

## Structural basis of DNA flexibility

Dhananjay Bhattacharyya and Rabi Majumdar\*

Biophysics Division, Saha Institute of Nuclear Physics, 37 Belgachia Road, Calcutta 700037, India

Accepted 3 October 2000

It has been recently shown by us, on the basis of crystal structure database that the flexibility of B-DNA double helices depends significantly on their base sequence. Our model building studies further indicated that the existence of bifurcated cross-strand hydrogen bonds between successive base pairs is possibly the main factor behind the sequence directed DNA flexibility. These cross-strand hydrogen bonds are, of course, weaker than the usual Watson-Crick hydrogen bonds and their bond geometry is characterized by relatively larger bond lengths and smaller bond angles. We have tried to improve our model structures by incorporating non-planarity of the amino groups in DNA bases due to the presence of lone pair electrons at the nitrogen atoms. Energy minimization studies have been carried out by using different quantum chemical methods, whereby it is found that in all cases of N—H...O type cross-strand hydrogen bonds, the bond geometry improves significantly. In the cases of N—H...N type hydrogen bonds, however, no such consistent improvements can be noticed. Perhaps the true picture would emerge only if all the other interactions present in the DNA macromolecule could be appropriately taken into account.

### Introduction

Base sequences are known to perturb the structure of DNA double helices and these perturbations are important for selective recognition of the DNA base sequences by regulatory proteins. The DNA macromolecule is a worm-like structure which can be straight or intrinsically curved depending on its base composition. It can also be curved due to external factors such as protein binding or supercoiling but such deformations are usually indifferent to base sequence. Here we consider only the intrinsic curvature or the sequence directed flexibility of DNA and the structural basis of such flexibility. We have recently shown from an analysis of the X-ray crystallographic structural database that the flexibility of DNA in its regular B form depends significantly on the base sequence of the polymer<sup>1</sup>. This conclusion was based on calculations of the macromolecular parameters such as, the persistence length (*P*) and the torsional rigidity (*C*) for some specific repeat sequences. It was shown, for example, that the sequences like poly(dA).poly(dT) and poly(dG).poly(dC) have rigid structures as their

persistence lengths are high in comparison with those for the usual mixed sequence polymers. On the other hand, the sequences like poly(dG-dC).poly(dG-dC) and poly(dA-dC).poly(dG-dT) are flexible as they have relatively much smaller persistence lengths and often smaller torsional rigidity values. In addition, it was also shown<sup>1</sup> that some triplet repeat sequences like poly(dG-dC-dC).poly(dG-dG-dC) and poly(dC-dT-dG).poly(dC-dA-dG) implicated in human genetic disorder are also quite flexible. This trend is seen to be consistent with some experimental results on DNA flexibility derived from nucleosomal packaging data<sup>2-6</sup>. Furthermore, in order to understand such variations in flexibility of DNA, depending on its base composition, we conducted model building studies<sup>1</sup> of the DNA double helices from the average base pair local doublet parameters as obtained from the crystallographic data. It was thus shown that the rigid sequences can form additional "bifurcated cross-strand hydrogen bonds" between the successive base pairs which are not possible in the case of flexible sequences. The strength of a hydrogen bond depends on its bond length and the bond angle. The bifurcated cross-strand hydrogen bonds were found to have larger bond lengths and smaller bond angles, thus making them weaker than the usual Watson-Crick type of hydrogen bonds. Also, the results of a quantum chemical calculation indicated that the energy minimum for these cross-strand hydrogen bonds occurred near the predicted bond lengths and the N—H...O type of hydrogen bonds were found to

\*Author for correspondence at this address:

CSIR Emeritus Scientist, Dept. of Biophysics, Molecular Biology & Genetics, University of Calcutta, 92, A.P.C. Road, Calcutta 700 009

E-mail: rabim@cubmb.ernet.in

Tel: (033) 351-0359

Fax: 91-33-351-9755 (Attn: Prof. Rabi Majumdar, Biophysics Dept.)

be more stable in comparison with N—H...N type of hydrogen bonds<sup>1</sup>.

