Biophysical Chemistry 15 (1982) 343–351
Elsevier Biomedical Press

INTRINSIC ELECTROSTATIC PROPERTIES AND BASE SEQUENCE EFFECTS IN THE STRUCTURE OF OLIGONUCLEOTIDES

Richard LAVERY, Bernard PULLMAN and Krystyna ZAKRZEWSKA

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique associé au C.N.R.S., 13, rue Pierre et Marie Curie, 75005 Paris, France

Received 7th October 1981
Accepted 2th December 1981

Key words: Oligonucleotide; Base sequence; Electrostatic potential; Steric accessibility

Molecular electrostatic potentials and steric accessibilities are calculated for Dickerson's dodecanucleotide CGCGAATTCGCG and compared with those for the 'inverted' sequence TATAGGCCTATA. The results are used to distinguish between properties due to base sequence (the location of the deepest potential minimum in the minor groove of A-T sequences and in the major groove of G-C sequences) and those due to the finite length of the oligonucleotide (location of the deepest potential in the central part of the oligonucleotide).

1. Introduction

The existence of sequence-dependent conformational variations in the structure of nucleic acids has recently attracted much interest, particularly in view of the role that such effects may play in the process of the recognition of nucleic acids by cellular proteins.

The existence of local conformations associated with specific base sequences has been clearly demonstrated by the elegant recent studies of Dickerson and co-workers [1–4] of the single crystal of the self-complementary dodecanucleotide d(CpGpCpGpApApTpTpCpGpCpG). Although this oligonucleotide has the global helical conformation of B-DNA many local modifications have been observed and are fully discussed in the articles cited.

As part of a series of studies in our laboratory on the properties of the nucleic acids involved in interactions with their usual partners, proteins, or with external agents such as carcinogens, mutagens, antibiotics, etc., we have investigated two features of Dickerson's dodecamer, namely, its molecular electrostatic potential and the steric accessibility of its reactive sites [5]. The study of these two properties, which, when considered together, have been demonstrated to be a very useful guide to the interactive behavior of the nucleic acids [6–8], enabled us to quantify the extent to which local conformational changes could influence their expression.

This study, however, has also brought to light a second powerful effect, namely, the influence of the finite length of the oligonucleotide and of the positioning of the bases with respect to its center (or extremities). It seems particularly important to clarify this influence before future results on oligonucleotides are extrapolated to longer nucleic acids.

We thus reconsider this problem in more detail in the present publication by comparing the potential and accessibility of the dodecanucleotide with Dickerson's base sequence with a related dodecamer in which the sequence has been altered by interchanging the purines, adenine and guanine, and the pyrimidines, thymine and cytosine, to yield TATAGGCCTATA (which we will term inverted Dickerson's dodecamer). As we are presently concerned essentially with the effect of the

finite length of these oligonucleotides versus the base sequence effect, and not with the relatively minor role of local structural changes, we have simplified the study of the problem by considering, in each case, a standard B-DNA conformation. For this reason the data concerning Dickerson's dodecamer are slightly different from those of ref. 5. A comparison between the results obtained for Dickerson's dodecamer in the two cases, the crystal and the standard one, show that all the main conclusions arrived at in the former study are completely preserved in the latter.

2. Method

The procedures for the calculation of the molecular electrostatic potential [6–9] and of the static steric accessibility [10] associated with oligo- and polynucleotides have been fully described in our previous publications and will not be repeated here.

We would just like to recall that by the term molecular electrostatic potential is meant the electrostatic potential created in the neighboring space by the nuclear charges and the electronic distribution of the system. For a given wave function with the corresponding electron density distribution $\rho[i]$, the value of such a potential $V[P]$ at a given point P in space is given by

$$V[P]=\sum_{\alpha}\frac{Z_{\alpha}}{r_{\alpha P}}-\int\frac{\rho[i]}{r_{Pi}}d\tau_i$$

where Z_{α} is the nuclear charge of nucleus α.

As to the static steric accessibility, this refers to the fraction of the area of the atomic van der Waals' sphere of the target atom in the macromolecule which can be reached by the probe representing the reactant, without the latter intersecting with the van der Waals' spheres associated with the remaining atoms of the macromolecule. The results described here relate to the accessibility toward a test sphere of radius 1.2 Å, representing a water molecule attacking through one of its hydrogen atoms [10] and representing thus, effectively, the upper limit of the atomic accessibilities, within the nucleic acids, towards molecular species.

