PROTEINS: Structure, Function, and Genetics 7:215–226 (1990)

Solution Structure of a Zinc Finger Domain of Yeast ADR1

Rachel E. Klevit,[1] Jon R. Herriott,[1] and Suzanna J. Horvath[2]
[1]Department of Biochemistry, SJ-70, University of Washington, Seattle, Washington 98195 and [2]Division of Biology, California Institute of Technology, Pasadena, California 91125

ABSTRACT **The "zinc finger" is a 30-residue repeating motif that has been identified in a variety of eukaryotic transcription factors. Each domain is capable of binding a Zn^{2+} ion through invariant Cys and His residues. We have determined the three-dimensional structure of a synthetic peptide that corresponds to one of the two zinc finger domains in the yeast transcription factor ADR1, using two-dimensional nuclear magnetic resonance spectroscopy. The Zn^{2+}-bound structure of the peptide consists of a loop containing the two Cys residues, a "fingertip," a 12- to 13-residue α-helix containing the two His residues, and a C-terminal tail. A majority of the interresidue contacts observed involve the seven conserved residues of the prototypic zinc finger (i.e., the four zinc ligands and the three hydrophobic residues), indicating that these residues are largely responsible for the three-dimensional structure of the domain and that all the zinc finger domains of the TFIIIA class will have similar structures. Potential DNA-binding residues are found throughout the structure, with the highest concentration of such residues on the external face of the α-helix.**

Key words: zinc finger, DNA-binding motif, two-dimensional NMR, synthetic peptides

INTRODUCTION

The recognition of specific sequences of DNA by proteins plays a pivotal role in the regulation of gene expression. Some understanding of these important molecular interactions has emerged from genetic, biochemical, and structural studies of one class of DNA-binding proteins, the prokaryotic repressor proteins. These proteins contain a structural motif, the helix-turn-helix, that is directly involved in their DNA binding. Much less is known about other protein motifs that interact with DNA. The zinc finger appears to represent a protein structural motif that is responsible for the sequence-specific DNA-binding activities of a variety of eukaryotic transcriptional factors.[1] First recognized as a tandemly repeating sequence that appears nine times in *Xenopus* transcription factor IIIA (TFIIIA),[2,3] these motifs have been the subject of much conjecture over the past few years.

Studies on purified TFIIIA indicated that the protein binds Zn^{2+} and that the tandemly repeating units behave as independent folding units in limited proteolysis experiments.[2] The prototypic TFIIIA-like zinc finger motif consists of 30 amino acid residues of which 7 positions are absolutely conserved: four putative zinc ligand residues (two Cys and two His), two aromatic, and one Leu residue (see Fig. 1), with the spacing between the four ligands residues being strongly conserved as well. No other positions in the 30-residue sequence are as highly conserved. Therefore, it seems likely that the 7 conserved residues are responsible to a large extent for the three-dimensional folding pattern of the zinc finger motif.

We have been applying ^1H NMR spectroscopy to synthetic peptides with zinc finger sequences from the transcriptional activator (ADR1) of the glucose-repressible alcohol dehydrogenase gene of the yeast *Saccharomyces cerevisiae*. ADR1 contains two tandem zinc finger sequences, which have been shown to be essential for the DNA-binding activity of the protein.[4] Our spectroscopic studies of the C-terminal finger peptide (designated ADR1a) showed that a single zinc finger sequence folds into a unique, compact structure in which the two Cys and two His sidechains form a tetrahedral complex with a Zn^{2+} ion.[5] Here we report the results of two-dimensional NMR (2DNMR) experiments on the other zinc finger sequence from ADR1, designated ADR1b, and shown in Figure 1.

MATERIALS AND METHODS

Materials

Peptide synthetic reagents were obtained from Applied Biosystems Inc. Deuterated Tris buffer was from MSD Isotopes (Montreal, Canada) and all other deuterated solvents were from Sigma.

Accepted August 9, 1989; revision accepted October 31, 1989.

Address reprint requests to Dr. Rachel E. Klevit, Department of Biochemistry, SJ-70, University of Washington, Seattle, WA 98195.

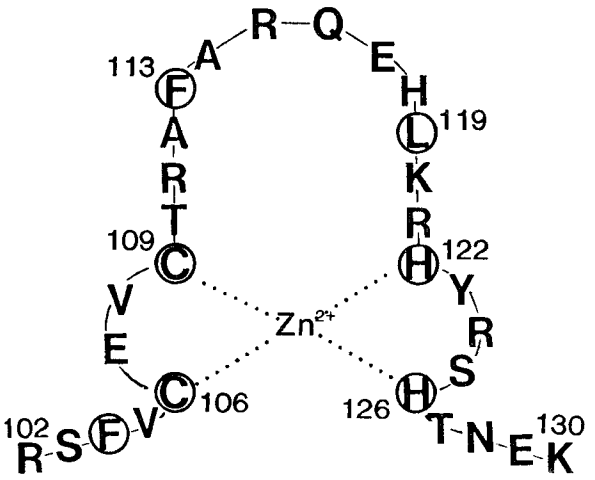


Fig. 1. Amino acid sequence of ADR1b. The sequence is numbered according to the sequence of the ADR1 protein and is drawn according to the model proposed by Miller et al.[2] The seven invariant residues are circled.

Peptide Synthesis

ADR1b peptide was synthesized by the stepwise solid-phase technique.[6,7] Purification was by reverse-phase HPLC. Characterization of the purified peptide was carried out using analytical HPLC, amino acid analysis, sequence analysis, and time-of-flight mass spectrometry.

Preparation of NMR Samples

Samples were prepared for NMR studies by dissolving purified, lyophilized material in 50 mM *d*-Tris, 25 mM *d*-acetic acid (pH 7.5, uncorrected for isotopes). A stoichiometric amount of $ZnCl_2$ was added and the samples were warmed to 70°C and allowed to cool to room temperature. For studies conducted in 90% H_2O/10% D_2O, the pH of the sample was then adjusted to pH 5.2. To prevent air oxidation, samples were purged with argon and stored at 4°C.

NMR Spectroscopy

2DNMR spectra were collected on a Bruker AM500 spectrometer. Routine parameters were 2.0 second presaturation of the H_2O resonance, 64 transients/t_1, 300 t_1 experiments per spectrum. Phase-sensitive spectra were collected in the TPPI mode and standard pulse sequences were used. The data were transformed and analyzed using FTNMR software (Hare Research, Woodinville, WA).

Distance Geometry

Metric matrix distance geometry was performed using DSPACE, provided by Dr. Dennis Hare, Hare Research, Woodinville, WA. Calculations were performed on a CONVEX C1 computer.

RESULTS AND DISCUSSION

Assignment of the Spectrum of ADR1b

The one-dimensional ^1H NMR spectrum of ADR1b in the presence of Zn^{2+} is shown in Figure 2. This spectrum contains the hallmarks of a Zn^{2+}-folded peptide spectrum, as originally described for ADR1a [5]: distinct histidine $C^{\epsilon}H$ resonances (~8 ppm), downfield-shifted $C^{\alpha}H$ resonances (~5–4.6 ppm), and upfield-shifted methyl resonances (0.9–0.5 ppm). Furthermore, the spectrum of Zn^{2+}-bound ADR1b contains only a single set of lines, indicating that a single, unique species (>95%) is present in solution, as was also observed for ADR1a.[5] Finally, since the Zn^{2+}-bound and Zn^{2+}-free forms are in slow exchange with each other, the lack of Zn^{2+}-free lines in this spectrum indicates that virtually all the peptide is in its Zn^{2+}-bound conformation under the conditions used.[8]

Representative regions of 2DNMR spectra used to obtain assignments are shown in Figure 3. The fingerprint region of a relayed coherence transfer (RELAY) spectrum is shown labeled with the sequence specific assignments (Fig. 3A). This region illustrates the chemical shift dispersion of ADR1b, which has relatively few overlapping resonances. This feature made obtaining the sequence-specific assignments quite straightforward. Since all of the assignments were obtained using conventional homonuclear 2DNMR methods, they will not be described in detail, but are summarized briefly below.

