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

Phys. Rev. E **90**, 022705 — Published 7 August 2014

DOI: 10.1103/PhysRevE.90.022705

Electronic Structure, Stacking Energy, Partial Charge, and Hydrogen Bonding in Four Periodic B-DNA Models

Lokendra Poudel[1], Paul Rulis[1], Lei Liang[1], and W.Y. Ching[1]

[1]Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City, Missouri 64110, USA

Abstract

We present a theoretical study on the electronic structure of four periodic B-DNA models labeled as $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$ with A=Adenine, T=Thymine, G=Guanine, and C=Cytosine. Each model has ten base pairs with Na counter-ions 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 have estimated 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 HOMO-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 counter-ions significantly affect the gap width, the conductivity, and the distribution of partial charge on the DNA backbone. Also evaluated quantitatively are the surface partial charge density on each functional group of the DNA models.

PACS number(s): 87.14.gk, 87.15.A-, 87.15.ag, 87.15.Fh

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 bonds to form the DNA backbone. There are two types of complementary base pairings in DNA: A-T and G-C that 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 last the 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 matters 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 on 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 ranging from 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 counter-ions 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 up to now[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 counter-ions (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), exact role of counter-ions, 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 used the following strategy: (1) We carefully constructed four DNA models that are periodic along the helical z-axis, have different base-pairs and stacking sequences, and are labeled as $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$. (2) Twenty sodium (Na) ions were added to the models to compensate for the twenty negatively charge PO_4 groups in the DNA backbone. (3) The structures of these models were 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, PO_4 + Na) are 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. The main results and their discussions are

presented in section IV. We end with a summary and conclusions section that includes plans for further applications.

II. B-DNA Models

DNA exists in several conformations such as A-DNA, B-DNA, Z-DNA, etc., 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 that a periodicity of 10 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 phosphate group or sugar group ($5^{'}$ and $3^{'}$ respectively) at the ends. We have built four different 10 base pair B-DNA models with periodicity in the axial direction (z-axis) and different stacking sequence. They are labeled as $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and $(AT\text{-}GC)_5$ and are shown in **Figure 1 (a)-(d).** 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 criteria 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 (FB), 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 (DOS). A minimal basis (MB) was used for the separate calculation of partial charge and bond orders. Please refer to reference 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 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 HOMO-LUMO gap where HOMO is the highest occupied molecular orbital and LUMO is the lowest unoccupied molecular orbital. The effective charge Q* (in units of electrons) on each atom in the models are 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:

$$Q_{\alpha}^{*} = \sum_{i,\alpha} \sum_{n} \sum_{occ} \sum_{j,\beta} C_{i\alpha}^{*n} C_{j\beta}^{n} S_{i\alpha,j\beta} \qquad (1)$$

where the $C^{n}_{j\beta}$ are the eigenvector coefficients of the n^{th} band, j^{th} orbital, and β^{th} atom. The $S_{i\alpha,j\beta}$ are the overlap integrals between the i^{th} orbital of the α^{th} atom and j^{th} orbital of the β^{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. It generally scales

with the bond length (BL) but also depends on the local environment of the bonding atoms. The bond order (BO) values for each pair of atoms α and β are calculated according to:

$$\rho_{\alpha\beta} = \sum_{n\,occ} \sum_{i,j} C_{i\alpha}^{*n} C_{j\beta}^{n} S_{i\alpha,j\beta} \qquad (2)$$

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, method such as Bader 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 counter ions which is always present in real biomolecular system. For example, recently, Lee et al[45] performed classical molecular dynamic simulation using Amber99 force field with the AMBER10 package (AMBER = Assisted Model Building with Energy Refinement)[46]. The electronic structure of dsDNA were obtained using "fragment orbital approach". (A fragment consists with of a pair of nucleotides (phosphate-deoxyribose-nucleobase) from DNA strand.). They have observed that the molecular orbitals of nucelotide fragments are intermittently switching from bases to backbone over the simulation time. The HOMO can temporarily have large weight on the backbones as a function of time. This finding is in line with our results that HOMO is located on phosphate groups (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 1.** $(AT\text{-}CG)_5$ 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 $(AT)_{10}$ and $(GC)_{10}$ models as reference energies for models with no stacking disorder, we can estimate the bilayer stacking energy. The TE of $(AT)_5(GC)_5$ 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 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/mole. 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 the 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 counter ions and sugar but the other calculation used an isolated base pair stack. We believe that the measured stacking energy will be highly dependent on the B-DNA structure and its local environment (e.g. stacking height, twisting angle, counter-ions, humidity, etc.) 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 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 subsection 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 -25 eV to 25eV (left column) and -1.0 eV to 6 eV (right column) are shown in **Figure 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 eV, 2.73 eV, 2.52 eV, 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, these are newly reported values. 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. (Ref 8) using the "electrostatic trapping" method for the

conductance of homogeneous sequence of 30 base-pairs (Poly(G)-Poly(C)) DNA. They found that band gap of short DNA sequence to be about 2.1 eV. Shapir et al. (Ref 9) studied long single poly(G)-poly(C) DNA molecules deposited on gold using scanning tunneling microscopy (STM). They also found that 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 **Figure 2**. To better trace the origin of these states, we resolve the TDOS into partial DOS (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 PDOS for the four B-DNA models are shown in **Figure 3**. The following facts are observed: (1) 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 counter ions to the PO_4 group which forms rather strong Na-O bonds with O atoms from the PO_4 group. (2) All of the lower conduction band states up to 5 eV are from base pairs. For $(AT)_{10}$, there are 5 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. (3) 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. (4) 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 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. (5) There is no significant participation of states below or near 5 eV from the sugar group. This simply reflects the fact that the anti-bonding 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 counter-ions 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 counter-ions play a pivotal role for electron transfer in DNA. This conclusion is important for understanding the conductivity properties of DNA.