The geometry employed for the B-DNA conformation was that due to Arnott and Hukins [11]. Each model dodecamer consists of 12 base-pairs and a total of 22 phosphates and 24 sugars in the phosphodiester backbones.

The potentials and accessibilities are calculated for the minima associated with the nucleophilic atoms of the bases [8,9]. Potentials are also calculated on the surface envelope surrounding each polynucleotide [12], this envelope being formed by the intersection of spheres centered on each atom of the macromolecule with radii proportional to their van der Waals' radii (a proportionality factor of 1.7 was employed, as in our previous studies [12]).

3. Results and discussion

The essence of our finding on the role of the finite length of the oligonucleotides on their molecular electrostatic poytential is the observation that this potential, which is dominated by the influence of the anionic phosphate groups of the backbones, is always more negative in the center of the oligonucleotide (i.e., midway along the double-helical segment) than at its extremities. This basic fact is illustrated for the case of a dodecanucleotide in fig. 1 where we reproduce the potential along the helical axis due only to the phosphodiester backbones, which is in fact mainly due to the anionic phosphate groups. The distances along the helical axis are measured from the

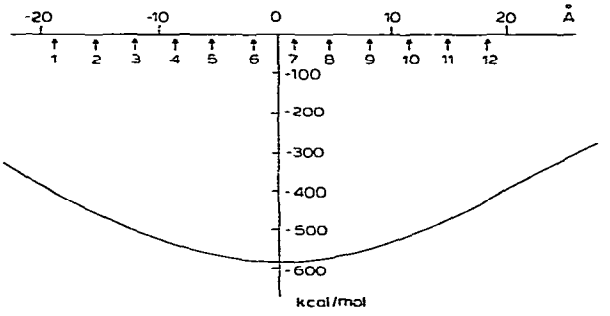


Fig. 1. Potential along the helical axis of a B-DNA due only to the phosphodiester backbones.

center of the dodecamer and the numbered arrows mark the positions occupied by the base-pairs in the B-DNA conformation. The result is very clear and shows a steady increase in the magnitude of the potential toward the center of the helix, this site being roughly 200 kcal/mol more negative than the ends of the dodecamer. Although we do not at present consider in detail the influence of counterions on the potential it is interesting to note that in the case of a rather extreme model screening, with an Na^+ bound to each phosphate group [13], the backbone potential at the center of the dodecamer still exceeds that at its ends by 50 kcal/mol.

When the potentials of the base pairs are added to those of the phosphodiester backbones to complete the dodecamers, this overall effect persists, as may be seen in fig. 2. In this figure, the range of the site potentials calculated for each base in one strand of the Dickerson (left-hand side) and inverted Dickerson (right-hand side) sequences is given. (The individual base site potentials are not presented here but are available on request. They refer as indicated above to a uniform B-DNA conformation and not to the true dodecamer conformations. However, the comparison with our previous study of Dickerson's dodecamer, in its crystal geometry, indicates that the influence of local conformational changes does not modify the corresponding site potentials by more than 15 kcal/mol, a small perturbation, indeed, in view of the overall values.)

The range of potential for each base again reflects the dominating influence of the phosphate groups so that the potentials associated with the central base-pairs of either sequence are roughly 200 kcal/mol more negative than those at the terminal base-pairs. Both sequences indeed give rise to rather similar graphics.

The effects of the individual bases are, on the contrary, much more distinct on the accessibilities which, summed for all the reactive atoms on each base, are illustrated in fig. 3. In this figure, there is little similarity between the two sequences, Dickerson and inverted Dickerson, apart from the noticeably larger accessibility of the terminal bases. The correlation is rather between bases of the same type, guanines having the highest accessibility, adenines and thymines being intermediate and cytosines being the least accessible. This ordering is, however, naturally, perturbed by the influence of the neighboring bases in the sequence.

