The EMBO Journal Vol.18 No.2 pp.470–479, 1999

Structure–function analysis of the Z-DNA-binding domain Zα of dsRNA adenosine deaminase type I reveals similarity to the (α + β) family of helix–turn–helix proteins

Markus Schade[1,2], Christopher J.Turner[3], Ky Lowenhaupt[1], Alexander Rich[1,4] and Alan Herbert[1]

[1]Department of Biology and [3]Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and [2]Forschungsinstitut für Molekulare Pharmakologie, 10315 Berlin, Germany

[4]Corresponding author
e-mail: alan@mit.edu

RNA editing alters pre-mRNA through site-selective adenosine deamination, which results in codon changes that lead to the production of novel proteins. An enzyme that catalyzes this reaction, double-stranded RNA adenosine deaminase (ADAR1), contains two N-terminal Z-DNA-binding motifs, Zα and Zβ, the function of which is as yet unknown. In this study, multidimensional NMR spectroscopy was used to show that the topology of Zα is α1β1α2α3β2β3. Long-range NOEs indicate that β1 and β3 interact with each other. Site-directed mutagenesis was used to identify residues in α3, β3 and the loop connecting β2 to β3 that affect Z-DNA binding. Also identified were 11 hydrophobic residues that are essential for protein stability. Comparison with known structures reveals some similarity between Zα and (α + β) helix–turn–helix proteins, such as histone 5 and the family of hepatocyte nuclear factor-3 winged-helix–turn–helix transcription factors. Taken together, the structural and functional data suggest that recognition of Z-DNA by Zα involves residues in both the α3 helix and the C-terminal β-sheet.
Keywords: DNA recognition/helix–turn–helix/NMR/
RNA editing/scanning mutagenesis

Introduction

RNA editing modifies the linear flow of information from DNA to protein (Bass, 1993). It was discovered in the mitochondria of trypanosomes, where uridines are inserted in, or deleted from, certain RNAs (Benne *et al*., 1986). Numerous other examples of RNA editing, where one base is substituted for another, have been found in organisms of all eukaryotic phyla (Chen *et al*., 1987; Covello and Gray, 1989; Hoch *et al*., 1991) and their viruses (Polson *et al*., 1996). In one type of substitution RNA editing, adenosine is deaminated to inosine in a site-specific manner. Inosine is translated as guanosine. This process is used to modulate the activity of neural glutamate (Sommer *et al*., 1991) and serotonin receptors (Burns *et al*., 1997) and is altered during disease (Brusa *et al*., 1995). In these cases, editing is guided by an intron that folds back onto an exon to form a double-stranded RNA (dsRNA) editing substrate (reviewed in Maas *et al*., 1996). The use of introns in editing reactions requires that RNA modification must occur before pre-mRNAs are spliced. The site-specific editing of RNA is catalyzed both *in vitro* and *in vivo* by ADAR1 and ADAR2 (dsRNA adenosine deaminases type I and II) (Melcher *et al*., 1996). However, both ADAR1 and ADAR2 can be quite non-specific in their action. *In vitro*, any dsRNA longer than 30 bp can act as a substrate (Hurst *et al*., 1995; Maas *et al*., 1996), suggesting that auxiliary factors may normally be required to regulate activity. Consequently, the mechanisms for site-selection and regulation of editing efficacy *in vivo* are yet to be determined (Herbert, 1996).

At the N-terminus, ADAR1 contains two left-handed DNA (Z-DNA)-binding motifs: Zα and Zβ (Herbert *et al*., 1997). These domains are conserved among the human, rat, bovine and frog enzymes and are predicted to belong to the helix–turn–helix (HTH) family of proteins, most likely the winged-helix variety (Herbert *et al*., 1997). The high affinity of Zα for linear poly(dCdG) stabilized in the Z-DNA conformation by chemical bromination and for supercoiled plasmids containing Z-DNA inserts (Herbert *et al*., 1997) links ADAR1-mediated RNA editing to Z-DNA.

A possible role for Z-DNA may be to target editing by ADAR1 to sites of active transcription. Z-DNA can form *in vivo* in segments of alternating purine–pyrimidine sequences under conditions of torsional stress resulting from the movement of RNA polymerase through a gene (Singleton *et al*., 1982; Haniford and Pulleybank, 1983; Liu and Wang, 1987). Studies in both prokaryotes and eukaryotes show that Z-DNA formation *in vivo* is initiated by transcription (Lipps *et al*., 1983). Anti-Z-DNA antibody cross-linking experiments in metabolically active, agarose-embedded nuclei revealed transcription-dependent Z-DNA formation in segments of the c-*myc* gene (Wittig *et al*., 1992). Recognition of Z-DNA thus may allow ADAR1 to associate with pre-mRNA prior to splicing. Potential Z-DNA-forming sequences are present in many genes, predominantly in the 5′ regions where generation of negative superhelicity is expected to be greatest (Schroth *et al*., 1992). Z-DNA formation by these DNA segments may be one way of modulating the site selectivity and editing efficiency of ADAR1 *in vivo* (Herbert and Rich, 1996).

In order to gain a molecular understanding of how the regulation of ADAR1-mediated RNA editing relates to the recognition of Z-DNA, we used a twin-tracked approach of structure/function analysis. We determined the secondary structure of Zα by multidimensional NMR spectroscopy, and showed that the topology is α1β1α2α3β2β3. By mutagenesis, we found that residues that are pivotal for Z-DNA binding cluster on one side of α3, making it likely that they form Z-DNA contacts. In addition, we identified residues in α1, α2 and α3 that probably form a hydro-

phobic core essential for protein stability. Zα shows some structural similarity to the core fold of hepatocyte nuclear factor (HNF)-3-like winged-HTH DNA-binding proteins, in which α3 is used as a recognition helix.

Results

Secondary structure and topology

Multidimensional nuclear magnetic resonance (NMR) spectroscopy is a powerful technique used to determine the atomic structure of proteins with a mol. wt of <35 kDa. Here we determined the secondary structure of human Zα (residues 134–200; 9.4 kDa) by an NMR approach, using ^{15}N-labeled protein to record well-resolved three-dimensional (3D) ^{15}N-edited NMR experiments. A set of 3D ^{15}N-edited TOCSY-HSQC, HNHA and NOESY-HSQC experiments (Marion *et al*., 1989; Vuister and Bax, 1993) allowed complete backbone assignment and yielded a continuous network of short- and medium-range nuclear Overhauser effect (NOE) distance restraints. These restraints were sufficient to elucidate the secondary structure of residues 134–200. The high quality of the spectra is exemplified by 2D strips extracted from the 3D ^{15}N-edited NOESY-HSQC spectrum (Figure 1), illustrating the sequential and secondary structure assignment process for residues I172–L179. NOE intensities show characteristic patterns for α-helices and β-strands (Figure 2) (Wuethrich, 1986), yielding three α-helices and three β-strands. Coupling constants of HN–Hα are consistent with these secondary structure assignments (Figure 2).

Analysis of long-range NH–NH and Hα–NH NOE restraints in 3D ^{15}N-edited NOESY-HSQC spectra (70, 150 and 250 ms mixing time) revealed that residues 185–189 in β2 and residues 194–198 in β3 form an antiparallel β-sheet at the C-terminus (Table I; Figure 4). Backbone NOE restraints observed between T156 in β1 and L194 and W195 in β3 indicate that β3 also contacts β1, making the β-sheet triple-stranded.

