Homopairing Possibilities of the DNA Bases Cytosine and Guanine: An ab Initio DFT Study

R. E. A. Kelly,*,† Y. J. Lee,‡ and L. N. Kantorovich†

Department of Physics, School of Physical Sciences and Engineering, King's College London, The Strand, London, WC2R 2LS, United Kingdom, and New Materials and Components Research Center, Research Institute of Industrial Science and Technology, P.O. Box 135, Pohang 790−330, South Korea

Received: September 14, 2005

All the planar homopairings of cytosine and guanine are reported for the first time in this study. The idea of *binding sites* suggested for the simple case of adenine homopairs (*J. Phys. Chem. B* **2005**, *109*, 11933) is shown to be applicable to more complicated molecules binding to each other via multiple hydrogen bonds and can be considered as a general method for constructing hydrogen bonding structures. As an example we consider homopairs formed by DNA bases cytosine and guanine, suggesting that there may be 13 cytosine and 17 guanine homopairs. However, only 11 cytosine and 15 guanine homopairs remain after atomic relaxation performed using ab initio density functional theory. Most of the homopairs obtained have not been studied before. The homopairs have significant binding energies, varying from −0.19 to −1.12 eV, that are explained by multiple hydrogen bonds formed between monomers in the pairs, up to four hydrogen bonds in most energetically favorable cases. The detailed information on all guanine and cytosine planar homopairs contained in this work can be used to construct various cytosine and guanine superstructures observed on different surfaces.

1. Introduction

All the DNA base molecules (adenine, cytosine, guanine and thymine) have the ability to form one- or two-dimensional superstructures when deposited onto various insulating or metal surfaces. The particular superstructures vary due to the particular base deposited, the surface and the temperature. In particular, when the interaction between the surface and the substrate is small, guanine[1−5] and adenine[3,4,6−12] have shown the ability to form chains and hexagonal monolayers on various surfaces. Cytosine has been studied to a much smaller extent but has the interesting property that it forms one-dimensional filaments when deposited onto various surfaces where surface-substrate interaction is weak.[3,4,10,12]

Self-assembling superstructures on solid surfaces are currently an interesting research area due to their potential uses in nanotechnology.[13] Unfortunately, it is not possible at present to resolve the atomistic structure of the superstructures experimentally, so that theoretical calculations are necessary to determine the actual geometries. Theoretical modeling often relied upon semiclassical and semiempirical approaches to predict the DNA base superstructures. However, previous ab initio calculations on pairs have highlighted the need for higher accuracy methods.[14,15]

Research into guanine tetrads has been of great interest because they have a functional role in nature (see ref 16 and references therein) and have been proposed as targets for drug design in cancer therapeutics.[17] Two-dimensional periodic guanine tetrads have been suggested to explain the observed images in recent scanning tunneling microscope (STM) experiments[5] formed by guanine on the gold(111) surface. A DFT modeling approach has been used to justify the single structural model suggested. At higher temperature a phase transition to a different structure was also observed; however, only a single structural model has been proposed to explain the STM image.

It has been previously shown[18] that one should be very careful when providing a model for a particular structure observed. There may be a numerous number of stable structures with similar lattice vectors. A comprehensive study should include all of these, in which case all possible pairings between monomers are to be carefully studied.

The previous theoretical calculations on the cytosine chains were performed using a semiempirical approach.[10] The authors suggested one particular chain based on information available on three dimer geometries. Note, however, as it has been already documented in the literature,[14,15] that semiempirical approaches perform poorly for the DNA base pairs, so that an ab initio approach is essential to get all the possible dimers, including their correct stabilization energies and geometries. This information will prove to be crucial for obtaining all possible cytosine superstructures.

All previous studies on cytosine and guanine superstructures include few pairs and no systematic approach has been applied to identify and study all possibilities. Therefore, in this study in section 3 we provide a complete analysis of *all* possible planar orientated cytosine and guanine homopairs. The ab initio approach we used is discussed in section 2. A short discussion of our results is given in section 4. To our knowledge, no previous work exists in which all the possible planar cytosine and guanine homopairs have been investigated.

2. Method

The ab initio calculations were performed using the SIESTA method[19−21] which is based on the density functional theory

* To whom correspondence should be addressed. E-mail: ross.kelly@kcl.ac.uk.
† King's College London.
‡ Research Institute of Industrial Science and Technology.

10.1021/jp055207z CCC: $30.25
Published on Web 10/22/2005

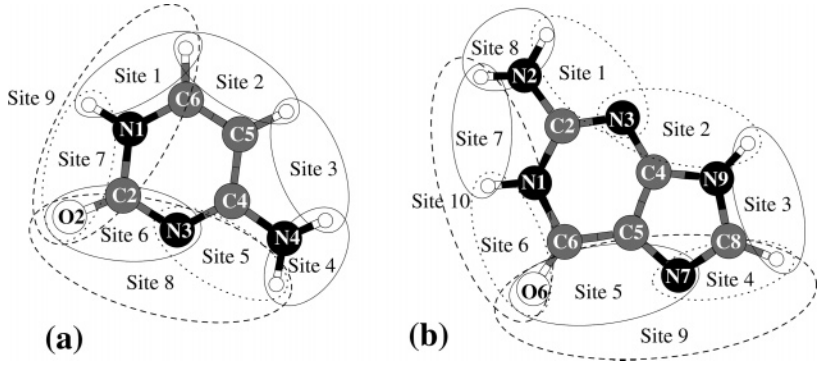

Figure 1. (a) Cytosine and (b) guanine molecules in configurations *C* and *G*, respectively. The configurations $\bar{C}$ and $\bar{G}$ are obtained by flipping configurations *C* and *G* in the molecular plane. All possible bonding sites which can participate in forming hydrogen bonds in cytosine and guanine homopairs are explicitly indicated. Three types of sites exist in these molecules: (i) with either two acceptors or two hydrogens, indicated by solid lines, (ii) with one hydrogen and one acceptor, indicated by dotted lines, and (iii) with three atoms that are indicated by dashed lines. The atom label nomenclature used is the same as in ref 24.

