Influence of the sequence on the *ab initio* band structures of single and double stranded DNA models

Ferenc Bogár[a,c], Attila Bende[b,c,*], János Ladik[c]

[a] *MTA-SZTE Supramolecular and Nanostructured Materials Research Group of the Hungarian Academy of Sciences, University of Szeged, Dóm tér 8., 6720 Szeged, Hungary*
[b] *Molecular and Biomolecular Physics Department, National Institute for Research and Development of Isotopic and Molecular Technologies, Str. Donath 65-103, C.P. 700, Cluj Napoca RO-400293, Romania*
[c] *Chair for Theoretical Chemistry and Laboratory of the National Foundation for Cancer Research, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstr. 3, 91058 Erlangen, Germany*

Abstract

The solid state physical approach is widely used for the characterization of electronic properties of DNA. In the simplest case the helical symmetry is explicitly utilized with a repeat unit containing only a single nucleotide or nucleotide pair. This model provides a band structure that is easily interpretable and reflects the main characteristic features of the single nucleotide or a nucleotide pair chain, respectively. The chemical variability of the different DNA chains is, however, almost completely neglected in this way. In the present work we have investigated the effect of the different sequences on the band structure of periodic DNA models. For this purpose we have applied the Hartree-Fock crystal orbital method for single and double stranded DNA chains with two different subsequent nucleotides in the repeat unit of former and two different nucleotide pairs in the latter case, respectively. These results are compared to simple helical models with uniform sequences. The valence and conduction bands related to the stacked nucleotide bases of single stranded DNA built up only from guanidine as well as of double stranded DNA built up only from guanidine-cytidine pairs showed special properties different from the other cases. Namely, larger conduction and lower valence band positions and this way larger band gaps and smaller widths of these bands. With the introduction of non-uniform sequences containing guanidine became more similar to each other and to the other ones. The maximal band widths of the non-uniform sequences are considerably smaller than in the case of uniform sequences implying smaller charge carrier mobilities both in the conduction and valence bands.

*Corresponding author. Phone: 0040-264-584037 ext. 194; Fax: 0040-264-420042
Email addresses: bogar@sol.cc.u-szeged.hu (Ferenc Bogár), bende@itim-cj.ro (Attila Bende), Janos.Ladik@chemie.uni-erlangen.de (János Ladik)

1. Introduction

The electronic properties of DNA attracts remarkable scientific interest for a long time. Since the first experimental evidences of charge transport through a DNA chain [1, 2] several aspects of this process was investigated in the last twenty years using different experimental setups in solution, on solid surfaces and also single molecular measurements (for a recent review see Ref. [3]). Independently on the methods used the experiments proved that the charge transfer is very sensitive to the structural and environmental fluctuations.

These effects allow coherent charge transport first of all in the short range case at room temperature. The transport mechanism at longer ranges show a complex behaviour [4, 5] and is still investigated. However, at lower temperatures for a single DNA helix fixed on a solid surface, the band-like description of a perfectly symmetric chain can be a proper approximation. Such an experimental arrangement was reported by Shapir *et al.*[6]. Scanning tunneling spectroscopy was applied to measure the current-voltage characteristics and to derive the local density of states (DOS) of long DNA molecules (built of subsequent guanine-cytosine nucleotide pairs) deposited on a gold surface, at 78 K. The DOS obtained this way was compared to the results of a density functional Car-Parrinello calculation of the same DNA chain with Na^+ counter-ions. The peaks of the experimental and theoretical DOS-s were found to be similar in the negative energy region. It was also suggested that the measured gap of ~ 2.5 eV originates from energy difference between the guanine-type HOMO and the unoccupied impurity bands related to the Na^+ ions. The electron transport properties of long natural or quasi-random sequences can be also treated theoretically using parametrized tight-binding Hamiltonian for the description of the electronic structure and Landauer-Büttiker formalism for the calculation of I-V characteristics (see *e.g.* [7]).

DNA is a possible component of nanoelectronical devices. The conduction through a single DNA chain was measured using metallic [8, 9] and covalently bound carbon nanotubes as contacts [10]. Nanowires were fabricated by inlaying different metal ions (like Ag [11] or Au [12]) in its minor and major grooves. Although the charge transfer mechanism in DNA is under a continuous debate [13] we believe that the solid state physical approach can contribute to the understanding of properties of these nanoelectronic devices.