Additional unambiguous long-range NOE constraints from the 3D ^{15}N-edited NOESY-HSQC spectra and also from a 2D NOESY build-up in D_2O restrain the tertiary fold of Zα further (Table I; Figure 4). Close contact between α3 and β3 is shown by the well-resolved side chain NOEs from the aromatic ring protons of W195 to the methyl protons of L176 and also to the aromatic protons of Y177. In addition, α2 folds against α1 to produce NOEs between the amide proton of L165 and the H2,6 protons of F146, as well as between the methyl protons of L165 and the amide proton of L147. Zα thus consists of a triple-stranded β-sheet in which β3 folds back against α3 and the N-terminus of α2. Helix α1 folds against α2. These results provide a coarse view of the tertiary structure and suggest that Zα has a topology similar to (α + β) HTH proteins such as HNF-3γ (Clark *et al*., 1993) (Figure 4).

Mutagenesis

The NMR-derived topology was set in a functional context by site-directed mutagenesis, a technique allowing high-resolution mapping of protein interactions with DNA (Cunningham and Wells, 1989; Brown *et al*., 1994). In many cases, alanine was substituted for the native residue. Alanine substitutions remove side chains beyond the β carbon, thereby eliminating hydrophobic, salt-bridged and hydrogen-bridged side chain interactions, while maintaining the peptide backbone. In addition, other amino acids were used to replace alanines present in Zα, and also to test the effect of more conservative substitutions on Zα function. A total of 44 amino acids between residues 139 and 198 of the human Zα domain were systematically replaced using PCR-based site-directed mutagenesis (Figure 3A). Mutant proteins were expressed as GST fusion proteins in *Escherichia coli* and purified by affinity chromatography, allowing recovery of full-length and proteolysed mutant protein. Mutant proteins were resolved by SDS–PAGE, blotted to a nitrocellulose membrane and probed for Z-DNA binding with radioactively labeled Z-DNA (Southwestern assay) (Figure 3B). Only protein with a fully intact Zα domain bound Z-DNA (data not shown). Subsequent probing of the blot with anti-GST made it possible to measure the extent of C-terminal proteolytic degradation, an indirect measure of how a mutation affects protein stability *in vivo* (Figure 3C). Each mutant was also tested in a bandshift assay using the same DNA probe. The bandshift assay allows more accurate quantitation and can be used over a 100-fold greater dilution range than the Southwestern. Dissociation constants (K_ds) for the most deleterious mutants were measured with a BIAcore instrument under equilibrium conditions using an immobilized Z-DNA ligand as a substrate. Concordant results between assays were found, except for some mutants discussed below.

Using the Southwestern assay, we identified four groups of mutants in Zα that have effects on Z-DNA binding and protein stability; first, a cluster of mutants that lie in the C-terminal β-sheet; secondly, a collection of mutants in α3 that alter Z-DNA binding but not protein stability; thirdly, a series of mutants that affect hydrophobic residues present in α1, α2 and α3 and disrupt protein stability; and fourthly, a number of mutants present in the turns between structural elements defined by NMR that reduce Z-DNA binding. A number of these mutants affect residues that are highly conserved between human, mouse, rat, bovine and *Xenopus* Zα sequences. Each set of mutants is discussed below.

The C-terminal β-sheet shows two distinct sides

The C-terminal β-sheet is composed of strands β2 and β3, connected by a four residue loop. The network of backbone NOEs between β2 and β3 shows that the side chains of K187, W195 and I197 are next to each other, projecting to one side of the β-sheet (side 1). The hydrophilic residues Q186, E188 and K196 project to the other side (side 2). Side chain NOEs show that side 1 of the β-sheet is in contact with α3.

Mutations to residues on side 1 of the β-sheet have profound effects. W195A is the most deleterious mutation of all Zα mutants made, regardless of the assay. Binding was diminished beyond detection and substantial proteolytic degradation was observed. These results confirm an important role for W195 in stabilizing Zα, and are consistent with the extensive long-range NOEs demonstrated for this residue by NMR (Table I; Figure 4). The more conservative W195Y mutant improved protein stability, but still caused an ~4-fold reduction in binding (bandshift assay, data not shown). In contrast, mutation

