A novel mode of DNA recognition by a β-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA

Mark D.Allen[1], Kazuhiko Yamasaki[1,2], Masaru Ohme-Takagi, Masaru Tateno[1] and Masashi Suzuki[1,3,4]

AIST-NIBHT Plant Molecular Biology Laboratory and [1]AIST-NIBHT CREST Centre of Structural Biology, Higashi 1-1, Tsukuba 305-0046 and [3]Graduate School of Human and Environmental Sciences, University of Tokyo, Komaba 3-8-1, Meguro, Tokyo 153-0041, Japan

[4]Corresponding author at AIST-NIBHT
[2]Corresponding author for NMR spectroscopy

M.D.Allen and K.Yamasaki contributed equally to this work

The 3D solution structure of the GCC-box binding domain of a protein from *Arabidopsis thaliana* in complex with its target DNA fragment has been determined by heteronuclear multidimensional NMR in combination with simulated annealing and restrained molecular dynamic calculation. The domain consists of a three-stranded anti-parallel β-sheet and an α-helix packed approximately parallel to the β-sheet. Arginine and tryptophan residues in the β-sheet are identified to contact eight of the nine consecutive base pairs in the major groove, and at the same time bind to the sugar phosphate backbones. The target DNA bends slightly at the central CG step, thereby allowing the DNA to follow the curvature of the β-sheet.
Keywords: *Arabidopsis thaliana*/DNA–protein interaction/DNA recognition/NMR/restrained molecular dynamic calculation

Introduction

Expression of pathogenesis-related proteins (Bowles, 1990) is induced in plants by stimuli including UV, salicylic acid and ethylene (Green and Fluhr, 1995), and leads to a defence response by inhibition of bacterial and fungal growth (Hammond-Kosack *et al.*, 1996; Ryals *et al.*, 1996). A consensus nucleotide sequence, AGCCGCC, known as the GCC-box, has been identified in the promoter region of the pathogenesis-related genes (Ohme-Takagi and Shinshi, 1995). Four cDNAs coding ethylene-responsive element binding proteins (EREBPs) from tobacco have been isolated, which specifically bind to the GCC-box (Ohme-Takagi and Shinshi, 1995). Equivalents of EREBPs, AtERF1-4 (M.Ohta, H.Shinshi and M.Ohme-Takagi, unpublished results) and AtEBP (Büttner and Singh, 1997), have been identified in *Arabidopsis thaliana*.

A region of ~60 amino acid residues (Figure 1B) is highly conserved among EREBPs (Jofuku *et al.*, 1994; Ohme-Takagi and Shinshi, 1995). The region is referred to here as the GCC-box binding domain (GBD). A large number of divergent genes in a wide range of plants contain the GBD domain (Weigel, 1995; Elliott *et al.*, 1996; Klucher *et al.*, 1996; Wilson *et al.*, 1996; Okamuro *et al.*, 1997). No animal or fungal protein has been reported to possess a GBD.

In this paper we describe the solution structure of the AtERF1 GBD in complex with the target DNA, and discuss the mode of the specific interaction of the two molecules. The GBD binds to the DNA via its β-sheet. This is somewhat unexpected, since this study reveals that GBD has an arrangement of secondary structural elements which resembles that of zinc fingers to some extent, although it is the α-helices of zinc fingers that are known to bind to DNA. The mode of DNA binding of the GBD β-sheet is different from that of the β-sheets of other known DNA-binding domains in the number and arrangement of β-strands, in the contacting pattern of amino acid residues and DNA bases, and in the number of base pairs contacted. The solution structure of GBD determined in the absence of DNA is also reported to enable better understanding of the structural requirements imposed on the protein structure for DNA recognition.

Results

Three dimensional (3D) structure of GBD

The solution structure of the AtERF1 GBD, Lys145-Val206, in the absence of DNA was determined using 1187 unambiguous conformational constraints (Table I) obtained from multidimensional nuclear magnetic resonance (NMR) experiments. Forty-six structures were selected after simulated annealing, which showed the smallest violations of the constraints (Figure 2A). The average root mean square (r.m.s.) deviation among the selected structures was 0.54 Å for the heavy atoms in the backbone, His146-Pro203 (Table I). The final structure (Figure 2B) was determined by energy minimization of the mean co-ordinates of the ensemble. All the residues were identified in the favored or additionally allowed regions in the (ϕ, ψ) space (Morris *et al.*, 1992).

The determined GBD structure consists of a three-stranded anti-parallel β-sheet comprising strand 1 (Val149-Arg152), strand 2 (Lys156-Asp163) and strand 3 (Ala169-Phe176), packed along an α-helix (Thr178-Arg194) (Figure 1A). The structure is stabilized by extensive hydrophobic contacts of the side-chains of Tyr146, Val149, Phe157, Ala159, Ile161, Val171, Leu173, Phe176, Ala179, Ala182, Ala183, Ala185, Tyr186, Ala189, Ala190, Ala198, Leu200 and Phe202 with each other. The geometry of the α-helix relative to the β-sheet appears to be determined by the interaction of the many Ala residues in the α-helix and the larger hydrophobic residues in the β-sheet, in particular, Phe157, Phe176, Val171 and Ile161, which clamp the
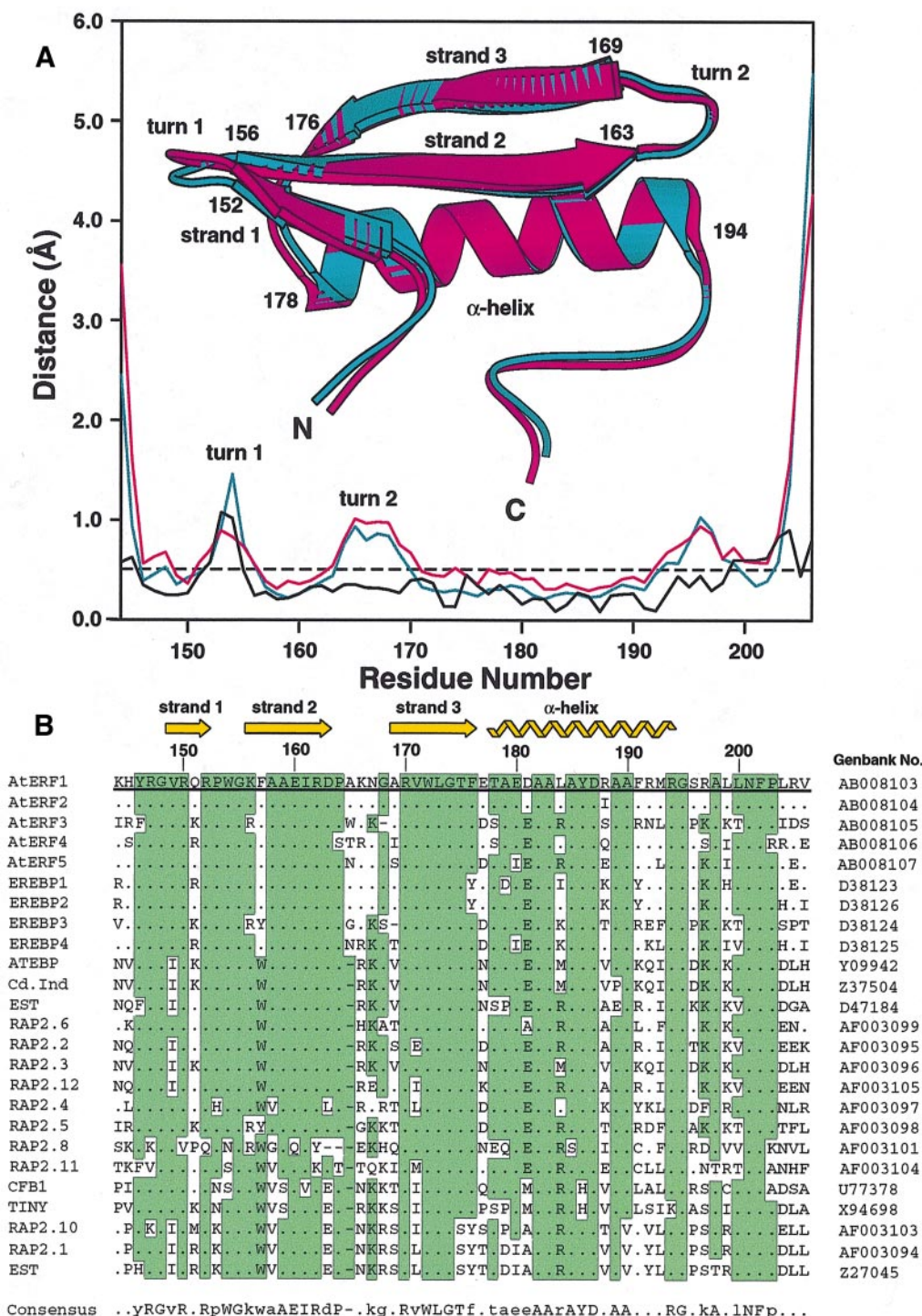
Fig. 1. The 3D structure and the amino acid sequence of GBD. (**A**) The r.m.s. deviation of the structures refined by simulated annealing in the absence of (blue line) or in the complex with (red line) the DNA. The two mean structures, one determined in the absence of (blue) and another determined in the complex with (red) the DNA are shown inset by the superposition of the atoms whose positional difference in the two structures (black line) is <0.5 Å (shown by a broken line). The secondary structural elements, e.g. the α-helix, are labeled with the amino acid positions at the ends, e.g. 178. The N- and C-termini are labeled. (**B**) An alignment of the amino acid sequences of selected GCC-box binding domains with the DDBJ/EMBL/GenBank (Benson *et al.*, 1997) accession numbers on the right. Strongly conserved residues are highlighted in green. The arrangement of the secondary structural elements (top) and the consensus sequence (bottom) are shown.

