INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

NMR Studies of the DNA-Binding Domain of ADR1

by

Mia Ruth Schmiedeskamp

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

1996

Approved by Rachel E. Klevit
(Chairperson of Supervisory Committee)

Program Authorized
to Offer Degree Biochemistry

Date December 18, 1996

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Mia J. Lucchesi

Date December 18, 1996

University of Washington

Abstract

NMR Studies of the DNA-Binding Domain of ADR1

by Mia Ruth Schmiedeskamp

Chairperson of the Supervisory Committee: Professor
Rachel E. Klevit
Department of Biochemistry

Nuclear magnetic resonance spectroscopy was used to study the DNA-binding domain of yeast transcription factor ADR1. A polypeptide containing the minimal DNA-binding region was overexpressed in *E. coli* and purified to homogeneity. The construct, ADR1z, contains two Cys_2-His_2 zinc fingers and an N-terminal sequence required for tight DNA binding. Resonance assignments are presented for ADR1z free and bound to specific DNA. The assignment strategy involved traditional NOESY methods, specific amino acid labeling, and triple resonance experiments. Structural data are presented in the form of chemical shift perturbation maps, ^{13}C chemical shift index analyses, NOE connectivities, and $^{3}J_{HN\alpha}$ coupling constants. The zinc fingers of ADR1z differ little in structure from single finger peptides and do not interact in the absence of DNA. The zinc fingers do not alter structure upon DNA binding. In contrast, the N-terminal flanking sequence is random coil

in the absence of DNA. This region becomes less mobile, less exposed to solvent, and more structured upon DNA binding. The N-terminus does not interact with exposed portions of the zinc fingers in the complex and likely enhances binding by contacting the DNA directly. No fully regular pattern of secondary structure is detected for the bound N-terminus. This region is not helical; it probably adopts a largely extended conformation. The orientation of the N-terminus with respect to the DNA site is deduced from perturbations induced by fixed paramagnetic cobalt. The N-terminus doubles back and runs antiparallel to the zinc fingers, approaching the center of the palindrome. Chemical shift perturbations are used to map the protein-DNA interface. In striking agreement with previous mutagenesis experiments, these data suggest that contacts arise from R115, H118, and R121 of finger 1 and R143 of finger 2. The fingers bind DNA in different orientations, with the entire helix of finger 1 but only the extreme N-terminus of the helix of finger 2 in proximity to DNA. Suboptimal contact by finger 2 likely necessitates the unusual third contact of finger 1 and the participation of the N-terminus in DNA binding.

Table of Contents

List of Figures

List of Tables

Abbreviations

ADH2	alcohol dehydrogenase II
βME	β-mercaptoethanol
CD	Circular dichroism
COSY	correlation spectroscopy
DBD	DNA binding domain
EMSA	electrophoretic mobility shift assay
EXAFS	extended x-ray absorption fine structure
H-bonds	hydrogen bonds
HSQC	heteronuclear single quantum correlation spectroscopy
NMR	nuclear magnetic resonance spectroscopy
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
RMSD	root mean square deviation
TOCSY	total correlation spectroscopy
TPPI	time proportional phase incrementation
Tris	Tris(hydroxymethyl)-aminomethane
UAS1	upstream activation sequence I

Acknowledgments

This work could not have been completed without the help of many members of the Department of Biochemistry at The University of Washington. I thank members of the Young and Davis labs for sharing their expertise; my committee, including especially Professor Jon Herriott, for lending advice and support; and past and present members of the Klevit lab, including especially Melissa Starovasnik and Beverly Castner, for giving precious encouragement. I thank Beverly Castner for purifying a number of the protein samples used in this study. I thank David Hyre for his work on the dynamics of ADR1z' and collaboration on the final stages of this project. Finally, I thank my mentors: Ponni Rajagopal, who spent countless hours implementing the NMR experiments used in this study and patiently teaching me to use them, and Rachel Klevit, who supported me greatly both personally and professionally during my time in her lab.

For my father, Jay Schmiedeskamp, my mother, Jane, and my sister, Kendra.

INTRODUCTION

I. Objectives and Summary of Thesis Research

The goal of this study is to increase understanding of the DNA binding function of the yeast transcription factor ADR1 from a structural perspective. The minimal ADR1 DNA-binding domain contains two Cys_2-His_2 zinc fingers and an additional N-terminal sequence required for high-affinity binding. The characterization of this domain took place in three main stages:

1) Development of an *in vitro* molecular system amenable to NMR, allowing study of an ADR1 DNA-binding fragment free in solution and bound to a specific DNA site.

2) Assignment of NMR spectra of the ADR1 fragment both free and bound to DNA.

3) Characterization of the structure of the ADR1 fragment both free and bound to DNA, with an emphasis on understanding the DNA binding function of the molecule.

The results of the first stage of research are presented in Chapter 1. An efficient method for producing a fragment of ADR1 containing the minimal DNA binding domain was developed. This

fragment, ADR1z', is appropriate for NMR study in terms of size, solubility, solution behavior, and ease of incorporation of NMR-active isotopic labels. Cognate DNA was designed that contains a specific binding site for ADR1 and that interacts strongly with the ADR1z' fragment. The resulting ADR1z'-DNA complex is also amenable to NMR and allows structural study of the protein-DNA interaction.

Chapter 2 describes the second stage of research, the assignment process. Resonance assignments are a prerequisite for the interpretation of NMR spectra to characterize structure, dynamics, and intermolecular interactions. Nearly complete assignments were obtained for the backbone and sidechain ^{1}H, ^{15}N, and ^{13}C resonances of ADR1z', both free and bound to DNA. The strategy involved traditional NOESY and TOCSY experiments, isotopic labeling of specific amino acids, and various triple resonance experiments.

The final stage of the project is addressed in Chapters 3, 4, and 5. Chemical shifts, NOE connectivities, and coupling constants are used to describe the structure of the ADR1z' polypeptide and to characterize changes upon DNA binding. Chemical shift perturbation maps are used to describe the protein-DNA interface, with an emphasis on comparison to existing mutagenesis data. The orientation of the protein on the DNA is described using fixed spin

labels that were incorporated in the metal binding sites of the zinc fingers of ADR1.

Chapter 6 presents a summary and discussion of the major conclusions of the research, and the final chapter describes the methodology used at each stage of the research in detail.

II. Introduction to Zinc Fingers

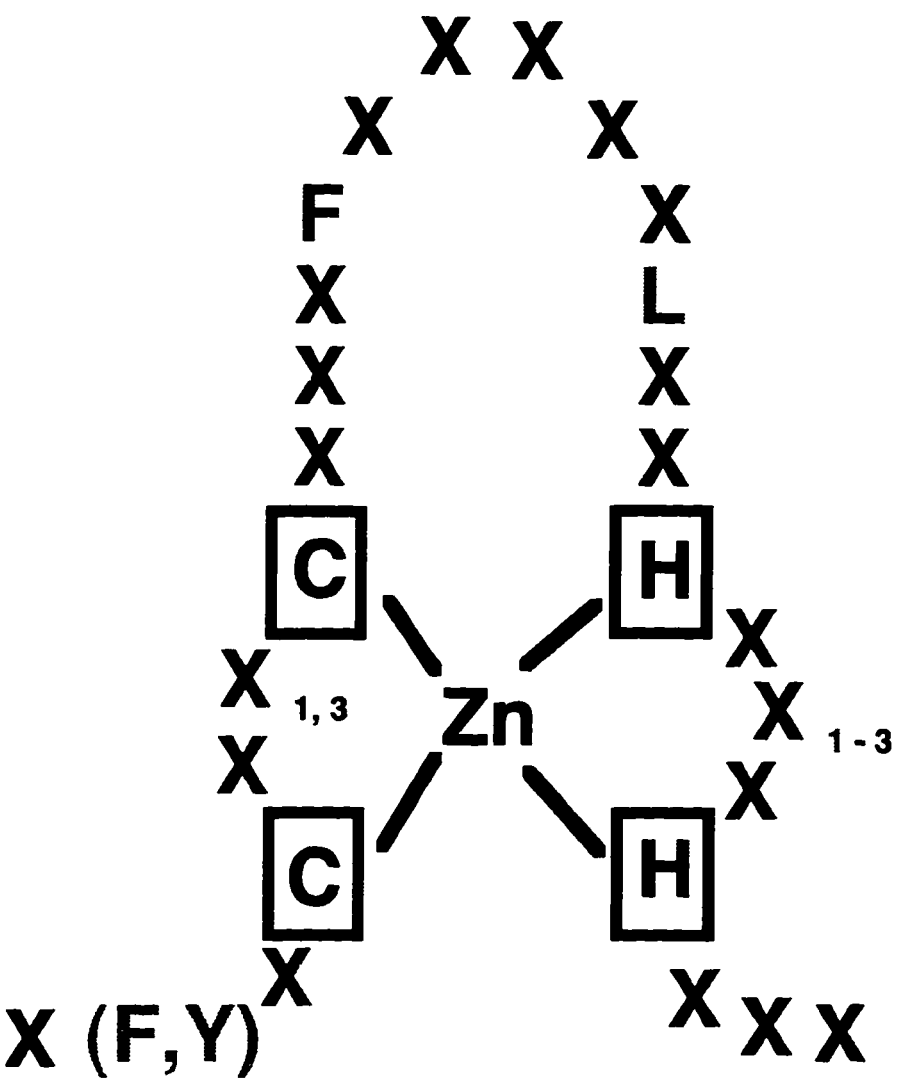
A. HISTORY

The DNA-binding domain of ADR1 contains two Cys_2-His_2 zinc fingers. This type of zinc finger is the most common DNA-binding motif found in eukaryotic cells. The motif was first recognized in a *Xenopus* transcription factor, TFIIIA, in the mid-1980s (1, 2). Klug and coworkers demonstrated that limited proteolysis of TFIIIA yielded stable fragments approximately thirty residues in length which seemed to be present in multiple copies (1). Inspection of the sequence revealed nine tandem repeats of a motif originally known as the TFIIIA-like zinc finger.

This motif is also referred to as the Cys_2-His_2 zinc finger, after the characteristic pattern of zinc ligands in the consensus sequence (see Figure 1). Well before this consensus was recognized it had been shown that TFIIIA binds zinc (3). Klug and coworkers

Figure 1. Consensus Cys_2-His_2 Zinc Finger Sequence.

Zinc ligands are shown in green, conserved hydrophobic residues in orange. Variable spacing between cysteines and histidines is indicated by subscript.


X X
X X
F X
X L
X X
X X
C H
X X
X 1, 3 Zn X 1 - 3
X X
C H
X X
X (F,Y) X X X

found 7 - 11 equivalents of zinc in purified TFIIIA (1). It seemed plausible that each zinc finger sequence could bind one zinc ion, using the conserved cysteine and histidine sidechains present in the motif. Tetrahedral coordination of zinc was later demonstrated for zinc finger peptides by EXAFS (4), and circular dichroism was used to show that isolated zinc finger motifs could fold in the presence of zinc (5). Since the original discovery of the motif in TFIIIA, thousands of potential Cys_2-His_2 zinc fingers have been identified on the basis of homology to the consensus sequence (Y,F)-X-C-$X_{2,4}$-C-X_3-F-X_5-L-X_2-H-X_{3-5}-H-X_{2-6} (6, 7). Very few of these putative fingers have been characterized in terms of zinc or DNA binding, however, and wild-type structures have been determined for fewer than two dozen.

B. STRUCTURE

i. Single Fingers

Based on the similarity of the consensus Cys_2-His_2 zinc finger sequence to molecules of known structure, including rubredoxin and thermolysin, it was predicted that the structure of the zinc finger would consist of a short, two-stranded β-sheet packed against an α-helix, each element of secondary structure

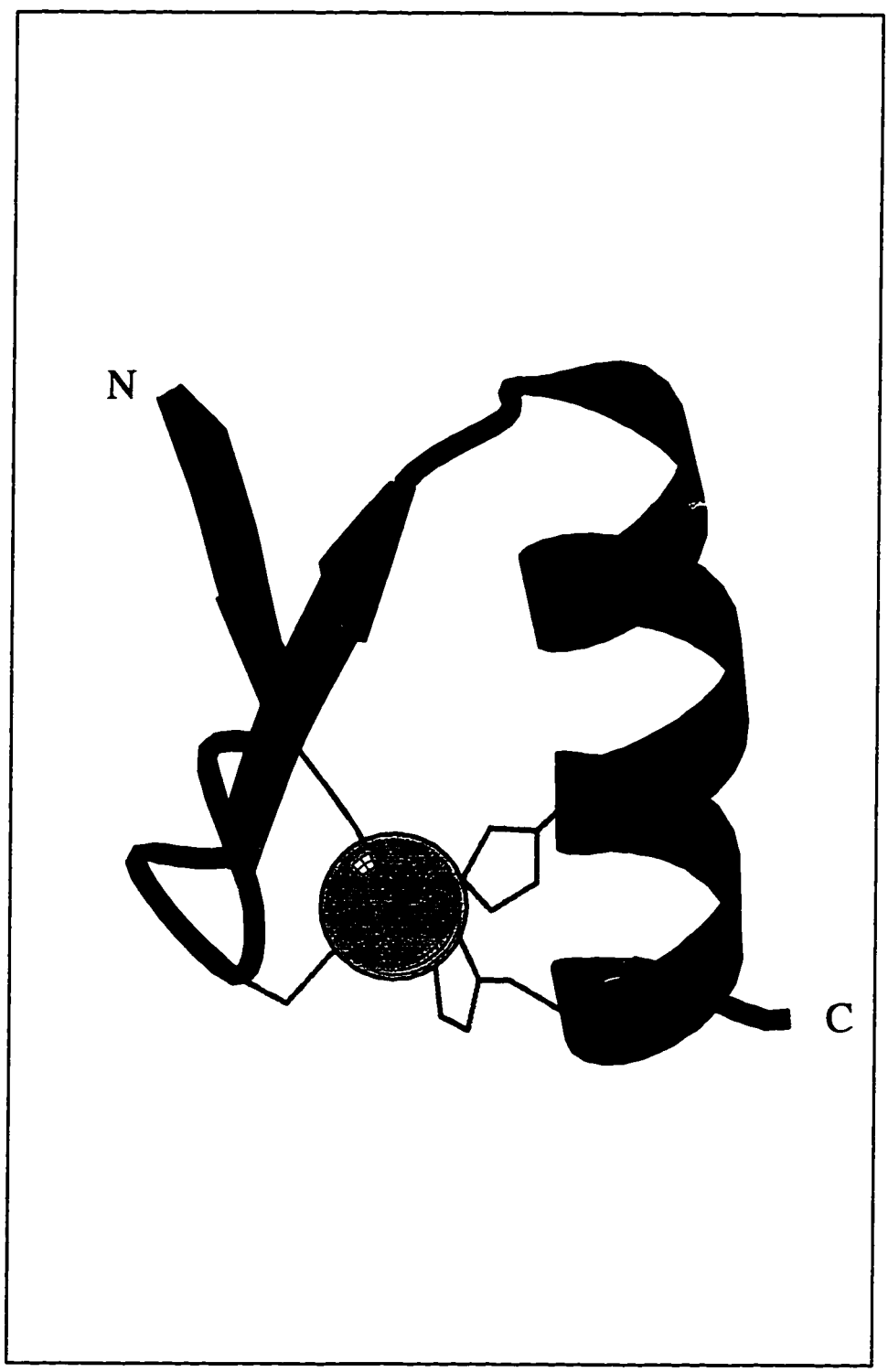
contributing two zinc ligands (8). This overall fold was in fact demonstrated for a single finger from a yeast transcriptional activator, ADR1, by Klevit and coworkers in 1988 (9).

Since that time, structures of more than two dozen wild-type and mutant zinc fingers have been solved by NMR and X-ray crystallography. Many have been solved as single finger peptides by solution NMR, as was the case for fingers from Xfin (10), ADR1 (11-13), MBP-1 (14), and ZFY (15, 16).

In each of these cases, the finger folds in the presence of zinc to what is now recognized as the canonical Cys_2-His_2 zinc finger structure (Figure 2) (reviewed in (17-19)). No sequence beyond the borders of the consensus is necessary for finger folding. The canonical fold involves a two-strand antiparallel β-sheet, which in some cases may be irregular, packed against a three-turn helix. Two conserved cysteine sidechains from the turn connecting the strands of the sheet and two conserved histidine sidechains from the C-terminus of the helix coordinate zinc tetrahedrally. The spacing of the histidine ligands is variable among zinc fingers, usually involving three or four intervening residues. In cases where only three residues intervene, a transition from α-helix to 3_{10} helix occurs; this may allow proper orientation of the N^{ε} of the histidines for tetrahedral metal coordination (10, 12, 20). The

Figure 2. Canonical Cys_2-His_2 Zinc Finger Fold.

The $\beta\beta\alpha$ fold of the second zinc finger of Zif268 is shown. The beta sheet region is also known as the Cys-Cys loop; the so-called "fingertip" connects the C-terminal strand of the beta sheet and the alpha helix. Zinc is coordinated tetrahedrally by two cysteines from the turn in the Cys-Cys loop and by two histidines from the helix. This figure and others were rendered using the graphics display program MOLSCRIPT (21). The coordinates of Zif268 were taken from the Brookhaven Protein Data Base, accession number 1zaa.


N
C

remaining conserved residues of the consensus form a small hydrophobic core that includes the histidine ligand imidazole rings.

ii. Double Fingers

While the single finger studies described above demonstrate that zinc fingers can fold as independent units, individual zinc fingers do not bind DNA specifically or with high affinity (5). Indeed, Cys_2-His_2 zinc finger motifs generally occur in nature in tandem arrays. While some proteins have been identified that contain only one or two zinc fingers, the large majority of arrays consist of at least three and up to 37 repeats (19). These repeats are connected by conserved sequences of four or five residues known as linkers.

Solution NMR structures of two-finger peptides from the yeast transcription factor SWI5 (22) and the human enhancer binding protein MBP-1 (23) have been published. In both cases each finger motif folds around a separate zinc ion, using as ligands the same conserved cysteine and histidine residues as do the single finger peptides. While the first finger of SWI5 has some unusual additional structure that will be discussed in detail later, in each peptide the consensus zinc finger sequences fold to the canonical zinc finger structure.

One question that arises for zinc fingers in tandem arrays is whether or not adjacent fingers adopt a unique orientation with respect to one another. Most DNA-binding proteins do not contain arrays of DNA binding motifs; instead, binding motifs are brought in proximity by protein oligomerization. Dimerization allows juxtaposition of two DNA binding motifs in an orientation appropriate for DNA recognition. It is possible that zinc fingers in tandem arrays could also adopt a fixed relative orientation useful for DNA binding.

The two fingers of the SWI5 peptide appear to be entirely structurally independent, however (24). NMR spectra were examined for chemical shift differences from single finger peptides and for alterations in NOE patterns that might indicate finger-finger contacts. No interaction between the fingers was detected.

The case of MBP-1 appears to be somewhat different. Evidence for limited interaction between the fingers was found in NOESY spectra in which a number of inter-finger constraints were identified for a lysine sidechain in the fingertip of the C-terminal finger (23). The NMR solution structure suggests a potential hydrogen bond involving this lysine and a threonine residue near the end of the helix of the N-terminal finger. Such an interaction might serve to orient the fingers with respect to one another for DNA binding. Given the counter-example of the non-interacting SWI5

fingers, however, this does not seem to be an absolute prerequisite for DNA binding.

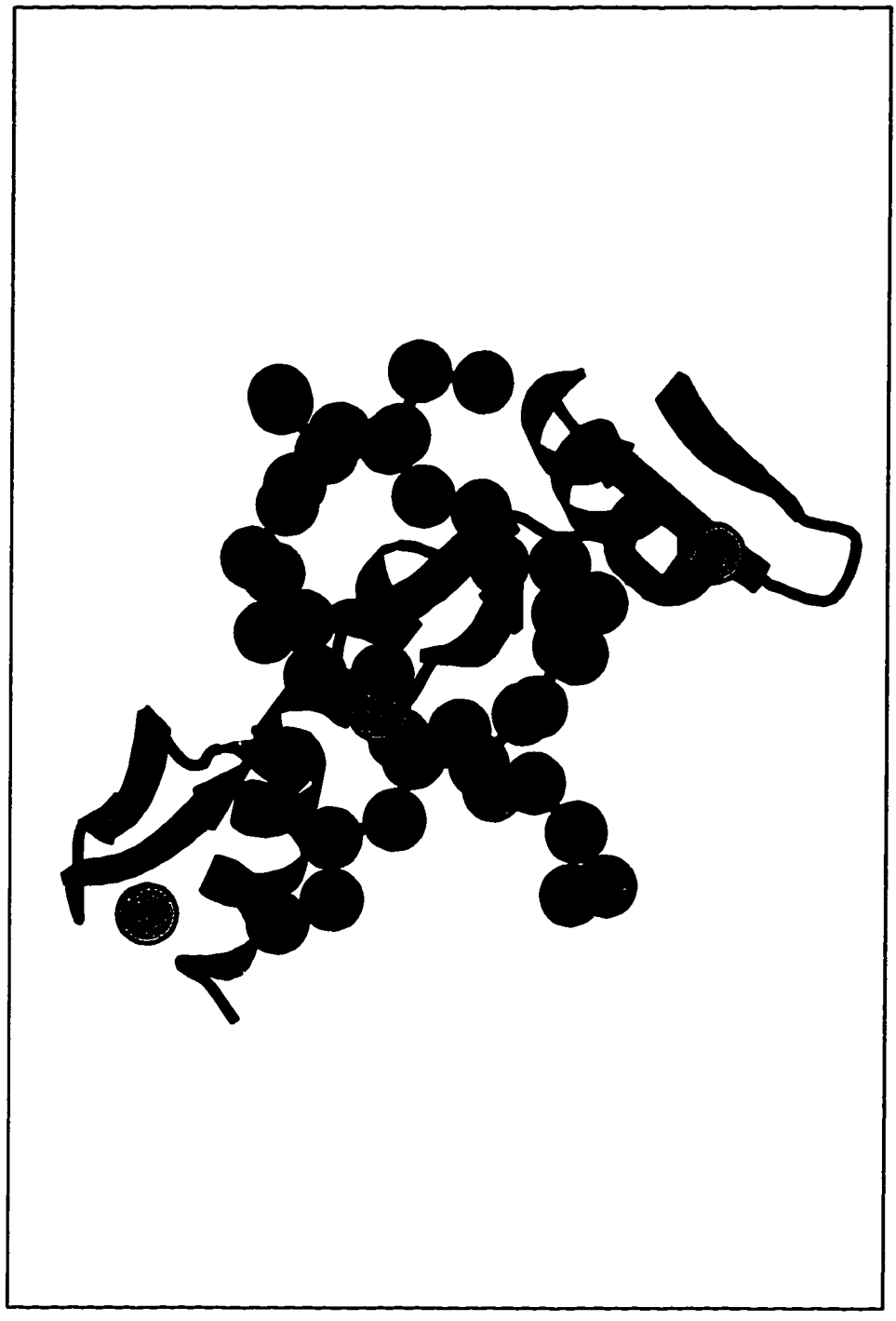
iii. X-ray Structures

The best information to date on zinc finger-DNA interactions comes from several X-ray crystallographic studies of multi-finger polypeptides bound to DNA. The first such study involved the three fingers of the mouse immediate early protein Zif268 cocrystallized with a 10-basepair specific DNA site (20, 25). The specifics of sidechain-base recognition will be addressed in a later section, but even in broad strokes the picture that the Zif268 structure reveals is striking.

Each of the three fingers of Zif binds the DNA in a similar orientation, contacting adjacent three-basepair subsites (20). The polypeptide chain winds around the DNA, following the course of the major groove (Figure 3). The orientation of the fingers with respect to the DNA is anti-parallel: the 5' end of the strand that receives the vast majority of contacts is bound by the C-terminal finger, while the 3' end is contacted by the N-terminal finger. In each case, the zinc finger contacts the major groove of the DNA with residues from the α-helix, making closest approach with the

Figure 3. Structure of Zif268 Zinc Fingers Bound to DNA.

The three zinc fingers of Zif268 bind in the major groove of DNA. Specific contacts arise from the finger helices. DNA strands are drawn in blue and cyan, the protein backbone in magenta.



N-terminal end of the helix. The Cys-Cys loop lies on the back of the helix, oriented away from the DNA.

X-ray structures of the Zif268 fingers bound to DNA are quite similar to structures determined for other, single fingers free in solution. One comparison has shown a mainchain RMSD of only 0.6 Å between the second finger of Zif and the single finger structure of the N-terminal finger of ADR1 (13).

The DNA-bound fingers appear to be modular in nature, in that each finger contacts a separate subsite and there is only limited interaction between the fingers. For each pair of neighboring fingers a single significant inter-finger contact is observed, at a similar position to that found in MBP-1 (20). These contacts involve a hydrogen bond between a backbone carbonyl from the fingertip of the more C-terminal finger of each pair and the sidechain of an arginine near the end of the helix of the N-terminal finger of the pair.

The cocrystal structure of the two fingers of *Drosophila* Tramtrack reiterates many of the patterns established for the three fingers of Zif (26). The helices lie in close contact to successive subsites in the major groove of the DNA. The two fingers again bind in a similar orientation, albeit a somewhat different orientation than that adopted by the fingers of Zif. In the

Zif structure, the linkers had been found to be ordered; in the case of Tramtrack, the linker residues were found to be fairly disordered, with high temperature factors (26).

Mutations in the highly conserved zinc finger linkers can alter DNA binding affinity (27). No direct contacts between DNA and the linkers were observed in the original Zif268 cocrystal structure, however, or in later cocrystal structures (20, 25, 26, 28, 29). It was thus proposed that the linkers might influence DNA binding through effects on the precise orientation of the fingers with respect to each other and the DNA (20). Recently, water-mediated linker-DNA contacts were detected in the Zif268 cocrystal structure solved to 1.6 Å (25). These contacts involve the conserved lysine sidechains of each linker and phosphate groups from the less-contacted DNA strand. Because no other Cys_2-His_2 zinc finger cocrystal structure has been solved to this level of resolution, it is impossible to know whether water-mediated linker-DNA contacts will be a general rule. However, the disorder of the linker in the Tramtrack cocrystal structure argues that no such contacts are present in that case; thus it seems unlikely that these interactions are essential for the orientation of the fingers or for DNA binding.

Recently, a third cocrystal structure that reiterates the lessons of Zif268 was published, for a three-finger construct

based on a consensus zinc finger sequence with a designed DNA specificity (29). Each of the general features of the Zif268 paradigm is present in this structure. The fingers have similar folds; they bind in similar orientations, using the same region of sequence to contact DNA; they contact one strand of DNA heavily, making contacts to adjacent three-basepair subsites; the binding to the primary contact strand is anti-parallel.

The conserved nature of the Cys_2-His_2 zinc finger structure, the consistent way in which the fingers interact with DNA, and the modular nature of the interactions suggested that zinc fingers might offer an optimal framework for the design of novel DNA-binding proteins. It was hoped that zinc fingers could be easily manipulated by retooling specificity and permuting arrays, to design proteins able to recognize desired DNA sequences. It was also hoped that the behavior of uncharacterized zinc fingers could be predicted based on the regularity of the early observations.

iv. Exceptions to the Canonical Structure

Life, of course, is not that easy. Soon after the perfectly simple and regular Zif268 structure was published, a number of other structures were solved that muddied the waters considerably.

First, the solution structure of the N-terminal zinc finger of SWI5 was determined (22). It was shown that this finger did not fold to a stable structure when only the consensus Cys_2-His_2 finger sequence was present (24). Approximately five additional residues N-terminal to the finger were required for folding. The three-dimensional structure showed that these residues fold to form a third strand to the β-sheet of the canonical structure (22). A similar fold was also found for the first finger of the Tramtrack protein (26). The additional strand does not contact the DNA in this cocrystal structure, but it is required for DNA binding, presumably because it stabilizes the structure of the finger itself.

Recently, it has been shown that residues N-terminal to the third, non-canonical strand of SWI5 finger 1 also increase DNA binding affinity. These residues fold to a helical structure that possibly makes direct contacts to DNA (30). The SWI5 and Tramtrack examples demonstrate that Cys_2-His_2 zinc fingers do not all fold to the same canonical structure, and that regions outside the consensus may participate, or even be required, in DNA binding.

