Calculation of the band structure of polyguanilic acid in the presence of water and Na^+ ions

J. Ladik[a)]
Institute for Theoretical Chemistry, Friedrich Alexander University, Erlangen-Nürnberg, Egerlandstrasse 3, 91058 Erlangen, Germany and Laboratory of the National Foundation for Cancer Research, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstrasse 3, 91058 Erlangen, Germany

A. Bende
National Institute for Research and Development of Isotopic and Molecular Technologies, Str. Donath 71-103, Caixa Postal 700, Cluj Napoca RO-400293, Romania, Institute for Theoretical Chemistry, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstrasse 3, 91058 Erlangen, Germany, and Laboratory of the National Foundation for Cancer Research, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstrasse 3, 91058 Erlangen, Germany

F. Bogar
Supramolecular and Nanostructured Materials Research Group of the Hungarian Academy of Sciences, University of Szeged, Dóm tér 8, H-6720 Szeged, Hungary, Institute for Theoretical Chemistry, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstrasse 3, 91058 Erlangen, Germany, and Laboratory of the National Foundation for Cancer Research, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstrasse 3, 91058 Erlangen, Germany

(Received 18 April 2007; accepted 4 June 2007; published online 3 August 2007)

Using the Hartree-Fock crystal orbital method with a combined symmetry (helix) operation, the band structure of polyguanilic acid was calculated in the presence of water and Na^+ ions. The water structure was optimized with the help of molecular mechanics.The obtained band structure shows that both the valence and conduction bands are purely guanine type. The three impurity bands in the 10.66 eV large gap are close to the conduction band and therefore cannot play any role in the assumed hole conduction of the system. Namely, according to detailed x-ray diffraction investigations of the nucleosomes in chromatin, there are possibilities of charge transfer from the negative sites of DNA to the positive ones in histones. Therefore most probably there is a hole conduction in DNA and an electronic one in the histone proteins. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2752806]

I. INTRODUCTION

Recently the crystal structure of nucleosomes (they build up the chromatins from which the chromosomes are formed) consisting of eight histone molecules in the nucleosome core particle (NCP)[1] and of a 147 base pair long DNA superhelix in its B form[2] was determined with the help of x-ray diffraction. These experiments were performed at rather large resolution (2.8 Å for NCP and 1.9 Å for DNA) by the Richmond Group at ETH.

According to their detailed data the DNA superhelix is wrapped around NCP. It has 1.67 turns[2] and is left handed. This complicated packing of the DNA superhelices around the NCPs causes in the nucleus of a single eukaryotic (differentiated) cell a 3.5 km long DNA chain to be accommodated.

At the larger resolution[2] it has turned out that contrary to an ideal 147 base pair long superhelix (B form of DNA, uniformly distributed base pairs) in the nucleosomes DNA is not bent uniformly because of the anisotropic flexibility of DNA, local structural deviations from the ideal form, etc. These features (irregular bending and twisting of the DNA superhelix causing excess DNA curvatures) most probably are the reason of the ability of the NCPs to "slide" along the DNA without releasing it.[3]

The interaction between the DNA and histone molecules in the nucleosomes is first of all caused by the charge transfer (CT) between the negative groups (phosphate groups, the O atoms of the thymine as well of the guanine nucleotide bases) and the positive side chains (arginine, lysine, and histidine) of the histone molecules. Additionally, there are attractive interactions between the PO_4^- groups of DNA and the dipole moments of the main chains of the proteins. Of course these interactions are significant only if the negative sites of DNA face the positive ones in the proteins.

On the basis of the detailed data given in Ref. 1 the number of charge transfers between DNA and histones via H bonds can be estimated to be roughly at 40 sites of the DNA superhelix that is at every fourth base pair.[1]

In a previous paper[4] we have described the calculation of CT between a PO_4^- group of DNA and a positive arginine or lysine side chain of a histone molecule. The *ab initio* calculations were performed with the aid of a sophisticated basis set (triple ξ+ polarization functions at the valence orbitals) for the HF and a subsequent Møller-Plesset second order many-body perturbation theoretical (MP2) computation. In

[a)]Electronic mail: janos.ladik@chemie.uni-erlangen.de

both cases the geometry of the systems was fully optimized starting from standard H-bond geometries. We have obtained a CT of $0.21e$ in the arginine case and $0.26e$ in the lysine case.