α-helix at the four corners. The N→C direction of the α-helix is approximately parallel to that of β-strand 2. This arrangement is slightly unusual, since an α-helix is typically tilted slightly with respect to β-strands (Janin and Chothia, 1980).

Turn 1, Arg152-Lys156, between β-strands 1 and 2 is classified to type 3:5, and turn 2, Asp163-Ala169, between strands 2 and 3 is classified to type 5:5 (Sibanda *et al.*, 1989). The N-terminus of the α-helix is capped by hydrogen bonds to and from Thr178. The N-terminal loop,

Table I. Structural constraints and characteristics of the models

Structural constraints	GBD in the absence of DNA		GBD–DNA complex	
Protein intra-residue NOEs	453[a]		474[a]	
sequential NOEs	228[a]		200[a]	
medium-range NOEs ($2 \leq |i-j| \leq 4$)	147[a]		102[a]	
long-range NOEs ($|i-j| > 4$)	292[a]		218[a]	
hydrogen bonds	61[a]		60[a]	
torsion (ϕ) angles	5[a]		0[a]	
DNA intra-residue NOEs	N.A.		372[c]	
sequential NOEs	N.A.		140[c]	
interstrand NOEs	N.A.		29[c]	
hydrogen bonds	N.A.		72[c]	
torsion angles	N.A.		187[c]	
Intermolecular NOEs	N.A.		37[a,c]	
Total	1186[a]		1891[a,c]	

Characteristics	Ensemble of 46 structures[f]	Final structure	Ensemble of 25 structures[f]	Final structure
R.m.s. deviation from constraints				
NOEs and hydrogen bonds of				
protein and DNA (Å)	0.0096 ± 0.0009[a]	0.0078[a]	0.0157 ± 0.0036[a,c]	0.0087[a,d]
Protein torsion angles (degrees)	0.050 ± 0.116[a]	0.000[a]	N.A.	N.A.
DNA torsion angles (degrees)	N.A.	N.A.	0.828 ± 0.603[c]	0.037[d]
van der Waals energy (kcal/mol)	3.77 ± 1.24[a,g]	5.38[a,g]	9.58 ± 4.10[a,c,g]	−985.5[a,d,h]
R.m.s. deviation from the ideal geometry				
Bond lengths (Å)	0.0020 ± 0.0001[a,g]	0.0024[a,g]	0.0061 ± 0.0002[a,c,g]	0.0051[a,d,h]
Bond angles (degrees)	0.53 ± 0.01[a,g]	0.59[a,g]	0.93 ± 0.02[a,c,g]	1.23[a,d,h]
Improper angles (degrees)	0.29 ± 0.02[a,g]	0.40[a,g]	0.37 ± 0.12[a,c,g]	0.41[a,d,h]
Average r.m.s. deviation among the ensemble (Å)				
Protein (N, C$_\alpha$, C)	0.54 ± 0.11[b]	N.A.	0.61 ± 0.11[b]	N.A.
Protein (all heavy atoms)	1.32 ± 0.21[b]	N.A.	1.21 ± 0.09[b]	N.A.
DNA (all heavy atoms)	N.A.	N.A.	0.86 ± 0.09[e]	N.A.
DNA and protein (all heavy atoms)	N.A.	N.A.	1.20 ± 0.12[b,e]	N.A.

[a]Values calculated with Lys144-Val206.
[b]Values calculated with His145-Pro203.
[c]Values calculated with G1–C13 in the coding strand and G14–C26 in the complementary strand.
[d]Values calculated with T3–C13 in the coding strand and G14–A24 in the complementary strand.
[e]Values calculated with T3–A11 in the coding strand and T16–A24 in the complementary strand.
[f]Refined by simulated annealing.
[g]Values calculated by using force-field parameters in X-PLOR.
[h]Values calculated by using force-field parameters in AMBER.
N.A., not applicable.

His145-Gly148, is better defined than the C-terminal loop, Gly195-Pro203 (Figure 1A).

DNA recognition by GBD

The solution structure of the GBD–DNA complex (Figure 3B) was determined by simulated annealing by using NOE constraints obtained from NMR spectra (Figure 3A) followed by restrained molecular dynamic (rMD) calculation (see Figure 4D for the nucleotide sequence and the scheme for the numbering of the bases).

The 25 structures selected after the simulated annealing (Figure 3A) had an average r.m.s. deviation of 0.61 Å for the heavy atoms in the protein backbone, His146-Pro203. The average r.m.s. deviation for all the heavy atoms in the polypeptide, His146-Pro203, nucleotides 3–11 in the coding strand, and nucleotides 16–24 in the complementary strand, was 1.20 Å.

During the rMD calculation the van der Waals energy term in the AMBER force field decreased by 684 kcal/mol. The average inter-base pair rise parameter in the GBD–DNA complex before the rMD calculation had a value of 4.02 ± 0.26 Å. In contrast the structure after the rMD calculation possessed a value of 3.21 ± 0.25 Å,

which was much closer to the values in the known crystal DNA structures. The improvement was due to consideration of the effects of water molecules, which resulted in the hydrophobic effects of base-stacking being included in the rMD calculation.

The orientation of the protein surface with respect to the target DNA was determined unambiguously on the basis of 37 intermolecular nuclear Overhauser effect (NOE) connectivities identified in homonuclear two dimensional (2D) NOE spectroscopy (NOESY) spectra: Arg162 to T16 and G17, Arg150 to C19, Trp154 to T3, A4 and C22, Trp172 to A4, G5 and C6, and Thr175 to A4 (Figures 5C and 6). The intermolecular cross-peaks were observed isolated from the other peaks in two sets of NOESY spectra measured at different temperatures, 298 and 303 K, and were assigned to the same protein–DNA connectivities.

By using ^{13}C, ^{15}N-labeled GBD in the complex with the unlabeled DNA, the 37 intermolecular NOE connectivites were studied further (Figure 7). Splitting of cross-peaks was expected in one of the two frequency dimensions only for intermolecular NOE connectivities (see Materials and methods). The splitting was observed for 16 out of
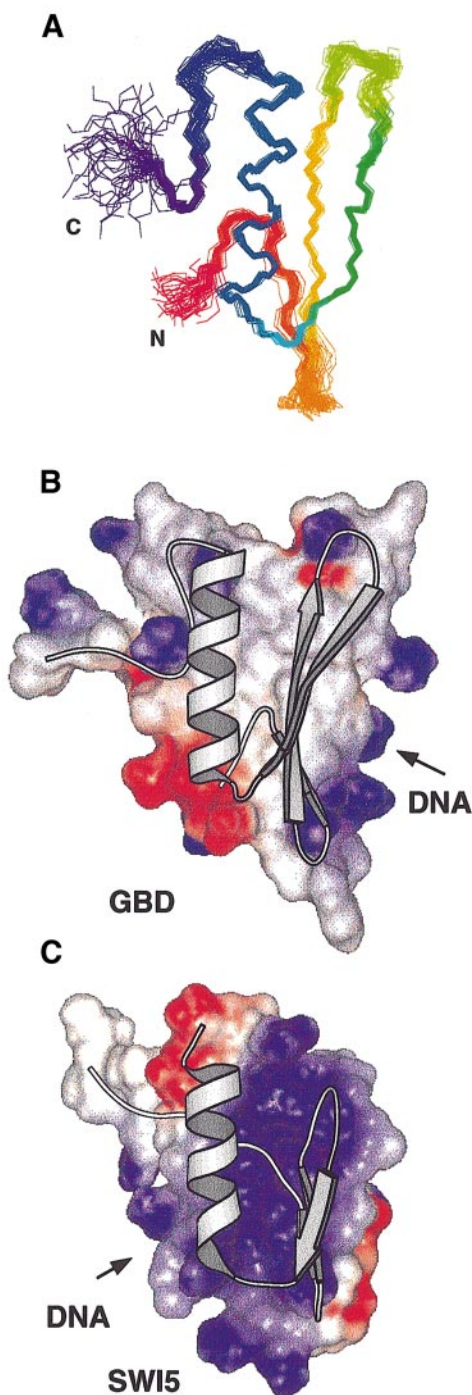
Fig. 2. Comparison of GBD with a zinc finger. (**A**) Superposition of the backbones of 46 GBD structures, Lys145–Val206, in the absence of the DNA refined by simulated annealing. Different colors are used for indicating the secondary structural elements. The N- and C-termini are labeled. (**B**) A diagrammatic drawing of the final structure made by energy minimization of the mean co-ordinates of the ensemble shown in (A), superimposed on a presentation of electropolarization. Positive charges are shown in blue and negative charges in red. The side that recognizes DNA is indicated with an arrow. (**C**) The 3D structure and electropolarization of the first zinc finger of SWI5 (Protein Data Bank code 1NCS) shown in the same way as in (B).

37 connectivities. For the rest (21 connectivities) the intensities of the cross-peaks were too weak for examination. The splitting of the 16 peaks was terminated by decoupling of the heteronuclear spin connectivities

(Figure 7C), which was further evidence that the peaks were those of intermolecular connectivities.

The GBD binds to the major groove of the GCC-box via the three-stranded β-sheet. The N→C direction of strand 2 is approximately parallel to the 5'→3' direction of the coding strand (Figure 3). Only one residue from the α-helix, Tyr186, is identified as binding to a phosphate, while no contact is identified to the bases.

The guanidyl groups of four arginine residues are identified to contact five guanine bases through hydrogen bonds (Figure 5B); Arg150 to O6 of G20, Arg152 to O6 and N7 of G5 and O6 of G21, Arg162 to N7 of G17, and Arg170 to N7 of G8. The stems of the arginine side-chains make hydrophobic interactions with pyrimidine bases; Arg150 to C19, Arg170 to C7, and Arg162 to T16. The side-chain Oε of Glu160 is positioned close to HN4 of C19. Although the distance, 3.11 Å, is slightly larger than would be expected for a standard hydrogen bond, no space remains for accepting a water molecule. The aromatic rings of Trp154 and Trp172 make hydrophobic interactions with T3 and A4, and G5 and C6, respectively. These interactions cover six base pairs in the conserved AGCCGCC sequence and thus serve as the specific recognition of the GCC-box by GBD.

