Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary investigations on a telomeric repeat sequence $C_4A_2C_4A_2$

G. Savitha,[a]* D. Leonidas,[b] K. R. Acharya[b] and M. A. Viswamitra[a,c]†

[a]Department of Physics, Indian Institute of Science, Bangalore 560 012, India,
[b]Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, England, and
[c]Centre for Human Genetics, IT Park, Whitefield Road, Bangalore 560 066, India

† Professor Viswamitra sadly passed away on 10 April 2001.

Correspondence e-mail:
savitha@chemie.fu-berlin.de

Received 22 December 2000
Accepted 5 March 2001

Repeat units based on the telomeric sequence of the ciliated protozoan *Tetrahymena*, d(C_4A_2)$_2$, have been crystallized. Cytosine-rich DNA stretches are known to reside in telomeres and centromeres of eukaryotic chromosomes, playing crucial roles in the structural stability of the chromosome in addition to their connection with cancer and aging. Preliminary investigations on the telomeric repeat sequence $C_4A_2C_4A_2$ from CD studies and X-ray crystal data suggest it to be a right-handed interdigitated tetraplex structure with hemiprotonated C·C$^+$ base pairs. The molecules appear to be packed one on top of another forming a discontinuous helix along *c* simulating a poly-C fibre, an arrangement which maximizes the number of cytosines stacked.

1. Introduction

Ever since the proposal of an intercalated tetraplex model from NMR studies by Gehring *et al.* (1993), the enigmatic C-rich sequences have seen more and more structures with the *i*-motif, sometimes with novel loops. The *i*-motif structure has now been corroborated with high-resolution crystal structures d(C_3T) (Kang *et al.*, 1994), d(C_4) (Chen *et al.*, 1994), d(TAACCC) (Kang *et al.*, 1995), d(CCCAAT) (Berger *et al.*, 1995), d(AACCCC) (Cai *et al.*, 1998) and d(ACCCT) (Weil *et al.*, 1999). Speculations on C-rich structures have long been made. Langridge & Rich (1963) proposed a right-handed parallel double-helical model for polycytidylic acid from X-ray fibre diffraction studies conducted at acidic pH, while poly(rC) and poly(2′-*O*-methyl-C) fibres drawn at neutral pH were interpreted as being single-stranded right-handed sixfold-helical structures stabilized by base-stacking interactions (Arnott *et al.*, 1976; Leslie & Arnott, 1978). A proposal for a duplex with the two polynucleotide strands intercalated instead of base-paired was also made from modelling the mononucleotide crystal structure of 5′-dCMP Na_2 (Viswamitra & Pandit, 1983). However, the self-pairing of cytosine bases *via* three hydrogen bonds has been a common feature in the crystal structures studied (Marsh *et al.*, 1962; Borah & Wood, 1976; Kistenmacher *et al.*, 1978; Egli *et al.*, 1993) and right-handed parallel structures containing both C·C$^+$ base pairs and neutral A·A self-pairs (Westhof *et al.*, 1980; Westhof & Sundaralingam, 1980; Cruse *et al.*, 1983; Krishnan *et al.*, 1991) in dinucleotide crystals. Although it is apparent that they assume non-canonical structures, it is not known what range of conformations is available for the C-rich strands to adopt. Neither is it known to what extent the local and surrounding sequences can influence the ability of C-rich regions to form unusual structures. In order to address this problem, the X-ray analysis of d(C_2A_4)$_2$, based on the telomeric sequence of the ciliated protozoan *Tetrahymena*, has been undertaken.

