# Electronic structure, stacking energy, partial charge, and hydrogen bonding in four periodic B-DNA models

Lokendra Poudel, Paul Rulis, Lei Liang, and W. Y. Ching
*Department of Physics and Astronomy, University of Missouri–Kansas City, Kansas City, Missouri 64110, USA*
(Received 4 May 2014; published 7 August 2014; corrected 8 August 2014)

We present a theoretical study of the electronic structure of four periodic B-DNA models labeled $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$ where A denotes adenine, T denotes thymine, G denotes guanine, and C denotes cytosine. Each model has ten base pairs with Na counterions to neutralize the negative phosphate group in the backbone. The $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models contain two and five AT-GC bilayers, respectively. When compared against the average of the two pure models, we estimate the AT-GC bilayer interaction energy to be 19.015 Kcal/mol, which is comparable to the hydrogen bonding energy between base pairs obtained from the literature. Our investigation shows that the stacking of base pairs plays a vital role in the electronic structure, relative stability, bonding, and distribution of partial charges in the DNA models. All four models show a highest occupied molecular orbital (HOMO) to lowest unoccupied molecular orbital (LUMO) gap ranging from 2.14 to 3.12 eV with HOMO states residing on the $PO_4$ + Na functional group and LUMO states originating from the bases. Our calculation implies that the electrical conductance of a DNA molecule should increase with increased base-pair mixing. Interatomic bonding effects in these models are investigated in detail by analyzing the distributions of the calculated bond order values for every pair of atoms in the four models including hydrogen bonding. The counterions significantly affect the gap width, the conductivity, and the distribution of partial charge on the DNA backbone. We also evaluate quantitatively the surface partial charge density on each functional group of the DNA models.

DOI: 10.1103/PhysRevE.90.022705 PACS number(s): 87.15.A−, 31.15.A−, 33.15.Fm

## I. INTRODUCTION

Deoxyribonucleic acid (DNA) is a macromolecule essential to all living species. It plays a pivotal role in biology as the carrier of genetic information [1]. The main structure of DNA is a double helix of simpler nucleobases [2]. Adenine (A), guanine (G), thymine (T), or cytosine (C) as well as alternating sugar (deoxyribose) and a phosphate group (phosphoric acid) connected by ester bond to form the DNA backbone. There are two types of complementary base pairings in DNA, A-T and G-C, which are stabilized by hydrogen bonds (HBs) between the base pairs and base-stacking interactions between the base layers. The A-T base pair contains two HBs and the G-C base pair has three HBs [3]. Although these HBs are weak, they contribute to the stability of the pairing and play a crucial role in coding genetic information, its transcription, and replication [4]. Beyond the simple base-pair interaction, other forms of interaction (e.g., between layers in the base-pair sequence) may have a significant effect on the higher-level structure of the DNA molecule.

Within the past two decades, research interest in DNA has extended beyond its biological relevance to include its potential application in molecular nanotechnology [5]. The electronic structure of DNA is of fundamental importance in both biological and materials science fields in order to understand subject matter such as DNA damage recognition and repair, DNA binding with proteins in biological cells, and electron transport through DNA [6]. It has been suggested that DNA or its derivatives may be used as a conducting molecular wire that is smaller and more efficient than those used in conventional silicon technology suitable for molecular electronic devices [7]. Currently, research and development in this area is still very primitive due to the lack of detailed information about electronic structure and interactions in different types of DNA structures.

The electronic structure of DNA has been investigated both experimentally [8–13] and theoretically [14–25] for quite some time. Scanning tunneling spectroscopy and direct measurements of electrical conductivity have been used to explore the electronic structure of DNA, but the results obtained were usually not consistent with each other due to differences in the specific experimental setups, the methods of sample preparation, and the ways that the experiments were conducted by different groups. Theoretically, the electronic structure of DNA models was predicted by using various computational methods including density-functional theory (DFT) [26,27] calculations, molecular mechanics, molecular dynamics, or combinations of different theoretical approaches. Results from these calculations also vary widely and in many cases are contradictory. This can be attributed partly to the different methods used in the calculations and variations in the underlying structural models (e.g., different numbers of base pairs, base-pair sequences [28], stacking heights [29], twisting angles, the presence or absence of counterions and water molecules, etc.). As may be expected, the electronic structure results were not consistent with each other and are still a hotly debated subject [30]. Nevertheless, the latest experimental and theoretical efforts have started to reach some limited consensus regarding issues such as the presence of a semiconducting band gap and the significant role played by counterions (if present) for electron transfer in DNA [31]. However, detailed quantitative information on the dependence on different DNA-structure models, base-pair stacking sequences and their relative stabilities, interatomic bonding (especially the hydrogen bonding), the exact role of

 

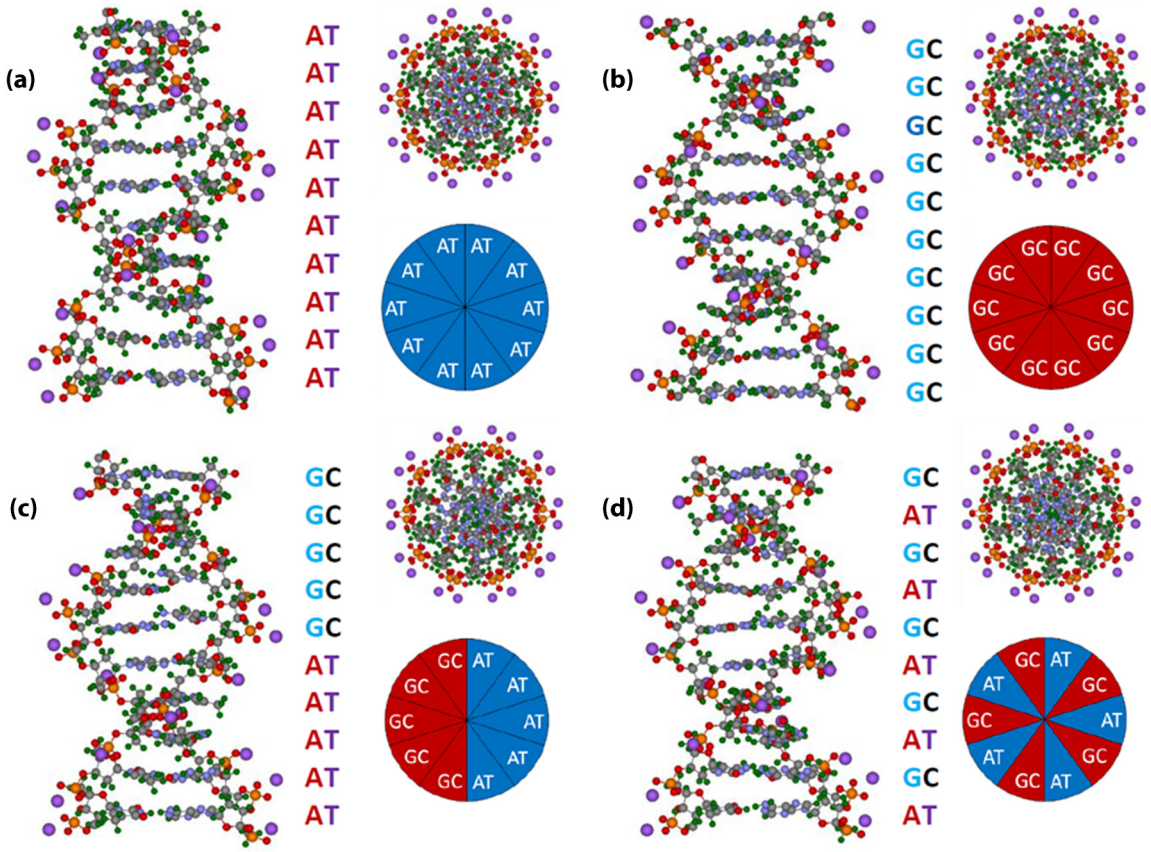


FIG. 1. (Color online) Axial ($z$ axis) and planer ($x$-$y$ plane) views of the four B-DNA models: (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, and (d) $(AT\text{-}GC)_5$.

counterions, charge transfer, and its implications on DNA in aqueous solutions, etc., remain to be explored.