All the Arg and Trp residues which contact the bases also bind to the sugar phosphate backbones either by ionic interactions to the phosphates or by hydrophobic interactions to the sugars, with the exception of Arg152 which does not bond to the backbone: Arg150 to C19 phosphate; Arg162 to G17 phosphate and T16 sugar; Arg170 to C7 phosphate; Trp154 to T3 sugar and A4 phosphate; and Trp172 to G5 phosphate and G5 sugar. Certain other residues bind to the sugar phosphate backbones: Lys156 to A4 phosphate; Thr175 to G5 phosphate and A4 sugar; Arg147 and Tyr186 to G17 phosphate; and the backbone amide of Gly148 to G18 phosphate. These interactions determine the geometry of GBD relative to the DNA and thereby comprise a framework for the specific base recognition.

Although some of the intermolecular contacts identified after the rMD calculation were not formed correctly before the rMD calculation, i.e. one of the four ionic contacts and eight of the 13 hydrogen bonds, the deviation of the candidates in question from the criteria used for identifying contacts was small. In one of the eight hydrogen bonds the distance criterion was satisfied, while in the other seven hydrogen bonds and in the single ionic contact in question the deviation from the distance criteria was 0.74 Å on average. Only two of the eight hydrogen bonds did not fully satisfy the angle criterion; the excess being ~44.7–45.6°.

For the final complex structure around the eight hydrogen bonds that were not correctly formed before the rMD calculation 33 intermolecular NOE connectivities are expected, where the intermolecular distances between the protons are <4 Å. Here connectivities that involve protons whose resonances are not expected to be observed at neutral pH are not included in the calculation. In the NOESY spectra 16 of the 33 were identified, i.e. the identification rate was 49%. Comapred with a similar rate of identification of intramolecular NOE connectivities calculated for the well-determined core of GBD, 58%, the
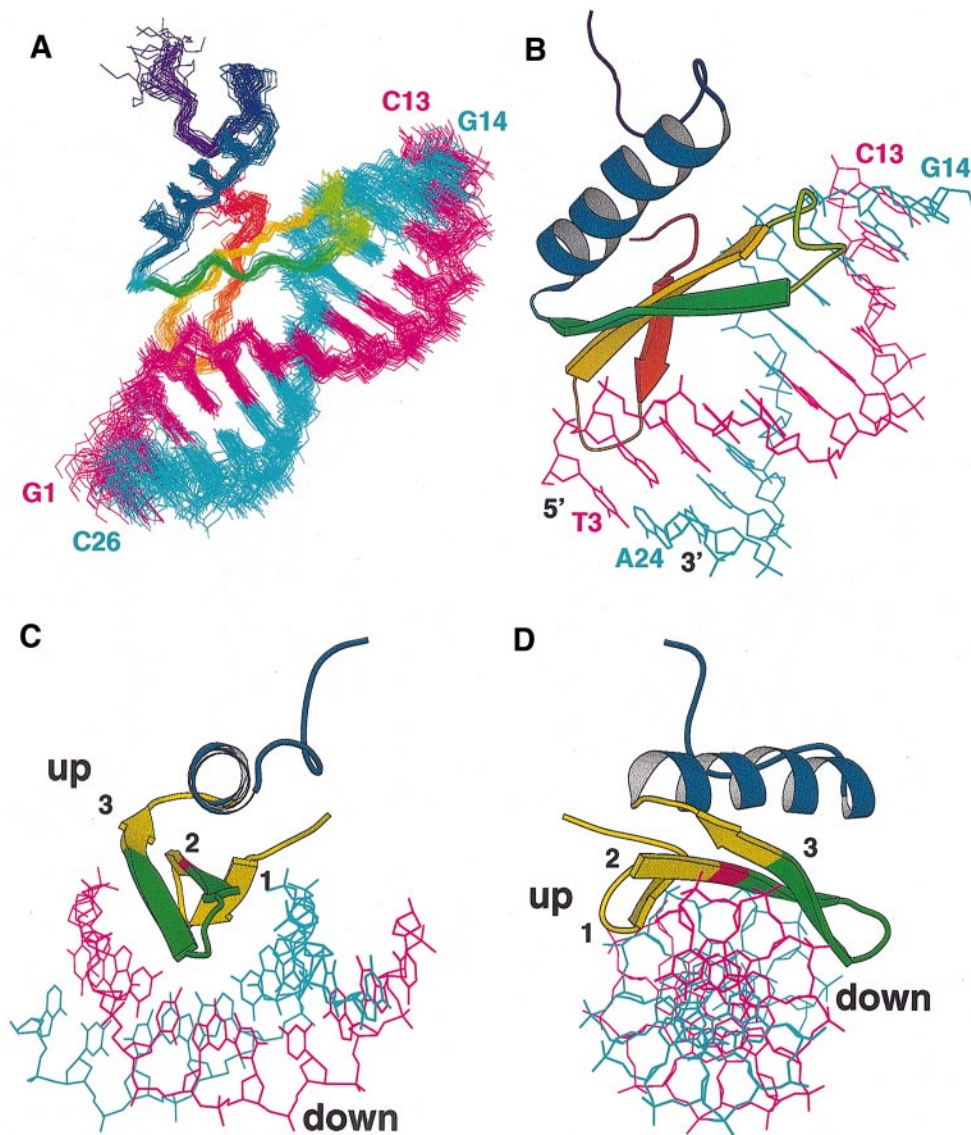
Fig. 3. The 3D structure of the GBD–DNA complex. (**A**) Superposition of 25 GBD–DNA complex structures refined by simulated annealing. Only the backbone is shown for the GBD moiety of the complex structures. Residues Lys145-Val206, nucleotides G1 to C13 in the coding strand (shown in crimson) and nucleotides G14 to C26 in the complementary strand (shown in cyan) of the DNA are included in the calculation. (**B**) A diagrammatic representation of the final structure determined by restrained molecular dynamic calculation using the mean co-ordinates of the ensemble shown in (A). Nucleotides 1 and 2 in the coding strand, and 25 and 26 in the complementary strand were not included in the calculation since these were not contacted by GBD and a meaningful experimental constraint was not observed with them. (C and D) A view of the complex viewed along the α-helix axis of GBD (**C**) and another view looking along the double helix axis of the DNA (**D**). Strands 1–3 are labeled. The position of Ala159 in strand 2 is highlighted in red, the position that separates the β-sheet into two parts (see Figure 5D). The upstream half of the β-sheet is shown in yellow, while the downstream half is in green. The up–down transcription direction is indicated. Note that the direction coincides with the N→C direction of strand 2.

value appears to be acceptable to justify the formation of the hydrogen bonds.

Changes in the two molecules upon binding

The overall structure of the DNA in the complex is close to the standard B-DNA structure, with the r.m.s. difference of 0.62 Å, but it is kinked by ~20° around the major groove at the CG step (Figure 4A). The roll value of the CG step is the largest among the steps, while the helical twist value is the smallest (Figure 4C). The width of the major groove around the CG step is slightly narrower than that expected for the standard B-DNA, while the minor groove is distinctively wider (Figure 4B). These

observations are consistent with the previous analysis of the DNA bending at pyrimidine–purine steps in crystal structures (Suzuki and Yagi, 1995a; Suzuki *et al.*, 1997).

The plot of the major groove width reveals a W-shaped feature around the CG step (Figure 4B). This seems to be a consequence of two contradictory requirements; namely, a primary requirement that the major groove needs to become narrower globally to allow a better fit around the β-sheet, which is carried out by the rolling of the CG step in the center, and a secondary requirement that the β-sheet is widest around Ala159, which is close to the CG step (Figure 5D), and thus the groove becomes wider locally and approaches its standard size around the step.
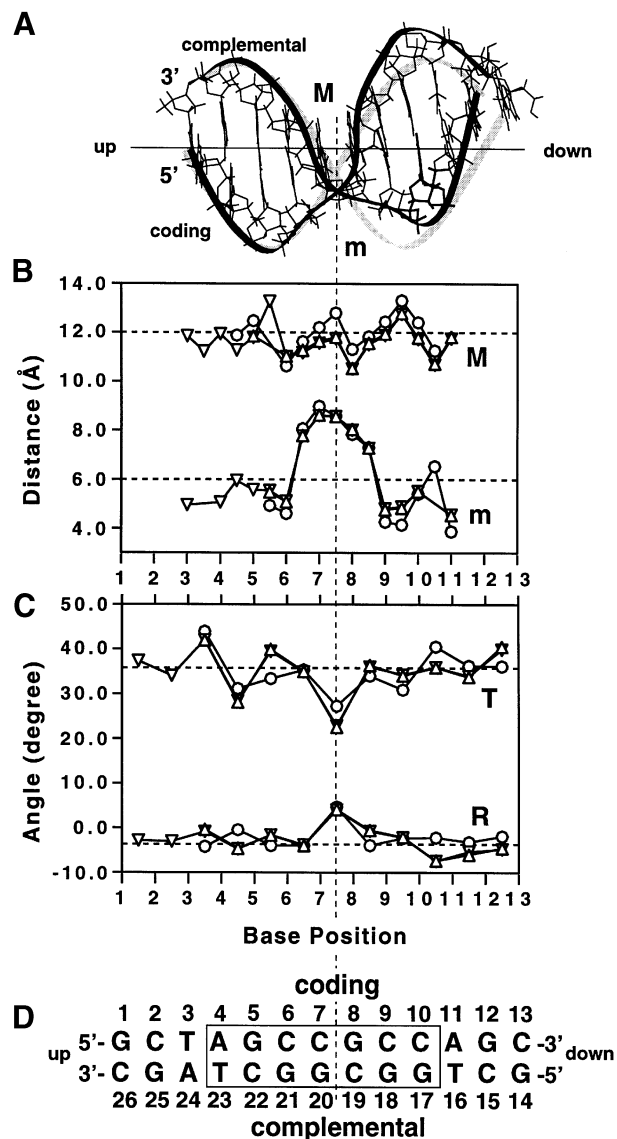
Fig. 4. The DNA conformation in the complex. (**A**) The DNA conformation (shown by darker ribbons) in comparison with the standard B-DNA (shown by lighter ribbons) superimposed at the upstream half, TAGCC/GGCTA. The helix axis of the B-DNA is shown by the straight line running horizontally. The CG step is indicated by a broken line running vertically. The major (M) and minor (m) grooves, and the coding and complementary strands are labeled. The up–down transcription direction is indicated. (B and C) Widths of the major (M) and minor (m) grooves (**B**), and the roll (R) and helical twist (T) angles (**C**). Symbols are used to represent different structures; (Δ) the final rMD structure, (∇) a reference structure made by restrained molecular dynamic calculation using a straight DNA, and (O) another reference structure made by restrained molecular dynamic calculation using a largely bent artificial DNA. The values expected for the standard B-conformation are shown by dotted lines. (**D**) Nucleotide sequences of the two nucleotide strands and the numbering scheme for the bases used in this study. The seven conserved base pairs are shown boxed. The up–down transcription direction is indicated.