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 counter-ions, and the weak but not entirely negligible bonding from the next nearest neighbor atoms (NNN) with BLs less than 2.6 Å. **Table 1** 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 $(GC)_{10}$ 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 (Na-O BO), and the next nearest neighbor bond order (NNNBO). **Table 1** shows that the TBO contribution from the CBO for $(AT)_5(GC)_5$ is 267.45 and for $(AT\text{-}GC)_5$ it is 267.59, which are higher than the average value 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 1**, we have found that the NNNBO values are very small, but those from Na-O BO is actually larger than that of from HB, indicating that the counter-ion 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 **Figure 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. It can be seen that: (1) 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. (2) 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 wider distributions in the stacked models with a concomitant decrease in BO. (3) 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. (4) As pointed out before, the bonding from the NNN pairs are small but not totally negligible.

In **Table 2**, we list the average BL and BO values for each specific HB between the base pairs in the four models. Table 2 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 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:

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 on this 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

which are displayed in **Figure 5.** The atomic partial charge is the deviation of effective charge Q* of Eq. (1) from the charge on the neutral atom Q_0 in units of electron, or $\Delta Q = (Q^* - Q_0)$ (i.e. -ΔQ = gain of electron and +ΔQ = loss of electron). **Figure 5** lists ΔQ 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 **Figure 5**, the plot of Q* for P is reduced by 1 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. Thirdly, 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 key roles in hydrogen bonding as more bilayers of (AT-GC) are created. The increased variations in ΔQ for Na and P in the stacked models are also very obvious.

By adding the ΔQ 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 surface partial charge density in the unit of electron/$(nm)^2$ for the four models are listed in **Table 3**. **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 ΔQ values in **Table 3** and the actual areas for each group. We believe that this is the first time such information has been provided and that it could be very useful for quantitative evaluation of electrostatic forces involving B-DNA.

V. Summary and Conclusions

We have built four periodic B-DNA models $(AT)_{10}$, $(GC)_{10}$, $(AT)_5(GC)_5$, and (AT-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 counter-ions 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, 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.

VI. Appendix: Construction of the B-DNA models

The $(AT)_{10}$ and $(GC)_{10}$ DNA models were created by the using Nucleic Acid Builder (NAB), as a part of the tools in the Assisted Model Building with Energy Refinement (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 twenty sodium counter-ions (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 (θ), stacking height (d), x- and y-shift 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^o, 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 Å × 30 Å × 33.8 Å. After fully relaxing the atomic positions (see method section below) and varying the stacking distances, we obtained minimum-energy $(AT)_{10}$ and $(GC)_{10}$ models at a stacking distance of 3.3785 Å so that the lattice parameters of our simulation cell were set to: a = 30 Å, b = 30 Å, c = 37.856 Å, $\alpha = \beta = \gamma = 90^o$. 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 are a total of 660 atoms in the $(AT)_{10}$ model and 650 atoms in the $(GC)_{10}$ model. **Figure 1 (a)** and **Figure 1(b)** respectively show the axial (z-axis) and planar (x-y plane) view 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. Firstly, we removed 5-base pairs from the $(AT)_{10}$ model and replaced them with 5-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. Firstly, 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 co-ordinates 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. **Figure 1 (c)** and **Figure 1(d)** respectively show the axial (z-axis) and plane (x-y plane) view 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 5 such bilayers.

Acknowledgement

This work is supported by the US DOE-Office of BES, Division of Materials Science and Engineering under the grant DE-SC008176. This research used the resources of NERSC supported by the Office of Science of 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 DEFG-0284DR45170.

Table 1: Total energy and total bond order for B-DNA models

Models	No. of Atoms	No. of valence electron	Total Energy (eV)	Bond Order calculations				
				TBO	CBO	HBO	Na-O BO	NNNBO (<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

Table 2: Bond length and bond order values for hydrogen bonds in four B-DNA models

Models	H-Bonds	Average B.L.(Å)	Average B.O.
$(AT)_{10}$	H_A-O_T*	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	0.507
$(AT)_5(GC)_5$	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

* The sub-script indicates the corresponding base of the atom.