The detailed distribution of the potential in the oligonucleotides is also influenced by the base sequence. In order to see this phenomenon clearly we turn to the representation of the surface potentials. These potentials were calculated for both oligonucleotides and we present for each of them two views, the first obtained when looking toward the minor groove in the center of the dodecamer

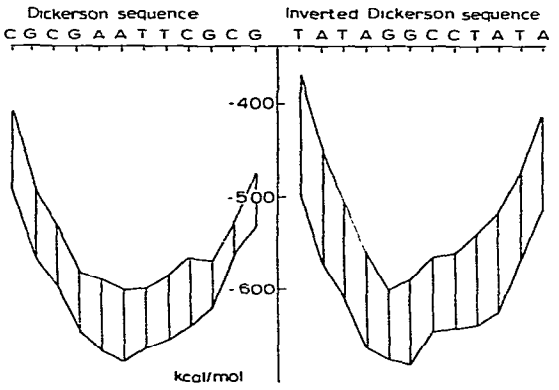


Fig. 2. Range of base site potentials in the dodecamers with Dickerson and inverted Dickerson sequences (kcal/mol).

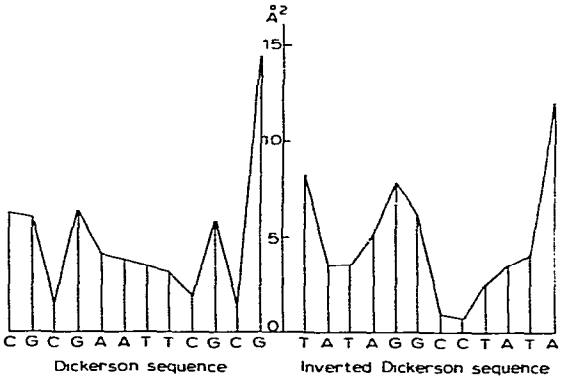


Fig. 3. Range of base accessibilities in the dodecamers with Dickerson and inverted Dickerson sequences (Å²).

and the second, diametrically opposed about the helical axis, obtained when looking toward the major groove. Figs. 4 and 5 refer to the Dickerson's sequence and figs. 6 and 7 to the inverted Dickerson's sequence. Each figure contains a molecular graphic of the relevant dodecamer in part a and the corresponding surface potentials in part b. The shadings used for the surface potentials are given in table 1.

These results confirm in the first place that the most negative surface potentials occur in the center of the oligonucleotides, as might have been de-