The above described deviations of the DNA superhelix from its ideal form, as mentioned above,[2] increase its curvatures, which very probably cause still more opportunities for CT between the two macromolecules. On the other hand the CT between the O atoms of DNA and the positive sites of the histone molecules as well the CT between a dipole moment of the main chain and negative sites of DNA are smaller than the above given values of $0.26e$ and $0.21e$, respectively.[4] Therefore as a conservative estimate we assume as average $0.10e$ CT from DNA to the histones at every fourth base pair.

The occurrence of this CT in a nucleosome causes a hole current in DNA and an electronic one in the histones. For proteins, such as the histones, one can assume infinite crystal orbitals—as suggested by Coulson[5]—through the N–H···O=C H bonds which are perpendicular to the directions of the main chain of a protein (if it forms an α helix or is in a β pleated sheet).

On the other hand if by some external disturbances (binding of foreign chemical molecules including carcinogens, radiations, etc.) the mutual positions of the negative groups in DNA and the positive ones in the histones can change in a way that charge transfers between them cannot occur or occur in a much lesser degree. In that case the hole and electron conductivities, respectively, could diminish so much that they become insulators. If this happens, as Luger *et al.* pointed out,[1] the DNA superhelix will move away from the NCPs (the eight histone molecules) and therefore all its genetic information becomes readable. If this happens, at the same time at several near lying nucleosomes, this causes a large perturbation via the well known biochemical mechanisms[6,7] of the self-regulation of the cell, so that it can go over to an unusual stationary state. Some of these unusual stationary states have the properties of a precancerous state, that is, these events can lead to cancer initiation.

To look more into the details of the above qualitatively described course of events, we have started a series of calculations of the band structures of the four polynucleotides in the presence of water and Na^+ ions.

On the basis of the band structures we plan to compute—using the simple deformation potential approximation[8]—also their mobilities due to the interaction of the electrons with the acoustic phonons belonging to the dilatation and contraction of the stack and to the twisting motion of the helix, respectively. From these mobilities by multiplying them with the number of charge carriers and the elementary charge we are going to determine the dc hole conductivities of the polynucleotides.

As a first step in this program we have calculated the band structure of polyguanilic acid in the presence of water and Na^+ ions. (It should be mentioned that the motilities of this system without water and Na^+ ions have been calculated and published before.[9,10])

II. METHODS

As first steps in the research program outlined in the Introduction, after the computation of the CT from the PO_4^- groups of DNA to the arginine and lysine side chains of the histones,[4] the band structure of polyguanilic acid in the presence of water and Na^+ ions was calculated with the aid of the *ab initio* Hartree-Fock (HF) crystal orbital (CO) method in its linear combination of atomic orbitals (LCAO) form.[11–13] This method is a generalization of the calculation with a simple Bloch function (one orbital per unit cell) to the case of arbitrary number of basis functions in the unit cell. Therefore a crystal orbital contains besides the summation according to the unit cells also a summation over the basis functions. In other words in a CO there is a summation over the Bloch functions, each one multiplied with a coefficient (LCAO approximation). This procedure leads to a matrix eigenvalue formalism to determine the LCAO coefficients. In this algorithm one has to perform the self-consistent-field (SCF) iterations for every k value separately. The different k-dependent self-consistent eigenvalues provide then the band structure.[11–13]

The method takes into account the translational symmetry, and by using periodic boundary conditions one can calculate infinite (or very long) chains. As a further step it was realized that the method is not restricted to simple translations, but it can be used (as was proven with the help of group theory) in the case of any kind of periodic symmetry of a one-dimensional system.[14] It can be shown that in the case of a helix operation (translation+rotation) from going from one unit to the next one has to put not only the nuclei in the right position but also those basis functions which are not parallel with the helix axis have to be rotated with the torsion angle of the helix (in the case of DNA by 36°). In our polymer program this modification, in contrary to the usual programs, is built in.