During the fifties and the sixties of the last century the *ab initio* Hartree-Fock Crystal Orbital (HFCO) theory was developed in a great extent [14–18]. Subsequently the theory was also generalized for the helical systems [19]. The application of these methods made possible a great success in the progress of the description of the electronic structure of polymers (including biopolymers) [18]. In two recent papers [20, 21] we have applied the HFCO method for calculation of the energy band structures of the four single stranded as well as the double stranded periodic DNA models [22] in the presence of water and Na^+ ions.

In the present paper we report Hartree-Fock crystal orbital energy band structure calculations for DNA using two different single nucleotides in the repeat unit of single stranded and two different nucleotide pairs in double stranded DNA model, respectively in the presence of counter ions. The water is not taken into account because the presence of water molecules induce only small changes in the framework of the HFCO model we apply as it was pointed out in a recent paper where we investigated the influence of the water molecules in the vicinity of Na^+ ions on the band structure of double stranded DNA built from a single nucleotide pair [22]. The results obtained this way are compared to those obtained for a simple helical model having a single nucleotide in the repeat unit of single stranded and a nucleotide pair in the double stranded model.

2. Methods

As it is well known the building blocks of DNA are the nucleotides built up from nucleotide bases (*i.e.* adenine (A), guanine (G), thymine (T), cytosine (C)) and the sugar-phosphate backbone that forms the framework of the double helix. In the followings we denote the nucleotides with the one letter code of the corresponding base with a tilde on it (*i.e.* $\tilde{A}$, $\tilde{G}$, $\tilde{T}$, $\tilde{C}$). The simplest periodic DNA model is a single stranded (SS) homopolynucleotide containing a single nucleotide (SN) in the repeat unit (we refer to it later as: SS-SN). Four different models belong to this class, we denote them by $p(\tilde{A})$, $p(\tilde{G})$, $p(\tilde{T})$, $p(\tilde{C})$, where p refers to the polymer. The corresponding double stranded (DS) model is built up from a single nucleotide pair (SNP) (*i.e.* $\tilde{G}-\tilde{C}$ or $\tilde{A}-\tilde{T}$) (DS-SNP). Only two different models of this kind can be formed: $p(\tilde{G}-\tilde{C})$ and $p(\tilde{A}-\tilde{T})$. These models reflect important characteristic features of the DNA structure but mostly neglect its chemical variability because of having a monotonic base sequence. A step towards the understanding the influence of the sequential variability on the electronic structure is to place two different building blocks in the repeat unit, that is two different nucleotides in the single stranded (SS-DN: single strand-double nucleotide) and two different base pairs in the double stranded (DS-DNP: double strand- double nucleotide pair) model. The SS-DN model family is the most populated one, having six members in the form of $p(\tilde{X}/\tilde{Y})$, where $\tilde{X}/\tilde{Y}$ are subsequent nucleotides along the single strand with the following possible cases: $\tilde{G}/\tilde{A}$, $\tilde{G}/\tilde{C}$, $\tilde{G}/\tilde{T}$, $\tilde{A}/\tilde{T}$, $\tilde{A}/\tilde{C}$, $\tilde{T}/\tilde{C}$. There are only four members belonging to the DS-DNP model system: $p(\tilde{G}/\tilde{A}-\tilde{C}/\tilde{T})$, $p(\tilde{G}/\tilde{T}-\tilde{C}/\tilde{A})$, $p(\tilde{G}/\tilde{C}-\tilde{C}/\tilde{G})$ and $p(\tilde{A}/\tilde{T}-\tilde{T}/\tilde{A})$. Here we used the convention that the first nucleotide in the first strand forms H-bonds with the first one in the second strand (i.e. $\tilde{G}$ with $\tilde{C}$ and $\tilde{A}$ with $\tilde{T}$ in the case of $p(\tilde{G}/\tilde{A}-\tilde{C}/\tilde{T})$).