Table 3: Average net partial charge and surface partial charge densities on functional groups of four B-DNA models

Models	Functional Group	Average ΔQ (e^-)	Surface charge density (e^-/nm^2)
$(AT)_{10}$	Adenine (A)	-0.192	-0.245
	Thymine (T)	-0.252	-0.364
	Sugar (S)	+0.715	+1.034
	PO_4 + Na	-0.494	-1.559
$(GC)_{10}$	Guanine (G)	-0.279	-0.355
	Cytosine (C)	-0.162	-0.212
	Sugar (S)	+0.715	+1.029
	PO_4 + Na	-0.494	-1.574
$(AT)_5(GC)_5$	Adenine (A)	-0.192	-0.248
	Thymine (T)	-0.247	-0.312
	Guanine (G)	-0.261	-0.358
	Cytosine (C)	-0.210	-0.266
	Sugar (S)	+0.720	+1.102
	PO_4 + Na	-0.493	-1.761
$(AT\text{-}GC)_5$	Adenine (A)	-0.181	-0.238
	Thymine (T)	-0.286	-0.376
	Guanine (G)	-0.289	-0.473
	Cytosine (C)	-0.117	-0.145
	Sugar (S)	+0.699	+1.004
	PO_4 + Na	-0.481	-1.733

Figure Captions:

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

Figure 2. (Color online) Total density of states (TDOS) for the four B-DNA models in the energy range -25 eV to 25 eV (Left panel) and -1 eV to 6 eV (Right panel): (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, (d) $(AT\text{-}GC)_5$.

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

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

Figure 5. (Color online) Calculated atomic partial charges on all atoms in the four B-DNA models: (a) $(AT)_{10}$, (b) $(GC)_{10}$, (c) $(AT)_5(GC)_5$, (d) $(AT\text{-}GC)_5$. For P, the plotted data ΔQ is reduced by 1 electron in order to have a clearer display.

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

References:

1. V. A. Bloomfield, D. M. Crothers and I. Tinoco, Herndon, VA: University Science Books (1996).
2. F. H. Crick and J. D. Watson, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **223** (1152), 80-96 (1954).
3. E. T. Kool, Annual review of biophysics and biomolecular structure **30** (1), 1-22 (2001).
4. G. A. Jeffrey and G. A. Jeffrey, *An introduction to hydrogen bonding*. (Oxford university press New York, 1997).
5. N. C. Seeman, Molecular biotechnology **37** (3), 246-257 (2007).
6. D. C. Young, *Computational drug design: a guide for computational and medicinal chemists*. (John Wiley & Sons, 2009).
7. A. V. Pinheiro, D. Han, W. M. Shih and H. Yan, Nature nanotechnology **6** (12), 763-772 (2011).
8. D. Porath, A. Bezryadin, S. De Vries and C. Dekker, Nature **403** (6770), 635-638 (2000).
9. E. Shapir, H. Cohen, A. Calzolari, C. Cavazzoni, D. A. Ryndyk, G. Cuniberti, A. Kotlyar, R. Di Felice and D. Porath, Nature Materials **7** (1), 68-74 (2008).
10. H. Wadati, K. Okazaki, Y. Niimi, A. Fujimori, H. Tabata, J. Pikus and J. P. Lewis, Applied Physics Letters **86** (2), - (2005).
11. K. Kummer, D. V. Vyalikh, G. Gavrila, A. B. Preobrajenski, A. Kick, M. Bo□nsch, M. Mertig and S. L. Molodtsov, The Journal of Physical Chemistry B **114** (29), 9645-9652 (2010).
12. M. Xu, S. Tsukamoto, S. Ishida, M. Kitamura, Y. Arakawa, R. Endres and M. Shimoda, Applied Physics Letters **87** (8), 083902 (2005).
13. L. A. Jauregui, K. Salazar-Salinas and J. M. Seminario, The Journal of Physical Chemistry B **113** (18), 6230-6239 (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, Physical Review Letters **85** (23), 4992-4995 (2000).
15. F. L. Gervasio, P. Carloni and M. Parrinello, Physical review letters **89** (10), 108102 (2002).
16. J. Ladik, G. Biczo and G. Elek, The Journal of Chemical Physics **44** (2), 483-485 (2004).
17. J. P. Lewis, T. E. Cheatham, E. B. Starikov, H. Wang and O. F. Sankey, The Journal of Physical Chemistry B **107** (11), 2581-2587 (2003).
18. J. P. Lewis, P. Ordejon and O. F. Sankey, Physical Review B **55** (11), 6880 (1997).
19. E. Maciá, Physical Review B **74** (24), 245105 (2006).
20. P. Maragakis, R. L. Barnett, E. Kaxiras, M. Elstner and T. Frauenheim, Physical Review B **66** (24), 241104 (2002).
21. M. Taniguchi and T. Kawai, Physical Review E **70** (1), 011913 (2004).
22. H. Wang, J. P. Lewis and O. F. Sankey, Physical review letters **93** (1), 016401 (2004).
23. P. Xie, K. Liu, F. Gu and Y. Aoki, International Journal of Quantum Chemistry **112** (1), 230-239 (2012).
24. P. Xie, H. Teramae, K. Liu and Y. Aoki, International Journal of Quantum Chemistry **113** (4), 489-496 (2013).
25. H. Yamada and K. Iguchi, Advances in Condensed Matter Physics **2010** (2010).
26. P. Hohenberg and W. Kohn, Physical Review **136** (3B), B864-B871 (1964).
27. W. Kohn and L. J. Sham, Physical Review **140** (4A), A1133-A1138 (1965).
28. P. Yakovchuk, E. Protozanova and M. D. Frank-Kamenetskii, Nucleic acids research **34** (2), 564-574 (2006).
29. J. MacNaughton, A. Moewes, J. Lee, S. Wettig, H.-B. Kraatz, L. Ouyang, W. Ching and E. Kurmaev, The Journal of Physical Chemistry B **110** (32), 15742-15748 (2006).
30. D. M. York, T.-S. Lee and W. Yang, Physical Review Letters **80** (22), 5011-5014 (1998).
31. Y.-W. Kwon, C. H. Lee, D.-H. Choi and J.-I. Jin, Journal of Materials Chemistry **19** (10), 1353-1380 (2009).