To confirm the DNA bending the rMD calculation was repeated twice more by replacing the initial DNA structure by the straight standard B-conformation and the standard B-conformation but modeled to bend to a large degree by inserting an untwisted-rolled CG step found in a crystal DNA structure (Hegde *et al.*, 1992), respectively. The two

rMD calculations resulted in structures very similar to those described above (Figure 4B and C).

The overall GBD structure does not change much upon binding to the target DNA (Figure 1A). The average r.m.s. difference between the mean structure determined in the complex with the target DNA and that determined in the absence of the target DNA is 0.78 Å for the backbone heavy atoms, His146-Pro203, and 1.38 Å when the side-chain heavy atoms were included. These values are comparable with the average r.m.s. deviation in each of the two ensembles (Table I). Locally, turn 1 bends slightly towards the DNA, thereby enhancing the concave curvature of the DNA-binding surface (Figure 1A, inset).

Higher disorder is observed at turn 1, turn 2 and the C-terminal loop in the GBD structure (blue and red lines in Figure 1A) compared with that observed in the GBD–DNA complex. At turn 2 the difference between the two mean structures is subtle (black line in Figure 1A) with only a slight change in the disorder (compare blue and red lines in Figure 1A). However, at turn 1 disorder decreases upon DNA binding, and the r.m.s. difference between the two mean structures is more substantial. It is likely that the changes observed are induced either for or by the binding of Arg152, Trp154 and Lys156 to the DNA.

Discussion

Comparison of GBD with some other DNA-binding domains

The DNA-binding domains of MetJ and Arc repressors interact with the target DNA via β-sheets (Breg *et al.*, 1990; Somers and Phillips, 1992; Raumann *et al.*, 1994). The β-sheets are composed of two identical strands, one from each subunit, and form the dimerization interface. Their target nucleotide sequences are essentially palindromic. In contrast, GBD is monomeric, its β-sheet is three-stranded, and it interacts with the nucleotide of a non-palindromic sequence. In the β-sheet of GBD many residues contact both bases and the backbones, while in Arc and MetJ repressor residues are differentiated for the two types of function. Bases in six consecutive base pairs can be recognized by a β-sheet of the MetJ–Arc type (Suzuki, 1995), while bases in nine consecutive base pairs are recognized by the GBD β-sheet. Thus, the interaction of GBD with the target DNA reported here represents a novel mode of protein–DNA interaction.

The first zinc finger of SWI5 (Dutnall *et al.*, 1996) has the same arrangement of the secondary structural elements as that in GBD. However, zinc fingers bind to the DNA via the α-helix and not the β-sheet. Calculations of electrostatic potentials suggest that the combination of positive charges on the β-sheet and negative charges on the α-helix in GBD (Figure 2B), and that of negative charges on the β-sheet and positive charges at the N-terminus of the α-helix in the zinc finger (Figure 2C) are important for creating the different DNA-binding mechanisms. All zinc fingers of the C_2H_2 type are substantially smaller than GBD (compare Figure 2B and C). Zinc co-ordination to the four amino acid residues can stabilize the core of zinc fingers as effectively as the extensive hydrophobic interactions in the core of GBD.
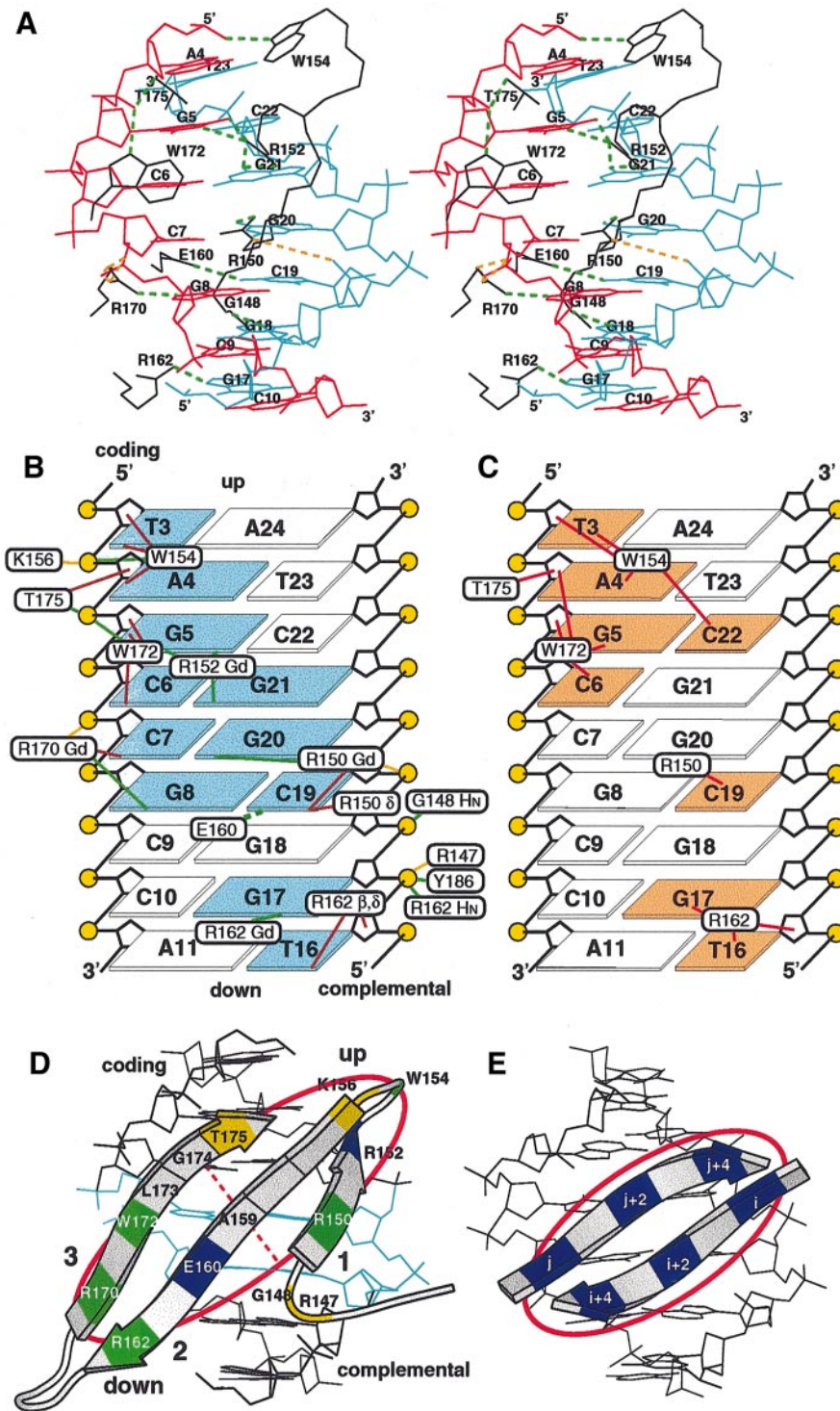
Fig. 5. Contacts between GBD and the GCC-box. (**A**) A stereo subfigure of part of the GBD–DNA complex structure. The contacts are indicated by broken lines with different colors; green for hydrogen bonds and yellow for ionic contacts. Different colors are used for the coding (crimson) and complementary (cyan) strands. (**B**) A diagrammatic representation of the contacts identified after restrained molecular dynamic calculation. The DNA is drawn by looking into the major groove. Yellow circles represent the phosphate groups. The contacted bases are highlighted in cyan. The same color code as in (A) is used for typing the contacts and, in addition, brown for hydrophobic contacts. The distance between Glu160 and C19 is slightly larger than the standard hydrogen bonding distance (shown by a broken line). The up–down transcription direction is indicated. The coding and complementary strands are labeled. Gd: guanidyl group. (**C**) A diagrammatic representation of the NOEs observed between amino acid residues and the target DNA. The bases with which the NOEs are observed are highlighted in orange. (**D**) The three stranded β-sheet of GBD. Residues are colored differently depending on the function; base-contacting only (blue), backbone-binding only (yellow) and having both functions (green). An ellipsoid is drawn by connecting the green positions, which is divided into two halves by the broken line that crosses Ala159. The CG step is highlighted in cyan. The up–down transcription direction is indicated. The coding and complementary strands are labeled. (**E**) A two stranded β-sheet of the MetJ–Arc type. The figure was made using the co-ordinates of the MetJ–DNA complex (Protein Data Base code 1CMA). The six amino acid positions used for base recognition by the MetJ–Arc family are in blue. An ellipsoid is drawn enclosing the blue residues. Compare the ellipsoid with the larger ellipsoid shown in (D) for an appreciation of the difference in size of the two interaction sites.
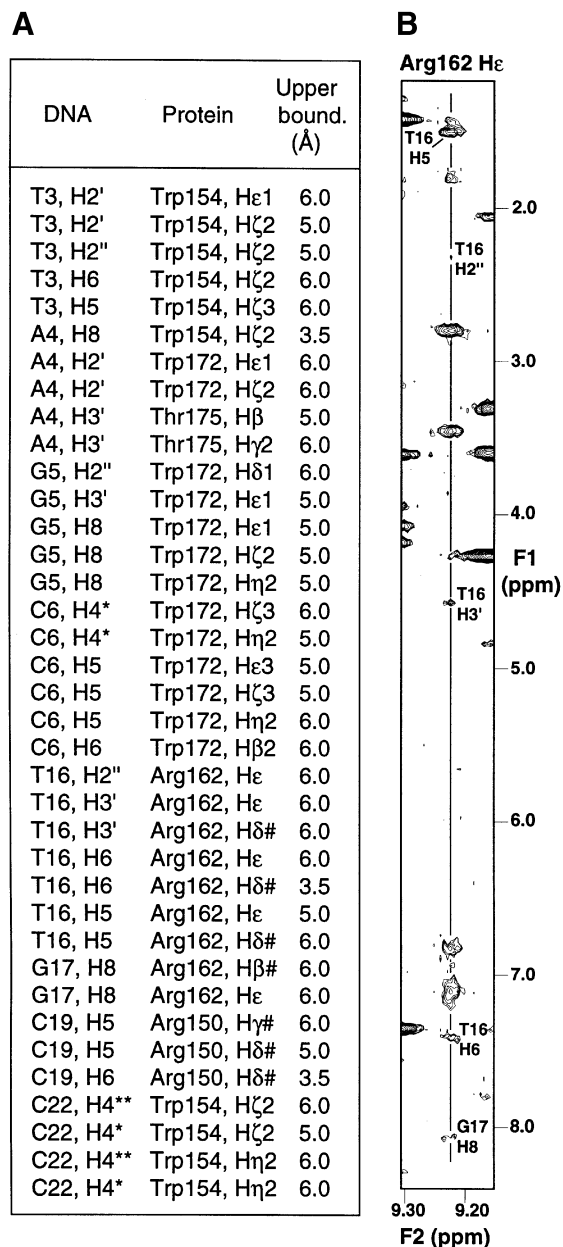
A