For the calculations Gianolo and Clementi's double ξ basis set[15] was applied. We are aware of the fact that simple HF calculations without correlation gives far too large gaps between the bands and usually too large bandwidths. Since, however, experimental ionization potentials and electron affinities are available for the nucleotide bases (for more details and references see Sec. III) one can correct at least the positions of the valence and conduction bands.

For the first half Brillouin zone 30 different k values were applied.

To determine the water structure around the guanilic acid units the following procedure was applied.

(1) The starting geometry of a triple G-C system with sugar and phosphate+Na^+ was generated with the help of the SYBYL program.[16]
(2) 263 H_2O molecules were put randomly around this system.
(3) Next, the geometry of the system was optimized using molecular mechanics [Amber force field from GAUSSIAN03 (Ref. 17)]. The reason for using two different methods (which are approximately equivalent in their speed and accuracy) is that steps (1) and (3) were performed at two different places (at Szeged the SYBYL

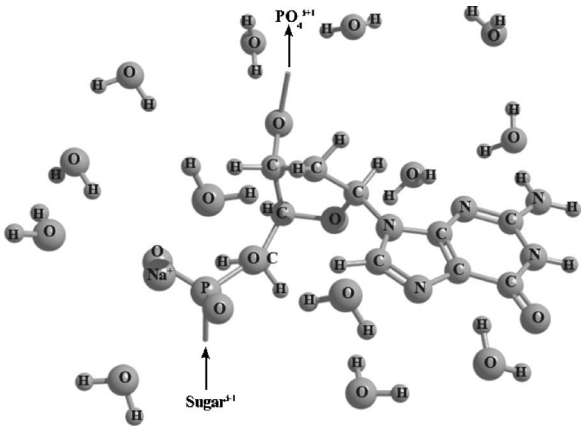

FIG. 1. The unit cell of polyguanilic acid together with the water structure around it (13 H_2O molecules) and the position of the Na^+ ion.

program is available and in Erlangen GAUSSIAN03).

(4) After this the upper and lower G-C nucleotide pairs together with the water molecules which have belonged to them were eliminated.

(5) Also the H_2O molecules on the outer part of the cylindrical mantle were eliminated until only 27 H_2O molecules remained.[18]

(6) Finally the cytidine molecules with their water molecules and Na^+ ions were left out. The remaining system consisted of (as mentioned before) as unit cell a guanilic acid+13 H_2O molecules+1 Na^+ ion (the Na^+ ion is between the two O atoms of PO_4^- which do not take part in H bonds, in the plane determined by the P atom and the two O atoms).

For this system the HF band structure has been determined. To reach self-consistency about 240 h were needed on an FSC Primergy Fujitsu-Siemens computer using one of its AMD Opteron processors. This very large CPU time is first of all due to the large number of basis functions (443). This causes a very large number of two-electron integrals. The threshold used for them was 10^{-6} a.u. The number of SCF iterations was 45 (with a SCF criterion of 10^{-6} in the elements of the charge-bond order matrix). About one-half of the CPU time was used for the generation of the integrals and the other half for the SCF iterations. Though the computer had a core memory of 4 GByte, because of the very large amount of data, one has all the time to transfer them to the hard disk or bring them back to the memory. These data transfers slow down very much the calculations.

We are aware that there are methods built in into the GAUSSIAN03 program which accelerate very much the HF calculations, but they are not implemented yet in our polymer program.

In Fig. 1 we show the unit cell of polyguanilic acid together with the 13 water molecules and Na^+ ions. Figure 2 shows the chemical formula of a cytidine (cytosine+sugar +phosphate) nucleotide.