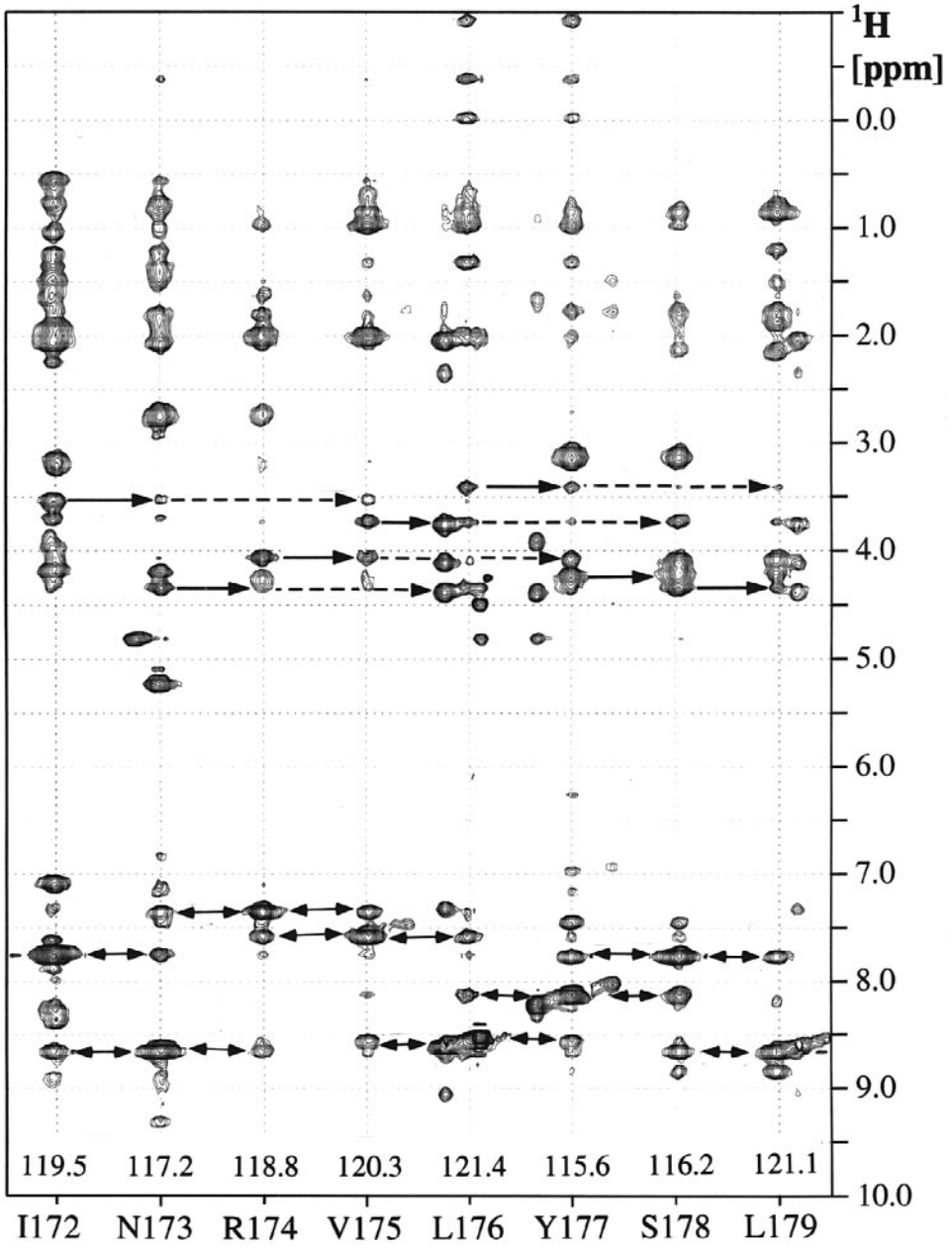

Fig. 1. Sequential and secondary structure assignment in α3. Two-dimensional strips of α3 residues extracted from a 3D ^{15}N-edited NOESY-HSQC spectrum of Zα depict sequential NOEs between amide protons (marked by double arrows in the 7–9.5 p.p.m. region), sequential Hα(i)→NH(i+1) NOEs (solid arrows in the 3.4–5.3 p.p.m. Hα region) and medium range Hα(i)→NH(i+3) NOEs (dashed arrows in the 3.4–5.3 p.p.m. Hα region) characteristic for α-helices. Backbone ^{15}N chemical shifts are shown above each residue label at the bottom.

of the other hydrophobic residue in β3, I197, to alanine had no effect on binding or protein stability. Changing the non-conserved residue K187 in β2 to alanine caused a mild but significant reduction in binding, raising the possibility that this flexible, positively charged side chain contacts DNA directly.

When residues on side 2 of the β-sheet were mutated, it was found that K196A enhanced Z-DNA binding while Q186A and E188A had wild-type protein stability and Z-DNA binding. Thus it appears that residues on this side of the β-sheet do not affect protein stability, but can influence DNA binding indirectly, for example by modifying the dipole moment of the β-sheet.

The loop (L1) between β2 and β3 contains two highly conserved residues: P192 and P193. Mutant P192A had significantly diminished binding, while P193A showed only a slight reduction. As prolines often bend the protein backbone to form loops, the effects of these proline mutants may arise indirectly by disrupting β-sheet folding crucial for interaction with Z-DNA. Thus, P192 could be important in stabilizing the fold while P193 only makes minor contributions. However, a direct interaction between these residues and Z-DNA is also possible and cannot be excluded by this study. Mutation of the non-conserved L1 residue, T191, to alanine also diminished binding, suggesting that the T191 side chain stabilizes loop structure, while L194G caused only a mild diminution in binding. The effects of mutating the other loop residue, G190, were not tested. Thus, of the four L1 residues mutated, all have some effect on DNA binding by Zα.

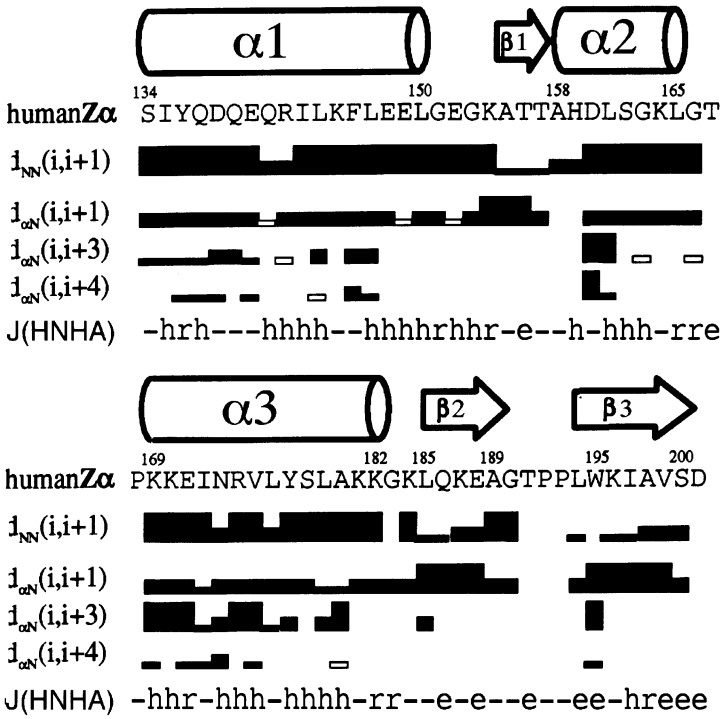

Fig. 2. The secondary structure of human Zα. The sequence of Zα from residues 134–200 is shown. d_{NN} and $d\alpha_N$ represent backbone NH–NH and Hα–NH NOEs and are indicated by strong (thick bar), medium (intermediate bar) and weak (thin bar) intensity. $d\alpha_N(i,i+3)$ and $d\alpha_N(i,i+4)$ NOEs are characteristic for α-helices, and strong $d\alpha_N(i,i+1)$ NOEs are diagnostic of β-strands. The line J(HNHA) shows the quantative analysis of backbone HN–Hα coupling constants. Coupling constants <6 Hz are characteristic of α-helices (identified by an h) while those >8 Hz suggest an extended chain conformation (marked with an e). Those residues with a coupling constant between 6 and 8 Hz are marked with an r.

Table I. Long-range NOEs in Zα

A β-sheet NOE				**B** Fold restraining NOE	
β2	β3	β1	β3		
L185.Hα	A198.NH	T156.NH	L194.Hα	W195.H6	Y177.H2/6
Q186.NH	A198.NH	T156.NH	W195.NH	W195.H5/7	L176.Hδ2
Q186.NH	I197.Hα	W195.Hβ1/2	T157.NH	W195.H2	A158.NH
Q186.NH	K196.NH			A158.Hβ	L176.Hδ2
K187.Hα	K196.NH			T156.NH	E152.Hα
E188.NH	W195.Hα			L165.NH	F146.H2/6
E188.NH	L194.NH			L165.Hδ1/2	L147.NH

The backbone NOE restraints observed in the C-terminal β-sheet and between β-strand β1 and β3 are summarized in A, e.g. the Hα of L185 and the backbone NH of A198 are connected through an NOE. B shows other unambiguous long-range NOEs that determine the tertiary fold, obtained from 3D ^{15}N-separated NOESY-HSQC and 2D NOESY (D_2O) build-up experiments. For each NH–NH NOE, two independent cross-peaks were observed.

α3 has characteristic properties of a recognition helix

Comparing results from NMR studies and alanine-scanning mutagenesis shows that a striking cluster of deleterious mutations occurs in α3 (residues 169–182). An α-helical net diagram (Figure 5) segmenting α3 into three faces is useful in interpreting these mutants. Face 1 contains charged and polar residues. Mutations to these residues diminish Z-DNA binding but do not significantly affect protein stability. Face 2 is composed of hydrophobic residues. Changes to these amino acids are poorly tolerated, and cause increased protein degradation within *E.coli*. Face 3 is occupied by residues that do not reduce DNA binding or affect protein stability. In some cases, mutation to these residues enhances Z-DNA binding. The three functionally distinct faces are discussed separately.

