ELSEVIER

Biophysical Chemistry 52 (1994) 35–43

Biophysical Chemistry

Canonical analysis of correlated atomic motions in DNA from molecular dynamics simulation

F. Briki, D. Genest

Centre de Biophysique Moléculaire, CNRS, 1A Avenue de la Recherche Scientifique, 45071 Orleans Cedex 02, France

Received 5 January 1994; accepted in revised form 28 March 1994

Abstract

We report a method for analyzing atomic correlated motions in biopolymers from trajectories obtained by molecular dynamics simulation. A correlation coefficient based on the canonical analysis of data is defined which is independent on the relative orientation of atomic displacement. To illustrate the method we studied correlation between positional fluctuations of protons in the double-stranded self complementary oligonucleotide d(CTGATCAG), deduced from a 200 ps molecular dynamics simulation in the presence of explicit water molecules and counterions. It is found that on this time scale the motions of protons belonging to different residues are poorly coupled while the motion of a base proton is correlated to the motion of the sugar ring protons of the same nucleotide. Such a method may be generalized to study correlated motions of two distinct domains of a macromolecule.

Key words: Canonical analysis; Computer Simulation; Correlated motions; DNA flexibility; Molecular dynamics

1. Introduction

Molecular dynamics (MD) simulation is widely used to determine the structure of biological macromolecules such as proteins or nucleic acids, from experimental data given by NMR or X-ray crystallography. It is also a powerful tool for investigating the internal flexibility of macromolecules whose average structure is known [1–3]. Much information may be extracted from simulations: conformation, root-mean-square fluctuations or correlation times. A property which has been less studied is the correlation between fluctuations of different structural parameters of a macromolecular system.

A first interest for the analysis of correlated fluctuations is to evidence how the variation of a structural parameter may affect other structural parameters. The study of cross-correlations is also an interesting tool for analyzing collective motions of atoms, which are usually low-frequency motions of high biological significance [1,2]. Finally the knowledge of atomic correlated motions may be of some help for interpreting NMR experiments, for example NMR spectroscopy which estimates interatomic distances [4].

Mc Cammon [5] used such an approach to study collective motions in cytochrome c. An interesting two-dimensional representation of the displacement cross-correlation between the

SSDI 0301-4622(94)00063-P

residues was given. The residue displacements were evaluated as the displacements of the residue centroids.

Ichiye and Karplus [6] and Swaminathan et al. [7] used a similar approach for evidencing correlated motions in BPTI and HIV-1 protease dimer respectively. These two studies focused on backbone atoms and they led to a fine description of the internal dynamics of these proteins. The method used in these works for identifying collective motions, makes use of the normalized covariance, C_{ij}, of the positional fluctuations, $\delta \boldsymbol{r}_i$ and $\delta \boldsymbol{r}_j$, of atoms i and j respectively. C_{ij} is proportional to the mean scalar product between $\delta \boldsymbol{r}_i$ and $\delta \boldsymbol{r}_j$. For totally correlated (or anticorrelated) motions, $C_{ij} = 1$ (or -1), while for uncorrelated motions $C_{ij} = 0$. However, as explained by Ichiye and Karplus [6] this is true only if $\delta \boldsymbol{r}_i$ and $\delta \boldsymbol{r}_j$ are collinear vectors, since C_{ij} depends on the angle between both vectors. This is due to the fact that this approach does not take into account possible correlations between the different components of the fluctuations of a given atom on one hand and between non-homonymic components of the fluctuations of different atoms on the other hand. In other words the value of C_{ij} expresses at the same time the correlation and the relative orientation of the fluctuations, making the interpretation of C_{ij} ambiguous. Consequently the analysis of translational collective motions is fully justified but analysis of rotational motions is not easy.

The aim of this paper is to propose a method which avoids the limitation mentioned above. It is based on the canonical analysis approach used in statistics for comparing different groups of variables. Here the variables are the components of normalized fluctuations. The method consists in defining new variables (canonical variables) which satisfy the following conditions:

- for a given atom there is no correlation between the canonical variables.
- a canonical variable of an atom is correlated, at most, to only one canonical variable of the other atom.

With this method a correlation coefficient between positional fluctuations may be calculated easily which can be interpreted without ambiguity.

