J. Mol. Biol. (1998) **277**, 621–634

JMB

The Solution Structure of the Leu22 → Val Mutant AREA DNA Binding Domain Complexed with a TGATAG Core Element Defines a Role for Hydrophobic Packing in the Determination of Specificity

Mary R. Starich[1], Mats Wikström[1], Silke Schumacher[1] Herbert N. Arst Jr[2], Angela M. Gronenborn[1]* and G. Marius Clore[1]*

[1]*Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases National Institutes of Health Bethesda, MD 20892-0520 USA*

[2]*Department of Infectious Diseases and Bacteriology Imperial College School of Medicine at Hammersmith Hospital, Du Cane Road London W12 ONN, UK*

**Corresponding authors*

The seemingly innocuous leucine-to-valine mutation at position 22 of the AREA DNA binding domain results in dramatic changes in the *in vivo* expression profile of genes controlled by this GATA transcription factor. This is associated with a preference of the Leu22 → Val mutant for **T**GATAG sites over (**A**/**C**)GATAG sites. Quantitative gel retardation assays confirm this observation and show that the Leu22 → Val mutant AREA DNA binding domain has a ~30-fold lower affinity than the wild-type domain for a 13 base-pair oligonucleotide containing the wild-type **C**GATAG target. To gain insight into the measured affinity data and further explore sequence specificity of the AREA protein, the solution structure of a complex between the Leu22 → Val mutant AREA DNA binding domain and a 13 base-pair oligonucleotide containing its physiologically relevant **T**GATAG target sequence has been determined by multidimensional nuclear magnetic resonance spectroscopy. Comparison of this structure with that of the wild-type AREA DNA binding domain complexed to its cognate **C**GATAG target site shows how subtle changes in amino acid side-chain length and hydrophobic packing can affect affinity and specificity for GATA-containing sequences, and how changes in DNA sequence can be compensated for by changes in protein sequence.

Keywords: AREA; GATA transcription factors; L22V mutant; NMR structure; zinc finger

Introduction

The AREA transcription factor, from *Aspergillus nidulans*, possesses a single class IV, Cys_2Cys_2 zinc finger DNA binding domain (DBD) which serves as a primary regulator of nitrogen metabolism *via* recognition of **H**GATA**R** (**H**= A, T, C; **R**= A, G) promoter elements (Starich *et al.*, 1998 and references therein). Lowered intracellular levels of L-glutamine trigger AREA activity, which directs the expression of more than 100 structural genes associated with catabolism of alternative nitrogen sources (Arst & Cove, 1973; Kudla *et al.*, 1990; Wiame *et al.*, 1985). Extensive genetic characterization of naturally selected mutations observed for AREA has identified residues in the DNA that are critical for normal protein function (Arst and Cove 1973; Hynes 1975; Kudla *et al.*, 1990; Langdon *et al.*, 1995; Platt *et al.*, 1996a,b; Wiame *et al.*, 1985). Among these mutations is an apparently conservative substitution of valine for a leucine residue at position 22 of the AREA DBD. This single substitution

Present address: M. Wikström, Pharmacia & Upjohn, Department of Structural Chemistry, S-112 87 Stockholm, Sweden.

The first two authors (M.R.S. and M.W.) contributed equally to this work.

Abbreviations used: DBD, DNA binding domain; wtAREA, wild-type AREA; mutAREA, Leu22 → Val mutant AREA; NOE, nuclear Overhauser enhancement; HSQC, heteronuclear single quantum coherence; 3D, three-dimensional.

0022–2836/98/130621–14 $25.00/0/mb971626

generates a striking phenotypic response in the expression profile of genes under AREA control, profoundly increasing expression of the *amd*S gene (encoding acetamidase) and nearly eliminating expression of the *uap*A and *uap*C genes (encoding uric acid-xanthine permease and purine permease, respectively; Arst & Scazzocchio, 1975; Diallinas *et al.*, 1995; Gorfinkiel *et al.*, 1993; Hynes, 1975; Scazzocchio & Arst, 1978). Consequently, the abnormally low levels of the *uap*A transcript result in the inability of *A. nidulans* to grow on uric acid or xanthine as its primary nitrogen source.

Cis-acting mutations which change the target *uap*A promoter element from **C**GATAG to **T**GATAG restore growth on uric acid and xanthine, effectively suppressing the Leu22 → Val mutation (Ravagnani *et al.*, 1997). The Leu22 → Val mutation affects specific recognition of the first base of the HGATAR sequence, effecting preferential binding to **T**GATAR targets, but not to **C**GATAR or **A**GATAR targets (Ravagnani *et al.*, 1997). Further, characterization of the promoter region for the *amd*S gene uncovered three **T**GATAR targets and suggests a physiologically relevant role for the most downstream site in the presence of mutant AREA (Hynes *et al.*, 1988; Ravagnani *et al.*, 1997).

Here we report equilibrium dissociation constants which indicate that the 66-residue Leu22 → Val mutant AREA DNA binding domain (mutAREA DBD) prefers **T**GATAR targets, and demonstrates ~30-fold lower affinity for a 13 base-pair (bp) oligonucleotide containing the single **C**GATAG site as compared to the wtAREA DBD. To gain insight into the measured affinity difference and to understand better the mechanisms underlying sequence-specific recognition for this class of zinc finger proteins, we have solved the three-dimensional solution structure of the mutAREA DBD complexed with a 13 bp oligonucleotide containing its physiologically relevant **T**GATAG target sequence using multidimensional nuclear magnetic resonance (NMR). The results provide a direct view of how minor changes in side-chain length and base character affect local structure. Comparison of this structure to the wtAREA DBD·**C**GATAG complex (Starich *et al.*, 1998) suggests how hydrophobic contacts and steric effects play a critical role in determining the specificity and affinity of the AREA protein for HGATAR sequence elements.

Results and Discussion

mutAREA DBD affinity for GATA sites

Previous affinity studies carried out for the AREA system utilized His fusion proteins containing a 262-residue AREA construct and a series of 52 bp probes containing the GATA element in the context of double sites (Ravagnani *et al.*, 1997). Prior to proceeding with the structure determination of the mutAREA DBD·DNA complex, we wished to confirm that the shorter

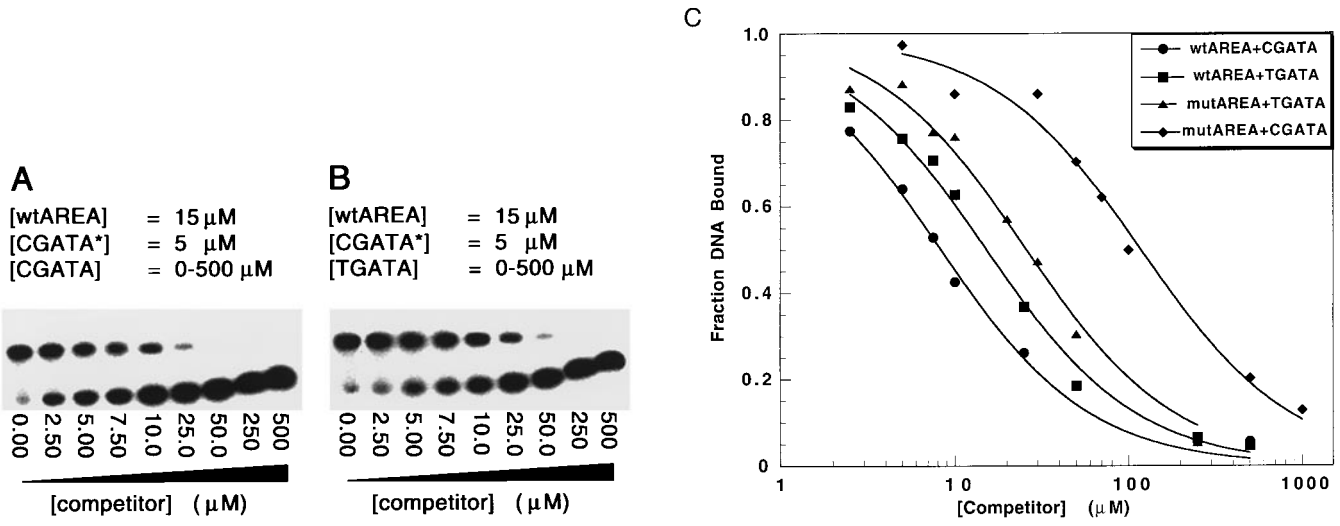

Figure 1. Binding of wild-type and Leu22 → Val mutant AREA DBDs to **C**GATA and **T**GATA containing sequences. A, Autoradiograph of a homologous competition gel retardation assay in which constant concentrations of wtAREA DBD (15 μM) and radiolabeled DNA (**C**GATA*; 5μM) were titrated with varying amounts of unlabeled **C**GATA competitor (0 to 500 μM, lanes 1 to 9, respectively). B, Autoradiograph of a heterologous competition experiment in which constant concentrations of wtAREA DBD (15 μM) and radiolabeled DNA (**C**GATA*; 5μM) were titrated with varying amounts of unlabeled **T**GATA competitor (0 to 500 μM, lanes 1 to 9, respectively). C, Plots of fraction radiolabeled DNA bound as a function of DNA competitor concentration for four AREA complexes tested. The best fit curves to the experimental data, obtained by non-linear least-squares optimization as described in Materials and Methods, are shown as continuous lines. The two double-stranded 13 bp oligonucleotides employed are 5′d[CAG**C**GATAGAGAC].5′d[GTCTCTATC**G**CTG] and 5′d[CAG**T**GATAGAGAC].5′d[GTCTCTATC**A**CTG].