Residues K169, K170, N173, Y177 and K181 form face 1 (Figure 5). Residues K169, N173 and Y177 show a high degree of conservation among the human, rat, bovine and frog Zαs, suggesting that they have an important role in Zα function. Residue 181 (K in humans) is poorly conserved. In the Southwestern and bandshift assays, mutant N173A displays the most dramatic decrease in binding of all α3 mutations. Mutants K169A and Y177A show only a minor effect on binding in the Southwestern assay, but show significant loss of activity in the bandshift assay, while mutant K181A lacks activity in both the Southwestern and bandshift assays. CD spectra (data not shown) of these mutants are similar to wild-type, suggesting that secondary structure is also similar. When analyzed using BIAcore measurements, N173A reduces binding 160-fold, K169A 37-fold, Y177A 26-fold and

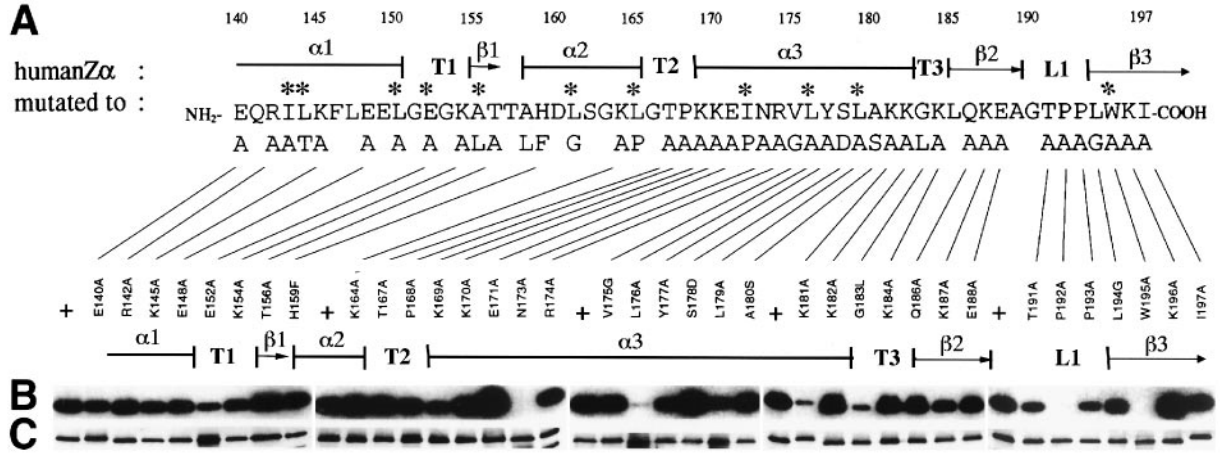

Fig. 3. Site-directed mutagenesis of human Zα. (**A**) Summary of mutants tested. Asterisks (*) indicate substantial (>80%) proteolytic degradation. (**B**) Southwestern assay. Autoradiogram of Z-DNA affinity assays, identifying those mutants that cause diminished Z-DNA binding. The panel is a collage of five different membranes. The lanes marked + are positive controls using wild-type Zα protein. (**C**) Western blot. Filters used for Southwestern assays were probed using antibodies to GST in order to quantitate protein loadings. Comparable band intensities show that equal amounts of full-length protein were assayed in (B) (Southwestern). Mutants sensitive to proteolysis, e.g. E152A, L176A and L179A, show a smearing downwards of the band, which results from overloading the SDS gel in order to ensure equal amounts of full-length protein. Because of the need to avoid extensive overloading of the gel, some mutants that exhibit severe proteolysis were excluded from the Southwestern.

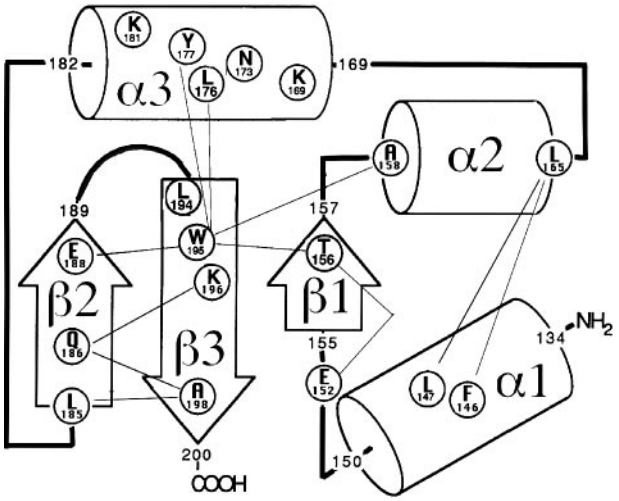

Fig. 4. The topology and candidate Z-DNA contacts of Zα. The data show that the topology and location of candidate DNA contacts of Zα are in some respects similar to those of histone H5 and HNF-3γ. The positions of α-helices and β-strands are indicated with boxes connected by thick lines. Numbers correspond to amino acid residues. Long-range NOEs given in Table I are indicated with thin lines and show the interactions between the C-terminal β-sheet and the α-helices of Zα. By analogy with HNF-3, β1 is shown as antiparallel to β3 because their relative orientation could not be determined unambiguously with the NOEs currently known. Residues (K169, N173, Y177, K181) on the face of α3 that contacts Z-DNA are indicated, as are the extensive contacts made by W195. W195A was the most deleterious mutant characterized.

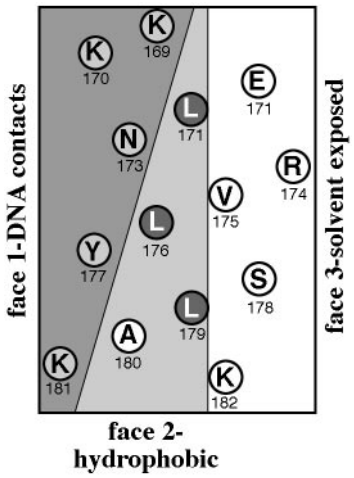

Fig. 5. α-helical net diagram of α3. The putative DNA-contacting face 1, the hydrophobic face 2 and the proposed solvent-exposed face 3 are colored dark gray, light gray and white, respectively. Mutations to K169, N173, Y177 and K181 affect DNA binding without affecting protein stability. Changes to I172, L176 and L179 greatly diminished protein stability. Both E171A and S178D increased Z-DNA binding.

K181A 4-fold, relative to the wild-type (Table II). The K_d for K181A is not reduced as much as would be predicted from bandshift and Southwestern assays. A possible reason for this discrepancy is suggested by inspection of the BIAcore sensorgram shown in Figure 6. During the dissociation phase of the measurement, K181A shows a faster drop in signal than wild-type protein. The rapid breakdown of protein–DNA complexes during electrophoresis may account for loss of bandshifting activity of K181A, while a similar breakdown during washing may explain the diminished binding seen in the Southwestern assay. The alignment of K169A, N173A, Y177A and K181A mutants in a row on one face of a helix is reminiscent of DNA contacts commonly found in the recognition helices of HTH DNA-binding proteins (Pabo and Sauer, 1992; Suzuki *et al.*, 1995). Therefore, the face 1 residues are good candidates to form direct or water-mediated DNA contacts.