Spin system identifications were made by analysis of J-correlated (COSY), RELAY, and total coherence (TOCSY) spectra. Two Val, two Ala, two Thr, one Leu, and 11 AMQX spin systems were unambiguously identified, and the larger spin systems were assigned at least to their $C^{\gamma}H$ resonances. In the nuclear Overhauser effect spectrum (NOESY), there is only one stretch of continuous strong $d_{\alpha N}$ connectivities, from Cys-106 through Phe-113. Another stretch of weaker $d_{\alpha N}$ connectivities was observed from Lys-120 through Lys-130, with a dipeptide break at His-122–Tyr-123. In addition, several dipeptide pairs also gave $d_{\alpha N}$ NOESY peaks. In the region of the NOESY spectrum that contains cross-peaks between amide resonances, a long continuous stretch of d_{NN} connectivities was identified, starting with Gln-116 and ending with Glu-129 (Fig. 3B). As well, this region of the sequence also gave $d_{\beta N}$ connectivities. There are two dipeptides in the sequence for which no nearest-neighbor connectivities were observed: Val-105–Cys-106 and Ala-114–Arg-115 (see Fig. 3). Curiously, there is also a break in the connectivities in the dipeptide homologous to Ala-114–Arg-115 in ADR1a.[9] These breaks in nearest-neighbor connectivity do not compromise the sequential assignments, however, as the connectivities observed in either direction from the dipeptides are unambiguous. The complete set of nearest-

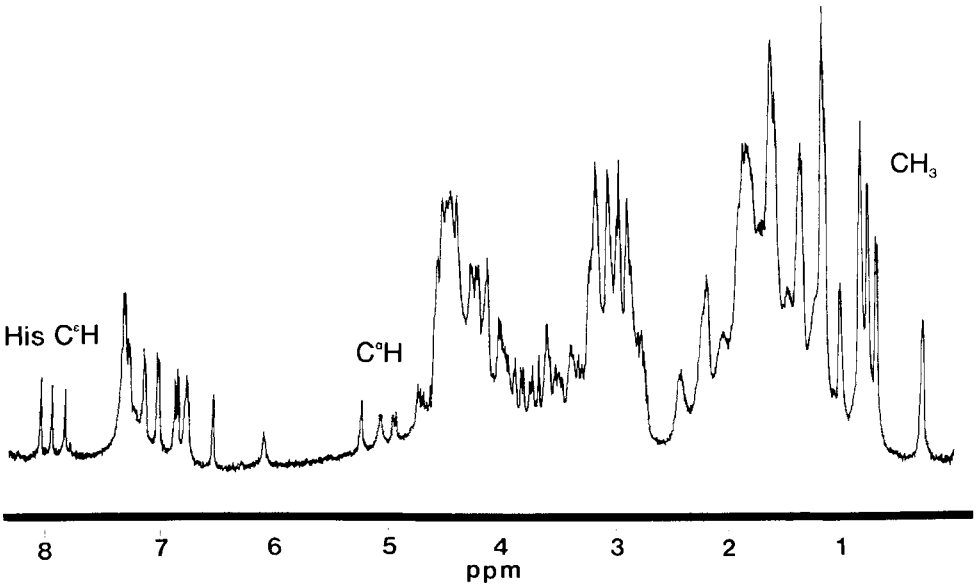


Fig. 2. 500 MHz NMR spectrum of ADR1b in D_2O reconstituted in the presence of Zn^{2+}. The resonances that signify the folded form of the peptides are noted.

neighbor connectivities observed is summarized in Figure 4.

Assignment of the Histidine Resonances

The correct assignment of the histidine imidazole proton resonances is crucial to the outcome of the structure determination described later. Unlike all other types of amino acid residues, side chain assignments for aromatic proton resonances cannot be obtained from coherence experiments (i.e., COSY, RELAY, TOCSY) because those protons are not directly coupled to the other protons in the residue. Assignments are made on the basis of a nuclear Overhauser effect (NOE) between the aromatic C^{δ} protons and (what is assumed to be) the C^{α} and/or C^{β} protons of that aromatic residue. We were concerned that certain ligand geometries in a tetrahedral zinc complex might result in the $C^{\delta}H$ of one His imidazole ring giving an NOE to the C^{α} or C^{β} protons of the other His residue, thereby confusing the assignments. However, since each His $C^{\delta}H$ gave NOEs to only a single set of C^{α} and C^{β} resonances, we assigned the imidazole protons according to conventional wisdom and then tested the assignments for self-consistency.

The identification of the nonligand (His-118) imidazole protons was straightforward. A pH titration of ADR1a (which contains only two His residues, both of which are Zn^{2+} ligands) revealed that the chemical shifts of the imidazole protons do not titrate with pH in the presence of Zn^{2+}.[8] In a pH titration of ADR1b, one set of imidazole protons shifted with pH while the other two did not, identifying the pH-sensitive resonances as His-118 (data not shown). The two ligand His assignments (His-122 and His-126) were tested later in our analysis, during the distance geometry calculations (see below). Briefly, our protocol involved switching the identities of the two sets of imidazole resonances and submitting a "switched" set of distance constraints to the distance geometry algorithm. That is, all interproton distances involving the $C^{\delta}H$ and $C^{\epsilon}H$ originally identified as His-122 were changed to His-126 in the input and vice versa, with all other distance bounds kept the same. Distance geometry calculations from independent embeds using the "switched" bounds consistently resulted in structures with a greater number of unsatisfied distance constraints, with major problems near the zinc site. Therefore, it appears that our original histidine assignments were correct since they enable the distance geometry algorithm to find structures that are more consistent with the distance estimates.

Secondary Structure of ADR1b

Regions of regular secondary structure can be identified on the basis of the patterns of short-range and medium-range NOEs observed among backbone protons.[10] β-Strands give strong $d_{\alpha N}$ connectivities and no d_{NN} connectivities, while α-helices give both d_{NN} and $d_{\beta N}$ as well as $d_{\alpha_i N_{i+3}}$ and $d_{\alpha_i \beta_{i+3}}$ connectivities. Certain types of reverse turns can also be identified by their expected patterns of connectivities.

