

Electronic Structure of Wet DNA

Francesco Luigi Gervasio

*Università di Firenze, Dipartimento di Chimica, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy
and Centro Svizzero di Calcolo Scientifico, via Cantonale, CH-6928 Manno, Switzerland*

Paolo Carloni

*International School for Advanced Studies (SISSA/ISAS) and Istituto Nazionale di Fisica della Materia (INFN),
via Beirut 4, I-34014, Trieste, Italy*

Michele Parrinello

*Centro Svizzero di Calcolo Scientifico, via Cantonale, CH-6928 Manno, Switzerland
and Physical Chemistry, ETH Zurich, Hönggerberg, CH-8093 Zurich, Switzerland
(Received 23 January 2002; published 15 August 2002)*

The electronic properties of a Z-DNA crystal synthesized in the laboratory are investigated by means of density-functional theory Car-Parrinello calculations. The electronic structure has a gap of only 1.28 eV. This separates a manifold of 12 occupied states which came from the π guanine orbitals from the lowest empty states in which the electron is transferred to the Na^+ from PO_4^- groups and water molecules. We have evaluated the anisotropic optical conductivity. At low frequency the conductivity is dominated by the $\pi \rightarrow \text{Na}^+$ transitions. Our calculation demonstrates that the cost of introducing electron holes in wet DNA strands could be lower than previously anticipated.

DOI: 10.1103/PhysRevLett.89.108102

PACS numbers: 87.15.Aa, 81.05.Zx, 87.14.Gg

The computational study of nucleic acids and other biopolymers in laboratory-realizable conditions has until now focused mainly on their structural properties, while the nature of the electronic structure has received far less attention. Yet the electronic states play an important role in determining the interatomic forces, as they lead to electronic polarization effects and many-body forces and provide an accurate description of the dielectric properties. Furthermore, they are crucial in phenomena such as radiation-induced damage [1]. More recently, there has been great interest in the nature of the electronic structure of DNA because of its potential applications to nanotechnology. Indeed, DNA has several properties that make it attractive in a variety of applications. It is a stable polymer and can easily be handled and modified almost at will. Furthermore, its one-dimensional character and regular stacking of π -bases have suggested the possibility of using DNA as a nanoscale conductor [2–7]. Unfortunately, experiments have provided very contradictory results. Such experiments are obviously very difficult since they require the handling of single molecules and the contact to the metallic leads can play an important role. Theory could be of great help in understanding these phenomena, but theoretical efforts to date have mostly been limited to the study of small fragments in the gas phase [8]. Only recently has an effort been made to study periodically infinite double strands [9]. However, in this instance a limited basis set was used and the solvation water and counterions, which are crucial for the stability of the DNA, were neglected [10]. In our view, it is extremely important rather to study systems that can actually be synthesized and studied in the laboratory.

Out of computational considerations we focus here on the smallest nucleotide crystal structure that can be synthesized in the laboratory, namely, $d(\text{GpCpGpCpGpCpGpCpGpCpGpCpGpCp})$ (Fig. 1) [3].

This is an infinitely repeated biopolymer, which in the unit cell contains a guanine:cytosine (Gua:Cyt) dodecamer in the Z conformation. This structure is very attractive for nanotechnology since the absence of disorder in the base-pair sequence can in principle lead to a higher conductivity. It is also interesting for radiation damage studies since it is rich in Gua bases, which are the most prone to UV-induced radical formation. Although not biologically active, it has all the ingredients of an active DNA, namely, the double strand, the counterions, and the solvation waters.

The system crystallizes in the hexagonal space group $P6(5)22$, with cell dimensions $a = b = 18.08 \text{ \AA}$, $c = 43.10 \text{ \AA}$. The coordinates were taken from the Protein Data Bank [12]. The hydrogen atom positions, not resolved in the x-ray structure, were assigned by hand, respecting only the constraint of standard bond angles and bond lengths. Then the position of all the atoms was optimized with a quenched *ab initio* molecular dynamics, which led to the spontaneous formation of a hydrogen bond network. Also the sodium counterions were not resolved in the crystal structure. After careful inspection their initial position was established in the following way. Twelve Na^+ were positioned in the middle of 12 symmetric solvent excluded cavities of 1.6 \AA radius existing in the structure. The excluded volumes were calculated with the method of Connolly [13]. Since we could not find enough empty space in the structure to position the remaining 12 sodium ions, 12 symmetrical water molecules in the proximity of