32. G. Kresse and J. Furthmüller, Physical Review B **54** (16), 11169-11186 (1996).
33. G. Kresse and J. Furthmüller, Computational Materials Science **6** (1), 15-50 (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 Professional Publishing, 1994).
36. S. Gregory, K. Barlow, K. McLay, R. Kaul, D. Swarbreck, A. Dunham, C. Scott, K. Howe, K. Woodfine and C. Spencer, Nature **441** (7091), 315-321 (2006).
37. J. C. Wang, Proceedings of the National Academy of Sciences **76** (1), 200-203 (1979).
38. J. P. Perdew, K. Burke and M. Ernzerhof, Physical Review Letters **77** (18), 3865-3868 (1996).
39. W. Y. Ching, Journal of the American Ceramic Society **87** (11), 1996-2013 (2004).
40. L. Liang, P. Rulis, B. Kahr and W. Y. Ching, Physical Review B **80** (23), 235132 (2009).
41. L. Ouyang, L. Randaccio, P. Rulis, E. Kurmaev, A. Moewes and W. Ching, Journal of Molecular Structure: THEOCHEM **622** (3), 221-227 (2003).
42. J. Eifler, P. Rulis, R. Tai and W.-Y. Ching, Polymers **6** (2), 491-514 (2014).
43. P. Adhikari, A. M. Wen, R. H. French, V. A. Parsegian, N. F. Steinmetz, R. Podgornik, and W.Y. Ching, Scientific reports (accepted and in press).
44. R. S. Mulliken, The Journal of Chemical Physics **23** (10), 1833-1840 (1955).
45. M. Lee, S. Avdoshenko, R. Gutierrez and G. Cuniberti, Physical Review B **82** (15), 155455 (2010).
46. J. W. Ponder and D. A. Case, Advances in protein chemistry **66**, 27-86 (2003).
47. V. R. Cooper, T. Thonhauser, A. Puzder, E. Schröder, B. I. Lundqvist and D. C. Langreth, Journal of the American Chemical Society **130** (4), 1304-1308 (2008).
48. J. Šponer, P. Jurecka and P. Hobza, Journal of the American Chemical Society **126** (32), 10142-10151 (2004).
49. C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders and E. J. Baerends, Journal of the American Chemical Society **122** (17), 4117-4128 (2000).
50. R. H. French, V. A. Parsegian, R. Podgornik, R. F. Rajter, A. Jagota, J. Luo, D. Asthagiri, M. K. Chaudhury, Y.-m. Chiang and S. Granick, Reviews of Modern Physics **82** (2), 1887 (2010).
51. I. Radhakrishnan and D. J. Patel, Journal of molecular biology **241** (4), 600-619 (1994).
52. G. N. Parkinson, M. P. Lee and S. Neidle, Nature **417** (6891), 876-880 (2002).

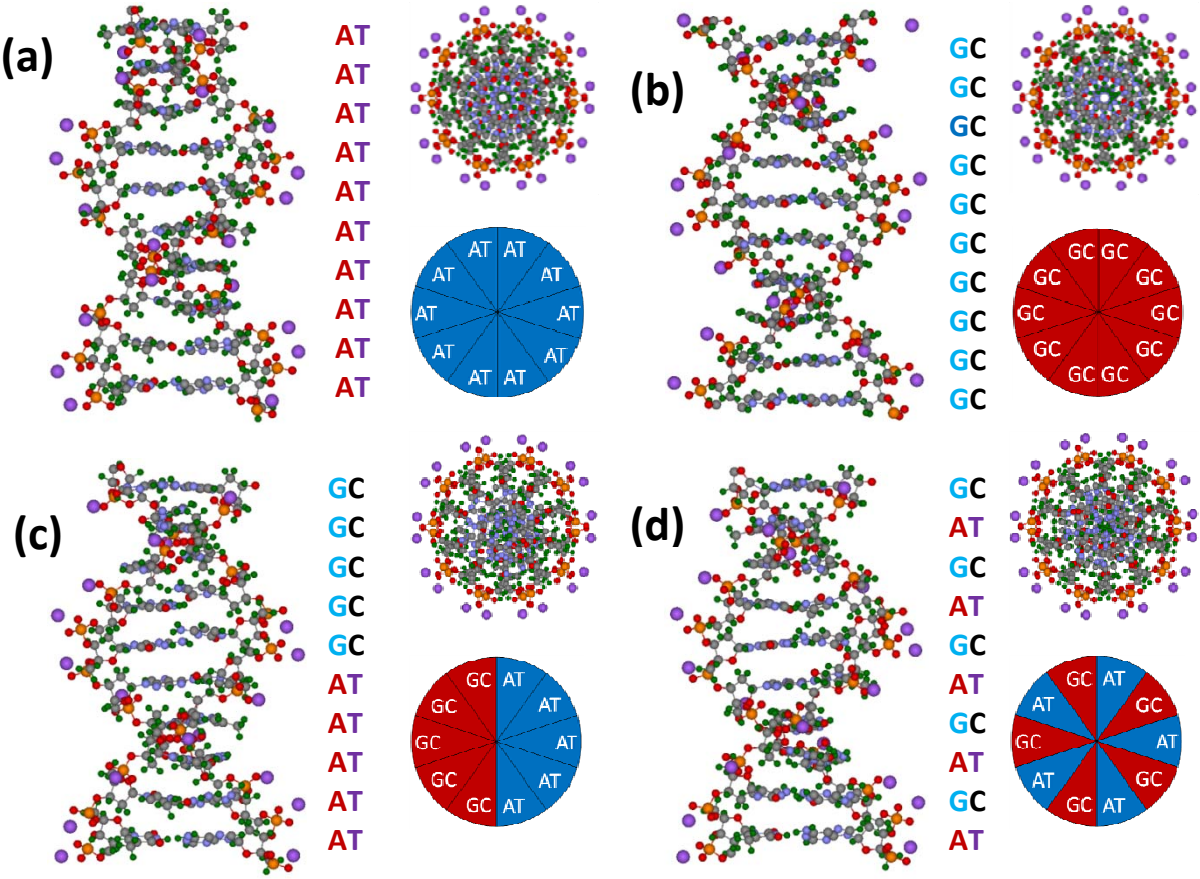
(a)
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
AT
(b)
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
GC
(c)
GC
GC
GC
GC
GC
AT
AT
AT
AT
AT
GC
GC
GC
GC
AT
AT
AT
AT
GC
GC
(d)
GC
AT
GC
AT
GC
AT
GC
AT
GC
AT
GC
AT
GC
AT
GC
AT
GC
AT
GC
AT