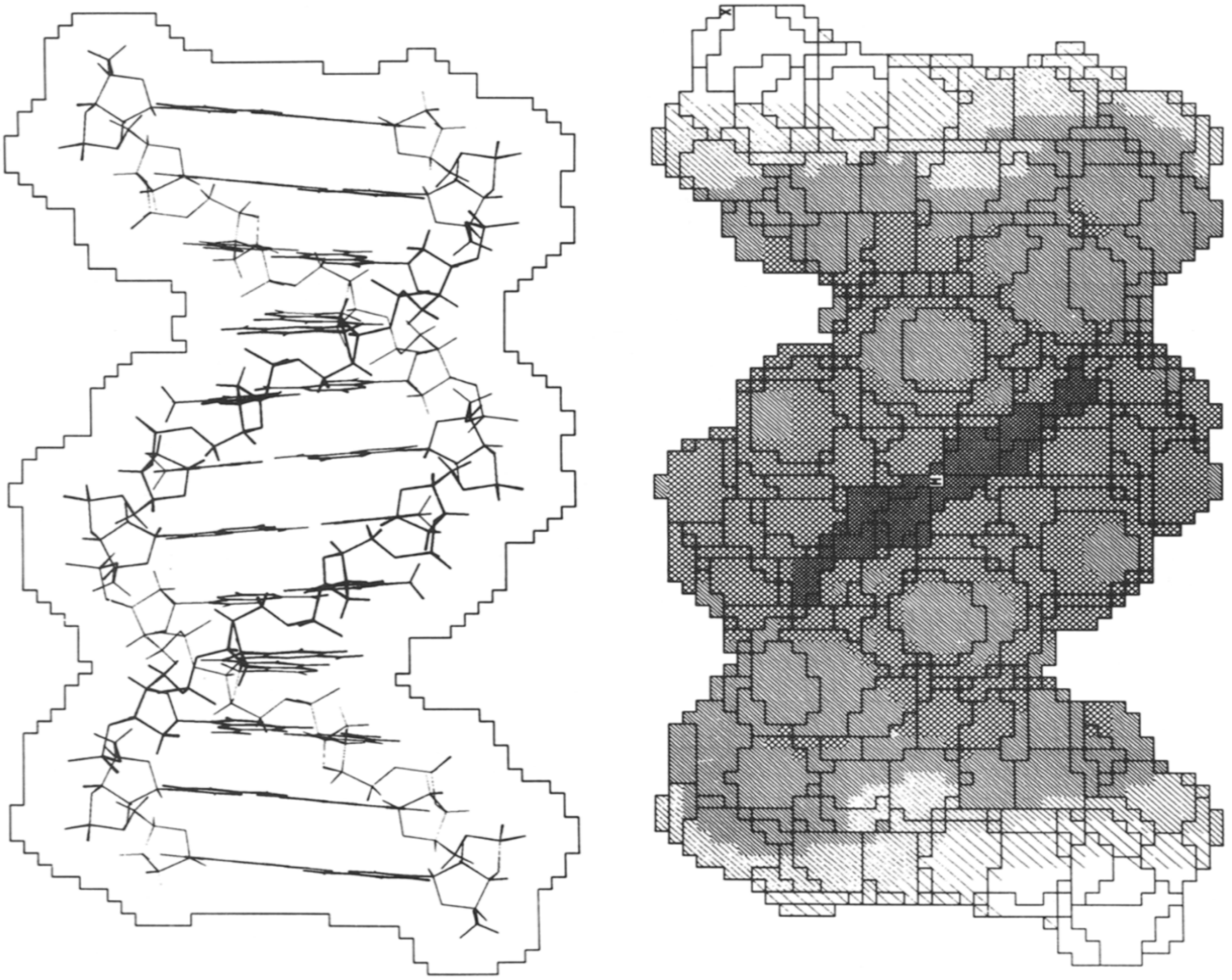


Fig. 4. View on the side of the central minor groove of the dodecamer with Dickerson's sequence: (a) molecular graphic; (b) surface potential.