66-residue constructs chosen for NMR studies (Starich *et al*., 1988) mimicked the qualitative results obtained previously. More importantly, quantification of the equilibrium dissociation constants for the wtAREA DBD and mutAREA DBD would reveal any significant difference in the affinity of these domains for the **C**GATAG core element. With the exception of a Leu22 → Val mutation, the 66 amino acid construct chosen for study was identical with that described for the wtAREA DBD in the accompanying paper (cf. Figure 1 of Starich *et al*., 1998). Two 13 bp target sequences, 5′-d[CAG**C**GATAGAGAC] (wtDNA) and 5′-d[CAG**T**GATAGAGAC] (mutDNA), were chosen for competitive gel-retardation assays. A representative pair of gels run for the radiolabeled wtAREA DBD·**C**GATA* complex indicates that the unlabeled **C**GATA site completely outcompetes the labeled CGATA* site at a lower concentration (~50 μM; Figure 1A) than the unlabeled **T**GATA site (~250 μM; Figure 1B). Analogous results were observed for the mutAREA DBD·**T**GATA* complex for which the unlabeled **T**GATA site was the most effective inhibitor (data not shown). Non-linear least-squares analysis of the data based on a weak-binding model (Swillens, 1995) yielded the best-fit curves depicted in Figure 1C. The equilibrium dissociation constants (K_D) of the wtAREA DBD for both **C**GATA and **T**GATA sites are very similar (3.1(±1.2) μM and 5.8(±0.8) μM, respectively); the K_D for the interaction of the mutAREA DBD and its cognate site is also in the low μM range (20(±3) μM). Interestingly, the K_D for the mutAREA DBD·**C**GATA complex (96(±13) μM) is substantially different, revealing a ~30-fold weaker binding to **C**GATA sites for the mutant relative to the wild-type DBD.

The equilibrium dissociation constants obtained here are consistent with recently published work which demonstrates that altering the residue at position 22 produces observable differences in the relative affinities for HGATAR (H = A, T, C; R = A, G) sites (Ravagnani *et al*., 1997). Specifically, our results confirm and quantify the preference of the Leu22 → Val mutAREA DBD for **T**GATAG core elements over **C**GATAG elements. In addition, the determination of equilibrium dissociation constants for single-site binding not only permits the linking of the functional data directly with the structural studies, but also provides the basis for further quantification of cooperative interactions observed in the AREA system.

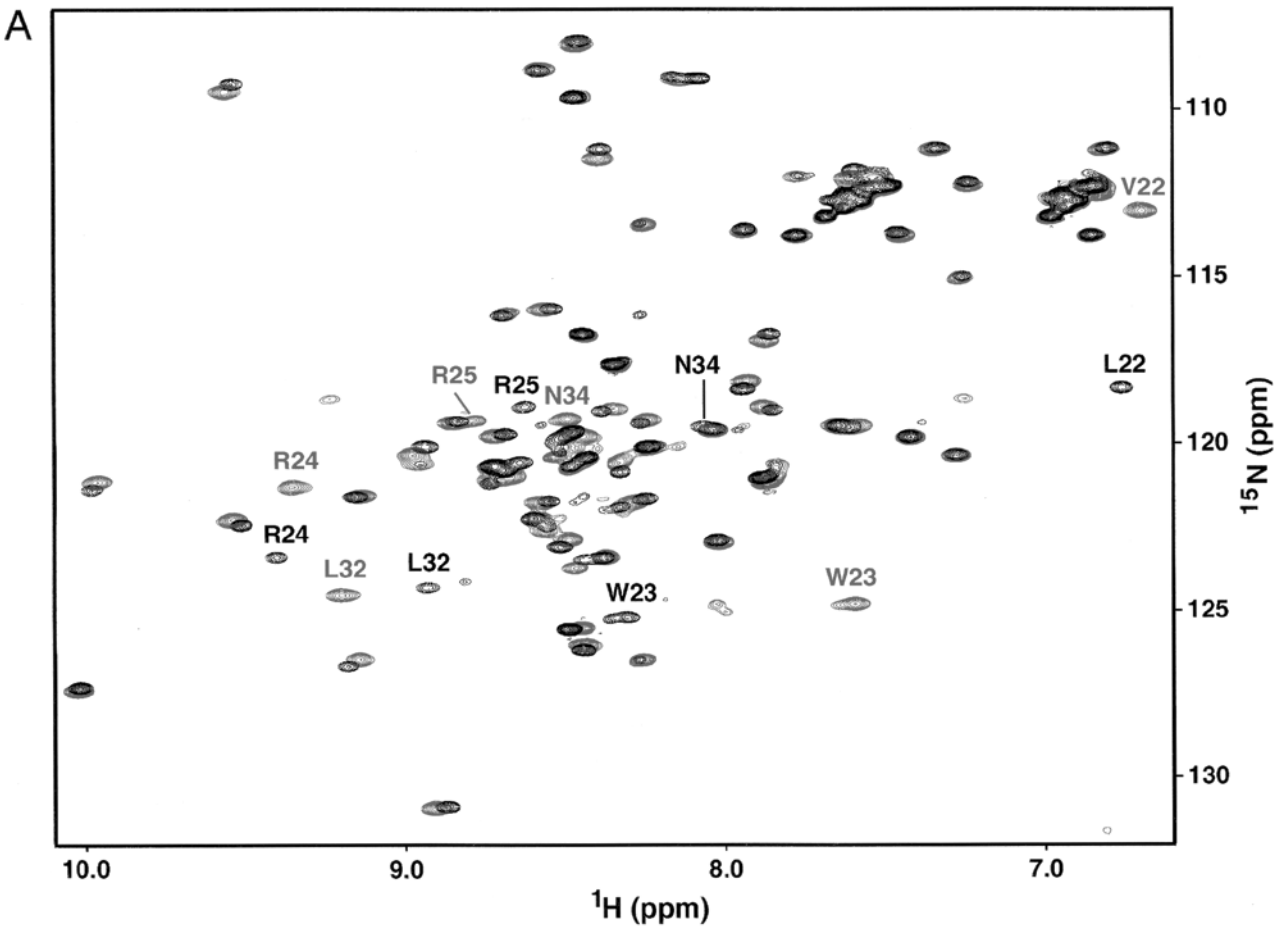

Figure 2A (*legend on page* 624)