The *ab initio* HF crystal orbital (CO) method in the linear combination of atomic orbitals (LCAO) approximation [15–17, 23], was used for the characterization of electronic properties of the above described DNA models. This method was also extended for general periodicity [19] (for instance helix operation in DNA B: translation + simultaneous

rotation both of the nuclei and the basis functions in the plane perpendicular to the long axis of the double helix) using the isomorphism of the translational group and the symmetry group belonging to the helix operation. This generalized theory was implemented in an extended version of J. Mintmire's PolyXa computer code [24], originally designed for the local density functional treatment of quasi 1D polymers.

As it is well known HF method overestimates the band gap periodic systems (see e.g. Chapter 2 of Ref. [18]). In a paper by two of us [25] the influence of the electron correlation (in the form of MP2 quasi-particle method) on the band structure of nucleotide base stacks was investigated. It was shown that valence bands were shifted upwards, the conduction bands downwards and consequently, the band gap decreased by 2-3 eV. In spite of the considerable changes in the numerical values the main characteristic features of the band structure (like *e.g.* the order of band gap values for the four nucleotide base) remained highly similar. From these results we can infer that the effect of electron correlation would give similar corrections for our systems. Therefore, the trends found in our HF investigations would remain valid for the correlation corrected band structures, as well.

The input geometries for the single and double stranded helices with different sequences used in our calculations were generated by the Ascalap Designer program [26]. In the case of single repeat units the helix operation in DNA B geometry going from one nucleotide base to the other one we shifted the atomic nuclei by 3.36 $\mathring{A}$ along the main axis of the helix and rotated them by 36^o around the same axis, for double repeat units these values were also doubled. As in our earlier calculations we used Clementi's double ζ basis set [27] in the LCAO expansion and 25 k-points in the first Brillouin zone for the k-space integration.

Mulliken population analysis was used for the identification of the dominating contributions to the orbital populations in order to distinguish the charge transport through the stacked nucleotide bases and the role of DNA backbone and counter ions.

3. Results and discussion

In our earlier studies we investigated several DNA models from the simple nucleotide base stacks to double stranded polynucleotide models in the presence of Na^+ counter ions and water molecules of the first solvatation shell (for a comparative evaluation see Ref. [22]). Here we recollect only the conclusions related to the topic of the present work namely the differences of the single stranded and double stranded polynucleotide models in the presence of Na^+ ions. We have found that all the bands of the DNA helix built up from nucleotide pairs $\tilde{G}-\tilde{C}$, $\tilde{A}-\tilde{T}$, ($p(\tilde{G}-\tilde{C})$ and $p(\tilde{A}-\tilde{T})$ using the notation introduced earlier) are shifted considerably in all cases upwards as compared to the single chain results ($p(\tilde{G})$, $p(\tilde{C})$, $p(\tilde{A})$ and $p(\tilde{T})$). This effect was explained by the $\sim 0.2e$ charge transfer from the sugars of both chains to the nucleotide bases. The fundamental gaps

between the nucleotide base-type highest filled and lowest unfilled bands were decreased by 1-3 eV in both cases, because the valence bands are purine-type and the conduction bands pyrimidine-type, respectively, while in the case of single homopolynucleotides they belong to the same base. We also pointed out that the lowest unoccupied orbital is mainly Na^+-type in both investigated cases and several unoccupied bands (belonging to the Na^+ ions and somewhat to the phosphate group) can be found between this and the first unoccupied pyrimidine-type empty band. For the better comparability we have recalculated these results using the geometry constructed with the same software [26], that is slightly different from that we had used earlier. The results are listed in Tables 1 and 2.

Table 1 and 2

The numerical values of valence and conduction bands as well as their widths are somewhat different but the above conclusions remained fully valid. In these tables as well as in Tables 3,4 we listed the energy band data of the highest occupied and lowest unoccupied crystal orbitals (HOCO, LUCO) having orbital populations mainly from the nucleotide bases (base-type crystal orbital). The short characterization of the crystal orbital populations can be found in the last column of these tables. The bands in the base-type gap belonging to the unoccupied orbitals with non-negligible contribution from the sugar-phosphate backbone and/or from the Na^+ counter ions are also listed in these tables.