The goal of this paper is to address the specific issues raised above. To this end, we use the following strategy. (i) We carefully construct four DNA models that are periodic along the helical $z$ axis, have different base pairs and stacking sequences, and are labeled $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$. (ii) Twenty sodium (Na) ions are added to the models to compensate for the twenty negatively charge $PO_4$ groups in the DNA backbone. (iii) The structures of these models are then fully relaxed with very high accuracy using the Vienna *ab initio* simulation package (VASP) [32,33]. (4) The electronic structure and bonding of these four relaxed models are calculated and analyzed by the *ab initio* orthogonalized linear combination of atomic orbitals (OLCAO) method [34]. (5) Finally, the surface partial charge density for each of the functional groups (A, T, G, C, sugar, and $PO_4$ + Na) is calculated and dissected to give insight regarding their dependence on the variations in the DNA structure. The use of the same strategy and computational methods for four different models enables us to get more meaningful quantitative conclusions since any relative errors due to unavoidable assumptions or approximation introduced are same. The presentation of this paper is outlined as follows. In the next section we briefly describe the construction of the four DNA models. This is followed by the description of the computational methods used in Sec. III. The main results and a discussion thereof are presented in Sec. IV. We end with a summary and conclusions in Sec. V, which includes plans for further applications.

## II. B-DNA MODELS

DNA exists in several conformations such as A-DNA, B-DNA, and Z-DNA, each with different helical structures, twist angles, stacking heights, and helix diameters [35]. B-DNA is the most common and well-studied structure because it is the form that is most commonly found in a living cell. In nature, B-DNA is a double-stranded helical molecule that is very long and may contain billions of base pairs [36]. Therefore, periodic models are more realistic for describing the structure of B-DNA and are more amenable to accurate quantum mechanical calculation. The structure of B-DNA in its optimized state circles the axis of the double helix every 10.4 base pairs [37], so a periodicity of ten base pairs contains very little unrealistic distortion while keeping the model size reasonable. Also, the periodic model avoids the need to terminate the B-DNA strand with a phosphate group or sugar group (5′ and 3′, respectively) at the ends. We have built four different ten-base-pair B-DNA models with periodicity in the axial direction ($z$ axis) and different stacking sequence. They are labeled $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$ and are shown in Figs. 1(a)–1(d), respectively. A more detailed description of these four models is presented in the Appendix.

## III. COMPUTATIONAL METHODS

We have used two *ab initio* quantum mechanical methods to study the structure and properties of the B-DNA models. VASP was used to relax the structures and the OLCAO method was used to calculate the electronic and bonding properties. VASP is based on DFT and has been highly successful for atomic relaxation and geometric optimization and we used it purely for the relaxation of the structures. In the present study we used the projector augmented wave method with the Perdew-Burke-Ernzerhof potential [38] for exchange correlation functional within the generalized gradient approximation. For electronic relaxation, a relatively-high-energy cutoff of 500 eV is adopted with the electronic convergence criterion set at $10^{-5}$ eV. For ionic relaxation, we set the force convergence criterion to be $10^{-3}$ eV/Å. Since a large periodic supercell is used in the calculation, we used one $k$ point at the zone center for a single-point calculation, which is more than sufficient for a large biomolecule such as DNA. All VASP calculations were carried out on Edison at the National Energy Research Scientific Computing (NERSC) facility at Lawrence Berkeley Laboratory.

The electronic structure of the relaxed DNA models was calculated using the OLCAO method, which is an all-electron method based on the local-density approximation of DFT. It is extremely efficient and versatile for the calculation of electronic structures, especially for large and complex systems due to the flexible choice of the basis set. This method has been successfully employed in the study of many complex systems such as inorganic [39] and organic materials [40] and biomolecules [41–43]. In the present calculation, a full basis, which consists of the core orbitals, occupied valence orbitals, and the next empty shell of unoccupied orbitals for each atom, was used for the determination of the self-consistent potential and calculations of the density of states. A minimal basis was used for the separate calculation of the partial charge and bond orders. The reader is referred to Ref. [34] for a complete description of the form of the basis set in the OLCAO method. The total density of states (TDOS) is obtained from the energy eigenvalues after the solution of the final Kohn-Sham equation. The TDOS is further resolved into a partial density of states (PDOS) for each functional group in the DNA models. The most important physical quantity in the electronic structure is the band gap or the highest occupied molecular orbital (HOMO) to lowest unoccupied molecular orbital (LUMO) gap. The effective charge $Q^*$ (in units of electrons) on each atom in the models is calculated according to the Mulliken population analysis [44] to provide information on the nature of the bonding and charge transfer between atoms. The effective charges are calculated from the formula

$$Q^*_\alpha = \sum_{i,\alpha} \sum_{occ} \sum_{j,\beta} C^{*n}_{i\alpha} C^n_{j\beta} S_{i\alpha,j\beta}, \tag{1}$$

where the $C^n_{j\beta}$ are the eigenvector coefficients of the $n$th band, $j$th orbital, and $\beta$th atom. The $S_{i\alpha,j\beta}$ are the overlap integrals between the $i$th orbital of the $\alpha$th atom and $j$th orbital of the $\beta$th atom. Also calculated are the bond order (BO) values $\rho_{\alpha\beta}$ for every pair of atoms including the HBs. The bond order values give a quantitative measure of the strength of the bonds, which generally scales with the bond length (BL), but also depends on the local environment of the bonding atoms. The BO values for each pair of atoms $\alpha$ and $\beta$ are calculated according to

$$\rho_{\alpha\beta} = \sum_{nocc} \sum_{i,j} C^{*n}_{i\alpha} C^n_{j\beta} S_{i\alpha,j\beta}. \tag{2}$$

It should be mentioned that there are many different ways to define bond order or overlap populations. We believe that the use of Mulliken analysis is the most effective one for large complex systems. For example, the Bader method depends on the topological analysis of local geometry and the use of numerical sampling to evaluate the BO between different pairs of atoms. It would be almost impossible to apply to systems such as the B-DNA models presented in this paper with different types of bonds and extremely complex structure.

## IV. RESULTS AND DISCUSSION

Before we describe our results, we would like point out that our calculations are based on the relaxed structure at zero temperature and contain no dynamic effects for the random motion of the counterions that is always present in real biomolecular system. For example, Lee *et al.* [45] performed classical molecular dynamic simulation using the AMBER99 force field with the AMBER10 package (where AMBER denotes assisted model building with energy refinement) [46]. The electronic structure of dsDNA was obtained using the fragment orbital approach. [A fragment consists with of a pair of nucleotides (phosphate-deoxyribose nucleobase) from the DNA strand.] They have observed that the molecular orbitals of nucelotide fragments are intermittently switching from bases to the backbone over the simulation time. The HOMO can temporarily put a large weight on the backbones as a function of time. This finding is in line with our results that the HOMO is located on phosphate groups (the backbone of DNA), to be discussed below.

