Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain

A Nagadoi[1,2], S Morikawa[3], H Nakamura[3], M Enari[1], K Kobayashi[1], H Yamamoto[2], G Sampei[4], K Mizobuchi[4], MA Schumacher[5], RG Brennan[5] and Y Nishimura[1]*

[1]Graduate School of Integrated Science, Yokohama City University, 22-2 Seto, Kanazawa-ku, Yokohama, 236 Japan, [2]Faculty of Science, Kanagawa University, 2946 Tsuchiya, Hiratsuka, 251-19 Japan, [3]Protein Engineering Research Institute, Furuedai, Suita, Osaka, 565 Japan, [4]Department of Applied Physics & Chemistry, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182 Japan and [5]Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, Portland, OR 97201-3098, USA

Background: The purine repressor (PurR) regulates genes that encode enzymes for purine biosynthesis. PurR has a two domain structure with an N-terminal DNA-binding domain (DBD) and a C-terminal corepressor-binding domain (CBD). The three-dimensional structure of a ternary complex of PurR bound to both corepressor and a specific DNA sequence has recently been determined by X-ray crystallography.

Results: We have determined the solution structure of the PurR DBD by NMR. It contains three helices, with the first and second helices forming a helix-turn-helix motif. The tertiary structure of the three helices is very similar to that of the corresponding region in the ternary complex. The structure of the hinge helical region, however, which makes specific base contacts in the minor groove of DNA, is disordered in the DNA-free form.

Conclusion: The stable formation of PurR hinge helices requires PurR dimerization, which brings the hinge regions proximal to each other. The dimerization of the hinge helices is likely to be controled by the CBD dimerization interface, but is induced by specific-DNA binding.

Structure 15 November 1995, **3**:1217–1224

Key words: DNA-induced α helix, lactose repressor, NMR, purine repressor

Introduction

In *Escherichia coli*, the genes encoding the 13 enzymes required for *de novo* purine biosynthesis form 10 polycistronic and monocistronic operons and their transcription is regulated by the purine repressor, PurR, a DNA-binding protein, 341 amino acids in length [1–4]. In addition, PurR is responsible for the regulation of genes involved in pyrimidine biosynthesis and salvage [5,6] and for the autoregulation of the *purR* gene [7]. The DNA-binding ability of PurR is activated by binding a purine corepressor, either hypoxanthine or guanine [8], which leads to the repression of the *pur* regulon.

PurR is a member of the LacI family, that contains over 20 repressors, including the cytidine repressor [9], the galactose repressor [10], the maltose repressor [11], the raffinose repressor [12], and the fructose repressor [13]. Proteins of this family consist of two functional domains: an N-terminal DNA-binding domain (DBD) [14] and a larger C-terminal effector-binding/oligomerization domain (CBD) [15]. In accord with the two domain structure of LacI family members, PurR can be cleaved specifically, by limited trypsin and chymotrypsin proteolysis, between Arg52 and Ser53, and Leu54 and Lys55, respectively [15,16], thereby leaving most of the DBD and a slightly extended (about 10 amino acids longer) CBD (Fig. 1).

Recently, the three-dimensional (3D) structure of a PurR–hypoxanthine–*purF* operator site ternary complex in which a PurR dimer recognizes, with a twofold symmetry, the nearly palindromic *purF* operator sequence, has been determined by X-ray crystallography [17]. In the complex the PurR DBD was defined as the N-terminal 56 amino acids, which comprised four helices. The first and second helices form a helix-turn-helix (HTH) motif [14,18,19], which, along with the following loop, makes specific base contacts to the major groove of the DNA, in a manner similar to that observed for other classical HTH proteins [18–20]. Additionally, the fourth (hinge) helix (Ser48–Val56) lies deeply in the minor groove of the DNA, and contributes to specificity by making hydrogen bonds and van der Waals contacts to an adenine [17]. The study also proposed that the hinge helix is disordered in its DNA-free state and that helix formation is induced by specific DNA binding. Such a coil-to-helix transition is reminiscent of the conformational changes observed for the Jun/Fos heterodimer, as well as for GCN4 and other bZip proteins, in which recognition-helix formation is induced by specific DNA binding [21–26].

To clarify the structure of the DNA-free form of the PurR DBD, we have determined the solution structures of two PurR DBD by nuclear magnetic resonance (NMR) spectroscopy. One consists of N-terminal residues 1 to 56 and the other of residues 1 to 62 which includes the linker to the CBD (PurRN56 and PurRN62, respectively [27], Fig. 1.), and these have been

*Corresponding author.

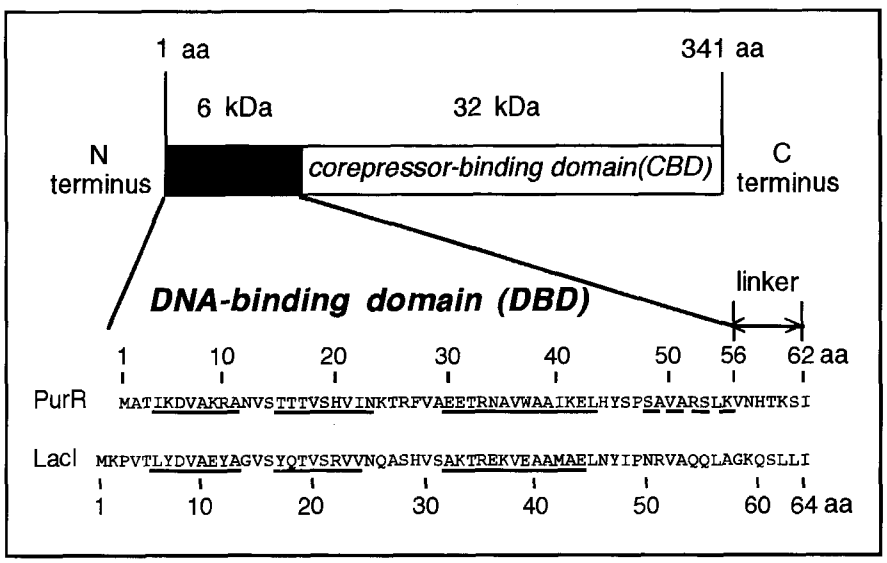

Fig. 1. Amino-acid sequence of the DNA-binding region of PurR. The amino-acid sequence of LacI was aligned, for comparison. The three underlined regions of PurR and LacI are helical regions both in the DNA-free and DNA-bound states. The dashed-underlined region of PurR is a hinge helical region only in PurR–DNA complex.