(DFT) and was described previously in our work on adenine dimers.[15] Briefly, we use the SIESTA code which employs localized numerical atomic orbital basis sets, norm-conserving pseudopotentials and periodic boundary conditions. The Perdew, Becke, and Ernzerhof (PBE)[22] exchange-correlation functional was found to be more than suitable for DNA base pairs[15] and therefore we used this functional in our work. The double-ζ plus polarization (DZP) basis set was used in all calculations with the energy cutoff of 10 meV.[21] We stress that a large basis set is essential to obtain realistic intermolecular bonding. The pseudopotentials were generated in the same way as was stated in ref 15.

Each pair geometry was obtained from a full atomic relaxation with the forces on atoms not greater than 0.01 eV/Å. In all calculations, the cell size was very large (which is not critical for a local basis set calculation) to avoid unphysical interaction between neighboring images and therefore only one (γ) **k**-point was required in all our calculations.

Some properties of the cytosine and guanine molecules, such as the intramolecular bond lengths and angles, were checked with previous DFT[23] and quantum chemical (QC) calculations.[24] Intramolecular bond lengths and angles compare well with these previous ab initio results. The dipole moment of cytosine was found to be 6.2 D, in good agreement with the previous ab initio DFT calculations (6.5 D[23]), QC calculations (6.4 D[25]), and experiment (7.0 D[26]). Guanine was found to have an even larger dipole moment of 6.4 D which is also in good agreement with the previous DFT (6.9 D[23]) and QC (6.6 D[25]) calculations and experiment (7.1 D[27]).

Amino groups in both molecules are not exactly within the molecule plane: both hydrogen atoms go out of the plane and the Nitrogen atom deviates in the opposite plane direction. This so-called *pyramidilization* was found by Sponer and Hobza[24] to be due to sp^3 hybridization. We also found this effect in our calculations of the individual monomers, the nonplanarity of guanine being the highest.[24] Note also that the energy barrier to the planar configuration was found in ref 24 to be rather small. As this point is essential in understanding equilibrium structures of guanine and cytosine pairs, we shall also consider it in section 3.4.

To analyze the stability of cytosine and guanine homopairs, we calculated a number of useful energies.[25] First, the *stabilization energy*, E_{stab}, is defined as the total energy of the relaxed combined system (a pair) minus the total energies of all its individual relaxed components (two cytosine or guanine molecules in this case). $E_{stab} < 0$ for stable pairs. To characterize the interaction between the two molecules, we calculate the *interaction energy*, E_{int}, which is defined as the total energy of the pair minus the energies of each individual molecule calculated in the geometry of the pair (i.e., without relaxation). Finally, we define the *deformation energy*, $E_{def} = E_{stab} - E_{int}$, which represents the sum of losses in the total energies of the two molecules due to their relaxation in the pair. Obviously, as $E_{def} > 0$, the interaction energy E_{int} must be negative, similar to E_{stab}.

When calculating stabilization energies, care should be taken due to the localized basis set used. The basis set superposition error (BSSE[28]) correction to the stabilization energies has to be included to account for the different basis set used in calculating the pair and each of the individual molecules. These corrections have been calculated by the standard Boys−Bernardi counterpoise correction method.[28]

3. Results

3.1. Cytosine and Guanine: Binding Sites. It is a rather general assumption that in order to form a stable dimer between two nearly planar molecules, at least two hydrogen bonds must exist between them. In constructing all possible adenine homopairs,[15] we found it extremely useful first to identify all *bonding sites* (or *sites* for short) in the molecule. Note that, in the case of adenine homopairs, the choice of binding sites is straightforward and all of them are of the same type of bonding, namely each site has one donor and one acceptor. Then, every pair can be constructed by connecting each site of one molecule with every site of the other flipping the second molecule if necessary. This method was found to be very general, as it can be applied to any DNA base.

As will be clear later on in this section, the cases of cytosine and guanine are much more complicated. However, we shall show that, even in these cases, the idea of binding sites can also be employed with some considerable generalizations, so that all possible homopairings of cytosine and guanine can be constructed.

For convenience, the orientation of the molecules shown in Figure 1, parts a and b will be denoted *C* and *G* respectively, while their nonchiral counterparts (flipped with respect to the molecule plane) will be called $\bar{C}$ and $\bar{G}$, respectively. Following notations of the previous paper,[15] the relevant sites of each molecule that are involved in the bonding will also be indicated explicitly, e.g., C_2C_6 or $G_2\bar{G}_1$.

We shall first consider a cytosine molecule. There are seven exposed atoms which can participate in hydrogen bonds with another cytosine molecule, see Figure 1a. We can thus identify

nine sites which have the ability to form at least a double hydrogen bond. These are indicated explicitly in the same Figure. The sites 1−4 have two neighboring hydrogens, while site 6 has two neighboring acceptors O2 and N3. Thus, sites 1−4 of one monomer can form a double hydrogen bond with site 6 of another. Note that, because of the symmetry of each site upon flipping the molecule, there are two dimer possibilities when bonding sites 1−4 with site 6, the other dimer is obtained in each case by flipping one of the molecules. For example, by flipping the second molecule in the pair C_2C_6, another nonequivalent pair $C_2\bar{C}_6$ is obtained. Altogether, there are 8 possibilities here.

Sites 5 and 7 can only bind with each other as they contain only one hydrogen and one acceptor in each site (note that all sites in the adenine molecule are of this type[15]). The correct chirality of the second monomer is imposed by the first one, so that only one pairing possibility is possible for each site−site combination giving three different dimers: C_5C_5, $C_7\bar{C}_5$ and C_7C_7.

Site 8 consisting of one of the N4 hydrogen atoms and the two acceptors O2 and N3 was also identified as another possible binding site. Indeed, one can see that the atom N3 is slightly pulled into the molecule. Therefore, it may not be an obstacle in attaching sites 8 of two molecules together via the distant hydrogen atoms and the acceptors O2. Of course, only a single possibility exists here, C_8C_8.

Finally, there is also site 9 containing two hydrogens of site 1 and the O2 acceptor atom. This three-atom site is special as it may bind to the other three-atom site 8, i.e., we anticipate that the pair $C_8\bar{C}_9$ may also exist. This pair may be expected to have *three* hydrogen bonds formed between two molecules. Note, however, that the atomic arrangement in this particular pair is similar to that in the pair $C_7\bar{C}_5$, so that one may expect that only one of these will survive.

Therefore, in total there are 13 possible cytosine homodimer possibilities.