### A. Relaxed structure and stacking energy

The calculated total energies (TEs) for the four B-DNA models are listed in Table I. The $(AT\text{-}CG)_5$ model has a lower TE than $(AT)_5(GC)_5$, which implies that an increase in the number of alternating AT-GC base pairs increases the stability due to enhanced base-pair interactions. Using the average TE of the $(AT)_{10}$ and $(GC)_{10}$ models as reference energies for models with no stacking disorder, we can estimate the bilayer stacking energy. The TE of the $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models are lower than the average TE of $(AT)_{10}$ and $(GC)_{10}$ by −1.7603 eV (–40.5942 Kcal/mol) and −3.8448 eV (–88.6649 Kcal/mol), respectively. Assuming that the lowering of energies is due to the interaction between stacked AT-GC bases in the models, the $(AT)_5(GC)_5$ model with two bilayers of AT-GC gives a stacking energy per stack of 20.297 Kcal/mol. Similarly, the $(AT\text{-}GC)_5$ with five bilayers gives a stacking energy per stack of 17.733 Kcal/mol. The average of the two estimations gives a stacking energy of 19.015 Kcal/mol, quite close to previous calculations reported in the literature that the stacking energies range from 13.5 to 18.2 Kcal/mol [47]. Our stacking energy is higher than the other reported value because our calculation includes counterions and sugar but the other calculation used an isolated base-pair stack. We believe that

TABLE I. Total energy and total bond order for B-DNA models.

| Model | No. of atoms | No. of valence electrons | Total energy (eV) | Bond order calculations | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | TBO | CBO | HBO | Na-O BO | NNN BO (<2.6 Å) |
| $(AT)_{10}$ | 660 | 2220 | –4251.552 | 278.13 | 269.75 | 1.07 | 3.41 | 0.044 |
| $(GC)_{10}$ | 650 | 2220 | –4210.215 | 273.67 | 264.86 | 1.46 | 3.44 | 0.053 |
| $(AT)_5(GC)_5$ | 655 | 2220 | –4232.643 | 275.91 | 267.45 | 1.05 | 3.58 | 0.063 |
| $(AT\text{-}GC)_5$ | 655 | 2220 | –4234.728 | 276.46 | 267.59 | 0.99 | 3.98 | 0.095 |

the measured stacking energy will be highly dependent on the B-DNA structure and its local environment (e.g., stacking height, twisting angle, counterions, and humidity). Our calculated stacking energy is of the same order of magnitude as the hydrogen bonding energies between nucleobases (15 Kcal/mol for AT and 27.5 Kcal/mol for GC) [48]. Our results indicate that the stacking interaction plays a substantial role in the stability and functionality of the DNA. We will return to this point later in the next section when we show that this interpretation is further supported by the BO analysis in the four models.

### B. Total and partial density of states

The calculated TDOS for the four DNA models using the OLCAO method in the energy range from −25 to 25 eV (left column) and from −1.0 to 6 eV (right column) are shown in Fig. 2. The calculated HOMO-LUMO gaps for the four B-DNA models $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$ are 3.12, 2.73, 2.52, and 2.14 eV, respectively. The band gaps of $(AT)_{10}$ and $(GC)_{10}$ agree with the gap values reported by other recent experimental [8,9] and theoretical work [20,22], while for the $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models, we found no other reported values with which to compare the present results. It should be noted that the calculated values from DFT generally underestimate the band gap and also depend on the methodology and potential used. On the other hand, experimentally quoted values also have uncertainty depending on the actual sample used and the nature of the experiment conducted. Porath *et al.* [8] used the electrostatic trapping method for the conductance of homogeneous sequence of 30 base pairs of [poly(G)-poly(C)] DNA. They found that the band gap of a short DNA sequence to be about 2.1 eV. Shapir *et al.* [9] studied long single poly(G)-poly(C) DNA molecules deposited on gold using scanning tunneling microscopy. They also found that the band gap of DNA was about 2.5 eV. These experimental band gap lies in the range of our finding. From these results, we can surmise that the electrical conductivity of DNA at finite temperature and under an applied voltage with only GC base pairs will be higher than B-DNA with only AT base pairs and that the electrical conductivity of B-DNA with mixed base pairs will be even higher than B-DNA with pure AT or GC base pairs due to the smaller band gaps. Also, the electrical conductivity should increase with increased mixing of the base pairs. The overall features of the TDOS for the four models are very similar. The most revealing differences between the models are found in the unoccupied region close to the LUMO and the top of the occupied states close to the HOMO. These states are far more sensitive to the structural differences between the four models as shown in the right column of Fig. 2. To better trace the origin of these states, we resolve the TDOS into a PDOS according the functional groups in the B-DNA models. This is particularly effective with the OLCAO method where the wave functions are expanded in terms of atomic orbitals centered on each atom. The PDOS for each functional group can be easily obtained by adding the atomic PDOS from atoms within each group. The functional-group-resolved PDOSs for the four B-DNA models are shown in Fig. 3. The following facts are observed. (i) The HOMO states in all four models comes predominately from the phosphate group, which is at variance with some of the existing calculations in the literature. We believe this is due to the proximity of the Na counterions to the $PO_4$ group, which form rather strong Na-O bonds with O atoms from the $PO_4$ group. (ii) All of the lower conduction band states up to 5 eV are from base pairs. For $(AT)_{10}$, there are five peaks below 5 eV from thymine and adenine. The Na peak is slightly above 5.0 eV. Similarly, for the $(GC)_{10}$ model, there are two peaks from cytosine and three peaks from guanine. (iii) Moving to the stacked $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models, the peak structure and distributions are more complicated with all nucleobases A, T, G, and C participating. Their positions and heights differ from those in the $(AT)_{10}$ and $(GC)_{10}$ models, reflecting the complex interactions occurring when the four bases interact in the stacked models. (iv) The sharp peaks at 5.38, 5.31, 5.11, and 4.82 eV for the $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$ models, respectively, are exclusively from the Na ion and a smaller peak at the same location from the $PO_4$ group indicates that the $PO_4$ + Na should be considered as a single functional group. It should be pointed out that the shifts of the Na peaks in the four models signify that the accurately relaxed B-DNA structures reflect the subtle changes in atomic positions of the base-pair atoms in the four different models. (v) There is no significant participation of states below or near 5 eV from the sugar group. This simply reflects the fact that the antibonding states from the strong covalently bonded sugar units in the DNA backbone are at a much higher energy. Hence, it is believed that electron transport in B-DNA makes use of both the base pair and the B-DNA backbone. Although, this finding contradicts some previous results [22], the counterions were absent in their models. They found that the HOMO-LUMO gap resides only at the base pair and not the backbone. It is obvious that the counterions play a pivotal role for electron transfer in DNA. This conclusion is important to the understanding of the conductivity properties of DNA.