Finally, one should answer the following question: Why a simple HF and not a correlation containing method [density functional theory (DFT) or MP2] was used for the calculation? Our program was developed during a longer time period (building in to it the case of combined symmetry operations and allowing an arbitrary number of basis functions). Both these options are missing in the polymer programs offered even by the latest GAUSSIAN package. Therefore one cannot say that in our calculations we were running a routine program.

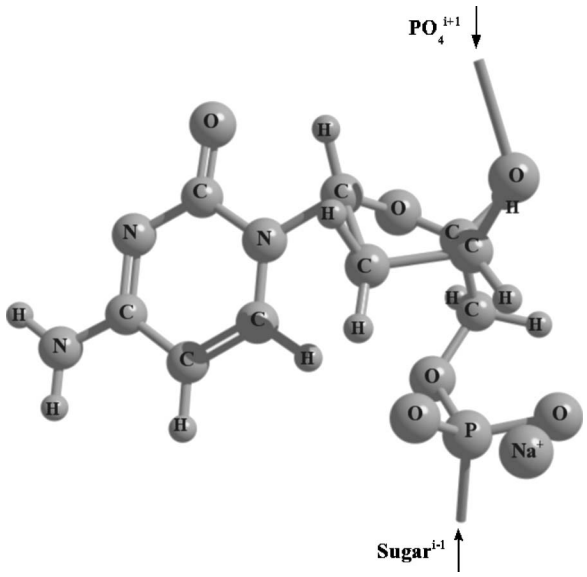

FIG. 2. The chemical formula of the cytidine nucleotide.

The program contains also the application of the local density approximation (LDA), PBE, BLYP, and B3LYP DFT methods. The reasons why we could not use them are as follows: (1) The DFT methods with exception of the hybrid ones (such as B3LYP) give far too small gaps in periodic chains,[19] and generally they are not applicable for systems with many H bonds and other weak interactions.[20]

We are thinking to apply the MP2 version of the program, but the workstations presently available to us are not large enough (by core and speed) to perform calculations on a such huge system as a homopolynucleotide with water environment even if we would introduce methods which accelerate the four-index transformation. Perhaps, this will be possible with the next (or second-next) generation of workstations. For the same reason we cannot calculate a double strand of DNA instead of a single helix with its environment even at the HF level.

Finally, we should like to point out that we permanently improve our program and not just run one which exists a longer time.

III. RESULTS AND DISCUSSION

Table I shows the physically interesting bands of polyguanilic acid in the presence of water molecules and Na^+ ions.

As one can see the fundamental gap is 10.66 eV (see also Fig. 3). Both the valence and the conduction bands (as the investigation of the eigenvectors belonging to them shows) are purely guanine-type bands. On the other hand the gap is considerably larger than in the case of polyguanilic

TABLE I. The physically interesting bands of polyguanilic acid in the presence of water. In the first column the role of the bands are indicated, and in the second one the positions of their upper limits (u.l.) and of the lower limits (l.l.) (all in eV s) are given with the corresponding k values and the bandwidths (W) (again in eV s). In the third column the origins of the bands are indicated.

The first dominantly phosphate band[d]	u.l.	-10.52 ($k=7/30\ \pi/a$)	Dominantly P[a] with some admixture of S[b]
	l.l.	-10.69 ($k=23/30\ \pi/a$)	
	W	0.17	
Valence band	u.l.	-6.81 ($k=\pi/a$)	Purely: G
	l.l.	-7.17 ($k=0$)	
	W	0.36	
(1). Imp. band[e]	u.l.	2.32 ($k=\pi/a$)	Mostly W[c] and Na^+
	l.l.	1.74 ($k=0$)	
	W	0.58	
(2). Imp. band	u.l.	3.07 ($k=\pi/a$)	Mostly Na^+ with W and some
	l.l.	2.66 ($k=0$)	
	W	0.41	
(3). Imp. band[f]	u.l.	3.72 ($k=0$)	Mostly Na^++W and some S and P contributions
	l.l.	3.85 ($k=\pi/a$)	
	W	0.44	
Conduction band[g]	u.l.	4.43 ($k=\pi/a$)	Purely: G
	l.l.	3.85 ($k=0$)	
	W	0.58	

[a]P=phosphate.
[b]S=sugar.
[c]W=water.
[d]The distance of the first dominantly P-type filled band from the valence band is 3.35 eV.
[e]The distance of the lowest unfilled impurity band from the valence band is 8.55 eV.
[f]The distance of the highest unfilled impurity band from the conduction band is only 0.13 eV.
[g]The fundamental gap is 10.66 eV.