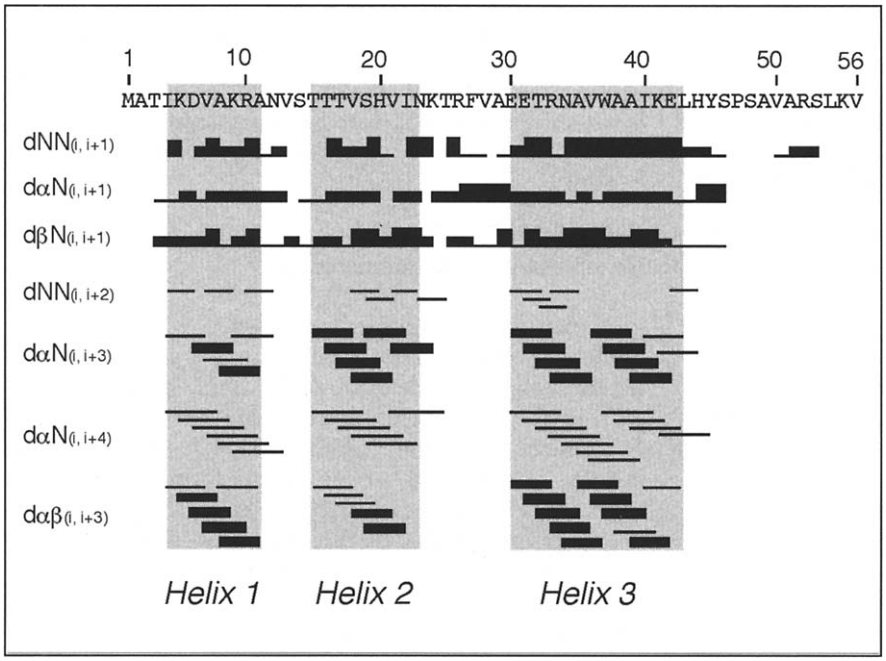

Fig. 2. Summary of the sequential and short-range NOEs observed for PurRN56. The thick, medium and thin bars correspond to strong, medium and weak NOEs, respectively. Helical regions are shaded.

compared with the X-ray structure of the DNA-bound form of PurR. PurRN56 was chosen because these residues were defined as the DBD in the ternary complex structure [17], and PurRN62, was studied to examine any C-terminal effects on the PurRN56 structure. After comparing the NMR spectra of both proteins, we propose that the structure of the PurRN56 is that adopted by the DBD in its DNA-free state.

Results and discussion

NMR spectroscopy of PurRN56 and PurRN62

The usual 2D NMR experiments, including double quantum filtered correlation spectroscopy (DQF-COSY) [28], total correlation spectroscopy (TOCSY) [29,30] and nuclear Overhauser enhancement spectroscopy (NOESY) [31] were carried out using PurRN56 or PurRN62 in 90% H_2O/10% D_2O and using PurRN56 in 99% D_2O. Nearly every signal in the NOESY spectra of PurRN62 could be identified with a corresponding signal in NOESY spectra of PurRN56, which indicates strongly that the structure of the PurRN56 portion of PurRN62 adopts the same conformation as PurRN56 alone. Residues Asn57–Ile62 of PurRN62 appear to be completely disordered, as even their sequential NOEs could not be obtained under the present condition.

Figure 2 shows the sequential and short-range NOE connectivities of PurRN56. Sequence-specific assignments could be completed from Thr3 to Ser48 and from Val50 to Arg52, and are summarized in Table 1. Residues 1 and 2 and 53 to 56 could not be identified in NOESY spectra, although five additional vague spin systems, that could not be assigned to specific residues, were observed in TOCSY spectra. The connectivities reveal that PurRN56 contains three helices (Ile4–Ala11, Thr15–Asn23, and

Table 1. ^{1}H NMR chemical shifts of PurRN56 in aqueous solution at pH 6.3, 27°C.