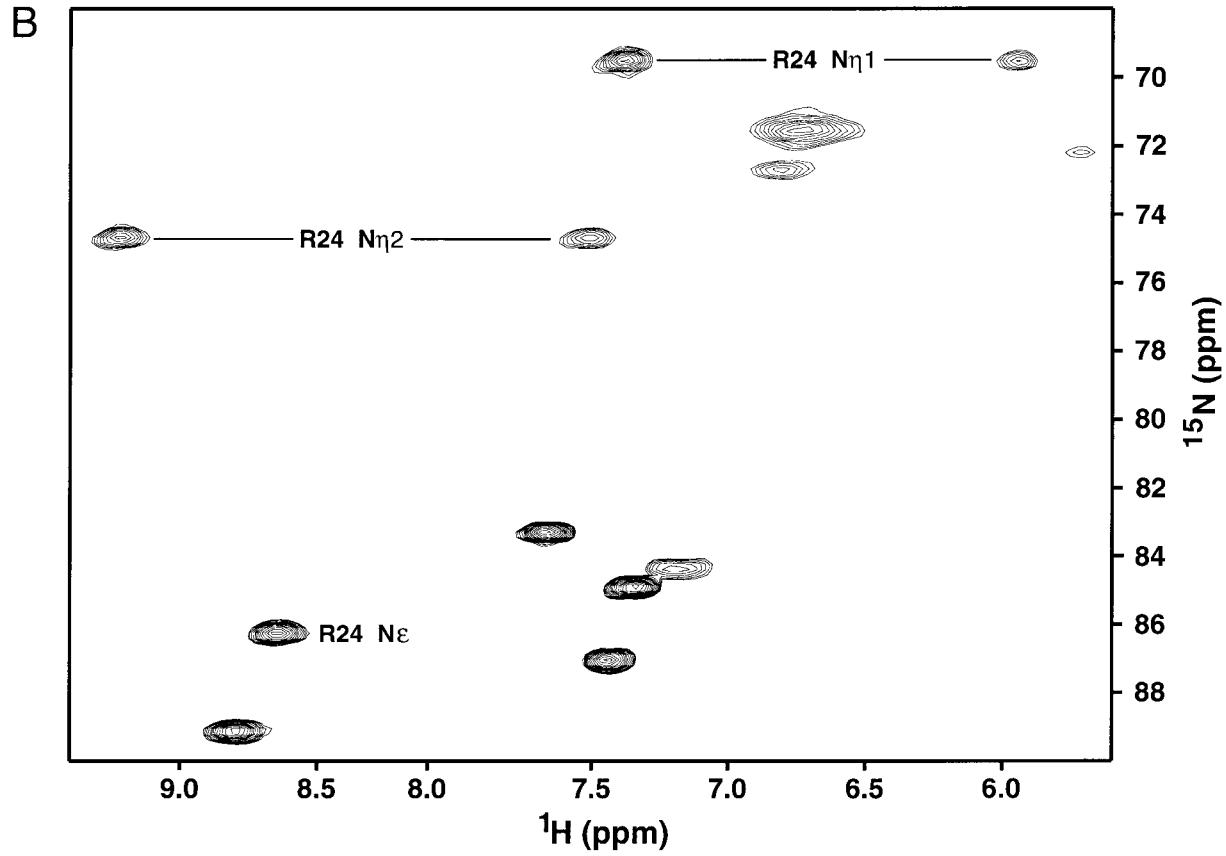

Figure 2. ^{1}H-^{15}N correlation spectra of Leu22 → Val mutant and wild-type AREA DBD·DNA complexes. A, Overlay of the ^{1}H-^{15}N-HSQC spectra for the wtAREA DBD·**C**GATAG complex (black) and the mutAREA DBD·**T**GATAG complex (red). The red labels designate residues in close spatial proximity to the Leu22 → Val mutation site which display ^{1}H chemical shift (Trp23, Arg25, Leu32 and Asn34) or ^{15}N chemical shift (Arg24) changes greater than 0.2 ppm and 2.0 ppm, respectively, relative to wild-type backbone amide correlation peaks (black labels). B, Region of the ^{1}H-^{15}N HSQC spectrum collected for the mutAREA DBD·**T**GATAG complex showing correlations between the Arg guanidino nitrogen atoms and their attached protons. The distinct resonance pairs observed for the guanidino N$^{\eta}$-H groups are labeled, and their presence is indicative of motional restriction of the Arg24 functional group due to formation of buried hydrogen bonding interactions.

Initial comparison of mutant and wild-type complexes

Prior to structure determination, preliminary comparison of the ^{1}H-^{15}N correlation spectra (Figure 2A) collected for the wtAREA DBD·**C**GATAG and mutAREA DBD·**T**GATAG complexes showed noticeable changes in the ^{1}H or ^{15}N chemical shifts of five residues in close proximity to the site of the Leu22 → Val substitution. For four of these residues (Trp23, Arg25, Leu32 and Asn34) perturbations in ^{1}H chemical shift of 0.2 to 0.6 ppm were observed, while the fifth (Arg24) experienced a 2.3 ppm change in ^{15}N chemical shift. Examination of the 3D ^{15}N-separated nuclear Overhauser enhancement (NOE) data revealed similar short and long range intramolecular contacts for the backbone protons of residue 22 in both the wild-type and mutant DBDs. However, important differences in the observed intermolecular NOEs for the Leu22 and Val22 side-chains were noted in the 3D ^{13}C-separated/^{12}C-filtered NOE data which correlate NOE interactions between protein protons attached to ^{13}C and non-exchangeable DNA protons attached to ^{12}C (Figure 3). Specifically, strong intermolecular NOEs were observed between the H3′ of the C4 sugar and both methyl groups of Leu22 in the wtAREA DBD·**C**GATAG complex (Figure 3A; Starich *et al.*, 1998) that were not present in the spectrum collected for the mutant complex (Figure 3B). Conversely, an NOE equivalent to the strong NOE between the methyl group of T4 and the γ1 methyl group of Val22 in the mutAREA DBD·**T**GATAG complex (Figure 3B) was not observed for the wild-type complex (Figure 3A). Interestingly, an additional weak NOE between the H2′/H2″ protons of G3 and the γ2 methyl group of Val22 was observed for the mutant complex (Figure 3B), but no similar NOE could be identified in the spectrum of the wild-type complex (Figure 3A; Starich *et al.*, 1998).

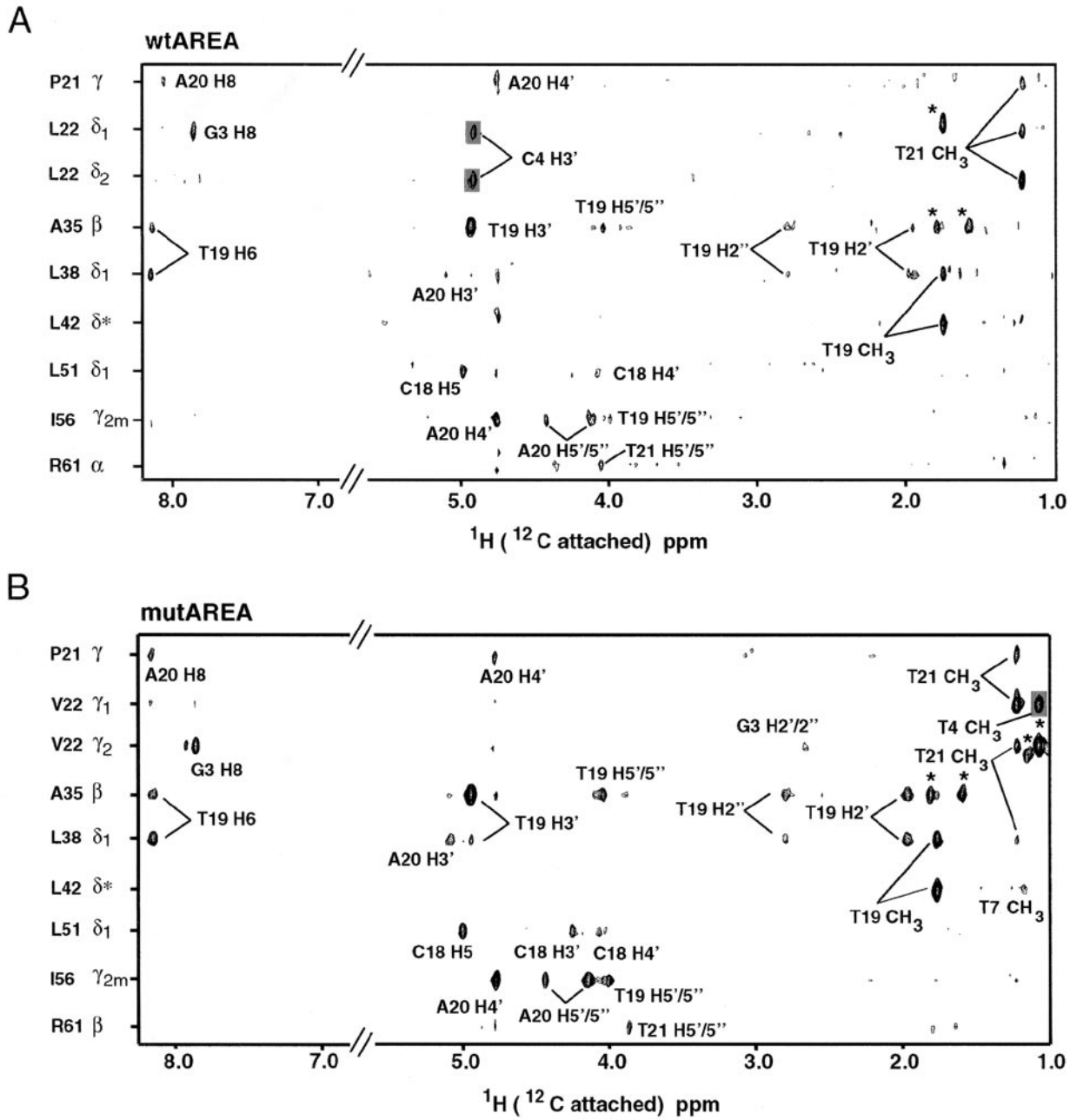

Figure 3. Composite of ^{13}C-H strips selected from the 3D ^{13}C-separated/^{12}C-filtered NOE spectrum for (A) the wtAREA DBD·**C**GATAG complex and (B) the Leu22 → Val mutAREA DBD·**T**GATAG complex. Boxed cross-peaks are intermolecular NOEs exclusive to either the wild-type or mutant complex. The δ methyl protons of Leu22 (wtAREA) exhibit unique NOEs to the H3′ sugar proton of C4, while the γ1 methyl protons of Val22 (mutAREA) exhibit a unique NOE to the methyl group of T4. Asterisks indicate autocorrelation cross-peaks in both spectra which correspond to incompletely filtered protons attached to ^{13}C.

Structure determination

The solution structure of the mutAREA DBD bound to its cognate **T**GATAG site was solved using multidimensional heteronuclear-filtered and heteronuclear-edited NMR spectroscopy (Clore & Gronenborn, 1991; Gronenborn & Clore, 1995; Bax & Grzesiek, 1993; Bax *et al.*, 1994). The structure was determined on the basis of 958 experimental NMR restraints, including 58 intermolecular NOEs. The somewhat larger number of observed intermolecular NOEs relative to the wild-type AREA DBD·DNA complex (Starich *et al.*, 1998) is due to the generally higher quality of the spectra obtained for the mutant AREA DBD·DNA complex. A summary of the structural statistics is provided in Table 1 and a superposition of the final 35 simulated annealing structures is shown in Figure 4.