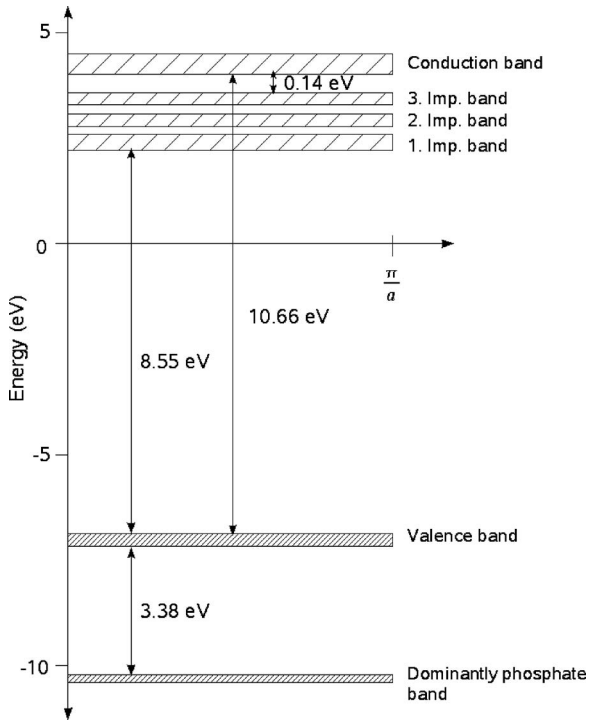

FIG. 3. The positions of the valence and conduction bands together with the three unfilled impurity bands and the highest lying filled dominantly phosphate-type band in polyguanilic acid in the presence of water and Na^+ ions.

acid in the absence of H_2O molecules and Na^+ ions calculated with the same basis set (7.76 eV, see Table I of Ref. 10).

In the gap there are three impurity bands in the neighborhood of the conduction band belonging first of all to the water molecules and Na^+ ions and in a lesser extent to the sugar and phosphate groups. On the other hand their eigenvectors do not contain any components belonging to the guanine molecules. The highest lying impurity band lies only by 0.14 eV lower than the lower edge of the conduction band. On the other hand there is no impurity band in the neighborhood of the valence band (the nearest lies by 8.55 eV higher than the upper edge of the valence band). From all this follows that there is no possibility for hole conduction due to the water and Na^+ ions impurities. [Since the transfer of a negative charge from polyguanilic acid to the histone molecules of the nucleosome particles occurs from DNA to these proteins, there is no possibility for electronic conduction in polyguanilic acid (unless electrons are injected into it from outside).] On the other hand one would expect a hole conduction in DNA because of this CT.

The upper edge of the valence band at -6.81 eV lies only in a small degree higher than in the case of the absence of water molecules [-7.14 eV (Ref. 10)]. On the other hand the lower edge of the conduction band of +3.85 eV is higher than in the absence of water [+3.20 eV (Ref. 10)]. Most probably the location of the three impurity bands causes this upward shift of the conduction band.

The width of the valence band (0.36 eV) has changed only little as compared to the case of naked polyguanilic acid [0.32 eV (Ref. 10)], while the width of the conduction band is non-negligibly larger (0.58 eV) than in the case of the absence of water [0.43 eV (Ref. 10)]. The three impurity bands in the gap (0.58, 0.41, and 0.44 eV) have similar widths as the valence band.