A similar analysis can be performed for the guanine molecule as well. It has 8 external atoms that can participate in hydrogen bondings, so that 10 binding sites can be identified as shown in Figure 1b. Sites 3, 7, and 8 consist of two neighboring hydrogens, while site 5 can provide two neighboring acceptors, O6 and N7. As in the previous case, both orientations of the second molecule are possible, giving 6 pairs in total.

Sites 1, 2, 4, and 6 all have only one hydrogen and one acceptor each, so that only combinations of these sites can bond together. The chirality of the second molecule is imposed on the bonding and therefore only one pairing possibility is possible for each of these site−site combinations leading in total to 10 pairs.

It is interesting that, similarly to the cytosine case, long bonding sites also exist in a guanine molecule as well, although they have a rather different structure and provide a new functionality. Indeed, site 9 contains two adjacent acceptors O6 and N7 and a single H atom attached to C8. On the other hand, site 10 has two adjacent H atoms (attached to atoms N2 and N1, respectively) and a single acceptor O6. One can consider each of these sites as a combination of two sites: site 9 is combined of sites 4 and 5, while site 10 is equivalent to using sites 6 and 7 at the same time. Only one possibility, $G_9\bar{G}_{10}$, exists here since one of the molecules should be flipped. In this special case three rather than two hydrogen bonds are to be expected to be formed.

Thus, summarizing, there are 17 possible guanine homodimer possibilities: sites 1−8 have the ability to form double hydrogen bonds and sites 9, 10 can participate in a triple hydrogen binding.

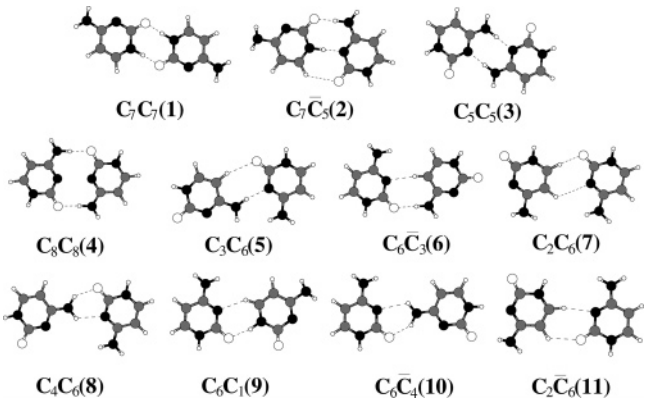

Figure 2. Relaxed cytosine homopairs in order of stability (shown by the number in the round brackets). The notations C_nC_m and $C_n\bar{C}_m$ correspond to chiral and nonchiral cytosine dimers, respectively, the indices *n* and *m* indicating explicitly the sites of the two molecules (as adopted in Figure 1) engaged in the hydrogen bonds in each pair.

Note that if one of the amino group hydrogens is involved in the bonding in a dimer, the pyramidilization of the group can in principle affect the chemical ability of the site, if the hydrogen is significantly out of the molecule plane. In this case, the out-of-plane hydrogen may expose the acceptor atom it is attached to whereby facilitating formation of another type of the site. Since, the pyramidilization is only significant in a guanine molecule, we have looked at this affect in more detail for the guanine only. In this case we find that the hydrogen atom common to both sites 7 and 8 (see Figure 1b) deviates more from the molecular plane than the other hydrogen of the amino group. In this case, site 7 has the ability to turn into a different site, type 7*, that contains one H atom (common to sites 7 and 6) and atom N2 as an acceptor. This site would then be able to bind to any of the sites 1, 2, 4, 6, and 7* of the other molecule. We shall consider these cases separately in section 3.4 as they lead to nonplanar dimers. At the same time, the described ability of site 7 (7*) to be either two hydrogen or one hydrogen and one acceptor site does play a certain role in some of the planar guanine dimers as we shall discuss in more detail in section 3.4 as well.

In fact, this systematic approach to construct planar pairs by identifying binding sites, can be applied to almost any system where the hydrogen bonding is the main binding mechanism.

3.2. Cytosine and Guanine Homodimers: Energetics. The geometries of all relaxed planar pairs of cytosine and guanine are shown in Figures 2 and 3, respectively. 11 cytosine pairs (out of 13 predicted in section 3.1 by considering all sites) were found to be stable after geometry relaxation. These include three chiral centrosymmetric pairs (C_5C_5, C_7C_7, C_8C_8) built using the same sites on each monomer, so that they each have C_2 symmetry. We found two pairs less since three of the predicted pairs, namely $C_6\bar{C}_1$, $C_7\bar{C}_5$ and $C_8\bar{C}_9$, relaxed in fact to the *same* configuration facilitated by a favoring arrangements of the adjacent sites 1 and 7 of one molecule with the adjacent sites 5 and 6 of the other (which is flipped). As will be shown in section 3.5, in this case two strong and one weak hydrogen bonds are formed. Since the bonding between sites 7 and 5 is the most prominent, we shall denote this particular resultant configuration as pair $C_7\bar{C}_5$.

A total of 15 guanine pairs (out of 17 predicted in the previous section) were successfully relaxed including four chiral centrosymmetric pairs (G_1G_1, G_2G_2, G_4G_4, and G_6G_6). Similar to

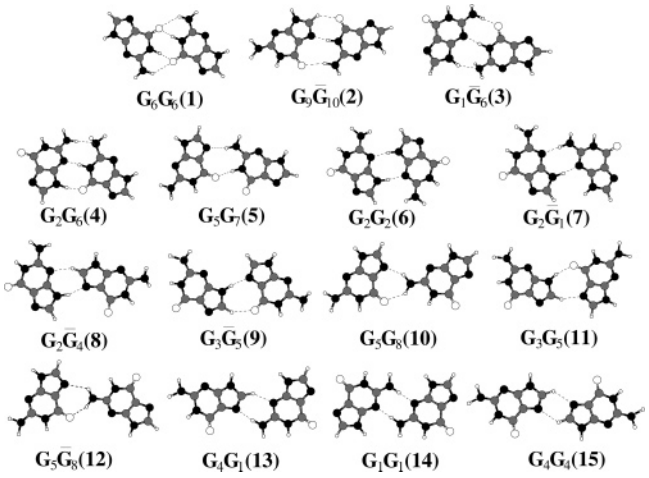

Figure 3. Relaxed guanine pairs in order of stability. Notations are analogous to those in Figure 2.

the case of the cytosine pairs $C_6\bar{C}_1$, $C_7\bar{C}_5$, and $C_8\bar{C}_9$, three of the predicted guanine homopairs, $G_4\bar{G}_6$, $G_5\bar{G}_7$, and $G_9\bar{G}_{10}$, also relaxed to the same configuration. Upon careful analysis of the charge density (see section 3.5), we have decided to call this structure $G_9\bar{G}_{10}$ as it offers *three* rather than two hydrogen bonds with substantial energy benefit. Note that in our method the molecular configurations containing more than two hydrogen bonds appear naturally without any special consideration.