Residue*	Chemical shift (ppm)				Residue*	Chemical shift (ppm)			
	NH	αH	βH	Others		NH	αH	βH	Others
Thr3	9.51	4.84	5.07	$C\gamma H_3$ 1.40	Phe27	8.57	4.40	3.19, 2.90	^{2,6}H 7.21; ^{3,5}H 7.37; ^{4}H 7.28
Ile4	8.97	3.30	1.96	$C\gamma H_2$ 1.34, 1.04; $C\gamma H_3$ 0.93; $C\delta H_3$ 0.82	Val28	7.34	3.97	1.64	$C\gamma H_3$ 0.90, 0.88
					Ala29	8.49	4.13	1.63	
Lys5	7.60	3.86	1.79, 1.53	$C\gamma H_2$ 1.43, 1.25; $C\delta H_2$ 1.74, 1.52; $C\epsilon H_2$ 3.05	Glu30	9.04	3.84	2.21, 2.09	$C\gamma H_2$ 2.38
					Glu31	9.49	4.21	2.14	$C\gamma H_2$ 2.43
Asp6	7.48	4.38	3.02, 2.74		Thr32	7.16	4.09	4.16	$C\gamma H_3$ 1.14
Val7	7.45	3.22	1.90	$C\gamma H_3$ 0.69, 0.50	Arg33	8.51	3.74	1.96, 1.91	$C\gamma H_2$ 1.74, 1.48; $C\delta H_2$ 3.42; $N\epsilon H$ 7.28
Ala8	8.16	3.53	1.42						
Lys9	7.94	4.17	1.96, 1.93	$C\gamma H_2$ 1.40, 1.59; $C\delta H_2$ 1.77, 1.50; $C\epsilon H_2$ 3.05	Asn34	8.72	4.68	3.01, 2.97	$N\gamma H_2$ 7.73, 7.14
					Ala35	7.78	4.34	1.70	
Arg10	7.73	4.21	1.99, 1.32	$C\gamma H_2$ 1.69, 1.55; $C\delta H_2$ 3.43, 3.21; $N\epsilon H$ 8.16	Val36	8.31	3.38	2.12	$C\gamma H_3$ 0.87,0.56
					Trp37	8.73	4.62	3.40, 3.31	^{2}H 7.17; ^{4}H 7.70; ^{5}H 7.21; ^{6}H 7.32; ^{7}H 7.54; NH 10.04
Ala11	8.08	4.38	1.24						
Asn12	7.98	4.41	3.33, 2.68	$N\gamma H_2$ 7.55, 6.80	Ala38	8.42	4.32	1.61	
Val13	8.20	4.44	2.26	$C\gamma H_3$ 0.88, 0.64	Ala39	7.76	4.18	1.58	
Ser14	8.35	4.73	4.44, 4.07		Ile40	8.55	3.53	2.27	$C\gamma H2$ 1.87, 0.82; $C\gamma H_3$ 1.32; $C\delta H_3$ 0.32
Thr15	9.06	3.64	4.16	$C\gamma H_3$ 1.25					
Thr16	7.63	3.74	3.99	$C\gamma H_3$ 1.08	Lys41	8.09	4.23	2.10, 2.05	$C\gamma H_2$ 1.57; $C\delta H_2$ 1.80, 1.72; $C\epsilon H_2$ 3.07
Thr17	7.83	3.85	4.29	$C\gamma H_3$ 1.13					
Val18	7.56	3.25	2.17	$C\gamma H_3$ 0.79, 0.67	Glu42	8.46	4.02	2.23, 2.07	$C\gamma H_2$ 2.46, 2.26
Ser19	7.81	4.05	3.89		Leu43	8.04	4.33	1.72, 1.51	$C\gamma H$ 1.79; $C\delta H_2$ 0.91, 0.74
His20	8.25	4.89	3.22, 2.90	^{2}H 7.62; ^{4}H 6.93	His44	7.84	4.35	3.54, 3.41	^{2}H 8.48; ^{4}H 7.30
Val21	8.13	3.76	2.21	$C\gamma H_3$ 1.00, 0.61	Tyr45	8.20	4.50	2.79, 2.58	^{2,6}H 6.77; ^{3,5}H 6.63
Ile22	8.16	3.86	1.82	$C\gamma H_2$ 1.52; $C\gamma H_3$ 0.92; $C\delta H_3$ 0.81	Ser46	7.74	4.70	3.72, 3.66	
					Pro47		4.33	2.34, 2.12	$C\gamma H_2$ 2.21, 2.00; $C\delta H_2$ 3.78, 3.59
Asn23	8.62	4.87	2.95	$N\gamma H_2$ 7.50, 7.07					
Lys24	7.98	4.32	1.88, 1.76	$C\gamma H_2$ 1.61; $C\delta H_2$ 3.20	Ser48	8.28	4.65	2.76, 2.64	
Thr25	8.01	4.22	4.46	$C\gamma H_3$ 1.22	Ala49				
Arg26	7.40	4.30	1.14, 1.11	$C\gamma H_2$ 1.31, 1.18; $C\delta H_2$ 2.90, 2.86; $N\epsilon H$ 7.05	Val50	8.16	4.43	1.69	$C\gamma H_3$ 0.98, 0.92
					Ala51	7.73	4.10	1.34	
Arg52	8.21	4.43	1.90, 1.83	$C\gamma H_2$ 1.73, 1.50; $C\delta H_2$ 3.28;					ε-NH 7.21

*Residue 1 and 2 and 53 to 56 could not be identified in NOESY spectra. The atoms of the aromatic rings are numbered according to the IUPAC-IUB Joint Commission on Biological Nomenclature (JCBN) (1993).

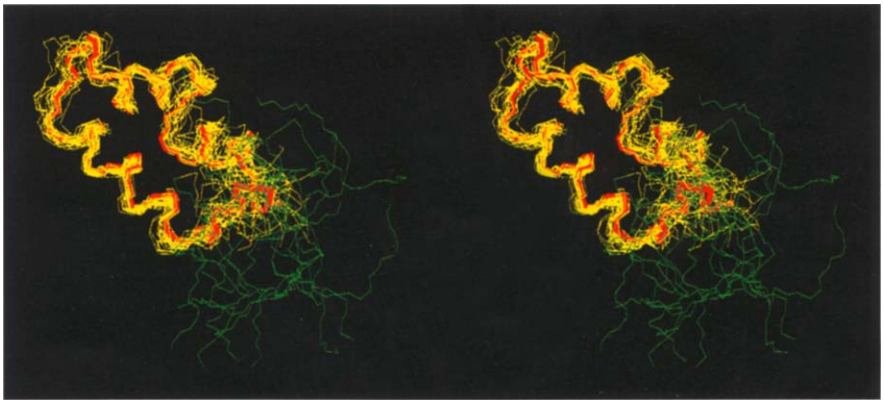

Fig. 3. Stereoview of the best-fit superpositions of the 20 PurN56 structures, along with the refined average structure. Met1–Pro47 is shown in yellow for the 20 structures and in red for the average structure, and Ser48–Val56 is green for the 20 structures. No average structure is shown for the C-terminal residues.

Glu30–Leu42) and that the conformation of C-terminal region (Tyr45–Val56) is not fixed (Fig. 2).

From the several NOESY spectra, we have obtained a set of 412 distance constraints: 187 sequential, 174 short range and 51 long range. According to these constraints, we have carried out distance geometry calculations using 4D simulated annealing (SA) [32], as reported previously [33–35].