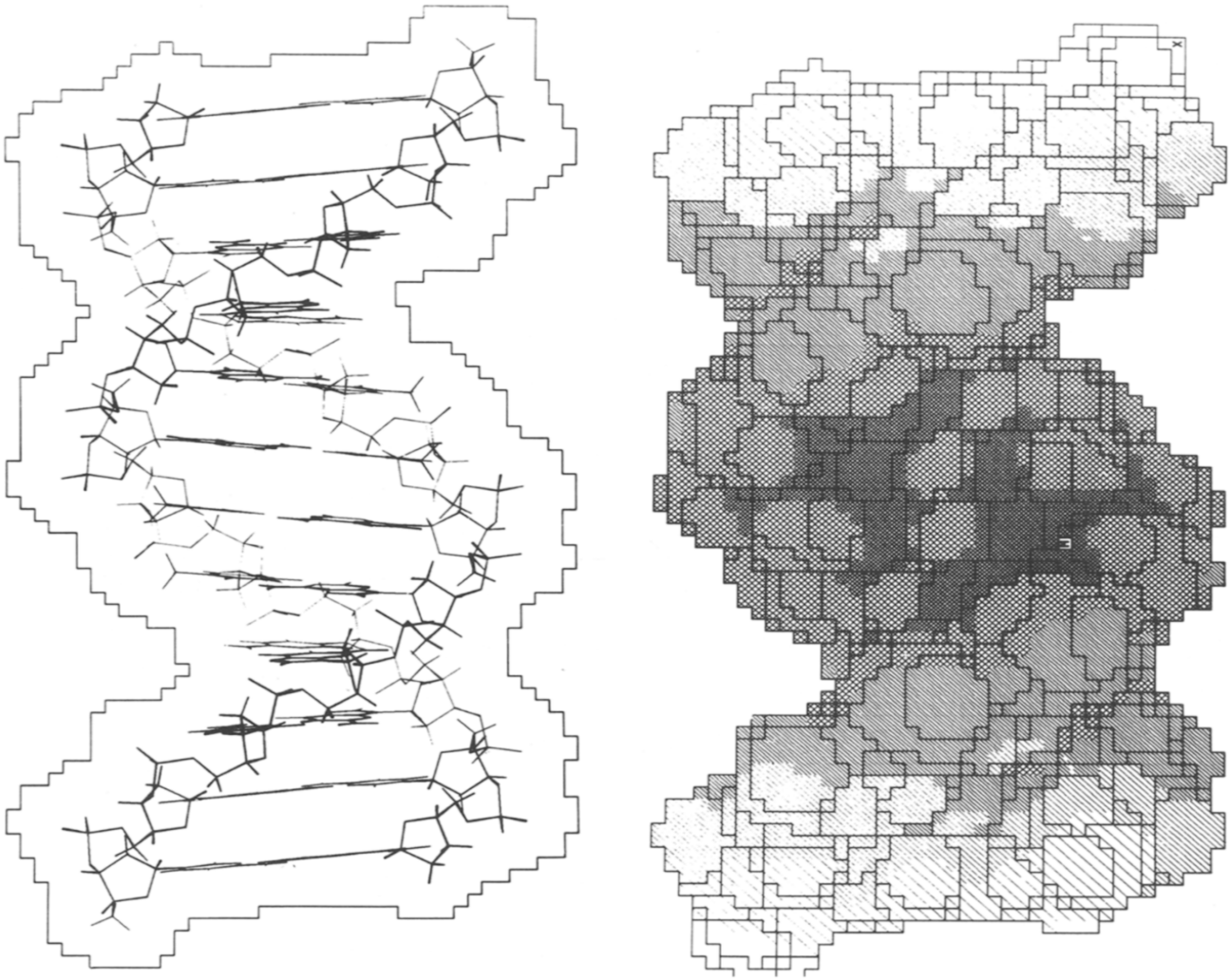


Fig. 5. View on the side of the central major groove of the dodecamer with Dickerson's sequence: (a) molecular graphic; (b) surface potential.

duced from the preceding discussion. They also indicate that the potentials in the grooves of the helices are more negative than those along their backbones. This latter finding, which is common to all double-helical B-DNAs, has been discussed in our previous publications [6–8,12] and is important in explaining the many interactions which occur at the bases of the nucleic acids.

Considering the graphics in detail we see for the Dickerson's sequence, in fig. 4, a narrow but con-

