Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

When pseudosymmetry and merohedral twinning come across: the case of the d(ApTpApTpApT) oligonucleotide in a hexagonal lattice

Nicola G. A. Abrescia*† and Juan A. Subirana

Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Avenida Diagonal 647, E-080028 Barcelona, Spain

† Present address: Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, England.

Correspondence e-mail: nicola@strubi.ox.ac.uk

Received 24 June 2002
Accepted 27 August 2002

The d(ApTpApTpApT) deoxyhexanucleotide packs in two different lattice systems: monoclinic and hexagonal. In this study, it is shown that in the hexagonal crystals twinning and pseudosymmetry coexist. The presence of both components resulted in a puzzling space-group determination. Only when a correct model became available [Hoogsteen-DNA; Abrescia *et al.* (2002), *Proc. Natl Acad. Sci. USA*, **99**, 2806–2811] was it possible to discriminate between twinning and pseudosymmetry. Here, the steps that led to a correct space-group assignment and to the solution of the oligonucleotide structure in the hexagonal lattice by molecular replacement are described.

1. Introduction

In the last 25 years, many deoxyoligonucleotide fragments have been crystallized either alone or as complexes with either proteins or drugs (Berman *et al.*, 1992). However, the structural information available at high resolution on AT-rich oligonucleotides, alternating or not, is very limited. Only a single 'naked' DNA sequence with only adenine and thymine base pairs has been crystallized so far (Viswamitra *et al.*, 1978). On the other hand, AT-rich sequences are strongly polymorphic (Arnott *et al.*, 1974), so that their study offers a particular interest. Therefore, we decided to study the deoxyhexanucleotide d(ApTpApTpApT). We first obtained crystals in a hexagonal lattice. They presented an unclear space group with indications of twinning. We subsequently obtained crystals in the $P2_1$ space group. Only when the latter structure was solved by the MAD method (Abrescia *et al.*, 2002) were we able to finalize the study of the d(ApTpApTpApT) crystals in the hexagonal lattice as described in this paper.

2. Crystallization procedure and diffraction experiments

The self-complementary deoxyhexanucleotide d(ApTpApTpApT) was synthesized on an automatic synthesizer by the phosphoramidite method and purified by gel filtration and reverse-phase HPLC. The ammonium salt of the hexamer was prepared by ion-exchange chromatography.

Crystallization conditions were tested using the phase-diagram method (Malinina *et al.*, 1987) at room temperature with MPD and NaCl solutions. Only two drops gave rise to unstable hexagon-shaped crystalline forms. Crystallization trials were carried out at three different temperatures: 277, 286 and 295 K.

The best crystals were grown by the vapour-diffusion technique at 286 K from hanging drops using a Linbro multi-well tissue plate. For the hexagonal crystals, 10 µl drops were prepared in 25 m*M* sodium cacodylate buffer pH 6.0 with 0.8 m*M* DNA (duplex), 18 m*M* NaCl and 10% 2-methyl-2,4-pentanediol (MPD) precipitant equilibrated against a reservoir containing 15% MPD. MPD also acts as a cryoprotecting agent. The MPD concentration in the reservoir was increased weekly up to 30–35% until crystals appeared. It was possible to obtain crystals over a wider range of DNA, MPD and NaCl concentrations and also to substitute NaCl with KCl. Nevertheless, good-quality crystals (Fig. 1) were produced by transferring the nucleating crystals for 24–36 h at room temperature or at 310 K, allowing

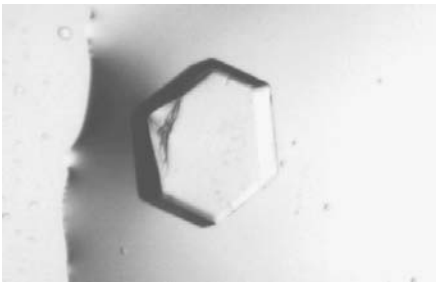

Figure 1
Hexagonal crystal of d(ApTpApTpApT) grown by vapour diffusion at 286 K (approximate dimensions 400 × 400 × 200 µm).

them to dissolve and recooling the drop solution at 286 K. The same protocol was followed for the analogous brominated sequence d(ApTpAp^{Br}UpApT).