The highest occupied (valence) and the lowest unoccupied (conduction) bands of the four investigated models (SS-SN, SS-DN, DS-SNP, DS-DNP) showed very similar character. The orbital population of the highest occupied crystal orbital is mainly from the base(s) (base-type). We can also identify the base-type empty band with the lowest energy (referred as base type conduction band). Between these two bands there are several bands belonging to a crystal orbital that has mostly contributions from the Na^+ counter ions and the phosphate groups. The one with the lowest energy will be referred later as non-base-type conduction band (for the schematic representation see Fig. 1).

Figure 1

In Table 3 we present the main characteristic features of the base-type highest occupied and lowest unoccupied bands of the single stranded model with two different, subsequent nucleotides in the repeat unit (SS-DN). In the single stranded case with a single nucleotide in the repeat unit (SS-SN) the $p(\tilde{G})$ showed exceptional behavior in a sense. The lower limit of the conduction band was the lowest (3.497 eV) and the upper limit of the valence band was the highest (-6.237 eV) and this way also the base-type gap was the lowest (9.734 eV) among the four homopolynucleotides. Changing every second guanine to an other base the band gap increases to $\sim$11 eV ($p(\tilde{G}/\tilde{A})$, $p(\tilde{G}/\tilde{T})$, $p(\tilde{G}/\tilde{C})$) even in the case

of adenine which is also a purine type base like guanine. This comes mainly from the upward shift of the conduction band due to its mixed character of LUCO. This orbital has contributions besides from guanine also from the other base (A, T or C) and also from the sugar phosphate backbone. For the remaining three sequences ($p(\tilde{A}/\tilde{T})$, $p(\tilde{A}/\tilde{C})$, $p(\tilde{T}/\tilde{C})$) we have obtained similar results with a somewhat larger gap between 11.427 eV and 12.065 eV.

The band widths also changed with the introduction of non-monotonic sequence in the single stranded case. For $p(\tilde{G})$, $p(\tilde{C})$, $p(\tilde{A})$ and $p(\tilde{T})$ we obtained 0.557, 0.104, 0.220 and 0.776 eV for the highest occupied and 0.487, 0.438, 0.078 and 0.093 eV for the lowest unoccupied base-type bands, respectively. In the non-monotonic case we calculated the largest conduction band widths for the $p(\tilde{T}/\tilde{C})$ and $p(\tilde{G}/\tilde{A})$ cases (0.116 eV and 0.114 eV) and the largest valence band width for $p(\tilde{G}/\tilde{C})$ (0.175 eV). This is the maximal band width in these systems implying that the expected charge carrier mobility is smaller for the non-monotonic sequences than for the monotonic ones in the single stranded case.

4-5 non-base type empty bands are between the valence and base-type conduction bands in the case of SS-DN model. The lower limits of the lowest lying non-base-type bands (having population from the phosphate groups and the counter ions) are around 0 eV (see Table 3), their widths are between 0.082 – 0.097 eV. In the SS-SN model the position of these bands are very similar but their width is larger, 0.263 – 0.284 eV.

Table 3

In Table 4 we listed the main characteristic features of the double stranded model with two different, subsequent nucleotide pairs in the repeat unit (DS-DNP). In the case of DS-SNP model in our earlier calculations we obtained characteristic differences between the $p(\tilde{G}-\tilde{C})$ and $p(\tilde{A}-\tilde{T})$ models. The base type valence band was positioned higher (with more than 1 eV) and the conduction band somewhat lower (with around 0.2 eV) in the former than in the latter case (see Table 2). This way the base-type band gap is also smaller for $p(\tilde{G}-\tilde{C})$ (9.00 eV) than the 10.40 eV obtained in the case of $p(\tilde{A}-\tilde{T})$.

Dropping the monotonous sequence used in the DS-SNP model (Table 4) caused gradual shift from the values obtained for $p(\tilde{G}-\tilde{C})$ in the direction of the value of $p(\tilde{A}-\tilde{T})$. To be more specific the upper limit of the valence band for the $p(\tilde{G}-\tilde{C})$ chain is -4.17 eV. Substituting every second $\tilde{G}-\tilde{C}$ nucleotide pair with an $\tilde{A}-\tilde{T}$ this quantity is shifted down to -4.62 eV. This relatively small change is due to fact that we substituted the purine (G) base with an other purine (A) and the C with an other pyrimidine base T. This way the change in the overlapping π-electron systems is small, which is supported by the dominantly G+A character of the valence band obtained from the population analysis. If we introduced pyrimidine type base after the guanine $p(\tilde{G}/\tilde{T}-\tilde{C}/\tilde{A})$, $p(\tilde{G}/\tilde{C}-\tilde{C}/\tilde{G})$, respectively, the upper limit of the valence band shifts further to the values of -4.96 eV and -4.86 eV, respectively. It is interesting to mention that in the former case the orbital

