

Structure of the A-DNA Octamer d(GGCATGCC)

CHRISTINE M. NUNN AND STEPHEN NEIDLE

*The CRC Biomolecular Structure Unit, The Institute of Cancer Research,
Sutton, Surrey SM2 5NG, England. E-mail: c.nunn@icr.ac.uk*

(Received 28 August 1996; accepted 27 November 1996)

Abstract

The crystal structure of the self-complementary DNA octamer d(GGCATGCC) has been determined to a resolution of 2.38 Å and *R* factor of 17.1%. The structure is an A-DNA octamer duplex with Watson–Crick base pairing along its length. This sequence extends the investigation of octamer duplexes of the type d(GGX₄CC) which assume *P*6₁ packing within the crystallographic unit cell. The A-DNA duplex is non-linear with a helix axis curvature of 7° and bending towards the DNA minor groove. In view of the biological importance of protein recognition of some A-DNA sequences the minor-groove geometry and helix-axis curvature for this structure is compared with all others of the type d(GGXATXCC).

1. Introduction

The structures of a number of native DNA octamer sequences have been determined by X-ray crystallography. With the exception of three sequences which contain alternating purine and pyrimidine bases and are consequently of Z-DNA type, all of these sequences crystallize as A-DNA duplexes. The vast majority of these A-DNA duplexes exist in two crystalline forms, in either hexagonal *P*6₁ or tetragonal *P*4₃2₁2 space groups. Whilst some sequences such as d(GGGCGCCC) exhibit both *P*6₁ and *P*4₃2₁2 crystal packing (Shakke, Guenstein-Guzikevich, Eisenstein, Frolow & Rabinovich, 1989), all of the known *P*6₁ structures are of sequences of the type d(GGX₄CC) with the four central bases comprising C, G, A, T, U and I.

Structural studies of protein–DNA complexes display a wide range of DNA conformations. For the structure of a TATA box-binding protein (TBP) complexed with a 14-base-pair oligonucleotide (Kim, Geiger, Hahn & Sigler, 1993; Kim, Nikolov & Burley, 1993; Kim & Burley, 1994) and the more recent structures of the transcription factor IIa with TBP and DNA (Tan, Hunziker, Sargent & Richmond, 1996; Geiger, Hahn, Lee & Sigler, 1996) the conformation of the DNA in these crystal structures has been found to be severely distorted from that of B-DNA type. In these structures the minor groove of the DNA has become wide and underwound to provide a large surface for interaction with the TBP assuming a conformation somewhat akin to features that one

observes in A-DNA. Analysis of the conformation in the TATA box (Guzikevich-Guerstein & Shakke, 1996) has shown that the conformation differs from canonical A-DNA by solely a rotation around each glycosidic bond along the DNA backbone. The structure of the TATA box containing sequence of the type d(GGX₄CC), where X₄ is TATA has been determined (Shakke *et al.*, 1981) and is an A-DNA duplex, with *P*6₁ packing within the unit cell. The close structural relationship shown for the TATA sequence in the native structure and upon protein binding highlights the importance of structural studies for sequences of this type.

In this paper the structure of the octamer sequence d(GGCATGCC) is presented. This sequence is of the type d(GGX₄CC) where X₄ is the alternating pyrimidine–purine sequence CATG. A recent structural study from this laboratory reported the structure of the decamer d(AGGCATGCCT) (Nunn & Neidle, 1996) in which the central eight bases form an A-DNA octamer duplex.

2. Materials and methods

2.1. Crystallization and data collection

The DNA octamer d(GGCATGCC) was purchased from the Oswel DNA Service (University of Southampton, Southampton, UK). Crystals were grown by vapour diffusion at both 277 and 287 K from sitting drops containing 5 µl 3 mM DNA in 30 mM sodium cacodylate (pH 7.0), 2 µl 35%(v/v) NH₃ solution, 2 µl 10 mM hexaminecobalt(III) trichloride and 2 µl 40%(v/v) 2-methyl-2,4-pentanediol, equilibrated against a 1 ml reservoir of 40%(v/v) 2-methyl-2,4-pentanediol. Pale brown crystals grew with hexagonal cross-section within a few days, growing to an optimal size over a period of 7–10 d. A hexagonal prism grown at 287 K and measuring 0.4 × 0.4 × 0.4 mm was selected for data collection. Data collection was carried out using a Siemens–Xentronic multiwire area detector equipped with a rotating-anode X-ray generator (70 mA, 40 kV) and graphite monochromator. A crystal-to-detector distance of 10 cm and swing angle of 15° was used. Data were collected with χ set at 45°, while the crystal was rotated through 100° in ω . 180 second frames were recorded every 0.2° steps.