Three sets of diffraction data were collected at 110 K. For one of them, an in-house X-ray rotating-anode generator operated at 40 kV and 110 mA was used. In the second case and for the MAD data, images were collected using synchrotron radiation (DESY, Hamburg, Germany, X11 and X31 beamlines) on a MAR Research imaging plate. Crystals were rapidly mounted at room temperature in a nylon fibre loop and flash-frozen in a nitrogen cryostream.

The images were processed and merged using the *HKL* package (Otwinowski & Minor, 1997) and *SCALEIT* (Collaborative Computational Project, Number 4, 1994). Unit-cell parameters, data-collection and reduction statistics for the native data as well as for the derivative data at the remote point are given in Table 1.

3. Space-group ambiguity

Analysis of the systematic absences revealed the presence of a 6_1 axis, which led to four possible space groups: $P6_{1(5)}$ or $P6_{1(5)}22$. No clear distinction was possible by comparing the χ^2 and R_{sym} of the data reduced in $P6_{1(5)}$ and $P6_{1(5)}22$. The self-rotation function (SRF) for $\kappa = 180^\circ$ indicates the presence of strong peaks at $\psi = 0, 30, 60, 90^\circ$ and $\varphi = 0^\circ$ in addition to the crystallographic sixfold at $\psi = 90^\circ$ and $\varphi = 90^\circ$ (Fig. 2*a*). These observations suggested point group 622 for our crystals. Calculation of the solvent content (volume/base pairs) indicated either point group 6 with two independent duplexes in the asymmetric unit or point group 622 with two strands in the asymmetric unit. We then attempted phasing *via* the molecular-replacement method in the four possible space groups using the *AMoRe* program (Navaza, 1994). Because of the values of *a* and *b* dimensions of 32.3 Å, we expected a repetitive unit of ten base pairs along these axes (the helix rise per pair of an ideal B-DNA is 3.4 Å). Consequently, as searching models we used either single- or double-stranded hexamers, pentamers and decamers in ideal B-DNA. The best solution, based on the comparison of correlation coefficients and on molecular fitting, was obtained in space group $P6_5$ with the data at low resolution using a hexamer duplex. This solution showed a clash of the last two terminal bases of the two duplexes present in the asymmetric unit. Although we removed these two overlapping bases, the refinement did not converge. More strangely, this solution showed negative correlation coefficients when the data at high resolution were added. All other attempts at molecular replacement failed.

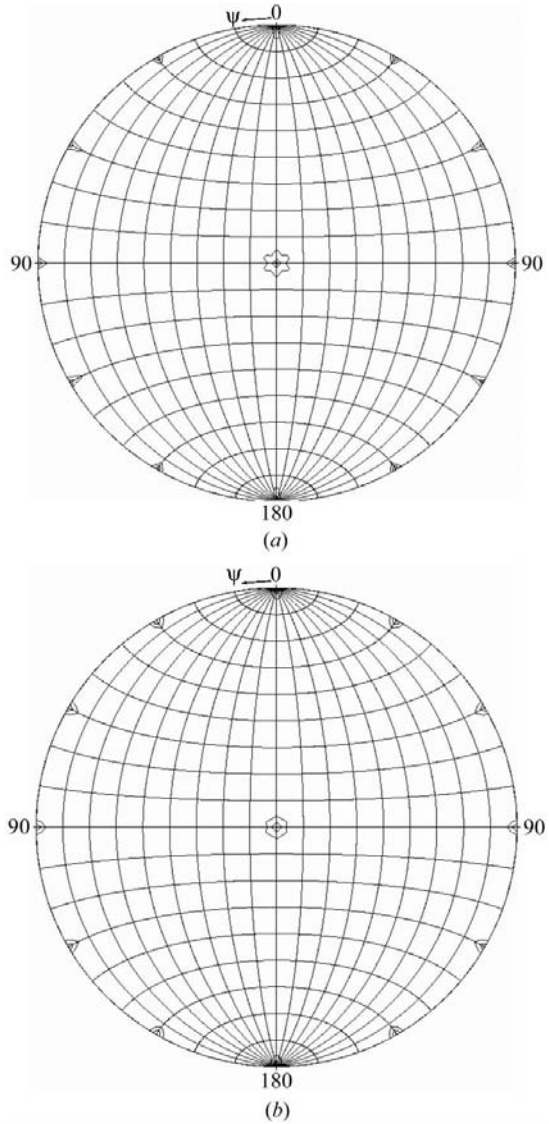

Figure 2
Self-rotation function plots in the resolution range 15–2.5 Å, integration radius 5–45 Å. Peaks are shown for sections at $\kappa = 180^\circ$. (*a*) From the experimental data, (*b*) from the calculated structure factors. Contours start at the 2σ level and increase in 2σ intervals. The highest peak corresponds to the sixfold axis. For the experimental data it is 7.13σ, while for the calculated data it is 6.45σ. Drawings were prepared with the *GroPat* program (Esnouf, 1999).

