Volume 16 Number 17 1988 Nucleic Acids Research

Structure of d(CACGTG), a Z-DNA hexamer containing AT base pairs

Miquel Coll, Ignacio Fita, Joaquim Lloveras, Juan A.Subirana*, Francesc Bardella[1], Tam Huynh-Dinh[1] and Jean Igolen[1]

Unidad de Química Macromolecular, Escuela de Ingenieros Industriales, Diagonal 647, 08028 Barcelona, Spain and [1]Unité de Chimie Organique, Département de BGM, ERA-CNRS 927, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France

Received April 11, 1988; Revised and Accepted July 6, 1988

ABSTRACT

The left-handed Z-DNA conformation has been observed in crystals made from the self-complementary DNA hexamer d(CACGTG). This is the first time that a non disordered Z form is found in the crystal structure of an alternating sequence containing AT base pairs without methylated or brominated cytosines.The structure has been determined and refined to an agreement factor R=22.9% using 746 reflections in the resolution shell 7 to 2.5 A. The overall shape of the molecule is very similar to the Z-structure of the related hexamer $d(CG)_3$ confirming the rigidity of the Z form. No solvent molecules were detected in the minor groove of the helix near the A bases. The disruption of the spine of hydration in the AT step appears to be a general fact in the Z form in contrast with the B form. The biological relevance of the structure in relation to the CA genome repeats is discussed.

INTRODUCTION

The conversion of rigth-handed B-DNA to left-handed Z-DNA was first detected by changes in the circular dichroism of poly (dG-dC) in high salt solution (1). The atomic organization of left-handed Z-DNA was first characterized in an X-ray diffraction study of an hexamer DNA fragment d(CGCGCG) (2). After a decade the biological role of the Z form remains unknown. The fact that alternating $d(CG)_n$ sequences are uncommon in biological systems adresses the intriguing question of whether or not the cell ever uses this high energy form of DNA. However other alternating sequences containing AT base pairs such as $d(CA)_n$ are widely found in eukaryotic genomes. Indeed the Z conformation has been observed in a number of polynucleotides with alternating purines and pyrimidines containing AT base pairs in their sequences $d(CA/GT)_n$ (3), $d(AS^4T)_n$ (4), and $d(AT)_n$ using Ni^{2+} as

counterion (5) . And also for a variety of deoxyoligonucleotides (6,7). However when AT base pairs are introduced into the sequence the B-Z conversion becomes more difficult (8,9). Prior to this work, Z-form structures had been reported in crystals of oligomers containing AT base pairs with either 5-methylated or 5-brominated cytosines (10) that are known to strongly stabilize the Z-DNA conformation (11) or in crystals with pronounced disorder (12) and containing cobalt hexamine which is a strong Z-DNA enhancer (13).Here we report the crystal analysis of a DNA hexanucleotide pentaphosphate d(CACGTG) duplex at 2.5A resolution which also has a Z conformation.

Comparison of the structure presented here with those determined for the related hexamers d(CGCGCG)(2) and d(CGCGCG)(9) , where D indicates 2-aminoadenine may contribute to better understand the problem why AT base pairs form Z-DNA less readily than GC base pairs.

EXPERIMENTAL

a)SYNTHESIS OF OLIGONUCLEOTIDES AND CRYSTALLIZATION

The oligonucleotide d(CACGTG) was synthesized by the phosphotriester method in solution, from dimers with triisopropylsulfonyl nitrotriazole as the coupling reagent (14). The purification was performed using Sephadex G-25 gel column chromatography and preparative high-performance liquid chromatography on reverse-phase Zorbax OMS(9.3-mm. column).

Crystals were grown using the vapor diffusion method with 2-methyl-2,4-pentanediol (MPD) as a precipitating agent (Table I). Two different cell dimensions were obtained depending on the presence or absence of spermine in the crystallization solution. In both cases the crystals grew very slowly. Due to the scarcity of crystals seeding was not tried. Both types of crystals examined by still and precession X-ray diffraction photographs were found to have the same orthorrombic space group $P2_12_12_1$.

b) DATA COLLECTION

The crystals grown in the presence of spermine diffracted to a higher resolution and were used for data collection. These crystals diffracted up to about 2A, though beyond 2.5A the