DNA	Protein	Upper bound. (Å)
T3, H2'	Trp154, Hε1	6.0
T3, H2'	Trp154, Hζ2	5.0
T3, H2"	Trp154, Hζ2	5.0
T3, H6	Trp154, Hζ2	6.0
T3, H5	Trp154, Hζ3	6.0
A4, H8	Trp154, Hζ2	3.5
A4, H2'	Trp172, Hε1	6.0
A4, H2'	Trp172, Hζ2	6.0
A4, H3'	Thr175, Hβ	5.0
A4, H3'	Thr175, Hγ2	6.0
G5, H2"	Trp172, Hδ1	6.0
G5, H3'	Trp172, Hε1	5.0
G5, H8	Trp172, Hε1	5.0
G5, H8	Trp172, Hζ2	5.0
G5, H8	Trp172, Hη2	5.0
C6, H4*	Trp172, Hζ3	6.0
C6, H4*	Trp172, Hη2	5.0
C6, H5	Trp172, Hε3	5.0
C6, H5	Trp172, Hζ3	5.0
C6, H5	Trp172, Hη2	6.0
C6, H6	Trp172, Hβ2	6.0
T16, H2"	Arg162, Hε	6.0
T16, H3'	Arg162, Hε	6.0
T16, H3'	Arg162, Hδ#	6.0
T16, H6	Arg162, Hε	6.0
T16, H6	Arg162, Hδ#	3.5
T16, H5	Arg162, Hε	5.0
T16, H5	Arg162, Hδ#	6.0
G17, H8	Arg162, Hβ#	6.0
G17, H8	Arg162, Hε	6.0
C19, H5	Arg150, Hγ#	6.0
C19, H5	Arg150, Hδ#	5.0
C19, H6	Arg150, Hδ#	3.5
C22, H4**	Trp154, Hζ2	6.0
C22, H4*	Trp154, Hζ2	5.0
C22, H4**	Trp154, Hη2	6.0
C22, H4*	Trp154, Hη2	6.0

B

Fig. 6. Intermolecular NOE connectivities. (**A**) All the intermolecular NOE connectivities are identified. The upper boundaries beyond which the distance constraints were imposed are shown in the third column. The two H4 protons in C6 or C22 are differentiated from each other by labeling the proton that hydrogen bonds to the partner G base with *, and the other with **. The # symbol indicates that the proton is one of the two in each methylene group, but is not identified uniquely to which one. (**B**) Part of a NOESY spectrum of the unlabeled GBD–DNA complex measured by a 750 MHz NMR spectrometer at 298 K. The mixing time used was 100 ms. Intermolecular NOE connectivities from Arg162 Hε in GBD are labeled with the partner nucleotide protons in the DNA.

Comparison of the DNA-binding modes of the β-sheets

In GBD, β-strand 2 is twisted slightly at Ala159 to follow strand 3 (Figure 5D), whose biphasic regularity is disturbed by the insertion of Gly174; Ala159 makes hydrogen bonds with Leu173 and Gly174. It has been reported that mutation of Gly174 in another GBD caused malfunction of the protein (Jofuku *et al.*, 1994). In one half of the protein–DNA interaction site, downstream of Ala159, strand 3 contacts the bases in the coding strand, while strand 2 contacts bases in the complementary strand (Figure 5D). In the other half, upstream of Ala159, strand 1 contacts the bases in the both strands, but strand 2 contacts no base. At no point do the three strands fit side by side in the center of the major groove as strands 2 and 3 in the upstream half are separated slightly from the groove (Figure 3D). Consequently, the DNA groove accepting the three-stranded β-sheet is not much wider or smaller than that accepting two-stranded β-sheets of MetJ–Arc (Suzuki and Yagi, 1995b).

A two-stranded β-sheet is not totally flat but has a curvature. Thus, there are essentially two ways of placing a two-stranded antiparallel β-sheet in the DNA groove; in one mode the convex side of the β-sheet faces the DNA, while in the other way it is the concave side (Suzuki, 1995). If the N→C direction of each strand follows the 5'→3' direction of the nearest nucleotide strand the β-sheet forms a concave mode, but if the N→C direction of each strand follows the 3'→5' direction of the nearest nucleotide strand, the β-sheet forms a convex mode. A convex mode, which is used by the MetJ–Arc family, is suitable for contacting bases since the major groove is concave (Figure 3C), but it is unable to follow the groove for more than six consecutive base pairs as it does not curve around the DNA helix axis (Figure 5E). In contrast, a concave mode is appropriate for following the DNA backbone, which is convex if viewed along the double helix axis (Figure 3D). However, the concave mode prevents a β-sheet completely entering the groove, and hence residues are unable to contact the DNA bases effectively (Suzuki, 1995). The concave mode has been found in β-sheets which bind to the minor groove, which is shallower than the major groove, by following the DNA backbones (Vis *et al.*, 1995; Rice *et al.*, 1996).

A convex mode is observed in the downstream half of the GBD β-sheet (Figure 5D), while the β-sheet is extended upstream by adopting a concave mode, where strand 2 and irregular strand 3 bind to the DNA via the sugar phosphate backbone, creating a curvature appropriate for following the DNA (Figure 5D). Hence, the GBD β-sheet represents a better design by combining the different modes in two halves, resulting in contact with a larger number of base pairs. The curvature of the β-sheet closely follows the DNA helix axis (Figure 3D) as well as closely fits into the major groove (Figure 3C).

Implication for GBD–protein interaction

Many amino acid residues are identical among the GBD sequences and most of the changes are conservative (Figure 1B). To confirm the formation of essentially the same 3D structure by other GBDs, GBD of C-repeat/DRE binding factor 1 (CBF1) protein (Stockinger *et al.*, 1997) was modeled. The side chains in the AtERF1 GBD were replaced with those present in CBF1 followed by molecular dynamic simulation in water (M.Tateno, K.Yamasaki and M.Suzuki, unpublished results). The same arrangement of the secondary structural elements was observed in the modeled CBF1 structure; the average r.m.s. difference between the modeled and original structures was 2.59 Å.