Table 1. Structural statistics

	$\langle SA \rangle$	$(\overline{SA})r$
Structural statistics		
rms deviations from NOE interproton distance restraints (Å)[a]		
All (578)	0.038 ± 0.002	0.030
Protein		
Interresidue sequential ($\|i-j\|=1$) (131)	0.038 ± 0.010	0.027
Interresidue short range ($1<\|i-j\|\leqslant 5$) (63)	0.021 ± 0.007	0.023
Interresidue long range ($1<\|i-j\|\leqslant 5$) (67)	0.042 ± 0.011	0.030
Intraresidue (38)	0.003 ± 0.006	0.001
DNA		
Intraresidue (75)	0.013 ± 0.004	0.012
Sequential intrastrand (124)	0.056 ± 0.004	0.045
Interstrand (22)	0.041 ± 0.008	0.036
Protein-DNA (58)	0.041 ± 0.009	0.039
rms deviation from hydrogen bonding restraints (Å)		
Protein (20)[b]	0.044 ± 0.009	0.028
DNA (63)[b]	0.014 ± 0.005	0.006
Protein-DNA (4)[c]	0.041 ± 0.007	0.039
rms deviations from distance restraints to phosphates (2)[d]	0.046 ±0.033	0.015
rms deviations from "repulsive" restraints (Å) (8)[d]	0.002 ± 0.007	0.003
rms deviations from exptl dihedral restraints (deg) (290)[a]	0.19 ± 0.09	0.32
rms deviations from exptl $^3J_{HN\alpha}$ (Hz) (39)	0.92 ± 0.05	0.92
rms deviations from exptl residual one-bond		
^{15}N-^{1}H dipolar couplings (Hz) (49)	0.106 ± 0.006	0.16
rms deviations from exptl ^{13}C shifts		
$^{13}C^{\alpha}$ (ppm) (41)	0.85 ± 0.03	0.81
$^{13}C^{\beta}$ (ppm) (34)	0.82 ± 0.03	0.91
Deviations from idealized covalent geometry		
Bonds (Å) (1930)	0.005 ± 0.0001	0.006
Angles (deg.) (3499)	1.001 ± 0.008	1.087
Impropers (deg.) (971)	0.451 ± 0.027	0.653
Measures of structural quality		
$E_{L\text{-}J}$ (kcal mol^{-1})[e]	−595 ± 10	−507
PROCHECK[f]		
% Residues in most favorable region of Ramachadran map	81.5 ± 2.5	81.4
Number of bad contacts/100 residues	6.3 ± 1.8	7.7
Coordinate precision[g]		
Protein backbone plus DNA (Å)	0.46 ± 0.11	
All protein atoms plus DNA (Å)	0.75 ± 0.15	
Protein backbone (Å)	0.35 ± 0.10	
All protein atoms (Å)	0.91 ± 0.18	
DNA (Å)	0.42 ± 0.09	

The notation of the NMR structures is as follows: $\langle SA \rangle$ are the final 35 simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual *SA* structures best fitted to each other (with respect to residues 10 to 61 and the zinc atom of the protein and base-pairs 2 to 11 of the DNA); $(\overline{SA})r$ is the restrained regularized mean structure obtained by restrained regularization of the mean structure $\overline{SA}$. The number of terms for the various restraints is given in parentheses.

[a] None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, $^3J_{HN\alpha}$ coupling constant violations greater than 3 Hz. There are 120 torsion angle restraints for the protein (53 ϕ, 13 ψ, 39 χ_1, 14 χ_2 and 1 χ_3). There are also 170 broad torsion angle restraints for the DNA backbone as described by Starich *et al.* (1998).

[b] Hydrogen bonding restraints for the DNA (Watson-Crick base-pairing) and protein backbone as described by Starich *et al.* (1998).

[c] Intermolecular hydrogen bonding restraints between the protein side-chain of Arg24 and G5 were only added in the final stage of refinement based on the observation of four distinct resonances for the guanidino protons of Arg24 in ^{1}H-^{15}N HSQC and ^{15}N-separated NOE spectra.

[d] The "repulsive" distance restraints and intermolecular restraints involving the phosphates are as described by Starich *et al.* (1998).

[e] $E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy and is not included in the target function for simulated annealing or restrained minimization.

[f] The PROCHECK (Laskowski *et al.*, 1993) statistics relate to the ordered region of the polypeptide chain (residues 10 to 61). There are no residues whose ϕ/ψ angles fall in the disallowed region of the Ramachandran plot.

[g] The precision of the coordinates is defined as the average atomic rms difference between the 35 individual simulated annealing structures of each complex and the mean coordinates $\overline{SA}$. The values refer to residues 10 to 61 of the mutAREA DBD, the zinc atom and base-pairs 2 to 11 of the DNA.

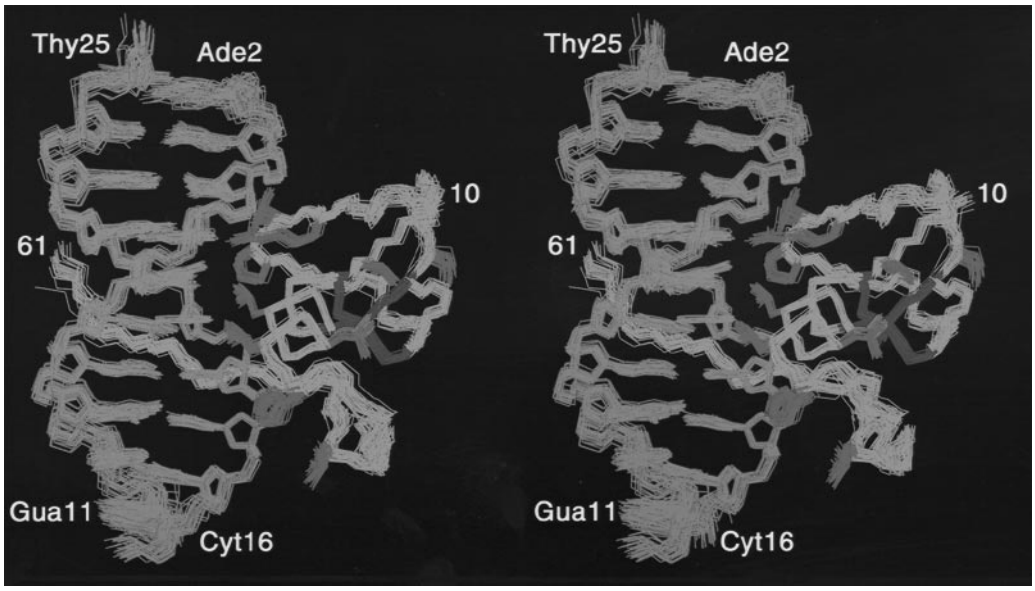

Figure 4. Stereoview showing the best fit superposition of the 35 simulated annealing structures of the Leu22 → Val mutAREA DBD·**T**GATAG complex. The backbone atoms (N, C^{α}, C) of the AREA DBD (residues 10 to 61) are shown in yellow, side-chains are shown in red, and the Zn and coordinating cysteine residues are shown in green. All non-hydrogen atoms of the DNA (residues 2 to 11 and 16 to 25), with the exception of the O1P and O2P phosphate oxygen atoms, are shown in blue.

Just as in the case of the wild-type complex (Starich *et al*., 1998), residues 1 to 9 and 62 to 66 at the N and C termini, respectively, are disordered in solution. The precision of the coordinates for the complex (backbone of residues 10 to 61 of the mutAREA DBD and basepairs 2 to 11 of the DNA) is ~0.5 Å.

Comparison of wtAREA and mutAREA structures

The global fold and DNA binding mode of the mutAREA DBD·**T**GATAG and wtAREA DBD·**C**GATAG (Starich *et al*., 1998) complexes are very similar (Figures 5 and 6A). The atomic

