Base–base interactions in nucleic acids containing A–T base pairs

Structure of poly[d(A–T)]

Malini Rajagopalan, Goutam Gupta* and V. Sasisekharan+

Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560 012, India

Received 24 May 1983

The Watson–Crick type of base pairing is considered to be mandatory for the formation of duplex DNA. However, conformational calculations carried out in our laboratory, have shown that some combinations of backbone torsion angles and sugar pucker lead to duplexes with Hoogsteen type of base pairing also. Here we present the results of energy calculations performed on A–T containing doublet sequences in the D-form with both Hoogsteen and Watson–Crick type of base pairing and the 3 viable models for the A–T containing polynucleotide duplex poly[d(A–T)].

Base–base interaction *Watson–Crick basepair* *Hoogsteen basepairs* *Poly [d(A–T)]*

1. INTRODUCTION

It has been generally accepted that the Watson–Crick type of base pairing [1,2] is mandatory for the formation of the double helical structure of DNA. However, conformational calculations carried out in our laboratory [3] have shown that not all allowed combinations of the backbone torsion angles and sugar pucker lead to duplexes with Watson–Crick type of base pairing; a few lead to base pairs of the Hoogsteen type. This different type of hydrogen bonding scheme was first observed by Hoogsteen [4] in single crystals having pairing of adenine with thymine. Further X-ray studies [5–7] have shown that base pairing in crystals containing adenine with thymine or uracil have four types of hydrogen bonding schemes: Watson-Crick; Hoogsteen; reverse Watson-Crick; and reverse Hoogsteen. Hoogsteen pairing has been implicated [8] in the fiber-type molecular structure of thymidylyl 3′,5′-deoxyadenosine. The fiber pattern obtained by them is consistent with a 7 residues/turn left-handed Hoogsteen paired structure with parallel chains. Single crystal studies on adenosyl 3′,5′-uridine phosphate (ApU) [5] and [5′-*P*-adenylyl-(3′–5′)-thymidylyl-(3′–5′)-adenylyl-(3′–5′)thymidine] (dApdTpdApdT) [9] however, have shown that these compounds form mini double helices with Watson-Crick base pairing. It is not surprising that adenine and thymine containing structures show both kinds of pairing (Watson–Crick [5,9] and Hoogsteen [8]), since both these bases retain the same tautomeric forms for the two types of pairing. This is in contrast to the case of guanine and cytosine where the bases necessarily differ in their tautomeric forms, so as to give rise to the two types of structures. In addition, the 3 hydrogen-bonded Watson–Crick paired G–C is more stable than the corresponding two hydrogen-bonded Hoogsteen pair.

A 7-fold left-handed double helical model for the D-conformation based on Hoogsteen type of pairing for DNA containing A–T and I–C base

+ Chemistry Department, SUNY at Albany, NY 12222, USA

* To whom reprint requests should be addressed

Published by Elsevier Science Publishers B.V.
00145793/83/$3.00

pairs has been proposed [10]. However, improved fiber patterns confirmed the fact that for poly[d(A–T)] [11] as well as poly[d(I–C)] [12], the D-form is most stable and there are 8 residues/turn, both right-handed and left-handed structures being possible.

We report here the results of the energy calculations for A–T containing doublets in the 8-fold D-form with both types of base pairing, and for the 3 types of stacking arrangements, viz., right-handed Watson-Crick, left-handed Watson-Crick and left-handed Hoogsteen models, with anti-parallel arrangement of the chains. All the 3 base arrangements lead to viable models for the D-form of poly[d(A–T)]. Details of the results for the other polymorphous forms and their implications will be described elsewhere.

2. METHODOLOGY BASE–BASE INTERACTIONS

In the base–base interactions, both base pairing and base stacking play a dominant role. Calculations performed in our laboratory [13] have shown that the energy contribution from the backbone is small and is similar for the different allowed backbone conformations. Hence, the essential difference comes from base stacking and base pairing. The methodology for the computation of the base–base interactions has been reported [14,15]. Interaction energy for doublets containing A–T in the D-form were computed as a function of the base parameters (D, θ_x and θ_y) [16]. The ranges of the base parameters for the D-form with mononucleotide as a repeat, which give rise to right and left double helical structures are shown in table 1.

Table 1

Ranges of the base parameters for D-DNA giving rise to right and left double helical structures in the allowed sugarphosphate conformations

	Right	Left
D(Å)	−4.0 to −1.5	−4.0 to −1.5
θ_x (deg.)	−12 to −24	+8 to −4
θ_y (deg.)	−6 to +6	−6 to +6

3. RESULTS AND DISCUSSION

Table 2 shows the interaction energy for the $\downarrow_{\mathrm{T}}^{\mathrm{A}}$ $_{\mathrm{A}}^{\mathrm{T}}\uparrow$ and the $\downarrow_{\mathrm{A}}^{\mathrm{T}}$ $_{\mathrm{T}}^{\mathrm{A}}\uparrow$ sequences in the D-form for the 3 models: left-handed Hoogsteen; left-handed Watson-Crick; and right-handed Watson-Crick type in the helical arrangement. Values of the minimum interaction energy are in kcal/2 mol base pair as a function of (D, θ_x). θ_y values are also indicated. Crosses in the boxes indicate that for particular (D, θ_x), irrespective of the θ_y value, stacking arrangement is not possible. From our calculations, it is seen that the $\downarrow_{\mathrm{A}}^{\mathrm{T}}$ $_{\mathrm{T}}^{\mathrm{A}}\uparrow$ sequence prefers left stacking, especially for the Hoogsteen base-paired structures. On the other hand, the $\downarrow_{\mathrm{T}}^{\mathrm{A}}$ $_{\mathrm{A}}^{\mathrm{T}}\uparrow$ sequence favours the right stacking arrangement.

3.1. *Poly[d(A–T)] in the D-form*

The relevance of these results to the polymer structure of poly[d(A–T)] was then investigated. A detailed analysis was performed using the data in [17]. For this purpose, 3 types of models [right-handed Watson-Crick R(W.C.), left-handed Watson-Crick L(W.C.) and left-handed Hoogsteen L(H.)] paired structures with anti-parallel chain arrangements were generated using linked atom least squares (LALS) procedure [18]. Table 3 shows the conformational parameters (backbone torsion angles and glycosidic torsion) obtained for the 3 models of poly[d(A–T)] in the D-form which are stereochemically satisfactory and are in general agreement with the X-ray data. As shown in [19], the backbone torsion angles are similar. Only the glycosidic torsion (X) for the left-handed Watson–Crick structure is 60° less than the right-handed Watson–Crick structure.

The essential difference between an anti-parallel Watson-Crick structure and an anti-parallel Hoogsteen paired structure is that the glycosidic torsion (X) for both purine and pyrimidine bases is identical in the former but $X_{\mathrm{purine}} = 180 + X_{\mathrm{pyrimidine}}$ in the latter. Model building studies [20] have shown that for a right-handed structure, an anti-glycosidic torsion is favourable as compared to a low anti-glycosidic torsion for the left-handed structure.

The base parameters for the 3 models have also been indicated in table 3. These parameters are different from the parameters corresponding to minimum energy values given in table 2. This dif-

ference is due to the fact that the best 'R' factor is achieved for the base parameters given in table 3 and not the parameters corresponding to minimum energy values of table 2.

The interaction energy values for the 3 models are comparable and the values are $E_{R(W.C.)} \simeq -20$ kcal/mol, $E_{L(W.C.)} \simeq -23$ kcal/mol and $E_{L(H.)} \simeq -22$ kcal/mol, respectively.

The crystallographic *'R'* factors obtained from the three models are also comparable (table 3) indicating that all three models agree equally well with the X-ray data. From the limited data available, further refinement is not possible to discriminate between the 3 models.

The stacking arrangements of the $\downarrow_{\mathrm{T}}^{\mathrm{A}}\,{}_{\mathrm{A}}^{\mathrm{T}}\uparrow$ and the $\downarrow_{\mathrm{A}}^{\mathrm{T}}\,{}_{\mathrm{T}}^{\mathrm{A}}\uparrow$ sequences in the D-form of poly[d(A–T)] have been obtained. Fig. 1 shows the $\downarrow_{\mathrm{T}}^{\mathrm{A}}\,{}_{\mathrm{A}}^{\mathrm{T}}\uparrow$ sequence stacking arrangements for the 3 models. Although there are considerable differences in the stacking arrangements for the $\downarrow_{\mathrm{T}}^{\mathrm{A}}\,{}_{\mathrm{A}}^{\mathrm{T}}\uparrow$ (fig. 1) and the $\downarrow_{\mathrm{A}}^{\mathrm{T}}\,{}_{\mathrm{T}}^{\mathrm{A}}\uparrow$ (not shown) sequences, as mentioned earlier, all the models are energetically equally favourable and their *'R'* factors are comparable.

Table 2

Interaction energy E in kcal/2 mol base pair for the $\downarrow_{\mathrm{T}}^{\mathrm{A}}\,{}_{\mathrm{A}}^{\mathrm{T}}\uparrow$ and the $\downarrow_{\mathrm{A}}^{\mathrm{T}}\,{}_{\mathrm{T}}^{\mathrm{A}}\uparrow$ sequences in the D-form for the left-handed Hoogsteen L(H.), left-handed Watson-Crick L(W.C.) and the right-handed Watson-Crick R(W.C.) paired structures

(a) Interaction energy for the $\downarrow_{\mathrm{T}}^{\mathrm{A}}\,{}_{\mathrm{A}}^{\mathrm{T}}\uparrow$ sequence for L(H.) and L(W.C.)

θ_y (deg) \ D(Å)		−4.0	−3.5	−3.0	−2.5	−2.0	−1.5	
8.0	L(H.)	−33.0[a]	−33.0[a]	−33.0[a]	−33.2[a]	−33.3[a]	−33.3[a]	
	L(W.C.)	−39.6[a]	−40.6[a]	−41.7[a]	−43.0[a]	−44.5	−46.4	
4.0	L(H.)	−35.0[a]	−35.0[a]	−35.1[a]	−35.1	−35.0[a]	−34.8	
	L(W.C.)	−41.2[a]	−42.2[a]	−43.1[a]	−44.3	−45.6	−47.2[b]	
0.0	L(H.)	−36.2	−36.6	−37.0	−37.2	−37.1	−36.8	
	L(W.C.)	−43.7	−44.7	−45.7	−46.6	−47.2	−48.5[b]	
−4.0	L(H.)	X	−35.5[b]	−40.2	−40.5	−40.3	−39.7	
	L(W.C.)	−46.8	−48.0	−49.0	−49.4	−50.2	−50.6[b]	

E_{min} L(H.) = −40.5 kcal/2 mol; E_{min} L(W.C.) = −50.6 kcal/2 mol.

Energy values obtained are due to the bond polarizability method [15]. θ_y values for the interaction energies and 0 unless otherwise specified: [a]$\theta_y = -3$; [b]$\theta_y = +3$; and [c]$\theta_y = +6$. The minimum interaction energy values have also been indicated

(b) Interaction energy for the $\,{}_{\mathrm{T}}^{\mathrm{A}}\,{}_{\mathrm{A}}^{\mathrm{T}}\uparrow$ sequence for the R(W.C.)

θ_y (deg) \ D(Å)	−4.0	−3.5	−3.0	−2.5	−2.0	−1.5
−24.0	−37.4[b]	−39.1[b]	−41.6[c]	−44.0[b]	−47.0[b]	−50.0[b]
−20.0	−37.1[b]	−38.7[b]	−40.6[b]	−43.0[b]	−45.6[b]	−48.4[b]
−16.0	−37.3[b]	−38.8[b]	−40.5[b]	−42.4[b]	−44.6[b]	−47.1
−12.0	−38.1[b]	−39.4[b]	−40.8[b]	−42.4[b]	−44.1	−46.5

E_{min} R(W.C.) = −50.0 kcal/2 mol

(continued)

Table 2 (continued)
(c) Interaction energy for the $\overset{T}{A}\,\overset{A}{T}\uparrow$ sequence for the L(H.) and L(W.C.)

θ_x(deg) \ D(Å)		−4.0	−3.5	−3.0	−2.5	−2.0	−1.5
	L(H.)	−45.0[a]	−46.1[a]	−47.4[a]	−48.5[a]	−49.6[a]	−51.0
8.0							
	L(W.C.)	−40.0[a]	−40.0[a]	−40.4	−41.5	−42.7	−44.2[b]
	L(H.)	−48.7[a]	−49.6[a]	−50.4[a]	−51.2	−51.9[a]	−52.7
4.0							
	L(W.C.)	−43.1[a]	−42.4[a]	−42.5	−43.0	−43.7	−44.9[b]
	L(H.)	−48.4	−50.0	−51.6	−53.0	−54.9[a]	−55.2
0.0							
	L(W.C.)	−47.6[a]	−46.0[a]	−45.3	−45.1	−45.1	−45.9[b]
	L(H.)	−53.6	−55.0	−56.1	−57.1	−58.0	−58.2
−4.0							
	L(W.C.)	−51.3	−50.1	−49.0	−47.8	−47.0	−47.3[b]

E_{min} L(H.) = −58.2 kcal/2 mol; E_{min} L(W.C.) = −51.3 kcal/2 mol.

(d) Interaction energy for the $\downarrow\overset{T}{A}\,\overset{A}{T}\uparrow$ sequence for the R(W.C.)

θ_x(deg) \ D(Å)	−4.0	−3.5	−3.0	−2.5	−2.0	−1.5
−24.0	−37.1[b]	−38.3[b]	−39.8[b]	−42.0[b]	−44.2[b]	−46.0
−20.0	−37.0[b]	−37.9[b]	−39.3[b]	−41.0[b]	−43.1[b]	−45.4
−16.0	−37.1[b]	−38.0[b]	−39.0[b]	−40.5[b]	−42.3	−44.7
−12.0	−38.0[b]	−38.6[b]	−39.3[b]	−40.5	−42.3	−44.2

E_{min} R(W.C.) = −46.0 kcal/2 mol

Table 3
Conformational parameters (backbone torsion angles and glycosidic torsion), base parameters and 'R' factor for the three models of the D-form of poly[d(A–T)][a]

	R(W.C.)	L(W.C.)	L(H.)
α	200	216	237
β	217	220	210
γ	313	263	273
δ	149	166	131
ϵ	56	30	31
ζ	148	159	132
χ_T	63	14	1
χ_A	63	14	181
D(Å)	−2.5	−2.5	−2.8
θ_x(deg.)	−20.8	1.5	0.6
θ_y(deg.)	−0.5	0.0	0.3
R[b]	0.358	0.323	0.328

[a]Atomic coordinates for the models are available from the authors
[b]Conventional crystallographic residual 'R'

4. CONCLUSION

The main conclusions of the present study are:
1. Base–base interactions depend both on the helical parameters and on the base parameters.
2. Hydrogen bonding is important in deciding not only the stacking pattern but also the pairing scheme.
3. All the 3 models are energetically equally favourable and consistent with the X-ray data.
4. The $\downarrow\overset{T}{A}\,\overset{A}{T}\uparrow$ sequence prefers the left stacking arrangement and the $\downarrow\overset{A}{T}\,\overset{T}{A}\uparrow$ sequence prefers the right stacking arrangement. The significance of the preferences of these two sequences is being investigated in our laboratory.

ACKNOWLEDGEMENT

This work was supported by the Department of Science and Technology, New Delhi.

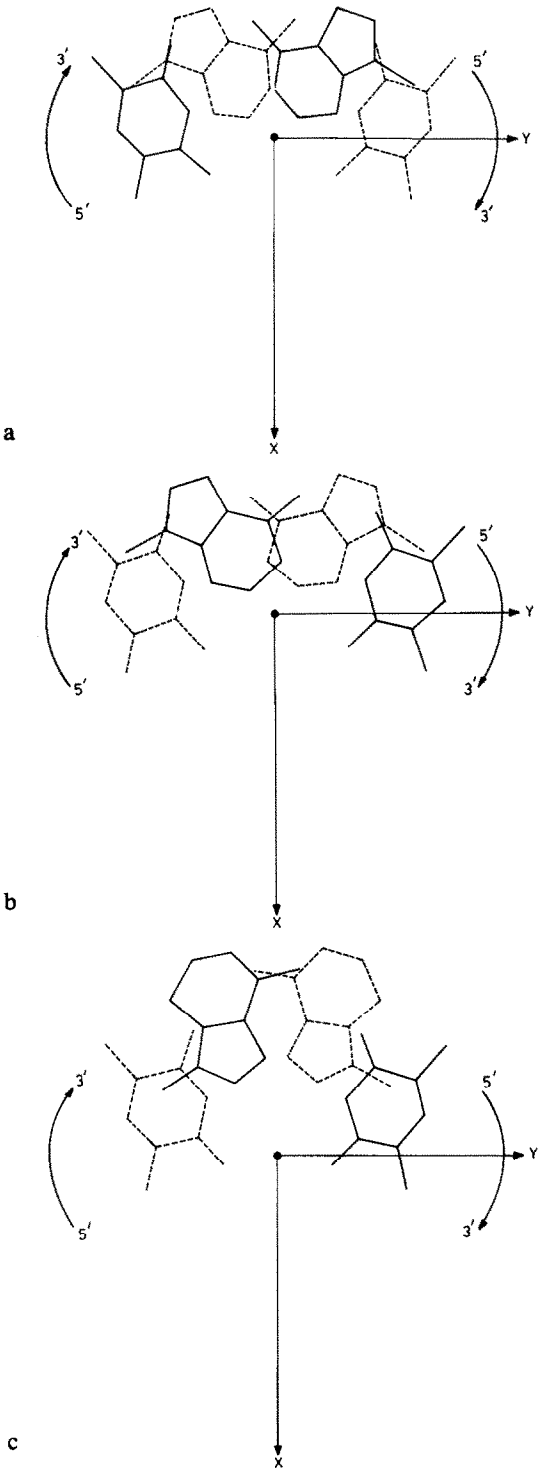

Fig. 1. Stacking patterns of the $\downarrow\overset{A}{T}\overset{T}{A}\uparrow$ sequence for the D-form of poly[d(A–T)] for the 3 models: (a) R(W.C.); (b) L(W.C.); (c) L(H.).

REFERENCES

[1] Watson, J.D. and Crick, F.H.C. (1953) Nature 171, 737–738.
[2] Watson, J.D. and Crick, F.H.C. (1953) Nature 171, 964–967.
[3] Rao, S.N. and Sasisekharan, V. (1983) Int. J. Biol. Macromol. in press.
[4] Hoogsteen, K. (1959) Acta Cryst. 12, 822–916.
[5] Rosenberg, J.M., Seeman, N.C., Kim, J.J.P., Sudath, F.L., Nicholas, H.B. and Rich, A. (1973) Nature 243, 150–154.
[6] Quigley, G.J. and Rich, A. (1976) Science 194, 796–806.
[7] Jack, A., Ladner, J.E. and Klug, A. (1976) J. Mol. Biol. 108, 619–649.
[8] Radwan, M.M. and Wilson, H.R. (1982) Int. J. Biol. Macromol. 4, 145–149.
[9] Viswamitra, M.A., Kennard, O., Shakked, Z., Jones, P.G., Sheldrick, G.M., Salisbury, S. and Falvello, L. (1978) Nature 273, 687–688.
[10] Drew, H.R. and Dickerson, R.E. (1982) EMBO J. 1, 663–667.
[11] Mahendrasingam, A., Rhodes, N.J., Goodwin, D.C., Nave, C., Pigram, W.J. and Fuller, W. (1983) Nature 301, 535–537.
[12] Ramaswamy, N., Bansal, M., Gupta, G. and Sasisekharan, V. (1982) Proc. Natl. Acad. Sci. USA 79, 6109–6113.
[13] Sasisekharan, V. (1973) in: Jerusalem Symp. Quant. Chem. Biochem. (Pullman, B. and Bergmann, E. eds.) vol. 5, pp. 245–260, Academic Press, New York.
[14] Gupta, G. and Sasisekharan, V. (1978) Nucleic Acid Res. 5, 1639–1654.
[15] De Voe, H. and Tinoco, I. (1962) J. Mol. Biol. 4, 500–517.
[16] Arnott, S. and Hukins, D.W.L. (1973) J. Mol. Biol. 88, 93–105.
[17] Arnott, S., Chandrasekharan, R., Hukins, D.W.L. Smith, P.J.C. and Watts, L. (1974) J. Mol. Biol. 88, 523–533.
[18] Arnott, S., Dover, S.D. and Wonacott, A.J. (1969) Acta Cryst. Sect. B25, 2192–2206.
[19] Gupta, G., Bansal, M. and Sasisekharan, V. (1980) Int. J. Biol. Macromol. 2, 368–380.