Structure of the PurR DBD

The calculated 20 structures and the refined average structure of PurRN56 are shown in Figure 3. Figure 4 shows the root mean square (rms) deviation, for the back-

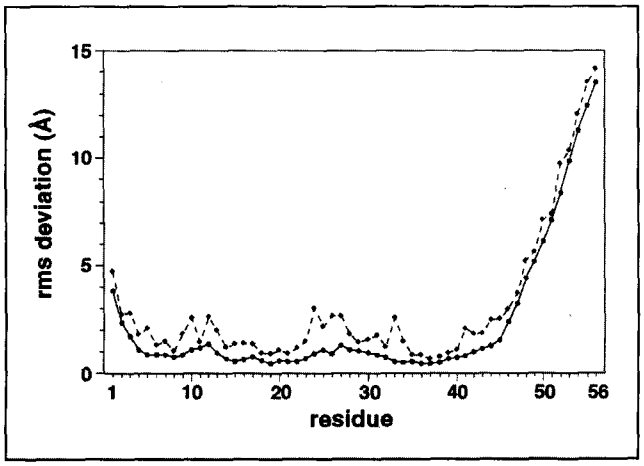

Fig. 4. The rms deviation, at each residue, between coordinates for the individual and the mean structures. Results for the backbone atoms are shown as a solid line and for the heavy atoms are shown as a dotted line.

bone-atoms and for all heavy atoms between the coordinates for individual and the mean structure. The conformations of N- and C- termini were less well defined. The rms deviations for Lys5 to Lys41, between the individual structures and the mean coordinates, are 0.84 ± 0.28 Å for the backbone atoms and 1.53 ±0.45 Å for all heavy atoms. These, and other relevant statistics, are summarized in Table 2. Although the N- and C-terminal residues are poorly defined, the three helical regions are well defined. The first and second helices form an HTH motif and the third helix lies nearly parallel to the first. The structures of the three helices are maintained by a hydrophobic core formed by residues Ile4, Val7, Ala8, Ala11, Val18, Val21, Ile22, Val28, Val36, Ala39, and Ile40 (Fig. 5), with residues Ile4 and Ala8 (from Helix 1) and Val18 (from Helix 2) stabilizing the HTH structure.

In most classical HTH proteins, a glycine residue is conserved in the first position of the HTH turn [14, 18–20]. In PurR, however, the conserved glycine is replaced by an asparagine residue (Asn12). The rms deviations between the HTH motifs of the 20 PurRN56 structures (Lys5–Asn23) and those of the HTH proteins LacI [20] and λ repressor [36], are 1.25 ± 0.25 Å and 1.13 ± 0.24 Å, respectively. Therefore, it seems that the 'conserved' glycine residue in the classical HTH can be replaced by asparagine residue without causing significant conformational changes.

Figures 3 and 4 reveal clearly that the structure of the hinge-helix region (Ser48–Val56) is completely disordered in the DNA-free form. This is not caused by the proximity of the terminal carboxylate (a negative charge) of Val56, because the structure of this region is also disordered in PurRN62, which contains six additional residues, Asn57 to Ile62, that link the DBD to the CBD.

Comparison with the DNA-bound DBD

Figure 5 shows the superimposition of the structures of the DNA-free DBD (this work) and the DNA-bound DBD (that was determined by X-ray crystallography [17]). The backbone architectures of the first three helices, together with the conformations of the hydrophobic core residues, are very similar between the free and DNA-bound forms. The rms deviation of the backbone atoms between the refined average structure of the DNA-free form (residues Lys5 to Lys41) and the

Table 2. Numbers of parameters and results of distance-geometry calculation for the DNA-binding domain of PurR.

Structural parameter	No. of parameters	SA_i* (rms deviations)	$\langle SA \rangle_{ref}$†
Rms distance deviations (Å)‡ total	412	0.0170 (0.0027)	0.0150
intraresidue	0	–	–
interresidue sequential (|i-j|=1)	187	0.0211 (0.0056)	0.0209
interresidue short range (1 < |i-j| ≤ 5)	174	0.0116 (0.0046)	0.0077
interresidue long range (5 < |i-j|)	51	0.0077 (0.0035)	0.0041
Deviations from idealized geometry			
bonds (Å)	936	0.005 (0.0002)	0.005
angles (°)	1692	0.659 (0.0204)	0.595
impropers (°)	164	0.367 (0.0354)	0.313
omega (°)	52	1.488 (0.144)	1.196
Energetic statistics (kcal mol^{-1})			
E_{dist}§	–	4.45 (1.21)	3.53
E_{L-J}	–	–203.32 (7.34)	-230.48
E_{total}#	–	–259.75 (28.54)	-364.00

*The average values for the variables obtained from the final 20 refined structures. Numbers in parentheses are the standard deviations. †$\langle SA \rangle_{ref}$ represents the average structure of the final 20 structures. They were least-square fitted each other using the coordinates of residues Lys5 to Lys41 because of the disordered structures of the N- and C-termini (see the text). ‡The maximum value of the distance violation for the final 20 structure is 0.299 Å. The average value, ± the standard deviation of the maximum distance violations for the individual 20 structures, is 0.201 ± 0.042 Å. §E_{dist} is the distance restraints term with a weight of 0.6 kcal $mol^{-1}Å^4$, and E_{L-J} is the non-bonded Lennard-Jones interaction term using AMBER all-atom force field [45]. #E_{total} is the total energy including all the distance restraints and AMBER al-atom force field with the electrostatic energy (see text).

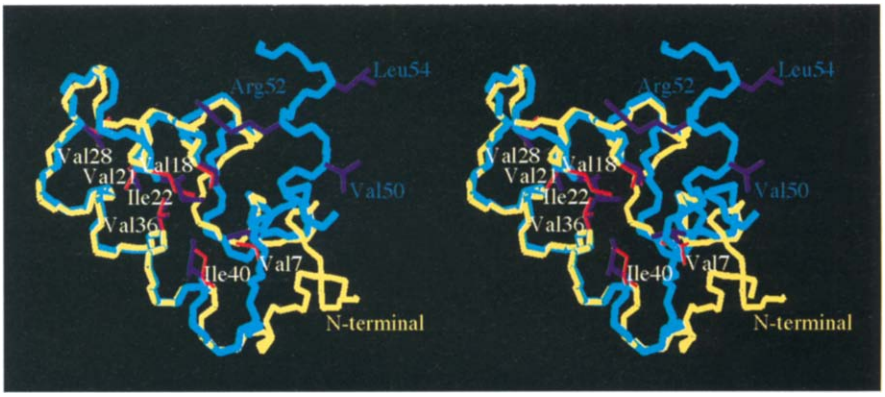

Fig. 5. Stereoview of a best-fit superposition of the free and DNA-bound structures of the PurR DBD. The side chains of amino acids forming the hydrophobic core (except alanines) are indicated; the side chains of Val50, Arg52 and Leu54 in the DNA-complexed DBD are also shown. The backbone and side-chain atoms of the DNA-free DBD (Met1–Ser48) are shown by yellow and red lines, respectively, and those of the DNA-bound DBD (Met1–Val56) are shown by light-blue and dark-blue lines, respectively.