4. Twinning symptoms

Given the discrepancy of the results obtained with the data at low and high resolution, we carried out a more detailed analysis of the collected intensities. In order to determine possible anomalies, we calculated the R_{merge} between the three different data sets (Table 1) scaled according to the $P6_5$ space group, taking care to check the two possible indexing orientations used by *DENZO* (Otwinowski & Minor, 1997). Although the R_{merge} resulting from this cross-merging was under 10%, its value doubled and the χ^2 stood out at above 2. Surprisingly, these values appeared to be independent of the two available indexing choices. These results indicated the possibility of twinning. The intensity statistics, the $N(Z)/Z$ plot (Fig. 3) and the ratio $\langle I^2\rangle/\langle I\rangle^2$ (Table 2) for the acentric reflections reduced in $P6$ were then inspected. All these indicators supported the hypothesis that the crystals could indeed be partially merohedrally twinned. Twinning by merohedry is a crystal-growth disorder and occurs when the differently oriented domains of which the specimen is composed have symmetry lower than the lattice symmetry (Catti & Ferraris, 1976). Therefore, we submitted the data (in $P6_5$) to the twinning server (Yeates, 1997). In all three cases the twinning analysis showed a twin fraction α of 27.8% for the low-resolution data (calculated in the range 20–2.7 Å), 39.6% for higher resolution data and 46.2% for the derivative data collected at

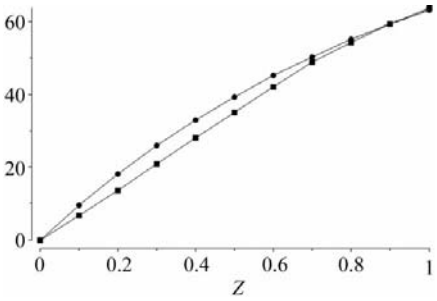

Figure 3
Graphical representation of the cumulative distribution function $N(Z)$ relative to Z (Z represents the intensity relative to the mean intensity) calculated with *TRUNCATE* (Collaborative Computational Program, Number 4, 1994). The square-dotted line is relative to our experimental distribution and shows a slightly sigmoid profile lying below the theoretical values (circles).

Table 1
Data collection.

Values in parentheses correspond to the outermost resolution shell.

	ApTpApTpApT		ApTpAp^{Br}UpApT
λ (Å)	1.5418	0.9076	0.8266 (remote)
Unit-cell parameters (Å), space group $P6_5$	$a = b = 32.43$, $c = 119.91$	$a = b = 32.31$, $c = 117.77$	$a = b = 32.41$, $c = 120.7$
Rotation per frame (°)	1	0.9	1
Total oscillation (°)	101	87.3	101
Resolution range (Å)	20.0–2.64	20.0–2.17	20.0–2.50
Unique reflections	2116	3630	2516
Completeness (%) [reflections with $I/\sigma(I) \geq 2$]	92.6 (74.8)	87.6 (61.9)	100 (68.6)
Overall redundancy†	13.5	11.7	11.4
$I/\sigma(I)$	36.2 (4.87)	30.3 (3.08)	29.5 (4.60)
R_{sym}‡ (%)	0.044 (0.391)	0.047 (0.456)	0.057 (0.357)
R_{sym} in $P6_522$ (%)	0.059 (0.456)	0.075 (0.521)	0.063 (0.378)

† Total reflections registered divided by the number of unique reflections. ‡ $R_{sym}(I) = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, calculated for the whole data set.

Table 2
Twin analysis (λ = 0.9076 Å).

Analysis was carried out with *CNS*. The expected values for untwinned data are 2 for $\langle|I|^2\rangle/\langle|I|\rangle^2$ and 0.785 for $\langle|F|^2\rangle/\langle|F|\rangle^2$ and are 1.5 and 0.865, respectively, for perfect twinning.