Figure: 1

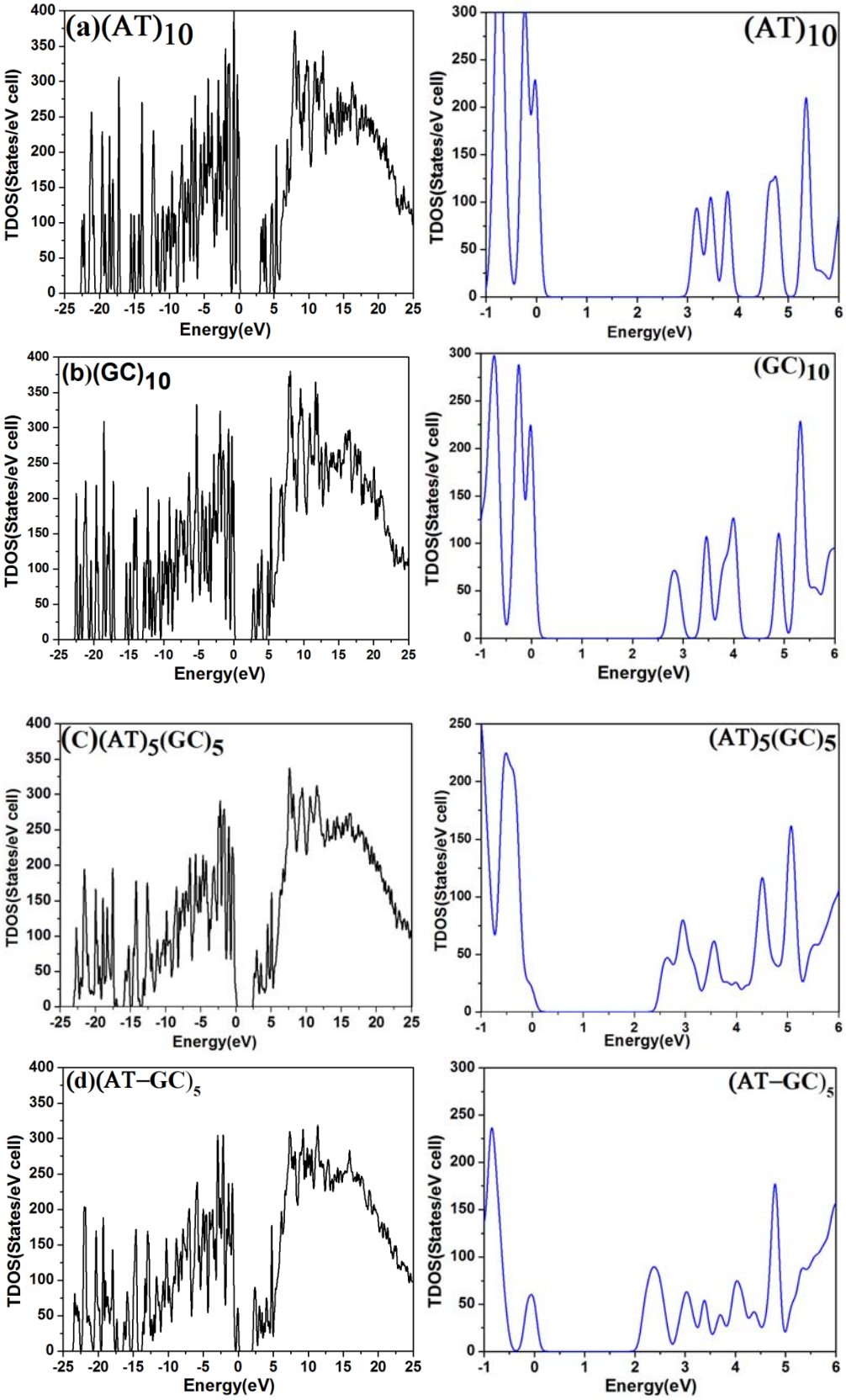
(a)(AT)10
(AT)10
(b)(GC)10
(GC)10
(C)(AT)5(GC)5
(AT)5(GC)5
(d)(AT–GC)5
(AT–GC)5
TDOS(States/eV cell)
Energy(eV)