population is dominated by G with some further contribution from the subsequent T, but in the latter case the valence band has population from the guanines placed in zig-zag form on the two chains. For the only DS-DNP sequence without guanine $p(\tilde{A}/\tilde{T}-\tilde{T}/\tilde{A})$ the upper limit of the valence band is shifted further to -5.59 eV which is even lower than the -5.40 eV obtained for the $p(\tilde{A}-\tilde{T})$ model. The orbital population is dominated by the two adenines on an analogous way as it was described for the $p(\tilde{G}/\tilde{C}-\tilde{C}/\tilde{G})$.

The lower limit of conduction band shifted upwards from the 4.83 eV ($p(\tilde{G}-\tilde{C})$) to the values of 5.04, 5.26, 5.41 and 5.41 eV for the $p(\tilde{G}/\tilde{A}-\tilde{C}/\tilde{T})$, $p(\tilde{G}/\tilde{T}-\tilde{C}/\tilde{A})$, $p(\tilde{G}/\tilde{C}-\tilde{C}/\tilde{G})$ and $p(\tilde{A}/\tilde{T}-\tilde{T}/\tilde{A})$ models, respectively. The conduction band has a no-negligible population from the sugar-phosphate backbone in every case. The base contributions show the same pattern as in the valence band case. The conduction band of $p(\tilde{G}/\tilde{A}-\tilde{C}/\tilde{T})$ has a C+T, $p(\tilde{G}/\tilde{T}-\tilde{C}/\tilde{A})$ a C+A character. In case of $p(\tilde{G}/\tilde{C}-\tilde{C}/\tilde{G})$ it is dominated by the two cytosines and for $p(\tilde{A}/\tilde{T}-\tilde{T}/\tilde{A})$ by the two thymines (last column of Table 4).

The band widths in the DS-DNP model are in general smaller than in single pair case (DS-SNP). The highest value was obtained for $p(\tilde{G}/\tilde{A}-\tilde{C}/\tilde{T})$, 0.106 eV and 0.185 eV for the valence and conduction bands, respectively (Table 4). The corresponding values for DS-SNP model are 0.62 eV and 0.31 eV in the $p(\tilde{G}-\tilde{C})$ case.

Nine to ten non-base type empty bands are situated in the gap in the case of the DS-DNP model. The lower limit of the lowest lying non-base-type bands are between 0.1 – 0.3 eV their widths are between 0.082 – 0.094 eV, their population in mainly coming from the phosphate groups and the counter ions (Table 4). The positions of these bands in the DS-SNP model are very similar but they are somewhat wider 0.265 eV and 0.257 eV in the cases of $p(\tilde{G}-\tilde{C})$ and $p(\tilde{A}-\tilde{T})$, respectively. The band gap between the base type valence and the non-base-type conduction band of the chains in the DS-DNP model is approximately equal to the negative of the upper limits of the valence bands and this way considerably smaller than the base-type gap.

Table4

It is worth mentioning that non-base type empty bands are situated in the gap probably play an important role in charge transport at low temperature for a DNA chain fixed on a solid surface as it was suggested by Porath and his coworkers [6]. However in water solution at normal temperature the fluctuations of ions and the surrounding water molecules most probably excludes these ion related states from these processes. At the same time the interaction of DNA with the hydrated counter ions results in changes of geometry and electronic structure of DNA that considerably influences the charge transport, as well [28].

Conclusion

In this study we investigated the influence of base sequence on the Hartree-Fock crystal orbital energy band structure of single and double stranded DNA models. For this purpose we compared the cases when the repeat unit contained one or two different single nucleotides for single stranded further one or two different nucleotide pairs in the double stranded DNA model in the presence of Na^+ counter ions. The band structure of single stranded DNA built up only from guanidine as well as of double stranded DNA built up only from guanidine-citidine pairs showed special properties different from the other sequences. Namely, they have the highest lying valence bands and the smallest fundamental gaps. If one introduce non-monotonic sequences containing $\tilde{G}$, but also other-type of nucleotides, the characteristic quantities determining the conduction properties (mobilities) change in a way that the unique properties of $p(\tilde{G})$ become less significant.