Data processing was carried out using the program *XDS* (Kabsch, 1993) and the *CCP4* program suite (Collaborative Computational Project, Number 4, 1994). A

Table 1. *Helical parameters for d(GGCATGCC)*

Calculated using program *CURVES* (Lavery & Sklenar, 1988). Local parameters are shown in bold.

Base pair	Tip (°)	Incline (°)	Buckle (°)	Propeller twist (°)	Base step	Tilt (°)	Roll (°)	Twist (°)	Base-pair rise (Å)
G1-C16	1	4	-7	12	G1-C16				
G2-C15	1	6	6	10	G2-C15	4 4	2 5	27 28	2.8 3.1
C3-G14	1	9	7	-1	C3-G14	1 1	-4 0	32 32	3.4 3.5
A4-T13	1	11	2	-4	A4-T13	-1 -1	2 7	32 31	3.0 3.4
T5-A12	0	12	10	1	T5-A12	2 2	0 7	32 31	2.9 3.3
G6-C11	1	13	0	-4	G6-C11	3 3	4 11	33 32	3.0 3.7
C7-G10	1	11	-5	7	C7-G10	-1 -2	-2 5	31 30	3.2 3.6
C8-G9	2	12	5	5	C8-G9	1 0	4 10	33 32	2.6 3.3

Table 2. *Backbone torsion angles (°) and furanose ring conformations for the octamer nucleotides*

	α	β	γ	δ	ϵ	ζ	χ	P
G1 G9			54 19	96 94	210 205	291 284	190 191	12 20
G2 G10	297 288	176 186	55 53	96 97	180 192	281 291	207 203	44 21
C3 C11	289 295	193 185	46 45	94 88	191 196	286 286	211 205	28 31
A4 A12	294 299	180 183	56 48	93 94	209 204	284 282	201 197	21 25
T5 T13	288 288	175 175	53 59	91 98	201 196	283 283	209 204	32 31
G6 G14	287 284	171 174	56 56	87 92	205 198	284 281	198 196	26 31
C7 C15	295 287	180 179	51 58	91 96	200 200	281 285	194 195	30 26
C8 C16	278 284	182 177	68 67	89 88			201 197	32 40

total of 5252 reflections were collected. Of a possible 2167 unique reflections, 2104 were observed to a resolution limit of 2.38 Å representing a 97% complete data set. The final merged data set has cell parameters of $a = 46.29$, $c = 42.97$ Å and $R_{\text{merge}}(I)$ of 3.9%. A second data set collected at 100 K using the same crystal showed a shrinkage in unit cell, however the resolution limit and crystal quality had deteriorated. This data was not used for the structure refinement.

2.2. Structure refinement

The structure crystallizes in the hexagonal space group $P6_1$ with cell parameters close to those reported

for a number of octameric sequences suggesting an isomorphous structure. These structures contain an octamer duplex within the asymmetric unit and are of A-DNA type. The coordinates for the structure of d(GGTATACC) (Shakked *et al.*, 1981) were used as a starting model for the refinement with the central four bases replaced by those of the correct type.