Additional mutants were made to examine the nature of the interaction of K181 and Y177 with DNA. A K181R mutant binds indistinguishably from wild-type, suggesting that K181 recognizes Z-DNA in a flexible, non-restrained mode, e.g. through hydrogen bonds with the phosphate backbone. This result perhaps explains why K181 is replaced by a charged residue such as arginine or glutamic acid in other family members. A Y177K mutant was

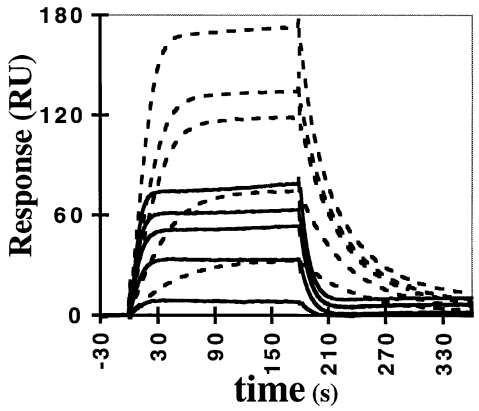

Fig. 6. BIAcore sensorgrams of wild-type and mutant K181A. For wild-type Zα (dashed lines), the curves represent from top to bottom injections of 100, 75, 50, 25 and 10 nM protein. For K181A, the curves (solid line) show injections of 100, 80, 60, 40 and 10 nM protein. Response units (RU) are an arbitrary measurement that is proportional to the mass of protein bound to Z-DNA polymer immobilized on the chip surface. The >2-fold lower steady-state response (plateau region) of mutant K181A reflects its lower affinity for Z-DNA. The rapid drop of response signal in the dissociation phase of K181A suggests a rapid breakdown of the DNA–protein complex.

Table II. Binding affinities and rates relative to wild-type

Protein	Relative K_d
Wild-type	1
K169A	37
N173A	168
L176A	43
Y177A	26
K181A	4
P192A	13
W195A	>10 000
Engrailed	286

The K_d was determined using BIAcore measurements to determine the amount of protein bound to a Z-DNA-coated chip surface after steady state was reached. Engrailed protein was used as a negative control for non-specific protein binding. The binding by Engrailed exceeded that of the most deleterious mutant W195A, for which a K_d could not be determined.

studied because the side chain of lysine is flexible enough to reach any putative hydrogen-bonding partner that interacts with the hydroxyl of tyrosine. Mutant Y177K bound better than Y177A, but still worse than wild-type, showing that lysine can mimic some but not all the binding contributions mediated by the tyrosine side chain. In addition to hydrogen bonding via its hydroxyl group, Y177 is also likely to make essential contacts through its rigid aromatic ring system. The importance of hydrophobic contacts is suggested by the substitution of Y177 with another hydrophobic residue, isoleucine, in Zβ. Taken together, the mutational analysis revealed that residues K169, N173, Y177 and K181 are pivotal to binding, suggesting that face 1 of α3 interacts directly with Z-DNA and thus is a recognition helix.

The hydrophobic residues on face 2, I172, L176 and L179, are also highly conserved, with L176 being absolutely conserved between Zα and Zβ of all species. The result of mutating each of these residues was determined. A I172P mutation disrupted the protein fold so severely that no full-length protein could be recovered from *E.coli*. A more conservative L176A mutation also caused extreme proteolytic degradation, allowing recovery of only 5% full-length protein. Even after correcting for the reduced yield of full-length protein, Z-DNA binding by this mutant was diminished 50-fold (Table II). Replacement of the third conserved residue, L179, with alanine results in slightly less protein degradation than L176A (~10% is full-length protein) and Z-DNA binding assayed by Southwestern and bandshift assays was somewhat better. These results show that mutation of hydrophobic residues on face 2 of α3 causes tremendous proteolytic degradation, suggesting that they form part of a hydrophobic protein core stabilizing Zα.

Face 3 of α3 carries three charged and one polar amino acid, none of which are conserved. Alanine mutants of the two positively charged residues R174 and K182 were indistinguishable from wild-type. The mutants E171A and S178D revealed a slightly enhanced Z-DNA affinity. Thus, these four mutants on face 3 are dispensable for Z-DNA binding and are probably solvent exposed.

α1 and α2 have essential hydrophobic faces

Mutation of α1 and α2 revealed conserved residues that form a hydrophobic face on each helix. In α1, the mutations I143A and L144T disrupt both binding and protein stability, while the L150A mutation increases proteolytic degradation without affecting binding. Together with L147, which was not investigated, these three residues make up a hydrophobic face on α1, suggesting that they are involved in contacts that correctly align this face within the protein. In contrast, mutation of α1 residues that lie on other faces of the helix, such as R142A, K145A and E148A, had no effect.

A similar face of hydrophobic residues is present in α2. Mutants A158L, L161G and L165P were found to be deleterious to both binding and protein stability. Replacement of A158 with a bulky leucine residue caused a milder phenotype than the severely destructive L161G and L165P mutations, suggesting that the environment of A158 at the joint of β1 and α2 is less constrained than that of the two leucines. In summary, the hydrophobic faces of α1 and α2 probably contribute to a hydrophobic core essential for correct folding of Zα, similar to that found in other HTH proteins (Pabo and Sauer, 1992; Suzuki *et al*., 1995).

Essential turn residues

Another class of mutants maps to the turns deduced from the NMR secondary structure of Zα. T1 connects α1 and β1 and contains residues E152 and K154. Mutant E152A clearly has diminished binding and protein stability, suggesting that contacts formed by the glutamate side chain are essential for correct folding of Zα. NOEs that show contact between this residue and T156, which lies in β1 (Table I; Figure 4), are consistent with such a role for E152. The mutant K154A binds Z-DNA almost as tightly as wild-type. T2, between α2 and α3, consists of residues G166, T167 and P168, of which only the latter showed a minimal reduction in binding upon mutation to alanine, indicating that the main chain bend of P168 makes some contribution to the turn. Direct or water-mediated contact of these side chain residues with DNA, as seen with other

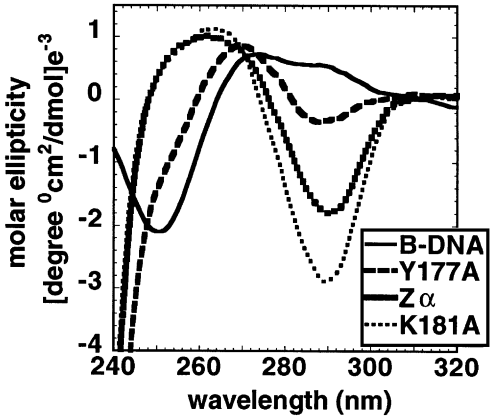

Fig. 7. Circular dichroism (CD) spectra showing the B–Z transition of poly(dCdG) stabilized by Zα, Y177A and K181A. The CD titration of poly(dCdG) with α3 face 1 mutants shows differences in their ability to promote formation of Z-DNA. The B-DNA spectrum of the polymer in the absence of protein is shown as a reference. The transition to Z-DNA in the presence of protein is shown by a loss of the negative peak at 250 nm and the appearance of a trough at 294 nm. Wild-type Zα, Y177A and K181A were used at 1:1 for base pair:protein stochiometry (50 μM) and spectra were measured after a 10 min equilibration period. The spectrum between 240 and 320 nm is presented to show changes in DNA conformation. Protein alone has a strong negative spectrum below 250 nm, but does not show a signal between 250 and 320 nm. Thus no correction for protein absorption has been made to the spectra shown.