In a previous discussion (Okamuro *et al.*, 1997) the GBD sequence was divided into two halves, the YRG element in the N-terminal half and the RAYD element in
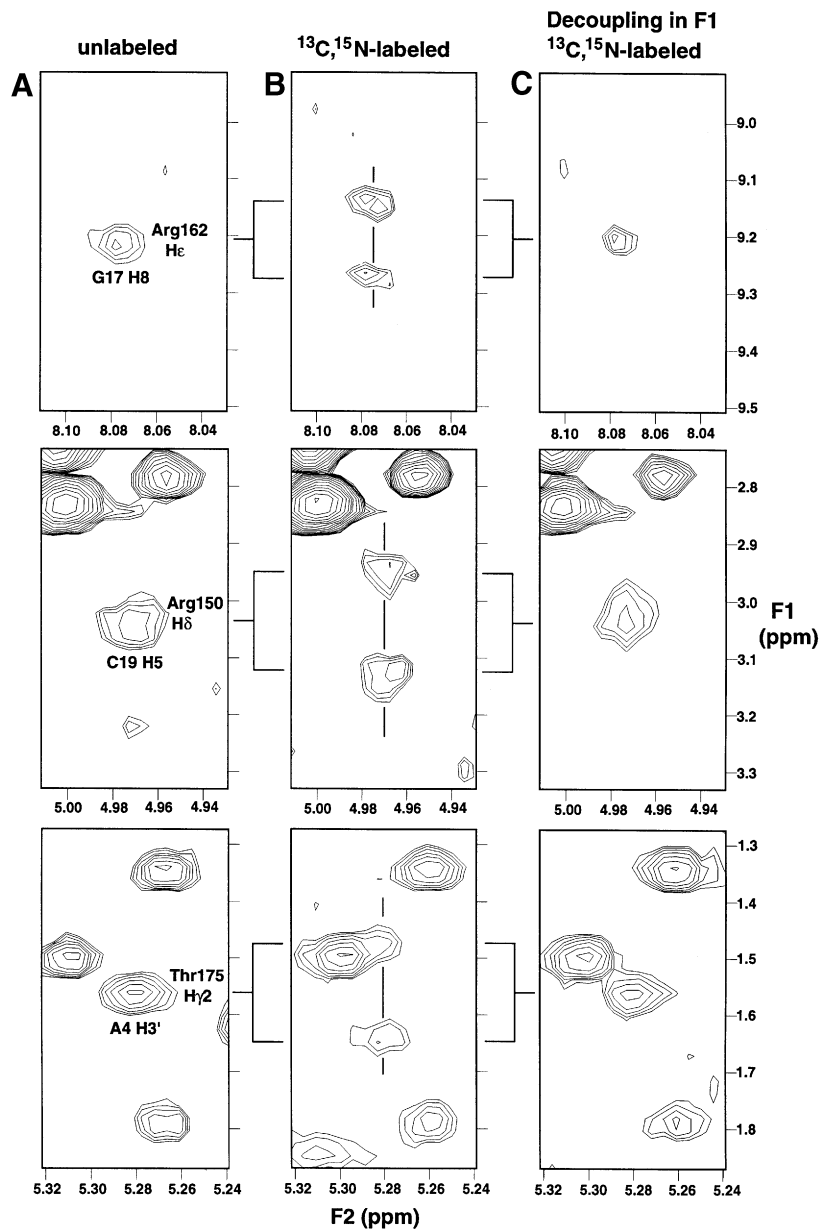
Fig. 7. Examples of intermolecular NOE connectivities observed from the unlabeled (A) and ^{13}C, ^{15}N-labeled (B and C) GBD to the DNA. (**A**) Part of the homonuclear NOESY spectrum of the unlabeled complex. (**B**) Equivalent part of the homonuclear NOESY spectrum of the complex of ^{13}C, ^{15}N-labeled GBD and the unlabeled DNA. (**C**) Equivalent part of homonuclear NOESY spectra of the complex of ^{13}C, ^{15}N-labeled GBD and the unlabeled DNA as in (B) but measured by decoupling the spin connectivities between ^{13}C and ^{1}H (middle and lower), or ^{15}N and ^{1}H (upper) in the F1 dimension. The resonances of the protons in GBD are shown in the F1 dimension, while the resonances of the protons in the DNA in the F2 dimension. The cross-peaks that reflected three intermolecular NOE connectivities are focussed on; the connectivity between Arg162 Hε and G17 H8 in the upper line, that between Arg150 Hδ and C19 H5 in the middle line, and that between Thr175 Hγ2 and A4 H3′ in the lower line. The cross-peaks were split along the F1 dimension, when the labeled GBD was used (B), by the effects of the spins of ^{15}N (upper) or ^{13}C (middle and lower) in the protein, while similar splitting was not observed in the F2 dimension, since the nucleotide protons were not covalently bonded to ^{15}N or ^{13}C in the unlabeled DNA. The splitting was terminated by decoupling the heteronuclear spin connectivities (C). The spectra were measured by a 750 MHz spectrometer at 298 K. The mixing time used for measuring (A) and (B) was 100 ms, while for (C) mixing time was either 100 ms (upper) or 50 ms (middle and lower).

the C-terminal half, and it was predicted that the RAYD element would fold into an α-helix. The two elements do not, however, exactly correspond to the β-sheet and the α-helix, respectively, as reported in this paper. The YRG element corresponds to the N-terminal loop and strands 1 and 2, while the RAYD element corresponds to strand 3, the α-helix and the C-terminal loop. Two alternative possibilities have been discussed for the function of the

predicted α-helix of the RAYD element, DNA-binding or interaction with another protein (Okamuro *et al.*, 1997).

Amino acids of the AtERF1 GBD identified to contact DNA bases are well conserved among the GBD sequences. Thus, it is likely that the DNA-binding specificity of the known GBDs is essentially the same. In the GBD–DNA complex many amino acid residues bind to both bases and the sugar phosphate backbones. Alternation of these

residues could cause truncation in the binding geometry. Thus, it seems more difficult to design a DNA-binding domain of different binding specificity on the basis of the GBD structure than that of some other DNA-binding domains.

The majority of the conserved residues in the GBD sequences can be classified into two categories: those that stabilize the protein structure, and those that are responsible for DNA recognition. However, a few residues appear to have no attributable function. Three such residues, Asp187, Arg194 and Phe202, are accessible to the solvent. It is conceivable that the α-helix or the C-terminal loop might be used for interaction through these residues with another domain, either in the same protein or another protein. *In vitro* studies have demonstrated the ability of AtEBP to interact with some other transcription factors (Büttner and Singh, 1997).

Materials and methods

Preparation of the protein and the DNA
The GBD region of the *AtERF1* gene, corresponding to Gly143-Glu210, was cloned into vector pAF104 (Fukuoh *et al*., 1997) and expressed by the *Escherichia coli* strain BL21(DE3). For isotope-labeling, K-MOPS minimal medium (Neiderhardt *et al*., 1974) containing 10 mM ^{15}N-labeled NH$_4$Cl and/or 0.2% (w/v) [^{13}C]glucose was used for the culture. The protein was purified by cation exchange chromatography (Pharmacia Resource™ S) and gel filtration (Superdex™ 75). By electrospray mass spectrometry it was indicated that the N-terminal methionine had been removed, possibly after translation. The GBD and the GBD–DNA complex were dissolved into a 50 mM potassium phosphate buffer (pH 5.0 or 6.0) containing 40 mM KCl in the different D$_2$O/H$_2$O ratio, depending on the type of the NMR measurements to yield 2.5 and 1.5 mM solutions, respectively.

Spectroscopic measurements
The NMR spectra were recorded on Bruker DMX-750 (750.13 MHz for ^{1}H, 188.64 MHz for ^{13}C and 76.02 MHz for ^{15}N) and DMX-500 (500.13 MHz for ^{1}H, 125.75 MHz for ^{13}C and 50.68 MHz for ^{15}N) spectrometers at 283–303 K; 2D NOESY (Jeener *et al*., 1979), 2D total correlation spectroscopy (TOCSY) (Braunschwieler and Ernst, 1983), 2D double-quantum-filtered (DQF) correlation spectroscopy (COSY) (Braunschwieler *et al*., 1983), 2D ^{1}H–^{15}N HSQC (Bodenhausen and Ruben, 1980), 3D ^{1}H–^{15}N NOESY-heteronuclear multiple-quantum coherence (HMQC) (Marion *et al*., 1989a), 3D ^{1}H–^{15}N TOCSY–HMQC (Marion *et al*., 1989b), 3D HNCA (Ikura *et al*., 1990) and 3D HN(CO)CA (Bax and Ikura, 1991). Quadrature detection in indirect dimensions was carried out by the time-proportional phase incrementation method (Marion and Wüthrich, 1983). During the detection of amide protons, GARP1 ^{15}N-decoupling (Shaka *et al*., 1985) was used. The mixing time chosen for TOCSY was 50 ms, and for NOESY, 25–150 ms. Spectra were referenced relative to external sodium 2, 2-dimethyl-2-silapentane-5-sulfonate for proton and carbon signals, or liquid ammonium for that of nitrogen.

The resonances were assigned using sequential inter-unit NOE connectivities by the standard procedures (Wüthrich, 1986). The ^{13}C$_\alpha$–^{15}N-coupling connectivities observed in the HNCA and HN(CO)CA spectra were also used for the sequential assignment. In the NOESY spectra of the complex the resonances due to the protein moiety and those due to the DNA moiety were identified and separated from each other using the ^{13}C- and ^{15}N-filtration method (Otting and Wüthrich, 1990; Ikura and Bax, 1992). Approximately half of the Hβ resonances were assigned stereospecifically by analyzing the NOESY and TOCSY spectra (Wüthrich, 1986).

Intermolecular NOE connectivities were identified by homonuclear NOESY spectroscopy (Figures 6 and 7) to the same protein–DNA connectivities at two different temperatures, 298 and 303 K. The intermolecular NOE connectivities were further studied by using ^{13}C, ^{15}N-labeled GBD in the complex with the unlabeled DNA. Intermolecular NOE cross-peaks were expected to be split in one of the two frequency dimensions in homonuclear 2D NOESY spectra, since the resonances of protons which were covalently bonded to ^{13}C or ^{15}N in GBD would be split by the coupling effects of ^{13}C or ^{15}N spins, while such splitting was not expected with the resonances of protons in the unlabeled DNA.

Amide hydrogen–deuterium exchange experiments were carried out at 283 K (pD = 5.0) in the absence of the DNA, and at 298 K (pD = 6.0) in the complex with the DNA, in order to identify hydrogen bond donors. Thirty hydrogen bond donors were identified for GBD in the absence of the DNA. For GBD in complex with DNA, 30 hydrogen bond donors were identified, among which 28 were the same as those identified in GBD in the absence of DNA.

Determination of the 3D structure of GBD in the absence of the DNA
All the spectroscopic characteristics suggested that Gly143 and Asn207-Glu210 in GBD were unfolded. Thus, residues Lys144-Val206 were included in all calculations.