Crystallographic refinement was carried out using the program *X-PLOR* version 3.1 (Brünger, Kuriyan & Karplus, 1987). Initial rigid-group refinement of the model as firstly one then 16 independent groups was carried out, followed by atomic positional refinement for data in the range 8–3 Å. The data range was extended to include data to the resolution limit of 2.38 Å and in

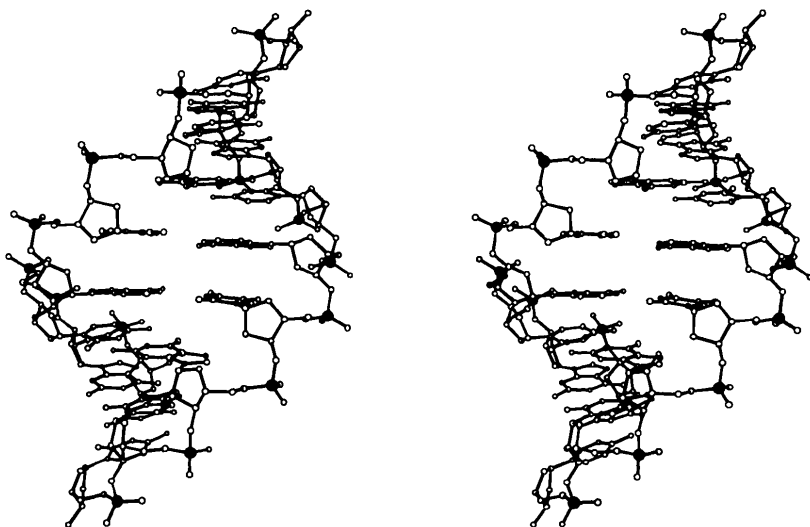


Fig. 1. Stereoview of the d(GGCATGCC) duplex.

addition to positional refinement, atomic temperature-factor refinement was carried out. After positional and temperature-factor refinement of all DNA atoms the R factor was 21.7%. Solvent molecules were located from difference electron-density maps and assigned as water molecules in the refinement. No hexamminecobalt(III) ions were located anywhere in the structure, and are presumably delocalized throughout the unit cell.

The refinement converged after the inclusion of 65 water molecules to give a final R factor of 17.1% for data within the range 8–2.38 Å. Mean values for the thermal parameters (Å²) are 27, 26, 26, and 34 for the phosphate groups, sugars, bases and solvent molecules, respectively. The X -PLOR refinement and R -factor calculations used 2045 data with $F > 2\sigma(F)$. The root-mean-square deviations from target values for bond lengths and angles are 0.009 Å and 1.48°, respectively. The final atomic coordinates have been deposited in the Nucleic Acid Database.*

3. Results and discussion

The structure is an A-DNA octamer duplex (Fig. 1) comprising two self-complementary strands of the octamer within the crystallographic asymmetric unit with Watson–Crick base pairs along the entire length of the duplex. The bases are labelled G1 to C8 in the 5' to 3' direction for the first strand and G9 to C16 for the second. The number of base pairs per turn extrapolates to 11.3, which is typical for A-DNA. The helix axis is non-linear with a total angle of curvature of 7.1°, bending in the central region of the helix towards the DNA minor groove.

The global and local (shown in bold) helical and base-pair parameters for the octamer duplex are shown in Table 1. The terminal GpG step has a small helical twist with both of these G-C base pairs exhibiting large values of propeller twist. The propeller twist for the terminal two base pairs at each end of the duplex are larger than the central four base pairs. Backbone torsion angles for the octamer are similar along the length of the duplex (Table 2) and typical of values seen for A-DNA. Sugar puckers for the nucleotides are C3'-endo for all but one guanosine (G2) and one cytosine (C16) for which the pseudorotation angles are slightly larger and the conformation is C4'-exo. Intrastrand P–P separations vary along the duplex, ranging from 5.7 to 6.3 Å with the largest values occurring at the positions of C4'-exo sugar pucker.

The stacking for each of the the dinucleotide steps for the octamer duplex is the same as that observed previously for A-DNA octamer duplexes of this type

(Lauble, Frank, Bloecker & Heinemann, 1988). The terminal stacking interactions at both ends of the duplex show no stacking interactions between the cytosine bases and very little for the guanine bases. Stepping through the duplex the base-pair stacking shows alternating poor and good overlap, with good overlaps observed for the two GpC steps and the central ApT step – the (pyrimidine)p(purine) steps and poor overlaps for the other four steps. This feature is also seen in the structure of d(AGGCATGCCT) (Nunn & Neidle, 1996). The DNA minor-groove width for d(GGCATGCC) is shown in Fig. 2 along with values for the corresponding sequence within the structure of d(AGGCATGCCT) (Nunn & Neidle, 1996). The minor-groove width increases at the centre of the duplex though not to the degree shown for d(AGGCATGCCT).