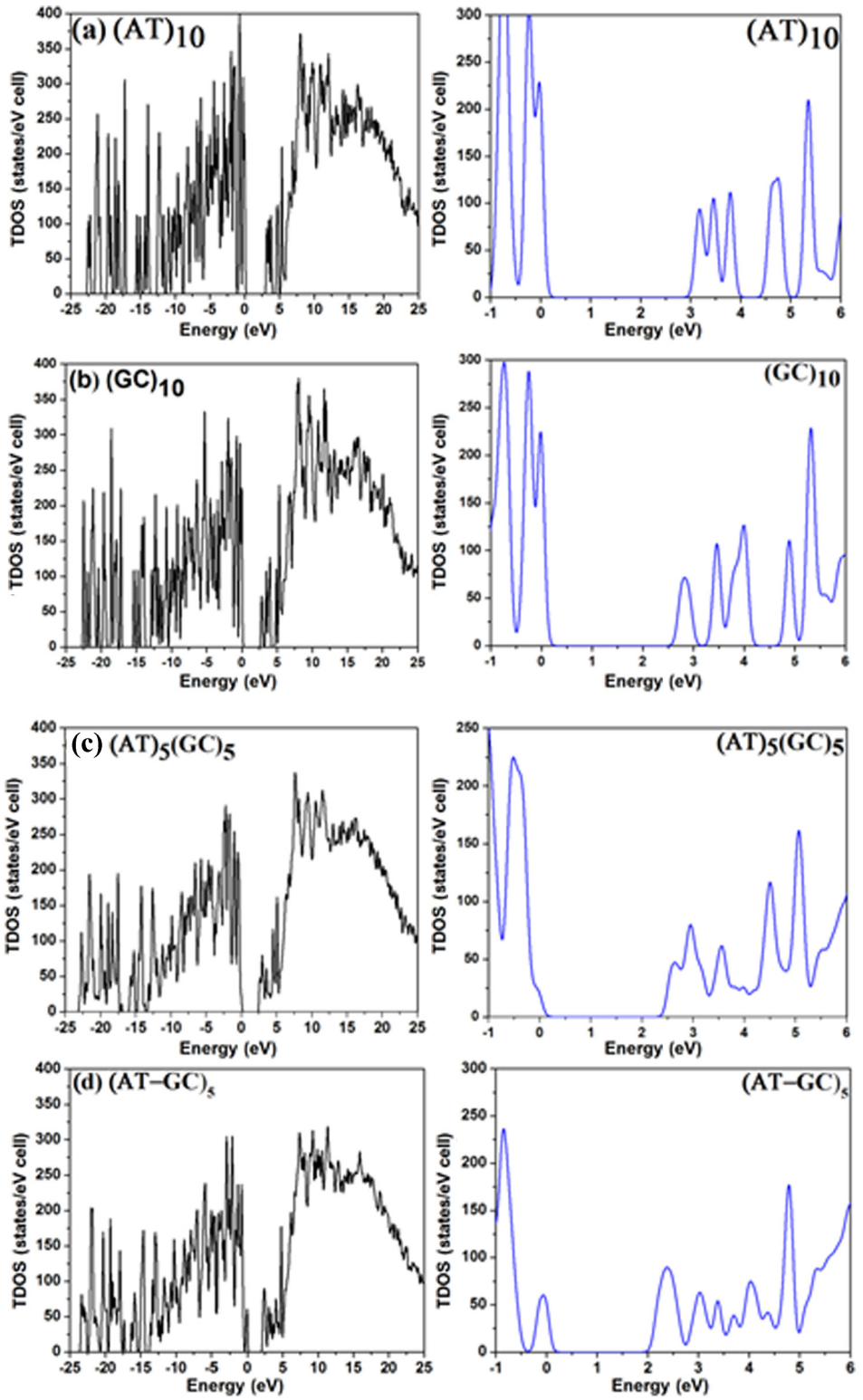


FIG. 2. (Color online) The TDOS for the four B-DNA models in the energy range from −25 to 25 eV (left panels) and −1 to 6 eV (right panels): (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, and (d) $(AT\text{-}GC)_5$.

### C. Bond order and hydrogen bonding

We have calculated the BO between all pairs of atoms in the four B-DNA models. These BO values are divided into four groups. The strong covalently bonded pairs, the HBs between base pairs, the fairly strong Na-O bonding from the counterions, and the weak but not entirely negligible bonding from the next-nearest-neighbor (NNN) atoms with BLs less than 2.6 Å. Table I lists the sum of all BOs, which we call the total bond order (TBO), and its breakdown into the above four groups. The average of the TBO of the $(AT)_{10}$ and $(GC)_{10}$

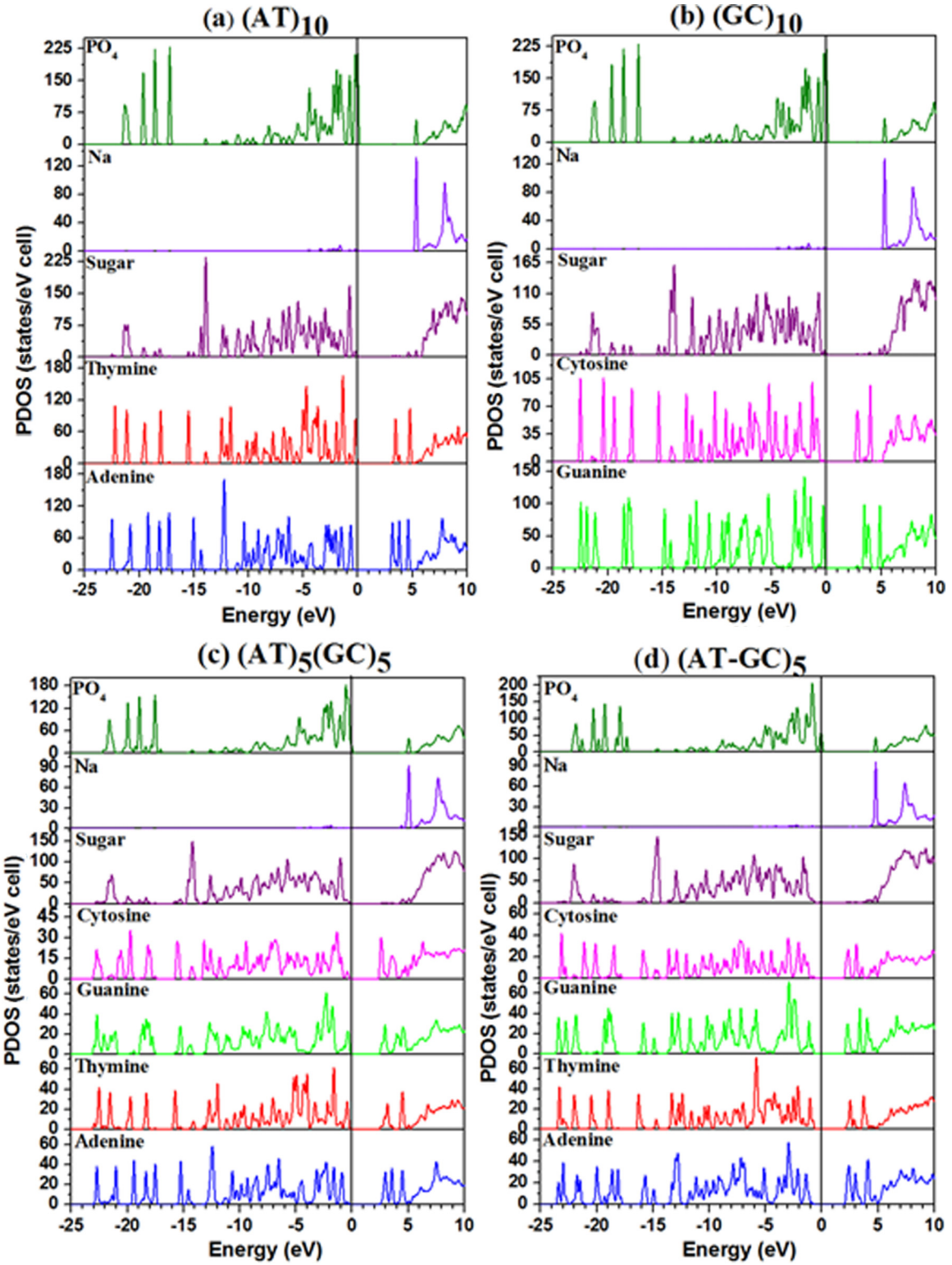


FIG. 3. (Color online) The PDOS of functional groups in four models: (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, and (d) $(AT\text{-}GC)_5$.

