Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of the DNA-binding site of Runt-domain transcription regulators

Malka Kitayner, Haim Rozenberg, Dov Rabinovich and Zippora Shakked*

Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence e-mail: zippi.shakked@weizmann.ac.il

Received 26 September 2004
Accepted 7 December 2004

PDB References: Runt-domain transcription regulator DNA-binding site, 1xjx, r1xjxsf; 1xjy, r1xjysf.

Runt-domain (RD) proteins are transcription factors that play fundamental roles in various developmental pathways. They bind specifically to DNA sequences of the general form PyGPyGGTPy (Py = pyrimidine), through which they regulate transcription of target genes. The DNA duplex T**CTGCGGTC**/T**GACCGCA**G, incorporating the binding site for the RD transcription factors (bold), was crystallized in space group $P4_3$. X-ray analysis of two crystals diffracting to 1.7 and 2.0 Å resolution, which had slight variations in their unit-cell parameters, revealed two distinct conformations of the A-DNA helix. The two crystal structures possessed several structure and hydration features that had previously been observed in A-DNA duplexes. A comparative analysis of the present A-DNA structures and those of previously reported B-DNA crystal structures of RD-binding sites in free and protein-bound states showed the various duplexes to display several common features. Within this series, the present A-DNA duplexes adopt two conformations along the pathway from the canonical A-DNA to the B-DNA forms and the protein-bound helices display conformational features that are intermediate between those of the current A-DNA structures and that of the B-DNA-type helix of the free RD target. Based on these data and energy considerations, it is likely that the propensity of the RD-binding site to adopt the A-DNA or B-DNA conformation in solution depends on the sequence context and environmental conditions, and that the transition from either DNA form to the protein-bound conformation involves a small energy barrier.

1. Introduction

High-resolution crystal structures of DNA-binding proteins in complexes with their DNA-binding sites have shown that in the majority of cases the DNA target is essentially of the B-DNA form with various degrees of bending and deformation (Dickerson, 1998; Garvie & Wolberger, 2001). A detailed analysis of the DNA conformation in such complexes has demonstrated that the association of DNA with the protein can also induce a partial or complete conversion of the B-type helix to the A-form (Lu *et al.*, 2000).

In many cases, the specificity of interaction could not be explained by direct hydrogen-bonding and non-polar contacts between the protein and the bases of its target DNA, but rather by indirect structural effects such as intrinsic structure and flexibility as well as by water-mediated interactions. The structural role or 'indirect readout' of the DNA-binding site in providing specificity to protein–DNA interactions was demonstrated in several systems where data were available on both the free and the protein-bound DNA conformations

Table 1
Crystallographic data and intensity statistics.

Values for the highest resolution shell are given in parentheses.

	TC**TGCGGTC**	
Sequence	I	II
Space group	$P4_1$ or $P4_3$	
Unit-cell parameters (Å)	$a = b = 41.0$, $c = 23.8$	$a = b = 42.3$, $c = 24.3$
Unit-cell volume (Å^3)	40072	43527
No. of independent DNA duplexes	1	
Volume per base pair (Å^3)	1252	1360
Resolution limits (Å)	18.4–1.7 (1.73–1.70)	19.0–2.0 (2.03–2.00)
No. of measured reflections with $I > 0$	41141	48713
No. of unique reflections with $\|F\| > 2\sigma(F)$	4180 (143)	2891 (118)
No. of unique reflections with $I > 0$	4315 (163)	2975 (137)
Completeness of data (%)	95.9 (78.4)	100 (100)
$\langle I/\sigma(I)\rangle$	39.3 (4.9)	52.4 (8.9)
$R_{\rm sym}(I)$ (%)	3.9 (23.6)	4.2 (26.1)

(Shakked *et al.*, 1994; Rozenberg *et al.*, 1998; Hizver *et al.*, 2001). More specifically, these studies demonstrated that the selection of dinucleotide or longer segments with appropriate conformational characteristics, when positioned at correct intervals along the DNA helix, can constitute a structural code for DNA recognition by regulatory proteins. This structural code facilitates the formation of a complementary protein–DNA interface that can be further specified by hydrogen bonds and nonpolar interactions between the protein amino acids and the DNA bases (Rozenberg *et al.*, 1998; Hizver *et al.*, 2001). More recently, the prediction of indirect effects based on global DNA curvature and flexibility has been achieved (Zhang *et al.*, 2004).

The Runt domain–DNA regulatory complex is one of the systems currently being studied by us in an attempt to gain further insight into the structural basis of DNA recognition by gene-regulatory proteins (Bartfeld *et al.*, 2002).

The Runt-domain proteins (also known as RUNX, AML, CBFα and PEBP2α) belong to a small family of transcription factors with important roles in processes ranging from segmentation in *Drosophila* to haematopoiesis and leukaemia in mouse and human (see reviews by Speck & Stacy, 1995; Downing, 1999; Ito, 2004). All members of this family contain a highly conserved region of 128 amino acids designated the 'runt domain' (RD) that conducts two central functions: it directs the binding of the protein to a target DNA sequence, PyGPyGGTPy (Py = pyrimidine), through which it regulates transcription of target genes, as well as mediating protein–protein interaction with an unrelated partner protein CBFβ, which alone does not bind DNA but increases the DNA-binding affinity of RD (Speck & Stacy, 1995; Downing, 1999; Ito, 2004).

Crystal structures of RD complexed with either DNA or CBFβ or with both have provided a detailed mapping of the RD/CBFβ dimerization region and of the RD/DNA interface (Warren *et al.*, 2000; Bravo *et al.*, 2001; Tahirov *et al.*, 2001). The crystal structures of the free Runt domain and one of its consensus DNA targets have been determined by us (Bartfeld *et al.*, 2002). A comparative structural analysis of the protein and DNA components in their free state and the previously reported crystal structures of the corresponding complexes demonstrated that the Runt domain undergoes significant conformational changes upon interacting with CBFβ or with its DNA-binding site and elucidated the mechanism by which CBFβ, which does not contact DNA, enhances DNA binding by the Runt domain (Bartfeld *et al.*, 2002). Similar results were reported on the basis of the ultrahigh-resolution structure of a modified Runt domain (Backstrom *et al.*, 2002). The free DNA-binding site was shown to adopt a B-DNA-type conformation compatible with that required to form a stable protein–DNA interface, thus corroborating the role of the intrinsic DNA conformation in the recognition process (Bartfeld *et al.*, 2002).

Here, we present the crystal structures of a 9-mer DNA fragment incorporating the consensus double-helical RD-binding site, TC**TGCGGTC**/T**GACCGCA**G (binding site in bold). We show that the DNA molecules adopt two conformational variants of the A-DNA-type structure that may be regarded as two intermediate states along the A-to-B transition pathway.

2. Materials and methods

2.1. Oligonucleotide synthesis, purification and crystallization

DNA oligonucleotides were purchased from the Keck Foundation resource laboratory (Yale University). The samples were purified by reverse-phase high-pressure liquid chromatography. The two complementary DNA strands were mixed in a 1:1 molar ratio and annealed.

Crystals of TC**TGCGGTC** and its complementary strand (denoted R9) were obtained at 292 K by the hanging-drop method from 4 µl drops containing 2 µl RD–DNA complex solution (see below) and 2 µl reservoir solution containing 400 m*M* MES pH 6.5, 40 m*M* NaCl, 20 m*M* $MgCl_2$, 40% PEG 200 (100%) equilibrated against 1 ml reservoir solution. The RD–DNA complex solution contained 10 mg ml^{-1} RD–DNA complex (1:1 molar ratio), 50 m*M* HEPES buffer pH 7.5, 150 m*M* NaCl and 20 m*M* dithiothreitol (DTT).

Two tetragonal crystals that differed significantly in their unit-cell parameters (see below) were analyzed and found to contain only the DNA component of the RD complex used for crystallization experiments.

2.2. Data collection, structure solution and refinement

Diffraction data from each crystal of R9 (referred to as I and II), covered with Exxon Paratone oil as cryoprotectant and flash-cooled to 100 K, were measured on a Rigaku R-AXIS IV image-plate detector mounted on a Rigaku RU-H3R rotating-anode generator with Cu $K\alpha$ radiation focused by Osmic confocal mirrors. Data were processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Crystal data and intensity statistics are given in Table 1.