corresponding atoms of the DNA-bound structure is 1.03 Å; once the side chain heavy atoms of the hydrophobic core residues are included the value increases to 1.11 Å. In contrast, two hydrophobic residues in the hinge-helix region, Val50 and Leu54, do not adopt a rigid conformation in the DNA-free form, and are likely to be solvent-exposed (Fig. 3). In the PurR–DNA complex, these two residues participate in the dyadic hydrophobic interactions between hinge helices [17]. Gel filtration indicates that the free PurR DBD (PurRN56) behaves as a monomer rather than a dimer. These direct interactions between the hinge helices, in the PurR-DNA complex are therefore likely to be induced by specific DNA-binding and by the dimerization of the CBD.

DNA-binding experiments of PurR DBDs

We have examined the DNA-binding activities of PurRN56 and PurRN62. Figure 6 shows electrophoretic mobility-shift assays for both proteins. Even at concentrations of 5 μM, neither protein shows a singularly shifted band, which is evidence that both PurRN56 and PurRN62 cannot bind specifically to the *purF* operator. These data suggest that, for the PurR DBD, the HTH motif alone is not sufficient for specific DNA binding and that the hinge-helical region of the isolated DBD is disordered even in the presence of specific DNA. Therefore, hinge-helix formation requires PurR dimerization, which requires the CBD, and is induced by specific DNA binding. This supports the earlier suggestion that hydrogen bonds from the CBD, which requires that PurR is dimeric, are critical for hinge-helix formation [17]. If this was not the case, the 'dimer' interface between the hinge helices would probably be sufficient to bring about DBD dimerization and therefore specific DNA binding.

Hinge-helix formation

The structure of the PurR DBD in its DNA-free form consists of only three helices, that form a small globular domain and a disordered structure. The fourth helix, the hinge helix, was only observed in the PurR–hypoxanthine–*purF*-operator complex [17]. Figure 7 shows the superposition of the DBD structure in its DNA–free form onto the ternary complex structure. Our results, together with earlier experiments that showed specific proteolytic cleavage at Arg52 and Leu54 of the intact PurR in its free state [15,16] suggest that the structures of hinge-helix regions of the PurR dimer in the DNA-free state are disordered and without sustained direct interactions between the two regions. In addition, the specific cleavage of the hinge helical region by trypsin was found to be largely inhibited by adding specific, but not non-specific, DNA [16]. Hinge-helix formation is the likely product of both specific DNA binding and dimerization of the CBD. The dimer of intact PurR in the DNA-free state is stabilized by the bridge, of a large surface area, formed by the CBD. In the DNA complex, the two hinge helices of the PurR dimer contact each other directly and are stabilized, in part, by hydrophobic interactions [17]. These findings lead to a model in which the PurR dimer binds specific DNA in a two step cooperative manner; first, each HTH of the

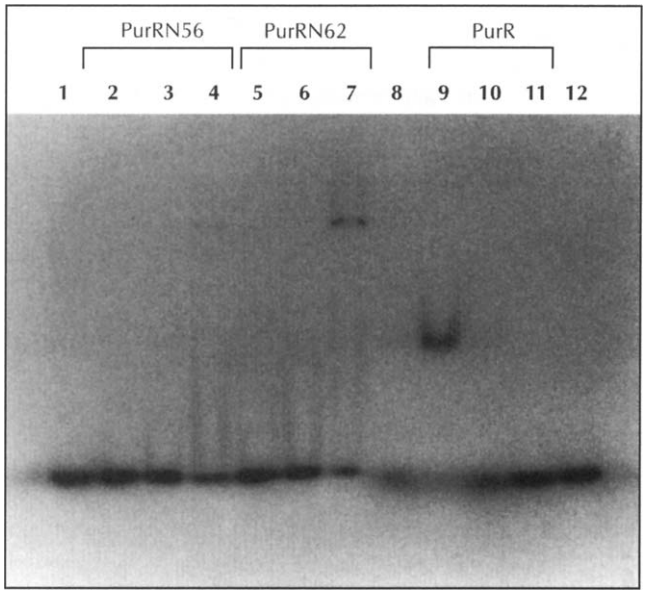

Fig. 6. Electrophoretic mobility-shift assays for PurRN56, PurRN62 and intact PurR. Electrophoretic mobilities of labeled *purF* operator DNA, a 30mer, are shown following additions of various different proteins: in lane 1 no protein was added; lanes 2, 3 and 4 represent 50 nM, 500 nM and 5 μM of PurRN56, respectively; lanes 5, 6 and 7 represent 50 nM, 500 nM and 5 μM PurRN62, respectively; lane 8 represents *E. coli* (MC4100) lysate; lanes 9, 10 and 11 represent *E. coli* (MC4100PurR) lysate with 0 pmole, 10 pmole and 100 pmole non-labeled *purF* operator DNA, respectively; and lane 12 represents the lysate of a *purR*-deleted *E. coli* mutant.

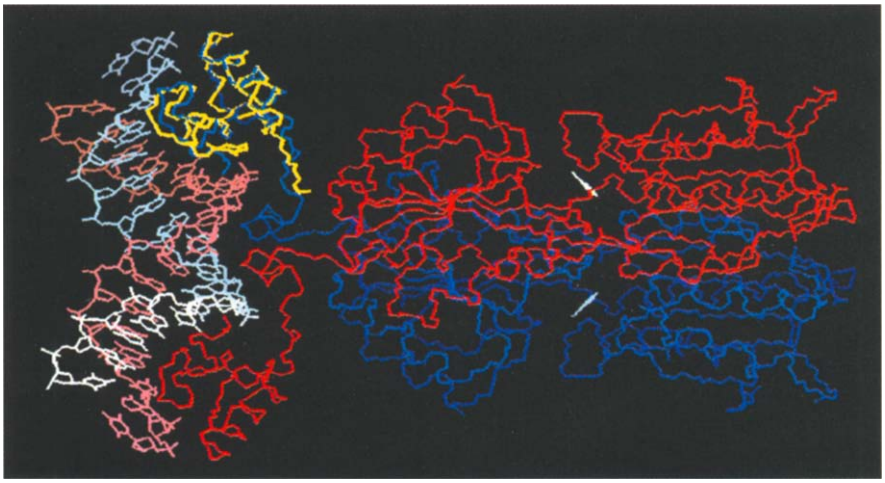

Fig. 7. A best-fit superposition of the free PurR DBD structure onto the ternary complex structure of the PurR–hypoxanthine–*purF*-operator. The backbone atoms of the free PurR DBD (Met1–Ser48) are shown in yellow, and the backbone atoms of PurR dimer in the ternary complex are shown in blue and red, the two colours distinguishing the two monomers. The DNA atoms are shown in pink and white. Two hypoxanthine molecules bound to PurR dimer are also shown in white.

