Protein Science (1993), *2*, 951–965. Cambridge University Press. Printed in the USA.

Structures of DNA-binding mutant zinc finger domains: Implications for DNA binding

ROSS C. HOFFMAN,[1,3] SUZANNA J. HORVATH,[2] AND RACHEL E. KLEVIT[1]

[1] Department of Biochemistry, University of Washington, Seattle, Washington 98195
[2] Division of Biology, California Institute of Technology, Pasadena, California 91125

(RECEIVED September 10, 1992; REVISED MANUSCRIPT RECEIVED February 12, 1993)

Abstract

Studies of Cys_2–His_2 zinc finger domains have revealed that the structures of individual finger domains in solution determined by NMR spectroscopy are strikingly similar to the structure of fingers bound to DNA determined by X-ray diffraction. Therefore, detailed structural analyses of single finger domains that contain amino acid substitutions known to affect DNA binding in the whole protein can yield information concerning the structural ramifications of such mutations. We have used this approach to study two mutants in the N-terminal finger domain of ADR1, a yeast transcription factor that contains two Cys_2–His_2 zinc finger sequences spanning residues 102–159. Two point mutants at position 118 in the N-terminal zinc finger (ADR1b: 102–130) that adversely affect the DNA-binding activity of ADR1 have previously been identified: H118A and H118Y. The structures of wild-type ADR1b and the two mutant zinc finger domains were determined using two-dimensional nuclear magnetic resonance spectroscopy and distance geometry and were refined using a complete relaxation matrix method approach (REPENT) to improve agreement between the models and the nuclear Overhauser effect spectroscopy data from which they were generated. The molecular architecture of the refined wild-type ADR1b domain is presented in detail. Comparisons of wild-type ADR1b and the two mutants revealed that neither mutation causes a significant structural perturbation. The structures indicate that the DNA binding properties of the His 118 mutants are dependent on the identity of the side chain at position 118, which has been postulated to make a direct DNA contact in the wild-type ADR1 protein. The results suggest that the identity of the side chain at the middle DNA contact position in Cys_2–His_2 zinc fingers may be changed with impunity regarding the domain structure and can affect the affinity of the protein–DNA interaction.

Keywords: DNA binding; molecular architecture; NMR; protein–DNA interaction; zinc finger

Analysis of the deduced amino acid sequence of the *Xenopus* transcription factor TFIIIA in the mid-1980s revealed that it contains nine tandem repeats of the consensus sequence (Y,F)-X-C-$X_{2,4}$-C-X_3-F-X_5-L-X_2-H-$X_{3\text{-}5}$-H-$X_{2\text{-}6}$ (Brown et al., 1985; Miller et al., 1985). It was further noted that when purified in the absence of metal-chelating agents, 7–11 moles of zinc co-purified per mole of protein (Miller et al., 1985) and that zinc was required for sequence specific DNA binding (Hanas et al., 1983). These and other findings were synthesized in the zinc finger hypothesis that proposed each of the consensus repeats binds a zinc ion through the conserved cysteine and histidine residues to form a stable structural domain that is responsible for the DNA-binding activity of the transcription factor (Brown et al., 1985; Miller et al., 1985). Since that time, a large number of Cys_2–His_2 zinc finger sequences have been identified in eukaryotic DNA-binding proteins known to regulate transcription in either a positive or negative fashion (for review, see El-Baradi & Peiler [1991]). Hybridization cloning based on the consensus sequence has also revealed a growing number of proteins with no characterized function that contain zinc finger sequences (Berg, 1990).

Initial biophysical studies utilizing EXAFS on TFIIIA (Diakun et al., 1986) and a combination of circular di-

Reprint requests to: Rachel E. Klevit, Department of Biochemistry, University of Washington, Seattle, Washington 98195.
[3] Present address: Immunex Research and Development Corporation, Department of Protein Chemistry, 51 University Street, Seattle, Washington 98101.

Abbreviations: EXAFS, extended X-ray absorption fine structure; 2D-NMR, two-dimensional nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; P.E.COSY, primitive exclusive correlation spectroscopy; H-bonds, hydrogen bonds; COSY, correlation spectroscopy; RELAY, relayed coherence transfer spectroscopy; TOCSY, total correlation spectroscopy; RMS, root mean square.

chroism and visible absorption spectroscopy on a single zinc finger domain isolated from TFIIIA by cleavage with trypsin (Frankel et al., 1987) indicated that zinc fingers fold in a metal-dependent manner into thermostable domains with tetrahedral coordination of the metal by cysteine and histidine residues. Models for the three-dimensional zinc finger structure were proposed by combining structural elements from metalloproteins with sequence homology to the zinc fingers (Berg, 1988; Gibson et al., 1988). The essential features of these models were confirmed in the first experimentally derived model of a zinc finger domain from yeast ADR1 using 2D-NMR (Párraga et al., 1988) and then in more detail in the structure of a finger domain from Xfin calculated from NMR data (Lee et al., 1989). The structure consists of an antiparallel loop with a reverse turn that includes the two cysteine residues packed against an α-helix that includes the two histidine residues. This arrangement brings the conserved aromatic and leucine residues into close proximity with one another. The structures of a number of single and multiple zinc finger domains have now been determined and all exhibit the topology described above (Lee et al., 1989; Klevit et al., 1990; Omichinski et al., 1990, 1992; Kochoyan et al., 1991a–d; Pavletich & Pabo, 1991; Neuhaus et al., 1992).

Although the global fold of the zinc finger domain is nearly the same in all published structures, variations in local structure have been noted. Unfortunately, it is not known whether such variations represent real sequence-dependent differences or are due to differences in the structure determination methods used. The present study permits a detailed comparison of three zinc finger domain structures generated by identical procedures. Structures of the zinc finger domains were determined using ^{1}H-homonuclear 2D-NMR spectroscopy, distance geometry, and a refinement protocol that uses a complete relaxation matrix approach called REPENT (Hoffman et al., 1993). The refinement of NMR protein structures to improve agreement between the observed and model-calculated NOE intensities is not yet commonplace. Only two protein structures have been refined in this manner to date: motilin (Edmondson et al., 1991) and a type II collagen-binding domain (Constantine et al., 1992).

ADR1 is a two-zinc finger transcription factor from *Saccharomyces cerevisiae* that acts as a positive regulator of *ADH2* transcription (Denis & Young, 1983). The protein contains 1,323 amino acids with residues 102–159 comprising two Cys_2–His_2 zinc finger sequences. Mutations at residue 118 have been identified that affect the ability of ADR1 to bind to UAS1, its cognate DNA sequence. Mutation of His 118 to alanine (H118A) results in a 20-fold decrease in DNA-binding affinity, while mutation of the same residue to tyrosine (H118Y) results in barely detectable DNA binding (Blumberg et al., 1987; Thukral et al., 1991). Residue 118 corresponds to position 3 of the α-helix of the N-terminal zinc finger in ADR1 and the wild-type histidine side chain is postulated to make a specific contact with the UAS1 DNA sequence (Klevit, 1991; Thukral et al., 1991). A decrease in binding affinity may arise from a loss of a protein–DNA contact and/or from a structural change in the protein caused by the mutation.

Structural analysis of three zinc fingers from Zif268 bound to their cognate DNA sequence in the X-ray crystal structure reveals that the conformation of DNA-bound zinc fingers and the conformations of free single and multiple zinc finger domains in solution are strikingly similar (Kochoyan et al., 1991b–d; Pavletich & Pabo, 1991; Omichinski et al., 1992). This suggests that structures derived from single zinc finger studies are excellent representatives of the active zinc finger conformation. Therefore, studying the structural ramifications of point mutations known to affect DNA binding by making amino acid substitutions in single zinc finger domains is a viable approach to understanding better the relationship between the domain structure and DNA binding. We have determined the structures of synthetic single zinc finger peptides comprising the wild-type ADR1 sequence from residues 102–130 (ADR1b) and the two mutant sequences, H118A and H118Y, in order to distinguish whether the modulation of DNA-binding activity caused by mutation of His 118 is a result of structural perturbations in the zinc finger domain or to loss of a DNA contact.

The solution structure of ADR1b has been previously reported (Klevit et al., 1990). However, this structure was determined based on a less complete set of distance constraints and was not refined against the measured NOE intensities. Also, in the process of assigning the spectra of mutant peptides, we made several corrections to the original wild-type resonance assignments (see assignments in the Supplementary material on the Diskette Appendix). We believe that the refined structure represents a significant improvement over the earlier model. The structural details of the refined ADR1b structure and a comparison of the wild-type domain with the two mutant domains are presented here.

Results and discussion

The sequence of wild-type ADR1b is shown at the bottom of the page, where the residue at position 118 (using the numbering system from the ADR1 protein sequence) is highlighted in a box and the four zinc ligands are in boldface and underlined. Residue 118 is histidine in the wild type and alanine or tyrosine in the H118A and H118Y mutants, respectively. Complete 2D-NMR data sets were

```
102                 118                 130
R S F V C E V C T R A F A R Q E [H] L K R H Y R S H T N E K
```

collected for each of the three zinc fingers and the spectra were assigned. Assignment of the mutant domains was simplified by the fact that there were relatively few chemical shift perturbations relative to wild-type ADR1b. ^{1}H resonance assignments for the three sequences are included as Supplementary material on the Diskette Appendix. Very few protons were unable to be assigned in these peptides: unassigned in ADR1b, the NH of R102 and S103, $C^{\alpha}H$ of F113 and N128, $C^{\delta}H$ of R115 and K120, and the side chain protons of K130; unassigned in H118A, the NH of R102 and the $C^{\gamma}H$, $C^{\delta}H$, and $C^{\epsilon}H$ of K130; unassigned in H118Y, the NH of R102, the $C^{\delta}H$ of K120, and the $C^{\gamma}H$, $C^{\delta}H$, and $C^{\epsilon}H$ of K130. Stereospecific assignments for prochiral $C\beta H$ proton pairs were determined for 10 of the 23 in wild-type ADR1b, 10 of the 23 in H118Y, and 8 of the 22 in H118A. The prochiral methyl groups of V105, V108, and L119 were stereospecifically assigned in all three domains, as described in Materials and methods. The stereospecific assignments are included in the Supplementary material on the Diskette Appendix.