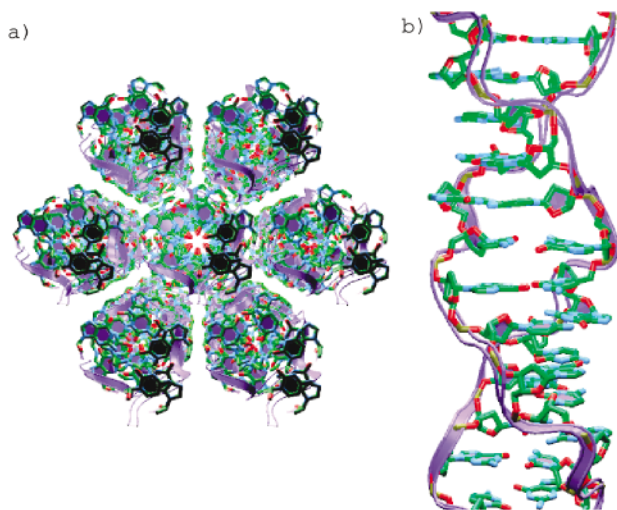


FIG. 1 (color). View of the three-dimensional structure of the Gua:Cyt dodecamer [11] (a) along, (b) orthogonal to the $c(z)$ axis. Water molecules, counterions, and hydrogens have been removed for clarity. The sugar-phosphate backbone is represented as ribbons.

the phosphates were substituted with sodium ions. The position of the sodium ions was optimized with a quenched *ab initio* molecular dynamics in which all the symmetry restraints, with the exception of the periodic boundary conditions, were relaxed. Our model contained 654 heavy atoms and 540 hydrogen atoms [the molecular formula is $C_{228}N_{96}O_{144}P_{24}Na_{24}H_{264} \cdot 138(H_2O)$].

The calculations were performed within the framework of density-functional theory. Gradient corrected exchange and correlation functionals with the BLYP [14,15] parametrizations were used. We treated explicitly the 3960 valence electrons, since in the case of sodium we treat explicitly the semicore states. The interaction between valence electrons and ionic cores was described by Martin-Troullier pseudopotentials [16]. The Kohn-Sham orbitals were expanded up to an energy cutoff of 70 Ry. The total number of PW treated was 408 238. We used the CPMD code [17], and optimized the structure using a molecular dynamics/quenching scheme for ionic relaxation. The cell parameters were not optimized. We sample only the Γ point of the Brillouin zone, which given the cell size is an accurate approximation. Coulomb interactions are evaluated by the Ewald summation method. No symmetry restriction was imposed on the calculation. We stopped the relaxation when the root-mean-square value of the force was less than 10^{-3} au. We estimate that the resulting uncertainty in the position is less than the experimental error. The chemical bonding and the polarization effects were characterized with the method of the maximally localized Wannier orbitals [18,19].

Solvation waters are partitioned in two linear structures and six droplets. The first string of waters is placed inside the helix where they bridge the Cyt carbonyl oxygens. The second is placed in the major groove and the waters interact

with the phosphate groups and the Na^+ counterions. In the pockets of the helices, i.e., in the mouth of the major and minor grooves [20], the water molecules form “droplets” of hydrogen-bonded molecules that have a peculiar structure (inset of Fig. 2).

The link between one water cluster and the other is provided by the phosphate groups. It is clear that the bonding pattern for all the water in the structure is unusual and rather different from that of bulk water. This leads to a wide distribution of dipole moments (Fig. 2), which ranges from the value 1.7 D close to the gas phase of the molecules inside the helix to the very high value of 3.8 D for the fivefold coordinated molecules in the droplets. The spread of the dipole moments and the unusual water structure confirms the widely held view that water close to the nucleic acids has a different character from that in the bulk. It therefore seems difficult that rigid ion models based on bulk properties can accurately describe the properties of nucleic acid solutions. To test this hypothesis we compared the hydrogen bond network obtained by full *ab initio* optimization with that obtained with molecular mechanics calculations on the same model system using the AMBER potential [23]. The optimized structure was