dimer recognizes, via the major groove, one half-site, and second, the two hinge regions of the dimer, which now would be located near the DNA minor groove, fold to become α helices that are stabilized by mutual hydrophobic interactions, as well as by key hydrogen bonds from the CBD. The formation of the hinge helices is the most probable cause of DNA-bending, as the side chains of Leu54 and Leu54′ can partially interdigitate between the bases of the central CpG step. Furthermore, the hinge helices are critical to specific DNA binding, as Lys55 makes a specific base contact in the minor groove [17]. The tight binding between specific DNA and PurR is then complete.

Comparison with LacI

As described previously, the LacI DBD is composed of an HTH motif that is followed by a third helix [20]. Residues of the LacI DBD are disordered beyond this helix. The conservation, in LacI, of key hinge-helix residues, Val52, Ala53 and Leu56, (corresponding to PurR residues Val50, Ala51 and Leu54) supports the concept [17,37] that hinge-helix formation, and also the minor-groove binding by this helix, are likely to occur in LacI–DNA complexes as well. Additionally, PurR hinge-helix residues, Arg52 and Ser48, that interact with the phosphate backbone, are replaced by Gln54 and Asn50 in LacI, both of which could also interact with DNA phosphate groups. There is, however, still a small but significant difference between the two repressors. In the PurR–DNA complex, Lys55 from the hinge helix recognizes a specific base in the minor groove [17]. The corresponding amino acid in LacI is Ala57, which, from modeling studies, cannot distinguish specific base pairs. Thus, the hinge-helix formation in LacI does not appear to be responsible for specific recognition of DNA base pairs through minor groove interactions. Instead it seems to increase DNA binding through 'nonspecific' sugar–phosphate interactions and to induce DNA bending via the intercalation of two Leu56 residues. In contrast to the PurR DBD, the LacI DBD can bind to DNA in a sequence-specific manner [20, 38–40] and appears to require only the first three DBD helices [20]. This might be due to the DNA-binding ability of the LacI HTH. In comparison with the PurR–DNA complex, more residues of the LacI HTH are involved in major-groove interactions, which might preclude the necessity for any specific interactions with the minor groove [20]. For PurR, specific hinge-helix recognition in the minor groove seems to be essential for the specific binding.

Biological implications

The primary function of purine repressor (PurR) of *Escherichia coli* is to regulate the transcription of genes which encode enzymes for *de novo* purine biosynthesis. PurR, a member of the LacI family, has a two-domain structure, comprising an N-terminal DNA-binding (DBD) domain and a larger C-terminal corepressor-binding/dimerization domain (CBD). The PurR dimer binds to palindromic sequences such that the complex displays nearly twofold symmetry. In the complex with DNA, the PurR DBD consists of four helices. The first and second helices form a helix-turn-helix (HTH) motif and bind to the major groove of DNA, and the fourth helix, the hinge helix, recognizes a specific base pair via a hydrogen bond in the minor groove.

In the DNA-free form, the PurR DBD contains only three helices and the structure in the region of the fourth helix is disordered. In this form it is unable to bind specifically to DNA. Hinge-helix formation, induced by a specific DNA binding and requiring the dimerization of the CBD, seems to be essential for specific DNA recognition by PurR. In the PurR–DNA complex, the hinge helix from each dimer is stabilized, in part, by mutual hydrophobic interactions between amino-acid residues within these helices, as well as by hydrogen bonds from the CBD. The spatial arrangement of the two hinge helices is likely to be controled by the dimerization interface of the CBD and removal of corepressor from PurR should alter this interface. Such an anticipated rearrangement of the two

DNA-binding domains of PurR would lead to the dissolving of the hinge helices by loss of the stabilizing van der Waals interactions between them. This would then lead to release of the PurR dimer from the specific operator site.

Materials and methods

Sample preparation

The plasmids, which express PurRN56 or PurRN62 proteins, were constructed by using a modified T7 expression vector, pAR2156NcoI [41], in which an Nco I site was introduced, by site-specific mutagenesis [42], into a site overlapping the ATG initiation codon. The Nco I fragments encoding either PurRN56 or PurRN62, were prepared by oligonucleotide-directed mutagenesis [42] and cloned into the Nco I site of pAR2156NcoI. A fresh overnight preculture of *E. coli* BL21(DE3), harboring one of the expression plasmids described above, was diluted into a 3 liter superbroth medium [43] containing ampicillin (400 mg l^{-1}) and was incubated at 37 °C. After the OD_{600} reached about 1.0, isopropyl-β-thiogalactoside (IPTG) was added to a final concentration of 0.24 g l^{-1} and the bacterial growth was continued for 3 h. Three liters of cultivated cells were harvested, washed and resuspended in phosphate buffer (100 mM KPB pH 7.1, 0.5 mM EDTA, 0.1 mM PMSF). The cells were lysed, on ice, by sonication and were centrifuged at 48400 *g* for 90 min. The supernatant was loaded on a CM cellulose column (CM52, Whatman), equilibrated with the phosphate buffer and eluted using the same phosphate buffer. Fractions were collected, and precipitated by adding ammonium sulfate. The precipitant was resuspended with a small volume of phosphate saline buffer (1 M NaCl, 50 mM KPB pH 6.3) such that sample volume was below 2 ml. The sample solution was then loaded on a column of Superdex 75pg (Pharmacia) and sample fractions were collected. The fractions were concentrated by an ultrafiltration concentrator, (Centriprep, Amicon). A final buffer exchange was carried out by a ultrafiltration concentrator and the proteins were stored in 10 mM KPB pH 6.3, 200 mM KCl, and 0.5 mM NaN_3.