As an illustration of this approach, we present a study of the correlations between fluctuations of some protons belonging to the double-stranded oligonucleotide d(CTGATCAG). The choice of the protons is guided by their importance in the elucidation of conformation from two-dimensional NMR NOESY experiments [8,9]. The atomic fluctuations are obtained from a 200 ps MD simulation in the presence of explicit solvent and counterions. The different correlation coefficients are compared to the root-mean-square fluctuations of the corresponding interproton distances which gives qualitative information about the phase of the correlation.

2. Materials and methods

2.1. Molecular dynamics simulations

The atomic trajectories used in the present study have been obtained from a 200 ps molecular dynamics simulation in the presence of 1534 explicit water molecules and 14 Na^+ counterions. The details of the simulation have been described previously [10]. In short, after a heating and equilibration period of 34 ps the 200 ps production period was performed using the SHAKE algorithm [11] to keep the bond lengths constant. The time increment during the simulation was 0.002 ps and a weak coupling to a thermal bath was used to keep the temperature close to 300°K. 10000 conformations were stored, separated by time intervals of 0.02 ps. The simulations were performed on an IBM 3090 computer located at CIRCE (Orsay-FRANCE) with the program GROMOS [12]. The analysis was done on an IRIS 4D35 Silicon Graphics work station.

GROMOS uses the concept of united atom model, and apolar proton positions are not immediately available. It was therefore necessary to recalculate their trajectories by using the same method as in the subroutine DISRE of GROMOS. This consists on a simple geometrical calculation which imposes the correct bond lengths and bond angles of CH, CH2 and CH3 groups for which the coordinates of the carbon atoms and of their direct neighbours are known from the simulation.

For each atom of the molecule the mean Cartesian coordinates $\langle x \rangle$, $\langle y \rangle$ and $\langle z \rangle$ are calculated together with the corresponding components of the instantaneous normalized fluctuations $\delta\alpha = \alpha - \langle \alpha \rangle)/[\langle(\alpha - \langle \alpha \rangle)^2\rangle]^{1/2}$ ($\alpha = x$, y or z).

For two protons 1 and 2 with average distance $\langle r_{12} \rangle$ the root-mean-squares (RMS) of their distance fluctuations is given by $\mathrm{RMS} = [\langle(r_{12} - \langle r_{12} \rangle)^2\rangle]^{1/2}$, where the averages are taken on the 10000 atomic trajectories. We also define an average RMS for each type of atom pair by $\langle \mathrm{RMS} \rangle = 1/n \sum \mathrm{RMS}_k$, where the index k refers to the different pairs of this type, and $n = 7$ or 8 for the intra- or inter-residue cases respectively.

2.2. Spatial correlations

We outline the canonical analysis approach which has been used in the present work to establish correlations between two groups of individuals described by three variables. In the present paper, the groups are related to two atoms referenced as 1 and 2 respectively. The individuals correspond to the N conformations, $N = 10000$, resulting from the MD simulation and the variables are the cartesian components of positional normalized fluctuations δx, δy and δz. As one is interested in equal time correlation, each variable may be considered as a vector defined in a N-dimensional space. In this space we define a scalar product between two vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ by:

$$\langle \boldsymbol{U} \cdot \boldsymbol{V} \rangle = (1/N) \sum U_k V_k,$$

which represents the normalized correlation between $\boldsymbol{U}$ and $\boldsymbol{V}$, U_k and V_k ($k = 1, N$) being the components of both vectors respectively.

In the general case the vectors δx, δy and δz of each group are not orthogonal in the N-dimensional space, nor are vectors belonging to different groups. This is the cause of the difficulties mentioned in section 1. In order to circumvent this problem we use in each group new variables called canonical variables, which are linear combinations of the original variables δx, δy and δz. The canonical variables are defined as follows:

Let W1 and W2 be the sets of the N-dimensional vectors obtained by linear combination of the original vectors of groups 1 and 2 respectively, and $A1$ and $A2$ the projection operators onto W1 and W2 respectively. $A1A2$ and $A2A1$ are symmetric with dimensions 3×3 and have the same set of eigenvalues λ_i ($i = 1,2$ and 3). The canonical variables $\boldsymbol{X}1_i$ and $\boldsymbol{X}2_i$ of groups 1 and 2 are the eigenvectors of $A1A2$ and $A2A1$ respectively, associated to λ_i. It can be demonstrated [13] that they verify the following conditions:

- $\langle \boldsymbol{X}1_i \cdot \boldsymbol{X}1_j \rangle = \delta_{ij}$,
- $\langle \boldsymbol{X}2_i \cdot \boldsymbol{X}2_j \rangle = \delta_{ij}$,
- $\langle \boldsymbol{X}1_i \cdot \boldsymbol{X}2_j \rangle = 0$ if i and j are different,
- $\langle \boldsymbol{X}1_i \cdot \boldsymbol{X}2_i \rangle = \lambda_i^{1/2}$,
- λ_i represents the square of the correlation coefficient between $\boldsymbol{X}1_i$ and $\boldsymbol{X}2_i$.

We could define with the canonical variables a correlation coefficient, C, as

$$C = \tfrac{1}{3} \sum [\lambda_i]^{1/2}.$$

However this definition needs to explicitate the projection operators $A1$ and $A2$ and the determination of the eigenvalues of $A1A2$.

An other possibility is to define a correlation coefficient, M, as:

$$M = \left[\tfrac{1}{3} \sum \lambda_i\right]^{1/2}.$$

Let us consider the four correlation matrices R11, R12, R21 and R22, where the indices 1 and 2 are related to atoms 1 and 2 respectively. These matrices are defined as follows:

$$\mathrm{R}nm = \begin{matrix} \langle \delta x_n \delta x_m \rangle & \langle \delta x_n \delta y_m \rangle & \langle \delta x_n \delta z_m \rangle \\ \langle \delta y_n \delta x_m \rangle & \langle \delta y_n \delta y_m \rangle & \langle \delta y_n \delta z_m \rangle \\ \langle \delta z_n \delta x_m \rangle & \langle \delta z_n \delta y_m \rangle & \langle \delta z_n \delta z_m \rangle \end{matrix}$$

In these expressions the angular brackets $\langle\rangle$ represent an average over all the configurations, δx, δy, δz are the original variables and n and m are equal to 1 or 2. It may be noted that except for the case $n = m$, Rnm is not symmetric.

It can be easily demonstrated [13] that the λ_i's are also the eigenvalues of the 3×3 matrices $\mathrm{R}a = \mathrm{R11}^{-1}\ \mathrm{R12}\ \mathrm{R22}^{-1}\ \mathrm{R21}$ and $\mathrm{R}b = \mathrm{R22}^{-1}\ \mathrm{R21}\ \mathrm{R11}^{-1}\ \mathrm{R12}$, where $\mathrm{R11}^{-1}$ and $\mathrm{R22}^{-1}$ are the inverse matrices of R11 and R22 respectively.

Therefore, M may be obtained from Ra (or Rb) instead of the projection operators. It is not necessary to explicitly determine the canonical variables, nor to diagonalize Ra (or Rb) since one needs only the sum of its eigenvalues which is the trace of Ra (or Rb). The four Rnm matrices are simply obtained from the trajectories using the initial variables, R11 and R22 are inverted and Ra (or Rb) is calculated. The canonical correlation coefficient is thus given by:

$$M = \left[\tfrac{1}{3}(\mathrm{Trace}(\mathrm{R}a))\right]^{1/2}$$

$$\text{or } M = \left[\tfrac{1}{3}(Trace(\mathrm{R}b))\right]^{1/2}.$$

M is defined for each atom pair. For fully correlated (or anticorrelated) atomic fluctuations, M is equal to 1 independently of the relative directions of each motion, while $M = 0$ occurs only if the motions are not correlated. We can also calculate an average correlation coefficient for the different atom pairs of a given type: $\langle M \rangle = (1/n)\Sigma M_k$ where the index k represents the different pairs of this type along the sequence and n is either equal to 8 for intra-residue pairs or 7 for inter-residue pairs.