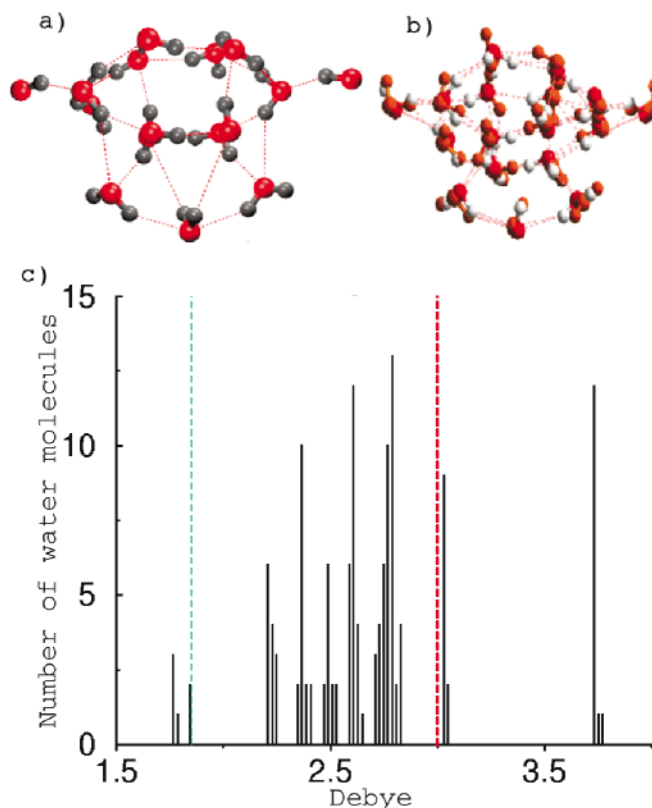


FIG. 2 (color). (a) H bond pattern of water droplets present in the minor grooves; (b) Comparison of the H bond pattern obtained by *ab initio* (white and red molecules) and molecular mechanics (orange molecules); and (c) Histogram of the dipole moment distribution of the water molecules; the dipole of water in the bulk [21] and gas [22] phase are represented in green and red, respectively.

obtained from the same initial configuration used for the *ab initio* optimization by a simulated annealing-quench procedure (6 ps of molecular dynamics at 200 K followed by a quench to 0 K in 3 ps). The final structure, while in general similar to the *ab initio* one, has a different bonding pattern, particularly in the proximity of water molecules that the *ab initio* calculation showed to be the most polarized compared to bulk phase [Fig. 2(b)].

We now turn to the description of the electronic properties of this interesting structure. Twelve quasidegenerate states are positioned at the top of the valence band. The noncomplete degeneracy is due to the residual disorder left in the structure. These states have a π character and are mostly localized on the Gua nucleobases. In order to quantitatively prove this point we integrate around the region that surrounds the Gua bases. The density summed over this region is 96% of the total. Each individual state of this manifold is spread over several Gua bases (see Fig. 3). The summed density of the top 12 states is also shown and has the full symmetry. The states immediately below the top of the valence band are relatively easy to assign to states localized also on the Gua. The first Cyt localized state is at 0.78 eV below the top. As we move further down the spectrum it becomes difficult to determine the nature of the states, as they become rather intermixed.

The DFT gap between empty and occupied states is particularly small, being only 1.28 eV. This very small value heralds the rather novel nature of the states at the bottom of the conduction band, which is a charge transfer state where one electron has been moved outside the helix mostly on the Na^+ counterions and on the PO_4^- groups [Fig. 4(b)]. This result is in agreement with the calculations of Enders *et al.* [24] which reaches similar conclusions about the reduced gap in the presence of counterions, and possibly with the experimental work of Tran *et al.* [25]

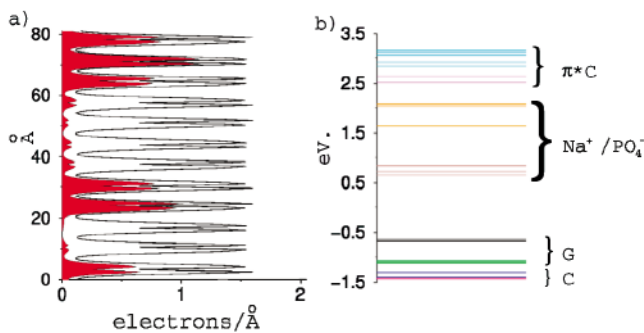


FIG. 3 (color). (a) Electronic charge density $\rho_e(z)$ along the z axis integrated over the x and y directions. The integrals are performed only in regions surrounding the Gua basis. These regions are defined as the union of all the cubes of side 2.1 Å which have a guanosine atom at their center. In red we show the $\rho_e(z)$ of the top state. The black line give the total $\rho_e(z)$ of the manifold. (b) Schematic level diagram around the Fermi level. The Fermi level positioned in the middle of the gap has been chosen as the zero of energy.

To assess the effect of water molecules on this state and on the value of the gap we repeated the calculation by removing the water molecules but otherwise leaving the geometry of the DNA and counterions unchanged. The gap is much reduced. This reveals the electrostatic nature of the charge transfer states and the fundamental role of water in shielding the DNA from the electrostatic field of counterions.