The stabilization, interaction, deformation, and BSSE energies of all the cytosine and guanine homodimers are given in Table 1. In all these calculations, the BSSE corrections were found to be significant.

The most stable cytosine, C_7C_7, and guanine, G_6G_6, pairs are both centrosymmetric. In total there are three cytosine pairs with similar high stability ($E_{stab} < -0.8$ eV) which are all derived from sites 5 and 7 (C_7C_7, $C_7\bar{C}_5$ and C_5C_5). The most stable guanine pair, G_6G_6, has very significant stabilization energy of -1.1 eV. This is investigated further in the next two sections. In fact we find that all the cytosine and guanine homopairs have substantial stabilization energies ($E_{stab} < -0.2$ eV).

One can also see that the most stable cytosine dimers are formed with sites 7 and 5, while the most stable guanine pairs involve sites 6, 9, and 10. Thus, in most cases the oxygen acceptor atoms are preferred to nitrogen atoms as may be expected on chemical grounds due to the higher electronegativity of the oxygen atoms.

The deformation energies for the most stable pairs are significant (reaching 0.2−0.3 eV); however, in all other cases the monomers deformation is found to be not very substantial and can be neglected. Obviously, this approximation is not applicable for the few most stable pairs.

Unfortunately, there is very little ab initio data available to date on the cytosine and guanine homopairs; a few available energies are presented in Table 1. The recent benchmark ab initio results for DNA base pairs[29] do not include any cytosine−cytosine base pairs and only one guanine−guanine pair. However, the stabilization energy of this guanine pair and of the C_5C_5 pair relaxed using the MP2 method[30] (presented in Table 1) agrees extremely well with our results. Note that the MP2 energy calculated using the Hartree−Fock geometry optimization[30] has lower energy, which is expected because the geometry does not correspond to the correct MP2 minimum. Note also that energetic ordering of the pairs calculated in ref 30 is the same as ours.

As mentioned in section 1, semiempirical schemes have been reported to be unsuitable for DNA base pairs and again this is evident when comparing our ab initio results with those calculated using the PM3 method[10] (see the last column in Table 1). The stabilization energies are largely underestimated and the energetic ordering of the pairs is inconsistent with our ab initio work.

TABLE 1: Cytosine and Guanine Homodimer Energies (in eV)[a]

	PBE				E_{stab}(literature)		
pair	E_{BSSE}	E_{int}	E_{def}	E_{stab}	ref 29	ref 30	ref 10
$C_7\bar{C}_7(1)$	0.21	−1.32	0.32	−0.99			−0.48
$C_7\bar{C}_5(2)$	0.20	−1.20	0.24	−0.96			
$C_5C_5(3)$	0.15	−1.05	0.18	−0.87		−0.76/−0.88	−0.40
$C_8C_8(4)$	0.23	−0.68	0.17	−0.51			−0.05
$C_3\bar{C}_6(5)$	0.07	−0.48	0.04	−0.44			
$C_6\bar{C}_3(6)$	0.08	−0.47	0.03	−0.43			
$C_2C_6(7)$	0.08	−0.41	0.01	−0.39			
$C_4C_6(8)$	0.09	−0.44	0.06	−0.38			
$C_6\bar{C}_1(9)$	0.11	−0.42	0.05	−0.37			
$C_6\bar{C}_4(10)$	0.09	−0.31	0.04	−0.27			
$C_2\bar{C}_6(11)$	0.10	−0.21	0.01	−0.19			
$G_6\bar{G}_6(1)$	0.21	−1.41	0.30	−1.12		−0.96	−0.46
$G_9\bar{G}_{10}(2)$	0.17	−0.85	0.10	−0.75	−0.80	−0.74	−0.35
$G_1\bar{G}_6(3)$	0.21	−0.97	0.22	−0.74			
$G_2G_6(4)$	0.21	−0.93	0.21	−0.72			
$G_5G_7(5)$	0.15	−0.59	0.07	−0.52			−0.11
$G_2\bar{G}_2(6)$	0.17	−0.61	0.09	−0.52			−0.22
$G_2\bar{G}_1(7)$	0.16	−0.59	0.14	−0.45			
$G_2\bar{G}_4(8)$	0.11	−0.50	0.05	−0.45			
$G_3\bar{G}_5(9)$	0.11	−0.48	0.04	−0.44			
$G_5G_8(10)$	0.11	−0.50	0.07	−0.43			
$G_3\bar{G}_5(11)$	0.11	−0.46	0.03	−0.43			
$G_5\bar{G}_8(12)$	0.11	−0.47	0.06	−0.41			
$G_4G_1(13)$	0.11	−0.48	0.09	−0.39			
$G_1G_1(14)$	0.16	−0.53	0.14	−0.39		−0.40	−0.20
$G_4G_4(15)$	0.07	−0.27	0.02	−0.25			

[a] E_{int}, E_{def} and E_{stab} are the interaction, deformation, and stabilization energies, respectively, while E_{BSSE} corresponds to the BSSE correction. Results of our calculations are compared with other available work presented in the last three columns. The two values for E_{stab} presented in the second to the last column for the $C_5C_5(3)$ Pair correspond to the MP2 energies calculated using the Hartree−Fock and MP2 optimized geometries, respectively.

Contrary to the QC study[30] and in agreement with the earlier work,[31] we found that the G_5G_7 dimer is stable with a significant stabilization energy.

3.3. Cytosine and Guanine Homodimers: Geometry. The hydrogen bonds are said to be *cooperative*[32] which means that each bond tries to achieve a donor-hydrogen-acceptor bond angle approaching 180° with a preferred bond length. However there is an intrinsic rigidity[32] of the DNA bases so that this is not always achieved. It has been reported that the hydrogen bond strength cannot be entirely analyzed using just the geometry of the hydrogen bonds because all atoms should be taken into account.[31,30] In the cases of cytosine and guanine, between two to four hydrogen bonds may be formed, although the strength of each bond in the same pair may significantly vary.