The method may be easily generalized to the case of more than 3 variables in each group. For example each group of variables may consist of the Cartesian (or internal) coordinates of a set of atoms belonging to the same domain of the macromolecule. If both groups have the same number K of variables, the correlation matrices Rnm have dimensions $K \times K$ and M is given by $M = [(1/K)\mathrm{Trace}(\mathrm{R}a)]^{1/2}$ (or $M = [(1/K)\mathrm{Trace}(\mathrm{R}b)]^{1/2}$. Both groups may as well contain a different number of variables $K1$ and $K2$ respectively. In this case the square matrices Ra and Rb have dimensions $K1 \times K1$ and $K2 \times K2$ respectively. It can be demonstrated [13] that the operators product $A1A2$ and $A2A1$ have, at most, K eigenvalues different from 0 where $K = \min[K1, K2]$. It is thus sufficient to calculate the trace of the matrix Ra or Rb having the smallest dimensions. The correlation coefficient M is obtained exactly in the same manner as in the preceding case. Therefore M expresses directly the level of correlation between the fluctuations of two sets of atoms.

3. Application

All data presented in this paragraph concern protons of strand I. We have obtained the same results for strand II (not shown). Indeed both strands are identical so that the conformation of the duplex is symmetric. However the helical parameters of one strand, such as rise, twist, tilt and roll, defined at the EMBO workshop on DNA curvature and bending [14], vary along the sequence leading to significant differences between nucleotides. No data corresponding to inter-strand proton pairs are presented.

Correlated motions have been studied for proton pairs belonging to a same nucleotide (intra residue) and for proton pairs belonging to different residues (inter residue) in d(CTGATCAG). In the first category we have analyzed the aromatic (H8 or H6)-H1′, aromatic-H2′ and aromatic-H2″ pairs. In the second category we have analyzed aromatic (H8 or H6)-aromatic, aromatic-H1′, aromatic-H2′ and aromatic-H2″ pairs both in the 3′–5′ and in the 5′–3′ direction. We have also analyzed correlations between the H8 or H6 of each extremity with the H8 or H6 of the other residues of the sequence. Examples of correlation matrices obtained with the original variables are given in Table 1. It can be seen that off-diagonal components may be of the same order of magnitude as the diagonal components. Figs. 1a and 1b show the value of M corresponding to H8 or H6 of the extremities and the other aromatic protons of the sequence. It is observed that the correlation factor is rather small for non-adjacent residues ($M = 0.12 \pm 0.07$) and independent of the residue. For the adjacent

Table 1
Correlation matrices R12 calculated with the original Cartesian coordinates. Upper part: intra-residue (T5) H6–H1′ proton pair; lower part: inter-residue (G3–A4) H6–H1″ proton pair

0.72	0.19	0.39
−0.46	0.25	0.47
0.26	0.09	0.84
−0.24	0.28	−0.09
0.33	0.60	−0.31
−0.25	−0.44	0.42

bases the correlation is more important since the values of M are about two times higher. This last feature does not seem to hold for all adjacent bases of the sequence. As shown in Fig. 1c only 5 adjacent bases have correlation factors lying between 0.2 and 0.3 while 2 have small correlation factors (0.08 and 0.15).