The distance constraints derived from the NOESY spectra were classified into four categories corresponding to inter-proton distance constraints of 1.8–2.8, 1.8–3.5, 1.8–5.0, and 1.8–6.0 Å, respectively. To the protons whose resonances were not assigned stereospecifically, i.e. all the protons in the methyl groups, and some protons in the methylene groups and the aromatic rings of Phe and Tyr, the constraints were imposed by using the $<\mathrm{r}^{-6}>$ averaged distances from all the identical protons in the groups or rings (Brünger *et al*., 1986). To maintain a hydrogen bond, a constraint of 1.5–2.5 Å was imposed on the distance between the hydrogen and the acceptor oxygen, while another constraint of 2.5–3.5 Å was imposed on the distance between the donor nitrogen and the acceptor oxygen. A force constant of 50 kcal/mol/Å^2 was used in order to impose the distance constraints outside the allowed regions using the method of Brünger (1992).

Five torsion angle constraints were obtained by HMQC-J (Kuboniwa *et al*., 1994) and were imposed on φ angles in GBD in the absence of the DNA. The angle constraints were classified into three categories, −120 ± 60°, −120 ± 50°, and −120 ± 40°, corresponding to the ^{3}J$_{\alpha N}$ coupling values, <7.5, 7.5–8.5 and >8.5 Hz, respectively. A force constant of 200 kcal/mol/rad^2 was used following the method of Brünger (1992).

Dynamic simulated annealing (Nilges *et al*., 1988a,b) was carried out using the algorithm implemented in the X-PLOR package (Brünger, 1992). Eighty initial structures were created in the form of random arrays of atoms (Nilges *et al*., 1988a). Forty-six structures were selected, where no distance violation was >0.2 Å and no torsion angle violation was >3°; the rate of acceptance was 57.5%. The mean co-ordinates of the ensemble was submitted to 400 cycles of energy minimization (Powell, 1977). Differences between the 46 structures were characterized by calculating the average r.m.s. deviations of the co-ordinates from the unminimized mean structure of the ensemble (Table I).

The simulated annealing was repeated twice. In the first round no hydrogen bond constraints were imposed. Of the 54 structures accepted after the first round, candidates for the acceptors were identified using part of a program NAOMI (Brocklehurst and Perham, 1993) for the hydrogen bond donors that were identified by the H–D exchange experiments. When two or more candidates of acceptors were found for the same donor in different structures, the most frequently occurring candidate was selected. In all such cases the pairs selected possessed the most favorable pseudo-energy values. A hydrogen bond donor, the δ nitrogen of Asn201, was identified to form two hydrogen bonds by using both of the two hydrogens to the same acceptor atom, producing three constraints. For the 30 hydrogen bond donors identified by the H–D exchange experiments, 61 constraints were made. The hydrogen bond constraints were imposed in the second round of the simulated annealing in order to maintain the selected donor–acceptor pairs. The method used was essentially the same as that of Kalia *et al*. (1993).

Determination of the 3D structure of the GBD–DNA complex
Nucleotides 1–13 in the coding strand and 14–26 in the complementary strand were included in the simulated annealing of the complex. In the rMD calculation, nucleotides 3–13 in the coding strand and 14–24 in the complementary strand were included, since the upstream end of the DNA was not contacted by the protein and a meaningful experimental constraint was not observed with the end section. Residues Lys144-Val206 of GBD were included in all the calculations.

The simulated annealing of the GBD–DNA complex was carried out essentially in the same way as in that of GBD in the absence of the DNA. The initial structures of DNA were random arrays of atoms, as were the initial structures of GBD. In addition to the 1632 experimental constraints the following theoretical constraints were imposed: 72 constraints on the intra-base pair hydrogen bonds in order to maintain the base-pairing (Werner *et al.*, 1995) with a force constant of 200 kcal/mol/Å^2; 65 constraints on the rotation of bases around the intra-base pair hydrogen bonds in order to ensure good stereochemistry in terms of propeller-twisting and base-rolling (Omichinski *et al.*, 1997) with a force constant of 200 kcal/mol/rad^2; and 122 constraints on the rotation of the atoms around the covalent bonds in the DNA backbones in order to prevent problems associated with the mirror image (Omichinski *et al.*, 1993) with a force constant of 200 kcal/mol/rad^2.

Starting with 200 random structures, 25 structures were refined, where no distance violation in the protein structure was >0.2 Å, no distance violation in the DNA structure was >0.3 Å, no GBD–DNA distance violation was >0.3 Å, and no torsion angle violation was >3°. Although the rate of accepting structures after the simulated annealing of the complex of 12.5% was different from that of GBD in the absence of the DNA (57.5%), the calculation algorithm and the criteria for the selection of structures were kept essentially the same. In fact, the r.m.s. deviation values of GBD structures in the two ensembles are found to be similar, i.e. 1.21 ± 0.09 Å for all the heavy atoms of GBD in the complex with the DNA, and 1.32 ± 0.21 Å for those of GBD in the absence of DNA (Table I). The difference between the two acceptance rate values is likely to be due to the difference in molecular weight.

After energy minimization in the AMBER force field the total energy value of the minimized mean structure became significantly lower than the smallest among the values calculated with the ensemble structures. The minimized mean structure was submitted to the rMD calculation.

The rMD calculation was carried out by using the force field (Cornell *et al.*, 1995) implemented in the AMBER package (Pearlman *et al.*, 1994) by imposing only the NOE distance constraints and not by imposing a torsion angle constraint or a hydrogen bond constraint. A force constant of 30 kcal/mol/Å^2 was used in order to impose the NOE constraints. The net charge of the protein and the DNA were neutralized by counterions. Approximately 5000 water molecules were modeled by the TIP3P method (Jorgensen *et al.*, 1983). Approximation of the electrostatic force by truncated Coulombic potentials, which ignores interactions beyond a chosen cut-off length, can distort the molecular system significantly. In this study, long-ranged electrostatic forces were taken into consideration by the particle mesh Ewald method (Darden *et al.*, 1993). The lengths of covalent bonds from hydrogens were kept constant using the SHAKE method (van Gunsteren and Berendsen, 1977).

Trajectories were generated every 0.002 ps under the pressure of 1 atm. For the first 20 ps the co-ordinates of the atoms in the protein and the DNA, and of the counterions were kept the same, while only the co-ordinates of the atoms of the water molecules were changed. Temperature was kept constant at 300 K. For the following 6 ps only the co-ordinates of counterions and water molecule atoms were changed for every 2 ps at 100, 200 and 300 K. For the following 8 ps the calculation was continued by changing the co-ordinates of all the atoms every 2 ps at 100, 150, 200 and 250 K. The calculation was further continued for 302 ps at 300 K. The co-ordinates of 200 complex structures were selected, that were produced at each 1 ps during the last 200 ps, and the mean co-ordinates were subjected to energy minimization.

Analysis of the structures

The secondary structural elements in GBD were identified on the basis of NOE profiles characteristic of the elements and the observed protection of the amide protons from the exchange to deuterons (Wüthrich, 1986). The φ and ψ dihedral angles were analyzed using the Procheck–NMR program (Laskowski *et al.*, 1996). Diagrammatic representations were made using MOLSCRIPT (Kraulis, 1991). Electrostatic potentials were essentially calculated by the method of Guenot *et al.* (1994). Parameters for describing the geometry of dinucleotide steps in the DNA were calculated using a program from Babcock *et al.* (1994). The widths of the major and minor grooves were calculated according to the method of Suzuki and Yagi (1996).

The criterion used in this study for identifying an ionic interaction between Arg or Lys and a DNA phosphate was a distance of <5.0 Å

from the side-chain HN to the phosphate oxygen (Billeter *et al.*, 1993). The criteria for an intermolecular hydrogen bond was a donor–acceptor distance of <3.4 Å and a donor–proton–acceptor angle of >110°, while the criterion for a hydrophobic interaction was a C–C distance of <4 Å (Chuprina *et al.*, 1993). The criteria used for identifying hydrogen bonds in the protein were the same as those described by Brocklehurst and Perham (1993).

The co-ordinates of the structures determined and a list of the experimental constraints have been deposited in the Protein Data Bank (Bernstein *et al.*, 1977). The accession code for the final structure of the GBD–DNA complex is 1GCC, while those for the GBD in the absence of DNA is 2GCC for the minimized mean structure, and 3GCC for the ensemble after the simulated annealing.

References

Babcock,M.S., Pednault,E.P.D. and Olson,W.K. (1994) Nucleic acid structure analysis: mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.*, **237**, 125–156.

Bax,A. and Ikura,M. (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α-carbon of the proceeding residue in uniformly ^{15}N/^{13}C enriched proteins. *J. Biomol. NMR*, **1**, 99–104.

Benson,D.A., Boguski,M.S., Lipman,D.J. and Ostell,J. (1997) GenBank. *Nucleic Acids Res.*, **25**, 1–6.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F. Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.

Billeter,M., Qian,Y.Q., Otting,G., Müller,M., Gehring,W. and Wüthrich,K. (1993) Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain–DNA complex. *J. Mol. Biol.*, **234**, 1084–1097.

Bodenhausen,G. and Ruben,D. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **569**, 185–200.

Bowles,D.J. (1990) Defense-related proteins in higher plants. *Annu. Rev. Biochem.*, **59**, 873–907.

Braunschwieler,L. and Ernst,R.R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.*, **53**, 521–528.

Braunschwieler,L., Bodenhausen,G. and Ernst,R.R. (1983) Analysis of networks of couple spins by multiple quantum NMR. *Mol. Phys.*, **48**, 535–560.

Breg,J.N., Opheusden,J.H., Burgering,M.J., Boelens,R. and Kaptain,R. (1990) Structure of Arc repressor in solution: evidence for a family of beta-sheet DNA-binding proteins. *Nature*, **346**, 586–589.