The absence of strong intensity data in the range 3.3–3.4 Å and the high similarity of the present unit-cell parameters

Table 2
Refinement statistics using *SHELX*.

	pCTGCGGTC/pGACCGCAG†	
Sequence	I	II
Space group	$P4_3$	
No. of independent DNA duplexes	1	
Resolution (Å)	15.0–1.7	15.0–2.0
R factor/R_{free}‡ (%)	16.2/25.1	14.1/18.8
No. of non-H DNA atoms	330	
No. of water molecules	102	90
R.m.s. deviations		
Bond length (Å)	0.008	0.007
Bond angle (Å)	0.025	0.024
$\langle B \rangle$ factor of non-H DNA atoms (Å^2)	36.6	38.4
$\langle B \rangle$ factor of water molecules (Å^2)	45.5	46.6

† The model includes the 5′-phosphate groups of residues C2 and G11 but not the terminal 5′-nucleosides (T1 and T10). ‡ The R factor is defined as $\sum ||F_o| - |F_c|| / \sum |F_o|$.

(Table 1) to those of other DNA crystal structures (Eisenstein *et al.*, 1988; Hunter *et al.*, 1989; Fernandez *et al.*, 1997) indicated an octameric A-DNA-type structure in space group $P4_3$ with one duplex in the asymmetric unit. Accordingly, a search model based on an A-DNA octamer was used to solve the structures using the multi-dimensional/multi-solution search procedure *ULTIMA* (Rabinovich & Shakked, 1984).

The top model in each case was first submitted to a simulated-annealing procedure with the program *CNS*, gradually increasing the resolution using bulk-solvent correction and overall anisotropic temperature-factor optimization (Brünger, 1990). In order to test the validity of each step of the refinement procedure, the free R factor was monitored throughout the process (Kleywegt & Brünger, 1996). The simulated-annealing procedure was applied with a starting temperature of 5000 K, followed by alternated cycles of energy minimization and refinement of isotropic individual temperature factors. Solvent peaks satisfying hydrogen-bonding criteria and adequately represented in $F_o - F_c$ (above 3σ) and $2F_o - F_c$ (above 1σ) electron-density maps were included in the model as O atoms.

The relatively large R_{free} value (more than 12% above the R factor) and the electron-density shapes of several base pairs outside the central C–G doublet indicated orientational disorder or crystal twinning. This was also supported by the intensity statistics, which corresponded to space group $P4_32_12$. Such crystal symmetry would require a symmetric duplex occupying a twofold axis or an asymmetric unit comprising a single DNA strand. The two strands of the R9 duplex are different except for the central base-pair doublet (C5–G6 in the first strand and C14–G15 in the second strand, as shown in Fig. 1). The use of orientational disorder in the refinement with two R9 duplexes related by a twofold axis did not lead to a significant improvement of the model.

Hence, the refinement was continued with *SHELX*97 (Sheldrick, 1997; Sheldrick & Schneider, 1997) with the option TWIN. The refinement was based on F^2 including all measured data to 1.7 and 2 Å for I and II, respectively. The solvent structure was rebuilt in several rounds using the automated water-location option with restraints on the water geometrical arrangement. Between rounds, the density maps were examined and the model was manually corrected using *O* (Jones *et al.*, 1991). The phosphate groups of residues 2 and 11 were modelled into the electron-density map, whereas the nucleosides at the 5′-ends (T1 and T10) were disordered and not included in the refinement. Refinement statistics are given in Table 2.

3. Results and discussion

3.1. Molecular packing and hydration

The overall packing of I and II in space group $P4_3$ is similar to that observed previously for symmetric octamers crystallized in the higher symmetry space group $P4_32_12$ (*e.g.* Haran *et al.*, 1987; Eisenstein *et al.*, 1988). The molecules pack in the crystal by the commonly observed motif in all crystalline A-DNA duplexes in which the terminal base pair of one duplex stacks against the shallow minor groove of a neighbouring molecule related by a twofold, fourfold or sixfold screw axis, depending on the specific space group (Eisenstein & Shakked, 1995; Tippin & Sundaralingam, 1997). In the tetragonal packing, each duplex interacts with four neighbouring molecules related by two 4_3 axes (see Fig. 2*a*).

In the current structures, the molecular packing is stabilized by four direct minor-groove to end-base-pair interactions shown in Figs. 2(*b*) and 2(*c*) for I and II, respectively. The N2 amino group of guanine G6 at the centre of the minor groove is engaged in hydrogen-bonding contact with the O4′ sugar-ring atom of guanine G11 (shown in cyan in Figs. 2*b* and 2*c*). The equivalent interaction formed by a symmetry-related molecule (shown in grey) is between the N2 amino group of guanine G15 and the O4′ sugar-ring atom of cytosine C2. These interactions were observed previously in A-DNA structures employing fourfold screw axes (Eisenstein & Shakked, 1995). The other two interactions involve N2 of guanine G7 and N3 of guanine G11 (cyan) at one side of the minor groove, and N2 of guanine G4 and O2 of cytosine C2 (grey) at the other side of the minor groove of the central duplex. Similar interactions have been demonstrated previously in the crystal structures of GGCCGGCC, GCCCGGGC and ^{I}CCGG/^{I}CCGG duplexes, as illustrated by Eisenstein & Shakked (1995).

The molecular packing of I and II is also stabilized by water-mediated interactions involving the minor-groove and backbone O atoms. A prominent feature of the minor-groove hydration is the chain of water molecules across the central C–G doublet that links the minor groove of one duplex to the

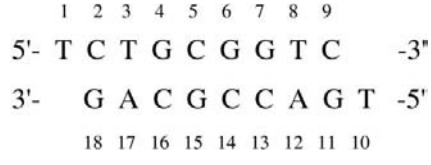
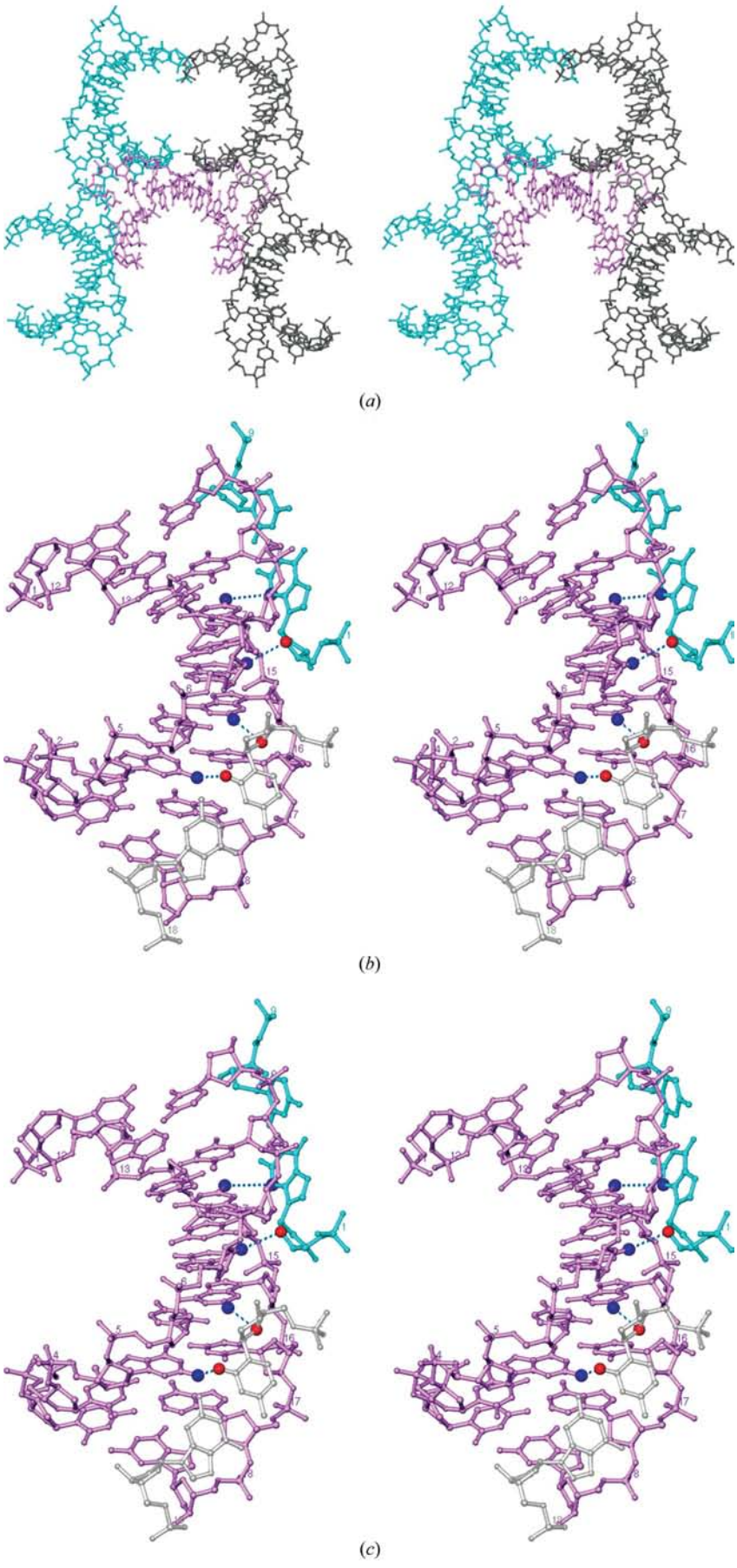