Moving higher in energy we find a number of states which have a similar character. It is only at 2.85 eV above the Fermi level that we find the first excited states with a strong base character and that is a π^* Cyt kind of state. The first π^* Gua state is instead at 3.18 eV. The $\pi \rightarrow \pi^*$ gap is 3.94 eV for Cyt and 3.82 eV for Gua, respectively. These values are rather similar to those of the Cyt:Gua monomer in the gas phase, obtained with the same computational setup (3.87 and 3.80 eV, respectively). In a first approximation these values can be compared to the excitation energies observed in experiments, namely, 4.6 and 4.5 eV for Cyt and Gua, respectively [26]. This comparison shows that our underestimation of the gap due to the use of DFT theory is less than 1 eV and that the relative values of the gaps are very well reproduced.

We have calculated the diagonal elements of the optical conductivity tensor $\sigma_{\alpha\alpha}(\omega)$ using the Kubo-Greenwood expression [27,28]:

$$\sigma_{\alpha\alpha}(\omega) = \frac{2\pi e^2}{3m_e^2 \Omega \omega} \sum_{v,c} |\langle \psi_v | p_\alpha | \psi_c \rangle|^2 \delta(E_c - E_v - \hbar\omega),$$

where E_v , E_c and ψ_v , ψ_c are the energies and wave functions of the valence and conduction bands, respectively, and Ω is the volume of the supercell.

In Fig. 5(a) we plot the average optical conductivity $\sigma(\omega) = \frac{1}{3} \text{Tr}[\sigma_{\alpha\alpha}(\omega)]$. This shows a sharp rise at about 4 eV; however, before that there is a tail of low-intensity states which extends to low energies. These are due to the

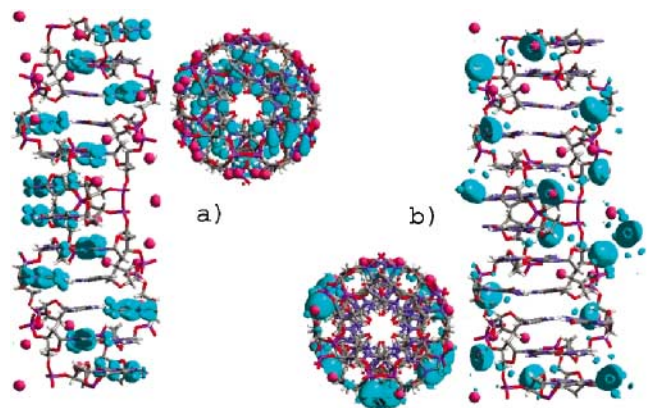


FIG. 4 (color). (a) Side and top view of the electronic density isosurface associated with the manifold of the 12 top states of the valence band. (b) Top and side view of the electronic density isosurface associated with the manifold of the 12 low states of the conduction band. The isosurfaces represented have a value of 10^{-2} electrons \AA^{-3} .

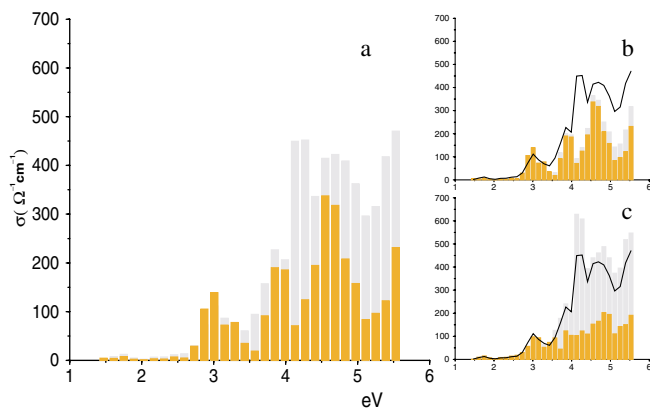


FIG. 5 (color). Optical conductivity ($\Omega^{-1}\text{cm}^{-1}$) versus frequency (eV). (a) $\sigma_{\text{average}}(\omega)$, (b) $\sigma_{\parallel}(\omega)$, and (c) $\sigma_{\perp}(\omega)$. The orange histograms are the contribution to the total conductivity (gray) due to excitations ending on the Na^+ electron transfer virtual states. The continuous black line in the right-hand figures is the average conductivity.

$\text{Gua} \rightarrow \text{Na}^+$ kind of transitions. They have low intensity because the dipole moment matrix elements connects two states that are centered on different positions.