2. Materials and methods

2.1. Sample preparation and characterization

Oligonucleotide samples were custom made on an automated Applied Biosystems (ABI) 380B DNA synthesizer on a 10 µ*M* scale using the trityl-on protocol. Purification was peromed by FPLC with the reverse-phase chromatography method using a PepRPC HR10/10 column. Buffers *A* and *B* consisting of 100 m*M* TEAA pH 7.0 in 5% acetonitrile against 100 m*M* TEAA pH 7.0 in 40% acetonitrile gave good resolution of the peaks. The sample was dissolved in deionized water and freeze dried, detritylated and desalted with the Fast Desalting column on the FPLC. CD spectra were recorded on a JASCO spectrometer under conditions close to that of the crystallization drop, without spermine and MPD, which would precipitate DNA.

2.2. Crystallization, data collection and preliminary X-ray analysis

The largest crystals grew when equal volumes of 20 m*M* sodium cacodylate pH 7.0, 20 m*M* $MgCl_2$, 2 m*M* spermine and 2.5% MPD were added sequentially to a droplet of 2 m*M* DNA and equilibrated against 25% MPD. The crystals typically grew over a period of 3–4 weeks at room temperature (289 K).

A Siemens Nicolet Area Detector system supported by a rotating-anode X-ray source (Cu $K\alpha$, $\lambda = 1.5418$ Å) with a power rating of 45 kV and 80 mA was used to collect the intensity data. The crystal-to-film distance was maintained at 180 mm with an exposure time of 120 s per frame, an oscillation angle of 0.25° and $2\theta = 15°$. The data has an overall $R_{\rm merge}$ of 4.8% for 957 unique reflections and 706 reflections greater than 2σ from the 4427 measured reflections, with a completeness of 90% for the whole data. The Patterson map was calculated using the *FFTBIG* routine from the *CCP*4 package. Rotation–translation searches were carried out using *MERLOT* (Fitzgerald, 1988), *X-PLOR* (Brünger, 1996) and *AMoRe* (Navaza, 1994).

3. Results and Discussion

The crystals of $C_4A_2C_4A_2$ (Fig. 1) belong to the hexagonal space group $P6_{2(4)}22$, with unit-cell parameters $a = b = 27.35$ (3), $c = 184.46$ (5) Å and 12 nucleotides in the asymmetric unit. Exploratory precession photographs of the $h0l$ zone display the characteristic meridional reflection at 3.2 Å suggesting a stacking of the molecule along $00l$ as continuous helical rods with a missing phosphate between the 12-mers, which was subsequently also confirmed in the Patterson sections. The native Patterson map displays a stack of peaks along c (Fig. 2), with the strongest peak at 6.37 Å along c which is consistent with the approximate 6.4 Å rise per base pair of each duplex in a tetraplex. The Patterson map implies that the helix and crystal symmetry coincide along this direction. As twofold symmetry is a subgroup of $6_{2(4)}22$, one strand could be related to the other by a crystallographic twofold. As CD studies indicate a right-handed helix, the space group appears to be $P6_422$ and not its enantiomer $P6_222$. Thus, the crystal symmetry itself demands a parallel-stranded right-handed helical structure. Further, the occupation of the special position $(0, 0, z)$ in the cell provides clues about the molecular structure of $C_4A_2C_4A_2$ and pointers for model building. Structure determination by the molecular-replacement method would involve only two unknowns, namely the rotation about and translation along c in the cell (Baikalov & Dickerson, 1998).

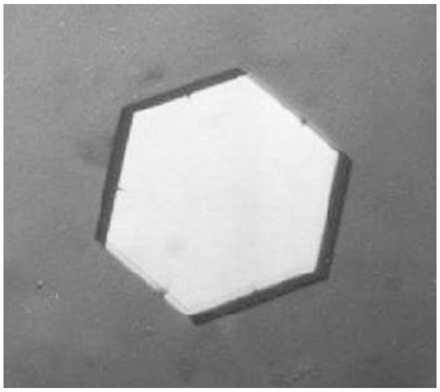

Figure 1
A typical crystal of $C_4A_2C_4A_2$.