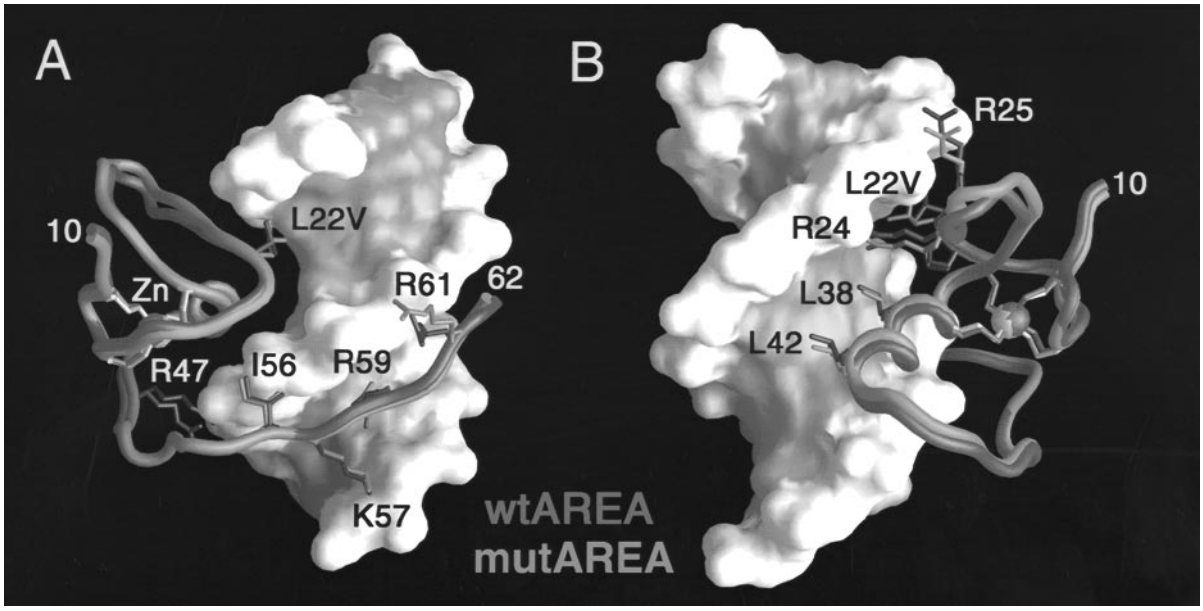

Figure 5. Two views showing a comparison of the wtAREA (red worm) and Leu22 → Val mutAREA (gold worm) DBDs complexed to DNA. Residues 10 to 61 of both domains are superimposed in the views shown. The surfaces representation of only one of the DNA models (the **T**GATAG site for the mutant complex) is shown for clarity. The surfaces of the bases in the major groove are colored in light blue while those in the minor groove are in light red. Side-chains for the wtAREA and mutAREA DBDs are shown in blue and green, respectively. The zinc atoms and their coordinating ligands are shown in purple (wtAREA) and pink (mutAREA). The coordinates of the wtAREA DBD complex are taken from Starich *et al*. (1998).

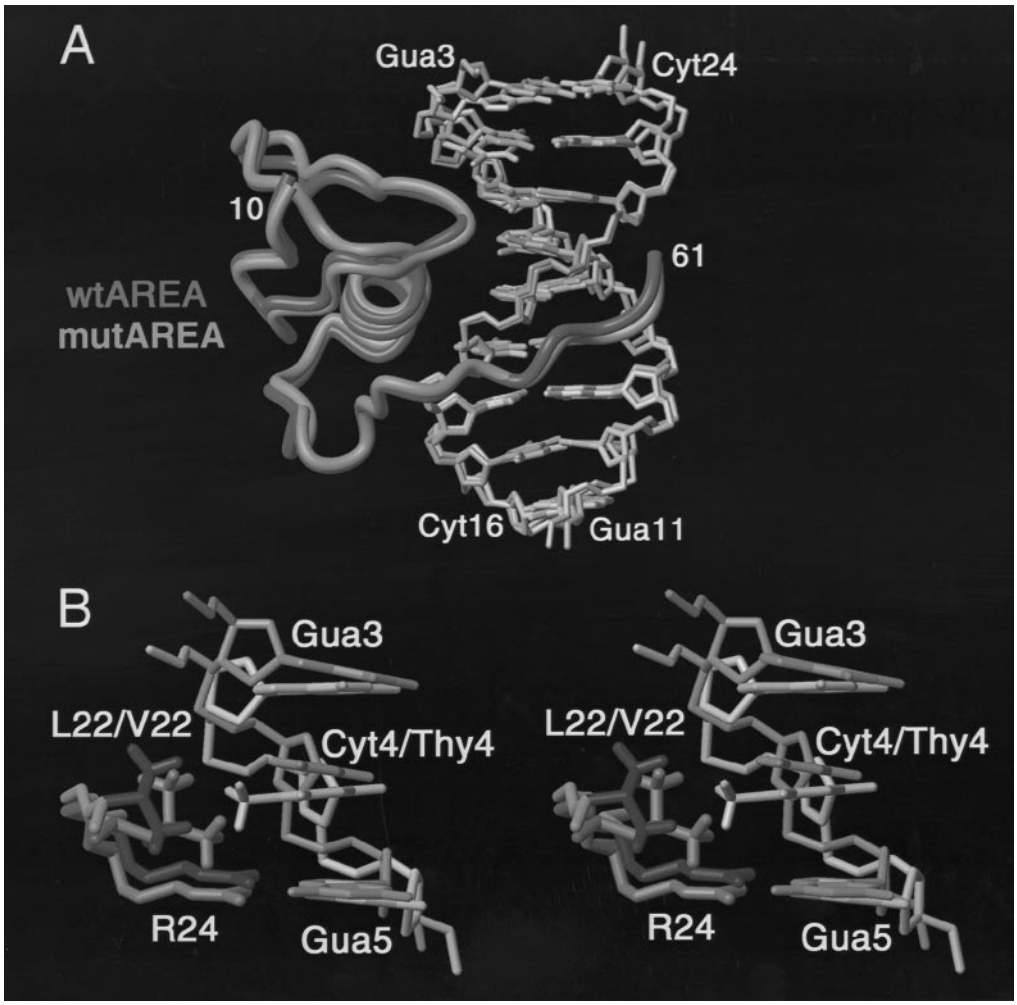

Figure 6. Comparison of the Leu22 → Val mutAREA DBD·**T**GATAG and wtAREA DBD·**C**GATAG complexes. A, Superposition of the protein backbone and DNA for the wild-type and mutant AREA DBD complexes. The protein backbones are depicted as red (wtAREA) and gold (mutAREA) worms. Bonds between heavy atoms of the DNA (base-pairs 3 to 11) are represented as light red (CGATA site) or tan (TGATA site) sticks. B, Stereoview showing protein-DNA interactions for residues at positions 22 and 24. The wtAREA DBD backbone and side-chains are shown in red and blue, respectively, while the mutAREA DBD backbone and side-chains are shown in gold and green, respectively. The G3-C4-G5 element of the **C**GATAG site is shown in light red, and the G3-T4-G5 element of the **T**GATAG site is shown in tan. Only bonds between heavy atoms for the protein and DNA are represented with the exception of C–H bonds for all methyl groups shown (Leu22 δs, Val22 γs and the methyl group of T4). The coordinates of the wtAREA DBD complex are taken from Starich *et al*. (1998).

rms differences between the coordinates of the mutAREA and wtAREA DBD·DNA complexes for residues 10 to 61 of the DBDs and bp 3 to 11 of the DNA (which are in contact with the DBDs) are ~0.6 Å for the protein backbone, ~0.8 Å for all protein atoms, ~0.8 Å for all DNA atoms and ~1.0 Å for all protein atoms plus all DNA atoms. These values are comparable to the precision of the coordinates (Table 1). In most cases, the side-chain orientations observed in both complexes are identical, as evidenced by essentially the same pattern of NOEs for the non-mutated amino acids. The backbone of the C-terminal tail runs parallel with the sugar phosphate backbone of the DNA along the edge of the minor groove in the two complexes (Figure 5A), and a best-fit superposition of the GATA DNA sites (Figure 6A) shows similar positions for all sugars and bases with the exception of those directly surrounding the C/T mutation (G3, C4/T4, G5 and C24). Further, the average values for the various structural parameters (helical twist and rise, propellor twist, local inter-base-pair tilt and roll) describing the DNA, which is *B*-like, are the same in the mutant and wild-type complexes (Starich *et al*., 1998).

The most significant perturbations to the protein structure occur around the Leu22 → Val mutation site (Figure 6B). The protein backbone shifts slightly in this region, showing the greatest displacements relative to the wild-type coordinates for the C^{α} atoms of Val22 (1.02 Å) and Arg24 (1.00 Å). The position of the Val22 side-chain relative to the base at position 4 is clearly altered as compared to

the longer Leu22 functional group (Figure 6B). For example, the distances between the $C^{\gamma 1}$ and $C^{\gamma 2}$ carbons of Val22 and the C5 ring carbon of T4 are 5.7 Å and 5.3 Å, respectively, compared to 4.5 and 4.2 Å, respectively, between the $C^{\delta 1}$ and $C^{\delta 2}$ carbons of Leu22 and the C5 ring carbon of C4. (Note that, as can be seen in Figure 3A, we do not observe NOEs between the methyl groups of Leu22 and the H5 proton of C4 despite the fact that the corresponding interproton distances are ~3.5 Å; this is due to line broadening of the H5 resonance of C4; the position of the methyl groups of Leu22 relative to the DNA, however, is well defined by five NOEs from the methyl groups of Leu22 to the methyl protons of T21, the H8 proton of G3 and the H3′ proton of C4.) As expected, the distances between the $C^{\gamma 1}$ and $C^{\gamma 2}$ atoms of Val22 and the C5 methyl carbon of T4 are shorter (4.2 Å and 3.8 Å, respectively). Closer examination of the mutation site shows that the $C^{\delta 1}$ and $C^{\delta 2}$ atoms of Leu22 are quite close to the H3′ and C3′ atoms of C4, each methyl carbon atom lying within ~5–6 Å of these sugar atoms. In contrast, the $C^{\gamma 1}$ and $C^{\gamma 2}$ atoms of Val22 lie more than 8.3 Å from the H3′ and C3′ atoms of the T4 sugar. These distances are a direct reflection of the observed differences in the intermolecular NOE data for the two complexes (Figure 3). The orientation of side-chain methyl groups also differs; the methyl groups of Val22 point backwards with respect to the plane of the page whereas those of Leu22 are directed towards the reader in the stereoview shown in Figure 6B. Side-chain orientations for both residues, determined unambiguously from $^3J_{C\alpha C\delta}$, $^3J_{NC\gamma}$ and $^3J_{COC\gamma}$ coupling constants and the relative intensities of intraresidue NOE cross-peaks are also different: specifically, Val22 has a χ_1 angle in the g^+ conformation (with a value of 60(±4)°), while the χ_1 and χ_2 angles of Leu22 are in the t and g^+ conformations, respectively (with values of 170(±3)° and 57(±3)°, respectively). As a result the γ2 methyl of Val22 occupies a very similar position to that of the δ1 methyl of Leu22 (atomic rms diplacement of ~1 Å), while the displacement between the γ1 methyl of Val22 and the δ2 methyl of Leu22 is much larger (~2.3 Å; Figure 6B).