Figure 1
Numbering scheme for the R9 duplex.



backbone edge of its neighbour (Fig. 3). The duplex is positioned on a pseudo-twofold axis resembling that of the other A-DNA structures that occupies a crystallographic twofold axis (Eisenstein & Shakked, 1995). This finding is likely to be a consequence of the symmetry of the central base-pair doublet, C–G/C–G, enabling the formation of analogous hydrogen bonds with adjacent molecules. The four water molecules at the centre, two of which link the N-3 acceptors of the middle guanine bases, are highly conserved in all tetragonal structures (Eisenstein & Shakked, 1995). The bases of the central C–G step are more hydrated in the major groove, where the first and second shells of hydration form intrastrand and interstrand chains lining the deep major groove (Fig. 3).

This hydration pattern has been considered to be a characteristic feature of the C–G step in the tetragonal structures, contributing to the stabilization of its particular geometry (Shakked *et al.*, 1989, 1990). In the present structures, the geometry of the four central water molecules at the minor groove and those comprising the first hydration shell at the major groove are similar to the previously studied structures. However, as expected the arrangement of these water molecules is not perfectly symmetric as in the previous structures that incorporate crystallographic twofold axes. In addition to the non-symmetrical environment, such differences can be also influenced by the presence of the unpaired T residues at the 5′-ends in the two R9 structures.

3.2. DNA conformation and comparisons with other structures

3.2.1. Global and local conformations. Several views of I and II are shown in Figs. 4(*a*), 4(*b*) and 4(*c*). The two R9 structures differ in several global and local parameters as discussed below. The r.m.s. deviation between I and II using all atoms is 1.0 Å. The corresponding values between canonical A-DNA

Figure 2
(*a*) A stereoview of the five duplexes related by 4_3 axes. The molecule chosen for the asymmetric unit is shown in magenta, the molecules which interact *via* the 4_3 axis at (0, 0, 0) are shown in cyan and those *via* the 4_3 axis at (1/2, 1/2, 0) are in grey. Based on the crystal structure of II. (*b*) and (*c*) Stereoviews of minor-groove to end-base-pair interactions that occur between one duplex (magenta) and two symmetry-related DNA duplexes (cyan and grey) in each of the crystal structures (I and II).

Table 3
Average helix parameters.

Parameters in parentheses are averaged values excluding the terminal base pairs. All parameters were calculated with *FREEHELIX* (Dickerson, 1998).

DNA sequence†	Crystal symmetry	Volume per base pair (Å^3)	Inclination (°)	Roll (°)	Helix twist (°)	Rise (Å)	Slide (Å)	D_x (Å)
(1) TCTGCGGTC/TGACCGCAG (I)	$P4_3$	1252	12.8 (12.3)	8.3 (7.9)	32.4 (31.8)	3.3 (3.2)	−1.4 (−1.4)	−3.5 (−3.5)
(2) TCTGCGGTC/TGACCGCAG (II)	$P4_3$	1360	8.1 (7.1)	6.6 (6.4)	31.2 (31.2)	3.3 (3.2)	−1.4 (−1.5)	−3.5 (−3.5)
(3) GGGCGCCC (I)	$P4_32_12$	1403	7.6 (6.9)	7.8 (6.1)	31.5 (32.2)	3.2 (3.2)	−1.6 (−1.7)	−3.6 (−3.6)
(4) GGGCGCCC (II)	$P6_1$	1645	13.8 (14.8)	6.7 (7.2)	31.2 (31.6)	3.3 (3.3)	−1.3 (−1.2)	−3.5 (−3.5)
(5) CCGGGCCCGG (I)	$P2_12_12_1$	1059	18.2 (20.5)	5.4 (7.9)	31.1 (31.4)	3.4 (3.3)	−1.6 (−1.6)	−4.6 (−4.3)
(6) CCGGGCCCGG (II)	$P2_12_12_1$	1345	16.0 (17.4)	5.8 (7.6)	30.2 (29.6)	3.3 (3.3)	−1.8 (−1.8)	−5.1 (−5.0)
(7) CCGGGCCm^{5}CGG (I)	$P2_12_12_1$	1048	18.2 (20.8)	5.4 (7.7)	31.3 (31.1)	3.4 (3.3)	−1.6 (−1.5)	−4.6 (−4.3)
(8) CCGGGCCm^{5}CGG (II)	$P2_12_12_1$	1325	16.2 (17.5)	6.2 (7.4)	30.4 (29.5)	3.3 (3.3)	−1.8 (−1.8)	−5.1 (−5.0)
(9) CCGGGCCm^{5}CGG (III)	$P6_1$	2005	10.7 (10.6)	5.0 (5.2)	30.4 (30.3)	3.3 (3.3)	−1.8 (−1.8)	−4.4 (−4.4)
(10) Cm^{5}CGGGCCm^{5}CGG (I)	$P2_12_12_1$	1296	16.6 (18.1)	6.2 (7.3)	30.3 (29.8)	3.3 (3.3)	−1.8 (−1.8)	−5.2 (−5.0)
(11) Cm^{5}CGGGCCm^{5}CGG (II)	$P6_1$	2069	11.1 (10.6)	5.6 (6.0)	30.3 (30.3)	3.3 (3.3)	−1.7 (−1.7)	−4.0 (−4.0)
(12) CCGGG(Br)^{5}CCCGG (I)	$P2_12_12_1$	1330	16.3 (17.6)	5.6 (7.2)	30.2 (29.5)	3.3 (3.3)	−1.9 (−1.9)	−5.1 (−5.0)
(13) CCGGGCC(Br)^{5}CGG (II)	$P6_1$	1980	10.0 (9.3)	5.0 (5.3)	30.3 (29.8)	3.3 (3.3)	−1.8 (−1.8)	−4.3 (−4.3)
(14) CATGGGCCCATG	$P4_12_12$	1302	5.6 (5.5)	3.6 (2.5)	31.0 (31.8)	3.4 (3.3)	−1.6 (−1.6)	−3.1 (−2.9)
(15) E-DNA (GGCGm^{5}CC)	$P4_32_12$	1953	0.4 (0.3)	1.9 (1.5)	28.3 (29.2)	3.5 (3.4)	−2.1 (−2.1)	−4.1 (−4.0)
A-DNA			21.8	12.0	30.3	3.3	−1.4	−4.4
B-DNA			2.9	1.5	36.0	3.4	0.5	0.6

† Sequences (1) and (2) are from the present study, (3) from Eisenstein & Shakked (1995), (4) from Shakked *et al.* (1989), sequences (5) to (13) are from Tippin & Sundaralingam (1997), (14) from Ng & Dickerson (2002) and (15) from Vargason *et al.* (2000). A-DNA and B-DNA were created from fibre-derived coordinates (Arnott *et al.*, 1983; Chandrasekaran & Arnott, 1996). m^{5}C, 5-methylcytosine; (Br)^{5}C, 5-bromocytosine.