Recently it has been pointed out that the amino groups of DNA bases can have non-planar character due to lone pair electrons at the nitrogen atom<sup>7-9</sup>. In the present work we have shown by quantum chemical calculations that such non-planarity of the amino groups of DNA bases improves the cross-strand hydrogen bond geometry significantly, at least for the N—H...O types of hydrogen bonds. There is, however, no significant improvement for the N—H...N types of hydrogen bonds. We have carried out quantum chemical calculations starting from the semi empirical methods to the *ab initio* methods with split basis set to show that similar trends are observed in each case.

### Methodology

DNA double helices have been modelled on the basis of NUCGEN<sup>10</sup> by using the average base pair local doublet parameters, namely tilt, roll, twist, shift, slide, rise as well as propeller twist. These average values of the local doublet parameters were obtained by analyzing all the available B-DNA crystal structures of oligonucleotides<sup>1,11</sup>. These models have all amino groups in planar conformation. The presence of cross-strand hydrogen bonds were detected using a fairly long cut-off distance of 3.2 Å between the hydrogen and the acceptor atoms as used by several other groups<sup>12,13</sup>.

In the present work, we have tested whether the non-planarity of the amino groups may change the hydrogen bond geometry substantially. We have optimized the geometry of these structures by keeping the heavy atoms at their original positions. In other

words, only the bond length, bond angles and dihedral angles of all the hydrogen atoms are energy minimized. We have used different quantum chemical procedures for the energy minimization, namely the semi empirical AM1<sup>14</sup> and PM3<sup>15</sup> methods as adopted in MOPAC<sup>16</sup>, and the *ab initio* STO-3G and 3-21G methods as adopted in GAMESS<sup>17</sup>.

### Results and Discussion

In general, the cross-strand hydrogen bonds as reported in Table 1 are quite long as compared to the usual strong hydrogen bond lengths ( $2.0 \pm 0.3$  Å). The angles subtended at the hydrogen atom, i.e. the N—H...A angles or the bond angles, where A is the acceptor atom, are also quite small as compared to those for the strong hydrogen bonds ( $180^\circ \pm 30^\circ$ ). Since the hydrogen bonds in these cases occur between two bases across the base pairs, these can never be linear and strong as in the case of Watson-Crick base pairs or in protein secondary structures. Such long and non-linear hydrogen bonds have also been reported in literature for the DNA-ligand complexes<sup>13</sup>.

In the present work, we find that the cross-strand hydrogen bond geometry for the N—H...O type bonds improves substantially when the non-planar character of the base amino groups are incorporated into the structure, as indicated in Table 1. As a typical example, the cross-strand hydrogen bonded adenine and thymine bases in poly(dA).poly(dT) are shown in Fig. 1 before and after optimization using MOPAC with AM1 hamiltonian. It is observed, in this case that hydrogen bond length is reduced from 2.99 Å to 2.82 Å and the bond angle increases from  $101.3^\circ$  to  $112.9^\circ$  as a result of non-planarity of the amino group. Similar

Table 1—The geometry of cross-strand hydrogen bonds in the model structures (in terms of bond lengths and bond angles) for the six polymeric DNA sequences with planar and non-planar amino groups. [The non-planar geometry of the amino groups are obtained by energy minimization using MOPAC with AM1 hamiltonian]

| Polymer sequence                | N—H...A Distance (Å)              |                                       | N—H...A Angle (°)                 |                                       | Atoms involved |
|---------------------------------|-----------------------------------|---------------------------------------|-----------------------------------|---------------------------------------|----------------|
|                                 | with planar NH <sub>2</sub> group | with non-planar NH <sub>2</sub> group | with planar NH <sub>2</sub> group | with non-planar NH <sub>2</sub> group |                |
| poly(dG).poly(dC) (Rigid)       | 3.05                              | 2.92                                  | 92.2                              | 99.9                                  | H42 – O6       |
| poly(dA).poly(dT) (Rigid)       | 2.99                              | 2.82                                  | 101.3                             | 112.9                                 | H62 – O4       |
| poly(dA-dG).poly(dC-dT) (Rigid) | 3.02                              | 3.04                                  | 99.6                              | 98.1                                  | H42 – N6       |
| poly(dA-dT).poly(dA-dT)         | 2.89                              | 2.89                                  | 94.0                              | 94.1                                  | H62 – N6       |
| (Rigid)                         | 2.87                              | 2.90                                  | 94.7                              | 93.6                                  | H62 – N6       |
| poly(dA-dC).poly(dG-dT)         | 3.09                              | 2.91                                  | 89.1                              | 99.6                                  | H62...O6       |
| (Flexible)                      |                                   |                                       |                                   |                                       |                |
| poly(dG-dC).poly(dG-dC)         | ----                              |                                       |                                   |                                       | ----           |
| (Flexible)                      |                                   |                                       |                                   |                                       |                |