Detailed information on the hydrogen bond lengths and the angles between the donor, the hydrogen and the acceptor atom are given in Tables 2 and 3 for the cytosine and guanine pairs, respectively. The cooperativity of the bonds can be seen, especially in the cases of strongly interacting molecules. If present, the N−H−O bonds are mostly closer to 180° and have shorter bond lengths than their closest rivals (C/N)−H−N bonds. Note that because the symmetry of centrosymmetric pairs is broken in our calculations implementing periodic boundary conditions, the bond lengths and angles of centrosymmetric pairs are presented as averages in Tables 2 and 3. The deviation from perfect symmetry in these cases was found to be within 1.0%

TABLE 2: Geometrical Characteristics of the Hydrogen Bonds in Cytosine Homopairs[a]

		PBE	
pair	H-bond	length	angle
$C_7C_7(1)$	N1−H−O2	2.64	177.1
	N1−H−O2	2.64	177.1
$C_7\bar{C}_5(2)$	N1−H−N3	2.74	178.3
	N4−H−O2	2.69	179.6
	C6−H−O2	3.61	128.0
$C_5C_5(3)$	N4−H−N3	2.81	175.7
	N4−H−N3	2.81	175.7
$C_8C_8(4)$	N4−H−O2	2.95	166.1
	N4−H−O2	2.95	166.1
$C_3C_6(5)$	N4−H−N3	3.05	159.7
	C5−H−O2	3.35	168.6
$C_6\bar{C}_3(6)$	N4−H−O2	2.86	174.9
	C5−H−N3	3.63	167.1
$C_2C_6(7)$	C6−H−O2	3.23	175.6
	C5−H−N3	3.93	122.2
$C_4C_6(8)$	N4−H−O2	2.87	138.2
	N4−H−N3	3.03	123.9
$C_6C_1(9)$	N1−H−O2	2.85	164.4
	C6−H−N3	3.57	142.0
$C_6\bar{C}_4(10)$	N4−H−O2	2.86	144.8
	N4−H−N3	3.18	116.6
$C_2\bar{C}_6(11)$	C5−H−O2	3.21	123.5
	C6−H−N3	3.52	169.7

[a] Each bond is presented via the distance (in Å) between the H donor atom of one molecule and the corresponding acceptor (either N or O) atom of another. The angle (in degrees) between the X−H atoms of the first molecule and the N or O atom of another is also shown, where X, depending on the particular dimer, is either C or N. Numbering of atoms is identical to that adopted in Figure 1a.

TABLE 3: Geometrical Characteristics of the Hydrogen Bonds in Guanine Homopairs[a]

		PBE		ref 30		
pair	H-bond	length	angle	length	angle	ref 29 length
$G_6G_6(1)$	N1−H−O6	2.66	172.5	2.87	178.1	
	N1−H−O6	2.66	172.5	2.87	178.1	
	N2−H−O6	3.23	132.4			
	N2−H−O6	3.23	132.4			
$G_9\bar{G}_{10}(2)$	N1−H−N7	2.78	170.8	2.96	172.5	2.87
	N2−H−O6	3.04	166.5	3.27	166.2	3.15
	C8−H−O6	3.15	123.8			
$G_1\bar{G}_6(3)$	N2−H−O6	2.75	177.4			
	N1−H−N3	2.76	175.2			
	N9−H−N2	3.14	144.2			
$G_2G_6(4)$	N9−H−O6	2.73	167.6			
	N1−H−N3	2.78	176.8			
	N2−H−N2	3.19	158.5			
$G_5G_7(5)$	N2−H−N7	3.05	164.0			
	N1−H−O6	2.87	174.4			
$G_2G_2(6)$	N9−H−N3	2.86	173.0			
	N9−H−N3	2.86	173.0			
$G_2\bar{G}_1(7)$	N2−H−N3	2.93	176.7			
	N9−H−N3	2.83	173.2			
$G_2\bar{G}_4(8)$	N9−H−N7	2.85	164.0			
	C8−H−N3	3.24	145.3			
$G_3\bar{G}_5(9)$	C8−H−O6	3.29	137.8			
	N9−H−N7	2.93	174.6			
$G_5G_8(10)$	N2−H−O6	2.94	135.5			
	N2−H−N7	2.95	150.7			
$G_3G_5(11)$	N9−H−O6	2.84	171.8			
	C8−H−N7	3.44	142.4			
$G_5\bar{G}_8(12)$	N2−H−O6	2.91	155.0			
	N2−H−N7	3.12	130.7			
$G_4G_1(13)$	N2−H−N7	2.88	171.1			
	C8−H−N3	3.19	139.4			
$G_1G_1(14)$	N2−H−N3	2.87	177.0	3.15	178.8	
	N2−H−N3	2.87	177.0	3.15	178.8	
$G_4G_4(15)$	C8−H−N7	3.22	143.8			
	C8−H−N7	3.22	143.8			

[a] The notations are the same as in Table 2. Numbering of atoms is identical to that adopted in Figure 1b. In the last three columns results of our calculations are compared with those found in refs 29 and 30, using the Hartree−Fock and MP2 geometry optimization methods, respectively. The bond lengths and angles of centrosymmetric pairs are averaged.

and 0.3% for bond lengths and angles respectively, similar to the adenine homopairs.

The G_6G_6 pair, seen in Figure 2, visually seems to have *four* hydrogen bonds; this conclusion can be drawn by noticing short N−H−O distances given in Table 3. This point is also confirmed by a detailed analysis of the electron density change made in section 3.5 and explains why the stabilization and interaction energies are the highest for this particular pairing.

The distances between donor-hydrogen−acceptor atoms in the C_5C_5 pair were reported previously[30] using quantum chemistry methods: the bond lengths are shorter after the MP2 geometry optimization (2.92 Å) as compared with the HF geometry optimization (3.05 Å). Our result (2.81 Å) agrees well with the MP2 result of ref 30. In addition, the angle of 173.2° found in ref 30 is very close to our result of 175.7°. Similar agreement can be found between our results and previous quantum chemistry calculations for the $G_9\bar{G}_{10}$ pair; see Table 3.