and each of the current octamers are 1.3 and 1.9 Å for I and II, respectively. Superposition of I and II is shown in Fig. 4(*d*). Based on the global axes of the two helices, it appears that helix I is slightly bent towards the major groove whereas helix II is essentially straight. The average helical parameters of the two duplexes along with the corresponding parameters for previously reported A-DNA crystal structures and standard fibre-based A-DNA and B-DNA structures are given in Table 3. Also included in Table 3 is the crystal structure of a hexamer displaying a new helical variant termed E-DNA for eccentric DNA (Vargason *et al.*, 2000). The definition and nomenclature of the various helical parameters are as reported previously (Dickerson *et al.*, 1989; Dickerson, 1998; Olson *et al.*, 2001). Table 3 includes two global parameters, inclination and *X*-displacement (D_x), which were calculated with respect to a global axis and found to be most discriminating in the comparison of right-handed helical structures (Shakked & Rabinovich, 1986). Inclination measures the rotation of the base-pair long axis with respect to a plane perpendicular to the global helix axis. *X*-displacement (D_x) is the perpendicular distance from the C6–C8 vector of a base pair to the global helix axis and is indicative of the major-groove depth. All other parameters were calculated with respect to a local axial system. Roll measures the closing of adjacent base pairs towards the major groove (positive roll) or the minor groove (negative roll). Helix twist is the relative rotation of two successive base pairs about the helix axis. Rise is the relative translation of two successive base pairs parallel to the helix axis. Slide is the relative displacement of the base pairs along their long axes. Selected groove and backbone parameters are given in Table 4. The Z_P parameter was found recently to be the most appropriate to distinguish between A-DNA-type and B-DNA-type base-pair steps (Lu *et al.*, 2000). It is defined as the mean *z* coordinate of the backbone P atoms of the base pair with respect to the base-pair dimer reference frame. The values of Z_P are greater than 1.5 Å for A-DNA steps and less than 0.5 Å for B-DNA steps.

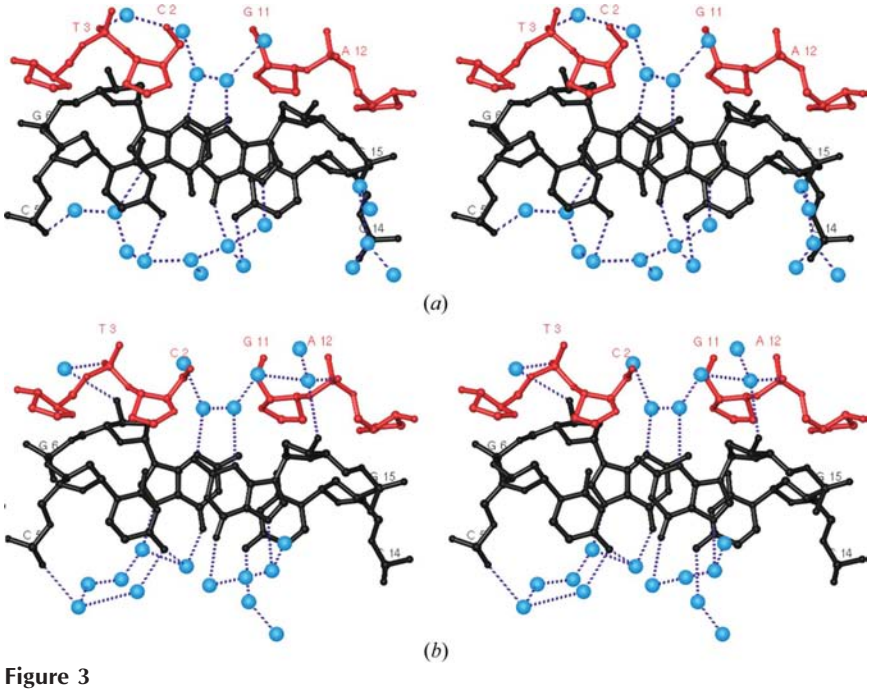


Figure 3
Stereoviews of the hydration of C–G steps in I (*a*) and II (*b*).

Table 4
Groove dimensions and backbone geometry.

Parameters in parentheses are averaged values excluding the terminal base pairs. All parameters were calculated with *FREEHELIX* (Dickerson, 1998), except for Z_P, which was calculated using *3DNA* (Lu *et al.*, 2000).

	Groove width† (Å)		Groove depth‡ (Å)				
DNA sequence§	Major	Minor	Major	Minor	Helix radius¶ (Å)	Intrastrand P—P distances (Å)	Z_P (Å)
(1) CTGCGGTC/GACCGCAG (I)	6.3	9.6	12.5	4.6	9.4	6.1	2.2 (2.2)
(2) CTGCGGTC/GACCGCAG (II)	9.2	9.5	12.8	4.8	9.6	6.2	2.2 (2.3)
(3) GGGCGCCC (I)	8.8	9.5	12.9	4.7	9.6	6.2	2.6 (2.8)
(4) GGGCGCCC (II)	5.0	10.3	12.3	4.3	9.1	5.9	2.3 (2.2)
(5) CCGGGCCCGG (I)	2.0	10.0	13.6	3.5	9.9	6.1	2.3 (2.2)
(6) CCGGGCCCGG (II)	5.3	9.9	14.7	3.5	9.4	6.0	2.4 (2.4)
(7) CCGGGCCm^{5}CGG (I)	2.0	9.9	13.7	3.5	9.4	6.1	2.2 (2.2)
(8) CCGGGCCm^{5}CGG (II)	5.2	9.8	14.7	3.5	9.9	6.0	2.5 (2.4)
(9) CCGGGCCm^{5}CGG (III)	8.5	9.7	14.1	4.3	10.0	6.1	2.2 (2.1)
(10) Cm^{5}CGGGCCm^{5}CGG (I)	5.0	9.9	14.8	3.4	9.9	6.0	2.4 (2.4)
(11) Cm^{5}CGGGCCm^{5}CGG (II)	8.2	9.8	13.5	4.4	9.8	6.1	2.2 (2.0)
(12) CCGGG(Br)^{5}CCCGG (I)	4.8	9.8	14.8	3.5	9.9	6.0	2.4 (2.3)
(13) CCGGGCC(Br)^{5}CGG (II)	9.2	9.5	13.9	4.3	9.9	6.1	2.3 (2.2)
(14) CATGGGCCCATG	8.8	9.5	12.0	4.7	9.2	6.1	2.5 (2.5)
(15) E-DNA (GGCGm^{5}CC)	—	8.9	13.9	4.7	10.1	6.1	2.4 (2.5)
A-DNA	3.7	11.0	12.7	2.9	8.6	5.5	2.6
B-DNA	11.4	5.9	8.4	8.5	9.2	6.7	−0.4

† The width of the grooves is estimated as the shortest P—P vectors across the groove less 5.8 Å (the van der Waals radius of phosphate group). The major-groove width of E-DNA cannot be estimated on the basis of six base pairs. ‡ The depth of the grooves is calculated from the *X*-displacement values (D_x) and the helix radius (R): depth (minor) = $R + D_x - 1.3$ and depth (major) = $R - D_x - 0.3$ (Heinemann *et al.*, 1992). § Sequences (1) and (2) are from the present study, (3) from Eisenstein & Shakked (1995), (4) from Shakked *et al.* (1989), sequences (5) to (13) are from Tippin & Sundaralingam (1997), (14) from Ng & Dickerson (2002) and (15) from Vargason *et al.* (2000). A-DNA and B-DNA were created from fibre-derived coordinates (Arnott *et al.*, 1983; Chandrasekaran & Arnott, 1996). m^{5}C, 5-methylcytosine; (Br)^{5}C, 5-bromocytosine. ¶ Average distance of the phosphate groups from the helix axis.

The average helical parameters exhibited by the R9 structures show that the structures belong to the A-DNA family and are similar to other A-type crystal structures characterized by helix-twist values ranging from 30.2 to 31.5°, a deep major groove and a shallow minor groove as well as Z_P values typical of C3′-*endo* sugar pucker. However, like CATGGGCCCATG (Ng & Dickerson, 2002), the present structures show some features that would classify them as intermediate conformational states between A-DNA and B-DNA. The E-DNA structure is the most underwound helix (average helix twist of 28°) with the lowest average values of slide (−2.1 Å) and roll (1.9°) of the crystal structures in the table and thus deviates largely from either A-DNA or B-DNA conformations.