(α + β) HTH proteins (Kaufmann and Knoechel, 1996), is unlikely. T3 residues G183 and K184 link the antiparallel β-sheet to α3. G183 is absolutely conserved in Zα and Zβ, while K184 is variable. The NMR data suggest that T3 must form a sharp turn to enable the β-sheet to fold back against α3. Mutagenesis supports this suggestion. The introduction of a bulky leucine residue, which reduces the main chain flexibility, in place of G183 diminished Z-DNA binding substantially.

Circular dichroism (CD) analysis of Y177A and K181A

To understand further the effect of Y177 and K181 on the interaction of Zα with Z-DNA, CD experiments were performed to test the effects of these mutants on the B–Z transition of poly(dCdG). The spectra collected after 10 min of incubation at 30°C for wild-type, K181A and Y177A at a molar ratio of one peptide to one base pair are shown in Figure 7. Also shown is the reference spectrum for the polymer in the B-DNA conformation obtained in buffer without any added protein. The spectra have not been corrected for absorption by the protein because this is absent above 250 nm and does not affect the interpretation of the DNA structural transition (data not shown). Below 250 nm, the signal from the peptide predominates. In Figure 7, the region between 240 and 320 nm is presented, showing the DNA transition from B-DNA to Z-DNA in the presence of protein (Herbert *et al*., 1997). The spectrum of Y177A shows a significantly reduced B- to Z-DNA transition compared with wild-type, while the K181A spectrum shows an enhancement. These results are somewhat paradoxical as it was expected that wild-type protein with the highest binding constant should be more effective in stabilizing Z-DNA than either of the mutants. However, CD is very sensitive to changes in the close environment of DNA, and these results may reflect slight variations in the Z-DNA conformation stabilized by each protein. Alternatively, K181A, which has only a 4-fold reduced affinity for Z-DNA (relative to Zα) and is used at high concentration in this assay (50 μM in Figure 7), may be more effective in promoting the B–Z transition of the polymer than wild-type.

Discussion

In this structure/function study, we determined the secondary structure of Zα, and assigned functions to the various motifs found in Zα. Evidence is provided suggesting that Zα belongs to the (α + β) HTH family of DNA-binding proteins, showing a similar arrangement of α-helices and β-sheets to that found in histone H5 (GH5) (Ramakrishnan *et al*., 1993) and HNF-3γ (Clark *et al*., 1993). A sequence alignment of Zα with HNF-3γ and GH5 (Figure 8) shows that helical regions have a similar distribution of hydrophobic residues. The hydrophobic residues in Zα that were shown by mutagenesis to be essential for protein stability (Figure 3) align well with the residues in HNF-3γ shown by X-ray crystallography to lie within its hydrophobic protein core (Figure 8). Furthermore, the mutagenesis data suggest that α3 of Zα acts as a recognition helix for Z-DNA, playing a role analogous to that of α3 in HNF-3γ and GH5 in the recognition of B-DNA. Zα, HNF-3γ and GH5 also appear to have C-terminal triple-stranded β-sheets that are similar in topology. In Zα and HNF-3γ, a conserved tryptophan present in β3 makes contact with α3. By analogy with HNF-3γ and consistent with the mutagenesis data presented here, the tryptophan of Zα may be hydrogen bonded with DNA phosphates through its imino group. The extensive contacts made by W195 shown in Figure 4 indicate that this residue is also essential for stabilizing the Zα fold. All of the above results suggest that the global protein fold of Zα resembles that of (α + β) HTH family members HNF-3γ and GH5. Detailed structural studies are under way to probe this relationship further.

Each member of the (α + β) HTH family has a unique orientation of the recognition helix to DNA, arising from differences in primary sequence, helix length, the number of residues in loops and the composition of β-sheets. As a consequence, this family of proteins can bind a wide variety of sequences and right handed DNA conformations, including bent DNA (Schulz *et al*., 1991; Pierrou *et al*., 1994; Escalante *et al*., 1998). The primary sequence of family members is quite divergent. For example, GH5 and HNF-3γ lack significant sequence homology, having identity in only nine of 72 residues (Figure 8; Clark *et al*., 1993). However, their Cα backbone shows a root mean standard deviation of 1.3 Å. Zα is the first family member known to bind left handed Z-DNA; Zα is not closely related to either GH5 or HNF-3/fork head proteins at the primary sequence level (Figure 8), even though they appear to share a common protein fold. Zα has many other significant differences from these two proteins. These differences are most apparent when it is compared with HNF-3γ, for which the structure bound to DNA is available (Clark *et al*., 1993). For example, the loop in HNF-3γ between α2 and the recognition helix α3 is eight residues while the turn T2 of Zα is only three. This loop in HNF-

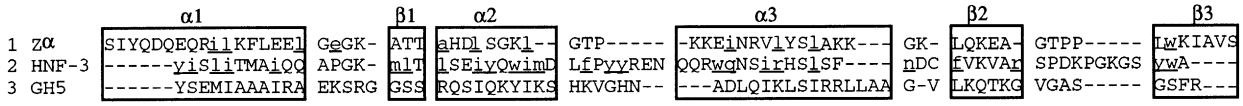

Fig. 8. A primary sequence alignment of Zα with HNF-3γ and GH5 using the secondary structure information described here, and the crystal structure data presented in Figure 5b of Clark *et al*. (1993). The residues in lower case and underlined in the Zα sequence are those that greatly affect protein stability (mutation of these residues results in <20% recovery of full-length protein from *E.coli*). HNF-3γ residues shown in lower case and underlined are buried amino acids contributing to its hydrophobic protein core.

3γ is important in determining sequence specificity and for the relative orientation of α3 relative to α2 (Kaufmann and Knoechel, 1996), but there is no evidence that the side chains of T2 residues in Zα are involved in DNA recognition. In addition, binding by HNF-3γ to DNA involves a single residue in the eight residue loop between β2 and β3 (called W1, equivalent to L1 in Zα) and a number of residues in the C-terminal loop, W2, which extends beyond the C-terminal β-sheet. Zα does not appear to require a region equivalent to W2 for tight binding to Z-DNA, while mutation of three of the four residues in L1 (G190 was not tested) affects DNA binding. Furthermore, in Zα, both the N- and C-termini of the putative recognition helix, α3, have lysine residues that affect binding to Z-DNA, whereas in HNF-3γ there is no equivalent. The above differences between Zα and HNF-3γ are consistent with their distinct modes of interaction with left- and right-handed DNA.

It is thus possible that (α + β) HTH motifs recognize Z-DNA in addition to B-DNA. A wide range of biological processes are regulated by the (α + β) family of HTH proteins, such as embryogenesis, tumorigenesis, longevity, insulin-related metabolic pathways and maintenance of differentiated cell states (Kaufmann and Knoechel, 1996; Ogg *et al*., 1997). The relationship of Zα to this family of proteins raises the intriguing question of whether other members of this family also bind to the left-handed DNA conformation.