As mentioned above, the spectra of the two point mutant peptides showed only limited chemical shift perturbations relative to the spectrum of wild-type ADR1b. In H118A, eight NH or $C^{\alpha}H$ resonances are shifted by more than 0.1 ppm, with only the NH of A114 and the $C^{\alpha}H$ of A118 being shifted by more than 0.2 ppm (+0.41 ppm and +0.36 ppm, respectively). In H118Y, four backbone resonances are shifted by more than 0.1 ppm and only the NH of R115 shifts by more than 0.2 ppm (+0.32 ppm). Since chemical shifts are extremely sensitive to environment, these results indicate that the two mutations at position 118 do not result in global structural changes and any change in structure that may have occurred is likely to be localized near the site of mutation.

In order to ascertain the detailed structural consequences of the position 118 mutations, high-quality solution structures are required. Toward this end, NOESY build-up sets were collected for each of the three zinc finger peptides and structures were generated. The quality of the spectral data can be judged from sections of 100-ms NOESY spectra shown in Figure 1. Interproton distance constraints obtained from NOESY spectra and X_1/X_2 dihedral angle constraints obtained from analysis of P.E.COSY and NOESY spectra were used to determine the structures, following the protocol described below (see Table 1 for a summary of the constraints used; actual constraint files are included as Supplementary material on the Diskette Appendix).

Initial structures were generated by the variable target function program, DIANA, with wide distance bounds for all unambiguous experimental constraints (1.8–5.0 Å) and these were refined using the REPENT protocol (Hoffman et al., 1993). In this procedure, the interproton distance constraints used to calculate structures are iteratively adjusted to bring about better agreement between the observed NOE intensities (I_{obs}) and NOE intensities calculated from the structural models (I_{calc}) using a complete relaxation matrix algorithm, CORMA (Borgias & James, 1989; James & Borgias, 1990). The agreement between I_{obs} and I_{calc} is quantified using the residual error function $R\langle 1/6\rangle$ (defined in Materials and methods).

Table 1. *Structural parameters for three zinc finger domains*

Parameter	ADR1b	H118A	H118Y
Ave.[a] DIANA target function[b], Å^2	3.88 ± 0.44	3.16 ± 0.73	3.82 ± 1.11
RMSD[c] (N,CA,C,O), Å	0.36	0.36	0.49
RMSD[c] (all heavy atoms), Å	1.45	1.34	1.50
Ave.[a] $R\langle 1/6\rangle$, %	9.2 ± 0.1	8.4 ± 0.2	9.5 ± 0.1
Number of interproton distance constraints[d]			
Intraresidue	141	144	162
$i-i+1$	69	79	75
$i-i+2$	13	17	14
$i-i+3$	24	17	25
$i-i+4$	5	7	9
Long range	68	73	54
Total	320	337	339
Number of dihedral angle constraints[e]			
X_1	12	10	12
X_2	1	1	1

[a] Averages are reported for the 10 refined structures in the family.
[b] The target function is a residual error function that reports upper and lower bound violations, dihedral angle constraint violations, and van der Waals violations for each calculated structure.
[c] All RMSDs are reported as the average pairwise RMSDs among the 10 refined structures for the residue range 104–126.
[d] The upper and lower bounds files used in the structure calculations are included in the Supplementary material on the Diskette Appendix.
[e] The dihedral angle constraints used in the structure calculations are included in the Supplementary material on the Diskette Appendix.

For a comparison of structures to be meaningful, the structures must be of similar quality. Each zinc finger solution structure is represented by a family of the 10 structures with the lowest $R\langle 1/6\rangle$ after refinement, shown in Figure 2 and Kinemage 1. Statistics for the three zinc finger domains (Table 1) show that all three structure families were calculated with a similar number of constraints and have low DIANA target functions, low RMS deviations, and low $R\langle 1/6\rangle$ values, indicating that all three domains are both well defined (i.e., high precision) and well refined (i.e., high accuracy). Structures were also generated using the commonly employed practice of categorizing NOE cross peaks as strong, medium, or weak and assigning distance constraints based on the categories. The resulting structures have higher $R\langle 1/6\rangle$ values associated with them, signifying that the REPENT-generated structures reported here are more consistent with the NMR measurements (Hoffman et al., 1993) Thus, a detailed comparison of these solution structures is justified.

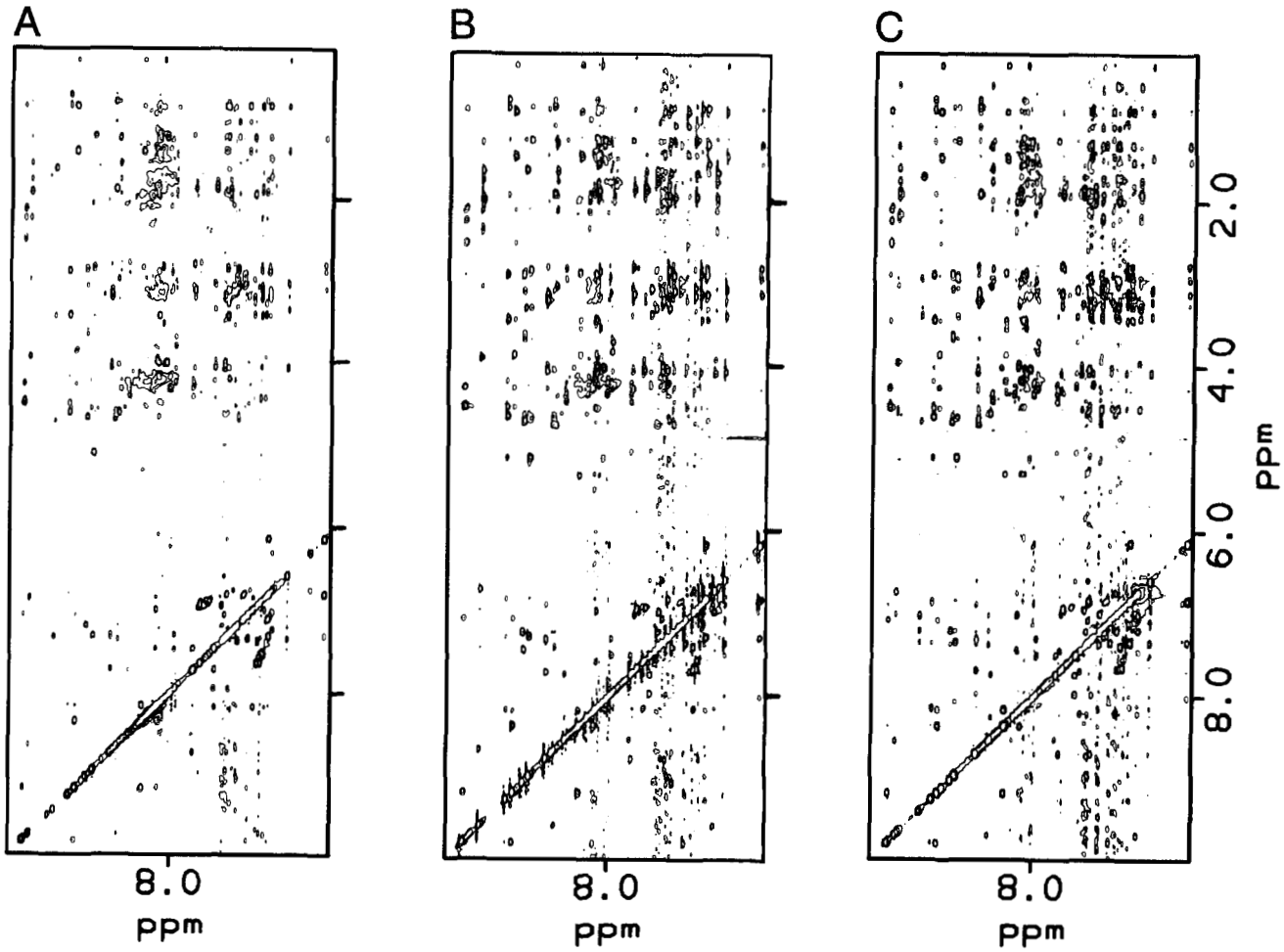

Fig. 1. One hundred-millisecond H_2O NOESY spectra of the three zinc finger domains. **A:** Wild-type ADR1b. **B:** H118A. **C:** H118Y. The horizontal axis includes the resonances of amide and aromatic protons, and the vertical axis represents the entire spectral range of resonances. Spectral parameters are given in Materials and methods. All three spectra were plotted at the same level and with identical plot parameters.

In the structures presented here, no distance constraints were used to enforce tetrahedral coordination geometry or H-bonds in the structure calculations. The coordination geometry and H-bonds predicted in the structures presented are therefore the result of experimentally measured interproton distances and the atomic van der Waals radii used in the DIANA calculations. The structural details of the wild-type ADR1b domain will be presented first and the two mutant domains will be discussed in the context of the wild-type structure.

Description of the ADR1b structure

The zinc finger is a small, compact domain that does not resemble a finger so much as it resembles a rectangular box; its approximate dimensions are 18 × 13 × 6 Å, with the helical face of the molecule constituting the smallest dimension. The topology of the molecule is a right-handed $\beta\beta\alpha$ unit consisting of a two-stranded antiparallel β-sheet with a reverse turn connecting the strands, a 90° turn in the fingertip region followed by a helix of ~11 residues. These three structural elements are shown in Figure 3 (see also Kinemage 2) and are referred to as the Cys–Cys loop (residues 104–113), the fingertip (residues 114–115), and the helix (residues 116–126) throughout the paper. The folding topology places the N- and C-termini at opposite ends of the domain, separated by approximately 22 Å (from Ser 103 CA to Thr 127 CA). The molecular architecture of ADR1b will be discussed in terms of the forces that stabilize the structure: the coordination of the zinc ion, H-bonds, and hydrophobic interactions.

