COMPUTER ANALYSIS OF CONFORMATIONAL POSSIBILITIES OF DOUBLE-HELICAL DNA

V. B. ZHURKIN, Yu. P. LYSOV and V. I. IVANOV
Institute of Molecular Biology, USSR Academy of Sciences, Moscow 117312, U.S.S.R.

Received 23 June 1975
Revised version received 21 August 1975

1. Introduction

X-ray analysis has revealed the existence of several forms of the double-helical nucleic acid: A, B, C [1], T [2] and D [3]. It is therefore important to obtain a complete description of all possible conformations of the complementary helix.

Tumanyan and Essipova [4] have suggested an algorithm which allows one to obtain the base pairs position, if four of the six independent dihedral angles, controlling DNA geometry, are known.

Here we present an alternative approach which permits, starting from positions of the bases (Arnott's parameters [1]), to find all the possible structures of the sugar-phosphate backbone and to calculate energy values of the helices. With this method we are able to suggest a general classification of the forms of the regular double-helical nucleic acid and to obtain interrelationships between the variables, characterizing the spatial structure of the double helix.

2. Methods

Only regular complementary helices with frozen valency angles and bonds and rigid deoxyribose and base rings are considered. The method used is a modification of that for closure of a simple chain proposed by Go and Sheraga [5]. Their algorithm cannot, however, be directly applied to DNA, because the sugar-phosphate chain between the bases contains the non-planar furanose rings and the dihedral angles around the glycosidic bonds, χ, in the regular helices are equal to each other.

The principle of the method is the following: one of the valency angles is allowed to vary and the chain is locked by the changing of five dihedral angles (except χ). Then, by simultaneous variation of the two χ angles the above valency angle is forced to have the required value. For a given disposition of bases the number of possible conformations is finite (in our experience it was not more than 10). A detailed description of the algorithm is given elsewhere [6].

For all the conformations thus found the helix energy per nucleotide pair was calculated as the sum of Van der Waals [7] and electrostatic [8] interactions, the dielectric constant, ϵ, was taken equal to 4. The interaction between a pair and the two closest neighbours only were taken into consideration. The bond lengths and valency angles for phosphate and bases were taken from [9]. Deoxyribose was used in two standard conformations, C2-*endo* [10] and C3-*endo* [9]. The locations of protons were calculated assuming the C–H distance to be 1.0 Å. The winding angle of the helix, τ (i.e. an angle of rotation between the successive base pairs), was varied from 15° to 50° at intervals of 2.5 Å. The distances of a base pair from the axis, D, were as follows (Å): –7.5; –5; –2.5; 0; 2.5; 4.5; 6.5; 10. The pair twist, TW, was 0° and ± 5°. The axial rise per nucleotide, H (Å), and the tilt of the bases, TL, were changed as follows: (TL = 0°, H = 3.4), (TL = ± 5°, H = 3.3); (TL = ± 10°, H = 3.0 and 3.2), (TL = ± 20°, H = 2.6; 2.8 or 3.0), (TL = 30°, H =

For correspondence: Dr V. B. Zhurkin, Institute of Molecular Biology, USSR Academy of Sciences, Vavilov Str. 32, Moscow, U.S.S.R. 117312.

2.6 and 2.8). The signs of TW and TL were such that in A-DNA [9] TW > 0°, TL < 0°.

The obtained conformations with the least energy values (≤–40 kcal/mol pairs) were used as the initial points of the minimization procedure [6].

The calculations were done for dA : dT polymer.

3. Results and discussion

The representation of energetically favourable structures of the double helix, calculated by us, in the form of the points on the map $\{D,\tau\}$ (fig.1) makes it possible to suggest a natural classification of the complementary helices. One may see that there are rather isolated regions which comprise the known forms of the $\bar{A}$- and $\bar{B}$-families (fig.1a; table 1, lines 1 and 4). There are also a number of forms, energetically favourable, which have the dihedral angles radically different from those for the known $\bar{A}$- or $\bar{B}$-forms (table 1, lines 2,3 and 5). We shall define them as the non-canonical forms. Nevertheless in the space of base parameters these forms are adjoined the canonical ones (fig.1b).

3.1. *B-family (C2-*endo *sugar)*

The bottom of a $\bar{B}$-ravine lies near to the positions of known models (fig.1a), the region of the most favourable calculated structures being in a very good accordance with the limits of τ and D variation in the known forms (B, C, D, T). It is noteworthy that the forms of relatively low energy are possible with the base pairs pressed in toward the wide groove ($D \approx -5$ Å).

The $\bar{B}$-region of the map is stretched out in such a way, that, while τ increases, the parameter D decreases. The reason is that the sugar-phosphate chain is lengthening as τ increases, and the strain thus created can be compensated by the base pairs shift. (This allows one to understand why the winding angle in the $\bar{A}$-forms is less than that in the $\bar{B}$-family).

The depedences of the base inclination, TL, of the

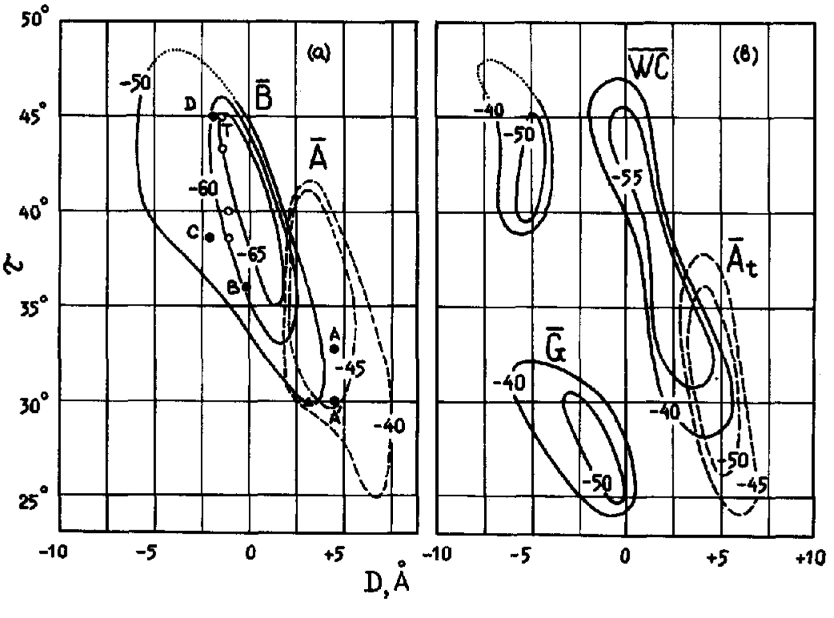

Fig.1. The regions for different families of the canonical (α) and non-canonical (β) forms of DNA in co-ordinates $\{D,\tau\}$. The energy levels in kcal/mol pairs are indicated. The dotted line is for the region with C3-*endo* sugar puckering, the solid one is for C2-*endo* pucker. The upper parts of some regions with C2-*endo* sugar puckering are drawn by points to show that the size of narrow groove is less than 6A. The forms belonging to these areas are unfavourable energetically due to a great electrostatic repulsion between phosphates of the opposite chains. The known forms are labelled as follows: (•) DNA and RNA, (○) glucosylated DNA [2], (▲) dT:dA:dT [16].

Table 1
Dihedral angles for some helices from the nucleic acid families of fig.1

Family	χ, N-Cl	ξ, C4-C5	θ, C5-04	ψ, 04-P	φ, P-01	ω, 03 - C3	Sugar
1. $\overline{B}$	115 ÷ 155[a] (129)[b]	25 ÷ 75 (53)	135 ÷ 200 (181)	−35 ÷ −85 (−65)	−100 ÷ −160 (−114)	170 ÷ 240 (182)	C2-*endo*
2. $\overline{G}$	165 ÷ 180 (172)	−10 ÷ −40 (−16)	100 ÷ 130 (119)	0 ÷ −30 (−26)	110 ÷ 135 (133)	−60 ÷ −110 (−75)	C2-*endo*
3. $\overline{WC}$	95 ÷ 110 (97)	160 ÷ 230 (170)	150 ÷ 210 (183)	85 ÷ 170 (159)	−90 ÷ −125 (−109)	175 ÷ 215 (191)	C2-*endo*
4. $\overline{A}$	55 ÷ 90 (81)	40 ÷ 100 (52)	160 ÷ 205 (190)	−50 ÷ −130 (−73)	−55 ÷ −85 (−63)	−120 ÷ −175 (−161)	C3-*endo*
5. $\overline{A_t}$	45 ÷ 65 (58)	165 ÷ 190 (−103)	−95 ÷ −115 (92)	85 ÷ 125 (−135)	−130 ÷ −145 (−79)	−75 ÷ −105	C3-*endo*

[a] The limits of variation of a given angle in the different forms of the family.

[b] The values of angles for a form of least energy.

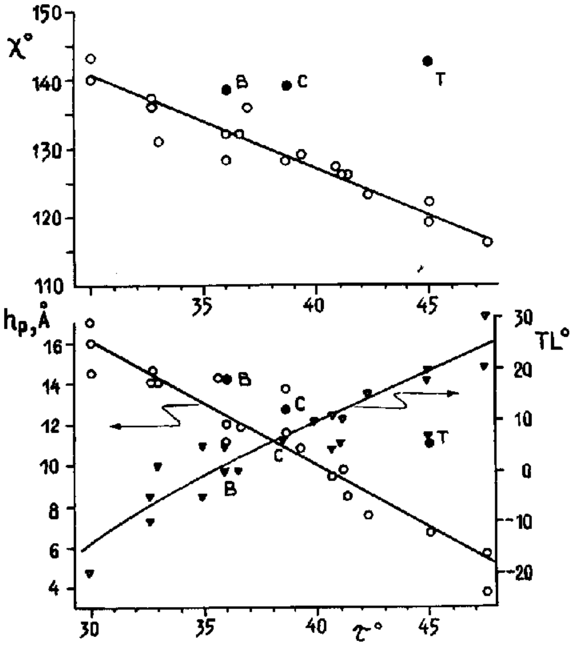

Fig.2. Dependences of χ (○), h_p (○) (left-hand ordinate) and TL (▾) (right-hand ordinate) from τ in the $\overline{B}$-family along the ravine bottom (fig.1a). h_p is the height of narrow groove, determined as a shift along DNA axis between the helical lines drawn through phosphorus atoms. Each point corresponds to an individual form. The filled symbols are for the known models: B [1], C [1], T [2].

height, h_p, of the narrow, or sugar groove (i.e. the one, which is faced by the glycosidic bonds), and of the χ angle from the winding angle, τ, are given in fig.2. As h_p characterizes the distance between negatively charged phosphate groups of opposite chains, the change of h_p is of interest for the consideration of ionic effects upon DNA structure [11,12]. The monotonic change of χ or TL (fig.2), perhaps, is reflected in monotonic decrease in the magnitude of the positive band of the circular dichroism spectrum of DNA during the B→C→T transition [11].

In the $\overline{B}$-family χ angle is high in comparison with the mean value in the monomers [1]. This is due to the unfavourable contact of 3′-sugar with 5′-base at low χ values ($\lesssim 110^\circ$) and $D \lesssim 0$. The increase in D to ∿ 2.5 Å and more weakens considerably the given contact and forms with a smaller χ angle, which are similar to the original Watson-Crick model of 1954 [13], become possible (table 1, line 3; fig.1b).

3.2. *$\overline{A}$-family (C3-*endo *sugar)*

With this sugar ring pucker the chain locks at angles which are by several dozens degrees less than those for C2-*endo* pucker [6]. Therefore, the contact of 3′-sugar with 5′-base prohibits the existence of DNA forms of low energy with the C3-*endo* sugar at $D \approx 0$. This circumstance is responsible for the sharp vertical left boundary in the $\overline{A}$ region (fig.1a).

At the increase in τ the narrow groove of the

$\overline{\mathrm{A}}$-form (now it is non-sugar groove) decreases as in the $\overline{\mathrm{B}}$-family. Concomittantly, base tilt increases, but unlike the $\overline{\mathrm{B}}$-structures its sign is negative.

One more distinction of the C3-*endo* helices is that in this case the canonical and non-canonical forms (table 1, lines 4 and 5) coexist in one and the same region of base parameters (cf. figs.1a and 1b). This poses the urgent question about the real existence of the non-canonical forms, all the more so because according to our calculations these forms can be advantageous energetically (figs.1a,b).

3.3. *A possible functional significance of the different forms of the double helix*

The ability of the nucleic acid helix to exist in different forms and to change smoothly its conformation within the families could be of importance in recognition by proteins of double helical nucleic acids allowing realization of the different structural schemes.

Accessibility of the chemical groups of the bases, exposed to the sugar groove of the double helix, and those of the non-sugar groove depends on the value and sign of the parameter D, e.g. on the distance of the base pairs from the helix axis.

Thus, in recognition by the groups, exposed to the non-sugar groove, the forms with $D<O$ would be stabilized ($\overline{\mathrm{B}}$-family) whereas in recognition by the sugar groove these would be the ones with $D>O$, e.g. the $\overline{\mathrm{A}}$ and $\overline{\mathrm{WC}}$ forms (table 1, line 3).

The forms with small τ and negative D (fig.1b, $\overline{\mathrm{G}}$-region) are possibly realized in the complexes of some dyes, diaminosteroids and antibiotics which according to the model of Gursky [14] unwind DNA with no intercalation.

The non-canonical $\overline{\mathrm{A}}$-forms with *trans-gauche* orientation around C4'-C5' sugar bond are supposed to be important in RNA polymerase operation [15].

References

[1] Arnott, S. (1970) Progress Biophys. Mol. Biol. 21, 265–319.

[2] Mokulsky, M. A., Kapitonova, K. A. and Mokulskaya, T. D. (1972) Molekul. Biol., 6, 883–901.

[3] Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C. and Watts, L. (1974) J. Mol. Biol. 88, 523–533.

[4] Tumanyan, V. G. and Essipova, N. G. (1974) Dokl. Akad, Nauk SSSR 218, 1222–1225.

[5] Go, N. and Scheraga, H. A. (1970) Macromolecules 3, 188–195.

[6] Zhurkin, V. B., Lysov, Yu. P. and Ivanov, V. I. (1975) Biopolymers, submitted for publication.

[7] Lakshminarayanan, A. V. and Sasisekharan, V. (1969) Biopolymers 8, 475–478.

[8] Renugopalakrishnan, V., Lakshminarayanan, A. V. and Sasisekharan, V. (1971) Biopolymers 10, 1159–1167.

[9] Arnott, S. and Hukins, D. W. L. (1972) Biochem. Biophys. Res. Commun. 47, 1504–1509.

[10] Arnott, S. and Hukins, D. W. L. (1972) Biochem. J. 130, 453–465.

[11] Ivanov, V. I., Minchenkova, L. E., Schyolkina, A. K. and Poletayev, A. I. (1973) Biopolymers 12, 89–110.

[12] Zhurkin, V. B., Lysov, Yu. P., Ivanov, V. I. and Malenkov, G. G. (1975) Molek. Biol. (SSSR) 9, 95–104.

[13] Crick, F. H. C. and Watson, J. D. (1954) Proc. Roy. Soc. A 223, 80–96.

[14] Gursky, G. V. (1970) Studia biophysika 24/25, 265–275.

[15] Florentiev, V. L. and Ivanov, V. I. (1970) Nature 228, 519–522.

[16] Arnott, S. and Selsing, E. (1974) J. Mol. Biol. 88, 509–521.