Acknowledgments

We should like to express our gratitude to Professor Enrico Clementi by giving his unpublished basis sets for Na and P atoms. Further we are indebted to Professor Ferenc Beleznay advising us many times about different solid state physics problems. The numerical work has been carried out using the facilities of the Hungarian National Information Infrastructure Development Institute and the DataCenter of INCDTIM Cluj-Napoca. We gratefully acknowledge their support.

References

[1] D. D. Eley, D. I. Spivey, Trans. Faraday Soc. 58 (1962) 411. doi:10.1039/TF9625800411

[2] C. J. Murphy, M. R. Arkin, Y. Jenkins, N. D. Ghatlia, S. H. Bossmann, N. J. Turro, J. K. Barton, Science 262 (1993) 1025. doi:10.1126/science.7802858.

[3] N. B. Muren, E. D. Olmon, J. K. Barton, Phys. Chem. Chem. Phys. 14 (2012) 13754. doi:10.1039/c2cp41602f.

[4] R. G. Sarmento, G. A. Mendes, E. L. Albuquerque, U. L. Fulco, M. S. Vasconcelos, O. Ujsághy, V. N. Freire, E. W. S. Caetano, Phys. Lett. A 367 (2012) 2413. doi:10.1016/j.physleta.2012.05.058

[5] R. G. Sarmento, E. L. Albuquerque, P.D. Sesion Jr., U.L. Fulco, B.P.W. de Oliveira, Phys. Lett. A 373 (2009) 1486 doi:10.1016/j.physleta.2009.02.043

[6] E. Shapir, H. Cohen, A. Calzolari, C. Cavazzoni, D. A. Ryndyk, G. Cuniberti, A. Kotlyar, R. Di Felice, D. Porath, Nat. Mater. 7 (2008) 68. doi:10.1038/nmat2060.

[7] L.M. Bezerril, D.A. Moreira, E.L. Albuquerque, U.L. Fulco, E.L. de Oliveira, J.S. de Sousa, Physics Letters A 373 (2009) 3381. doi:10.1016/j.physleta.2009.07.028

[8] D. Porath, A. Bezryadin, S. de Vries, C. Dekker, Nature 403 (2000) 635. doi:10.1038/35001029.

[9] E. Shapir, G. Brancolini, T. Molotsky, A. B. Kotlyar, R. Di Felice, D. Porath, Adv. Mater. 23 (2011) 4290. doi:10.1002/adma.201101100.

[10] X. F. Guo, A. A. Gorodetsky, J. Hone, J. K. Barton, C. Nuckolls, Nat. Nanotechnol. 3 (2008) 163. doi:10.1038/nnano.2008.4.

[11] S. A. F. Al-Said, R. Hassanien, J. Hannant, M. A. Galindo, S. Pruneanu, A. R. Pike, A. Houlton, B. R. Horrocks, Electrochem. Commun. 11 (2009) 550. doi:10.1016/j.elecom.2008.12.031.

[12] S. R. Quake, A. Scherer, Science 290 (2000) 1536. doi:10.1126/science.290.5496.1536.

[13] J. C. Genereux, J. K. Barton, Chem. Rev. 110 (2010) 1642. doi:10.1021/cr900228f.

[14] P.-O. Löwdin, Adv. Phys. 5 (1956) 1. doi:10.1080/00018735600101155

[15] G. Del Re, J. Ladik, G. Biczó, Phys. Rev. 155 (1967) 997. doi:10.1103/PhysRev.155.997

[16] J.-M. André, L. Gouverneur, G. Leroy, Int. J. Quantum Chem. 1 (1967) 427. doi:10.1002/qua.560010415

[17] J.-M. André, L. Gouverneur, G. Leroy, Int. J. Quantum Chem. 1 (1967) 451. doi:10.1002/qua.560010416

[18] J. Ladik, Quantum theory of polymers as solids, Plenum Press, New York and London, 1988.