models is 275.90 and that value can serve as the reference value to compare with the TBO of the $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models. On the basis of the TBO values, the $(AT\text{-}GC)_5$ model is more stable than the $(AT)_5(GC)_5$ model, in agreement with the total energy calculation. The TBO was then subdivided into contributions from different types of bonds on the basis of bond lengths and bond order values, including the covalent bond order (CBO), the hydrogen bond order (HBO), the Na-O bond order, and the next-nearest-neighbor bond order. Table I shows that the TBO contribution from the CBO for

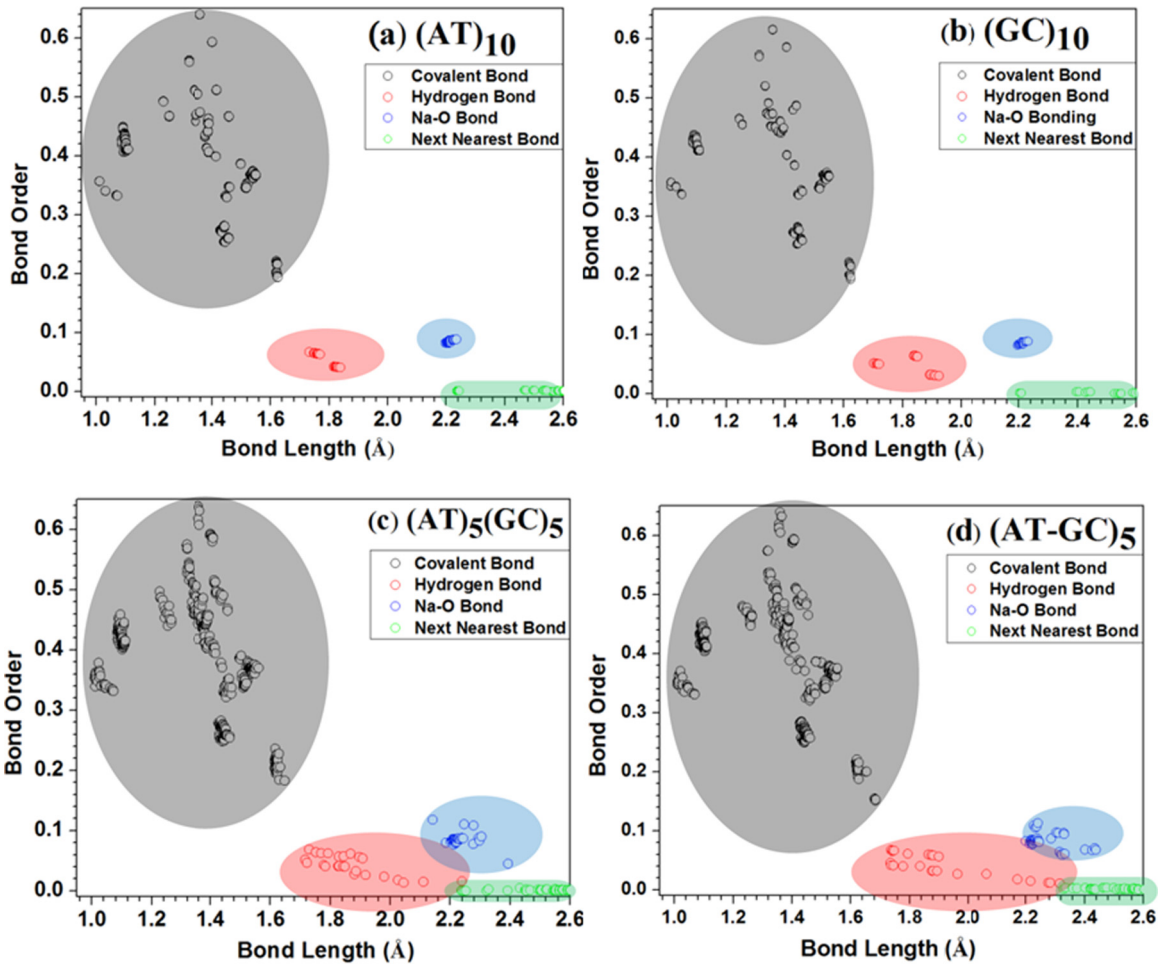


FIG. 4. (Color online) Bond length vs bond order values of covalent bonds (gray), HBs (pink), Na-O bonds (blue), and NNN bonds less than 2.6 Å (green): (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, and (d) $(AT\text{-}GC)_5$.

$(AT)_5(GC)_5$ is 267.45 and for $(AT\text{-}GC)_5$ it is 267.59, which are higher than the average values of the CBOs of the $(AT)_{10}$ and $(GC)_{10}$ models. The BOs of Na-O and NNN follow the same trend as the CBO, while the HBO is opposite. However, they are all relatively small compared to the CBO and so contribute only weakly to the TBO. From Table I we have found that the NNN BO values are very small, but those from the Na-O BO are actually larger than that from the HB, indicating that the counterion effect can be quite substantial in determining the stability of different models and in affecting the electron states (peaks in the PDOS).

In order to have a more detailed picture of the interatomic bonding in the four B-DNA models, we display in Fig. 4 the distribution of all BO pairs vs the BLs in four models with the four types of BOs depicted with different colors. The strong covalent bonds have BLs that range from 1.0 to 1.64 Å, the HB BLs range from 1.7 to 2.3 Å, and the Na-O BLs range from 2.2 to 2.4 Å with an increased range in the stacked models, $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$, versus the pure models. The following can be seen. (i) The covalently bonded pairs are very strong and have a larger distribution in BO values due to the presence of different types of covalent bonds in the backbone and base pairs. (ii) The HB contributions are substantial. The distribution is scattered from two HB per base pair in $(AT)_{10}$ and three HB per base pair in $(GC)_{10}$ to much

TABLE II. Bond length and bond order values for hydrogen bonds in four B-DNA models.

| Model | H bond | Average BL (Å) | Average BO |
|---|---|---|---|
| $(AT)_{10}$ | $H_A$-$O_T$[a] | 1.825 | 0.042 |
| | $N_A$-$H_T$ | 1.755 | 0.065 |
| $(GC)_{10}$ | $H_G$-$O_C$ | 1.906 | 0.031 |
| | $H_G$-$N_C$ | 1.844 | 0.064 |
| | $O_G$-$H_C$ | 1.713 | 0.051 |
| $(AT)_5(GC)_5$ | $H_A$-$O_T$ | 1.846 | 0.039 |
| | $N_A$-$H_T$ | 1.774 | 0.063 |
| | $H_G$-$O_C$ | 2.036 | 0.020 |
| | $H_G$-$N_C$ | 1.879 | 0.057 |
| | $O_G$-$H_C$ | 1.750 | 0.046 |
| $(AT\text{-}GC)_5$ | $H_A$-$O_T$ | 1.889 | 0.033 |
| | $N_A$-$H_T$ | 1.752 | 0.066 |
| | $H_G$-$O_C$ | 2.282 | 0.012 |
| | $H_G$-$N_C$ | 1.884 | 0.058 |
| | $O_G$-$H_C$ | 1.748 | 0.034 |

[a]The subscript indicates the corresponding base of the atom.