3.4. Flexibility of the Cytosine and Guanine Amino Groups. In pairs in which the amino groups did not participate in the bonding, the actual orientation of the groups did not change appreciably, i.e., the pyramidilization remained almost identical to that of isolated monomers. In most cases where amino groups did participate in the bonding, the relaxed geometries were found to be nearly perfectly planar, i.e., the pyramidilization is easily destroyed. This is explained by a very small energy required to rotate the amino group in order to position one of its hydrogens into the planar configuration (19 and 11° for the cytosine hydrogens and 34 and 15° for the two hydrogens of a guanine monomer). To estimate the required energies, we calculated the potential energy surfaces of an isolated cytosine and guanine as functions of the rotational angle of the two amino group hydrogen atoms about the C4−N4 (cytosine) and C2−N2 (guanine) axes. In both cases the two hydrogen atoms were rotated as a whole. The angle of rotation was measured from the equilibrium (out-of-plane) initial position of the amino group. Results of these calculations demonstrate clearly that quite large barriers are required to rotate the groups by 180° and that the barrier for the guanine is almost two times smaller than for the cytosine molecule. Note that in these calculations no atomic relaxation was performed, so that obtained curves (and the barriers) are upper estimates for the actual potential energy surfaces with respect to the rotational angles in question. Small rotations (10−35°) required to bring one of the two hydrogens into the molecular plane cost very little energy which is easily available due to hydrogen bond formation facilitated by this rotation.

As was already mentioned at the bottom of section 3.1, in some cases the amino group may actually facilitate the hydrogen bonding without going into the planar configuration by exposing the corresponding N atom which the two hydrogens are attached to. This effect is the most prominent in guanine where one of the H atoms (common to sites 7 and 8, see Figure 1) is rotated by 34° from the molecular plane significantly opening atom N2 to a hydrogen of another molecule. Two mechanisms with which this happens have been identified in our calculations on guanine dimers.

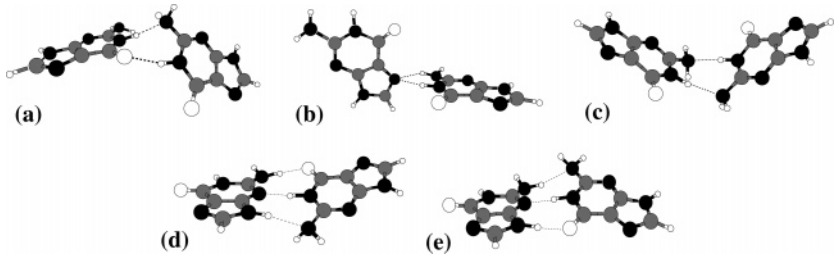

Figure 4. Relaxed nonplanar guanine pairs (a) $G_6\bar{G}_{7*}$, (b) G_4G_{7*}, (c) $G_{7*}G_{7*}$, (d) $G_1\bar{G}_6$, and (e) G_2G_6.

TABLE 4: Energies (in eV) of the Three Nonplanar Guanine Dimers[a]

	PBE			
pair	E_{BSSE}	E_{int}	E_{def}	E_{stab}
$G_6\bar{G}_{7*}$	0.14	−0.54	0.08	−0.46
G_4G_{7*}	0.08	−0.49	0.05	−0.44
$G_{7*}G_{7*}$	0.10	−0.35	0.16	−0.19

[a] See Figure 4, parts a, b, and c.

The first case is related to two planar pairs G_2G_6 and $G_1\bar{G}_6$ (see Figure 3, parts d and e) in which the main binding mechanism is via double hydrogen bonds 2−6 and 1−6 (indicated by the pair symbols). However, due to a particularly convenient spacial position of the amino group of one of the molecules that is opposite to a hydrogen atom of another, a third hydrogen bond is formed between this H atom and the N2 acceptor of the amino group (see also section 3.5). The extra third bond between the two molecules works toward increasing the stabilization energies of the two pairs. However the planar geometry is slightly distorted to accommodate this extra bond (see Figure 4, parts d and e). The amino group which allows the extra bonding with the hydrogen atom of the other molecule is found to be slightly distorted to facilitate this extra bonding: the average additional out-of-plane rotation of the amino group hydrogens was found to be 13.9 and 10.8° for the G_2G_6 and $G_1\bar{G}_6$ pairs, respectively.

The second case appears when atom N2 in site 7 is treated as a hydrogen-acceptor site 7* rather than a two hydrogen site, and is considered as one of the main hydrogen bonds stabilizing the pair. Five bonding possibilities then arise in this way, G_1G_{7*}, $G_2\bar{G}_{7*}$, G_4G_{7*}, $G_6\bar{G}_{7*}$, and $G_{7*}G_{7*}$, have been considered. We find that homopairs G_1G_{7*} and $G_2\bar{G}_{7*}$ relaxed respectively to slightly nonplanar geometries G_2G_6 and $G_1\bar{G}_6$, already discussed above. At the same time, the other three possibilities relaxed to distinctly nonplanar L-shape pairs which are shown in Figure 4 (*a*), (*b*) and (*c*). Note that only $G_6\bar{G}_{7*}$ and $G_{7*}G_{7*}$ pairs relaxed to the configurations that benefited from that hydrogen bonding with the amino group. The pair G_4G_{7*} relaxed to a completely different structure in which two hydrogen bonds are formed with the same acceptor atom. Note that this structure could as well be a saddle point. The stabilization, interaction, deformation and BSSE energies for the relaxed nonplanar pairs are given in Table 4. It is clear that these pairs do mostly have significant stabilization energies. Note, however, that nonplanar geometries will most probably play little role in possible super-structures observed on crystal surfaces.