[19] A. Blumen, C. Merkel, Phys. Status Solidi (b) 83 (1977) 425. doi:10.1002/pssb.2220830208

[20] J. Ladik, A. Bende, F. Bogár, J. Chem. Phys. 127 (2007) 055102. doi:10.1063/1.2752806.

[21] J. Ladik, A. Bende, F. Bogár, J. Chem. Phys. 128 (2008) 105101. doi:10.1063/1.2832860.

[22] A. Bende, F. Bogár, J. Ladik, Solid State Commun. 151 (2011) 301. doi:10.1016/j.ssc.2010.12.001.

[23] J. Ladik, Physics Reports 313 (1999) 171. doi:10.1016/S0370-1573(98)00087-8.

[24] J. W. Mintmire, in Density functional methods in chemistry, (Editors, J. Labanowski and J. Andzelm) Springer-Verlag, pp 125, 1991. doi:10.1007/978-1-4612-3136-3_9

[25] F. Bogar J. Ladik, Chem. Phys. 237 (1998) 273. doi:10.1016/S0301-0104(98)00261-4

[26] F. Miller, A. Vandome, M. John, VDM Publishing, 2010.

[27] L. Gianolo, E. Clementi, Gazz. Chim. Ital. 110 (1980) 79.

[28] R.N. Barnett, C.L. Cleveland, A. Joy, U. Landman, G.B. Schuster, Science 294 (2001) 567. doi:10.1126/science.1062864

Figure Captions:

Figure 1. Schematic representation of the band structure of the investigated systems, representing their common features.

Table 1: The influence of the sequence on the main characteristic features (l.l.:lower limit, w.:width and gap) of valence band (VB), base-type and non-base-type conduction bands (CB) of single stranded DNA models having a single nucleotide in the repeat unit. The main contributions of the corresponding crystal orbital populations are indicated in the last column (Type). All energy values are in eV-s. The widths of bands are written in bold face if they are larger than 0.1 eV (the thermal energy at 300K is 0.026 eV).

$p(\tilde{G})$	Base-type CB	l.l.[a] (w.[b])	3.497 (0.487)	G + PO_4+ Sugar
		Gap	9.734	
	Non-base-type CB	l.l. (w.)	0.105 (0.263)	Na + PO_4
		Gap	6.312	
	VB	u.l.[c] (w.)	-6.237 (**0.557**)	G
$p(\tilde{C})$	Base-type CB	l.l. (w.)	4.370 (0.438)	C + PO_4+ Sugar
		Gap	11.484	
	Non-base-type CB	l.l. (w.)	-0.168 (0.284)	Na + PO_4
		Gap	6.946	
	VB	u.l. (w.)	-7.114 (**0.104**)	C
$p(\tilde{A})$	Base-type CB	l.l.(w.)	5.104(0.078)	A + PO_4+ Sugar
		Gap	11.993	
	Non-base-type CB	l.l.(w.)	-0.023(0.276)	Na + PO_4
		Gap	6.863	
	VB	u.l.(w.)	-6.886(**0.220**)	A
$p(\tilde{T})$	Base-type CB	l.l. (w.)	4.214 (0.093)	T + PO_4+ Sugar
		Gap	11.765	
	Non-base-type CB	l.l. (w.)	-0.147 (0.282)	Na + PO_4
		Gap	7.833	
	VB	u.l. (w.)	-7.551 (**0.776**)	T

[a] lower limit
[b] width
[c] upper limit

Table 2: Same as Table 1 but for the double stranded DNA having one nucleotide pair in the repeat unit. All energy values are in eV-s.

$p(\tilde{G}-\tilde{C})$	Base-type CB	l.l. (w.)	4.835 (0.310)	C + PO_4+ Sugar
		Gap	9.004	
	Non-base-type CB	l.l. (w.)	0.108 (0.265)	Na + PO_4
		Gap	4.434	
	VB	u.l. (w.)	-4.169 (**0.624**)	G
$p(\tilde{A}-\tilde{T})$	Base-type CB	l.l. (w.)	4.995 (0.295)	T + PO_4+ Sugar
		Gap	10.400	
	Non-base-type CB	l.l. (w.)	0.243 (0.257)	Na + PO_4
		Gap	5.648	
	VB	u.l. (w.)	-5.405 (**0.208**)	A

Table 3: Same as Table 1 but for single stranded DNA having two different nucleotides on the top of each other in the repeat unit. All energy values are in eV-s.