Figure: 2

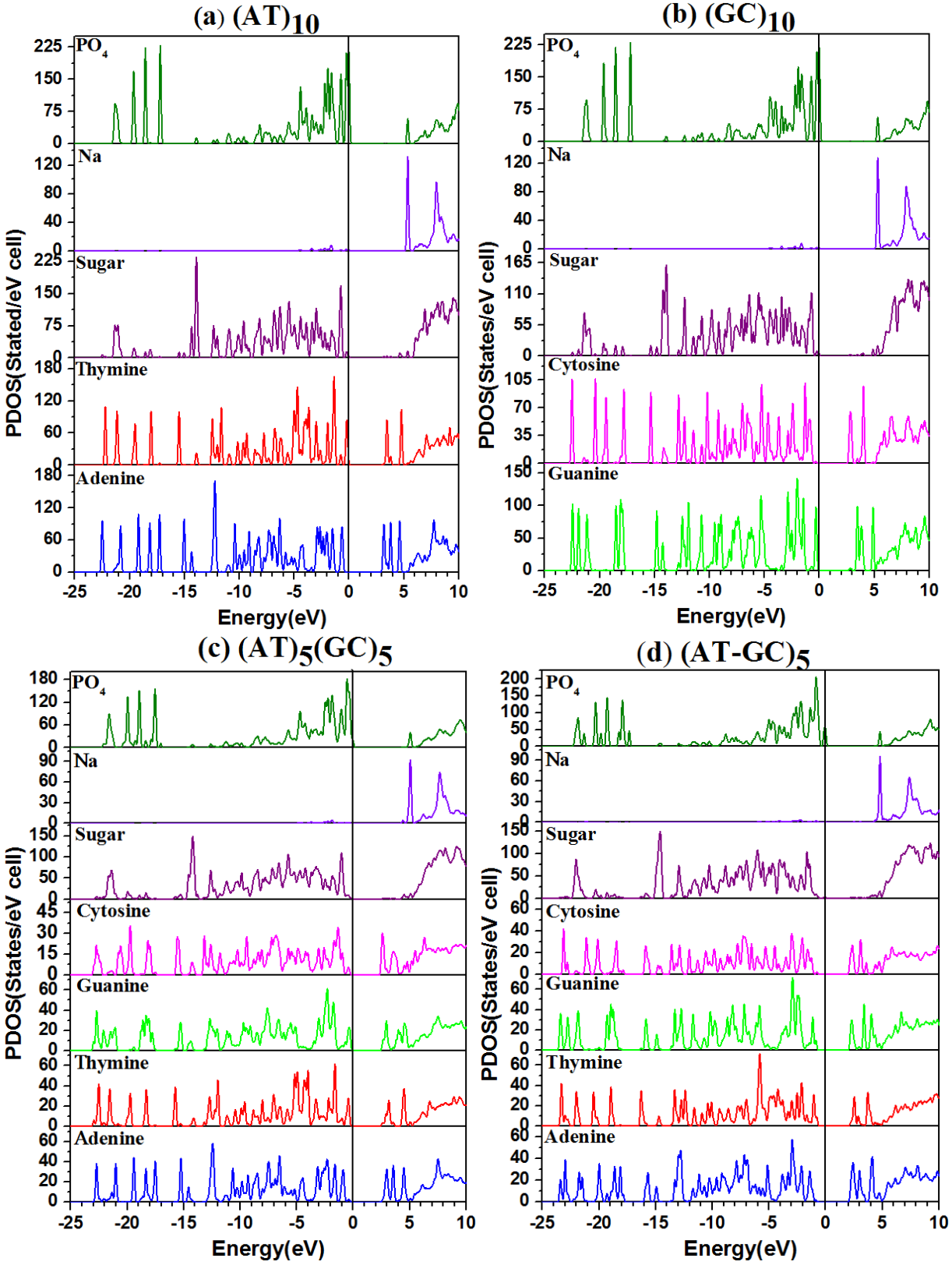
(a) (AT)10
(b) (GC)10
(c) (AT)5(GC)5
(d) (AT-GC)5
PO4
Na
Sugar
Thymine
Adenine
Cytosine
Guanine
PDOS(Stated/eV cell)
PDOS(States/eV cell)
Energy(eV)