The molecular packing of octamer duplexes within this $P6_1$ cell has been observed previously for all other octameric sequences of the type d(GGX₄CC). Presumably this packing arrangement requires two contiguous guanine and cytosine bases at the 5' and 3' end of the sequence, respectively. The terminal base pair of one duplex stacks against the minor groove of a neighbouring duplex within the unit cell, related by either a 6₁ or 2₁ screw axis. There is only one intermolecular hydrogen-bonding contact for d(GGCATGCC) from atom N2 of base G(2) to atom O4' of base G(9) of a symmetry-related sugar phosphate backbone of 2.81 Å.

65 solvent sites were included in the refinement. All four of the terminal hydroxyl groups 1O5', 8O3', 9O5' and 16O3' form hydrogen-bonding interactions with water molecules. Only limited hydration was observed in the DNA minor groove with water molecules directly bound to atom N3 of A(4) and G(10) and atom O2 of

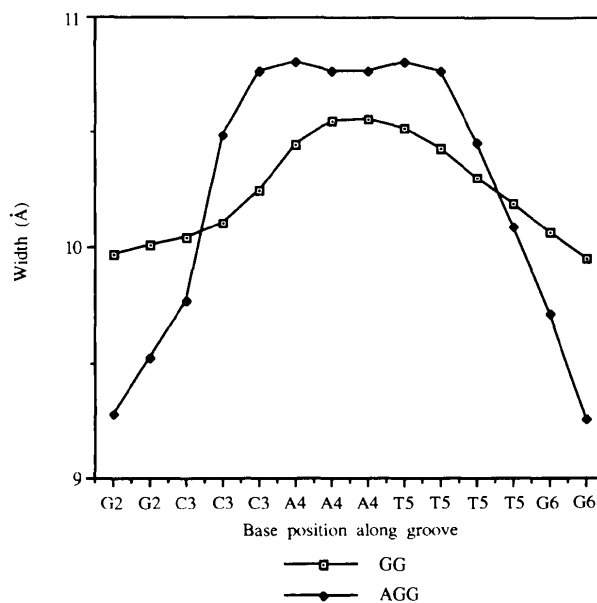


Fig. 2. Minor-groove width for d(GGCATGCC) together with the similar sequence in d(AGGCATGCCT) (Nunn & Neidle, 1996).

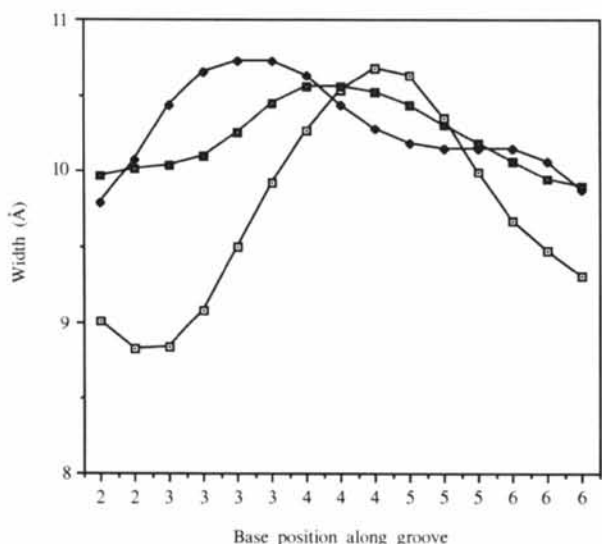
* Atomic coordinates and structure factors have been deposited with the Nucleic Acid Database (Reference: ADH078, RADH078SF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: li0241).