Table I
CRYSTALLIZATION

Crystal #1 without spermine	Crystal #2 with spermine
2 mM DNA 23 mM Na Cacodylate pH6 35 mM Mg Cl_2 3.5% MPD Reservoir: 70% MPD 4 months Setting drop	3 mM DNA 30 mM Na Cacodylate pH7 4.8 mM Mg Cl_2 25% MPD Reservoir: 55% MPD Several months Hanging drop
CRYSTAL DATA	
$P2_12_12_1$ a = 16.1 b = 29.7 c = 41.6 Resolution: 2.5 -3.0 A	$P2_12_12_1$ a = 17.6 b = 31.1 c = 44.4 Resolution: 2.0 -2.5 A

intensity decreased very rapidly with resolution.

Three dimensional X-ray diffraction data were collected at room temperature (20·C) in an oscillation Huber camera with 30mA, 40kv, 200μm focal cup 60 mm specimen to film distance using Cu-α radiation from an Elliot GX6 rotating anode generator.

Each photograph was taken by rotating the crystal 8· around the a axis and required a exposure of about 10 hours. 14 film packs (A&B) films) with 1· overlap between contiguous regions were obtained from a single crystal 0.6*0.4*0.1 mm. in size. Radiation dammage increased slowly, being quite apparent in the final photographs. The films were evaluated with the PURDUE system (15).

c) STRUCTURE SOLUTION AND REFINEMENT

The cell dimensions and the spot distribution in the (hol) reciprocal plane, visualized with a twenty degree (μ=20·) precession photograph, in particular the presence of a very strong (0,0,12) reflection are similar to those observed in other orthorrombic single crystals of Z-DNA hexamer duplexess (2,9). Therefore coordinates from the magnesium form of the d(CDCGTG) Z-DNA structure were used as an initial model. Cycles

of Hendrickson-Konnert refinement (16) alternating with visual inspection of ($2F_o$-F_c) maps were carried out in order to obtain the refined structure presented in this work. Coordinates will be deposited in the Brookhaven Data Bank.

In the last cycle the r.m.s. of the shifts was 0.009A for coordinates and 0.39A^2 for temperature factors. The standard R-factor was 22.9% and the weighted R_w factor was 24.4% for 746 structure factor observations with intensities greater than 2.5σ(I) in the resolution shell 7.0-2.5A. The r.m.s. for bond distances in the final model is 0.029A, the value of sigma used during the refinement was 0.027A.

At this stage of refinement 30 solvent molecules have been included in the model. This number is smaller than the number of solvent molecules in other Z-DNA structures, probably due to the lower resolution attained in this study (2.5A).

Only one tentative metal ion was assigned, the position of its strong electron density peak being close (<1.5A) to the hydrated magnesium atom located in the Z-DNA crystal of d(CDCGTG), where it bridges the wide grooves of two symmetrically related molecules at the level of the G10 base for the reference molecule (9). Solvent molecules with low B factors might indicate the presence of hydrated metal ions or clusters of water molecules. However since individual atoms could not be resolved, the corresponding electron density peaks were treated as single solvent molecules. Spermine molecules were not located.

RESULTS

The d(CACGTG) molecule has an overall Z-DNA conformation very similar to that of the related sequences d(CGCGCG) (2), d(CDCGTG) (9) and d(m^5CGTAm^5CG) (10), with an r.m.s. for the common non-hydrogen atoms of 0.8A, 0.6A and 1.1A respectively. As shown in fig.1, the two strands of the duplex form an antiparallel left-handed helix with Watson-Crick base pairing; the purine residues adopt the "syn" conformation while the pyrimidine residues adopt the "anti" conformation. Symmetry related molecules stack end-on-end along the c-axis. The deoxyribose rings of the purine nucleotides with a "syn"

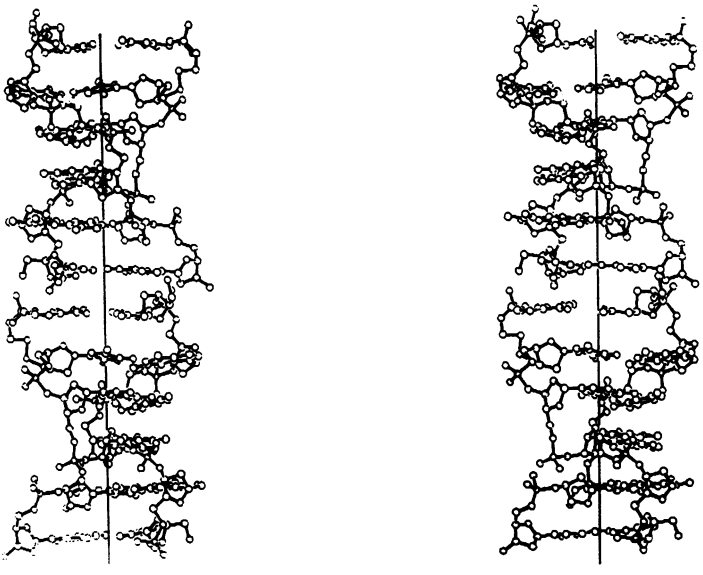

Fig.1. A (stereo) skeletal drawing, made with ORTEP (25) of the d(CACGTG) structure with two symmetry-related molecules stacked end-on-end along the crystallographic c-axis. The two molecules approximately complete a full turn of Z-DNA. The presence of two AT base pairs does not modify the global shape of the Z structure although it has a strong effect in its stability.

conformation are C3-endo, while the pyrimidine nucleotides in the "anti" conformation have a C2-endo sugar puckering as observed previously for Z-DNA structures (2,9) (table II). Only the sugar conformational parameters from nucleotide A2 appears to depart significatively from those observed in other Z DNA conformations such as d(CDCGTG) with an amplitude and phase of 16· and 70· respectively. This might be related to peculiarities in the stacking of the A2/C3 step (see below) or also to subtle differences in crystal packing interactions. The closest interaction between molecules related by translations along the a axis occurs between the phosphodiester chains of nucleotides A2 and T_{11}. Some main chain torsion angles (table III) are unusual for Z-DNA conformations. No restraints were applied during refinement and due to the low ratio between reflections and parameters it is difficult to establish how meaningful these differences might be.

Table II
SUGAR CONFORMATIONS

Residue	Phase (P)*	Amplitude (T_m)*	Conformation
C1	202	63	C2'-endo
A2	32	3	C3'-endo
C3	145	32	C2'-endo
G4	28	32	C3'-endo
T5	143	39	C2'-endo
G6	173	25	C2'-endo
C7	158	37	C2'-endo
A8	15	24	C3'-endo
C9	149	36	C2'-endo
G10	16	22	C3'-endo
T11	149	46	C2'-endo
G12	154	43	C2'-endo

* P and T_m are the deoxyrribose ring conformational parameters defined by Altona and Sundaraligam (26).

Stacking interactions of succesive base pairs in the molecule viewed down the helix axis (fig.2) show minimal differences when compared with the native $d(CG)_3$ structure. It can be seen that there is a small amount of stacking between pyrimidines of opposite strands in the "anti-P-syn" sequences and moderate stacking between adjacent purines and pyrimidines along the same strand in the "syn-P- anti" sequences. The absence of the N2 amino group in adenine decreases the amount of stacking in the A8/C9 pairs (fig.2E) with respect to both the

Table III
MAIN CHAIN TORSIONAL ANGLES*

Alpha	Beta	Gamma	Delta	Epsilon	Zeta
Strand 1					
-	-	-17	-179	-41	52
-29	150	122	127	-11	95
-7	34	178	132	-48	-171
27	-172	117	87	-55	11
-105	110	42	135	-33	40
21	157	98	126	-	-
Strand 2					
-	-	9	149	-23	59
-16	-179	114	97	-90	13
80	-162	49	136	-33	-174
-30	166	174	104	-83	18
55	-137	63	126	-74	-14
-3	156	24	138	-	-

* Main chain: P $\overset{\alpha}{\text{---}}$ O_5' $\overset{\beta}{\text{---}}$ C_5' $\overset{\gamma}{\text{---}}$ C_4' $\overset{\delta}{\text{---}}$ C_3' $\overset{\varepsilon}{\text{---}}$ O_3' $\overset{\zeta}{\text{---}}$ P

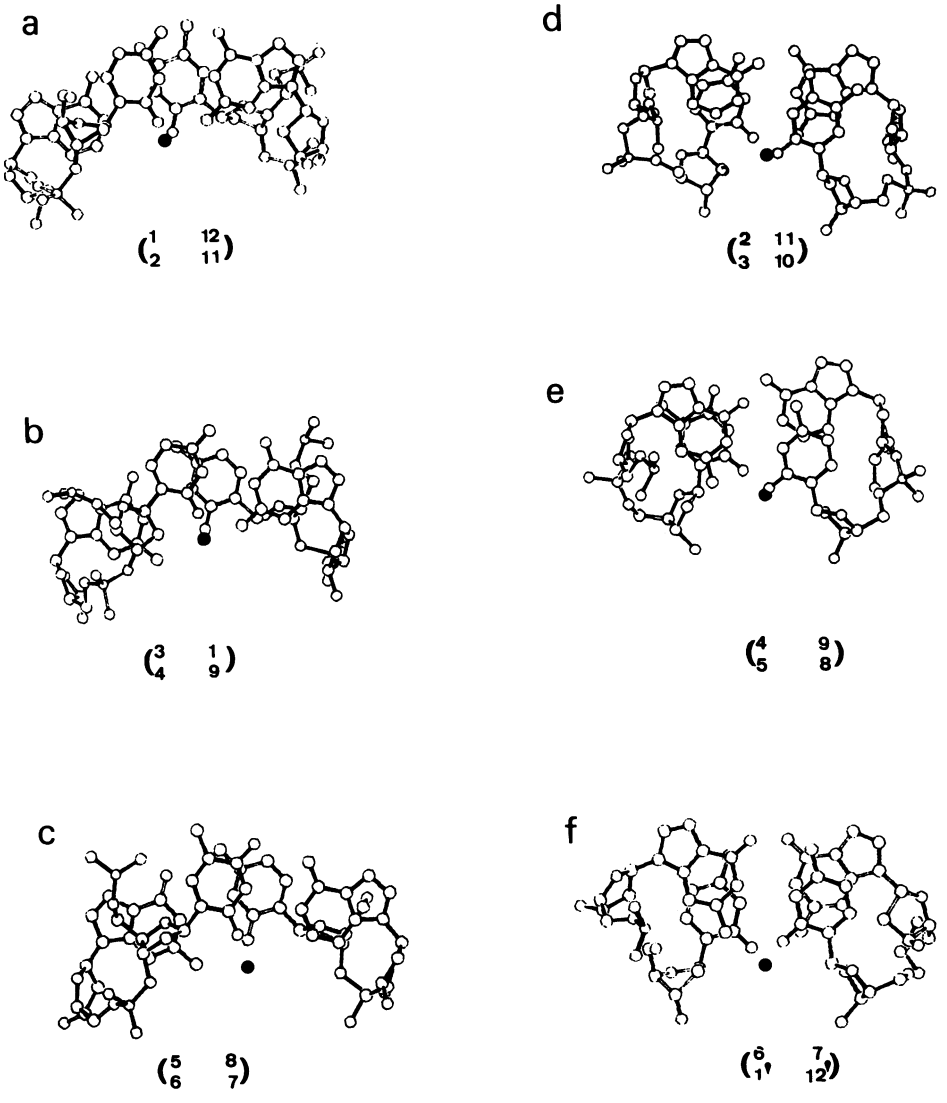

Fig.2. Skeletal drawings illustrating the overlap of successive base pairs in the molecule as viewed down the helix axis. The molecule contains nucleotides 1 to 6 in one chain and 7 to 12 in the other, in such a way that C1 is paired to G12, A2 to T11 and so on. The base pair with solid bonds stacks above the base pair drawn with empty bonds. The solid dot represents the position of the helix axis. Sequences with an "anti-P-syn" (a-c) and with a "syn-P-anti" (d-f) conformation are shown.Figure (f) corresponds to the step between two symmetrical-related molecules.

d(CDCGTG) and d(CGCGCG) structures. Instead the smaller stacking surface of an adenine with respect to a guanine is compensated in the A2/C3 step by a displacement of the A2 in such a way that the six membered ring of the adenine lies over the cytosine ring (Fig. 2D). This is related to a different orientation of the C3 phosphate group close to the ZII conformation, whereas in both d(CGCGCG) and d(CDCGCG) it is in Z_I. However there are no significant differences between the propeller twist of the AT base pairs and the GC base pairs, as was also the case for d(CGCGCG) and d(CDCGCG).

Though the solvent structure in this work has been only partially determined, it is clear, by examining the electron density maps, that no ordered solvent molecules are located in the narrow groove near the A bases. By contrast the N2 amino groups from guanines have a well defined hydrogen bonded water that bridges the adjacent phosphate.Isotropic tempearature factors of atoms in AT nucelotides are similar to those in GC residues, indicating that, whitin the crystal the mobility of the conformation does not relate with the sequence.

Finally it is interesting to note that this hexamer adopts the Z form in crystals in spite of having 33% AT pairs and only one CG sequence. Wang et al (7) have proposed that longer CG sequences are required for Z DNA formation, but our results show that other factors have to be taken into account. In particular packing interactions in the crystal may be of importance, an hexamer appears to be particularly suited to give the Z form.

DISCUSSION

The structure of d(CACGTG) shows that this sequence, under the crystallization conditions, is able to form a Z-DNA structure. Whereas the oligonucleotide d(CGCGCG) crystallizes easily as Z-DNA, d(CACGTG) needs a higher precipitant (MPD) concentration and a much longer crystal growth period. The d(CACGTG) crystals thus obtained diffract only to 2.5A resolution whereas the d(CGCGCG) crystals diffract to about 1A, indicating a higher order in the crystal lattice.Though is not possible to directly relate quality of crystals with stability of the structures obtained it is interesting that the crystal of

the hexamer d(CDCGTG) has an intermediate behavior between the d(CGCGCG) and d(CACGTG) crystals diffracting to 1.3A. In solution studies (14) the B to Z transition was not observed for the d(CACGTG) hexamer even at high ionic strength (unpublished observations).

The lower stability of the Z conformation when AT base pairs are introduced in an alternating sequence can be partially explained by the disruption of the spine of hydration in the minor groove in the region close to the adenines. In particular Wang et al (10) first pointed out that the water molecule that bridges the N2 amino group of the guanines and the nearby phosphate group and stabilizes the syn conformation of the purines was not observed in the corresponding region of the adenines. This is confirmed in our structure and strengthens the idea that the N2 amino group of purine plays a significant role in organizing the solvent in the groove. In sharp contrast, in B-DNA a well defined spine of hydration in the minor groove has been reported involving hydrogen bonds with N3 of adenine.

N3, from both adenine 2 and 8, secluded at the bottom of the groove, appear inacessible for cation coordination. Thus, also the electronic properties of bases (adenine and guanine in particular) may have a significant role in determining the relative stability of AT versus GC pairs (10) in the B to Z-DNA transition.

Another factor of instability might be the poorer stacking interaction in the ApC steps due to the absence of the N2 amino group and thus the smaller stacking surface of the adenines compared with the guanines. It is unclear why this is compensated in one of the two ApC steps with a ZII backbone conformation and a better A/C overlap but not on the other ApC step.

One can summarize the ability to form Z-DNA of the three hexamers compared as follows:

d(CGGCGCG) > d(CDCGTG) > d(CACGTG)

The structure of the d(CACGTG) molecule shows very small changes in comparison with both d(CGCGCG) and d(CDCGTG). For instance there are no significant differences in the width of the minor groove. The low flexibility atributed to the backbone

structure of the Z-DNA (2) agrees with these results. Even in the hexamers d(CGCGTG) (17) and d(CGCGFG) (18), F indicates flour deoxyuridine where wobble base pairing is present, the Z backbone remains almost identical to the original $d(CG)_3$ structure. Packing interactions should not be ignored considering that all these hexamers crystallize in the same kind of orthorrombic lattice. However other Z molecules crystallized in different lattices show the same basic structure. On the other hand B-DNA appears to be a more flexible structure very dependent on the base sequence. In that case different groove width have been observed in the CG streches and in the AT streches (19,21)or even with a high propeller twist in the A tracks (22,23).

Basically it seems that the introduction of AT base pairs in an alternating purine/pyrimidine sequences does not change substantially the Z form but it affects drastically its stability. The conformational properties of these sequences should be related to the functional role, recently suggested by direct studies and by the conservation of poly (GT) tracts in certain locations between evolutionarily diverse species (24).

ACKNOWLEDGEMENTS

We are very grateful to Greg Kramer for his help during film processing. This research was supported in part by a CAICYT grant (2679/83), by the spanish-US cooperation program (8402-039) and by a spanish-french cooperative project 98/86.