Selected local parameters of I and II at the base-pair and base-pair step level are compared in Figs. 5 and 6. The mean helical twist angles are 32.4 and 31.2°, similar to those of the other A-DNA structures in Table 3, with individual twist values showing similar fluctuations in a zigzag manner, where the minimum values are at the central C–G steps. The roll-angle patterns show significant differences between the two, averaging at 8.3 and 6.6° for I and II, respectively (Fig. 5*b*). The tilt-angle variations are also dissimilar for the two structures, with large fluctuations in I and negligible ones in II (Fig. 5*c*). In contrast to roll and tilt, the slide patterns are similar averaging at −1.4 Å for both structures (Fig. 5*d*). The local parameters at the base-pair level, propeller twist and buckle, are also different and exhibit large variations along each helix (Fig. 6).

The groove dimensions of the R9 helices as well as other global parameters are similar to those of GGGCGCCC (I) and CATGGGCCCATG (Table 4). The latter helix has been proposed to be intermediate between A-DNA and B-DNA (Ng & Dickerson, 2002). The four structures display the extreme values of several parameters, including the depth of the minor groove, inclination (except for I where the larger inclination is affected by slight bending) and *X*-displacement which are farthest from those of classical A-DNA, thus reflecting an intermediate conformation between A-DNA and B-DNA. In contrast, E-DNA is characterized by a significantly deeper major groove, the largest slide in the series, the narrowest minor groove of the A-DNA structures and inclination and roll angle close to zero as in B-DNA, whereas other parameters are similar to those of the A-DNA structures. Hence, E-DNA represents another type of A–B intermediate helix.

3.3. Comparison between the free and protein-bound binding site

The previously reported structure of a DNA oligomer incorporating the RD-binding site, AG**CTGCGGTC**AT, is of a B-DNA type (Bartfeld *et al.*, 2002), whereas the structures of R9 display A-DNA-type helices. The DNA dodecamer has the same central eight-base-pair sequence as that of R9 (**CTGCGGTC**). The B-form is assumed to be dominant in aqueous solution (Tolstorukov *et al.*, 2001). Depending on the change in the free energy of transition, ΔG_{BA}, some DNA

dimers or trimers can undergo a transition from the B-form to the A-form more easily than others, as a result of changes in water activity or upon protein binding. According to Tolstorukov and coworkers, an average value of ΔG_{BA} = <2.5 kJ mol^{-1} indicates an A-philic sequence and a value >3.3 kJ mol^{-1} indicates a B-philic one. The calculated values for the octamer and dodecamer based on their model are 2.55 and 2.76 kJ mol^{-1}. The corresponding numbers according to the model of Ivanov & Minchenkova (1994) are 2.34 and 2.51 kJ mol^{-1}, respectively. Hence, the octamer is more A-philic than the dodecamer on the basis of such dimeric models (Table 5). According to the dimeric model of Tolstorukov *et al.* (2001), the additional dimers in the dodecamer, AG, AT and CA, are B-philic (with high ΔG_{BA}), whereas GC is A-philic. The average ΔG_{BA} of these dimers, 3.10 kJ mol^{-1}, is larger than that of the whole dodecamer. According to the dimeric model of Ivanov and Minchenkova, AG and AT are A-philic, whereas CA and GC are B-philic and the average ΔG_{BA} of the four dimers, 2.80 kJ mol^{-1}, is again larger than that of the whole dodecamer, yet significantly smaller than the expected value for a B-philic sequence. By considering the sequence of R9 as A-philic, the B-type conformation of the dodecamer may be attributed to the additional transition energy barrier of the end base-pair steps according to the dimeric model of Tolstorukov *et al.* (2001). The B-type structure of the dodecamer cannot be accounted for by the trimeric model of Tolstorukov *et al.* (2001), since according to this model AGC is an intermediate between A and B and CAT is A-philic. As discussed below, the A- and B-type conformations of the various RD targets display several common features, which may explain the ambiguous results of the various prediction models.

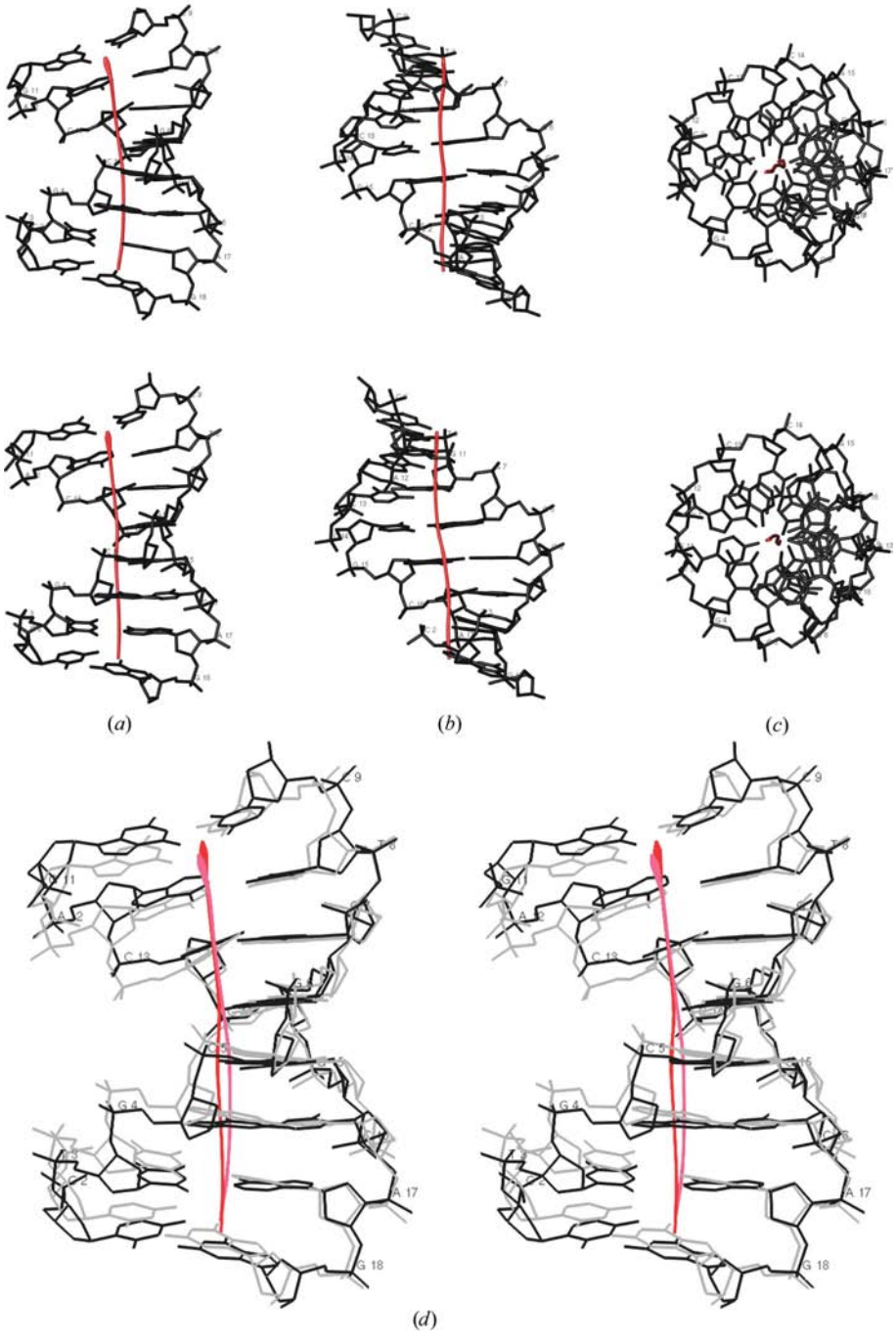


Figure 4
Different views of I (top) and II (bottom): (*a*) perpendicular to the helix axis and to the pseudo-twofold axis, (*b*) down the major groove, (*c*) along the helix axis. (*d*) Stereoview of superposition of I (grey and pink) and II (black and red). View as in (*a*). The global helix axis was calculated using *CURVES* (Lavery & Zakrzewska, 1999).