3.5. Cytosine and Guanine Homodimers: Electronic Density. The valence electron density of a dimer gives little information about the binding mechanism. To see more characteristics of each hydrogen bond in a pair one can use the interaction density.[15] Briefly, the interaction density is the difference between the total electron density of the pair minus the sum of monomer densities. The monomer densities must be in the pair geometry, otherwise density changes due to atomic relaxation will be visible. A similar method was used in another DNA base pair study.[33]

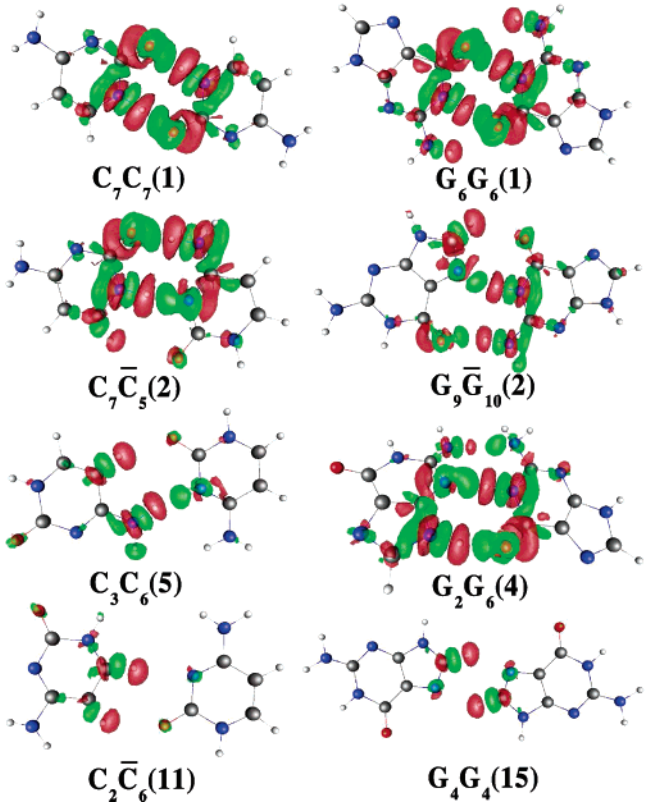

Figure 5. Interaction electron density of cytosine (on the left-hand side) and guanine (right) homopairs. The electron density difference corresponds to ±0.01 electrons in each picture. Green surfaces correspond to regions of positive electron density difference (i.e., charge excess areas) and red areas correspond to regions of negative electron density difference (charge depletion areas). A number inside the brackets by the cytosine (guanine) pair symbol above corresponds to the pair position within all cytosine (guanine) molecules ordered with respect to their stabilization energies, the most stable pairs go first (see Table 1).

There are many different and interesting hydrogen bonded pairs that we found in this study. We start from the most stable cytosine and guanine pairs C_7C_7 and G_6G_6. The difference electron densities of these are shown in Figure 5. Similar to the study on adenine homopairs,[15] we find characteristic alternating regions of excess and depletion in the charge density (±0.01 electrons) that are present across the hydrogen bonds resembling a "*kebab*" structure.

The most stable cytosine pair C_7C_7 is seen to have two bonds, whereas the most stable guanine pair G_6G_6 has four hydrogen bonds as suggested in section 3.3. In fact, one can appreciate that the main two hydrogen bonds in the most stable cytosine and guanine pairs look identical to one another because they involve the same hydrogen bonding atoms. These high stability pairs also have very similar bonding geometry (see section 3.3).

The two extra hydrogen bonds in the guanine pair G_6G_6 are due to planar amino groups of each molecule that bind to the acceptor O6 atom of another molecule already involved in the main hydrogen bonds; compare Figures 3 and 5.

The cytosine pair $C_7\bar{C}_5$, (see Figures 2 and 5), apart from two strong bonds between sites 7 and 5, has also established an additional, although much weaker, hydrogen bond between the acceptor O2 of one molecule and the hydrogen atom common to sites 1 and 2 of another. A similar picture is observed in the charge density difference plot of the guanine pair $G_9\bar{G}_{10}$: although it has been anticipated (opposite to the previous case of the pair $C_7\bar{C}_5$) that three bonds should be formed, only two bonds are, in fact, found to be relatively strong, as the third one made between atom O6 of site 10 of one molecule and the hydrogen of site 9 of another appeared to be much weaker.

As was already discussed in section 3.4, the guanine pairs $G_1\bar{G}_6$ and G_2G_6 exhibit an additional (third) hydrogen bond between the exposed acceptor N2 of the amino group of one molecule and a hydrogen atom of another. As an example, we show in Figure 5 the charge density difference plot of the pair G_2G_6 where the third bond is clearly visible.

In the cytosine pair C_3C_6, the charge density plot is also given in Figure 5, the amino group rotated into the planar geometry and established a rather strong hydrogen bonding with the other molecule; a relatively small stabilization energy of this pair (−0.44 eV, Table 1) is explained by the much weaker second bond in this pair.

Less stable structures are expected to show much less charge redistribution (cf. adenine pairs in ref 15) with incomplete "kebab" structures; however, a few alternating regions may be still present which appear to be enough to justify some stability of these pairs. This is clearly seen in Figure 5, where the charge density difference plots of the least stable pairs $C_2\bar{C}_6$ and G_4G_4 are shown.

Finally, one more observation can be added to the above analysis of the charge density of the cytosine and guanine pairs: interestingly, the pairs C_3C_6 and G_2G_6, see Figure 5, have the charge redistribution on some of their atoms which are located far from the bonding. Note that this cannot possibly be an artifact of our calculations since the interaction between adjacent images must be negligible due to massive cell sizes used to avoid this effect. It is reasonable to expect that this could be a direct quantum chemical evidence of the so-called *resonance assisted hydrogen bonding* (see for example ref 5 and references therein).

4. Discussion and Conclusions

In this work, we have presented all relaxed planar cytosine and guanine homopairs stabilized by at least two hydrogen bonds using an ab initio method. In analyzing all possible pairings between two molecules, we identified binding sites on each of the monomers that can bind to each other. We demonstrate in this paper that the method suggested previously in ref 15 for the simple case of adenine pairs, can be extended and is in fact quite general. It can be used to systematically identify all pairs bound together by multiple hydrogen bonds and then suggest a very good initial guess for their geometries before more complex calculations are performed.

We found that 11 cytosine and 15 guanine pairs are possible, all with significant stabilization energies ranging from −0.19 up to −1.12 eV. There are three cytosine pairs with very similar high stabilization energies ($E_{stab} < -0.80$ eV), whereas there are five guanine pairs with E_{stab} ranging between −0.50 and −0.75 eV and one pair, G_6G_6, with distinctively high stabilization energy of −1.12 eV. We attribute this very high stabilization energy to the four hydrogen bonds discussed in the previous sections. In all cases where more than two hydrogen bonds were formed, these were all obtained by only considering two hydrogen bonds between monomers.

There is a very little number of available ab initio studies on homopairings. Our results do compare extremely well with the recent ab initio benchmark calculations of DNA base pairs.[29,30] However, as to be expected from the previous work on DNA base pairs,[14,15] we find a rather bad agreement with the semiempirical methods, so that care should be exercised when applying semiempirical or semiclassical techniques to these systems.