*To whom correspondence should be addressed

REFERENCES

1. Pohl, F.M. and Jovin, T.M. (1972), J.Mol.Biol. 67, 375-396.
2. Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., Van Boom, J.H. Van der Marel, G.A. and Rich, A. (1979) Nature 282, 680-686.
3. Nordheim, A. and Rich, A. (1983), Proc.Nat.Acad.Sci.USA 80, 1821-1825.
4. Arnott, S., Chandrasekaran, R., Birdsall, D.L., Leslie, A.G.W. and Ratliff, R.L. (1980), Nature 283, 743-745.
5. Liquier, J., Ridoux, J.P., Bourtayre, P., Pizzorini, L. and Taillandier, E. (1987) In R.H. Sarma (ed.) pp. 11-12 "Book of Abstracts fifth conversation in Biomolecular Stereodynamics" Institute of Biomolecular Stereodynamics, Albany, N.Y.

6. Rich, A., Nordheim, A. and Wang, A.H.J. (1984), Ann. Rev. Biochem. 53, 791-846.
7. Wang, Y., Thomas, G.A. and Peticolas, W.L. (1987), Biochemistry 26, 5178-5186.
8. Zimmer, C., Tymen, S., Marck, C. and Guschlbauer, W. (1982), Nuc.Acid.Res. 10, 1081-1091.
9. Coll, M., Wang, A.H.J., Van der Marel, G., Van Boom, J.H. and Rich, A. (1986), J.Biomol.Str. and Dyn. 4, 157-171.
10. Wang, A.H.J., Hakoshima, T., Van der Marel, G., Van Boom, J.H. and Rich, A. (1984), Cell 37, 321-333.
11. Jovin, T.M., McIntosh, L.P., Arndt-Jovin, D.J., Zarling, D.A., Robert-Nicoud, M., Van de Sande, J.H., Jorgenson, K.F., and Eckstein, F. (1983), J.Biomol.Str. and Dyn. 1, 21-57.
12. Fujii, S., Wang, A.H.J., Quigley, G.J., Westerink, H., Van der Marel, G.A. and Rich, A. (1985), Biopolymers 24, 243-250.
13. Brennan, R.G., Westhof, E. and Sundaralingam, M. (1986), J.Biomol.Str. and Dyn. 3, 649-665.
14. Tran-Dinh, S., Neumann, J., Taboury, J., Huynh-Dinh, T., Renous, S., Genissel, B. and Igolen, J. (1983), Eur.J. Biochem. 133, 579-589.
15. Rossmann, M.G., Leslie, A.G.W., Abdel-Meguid, S.S. and Tsukihara, T. (1979), J.Appl.Cryst. 12, 571-581.
16. Hendrickson, W.A., Konnert, J. (1979) In Srinvasan, R. (ed.) "Biomolecular Structure Conformation, Function and Evolution" pp. 43-57, Pergamon Press, Oxford.
17. Ho, P.S., Frederick, C.A., Quigley, G.J., Van der Marel, G.A., Van Boom, J.H., Wang, A.H.J. and Rich, A. (1985), Embo J. 4, 3617-3623.
18. Coll, M., Saal, D., Lim, M., Aymamí, J., Wang, A.H.J., Manuscript in preparation.
19. Privé, G.G., Heinemann, U., Chandrasegarau, S., Kan, L., Kopka, M.L., Dickerson, R.E. (1987), Science 238, 498-504.
20. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakuda, K., Dickerson, R. (1980), Nature 287, 755-758.
21. Coll, M., Aymamí, J., Wang, A.H.J., Van der Marel, G.A., Van Boom, J.H. and Rich, A. (1988), Biochemistry, in press.
22. Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987), Nature 330, 221-226.
23. Coll, M., Frederick, C.A., Wang, A.H.J., and Rich, A. (1987), Proc.Natl.Acad.Sci.USA 84, 8385-8389.
24. Braaten, D.C., Thomas, J.R., Little, R.D., Dickson, K.R., Goldberg, I., Schlessinger, D., Ciccodicola, A. and D'Urso, M. (1988), Nucl.Ac.Res. 16, 865-881.
25. Johnson, C.K. (1965) ORTEP. Report ORNL-3794, Oak-Ridge National Laboratory, Tennessee.
26. Altona, C. and Sundaralingam, M. (1972), J.Amer.Chem.Soc. 94, 8205-8212.