Resolution range (Å)	No. of reflections	$\langle\|I\|^2\rangle/\langle\|I\|\rangle^2$	$\langle\|F\|^2\rangle/\langle\|F\|\rangle^2$
4.14–20.0	488	1.847	0.803
3.29–4.14	507	2.354†	0.766
2.88–3.29	505	4.469†	0.703
2.61–2.88	512	1.802	0.835
2.43–2.61	493	1.829	0.830
2.28–2.43	494	1.871	0.836
2.17–2.28	455	1.584	0.865

† These high values are probably a consequence of the anisotropy and the presence of the stacking reflections.

Table 3
Refinement statistics (λ = 0.9076 Å).

Resolution range (Å)	15.0–2.17
DNA atoms	480
Water molecules	9
R_{work}† (%)	22.90
No. reflections, work set ($F > 0$)	3162
R_{free}‡ (%)	25.92
No. reflections, free set ($F > 0$)	313
R.m.s.d. bond lengths (Å)	0.0089
R.m.s.d bond angles (°)	1.54

† R factor $= \sum_{hkl} ||F_o(hkl)| - k|F_c(hkl)|| / \sum_{hkl} |F_o(hkl)|$. ‡ R factor of reflections used for cross validation in the refinement.

the remote point (both calculated in the resolution range 20–2.5 Å). This latter α value discouraged us from pursuing phasing *via* heavy-atom location.

5. Molecular replacement and space-group assignment

Once the structure of d(ApTpApTpApT) in the space group $P2_1$ was solved showing the presence of Hoogsteen base pairs (Abrescia *et al.*, 2002), it became clear that the standard B-DNA models previously employed were not suitable for molecular replacement in the hexagonal case. Thus, using this new DNA as a search model form and with the information of possible twinning, we repeated the phasing *via* molecular replacement in the space group $P6_5$. Three peaks stood out employing the high-resolution data without detwinning, the first of them with a 75.1% correlation in amplitudes, an R_{factor} of 43.8% and a 70% correlation in intensities. Analysis in the $P6_1$ space group gave significantly lower values.

We also investigated the space group $P6_522$. We found a peak with a 71.3% correlation in amplitudes (R_{factor} = 47.0%; 67.6% correlation in intensity) using as a model a pentamer duplex built from the $P2_1$ asymmetric unit. However, in order to dissipate our doubts on a space-group misinterpretation or a fake twinning arising from the helical pseudosymmetry, we analysed all the solutions in the $P6_522$ space group either built manually (by a translation and rotation of a single strand of the $P6_5$ solution) or as obtained from the molecular-replacement protocol. We could not refine any of these structures.

Before starting the refinement of the solution obtained in $P6_5$, the structure factors of the search model and its self-rotation function were calculated using the *CNS* program (Brünger *et al.*, 1998). The self-rotation showed additional peaks indicating the presence of dyads (Fig. 2*b*), as had been found in the experimental data (Fig. 2*a*) and as was expected from the intrinsic NCS present in DNA duplexes. To check if the pseudosymmetry of the oligonucleotide gave rise to an apparent twinning, we sent the calculated structure factors to the twin server (Yeates, 1997). The resulting twin fraction was under 3% (20–2.5 Å resolution range). It is therefore clear that molecular pseudosymmetry does not give rise to an apparent twinning.

6. Structure refinement

The $P6_5$ solution was refined with *CNS* version 1.1 using input files for hemihedral twinning. The α fraction for the high-resolution data was recalculated in the resolution range 15–2.17 Å and used in the following steps. Apart from taking into account the twinning fraction and the twinning operator, the rest of the refinement process followed the standard protocol. A random free data set was created in order to avoid over-refinement and the least-squares residual target function was employed. The starting model corresponds to the refined structure at 2.5 Å of the same sequence crystallized in $P2_1$ but without the extra-helical bases. The starting twinned R_{work} was 31.3% and R_{free} was 37.7% (15–2.5 Å with bulk-solvent correction). After rigid-body refinement of independent base pairs, R_{work} dropped to 27.2% and R_{free} to 31.2% (15–2.17 Å with bulk-solvent correction). However, owing to the alternating base composition of the sequence, we decided to repeat the rigid-body refinement steps