Although a fair number of $d_{\alpha N}$ connectivities were observed for the N-terminal residues of ADR1b, this region of the peptide also gave d_{NN} peaks (see Fig. 4). Indeed, the only sequence which gave a classical β-strand-like connectivity pattern is the tetrapep-

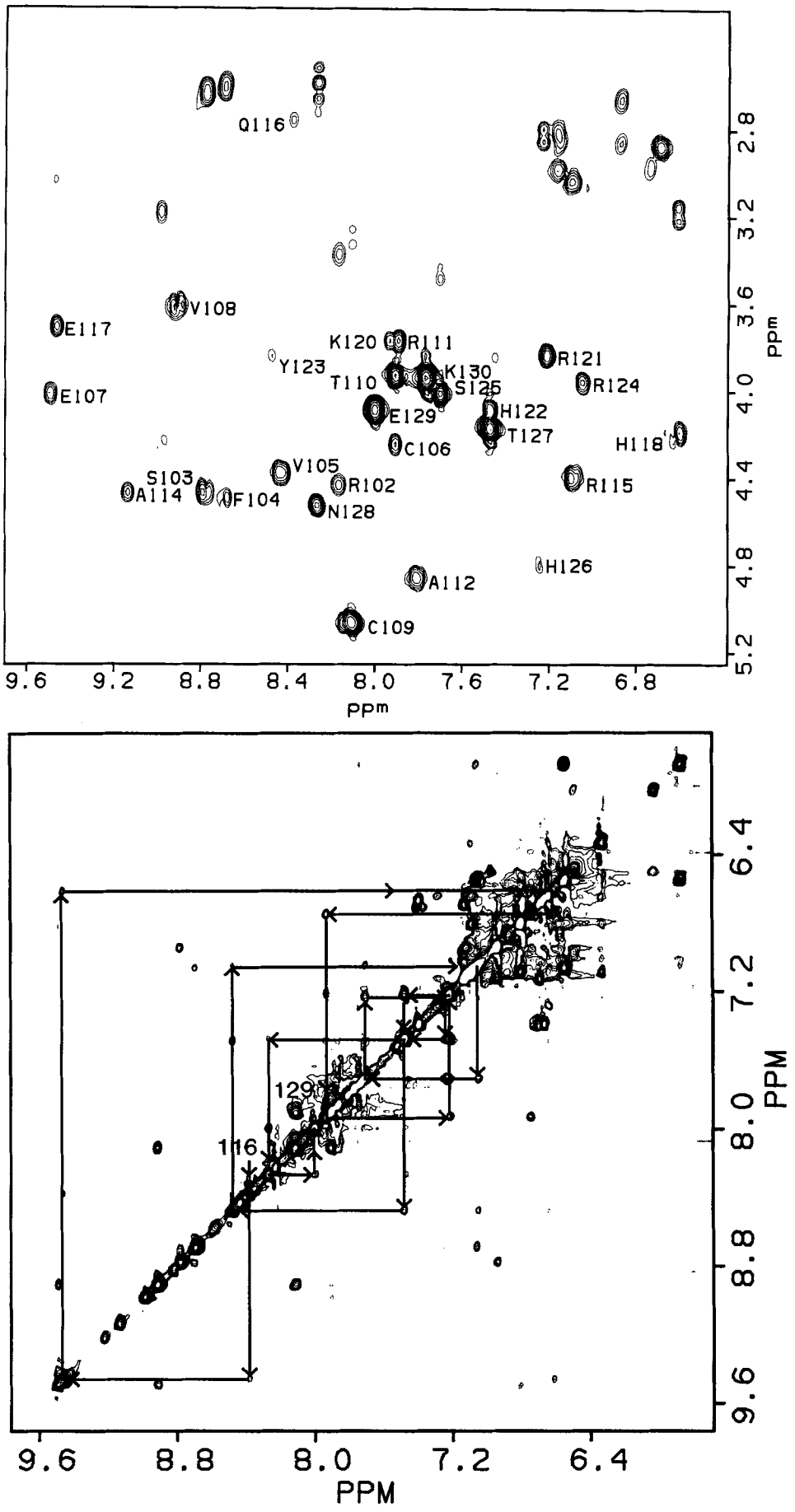

Q116
V108
E117
K120
R111
Y123
T110
K130
S125
R121
R124
E107
E129
H122
T127
C106
H118
V105
S103
A114
F104
R102
N128
R115
H126
A112
C109
9.6
9.2
8.8
8.4
8.0
7.6
7.2
6.8
ppm
2.8
3.2
3.6
4.0
4.4
4.8
5.2
ppm
129
116
9.6
8.8
8.0
7.2
6.4
PPM
6.4
7.2
8.0
8.8
9.6
PPM

Fig. 3.

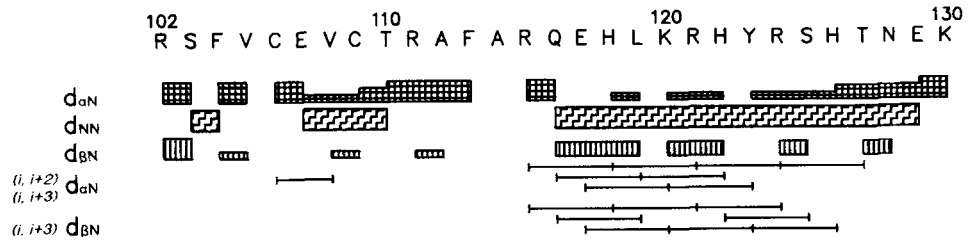


Fig. 4. Sequential connectivities for ADR1b. This represents a summary of the short-range and medium-range NOEs observed in the 200 msec NOESY. The width of the bars gives a qualitative indication of the strength of each cross-peak. The presence of medium-range NOEs is denoted by a line between the two residues.

tide Thr-110–Phe-113. However, as discussed below, this short segment is not part of a β-sheet structure. The preceding tetrapeptide, Cys-106–Cys-109, gave weak $d_{\alpha N}$ peaks and strong d_{NN} peaks. In addition, several medium-range NOEs ($i \rightarrow i+2$, $i \rightarrow i+3$) were observed, suggesting that a turn structure exists in this region. A predicted pattern of NOE connectivities was generated for the Cys-ligand-containing turn of rubredoxin[11] (residues 3–13 of rubredoxin), suggested by Berg[12] as a model for the Cys–Cys loop in the zinc finger. As shown in Figure 5, the predicted and observed pattern of NOEs agree qualitatively for residues Cys-106–Thr-110 of ADR1b.

There is a long stretch of d_{NN} connectivities from Gln-116 to Glu-129, as well as medium-range $i \rightarrow i+3$ connectivities without a break starting with Arg-115 and ending at Thr-127. This pattern of short- and medium-range NOEs is a strong indication that an α-helix extends between these residues. In a 2DNMR study of a synthetic zinc finger peptide from *Xfin*, $i \rightarrow i+2$ connectivities were observed for residues following the first His ligand,[13] suggestive of a 3_{10} helix at the C-terminal end of this peptide. No such NOEs were observed for ADR1b. We note that the *Xfin* peptide is shorter by two residues at each end than ADR1b, and the differences observed at the C-terminal end of the helix may be due to the differences in peptide length. In summary, a qualitative analysis of the connectivity patterns observed for ADR1b indicate that the peptide consists of a turn involving the two Cys residues, an extended 4-residue strand directly following the Cys–Cys turn, and a 12- to 13-residue α-helix that includes both His ligands. This is qualitatively similar to the folding topology described from 2DNMR studies on ADR1a, the other single finger peptide from ADR1.[5]

Fig. 3. 2DNMR spectra of ADR1b. 5 mM ADR1b in 50 mM *d*-Tris, 25 mM *d*-acetic acid, 5.5 mM $ZnCl_2$ pH 5.2 in 90% H_2O/10% D_2O at 25°C. **Top:** The fingerprint region of a 52 msec mixing time RELAY spectrum. The NH-$C^{\alpha}H$ for each residue is labeled. The unlabeled peaks are NH-$C^{\beta}H$ RELAY peaks. **Bottom:** The NH–NH region of a 200 msec mixing time NOESY spectrum. The beginning and end points for the long segment of d_{NN} connectivities are shown.

The Zinc Binding Site

Earlier studies on zinc finger domains and proteins have established that the Zn^{2+} is tetrahedrally coordinated by two Cys and two His residues.[5,8,14,15] However, these studies did not address which nitrogen atoms in each of the His imidazoles serves as the zinc ligand. Crystal structures of Zn^{2+}-binding proteins such as alcohol dehydrogenase[16] and thermolysin[17] illustrate that either of the nitrogens can serve as ligands in proteins. This aspect of the zinc site could be addressed directly by the interligand NOEs that were observed for ADR1b. Figure 6 illustrates the NOEs that were observed among ligand protons. There is a unidirectional path for the NOEs observed: Cys-106→His-122→His-126→Cys-109. These NOEs allow for the unambiguous identification of the nitrogen atoms that are the Zn^{2+} ligands, as only the complex in which both histidines use N^{ϵ} atoms as the ligands yields a tetrahedral complex that is consistent with the NOEs.

Long-Range Contacts

For proteins that consist of a high degree of regular secondary structure, the overall folding topology can be described by a limited number of long-range NOEs.[18] As described above, ADR1b contains a 4-residue stretch with β-strand-like connectivities, a rubredoxin-like turn, and a 12- to 13-residue α-helix. β-Strands that are part of a β-sheet structure show a well-defined pattern of NOEs between protons that are on opposite strands from each other. Indeed, there is a strong $C^{\alpha}H \leftrightarrow C^{\alpha}H$ NOE observed between Val-105 and Ala-112 in ADR1b. In addition, an NOE was identified between the $C^{\alpha}H$ of Val-105 and the NH of Phe-113. Figure 5 shows the 2-stranded antiparallel β-sheet that extends out of the Cys–Cys turn in rubredoxin.[11] The top figure illustrates all the proton pairs involving residues i and j where $i-j>1$ that are within 4 Å or less of each other (NOE-observable range). The bottom figure illustrates the proton pairs for which NOESY crosspeaks ($\tau_m = 200$ msec) were actually observed in ADR1b. Noteworthy are the long-range NOEs observed between Phe-104 and Ala-112 and between Thr-110 and Val-105, proton pairs that are farther

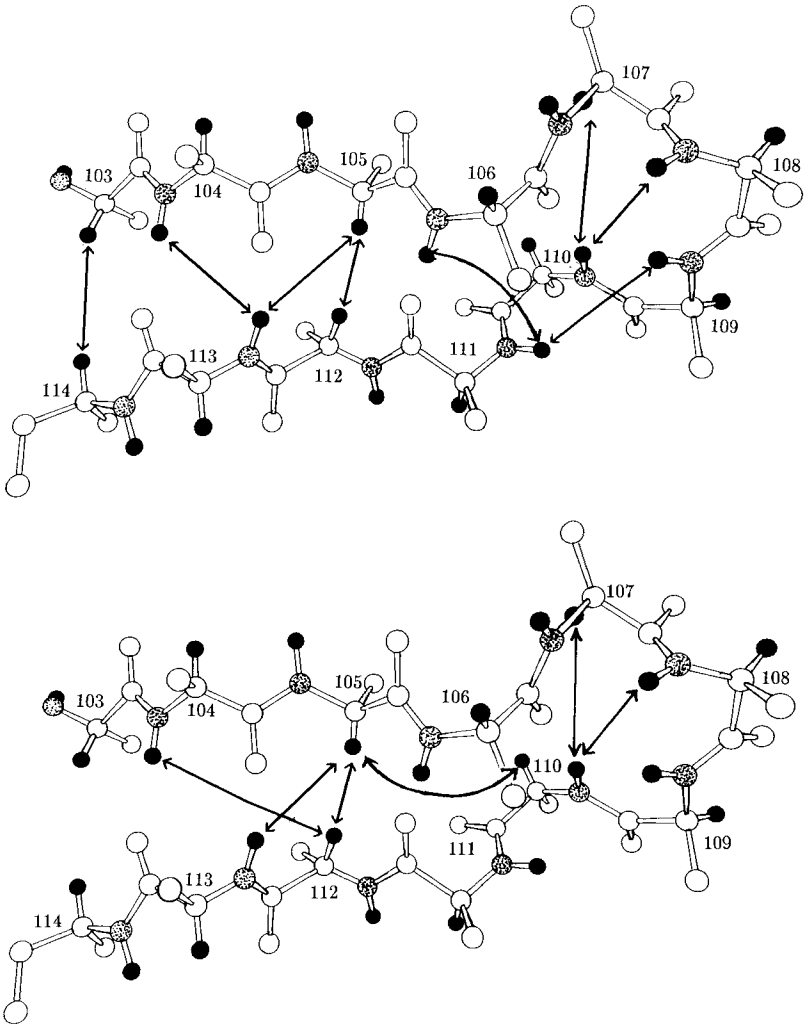


Fig. 5. β-Sheet and Cys loop structure from rubredoxin.[10] The numbering is for ADR1b, matching the positions of the two Cys residues with the rubredoxin structure. Protons have been added (dark circles) using standard bond lengths and geometries. Top: Arrows indicate the protons that are within 4 Å of each other in the rubredoxin structure, i.e., the predicted NOEs (only nonsequential NOEs are shown for the sake of clarity). Bottom: Arrows indicate the NOEs observed for ADR1b.

than 4 Å from each other in the β-sheet structure, and the absence of several predicted NOEs in the observed data. The pattern of observed NOEs indicates that while two segments of the polypeptide cross each other in an antiparallel direction at Val-105 and Ala-112, the tertiary structure of this region is not a regular 2-stranded β-sheet.

Figure 7 shows a diagonal contact plot in which the long-range NOEs observed for ADR1b are summarized.[19] Several conclusions can be drawn from this plot. Thirteen residues out of 30 are involved in tertiary contacts that result in NOE cross-peaks and these residues are found in patches: residues 104–109, 112–113, 119–120, 122–123, and 126. Neither the first two N-terminal nor last four C-terminal residues gave any long-range NOEs. An analysis of the residues involved revealed that each of the 13 residues is either one of the 7 conserved zinc finger residues (i.e., the four zinc ligands plus the 3 hydrophobics) or a nearest-neighbor to a conserved residue. Thus, the tertiary contacts that determine the three-dimensional structure of the Zn^{2+}-bound peptide are largely among the 7 residues common to all TFIIIA-like zinc finger domains, suggesting that the folding topology for all such domains will indeed be similar.

The stripe of contacts that lie parallel to the diagonal in Figure 7 from residues 115 to 127 are the medium-range NOEs indicative of the α-helix. In this part of the sequence, there is an alternating pattern of two residues each that do and do not participate in long-range contacts. The interpretation of this pattern is quite straightforward—one side of the α-helix faces toward the interior of the structure and has tertiary contacts while the opposite side faces out toward the solvent. As would be expected

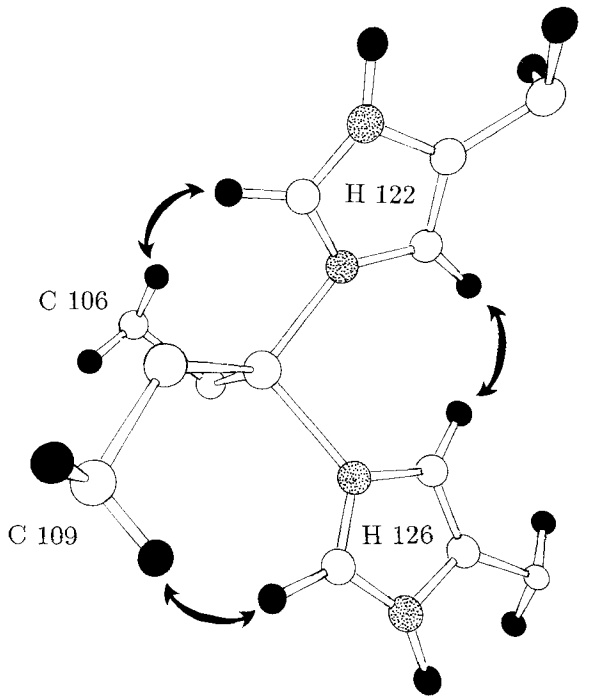


Fig. 6. Tetrahedral zinc complex in ADR1b. The interligand NOEs are illustrated on an ORTEP[25] drawing with ligand configuration as determined from the distance geometry calculations. H atoms are in black and N atoms are stippled.

for a soluble protein, the face containing hydrophobic residues is toward the interior allowing Leu-119 to interact with both Phe-104 and Phe-113. One helical turn away from Leu-119 is His-122 (a zinc ligand) which has tertiary contacts with Phe-113 as well as with residues in the Cys–Cys turn. Finally, one helical turn from His-122 is His-126, which also gives long-range NOEs to protons in the Cys–Cys turn.

The NOEs involving residues 103–112 result in a stripe that is roughly perpendicular to the diagonal in the contact plot (Fig. 7), which is the pattern expected for a turn conformation. We note that the stripe is not precisely a perpendicular line as would be expected for two antiparallel β-strands.

Three-Dimensional Structure of ADR1b

On the basis of the medium- and long-range NOEs described, a model for the overall three-dimensional folding pattern of ADR1b could be built, similar to that presented earlier for ADR1a.[5] In order to avoid introducing any of our own bias into the model, metric matrix distance geometry (DSPACE) was applied to ADR1b.[20] Upper and lower bounds were estimated from cross-peak volumes in a 200 msec mixing time NOESY, using a d_{NN} distance of 2.8 Å for protons in the α-helix as the ruler. In addition to the experimental distance estimates and the standard amino acid geometries used in DSPACE, several nonmeasured distances were added, namely Zn–S distances of 2.25–2.35 Å for the Cys ligands and Zn–N^{ϵ} distance of 2.08–2.09 Å for the His ligands. The upper and lower bounds for these were taken to be the shortest and longest distances observed for similar Zn^{2+} ligand pairs from examples found in the Protein Data Bank. However, the geometry of the Zn^{2+} complex that resulted when only the four Zn^{2+} ligand distances were input was significantly nontetrahedral, suggesting that there were not enough measured proton–proton distances around the Zn^{2+} center to constrain the ligand bond angles to near 109°. Therefore, six additional nonmeasured distances corresponding to the edges of a tetrahedron (calculated using the upper and lower bounds for the Zn ligand distances and tetrahedral geometry) were then included to better constrain the zinc geometry. Note, however, that none of these nonmeasured bounds defines the *configuration* of the four ligands around the zinc. Even so, the absolute ligand configuration around the zinc ion was reproduced in greater than 80% of the calculated structures that were right-handed, indicating that the collection of measured distance constraints involving protons on the ligand residues is more consistent with one of the two possible configurations. (Distances per se do not distinguish between overall handedness of structures. Therefore, approximately half of all structures calculated with distance geometry have a left-handed folding topology and must either be inverted or discarded.)

Although the spectrum of ADR1b is quite well-resolved, there are several frequencies at which more than one proton resonates, thereby rendering ambiguous the crosspeaks emanating from that frequency.* In initial calculations, only completely unambiguous cross-peaks were used. These early structures were then used to resolve some of the cross-peak ambiguities in the following manner. If a cross-peak lies at the cross section of two frequencies each of which has, for instance, two resonances, *a* and *b* along f_1 and *c* and *d* along f_2, the cross-peak could arise from any of four possible proton pairs: *a–c*, *a–d*, *b–c*, or *b–d*. The predicted distance for each of these pairs was calculated for each of the early structures and in cases where a given pair of protons

*It is worth noting that there are actually two assignment problems in NMR. The first is the assignment of resonance positions (i.e., along the diagonal) to individual protons in the molecule. This is accomplished by the sequential assignment method and is mainly concerned with cross-peaks arising between amide protons and other amide, C^{α}, or C^{β} protons. Once these assignments are obtained (they are reported in the literature as "complete resonance assignments"—a confusing term), the second assignment problem involves identifying the proton "parents" for each cross-peak observed in the NOESY spectrum. Due to the fact that the amount of spectral overlap tends to increase as one proceeds out the side chains, it is possible to encounter a fair amount of ambiguity in parent identification even in a spectrum for which "complete" resonance assignments have been determined.

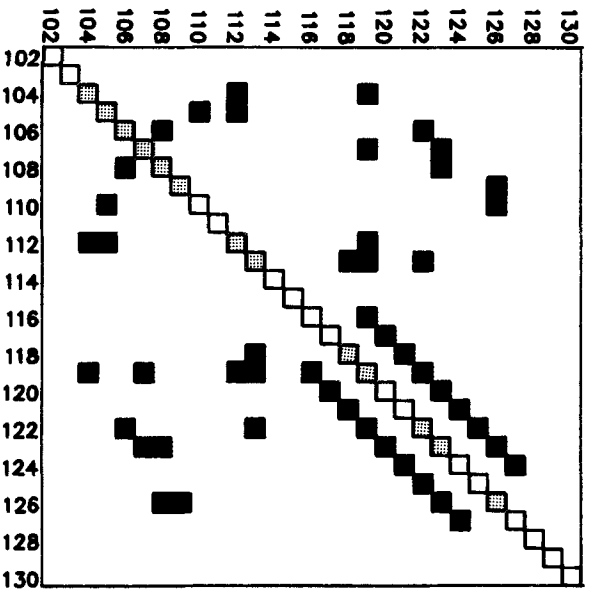


Fig. 7. Diagonal plot for ADR1b. Based on the contact plot suggested by Nishikawa et al.,[19] a dark square off the diagonal signifies that an NOE was observed between protons in the two residues shown along the axes. The cross-hatched squares along the diagonal indicate the residues that have long-range contacts and the open squares indicate the residues with no long-range contacts.

was predicted to be within 5 Å in *every* structure and *none* of the other possible pairs was within 5 Å in more than one out of six structures, the "ambiguous" peak was considered assigned and added to the bounds list used as input for new structure calculations.[21] Iterations of this process yielded several new "parent assignments" and resulted in significantly lower RMSD values among calculated structures.

Figure 8 shows one of the structures which was found to be consistent with the input distance constraints. Superimposed on the structure are the proton pairs that were used as distance constraints.[22] In this representation it is clear that ADR1b contains well-restrained regions, such as the α-helix and the interior core, and less constrained regions, such as the C-terminal four residues (127–130), the top of the finger (residues 114–115), and the first two N-terminal residues. Hence, these regions cannot be defined relative to the rest of the structure by NMR measurements. While it is not surprising that the two termini of the peptide domain are ill-defined in solution, the lack of NOEs in the "fingertip" may be a consequence of some dynamic process that may be of some mechanistic importance (see later).

Figure 9 shows 6 of the 10 most recent structures calculated. We choose to show individual structures as opposed to an "average" structure or a superposition of structures as we feel it gives a clearer view of both the variations and the similarities. These structures are shown as refined by DSPACE without further manipulation. While we feel that protocols such as energy minimization may refine these structures cosmetically, leading, for example, to better dihedral angles and other geometries, such protocols will not alter the main features of these structures. The structures generated by distance geometry, while not identical in detail, are all composed of the same four structural units: (1) the Cys-containing loop (residues 103–113), (2) the "fingertip" (residues 114–115), (3) the 3-turn α-helix (residues 116–126 or 127), and (4) the C-terminal tail (residues 127–130). As expected, the greatest variations among the structures are in the ill-defined regions—the C-terminal "tail" moves relative to the rest of the structure and the "fingertip" changes its position and shape. Such variation will not be removed by further refinement because it is inherent in the original measurements (i.e., the absence of NOEs.) While the tail structure is likely to be a consequence of the fact that we are studying a fragment of a protein, we feel that the main core of the structure, composed of the remaining three structural units is likely to represent the structure of TFIIIA-like zinc finger domains in this class of DNA-binding proteins.

Although at the present level of structure refinement, the side chain conformations and positions cannot be analyzed in detail, there are several general observations that can be made from the structures in Figure 9. In discussing and interpreting structural models generated by distance geometry, it is important to distinguish between elements of a structure that are well-restrained by the observed NOEs and those that are not. For example, although the resonances for most of the protons near the ends of the long hydrophilic side chains (such as Gln, Arg, and Lys) have been assigned, they rarely give interresidue NOE cross-peaks. Therefore, the exact location of such side chains relative to the rest of the structure cannot be defined. On the other hand, the side chain protons from the hydrophobic residues such as Phe, Leu, and Val participate in many interresidue NOEs, indicating that the conserved hydrophobic residues interact in the core, as suggested in the original zinc finger hypothesis.[2,3]

Although the precise location of many of the exterior side chains is not defined, their lack of NOEs indicates that they are indeed on the exterior of the structure and are probably fairly flexible. As summarized in the diagonal plot (Fig. 7), the residues that gave no long-range NOEs are (excluding the residues at the extreme termini) Arg-111, Ala-114, Arg-115, Gln-116, Glu-117, Lys-120, Arg-121, Arg-124, and Ser-125. The list contains virtually all the residues in the ADR1b sequence that would be recognized as potential DNA-binding side chains. A comparison of the list and the structures in Figure 9 shows that these residues are found in each of the three structural units, i.e., Arg-111 is in the Cys–Cys loop, Ala-114 and Arg-115 form the fingertip,

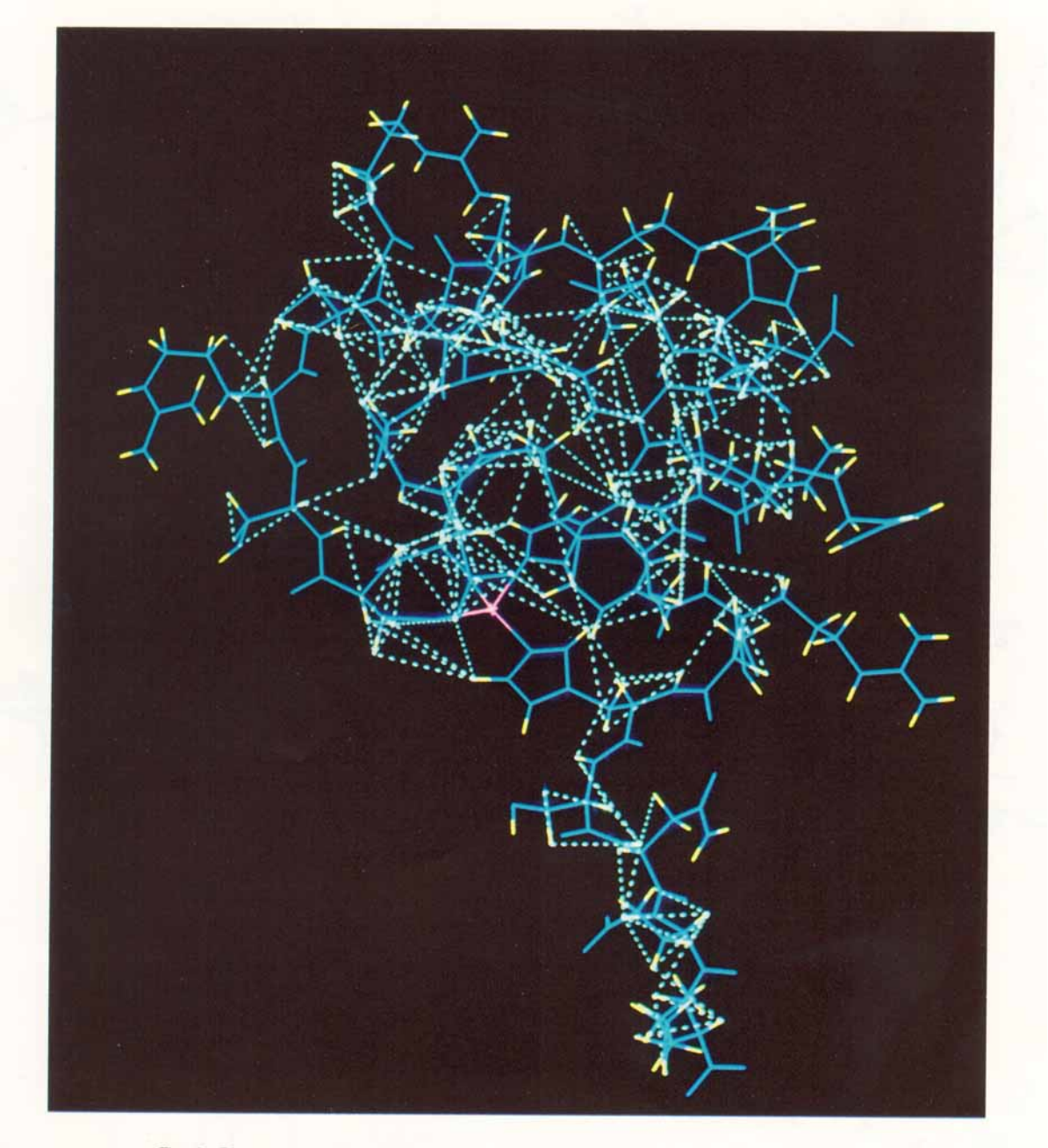
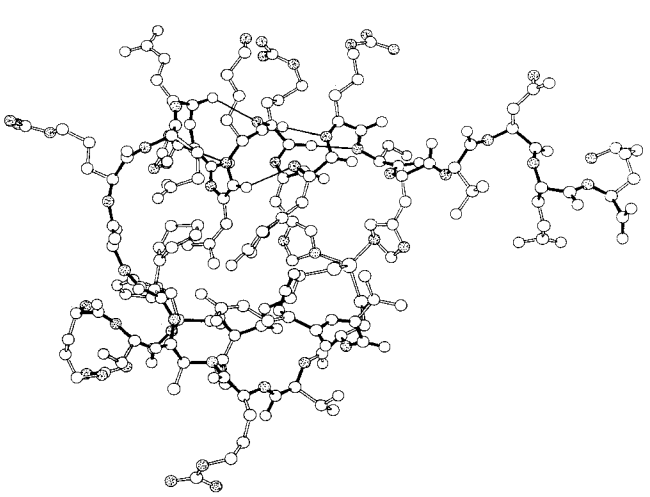
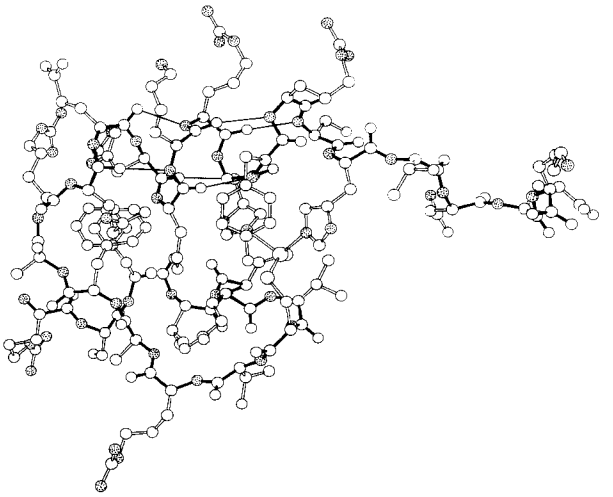


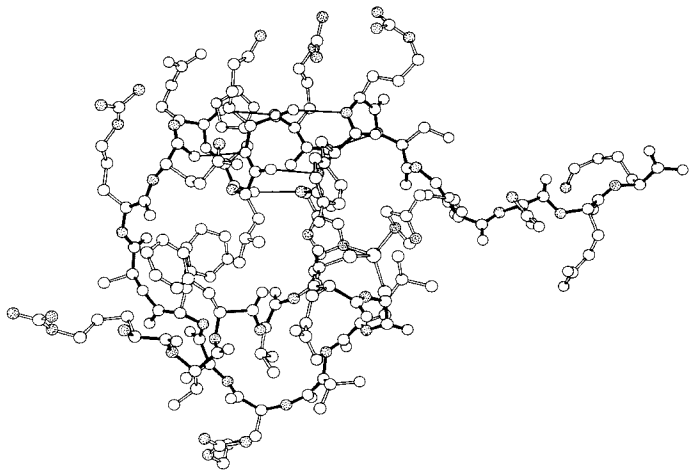
Fig. 8. Distance constraints, illustrated by dashed yellow lines, are shown superimposed on one of the calculated structures of the zinc finger peptide, ADR1b. The zinc atom is shown in pink. The model shown is the same as the numbered model in Figure 9. (Methyl groups are represented by the carbon only.)

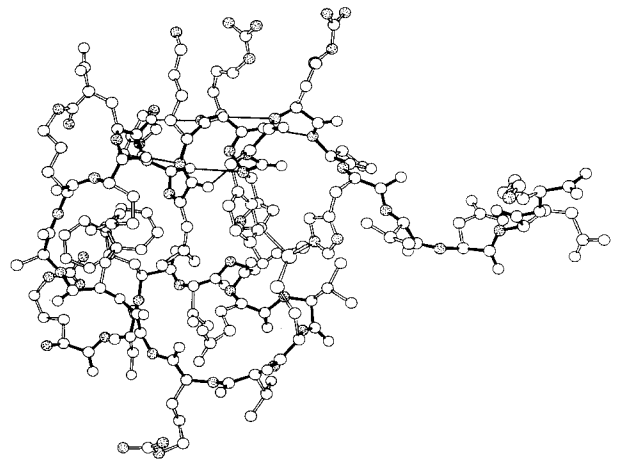
and the other side chains are on the outer face of the α-helix.

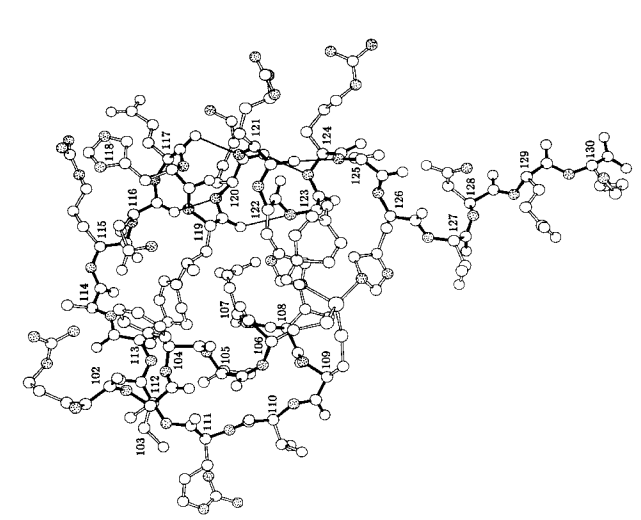
There is no direct evidence yet to indicate which residues are in contact with DNA when ADR1 protein binds to its dyad symmetrical upstream activator sequence. Certainly, the 3-turn α-helix appears to be a good candidate for this role as it contains several possible H-bond donors and acceptors. There are genetic and biochemical data that implicate the fingertip and the N-terminal region of the α-helix as important for ADR1 function. Blumberg and coworkers have identified ADR1 mutants in which Ala-114 was replaced by a Val as well as mutants in which the analogous position in the other finger, Thr-142, was replaced by an Ile.[4] DNA-binding measurements performed with these mutant proteins

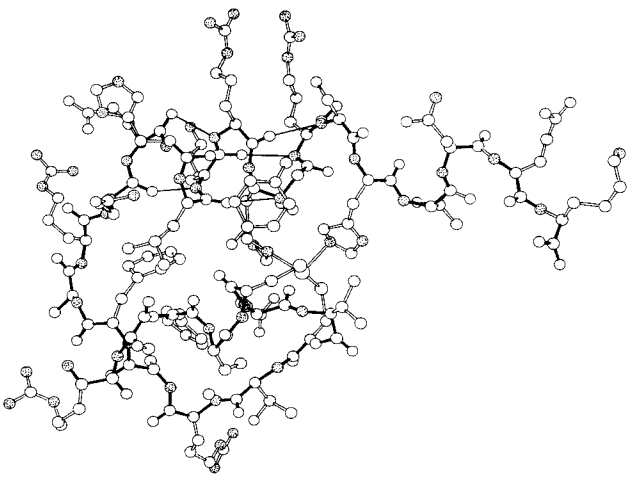









102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130



show that their affinity is decreased relative to the wild-type protein while the DNA sequence specificity is retained.[23] The other "fingertip" residue is an arginine in both fingers and is often either Arg, Lys, or Gln in other putative fingers. Thus, it is possible that the fingertip is involved in a nonspecific DNA contact and that the substitution of a bulky Val or Ile for the Ala or Thr in the first fingertip position disrupts this interaction. There are, however, a few putative zinc finger sequences that contain a Val or Ile in the position analogous to Ala-114, but since there are no DNA-binding data available for these proteins, it is impossible to draw further conclusions.

In their analysis of ADR1 mutants, Blumberg et al. also identified a mutation that substituted a Tyr for His-118, in the first turn of the α-helix.[23] This mutant protein has no detectable DNA binding activity. Such a result could be the consequence of either a significant structural perturbation that leads to the abrogation of DNA binding or of the loss of a specific side chain that is crucial to the DNA binding. Our preliminary analysis of an ADR1b peptide that has a Tyr as residue 118 indicates that the peptide maintains its structural integrity,[24] indicating that the latter is a more likely explanation.

Zinc fingers have been the subject of several structural predictions.[12,25] Berg's approach was to identify structural units of known zinc-binding proteins to serve as models for the zinc finger structure. The resulting model consisted of two units of secondary structure, each containing two of the four Zn^{2+} ligands, and connected at the "top" (what we call the fingertip) of the structure. The two Cys ligands were predicted to be in a 2-stranded antiparallel β-sheet and turn similar to that found in rubredoxin, while the two His ligands were predicted to be part of an α-helix, analogous to the thermolysin structure. The model offerred by Argos and co-workers was qualitatively similar. The residues that we have found to be in the α-helices of both ADR1b and ADR1a[5] are exactly those predicted by Berg. The turn containing the two Cys ligands also appears to be qualitatively similar. The loop that contains that turn and the polypeptide chains on either end of the turn, as determined from the NMR data, is a less regular structure that is different from both the antiparallel β-sheets in the predicted structures.[12,25] As shown in Figure 5, the NOEs observed for ADR1b do not match those predicted for the more regular β-sheet structure from rubredoxin. However, to investigate this point further, we prepared a model for ADR1b with a regular β-sheet structure by adjusting the dihedral angles of the residues in question to those of the predicted β-sheet and submitted this structure to distance geometry refinement using the observed NOEs. In all the resulting structures, the β-sheet structure was changed back to the less regular loop structure. Thus, we feel that this open looped structure may be peculiar to zinc fingers (and therefore would have been difficult to predict based on known structures.) Still, the overall qualitative similarity between the structures presented here and that presented by Berg indicates that the use of known substructures to predict new protein structures is a promising approach to the prediction of three-dimensional protein structure from primary structure.

It now seems clear that while TFIIIA-like zinc finger domains are responsible for sequence-specific DNA binding, this activity requires the presence of more than one finger domain in tandem (or perhaps, dimers of single finger proteins).[4,5,13] Thus, the precise determinants for sequence specificity probably arise from a combination of residues on neighboring fingers. Therefore, structural determination of multiple-finger-containing molecules will be important in advancing our understanding of the mechanism of zinc finger–DNA interactions. NMR data obtained for a double finger peptide containing the sequences of both ADR1b and ADR1a indicate that each of the finger sequences folds independently into discrete structural domains similar to the single finger structures of ADR1a[5] and ADR1b (this work, 24). Thus, the single finger structure reported here can be used to guide our thinking about zinc finger domains in proteins and their potential interactions with DNA. As more genetic and biochemical data become available on these proteins, it may be possible to begin to identify the components of the structure that are directly involved in the sequence-specific DNA binding activity. However, in light of the flexible nature of the side chains that are the most likely to be involved, we believe that a detailed description of zinc finger–DNA interactions must await the determination of a zinc finger–DNA complex either by NMR or X-ray diffraction methods.

Fig. 9. Six of the structures for ADR1b generated by the distance geometry algorithm. Protons have been removed for clarity. The mainchain is highlighted by black bonds. The Cys loop is on the left, the "fingertip" is at the top, the α-helix is on the right (H-bonds in thin black lines), and the tail is at the bottom of each of the structures. (Figure was rotated for publication.)

ACKNOWLEDGMENTS

This work was supported by NIH 2 PO1 32681 (REK). We wish to thank G. Drobny for the use of the NMR spectrometer, D. Hare for his advice on distance geometry, P. Weber for sharing the idea behind the IDNOE program, and our colleagues G. Parraga, E.T. Young, and L. Hood for their continued interest and support.

REFERENCES

1. Klug, A., Rhodes, D. "Zinc fingers": A novel protein motif for nucleic acid recognition. Trends Biochem. Sci. 12:464–469, 1987.

2. Miller, J. McLachlan, A.D., Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. 4:1609–1614, 1985.
3. Brown, R.S., Sander, C., Argos, P. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett. 186:271–274, 1985.
4. Blumberg, H., Eisen, A., Sledziewski, A., Bader, D., Young, E.T. Two zinc fingers of a yeast regulatory protein shown by genetic evidence to be essential for its function. Nature (London) 328:443–445, 1987.
5. Parraga, G., Horvath, S., Eisen, A., Taylor, W.E., Hood, L., Young, E.T., Klevit, R.E. Zinc-dependent structure of a single-finger domain of Yeast ADR1. Science 241:1489–1492, 1988.
6. Bruist, M.F., Horvath, S.J., Hood, L.E., Steitz, T.A., Simon, M.I. Synthesis of a site-specific DNA-binding peptide. Science 235:777–780, 1987.
7. Roise, D., Horvath, S.J., Tomich, J.M., Richards, J.H., Schatz, G. A chemically synthesized pre-sequence of an imported mitochondrial protein can form an amphiphilic helix and perturb natural and artificial phospholipid bilayers. EMBO J. 5:1327, 1986.
8. Parraga, G., Horvath, S., Hood, L., Young, E.T., Klevit, R.E. Spectroscopic studies of wildtype and mutant zinc finger peptides: Determinants of domains folding and structure. Proc. Natl. Acad. Sci. U.S.A., in press.
9. Parraga, G., Klevit, R.E., to be published.
10. Billeter, M., Braun, W., Wüthrich, K. Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations. J. Mol. Biol. 155:321–346, 1982.
11. Watenpaugh, K.D., Sieker, L.C., Jensen, L.H. Crystallographic refinement of rubredoxin at 1.2 Å resolution. J. Mol. Biol. 138:615–633, 1980. Coordinates used were from the Protein Data Bank, identification code 4RXN.
12. Berg, J.M. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. Proc. Natl. Acad. Sci. U.S.A. 85:91–102, 1988.
13. Lee, M.S., Cavanagh, J., Wright, P.E. Complete assignment of the ^{1}H NMR spectrum of a synthetic zinc finger from *Xfin*. Sequential resonance assignments and secondary structure. FEBS Lett. 254:159–164, 1989.
14. Frankel, D.A., Berg, J.M., Pabo, C.O. Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc. Natl. Acad. Sci. U.S.A. 84:4841–4845, 1987.
15. Diakun, G.P., Fairall, L., Klug, A. EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. Nature (London) 324:698–699, 1986.
16. Eklund, H., Norstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, T., Soderberg, B.O., Tapia, O., Branden, C.I. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. J. Mol. Biol. 102:27–59, 1976.
17. Holmes, M.A., Matthews, B.W. Structure of thermolysin refined at 1.6 Å resolution. J. Mol. Biol. 160:623–639, 1980.
18. Klevit, R.E., Waygood, E.B. Two-dimensional ^{1}H NMR studies of histidine-containing protein from *Escherichia coli*. 3. Secondary and tertiary structure as determined by NMR. Biochemistry 25:7774–7781, 1986.
19. Nishikawa, K., Ooi, T., Isogai, Y., Saito, N. Tertiary structure of proteins. I. Representation and computation of the conformations. J. Phys. Soc. Jpn. 32:1331, 1972.
20. Crippen, G.M. "Distance Geometry and Conformational Calculations" Research Studies. New York: Wiley, 1981.
21. This protocol is a variation on an unpublished algorithm, IDNOE, written by Dr. Paul Weber, Hare Research, Woodinville, WA.
22. This representation was suggested by Dr. Angela Gronenborn, National Institutes of Health.
23. Eisen, A., Young, E.T. Unpublished results.
24. Parraga, G., Horvath, S. Klevit, R.E. Unpublished results.
25. Gibson, T.J., Postma, J.P., Brown, R.S., Argos, P. 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, 1988.
26. Johnson, C.K. ORNL-3794 UC-4 Chemistry TID4500, 1965.