C(8), C(11) and C(16). In the major groove the degree of hydration is greater with 11 solvent molecules directly hydrogen bonding to the DNA bases. The hydration of the phosphate groups is typical of A-DNA (Eisenstein & Shakked, 1995) with 17 water molecules directly bound to phosphate O atoms. As a result of unit-cell packing the terminal base pairs G(1)·C(16) and C(8)·G(9) stack against symmetry-related duplexes. In this region of the

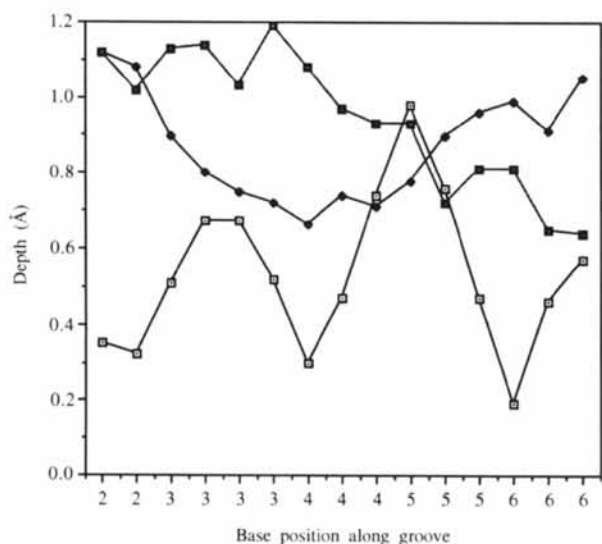
structure there are two solvent interactions between the adjacent minor grooves of the two symmetry-related helices. One water molecule binds to atom O2 of both C(11) and C(16) whilst there is a solvent-mediated interaction between atom O2 of C(8) and atom N3 of A(4).

The crystal structures of two other A-DNA octamer sequences with sequences close to d(GGCATGCC) have been reported. These sequences, d(GGGATCCC) (Laube *et al.*, 1988) and d(GGTATACC) (Shakked *et al.*, 1983) both contain two base-pair changes from the present structure (shown as bold type). The sequence d(GGCATGCC) is close to being an alternating purine-pyrimidine sequence, with just one base-pair change at either end of the duplex compared to the Z-DNA octamer d(CGATGCG) (Fujii *et al.*, 1985). Another sequence which has six similar bases is d(GCATGCT) (Leonard *et al.*, 1995). This sequence exhibits a quite different structure with a novel loop conformation which dimerizes to form a quadruplex structure.

The minor groove of A-DNA is recognised by proteins. It is of interest to examine the geometry of the DNA minor groove within the three similar $P6_1$ A-DNA structures of the type d(GGX₄CC) where X₄ = CATG, TATA and GATC. Because of the non-linear nature of the A-DNA helical axis in these structures the program CURVES (Lavery & Sklenar, 1988) was used to investigate the minor-groove geometry (Stofer & Lavery, 1994). The minor-groove width and depth for the three



(a)



—□— GATC
—●— TATA
—■— CATG

(b)

Fig. 3. Minor-groove width and depth for the structures of d(GGCATGCC), d(GGGATCCC) (Laube *et al.*, 1988) and d(GGTATACC) (Shakked *et al.*, 1983), shown as CATG, GATC and TATA, respectively.

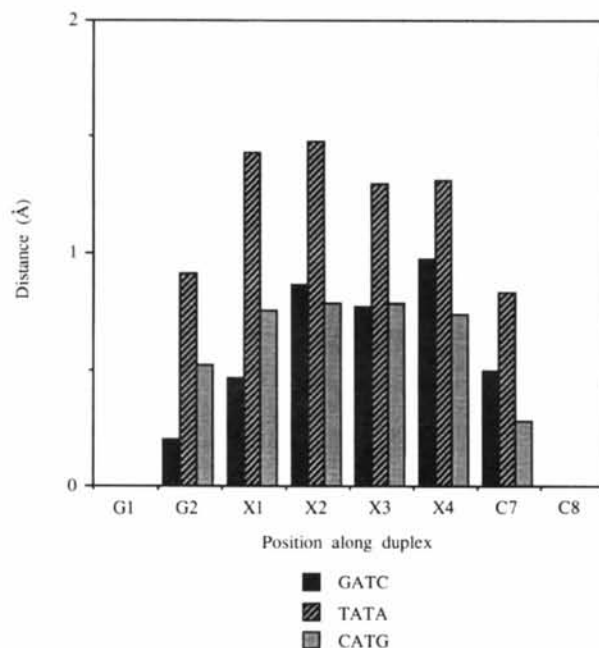


Fig. 4. Bar graph to show the base-pair offset from a straight line joining the terminal two base pairs for d(GGCATGCC), d(GGGATCCC) (Laube *et al.*, 1988) and d(GGTATACC) (Shakked *et al.*, 1983) octamer duplexes, shown as CATG, GATC and TATA, respectively.

structures are shown in Fig. 3. The maximum groove width for all three structures is similar in magnitude, however the position of maximum groove width and variation across the groove differs for each sequence. The minor-groove depth for d(GGGATCCC) is largest at the position of maximum groove width. In the case of d(GGCATGCC) and d(GGTATACC) this does not hold and the minor-groove depth is much greater for these two structures.