Zinc coordination

The zinc ion is liganded by the two conserved cysteine residues, Cys 106 and Cys 109, and the two conserved histidine residues, His 122 and His 126. Coordination bond lengths of 2.30 Å for the Sγ–Zn bond and 2.00 Å for the Nϵ–Zn bonds were imposed during the structure calcula-

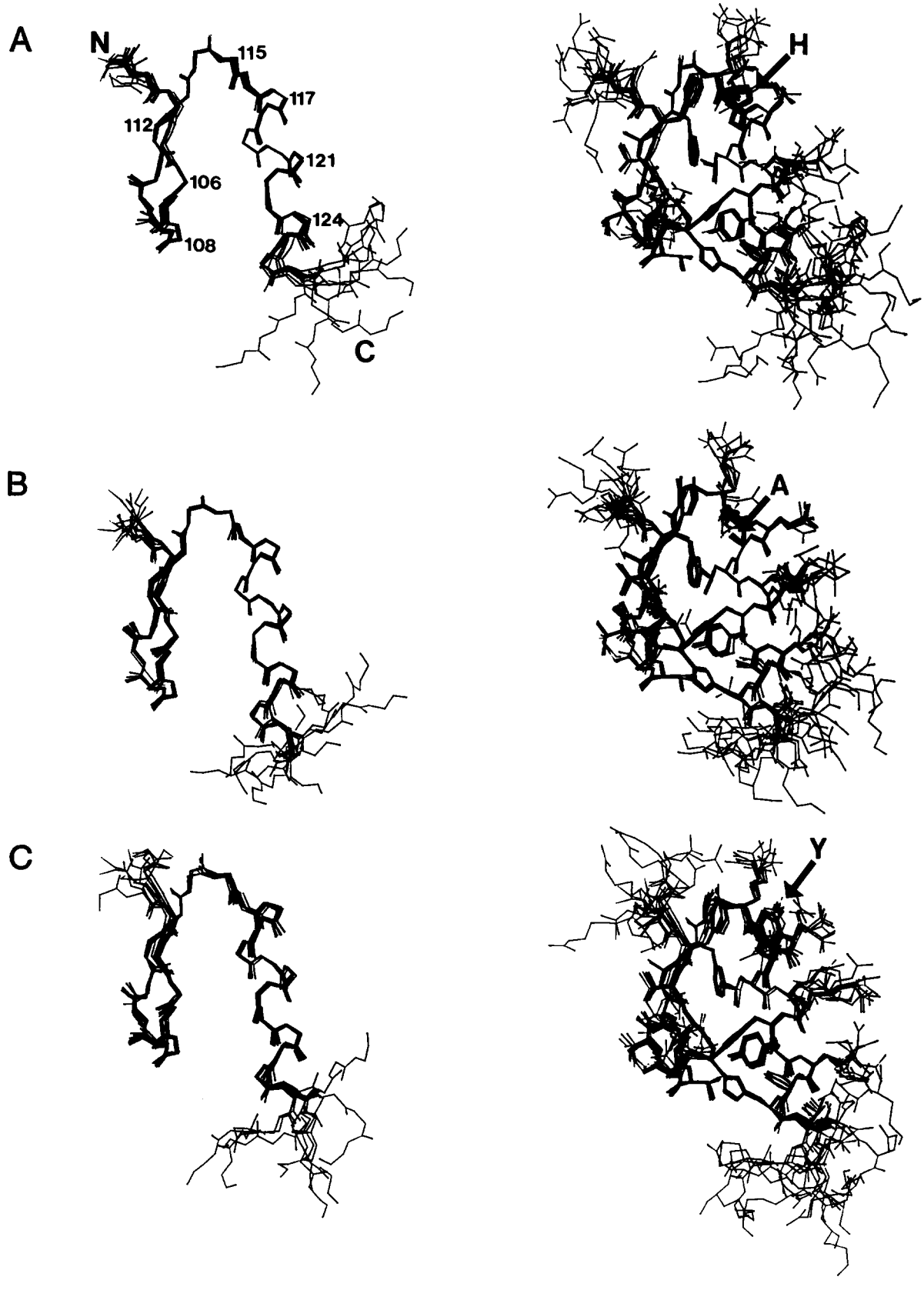

Fig. 2. Superpositions of the three zinc finger domain structure families. **A:** Wild-type ADR1b. **B:** H118A. **C:** H118Y. The left panels of each show backbone atoms only; the right panels show all heavy atoms. Each superposition contains the 10 refined structures in the family, superimposed using backbone atoms (N, CA, C, O) of residues 104–126. The position of residue 118 is denoted in the right panels.

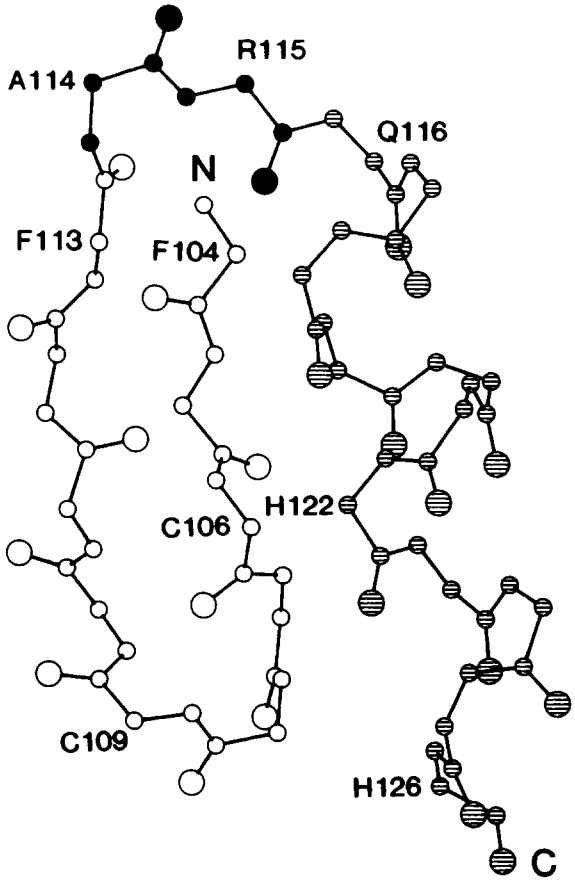

Fig. 3. Backbone atoms of a representative wild-type ADR1b structure. Residues 104–126 are shown, with residue numbers adjacent to their respective Cα. The Cys-Cys loop (Phe 104–Phe 113) is drawn with unfilled atoms, the fingertip (Ala 114–Arg 115) is drawn with filled atoms, and the helix (Gln 116–His 126) is drawn with stippled atoms. The amino- and carboxy-termini are labeled N and C, respectively.

Table 2. *Wild-type ADR1b zinc coordination parameters*[a]

Ligand atom pairs	Distance (Å)	Bond angle[b] (degrees)
Sγ 106–Sγ 109	3.9 ± 0.1	118 ± 6
Sγ 106–Nε 122	3.3 ± 0.1	99 ± 3
Sγ 106–Nε 126	3.3 ± 0.1	99 ± 3
Sγ 109–Nε 122	3.2 ± 0.1	98 ± 5
Sγ 109–Nε 126	4.0 ± 0.1	139 ± 4
Nε 122–Nε 126	2.9 ± 0.1	93 ± 2
Ligand residue	**X_1 (degrees)**	**X_2[c] (degrees)**
Cys 106	−171 ± 2	−151 ± 4
Cys 109	68 ± 4	−87 ± 4
His 122	156 ± 1	81 ± 3
His 126	−68 ± 7	−35 ± 4

[a] All values are reported as the average value for the 10 refined ADR1b structures.
[b] The bond angle reported is that formed by atom 1–zinc–atom 2.
[c] The X_2 dihedral angle reported for the two cysteines is that formed by Cα–Cβ–Sγ–Zn.

tions (Diakun et al., 1986); however, no interligand atom distances were included to impose tetrahedral coordination geometry. Stereospecific assignments were obtained for the prochiral CβH proton pairs in all four liganding residues and X_1 angle constraints with a 40° range centered about the staggered rotamer value (based on analysis of J-couplings and relative NOESY intensities) were included in the calculations. No side chain angle constraints beyond X_1 were used for these four residues. The resultant zinc coordination parameters are given in Table 2.

An earlier analysis of side chain conformations of cysteine residues involved in metal coordination in 12 metalloproteins revealed a trimodal distribution of X_1/X_2' values (where X_2' is defined by Cα–Cβ–Sγ–metal): 180°/90°, 180°/180°, and 60°/−90° (Chakrabarti, 1989). In ADR1b, Cys 106 has average X_1 and X_2' values of −171 ± 2° and −151 ± 4° respectively, placing it in the second class, whereas Cys 109 has average values of 68 ± 4° and −87 ± 4°, placing it in the third class of cysteine ligands. This pattern is consistent with the analogous cysteines in the rubredoxin family (Watenpaugh et al., 1979; Frey et al., 1987; Stenkamp et al., 1990; Adman et al., 1991) and also with the two Cys-X_2-Cys fingers of Zif268. It is interesting to note that the Cys-X_4-Cys finger of Zif268 has both coordinating cysteines in the fully extended conformation, i.e., 180°/180° (Pavletich & Pabo, 1991). The longer loop in Cys-X_4-Cys fingers may permit the second cysteine to adopt a more energetically favored *t* side chain conformation at X_1; in the *g*-conformation ($X_1 = 60°$) the bulky Sγ is wedged between its own amide and carbonyl moieties. This reduction in strain of the Cys side chain could contribute to a lower zinc dissociation constant for the Cys-X_4-Cys type zinc finger.

The side chain conformations of the liganding histidine residues can also be uniquely described by their X_1/X_2 pairs. His 122 and His 126 have average values of 156 ± 1°/81 ± 3° and −68 ± 7°/−35 ± 4°, respectively. The two histidine rings present their ring faces to one another, as opposed to a face-edge interaction, and both present their Nδ to the more solvent-exposed face of the zinc center. The histidine side chain conformations in ADR1b are similar to those in thermolysin, which also ligands a zinc ion using a His-X_3-His motif (Holmes & Matthews, 1982). They are also consistent with the histidine ligand conformations observed in the Zif268 co-crystals in which the Nδ of the first histidine ligand contacts the backbone of the DNA (Pavletich & Pabo, 1991).

The geometry of the zinc center in the refined ADR1b structures is not perfectly tetrahedral. However, four of the six angles deviate by only ~10° from the "perfect" value of 109.5°. The other two angles, Sγ109–Zn–Nε126

and Nε122–Zn–Nε126, are opened by ~30° and closed by ~17°, respectively, and represent the largest deviations from tetrahedral coordination. Although the magnitude of the deviations is not as great, the largest deviations from tetrahedral geometry in the third finger of Zif268 also correspond to these two angles (Pavletich & Pabo, 1991).

Backbone conformation and hydrogen bonding

Because no explicit hydrogen bond (H-bond) constraints were used to calculate the structures reported here, it is of interest to ascertain where H-bonds are predicted in the resulting refined structures. This, however, poses a dilemma when analyzing solution structures: one can analyze an average structure (either mathematical or geometric), calculated from the family of solution structures as representative, or one can treat the family as an ensemble and analyze the group statistically. The first method has the obvious attraction of being a smaller data set (i.e., one structure), but has the disadvantage of oversimplifying the true nature of solution structures. When the family of solution structures have low RMS deviations among themselves, the two types of analysis will lead to the same conclusions, except for instances where a predicted H-bond is just within the geometric limits set. Hydrogen bonds have been analyzed in the 10 refined structures that constitute the family of ADR1b solution structures; frequencies with which they are predicted in the 10 structures are reported as percentages for each H-bond and the geometric parameters in the average structure are given in Table 3. The cutoff values used to predict H-bonds are: distance between N and O ≤3.5 Å, CO–H angle ≥90°, and NH–O angle ≥120°, as these are the most generous limits currently in use and should encompass all reasonable H-bond geometries (Stickle et al., 1992).