Even when zinc fingers do fold to the canonical structure, they may not always interact with DNA in the expected manner. This was demonstrated graphically by the crystal structure of the five fingers of GLI bound to DNA (28). In this case, four of five

fingers bind DNA in a fairly typical orientation, but only the two C-terminal fingers make the expected extensive specific contacts with the DNA. Most surprisingly, the first zinc finger does not contact the DNA at all, interacting instead with the second finger. The observations in this complex may not be easily generalized to other systems, since the binding site used was selected *in vitro* from genomic DNA; while it binds GLI with high affinity, it is not known whether it is the optimal site or biologically relevant. However, this case demonstrates that a wide range of possible finger-DNA and finger-finger interactions are plausible.

The variability demonstrated by the Cys_2-His_2 zinc fingers is also seen for other DNA binding domains. For example, the conserved helix-turn-helix structural motif is sometimes augmented by accessory structure and sequences that bind DNA. The lambda repressor uses an additional extended arm to bind the major groove of DNA (31); the winged helix-turn-helix motif utilizes additional sidechains from neighboring loops (32-34); the homeodomains, eukaryotic analogues of the helix-turn-helix, bind the minor groove of DNA with an additional N-terminal arm (35). The RXR nuclear receptor DBD involves an additional helix beyond the conserved nuclear receptor fold (36), and the erythroid transcription factor GATA-1 binds DNA with a domain that is structurally homologous to the N-terminal portion of the nuclear

receptor DBDs, but that also includes a C-terminal arm which binds in the minor groove on the opposite face of the DNA (37).

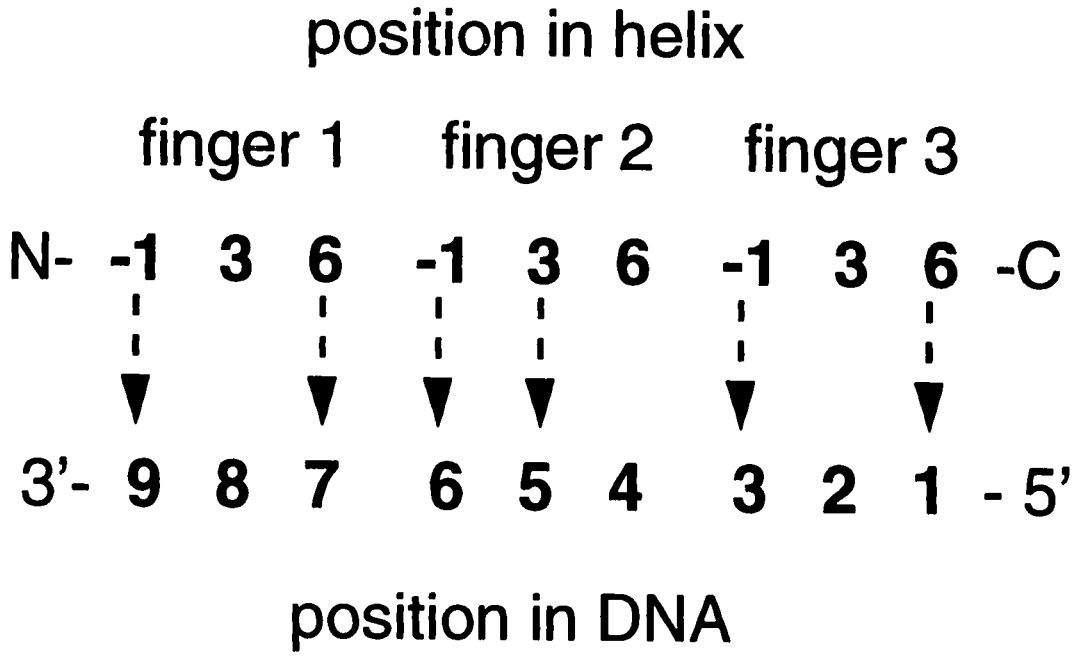
The variability of the zinc finger-based DNA binding domains is therefore not unique. It does complicate the prediction of structure and behavior of uncharacterized fingers, however. Further complications arise when considering the nature of the specific contacts made by the fingers to the bases of DNA.

C. DNA RECOGNITION

The Zif268 cocrystal structure revealed a striking pattern of specific sidechain-base contacts which suggested there might be general rules for zinc finger-DNA recognition (see Figure 4) (20). The three fingers contact adjacent three-base pair subsites along a single strand of the DNA using residues just prior to and in the alpha helix. The residues that make base-specific contacts are located at positions -1, 3, and 6 of the helix; when used, the -1 position contacts the 3' base of the subsite, position 3 contacts the central base of the subsite, and position 6 contacts the 5' base of the subsite. Each finger makes two such contacts to guanines in the subsites, using arginine and histidine sidechains to hydrogen bond to the edges of the bases (20).

Figure 4. Specific Zinc Finger-DNA Contacts.

Specific zinc finger-DNA contacts identified in the Zif268 cocrystal structure are shown (20). These contacts arise from the -1, 3, and 6 positions of the alpha helices. Neighboring zinc fingers contact adjacent three-basepair subsites in an antiparallel manner.


position in helix
finger 1
finger 2
finger 3
N-
-1
3
6
-1
3
6
-1
3
6
-C
3'-
9
8
7
6
5
4
3
2
1
- 5'
position in DNA

A number of mutational studies have since been performed to test the relevance of the -1, 3, and 6 positions of the helix to the binding affinity and specificity of various zinc fingers (reviewed in (38)). These studies tend to bear out the importance of those positions, lending credence to the thought that novel specificities might be generated simply by altering this small handful of residues.

A simple set of rules for zinc finger-DNA recognition is still well beyond our grasp, however. As more mutagenesis has been done and new cocrystal structures have been solved, it has become clear that the repertoire of finger contact positions, DNA subsites, and residue-basepair matches is extensive. In the Tramtrack cocrystal structure, specific contacts from helical position 2 were observed in addition to those from canonical base-contacting positions -1, 3, and 6 (26). One of these contacts was made to the opposite strand of the DNA from that normally contacted, overlapping a neighboring subsite. Similar contacts from position 2 have since been observed in several other crystal structures, and may represent a fourth canonical contact type (25, 28, 29).

For the two fingers of GLI that contact DNA specifically, the situation is even more complicated. The subsite for one finger overlaps that of another finger, numerous base contacts to both

strands are made, and base contacts arise from positions -1, 1, 2, 3, 5, and 6 of the fingers (28). Finally, in addition to the common arginine-guanine match, the following contacts have been observed in cocrystals: histidine with adenine or guanine, lysine with guanine, asparagine with adenine, aspartate with cytosine, glutamine with adenine, and serine with guanine, adenine or thymidine (20, 28, 29, 39).

An even wider repertoire of contacts has been generated by mutagenesis, site-selection, and phage-display studies (reviewed in (40)). Sidechain-base preferences seem to depend on the position of the residue in the finger, and also on the neighboring context in the finger. Attempts to alter zinc finger specificity in a directed manner have been met with mixed results (41-43). It is therefore unlikely that any set of simple rules will be useful to predict what type of DNA sequence a particular finger will prefer.

Variations thus exist in zinc finger structure, orientation with respect to DNA, positions of base-contacting residues, base-sidechain recognition pairs, subsites and strand of contact, and the use of accessory sequences to augment zinc finger DNA binding. These variations necessitate individual characterization of zinc fingers of interest. Many more zinc fingers must be studied before it will be apparent whether a set of reliable rules can be formulated for describing zinc finger behavior.

III. Introduction to ADR1

A. HISTORY

ADR1 is a transcriptional activator from the yeast *S. cerevisiae*. It was first identified as a positive regulator of expression of the glucose-repressible alcohol dehydrogenase, ADHII (44). ADR1 has also been shown to regulate expression of several genes involved in peroxisomal biogenesis and production of peroxisomal proteins (45). ADR1 acts through a 22 base pair palindromic enhancer sequence called UAS1 which is found upstream of these genes (Figure 5) (46).

The presence of two Cys_2-His_2 zinc finger motifs in ADR1 was noted shortly after the publication of Klug's original zinc finger hypothesis (47). The sequence of these fingers is shown in Figure 6, with the consensus elements highlighted. It was later confirmed that ADR1 binds to UAS1 in a zinc-dependent manner (48).

Intensive mutagenesis has been performed to identify the functions of various regions of ADR1 (49-51). The zinc fingers fall in a region identified early on as the DNA-binding domain (47). While the entire protein contains over 1300 residues, the DNA

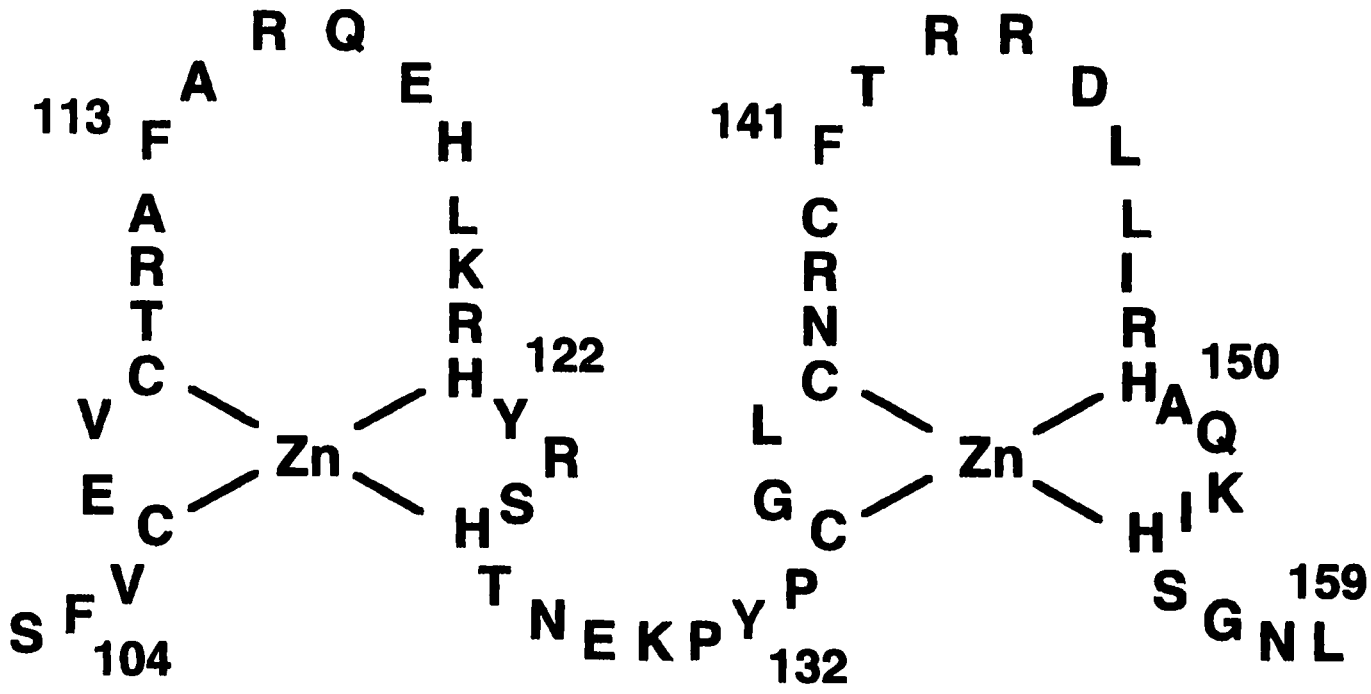
Figure 5. The UAS1 Sequence.

ADR1 binds specifically to a 22 basepair palindrome, UAS1, to activate expression of *ADH2* and several peroxisomal genes.


TCTCCAACTTA TAAGTTGGAGA
AGAGGTTGAAT ATTCAACCTCT

Figure 6. ADR1 Zinc Fingers.

The ADR1 sequence includes two tandem Cys_2-His_2 zinc finger motifs, drawn here with conserved residues highlighted as in Figure 1.


R Q
A
E
113 F
H
A
L
R
K
T
R
C
H
122
V
Y
Zn
R
E
S
C
H
V
T
S F
N E K P Y
104
132
R R
T
D
141 F
L
C
L
R
I
N
R
C
150
H A
L
Q
Zn
G
I K
C
H
P
159
S
G
N L

binding domain has been localized to 143 residues following the 16-residue nuclear localization signal (Figure 7). Hydroxylamine-induced mutations that completely obliterate ADR1 activity *in vivo* clustered in the zinc finger sequences, at the conserved cysteine and histidine ligands (52). These mutations demonstrated that both fingers of ADR1 are essential for its function.

B. MUTAGENESIS AND DNA RECOGNITION

Since the time of the initial identification of Cys_2-His_2 fingers in ADR1, the entire DNA binding domain has been characterized particularly thoroughly. Deletion mutagenesis has been used to define the minimal region of ADR1 that is sufficient to bind DNA. Full DNA binding affinity is found for a fragment that extends from residue 17 to residue 160 (53). This sequence encompasses the two repeats of the zinc finger motif, from residue 104 to residue 159, as well as a region N-terminal to the fingers that has no strong similarity to known DNA binding motifs (Figure 8). C-terminal deletion past residue 160 to residue 155 impinges on one of the zinc finger sequences, eliminating the final histidine zinc ligand. This deletion obliterates ADR1 activity *in vivo* and DNA binding *in vitro* (53).

Figure 7. Functional Regions of ADR1.

Functional regions of ADR1 have been delineated by mutagenesis. DNA-binding activity is localized to a region of fewer than 150 residues at the N-terminus of the 1323-residue ADR1 protein. This region contains the two zinc finger repeats, at residues 104 to 128 and 132 to 157. Random mutations found to diminish ADR1 activity *in vivo* are indicated by filled circles; a majority fall at the zinc ligands of the fingers. Data are from (49-53).

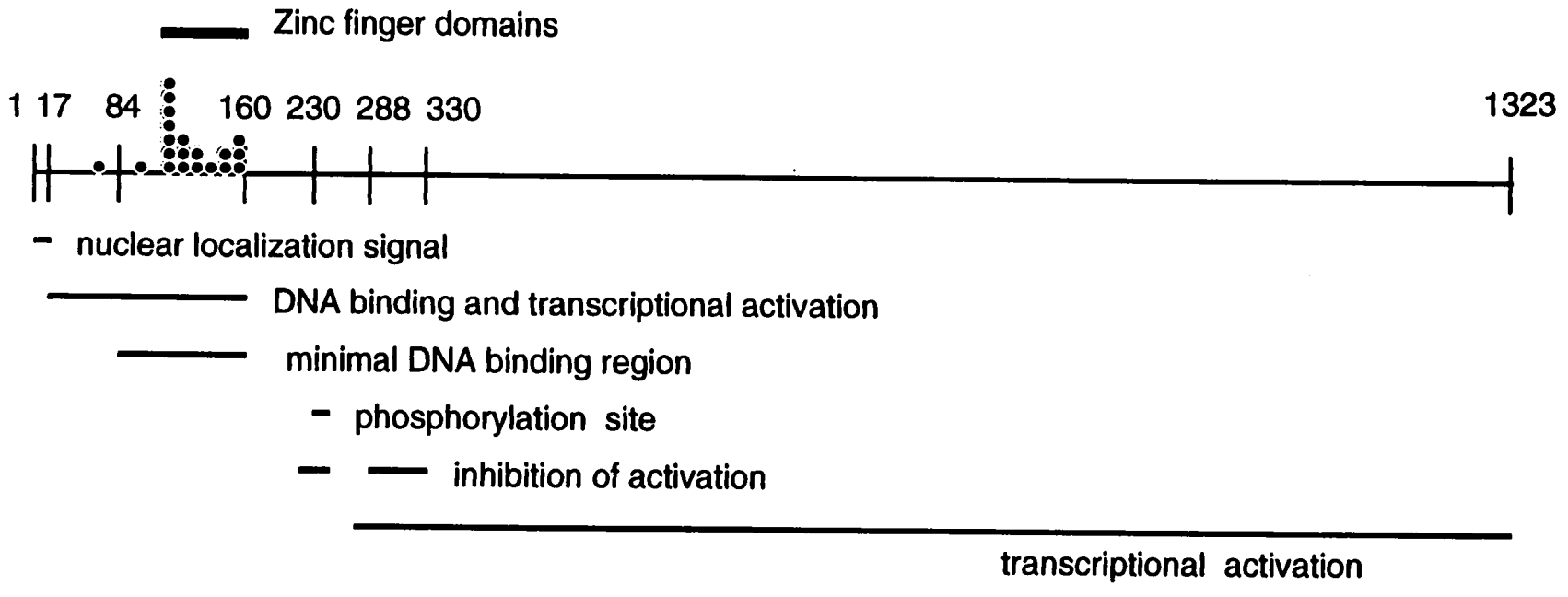
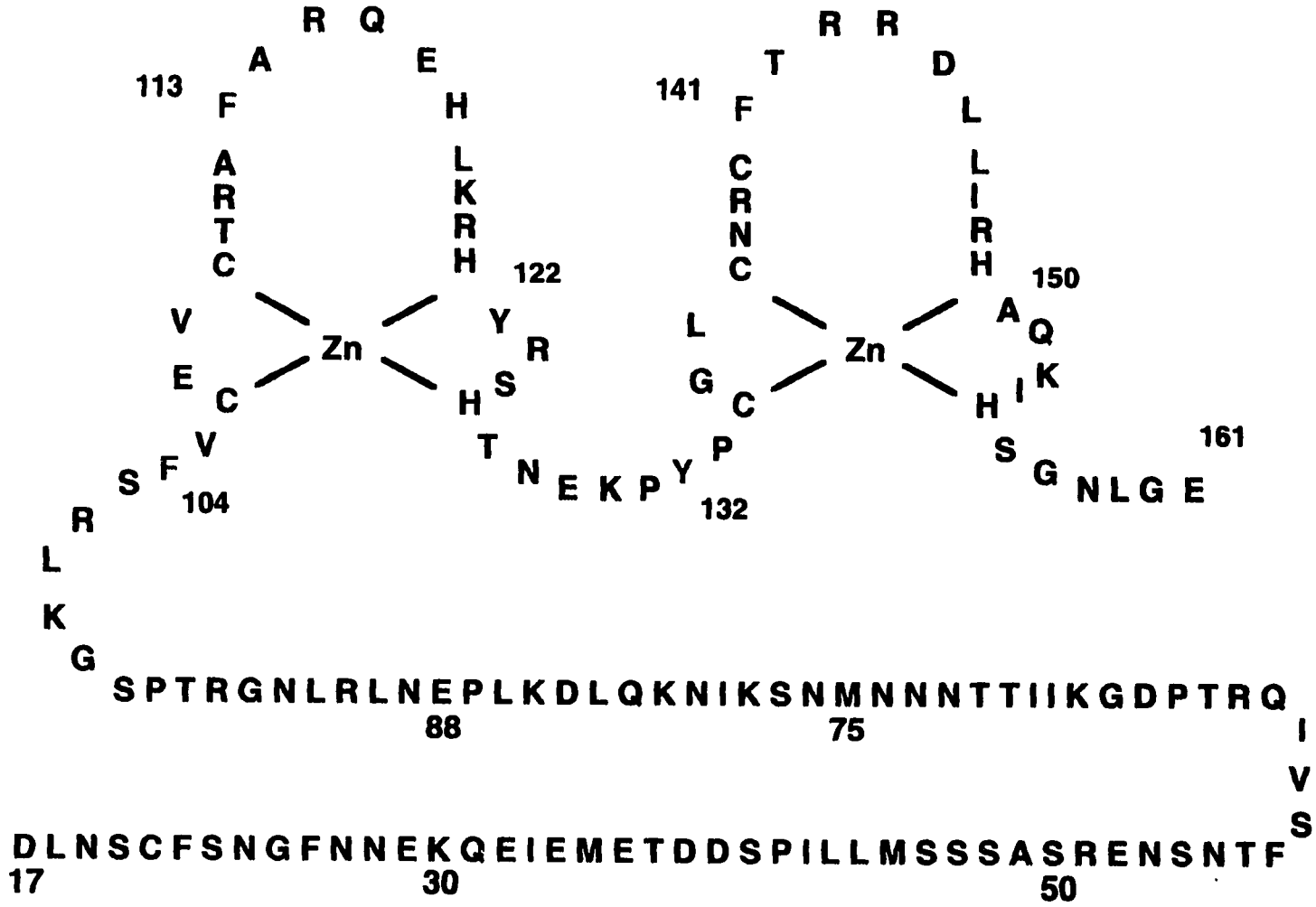

Zinc finger domains
1 17 84 160 230 288 330
1323
nuclear localization signal
DNA binding and transcriptional activation
minimal DNA binding region
phosphorylation site
inhibition of activation
transcriptional activation

Figure 8. The DNA-Binding Domain of ADR1.

The DNA-binding domain of ADR1 extends from residue 17 to residue 160. This region encompasses two zinc fingers, at residues 104 to 128 and 132 to 157, and an additional N-terminal region.


R Q
A E
113
F H
A L
R K
T R
C H
122
V Y
Zn R
E S
C H
V T
S F N
104
E K P Y
132
R R
T D
141
F L
C L
R I
N R
C H
150
L A
Zn Q
G K
C I
P H
S
161
G N L G E
R
L
K
G
S P T R G N L R L N E P L K D L Q K N I K S N M N N N T T I I K G D P T R Q
88
75
I
V
S
D L N S C F S N G F N N E K Q E I E M E T D D S P I L L M S S S A S R E N S N T F
17
30
50

More surprisingly, residues from the region N-terminal to the zinc fingers are also essential for DNA binding. A minor reduction in DNA binding activity is observed for truncation of the N-terminus to residue 84. ADR1 activity *in vivo* is reduced less than ten-fold by this mutation. Further truncation to residue 88 eliminates detectable DNA binding, however, and obliterates ADR1 activity *in vivo* (53).

Site-directed and random mutagenesis have been used to further dissect the DNA binding domain (54, 55). Many of the mutations that diminish or eliminate binding fall in the fingers themselves, including especially mutations at residues analogous to the primary DNA-contacting positions in known cocrystal structures. These residues include R115, H118, and R121, at the -1, 3, and 6 positions of the N-terminal finger, and R143, at the -1 position of the C-terminal finger. Mutations at R115, H118, and R121 have also been identified that change the DNA binding specificity of ADR1 (56, 57). This suggests that these residues make base-specific contacts to UAS1.

Several mutations in the region N-terminal to the fingers also diminish DNA binding, including the mutations P87L, P97L, and TT70,71II (53). Further evidence of the importance of the N-terminal region came from a search for mutations that suppress

defects in DNA binding. This effort revealed a mutation in the N-terminus, R91K, that was able to restore DNA binding activity to defective constructs bearing deleterious mutations in finger 1 or in finger 2 (58). The R91K suppressor mutation is not only able to compensate for a number of different mutations in different positions, but the mutation in the wild-type context augments normal DNA binding by three-fold. Mutation of R91 to a glycine causes a slight defect in DNA binding (58).

The UAS1 DNA sequence has also been extensively mutated to define which bases are necessary for sequence-specific binding of ADR1 (59). The region identified includes several bases from the center of each inverted repeat: 3'-AGAGGTT-5' (see Figure 5). Change-of-specificity studies in which mutations in the finger sequence are compensated for by complementary mutations in the DNA sequence revealed that finger one contacts the sequence 3'-AGAGGTT-5'. Residue R143 of the second finger is expected to contact 3'-AGAGGTT-5' (56). These contacts are exactly in line with expectations from cocrystal structures, in which the protein-DNA interaction is antiparallel, and neighboring fingers make contacts to adjacent three-base pair subsites (20, 26, 28).

No basepairs from the center of the palindromic UAS1 sequence are found to be essential for specificity (59). However, ethylation of two of the central basepairs of UAS1 interferes with

DNA binding, indicating the presence of phosphate contacts in this region (59). This interference with binding does not occur when only residues C-terminal to Met75 are present, indicating that the extreme N-terminus is directly or indirectly responsible for phosphate contacts in the center of UAS1.

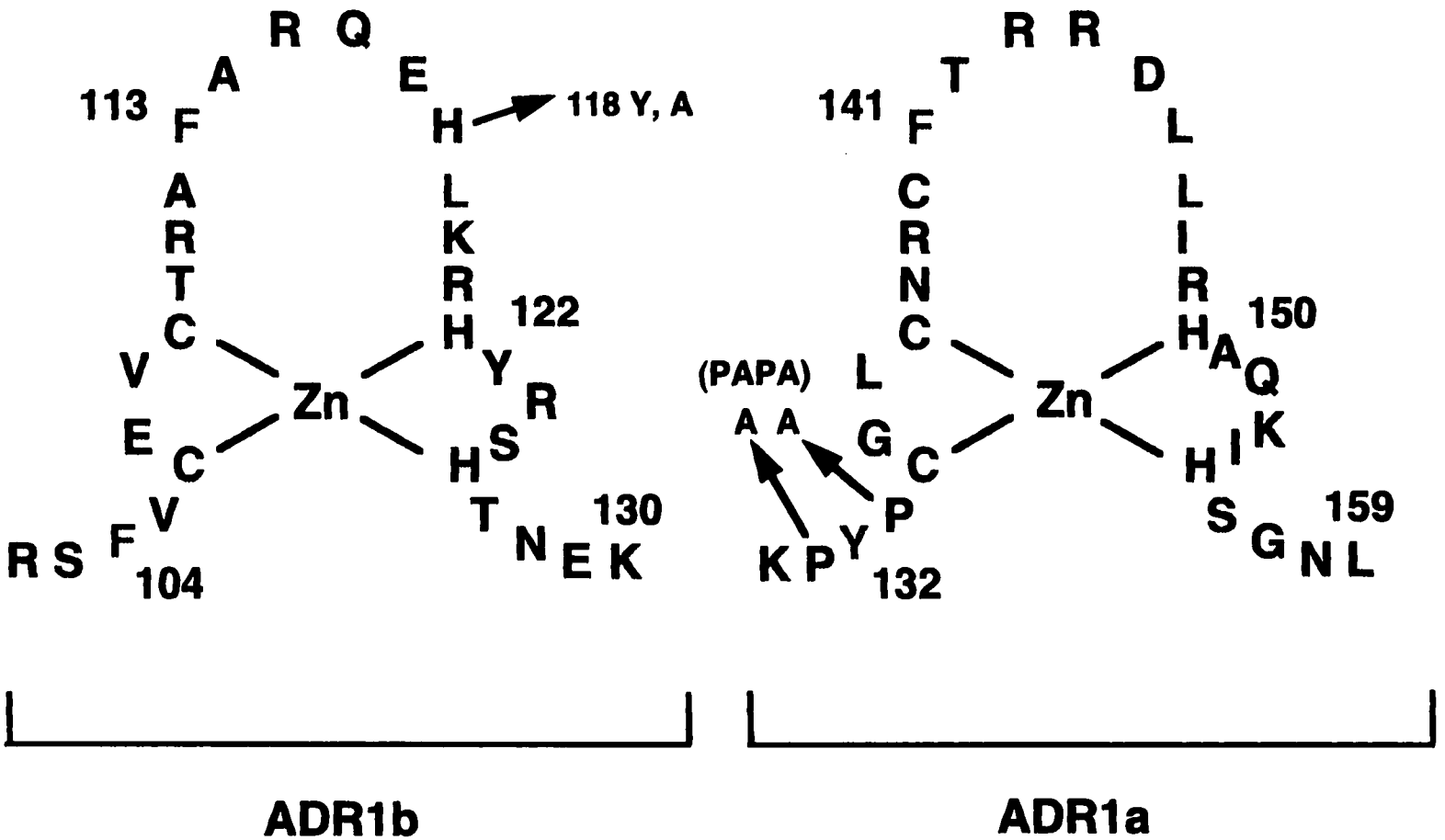
C. STRUCTURE

NMR studies have been conducted on several single zinc finger peptides from ADR1. The N-terminal finger has been studied in the form of a wild-type peptide (ADR1b) (11, 12) and two H118 mutants (B118Y and B118A) (12). The C-terminal finger has been studied as a wild-type peptide (ADR1a) (60) and a double mutant: P131A, P133A (PAPA) (13). These sequences are shown in Figure 9. Each of these single finger peptides folds to approximately the canonical zinc finger structure in a zinc-dependent manner.

Full structure determinations have been done for ADR1b (11, 12), its H118Y and H118A mutants (12), and for PAPA (13). These peptides each possess an N-terminal two-stranded β-sheet packed against a three-turn helix. The fingers are quite similar to each other, and to the DNA-bound fingers of Zif268, with mainchain RMSDs on the order of 0.64 Å (13). No structure has been calculated

Figure 9. ADR1b and ADR1a Sequences.

ADR1b and ADR1a are single zinc finger peptides corresponding to the N- and C-terminal fingers of ADR1, respectively. In addition to the wild-type peptides several mutants have been studied by NMR, including ADR1b H118Y and H118A and an ADR1a P131,133A (PAPA) double mutant.


R Q
A
E
113
F
118 Y, A
H
A
R
T
C
V
E
C
F
V
R S
104
Zn
L
K
R
H
122
H
Y
R
S
T
N
E K
130
(PAPA)
A A
K P
Y
132
L
G
C
P
R R
T
D
141
F
L
C
R
N
C
L
I
R
H
150
H
A
Q
K
I
S
G
N L
159
Zn
ADR1b
ADR1a

for the wild-type finger ADR1a, or for the N-terminal sequence also required for DNA binding.

The requirement for a region outside of the zinc fingers to bind DNA is unusual, but not without precedent. As discussed previously, The N-terminal fingers of SWI5 and Tramtrack both require several residues adjacent to the fingers for formation of a third strand to the β-sheet; the SWI5 finger fails to fold in the absence of these residues (24). Both fingers of ADR1, however, are competent to fold into the canonical zinc finger structure as isolated sequences (9, 11, 12, 60). Furthermore, the N-terminal region required in ADR1 is much more extensive than that required for a third strand of sheet, as the minimal domain encompasses twenty residues beyond the fingers (53).

One possible explanation for the necessity of the N-terminus is that it contacts DNA directly, augmenting the DNA-binding affinity of the protein. Most zinc finger proteins have at least three fingers, whereas two-finger examples such as ADR1 are quite rare. The presence of only two fingers in ADR1 is not enough to explain the requirement for the N-terminus, however. Tramtrack is also a two finger protein, and it binds DNA with high affinity without using contacts outside the canonical fingers (26). The difference between ADR1 and Tramtrack may lie in the number of

contacts made by each finger. In the Tramtrack crystal structure each finger contributes several base-specific contacts (26), while mutagenesis suggests that the second finger of ADR1 manages only one such contact (54). The zinc finger array of ADR1 thus appears to be less than optimal. The role of the N-terminus of ADR1 may be to provide DNA-binding affinity not supplied by the unusual finger 2.

Thus ADR1 presents some unique features, including the necessary N-terminus, the oddly underused second finger, and the presence of two proline residues in the second finger. Moreover, no other zinc finger system that is as well-characterized mutationally as ADR1 is is also well-characterized structurally. Structural insights on DNA binding will provide a rich opportunity to analyze the extensive functional data already available from mutagenesis. In order to fully understand the ADR1-DNA interaction, however, it will be necessary to study a construct containing both zinc fingers and the essential N-terminal region, in the presence and absence of DNA. No such NMR studies have yet appeared on any DNA-bound zinc finger.

IV. Introduction to NMR Methods

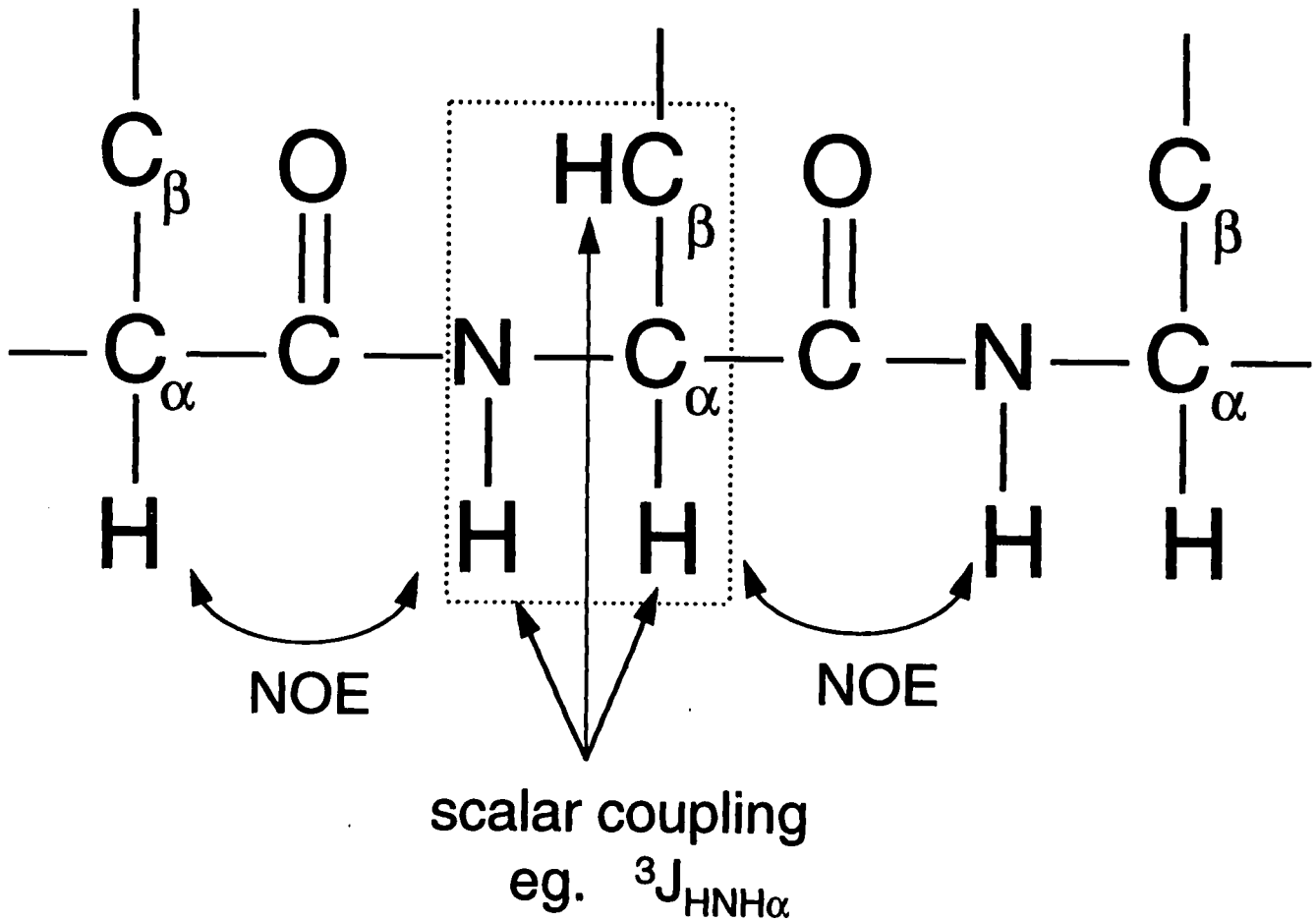
A. HISTORY

To this point, structural studies on the DNA binding region of ADR1 have been limited to homonuclear ^1H-^1H studies of synthetic single zinc finger peptides. This situation reflects the state of NMR spectroscopy during the 1980's.

During those years a revolution in the use of NMR for structural characterization of macromolecules occurred. The development of spectra resolved into two dimensions allowed the detailed study of relatively large molecules of less than about 10,000 daltons for the first time. These spectra provided the resolution necessary to interpret signals from hundreds of protons per molecule that resonate in the narrow ^1H chemical shift range of about 0 to 12 ppm. Two-dimensional techniques were introduced that correlate signals based on scalar couplings (COSY, TOCSY) or based on dipolar interactions (NOESY) (see (61)). The former provide information on interactions that occur through bonds, allowing identification of networks of protons that affect one another by virtue of bonding interactions. An example of such a network is an amino acid residue in a protein (see Figure 10). Assigning all proton resonances in a protein to individual amino

Figure 10. Amino Acid Spin Systems.

Potential routes for magnetization transfer within and between amino acid spin systems are shown. Magnetization may be transferred via scalar couplings within amino acid residues in experiments such as the COSY and TOCSY. These through-bond ^{1}H-^{1}H interactions do not extend beyond the carbonyl groups that flank the spin system. In contrast, dipolar magnetization transfer in NOESY experiments occurs through space, both within and between spin systems.


Cβ
O
HCβ
O
Cβ
Cα
C
N
Cα
C
N
Cα
H
H
H
H
H
NOE
NOE
scalar coupling
eg. 3JHNHα

acid spin systems and classifying those spin systems by type is the first step in determining the NMR structure of a peptide (61).

Subsequent steps rely on the second type of two-dimensional spectrum mentioned above, the NOESY. This spectrum provides information on interactions that occur through space. Examples of this type of interaction are local contacts formed when a protein assumes a regular secondary structure. Other examples are long-range, tertiary contacts that form when a protein folds into a globular structure, bringing distant parts of the primary sequence into proximity (Figure 11). Identification of the local sort of contact allows spin systems identified in scalar experiments to be organized in sequential order. This assignment of sequential NOE connectivities is the second step in NMR structure calculation (61).

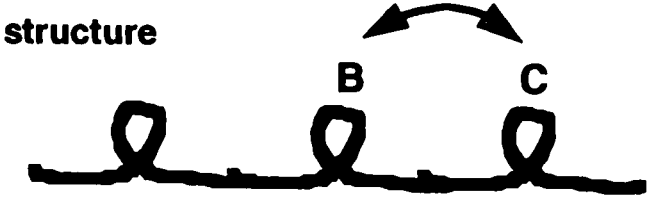
The final step of the calculation uses these sequential NOEs and also long-range, tertiary NOEs to generate a picture of the three-dimensional fold of a molecule. Because an NOE depends on the separation of two protons with a $1/r^6$ dependence, the intensity of a NOESY crosspeak is a measure of proximity of two protons. Large numbers of NOE-based pairwise distances are used to generate plausible solutions to the three-dimensional structure; the result of the calculation is a family of related structures.

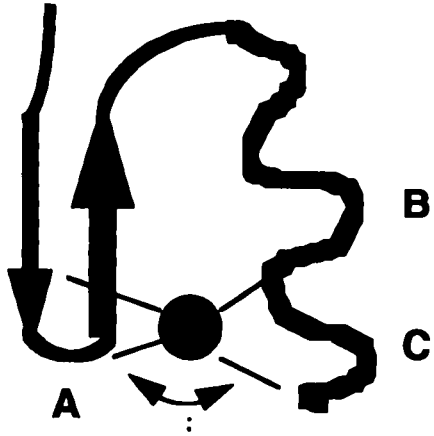
Figure 11. Local and Tertiary Contacts.

NOESY experiments depend on through-space magnetization transfer in a distance-dependent manner. NOEs are thus indicators for proximity of protons, either in local secondary structure or in long-range contacts due to folding of distant parts of the primary sequence into tertiary structure.

primary structure

local contacts in secondary structure





long-range contacts in tertiary structure

There are a number of limitations to the application of this valuable technique, however, that have practical consequences for the study of macromolecules. First, as the size of a macromolecule increases, so does overlap in its two-dimensional spectra. Overlap of resonances in COSY and TOCSY spectra can preclude complete and unambiguous assignment of protons to known positions in the sequence. Overlap in NOESY spectra can also prevent unambiguous assignment of the protons that originate NOE crosspeaks and can cause loss of distance information for those peaks that are superimposed. Increases in macromolecular size also lead to slower tumbling of the molecules in solution and thus broader linewidths. Because the homonuclear COSY and TOCSY experiments rely on very small ^{1}H-^{1}H couplings, these experiments become less efficient as linewidth increases.

Any NMR study of the DNA binding domain of ADR1 runs up against these limitations. The minimal polypeptide is on the order of 10 kD, at the limit of polypeptide size amenable to study by simple homonuclear 2D NMR methods. A minimal polypeptide-DNA complex will be about 20 kD, well beyond the limits of the methods of the 1980's. Furthermore, the minimal binding domain is longer than eighty residues, making the synthesis of this polypeptide a challenging and expensive project.

B. NEW STRATEGIES

Fortunately, a second revolution in NMR methodology has occurred in the last decade. New techniques are available that render feasible the study of macromolecules up to about 35 kD (62, 63) and offer an answer to the problem of studying the DNA binding domain of ADR1. The solution is to incorporate the NMR-active isotopes ^{15}N and ^{13}C into the protein and take advantage of a plethora of recently conceived heteronuclear experiments. These strategies offer a number of advantages over the 2D homonuclear proton methods of the past. First, the heteronuclear experiments relieve overlap by resolving resonances by a second and sometimes a third chemical shift. Extra resolution is also obtained by running the experiments in three or even four dimensions. The issue of broad linewidths is addressed by the use of large single bond heteronuclear couplings for transfers of magnetization, instead of small proton-proton couplings.

Beyond providing practical solutions to pressing size-related technical problems, heteronuclear NMR experiments offer a number of additional advantages. For example, if isotopic labeling of specific amino acids can be achieved, spectra can be simplified dramatically. Further, uniform incorporation of heteronuclei allows connectivities between amino acids to be established using

through-bond information, by following magnetization directly along the backbone from ^{13}C to ^{15}N to ^{13}C. This strategy avoids the use of more ambiguous proton NOE information to determine sequential assignments. Finally, isotopic labeling of only one component of a bimolecular complex can be exploited to simplify spectra, allowing selective observation of either the labeled or the unlabeled component, or the exclusive detection of intermolecular interactions (62).

C. PROTEIN-DNA COMPLEXES

All of these advantages have been exploited in recent NMR structural studies of protein-DNA complexes, offering hope for the study of an ADR1 DNA-binding domain-DNA complex. In earlier years, only a small handful of protein-DNA complexes had been studied by NMR. Among the first to receive a thorough treatment were the relatively small lac repressor headpiece (64), engrailed homeodomain (65), and nucleocapsid-type zinc finger (66). Initial studies focused on calculation of the structure of the proteins using data from homonuclear proton-based experiments. Only a handful of intermolecular NOEs were identified in each case, often for well-resolved aromatic and methyl-containing residues. Protein-DNA models were generated by docking the protein to regular DNA according to the intermolecular NOEs.

The first protein-DNA complex to be fully calculated based on numerous intra-protein, intra-DNA, and intermolecular constraints was the GATA-1 system published in 1993 (37). This effort made full use of the arsenal of heteronuclear techniques for the first time, and resulted in the identification of over one hundred protein-DNA NOEs. Since that time this method has been applied to the NMR structure determination of a number of protein-DNA complexes, including those of the *Antennapedia* homeodomain (67), LEF-1 (68), SRY (69), Myb (70), and ETS1 (71). No NMR study of a Cys_2-His_2 zinc finger-DNA complex has yet been reported, but the topic is ripe for study. The ADR1 DNA-binding domain is a particularly interesting subject for such a study, due to the numerous unusual features of the system.

V. NMR Structural Studies of the ADR1 DNA Binding Domain: Strategy and Rationale

The only cost-effective way to take advantage of the recent advances in heteronuclear NMR is to isotopically label proteins metabolically in bacterial overexpression systems. For this reason, the strategy for NMR studies of the ADR1 minimal DNA-binding domain begins with overexpression of the domain in *E. coli*, followed by purification. In order to minimize cost, overexpression and purification yields must be as high as possible.

Heteronuclear techniques will be used to assign the resonances of the expressed polypeptide both free and bound to DNA. Such assignments are a prerequisite to exploration of the structure and behavior of the zinc fingers and N-terminus in the context of the entire minimal DNA binding domain and to comparisons of the structure and behavior of these regions free and bound to DNA.

The preliminary study described in this thesis does not include full calculations of the ADR1 DNA binding domain structure. However, a qualitative picture of ADR1 DBD structure free and bound to DNA is presented, based on measurements of NOEs, coupling constants, and chemical shifts. These conventional methods each benefit from the added power imparted by the utilization of heteronuclear components.

CHAPTER 1. DEVELOPMENT OF SYSTEM

I. Subcloning and Overexpression Construct

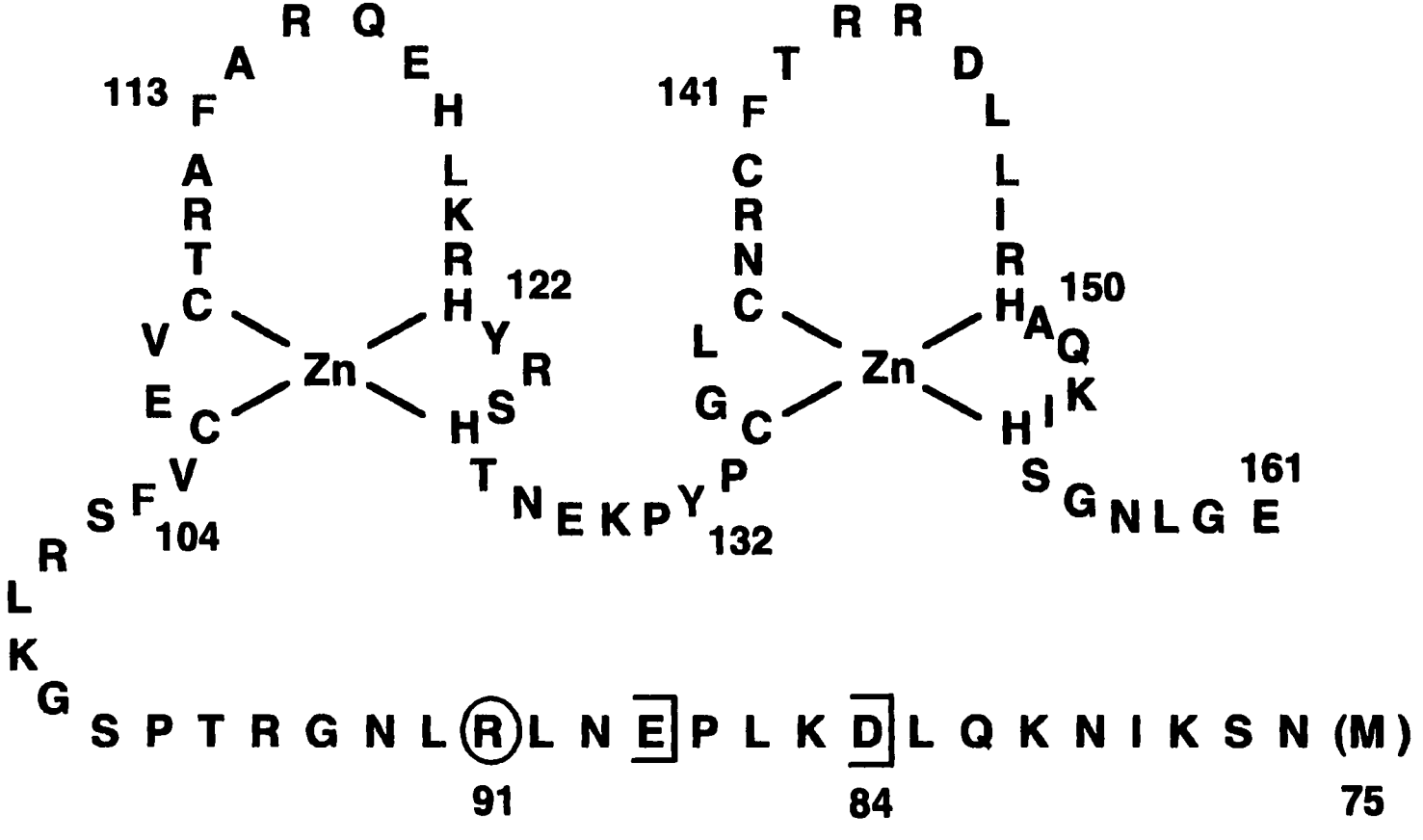
A. ADR1z' CONSTRUCT

The construct used in this study, ADR1z', spans residues 75-161 of ADR1, encompassing both zinc finger motifs as well as a region N-terminal to the first finger (Figure 12). This sequence was chosen based on deletion mutagenesis results that defined the minimal DNA-binding domain of ADR1 (53). While deletions C-terminal to the zinc fingers have no effect on DNA binding, deletions N-terminal to the fingers diminish binding affinity in electrophoretic mobility shift assay. N-terminal truncation to residue 84 of ADR1 results in a minor decrease in binding affinity. Further deletion to residue 88, however, results in loss of all detectable binding (53). The choice of Met 75 as a starting point for the construct allows inclusion of the pivotal region of residues 84 to 88 while minimizing the size of the polypeptide for NMR analysis.

For overexpression of ADR1z', a construct was subcloned that extends from codon 75 to 161 of ADR1. The details of the construction are supplied in Chapter 7. In brief, an engineered *Nde* I

Figure 12. ADR1z' Sequence.

The ADR1z' construct encompasses the entire minimal DNA-binding domain of ADR1, including the two zinc fingers and an additional N-terminal region. The shortest sequence that binds DNA with high affinity begins at residue 84; the longest sequence that fails to bind DNA with high affinity begins at residue 88. R91 is the location of a suppressor mutation that compensates for defects in DNA binding.


R Q
A
E
113 F
H
A
R
T
C
V
E
C
V
S F
104
L
K
G
L
K
Zn
H
Y
122
R
S
H
T
N E K P Y
132
R R
T
D
141 F
L
C
L
R
N
I
C
R
L
H
150
A
Q
G
C
K
P
I
H
S
G
161
N L G E
Zn
S P T R G N L R L N E P L K D L Q K N I K S N (M)
91
84
75

site takes advantage of the native ATG at codon 75, with a change at the preceding codon from AAT to CAT. An in-frame TAA stop codon was introduced immediately following codon 161, followed by a *Bam* H I site. Overexpression in *E. coli* was achieved using the Studier T7 polymerase system (72). The ADR1z' construct was introduced into plasmid pET11a (Novagen) between the unique *Nde* I and *Bam* H I sites and propagated in *E. coli* strain BL21(DE3). In this system the translation start codon is the ATG included in the *Nde* I site, and therefore the resulting polypeptide contains native sequence only.

B. OTHER CONSTRUCTS

Several other plasmids for overexpression were evaluated but did not perform as well. Fusions incorporating ADR1 75-175 into systems for production of calmodulin or trpE were constructed. While both systems were quite efficient for their original purpose, inclusion of the fused ADR1 sequences drastically reduced overexpression, indicating that expression of ADR1 DNA binding domain may be toxic. In fact, in the Studier T7 system that was eventually used, cells do continue growth after expression, but they do not reach stationary phase after more than twelve hours of induction.

In addition to the pET11a plasmid used in the final construct, another T7 polymerase-based plasmid was also tried: pT7-7. This plasmid produced barely detectable levels of ADR1z; it differs from pET11a in several ways, perhaps most importantly in the ribosome binding site upstream of the ATG start codon.

Finally, two different lengths of ADR1 DNA binding domain were made. ADR1z extends from residue 75 to 175, ADR1z' from 75 to 161. Both constructs expressed to high levels in the pET11a system and were easily purified. They differed little in solubility, affinity for DNA, and behavior at NMR concentrations. As will be shown later, the principle difference between these constructs is additional overlap in the NMR spectra for the longer molecule; therefore the shorter construct was chosen for all further studies.

II. Purification and Characterization of ADR1z'

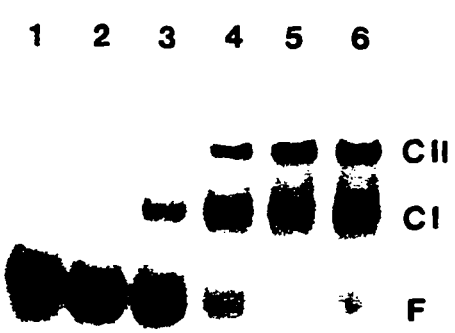
A. OVEREXPRESSION AND PURIFICATION PROTOCOL

Expression in *E. coli* allows efficient production of ADR1z' to approximately 5-10% of cellular protein (Figure 13). The inclusion of zinc in the medium is essential for high yield induction, indicating that the soluble, expressed zinc finger protein likely binds metal in the cell.

Figure 13. SDS-PAGE and EMSA Analysis of ADR1z.

The top panel is an SDS-PAGE gel. Lane four contains molecular weight markers; lane 3 shows the level of expression of ADR1z in induced cells, compared to uninduced cells in lane 2; and lane 1 shows the purified ADR1z polypeptide. The lower panel is an electrophoretic mobility shift assay (EMSA) demonstrating DNA binding by purified ADR1z. Each lane contains 1 ng radiolabeled UAS1 50mer DNA. Lanes 2 - 6 also contain 1, 2, 4, 6, and 8 ng of ADR1z protein in 20 μl. Two shifted complexes are formed, designated as CI and CII, corresponding to one or two molecules of ADR1z bound per UAS1 palindrome. Using the assumption that ADR1z binds as a monomer to each UAS1 halfsite, the ADR1z-DNA dissociation constant calculated from this experiment is 12 nM.


1
2
3
4


1
2
3
4
5
6
CII
CI
F

Purification proceeds through a simple two-column protocol. The extremely basic nature of ADR1z' allows purification by cation exchange column to approximately 85% purity. A high resolution reverse phase HPLC step completes the purification to 98 - 99% purity (Figure 13).

Final yields average 0.5 μmol per liter of culture, enough for a 1 mM NMR sample. At this level of expression, isotopic labeling with ^{15}N and ^{13}C is feasible. At date, one liter of uniformly ^{15}N, ^{13}C-labeled glucose-based medium costs about $800, meaning the typical 2 mM ADR1z' NMR sample costs $1600, well within reason.

B. SAMPLE QUALITY

The identity of the expressed polypeptide was confirmed during initial purification trials by Western blots using polyclonal antibodies to the zinc fingers of ADR1. In addition, gel mobility shift experiments were conducted on column fractions using radiolabeled DNA containing the ADR1 UAS1 target sequence.

EMSA experiments on purified ADR1z were also conducted by Nataly Kacherovsky and Sylvie Camier in the laboratory of Ted Young. A typical experiment is shown in Figure 13. The calculated dissociation constant for ADR1z-DNA from this experiment is on

the order of 12 nM, demonstrating high affinity binding for the fragment to be used in NMR studies. A final check on the integrity of the purified ADR1z' polypeptide was obtained by electospray mass spectrometry in the laboratory of Kenneth Walsh. The expected mass is 10,334.7 daltons. Two molecular masses of purified ^{15}N-labeled polypeptide were observed: 10,335 and 10,203 daltons, indicating partial post-translational removal of the N-terminal methionine.

III. Initial Spectra

A. SPECTRA OF ADR1z'

ADR1z' samples were typically 2 mM, folded in the presence of two equivalents of zinc. A typical 1D NMR spectrum of ^{15}N-labeled ADR1z' is shown in Figure 14. The presence of upfield shifted methyls, downfield shifted alpha protons, and downfield shifted amide protons is indicative of folded protein (61).

An ^{15}N-HSQC spectrum of ADR1z' is shown in Figure 15. This type of spectrum displays peaks for each backbone amide proton, as well as sidechain NH_2 shifts, resolved by the shift of the attached nitrogen. The spectrum of ADR1z' shows chemical shift dispersion

Figure 14. 1D Spectrum of ADR1z'.

Two portions of the 1D proton spectrum of ^{15}N-labeled ADR1z' are plotted. The highlighted downfield-shifted amide protons and upfield-shifted methyl protons are indicative of folded structure. This spectrum was acquired at pH 5.4 and 25 ^{o}C.

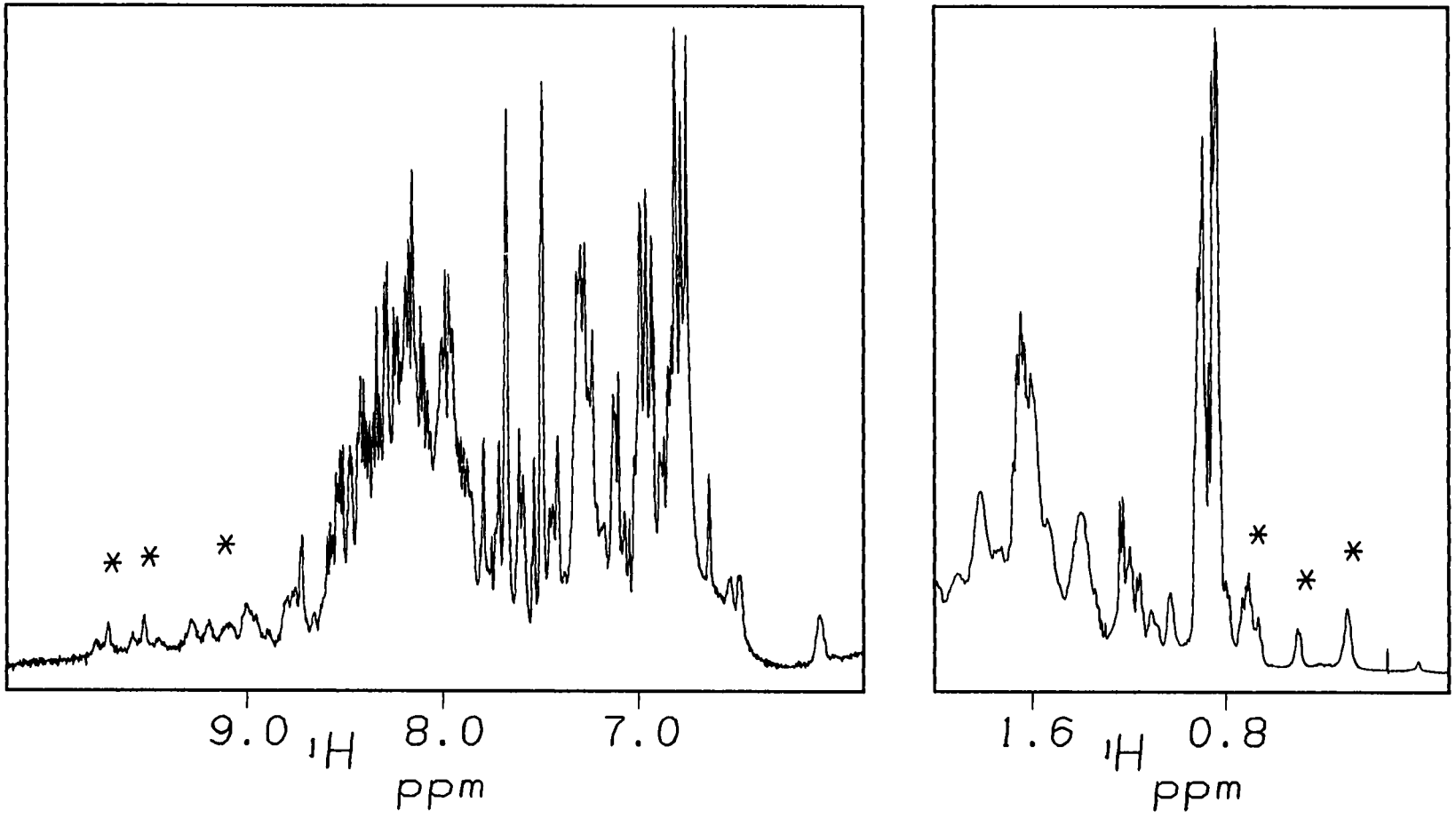
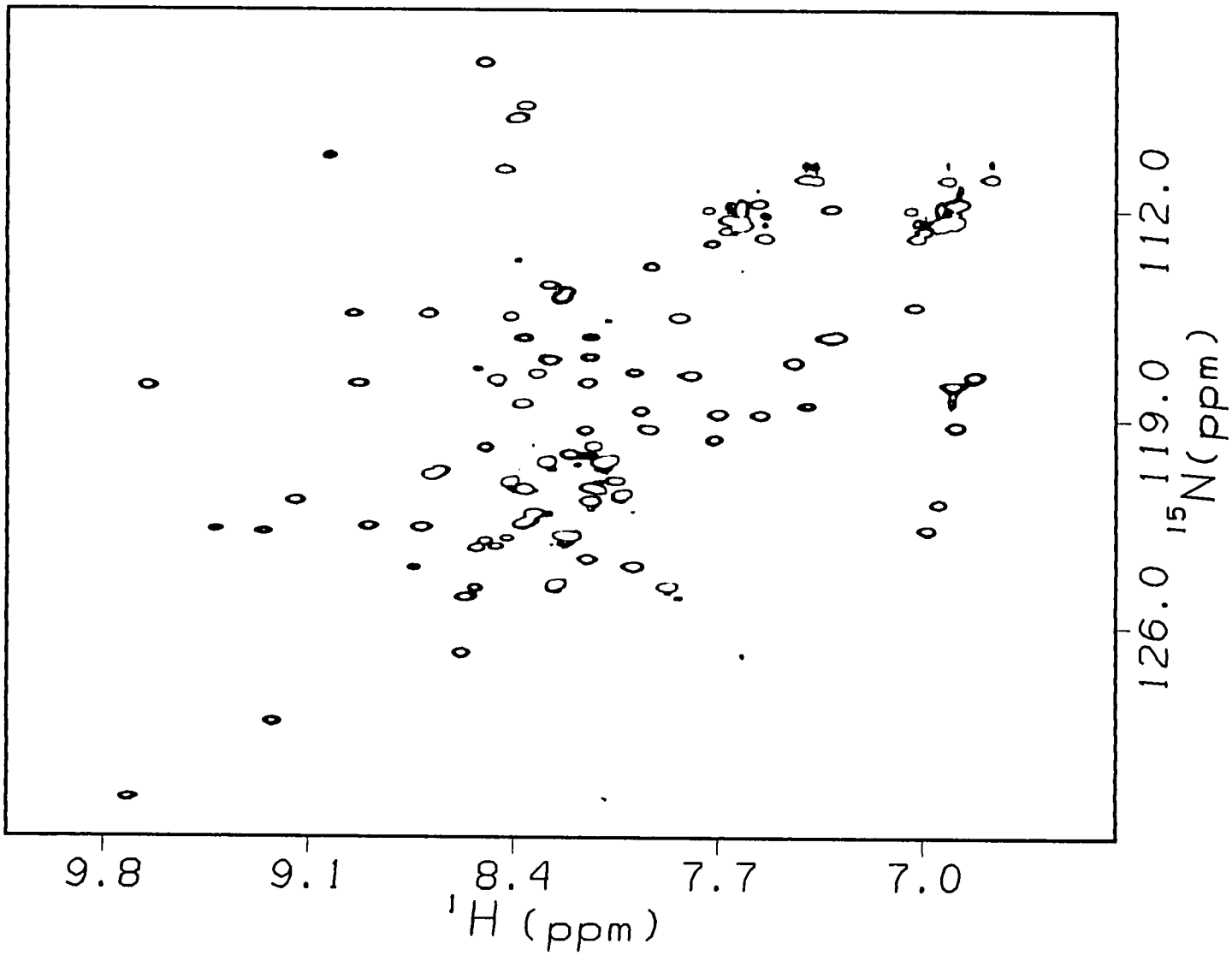

9.0
1H
8.0
7.0
ppm
1.6
1H
0.8
ppm

Figure 15. HSQC of ADR1z'.

The ^{15}N-HSQC experiment gives a peak for each backbone amide proton of an ^{15}N- labeled protein. The HSQC of ADR1z' displays a single set of lines for all residues except those at the extreme N- and C-termini. There are few regions of extreme overlap, an indication that experiments such as the ^{15}N-edited NOESY and triple resonance HNCA will give data of sufficient resolution to make assignments.


9.8
9.1
8.4
7.7
7.0
^1H (ppm)
126.0
119.0
112.0
^{15}N (ppm)

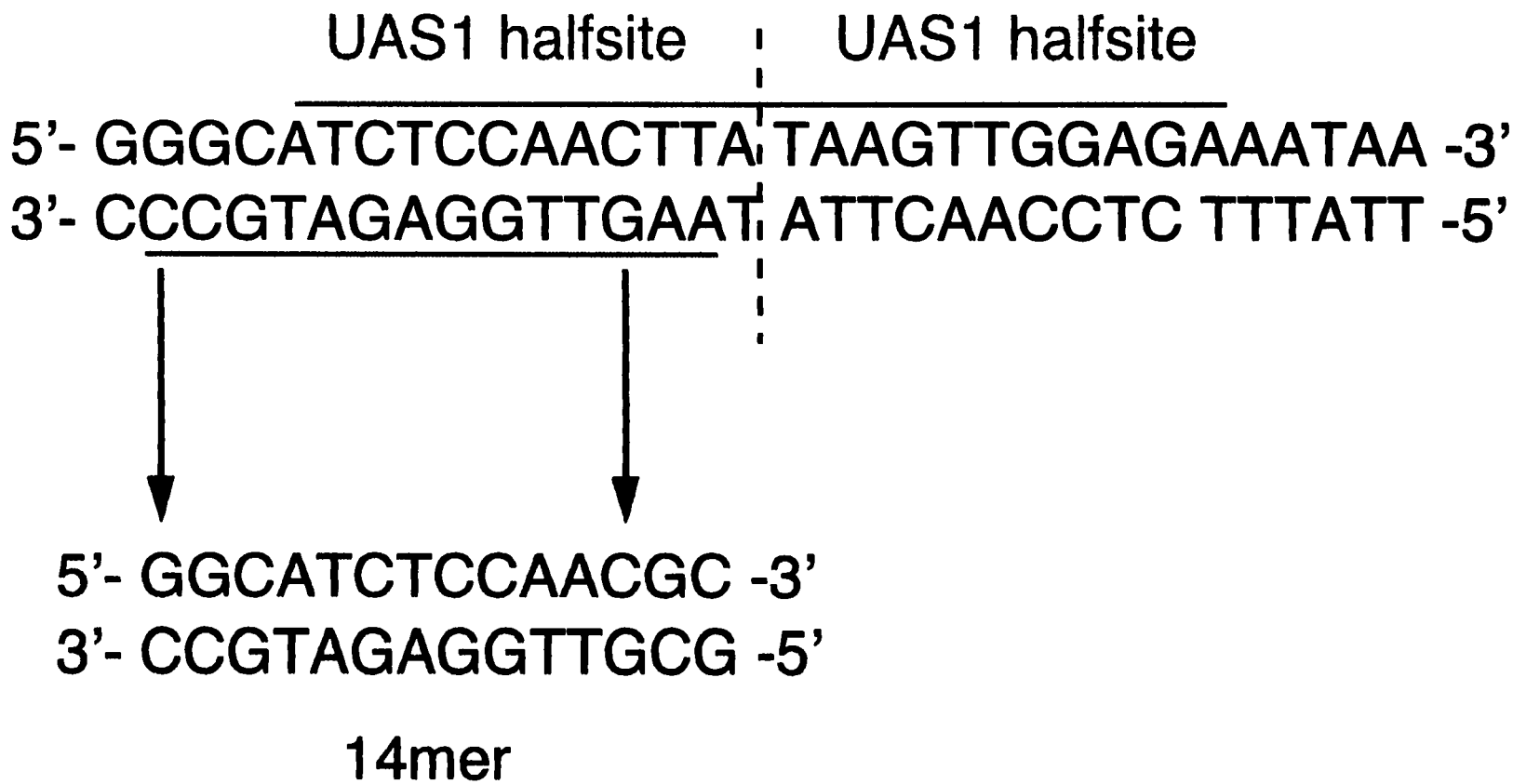
that is adequate for further steps in the NMR analysis. Peaks are detected for all but the two most N-terminal residues, which likely undergo rapid exchange with solvent on the timescale of the NMR experiment. Very limited doubling of peaks is detected, for the three most N-terminal residues that are detectable, K78, I79, and K80, and for the two C-terminal residues, G160 and E161. The N-terminal doubling is likely due to the heterogenous processing of Met 75. The C-terminal doubling is unexplained, but has been observed in ADR1a single finger peptides (60),(R. E. Klevit, personal communication), and in a SWI5 zinc finger peptide (30).

B. SPECTRA OF ADR1z'-DNA

Two monomers of ADR1 bind to a palindromic DNA site of 22 base pairs called UAS1 found upstream of *ADH2* and other yeast genes (53). To minimize the size of the final protein-DNA complex for NMR analysis, only a single half-site of UAS1 was included in the DNA sequence used for this study. Several versions of this sequence were evaluated, ranging in length from 12 to 19 basepairs; the chosen 14mer emerged as the minimal size to retain good solubility characteristics when bound to ADR1z' (Figure 16). This 14mer contains native UAS1 sequence, except for two A/T basepairs from the center of the palindrome which have been

Figure 16. 14mer DNA Oligo.

The DNA used in NMR studies of the ADR1z'-DNA complex is 14 basepairs long. This 14mer contains one UAS1 halfsite and should bind one molecule of ADR1z'. The resulting complex will have a molecular weight of about 20 kD, a size reasonable for NMR studies.


UAS1 halfsite
UAS1 halfsite
5'- GGGCATCTCCAACTTA TAAGTTGGAGAAATAA -3'
3'- CCCGTAGAGGTTGAAT ATTCAACCTC TTTATT -5'
5'- GGCATCTCCAACGC -3'
3'- CCGTAGAGGTTGCG -5'
14mer

replaced by G/C basepairs to avoid end fraying. Previous mutagenesis studies showed that ADR1 exhibits no base preference at these positions (59).

ADR1z' forms a complex with 14mer DNA that is soluble at 2 mM under NMR conditions. The 1D proton NMR spectrum of ^{15}N-labeled ADR1z'-DNA is shown in Figure 17. As for the free protein, this spectrum shows the hallmarks of folded protein. The methyl region in particular may be used to evaluate the presence of bound protein; differences from the free protein spectrum are highlighted in Figure 18.

An ^{15}N-HSQC spectrum of ADR1z'-DNA is shown in Figure 19. The spectrum displays a single set of resonances, except for the limited N-terminal doubling due to heterogenous ADR1z' processing discussed for the free protein. The dispersion is again adequate for further NMR analysis, and the quality of the spectrum is acceptable.

NMR spectra of the longer ADR1z construct were evaluated as well. ADR1z extends from 75 to 175 of ADR1, well past the C-terminal end of the zinc fingers, while ADR1z' includes only 75 to 161, ending just at the edge of the second finger. HSQC spectra for the two constructs bound to DNA are shown in Figure 20. Both constructs bind DNA to give spectra with dispersed peaks in nearly identical positions, indicating similar binding modes. Given this

Figure 17. 1D Spectrum of ADR1z'-DNA.

Two regions of the 1D proton spectrum of ^{15}N-labeled ADR1z' bound to DNA are shown. The upper spectrum was acquired at 37 °C, the lower one at 25 °C. The highlighted upfield-shifted methyls and downfield-shifted amide protons are indicative of folded structure.

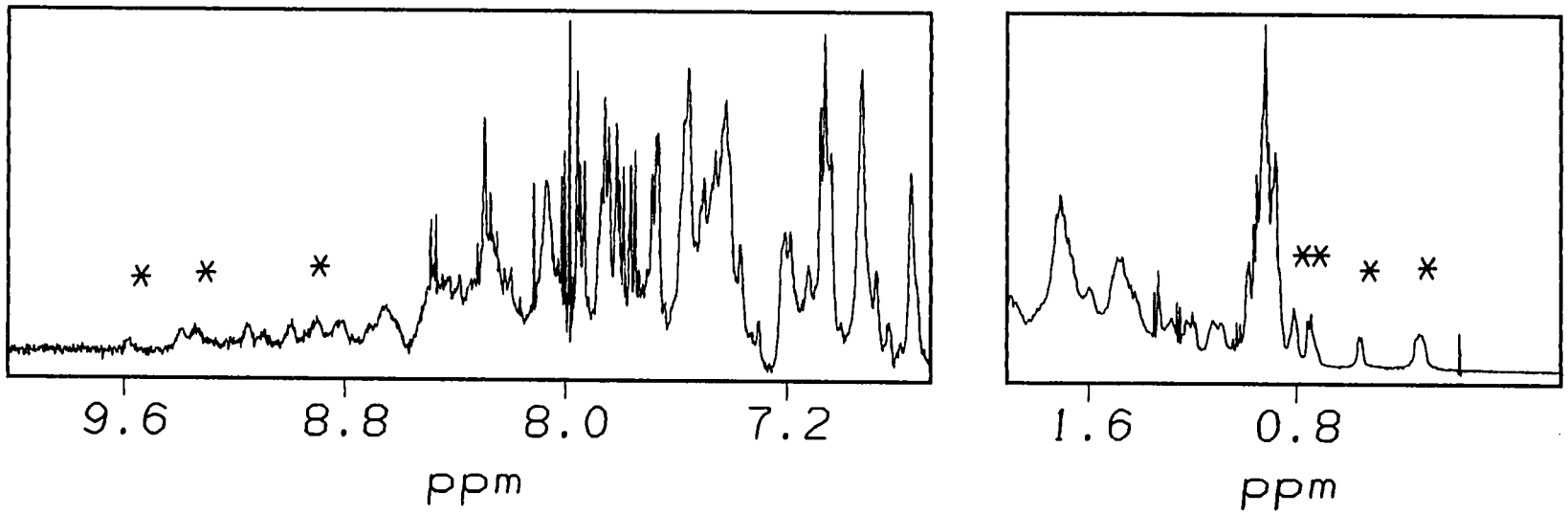
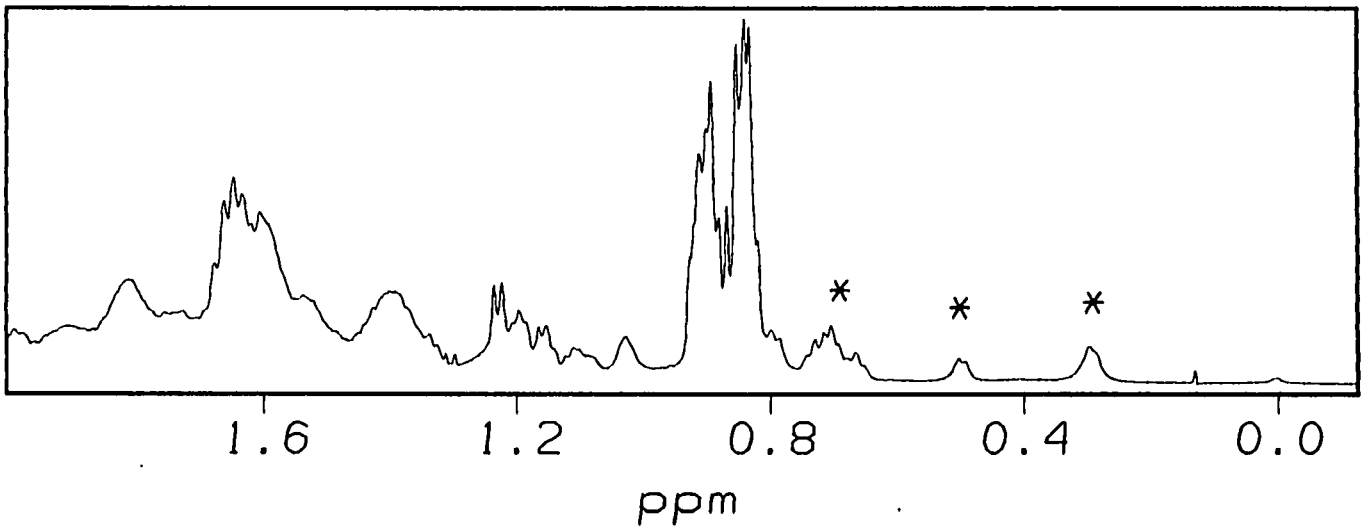

* * *
9.6
8.8
8.0
7.2
ppm
** * *
1.6
0.8
ppm
* * *
9.6
8.8
8.0
7.2
ppm
** * *
1.6
0.8
ppm

Figure 18. Comparison of 1D Spectra of ADR1z' Free and Bound to DNA.

Methyl regions of the 1D proton spectrum of free (upper panel) and bound ADR1z' at 25 ^{o}C are plotted. Differences in the upfield-shifted methyls diagnostic of complex formation are highlighted.


*
*
*
1.6
1.2
0.8
0.4
0.0
ppm

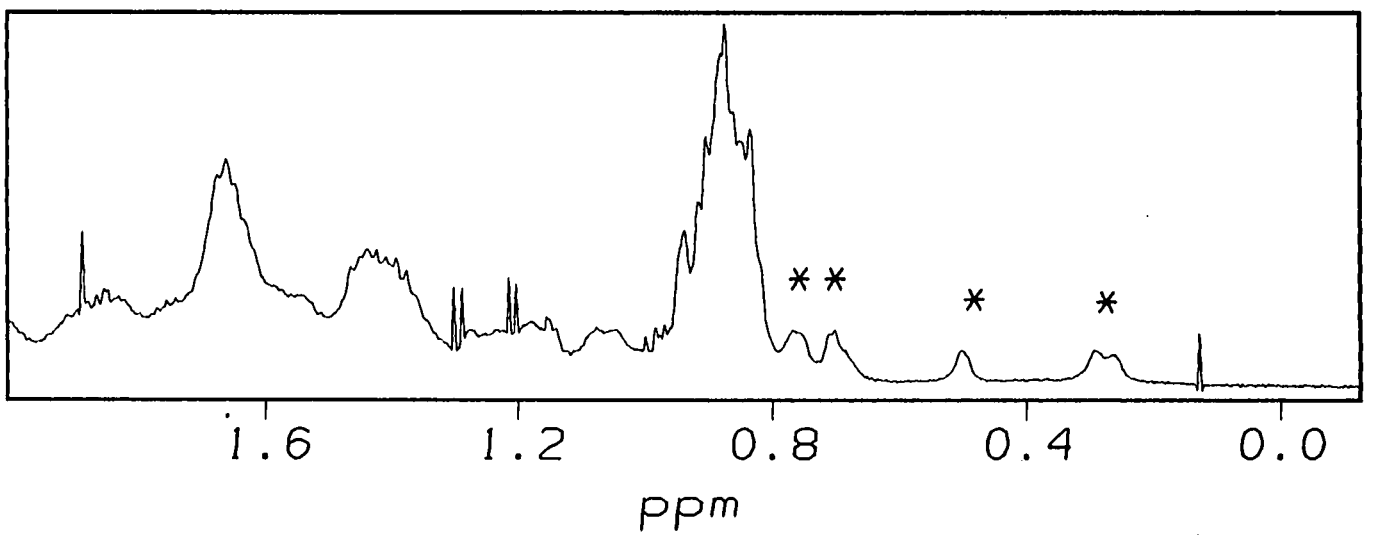

**
*
*
1.6
1.2
0.8
0.4
0.0
ppm

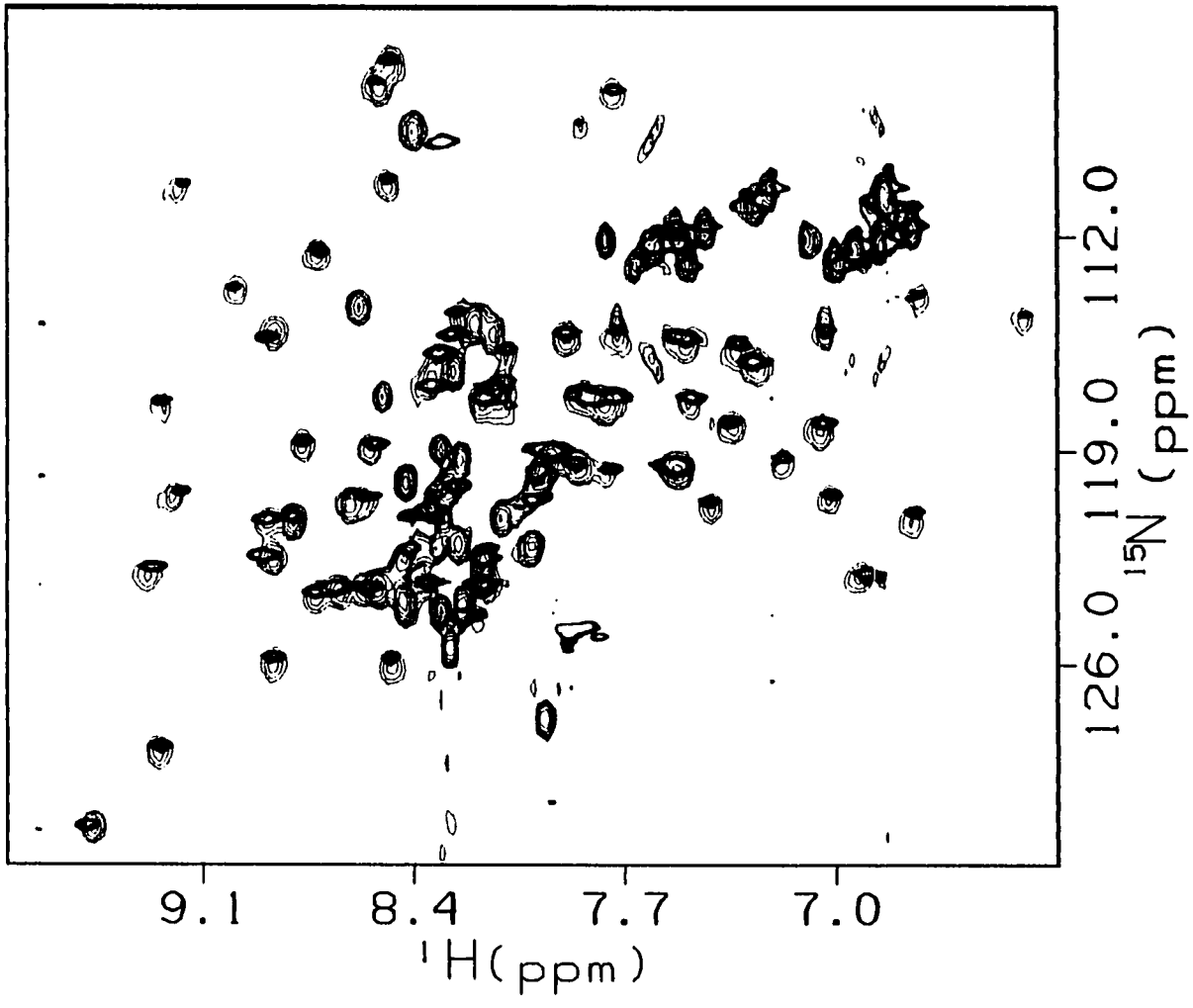
Figure 19. HSQC of ADR1z'-DNA.

The HSQC of ^{15}N-labeled ADR1z' bound to unlabeled 14mer DNA displays a single set of lines and resolution sufficient to proceed with the assignment process.

128.0
120.0
112.0
^{15}N (ppm)
10.5
9.8
9.1
8.4
7.7
7.0
^{1}H (ppm)

Figure 20. Overlap of HSQC Spectra of ADR1z-DNA andADR1z'-DNA.

ADR1z (75-175) and ADR1z' (75-161) give nearly identical HSQC spectra when bound to DNA. The ADR1z-DNA spectrum was acquired at 25 °C, pH 5.4 and is plotted in red. The ADR1z'-DNA spectrum was acquired at 25 °C, pH 7.0 and is plotted in blue. The principle differences between these spectra fall in the central region where random coil residues resonate. The significantly worse overlap for ADR1z-DNA in this region, along with the similarity of the spectra in other regions, argues for use of ADR1z' in further NMR studies. In addition, the nearly identical spectra obtained at different pH argue that comparison of the free protein at low pH to the complex at higher pH is not unreasonable.


9.1
8.4
7.7
7.0
^{1}H(ppm)
126.0
119.0
112.0
^{15}N (ppm)

similarity and the obvious complication of the spectra of ADR1z from the additional random coil peaks at the C-terminus, it was decided to use the less overlapped ADR1z' spectra for the remainder of the study.

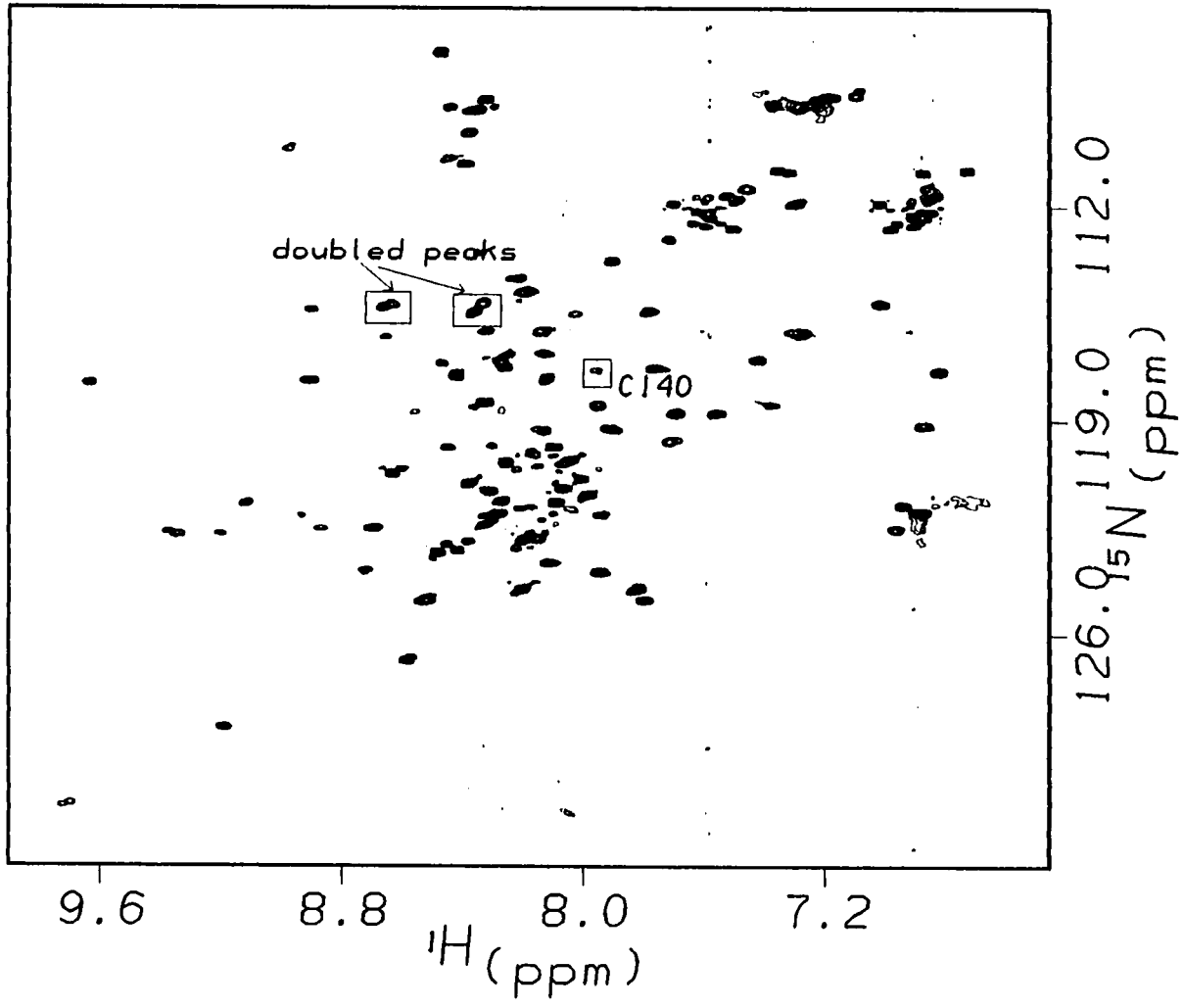
C. RULES FOR JUDGING QUALITY OF SPECTRA

The biggest threat to the quality of the NMR spectra of ADR1z' and its complex with DNA is oxidation. Partially oxidized protein samples look in the tube much like fully reduced protein. The HSQC spectrum, however, undergoes obvious changes that signal trouble. The first sign is the loss of several peaks from the spectrum that will later be shown to lie in the region of the sole non-zinc ligand cysteine in the sequence, C140. Further changes to the spectrum occur with more serious oxidation, especially at high pH and temperature. These include the appearance of peak doubling (Figure 21). Both symptoms of oxidation can be reversed by the addition to samples of fresh DTT at 5 mM.

The early signs of oxidation of the protein-DNA complex are more difficult to detect in HSQC spectra, as it manifests itself in a general deterioration of the quality of signal, as opposed to discrete peak loss or doubling. The complex is much less susceptible to oxidation in any case, likely due to partial

Figure 21. HSQC Spectra of ADR1z', oxidized ADR1z'.

An HSQC spectrum of oxidized ADR1z' is plotted in red, the spectrum of the same sample with added DTT in green. The principle change upon oxidation is the loss of the peak from the only non-zinc ligand cysteine, C140. Other changes include the appearance of peak doublings in some regions of the spectrum. Examples of these changes are shown by boxes in the spectrum. These spectra were acquired at pH 7.0, at 37 ^{o}C.


doubled peaks
C140
9.6
8.8
8.0
7.2
^1H (ppm)
126.9
119.0
112.0
^{15}N (ppm)

protection of C140 in the bound state. However, the methyl region of the 1D spectrum serves as an indicator of oxidation, as the resolved shoulder to the main methyl grouping tends to be subsumed into the main peak as oxidation progresses (Figure 22). This too can be reversed by the addition of 5 mM fresh DTT.

D. SLOW EXCHANGE BEHAVIOR

The ADR1z'-DNA complex exhibits slow exchange behavior on the NMR timescale, meaning that free and bound components to the equilibrium do not exchange during the course of the NMR experiment. This is demonstrated by a titration of excess DNA into an ADR1z'-DNA complex. Figure 23 shows the imino protons of the DNA during the course of the titration. As excess DNA is added, new lines appear at the position of free DNA resonances, while the bound DNA lines remain at their previous positions and intensities. This behavior reflects the presence of two populations of DNA in the sample that do not exchange during the course of the NMR experiment.

Such slow exchange behavior has serious implications for the assignment process, however. As shown in Figure 24, many resonances of ADR1z' undergo large chemical shift perturbations upon DNA binding, and simple inspection of the spectra does not

Figure 22. 1D Spectra of Oxidized and Reduced ADR1z'-DNA.

The methyl regions of 1D proton spectra of reduced (upper panel) and partially oxidized ADR1z'-DNA are compared. The principle detectable change is a broadening of the upfield methyl peaks. This change is reversible with addition of fresh DTT to the samples.

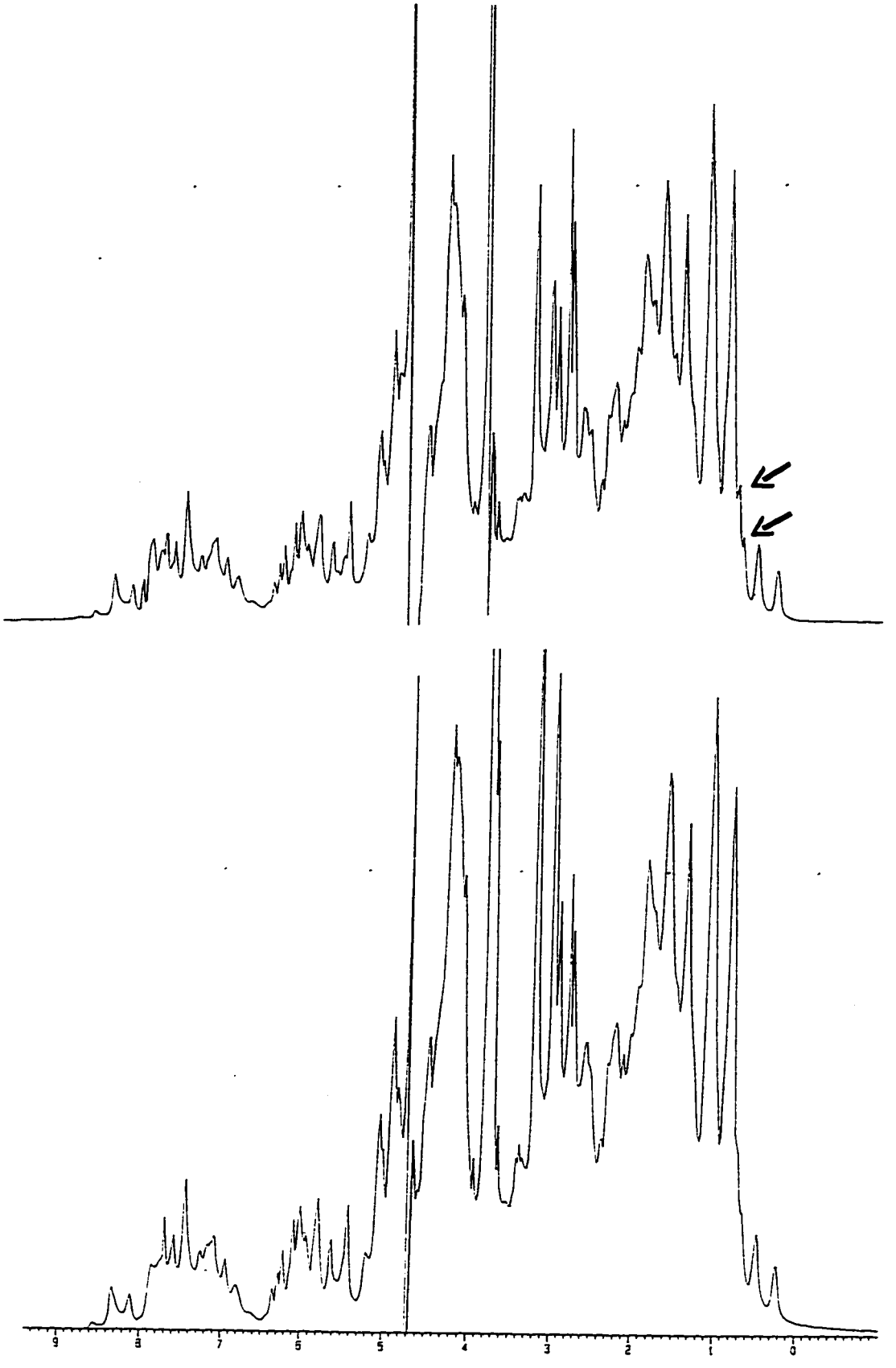

9
8
7
6
5
4
3
2
1
0

Figure 23. The ADR1z'-DNA Complex is in Slow Exchange.

1D proton spectra of the imino region of 17mer UAS1 DNA are plotted. The bottom spectrum is that of a stoichiometric complex of 17mer with ADR1z. Successive spectra contain increasing amounts of excess 17mer; the upper spectrum is that of 17mer alone. The intermediate spectra show two sets of lines corresponding to distinct populations of free and bound DNA. This slow exchange behavior indicates that the free DNA is not exchanging into the complex on the NMR timescale.

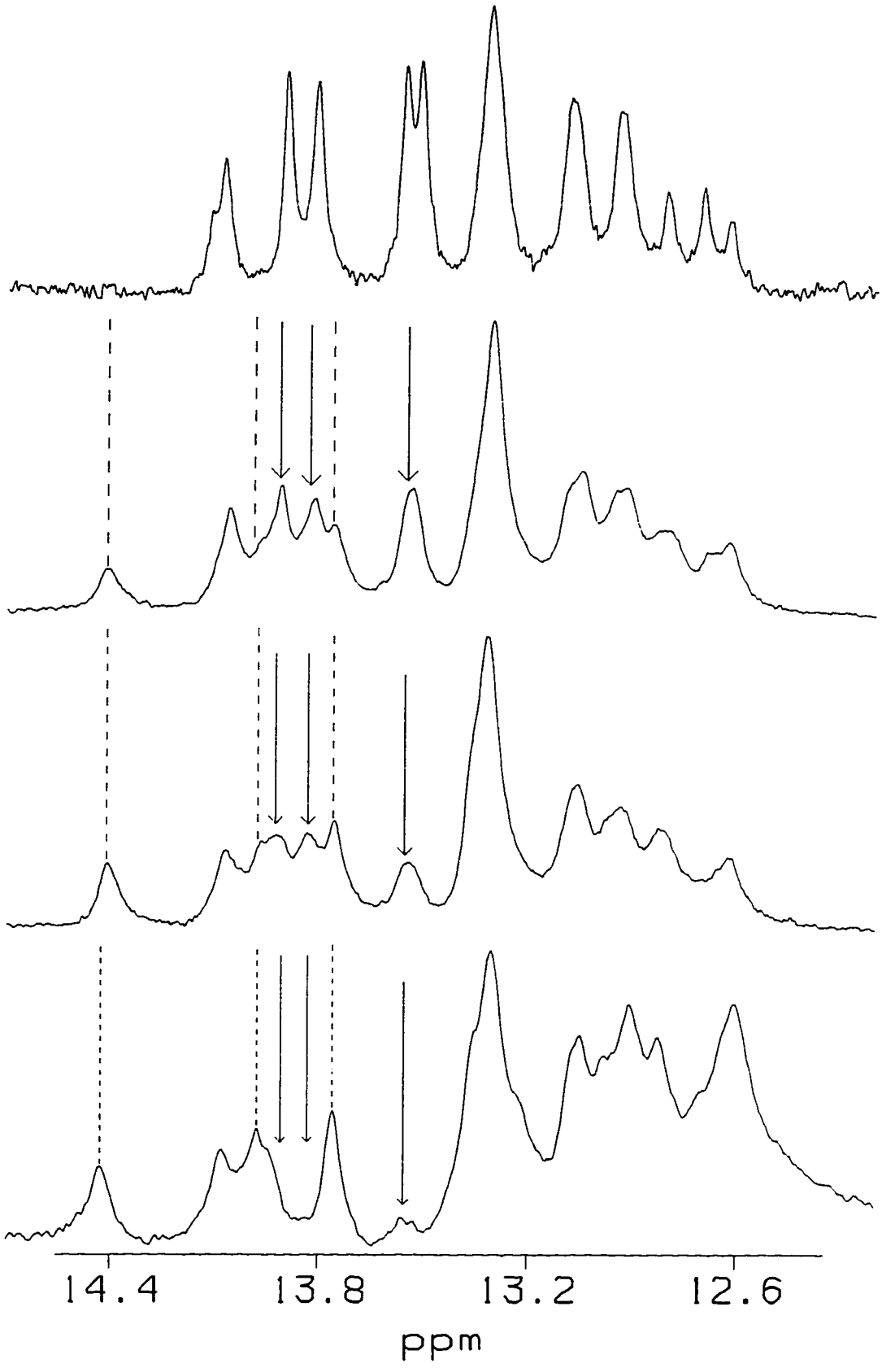
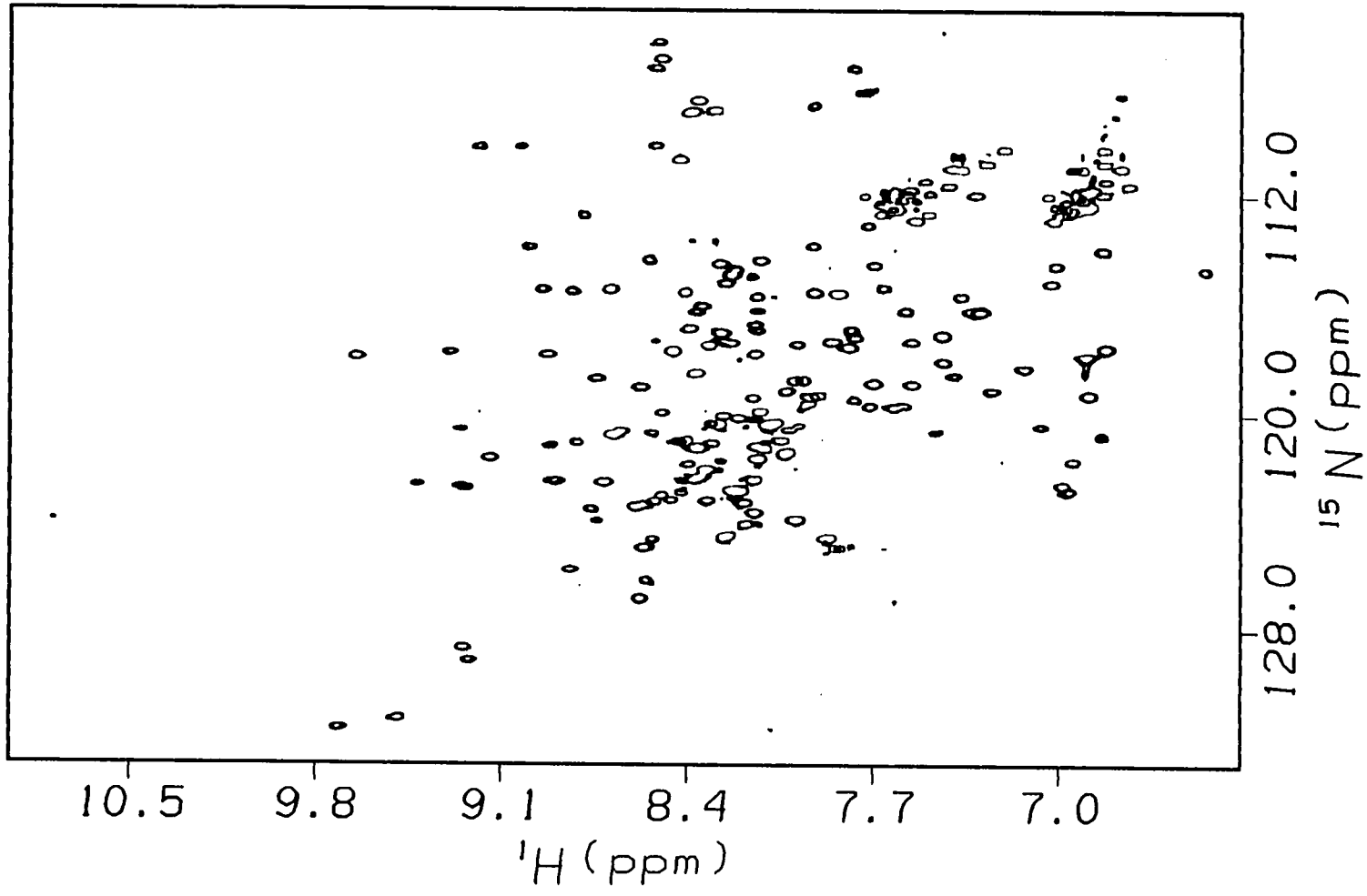

14.4
13.8
13.2
12.6
ppm

Figure 24. HSQC Spectra of Free and Bound ADR1z'.

The ^{15}N-HSQC spectrum of free ADR1z' (red) is overlapped with that of ADR1z' bound to 14mer DNA (blue). A large number of resonances undergo significant chemical shift changes upon DNA binding. The free spectrum was acquired at pH 5.4, 30 °C; the bound spectrum was acquired at pH 7.0, 37 °C. These differences in temperature and pH alone do not perturb the spectrum of free ADR1z' significantly, as shown in Figure 55.


10.5
9.8
9.1
8.4
7.7
7.0
^{1}H (ppm)
112.0
120.0
128.0
^{15}N (ppm)

reveal the identities of the bound resonances. Were fast exchange behavior to apply, assignment of the shifted resonances could easily be accomplished by titration of DNA into protein. The chemical shift of a protein resonance at each titration step would be a weighted average of the free and bound shifts, essentially marking a trajectory from initial (free) shift to final (bound) shift. Peaks in slow exchange cannot be followed in this kind of titration, however, as new peaks arise from the start at their final chemical shift. Thus ADR1z' assignments must be obtained separately for the free and the bound species, significantly complicating the assignment process.

CHAPTER 2. RESONANCE ASSIGNMENTS

I. Strategy and Rationale

Because of the slow exchange behavior of the ADR1z'-DNA complex, assignment of ADR1z' free and bound must take place independently. The ^{15}N-HSQC spectrum of each form is promising, in that there is not a great deal of overlap of the backbone amide resonances when separated by both proton and nitrogen chemical shift (Figure 24). This is essential because the amide positions of the backbone are pivotal to assignment strategies based either on traditional NOE connectivities or on triple resonance experiments.

Assignment of ADR1z' free and bound to DNA relied on both types of assignment tools, as well as on specific labeling of amino acids. The small amount of overlap of amide resonances is outweighed by fairly extensive overlap of alpha carbon and alpha proton resonances, which necessitates a multi-technique strategy. Finally, the size of the molecules, especially of the complex, precludes use of traditional proton-proton coupling experiments such as the TOCSY for assignment of sidechain resonances. The HCCH-TOCSY was used instead, in which larger carbon-carbon couplings provide the basis of transfer of magnetization along the sidechain.

II. Assignment of ADR1z'

A. TRIPLE RESONANCE

The cornerstone of the assignment process for free ADR1z' is the triple resonance HNCA and HN(CO)CA pair, which utilizes uniformly $^{15}N,^{13}C$-labeled sample (reviewed in (63)). These experiments provide connectivities between neighboring alpha carbons, resolved by the well-dispersed backbone nitrogen and amide proton resonances (see Figure 25). Given sufficient dispersion of alpha carbon, amide nitrogen, and amide proton chemical shift, the HNCA, HN(CO)CA pair can provide sequential assignments of backbone resonances. Examples of the data provided by these experiments for ADR1z' are shown in Figure 26, organized according to the final NH assignments.

Due to considerable redundancy of C^{α} chemical shift, however, additional experiments were required to obtain unambiguous backbone assignments for ADR1z'. Such is usually the case for proteins of a size that requires heteronuclear experiments; often the solution is to run an additional pair of triple resonance experiments that correlate residues through their carbonyl carbons (in the hope that these will not be overlapped as well). In the case of ADR1z', however, considerable data already

Figure 25. Connectivity Patterns in the HNCA and HN(CO)CA Triple Resonance Experiments.

The HNCA experiment correlates the amide proton and nitrogen of residue i with the alpha carbons of residues i and i-1. The HN(CO)CA selects only the i to i-1 correlation. The combination of these two experiments gives chemical shift data on the alpha carbons preceding and following any given amide group; in the absence of confounding chemical shift overlap, this technique provides a means to identify neighboring amino acid residues as part of the sequential assignment process.

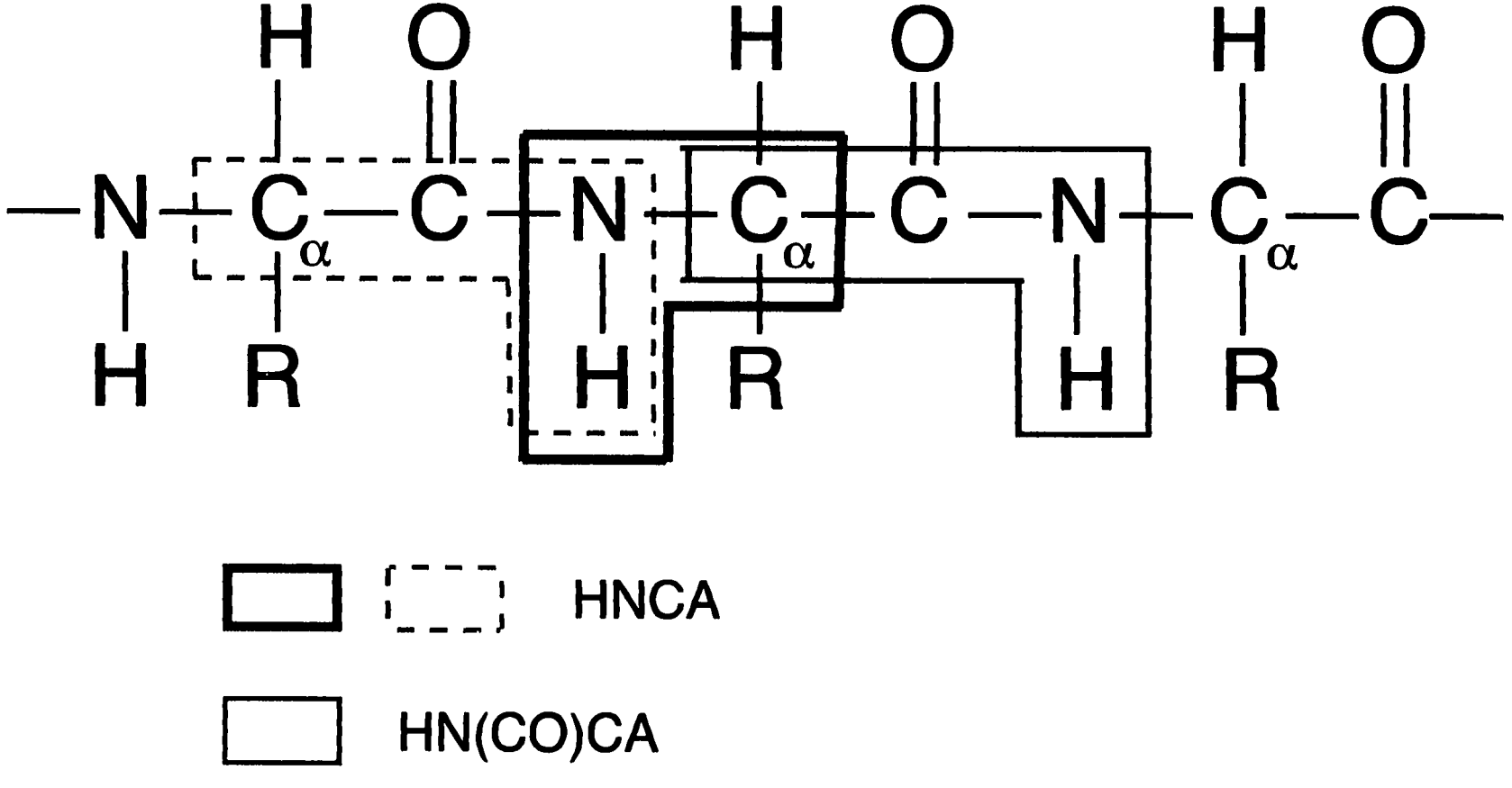
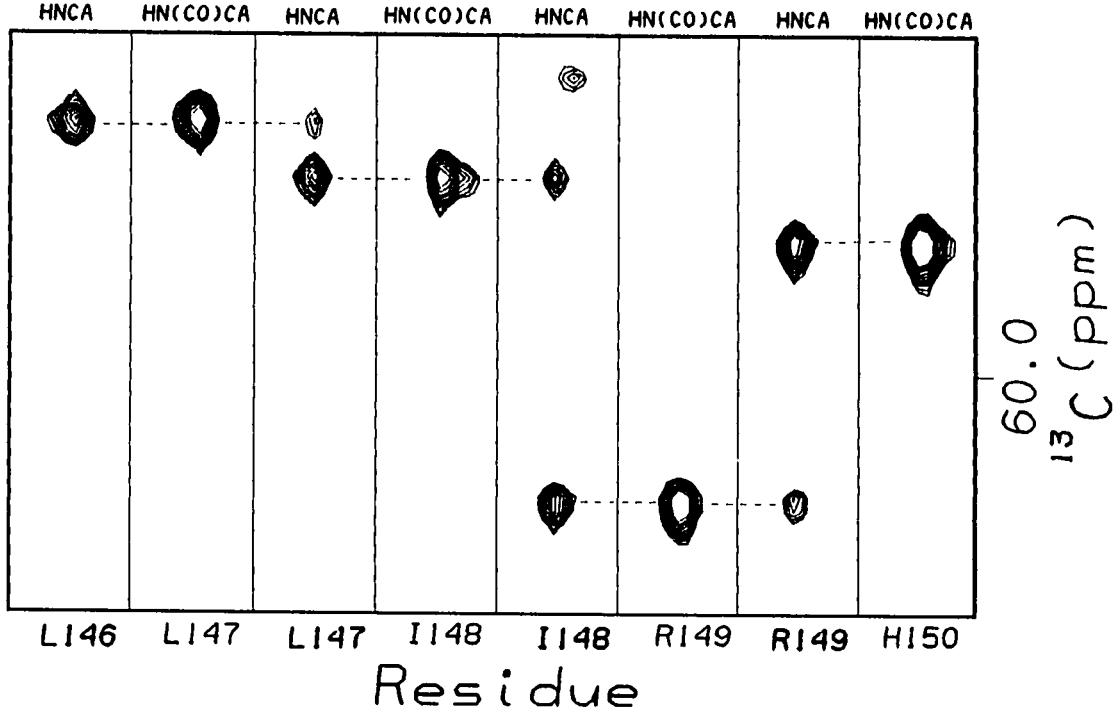

H
O
H
O
H
O
N
C
α
C
N
C
α
C
N
C
α
C
H
R
H
R
H
R
HNCA
HN(CO)CA

Figure 26. HNCA and HN(CO)CA Spectra of Free ADR1z'.

Sections of the 3D HNCA and HN(CO)CA spectra of ADR1z' are taken at the amide proton chemical shift and amide nitrogen chemical shift positions of several sequential residues. The HNCA and HN(CO)CA sections are interleaved. Two alpha carbons are correlated to each amide in the HNCA, corresponding to the i and i-1 residues. A single alpha carbon is correlated to each amide in the HN(CO)CA, corresponding to the i-1 residue.


HNCA
HN(CO)CA
HNCA
HN(CO)CA
HNCA
HN(CO)CA
HNCA
HN(CO)CA
60.0
^{13}C (ppm)
L146
L147
L147
I148
I148
R149
R149
H150
Residue

existed from isotopic labeling of specific amino acids, and these were used to resolve ambiguities instead.

B. SPECIFIC NITROGEN LABELING

ADR1z' samples labeled in *E. coli* with ^{15}N-leucine or ^{15}N-glutamate were analyzed to classify backbone amide peaks in HSQC experiments (Figure 27) (73). Due to differential scrambling of the isotope label to Leu, Ile, Val, and Ala in the leucine experiment, and to Asp, Glu, Gln, Phe, Tyr, Asn, Leu, Val, Ile, Ser, and Ala in the glutamate experiment, peaks were assignable to the following classes: Asp, Glu, Gln or Phe; Leu; Val or Ile; Ala; Asn, Ser or Tyr; and Cys, His, Lys, Gly, Arg or Thr (unlabeled).

Further clarification of amino acid type was obtained from samples grown on ^{15}N-labeled media to which ^{14}N-lysine or arginine was added (74). Because metabolic scrambling of the amide nitrogen of these residues is inefficient, HSQC spectra of these samples lack peaks for lysine and arginine, respectively, allowing their identification (Figure 28).

Figure 27. Specific ^{15}N-Labeling: ^{15}N-Leu- ADR1z.

An HSQC spectrum of ADR1z specifically labeled with ^{15}N-leucine is overlapped with a spectrum of uniformly ^{15}N-labeled ADR1z. Strong crosspeaks in the specific labeling experiment belong to the leucines in the polypeptide; weaker peaks also arise due to scrambling of ^{15}N to Ile and Val during the metabolic labeling process. The peaks for nine of ten leucines are evident for the specifically labeled sample. The tenth leucine is L136, which gives a very weak signal even in uniformly labeled spectra. The peak for L136 in the specifically labeled sample is below the level plotted here; the peak shows the same intensity relative to uniformly labeled spectra as do the other leucine peaks, however.

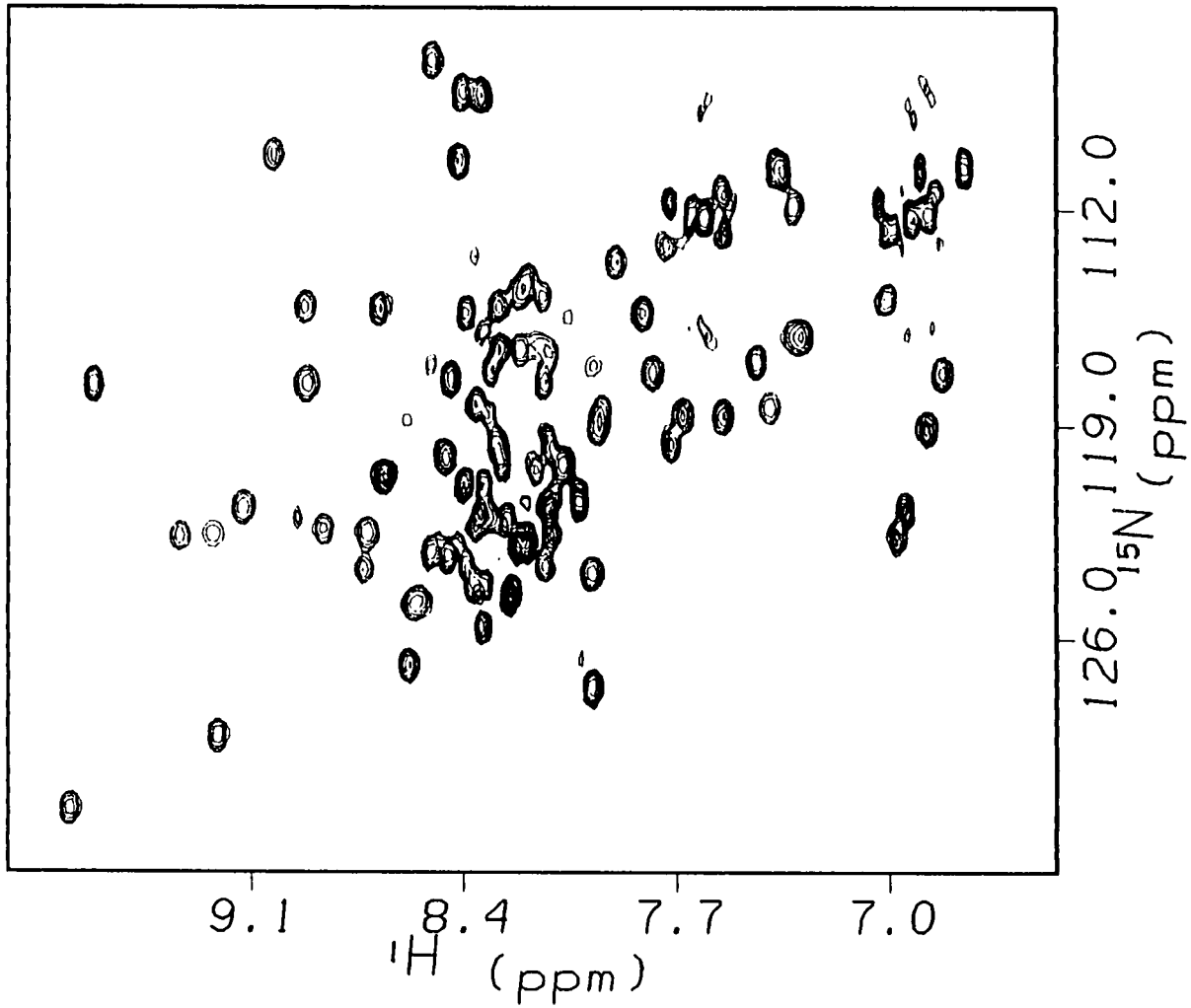
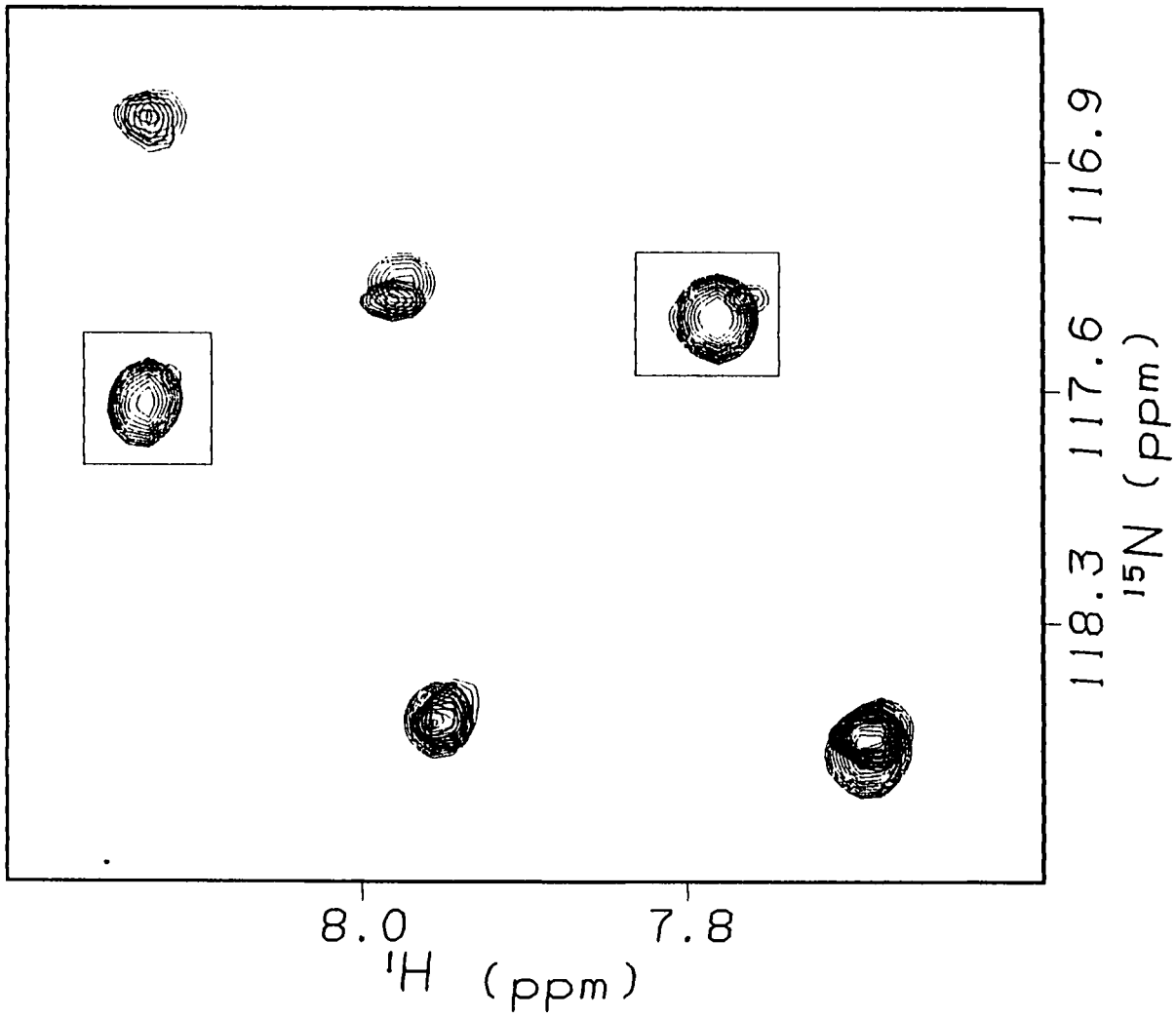

126.0
119.0
112.0
^{15}N (ppm)
9.1
8.4
7.7
7.0
^{1}H (ppm)

Figure 28. Specific ^{14}N labeling: ^{14}N-Lys-ADR1z'.

An ^{15}N-HSQC spectrum of ADR1z' specifically labeled with ^{14}N-lysine (red) is overlapped with that of uniformly ^{15}N-labeled ADR1z' (green). Two peaks which are much lower in intensity in the specifically labeled sample are highlighted. These peaks arise from lysine residues; all such ^{14}N-labeled lysine residues essentially lack ^{15}N-HSQC peaks.


116.9
117.6
118.3
^{15}N (ppm)
8.0
7.8
^{1}H (ppm)

C. SPECIFIC CARBON LABELING

Additional sequential information was obtained from specific ^{13}C labeling (74). Leucine, arginine, or lysine labeled at the backbone carbonyl position with ^{13}C was incorporated into uniformly ^{15}N-labeled ADR1z'. In ^{15}N-HSQC spectra of these samples, backbone amides neighboring the incorporated ^{13}C label yield crosspeaks split in the ^{15}N dimension (Figure 29), allowing identification of the C-terminal neighbors to Leu, Lys, and Arg.

Information on peak type from the ^{15}N labeling experiments was combined with information on C-terminal neighbor from the ^{13}C labeling experiments to identify dipeptides. Each dipeptide assigned was of the form XaaYaa, where the N-terminal amino acid Xaa was either Arg, Lys, Leu, or none of these; and where the C-terminal amino acid Yaa was in one of the categories defined by ^{15}N labeling. Because a third of the sequence of ADR1z' consists of Arg, Lys, or Leu residues, a large number of unambiguous dipeptides were identified.

The dipeptide assignments from specific labeling experiments, combined with the HNCA, the HN(CO)CA, and traditional NOESY-based connectivities obtained from an ^{15}N-edited TOCSY and NOESY pair (61), allowed complete assignment of the backbone N, NH, and C^{α} resonances of free ADR1z'.

Figure 29. Specific ^{13}C Labeling: $^{13}C_1$-Leu-ADR1z'.

ADR1z' was labeled uniformly with ^{15}N and specifically with $^{13}C_1$-leucine. The resulting sample gives HSQC spectra in which some amide crosspeaks are split in the nitrogen dimension (panel B); the same peaks are unsplit in sample labeled with ^{15}N only (panel A). This splitting arises because of coupling between the amide nitrogen and a neighboring ^{13}C-labeled carbonyl carbon. In the $^{13}C_1$-Leu specific labeling experiment, only leucines bear ^{13}C at their carbonyls; thus each split amide can be identified as lying C-terminal to a leucine.

^{15}N (ppm)
122.0
120.0
6.9
^1H (ppm)
B

(ppm)
122.0
120.0
6.9
^1H (ppm)
A

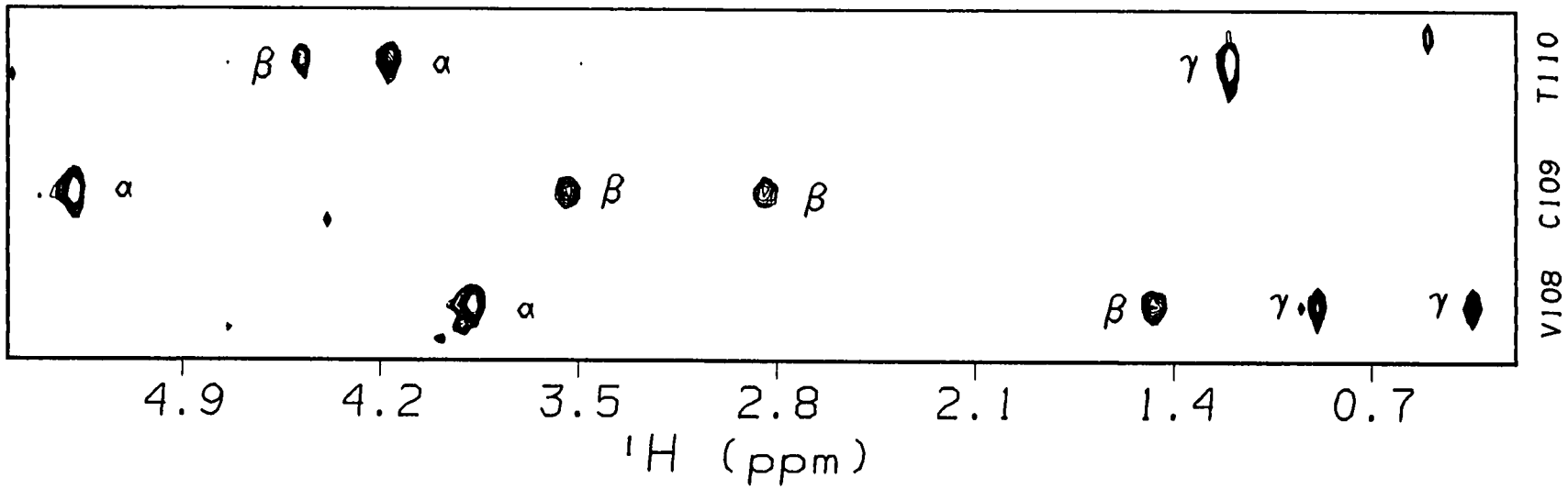
D. SIDECHAIN ASSIGNMENTS

Sidechain carbon and proton assignments for ADR1z' were obtained from 3D HCCH-TOCSY experiments on uniformly ^{13}C,^{15}N-labeled samples in D_2O (reviewed in (63)). These experiments rely on isotropic mixing of magnetization along the sidechain through large carbon-carbon couplings. In an ideal case, the result is a crosspeak between each carbon in a residue and every proton in that residue. Figure 30 shows examples of the data obtained. All sidechain carbon and proton resonances were assigned by this method, except in aromatic rings and sidechain NH_2 groups.

Proton assignments for aromatic spin systems in free ADR1z' were obtained by traditional means from homonuclear ^1H-^1H D_20 TOCSY and NOESY spectra (61). Asn and Gln sidechain NH_2 assignments were obtained from 3D ^{15}N-edited NOESY spectra, in which sidechain NH_2 protons NOE to Asn beta or Gln gamma protons. These assignments were confirmed in the HN(CO)CA spectrum, in which the sidechain NH_2 protons are weakly correlated with the Asn beta carbon or Gln gamma carbon through the sidechain carbonyl.

Figure 30. HCCH-TOCSY Spectrum of ADR1z'.

Sections of the 3D HCCH-TOCSY spectrum of ADR1z' were taken at the alpha carbon chemical shift and alpha proton chemical shift of several sequential residues. These strips demonstrate the correlation of an entire proton spin system to each alpha carbon. The success of this experiment is due to isotropic mixing of magnetization through the sidechain carbons. Similar spectra that rely on isotropic mixing of proton magnetization tend to display only partial spin systems for ADR1z'.


β
α
γ
α
β
β
α
β
γ
γ
V108 C109 T110
4.9
4.2
3.5
2.8
2.1
1.4
0.7
¹H (ppm)

E. PRESENTATION OF ADR1z' ASSIGNMENTS

The strategies described above resulted in complete backbone proton, nitrogen, and alpha carbon assignments, and complete sidechain assignments but for the carbons of aromatic rings. These assignments are presented in Table I. An HSQC spectrum of ADR1z' labeled with amide assignments is shown in Figure 31.

III. Assignment of DNA-Bound ADR1z'

A. INITIAL STRATEGY

An essentially identical assignment strategy was initially applied to the ADR1z'-DNA complex. An example of data from the HNCA experiment is shown in Figure 32, and is of reasonable quality compared to the free protein (Figure 26). Spectra for the specific labeling experiments are shown in Figure 33, and data from the HCCH-TOCSY in Figure 34.

Due to relatively weak signals from the arginine sidechain carbons in the ADR1z'-DNA HCCH-TOCSY experiment, a sample specifically labeled with $^{13}C_6$-arginine was prepared. 2D HCCH-

Table I. ADR1z' Assignments at pH 5.4, 303 K

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
N76		50.4, 51.0	36.4			
S77		56.6	61.7			
K78	123.2, 122.9	54.6				
I79	120.9, 121.2	59.1, 59.2	36.9	25.5 15.6	11.0	
N80	123.3, 123.0	51.1, 51.2	36.9			
K81	122.4	54.6	31.0	22.8	27.1	40.1
Q82	121.1	54.1	27.4	32.0		
L83	122.9	53.5	40.6	25.3	23.2 21.7	
D84	120.4	52.6	39.1			
K85	120.3	53.8	31.2	23.0	27.1	40.1
L86	124.5	51.3	39.9	25.1	23.3 21.5	
P87		61.5	30.2	25.6	48.6	
E88	120.8	55.9	28.2	34.3		
N89	117.6	51.7	36.3			
L90	121.4	53.5	40.6	25.3	23.2 21.7	
R91	120.3	54.6	28.6	25.3	41.4	
L92	121.7	53.5	40.6	25.3	23.2 21.7	

Table I. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
N93	118.5	51.5	36.9			
G94	108.5	43.5				
R95	119.9	54.0	29.2	25.3	41.4	
T96	117.4	58.1	67.6	19.7		
P97		61.8	30.2	25.6	49.1	
S98	115.6	56.7	61.7			
G99	110.6	43.2				
K100	120.3	54.4	31.0	22.8	27.1	40.1
L101	122.8	53.5	40.5	25.3	23.2 21.7	
R102	122.0	53.7	28.9	25.3	41.4	
S103	115.5	56.1	62.5			
F104	122.5	55.6	37.3			
V105	124.9	59.8	32.7	19.7 19.1		
C106	129.1	58.6	27.8			
E107	131.6	56.8	27.8	34.5		
V108	121.7	63.4	30.9	19.9 18.4		
C109	116.2	56.6	30.4			

Table I. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
T110	116.8	63.6	67.3	20.4		
R111	123.6	56.4	29.2	26.3	41.7	
A112	123.9	48.4	20.4			
F113	117.8	55.3	41.7			
A114	122.6	51.5	18.4			
R115	111.8	51.6	31.2	25.3	41.0	
Q116	126.8	57.5	25.8	31.7		
E117	117.9	57.7	26.6	34.0		
H118	117.5	54.9	29.7			
L119	121.8	55.8	38.4	25.6	24.8 20.5	
K120	117.7	58.1	30.0	23.3	27.4	39.9
R121	117.0	57.2	28.2	25.5	41.5	
H122	119.6	57.0	26.3			
Y123	120.6	59.6	36.3			
R124	116.2	56.4	27.9	25.5	41.4	
S125	113.8	58.7	61.2			
H126	118.5	53.1	26.8			

Table I. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
T127	113.2	60.6	67.8	19.7		
N128	119.8	51.8	36.8			
E129	120.0	55.2	28.4	34.3		
K130	121.2	51.9	31.2	22.7	27.4	40.1
P131		61.7	30.5	25.1	48.4	
Y132	116.1	53.9	37.1			
P133		60.3	30.5	24.8	49.1	
C134	123.9	58.1	27.6			
G135	115.5	44.3				
L136	122.7	53.5	40.9	25.3	23.5 21.2	
C137	118.5	56.8	28.6			
N138	114.8	52.1	36.1			
R139	121.3	54.7	29.7	26.1	40.9	
C140	117.4	55.3	28.2			
F141	120.2	55.3	41.7			
T142	110.0	61.5	67.5	20.7		
R143	115.3	52.6	34.5	25.3	41.7	

Table I. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
R144	124.7	57.5	27.8	25.8	41.2	
D145	115.4	55.0	37.1			
L146	119.1	54.8	39.7	25.5	24.1 20.2	
L147	122.5	56.0	38.6	26.1	24.6 20.7	
I148	119.3	63.1	35.3	27.1 15.3	10.4	
R149	118.7	57.5	28.2	25.5	41.5	
H150	118.8	57.1	26.3			
A151	122.5	53.8	15.9			
Q152	117.0	56.9	26.4	31.9		
K153	117.3	56.6	31.4	23.3	26.8	39.9
I154	114.8	58.0	35.8	25.6 15.3	8.2	
H155	116.2	51.5	27.6			
S156	114.4	57.3	60.7			
G157	106.9	42.9				
N158	119.3	51.5	37.1			
L159	121.3	53.1	41.2	25.3	23.3 22.0	
G160	108.8, 109.6	43.4				
E161	124.5, 124.9	55.9	29.4	34.5		

Table I. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
N76		4.33	3.00				
			2.96				
S77		4.52	3.92				
K78	8.42, 8.38	4.38					
I79	8.00, 8.07	4.16	1.87	1.48	0.87	0.93 (gme)	
				1.20			
N80	8.48, 8.45	4.74	2.86	112.2 (NH2)	7.62		
			2.77		6.88		
K81	8.33	4.30	1.87	1.46	1.71	3.02	
			1.79				
Q82	8.38	4.33	2.12	2.40	112.2 (NH2)	7.58	
			2.04			6.84	
L83	8.20	4.34	1.65	1.65	0.95		
					0.89		
D84	8.25	4.54	2.67				
K85	8.05	4.35	1.85	1.44	1.71	3.02	
			1.75				
L86	8.20	4.60	1.63	1.71	0.97		
			1.58		0.93		
P87		4.41	2.36	2.08	3.92		
			1.97		3.67		
E88	8.63	4.11	2.02	2.32			
N89	8.42	4.65	2.86	112.4 (NH2)	7.58		
			2.82		6.90		
L90	7.99	4.33	1.65	1.65	0.95		
					0.89		
R91	8.05	4.30	1.87	1.71	3.24		
			1.79				
L92	8.09	4.35	1.65	1.65	0.95		
					0.89		

Table I. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
N93	8.33	4.68	2.88	112.4 (NH2)	7.58		
			2.81		6.90		
G94	8.33	3.97					
R95	8.09	4.46	1.89	1.65	3.22		
			1.79				
T96	8.25	4.63	4.23	1.28			
P97		4.46	2.36	2.10	3.90		
			1.99	2.02	3.76		
S98	8.37	4.43	3.92				
G99	8.38	3.97					
K100	8.05	4.35	1.87	1.46	1.71	3.02	
			1.79				
L101	8.15	4.31	1.65	1.65	0.95		
					0.89		
R102	8.28	4.27	1.65	1.56	3.14		
				1.44			
S103	7.79	4.41	3.53				
F104	8.69	4.71	3.06		7.17	7.37	
			2.92				
V105	8.52	4.60	1.89	0.83			
				0.75			
C106	9.19	4.49	3.43				
			2.86				
E107	9.72	4.24	2.24	2.47			
			2.12				
V108	9.12	3.86	1.48	0.89			
				0.34			
C109	8.33	5.25	3.53				
			2.84				

Table I. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
T110	8.14	4.16	4.47	1.20			
R111	8.12	4.00	1.50	1.73	3.10		
			1.28		2.92		
A112	7.94	5.12	1.24				
F113	8.91	4.74	3.45		7.29	6.79	6.10
			2.84				
A114	9.37	4.68	1.65				
R115	7.28	4.60	1.79	1.48	3.10		
			0.83		3.04		
Q116	8.57	2.99	1.61	2.04	110.8 (NH2)	7.32	
			1.34	1.97		6.88	
E117	9.63	3.92	1.93	2.26			
H118	6.82	4.41	3.39		6.90	7.96	
			3.25		240.6 (Nd)	170.2 (Ne)	
L119	6.94	3.20	1.97	1.42	1.09		
			1.24		0.87		
K120	8.12	3.97	1.85	1.56	1.69	2.94	
				1.42			
R121	7.42	4.02	1.87	1.79	3.22		
				1.67			
H122	7.69	4.27	3.12		6.90	7.98	
			2.81		173.6 (Nd)	216.9 (Ne)	
Y123	8.61	4.08	3.20		7.33	7.06	
R124	7.28	4.16	1.97	1.91	3.27		
			1.91	1.77			
S125	7.90	4.22	3.78				
			3.68				
H126	7.37	5.01	3.35		6.57	8.11	
			3.08		170.5 (Nd)	211.9 (Ne)	

Table I. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
T127	7.72	4.33	4.29	1.24			
N128	8.45	4.68	2.88	111.8 (NH2)	7.49		
			2.79		6.85		
E129	8.17	4.27	2.08	2.32			
			2.00				
K130	8.09	4.52	1.63	1.32	1.60	2.92	
			1.54	1.26	1.54		
P131		4.38	2.06	1.91	3.66		
			1.54	1.85	3.57		
Y132	8.12	4.90	2.98		7.02	6.89	
			2.81				
P133		4.90	2.24	2.16	3.84		
			1.95	1.95	3.68		
C134	8.72	4.52	3.43				
			2.79				
G135	8.91	4.25					
		3.96					
L136	9.19	4.49	1.05	1.28	0.54		
			0.48		0.74		
C137	7.95	4.98	3.37				
			3.29				
N138	8.19	4.82	2.98	111.6 (NH2)	7.53		
			2.94		6.83		
R139	8.30	4.19	1.40	1.81	3.14		
			1.30	1.65			
C140	7.95	4.90	2.65				
			2.59				
F141	8.12	4.93	3.31		7.36	6.85	6.13
			2.96				
T142	8.97	4.43	4.52	1.40			
R143	7.02	4.76	2.16	1.77	3.33		
			1.61				

Table I. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
R144	8.50	3.12	1.59	1.44	3.14		
			1.26	1.34			
D145	8.66	4.13	2.69				
			2.59				
L146	6.88	4.02	1.97	1.65	1.15		
			1.58		0.89		
L147	6.97	3.29	2.10	1.56	1.07		
			1.44		1.13		
I148	8.13	3.61	1.83	1.56	0.77	0.87 (gme)	
				1.24			
R149	7.69	4.02	1.85	1.81	3.22		
			1.81	1.69			
H150	7.54	4.30	3.24		7.18	7.99	
			2.84		172.9 (Nd)	219.4 (Ne)	
A151	8.87	3.81	1.58				
Q152	8.27	3.81	2.26	2.59	110.8 (NH2)	7.37	
			2.06	2.40		6.73	
K153	7.77	4.16	1.91	1.54	1.69	2.98	
			1.85	1.44			
I154	8.19	4.13	0.97	1.32	0.72	0.33 (gme)	
				0.99			
H155	7.30	5.28	3.47		6.54	7.93	
			3.27		169.8 (Nd)	214.9 (Ne)	
S156	8.23	4.27	4.08				
			4.00				
G157	8.47	4.13					
		3.48					
N158	7.89	4.74	2.86	112.8 (NH2)	7.49		
			2.77		6.98		
L159	8.30	4.49	1.67	1.67	0.93		
					0.89		
G160	8.34, 8.38	3.86					
E161	7.82, 7.79	4.16	2.08	2.24			
			1.91				

Figure 31. ADR1z' Backbone Amide Assignments.