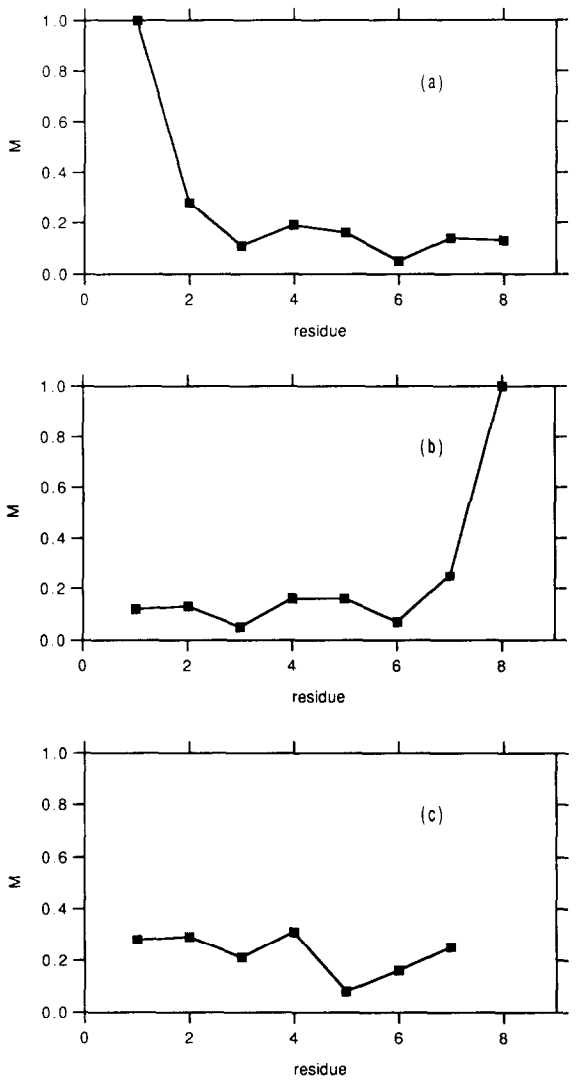

Fig. 1. Correlation coefficient, M, between aromatic protons H8 or H6 of strand I, as a function of residue number. (a) H6 of cytosine 1 with other aromatic protons. (b) H8 of guanine 8 with other aromatic protons. (c) Aromatic proton pairs belong to adjacent bases.

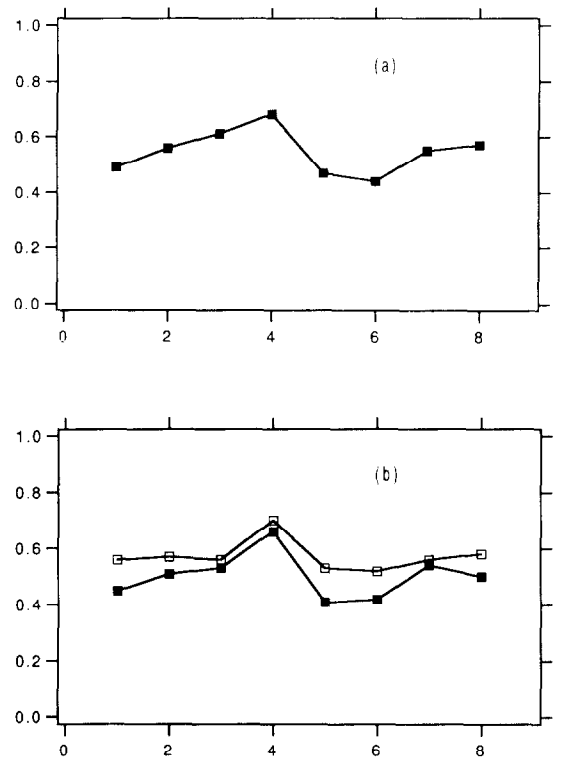

Fig. 2. Correlation coefficient, M, between aromatic protons H8 or H6 of strand I and sugar protons of the same nucleotide, as a function of residue number. (a) aromatic-H1′; (b) aromatic-H2′ (empty squares) and aromatic-H2″ (filled squares).

For the intra-residue H8 (or H6)-H1′ pairs (Fig. 2a) a rather high level of correlation is observed ($M = 0.55 \pm 0.1$). The same observation holds for the intra residues H8 (or H6)-H2′ and H8 (or H6)-H2″ proton pairs with M values in the range 0.5–0.6 (Fig. 2b).

On the contrary, for inter residue pairs involving an aromatic proton and a H1′, H2′ or H2″ proton, the correlation factors are smaller ($0.1 < M < 0.3$) (Figs. 3a–3c).

Table 2 gives $\langle M \rangle$, the average value of the correlation factor corresponding to some proton pairs types which are of importance for structural studies by 2D-NMR. It is seen that intra residue pairs have a much higher level of correlation than inter residue pairs.

In order to evidence a relationship between correlated motions of protons and the fluctua-