NMR measurements

All NMR spectra were recorded at 500 MHz on a Bruker AMX-500 spectrometer. The temperature during data acquisition was set to 300 K. Quadrature detection was made by the TPPI method. For water signal suppression, weak presaturation was used. DQF-COSY spectra [28], NOESY spectra [31] with mixing times of 50 ms and 150 ms, and TOCSY spectra [29] with a mixing time of 100 ms, were recorded and relaxation-compensated DIPSI-2 mixing schemes with z filtration [30] were used in the pulse sequence of the TOCSY experiments. In 90 % H_2O/10% D_2O solution, each measurement was performed with 2048 real data points along the t2 dimension over 8064 Hz spectral width, and 768 and 640 points along the t1 dimension for PurRN56 and PurRN62, respectively. In D_2O solution of PurRN56, each measurement was performed with 2048 real data points along the t2 dimension over 5263 Hz spectral width and 512 points along t1 dimension. After data acquisition, zero filling along t1 dimension was performed to produce 1024 datapoints in each experiment.

Structure calculations

Interproton distance constraints were derived from the cross-peak intensities of the NOESY spectra with mixing times of 50 ms and 150 ms, using assumptions similar to those of previous calculations [33–35]. From the NOE intensities, the distances between backbone protons were classified into three ranges: 2.0–3.0, 2.0–4.0 and 2.0–5.0 Å, corresponding to strong, medium and weak NOEs, respectively. The interproton distances involving side-chain protons were classified into two ranges, 2.0–4.0 Å and 2.0–5.0 Å, corresponding to strong and medium, and weak NOEs.

Structures were constructed from 100 random-coil conformations using the 4D simulated-annealing program EMBOSS [32] with the 412 experimentally derived distance constraints, as reported previously [33–35]. Out of 100 calculated structures, 30 structures were selected with no individual distance violation larger than 0.1 Å. Each of 30 conformations was energy minimized by 5000 conjugate gradient steps with the program PRESTO [44] and was subject to all of the distance restraints and AMBER all atom force field [45]. The electrostatic interactions were included, with a dielectric constant $2r_{ij}$, for the nonbonded atoms i and j, where r_{ij} is the distance between the i and j atoms. Finally, the 20 structures were selected were those without any individual distance violation larger than 0.3 Å, or total energy larger than 190 kcal mol^{-1}.

Electrophoretic mobility-shift assays

Both strands of a duplex DNA containing the *purF* operator, with a sequence of AATCCCTACGCAAACGTTTTCTTT-TTCTGT, were synthesized with a DNA synthesizer, ABI381A (Applied Biosystems). Both strands were mixed in an annealing buffer (20 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 1mM DTT) and the solution was heated to 70 °C, then slowly cooled to room temperature. The duplex DNA was purified by non-denaturing polyacrylamide gel electrophoresis, phosphorylated using [γ^{32}P] ATP and T4 polynucleotide kinase, and the labeled duplex was then precipitated by adding ethanol. Electrophoretic mobility-shift assays were performed for purified PurRN56 and PurRN62 proteins, and for the cell lysate of an *E. coli* strain (MC4100PurR), which contains the pTZ18RpurR plasmid that encodes the intact PurR protein. As controls, the cell lysate of an *E. coli* strain (MC4100) which does not contain the plasmid, as well as the cell lysate of a *purR* deleted mutant (MC4100Δ*purR*), were examined under the same conditions. The binding of each protein to 50 fmole of labeled DNA was carried out in 50 mM Tris-HCl pH 7.4, 70 mM KCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 7 mM $MgCl_2$ and 200 μg ml^{-1} BSA. The electrophoretic mobility-shift assays were done using non-denaturing, 6 % polyacrylamide gel electrophoresis.

The coordinates of the PurR DBD have been deposited in the Protein Data Bank as 1PRU for the average structure and as 1PRV for the 20 structures. The restraint data of the PurR DBD structure calculations was deposited as R1PRVMR.

Acknowledgements: We thank Yasuhiro Yamamoto for help with the preparation of PurRN56 sample. This work was supported by Grants in Aid for Scientific Research on Priority Areas (06276103 and 06270104 to YN) from the Ministry of Education, Science, and Culture of Japan and by grants from the National Institutes of Health (GM 49224 to RGB) and the Department of Defense (to MAS).

References

1. He, B., Shiau, A., Choi, K.Y., Zalkin, H. & Smith, J.M. (1990). Genes of *Escherichia coli pur* regulon are negatively controlled by repressor–operator interaction. *J. Bacteriol.* **172**, 4555–4562.
2. He, B. & Zalkin, H. (1992). Repression of *Escherichia coli purB* is by a transcriptional roadblock mechanism. *J. Bacteriol.* **174**, 7121–7127.