The ^{15}N-HSQC of ADR1z' is labeled with assignments for the backbone amide protons. The dotted lines connect peaks for the sidechain NH_2 groups of Asn and Gln residues.

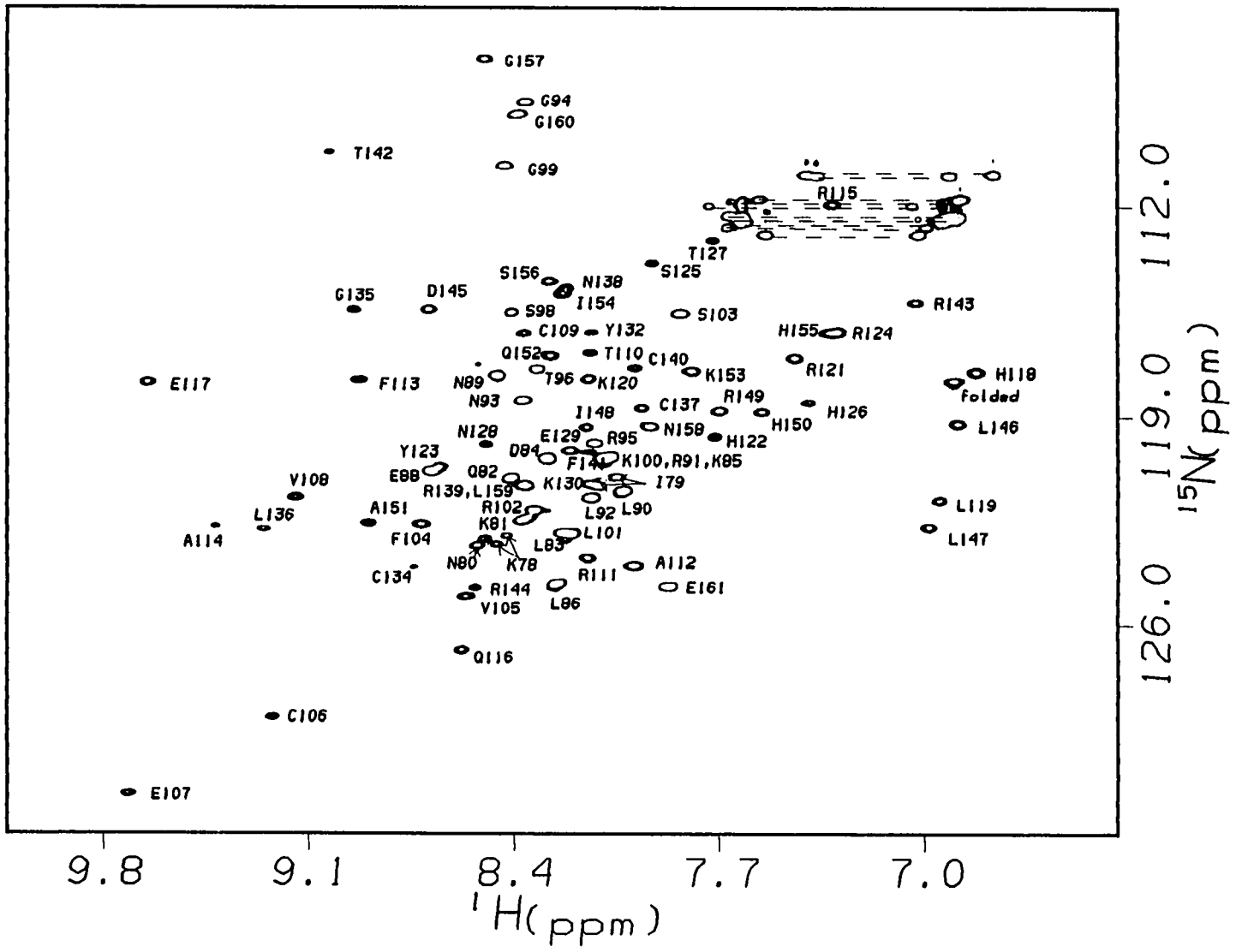

G157
G94
G160
T142
G99
R115
T127
S156
N138
S125
G135
D145
I154
R143
S98
S103
C109
Y132
H155
R124
Q152
T110
C140
R121
H118
E117
F113
N89
T96
K120
K153
folded
N93
R149
I148
C137
H126
N128
E129
R95
N158
H150
L146
H122
Y123
D84
K100,R91,K85
E88
Q82
I79
V108
R139,L159
K130
L119
A151
R102
L92
L90
L136
K81
L101
L147
A114
F104
L83
N80
K78
R111
A112
C134
E161
R144
L86
V105
Q116
C106
E107
9.8
9.1
8.4
7.7
7.0
^{1}H(ppm)
126.0
119.0
112.0
^{15}N(ppm)

Figure 32. HNCA Spectrum of ADR1z'-DNA.

Sections of the 3D HNCA spectrum of ADR1z'-DNA taken at the amide proton chemical shift and amide nitrogen chemical shift of various sequential residues are shown. The amide of each residue i is correlated to the alpha carbon of residues i and i-1. Where the alpha carbon chemical shifts are unique, sequential assignments can be made.

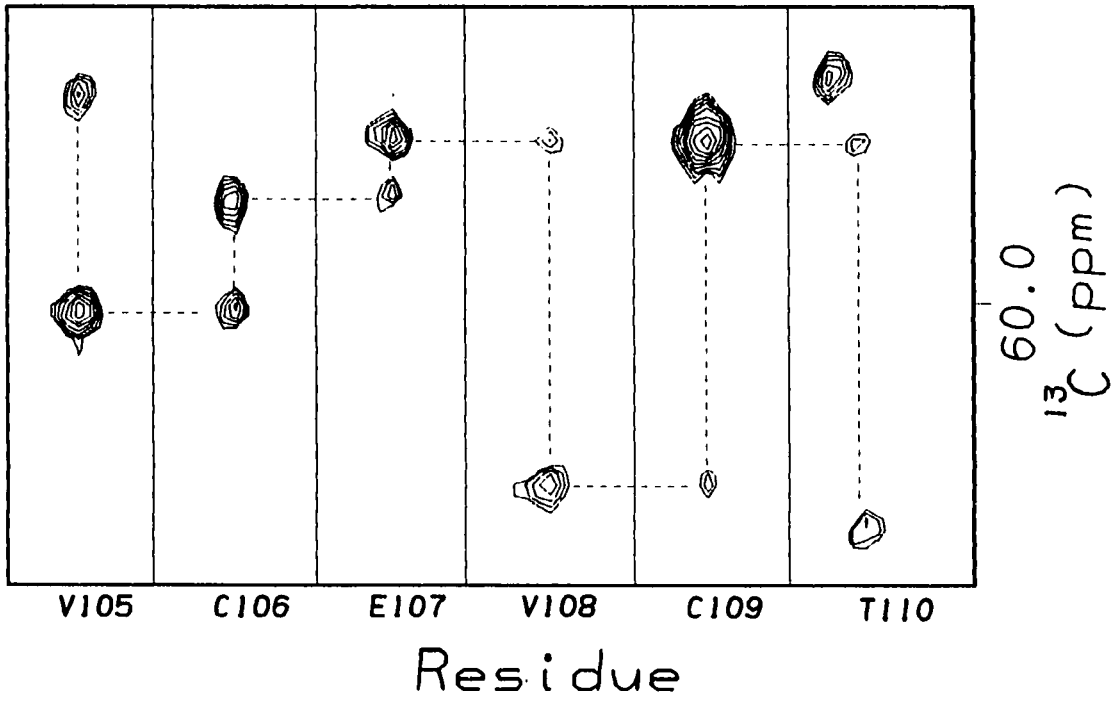

V105
C106
E107
V108
C109
T110
Residue
60.0
^{13}C (ppm)

Figure 33. Specific ^{13}C Labeling of ADR1z'-DNA: $^{13}C_1$-Leucine.

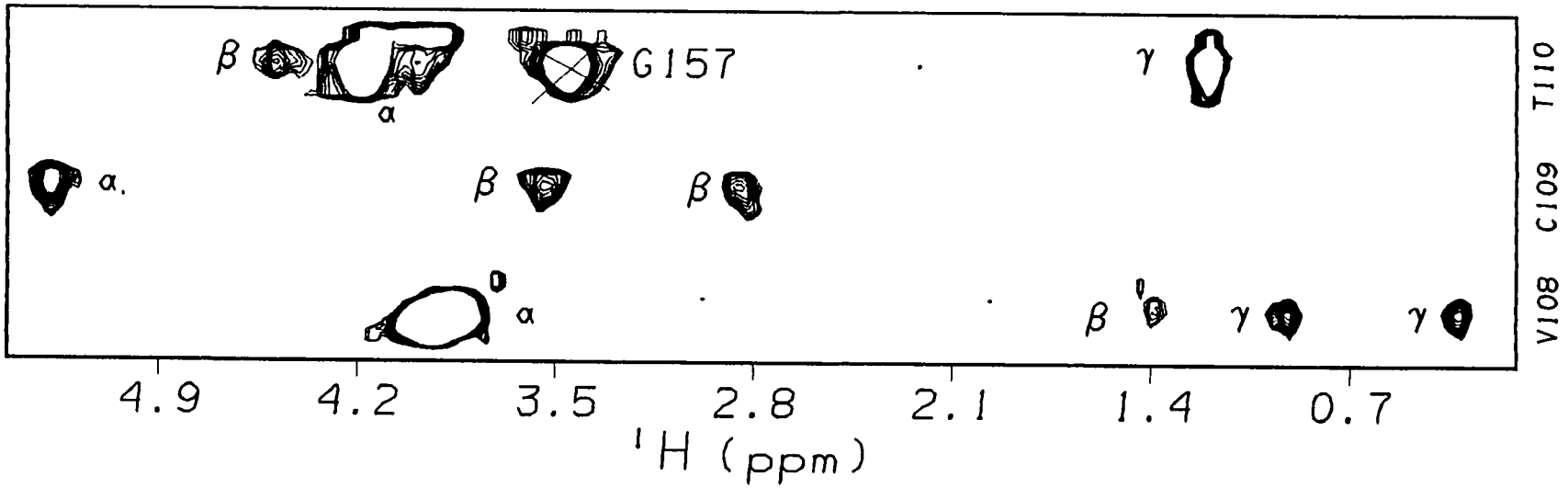
As in Figure 29, only those amides that neighbor leucine are split in this spectrum, allowing identification of the C-terminal neighbors of leucines.

^{15}N (ppm)
120.6
119.7
118.8
^1H (ppm)
8.6
8.8
B

(ppm)
120.6
119.7
118.8
^1H (ppm)
8.6
8.8
A

Figure 34. HCCH-TOCSY Spectrum of ADR1z'-DNA.

As in Figure 30, sections of a 3D HCCH-TOCSY are shown for several residues. The same spin systems are apparent in this spectrum of the bound protein as were for the free protein.


β
α
G157
γ
α.
β
β
α
β
γ
γ
T110
C109
V108
4.9
4.2
3.5
2.8
2.1
1.4
0.7
¹H (ppm)

TOCSY and HCCH-COSY experiments on this sample provided firmer assignments for these residues.

Sidechain NH_2 assignments were again obtained from 3D ^{15}N-edited NOESY experiments, now in combination with the CBCA(CO)NH. The latter provides weak coupling between sidechain NH_2 protons and the alpha and beta carbons of Asn or the beta and gamma carbons of Gln through the sidechain carbonyl.

B. ADDITIONAL TRIPLE RESONANCE EXPERIMENTS

The assignment strategy designed for free ADR1z' proved largely successful for the more challenging process of assigning the polypeptide when bound to DNA. Due to poorer signal from the bound ^{15}N-glutamate sample, however, the differential specific labeling results were not as quantitative as those for the free protein. This resulted in a number of remaining ambiguities in backbone assignment.

These ambiguities were resolved by the application of the HNCACB, CBCA(CO)NH triple resonance pair, which provides connectivities between neighboring C^{β} resonances as well as C^{α} resonances (Figure 35) (reviewed in (63)). Thus, at backbone positions where alpha carbon degeneracy leads to uncertainty in

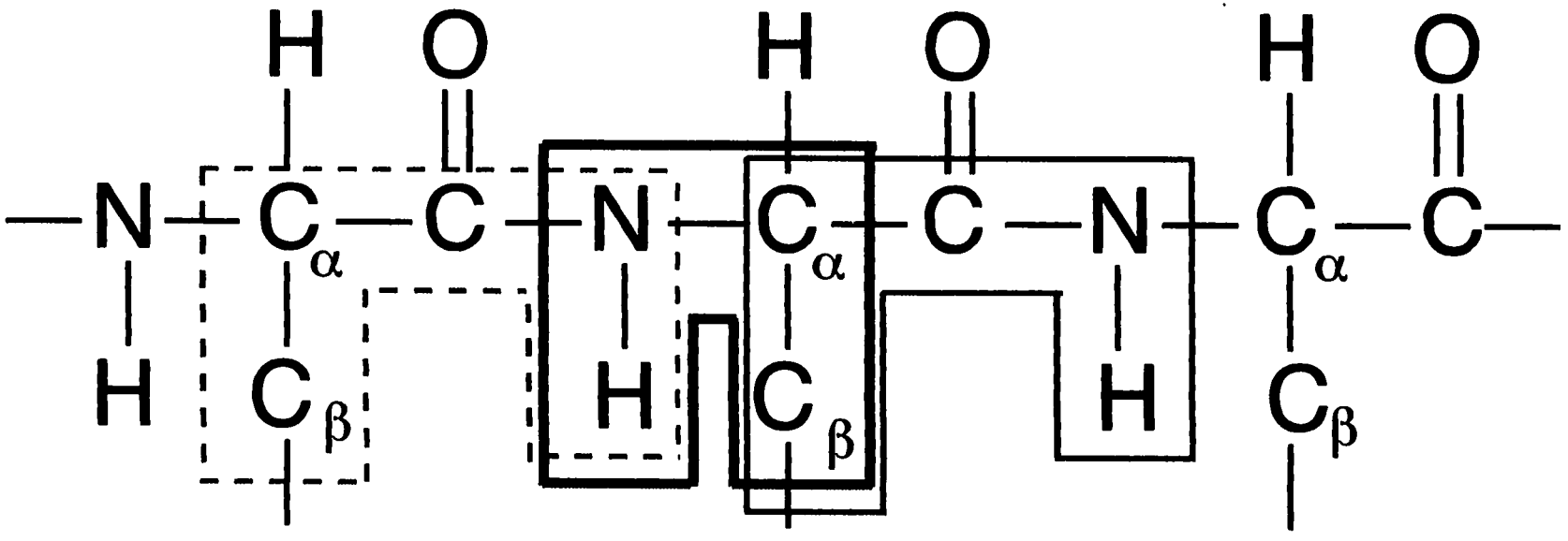
assignment, the ambiguity may be resolved if the beta carbons for these residues are not also degenerate. Examples of the data for ADR1z'-DNA are shown in Figure 36. These data resolved the remaining uncertainties in the backbone assignments of bound ADR1z'.

C. PRESENTATION OF ADR1z'-DNA ASSIGNMENTS

Nearly complete carbon, nitrogen, and proton assignments are presented for ADR1z' bound to DNA in Table II. Backbone carbonyl carbon and aromatic sidechain carbon and proton resonances remain unassigned. Other missing assignments are generally due to unavoidable overlap, particularly for the sidechain spin systems of K130, K100, and L101, which lie adjacent to very intense random coil spin systems. For several residues it is not possible to distinguish among the beta, gamma, and delta protons; these assignments are presented as groups bounded by brackets. The amide assignments of bound ADR1z' are labeled on the HSQC spectrum shown in Figure 37.

Figure 35. Pattern of Connectivities in the HNCACB and CBCA(CO)NH Experiments.

Just as the HNCA, HN(CO)CA pair correlates amide groups to both neighboring alpha carbons, the triple resonance HNCACB and CBCA(CO)NH experiments correlate amide groups to neighboring alpha and beta carbons.



HNCACB

CBCA(CO)NH

Figure 36. HNCACB and CBCA(CO)NH Spectra of ADR1z'-DNA.

Sections of the 3D HNCACB and CBCA(CO)NH spectra of ADR1z'-DNA are taken at the amide proton chemical shift and amide nitrogen chemical shift of several sequential residues. The HNCACB correlates the amide group of residue i to the alpha and beta carbon of residue i. The beta carbon peak is of opposite sign to the alpha carbon peak, allowing its identification. The CBCA(CO)NH correlates the amide group of residue i to the alpha and beta carbons of residue i-1. Due to folding of peaks in the carbon dimension in these spectra (to reduce spectral width) the strips are not comparable placed side-by-side. However, examination of the carbon chemical shifts of the peaks allows determination of sequential connectivities, as shown in the expanded regions of the spectra plotted below.

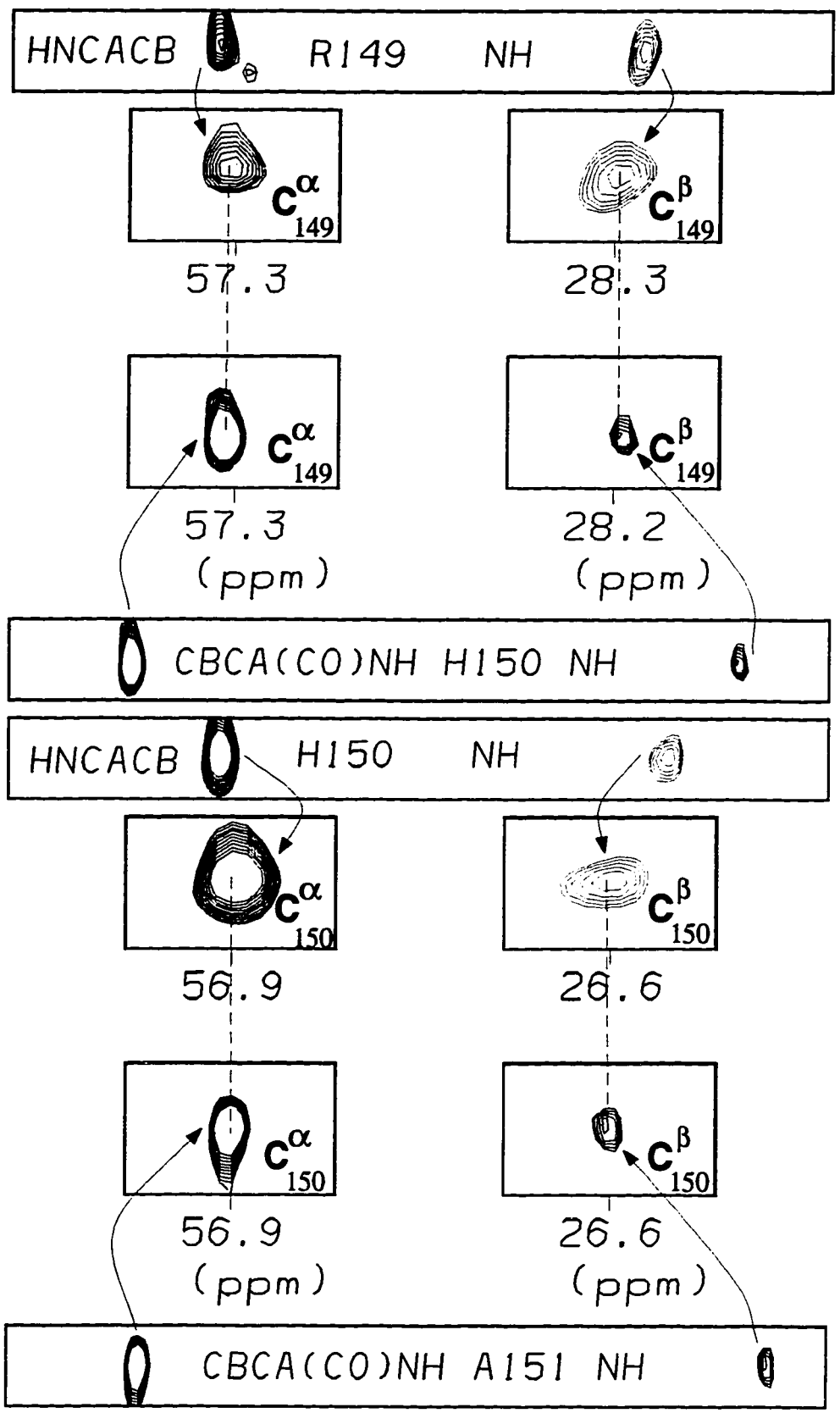

HNCACB R149 NH
C^{α}_{149}
C^{β}_{149}
57.3
28.3
C^{α}_{149}
C^{β}_{149}
57.3 (ppm)
28.2 (ppm)
CBCA(CO)NH H150 NH
HNCACB H150 NH
C^{α}_{150}
C^{β}_{150}
56.9
26.6
C^{α}_{150}
C^{β}_{150}
56.9 (ppm)
26.6 (ppm)
CBCA(CO)NH A151 NH
^{13}C

Table II. ADR1z'-DNA Assignments at pH 7.0, 310 K

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
N76		51.6	38.6			
S77		56.6	62.0			
K78	122.8, 122.5	54.7	31.0	23.1	27.2	40.1
I79	120.6, 120.4	59.4	36.4	25.5 15.5	11.1	
N80	122.5, 122.3	51.2	36.9			
K81	121.7	54.6	31.0	22.9	27.2	40.1
Q82	120.3	54.0	27.3	32.1		
L83	122.4	53.6	40.6	25.3	23.1 21.6	
D84	120.0	52.7	39.4			
K85	119.5	53.9	31.0	22.9	27.0	40.1
L86	124.1	51.1	39.9	25.2	23.5 22.4	
P87		60.8	30.4	25.7		
E88	121.1	57.9	27.6	34.5		
N89	114.3	53.6	35.9			
L90	116.1	52.6	40.9			
R91	118.2	53.5	29.9	25.3	41.4	
L92	120.7	55.0	39.7		23.7 20.8	

Table II. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
N93	118.6	54.1	34.7			
G94	106.3	42.7				
R95	115.5	52.1	32.8	25.7	41.6	
T96	113.9	58.7	65.4	21.2		
P97		61.4	30.1	25.7	48.8	
S98	108.6	57.2	61.4			
G99	110.1	43.6				
K100	120.6	53.9	30.0			
L101	123.4	53.6	41.5			
R102	123.4	53.6	29.6	27.0	40.4	
S103	119.0	56.8	62.7			
F104	121.0	55.7	37.5			
V105	125.8	60.1	32.8	20.9 18.5		
C106	128.6	57.9	27.3			
E107	131.3	56.6	27.6	34.5		
V108	121.2	63.5	30.6	20.1 18.3		
C109	116.0	56.7	30.3			

Table II. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
T110	117.3	64.3	67.7	20.3		
R111	123.3	56.7	29.5	27.7	42.1	
A112	122.4	47.5	21.3			
F113	117.6	55.5	42.5			
A114	124.0	51.4	18.8			
R115	110.9	52.1	33.0	26.0	41.7	
Q116	126.2	57.5	25.8	31.9		
E117	117.9	57.1	28.5	34.5		
H118	114.7	54.0	30.4			
L119	120.4	55.9	38.7		24.9 20.7	
K120	118.9	58.3	30.4	23.5	27.6	
R121	115.3	57.6	29.1	24.0	44.1	
H122	119.4	57.7	26.6			
Y123	120.5	59.9	36.6			
R124	115.6	57.0	26.6	28.1	42.9	
S125	115.6	59.4	60.8			
H126	117.2	54.1	26.6			

Table II. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
T127	107.3	60.0	68.3	20.3		
N128	116.6	52.7	35.9			
E129	119.6	55.3	29.1	34.6		
K130	123.2	51.5	32.5			
P131		61.8	29.7	25.8		
Y132	117.1	53.9	36.4			
P133		60.2	30.7	24.4	49.3	
C134	123.5	58.2	27.2			
G135	115.4	44.1				
L136	122.8	53.6	40.9	25.5	23.6 21.1	
C137	118.7	56.8	28.6			
N138	114.9	52.2	36.0			
R139	121.9	54.9	29.4	25.9	40.9	
C140	116.8	54.9	29.1			
F141	117.2	55.2	42.2			
T142	110.1	62.4	68.0	21.9		
R143	114.0	52.7	35.1	24.2	42.6	

Table II. (continued)

residue	amide nitrogen	alpha carbon	beta carbon	gamma carbon	delta carbon	epsilon carbon
R144	123.3	56.8	28.2	25.8	41.3	
D145	112.7	53.9	35.3			
L146	120.8	54.8	39.1			
L147	122.9	56.2	38.3	26.0	24.7 20.8	
I148	119.1	61.8	34.4	15.7		
R149	119.7	57.5	28.2	25.5	41.3	
H150	118.0	57.0	26.6			
A151	122.5	54.0	16.0			
Q152	116.8	56.9	26.3	31.9		
K153	117.2	56.4	31.0	23.2	27.0	40.1
I154	115.1	57.9	35.5	25.5 15.3	8.0	
H155	116.2	51.6	27.5			
S156	114.3	57.4	60.8			
G157	107.3	43.3				
N158	119.2	51.7	37.5			
L159	121.1	53.0	40.9	25.5	23.2 21.9	
G160	108.8	43.5				
E161	125.0	56.1	29.0	34.9		

Table II. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
N76		4.05	2.82	112.2 (NH2)	7.66		
					6.96		
S77		4.46	3.92				
K78	8.40, 8.39	4.35	1.89	1.48	1.71	3.04	
			1.81				
I79	8.00, 7.96	4.16	1.87	1.48	0.87	0.91 (gme)	
				1.20			
N80	8.41, 8.38	4.74	2.88	112.1 (NH2)	7.58		
			2.78		6.90		
K81	8.26	4.30	1.89	1.44	1.71	3.02	
			1.79				
Q82	8.30	4.33	2.14	2.39	111.9 (NH2)	7.55	
			2.06			6.83	
L83	8.14	4.33	1.67	1.67	0.97		
					0.91		
D84	8.25	4.57	2.67				
K85	7.93	4.35	1.85	1.42	1.67	3.02	
			1.73				
L86	8.16	4.60	1.65	1.71	0.93		
			1.53		0.93		
P87		4.44	2.43	2.12	4.03		
			1.98	2.02	3.59		
E88	8.80	3.75	2.06	2.35			
N89	8.53	4.63	2.98	114.5 (NH2)	7.69		
			2.76		7.00		
L90	7.56	4.27	1.61	1.73	0.91		
					0.83		
R91	7.11	4.40	1.81	1.65	3.19		
L92	8.52	3.59	1.59	1.50	0.89		
			1.40		0.85		

Table II. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
N93	8.73	4.71	3.15	111.9 (NH2)	7.48		
			3.12		6.83		
G94	8.49	4.37					
		4.00					
R95	7.90	5.61	1.75	1.67	3.04		
				1.50			
T96	8.98	4.79	4.76	1.48			
P97		4.41	2.39	2.06	3.90		
			2.00	2.00	3.62		
S98	7.91	4.60	4.11				
			3.96				
G99	8.50	4.39					
		3.66					
K100	7.45	4.41	[2.04	1.61	1.74]	3.02	
			[1.89	1.46	]		
L101	8.58	4.29	[1.74	1.46]	0.86		
			[1.61	]	0.86		
R102	8.54	3.94	1.91	1.75	3.02		
				1.48			
S103	7.24	4.46	3.37				
			3.23				
F104	8.41	4.87	3.08				
			2.86				
V105	8.83	4.44	1.96	0.93			
				0.87			
C106	9.23	4.68	3.59				
			2.86				
E107	9.48	4.27	2.30	2.47			
			2.20				
V108	8.90	3.84	1.40	0.91			
				0.32			
C109	8.33	5.28	3.53				
			2.84				