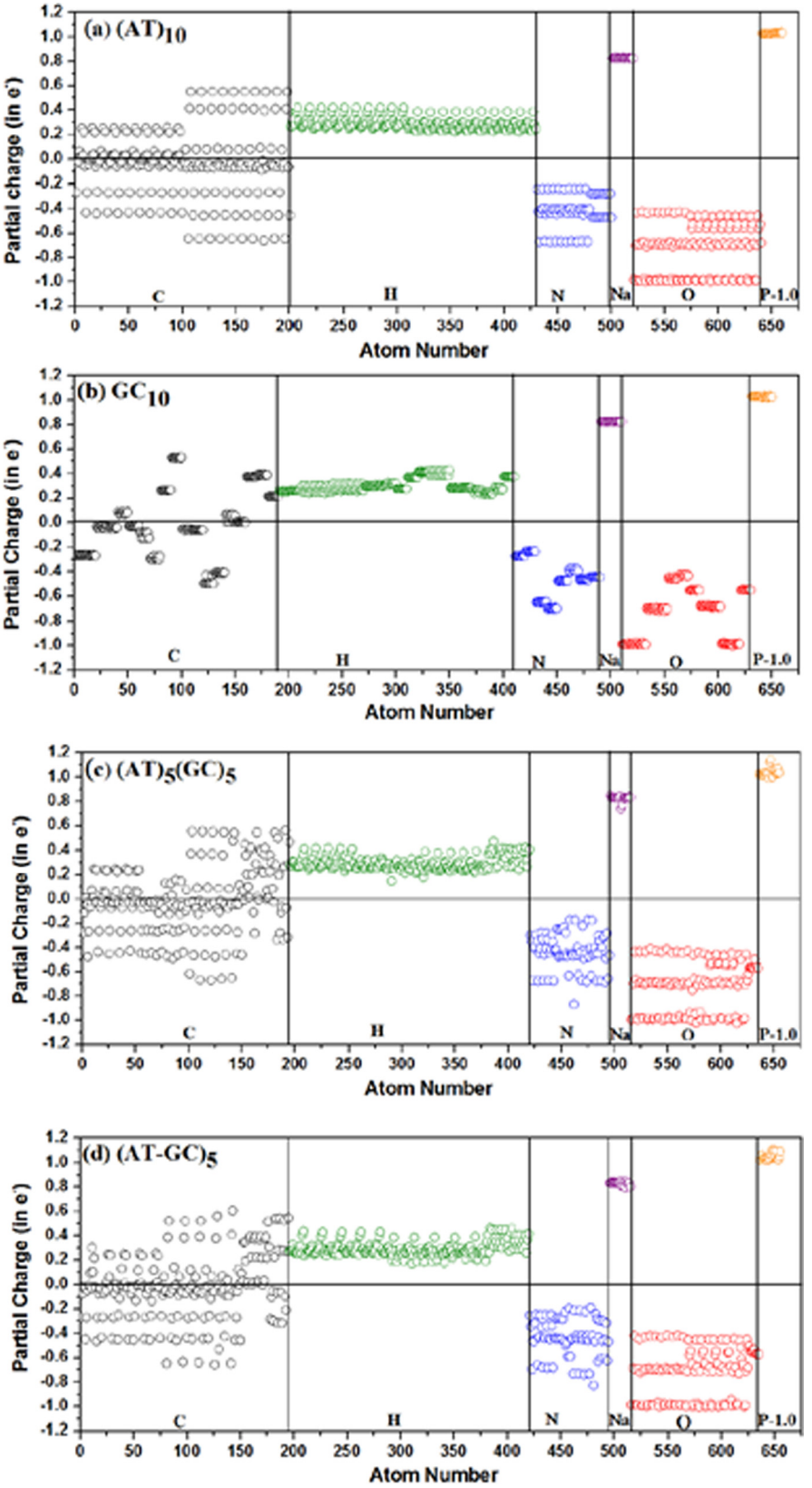


FIG. 5. (Color online) Calculated atomic partial charges on all atoms in the four B-DNA models: (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, and (d) $(AT\text{-}GC)_5$. For P, the plotted data $\Delta Q$ are reduced by one electron in order to have a clearer display.

TABLE III. Average net partial charge and surface partial charge densities on functional groups of four B-DNA models.

| Model | Functional group | Average $\Delta Q$ ($e^-$) | Surface charge density ($e^-$/nm$^2$) |
|---|---|---|---|
| $(AT)_{10}$ | adenine | –0.192 | –0.245 |
| | thymine | –0.252 | –0.364 |
| | sugar | 0.715 | 1.034 |
| | $PO_4$ + Na | –0.494 | –1.559 |
| $(GC)_{10}$ | guanine | –0.279 | –0.355 |
| | cytosine | –0.162 | –0.212 |
| | sugar | 0.715 | 1.029 |
| | $PO_4$ + Na | –0.494 | –1.574 |
| $(AT)_5(GC)_5$ | adenine | –0.192 | –0.248 |
| | thymine | –0.247 | –0.312 |
| | guanine | –0.261 | –0.358 |
| | cytosine | –0.210 | –0.266 |
| | sugar | 0.720 | 1.102 |
| | $PO_4$ + Na | –0.493 | –1.761 |
| $(AT\text{-}GC)_5$ | adenine | –0.181 | –0.238 |
| | thymine | –0.286 | –0.376 |
| | guanine | –0.289 | –0.473 |
| | cytosine | –0.117 | –0.145 |
| | sugar | 0.699 | 1.004 |
| | $PO_4$ + Na | –0.481 | –1.733 |

wider distributions in the stacked models with a concomitant decrease in BO. (iii) Although the Na-O BLs are larger than the BLs of the HBs, their BO values are actually slightly larger. They are also more dispersed in the $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models. (iv) As pointed out before, the bonding from the NNN pairs is small but not totally negligible.

In Table II we list the average BL and BO values for each specific HB between the base pairs in the four models. Table II shows that N-H bonding is stronger than O-H bonding in both AT and GC base pairs for all DNA models. The N-H bonding of the AT base pair is stronger than that of the GC base pair. In the GC base pair, H-O bonding with H at C and O at G is stronger than H at G and O at C for all of our DNA models. For the $(AT)_{10}$ and $(GC)_{10}$ models, these values agree in general with those reported in the literature [49]. For $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models, our results show that the H bonding changes considerably.

### D. Partial charge distribution on functional groups

The partial charge distribution on biological macromolecules such as proteins and peptides is an important segment of biophysical research because of its implications on long-range electrostatic and polar interactions. However, they are usually presented in a rudimentary manner (positive, negative, or neutral) without any quantitative measures. We demonstrate that we can obtain quantitative values for surface partial charges on each of the functional groups in B-DNA models, which provide additional insight into these important biomolecules that is so far missing. We start with the calculation of atomic partial charges on every atom in the four B-DNA models that are displayed in Fig. 5. The atomic partial charge is the deviation of the effective charge $Q^*$ of Eq. (1) from the charge on the neutral atom $Q_0$ in units of electrons, or $\Delta Q = (Q^* - Q_0)$ (i.e., $-\Delta Q$ is equal to the gain of an electron and $\Delta Q$ the loss of an electron). Figure 5 lists $\Delta Q$ data for every atom (C, H, N, Na, O, and P) in the four models and contains a wealth of information that corroborates the other electronic structure results presented earlier. (In Fig. 5 the plot of $Q^*$ for P is reduced by one electron in order to have a clearer display.) In all four models, H, Na, and P always gain charge and N and O always lose charge, whereas C can either gain or lose charge depending on its local bonding characteristics in the structure. The distribution of the partial charges for each type of atom is most regular in $(AT)_{10}$ and slightly more distributed in $(GC)_{10}$, reflecting the fact that the atomic scale structure in $(GC)_{10}$ is slightly more complicated than $(AT)_{10}$ due to different types of base pairs and the number of HBs. Further, the distributions in the two stacked models $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ are far more dispersed. This is more evident in the case for H, which plays a key role in hydrogen bonding as more bilayers of AT-GC are created. The increased variations in $\Delta Q$ for Na and P in the stacked models are also obvious.