Figure: 3

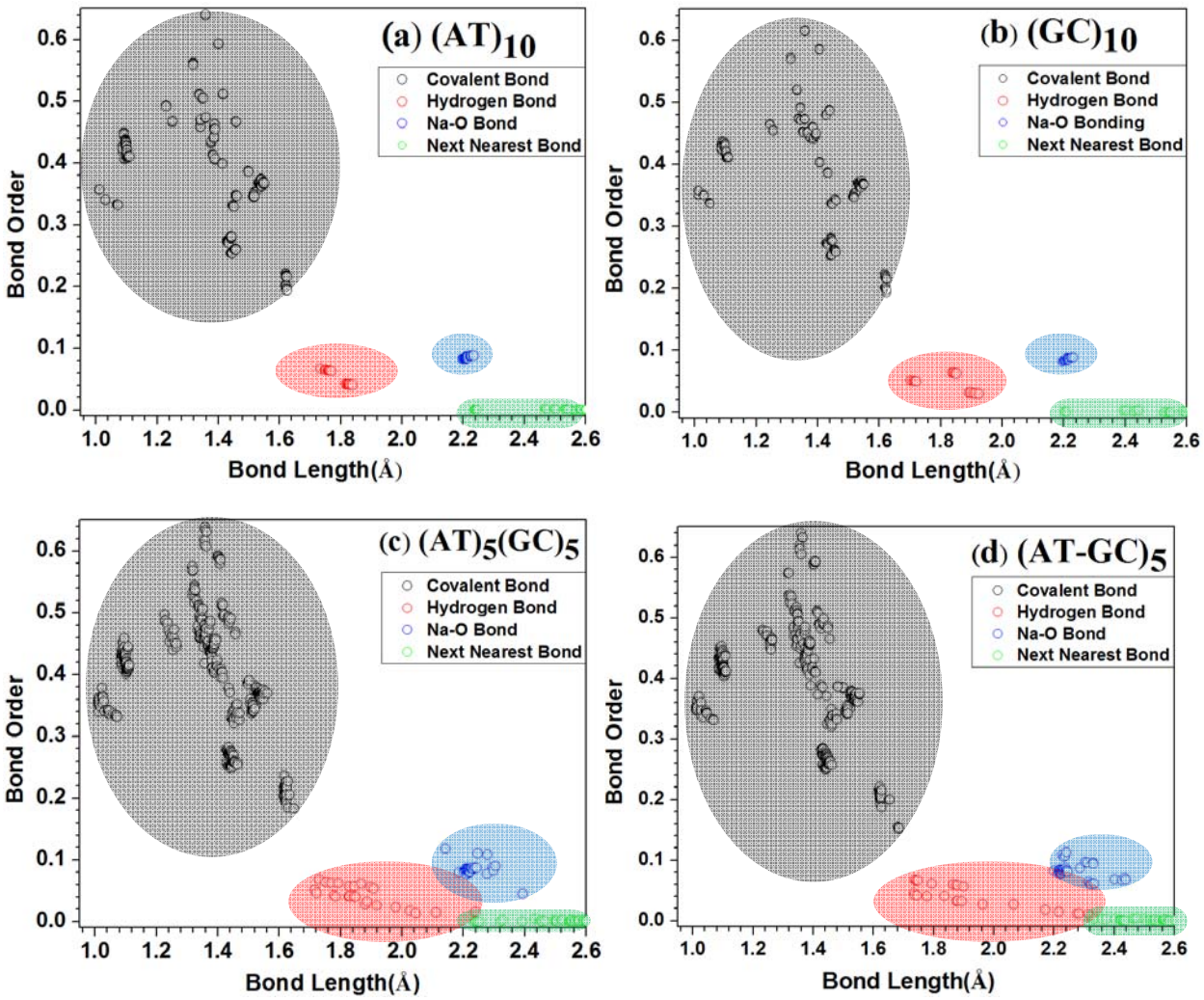
(a) (AT)10
Covalent Bond
Hydrogen Bond
Na-O Bond
Next Nearest Bond
(b) (GC)10
Covalent Bond
Hydrogen Bond
Na-O Bonding
Next Nearest Bond
(c) (AT)5(GC)5
Covalent Bond
Hydrogen Bond
Na-O Bond
Next Nearest Bond
(d) (AT-GC)5
Covalent Bond
Hydrogen Bond
Na-O Bond
Next Nearest Bond
Bond Order
Bond Length(Å)

Figure 4:

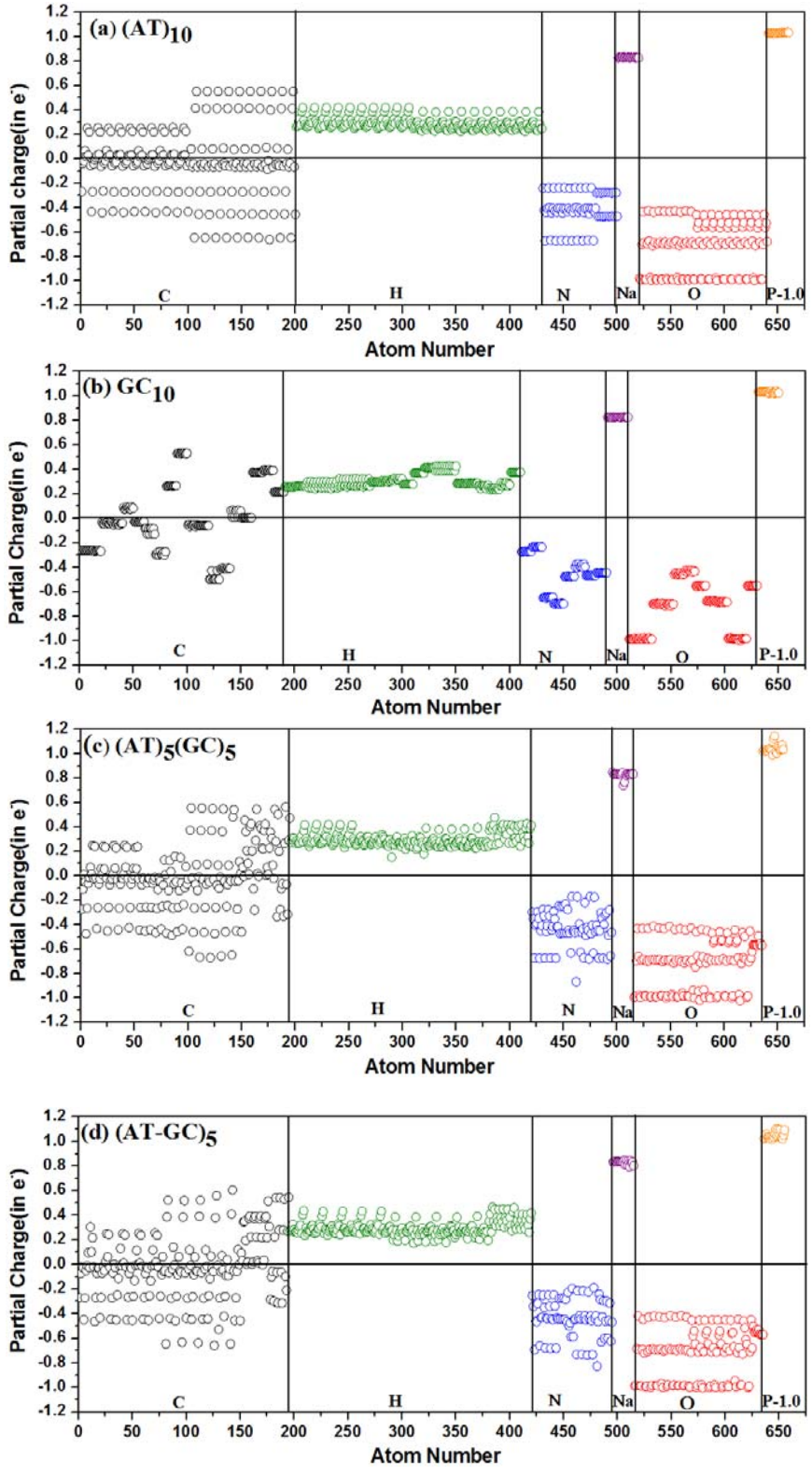
(a) (AT)10
(b) GC10
(c) (AT)5(GC)5
(d) (AT-GC)5
Partial charge(in e)
Partial Charge(in e)
Atom Number
C
H
N
Na
O
P-1.0

Figure 5:

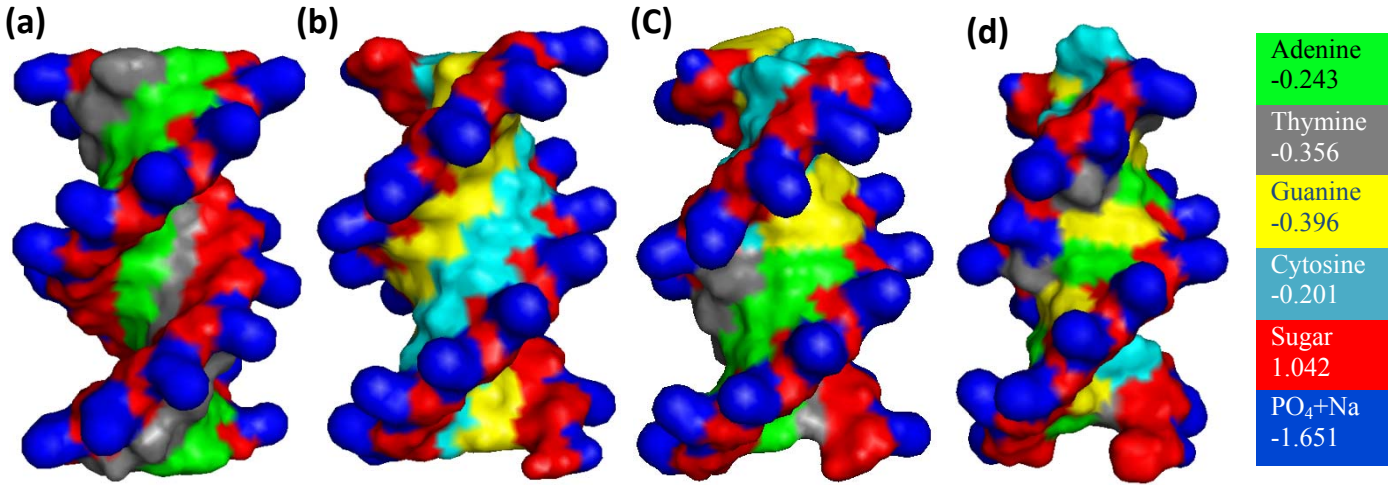
(a)
(b)
(C)
(d)
Adenine
-0.243
Thymine
-0.356
Guanine
-0.396
Cytosine
-0.201
Sugar
1.042
PO4+Na
-1.651

Figure 6: