Volume 12 Number 11 1984 Nucleic Acids Research

The structure of d(CGCGAAT[ ]TCGCG)·d(CGCGAATTCGCG): the incorporation of a thymine photodimer into a B-DNA helix

Shashidhar N.Rao, Joe W.Keepers and Peter Kollman

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA

Received 23 January 1984; Revised 6 April 1984; Accepted 7 May 1984

ABSTRACT

In the light of the biological significance of thymine photodimers, studies of the energetics of the dodecanucleotide fragment d(CGCGAATTCGCG)$_2$ have been carried out using the methods of molecular mechanics, with and without incorporation of a thymine dimer in the cis-syn configuration. The results of the calculations suggest that the thymine dimerized structures show no gross distortion in the double helix with the conformational changes relative to the normal B-DNA double helix restricted largely to the dimer region. The energetics of dTp[ ]dT reveal a number of conformers which are energetically almost equally favorable and are, as a group, qualitatively consistent with NMR studies on this molecule. The biological implications of the results of the conformational studies, reported here, have been examined vis-a-vis the currently available models for the recognition of DNA "damage" by repair enzymes.

INTRODUCTION

Genetic mutations in higher organisms are known to play decisive roles in their development at various levels. Several types of mutations are known to take place in the DNA of higher organisms. Chemical modifications, transitions and transversions of bases are some of the commonly known causes of mutations. DNA is also damaged by absorption of photons of ultraviolet or shorter wavelengths. In the ultraviolet range, absorption mainly affects the pyrimidines, especially thymines, by activating the ethylene bond in the ring. If the activated pyrimidine is next to another activated pyrimidine, bonds may form between them, creating a cyclobutane-type dimer, either on the same strand or between strands. In humans, such damages are known to lead to skin cancer. Of particular interest is the condition known as xeroderma pigmentosum, a genetic disease marked by unusual sensitivity to sunlight that leads to malignancies (1).

In view of their significance, pyrimidine photodimers have been the target of analyses by several investigators involved in understanding their implications for a DNA structure-function correlation. Photochemical and photobiological studies indicate the possibilities of four isomeric forms of dimers (2).

Of these, the cis-syn isomers are known to be biologically the most important product. Several crystallographic analyses have elucidated the conformations of the dimerized pyrimidine bases (3-7). Crystal structures of both uracil and thymine dimers with cis-syn configurations have been carried out (7). Nuclear magnetic resonance (NMR) investigations on dTp[ ]dT[ ] ([ ] denotes the dimer) suggest the possibility of a number of conformations with different sugar puckers, C4'-C5' rotations and glycosydic orientations (8). However, experimental analyses provide limited information on the possible backbone and base conformations.

Recently, conformation energy calculations have been reported on the cis-syn isomer of uracil dimer, dUp[ ]duU (9). The backbone conformations of the energy minimized conformer have been indicated to produce a severe bend in the helix when incorporated into a model B-DNA helix. The present investigations have looked into the conformational aspects of a larger B-DNA double helix in which a thymine dimer has been incorporated. Specifically, we focus on a thymine dimerized structure of $d(CGCGAATTCGCG)_2$. The results of these investigations have been compared with the corresponding ones on the dodecamer without an incorporated thymine dimer. Also, energy minimization studies were carried out on a dinucleoside monophosphate, dTp[ ]dT, with the two thymines dimerized in the cis-syn configuration. The conformational features of these investigations have been compared with those obtained from earlier theoretical and experimental studies.

METHODS

The conformational analyses were carried out using the methods of molecular mechanics wherein energy calculations were performed with the program AMBER (Assisted Model Building with Energy Refinement; see Reference 10). Two different force fields presented by Kollman et al. (11) (F_1) and Weiner et al. (12) (F_2) were employed. These force fields have both been shown to give a reasonable representation of structures and energies of nucleic acids (11,12) but differ substantially in the atomic partial charges used. Carrying out simulations with two different force fields enables one to attempt to assess which results of the calculations are dependent on force field and which are force field independent. The molecular mechanical energy was determined using equation (1) given below and the structures were refined until the root mean square gradient was <0.1 kcal/mole Å. Each of these force fields was supplemented by appropriate parameters for the additional atom types created when the cyclobutane-like thymine dimer was formed, using simple and standard force

constants and equilibrium bond lengths and bond angles. For example, for all the newly created sp^3 carbons in the cyclobutane dimer part in F_1, the values of K_r and K_θ used were respectively 300 kcal/mole/Å^2 and 46.5 kcal/mole, while those of R_0 and θ_0 were 1.53Å and 109.5°, respectively. These values are identical to those of the other sp^3 carbon atoms in the nucleotide. sp^2 atoms were also treated in a similar standard fashion. A list of added parameters is available from the authors. The energy functions are listed in equation (1); in all the

$$E_{total} = \sum_{bonds} K_r(R-R_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 - \cos(n\phi-\gamma)] + \sum_{i<j}^{non\ bonded} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] + \sum_{H\text{-}bonds} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \quad (1)$$

calculations, we used a distance dependent dielectric constant, $\varepsilon = R_{ij}$.

For the $d(CGCGAATTCGCG)_2$ molecule, we have used two different sets of geometrical parameters. Initially, the force field F_1 was used to study torsional variants in this fragment (13). Subsequently, we have employed the structure with the lowest energy after extensive optimization. These geometrical parameters have been used in the studies on the dodecanucleotide with a thymine dimer, carried out in the framework of F_1. We have recently refined the dodecanucleotide using F_2 starting with the B-DNA model proposed by Arnott *et al.* (14). This refined geometry was used in calculations employing F_2. We have created two different cyclobutane-like thymine dimer geometries using each force field. We began by adding to the molecular mechanical energy (Eq. 1) two harmonic constraint energies $K_r(R-R_0)^2$ between the C5 and C6 atoms of thymines 7 and 8 in the dodecamer fragment. We used K_r = 100 kcal/mole Å^2 and R_0 = 2.0 Å to force these atoms close together prior to the formation of the "covalent" bond between them. In the next stage, we have treated the T7C5-T8C5 and T7C6-T8C6 linkages as normal covalent bonds and continued the refinement until convergence. We dimerized the thymines in two stages so that the addition of the extra covalent bonds would not cause excessive distortions during refinement, which could happen in the presence of very large localized energy gradients.

We created the second type of the cyclobutane-like dimer structure in the dodecamer in the following way: We started with the above-described dodecamer (with thymine dimer) and forced the local geometry in the cyclobutane-containing nucleotide to have the parameters suggested by Broyde *et al.* (9) in their studies on dUp[ ]dU. We did this by constraining the backbone dihedral angles of T7 and T8 containing dinucleotide to the values of 61° (C4'-C5'), 154° (C3'-

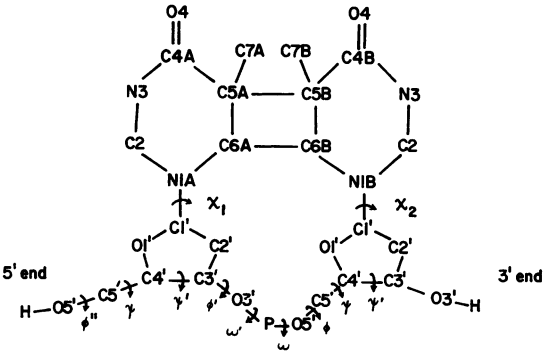

Figure 1: Schematic representation of the dinucleoside monophosphate dTp[ ]dT ([ ] indicates the dimer), investigated in the present study.

O3'), 317° (O3'-P), 251° (P-O5'), 116° (O5'-C5'), 158° (C5'-C4') and 148° and 1°, respectively, about the glycosydic bonds at the 5' and 3' ends (see figure 1 for schematic drawing of a dinucleoside monophosphate). To start with, a constraint weight of 100 kcal/mole was used and after convergence the refinement was continued with the constraints removed. A parallel set of calculations was also carried out on dTp[ ]dT itself, starting the refinements both in the structures which came from our two dodecamer refinements and the refinement by Broyde et al. (9).

RESULTS

Our calculations find that the bond lengths, bond angles and dihedral angles in the cyclobutane fragment of all the thymine dimers, as obtained from the refined structures of the present investigations and crystal structure analyses (3-7), are not significantly different. The specific values for these were in a previous version of this manuscript but have been deleted to fit the manuscript within the page limit. Tables of bond lengths, bond angles and dihedral angles or the actual cartesian coordinates of any of these structures can be obtained from the authors on request. We have designated the six dodecanucleotide models discussed here as D_1 to D_6. Of these, D_1, D_2 and D_3 correspond to using F_1 in the refinement while D_4, D_5, and D_6 correspond to using F_2. D_1 and D_4 represent the dodecamer without the thymine dimer, D_2 and D_5 that with the thymine dimer incorporated into the dodecamer by the two-stage minimization described above. D_3 and D_6 represent the structures in which D_2 and D_5, respectively, have had their conformations forced to values near the dihedral angles obtained in the studies on dUp[ ]dU (9).

TABLE 1. SUGAR PUCKER, GLYCOSYDIC TORSION AND P-O3' TORSION IN THE DODECANUCLEOTIDE MODELS D_1 TO D_6 OBTAINED IN THE PRESENT INVESTIGATIONS. (Only those nucleotides whose conformations are significantly different in the models have been listed.)

Nucleotide Number	SUGAR PUCKER				P-O3' TORSION (ω')			Glycosidic Torsion
	E^3	E_4	E^0, E_1	E^2	trans (180°–240°)	gauche⁻ (240°–270°)	(270°–300°)	
G2 (*)				*			D_4, D_5, D_6	D_1, D_2, D_3 anti
C3 (*)				D_1	**		D_4, D_5, D_6	D_1, D_2, D_3 anti
G4				D_1, D_2	***		***	D_1, D_2 anti (*)
A5 (*)			*				D_4, D_5, D_6	D_1, D_2, D_3 anti
A6				D_4, D_5, D_6	D_1, D_2, D_3	D_1, D_2, D_3	D_4, D_5, D_6	anti (*)
T7 (D_1, D_2, D_4)		D_1, D_2		D_4	D_3, D_5, D_6			*anti high anti (D_3, D_5, D_6
T8 (D_1, D_4, D_5) (D_2,		D_1, D_2	D_3, D_6	D_4, D_5			D_5	D_1, D_2, D_3, anti D_4, D_6 low anti D_3, D_6)
C9				D_4	D_1, D_2, D_3, D_5, D_6		***	D_1, D_2 anti (*)
G10					*	*		anti (*)
C11					*		***	D_1, D_2 anti (*)
A5'			D_2		D_1, D_3, D_4, D_5, D_6		D_4, D_5	D_1, D_2, D_3, D_6 anti (*)
A6'		D_1, D_2	D_3	D_4, D_5, D_6				*anti (*)
T7'				D_4, D_5	D_1, D_2, D_3, D_6		*	anti (*)
T8'		D_1, D_2	D_3	D_4, D_5, D_6				*anti (*)

* D_1, D_2, D_3, D_4, D_5, D_6

** D_2, D_3, D_4, D_5, D_6

*** D_3, D_4, D_5, D_6

In all these models, the conformations about C3'-O3' and C5'-O5' bonds are trans with the one exception in the T7 residues in D_3 and D_6, where the C5'-O5' torsion is in the gauche$^+$ region. Also, the P-O5' conformation is gauche$^-$ in D_1 to D_6. The C4'-C5' bond torsion in these models is mainly confined to a narrow range of 45° to 60° in the gauche$^+$ region. The only exceptions to this feature are found in the thymine dimer part of D_3 and D_6 where the C4'-C5' torsion corresponding to the T8 residue in these models, have trans conformations, with the values of 175° and 170°, respectively. Most of the changes brought by the thymine dimerization seem to be manifested in the modifications of the sugar pucker (ψ'), glycosydic torsion (χ) and the O3'-P torsion (ω'). These conformational parameters vary significantly compared to the abovementioned ones.

Table 1 lists these three parameters as obtained in the models D_1 to D_6. Here, the sugar puckers have been divided into four classes, namely, (C2' endo), (C1' exo, O1' endo), (C4' exo) and (C3' endo). The ranges of phases (W) corresponding to these sugar puckers are (0° to 54°), (54° to 90°), (90° to 144°) and (144° to 190°), respectively. The glycosydic conformation has been classified into anti ($+40° < \chi < 80°$ for sugars with C2' endo, C1' exo and O1' endo puckers; $20° < \chi < 40°$ for C3' endo and C4' exo sugar puckers), high anti ($90° < \chi < 180°$) and low anti ($-60° < \chi \sim 20°$). ω' has been classified into three regions which occupy ranges of (180° to 240°); (240° to 270°) and (270° to 310°). For a given nucleotide in the dodecamer sequence, the appropriate models amongst D_1 to D_6 have been entered under the heading of the three parameters. For example, all six models have their G10 residues with C2' endo sugars, trans conformations about P--O3' at the 3' end and anti glycosydic orientations. It may be noted that the nucleotides on the second strand have been listed only when they differ significantly from their corresponding first strand counterparts. For example, G10' is not listed because it is similar in its conformational features to G10. On the other hand, the conformations of T7 and T8 residues are different from those of T7' and T8' residues, respectively. Hence, these four residues have been listed in Table 1. This is also true of A5' and A6'. In the following paragraphs, the sugar puckers, glycosydic orientations and P--O3' torsions in D_1 through D_6 have been briefly discussed. Figures 2 and 3 pictorically summarize the models D_4, D_5 and D_6. In them are shown superpositions of (D_4 and D_5) and (D_5 and D_6), respectively.

Sugar Puckers and Glycosydic Orientations

As seen from Table 1, most of the nucleotides in the six models have their sugar puckers in the C1' exo, O1' endo and C2' endo regions. The intermediate sugar puckers are associated with χ values less by about 10° to 15° than those

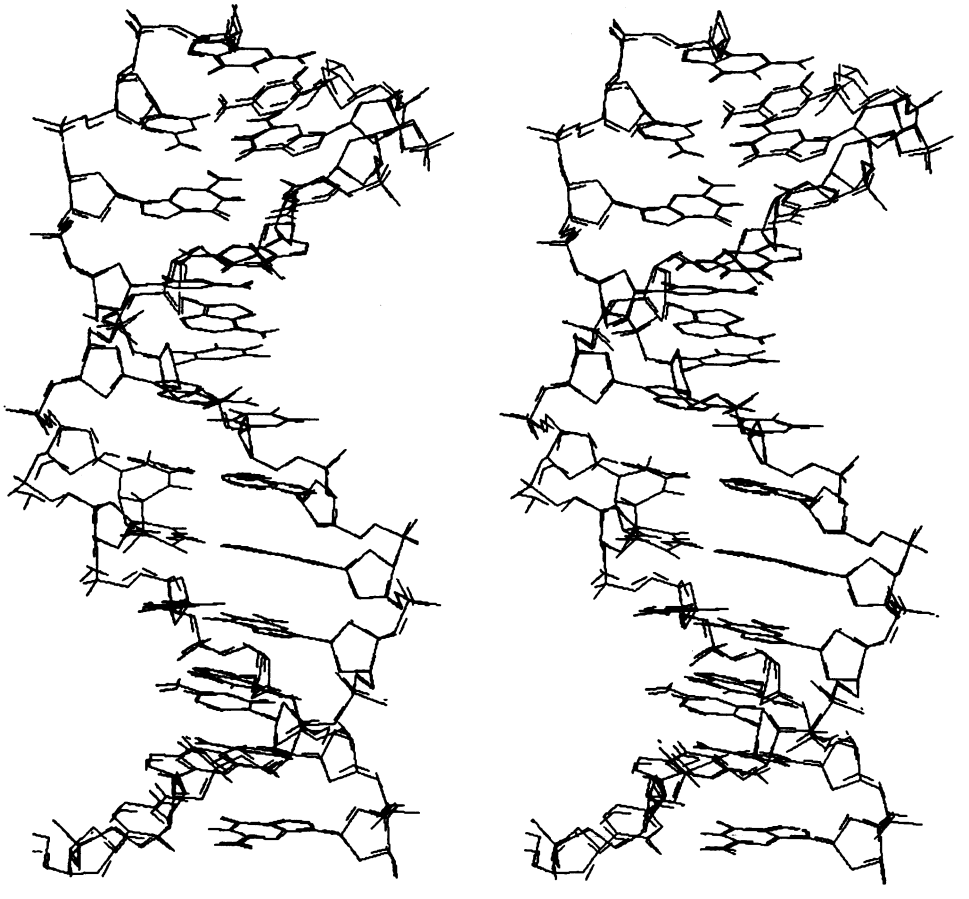

Figure 2: The superposition of dodecamer double-helical structures with (solid lines; D_5) and without (dashed lines; D_4)thymine dimer incorporated into them. It is seen that the two structures are grossly the same, while the region encompassing the central two T-A base pairs [d(ATTC):d(GAAT)] have significantly different conformational features.

corresponding to C2' endo sugars. The C3' endo and C4' exo sugars are found to occur only in the models refined with the old force field parameters, the only exception being the T8 residue in D_6. Interestingly, these puckers are confined to regions containing the thymine dimer and its neighboring nucleotides. It may be noted that in the anti range, the χ values are smaller for C3' endo sugars by about 30° than those for C2' endo sugars. Typically, the anti orientations for the C3' endo sugars correspond to χ values of between 20° and 30°. However, in D_2, thymine dimerization tends to push χ to higher and lower values (at 5' and

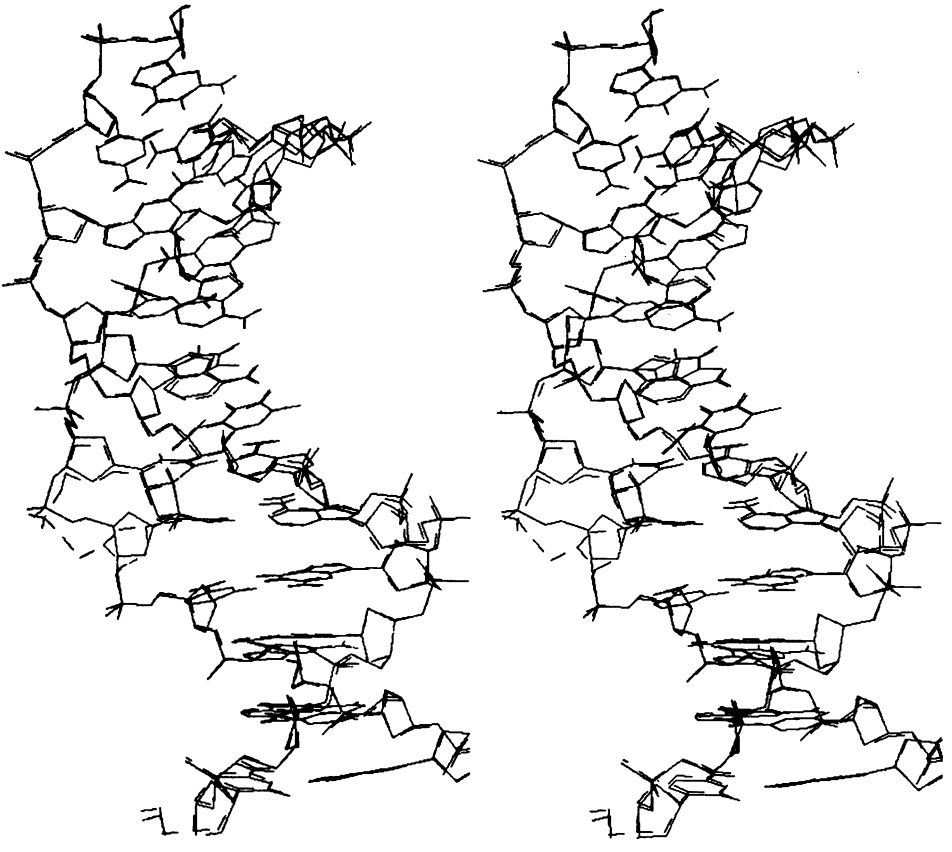

Figure 3: The superposition of the two thymine dimerized structures D_5(solid lines) and D_6 (dashed lines). It is seen that the differences between the two structues are only in the backbone between T7 and T8. It is interesting to not that despite photodimerization, the hydrogen bonding arrangement between T7 and A7' is not significantly altered. T8 is less significantly distorted than is T7. Also, the base-pairs at the 3' and 5' ends of the thymine dimerized regions are essentially as in D_4 (see text for further details).

3' ends respectively) despite the sugar pucker being C3' endo. The glycosydic orientations of the thymine bases in T7 and T8 are significantly altered in models D_5 and D_6 compared to D_4. In D_5, the 5' and the 3' end bases have, respectively, high anti (χ = 114) and anti (χ = 51) orientations. In D_6, the corresponding orientations are high anti and low anti, with accompanying sugar puckers of C2' endo and C4' exo, respectively.

P—O3' Torsion (ω)

The P--O3' torsion ω' varies considerably from trans to gauche$^-$ confor-

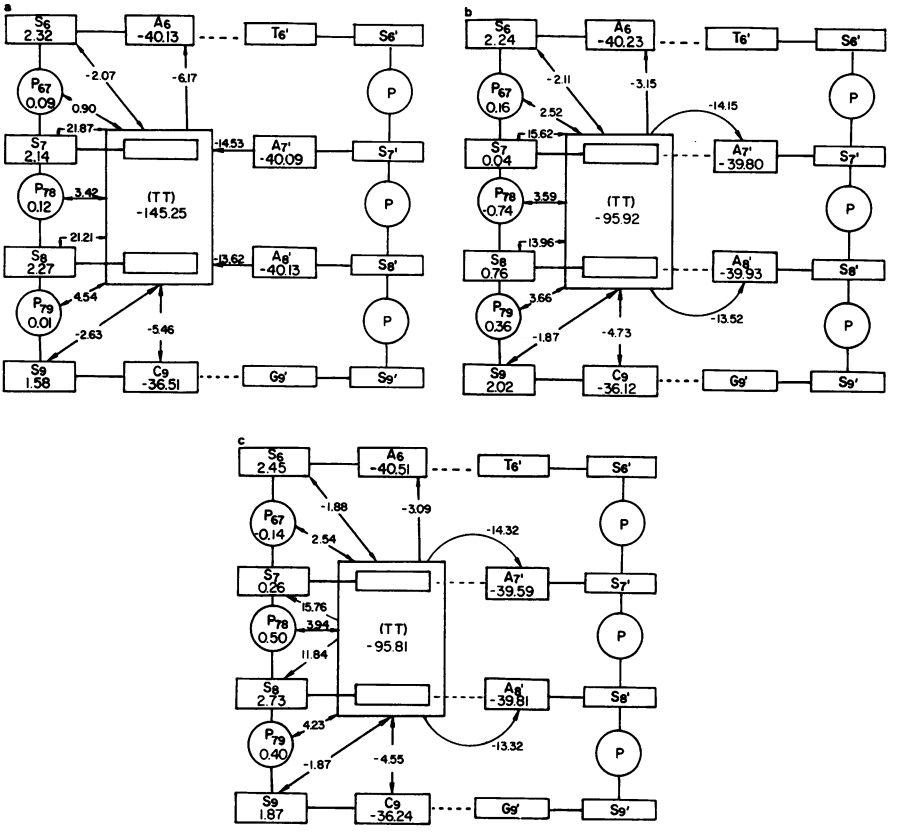

Figure 4: Component analysis of the energetics of the three dodecanucleotide structures D_4 (a), D_5 (b) and D_6 (c). The thymine dimer region (T7 and T8) have been considered together as a single unit (TT). The fragment considered has the central four base pairs [d(ATTC):d(TAAG)]. The bases are marked A6, C9, T6', A7', A8' and G9' apart from (TT). The sugars are denoted by S_7, S_8...etc.

mations over several residues in the models D_1 to D_6. However, ω' is largely confined to gauche$^-$ (280° to 300°) and near gauche$^-$ (240° to 270°) values. The latter range generally corresponds to the intermediate sugar puckers while the former to predominantly C2' endo sugars. The C3' endo sugars are associated with high values of ω'. This type of interdependence between sugar puckers and phosphodiester conformations in double-helical structures of polynucleotides has been observed by several earlier investigators (13,15). Only residues A6 (in D_1, D_2 and D_3) and G10 (in all the models) have their corresponding ω' in the trans domains. These values are mostly around 220° to 240°, deviating to a fair extent from the standard trans conformational value of 180°. Thus, in the

majority of nucleotides in these models, the phosphodiester conformations are (gauche$^-$,gauche$^-$).

Energetics

The total energies of D_1, D_2 and D_3 are, respectively, -562.1, -493.7 and -483.5 kcal/mole; while those of D_4, D_5 and D_6 are -1054.3, -1011.5 and -1013.3 kcal/mole, respectively. The structures without the thymine dimers (D_1 and D_4) are more stable than the dimerized structures. The total energies of D_5 and D_6 are nearly similar, while those of D_2 and D_3 differ by about 10 kcal/mole. For the sake of better understanding of various contributions to the total energies of D_4, D_5 and D_6, a set of component analysis diagrams have been presented in Figs. 4a, b and c, respectively. Here, the sequence of central four base pairs d(ATTC)•d(TAAG) has been considered. In all three figures, the two thymines have been represented as a single moiety called (TT). The intra-group energies have been listed in the boxes, while the arrows indicate inter-group interactions. The other bases are marked A6, C9, T6', A7', A8' and G9'. The sugar moieties are denoted by S_6, S_7...etc.

The major contributors to the difference in energies of D_4 and D_5 (or D_6) (about 40 kcal/mole) are the cyclobutane-like dimer between T7 and T8, distortion of the planarity of the bases and destacking of the bases in the process of dimerization. The difference in energy of about 50 kcal/mole between the blocks (TT) in D_4 and D_5 is mainly due to bond angle deformation in the cyclobutane ring. The endocyclic bond angles of the ring have values ranging between 86° and 94° and thus, deviate by about 15° to 20° from the standard tetrahedral bond angle of 109.5°. Besides, there is a loss of stacking energy between the thymines and the adjacent bases at the 5' (adenine) and 3' (cytosine) ends of the dimer region. As is seen from Figs. 4a and b, the (TT)--(A6) total interaction energies are -6.17 and -3.09 kcal/mole, respectively in D_4 and D_5; while those between (TT)--(C9) are -5.46 and -4.55 kcal/mole, respectively. These energies indicate that T7 is more significantly reoriented in D_5 and D_6 compared to D_4 than is T8. This is reflected in the fact that the orientation of T7 is high anti compared to the anti orientation of T8.

The conformation variations seem to affect the nonbonded and electrostatic interactions of the (TT) dimer with the sugars attached to A6, T7,T8 and C9 and the intervening phosphates. In particular, it is seen that these interaction energies between (TT) and sugars attached to T7 and T8 are higher in D_4 than in D_5 or D_6 by about 6-8 kcal/mole. On a detailed examination of the atom-atom interaction energies in these fragments, it was found that a significant part of this difference was due to stronger electrostatic attractive interactions

between the sugar-base combinations (sugar 7-T8) and (sugar 8-T7) in D_5 (or D_6) than in D_4. This in turn was due to the fact that on dimerization, the C1' atoms of sugars 7 and 8 (charge:~0.5) are dragged closer to atoms N1, C6, C5 and C4 (with charges: -0.74, 0.55, -0.60 and 0.98) on T8 and T7, respectively (see Fig. 3) in D_5 than they are in D_4, with the N1...C1' interactions being the strongest.

Further, it is interesting to note that there is no large change in the interaction energies between the dimerized thymines and the adjacent hydrogen bonded adenines. On a detailed examination, it was found that the electrostatic and nonbonded contributions to the interaction energies between T7 and A7' in D_4 and D_5 differed by about 0.5 and 0.25 kcal/mole, respectively. Thus, the total interaction interaction energies between the bases T7 and T7' are different in D_4 (-12.82 kcal/mole) and D_5 (-12.08 kcal/mole) by less than 1 kcal/mole. So also is the case with the T8-A8' interactions whose total energies in D_4 and D_5 are, respectively, -12.82 and -12.33 kcal/mole. These small differences in energies are reflected in the hydrogen bond parameters in the T7-A7' and T8-A8' pairs in D_5 compared to D_4. In D_5, the hydrogen bond lengths O4...N6 and N3...N1 are respectively, 2.83 and 2.82Å and the corresponding hydrogen bond angles (O4...N6--H6) and (N1...N3--H3) are, respectively, 14.3° and 14.8°, for T7-A7'. In T8-A8', these parameters have the values of 2.80Å, 2.87Å, 14.8° and 6.0°, respectively. In D_4, the corresponding parameters for the A-T base pairs are 2.80 Å, 2.87 Å, 5.4° and 2.6°. In addition, on dimerization, there is a slight increase in the electrostatic attractive potential energy between the cross bases (T7 and A8') and (T8 and A7') in the dimerized regions.

These facts indicate that the bases are not significantly reoriented relative to the plane containing the base pairs and perpendicular to the helix axis (see Fig. 3). Also, T7 is reoriented more than T8 in the thymine dimerized structures, D_5 and D_6. This implies a greater variation in the glycosydic torsion for T7 than for T8. Interestingly, it is seen that despite the above discussed differences in the energetics of the dodecanucleotides with and without an incorporated thymine dimer, the structures of the two fragments are grossly similar. This is demonstrated in Fig. 2.

Studies on dTp[ ]dT

Table 2 lists the conformational parameters of dTp[ ]dT on which energy minimization studies were carried out as detailed earlier. Here, A and B refer to the dimer fragment from the dodecanucleotide $d(CGCGAATTCGCG)_2$ refined with both the old and the new force field parameters, F_1 and F_2. C refers to the fragment in which the backbone torsions were initially constrained to the values

TABLE 2. MINIMUM ENERGY CONFORMERS OF dTp[ ]dT USING THE OLD AND NEW FORCE FIELD PARAMETERS.

Dihedral[d] Angle	Conformation[a]					
	A_1[b]	A_2[c]	B_1[b]	B_2[c]	C_1[b]	C_2[c]
ϕ''	197.4	280.6	180.4	179.8	182.5	179.9
ψ_1	56.2	41.7	58.6	58.5	(61) 58.2	58.3
ψ_1'	71.4 [4.2]	74.7 [6.8]	140.7 [164.3]	136.2 [148.3]	(C4'exo) 138.2 [152.7]	138.0 [145.9]
ϕ_1'	216.1	241.9	169.7	174.0	(154) 197.3	194.8
ω_1'	304.8	295.5	284.8	277.0	(317) 286.6	286.8
$^1_1$	299.0	293.9	295.5	288.0	(251) 275.7	267.7
ϕ_1	156.1	153.5	170.0	169.5	(116) 62.6	65.2
ψ_2	62.7	59.5	46.4	51.9	(158) 185.4	153.4
ψ_2'	80.4 [4.3]	83.1 [29.3]	119.3 [122.4]	118.3 [122.8]	(C4'exo) 134.7 [145.5]	136.8 [144.1]
χ_1	73.5	53.0	121.6	117.7	(148) 138.5	136.4
χ_2	7.3	31.9	61.6	71.2	(1) 6.6	13.5
E	16.1	-80.7 (-72.9*)	16.8	-72.9	16.2	-74.5

[a]The three conformations of the thymine dimer are designated as A, B, and C (see text for details)

[b]Old force field parameters (F_1).

[c]New force field parameters (F_2).

[d]ψ, ψ', ϕ', ω', ω, ϕ, and χ refer to the torsions about the C5'-C4', C4'-C3', C3'-O3', O3'-P, P-O5', O5'-C5' and C1'-N1 bonds, respectively. The phases of the sugars are indicated in brackets next to the ψ' torsion values. The numbers within paranthesis on the left of column 5 are the torsion angles obtained from the energy minimization studies on dUp[ ]dU (9). All the conformational parameters are in degrees. The torsion ϕ'' is defined as H--O5'--C5'--C4'. E is the total energy of the conformers in kcal/mole.

*Energy after the ϕ'' torsion was restrained to 180°.

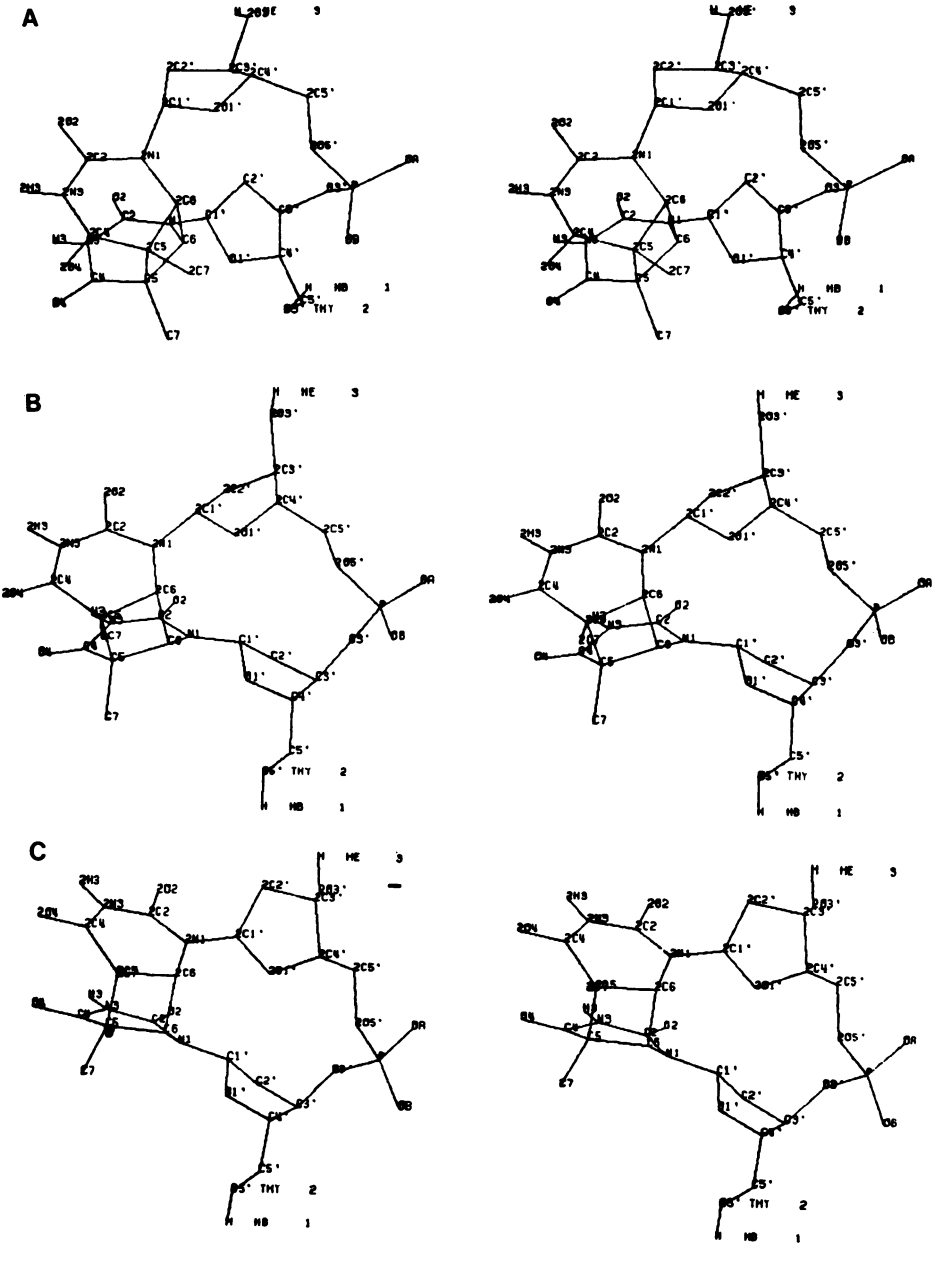

Figure 5: Stereo pairs of the three dTp[ ] dT structures A_2 (a), B_2 (b) and C_2 (c). The O--H....O hydrogen bond between the 5'-OH and one of the phosphate group oxygens is seen in A_2, but absent in the other two structures.

obtained in the energy minimization studies on dUp[ ]dU (9) and then the constraints removed. These are indicated in the parentheses on the left side of column 5. The parameters in columns 1, 3 and 5 correspond to the refined dTp[ ]dT using F_1, while the rest correspond to F_2.

Interestingly, the dimer fragments with the C3' endo sugar puckers are found to be energetically favored over those with the C2' endo and C1' exo sugars. The former have their bases oriented in anti conformation (column 2) as against high anti (5' end) and anti or low anti (3' end) in the latter (columns 4 and 6). However, calculations done in the framework of F_1 reveal that the three structures have nearly the same total energy. On a detailed investigation, it was found that structure A obtained with F_2 had a gauche$^-$ conformation about the C5'-O5' bond. This causes the 5' hydroxyl group to orient in such a way as to promote good hydrogen bonding (O--H.....O) interaction with one of the pendant oxygens in the phosphorus atom intervening the two sugars. However when the H--O5'--C5'--C4' torsion was constrained to trans and gauche$^+$ values, the resultant structure was energetically similar to the other structures listed in Table 2. It may be noted that these have trans conformations about the C5'-O5' bond. Figs. 5a, b and c show the energy minimized conformers of dTp[ ]dT with the three sets of conformational parameters listed in columns 2, 4 and 6, respectively.

The sugar puckers of the dimer fragment in the dodecanucleotide are confined to either C2' endo or intermediate sugar puckers such as C1' exo and O1' endo. Conformers A and B have a gauche$^+$ conformation about C4'-C5' and trans conformations about C3'-O3' and C5'-O5'. Also, the phosphodiester conformations are (gauche$^-$,gauche$^-$) in both the cases, as observed for most phosphodiesters in the dodecanucleotide. However, in C, the 3' end C4'-C5' conformation is trans as against gauche$^+$. These calculations emphasize the significance of conformational flexibility in generating families of energetically favored structures for the dimer fragment and, hence, oligomers with these fragments incorporated into them.

DISCUSSION

Our parallel simulations of d(CGCGAATTCGCG)$_2$ and d(CGCGAAT[ ]TCGCG)•d(CGCGAATTCGCG) have found reasonable conformations for incorporating a thymine dimer fragment into a B-DNA double-helix. In fact, three such conformations have emerged from this study; one from smoothly "forming" the thymine dimer using the lowest energy conformation of the B-DNA helix from after extensive and conformational variation using molecular mechanical parameters F_1

(11), that is D_2; a second from a similar operation on an energy refined B-DNA helix using parameters F_2 (D_5) and a third derived from an earlier study by Broyde et al. (9) which we implemented using both parameter models (D_3 using F_1 and D_6 using F_2). The somewhat surprising result is that all the models are similar in energy and none show a great "distortion" of the B-DNA helix. Energetically, D_2 is ~10 kcal/mole more stable than D_3 using force field F_1, but D_5 is ~2 kcal/mole less stable than D_6, using F_2. Given the simplicity of our molecular mechanical model (due to lack of explicit representation of solvent and counterion effects) and because the two force fields do not consistently predict one conformation lower than the other, we cannot place too much significance in the abovementioned quantitative differences, other than to note that all the conformations should be considered in the analysis of thymine dimer structures incorporated into DNA helices. We note that no definitive structural studies (by X-ray crystallography or NMR) on a thymine dimer incorporated into a long B-DNA helix (such as the dodecamer), have so far been reported and hope that our calculations will provide an impetus for such a study.

We have also studied the above conformations in a dinucleotide model dTp[ ]dT and the results are similar to those found in the longer helix; that is, all the three conformations are similar in energy with force field F_1 but conformation A (Table 2; derived from D_4) is lower than B and C using F_2. Again, this energy difference is somewhat of an artefact due to the O5'-H5'...OA (one of the pendant oxygens in the phosphate group) hydrogen bond formed in model A_2.

Comparing the salient conformational features of our studies on dTp[ ]dT with those of the pmr studies by Hruska et al. (8), the following points of interest are brought forth. The solution studies indicated that both C2' endo and C3' endo sugars are observed in dTp[ ]dT. A tendency towards intermediate mixed sugar puckers was also indicated. This is generally in agreement with the present investigations which have resulted in models with C2' endo, C1' exo and C3' endo sugar puckers. The C4'-C5' torsion was indicated to take up both gauche$^+$ and trans conformations (8). Our models A and B have the C4'-C5' torsion gauche$^+$, model C has the C4'-C5' torsion trans. Since A, B and C are of comparable energy, the calculations are in reasonable agreement with experiment.

The solution studies show that glycosydic torsion in both the nucleosides of the dimer fragment need to undergo changes to bring the bases into alignment for photodimerization. High anti orientation was suggested at the 3' end and syn conformation was eliminated (8). On the other hand, we have obtained models with both nucleosides having anti, (high anti and anti) and (high anti and low

anti) orientations, which are all energetically almost equally favorable. The high anti orientations of the bases are found only at the 5' end of dTp[ ]dT.

The C3'-O3' and C5'-O5' torsions were indicated to prefer trans conformation and this is in agreement with the models presented in this paper. These points of difference and similarity between the experimental results and theoretical calculations would lead to the suggestion that the intervening phosphodiester conformations should be preferrably detected as (gauche$^-$, gauche$^-$), in further solution studies using ^{31}P NMR on dTp[ ]dT or its derivatives with the cis-syn configuration. Thus, we conclude that there can be significant torsional variation even in so contrained a structure as a thymine dimer.

How can one reconcile the structures obtained in this study, which suggest rather small perturbations in the double helix, with melting temperature and chemical accessibility studies which suggest significant destabilization of thymine dimer containing helices relative to corresponding DNA sequences not photodimerized? It is clear from proton exchange experiments that base pair open states are populated even well below the actual melting temperature of a DNA helix. Even a rather small destabilization of the Watson Crick structure can substantially shift the equilibrium toward such base open states. Thus, available experimental melting and accessibility data are not necessarily inconsistent with the suggestion from the studies reported here that the gross structure of the Watson-Crick double helix is little altered upon photodimerization; it is clear, however, that such data require a destabilization of the double helical structure relative to base open states. Our calculations find such a destabilization, although they are not sufficiently accurate to determine a realistic quantitative value for it.

A referee has pointed out (and we agree) that our minimization might not have found low energy kinked structures, given that we started with non-kinked structures. This refereee also suggested that conformations similar to the D_3, D_6, C_1 and C_2 models considered here but with (ω', ω, ψ) g^-, g^-, t on both strands, would be kinked, as noted earlier by Crick and Klug. (20). We thus examined structure D_6 in more detail and forced the ψ between A_5' and A_6' to a trans conformation by constrained minimization. This led to a conformation (D_6') with ω', ω, ψ = g^-, g^-, t on both strands, which had comparable energy to D_6. Interestingly, both D_6 and D_6' had rather minimal kinks in that the average helix axis for the first four base pairs makes an angle of only ~ 10° with the helix axis of the last four base pairs. A similar calculation on the normal B-DNA dodecamer (D_2) in which ψ between T7 and T8 and between A5' and A6' were both forced to be trans led to a structure D_2' which had a comparable energy to

D_2 and virtually no kink. Thus we conclude that the presence of a dihedral angle combination (ω', ω, ψ = g^-, g^-, t) is not a sufficient condition for DNA to be kinked, in that, acceptable non-kinked reasonable low energy structures can be found without kinking. We feel this is because changes in other torsional angles (13) can compensate for the ψ ($g^+ \rightarrow$ t) conformational change.

What do our calculations say about the biological implications of thymine dimerization and its repair? The calculations presented here suggest that, while the local environment around the thymine dimer is altered, the gross structural changes may be relatively small. In the dimer region, changes in the glycosidic torsion and base geometries are more significant than those in the other conformational parameters. How can such a result be squared with the fact that repair enzymes must recognize and repair thymine dimers? If such enzymes (17) recognize "kinked" helices, it is likely that the thymine dimer structures presented here would escape detection. A way out of this dilemma would be to suggest that thymine dimerization stabilizes other local minima (kinks, looped out structures), which, because of our limited scan of allowed structures, have not been found in the present investigations. If such structures contribute significantly to the Boltzmann average of conformations, the repair enzymes would be able to recognize and excise them. Even if we were to create such "looped out" or kinked structures (using techniques such as those employed in ref. 13), we cannot calculate energies sufficiently accurately to evaluate whether they would contribute to the Boltzmann average.

However, recent experimental work has suggested an alternate mechanism for thymine dimer repair (18-20). This mechanism, which is specific for the repair of pyrimidine dimerized structures, has two distinctly different activities. The first of them is an N-glycosylase activity that cleaves the N-glycosyl bond between the 5' pyrimidine of a photodimer and the corresponding sugar. The other is an AP (apyrimidinic) endonuclease activity that cleaves a phosphodiester bond 3' to the newly created apyrimidinic site.

Our structural results are not inconsistent with this repair mechanism, since the N-glycosylase activity of the enzyme would be facilitated by protonation of the pyrimidine N1 nitrogen. Such a process may in turn be facilitated by the geometrical distortions around that atom, as are found in models D_2, D_3, D_5 and D_6. In fact, our calculations find a greater distortion of the 5' thymine than the 3' thymine, consistent with the enzymatic cleavage pattern. The torsion angles around the N1 atom in T7 and T8 of these models deviate by about 15 to 20° from planarity; whereas in D_1 and D_4, the bases are essentially planar. In the distorted base, the proton affinity of N1 would be much larger

and subsequent to this protonation, cleavage of the C1'-N1 bond would be expected to be facile. In the absence of the thymine dimers, the above-mentioned geometrical distortions around N1 will be absent and this will prevent N-glycosylase activity of the repair enzymes on normal pyrimidines.

We should stress that our model does not offer any illumination on the activities of a newly discovered repair enzyme not specific for pyrimidine photodimers (21). Such enzymes apparently cleave the DNA eight nucleotides from the 5' end and four nucleotides from the 3' end of the altered region (such as a dimer) of the DNA structure. One cannot rule out that any subtle changes in the single-double strand equilibrium or helix repeat could stimulate the action of these enzymes, but it may be pointed out that, unlike the previously discussed mechanism, such a repair mechanism is not specific for a pyrimidine photodimer.

CONCLUSION

Molecular mechanics calculations have found a number of low energy conformations for thymine dimers incorporated into a B-DNA double-helix, none of which involve large distortions or kinks of the nucleic acid structure. We hope such calculations will stimulate single crystal diffraction and NMR studies on oligonucleotides containing thymine photodimer. Further theoretical studies are also in order. We plan a detailed analysis of the possible structures of thymine dimerized nucleic acid fragments, molecular dynamics studies on a hexamer which incorporates the dimer and finally an exhaustive conformational analysis of dinucleoside monophosphate dXp[ ]dX (X:pyrimidine) and their derivaties. These studies are a step towards developing a more complete understanding of the role of pyrimidine photodimers in nucleic acids.

ACKNOWLEDGEMENTS

This investigation was supported by PHS Grant Number CA-25644, awarded by the National Cancer Institute, DHHS. We are pleased to acknowledge the use of the computer graphics facility at UCSF, supported by RR-1081, R. Langridge, director and T. Ferrin, systems manager. We also are grateful for useful discussions with Suse Broyde and Don Eden.

REFERENCES

1. Marx, J. (1978) Science 200, 518-521.
2. Patrick, M. and Rahn, R. (1976) in Photochemistry and Photobiology of Nucleic Acids, Wang, S. Ed., Vol. II, pp 35-91, Academic Press, New York.
3. Adman, E. and Jensen, L. (1970) Acta Crystallogr. Sect B 26, 1326-1334.
4. Gibson, J. and Karle, I. (1971) J. Cryst. Mol. Struct. 1, 115-121.
5. Camerman, N. and Camerman, A. (1970) J. Am. Chem. Soc. 92, 2523-2527.
6. Flippen, J. and Karle, I. (1971) J. Am. Chem. Soc. 93, 2762-2768.

7. Karle, I. (1976) in Photochemistry and Photobiology of Nucleic Acids, Wang, S. Ed., Vol. I, pp 483-518, Academic Press, New York.
8. Hruska, F., Wood, D., Ogilvie, K. and Charlton, J. (1975) Can. J. Chem. 53, 1193-1203.
9. Broyde, S., Stellman, S. and Hingerty, B. (1980) Biopolymers 19, 1695-1701.
10. Weiner, P.K. and Kollman, P.A. (1981) J. Comp. Chem. (1981) 2, 287-303.
11. Kollman, P.A., Dearing, A. and Weiner, P.K. (1981) Biopolymers 20, 2583-2621.
12. Weiner, S.J., Kollman, P.A., Case, D., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Jr. and Weiner, P.K. (1983) J. Am. Chem. Soc. 106, 785.
13. Keepers, J.W., Kollman, P.A., Weiner, P.K. and James, T.L. (1982) Proc. Natl. Acad. Sci. USA 79, 5537-5541.
14. Arnott, S., Campbell-Smith, P. and Chandrasekharan, R. (1976) in CRC Handbook of Biochemistry, Fasman, G.D. E., Vol. II, pp 411-422, CRC, Cleveland, Ohio.
15. Sasisekharan, V. and Pattabiraman, N. (1978) Nature 275, 159-162.
16. Crick, F.H.C. and Klug, A. (1975) Nature 255, 530-53.
17. Haseltime, W.A. (1983) Cell 33, 13-17.
18. Haseltine, W.A., Gordon, L.K., Lindan, C.P., Grafstrom, R.H., Shaper, N.L. and Grossman, L. (1980) Nature 285, 240-244.
19. Gordon, L.K. and Haseltine, W.A. (1980) J. Biol. Chem. 255, 12047-12050.
20. Gordon, L.K. and Haseltine, W.A. (1981) J. Biol. Chem. 256, 6608-6616.
21. Sancar, A. and Rupp, W.D. (1983) Cell 33, 249-260.