By adding the $\Delta Q$ values for all of the atoms within each of the functional groups and dividing by the solvent excluded surface area for these groups, we can have a more vivid picture as to how the surface partial charge density looks like as a function of their structures obtained from realistic quantum calculations. For that purpose, Na + $PO_4$ is considered as a single unit. The calculated values of the surface partial charge

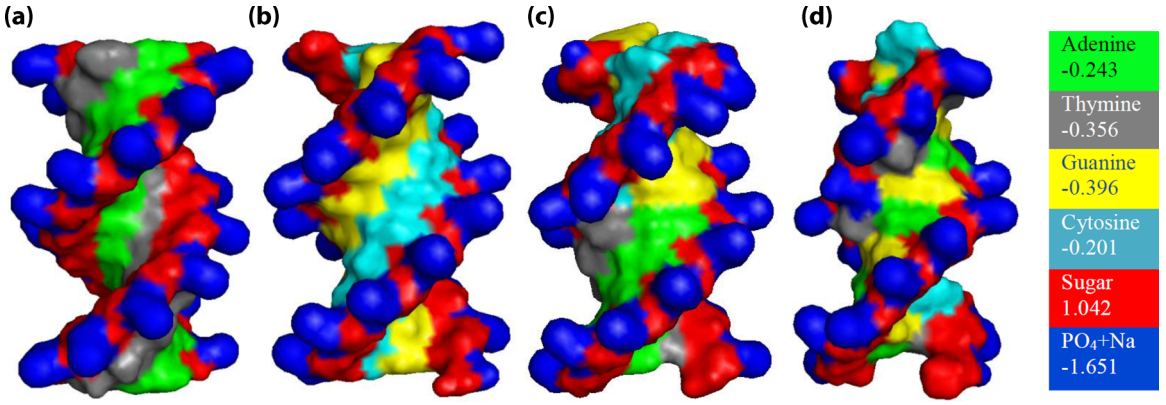


FIG. 6. (Color online) Partial charge density in the solvent excluded surface of the four B-DNA models: (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, and (d) $(AT\text{-}GC)_5$.

density in the unit of electrons/(nm)$^2$ for the four models are listed in Table III. Figure 6 shows a color-coded plot of the solvent excluded surface for each functional group as is commonly used for biomolecular systems. It can be seen that sugar is highly positively charged and that all of the base pairs and Na-compensated $PO_4$ units are negatively charged. It is interesting to note that although the changes in the A, T, G, and C surface partial charge densities in the four models are small and negative, it is still clear that, in magnitude, G>T>A>C. This relative order depends on the calculated $\Delta Q$ values in Table III and the actual areas for each group. We believe that this information could be very useful for quantitative evaluation of electrostatic forces involving B-DNA.

## V. CONCLUSION

We have built four periodic B-DNA models $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$ and calculated their electronic structures and the AT-GC base pair stacking energy in B-DNA to be 19.015 Kcal/mol. Our results show that stacking of different base pairs plays a significant role in the stability and electronic properties of B-DNA with stacking energy comparable to the HB energy between base pairs. Our calculation also shows that the stacking interaction affects the interatomic bonding as revealed by the change in bond order distributions. It is concluded that AT-GC base-pair stacking will enhance the electrical conductivity of B-DNA and should increase with increasing number of AT-GC base pairs. All four periodic B-DNA models have semiconducting band gaps. The calculated TDOS and functional-group-resolved PDOS revealed that the HOMO states originate from Na-compensated phosphate groups and the LUMO and the lower excited states are all from the base pairs. Furthermore, we are able to evaluate quantitatively the surface partial charge density on each functional group in the four B-DNA models from the effective charges of the individual atoms in the model. The base pairs and phosphates with counterions are always electronegative, but sugar is always electropositive.

The work presented in this paper paves the way for further investigations including modeling of the solvent effect by including water molecules in the simulation box and the creation of models with defect layers or abnormal base pairing to mimic those found in defective genes and their implications in relation to rational drug design. Based on the electronic structures and the *ab initio* wave functions already obtained, optical properties of these B-DNA models can be calculated and used to estimate the long-range van der Waal–London interactions based on Lifshitz theory [50] with possible applications to mesoscale nanotechnology. The computational techniques and capabilities demonstrated in this paper can be extended to more complex DNA structures such as triplex [51] and quadruplex [52] DNA.

## ACKNOWLEDGMENTS

This work was supported by the U.S. DOE, Office of BES, Division of Materials Science and Engineering under Grant No. DE-SC008176. This research used the resources of NERSC supported by the Office of Science of the DOE under Contract No. DE-AC03-76SF00098. In the very early stage of research on modeling the $(AT)_{10}$ and $(GC)_{10}$ models, the work was partially supported by Grant No. DEFG-0284DR45170.

## APPENDIX: CONSTRUCTION OF THE B-DNA MODELS

The $(AT)_{10}$ and $(GC)_{10}$ DNA models were created by the using NAB (which denotes nucleic acid builder) as a part of the tools in the AMBER program. The $(AT)_{10}$ model contains only the A and T nucleobases and the $(GC)_{10}$ model contains only the G and C nucleobases. Because the phosphate group is negatively charged we added 20 sodium counterions ($Na^+$) near the phosphate groups to promote a normal charge distribution as might be found in solution (although no $H_2O$ molecules were added). There are six parameters that can be used to describe the structure of the model: twist angle ($\theta$), stacking height ($d$), $x$ and $y$ shifts in the plane perpendicular to the axis of the helix, and tilt and roll angles. Because our model is periodic in the axial direction, the number of free parameters is reduced. Initially, our DNA models have a twist angle of 36°, a stacking distance of 3.38 Å, and a radius in the $x$-$y$ plane of 10 Å, which are the parameters of the conventional B-DNA structure. The periodic DNA model was placed in a rectangular box of $30 \times 30 \times 33.8$ Å$^3$. After fully relaxing the atomic positions and varying the stacking distances, we obtained minimum-energy $(AT)_{10}$ and $(GC)_{10}$ models at a stacking distance of 3.3785 Å, so the lattice parameters of our simulation cell were set to $a = 30$ Å, $b = 30$ Å, $c = 37.856$ Å, and $\alpha = \beta = \gamma = 90°$. The simulation box in the $x$ and $y$ directions is sufficiently large to avoid any interactions between the B-DNA and its neighbors in the replicated periodic cells. There is a total of 660 atoms in the $(AT)_{10}$ model and 650 atoms in the $(GC)_{10}$ model. Figures 1(a) and 1(b) show, respectively, the axial ($z$ axis) and planar ($x$-$y$ plane) views of these two models.