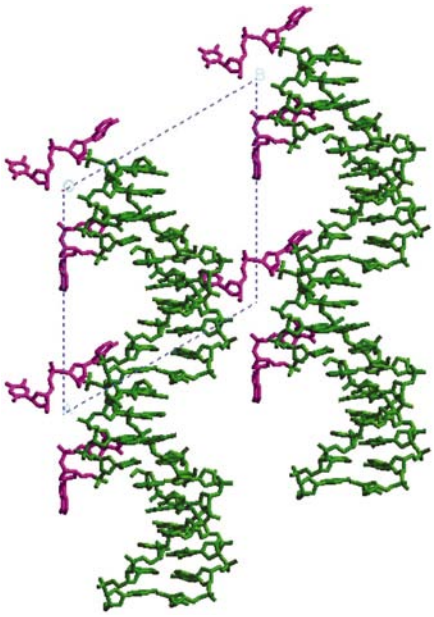

Figure 4
View of one layer of DNA molecules in the crystal. The unit cell and four asymmetric units are shown. Hoogsteen DNA is presented in green. The extra-helical bases which interact with neighbouring duplexes in either the same layer or the neighbouring layer are shown in magenta.

without phosphate groups. We then calculated the $(F_o - F_c)$ electron-density map. By visual inspection of the model and of the positive peaks in the $(F_o - F_c)$ map, we rebuilt the connectivity between the stacked base pairs. We also added the extra-helical bases, although two of them appeared to be very mobile and could not be accurately located. A summary of the current refinement statistics is given in Table 3.

The structure obtained shows the same essential features as the structure previously found in the $P2_1$ space group (Abrescia *et al.*, 2002). The bases are paired according to the Hoogsteen scheme. The duplexes stack as columns of alternating hexamers and tetramers. The tetramers have two overhanging bases at the 3′ end of both strands, which penetrate the minor grooves of neighbouring DNA duplexes. However, in $P2_1$ the DNA columns are all parallel, whereas in $P6_5$ they lie in planes which are rotated by 60°. Details of the molecular conformation, hydration and other interactions will be presented elsewhere.

The location of the extra-helical terminal AT bases supports the fact that the $P6_522$ space group is not correct, since in the latter case both terminal AT pairs should be identical, whereas in practice this is not the case.

7. Discussion

We demonstrate in this paper that the crystals of the d(ATATAT) oligonucleotide in the hexagonal lattice are affected by merohedral twinning and an apparent 622 point group is generated both by the contribution of the twin and by the pseudosymmetry of DNA. In fact, it is always possible to find a pseudo-twofold axis orthogonal to the helical axis of DNA. In particular, when the helix axis is perpendicular to the *c* direction the crystals appear to have a higher crystallographic point group. In other words, the pseudo-twofold axis of our helical structure runs parallel to the expected crystallographic dyad (Fig. 4) and a false 622 point group appears to be present. We have shown that when pseudosymmetry and twinning by merohedry coincide, as in the case presented here, the assignment of the space group is not straightforward. We could only explain the anomalies displayed by the experimental data and solve this structure by molecular replacement when an adequate Hoogsteen DNA model became available.

We are thankful to Drs J. Navaza and I. Fita for useful discussions. This work has been supported by grants from the Generalitat de Catalunya, the Ministerio de Ciencia y Tecnología (Grant PM98-0135) and the European Commission (Marie Curie Program for NGAA and access to the DESY Synchrotron, Hamburg).

References

Abrescia, N. G. A., Thompson, A., Huynh-Dinh, T. & Subirana, J. A. (2002). *Proc. Natl Acad. Sci.* **99**, 2806–2811.
Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C. & Watts, L. (1974). *J. Mol. Biol.* **88**, 523–533.
Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). *Biophys. J.* **63**, 751–759.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Catti, M. & Ferraris, G. (1976). *Acta Cryst.* A**32**, 163–165.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Esnouf, R. (1999). *GroPat* v.2. Unpublished program.
Malinina, L., Makhaldiani, V. V. & Tereshko, V. (1987). *J. Biomol. Struct. Dyn.* **5**, 405–433.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Viswamitra, M. A., Shakked, Z., Jones, P. G., Sheldrick, G. M., Salisbury, S. A. & Kennard, O. (1978). *Nature (London)*, **273**, 687–688.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.