Residues 104–113 form an antiparallel β-sheet and reverse turn that comprise the first structural unit of the zinc finger domain, the Cys–Cys loop, shown in Figure 4. The two conserved aromatic residues, Phe 104 and Phe 113, lie opposite one another in the sheet with their side chains on the interior face of the sheet. Val 105 and Ala 112 also lie opposite one another in the sheet with their side chains on the solvent-exposed face, and with Phe 104 and Phe 113 these four residues comprise the "classical" β-sheet region of the molecule. An H-bond is predicted in 60% of the structures between the carbonyl of Phe 104 and the amide of Phe 113. An H-bond is not predicted, however, between the amide of Phe 104 and the carbonyl of Phe 113 in any of the structures because the donor and acceptor atoms are too far apart (N_{104} to O_{113} ranges from 4.32 to 5.09 Å) and because the amide of Phe 104 does not lie in the plane of the sheet. The analogous H-bond is also not predicted in the NMR structure of the N-terminal finger of MBP-1 (Omichinski et al., 1990), but is predicted in the NMR structures of Xfin31 (Lee et al., 1989) and ZFY-swap (Kochoyan et al., 1991b) and in the three zinc finger domains in the crystal structure of Zif268 (Pavletich & Pabo, 1991). The lack of a predicted H-bond in the ADR1b structures is based on experimental observations rather than a lack of constraints: long-range NOE

Table 3. *Predicted hydrogen bonds for three zinc finger domains*

	ADR1b[a]				H118A[a]				H118Y[a]			
H-bond: HN–CO	dNO (Å)	CO–H (deg)	NH–O (deg)	Freq. (%)	dNO (Å)	CO–H (deg)	NH–O (deg)	Freq. (%)	dNO (Å)	CO–H (deg)	NH–O (deg)	Freq. (%)
6–11	**2.57**	**133**	**158**	100	**2.69**	**119**	**129**	100	**2.91**	**123**	**157**	100
10–6	**2.15**	**167**	**123**	60	2.40	129	102	20	**2.42**	**142**	**116**	30
11–6	**3.37**	**105**	**125**	90	3.73	112	101	10	3.84	117	108	10
13–4	**3.20**	**156**	**144**	100	**2.80**	**153**	**153**	100	**2.98**	**144**	**141**	100
18–15	3.60	88	124	20	**2.76**	**108**	**124**	100	**3.09**	**119**	**118**	50
19–15	3.82	130	151	0	**3.32**	**138**	**132**	90	**3.39**	**140**	**155**	90
19–16	**2.64**	**115**	**117**	80	**2.61**	**107**	**134**	100	**2.85**	**92**	**110**	30
20–16	**2.97**	**152**	**149**	100	**2.90**	**158**	**147**	100	**2.41**	**171**	**170**	100
21–17	**2.63**	**161**	**155**	100	**2.94**	**154**	**165**	100	**2.82**	**156**	**169**	100
22–18	**2.69**	**130**	**135**	100	**3.22**	**116**	**142**	100	**2.72**	**132**	**156**	100
23–19	**2.60**	**162**	**155**	100	**2.71**	**156**	**154**	100	**2.81**	**150**	**140**	100
24–20	**3.37**	**148**	**141**	90	**3.07**	**152**	**151**	100	3.64	126	118	20
24–21	**3.23**	**96**	**133**	90	**3.12**	**95**	**117**	50	**2.50**	**122**	**153**	100
25–22	**2.67**	**140**	**165**	100	**2.70**	**115**	**140**	100	**3.03**	**113**	**147**	100
26–23	**2.31**	**134**	**139**	70	**2.68**	**111**	**130**	100	**2.94**	**99**	**140**	90
27–23	**3.20**	**148**	**120**	50	**3.32**	**143**	**134**	70	3.65	129	101	10
27–24	**3.47**	**92**	**153**	70	**3.12**	**93**	**129**	40	**3.08**	**105**	**151**	90

[a] The first three columns for each domain list the H-bond donor-acceptor distance, the carbon-oxygen-hydrogen angle, and the nitrogen-hydrogen-oxygen angle for the three average structures derived from each family of 10 structures. Geometries that are within the limits set are highlighted in bold. The fourth column for each domain gives the frequency with which the H-bond is predicted in individual members of the family.

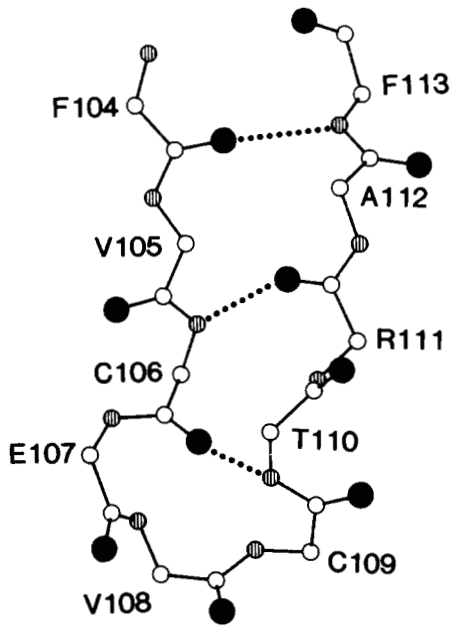

Fig. 4. Backbone atoms from the Cys–Cys loop (Phe 104–Phe 113) of a representative structure of the refined wild-type ADR1b family, showing three predicted H-bonds. Residue numbers are shown adjacent to their Cα. Carbon atoms are unfilled, nitrogen atoms are stippled, and oxygen atoms are filled.

interactions between Cβ protons of Ser 103 and protons of Val 105 and Ala 112 result in a turn in the main chain at the peptide bond between Ser 103 and Phe 104 and preclude the amide of Phe 104 from donating an H-bond. The absence of this H-bond in the peptide ADR1b may indicate a real structural feature in the N-terminal finger of the protein ADR1. Recent studies have suggested that the N-terminal fingers in the two-finger protein Ttk and the three-finger protein SWI5 have a third strand of β-sheet involving the residues preceding the first conserved aromatic residue (Fairall et al., 1992; Neuhaus et al., 1992). If this structure is also present in the protein ADR1, the main chain would have to adopt a turn conformation N-terminal to Phe 104. The long range NOEs involving Ser 103 are consistent with a turn in this region.

Cys 106 lies opposite both Thr 110 and Arg 111 and together these three residues form a structure similar to a G1 type β-bulge in the sheet (Richardson, 1981). Two H-bonds, predicted between the amide of Cys 106 and the carbonyl of Arg 111 (100%) and between the carbonyl of Cys 106 and the amide of Thr 110 (60%), may serve to stabilize the bulge (see Fig. 4). In a classical G1 β-bulge, the residue at position 110 would be a glycine, and indeed this is the most commonly observed amino acid at this position in the Cys_2–His_2 class of zinc fingers (Krizek et al., 1991). The average ϕ and ψ values for such a glycine residue would be 85° and 0°, respectively (Richardson, 1981). The average ϕ and ψ angles observed for Thr 110 are 58 ± 25° and 45 ± 12°, placing the residue in the left-handed helical region of Ramachandran space, which is accessible to residues with β-carbons (Richardson, 1981). A positive ϕ angle for a non-glycine residue in this position is also observed in the second finger of Zif268, Met 141 (Pavletich & Pabo, 1991). The ϕ and ψ angles of Met 141 are closer to the classic G1 β-bulge values than those of Thr 110; this backbone conformation may be more accessible to residues that are not branched at the β-carbon. The β-bulge projects the side chains of Cys 106 and Arg 111 toward the interior side of the sheet and places the side chain of Thr 110 on the solvent-exposed face. The bulge also results in a pronounced right-handed twist in the sheet that brings the two cysteine side chains (Cys 106 and Cys 109) into close proximity to ligand the zinc. Any unfavorable strain resulting from having a non-glycine residue after the second cysteine adopt a positive ϕ angle is likely compensated by the gain in stability afforded by zinc binding.

As well as being involved in the β-bulge, Thr 110 participates with Glu 107, Val 108, and Cys 109 in forming a reverse turn. Although the backbone dihedral angles do not classify the turn as one of the canonical types, it is most similar to a type I reverse turn (Val 108 $\phi = -76 \pm 11°$, $\psi = -71 \pm 2°$; Cys 109 $\phi = -95 \pm 6°$, $\psi = 1 \pm 22°$). The side chains of Glu 107, Val 108, and Cys 109 are disposed toward the interior face of the sheet. The turn is stabilized by amide–sulfur H-bonds: the amide protons of Val 108 and Cys 109 are predicted to H-bond to the Sγ of Cys 106 and the amide of Arg 111 may participate in a weak H-bond with the Sγ of Cys 109. Consonant with the predicted H-bonds, the amide protons of Val 108 and Cys 109 are the most resistant to solvent-exchange in ADR1b (data not shown). The strength of these amide–sulfur H-bonds based on the nitrogen–sulfur distance and NH–S bond angle agree in a qualitative fashion with the recent results of Blake and coworkers (1992a), who measured single bond heteronuclear couplings between amide ^{1}H and ^{113}Cd in Cd-substituted rubredoxin. The Cys–Cys loop structure of ADR1b (from Phe 104 to Phe 113) is strikingly similar to the rubredoxin family of proteins (see Kinemage 3). They all show the same main chain and amide–sulfur H-bond pattern and β-sheet, β-bulge, and reverse turn structural elements (Watenpaugh et al., 1979; Frey et al., 1987; Stenkamp et al., 1990; Adman et al., 1991; Blake et al., 1992b; Day et al., 1992). The average RMS deviation between the average refined ADR1b structure and three published rubredoxin structures for main chain atoms (N, CA, C, O) for residues 104–113 is 0.56 Å (Watenpaugh et al., 1979; Frey et al., 1987; Stenkamp et al., 1990).