From Fig. 5(b) we see that the optical activity in the strand direction is almost entirely due to the $\text{Gua} \rightarrow \text{Na}^+$ transition described here. In $\sigma_{\perp}(\omega)$, on the other hand, are also admixed the in-plane $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions described in the literature for the gas phase basis [29–32]. Note that the counterions Z coordinate lies between successive bases and therefore the transition dipole moments for the $\text{Gua} \rightarrow \text{Na}^+$ transitions have parallel as well as perpendicular components.

The small value of the gap suggests the possibility that DNA can be doped for instance by divalent or trivalent counterions. One could also speculate that the contradictory experimental evidence on DNA conductions may be due to incidental doping. From our calculations it emerges clearly that for a proper understanding of DNA electronic properties it is imperative to include solvation effects.

We acknowledge a generous grant from the Leibniz-Rechenzentrum-München on the Hitachi SR8000 super-computer which has made this calculation possible. We are very grateful to Dr. M. Boero and Dr. A. Seitsonen for their useful suggestions.

- [1] D. T. Odom and J. K. Barton, *Biochemistry* **40**, 8727–8737 (2001).
- [2] A. P. Alivisatos, K. P. Johnson, T. E. Wilson, C. J. Loveth, M. P. Bruchez, and P. G. Schulz, *Nature (London)* **382**, 609–611 (1996).
- [3] E. Braun, Y. Eichen, U. Silvan, and G. Ben-Yoseph, *Nature (London)* **391**, 775–778 (1998).

- [4] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff, *Nature (London)* **382**, 607–609 (1996).
- [5] Y. Okata, T. Kobayashi, K. Tanaka, and M. Shimomura, *J. Am. Chem. Soc.* **120**, 6165–6166 (1998).
- [6] J. J. Storhoff, *Chem. Rev.* **99**, 1849–1862 (1999).
- [7] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, *Nature (London)* **394**, 539 (1998).
- [8] P. Hobza and J. Sponer, *Chem. Rev.* **99** 3247–3276 (1999).
- [9] 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**, 4992–4995 (2000).
- [10] D. L. Beveridge and K. J. McConnell, *Curr. Opin. Struct. Biol.*, **10**, 182–196 (2000).
- [11] C. Ban, B. Ramakrishnan, and M. Sundaralingam, *Biophys. J.* **71**, 1215–1221 (1996).
- [12] F. C. L. Bernstein, T. F. Koetzle, G. J. B. Williams, E. E. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.* **112**, 535–542 (1977).
- [13] M. L. Connolly, *J. Appl. Crystallogr.* **16**, 548–558 (1983).
- [14] A. D. Becke, *Phys. Rev. A* **38**, 3098–3100 (1988).
- [15] C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B*, **37**, 785–789 (1988).
- [16] N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993–2006 (1991).
- [17] CPMD v3.5. Copyright IBM Corp 1990–2001. Copyright MPI fuer Festkoerperforschung Stuttgart, 1997–2001.
- [18] N. Mazari and D. Vanderbilt, *Phys. Rev. B* **56**, 12 847–12 865 (1997).
- [19] P. L. Silvestrelli, N. Marzari, D. Vanderbilt, and M. Parrinello, *Solid State Commun.* **107**, 7 (1998).
- [20] M. Blackburn and M. J. Gait, *Nucleic Acids in Chemistry and Biology* (Oxford University Press, New York, 1996).
- [21] P. L. Silvestrelli and M. Parrinello, *Phys. Rev. Lett.* **82**, 3308–3311 (1999).
- [22] S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman, *J. Chem. Phys.* **59**, 2254 (1973).
- [23] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, Jr., David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollmann, *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
- [24] R. G. Enders, D. L. Cox, and R. R. P. Singh, *cond-mat/0201404*, pp. 1–23.
- [25] P. Tran, B. Alavi, and G. Gruner, *Phys. Rev. Lett.* **85**, 1564–1567 (2000).
- [26] T. Yamada and H. Fukutome, *Biopolymers* **6**, 43 (1968).
- [27] D. A. Greenwood, *Proc. Phys. Soc. London* **71**, 585 (1958).
- [28] R. J. Kubo, *Phys. Soc. Jpn.* **12**, 570 (1957).
- [29] F. Zaloudek, J. S. Novros, and L. B. Clark, *J. Am. Chem. Soc.* **107**, 7344 (1985).
- [30] M. P. Fulscher, L. Serrano-Andres, and B. O. Roos, *J. Am. Chem. Soc.* **119**, 6168–6176 (1997).
- [31] M. P. Fulscher and B. O. Roos, *J. Am. Chem. Soc.* **117**, 2089–2095 (1995).
- [32] Y. Matsuoka and B. J. Norden, *J. Phys. Chem.* **86**, 1378 (1982).