Above the conduction band there are five impurity bands and a sugar-type band spreading over 1.2 eV s until one reaches the next dominantly G-type empty band which has a width of 0.44 eV. These impurity bands have widths between 0.10 and 0.66 eV.

Under the valence band there are eight dominantly guanine-type bands until one reaches a dominantly phosphate-type band (with a small width of 0.02 eV) at 3.35 eV below the lower edge of the valence band. All these band have small widths (0.01–0.16 eV).

From all these band structure data we can conclude that in polyguanilic acid in the presence of water and Na^+ there is a possibility of hole conduction if the system can give away some negative charges, but impurities do not play any role in the conduction.

As we have mentioned before the HF method provides too large gap values and usually also the bandwidths are overestimated. In the case of the four nucleotide bases there are experimentally determined electron affinities (besides the

TABLE II. Gianolo and Clementi's ζ HF highest unoccupied molecular orbital–lowest unoccupied molecular orbital energy difference of a single guanine molecule scaled down with the help of the experimental value, the fundamental gaps of a guanine stack, and of polyguanilic acid in the absence and presence of water, respectively (in eV s).

	G	G stack	Polyguanilic acid (no H_2O)	Polyguanilic acid (with H_2O)
Calculated HF value	11.96	10.43[a].	10.34	10.66
Scaled down value with the aid of the factor of 0.55	6.54	5.74	5.69	5.86

[a]In Ref. 10 erroneously the first nonguanine-type band (dominantly Na^+ type) was taken as conduction band. If we take the first G-type empty band as conduction band (in the case with water its lower limit lies at 3.85 eV) on the basis of the shift of the upper limit of the valence band of a G stack [from −7.76 (Ref. 9) to −7.14 eV in the case of polyguanilic acid without water (Ref. 10) one can estimate the position of the lower limit of the conduction band to be at 3.20 eV (2.57+0.60 eV≈3.20 eV).

ionization potentials).[21] For guanine the experimental ionization potential is 7.77 eV and its vertical electron affinity is −1.23 eV.[21] Their difference is [the ionization potential has to be taken here with a negative sign, because it corresponds to the highest occupied molecular orbital (HOMO) energy] −7.77−(−1.23)=−6.54 eV. In the case of a stack one would expect that due to band broadening effects this value has to be somewhat smaller. On the other hand for a guanine stack the double ξ HF calculation has provided a gap of 10.27 eV.[9] If one takes the ratio of 6.54/11.96=0.55, one has to multiply by this factor to the theoretical gap value of 10.43 eV of a G stack,[9] of 10.34 eV of polyguanilic acid in the absence of water, and finally in this paper the calculated gap of 10.66 eV for polyguanilic acid with water (see Table II).

The results of Table II should be taken with caution because a scaling factor established on a single molecule is not necessarily valid for a complicated chain containing also this molecule.

In this way one would obtain a gap for a G stack 10.43.0.55 of 5.74 eV. Of course this result should be taken with caution because a factor valid for a single molecule is not necessary valid for a stack. Further it is very questionable whether the bandwidths have to be scaled down in a similar way. This problem of course requires further investigations, which could be started by a HF and HF+MP2 calculation for a guanine stack using again Gianolo and Clementi's double ξ basis set.[15]

IV. CONCLUSION

The obtained band structure for polyguanilic acid in the presence of water and Na^+ ions indicates that in this system there could be a hole conduction if CT of electrons from DNA to the histone molecules takes place. One should not forget, however, that real DNA is aperiodic. On the other hand there are a few facts which point in the direction that DNA B is not completely aperiodic. (1) It is well known that there are several places in DNA in which the sequence is periodic for a few tenths of base pairs, and in many cases there are three to seven subsequent purine or pyrimidine bases, respectively. (2) Though the HOMO levels of G and A (purine-type bases) and those of C and T (pyrimidine-type bases) are rather similar, if in a sequence[1] purine and pyrimidine-type bases occur randomly, there cannot be a coherent (Bloch-type) conduction along this sequence. G-C and A-T base pairs, however, do not differ much. Namely, the general structure of one pyrimidine-type base coupled by H bonds to a purine-type base makes them quite similar. In addition to this both base pairs have 24 π electrons (if one takes into account the hyperconjugation at the methyl group of thymine). All this is reflected also in the rather similar HF band structures of the stacked base pairs.[22]

It will be possible to settle this problem, when we have performed similar band structure calculations for the other three polynucleotides in the presence of water and ions. Later as still larger computational power becomes available one could perform band structure computations for polynucleotide pairs in the presence of water and ions. The results of these very demanding calculations will show whether these band structures are really similar.