The strongly bound cytosine and guanine homopairs do have quite significant deformation energies, so that one may not consider them as rigid. The strongest pairs also have short linear donor-hydrogen-acceptor distances.[24,31] We find that the idea of *binding sites* for cytosine and guanine suggested in this paper works quite well, the strongest hydrogen bonds are established according to the preliminary analysis based on the available sites; however, we have also found that in some cases pairs manage to form weak additional hydrogen bonds due to convenient arrangement of neighboring donor and acceptor atoms. This results in formation of three and even four hydrogen bonds between the molecules in some cases.

In most cases, out-of-plane amino group hydrogens become planar to facilitate bonding between planar molecules as this costs very little energy. In some other cases, the out-of-plane hydrogen of the amino group of one molecule in the pair facilitated an extra but rather weak binding, by exposing the acceptor atom it is attached to, to a hydrogen atom of the other molecule. In this case, the donor nature of the binding site changed so that the amino group became an acceptor site instead, but these two guanine homopairs did not relax to approximately planar pairs. By exploiting this idea even further, we have also looked at a possibility where the amino group acceptor binding site is used as one of the main binding sites. Five possibilities arise in this case, and three of them resulted in new stable configurations which were found to be extremely nonplanar. Note that the added stability of binding to a surface, however small, would most probably favor the planar homopairings.

We analyzed the interaction electron density for all pairs. We find the characteristic charge redistribution resembling a "kebab" structure that was previously observed in adenine pairs.[15] Stronger bonds show more charge redistribution than weaker bonds. The highest stability guanine pair is shown to have a total of four hydrogen bonds. However in comparison to the geometrically similar highest stability cytosine pairing, this guanine pairing has only an additional −0.13 eV stabilization energy. Interestingly this stable guanine pairing is not present in G-tetrads and was not suggested as a basis of superstructures on the gold(111) surface,[5] opposite to the dimer formation mechanisms[10] proposed in previous studies. Resonance assisted hydrogen bonding (RAHD) has been suggested in ref 5 to explain the observed guanine networks in scanning tunneling microscopy (STM) experiments.

The detailed information about cytosine and guanine pairs presented in this study should help in building up all gas-phase superstructures, for example, 2D guanine monolayers and 1D cytosine filaments observed in STM experiments,[5,12] in a similar way to that proposed earlier by us in ref 18 for constructing all adenine monolayers. The highest stability cytosine and guanine pairs are both centrosymmetric and are therefore most likely to initiate superstructure formation. However, one should remem-

ber that superstructures cannot be formed entirely by using the most favorable pairings, so that weaker pairings must also be considered to understand the various possibilities to produce 1D or 2D superstructures.

Acknowledgment. We would like to thank Roberto Otero, Maya Stöck, and Flemming Besenbacher for many useful discussions. We would like to acknowledge the computer time on the HPCx supercomputer provided via the Materials Chemistry Consortium and the technical support provided by the HPCx supercomputer helpdesk. R.E.A.K. is also grateful to the EPSRC for financial support (Grant GR/P01427/01).

References and Notes

(1) Srinivasan, R.; Murphy, J. C.; Fainchtein, R.; Pattabiraman, N. *Ultramicroscopy* **1991**, *312*, 293.
(2) Tao, N. J.; Shi, Z. *J. Phys. Chem.* **1994**, *98*, 1464.
(3) Tanaka, H.; Nakagawa, T.; Kawai, T. *Surf. Sci.* **1996**, *364*, L575.
(4) Kawai, T.; Tanaka, H.; Nakagawa, T. *Surf. Sci.* **1997**, *386*, 124.
(5) Otero, R.; Schöck, M.; Molina, L. M.; Laegsgaard, E.; Stensgaard, I.; Hammer, B.; Besenbacher, F. *Angew. Chem., Int. Ed.* **2005**, *44*, 2270.
(6) Sowerby, S. J.; Heckl, W. M.; Petersen, G. B. *J. Mol. Evol.* **1996**, *43*, 419.
(7) Freund, J. E.; Edelwirth, M.; Krobel, P.; Heckl, W. M. *Phys. Rev. B* **1997**, *55*, 5394.
(8) Furukawa, M.; Tanaka, H.; Kawai, T. *Surf. Sci.* **1997**, *392*, L33.
(9) Furukawa, M.; Tanaka, H.; Kawai, T. *Surf. Sci.* **2000**, *445*, 1.
(10) Furukawa, M.; Tanaka, H.; Kawai, T. *J. Chem. Phys.* **2001**, *115*, 3419.
(11) Chen, Q.; Frankel, D. J.; Richardson, N. V. *Langmuir* **2002**, *18*, 3219.
(12) Otero, R.; Schöck, M.; Besenbacher, F. Private communication.
(13) De Feyter, S.; De Schryver, C. *Chem. Soc. Rev.* **2003**, *32*, 139.
(14) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136.
(15) Kelly, R. E. A.; Lee, Y. J.; Kantorovich, L. N. *J. Phys. Chem. B* **2005**, *109*, 11933.
(16) Davis, J. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 668.
(17) Mergny, J. L.; Helene, C. *Nat. Med.* **1998**, *4*, 1366.
(18) Kelly, R. E. A.; Kantorovich, L. N. *Surf. Sci.* **2005**, *589*, 139.
(19) Ordejon, P.; Artacho, E.; Soler, J. M. *Phys. Rev. B* **1996**, *53*, R10441.
(20) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejon, P.; Sanchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745.
(21) Artacho, E.; Sanchez-Portal, D.; Ordejon, D.; Garcia, A.; Soler, J. M. *Phys. Status Solidi B* **1999**, *215*, 809.
(22) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(23) Preuss, M.; Schmidt, W. G.; Seino, K.; Furthmuller, J.; Bechstedt, F. *J. Comput. Chem.* **2004**, *25*, 112.
(24) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.
(25) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3.
(26) Weber, H. P.; Craven, B. M. *Acta Crystallogr.* **1990**, *B46*, 532.
(27) DeVoe, H.; Tinoco, I. J. *J. Mol. Biol.* **1962**, *4*, 500.
(28) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(29) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.
(30) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.
(31) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.
(32) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441.
(33) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem.—Eur. J.* **1999**, *5*, 3581.