The average helical parameters and groove dimensions of the various RD DNA structures in the free and protein-bound states are given in Table 6. The RD-binding site used here for the free DNA crystal structure of R9 is TGCGGTC incorporated in TC**TGCGGTC** (bold). The structures of R9 (I and II) adopt A-type helices, as shown by the average helical parameters (Table 6) and discussed above in relation to other A-DNA structures. The same RD-binding site was incorporated in the dodecamer AGC**TGCGGTC**AT, which crystallized as B-DNA (Bartfeld *et al.*, 2002). The binding sites used for the crystal structures of the RD–DNA and RD–DNA–CBFβ complexes were TGTGGTT and TGCGGTT incorporated in 16- and 10-mers, respectively (Table 6). The average helical parameters of the three structures demonstrate that they adopt a B-DNA

conformational variant observed previously for the DNA-binding sites of the E2 transcription regulator of bovine papillomavirus (Rozenberg *et al.*, 1998). This form, referred to as WB-DNA (for writhed B-DNA), is characterized by a gentle roll-induced writhe, so that the base pairs are positively inclined to the helix axis by 4–7° and the mean local roll angle is in the range 3–5°, a helical repeat greater than 10 bp per turn, a deep major groove and a wide minor groove in comparison with regular B-DNA. The R9 structures, II in particular, display average inclination and roll-angle values closer to those of WB-DNA. However, the values of other parameters are in between those of canonical A-DNA and WB-DNA (Table 6).

The major- and minor-groove dimensions of I and II are also midway between A-DNA and WB-DNA, as shown by the corresponding values in Table 6. The variations in the depth of the major groove along the helix can be described by the distances of the base pairs from the helix axis (*X*-displacement), which is a major criterion for distinguishing between A-DNA and B-DNA configurations (Shakked & Rabinovich, 1986). The *X*-displacement values (also referred to as D_x in Table 6) shown in Fig. 7 demonstrate that the complexed DNA targets adopt intermediate conformations between that of the R9 structures and the free dodecamer. The major groove of each of the complexed DNA molecules on average is shallower by nearly 2.5 Å and wider by 2–3 Å relative to that of the free R9 (II) target, respectively, but deeper and slightly wider than that of the free B-type dodecamer. The minor groove in the complexed DNA molecules is wider by nearly 1 Å than that in the free dodecamer and narrower by nearly 2 Å than that in the R9 molecules (Table 6). The changes in the complexed DNA duplexes are associated with enhanced bending into the major groove. Such features appear to optimize the interactions between critical amino acids of the Runt domain and the specific base pairs at the major and minor grooves, as discussed previously (Bartfeld *et al.*, 2002). The gradual transition from A-DNA to B-DNA in this system is illustrated by the various structures in Fig. 8, showing two views of the central base pairs of the free and bound DNA structures. Based on the present comparative analysis, it is likely that the transition from A-DNA to B-DNA helix in the RD-binding site involves a small energy barrier and hence the protein can recognize both A-DNA-type and B-DNA-type conformations.

Table 5
Predicted average free-energy changes (kJ mol^{-1}) for B→A transitions.

The values are averaged over dimeric base pairs or trimeric base pairs.

DNA sequence	Dimeric model (Tolstorukov *et al.*, 2001)	Dimeric model (Ivanov & Minchenkova, 1994)	Trimeric model (Tolstorukov *et al.*, 2001)
CTGCGGTC	2.55	2.34	2.09
AGCTGCGGTCAT	2.76	2.51	1.92
A-philic ΔG_{BA}	≤2.5	≤2.5	≤1.3
B-philic ΔG_{BA}	>3.3	>3.3	>4.6

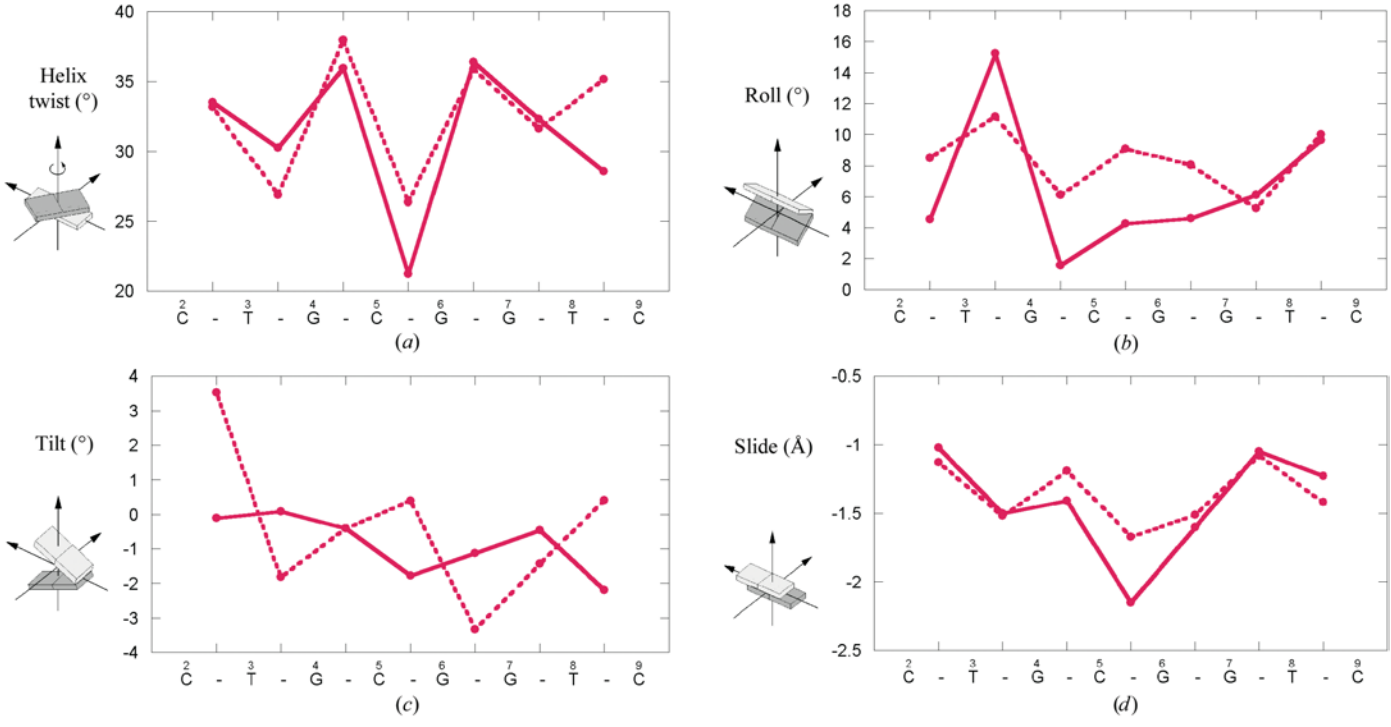


Figure 5
Plots of base-pair step parameters: (*a*) helix twist, (*b*) roll, (*c*) tilt, (*d*) slide. The variations are shown by broken and continuous lines for I and II, respectively. The definitions of the parameters are illustrated by the cartoon at the left of each figure. Only the sequence of the first strand is shown (numbering as in Fig. 1).

Table 6
Average helix parameters, groove dimensions and backbone geometry.

All parameters were calculated with *FREEHELIX* (Dickerson, 1998), except for Z_P which was calculated using *3DNA* (Lu *et al.*, 2000). Values in parentheses refer to the seven base-pair fragments in bold.

(*a*) Average helix parameters.

DNA sequence	Inclination (°)	Roll (°)	Helix twist (°)	Rise (Å)	Slide (Å)	D_x (Å)
TC**TGCGGTC** (I)	12.8 (12.3)	8.3 (7.9)	32.4 (31.8)	3.3 (3.2)	−1.4 (−1.4)	−3.5 (−3.5)
TC**TGCGGTC** (II)	8.1 (7.1)	6.6 (6.4)	31.2 (31.2)	3.3 (3.2)	−1.4 (−1.5)	−3.5 (−3.5)
AGC**TGCGGTC**AT†	5.2 (4.9)	3.7 (2.3)	34.7 (35.1)	3.3 (3.3)	0.1 (0.1)	−0.2 (−0.4)
–CTC**TGTGGTT**GC–‡	6.6 (6.4)	4.1 (3.0)	33.8 (33.0)	3.3 (3.3)	−0.2 (−0.5)	−1.1 (−1.4)
GT**TGCGGTT**G§	6.4 (7.4)	4.0 (3.4)	33.0 (33.5)	3.3 (3.4)	−0.1 (−0.2)	−0.9 (−0.9)
A-DNA¶	21.8	12.0	30.3	3.3	−1.4	−4.4
B-DNA††	2.9	1.5	36.0	3.4	0.5	0.6

(*b*) Groove dimensions and backbone geometry. See Table 4 for definitions.