One should point out, however, that real DNA is a double helix and is aperiodic. Therefore all these planned calculations can only serve as a starting point for a general theory of the dc conductivity in DNA. In such a theory a Bloch-type conduction is coupled with tunneling and first of all with hopping from one strand to the other. Only in this way we can understand the results of O'Neil and Barton who were doping aperiodic DNA with intercalating rhodium complexes which act as electron acceptors. They could detect positive charges (holes) at about 60 base pairs (~200 Å) away from the position of the doping agent.[23] We are going to return to this problem in later publications.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Professor E. Clementi for providing the parameters for the P atom of his basis set. Furthermore the authors are indebted to the High Performance Computing Group of the University of Szeged (Hungary) for providing free CPU time.

[1] K. Luger, W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, Nature (London) **389**, 251 (1997).
[2] T. J. Richmond and C. A. Davey, Nature (London) **423**, 145 (2003).
[3] S. C. R. Elgin and J. L. Workman, *Chromatin Structure and Gene Expression* (Oxford University Press, Oxford, 2000).
[4] A. Bende, F. Bogár, and J. Ladik, Chem. Phys. Lett. **437**, 117 (2007).
[5] C. A. Coulson (private communication).
[6] See, for instance, J. Ladik, Int. J. Quantum Chem. **78**, 450 (2000).
[7] J. Ladik and W. Förner, *The Beginning of the Cancer in The Cell* (Springer, Heidelberg, 1994), pp. 29–43.
[8] See, for instance, W. Shockley, *Electron and Holes in Semiconductors*

(Van Nostrand, New York, 1950), p. 27.
[9] F. Beleznay, F. Bogar and J. Ladik, J. Chem. Phys. **119**, 5690 (2003).
[10] F. Beleznay, F. Bogar, Zs. Szekeres, and J. Ladik, J. Chem. Phys. **124**, 074708 (2006).
[11] P.-O. Löwdin, Adv. Phys. **5**, 1 (1956).
[12] G. Del Re, J. Ladik, and G. Biczo, Phys. Rev. **155**, 997 (1967).
[13] J. M. Andre, L. Gouverneur, and G. Leroy, Int. J. Quantum Chem. **11**, 427 (1967); **11**, 451 (1967).
[14] A. Blumen and C. Merkel, Phys. Status Solidi B **83**, 425 (1977).
[15] L. Gianolo and E. Clementi, Gazz. Chim. Ital. **110**, 79 (1980).
[16] SYBYL 7.3 program (Tripos Inc., St. Louis, 2006).
[17] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN03, Revision C.02 (Gaussian, Inc., Wallingford, CT, 2004).
[18] G. Albiser, A. Lamiri, and S. Premilat, Int. J. Biol. Macromol. **28**, 199 (2001).
[19] Zs. Szekeres, F. Bogar, and J. Ladik, Int. J. Quantum Chem. **102**, 422 (2005).
[20] M. Piacenza and S. Grimme, ChemPhysChem **6**, 1554 (2005).
[21] S. D. Wetmore, R. J. Boyd, and L. A. Ericson, Chem. Phys. Lett. **322**, 129 (2000).
[22] See, for instance, P. Otto, E. Clementi, and J. Ladik, J. Chem. Phys. **78**, 4547 (1983).
[23] M. A. O'Neil and J. Barton, Top. Curr. Chem. **236**, 67 (2004), and references cited therein.