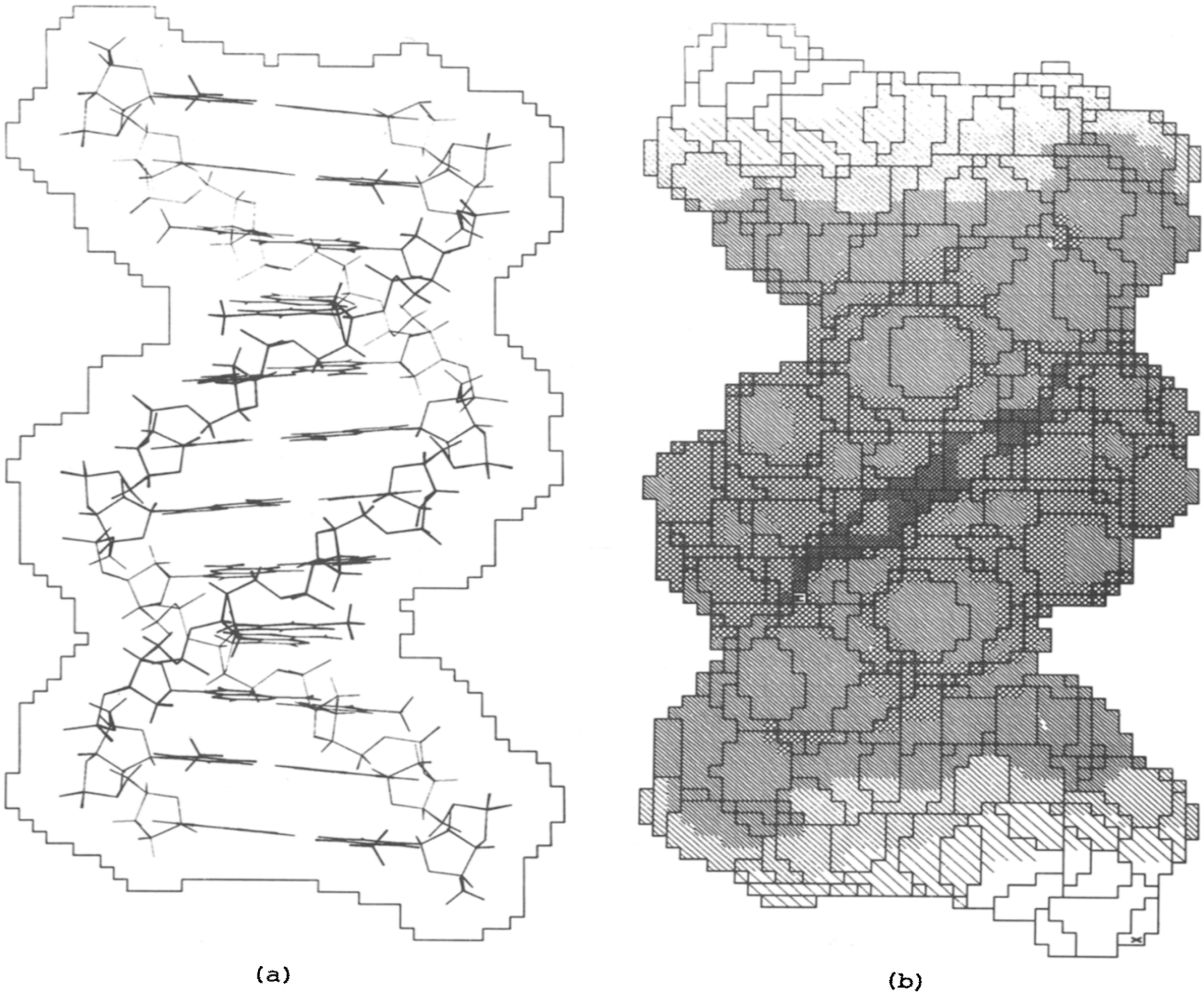


Fig. 6. View on the side of the central minor groove of the dodecamer with inverted Dickerson's sequence: (a) molecular graphic: (b) surface potential.

tinous band of strongly negative potential located in the central minor groove, which corresponds to the base sequence AATT. In the major groove of Dickerson's sequence, in fig. 5, the most negative potentials, although again located in the central part of the oligomer and thus associated also with the AATT sequence, form two zones, separated by a vertical band of weaker potential, which is due to the 'disruptive' effect of the adenine N6 amino groups (notation N6(A)) [14] located between the more electrophilic sites, N7(A) and O4(T). From the minimal potentials calculated in each surface

view, the minor groove is seen to be deeper than the major one by 30 kcal/mol (table 2).

Turning to the inverted Dickerson's sequence (figs. 6 and 7), we see that here again the deepest potentials are located in the central part of the oligomer. The situation between the two views is, however, inverted with respect to figs. 4 and 5. Thus, in fig. 6 which views the central minor groove, involving now the sequence GGCC, the most negative potentials are again in this groove. They do not form, however, a continuous zone but are interrupted by weaker potential spots due to the disruptive character of the guanine amino groups. Fig. 7, on the other hand, which views the