Changes are also noted for the **T**GATAG sequence element and include a shift in position for both the sugar and bases of G3, T4 and C24 relative to wild-type **C**GATAG coordinates (Figure 6A and B). The carbon atoms belonging to the sugar ring of G3 are displaced by more than 2.1 Å (with a range of 2.1 to 2.7 Å) relative to those in the wild-type structure, while those belonging to the sugar ring of T4 are displaced by more than 1.8 Å (with a range of 1.8 to 2.3 Å) relative to the analogous atoms of C4. The atomic rms shift for the sugar ring of C24 is not as pronounced, with an average displacement of ~1.0 Å. The most pronounced displacements for mutant base coordinates relative to wild-type coordinates range from 1.5 to 2.4 Å for individual atoms, and involve the bases of G3 and T4. When considering, however, the impact of these differences between the two DNA sites, it should be noted that the positions of individual DNA bases are primarily determined by a small number of intermolecular NOEs which provide long range order, supplemented by restraints that are limited to short range order and consist of the hydrogen bonds associated with Watson-Crick base-pairing, and intraresidue base-sugar and sequential base-base and base-sugar NOEs. Thus, the absence or presence of intermolecular contacts observed for the mutation site becomes a major contributor to base position.

Interestingly, little displacement is noted for the adjacent base of G5. This may be attributed to the preservation of the intermolecular hydrogen bonds between the guanidino $N^{\eta}H_2$ groups of Arg24 and the O6 and N7 atoms of G5 in the major groove. Analogous to signals observed for Arg24 in the wild-type complex, four distinct ^1H-^{15}N correlations for the guanidino N^{η}-H pairs of Arg24 are observed in the ^1H-^{15}N HSQC spectrum collected at 25°C for the mutAREA DBD·**T**GATAG complex (Figure 2B). The chemical shifts of these signals are quite similar for both complexes, with the exception of a ~2 ppm upfield shift in the ^{15}N resonance for the $N^{\eta 1}$ nitrogen, relative to the wild-type chemical shift. Compared to $N^{\eta 2}$, this nitrogen atom is closer in proximity to the Leu22 → Val and C4 → T substitutions and is more likely to experience chemical shift perturbations. Observation of these signals for the mutant complex indicates motional restriction of the guanidino functional group of Arg24 and is consistent with the maintenance of a buried hydrogen bonding interaction.

Hydrophobic packing, affinity and specificity

Understanding the mechanisms which contribute to sequence-specific binding of DNA not only requires the study of new DNA binding motifs, but also necessitates the systematic examination of the roles played by individual amino acids and bases involved in the recognition event. The combination of biochemical, genetic and structural data now available for the AREA system permits direct analysis of how a subtle change in hydrophobic packing affects specificity and affinity. To assess further the effects of mutations at position 22 on local structure, space-filling representations of the wt and mutAREA DBD·(**C**/**T**)GATAG structures were generated from the regularized mean coordinates (Figure 7A and D). Four additional models, created by substitution and subsequent regularization of the resulting coordinates, illustrate the introduction of Val22 into the coordinates of the wild-type complex (Figure 7B), Leu22 into the coordinates of the mutant complex (Figure 7E), and Met22 into the coordinates of both complexes (Figure 7C and F).

Inspection of space-filling representations for the two complexes shows that both side-chains

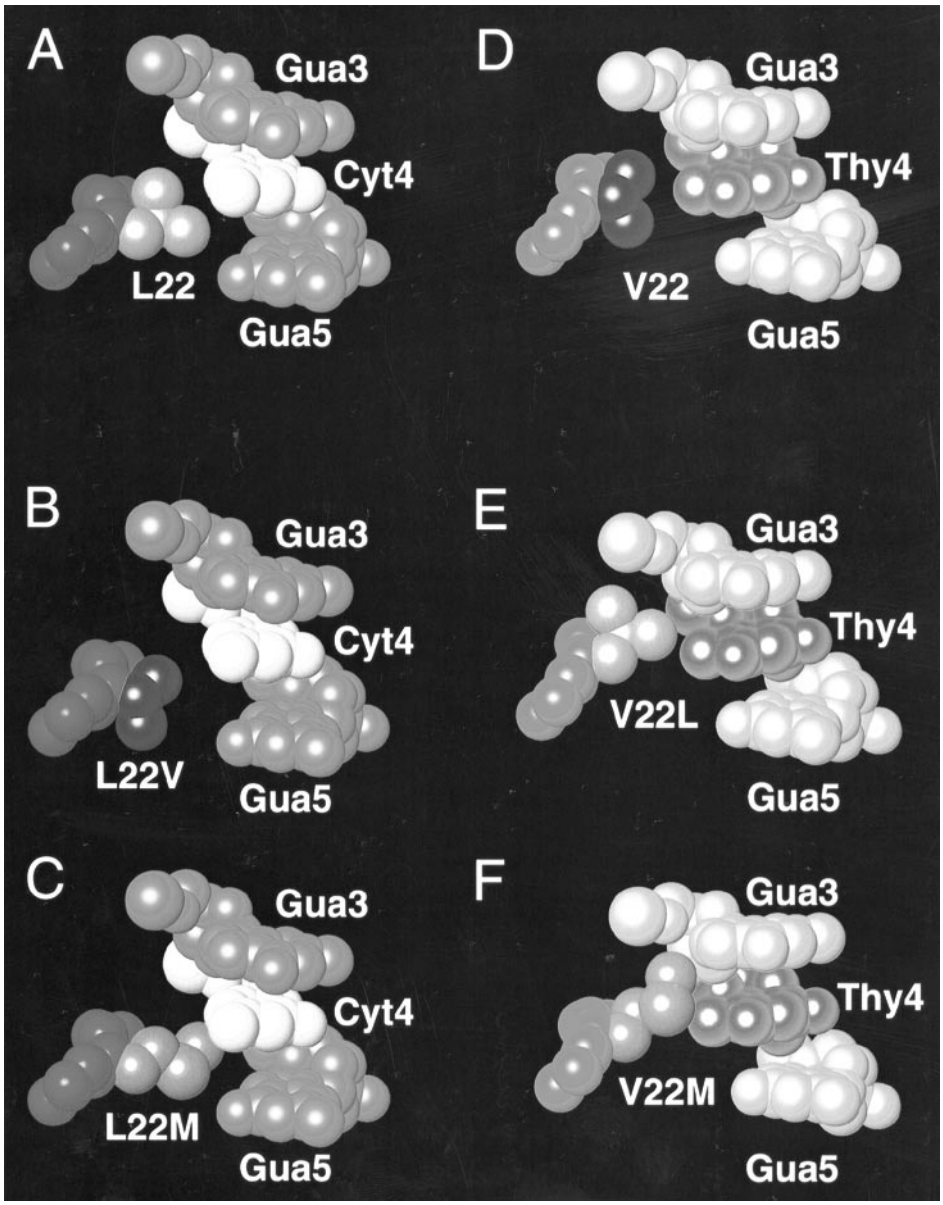

Figure 7. Space-filling models detailing the interaction of residue 22 with DNA. Restrained regularized mean coordinates of (A) the wtAREA DBD·**C**GATAG complex and (D) the mutAREA DBD·**T**GATAG complex; models derived from the structure of the wtAREA DBD·DNA complex showing the substitution of Leu22 with (B) valine and (C) methionine; models derived from the structure of the mutAREA DBD·DNA complex showing the substitution of Val22 with (E) leucine and (F) methionine. The wtAREA DBD backbone is shown in red (A,B,C), and the mutAREA DBD backbone is shown in gold (D,E,F). Leucine is always represented in deep blue (A,E), valine in bright green (B,D), and methionine in drab green (C,F). Bases belonging to the **C**GATAG site are shown in light red and light blue, while those belonging to the **T**GATAG site are shown in tan and purple. The coordinates of the wtAREA DBD complex are taken from Starich *et al*. (1998).

maintain approximately the same distance from the base at position 4 (~4.0 to 4.5 Å; Figure 7A and D). The center of the Leu22 side-chain sits slightly below the plane of the C4 base in the wild-type site, while the Val22 methyl groups are essentially centered with respect to the T4 methyl group. When valine is substituted into the wild-type coordinates (Figure 7B) the distance to the base carbon atoms of C4 visibly increases. Measured distances between valine methyl carbon atoms and the C5 atom of the cytosine are 4.9 and 6.6 Å (Figure 7B), while those observed for the leucine methyl carbon atoms are 4.5 and 4.2 Å (Figure 7A). Lacking the additional methylene group, the Val22 side-chain packs poorly against the **C**GATA site, and exploration of alternative χ_1 rotamers (g^- or t) does not reduce the gap between the side-chain and DNA. Conversely, the regularized model in Figure 7E demonstrates that substitution of leucine into the mutant coordinates maintains tight hydrophobic packing at the GTG interface. It is clear that a

TGATAG site provides an acceptable target for the wtAREA DBD, although it is probable that some rearrangement would be observed relative to the structure of the wtAREA DBD·**C**GATAG complex.