$p(\tilde{G}/\tilde{A})$	Base-type CB	l.l.(w.)	4.807 (0.156)	G + A +Sugar + PO_4
		Gap	11.161	
	Non-base-type CB	l.l. (w.)	0.051 (0.273)	Na + PO_4
		Gap	6.627	
	VB	u.l. (w.)	-6.354 (**0.114**)	G + A
$p(\tilde{G}/\tilde{T})$	Base-type CB	l.l. (w.)	4.638 (0.171)	G + T +Sugar + PO_4
		Gap	11.207	
	Non-base-type CB	l.l. (w.)	-0.003 (0.277)	Na + PO_4
		Gap	6.566	
	VB	u.l. (w.)	-6.569 (0.033)	G + few T
$p(\tilde{G}/\tilde{C})$	Base-type CB	l.l.(w.)	4.748 (0.175)	G + C +Sugar + PO_4
		Gap	10.973	
	Non-base-type CB	l.l. (w.)	-0.012 (0.276)	Na + PO_4
		Gap	6.213	
	VB	u.l. (w.)	-6.225 (0.006)	G + few C
$p(\tilde{A}/\tilde{T})$	Base-type CB	l.l. (w.)	4.938 (0.036)	A + T +Sugar + PO_4
		Gap	12.065	
	Non-base-type CB	l.l. (w.)	-0.086 (0.094)	Na + PO_4
		Gap	7.041	
	VB	u.l. (w.)	-7.127 (0.035)	A + T
$p(\tilde{A}/\tilde{C})$	Base-type CB	l.l. (w.)	4.591 (0.054)	A + C +Sugar + PO_4
		Gap	11.427	
	Non-base-type CB	l.l. (w.)	-0.094 (0.280)	Na + PO_4
		Gap	6.742	
	VB	u.l. (w.)	-6.836 (0.007)	A + few C
$p(\tilde{T}/\tilde{C})$	Base-type CB	l.l. (w.)	4.295 (0.088)	T + C +Sugar + PO_4
		Gap	11.696	
	Non-base-type CB	l.l. (w.)	-0.158 (0.099)	Na + PO_4
		Gap	7.243	
	VB	u.l. (w.)	-7.401 (**0.116**)	T + C

Table 4: Same as Table 1 but for the double stranded DNA model having two nucleotide pairs in the repeat unit. All energy values are in eV-s.

$p(\tilde{G}/\tilde{A}-\tilde{C}/\tilde{T})$	Base-type CB	l.l. (w.)	5.041 (0.185)	C + T + PO_4 + Sugar
		Gap	9.664	
	Non-base-type CB	l.l. (w.)	0.175 (0.091)	Na + PO_4
		Gap	4.798	
	VB	u.l. (w.)	-4.623 (**0.106**)	G + A
$p(\tilde{G}/\tilde{T}-\tilde{C}/\tilde{A})$	Base-type CB	l.l. (w.)	5.262 (0.147)	C + A + PO_4 + Sugar
		Gap	10.225	
	Non-base-type CB	l.l. (w.)	0.298 (0.097)	Na + PO_4
		Gap	5.261	
	VB	u.l. (w.)	-4.963 (0.019)	G + few T
$p(\tilde{G}/\tilde{C}-\tilde{C}/\tilde{G})$	Base-type CB	l.l.(w.)	5.408 (0.137)	C_1 + C_2+ PO_4 + Sugar
		Gap	10.271	
	Non-base-type CB	l.l. (w.)	0.370 (0.094)	Na + PO_4
		Gap	5.233	
	VB	u.l. (w.)	-4.863 (0.055)	G_1 + G_2
$p(\tilde{A}/\tilde{T}-\tilde{T}/\tilde{A})$	Base-type CB	l.l. (w.)	5.410 (0.109)	T_1 + T_2+ PO_4 + Sugar
		Gap	11.005	
	Non-base-type CB	l.l. (w.)	0.357 (0.082)	Na + PO_4
		Gap	5.952	
	VB	u.l. (w.)	-5.595 (0.089)	A_1 + A_2

Figure
Click here to download high resolution image