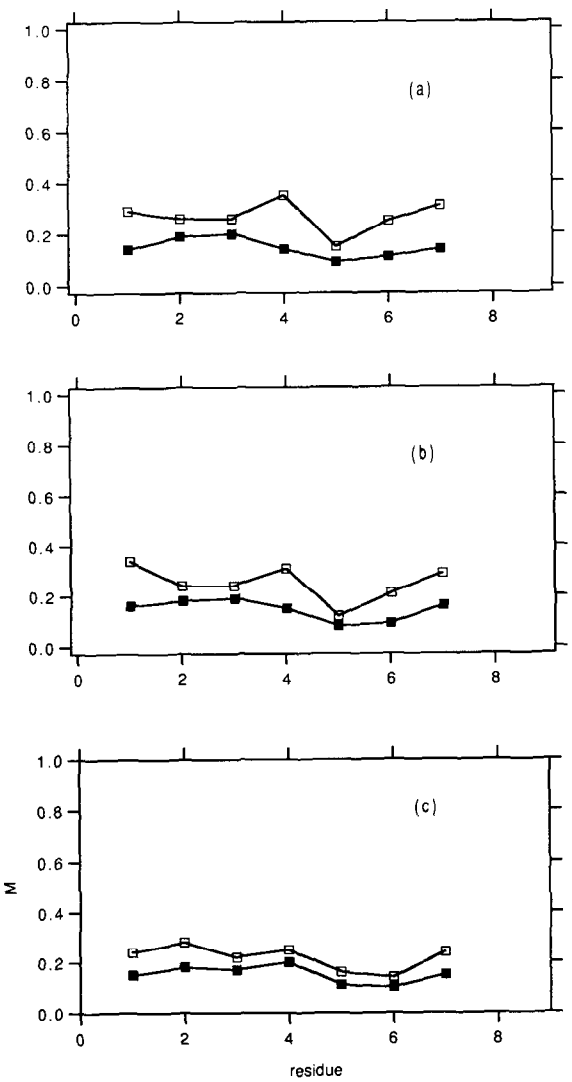

Fig. 3. Correlation coefficient, M, between aromatic protons H8 or H6 of strand I and sugar protons of the adjacent nucleotides, as a function of residue number. (a) aromatic-H2″; (b) aromatic-H2′; (c) aromatic-H1′. Filled squares are for the 5′–3′ direction and empty squares for the 3′–5′ direction.

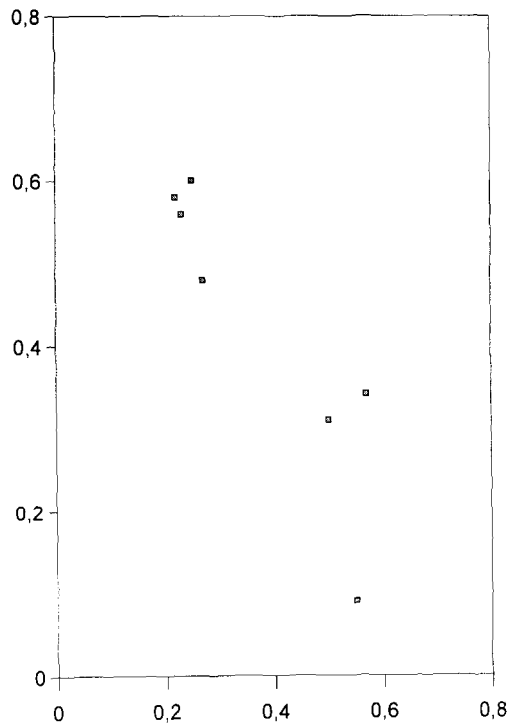

Fig. 4. ⟨RMS⟩ as a function of $\langle M \rangle$. (See materials and methods section for definition).

tions of their distances, we have plotted the average RMS of the distance fluctuations for different types of proton pairs as a function of $\langle M \rangle$ (Fig. 4). The same representation is shown in Fig. 5 for the individual proton pairs. It is obvious that the highest fluctuations correspond to the smallest correlation factors

4. Discussion and conclusion

In the present work we have adapted a canonical analysis approach to investigate correlated

Table 2
Average correlation factor ($\langle M \rangle$) for different types of proton pairs including an aromatic proton H8 or H6 and an other proton (H) belonging to the 5′ neighbour residue (a) or to the same residue (b). The average is taken along the sequence

H	H6/H8 (a)	H1′ (a)	H2′ (a)	H2″ (a)	H1′ (b)	H2′ (b)	H2″ (b)
⟨M⟩	0.23	0.22	0.25	0.27	0.55	0.57	0.50