Both the structures and the above described models correlate well with existing biochemical data, suggesting that AREA affinity and specificity for **C**GATAG elements may be altered by subtle changes in hydrophobic packing associated with residue 22. The structures further indicate that introduction of a **T** preceding the GATA sequence effectively compensates for the shorter valine side-chain of mutAREA, restoring a close-packed hydrophobic interface. These findings provide a structural rationale for *in vivo* results that show null expression of *uap*A and *uap*C genes for *A. nidulans* strains possessing Leu22 → Val mutant AREA. Further, obtaining enhanced expression levels of the *amd*S gene requires specific recognition of **T**GATAG sequences, which is more readily achieved by utilizing valine at position 22 of the DBD.

It is also of interest to consider the previously characterized Leu22 → Met mutation which generates a mirror image phenotype of the Leu22 → Val mutation *in vivo*. Characterization of this phenotype revealed that *A. nidulans* strains containing the Leu22 → Met mutation showed low level expression of the *amd*S gene product, but exhibited overexpression of the *uap*A and *uap*C gene products (Arst & Scazzocchio, 1975; Hynes, 1975; Gorton, 1983). Qualitative examination of model coordinates shows that a Leu22 → Met mutation in the presence of C4 is well tolerated (Figure 7C), while the Val22 → Met mutation in the presence of T4 results in steric clash (Figure 7F). Altering the side-chain dihedral angles of the methionine reduces steric clash with major groove bases but compromises hydrophobic packing and introduces less favorable side-chain conformations. Hence, both model building and *in vivo* results suggest that AGATA and CGATA are preferred sites for AREA possessing the Val22 → Met mutation.

Correlation with mutational data

A wide spectrum of mutational changes in AREA has been characterized using a variety of *in vivo* tests and, in some cases, more qualitative tests (Arst & Cove, 1973; Kudla *et al*., 1990; Langdon *et al*., 1995; Platt *et al*., 1996a,b; Ravagnani *et al*., 1997). Loss-of-function missense mutations have an intrinsic interest as they report directly on the functional impact of the altered residue. They are also invaluable tools for reversion as the revertants can enable identification of acceptable substitutes for the altered residue and second-site substitutions which report on functional synergy and compensation effects. The serendipitous recovery of the Ala35 → Pro mutation, which very likely disrupts the α-helix and would alter contacts with T19 and A20 in the major groove (Starich *et al*., 1998), in an otherwise wild-type (*vis-a-vis* AREA) background (Kudla *et al*., 1990; Platt *et al*., 1996a) provides a unique opportunity for comparisons of its behaviour in the Leu22 → Val (in which it was originally obtained) and wild-type backgrounds. The Leu22 → Val mutation markedly changes the reversion spectrum associated with second-site mutations. Although the direct Pro35 → Ser revertant substitution restores function in both wild-type (Leu22) and Leu22 → Val backgounds, only a single second-site substitution, Phe39 → Tyr, suppressing Ala35 → Pro in a wild-type background has been obtained, whereas five different alternative second-site mutational changes suppressing Ala35 → Pro in a Leu22 → Val background have been recovered, including mutation of Pro21 to Ser, Thr or Leu and Pro31 → Arg (Platt *et al*., 1996a; H.N.A. and T. Langdon, unpublished results).

The Phe39 → Tyr residue change is located on the same face of the helix as Ala35 (cf. Figure 4B of Starich *et al*., 1998), and one possible explanation for its effect is that the Ala35 → Pro substitution introduces a small kink and changes the direction of the helix slightly such that distances between the C^{ζ} of Phe39 on the one hand and the phosphate oxygen atoms of C18 and the guanidino group of Arg47 on the other (which are <4 Å in the wild-type structure) are increased. Substitution of Phe39 by the bulkier Tyr would fill the void and possibly permit the formation of a hydrogen bond(s) between the hydroxyl group of the tyrosine and either the phosphate of C18 and/or the guanidino group of Arg47. Indeed, most of the other GATA domains possess a highly conserved Tyr in this location.

The proline at position 21 also participates in major groove recognition of A20 (Starich *et al*., 1998), and it is likely that replacement of its side-chain by that of Ser, Thr or Leu increases the energetically favorable regions of ϕ/ψ space that can be sampled at position 21, allowing local structural rearrangements to restore nearly optimal packing in the major groove. Fine-structure recombination experiments have established that Pro21 → Leu is almost certainly incapable of suppressing Ala35 → Pro in a wild-type (Leu22) background (H.N. A. unpublished results), emphasizing the profound influence of the Leu22 → Val substitution. Although remote from the DNA, the Pro31 → Arg mutation might allow a subtle change in the relative orientation of the helix (residues 33 to 43) relative to the β3-β4 antiparallel sheet (residues 23 to 26 and 29 to 32), thereby compensating for the conformationally less flexible proline introduced at position 35. The ability of the revertant Pro35 → Ser substititution to restore function in both Leu22 → Val and wild-type backgrounds (Kudla *et al*., 1990) further supports the notion that proline at position 35 disrupts hydrophobic packing in the major groove.

Although the Gln30 → Pro mutation (Kudla *et al.*, 1990) exists only in a Leu22 → Val background, at least two of its second-site suppressors Asn26 → Ser and Asn62 → Lys (H.N.A. & T. Langdon, unpublished results) are of structural interest. The Gln30 → Pro substitution occurs adjacent to Pro31 and eliminates a hydrogen bond between Asn26-CO and Gln30-NH associated with the β3-β4 antiparallel sheet. Introduction of an alternative hydrogen bond involving Ser26-O$^{\gamma}$H and Glu28-O$^{\varepsilon}$ is one possible explanation for suppression by the Asn26 → Ser substitution.

The Asn62 → Lys mutation is of special interest, as it occurs just beyond the C-terminal tail of the AREA DBD and structural data indicate that Arg61 is the last residue of the domain to participate in DNA recognition. In this case, the structure of the chicken GATA-1 (cGATA-1) DBD complexed with DNA (Omichinski *et al.*, 1993) provides an important clue for determining how this remote mutation might compensate for the Gln30 → Pro substitution. As observed previously, the primary structural difference between the cGATA-1 and AREA DBDs lies in the binding mode of the C-terminal tail (Starich *et al.*, 1998). The cGATA-1 domain utilizes a lysine residue at position 62 to make key hydrophobic and hydrogen bonding contacts in the minor groove which contribute to higher affinity binding to GATA sites. Assuming a similar role for the Asn62 → Lys mutation in AREA, it is satisfying to note that lysine, which is highly conserved among mammalian factors in position 62, was selected to restore AREA function.

The current structure-function studies of the AREA system indicate that measurable differences in affinity and profound changes in specificity may be realized *via* subtle alterations in structure. This is emphasized by the results presented here for the Leu22 → Val mutAREA DBD·**T**GATAG complex. The catalog of structures now available for the GATA family of proteins extends our knowledge of how specificity and affinity are conferred within this family. Careful examination of these structures will aid in the design of future *in vitro* mutagenesis studies aimed at clarifying structure-function relationships for this family of DNA binding proteins.

Materials and Methods

Gel retardation assays

The wtAREA DBD was dissolved in 0.05% trifluoroacetic acid (TFA) to prepare a 0.23 mM stock solution (pH 6.5) as determined by ultraviolet absorption (UV) spectroscopy using a molar extinction coefficient of 3280 at a wavelength of 280 nM. Likewise, mutAREA DBD was dissolved in 0.05% TFA to prepare a 0.23 mM stock solution (pH 6.5) as determined by UV spectroscopy using a molar extinction coefficient of 5000 at 280 nM. Serial dilutions of these protein stocks were made so that a constant addition of 1 μl of protein solution to assay mixtures yielded final concentrations of 2.5 to 1000 μM AREA DBD. Also, 125:1 mixtures of unlabeled DNA to ^{32}P radiolabeled DNA were prepared as 50 μM stock solutions to allow a constant addition of 1 μl of DNA solution to assay mixtures, yielding a final concentration of 5 μM DNA. Titrations were carried out at 25°C by mixing 2 μl assay buffer (250 mM Tris-HCl, 0.0625% (v/v) Triton, 16% (w/v) Ficoll 400), 1 μl DNA stock, 1 μl protein solution, 0 to 5 μl unlabeled competitor DNA and 0 to 5 μl sterile water for a final reaction volume of 10 μl. Individual reactions were mixed thoroughly and incubated for ten minutes at 25°C before loading 8 μl of this volume onto a 10% (w/v) polyacrylamide gel and electrophoresing at 100 V for 1 to 1.5 hours. All gels were pre-run at 100 V for 30 minutes in running buffer (10 mM Tris-HCl, 10 mM Hepes-free acid form).