Materials and methods

Protein preparation and Southwestern assay

Site-specific Zα mutants were constructed by PCR-based site-directed mutagenesis (QuikChange site-directed mutagenesis kit, Stratagene, La Jolla, CA) and checked by DNA sequencing. Mutant proteins containing residues 120–197 of human ADAR1 (DDBJ/EMBL/GenBank accession No. U10439) were expressed as GST-fusion proteins (GST is N-terminal) in pGEX-5×1 vectors in *E.coli* DH-5α and purified as described before (Herbert *et al*., 1997). The GST fusion mutant proteins were probed for binding to Z-DNA by Southwestern assays as described (Herbert *et al*, 1995). The binding affinities from Southwestern assays were verified by bandshift assays, which were performed as described (Herbert and Rich, 1993) with both GST fusion mutants and with mutants whose GST affinity tag was removed by factor Xa on-column digestion in 20 mM HEPES (pH 8), 100 mM NaCl, 2 mM $CaCl_2$, 0.05% *n*-octylglucopyranoside at 22°C. Mutant proteins showed a varying extent of proteolytic degradation *in vivo*. The degree of proteolysis of each mutant was determined by Western blots (developed with anti-GST antibody) on the same membrane used for Southwestern assays (Herbert *et al*., 1995). Only full-length protein bound to Z-DNA, so the amount of full-length protein used for Southwestern assays was adjusted to give Western blot bands of equal intensity.

Mutant proteins for BIAcore assays and CD analysis (ADAR1 residues 122–197) were expressed in pET-28a vectors. The pET-28a-encoded construct differs from the pGEX-5×1-derived peptides in having a stop codon after I197, deleting 12 residues encoded by the pGEX-5×1 polylinker. In addition, there are five vector-encoded residues in the N-terminus (GSHMG) compared with three in pGEX-5×1 vectors (GIP). Proteins were expressed in *E.coli* strain Novablue and purified in two steps. First, the *E.coli* lysate was passed over a His tag affinity column (TALON Metal Affinity Resin, Clontech, Palo Alto, CA) in 20 mM Tris–HCl (pH 8) 150 mM NaCl, using a 0–300 mM imidazole gradient for elution. The His tag was removed by thrombin digestion in the same buffer. For the second purification step, protein was loaded on a cation exchange chromatography column (Mono S 5/S, Pharmacia) in 50 mM HEPES (pH 7.4), 25 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), 0.125 mM phenylmethylsulfonyl fluoride (PMSF) and eluted with a 25–1000 mM NaCl gradient, yielding homogeneous protein of wild-type molecular weight, as indicated by SDS gel analysis.

For NMR, a different pET-28a construct was used. This plasmid encoded residues 119–200 of ADAR1. The vector added two other residues, a serine at the N-terminus and an aspartic acid at the C-terminus. Purification was as for the other pET-28a constructs described above.

BIAcore analysis

Binding kinetics were measured on a BIAcore 2000 instrument as described (Herbert *et al*., 1997). Briefly, 270 response units (RU) of poly(dCdG) stabilized in the Z-DNA conformation by chemical bromination was immobilized on the chip surface through a biotin linkage. Protein solutions at different concentrations were then passed over the chip surface with a flow rate of 20 μl/min, and the change in surface plasmon resonance was determined at a data collection rate of 1 Hz. RU are an arbitrary measurement that is proportional to the mass of protein bound to the chip surface. Additionally, non-specific background binding to the chip surface was eliminated by subtracting the baseline obtained by traversing the protein solution in line over a second chip surface lacking immobilized DNA. The Zα protein does not bind to a random sequence B-DNA polymer of the same size and immobilized at the same density as the Z-DNA probe (Herbert *et al*., 1997). The protein concentrations used covered a >10-fold concentration range around the K_d of a mutant.

Binding curves were analyzed by non-linear least square steady-state fitting to Langmuir isotherms. Because on-rates varied substantially, some mutants showed mass transport effects, while others did not. Therefore, steady-state fitting-derived binding constants were considered more robust (Hallet *et al*., 1990; Ward *et al*., 1995) and were used to calculate relative affinities. The Zα protein (residues 122–197) derived from the pET-28a constructs showed an equilibrium K_d of 70 nM ± 30%. This K_d deviates about an order of magnitude from data obtained previously using the pGEX-5×1 constructs (Herbert *et al*., 1997). The Zα construct used for these experiments stops after residue 197, lacking the most C-terminal residues 198–200 of the β-sheet, while the construct used previously carries 12 pGEX-5×1 vector-encoded residues after residue 197. These vector-encoded residues may substitute for the wild-type residues in β-sheet III, resulting in a higher affinity. The effect of the five additional N-terminal vector-encoded residues in pET-28a constructs is unknown.

CD analysis

Protein CD spectra were recorded at 5 μM in 10 mM Na phosphate (pH 7.2), 137 mM NaCl and H_2O at 25°C on an Aviv model 60DS spectrometer using a cuvette with a 5 mm pathlength. For CD titrations, poly(dCdG) (50 μM bp concentration) was titrated with mutant and wild-type Zα in 1:6, 1:3 and 1:1 (mol bp):(mol Zα) ratios at 30°C in 50 mM Tris–HCl (pH .4), 25 mM NaCl and 0.1 mM EDTA.

NMR spectroscopy

Zα protein for NMR experiments (ADAR1 residues 119–200) was prepared in the same way as that used for BIAcore assays. ^{15}N-labeled protein was expressed in *E.coli* strain HM174(DE3) in M9 minimal medium containing 1 g/l $^{15}NH_4Cl$. 3D HNHA-, 3D TOCSY-HSQC and 3D NOESY-HSQC experiments were acquired with 1.8 mM samples in 10 mM Na phosphate (pH 5), 137 mM NaCl, 0.1 mM NaN_3, 10% D_2O at 25°C on 500, 600 and 750 MHz NMR spectrometers (Francis Bitter Magnet Laboratory, MIT, MA). Spectra were processed and analyzed

with the programs NMRpipe (Delaglio *et al*., 1995), Pipp and Plotseq (Garrett *et al*., 1991). In addition to the NH–NH and Hα–NH NOE restraints shown in Figure 2, Hβ–NH NOEs and further side chain–NH NOEs were used to verify the sequential assignment. For determination of long-range NOE constraints, a 2D NOESY build-up in D_2O with 20, 40, 80 and 150 ms mixing time was recorded on a sample of unlabeled Zα protein under the buffer conditions described above.

HNHα coupling constants, J(HNHA), were determined from a 3D ^{15}N-edited HNHA spectrum (ξ = 14 ms) by the method of Vuister and Bax (1993) using integrated diagonal and cross-peaks. Apparent coupling constants were multiplied by a factor of 1.11 to correct for Hα spin flips (Vuister and Bax, 1993). The J(HNHA) values were sorted into three groups: (i) J(HNHA) <6 Hz being diagnostic of α-helical secondary structure; (ii) J(HNHA) >8 Hz being diagnostic of extended chain conformation; and (iii) J(HNHA) between 6 and 8 Hz (Markley *et al*., 1998).

Acknowledgements

We thank Frank Delaglio, Dan Garrett and Stephan Grzesiek and the Laboratory of Chemical Physics at the National Institutes of Health for making available NMR analysis software, and Ernest Fraenkel for providing purified Engrailed protein. We thank Thomas Schwartz for his help in overexpressing pET28a constructs and for his critical reading of the manuscript. We also thank Yang-Gyun Kim for instruction on how to use the BIAcore instrument. Grant support was obtained from the National Institutes of Health, the National Science Foundation the Office for Naval Research and the Boehringer Ingelheim Fonds. The NMR facility is supported by NIH grant RR-00995.