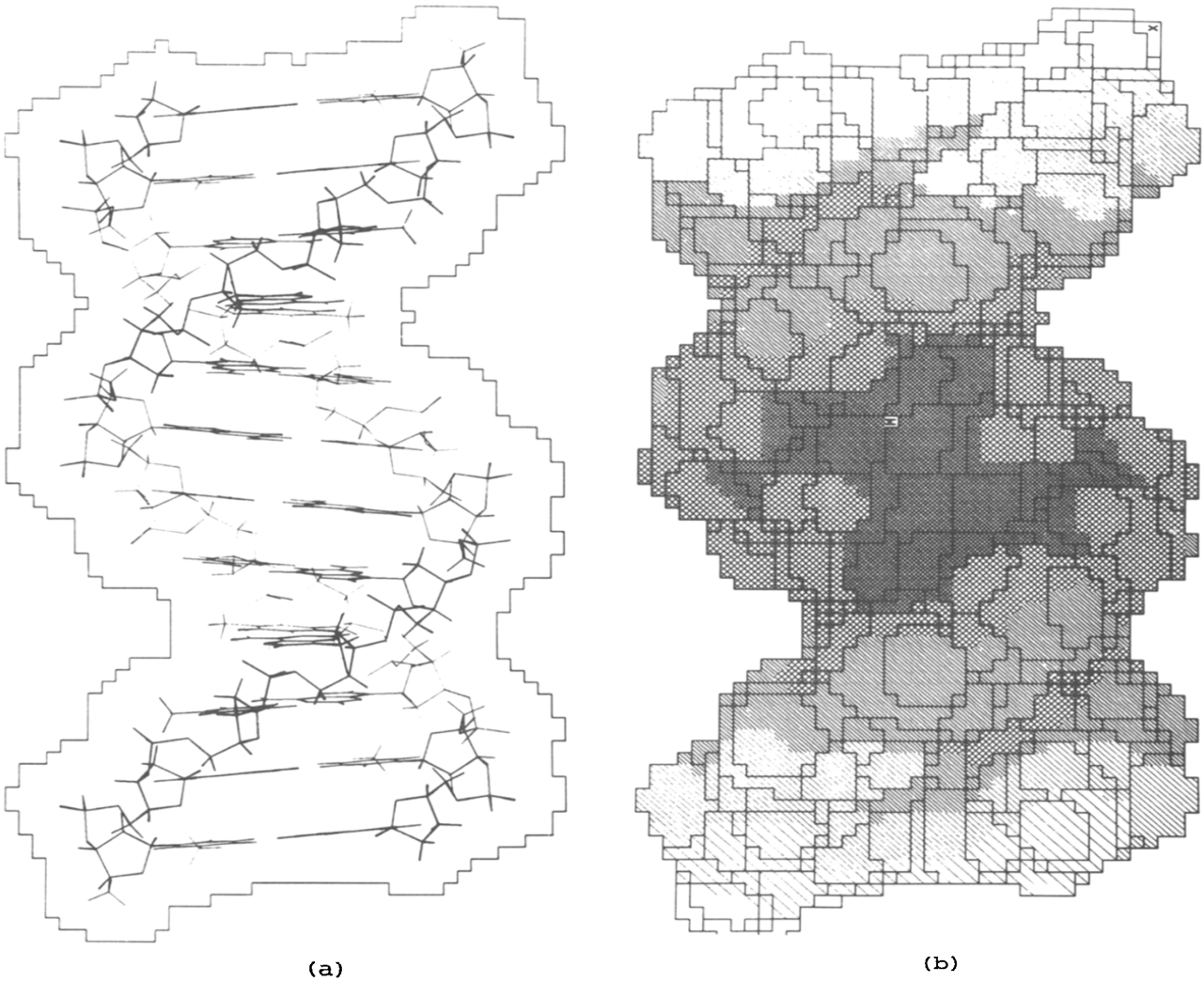
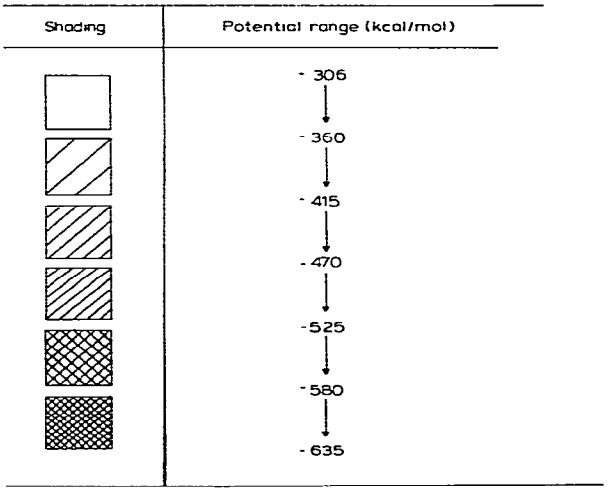


Fig. 7. View on the side of the central major groove of the dodecamer with inverted Dickerson's sequence: (a) molecular graphic; (b) surface potential.

Table 1

Shadings used in representing the surface envelope potentials



major groove of the same central GGCC part of the oligomer, shows a uniform zone of most negative potential located in that groove, extending essentially over the electrophilic sites N7(G) and O6(G). In table 2, the surface minima show that the potentials of this dodecamer are 30 kcal/mol more negative in the central major groove than in the corresponding minor groove. The distribution of the minima between the two grooves is thus almost exactly the reverse of the distribution found for the Dickerson's sequence.

It thus appears obvious that while the effect of the finite length of the oligomers consists of imposing in both cases the presence of the deepest potential zone in the center of the double helix, the nature of the base-pairs present in that central deepest zone is responsible for the preponderance of the potential in the minor groove, in the case of central A-T pairs, or in the major groove, in the case of central G-C pairs: the latter effect being due to the intrinsic tendency of these two pairs of complementary bases, when in the B conformation, to have the deepest minima in these respective grooves [6–8,12,15,16]. (A different situation would occur in A-DNA [17] or Z-DNA [18].)