The gels were dried and exposed to a Fuji imaging plate. A Fuji BAS 2000 system (Fuji Medical Systems) was used for subsequent processing of the autoradiographs and quantification of band intensities. The fraction of radiolabeled DNA bound (*FB*) was determined from relative intensities and plotted as a function of competitor concentration, [C]. For cases in which the unlabeled DNA competitor and radiolabeled DNA possessed the same sequence, data were fit by non-linear least-squares optimization to the equation $FB = FB_o/[1 + 10^{\{\log[C] - \log(K_D + [DNA])\}}]$ for homologous competition, where K_D is the equilibrium dissociation constant, *FB* is fraction of radiolabeled DNA bound, FB_o is *FB* in the absence of competitor; [C] is competitor concentration; and [DNA] is the concentration of DNA (Swillens, 1995). For cases in which the unlabeled DNA competitor and radiolabeled ligand had different DNA sequences, data were fit to the equation (K_{Dc}): $FB = FB_o/[1 + 10^{\{\log[C] - \log(K_{Dc} + K_{Dc}[DNA]/K_D)\}}]$ for heterologous competition, where K_{Dc} is the equilibrium dissociation constant for the competitor, and K_D the equilibrium dissocaition constant previously determined by homologous competition experiments (Swillens, 1995).

Sample preparation and structure determination

All mutAREA DBD samples required for NMR studies were prepared in an identical fashion to that described previously for the wtAREA DBD samples (Starich *et al.*, 1998). Spectra for the complex were recorded at 25°C on AMX360, AMX500, DMX500, AMX600, DMX600 and DMX750 Bruker spectrometers equipped with *x*, *y*, *z*-shielded gradient triple resonance probes. The same set of multidimensional heteronuclear experiments that were used in the structure determination of the wtAREA DBD·**C**GATAG complex for ^{1}H, ^{15}N and ^{13}C resonance assignments, analysis of intra- and intermolecular NOEs, analysis of intramolecular ROEs, measurement of homo- and heteronuclear three-bond couplings, and measurement of residual one-bond ^{15}N-^{1}H dipolar couplings was employed for the mutant complex, and is discussed in detail in the accompanying paper (Starich *et al.*, 1998). All spectra were processed with the NMRPipe package (Delaglio *et al.*, 1995), and analyzed using the programs PIPP, CAPP and STAPP (Garrett *et al.*, 1991). The structures were calculated using simulated annealing (Nilges *et al.*, 1988) with the program XPLOR-31 (Brünger, 1993), modified to incorporate pseudo-potentials for secondary ^{13}C shift, coupling constant, and residual dipolar couplings restraints, as well as a conformational database potential for proteins and nucleic acids, as described in the accompanying paper (Starich *et al.*, 1998, and references therein).

The models in Figure 7 that required substitution of the residues at position 22 or base-pair 4 were generated from the restrained regularized mean coordinates for either the wtAREA DBD·**C**GATAG or mutAREA DBD·**T**GATAG structures. The substitutions were carried out using the program MOLMOL (Koradi *et al*., 1996) and the resulting structures were subjected to regularization in XPLOR using the same protocol described for the regularization of the mean coordinates (Starich *et al*., 1998). Figures were generated with the programs MOLMOL (Koradi *et al*., 1996) and GRASP (Nicholls *et al*., 1991).

The coordinates of the 35 final simulated annealing structures of the mutAREA DBD·**T**GATAG complex, together with the coordinates of the restrained regularized mean structure, ⟨*SA*⟩*r*, and the complete list of experimental NMR restraints have been deposited in the Brookhaven Protein Data Bank (PDB accession codes 6GAT, 7GAT and 6GATMR).

Acknowledgements

The authors thank D. S. Garrett and F. Delaglio for software support; K. Mizuuchi for radioactive oligonucleotides and assistance with gel-retardation assays; R. Tschudin for technical support; and A. Bax, J. Huth, J. Kuszewski, J. Omichinski, T. Strzelecka, N. Tjandra and C. Trainor for useful discussions. M.W. gratefully acknowledges a postdoctoral fellowship from the Swedish Natural Sciences Research Council (NFR). This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C. and A.M.G.).

References

Arst, H. N., Jr & Cove, D. J. (1973). Nitrogen metabolite repression in *Aspergillus nidulans*. *Mol. Gen. Genet.* **126**, 111–141.

Arst, H. N., Jr & Scazzocchio, C. (1975). Initiator constitutive mutation with an 'up-promoter' effect in *Aspergillus nidulans*. *Nature*, **254**, 31–34.

Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Acc. Chem. Res.* **26**, 131–138.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. & Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**, 79–125.

Brünger, A. T. (1993). *XPLOR: A system for X-ray crystallography and NMR,* Yale University Press, New Haven.

Clore, G. M. & Gronenborn, A. M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science,* **252**, 1390–1399.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR,* **6**, 277–293.

Diallinas, G., Gorfinkiel, L., Arst, H. N., Jr, Cecchetto, G. & Scazzocchio, C. (1995). Genetic and molecular characterization of a gene encoding a wide specificity purine permease of *Aspergillus nidulans* reveals a novel family of transporters conserved in prokaryotes and eukaryotes. *J. Biol. Chem.* **270**, 8610–8622.

Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220.

Gorfinkiel, L., Diallinas, G. & Scazzocchio, C. (1993). Sequence and regulation of the *uap*A gene encoding a uric acid-xanthine permease in the fungus *Aspergillus nidulans*. *J. Biol. Chem.* **268**, 23376–23381.

Gorton, D. J. (1983). Genetical and biochemical studies of the uptake of purines and their degradation products in *Aspergillus nidulans*. PhD thesis, University of Essex, UK.

Gronenborn, A. M. & Clore, G. M. (1995). Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. *CRC Crit. Rev. Biochem. Mol. Biol.* **30**, 351–385.

Hynes, M. J. (1975). Studies on the role of the *are*A gene in the regulation of nitrogen catabolism in *Aspergillus nidulans*. *Aust. J. Biol. Sci.* **28**, 301–313.

Hynes, M. J., Corrick, C. M., Kelly, J. M. & Littlejohn, T. J. (1988). Identification of the sites of action for regulatory genes controlling the *amd*S gene of *Aspergillus nidulans*. *Mol. Cell. Biol.* **8**, 2589–2596.

Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

Kudla, B., Caddick, M. X., Langdon, T., Martinez-Rossi, N. M., Bennett, C. F., Sibley, S., Davies, R. W. & Arst, H. N., Jr (1990). The regulatory gene *are*A mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. *EMBO J.* **9**, 1355–1364.

Langdon, T., Sheerins, A., Ravagnani, A., Gielkens, M., Caddick, M. X. & Arst, H. N., Jr. (1995). Mutational analysis reveals dispensability of the N-terminal region of the *Aspergillus* transcription factor mediating nitrogen metabolite repression. *Mol. Microbiol.* **17**, 877–888.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Nichols, A. J., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Letters,* **229**, 317–324.

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

Platt, A., Langdon, T., Arst, H. N., Jr, Kirk, D., Tollervey, D., Sanchez, J. M. M. & Caddick, M. X. (1996a). Nitrogen metabolite signalling involves the C-terminus and the GATA domain of the *Aspergillus* transcription factor AREA and the 3′ untranslated region of its mRNA. *EMBO J.* **15**, 2791–2801.

Platt, A., Ravagnani, A., Arst, H. N., Jr., Kirk, D., Langdon, T. & Caddick, M. X. (1996b). Mutational

analysis of the C-terminal region of AREA, the transcription factor mediating nitrogen metabolite repression in *Aspergillus nidulans*. *Mol Gen Genet.* **250**, 106–114.

Ravagnani, A., Gorfinkiel, L., Diallinas, G., Langdon, T., Adjadj, E., Demais, S., Gorton, D., Arst, H. N., Jr & Scazzocchio, C. (1997). Subtle hydrophobic interactions between the seventh residue of the zinc finger loop and the first base of an HGATAR sequence determine promoter-specific recognition by the *Aspergillus nidulans* GATA factor AreA. *EMBO J.* **16**, 3974–3986.

Scazzocchio, C. & Arst, H. N., Jr (1978). The nature of an initiator constitutive mutation in *Aspergillus nidulans*. *Nature,* **274**, 177–179.

Starich, M. R., Wikström, M., Arst, H. N., Jr, Clore, G. M. & Gronenborn, A. M. (1998). The solution structure of a fungal AREA protein-DNA complex reveals an alternative binding mode for the basic carboxyl tail of GATA factors. *J. Mol. Biol.* **277**, 605–620.

Swillens, S. (1995). Interpretation of binding curves obtained with high receptor concentrations: practical aid for computer analysis. *Mol. Pharmacol.* **47**, 1197–1203.

Wiame, J. M., Grenson, M. & Arst, H. N., Jr (1985). Nitrogen catabolite repression in yeasts and filamentous fungi. *Advan. Microb. Physiol.* **26**, 1–88.

Edited by P. E. Wright

(Received 25 November 1997; received in revised form 30 December 1997; accepted 6 January 1998)