3. He, B. & Zalkin, H. (1994). Regulation of *Escherichia coli purA* by purine repressor, one component of a dual control mechanism. *J. Bacteriol.* **176**, 1009–1013.
4. Meng, L.M., Kilstrup, M. & Nygaard, P. (1990). Autoregulation of PurR repressor synthesis and involvement of *purR* in the regulation of *purB, purC, purL, purMN,* and *guaBA* expression in *Escherichia coli*. *Eur. J. Biochem.* **187**, 374–379.
5. Choi, K.I. & Zalkin, H. (1990). Regulation of *Escherichia coli pyrC* by the purine regulon repressor protein. *J. Bacteriol.* **172**, 3201–3207.
6. Wilson, H.R. & Turnbough, C.L. Jr. (1990). Role of the purine repressor in the regulation of pyrimidine gene expression in *Escherichia coli* K-12. *J. Bacteriol.* **172**, 3208–3213.
7. Rolfes, R.J. & Zalkin, H. (1990). Autoregulation of *Escherichia coli purR* requires two control sites downstream of the promoter. *J. Bacteriol.* **172**, 5758–5766.
8. Rolfes, R.J. & Zalkin, H. (1990). Purification of *Escherichia coli* purine regulon repressor and identification of corepressor. *J. Bacteriol.* **172**, 5637–5642.
9. Valentin-Hansen, P., Larsen, J.E.L., Hojrup, P., Short, S.A. & Barbier, C.S. (1986). Nucleotide sequence of CytR regulatory gene of *E. coli* K-12. *Nucleic Acids Res.* **14**, 2215–2228.
10. von Wiclken-Bergmann, B. & Muller-Hill, B. (1982). Sequence of *galR* gene indicates a common evolutionary origin of *lac* and *gal* repressor in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **79**, 2427–2431.
11. Redl, J., Romish, M., Ehrmann, M. & Boos, W. (1989). Mall, a novel protein involved in regulation of the maltose system of Escherichia coli, is highly homologous to the repressor proteins GalR, CytR, and LacI. *J. Bacteriol.* **171**, 4888–4899.
12. Aslanidis, C. & Schmitt, R. (1990). Regulatory elements of the raffinose operon: nucleotide sequences of operator and repressor genes. *J. Bacteriol.* **172**, 2178–2180.
13. Jahreis, K.P., Postma, W. & Lengeler, J.W. (1991). Nucleotide sequence of the *ilvH-fruR* gene region of *Escherichia coli* K12 and *Salmonella typhmurium* LT-2. *Mol. Gen. Genet.* **226**, 332–336.
14. Brennan, R.G. & Matthews, B.W. (1989). The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264**, 1903–1906.
15. Choi, K.Y. & Zalkin, H. (1992). Structural characterization and corepressor binding of the *Escherichia coli*. *J. Bacteriol.* **174**, 6207–6214.
16. Choi, K.Y. & Zalkin, H. (1994). Role of the purine repressor hinge sequence in repressor function. *J. Bacteriol.* **176**, 1767–1772.
17. Schumacher, M.A., Choi, K.Y., Zalkin, H. & Brennan, R.G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science* **266**, 763–770.
18. Harrison, S.C. & Aggarwal, A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* **59**, 933–969.
19. Pabo, C.O. & Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.
20. Chuprina, V.P., *et al.*, & 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.
21. Patel, L., Abate, C. & Curran, T. (1990). Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347**, 572–574.
22. Weiss, M.A. (1990). Thermal unfolding studies of a leucine zipper domain and its specific DNA complex: implications for Scissor's grip recognition. *Biochemistry* **29**, 8020–8024.
23. Weiss, M.A., Ellenberger, T., Wobbe, C.R., Lee, J.P., Harrison, S.C. & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, 575–578.
24. Saudek, V., Pasley, H.S., Gibson, T., Gausepohl, H., Frank, R. & Pastore, A. (1991). Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry* **30**, 1310–1317.
25. O'Neil, K.T., Schuman, J.D., Ampe, D. & DeGrado, W.F. (1991). DNA-induced increase in the α-helical content of C/EBP and GCN4. *Biochemistry* **30**, 9030–9034.
26. Santigo-Rivera, Z.I., Williamas, J.S., Gorenstein, D.G. & Andrisani, O.M. (1993). Bacterial expression and characterization of the CREB bZip module-circular dichroism and 2D H-1-NMR **[AU:OK?]** studies. *Protein Sci.* **2**, 1461–1471.
27. Rolfes, R.J. & Zalkin, H. (1988). Escherichia coli gene *purR* encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence, and interaction with the *purF* operator. *J. Biol. Chem.* **263**, 19653–19661.
28. Rance, M., *et al.*, & Wuthrich, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **171**, 479–485.
29. Bax, A. & Davis, D.C. (1985). MLEV-17 based two dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360.
30. Griesinger, C., Otting, G., Wuthrich, K. & Ernst, R.R. (1988). Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870–7872.
31. Macura, S. & Ernst, R.R. (1980). Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.* **41**, 95–117.
32. Nakai, T., Kidera, A. & Nakamura, H. (1993). Intrinsic nature of the three dimensional structure of proteins as determined by distance geometry with good sampling properties. *J. Biomol. NMR.* **3**, 19–40.
33. Ogata, K., *et al.*, & Nishimura,Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl. Acad. Sci. USA.* **89**, 6428–6432.
34. Ogata, K., *et al.*, & Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* **79**, 639–648.
35. Ogata, K., *et al.*, & Nishimura,Y. (1995). Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nat. Struct. Biol.* **2**, 309–320.
36. Jordan, S.R. & Pabo, C.O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. *Science* **242**, 893–899.
37. Sauer, R.T. (1995). Minor groove DNA-recognition by α helices. *Nat. Struct. Biol.* **2**, 7–9.
38. Geisler, N. & Weber, K. (1977). Isolation of the amino-terminal fragment of lactose repressor necessary for DNA binding. *Biochemistry* **16**, 938–943.
39. Ogata, R.T. & Gilbert, W. (1979). DNA-binding site of lac repressor probed by dimethylsulfate methylation of lac operator. *J. Mol. Biol.* **132**, 709–728.
40. Khoury, A.M., Nick, H.S. & Lu, P. (1991). *In vivo* interaction of *E. coli.* repressor N-terminal fragment with the lac operator. *J. Mol. Biol.* **219**, 623–634.
41. Tanikawa, J., *et al.*, & Sarai, A. (1993). Recognition of specific DNA sequences by the c-myb proto-oncogene product: Role of three repeat units in the DNA-binding domain. *Proc. Natl. Acad. Sci. USA.* **90**, 9320–9324.
42. Kunkel, T.A., Roberts, J.D. & Zakour, R.A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.
43. Ramsay, R.G., Ishii, S., Nishina, Y., Soe, G. & Gonda,T.J. (1989). Characterization of alternate and truncated forms of murine c-myb proteins. *Oncogene Res.* **4**, 259–269.
44. Morikami, K., Nakai, T., Kidera, A., Saito, M., & Nakamura, H. (1992). PRESTO: a vectorized molecular mechanics program for biopolymers. *Comput. Chem.* **16**, 243–248.
45. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986). All atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**, 230–252.

Received: **3 Aug 1995**; revisions requested: **29 Aug 1995**; revisions received: **11 Sept 1995**. Accepted: **12 Sept 1995**.