In addition to the variation of minor-groove geometry observed for the $P6_1$ octamer duplexes there is also marked variation in the curvature of the helix axis and the displacement of base pairs from the helix axis. Fig. 4 shows the offset of each base pair from a straight line joining the terminal two base pairs for the sequences d(GGCATGCC), d(GGGATCCC) and d(GGTATACC). The TATA sequence, which is known to undergo minor-groove binding with protein molecules and is present in a wide range of promoter sequences shows a large axis curvature with bending of 21° towards the DNA minor groove. The sequence d(GGGATCCC) containing a shorter AT region at the centre of the duplex exhibits much less helix curvature with bending of 11° , close to the value seen here for d(GGCATGCC) of 7° .

These results show that the central four base pairs within these octamer duplexes are conformationally variable within A-DNA structure and do not impose a rigid helical geometry. In relation to protein binding this DNA flexibility enables DNA and protein molecules to potentially interact with maximum interaction.

For the octamer duplex observed in the structure of d(AGGCATGCCT) (Nunn & Neidle, 1996) the helix-axis curvature is 17° towards the minor groove. The increased curvature for this structure is probably a result of crystal packing effects and base-pairing interactions of the terminal A and T bases. Within this structure cobalt hexammine ions bind to the DNA major groove via hydrogen-bonding interactions with GG bases at each end of the duplex in addition to a second binding site with backbone phosphate groups. Bending of DNA as a result of charge neutralization is known to occur

(Crothers, 1994) and could also explain the increase in curvature that we observe in this structure.

References

- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Crothers, D. M. (1994). *Science*, **266**, 1819–1820.
- Eisenstein, E. & Shakked, Z. (1995). *J. Mol. Biol.* **248**, 662–678.
- Fujii, S., Wang, A. H.-J., Quigley, G. J., Westerink, H., van Der Marel, G. A., van Boom, J. H. & Rich, A. (1985). *Biopolymers*, **24**, 243–250.
- Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. (1996). *Science*, **272**, 830–836.
- Guzikevich-Guerstein, G. & Shakked, Z. (1996). *Nature Struct. Biol.* **3**, 32–37.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Kim, J. L. & Burley, S. K. (1994). *Nature Struct. Biol.* **1**, 638–653.
- Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Nature (London)*, **365**, 520–527.
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature (London)*, **365**, 512–520.
- Lauble, H., Frank, R., Bloecker, H. & Heinemann, U. (1988). *Nucleic Acids Res.* **16**, 7799–7816.
- Lavery, R. & Sklenar, H. (1988). *J. Biomol. Struct. Dynam.* **6**, 63–91.
- Leonard, G. A., Zhang, S., Peterson, M. R., Harrop, S. J., Helliwell, J. R., Cruse, W. B. T., d'Estaintot, B. L., Kennard, O., Brown, T. & Hunter, W. N. (1995). *Structure*, **3**, 335–340.
- Nunn, C. M. & Neidle, S. (1996). *J. Mol. Biol.* **256**, 340–351.
- Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolov, F. & Rabinovich, D. (1989). *Nature (London)*, **342**, 456–460.
- Shakked, Z., Rabinovich, D., Cruse, W. B. T., Egert, E., Kennard, O., Sala, G., Salisbury, S. A. & Viswamitra, M. A. (1981). *Proc. R. Soc. London Ser. B*, **213**, 479–487.
- Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A. & Viswamitra, M. A. (1983). *J. Mol. Biol.* **166**, 183–201.
- Stofer, E. & Lavery, R. (1994). *Biopolymers*, **34**, 337–346.
- Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. (1996). *Nature (London)*, **381**, 127–134.