base pair of duplex **3** (B), from the base pairs 8 and 16, respectively, in the fully complementary duplex **1**. Crosses indicate the absence of the torsion angle as defined.<sup>17</sup>

changed their usual  $C_{2'}$ -endo or South conformation (pucker phase  $P \sim 160^\circ$ ) toward the  $C_{3'}$ -endo or North conformation ( $P \sim 20^\circ$ ). In duplex **2**, the thymidine nucleotide in the C-T combination also displayed a shift of sugar pucker, toward  $C_{1'}$ -exo ( $P \sim 130^\circ$ ). The pucker amplitude  $\nu_{\max}$  strongly increased for the cytidine residues in the C-T combinations of both duplexes.

The calculated changes in sugar conformation resulted in a decrease of the  $C_{1'}$ - $C_{1'}$  distance in the C-T mismatches, which is essential for base pair formation. This decrease was small for duplex **2** (from 10.9 to 9.6 Å), and only one hydrogen bond was present:  $N_4-H \cdots O_4$  (minimized  $H \cdots O$  distance 1.84 Å). The  $N_3 \cdots H-N_3$  hydrogen bond was not formed (minimized  $N \cdots H$  distance 2.59 Å). In duplex **3**, the  $C_{1'}$ - $C_{1'}$  distance decreased more strongly (from 10.9 to 8.4 Å), resulting in the formation of the two hydrogen bonds  $N_4-H \cdots O_4$  (minimized  $H \cdots O$  distance 1.86 Å) and  $N_3 \cdots H-N_3$ — (minimized  $N \cdots H$  distance 1.82 Å). That two hydrogen bonds are formed is in nice agreement with the experimental difference of  $\Delta G^\circ$  for **3** and **4** (vide supra). The helix diameter (measured as phosphate-phosphate distance across a base pair)

decreased only slightly (0.4 Å) in the C-T combination of duplex **2**, whereas in duplex **3** a change of 1.3 Å was seen for the C-T base pair. In duplex **2**, the decrease in  $C_{1'}$ - $C_{1'}$  distance due to the changed sugar pucker seems to be largely negated by the strong change in  $\zeta$ , which allows the phosphates to stay far apart.

The calculated changes in the  $C_{1'}$ -N torsion angle  $\chi$  led in both duplexes to an enhanced value for the propeller twist angle between the two pyrimidine bases, which is necessary to escape the steric and electronic clash of the carbonyl oxygens  $O_2$ . However, in duplex **2** the increase was much smaller than in duplex **3**, which again indicates that the cytosine and thymine bases cannot form a stable base pair in the center of a duplex.

Thus, the AMBER calculations suggest a way in which the terminal C-T base pair may be formed in spite of the two above-mentioned difficulties by changing the sugar pucker to decrease the base-base distance, and adapting the base orientation to avoid  $O_2-O_2$  close contacts. However, the central C-T combination lacks the freedom to adapt its conformation in this way. Concomitant with the sugar pucker change, the torsion angle  $\zeta$  changes to counteract the decrease of the helix diameter. Also, the neighboring base pairs prevent too large a change in the propeller twist angle.

## CONCLUDING REMARKS

The present experimental and theoretical findings show that an antiparallel C-T base pair may be formed in a DNA duplex, but only at the end of the system. Only there is the conformational freedom present to allow the small helix diameter and high propeller twist angle that are necessary. Interestingly, the work of Sugimoto et al.<sup>15</sup> shows that a terminal C-U base pair is not formed in RNA duplexes. It would seem reasonable that this is due to the well-known limited conformational freedom of RNA. Since it was already known that thymine may form a wobble pair in DNA duplexes with guanine,<sup>1</sup> it now appears that base pairs can be formed with all three other bases by thymine.

The new C-T base pair may be employed in hybridization probes as a terminal wobble pair, and may be useful for the site-directed mutagenesis technique. In the latter case, a natural DNA primer with a mismatched base is used to start the replication of a single-stranded phage. Thus, a new phage is obtained with a single, well-defined point mutation. Normally, more mismatches cannot be intro-

duced simultaneously, since then the primer would not hybridize efficiently. A primer that leads to a stable C-T base pair would allow an extra mutation (changing A into C, or G into T) to be incorporated in the newly synthesized phage, along with the normally present mismatched base. This would in effect double the mutation rate for the in vitro mutagenesis technique.

Use of the services and facilities of the Dutch CAOS/CAMM Center, under grant numbers SON-11-20-700 and STW-NCH-44.0703, is gratefully acknowledged.

## REFERENCES

1. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
2. Adler, A., Grossman, L. & Fasman, G. D. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 423-430.
3. Guschlbauer, W. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 1441-1448.
4. Van Genderen, M. H. P., Hilbers, M. P. & Buck, H. M. (1990) *Biochemistry* **29**, 7838-7845.
5. Koole, L. H., van Genderen, M. H. P. & Buck, H. M. (1987) *J. Am. Chem. Soc.* **109**, 3916-3921.
6. Quaedflieg, P. J. L. M., Broeders, N. L. H. L., Koole, L. H., van Genderen, M. H. P. & Buck, H. M. (1990) *J. Org. Chem.* **55**, 122-127.
7. Miller, J. H. (1977) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
8. Marky, L. A. & Breslauer, K. J. (1987) *Biopolymers* **26**, 1601-1620.
9. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr. & Weiner, P. K. (1984) *J. Am. Chem. Soc.* **106**, 765-784.
10. Arnott, S., Campbell-Smith, P. & Chandrasekharan, P. (1976) in *CRC Handbook of Biochemistry and Molecular Biology, Nucleic Acids*, Vol. 2, Fasman, G. D., Ed., CRC, Cleveland, OH, pp. 411-422.
11. Van Genderen, M. H. P., Koole, L. H., Aagaard, O. M., van Lare, C. E. J. & Buck, H. M. (1987) *Biopolymers* **26**, 1447-1461.
12. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. A. (1988) *Proc. Natl. Acad. Sci. USA* **83**, 3746-3750.
13. Quartin, R. S. & Wetmur, J. G. (1989) *Biochemistry* **28**, 1040-1047.
14. Sugimoto, N., Kierzek, R. & Turner, D. H. (1987) *Biochemistry* **26**, 4554-4558.
15. Sugimoto, N., Kierzek, R. & Turner, D. H. (1987) *Biochemistry* **26**, 4559-4562.
16. Hobza, P. & Sandorfy, C. (1987) *J. Am. Chem. Soc.* **109**, 1302-1307.
17. IUPAC-IUB Commission on Biochemical Nomenclature (1983) *Eur. J. Biochem.* **131**, 9-15.
18. Altona, C. & Sundaralingam, M. (1972) *J. Am. Chem. Soc.* **94**, 8205-8212.

Received October 31, 1990

Accepted April 28, 1991