Table 2

Potential minima on the surface envelopes

Sequence	Minor groove minimum	Major groove minimum
Dickerson's	−635	−604
Inverted Dickerson's	−605	−634

4. Conclusion

We have thus demonstrated here, through the comparison of the two dodecameric sequences, CGCGAATTCGCG and TATAGGCCTATA, both: (1) the influence of the base sequence on the molecular potential and accessibility of these oligonucleotides and (2) the important general effect in these double helices of their finite length which leads to the concentration of negative potential in the center of the oligonucleotides.

The implications of these findings for the interactive properties of the oligonucleotides may be considered for the example of their hydration. Dickerson and co-workers [4] have located 72 bound water molecules in the crystal around the dodecamer CGCGAATTCGCG. The majority of these molecules were found to form a regular spine in the minor groove of the central AATT sequence. Less water molecules were observed in the major groove and toward the ends of the dodecamer. This was interpreted by Dickerson in terms of the availability of binding sites on the base-pairs, namely, the appropriate location of the N3(A) and O2(T) sites of the AT pairs in the minor groove and of the blocking effect of the amino group of guanine in the minor groove of the GC pairs.

We have pointed out in our previous publication [4] that the molecular potentials may cast further light on this interpretation, the greatest concentration of bound water in the minor groove of the AT pairs, in the center of the oligonucleotide, correlating with the location of the most negative potentials in the same zone.

The question may now be posed as to how this

hydration scheme would change for the as yet experimentally unstudied inverted Dickerson's sequence, TATAGGCCTATA. Our results would suggest that despite the possibility of regular spines of water in the minor grooves of the AT base-pair segments at the extremities of the oligonucleotide, the overall potential distribution could nevertheless favor the center of the oligonucleotide, on the side of the major groove of the GC base-pairs, where two accessible and attractive binding sites N7(G) and O6(G) exist.

Acknowledgment

This work was supporeted by the National Foundation for Cancer Research to which the authors express their thanks.

References

1 R. Wing, H.R. Drew, T. Takano, C. Broka, S. Tanaka, H. Itakura and R.E. Dickerson, Nature 287 (1980) 755.
2 H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, Proc. Natl. Acad. Sci. U.S.A. 78 (1981) 2179.
3 R.E. Dickerson and H.R. Drew, J. Mol. Biol. 149 (1981) 761.
4 H.R. Drew and R.E. Dickerson, J. Mol. Biol. 151 (1981) 535.
5 R. Lavery and B. Pullman, Nucleic Acids Res. 9 (1981) 3765.
6 A. Pullman and B. Pullman: in Chemical applications of atomic and molecular electrostatic potentials, eds. P. Politzer and D.G. Truhlar, (Plenum Press, New York, 1981) p. 381.
7 A. Pullman: in New horizons in biological chemistry, eds. M. Koike, T. Nagatsu, J. Okuda and T. Ozawa (Japan Scientific Societies Press, Tokyo, 1980) p. 19.
8 A. Pullman and B. Pullman, Rev. Biophys. 14 (1981) 289.
9 B. Pullman, D. Perahia and D. Cauchy, Nucleic Acids Res. 6 (1979) 3821.
10 R. Lavery, A. Pullman and B. Pullman, Int. J. Quantum Chem. 20 (1981) 49.
11 S. Arnott and D.W.L. Hukins, Biochem. Biophys. Res. Commun. 47 (1972) 1504.
12 R. Lavery and B. Pullman, Int. J. Quantum Chem. 20 (1981) 259.
13 R. Lavery, D. Cauchy, O. de la Luz Rojas and A. Pullman, Int. J. Quantum Chem., Quantum Biol. Symp. 7 (1980) 323.
14 B. Pullman, A. Pullman and R. Lavery: in Contemporary aspects of biophysics (Naukova Dumka, Moscow, 1982) in the press.
15 D. Perahia, A. Pullman and B. Pullman, Int. J. Quantum Chem., Quantum Biol. Symp. 6 (1981) 353.
16 B. Pullman, Ann. New York Acad. Sci. 367 (1981) 182.
17 R. Lavery and B. Pullman, Nucleic Acids Res. 9 (1981) 4677.
18 K. Zakrzewska, R. Lavery, A. Pullman and B. Pullman, Nucleic Acids Res. 8 (1980) 3917.