DNA sequence	Groove width (Å)		Groove depth (Å)		Helix radius (Å)	Z_P (Å)
	Major	Minor	Major	Minor		
TC**TGCGGTC** (I)	6.3	9.6	12.5	4.6	9.4	2.2 (2.2)
TC**TGCGGTC** (II)	9.2	9.5	12.8	4.8	9.6	2.2 (2.3)
AGC**TGCGGTC**AT†	11.1	6.8	9.3	7.8	9.4	−0.1 (−0.2)
–CTC**TGTGGTT**GC–‡	11.6	7.7	10.5	7.4	9.8	0.0 (0.3)
GT**TGCGGTT**G§	12.4	7.4	10.4	7.5	9.8	−0.2 (−0.1)
A-DNA¶	3.7	11.0	12.7	2.9	8.6	2.6
B-DNA††	11.4	5.9	8.4	8.5	9.2	−0.4

† The free DNA dodecamer (Bartfeld *et al.*, 2002). ‡ The central 12-base pair fragment of the 16-mer used in the crystal structure of the RD–DNA complex (Tahirov *et al.*, 2001). § The DNA decamer used for the crystal structure of the RD–CBFβ–DNA complex (Bravo *et al.*, 2001). ¶ A-DNA (CTGCGGTC) derived from the standard A-DNA fibre coordinates (Arnott *et al.*, 1983). †† B-DNA from Chandrasekaran & Arnott (1996).

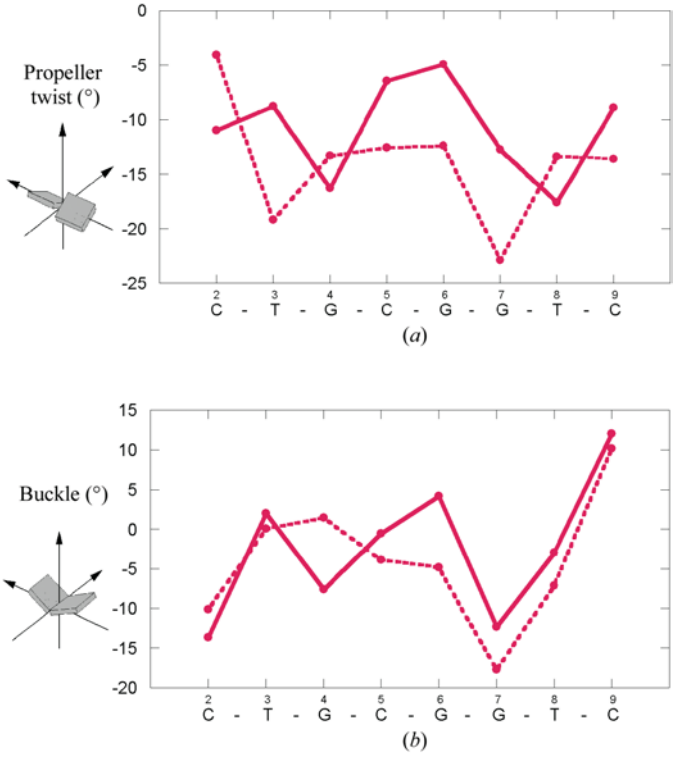


Figure 6
Plots of base-pair parameters: (*a*) propeller twist, (*b*) buckle. The definitions of the parameters are illustrated by the cartoon at the left of each figure.

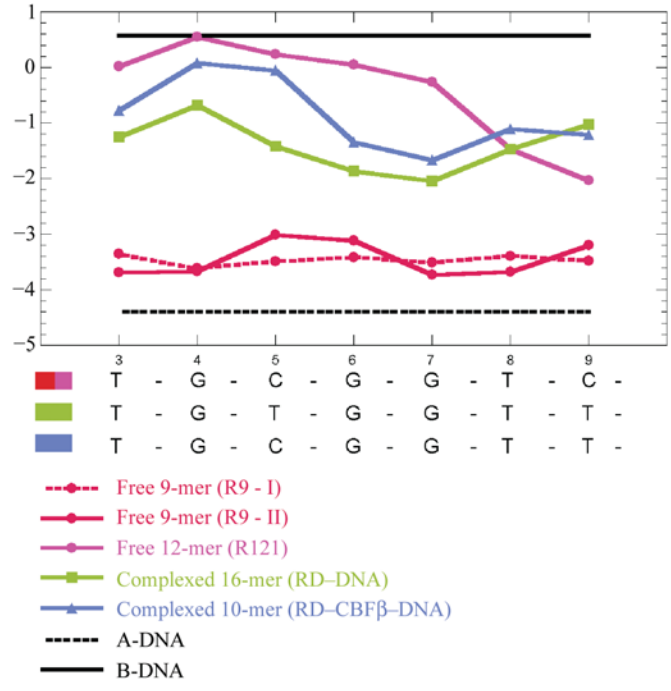


Figure 7
X-displacement plot of the seven-base-pair RD-binding sites. The numbering scheme is as in Fig. 1. Calculations were performed with the program *FREEHELIX* (Dickerson, 1998). The colour code is as shown. The free dodecamer is from Bartfeld *et al.* (2002), the complexed 16-mer is from Tahirov *et al.* (2001) and the complexed decamer is from Bravo *et al.* (2001). A-DNA and B-DNA were derived from the corresponding fibre-based coordinates (Arnott *et al.*, 1983; Chandrasekaran & Arnott, 1996).