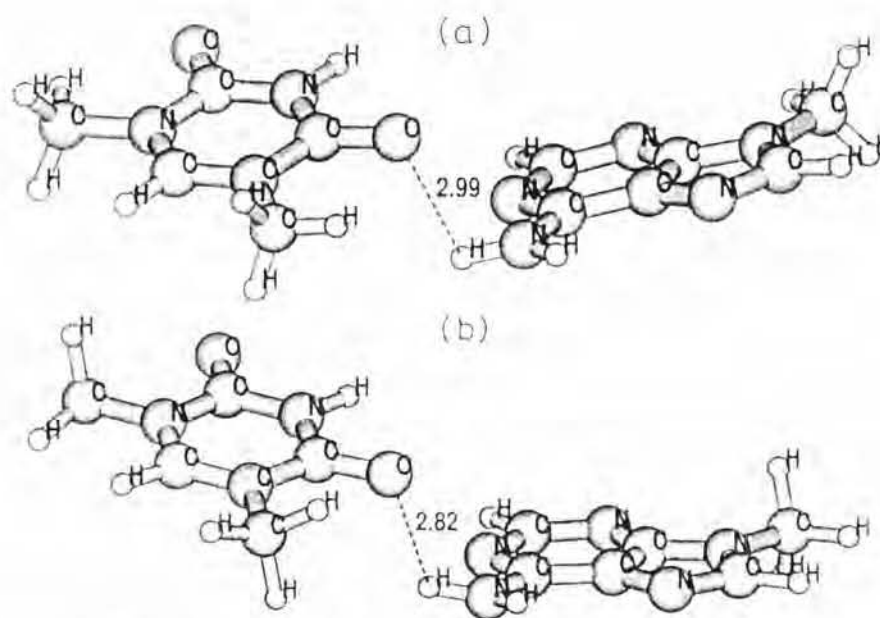


Fig. 1—Ball and stick representation of adenine and thymine bases in their cross-strand hydrogen bonded conformation in poly(dA)-poly(dT) structure with (a), planar amino group and (b), non-planar amino group imposed by energy minimization using MOPAC with AM1 hamiltonian. [The hydrogen bond lengths (in Å) are shown beside the bonds].

Table 2—The geometry of the various cross-strand hydrogen bonds (in terms of bond lengths and bond angles) in the structures optimized using different quantum chemical methods. [The amino group dihedral angles as measures of non-planarity of the amino groups are also listed]