Connecting the Cys–Cys loop and the helix are two residues forming the fingertip of the domain. Ala 114 adopts a backbone conformation in the right-handed helix region of Ramachandran space ($\phi = -84 \pm 5°$, $\psi = -18 \pm 9°$) and is followed by Arg 115 in an extended conformation ($\phi = -132 \pm 12°$, $\psi = 154 \pm 2°$). This combination of backbone dihedral angles leads to a 90° turn in the main

chain. Neither the amides nor carbonyls of these residues are predicted to participate in any H-bonds; the amide proton of A114 exchanges with solvent very rapidly, as evidenced by its low intensity in spectra collected using a presaturation pulse (data not shown). The main chain conformation through this region is virtually identical to that seen in the three fingers of Zif268 (Pavletich & Pabo, 1991).

Based on backbone dihedral angles, the helix in ADR1b extends from Gln 116 through His 126. The first strongly predicted H-bond involves the carbonyl oxygen of Gln 116, which is within H-bonding distance and geometry to the NH of Leu 119 in 80% of the structures and to the NH of Lys 120 in 100% of the structures. The carbonyls of residues 117, 118, and 119 are all predicted to participate in α-helical H-bonds (with the amides of residues 121, 122, and 123, respectively) in 100% of the ADR1b structures. The next two carbonyls, Lys 120 and Arg 121, point away from the helix axis, and although an H-bond between Lys 120 CO and Arg 124 NH is still predicted in 90% of the structures, the NO distance is longer than in the preceding H-bonds (3.37 Å). The distortion away from α-helical conformation is even greater at the carbonyl of Arg 121, which is not predicted to form an H-bond with the NH of residue 125 in any of the structures, but is predicted to be within H-bonding distance to Arg 124 NH in 90% of the structures. This 3_{10}-helical structure continues with strongly predicted H-bonds between His 122 CO and Ser 125 HN (100%) and between Tyr 123 CO and His 126 NH (70%). The helix of ADR1b terminates at residue 127, with its NH predicted to H-bond to the carbonyl of Tyr 123 (50%) and/or Arg 124 (70%). A similar α- to 3_{10}-helix transition was reported for Xfin31 (Lee et al., 1989), ZFY-switch (Kochoyan et al., 1991c), and the Zif268 zinc fingers (Pavletich & Pabo, 1991). A transition to a 3_{10}-helix at the C-terminal end of an α-helix is commonly observed in proteins and may therefore explain the observation of this structure in the zinc fingers (Richardson, 1981).

In thermolysin, the His-X_3-His motif is in the middle of a long helix, yet the helical parameters also change to 3_{10} in the region containing the zinc ligands, and return to more α-helical values following the motif (Holmes & Matthews, 1982). The transition from α-helical structure to 3_{10} structure results in an increase in the distance between the C^α atoms of the two histidine residues: this distance is 6.3 Å in an α-helix and 7.8 Å in a 3_{10}-helix. The average C^α–C^α distance between His 122 and His 126 in ADR1b is 7.1 Å, placing it intermediate between the two idealized structures. The increased C^α–C^α distance may be important to allow the two histidine ligands to assume the proper relative positions to facilitate zinc coordination.

In addition to the main chain H-bonds described above, an H-bond between the side chain hydroxyl of Ser 125 and the carbonyl of His 122 is predicted. The side chain of the Ser 125 exists as a unique rotamer of g^- ($X_1 = 60°$), as indicated by its intraresidue NOE pattern and two small $^3J_{\alpha\beta}$ coupling constants. Observation of a unique conformation in a hydrophilic, solvent-exposed side chain in solution is good evidence for a side chain being involved in an H-bond. Serines in helices are commonly observed to donate H-bonds from their side chain hydroxyl to the carbonyl of either the $n-3$ or $n-4$ residue (Baker & Hubbard, 1984).

Packing of the hydrophobic interior

The superposition of structures shown in Figure 1 reveals that the side chains in the interior of ADR1b are well defined. The packing of the conserved hydrophobic residues and zinc ligands is illustrated in Figure 5. The two conserved aromatics, Phe 104 and Phe 113, are on opposite sides of the β-sheet and have the same X_1 rotamer ($-60°$). Thus, they display twofold symmetry about an axis running perpendicular to the plane of the sheet. This arrangement of the two rings provides a large flat hydrophobic surface that packs against the N-terminal half of the helix. Both phenylalanine residues are in contact with the conserved leucine residue, Leu 119, which fits into a cleft formed by Phe 104, Phe 113, and Cys 106. The edge of the ring of Phe 113 also contacts the face of the first liganding histidine, His 122.

As shown in Figure 5, the side chains of Cys 106 and Cys 109 approach the zinc from the sheet above and the turn below, respectively, presenting the Sγ ligands to opposite sides of the zinc. This conformation provides another relatively flat hydrophobic surface that packs against the C-terminal half of the helix, leaving one face of the zinc exposed for the Nϵ ligands of His 122 and His 126. There are a number of van der Waals contacts between the zinc liganding residues and two of the conserved hydrophobic residues: Cys 106 and His 122 con-

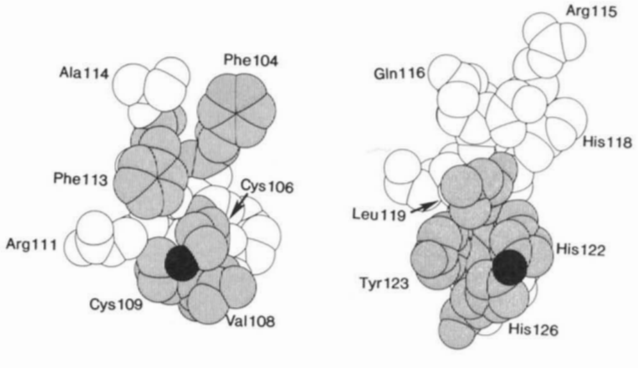

Fig. 5. Packing of the hydrophobic core. A representative of the refined wild-type ADR1b family in a space-filling representation is shown, split in half longitudinally and opened like a sandwich. The left panel shows residues Phe 104–Ala 114 and the right panel shows residues Arg 115–His 126. Hydrophobic residues are indicated by stippling and the zinc atom is filled. Residues are numbered adjacent to their side chains.

tact one another, Phe 113, and Leu 119, while Cys 109 has interactions with His 126. The packing angle between the long axis of the β-sheet and the helix axis is approximately 0°, the most commonly observed packing angle between helices and sheets (Chothia, 1984).

In addition to these seven conserved residues, two additional hydrophobic residues, Val 108 and Tyr 123, are present in the "mini-core" of the ADR1b domain and are well defined in the solution structure. Val 108 contacts both Tyr 123 and His 126. Tyr 123 acts a "backflap," shielding one face of the zinc center from solvent through interactions with Val 108 and Leu 119.

Comparison of the wild-type and mutant structures

No change in the overall architecture of the zinc finger domain results from replacing His 118 with either an alanine or a tyrosine. Superposition of structures for wild-type ADR1b and the H118 mutants (Fig. 6) reveals that the calculated structures are remarkably similar, with RMS deviation for backbone atoms (N, CA, C, O) from Phe 104 to His 126 (calculated between the geometric averages of each family) of 0.55 Å (H118A) and 0.53 Å (H118Y). As expected from the limited chemical shift perturbations observed, if there are differences in the structures, they are very small indeed. The ability to discern small but significant changes between two ensembles of NMR solution structures is dependent on the certainty with which atomic positions are known. The significance parameter, S, introduced by Hurle and coworkers (1992), represents an excellent method for analyzing such differences. Briefly, positional shifts in atoms are normalized by the positional variability of that atom in each family of structures. This variability represents the degree of certainty with which the position of the atom is known in the average structure and can be adjusted to different confidence levels by incrementing the number of standard deviations in position for the atom (see Materials and methods). At a given confidence level, an atom with $S \geq 1.0$ is said to have a significant shift in its position in the structures being compared.

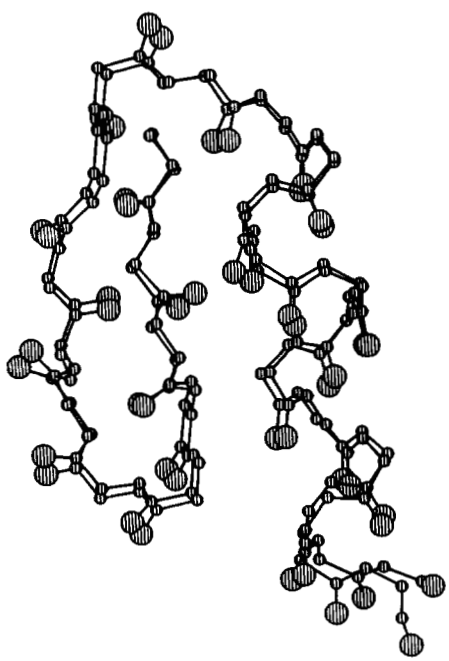

Fig. 6. Superposition of wild-type ADR1b and H118A. The average structures were superimposed, using residues 104–126. For the sake of clarity, only one mutant structure is superimposed, but as the root mean square deviations indicate, both structures are extremely similar to wild-type ADR1b.

As described in Materials and methods, average structures were calculated from each of the family of structures and these were superimposed to give the best agreement between the structures. These superpositions were used to calculate atomic positional shifts for each of the mutant structures, relative to wild-type ADR1b. S values were calculated at both the 95% confidence level (as suggested by Hurle et al. [1992]; this corresponds to 2σ in atomic position) and at the 99% confidence level (which corresponds to 3σ). The atomic positional shifts and corresponding S values for atoms with $S \geq 1.0$ are listed in Table 4, and these parameters for all backbone atoms are listed in the Supplementary material on the Diskette Appendix. When H118A is compared with ADR1b, there are only four backbone atoms with positional shifts ≥0.9 Å. Although a number of backbone atoms have $S \geq 1.0$ at the 95% confidence level, there are only six atoms with $S \geq 1.0$ when calculated using the more conservative value of 3σ. Similar results were obtained for the H118Y mutant. The structures of the two mutant zinc finger peptides are therefore deemed to be indistinguishable from the wild-type structure at the 99% confidence level.

H118A and H118Y were both analyzed for predicted H-bonds, as already described for the wild-type structure (see Table 3). As would be expected based on the close correspondence among these structures, the H-bonds predicted are, in general, also similar. The biggest difference involves the carbonyl group of Arg 115 (the second fingertip residue), which is strongly predicted to form H-bonds in the two mutant structures that are not predicted in the wild-type structures. Because this newly predicted H-bond involves the mutated residue, it is tempting to conclude that it is a real structural consequence of the mutation of His 118. However, the variability in predicted H-bonds in the NMR structures is more likely a consequence of imprecision in the structures, rather than significant structural differences. Indeed, similar variabilities in predicted H-bonding are present in the three zinc fingers of the Zif268 structure (solved at 2.8 Å resolution): the analogous fingertip carbonyl oxygen is predicted to participate in no H-bond in the first finger, an $i, i+4$ H-bond in the second finger, and both an $i, i+3$ and an $i, i+4$ H-bond in the third finger, even though fingers 1 and 3 have nearly identical sequences in this region. Hence,

Table 4. *Backbone atoms with S-factors ≥1.0 for two mutant ADR1b zinc fingers relative to wild-type ADR1b*

Atom		Positional shift (Å)	S-Factor 2σ[a]	S-Factor 3σ[b]
H118A backbone atoms				
V105	O	1.04	1.48	0.98
A114	O	0.80	1.01	0.67
R115	C	0.55	1.11	0.74
	O	0.96	1.55	1.03
A118	C	0.44	1.19	0.79
	O	0.71	1.53	1.02
L119	N	0.43	1.47	0.98
	CA	0.45	1.58	1.06
	C	0.34	1.28	0.85
	O	0.37	1.15	0.77
R121	C	0.51	1.02	0.68
H122	N	0.41	1.13	0.75
	CA	0.50	1.39	0.92
	C	0.60	1.60	1.07
	O	0.70	1.33	0.89
Y123	N	0.58	1.54	1.02
	CA	0.69	1.55	1.03
	C	0.74	1.43	0.95
R124	N	0.84	1.07	0.71
	CA	0.96	1.00	0.67
H126	N	0.63	1.04	0.69
	CA	0.80	1.13	0.75
	C	0.98	1.09	0.72
H118Y backbone atoms				
R121	O	0.46	1.32	0.88
H122	C	0.36	1.11	0.74
	O	0.72	1.39	0.93

[a] *S* calculated at the 95% confidence level using two standard deviations in atomic position (see Materials and methods).
[b] *S* calculated at the 99% confidence level using three standard deviations in atomic position (see Materials and methods).

such variabilities are not unique to NMR solution structures and their existence highlights the limitations of drawing conclusions regarding individual H-bonds on the basis of one structure.

Implications for DNA binding

Three amino acid residues in the N-terminal zinc finger of ADR1 (ADR1b) are postulated to make specific contacts with bases in the UAS1: Arg 115, His 118, and Arg 121, as indicated schematically in Figure 7 and Kinemage 2 (Klevit, 1991; Thukral et al., 1991). The importance of these three residues for binding to UAS1 was identified by scanning alanine mutagenesis performed on the ADR1 zinc finger sequences (Thukral et al., 1991). These residues may form H-bonds to a GAG purine triplet in UAS1 in a manner analogous to the contacts seen in the Zif268–DNA co-crystal structure. In this model, the guanidinium group of Arg 115 forms H-bonds to both the O_6 and N_7 of the 3′ guanine while Arg 121 forms similar

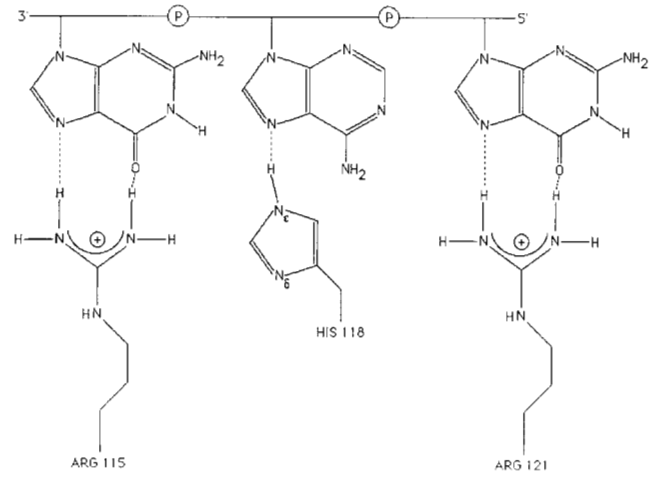

Fig. 7. Proposed contacts between ADR1b side chains and the GAG triplet in UAS1.

H-bonds to the 5′ guanine in the triplet. The histidine at position 118 provides an H-bond between its side chain Nε and the N_7 of the central adenine base, analogous to the His 149 Nε to guanine N_7 H-bond predicted in the second finger of the Zif268–DNA structure. It has been shown that the central base of the triplet in UAS1 can be changed from adenine to guanine with no loss of high-affinity binding by ADR1, consistent with a histidine at this position H-bonding to the N_7 of either type of purine base (Thukral et al., 1992).

When the histidine at position 118 is mutated to an alanine, the ability of ADR1 protein to bind to UAS1 decreases approximately 20-fold, as measured by gel shift assay (Thukral et al., 1991). Comparison of the refined solution structures of wild-type ADR1b and H118A reveals that this mutation does not result in any detectable structural changes. Thus, the most dramatic change between the two domains is the loss of the imidazole moiety itself. A 20-fold increase in the dissociation constant for the ADR1–UAS1 complex corresponds to a $\Delta\Delta G$ of −1.8 kcal/mol. This result is consistent with the loss of a single H-bond in the complex, suggesting that the side chains of Arg 115 and Arg 121 still form the proper H-bonds to UAS1 in the alanine mutant.

The mutation of His 118 to a tyrosine in ADR1 results in barely detectable DNA binding (Blumberg et al., 1987). Under the assay conditions used, this indicates at least a 3–4-orders of magnitude increase in K_D, corresponding to a $\Delta\Delta G$ of ca. −4 to −6 kcal/mol. Comparison of the refined solution structures of ADR1b and H118Y also reveals no significant structural change, and again the most obvious change between the two is replacement of the imidazole group with a phenoxyl moiety. Unlike the alanine mutant in which the wild-type side chain is essentially truncated at the Cβ, H118Y contains a bulkier side chain at this position, with a gain of 23 Å^3 for H118Y as

opposed to a loss of 51 Å^3 for H118A (volumes from Creighton [1992]). To ascertain if the larger aromatic side chain can be accommodated in a zinc finger–DNA complex at this position, His 149 in the second finger of the Zif268 crystal structure (which is structurally analogous to His 118 in ADR1b) was modeled with a tyrosine residue. When X_1 of the tyrosine is set to ca. −60° (as found for His 149 in the Zif268–DNA complex and for Tyr 118 in the solution structure of H118Y), steric clashes ($d \leq 2$ Å) occur between the phenoxyl oxygen of tyrosine and either the N_7 of the base or the substituent on C_6 (O in guanine, NH_2 in adenine) depending on the X_2 angle of the tyrosyl ring. A sampling of X_1/X_2 dihedral angle space revealed that no side chain conformation was free from van der Waals clashes between the tyrosine ring and either the DNA or other atoms in the zinc finger. Thus, the modeling indicates that a tyrosine at position 118 in ADR1 cannot be accommodated in a zinc finger–DNA complex analogous to the Zif268 complex without the loss of contact at either Arg 115 and/or Arg 121. It has been shown that an R115A mutation in ADR1 causes a 30-fold decrease in UAS1 binding, while an R121A mutation results in a ≥100-fold decrease (Thukral et al., 1991). If these effects were additive, the loss of an H-bond due to the substitution of His 118 with a bulky tyrosine side chain and concomitant loss of H-bonds from either Arg 115 or Arg 121 due to steric clashes could account for the loss of DNA binding observed for H118Y.

Conclusions

The Cys_2–His_2 class of zinc fingers possess a remarkably stable and yet plastic structure. It has been shown that single zinc finger domains can tolerate many sequence changes and retain their ability to fold in the presence of zinc into a stable structure. The loop between the two cysteine ligands can include two or four amino acid residues. Cys-X_4-Cys is observed in the N-terminal finger of the Zif268 crystal structure (Pavletich & Pabo, 1991). As well, the spacing between the two histidine ligands is variable and the domain can tolerate three, four, or even five residues between the conserved histidines as shown in the first experimentally derived structures of each type: Xfin-31 (Lee et al., 1989), ADR1a (Párraga et al., 1988), and the N-terminal finger of MBP-1 (Omichinski et al., 1990), respectively. Changes in the hydrophobic core can also be accommodated as illustrated by the ability to substitute the highly conserved phenylalanine at position 113 (ADR1b numbering scheme) with the less frequently observed tyrosine or with leucine and retain the same global fold (Mortishire-Smith et al., 1992). The phenylalanine in position 113 can also be substituted by a phenylalanine or tyrosine at position 111. The side chain rings occupy approximately the same space in the hydrophobic core whether they are found at position 111 or 113 (Weiss & Keutmann, 1990; Mortishire-Smith et al., 1992).

The most dramatic demonstration of the plasticity of the domain comes from the work of Berg and coworkers on a "minimalist" zinc finger called MZF (Michael et al., 1992). The MZF sequence preserves the two conserved aromatics at positions 104 and 113 (ADR1b numbering scheme), the conserved leucine at position 119, and the cysteine and histidine ligands, with the remaining residues being alanine (with lysine at positions 103, 116, and 128 to promote solubility). ^{1}H chemical shifts, sequential connectivities, and long-range NOEs from NMR data all indicate that the minimalist peptide assumes a topology identical to that of all other reported Cys_2–His_2 zinc finger structures.

In light of the effects that substitutions at position 118 have on DNA binding in ADR1 (Blumberg et al., 1987; Thukral et al., 1991), we felt it was crucial to ascertain the extent to which such substitutions altered details of the structure that might be relevant to the finger's ability to bind DNA. The structural analysis reported here shows that no significant structural perturbations occur when the histidine residue at position 118 is changed to either an alanine or a tyrosine, signifying that the observed loss of DNA binding affinity for these mutant proteins is due solely to the loss of a direct DNA contact.

The presumption that zinc fingers can be engineered to recognize specific DNA sequences by changing the residues at the direct DNA contact positions relies on the assumption that the domain is a solid scaffolding upon which DNA base-contacting residues can be hung. Our results suggest that one can make disparate amino acid substitutions at position 118, the middle DNA contact position in the zinc finger, with impunity regarding the overall domain structure. Results from mutagenesis studies indicate that the nature and size of the side chain at this and the other DNA contact positions will be critical in determining both the specificity and the affinity of a given finger domain. In particular, our results suggest that bulky side chains at the middle DNA contact position have the potential to disrupt other DNA contacts. This is supported by the observation of Desjarlais and Berg (1992) that the size of the side chains at the first two DNA contact positions (115 and 118 in our numbering system) affect the ability of a zinc finger to discriminate among different DNA sequences. Further detailed structural and functional studies will be required to determine whether sequence-dependent structural variations observed among zinc fingers have an impact on the manner in which these domains interact with DNA.

Materials and methods

NMR samples

The three zinc finger peptides were synthesized and purified using methods previously described (Párraga et al., 1990). The sequence of the peptide corresponds to the N-

terminal zinc finger in ADR1, spanning residues 102–130 with the site of mutation at position 118:

NH_3^+-RSFVCEVCTRAFARQE(H/A/Y)LKRHYRSHTNEK-COO^-

All peptides were analyzed by electrospray mass spectrometry and analytical reverse-phase high-performance liquid chromatography for purity. NMR samples of these peptides were prepared using the folding protocol previously described (Párraga et al., 1990). Samples were reconstituted with equimolar zinc in a 50 mM Tris/25 mM acetate buffer. H_2O samples were ~4 mM at pH 5.2 in 90% H_2O/10% D_2O (pH uncorrected for isotope effect). The pH was chosen for practical considerations, in particular to minimize amide proton exchange rates. At higher pH, significant cross saturation occurs as a result of the presaturation pulse used to suppress the solvent signal, yielding lower signal-to-noise in spectra collected in H_2O. pH titration of ADR1b revealed that only His 118 has a pK_a in the pH range over which the spectra were collected (pH 5.2–7.5), as the two histidines that serve as zinc ligands do not become protonated until lower pH (data not shown). Two-dimensional spectra were collected at pH 7.4 in D_2O to detect pH-dependent conformational changes. $^3J_{\alpha\beta}$ coupling constants and a NOESY spectrum collected indicate that the side chain of His 118 has the same X_1 rotamer (−60°) regardless of its protonation state.

NMR methods

All NMR experiments were carried out on a Bruker AM-500 spectrometer at 25 °C and a spectral width of 10,000 Hz (unless otherwise noted). Water suppression was achieved by presaturation of the water resonance during the 2.0-s relaxation delay. All experiments were processed using FTNMR (v.5.0 and v.5.2) from Hare Research, Inc. (Woodinville, Washington). COSY, RELAY, TOCSY, and NOESY experiments were performed on all three zinc finger domains and assignments were determined using conventional methods (Wüthrich, 1986). The resonance assignments of all three zinc finger peptides are available as Supplementary material on the Diskette Appendix.

$^3J_{\alpha\beta}$ coupling constants were obtained from P.E.COSY spectra recorded using the method suggested by Mueller (1987), with a flip angle of 45° for the mixing pulse. P.E.COSY experiments were collected with a spectral width of 6,410 Hz and transformed into 4,096 points in $t2$ for a resolution of 1.56 Hz/pt in the F2 dimension. Analysis of the $^3J_{\alpha\beta}$ coupling constants and the CαH-CβH and CβH-NH NOE intensity patterns in the 100-ms NOESY permitted simultaneous assignment of the X_1 rotamer and stereospecific assignment of the CβH protons (Wagner et al., 1987; Kim & Prestegard, 1990). In cases where the coupling constants could not be measured due to overlap or where the coupling constant values and NOE patterns were inconsistent with a single X_1 rotamer, no such assignments were made. Stereospecific assignments of prochiral $C\gamma H_3$ groups in the two valine spin systems (V105 and V108) were made on the basis of NOE intensity patterns in the intraresidue NOEs (Hyberts et al., 1987). Stereospecific assignment of the CβH resonances of Leu 119 also permitted stereospecific assignments for the $C\delta H_3$ groups on the basis of intraresidue NOEs. Torsion angle constraints and stereospecific assignments of prochiral groups were included in the structure calculations and are given in the Supplementary material on the Diskette Appendix. No X_1 torsion angle constraints were included for either of the two threonine residues.

A set of NOESY spectra with mixing times of 5, 50, 100, and 200 ms was recorded to measure NOE build-up rates. The mixing time was varied randomly by 10% to suppress transfer by scalar couplings in each NOESY experiment (Macura et al., 1981). In $t2$, FIDs were zero-filled to 2K, multiplied by a skewed sinebell function, and baseline-corrected using a fifth-order polynomial after Fourier transformation. In $t1$, the first point of the FID was multiplied by 0.5 to reduce $t1$ ridges (Otting et al., 1986), zero-filled to 4K, and multiplied by a skewed sinebell function. The transformed 2K × 2K matrices were further baseline corrected using the method of Dietrich and coworkers (1991), which greatly improved the quality of the spectra in terms of signal-to-noise and reproducibility of volumes for crosspeaks symmetric about the diagonal.

Build-up rates for each NOE crosspeak were determined as follows. NOE intensities were measured by volume integrals using FTNMR (v.5.0) for each of the four mixing times and displayed as a function of mixing time using in-house software. NOE build-up curves were fit with a single exponential curve when possible (Boelens et al., 1988), otherwise a simple linear fit was performed. The build-up rates obtained were used to calculate 100-ms NOE intensities for all observed NOEs.

Additionally, an isotope-labeled H118Y zinc finger peptide was synthesized with a deuterated phenyl ring at Phe 104 using N-t-BOC-phenylalanine-d_5 in order to resolve an overlap problem between the aromatic ring resonances of Phe 104 and Tyr 118. Spectra of this selectively labeled peptide provided unambiguous assignments of NOEs to the ring of Tyr 118 (similar to Kochoyan et al. [1991b]).

Structure calculations and refinement

All structure calculations were carried out on a Silicon Graphics 4D/35 Personal Iris and structures were visualized using QUANTA (Polygen Corp., Waltham, Massachusetts). Initial structures were calculated using 1.8–5.0-Å bounds for all observed NOEs. These initial structures were then submitted to the refinement protocol REPENT (described in Hoffman et al. [1993]), which employs the distance geometry algorithm DIANA (v.1.0), licensed from Dr. Kurt Wüthrich (ETH, Zürich) and the complete

relaxation matrix algorithm CORMA (v.3.0), obtained from Dr. Thomas James (UCSF, San Francisco). The goal of the refinement procedure is to improve agreement between the experimentally observed NOE intensities and those calculated from the structural models using CORMA. Agreement is quantified using a residual error function that is similar to the crystallographic R-value: the NMR residual error function, $R\langle 1/6\rangle$, where $R\langle 1/6\rangle \equiv \Sigma\{|(I_{obs})^{1/6} - (I_{calc})^{1/6}|\}/\Sigma(I_{obs})^{1/6}$ (Thomas et al., 1991). Improved models were generated iteratively through adjustment of the bounds used for each NOE constraint based on a comparison of the observed NOE intensity and the CORMA-calculated intensity. Thus, the refined models represent significant improvement over the initial models. Furthermore, no constraints for H-bonds or tetrahedral coordination of the zinc were included in the structure calculations. Structure calculation parameters for each of the refined zinc finger models are given in Table 1. Dihedral angle constraints, upper and lower distance constraints, and stereospecific assignments used in DIANA calculations are included as files in the Supplementary material on the Diskette Appendix.

Comparisons of wild-type and mutant structures

A procedure essentially identical to that of Hurle and coworkers (1992) was used to analyze differences between wild-type ADR1b and the mutants, H118A and H118Y. Average structures of each domain were calculated from the 10 models in each family based on superposition of the backbone atoms (N, CA, C, O) from residues 104–126. The two average structures to be compared were then superimposed using the same atoms employed to calculate the average structure. This allowed distances between equivalent atoms in each average structure to be calculated (i.e., Phe 104 N in $ADR1b_{ave}$ to Phe 104 N in $H118A_{ave}$). Positional shifts in atoms were then normalized by the positional variability of that atom in each of the average structures. This variability represents the degree of certainty with which the position of the atom is known in the average structure and can be adjusted to different confidence levels by incrementing the number of standard deviations in position for the atom. For example, if 2σ for atom $x = 0.50$ Å, then the position of atom x is 95% likely to be in a sphere of radius 0.50 Å about its position in the average structure; likewise 3σ for atom $x = 0.75$ Å, and thus its position is 99% likely to be in a sphere of radius 0.75 Å about its position in the average structure. The normalized positional shift is the significance parameter, S, where

$$S = \frac{\text{distance between atom } x \text{ and } x'}{[(q\sigma)^2 \text{atom } x + (q\sigma)^2 \text{atom } x']^{1/2}}$$

At the 95% confidence level, $q = 2$, whereas $q = 3$ at the 99% confidence level. If $S \geq 1.0$, then the positional shift of the atom between the two average structures is greater than the uncertainty in its position and is, at 95% or 99% certainty, a significant shift.

Acknowledgments

We thank Melissa Starovasnik and Mia Schmiedeskamp for critical reading of the manuscript, Jon Herriott for insightful discussions, and Nikola Pavletich and Carl Pabo for the Zif268–DNA crystal structure coordinates. R.C.H. was supported by an NIH training grant in molecular biophysics (T32/GM-08268). R.E.K. was supported by an NIH grant (P01/GM-32681) and an American Heart Association Established Investigatorship.

References

Adman, E.T., Sieker, L.C., & Jensen, L.H. (1991). Structure of rubredoxin from *Desulfovibrio vulgaris* at 1.5 Å resolution. *J. Mol. Biol. 217*, 337–352.

Baker, E.N. & Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol. 44*, 97–179.

Berg, J.M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Natl. Acad. Sci. USA 85*, 99–102

Berg, J.M. (1990). Zinc fingers and other metal-binding domains. *J. Biol. Chem. 265*, 6513–6516.

Blake, P.R., Park, J.-B., Adams, M.W.W., & Summers, M.F. (1992a). Novel observation of NH–S(Cys) hydrogen-bond-mediated scalar coupling in ^{113}Cd-substituted rubredoxin from *Pyrococcus furiosus*. *J. Am. Chem. Soc. 114*, 4931–4933.

Blake, P.R., Park, J.-B., Zhou, Z.H., Hare, D.R., Adams, M.W.W., & Summers, M.F. (1992b). Solution-state structure by NMR of zinc-substituted rubredoxin from the marine hyperthermophile archaebacterium *Pyrococcus furiosus*. *Protein Sci. 1*, 1508–1521.

Blumberg, H., Eisen, A., Sledziewski, A., Bader, D., & Young, E.T. (1987). Two zinc fingers of a yeast regulatory protein shown by genetic evidence to be essential for its function. *Nature 328*, 443–445.

Boelens, R., Konig, T.M.G., & Kaptein, R. (1988). Determination of biomolecular structures from proton–proton NOE's using a relaxation matrix approach. *J. Mol. Struct. 173*, 299–311.

Borgias, B.A. & James, T.L. (1989). Two-dimensional nuclear Overhauser effect: Complete relaxation analysis. *Methods Enzymol. 176*, 169–183.

Brown, R.S., Sander, C., & Argos, P. (1985). The primary structure of a transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett. 186*, 271–274.

Chakrabarti, P. (1989). Geometry of interaction of metal ions with sulfur-containing ligands in protein structures. *Biochemistry 28*, 6081–6085.

Chothia, C. (1984). Principles that determine the structures of proteins. *Annu. Rev. Biochem. 53*, 537–572.

Constantine, K.L., Madrid, M., Bányai, L., Trexler, M., Patthy, L., & Llinás, M. (1992). Refined solution structure and ligand-binding properties of PDC-109 domain b. *J. Mol. Biol. 223*, 281–298.

Creighton, T.E. (1992). *Proteins*, 2nd Ed. W.H. Freeman and Co., New York.

Day, M.W., Hsu, B.T., Joshua-Tor, L., Park, J.-B., Zhou, Z.H., Adams, M.W.W., & Rees, D.C. (1992). X-ray crystal structure of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Protein Sci. 1*, 1494–1507.

Denis, C.L. & Young, E.T. (1983). Isolation and characterization of the positive regulatory gene ADR1 from *Saccharomyces cerevisiae*. *Mol. Cell. Biol. 3*, 360–370.

Desjarlais, J.R. & Berg, J.M. (1992). Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl. Acad. Sci. USA 89*, 7345–7349.

Diakun, G.P., Fairall, L., & Klug, A. (1986). EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature 324*, 698–699.

Dietrich, W., Rüdel, C.H., & Neumann, M. (1991). Fast and precise automatic baseline correction of one- and two-dimensional NMR spectra. *J. Magn. Reson. 91*, 1–11.

Edmondson, S., Khan, N., Shriver, J., Zdunek, J., & Gräslund, A. (1991). The solution structure of motilin from NMR distance constraints, distance geometry, molecular dynamics, and an iterative full relaxation matrix refinement. *Biochemistry 30*, 11271–11279.

El-Baradi, T. & Peiler, T. (1991). Zinc finger proteins: What we know and what we would like to know. *Mech. Dev. 35*, 155–169.

Fairall, L., Harrison, S.D., Travers, A.A., & Rhodes, D. (1992). Sequence-specific DNA binding by a two zinc-finger peptide from the *Drosophila melanogaster* Tramtrak protein. *J. Mol. Biol. 226*, 349–366.

Frankel, A.D., Berg, J.M., & Pabo, C.O. (1987). Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc. Natl. Acad. Sci. USA 84*, 4841–4845.

Frey, M., Sieker, L., Payan, F., Haser, R., Bruschi, M., Pepe, G., & LeGall, J. (1987). Rubredoxin from *Desulfovibrio gigas*, a molecular model of the oxidized form at 1.4 Å resolution. *J. Mol. Biol. 197*, 525–541.

Gibson, T.J., Postma, J.P.M., Brown, R.S., & Argos, P. (1988). A model for the tertiary structure of the 28 residue DNA-binding motif ('zinc finger') common to many eukaryotic transcriptional regulatory proteins. *Protein Eng. 2*, 209–218.

Hanas, J.S., Hazuda, D.J., Bogenhagen, D.F., Wu, F.H.-Y., & Wu, C.-W. (1983). *Xenopus* transcription factor A requires zinc for binding to the 5S RNA gene. *J. Biol. Chem. 258*, 14120–14125.

Hoffman, R.C., Xu, R., Klevit, R.E., & Herriott, J.R. (1993). A simple method for the refinement of models derived from NMR data demonstrated on a zinc-finger domain from yeast ADR1. *J. Magn. Reson.*, in press.

Holmes, M.A. & Matthews, B.W. (1982). Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol. 160*, 623–639.

Hurle, M.R., Eads, C.D., Pearlman, D.A., Seibel, G.L., Thomason, J., Kosen, P.A., Kollman, P., Anderson, S., & Kuntz, I.D. (1992). Comparison of solution structures of mutant bovine pancreatic trypsin inhibitor using two-dimensional nuclear magnetic resonance. *Protein Sci. 1*, 91–106.

Hyberts, S.G., Márki, W., & Wagner, G. (1987). Stereospecific assignments of side-chain protons and characterization of torsion angles in eglin c. *Eur. J. Biochem. 164*, 625–635.

James, T.L. & Borgias, B.A. (1990). Determination of DNA and protein structures in solution via complete relaxation matrix analysis of 2D NOE spectra. In *Frontiers of NMR in Molecular Biology*, pp. 177–187. Alan R. Liss, Inc., New York.

Kim, Y. & Prestegard, J.H. (1990). Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins Struct. Funct. Genet. 8*, 377–385.

Klevit, R.E. (1991). Recognition of DNA by Cys_2,His_2 zinc fingers. *Science 253*, 1367, 1393.

Klevit, R.E., Herriott, J.R., & Horvath, S.J. (1990). Solution structure of a zinc finger domain from yeast ADR1. *Proteins Struct. Funct. Genet. 7*, 215–226.

Kochoyan, M., Havel, T.F., Nguyen, D.T., Dahl, C.E., Keutmann, H.T., & Weiss, M.A. (1991a). Alternating zinc fingers in the human male associated protein ZFY: 2D NMR structure of an even finger and implications for jumping-linker DNA recognition. *Biochemistry 30*, 3371–3386.

Kochoyan, M., Keutmann, H.T., & Weiss, M.A. (1991b). Alternating zinc fingers in the human male associated protein ZFY: Refinement of the NMR structure of an even finger by selective deuterium labeling and implications for DNA recognition. *Biochemistry 30*, 7063–7072.

Kochoyan, M., Keutmann, H.T., & Weiss, M.A. (1991c). Alternating zinc fingers in the human male-associated protein ZFY: HX_3H and HX_4H motifs encode a local structural switch. *Biochemistry 30*, 9396–9402.

Kochoyan, M., Keutmann, H.T., & Weiss, M.A. (1991d). Architectural rules of the zinc-finger motif: Comparative two-dimensional NMR studies of native and aromatic swap domains define a weakly polar switch. *Proc. Natl. Acad. Sci. USA 88*, 8455–8549.

Krizek, B.A., Amann, B.T., Kilfoil, V.J., Merkle, D.L., & Berg, J.M. (1991). A consensus zinc finger peptide: Design, high-affinity metal binding, a pH-dependent structure, and a His to Cys sequence variant. *J. Am. Chem. Soc. 113*, 4518–4523.

Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A., & Wright, P.E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science 245*, 635–637.

Macura, S., Huang, Y., Suter, D., & Ernst, R.R. (1981). Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J. Magn. Reson. 43*, 259–281.

Michael, S.F., Kilfoil, V.J., Schmidt, M.H., Amann, B.T., & Berg, J.M. (1992). Metal binding and folding properties of a minimalist Cys_2His_2 zinc finger peptide. *Proc. Natl. Acad. Sci. USA 89*, 4796–4800.

Miller, J., McLachlan, A.D., & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J. 4*, 1609–1614.

Mortishire-Smith, R.J., Lee, M.S., Bolinger, L., & Wright, P.E. (1992). Structural determinants of Cys_2His_2 zinc fingers. *FEBS Lett. 296*, 11–15.

Mueller, L. (1987). P.E.COSY, a simple alternative to E.COSY. *J. Magn. Reson. 72*, 191–197.

Neuhaus, D., Nakaseko, Y., Schwabe, J.W.R., & Klug, A. (1992). Solution structures of two zinc finger domains from SWI5, obtained using two-dimensional ^1H NMR spectroscopy: A zinc finger structure with a third strand of β-sheet. *J. Mol. Biol. 228*, 637–651.

Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K. & Gronenborn, A.M. (1990). High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry 29*, 9324–9334.

Omichinski, J.G., Clore, G.M., Robien, M., Sakaguchi, K., Appella, E., & Gronenborn, A.M. (1992). High-resolution solution structure of the double Cys_2His_2 zinc finger from the human enhancer binding protein MBP-1. *Biochemistry 31*, 3907–3917.

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986). Origin of t_1 and t_2 ridges in 2D NMR spectra and procedures for suppression. *J. Magn. Reson. 66*, 187–193.

Párraga, G.E., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T., & Klevit, R.E. (1988). Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science 241*, 1489–1492.

Párraga, G.E., Horvath, S.J., Hood, L., Young, E.T., & Klevit, R.E. (1990). Spectroscopic studies of wild type and mutant zinc finger peptides: Determinants of domain folding and structure. *Proc. Natl. Acad. Sci. USA 87*, 137–141.

Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger–DNA recognition: Crystal structure of a Zif268–DNA complex at 2.1 Å. *Science 252*, 809–817

Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem. 34*, 167–339.

Stenkamp, R.E., Sieker, L.C., & Jensen, L.H. (1990). The structure of rubredoxin from *Desulfovibrio desulfuricans* strain 27774 at 1.5 Å resolution. *Proteins Struct. Funct. Genet. 8*, 352–364.

Stickle, D.F., Presta, L.G., Dill, K.A., & Rose, G.D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol. 226*, 1143–1159.

Thomas, P.D., Basus, V.J., & James, T.L. (1991). Protein solution structure determination using distances from two-dimensional nuclear Overhauser effect experiments: Effect of approximation on the accuracy of derived distances. *Proc. Natl. Acad. Sci. USA 88*, 1237–1241.

Thukral, S.K., Morrison, M.L., & Young, E.T. (1991). Alanine scanning site-directed mutagenesis of the zinc fingers of transcription factor ADR1: Residues that contact DNA and that transactivate. *Proc. Natl. Acad. Sci. USA 88*, 9188–9192.

Thukral, S.K., Morrison, M.L., & Young, E.T. (1992). Mutations in the zinc fingers of ADR1 that change the specificity of DNA binding and transactivation. *Mol. Cell. Biol. 12*, 2784–2792.

Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N., & Wüthrich, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry: The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol. 196*, 611–639.

Watenpaugh, K.D., Sieker, L.C., & Jensen, L.H. (1979). The structure of rubredoxin at 1.2 Å resolution. *J. Mol. Biol. 131*, 509–522.

Weiss, M.A. & Keutmann, H.T. (1990). Alternating zinc finger motifs in the male-associated protein ZFY: Defining architectural rules by mutagenesis and design of an aromatic swap second-site revertant. *Biochemistry 29*, 9808–9813.

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