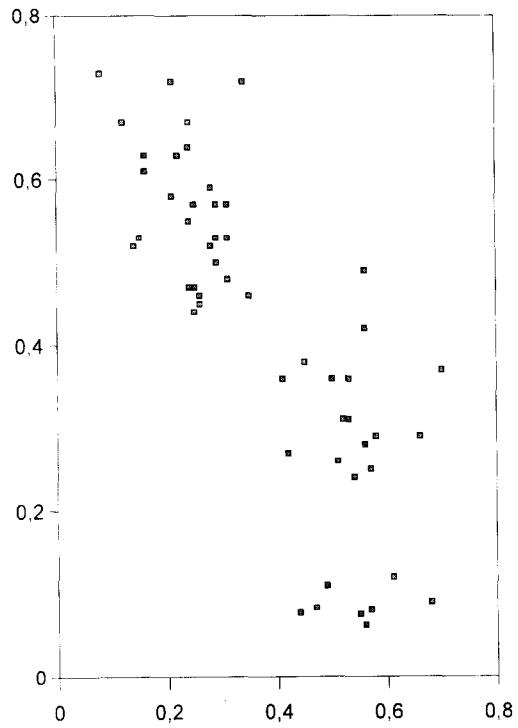

Fig. 5. RMS as a function of M. (See Section 2 for definition).

motions of atoms in d(CTGATCAG). A detailed description of correlation would require a tensorial representation. In particular, information about privileged directions can be obtained by determining the canonical variables. However such an analysis is much less immediate to use than a scalar one. Consequently, a scalar coefficient, M, has been defined in order to quantify the correlation between the fluctuations of atoms. It has the advantage of taking into account all the possible correlations between the different coordinates of both atoms. This is not the case for other correlation factors used previously [5–7]. An important feature of the coefficient defined here is that it does not depend on the relative orientations of the displacements of both atoms. Therefore its value describes really the actual level of correlation. On the other hand it does not indicate if the correlated motions are in phase or in antiphase. However, it is possible to get some idea about the phase or antiphase correlation by combining the RMS measurements and the correlation factor. A small RMS of the interatomic distance coupled to a high M value is consistent with in phase correlation while a high RMS coupled to a high M value favours antiphase correlation. Indeed, if both atoms move in the same direction in a correlated manner, it is expected that their mutual distance varies less than if they move in the opposite direction.

This analysis has been applied to the study of correlated motions between the aromatic protons H8 or H6 of each base and the aromatic protons of the other bases or the H1′, H2′ and H2″ protons of the deoxyribose ring belonging either to the same residue or to adjacent residues. The proton pairs investigated may be classified in 3 classes according to the values of their average correlation coefficient, $\langle M \rangle$.

A first class contains the intra residue pairs. The corresponding values of $\langle M \rangle$ are large, ranging from 0.50 to 0.57 which indicates that the motion of the base is highly coupled to the sugar ring deformation of its own deoxyribose. However, it should be pointed out that the correlation is far from its maximum value. About 40 to 50% of the aromatic motions are not coupled to the sugar motion. At the same time, the RMS of the distances between the aromatic protons and its H1′, H2′ or H2″ is small, indicating that the base and the sugar protons move in phase.

A second class includes the inter residue proton pairs corresponding to aromatic–sugar protons in the 3′–5′ direction. In this case, the average correlation factor, $\langle M \rangle$, has a considerably reduced value (0.22–0.27). Therefore the motions of the base are only moderately coupled to the motion of the adjacent sugar. Furthermore, the average fluctuations of the corresponding interproton distances are important (0.48–0.58 Å).

The aromatic–aromatic inter residue pairs also belong to the same category. These findings have to be related to fluorescence polarization data in DNA–ethidium bromide complex. It has been observed that the number of intercalated molecules does not affect this parameter, from which it was deduced that adjacent bases rotate independently of each other [15]. It was also found by analyzing the rotational motions of the bases from the same simulation as used in this

work that adjacent bases do not move in phase [10]. The present study confirms this observation.

The last class is characterized by a very small correlation factor (0.09–0.15). It concerns the aromatic–aromatic pairs separated by at least one residue and the inter-residue aromatic–sugar proton pairs in the 5′–3′ direction. For the aromatic–aromatic proton pairs it is interesting to note that the correlation does not vary along the sequence. It is possible that the values found for M are in the limit of the accuracy domain. Another possibility would be the existence of a very weak residual correlation between the motions of all the bases.