| Polymer sequence        | Q.M.<br>Method<br>adopted | H...A<br>Distance<br>(Å) | N—H...A<br>Angle<br>(°) | Amino group<br>dihedral angle |
|-------------------------|---------------------------|--------------------------|-------------------------|-------------------------------|
| Poly(dG),poly(dC)       | AM1                       | 2.92                     | 99.9                    | -8.85                         |
|                         | PM3                       | 2.85                     | 104.4                   | -13.8                         |
|                         | STO-3G                    | 2.84                     | 99.2                    | -13.9                         |
|                         | 3-21G                     | 2.93                     | 104.0                   | -8.3                          |
| Poly(dA),poly(dT)       | AM1                       | 2.82                     | 112.9                   | -13.2                         |
|                         | PM3                       | 2.77                     | 116.0                   | -17.6                         |
|                         | STO-3G                    | 2.74                     | 116.8                   | -20.4                         |
|                         | 3-21G                     | 2.85                     | 110.5                   | -11.4                         |
| poly(dA-dG),poly(dC-dT) | AM1                       | 3.04                     | 98.1                    | 2.2                           |
|                         | PM3                       | 3.20                     | 88.8                    | 12.5                          |
|                         | STO-3G                    | 3.25                     | 85.8                    | 14.8                          |
|                         | 3-21G                     | 2.99                     | 101.5                   | -2.0                          |
| poly(dA-dT),poly(dA-dT) | AM1                       | 2.89                     | 94.1                    | 0.6                           |
|                         |                           | 2.90                     | 93.6                    | 1.9                           |
|                         |                           | 3.11                     | 81.2                    | 16.0                          |
|                         |                           | 3.12                     | 80.8                    | 15.9                          |
|                         | STO-3G                    | 2.56                     | 114.4                   | -22.7                         |
|                         |                           | 3.20                     | 76.3                    | 20.3                          |
|                         | 3-21G                     | 2.85                     | 96.6                    | -3.0                          |
|                         |                           | 2.83                     | 97.5                    | -11.6                         |
| poly(dA-dC),poly(dG-dT) | AM1                       | 2.91                     | 99.6                    | -16.0                         |
|                         | PM3                       | 2.85                     | 103.7                   | -18.6                         |
|                         | STO-3G                    | 2.80                     | 105.7                   | -8.9                          |
|                         | 3-21G                     | 2.95                     | 97.0                    | -8.9                          |

trends are also seen in the structures optimized by using other quantum chemical procedures as indicated in Table 2, where the amino group dihedral angles (as measures of non-planarity) are also tabulated. For the weaker N—H...N bonds, however, no such consistent improvements in geometry are noticed due to non-planarity of the amino groups. Thus it appears from the present model that the N—H...O type bifurcated cross-strand hydrogen bonds are the ones which impart rigidity to the DNA polymer more effectively in comparison with the N—H...N type bonds. However, it is felt that a true picture would emerge only if all the other interactions present in DNA, specially the hydrophobic environment created by the neighbouring base pairs, could be appropriately incorporated into our model. These points are being currently investigated.

## References

- 1 Bhattacharyya D, Kundu S, Thakur AR & Majumdar R (1999) *J. Biomol Struct Dynam* 17, 289-300
- 2 Rhodes D (1979) *Nucl Acids Res* 6, 1805-1816
- 3 Simpson RT & Kuntzler P (1979) *Nucleic Acids Res* 6, 1387-1415
- 4 Kunkel, GR & Martinson HG (1981) *Nucleic Acids Res* 9, 6869-6888
- 5 Godde JS & Wolffe AP (1996) *J Biol Chem* 271, 15222-15229
- 6 Bacolla A, Geillibolian R, Shimizu M, Amirhaeri S, Keng S, Oshima K, Larson JE, Harvey SC, Stollar BD & Wells RD (1997) *J Biol Chem* 272, 16783-16792
- 7 Sponer J & Hobza P (1994) *J Am Chem Soc* 116, 709-714
- 8 Sponer J, Florian J, Hobza P & Leszczynski J (1996) *J Biomol Struct Dynam* 13, 827-833
- 9 Luisi B, Orozco M, Sponer J, Luque FJ & Shakked Z (1998) *J Mol Biol* 279, 1123-1136
- 10 Bansal M, Bhattacharyya D & Ravi B (1995) *CABIOS*, 11, 281-287
- 11 Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR & Seneider B (1992) *Biophys J* 63, 751-759
- 12 Mohan S & Yathindra N (1991) *J Biomol Struct Dynam* 9, 113-125
- 13 Tabernero L, Bella J & Aleman C (1996) *Nucleic Acids Res* 24, 3458-3466
- 14 Dewar MJS, Zoebisch EG, Healy EF & Stewart JJP (1985) *J Am Chem Soc* 107, 3902-3909
- 15 Stewart JJP (1989) *J Comp Chem* 10, 221-264
- 16 Stewart JJP (1990) *J Comp Aided Mol Design* 4, 1-105
- 17 Schmidt MW, Baldrige KK, Boatz JA, Elber ST, Gordon MS, Jensen AH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M & Montgomery Jr JA. (1993) *J Comp Chem* 14, 1347-1363