References

Bass,B.L. (1993) *RNA Editing: New Uses for Old Players in the RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Benne,R., Burg van den,J., Brakenhoff,J.P., Sloof,P., Van Boom,J.H. and Tromp,M.C. (1986) Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria. *Cell*, **46**, 819–826.

Brown,B.M., Milla,M.E., Smith,T.L. and Sauer,R.T. (1994) Scanning mutagenesis of the Arc repressor as a functional probe of operator recognition. *Nature Struct. Biol*., **1**, 164–168.

Brusa,R., Seeburg,P.H. and Sprengel,R. (1995) Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science*, **270**, 1677–1680.

Burns,C.M., Chu,H., Rueter,S.M., Hutchinson,L.K., Canton,H., Sanders-Bush,E. and Emeson,R.B. (1997) Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*, **387**, 303–308.

Chen,S.H. *et al*. (1987) Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science*, **238**, 363–366.

Clark,K.L., Halay,E.D., Lai,E. and Burley,S.K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, **364**, 412–420.

Covello,P.S. and Gray,M.W. (1989) RNA editing in plant mitochondria. *Nature*, **341**, 662–666.

Cunningham,B.C. and Wells,J.A. (1989) High-resolution epitope mapping of hGH–receptor interactions by alanine-scanning mutagenesis. *Science*, **244**, 1081–1085.

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

Egli,M. and Gessner,R.V. (1995) Stereoelectronic effects of deoxyribose O4′ on DNA conformation. *Proc. Natl Acad. Sci. USA*, **92**, 180–184.

Escalante,C., Yie,J., Thanos,D. and Aggarwal,A. (1998) Structure of IRF-1 with bound DNA reveals determinents of interferon regulation. *Nature*, **391**, 103–106.

Garrett,D.S., Powers,R., Gronenborn,A.M. and 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. Resonance*, **95**, 214–222.

Hall,D.R., Cann,J.R. and Winzor,D.J. (1996) Demonstration of an upper limit to the range of association rate constants amenable to study by biosensor technology based on surface plasmon resonance. *Anal. Biochem*., **235**, 175–184.

Haniford,D.B. and Pulleybank,D.E. (1983) Facile transition of poly[d(TG)·d(CA)] into a left-handed helix in physiological conditions. *Nature*, **302**, 632–634.

Herbert,A., Lowenhaupt,K., Spitzner,J. and Rich A. (1995) Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc. Natl Acad. Sci. USA*, **92**, 7550–7554.

Herbert,A., Alfken,J., Kim,Y.G., Mian,I.S., Nishikura,K. and Rich,A. (1997) A Z-DNA binding domain present in the human editing enzyme double-stranded RNA adenosine deaminase. *Proc. Natl Acad. Sci. USA*, **94**, 8421–8426.

Herbert,A. and Rich,A. (1996) The biology of left-handed Z-DNA. *J. Biol. Chem.*, **271**, 11595–11598.

Herbert,A.G. (1996) RNA editing, introns and evolution. *Trends Genet.*, **12**, 6–9.

Herbert,A.G. and Rich,A. (1993) A method to identify and characterize Z-DNA binding proteins using a linear oligodeoxynucleotide. *Nucleic Acids Res.*, **21**, 2669–2672.

Hoch,B., Maier,R., Appel,K., Igloi,G.L. and Kossel,H. (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, **353**, 178–180.

Hurst,S.R., Hough,R.F., Aruscavage,R.J. and Bass,B.L. (1995) Deamination of mammalian glutamate receptor RNA by *Xenopus* dsRNA adenosine deaminase: similarities to *in vivo* RNA editing. *RNA*, **1**, 1051–1060.

Kaufmann,E. and Knoechel,W. (1996) Five years on the wings of fork head. *Mech. Dev.*, **57**, 3–20.

Lipps,H.J., Nordheim,A., Lafer,E.M., Ammermann,D., Stollar,B.D. and Rich,A. (1983) Antibodies against Z-DNA react with the macronucleus but not the micronucleus of the hypotrichous ciliate *Stylonychia mytilus*. *Cell*, **32**, 435–441.

Liu,L.F. and Wang,J.C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl Acad. Sci. USA*, **84**, 7024–7027.

Maas,S., Melcher,T., Herb,A., Seeburg,P., Keller,W., Krause,S., Higuchi,M. and O'Connell,M.A. (1996) Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J. Biol. Chem.*, **271**, 12221–12226.

Marion,D., Driscoll,P.C., Kay,L.E., Wingfield,P.T., Bax,A., Gronenborn, A.M. and Clore,G.M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H–15N Hartmann–Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy application to interleukin 1β. *Biochemistry*, **28**, 6150–6156.

Markley,J., Bax,A., Arata,Y., Hilbers,C., Kapstein,R., Sykes,B. and Wuthrich,K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Mol. Biol*., **280**, 933–952.

Melcher,T., Maas,S., Herb,A., Sprengel,R., Seeburg,P.H. and Higuchi,M. (1996) A novel mammalian RNA editing enzyme. *Nature*, **379**, 460–464.

Ogg,S., Paradis,S., Gottlieb,S., Patterson,G.I., Lee,L., Tissenbaum,H.A. and Ruvkun,G. (1997) The fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C.elegans*. *Nature*, **389**, 994–999.

Pabo,C.O. and Sauer,R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.

Pierrou,S., Hellqvist,M., Samuelsson,L., Enerback,S. and Carlsson,P. (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. *EMBO J.*, **13**, 5002–5012.

Polson,A.G., Bass,B.L. and Casey,J.L. (1996) RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature*, **380**, 454–456.

Ramakrishnan,V., Finch,J.T., Graziano,V., Lee,P.L. and Sweet,R.M. (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**, 219–223.

Schroth,G.P., Chou,P.J. and Ho,P.S. (1992) Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of Z-DNA-forming sequences in human genes. *J. Biol. Chem.*, **267**, 11846–11855.

Schulz,S., Shields,G. and Steitz,T. (1991) The crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees. *Science*, **253**, 1001–1007.

Singleton,C.K., Klysik,J., Stirdivant,S.M. and Wells,R.D. (1982) Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature*, **299**, 312–316.

Sommer,B., Kohler,M., Sprengel,R. and Seeburg,P.H. (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell*, **67**, 11–19.

Suzuki,M., Brenner,S., Gerstein,M. and Yagi,N. (1995) DNA recognition code of transcription factors. *Protein Eng.*, **8**, 319–328.

Vuister,W. and Bax,A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J (HNHA) coupling constants in N15-enriched proteins. *J. Am. Chem. Soc.*, **115**, 7772–7777.

Ward,L.D., Howlett,G.J., Hammacher,A., Weinstock,J., Yasukawa,K., Simpson,R.J. and Winzor,D.J. (1995) Use of a biosensor with surface plasmon resonacnce detection for the determination of binding constants: measurement of interleukin-6 binding to the soluble interleukin-6 receptor. *Biochemistry*, **34**, 2901–2907.

Wittig,B., Wolfl,S., Dorbic,T., Vahrson,W. and Rich,A. (1992) Transcription of human c-*myc* in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J.*, **11**, 4653–4663.

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

Received June 6, 1998; revised and accepted November 17, 1998