Brocklehurst,S.M. and Perham,R.N. (1993) Prediction of the three-dimensional structures of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycogen cleavage system: a new automated method for the prediction of protein tertiary structure. *Protein Sci.*, **2**, 626–639.

Brünger,A.T. (1992) *X-PLOR Version 3.1: A System for Crystallography and NMR*. Yale University Press, New Haven, CT.

Brünger,A.T., Clore,G.M., Gronenborn,A.M. and Karplus,M. (1986) Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin. *Proc. Natl Acad. Sci. USA*, **83**, 3801–3805.

Büttner,M. and Singh,K.B. (1997) *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-indelible, GCC box binding protein interacts with an *ocs* element binding protein. *Proc. Natl Acad. Sci. USA*, **94**, 5961–5966.

Chuprina,V.P., Rullmann,J.A.C., Lamerichs,R.M.J.N., Van Boom,J.H., Boelens,R. and Kaptein,R. (1993) Structure of the complex of *lac* repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. Mol. Biol.*, **234**, 446–462.

Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

Darden,T.A., York,D. and Pedersen,L.G. (1993) Particle mesh Ewald: an N log (N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

Dutnall,R.N., Neuhaus,D. and Rhodes,D. (1996) The solution structure of the first zinc finger of SWI5: A novel structure extension to a common fold. *Structure*, **4**, 599–611.

Elliott,R.C., Betzner,A.S., Huttner,E., Oakes,M.P., Tucker,W.Q.J., Generates,D., Perez,P. and Smyth,D.R. (1996) AINTEGUMENTA, an APETALA2-like gene of *Arabidopsis thaliana* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell*, **8**, 155–168.

Fukuoh,A., Iwasaki,H., Ishioka,K. and Shinagawa,H. (1997) ATP-dependent resolution of R-loops at the *ColE1* replication origin by *Escherichia coli RecG* protein, Holliday junction-specific helicase. *EMBO J.*, **16**, 203–209.

Green,R. and Fluhr,R. (1995) UV-B induced PR-1 accumulation is mediated by active oxygen species. *Plant Cell*, **7**, 203–212.

Guenot,J., Fletterick,R.J. and Kollman,P.A. (1994) A negative electrostatic determinant mediates the association between the *Escherichia coli* trp repressor and its operator DNA. *Protein Sci.*, **3**, 1276–1285.

Hammond-Kosack,K.E. and Jones,J.D. (1996) Resistance gene dependent plant defence responses. *Plant Cell*, **8**, 1773–1791.

Hegde,R.S., Grossman,S.R., Laimins,L.A. and Sigler,P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.

Ikura,M. and Bax,A. (1992) Isotope-filtered 2D NMR of a protein–peptide complex: study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.*, **114**, 2433–2440.

Ikura,M., Kay,L.E. and Bax,A. (1990) A novel approach for sequential assignment of ^{1}H, ^{13}C and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy; application to calmodulin. *Biochemistry*, **29**, 4659–4667.

Janin,J. and Chothia,C. (1980) Packing of alpha-helices onto beta-pleated sheets and the anatomy of alpha/beta proteins. *J. Mol. Biol.*, **143**, 95–128.

Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) Investigation of exchange processes to two-dimensional NMR spectroscopy. *J. Chem. Phys.*, **71**, 4546–4553.

Jofuku,K.D., denBoer,B.G.W., Montagu,M.V. and Okamuro,J.K. (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell*, **6**, 1211–1225.

Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926.

Kalia,Y.N., Brocklehurst,S.M., Hipps,D.S., Appella,E., Sakaguchi,K. and Perham,R.N. (1993) The high-resolution structure of the peripheral subunit-binding domain of dihydrogenase multienzyme complex of *Bacillus stearothermophilus*. *J. Mol. Biol.*, **230**, 323–341.

Klucher,K.M., Chow,H., Reiser,L. and Fischer,R.L. (1996) The AINTEGUMENTA gene of *Arabidopsis thaliana* required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. *Plant Cell*, **8**, 137–153.

Kraulis,P.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Kuboniwa,H., Grzesiek,S., Delaglio,F. and Bax,A. (1994) Measurement of Hn-Hα J. couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR*, **4**, 871–878.

Laskowski,R.A., Rullmann,J.A.C., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.

Marion,D. and Wüthrich,K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H–^{1}H spin–spin coupling constants in proteins. *Biochem. Biophys. Res. Comm.*, **113**, 967–974.

Marion,D., Kay,L.E., Sparks,S.W., Torchia,D.A. and Bax,A. (1989a) Three-dimensional heteronuclear NMR of ^{15}N-labeled proteins. *J. Am. Chem. Soc.*, **111**, 1515–1517.

Marion,D., Driscoll,P.C., Kay,L.E., Wingfiels,P.T., Bax,A.,

Gronenborn,A.M. and Clore,G.M. (1989b) Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by the use of three-dimensional heteronuclear ^{1}H–^{15}N Hartmann–Hahn-multiple quantum coherence and nuclear overhauser-multiple quantum correlation spectroscopy: application on interleukin 1b. *Biochemistry*, **28**, 6150–6156.

Morris,A.L., MacArthur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins*, **12**, 345–364.

Neidhardt,F.C., Bloch,P.L. and Smith,D.F. (1974) Culture medium for enterobacteria. *J. Bacteriol.*, **119**, 736–747.

Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988a) Determination of the three-dimensional structures of proteins from inter-proton distance data by dynamic simulated annealing from a random array of atoms. *FEBS Lett.*, **239**, 129–136.

Nilges,M., Gronenborn,A.M., Brünger,A.T. and Clore,G.M. (1988b) Determination of the three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine protease inhibitor 2. *Protein Eng.*, **2**, 27–38.

Ohme-Takagi,M. and Shinshi,H. (1995) Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell*, **7**, 173–182.

Okamuro,J.K., Caster,B., Villarroel,R., Montagu,M.V. and Jofuku,K.D. (1997) The AP2 domain of *APETALA2* defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **94**, 7076–7081.

Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahi,S.J. and Gronenborn,A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438–446.

Omichinski,J.G., Pedone,P.V., Felsenfeld,G., Gronenborn,A.M. and Clore,G.M. (1997) The solution structure of a specific GAGA factor–DNA complex reveals a modular binding mode. *Nature Struct. Biol.*, **4**, 122–132.

Otting,G. and Wüthrich,K. (1990) Heteronuclear filters in two-dimensional [^{1}H,^{1}H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular comformation and intermolecular interactions. *Quant. Rev. Biophys.*, **23**, 39–96.

Pearlman,D.A., Case,D.A., Caldwell,J.W., Ross,W.S., Cheatham III,T.E., Ferguson,D.M., Seibel,G.L., Singh,U.C., Weiner,P.K. and Kollman,P.A. (1994) *AMBER 4.1*. University of California, San Francisco, CA.

Powell,M.J.D. (1977) Restart procedures for the conjugate gradient method. *Math. Program.*, **12**, 241–254.

Raumann,B.E., Rould,M.A., Pabo,C.O. and Sauer,R.T. (1994) DNA recognition by β-sheets in the Arc repressor–operator crystal structure. *Nature*, **367**, 754–757.

Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF–DNA complex: a protein-included DNA u-turn. *Cell*, **87**, 1295–1306.

Ryals,J.A., Neuenschwander,U.H., Willits,M.G., Molina,A., Steiner,H.-Y. and Hunt,M. (1996) Systemic acquired resistance. *Plant Cell*, **8**, 1809–1819.

Shaka,A.J., Barker,P.D. and Freeman,R. (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. *J. Magn. Reson.*, **64**, 547–552.

Sibanda,B.L., Blundel.T.L. and Thornton,J.M. (1989) Conformation of β-hairpins in protein structures: a systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.*, **206**, 759–777.

Somers,W.S. and Phillips,S.E.V. (1992) Crystal structure of the met repressor–operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature*, **359**, 387–393.

Stockinger,E.J., Gilmour,S.J. and Thomashow,M.F. (1997) *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proc. Natl Acad. Sci. USA*, **94**, 1035–1040.

Suzuki,M. (1995) DNA recognition by a β-sheet. *Protein Eng.*, **8**, 1–4.

Suzuki,M. and Yagi,N. (1995a) Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res.*, **23**, 2083–2091.

Suzuki,M. and Yagi,N. (1995b) DNA recognition by a β-sheet (2): stereochemical basis of DNA bending by the β-sheets of MetJ and Arc repressors. *Proc. Japan Acad.*, **B71**, 1108–1113.

Suzuki,M. and Yagi,N. (1996) An in-the-groove view of DNA structures in complexes with proteins. *J. Mol. Biol.*, **255**, 677–687.

Suzuki,M., Amano,N., Kakinuma,J. and Tateno,M. (1997) Use of a 3D structure data base for understanding sequence-dependent conformational aspects of DNA. *J. Mol. Biol.*, **274**, 421–434.

van Gunsteren,W.F. and Berendsen,H.J.C. (1977) Algorithms for macromolecular dynamics and constraints dynamics. *Mol. Phys.*, **34**, 1311–1327.

Vis,H., Mariani,M., Vorgias,C.E., Wilson,K.S., Kaptain,R. and Boelens,R. (1995) Solution structure of the HU protein from *Bacillus stearothermophilus*. *J. Mol. Biol.*, **254**, 692–703.

Weigel,D. (1995) The APETALA2 domain is related to a novel type of DNA binding domain. *Plant Cell*, **7**, 388–389.

Werner,H.M., Huth,J.R., Gronenborn,A.M. and Clore,G.M. (1995) Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, **81**, 705–714.

Wilson,K., Long,D., Swinburne,J. and Coupland,G. (1996) A dissociation insertion causes a semi-dominant mutant that increases the expression of TINY, an *Arabidopsis* gene related to APETALA2. *Plant Cell*, **8**, 659–671.

Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, New York.