The $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models with mixed base pairs were produced with the help of the $(AT)_{10}$ and $(GC)_{10}$ models. The $(AT)_5(GC)_5$ model is built from five base pairs of the $(AT)_{10}$ model and another five base pairs of the $(GC)_{10}$ model. We removed five base pairs from the $(AT)_{10}$ model and replaced them with five base pairs from the $(GC)_{10}$ mode with appropriate translation and rotation of atomic coordinates. Extra care was taken to ensure the correct rotation of base pairs in the $x$-$y$ plane and to preserve periodicity along the $z$ axis. The construction of the $(AT\text{-}GC)_5$ model is similar to that of the $(AT)_5(GC)_5$ model except with greater complexity. We alternately kept and removed an AT base pair from the $(AT)_{10}$ model and then we inserted the corresponding GC base pairs from the $(GC)_{10}$ with appropriate translation and rotation of coordinates of each of the A-T and G-C base pairs. Both $(AT)_5(GC)_5$ and $(AT\text{-}GC)_5$ models were placed in a simulation box of the same size as the $(AT)_{10}$ and $(GC)_{10}$ models and fully relaxed using VASP. It is noted that these relaxations take a far longer time to reach the required convergence than the $(AT)_{10}$ and $(GC)_{10}$ models because of the structural distortions introduced in the stacking models. Figures 1(c) and 1(d) show, respectively, the axial ($z$ axis) and plane ($x$-$y$ plane) views of these two models. Because of the periodicity, we can clearly see that $(AT)_5(GC)_5$ has two (AT-GC) bilayers, whereas $(AT\text{-}GC)_5$ has five such bilayers.

[1] V. A. Bloomfield, D. M. Crothers, and I. Tinoco Herndon, *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, 1996).
[2] F. H. Crick and J. D. Watson, Proc. R. Soc. London Ser. A **223**, 80 (1954).
[3] E. T. Kool, Annu. Rev. Biophys. Biomol. Struct. **30**, 1 (2001).
[4] G. A. Jeffrey, *An Introduction to Hydrogen Bonding* (Oxford University Press, New York, 1997).
[5] N. C. Seeman, Mol. Biotechnol. **37**, 246 (2007).
[6] D. C. Young, *Computational Drug Design: A Guide for Computational and Medicinal Chemists* (Wiley, New York, 2009).
[7] A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan, Nat. Nanotechnol. **6**, 763 (2011).
[8] D. Porath, A. Bezryadin, S. De Vries, and C. Dekker, Nature (London) **403**, 635 (2000).
[9] E. Shapir, H. Cohen, A. Calzolari, C. Cavazzoni, D. A. Ryndyk, G. Cuniberti, A. Kotlyar, R. Di Felice, and D. Porath, Nat. Mater. **7**, 68 (2008).
[10] H. Wadati, K. Okazaki, Y. Niimi, A. Fujimori, H. Tabata, J. Pikus, and J. P. Lewis, Appl. Phys. Lett. **86**, 023901 (2005).
[11] K. Kummer, D. V. Vyalikh, G. Gavrila, A. B. Preobrajenski, A. Kick, M. Bönsch, M. Mertig, and S. L. Molodtsov, J. Phys. Chem. B **114**, 9645 (2010).
[12] M. Xu, S. Tsukamoto, S. Ishida, M. Kitamura, Y. Arakawa, R. Endres, and M. Shimoda, Appl. Phys. Lett. **87**, 083902 (2005).
[13] L. A. Jauregui, K. Salazar-Salinas, and J. M. Seminario, J. Phys. Chem. B **113**, 6230 (2009).
[14] P. J. de Pablo, F. Moreno-Herrero, J. Colchero, J. Gómez Herrero, P. Herrero, A. M. Baró, P. Ordejón, J. M. Soler, and E. Artacho, Phys. Rev. Lett. **85**, 4992 (2000).
[15] F. L. Gervasio, P. Carloni, and M. Parrinello, Phys. Rev. Lett. **89**, 108102 (2002).
[16] J. Ladik, G. Biczó, and G. Elek, J. Chem. Phys. **44**, 483 (1966).
[17] J. P. Lewis, T. E. Cheatham, E. B. Starikov, H. Wang, and O. F. Sankey, J. Phys. Chem. B **107**, 2581 (2003).
[18] J. P. Lewis, P. Ordejon, and O. F. Sankey, Phys. Rev. B **55**, 6880 (1997).
[19] E. Maciá, Phys. Rev. B **74**, 245105 (2006).
[20] P. Maragakis, R. L. Barnett, E. Kaxiras, M. Elstner, and T. Frauenheim, Phys. Rev. B **66**, 241104 (2002).
[21] M. Taniguchi and T. Kawai, Phys. Rev. E **70**, 011913 (2004).
[22] H. Wang, J. P. Lewis, and O. F. Sankey, Phys. Rev. Lett. **93**, 016401 (2004).
[23] P. Xie, K. Liu, F. Gu, and Y. Aoki, Int. J. Quantum Chem. **112**, 230 (2012).
[24] P. Xie, H. Teramae, K. Liu, and Y. Aoki, Int. J. Quantum Chem. **113**, 489 (2013).
[25] H. Yamada and K. Iguchi, Adv. Condens. Matter Phys. **2010**, 380710 (2010).
[26] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[27] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[28] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, Nucleic Acids Res. **34**, 564 (2006).
[29] J. MacNaughton, A. Moewes, J. Lee, S. Wettig, H.-B. Kraatz, L. Ouyang, W. Ching, and E. Kurmaev, J. Phys. Chem. B **110**, 15742 (2006).
[30] D. M. York, T.-S. Lee, and W. Yang, Phys. Rev. Lett. **80**, 5011 (1998).
[31] Y.-W. Kwon, C. H. Lee, D.-H. Choi, and J.-I. Jin, J. Mater. Chem. **19**, 1353 (2009).
[32] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[33] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[34] W.-Y. Ching and P. M. Rulis, *Electronic Structure Methods for Complex Materials: The Orthogonalized Linear Combination of Atomic Orbitals* (Oxford University Press, Oxford, 2012).
[35] R. R. Sinden, *DNA Structure and Function* (Gulf, Houston, 1994).
[36] S. Gregory, K. Barlow, K. McLay, R. Kaul, D. Swarbreck, A. Dunham, C. Scott, K. Howe, K. Woodfine, and C. Spencer, Nature (London) **441**, 315 (2006).
[37] J. C. Wang, Proc. Natl. Acad. Sci. USA **76**, 200 (1979).
[38] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[39] W. Y. Ching, J. Am. Ceram. Soc. **87**, 1996 (2004).
[40] L. Liang, P. Rulis, B. Kahr, and W. Y. Ching, Phys. Rev. B **80**, 235132 (2009).
[41] L. Ouyang, L. Randaccio, P. Rulis, E. Kurmaev, A. Moewes, and W. Ching, J. Mol. Struct.: THEOCHEM **622**, 221 (2003).
[42] J. Eifler, P. Rulis, R. Tai, and W.-Y. Ching, Polymers **6**, 491 (2014).
[43] P. Adhikari, A. M. Wen, R. H. French, V. A. Parsegian, N. F. Steinmetz, R. Podgornik, and W.-Y. Ching, Sci. Rep. **4**, 5605 (2014).
[44] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).
[45] M. H. Lee, S. Avdoshenko, R. Gutierrez, and G. Cuniberti, Phys. Rev. B **82**, 155455 (2010).
[46] J. W. Ponder and D. A. Case, Adv. Protein Chem. **66**, 27 (2003).
[47] V. R. Cooper, T. Thonhauser, A. Puzder, E. Schröder, B. I. Lundqvist, and D. C. Langreth, J. Am. Chem. Soc. **130**, 1304 (2008).
[48] J. Šponer, P. Jurecka, and P. Hobza, J. Am. Chem. Soc. **126**, 10142 (2004).
[49] C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, and E. J. Baerends, J. Am. Chem. Soc. **122**, 4117 (2000).
[50] R. H. French *et al.*, Rev. Mod. Phys. **82**, 1887 (2010).
[51] I. Radhakrishnan and D. J. Patel, J. Mol. Biol. **241**, 600 (1994).
[52] G. N. Parkinson, M. P. Lee, and S. Neidle, Nature (London) **417**, 876 (2002).