CD studies provide clues to the handedness of the molecule and the base-pairing scheme in $C_4A_2C_4A_2$. CD spectra indicate that the molecule exists as an ordered right-handed helical structure in solution at room temperature and the band at 283 nm (Fig. 3) is characteristic of hemiprotonated $C{\cdot}C^+$ base pairs (Langridge & Rich, 1963; Akrinimiski *et al.*, 1963; Inman, 1964). The ordered structure is disrupted upon heating as indicated (Fig. 3) by the lowered amplitude of the CD signal and the blue shift of the maximum peak at 283–278 nm. This was also corroborated with thermal studies, which show that there is a sharp increase in absorbance occurring over a narrow temperature zone, indicating that the transition is cooperative and irreversible.

Models of $d(C_4)$ stretches which were single stranded (Arnott *et al.*, 1976; Leslie & Arnott, 1978), duplex (Langridge & Rich, 1963) and tetraplex (Gehring *et al.*, 1993; Kang *et al.*, 1994, 1995; Chen *et al.*, 1994; Berger *et al.*, 1995) with varying model parameters were generated and rotation–translation searches were carried out rigorously with different resolution ranges, radii of integration, artificial cell size, number of reflections and B factors. In all cases, the strongest peaks in the rotation search were with the tetraplex model as opposed to the duplex and single-stranded models with varying helical parameters.

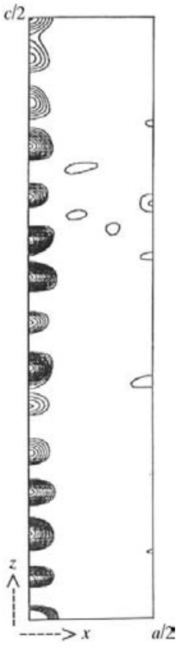

Figure 2
The $y = 0$ section of the Patterson map displaying 6.37 Å spaced vector peaks along the c axis.

Table 1
List of cross-rotation peaks calculated from *AMoRe* from the *CCP*4 package.

Minimum density in map = −37.7; maximum density = 51.4; RMS deviation from mean = 13.4.

No.	α	β	γ	Peak height
1	39.1	44.1	211.4	51.4
2	21.5	16.3	48.5	37.5
3	27.9	75.3	38.0	37.0
4	28.0	39.5	106.7	26.9

The rotation-function (RF) solutions for the resolution range 10–3.1 Å and a radius of integration of 13 Å are listed in Table 1 for a tetraplex model. The tetraplex model always gave RF solutions with inclinations less than 10° with c. In the example, peak No. 1 (Table 1) satisfies this condition and it is indeed the strongest peak. The correlation coefficients were also significantly higher for the $P6_422$ space group compared with $P6_222$, suggesting that $P6_422$ is the correct enantiomer. However, translation searches were difficult and did not yield clear-cut solutions with most of the models. Both RF and the Patterson maps thus point to a tetraplex structure for $C_4A_2C_4A_2$ with stacking along c. Formation of the *i*-motif, which would be intermolecular between four strands of $C_4A_2C_4A_2$ and staggered at half length to form a network-like packing arrangement, also appeared unlikely on the basis of other considerations. The various solved crystal structures have followed the principle that the number of C bases stacked are maximized in tetraplexes. Secondly, the crystal growth (Fig. 1) is maximum in the xy plane. The hairpin model is unlikely, as a hairpin lacks 222 symmetry and within the hairpin the strands would be antiparallel and not parallel (Chattopadhyaya *et al.*, 1988). Besides, the looped-out central adenines within a putative hairpin would not be stacked on the $C{\cdot}C^+$ base pair. Also, the Patterson shows evidence of a continuous stack along the c direction at regular intervals.