Concerning aromatic–sugar proton pairs it is obvious from Figs. 3a–3c that the correlation factor in the 5′–3′ direction is significantly smaller than in the 3′–5′ direction. Ichiye and Karplus reported for BPTI a decrease in the correlation when the number of covalent bonds between two atoms increases [6]. This cannot be the case here since the number of bonds between an aromatic proton and a H1′, H2′ or H2″ of an adjacent residue is the same in both directions. In our case the difference seems to be more related to a smaller distance in the 3′–5′ direction than in the 5′–3′ one, which is a consequence of the helical geometry of the DNA molecule. Although there are conformational differences between residues belonging to the same strand, no relation is found between helical parameters and correlation coefficient.

The present study, based on correlated motions analysis, shows that a nucleotide is a much more rigid entity than a dinucleotide, at least on a 200 ps time scale, although the correlation between a base and its sugar ring is not complete. Such an approach could be helpful for interpreting NMR data.

An alternative for analyzing correlation between atomic positions would be to calculate a scalar correlation coefficient, M', between δr_i and δr_j, the distance fluctuations from mean position of atoms i and j respectively. M' would not depend on the relative orientation of atomic motions. However the correlation in this case would have a different meaning since it concerns only the fluctuations norm. On the contrary the coefficient M used in the present study depends both on the correlation between norms and on the correlation between orientations of the fluctuations, although not on the value of the relative orientation. For the one-dimensional case, i.e. correlation between two scalar variables, M is the usual correlation coefficient, while M' is related to correlation between absolute values of the variables.

Of course, the correlation factor defined here may be used to study correlated motions in proteins. An important application would be to study relative motions between two molecular sub-units in a very simple way. In this case the two groups of variables would contain the coordinates of atoms which compose both sub-units.

All the present work deals with equal time correlations. It is also possible to assign atomic coordinates taken at the same time to the first group of variables while the second group is composed of atomic coordinates taken at a time t later. The determination of M as a function of t would give temporal correlation functions, either auto-correlation, if the variables of both groups are coordinates of the same set of atoms, or cross-correlation if they belong to different sets.

References

[1] C.L. Brook, M. Karplus and B.M. Pettit, Proteins: a theoretical perspective of dynamics, structure and thermodynamics (Wiley, New York, 1988).
[2] J.A. McCammon and S.C. Harvey, Dynamics of proteins and nucleic acids (Cambridge Univ. Press, Cambridge, 1987).
[3] W.F. van Gunsteren and H.J.C. Berendsen, Angew. Chem. Intern. Ed. Engl. 29 (1990) 992–1023.
[4] J.H. Noggle and R.E. Schirmer, The nuclear Overhauser effect (Academic Press, New York, 1971).
[5] J.A. MacCammon, Rept. Progr. Phys. 47 (1984) 1–46.
[6] T. Ichyie and M. Karplus, Proteins 11 (1991) 205–217.
[7] S. Swaminathan, W.E. Harte and D.L. Beveridge, J. Am. Chem. Soc. 113 (1991) 2717–2721.
[8] K. Wüthrich, NMR of proteins and nucleic acids (Wiley, New York, 1986).
[9] L.P.M. Orbons, G.A. van der Marel, J.H. van Boom and C. Altona, Eur. J. Biochem. 160 (1986) 131–139.
[10] F. Briki and D. Genest, J. Biomol. Struct. Dyn. 11 (1993) 43–56.
[11] J.P. Ryckaert, G. Ciccotti and H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327-341.

[12] W.F. van Gunsteren and H.J.C. Berendsen, GROMOS Program System (Biomos Biomolecular Software, University of Groningen, 1986).
[13] G. Saporta, Probabilites, Analyse des donnees et Statistiques (Technip, Paris, 1990).
[14] R.E. Dickerson, M. Bansal, C.R. Calladine, S. Diekmann, W.N. Hunter, O. Kennard, R. Lavery, H.C.M. Nelson, W.K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D.M. Soumpasis, C.S. Toung, E. von Kitzing, A.H.J. Wang and V.B. Zhurkin, J. Mol. Biol. 205 (1989) 787–791.
[15] D. Genest, P.A. Mirau and D.R. Kearns, Nucl. Ac. Res. 13 (1985) 2603–2615.