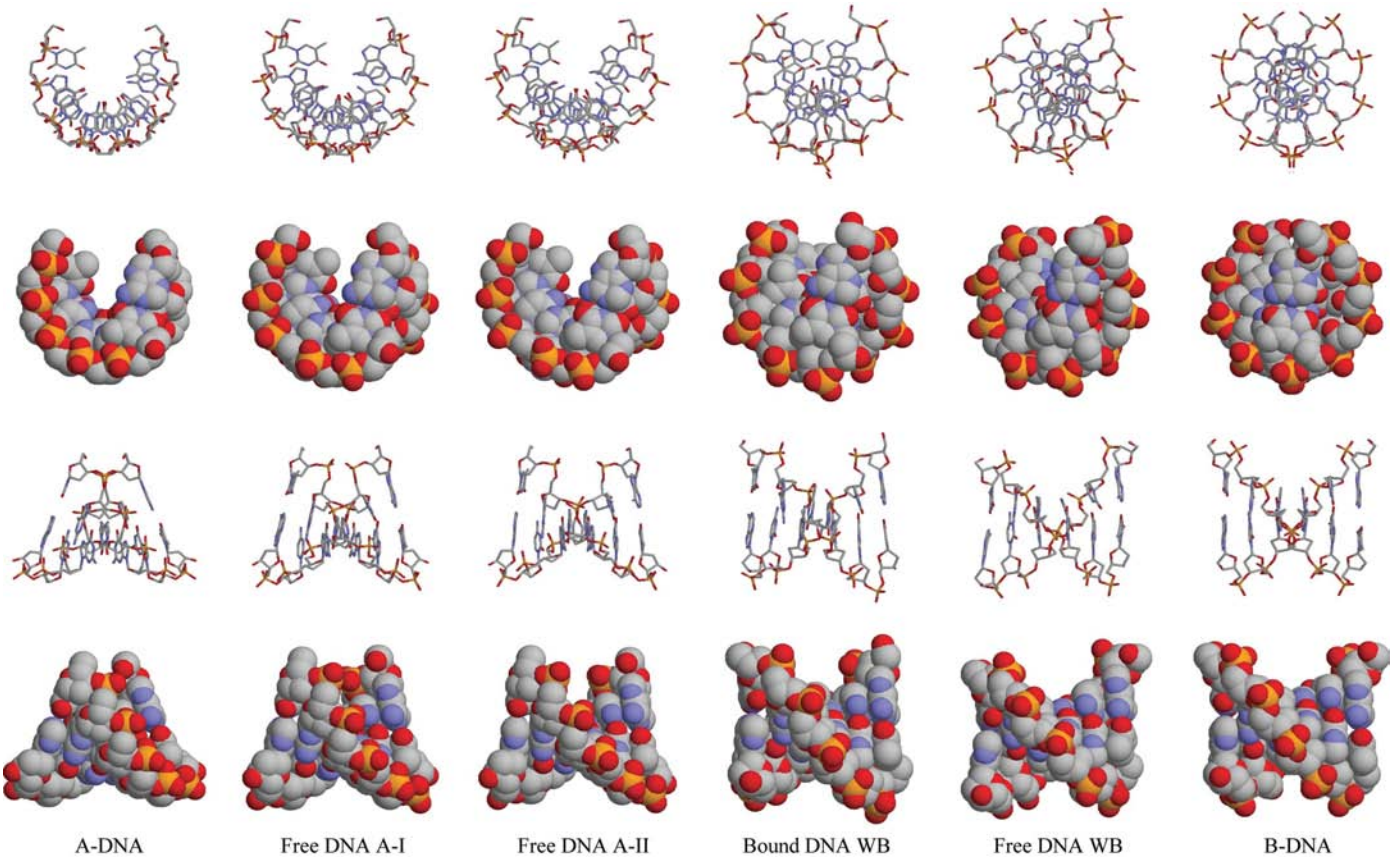


Figure 8
Six conformations of the double-helical TGCGGT fragments viewed down the helix axis (first two rows) and perpendicular to the helix axis (last two rows). Both stick and space-filling drawings are shown. The structures of A-DNA (Arnott *et al.*, 1983) and B-DNA (Chandrasekaran & Arnott, 1996) were derived from fibre-based coordinates. The free A-DNA structures are from the R9 crystal structures (I and II). The bound WB-DNA structure is from the DNA decamer of the RD–CBFβ–DNA complex (Bravo *et al.*, 2001), which is similar to that of the RD–DNA complex (Tahirov *et al.*, 2001) not shown here. The free WB-DNA structure is from the free dodecamer crystal structure (Bartfeld *et al.*, 2002).

4. Summary and conclusions

Site-specific protein–DNA interactions are frequently accompanied by conformational alterations of the protein, the DNA or both. The ability of the molecular partners to induce them in each other makes a significant contribution to the specificity of recognition. In biological systems where a regulatory DNA-binding protein is confronted by multiple versions of its consensus binding-site, this indirect mechanism of target selection can play an important role in controlling the transcriptional program of the organism. The regulatory system of the Runt-domain (RD) proteins represents such a system.

In the present work, two crystal structures of the DNA 9-mer: TC**TGCGGTC**/T**GACCGCA**G (R9), which incorporate the binding site of the RD proteins (bold), were studied and compared with several related DNA structures. The helical conformation patterns exhibited by the two structures are essentially of the A-DNA type with some B-DNA features, displaying intermediate conformations between the two right-handed forms. Based on the dimeric model of Tolstorukov *et al.* (2001), the RD-binding site (TGCGGTC) in the context of the present sequence (R9) is more A-philic than in the context of the B-DNA dodecamer AGC**TGCGGTC**AT studied previously; hence, it is likely that either conformation exists in solution depending on the sequences flanking the specific DNA target. A comparison between the free and bound DNA helices incorporating the consensus RD binding site shows that the complexed DNA helices display conformational features that are intermediate between those of the free A-type and B-type DNA targets, suggesting that the transition from either form to the protein-bound conformation involves a small energy barrier.

This work was supported by grants from the MINERVA Foundation with funding from the Federal German Ministry for Education and Research, and the Israel Science Foundation (ISF) founded by the Israel Academy of Sciences and Humanities. ZS holds the Helena Rubinstein Professorial Chair of Structural Biology at the Weizmann Institute of Science.

References

Arnott, S., Chandrasekaran, R., Hall, I. H., Puigjaner, L. C., Walker, J. K. & Wang, M. (1983). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 53–65.

Backstrom, S., Wolf-Watz, M., Grundstrom, C., Hard, T., Grundstrom, T. & Sauer, U. H. (2002). *J. Mol. Biol.* **322**, 259–272.
Bartfeld, D., Shimon, L., Couture, G., Rabinovich, D., Frolow, F., Levanon, D., Groner, Y. & Shakked, Z. (2002). *Structure*, **10**, 1395–1407.
Bravo, J., Li, Z., Speck, N. A. & Warren, A. J. (2001). *Nature Struct. Biol.* **8**, 371–378.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 585–593.
Chandrasekaran, R. & Arnott, S. (1996). *J. Biomol. Struct. Dyn.* **13**, 1015–1027.
Dickerson, R. E. (1998). *Nucleic Acids Res.* **26**, 1906–1926.
Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Mario Soumpasis, D., Tung, C.-S., Wang, A. H.-J. & Zhurkin, V. B. (1989). *EMBO J.* **8**, 1–4.
Downing, J. R. (1999). *Br. J. Haematol.* **106**, 296–308.
Eisenstein, M., Hope, H., Haran, T. E., Frolow, F., Shakked, Z. & Rabinovich, D. (1988). *Acta Cryst.* B**44**, 625–628.
Eisenstein, M. & Shakked, Z. (1995). *J. Mol. Biol.* **248**, 662–678.
Fernandez, L. G., Subirana, J. A., Verdaguer, N., Pyshnyi, D., Campos, L. & Malinina, L. (1997). *J. Biomol. Struct. Dyn.* **15**, 151–163.
Garvie, C. W. & Wolberger, C. (2001). *Mol. Cell*, **8**, 937–946.
Haran, T. E., Shakked, Z., Wang, A. H. & Rich, A. (1987). *J. Biomol. Struct. Dyn.* **5**, 199–217.
Heinemann, U., Alings, C. & Bansal, M. (1992). *EMBO J.* **11**, 1931–1939.
Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D. & Shakked, Z. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 8490–8495.
Hunter, W. N., D'Estaintot, B. L. & Kennard, O. (1989). *Biochemistry*, **28**, 2444–2451.
Ito, Y. (2004). *Oncogene*, **23**, 4198–4208.
Ivanov, V. I. & Minchenkova, L. E. (1994). *Mol. Biol.* **28**, 780–788.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Brünger, A. T. (1996). *Structure*, **4**, 897–904.
Lavery, R. & Zakrzewska, K. (1999). *Oxford Handbook of Nucleic Acid Structure*, edited by S. Neidle, pp. 39–76. Oxford University Press.
Lu, X. J., Shakked, Z. & Olson, W. K. (2000). *J. Mol. Biol.* **300**, 819–840.
Ng, H. L. & Dickerson, R. E. (2002). *Nucleic Acids Res.* **30**, 4061–4067.
Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C. & Berman, H. M. (2001). *J. Mol. Biol.* **313**, 229–237.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rabinovich, D. & Shakked, Z. (1984). *Acta Cryst.* A**40**, 195–200.
Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S. & Shakked, Z. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 15194–15199.
Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. (1989). *Nature (London)*, **342**, 456–460.
Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. (1990). *Structure Methods*, **3**, 55–72.
Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994). *Nature (London)*, **368**, 469–473.
Shakked, Z. & Rabinovich, D. (1986). *Prog. Biophys. Mol. Biol.* **47**, 159–195.
Sheldrick, G. M. (1997). *SHELXTL Reference Manual*. Bruker-AXS, Madison, WI, USA.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Speck, N. A. & Stacy, T. (1995). *Crit. Rev. Eukaryot. Gene Expr.* **5**, 337–364.
Tahirov, T. H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., Sato, K., Kumasaka, T., Yamamoto, M., Ishii, S. & Ogata, K. (2001). *Cell*, **104**, 755–767.
Tippin, D. B. & Sundaralingam, M. (1997). *J. Mol. Biol.* **267**, 1171–1185.
Tolstorukov, M. Y., Ivanov, V. I., Malenkov, G. G., Jernigan, R. L. & Zhurkin, V. B. (2001). *Biophys. J.* **81**, 3409–3421.
Vargason, J. M., Eichman, B. F. & Ho, P. S. (2000). *Nature Struct. Biol.* **7**, 758–761.
Warren, A. J., Bravo, J., Williams, R. L. & Rabbitts, T. H. (2000). *EMBO J.* **19**, 3004–3015.
Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z. & Crothers, D. M. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 8337–8341.