From the evidence of cross-rotation searches, which is based on the philosophy of superposition and maximal overlap when there is strong similarity between the model

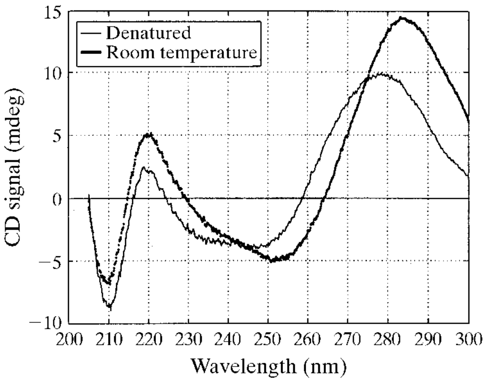

Figure 3
CD spectra of $C_4A_2C_4A_2$ before and after heat denaturation.

and the real structure, and the 6.37 Å peak along *c* in the Patterson maps, we are tempted to believe that $d(C_4A_2C_4A_2)$ is indeed a tetraplex structure. The tentative model of an interdigitated tetraplex for the C-stretch with the adenines looping out is only a fleeting glimpse of the actual structure, which remains to be unravelled.

We acknowledge the X-ray facility for Structural Biology (supported by the Departments of Science & Technology and Biotechnology) at MBU, Indian Institute of Science, India for the several trials made for data collection. The financial support by BRNS (India), DBT and INSA is gratefully acknowledged.

References

Akrinimiski, E. O., Sander, C. & Ts'o, P. O. P (1963). *Biochemistry*, **2**, 340–344.
Arnott, S., Chandrasekaran, R. & Leslie, A. G. W. (1976). *J. Mol. Biol.* **106**, 735–748.
Baikalov, I. & Dickerson, R. E. (1998). *Acta Cryst.* D**54**, 324–333.
Berger, I., Kang, C., Fredian, A., Ratliff, R., Moyzis, R. & Rich, A. (1995). *Nature Struct. Biol.* **2**, 416–425.
Borah, B. & Wood, J. L. (1976). *J. Mol. Struct.* **30**, 13–30.
Brünger, A. T. (1996). *X-PLOR Version* 3.85. Yale University, New Haven, USA.
Cai, L., Chen, L., Raghavan S., Ratliff, R., Moyzis, R. & Rich, A. (1998). *Nucleic Acids Res.* **26**, 4696–4701.
Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K. & Dickerson, R. E. (1988). *Nature (London)*, **334**, 175–179.
Chen, L., Cai, L., Zhang, X. & Rich, A. (1994). *Biochemistry*, **33**, 13540–13546.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A. & Viswamitra, M. A. (1983). *Biochemistry*, **22**, 1833–1839.
Egli, M., Lubini, P., Bolli, M., Dobler, M. & Leumann, C. (1993). *J. Am. Chem. Soc.* **115**, 5855–5856.
Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.
Gehring, K., Leroy, J.-L. & Gue'ron, M. (1993). *Nature (London)*, **363**, 561–565.
Inman, R. B. (1964). *J. Mol. Biol.* **9**, 624–637.
Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyziz, R. & Rich, A. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 11636–11640.
Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 3874–3878.
Kistenmacher, T. J., Rossi, M. & Marzilli, L. G. (1978). *Biopolymers*, **17**, 2581–2585.
Krishnan, R., Seshadri, T. P. & Viswamitra, M. A. (1991). *Nucleic Acids Res.* **19**, 379–384.
Langridge, R. & Rich, A. (1963). *Nature (London)*, **198**, 725–728.
Leslie, A. G. W. & Arnott, S. (1978). *J. Mol. Biol.* **119**, 399–414.
Marsh, R. E., Biersted, R. & Eichhorn, E. L. (1962). *Acta Cryst.* **15**, 310–316.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Viswamitra, M. A. & Pandit, J. (1983). *Curr. Sci.* **52**, 207–209.
Weil, J., Min, T. P., Yang, C., Wang, S., Sutherland, C., Sinha, N. & Kang, C. H. (1999). *Acta Cryst.* D**55**, 422–429.
Westhof, E., Rao, S. T. & Sundaralingam, M. (1980). *J. Mol. Biol.* **142**, 331–361.
Westhof, E. & Sundaralingam, M. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 1852–1856.