Table II. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
T110	8.21	4.16	4.48	1.20			
R111	8.18	3.86	1.61	1.69	3.16		
			1.34	1.20	2.98		
A112	8.14	5.56	1.01				
F113	9.28	4.90	3.96				
			2.88				
A114	10.78	5.12	1.67				
R115	6.93	4.60	[1.69	1.65]	2.91		
			[1.55	1.61]	2.87		
Q116	8.54	3.26	1.79	2.06	110.2 (NH2)	7.19	
			1.59	1.94		6.81	
E117	8.19	3.97	2.04	2.32			
			1.91				
H118	6.44	3.84	3.49		7.00	7.73	
			3.32		227.5 (Nd)	181.0 (Ne)	
L119	7.05	3.48	2.08	1.57	1.12		
			1.40		1.07		
K120	8.57	4.03	1.91	1.61	1.75	2.98	
				1.48			
R121	7.64	4.13	1.73	2.31	2.99		
			1.37	1.95			
H122	7.75	4.30	3.21		6.54	8.01	
			3.14		176.9 (Nd)	215.9 (Ne)	
Y123	9.24	4.11	3.41				
			3.25				
R124	7.35	4.33	[1.96	1.73]	3.23		
			[1.86	]	3.37		
S125	8.13	4.27	3.86				
			3.64				
H126	7.54	4.85	3.25		6.71	8.07	
			3.06		169.2 (nd)	212.7 (ne)	

Table II. (continued)

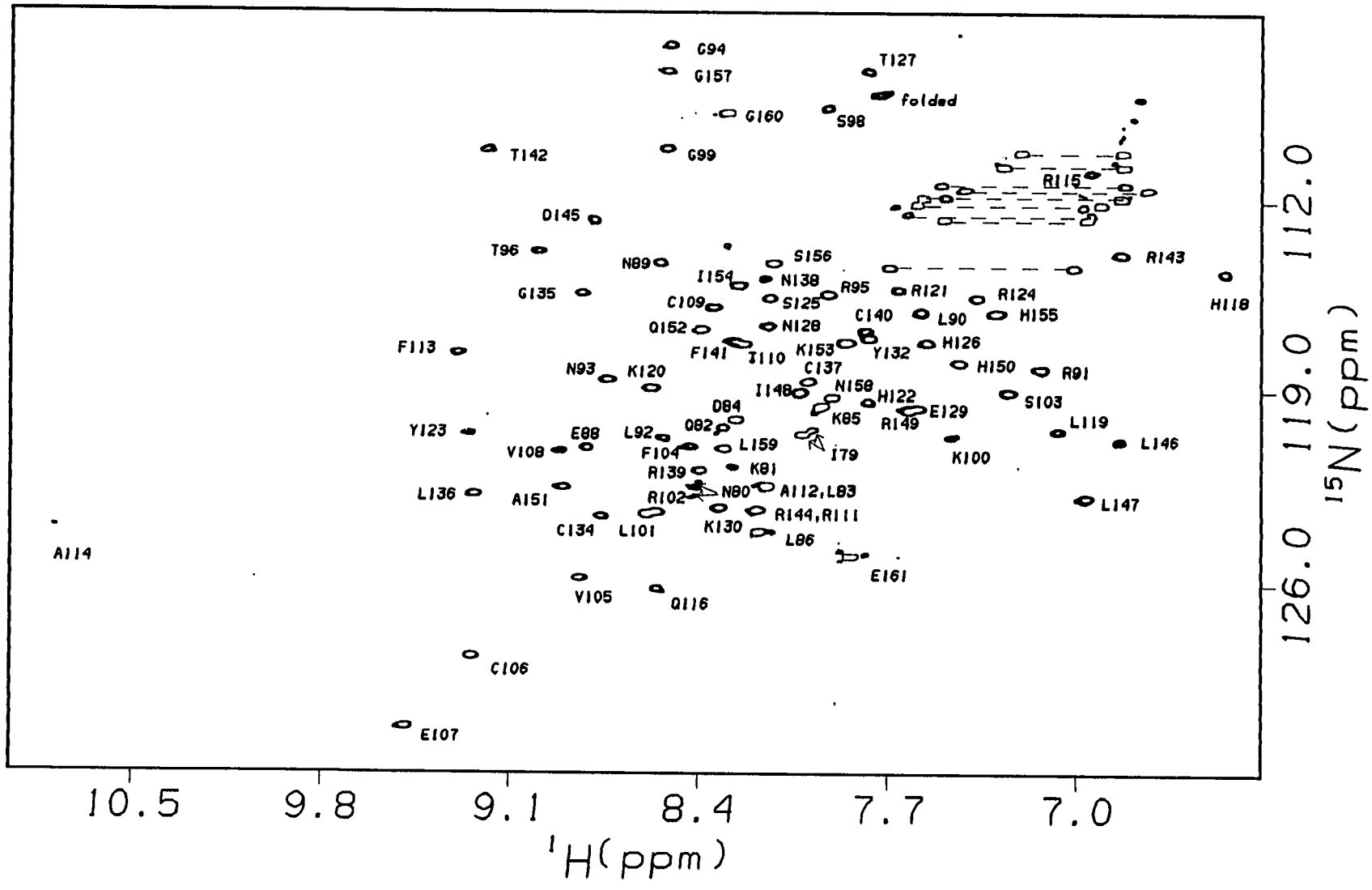
residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
T127	7.76	4.30	4.44	1.20			
N128	8.13	4.46	3.08	111.5 (NH2)	7.40		
			2.86		6.73		
E129	7.56	4.05	2.04	2.32			
			1.91				
K130	8.31	4.52	[1.65	1.23	1.49]	2.88	
			[1.51	1.17	]		
P131		4.44	2.08	2.03	3.90		
			1.53	2.01	3.78		
Y132	7.75	4.98	3.00				
			2.82				
P133		5.01	2.31	2.06	3.90		
				1.94	3.72		
C134	8.74	4.55	3.43				
			2.82				
G135	8.82	4.27					
		3.98					
L136	9.22	4.49	1.03	1.28	0.54		
			0.50		0.73		
C137	7.98	5.01	3.39				
			3.29				
N138	8.14	4.82	3.03	111.4 (NH2)	7.49		
			2.97		6.81		
R139	8.38	4.16	1.44	1.77	3.14		
			1.30	1.59			
C140	7.76	5.04	2.84				
			2.61				
F141	8.25	4.93	3.31				
			2.96				
T142	9.17	4.33	4.50	1.50			
R143	6.83	4.66	2.41	1.83	3.06		
			1.50	1.73	2.87		

Table II. (continued)

residue	amide proton	alpha proton	beta proton	gamma proton	delta proton	epsilon proton	zeta proton
R144	8.18	3.12	1.55	1.63	3.23		
			1.43	1.50			
D145	8.78	4.27	2.63				
L146	6.83	4.11	[1.87	1.74]	1.30		
			[1.51	]	0.93		
L147	6.95	3.23	2.08	1.55	1.10		
			1.44		1.10		
I148	8.01	3.70	1.92	1.61	0.76	0.95 (gme)	
				1.44			
R149	7.61	4.00	1.91	1.77	3.19		
			1.77	1.63			
H150	7.43	4.33	3.23		7.17	8.00	
			2.90		173.2 (Nd)	219.6 (Ne)	
A151	8.89	3.81	1.59				
Q152	8.38	3.81	2.26	2.59	110.7 (NH2)	7.26	
			2.06	2.39		6.81	
K153	7.84	4.16	1.96	1.53	1.71	3.02	
			1.89				
I154	8.24	4.14	0.97	1.36	0.74	0.32 (gme)	
				1.03			
H155	7.29	5.28	3.47		6.55	7.91	
			3.31		170.7 (Nd)	214.7 (Ne)	
S156	8.11	4.27	4.09				
			4.00				
G157	8.50	4.17					
		3.47					
N158	7.89	4.74	2.88	112.6 (NH2)	7.48		
			2.80		6.95		
L159	8.29	4.49	1.71	1.67	0.95		
			1.67		0.91		
G160	8.29	3.96					
		3.88					
E161	7.83	4.16	2.08	2.22			
			1.92				

Figure 37. Backbone Amide Assignments of ADR1z'-DNA.

An ^{15}N-HSQC spectrum of ADR1z'-DNA is labeled with backbone amide assignments. The dotted lines connect peaks for sidechain NH_2 groups of Asn and Gln residues.


G94
T127
G157
folded
G160
S98
T142
G99
R115
D145
T96
N89
S156
R143
I154
N138
R95
R121
R124
H118
G135
C109
S125
C140
L90
H155
Q152
N128
F113
F141
I110
K153
Y132
H126
N93
K120
C137
H150
R91
I148
N158
H122
S103
D84
K85
R149
E129
Y123
E88
L92
Q82
L159
I79
L119
L146
V108
F104
K100
R139
K81
L136
A151
R102
N80
A112,L83
L147
R144,R111
C134
L101
K130
L86
A114
E161
V105
Q116
C106
E107
10.5
9.8
9.1
8.4
7.7
7.0
^{1}H(ppm)
126.0
119.0
112.0
^{15}N(ppm)

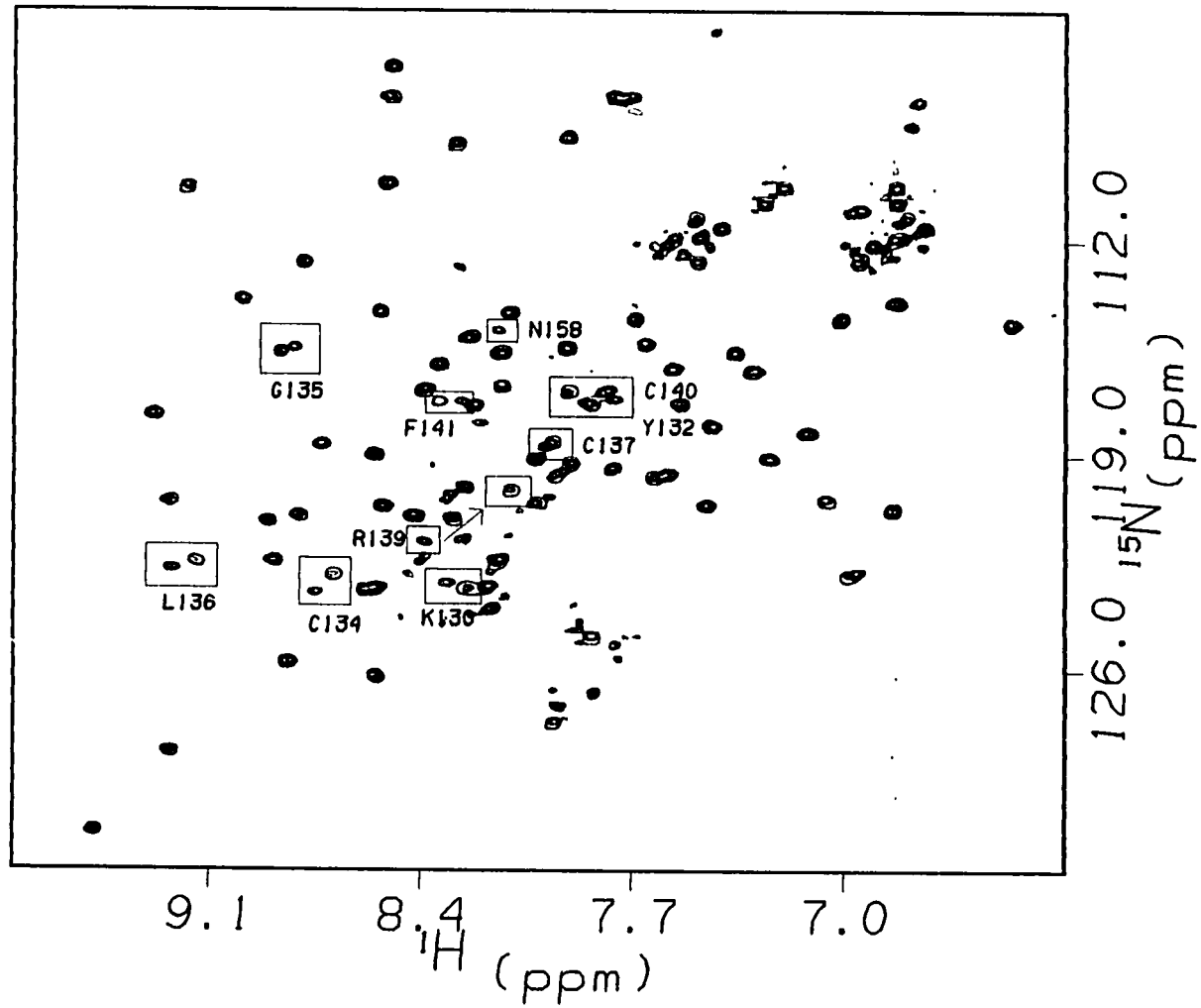
D. MINOR FORM OF ADR1z'-DNA

The single finger ADR1a peptide displays a minor conformation in solution that is detectable in NMR spectra as a doubling of peaks (60). This doubling is most pronounced for residues in the Cys-Cys loop. The free form of ADR1z' does not display such doublings. The only peaks that appear doubled in the reduced form of the polypeptide are from the extreme N-terminus, due to heterogenous post-translational processing, and at the two C-terminal residues, well away from the Cys-Cys loop.

The ADR1z'-DNA complex, when freshly prepared, also lacks any doubling analogous to that seen in the ADR1a spectra. However, after a number of months in solution, the complex begins to display a second set of lines for residues from K130 to F141 in the Cys-Cys loop of finger 2, exactly the region where ADR1a is doubled. Just as in ADR1a, these new lines are in slow exchange with the original lines, indicating the presence of two populations that exchange more slowly than the NMR timescale. This doubling persists even in the presence of added DTT. The spectra are otherwise identical in quality, peak shape, and chemical shift to the freshly made complex (Figure 38). The molecular mass of ADR1z' from samples that display the minor form is identical to the mass of fresh ADR1z'.

Figure 38. The Minor Form of ADR1z'-DNA.

ADR1z'-DNA exhibits limited peak doubling in the region of K130 - F141 that appears after months in solution. The new minor form is in slow exchange with the major form, as can be seen in the HSQC spectrum of a sample that has both forms (green). After further time in solution, the samples eventually convert entirely to the second form, as shown in the red spectrum. The original and second-form peaks for each residue from K130 to F141 are boxed.


N158
G135
C140
F141
Y132
C137
R139
L136
C134
K130
9.1
8.4
7.7
7.0
^1H (ppm)
112.0
119.0
126.0
^{15}N (ppm)

Unlike ADR1a, as the months progress, the second form of ADR1z'-DNA becomes increasingly prevalent, until it reaches 100% occupancy (see Figure 38). Backbone proton and nitrogen assignments for the second form are presented in Table III. While partial or complete spectral doubling have been observed in several ^{15}N-labeled samples, no spectra of ^{13}C-labeled samples have exhibited enough of the second form to obtain carbon assignments.

The chemical shift changes for ADR1z'-DNA are smaller than those observed for ADR1a. As is the case with ADR1a, it is not clear what structural change the minor form represents; it is also not clear that the change in ADR1z'-DNA is equivalent to that in the single finger peptide. However, the second conformational form appears to be the thermodynamically favored form for ADR1z'-DNA, although it apparently is not kinetically favored during the formation of the complex. That both ADR1a and ADR1z'-DNA exhibit conformational doubling in the Cys-Cys loop indicates that this region is conformationally flexible.

IV. Assignment of 14mer DNA Imino Protons, Free and Bound to ADR1z'

The imino protons of the 14mer DNA sequence used in this study have been assigned using NOESY methods. The assignments

Table III. Minor Form Assignments for ADR1z'-DNA

Residue	minor form NH [a]		minor form amide nitrogen	
K130	8.24	ppm	123.4	ppm
Y132	7.91		116.8	
C134	8.69		122.9	
G135	8.86		115.6	
L136	9.14		122.5	
C137	7.96		118.5	
N138	?		?	
R139	8.10		120.2	
C140	7.80		116.9	
F141	8.34		117.2	

a. Minor form assignments were obtained at pH 7.0, 37 °C.

for the free DNA were completed by Shamus Young and Lawrence Schaufler. The assignments for 14mer bound to ADR1z' were obtained by a 2D NOESY "imino walk". This strategy takes advantage of the fact that each imino proton in the sequence is within NOE range of both neighboring imino protons, and that iminos from thymidine residues fall downfield of iminos from guanine residues. The imino proton walk for 14mer is shown in Figure 39. No protein lines interfere in this region of the spectrum, due to the extreme downfield shift of imino protons.

1D spectra of the imino region of free and bound 14mer are shown in Figure 40, along with their assignments. Each base pair contains one imino proton, from either the G or the T base. The iminos from the basepairs at the ends of the DNA exchange very rapidly with solvent, and are never observed in these spectra. At room temperature, then, 12 imino protons are detected for the 14mer. At the higher temperature of the NOESY, only ten are detected, as now two bases at each end exchange efficiently with solvent. The imino assignments of 14mer free and bound to ADR1z' are cataloged in Table IV.

Figure 39. NOESY of the DNA Iminos of ADR1z'-DNA.

The imino region of a 150 msec NOESY spectrum of ADR1z'-DNA in water is plotted. Each imino proton gives NOEs to the two neighboring iminos, allowing sequential assignment as outlined on the figure.

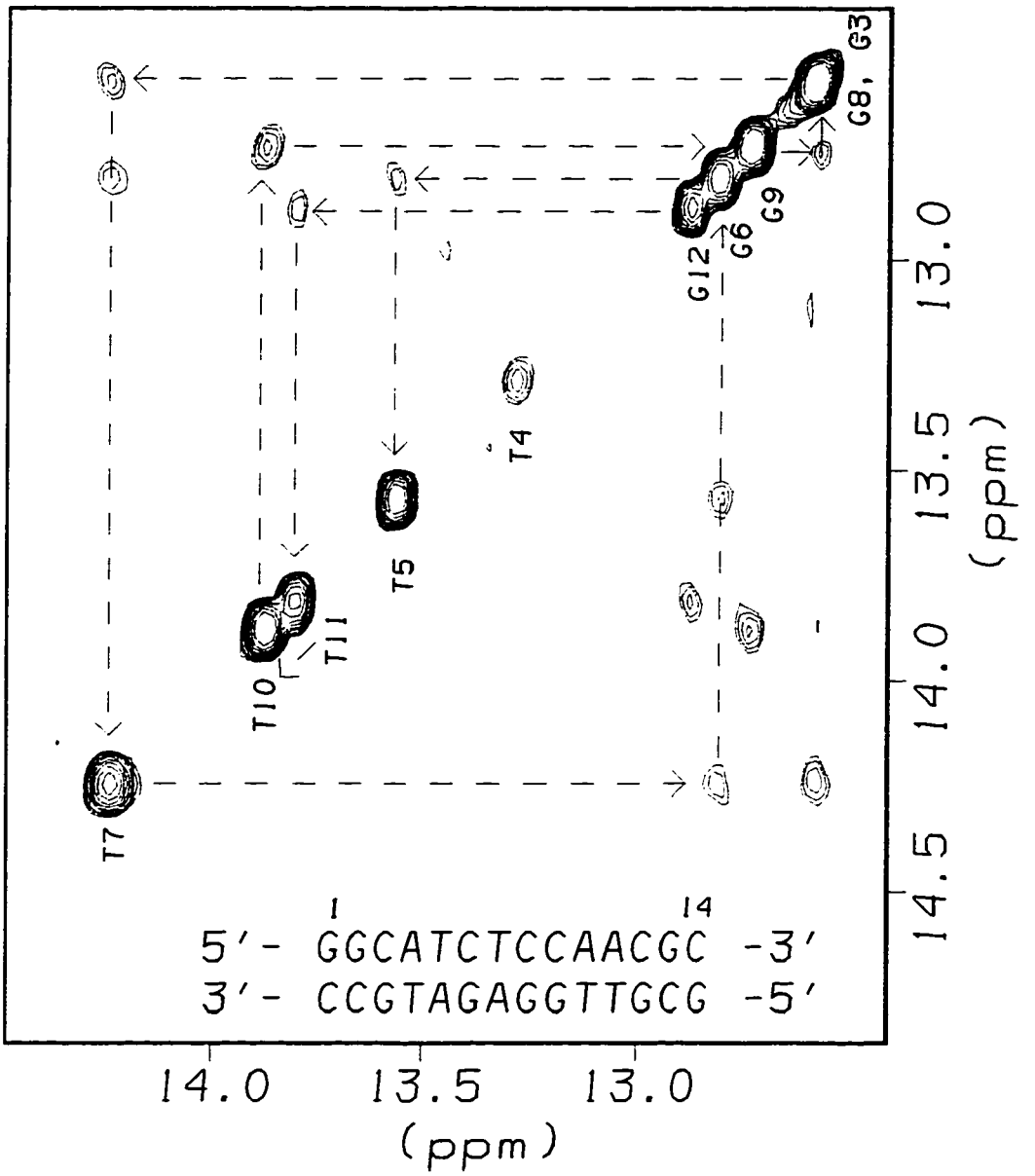

G3
G8,
G9
G6
G12
T4
T5
T11
T10
T7
1
14
5′- GGCATCTCCAACGC -3′
3′- CCGTAGAGGTTGCG -5′
14.0
13.5
13.0
(ppm)
13.0
13.5
14.0
14.5
(ppm)

Figure 40. Imino Assignments for 14mer free and bound to ADR1z'.

The imino regions of 1D proton spectra acquired in water at 22 °C are plotted for free 14mer (top) and for ADR1z'-bound 14mer (bottom). Assignments obtained from NOESY data are indicated.

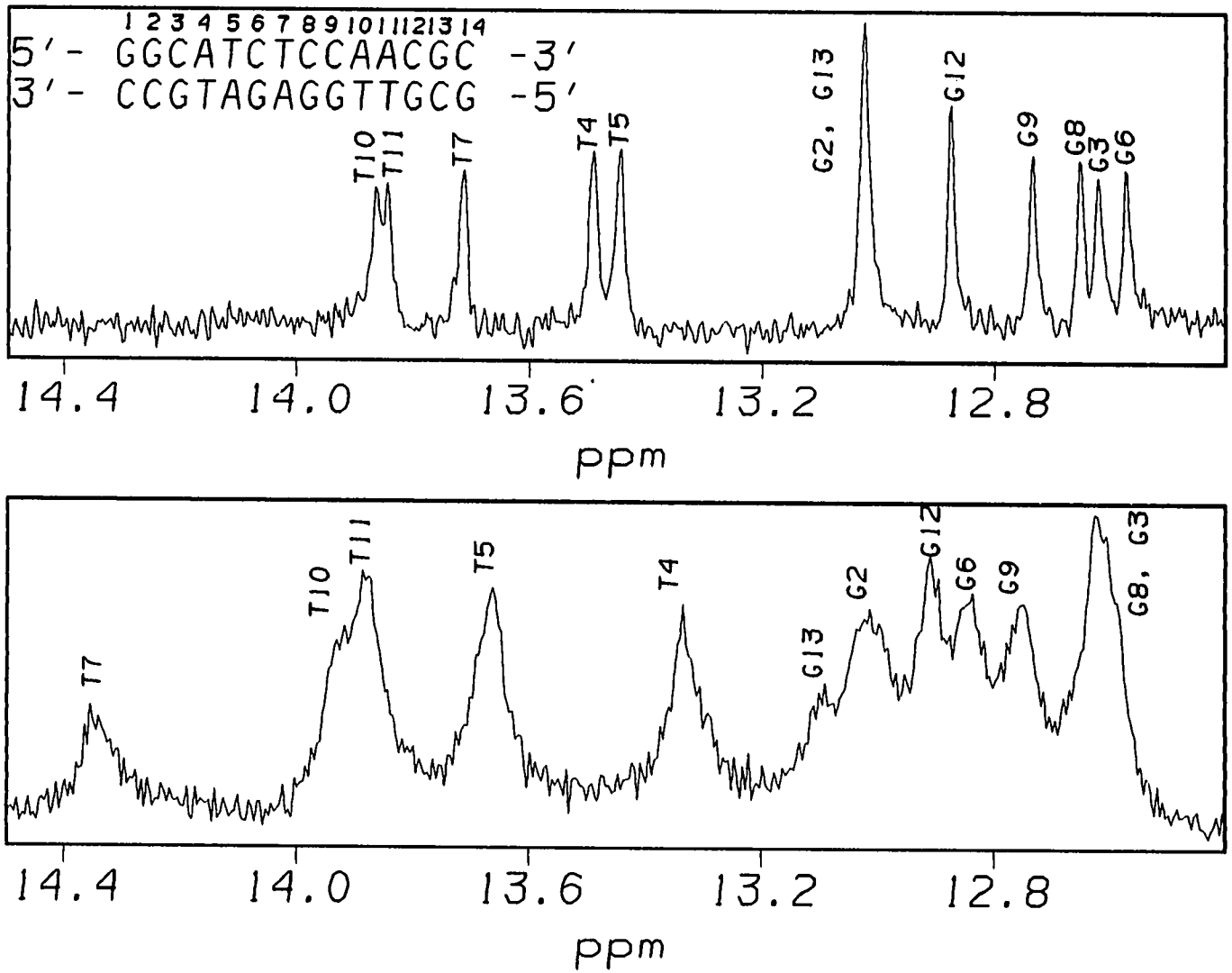

1 2 3 4 5 6 7 8 9 10 11 12 13 14
5′- GGCATCTCCAACGC -3′
3′- CCGTAGAGGTTGCG -5′
T10
T11
T7
T4
T5
G2, G13
G12
G9
G8
G3
G6
14.4
14.0
13.6
13.2
12.8
ppm
T7
T10
T11
T5
T4
G13
G2
G12
G6
G9
G8, G3
14.4
14.0
13.6
13.2
12.8
ppm

Table IV. Resonance Assignments for DNA Imino Protons

Base [a]	Free assignment [b]	Bound assignment [c]
G2	13.02 ppm	13.02 ppm
G3	12.62	12.58
T4	13.49	13.31
T5	13.44	13.64
G6	12.57	12.82
T7	13.72	14.36
G8	12.66	12.58
G9	12.74	12.74
T10	13.87	13.94
T11	13.84	13.89
G12	12.87	12.90
G13	13.02	13.09

a. Assignments are for the sequence

5'- GGCATCTCCAACGC -3'
3'- CCGTAGAGGTTGCG -5'

b. The free 14mer was assigned at pH 7.0, 22 °C.

c. The ADR1z'-bound 14mer was assigned at pH 7.0, 22 °C.

CHAPTER 3. STRUCTURE AND BEHAVIOR OF ADR1z'

I. Comparison to Single Zinc Finger Peptides

A. CHEMICAL SHIFT PERTURBATION MAPS

The two zinc fingers of ADR1 have been studied individually by NMR spectroscopy in the context of thirty-residue synthetic peptides spanning the canonical zinc finger sequences. Assignments have been published for the N-terminal finger (ADR1b; residues 102-130) (12), and for variants of the C-terminal finger, including a wild-type peptide (ADR1a; 130-159) (60) and a proline to alanine double mutant (PAPA) (13). A comparison of NMR chemical shift assignments for the fingers of ADR1z' with wild-type single finger assignments is shown in Figure 41, Figure 42, and Figure 43.

The overall pattern of chemical shift differences indicates that there is little if any perturbation of the individual fingers in the context of the two-finger construct. Few resonances exhibit chemical shift differences larger than the apparent ^1H linewidth of about 0.06 ppm, despite a 5 °C difference in temperature of acquisition. Of those resonances that do exhibit perturbations, the majority belong to residues that lie near the termini of the single finger peptides and are thus liable to end effects. For the remaining

Figure 41. ADR1z' Chemical Shifts Compared to Single Finger Shifts: NH Protons.

NH chemical shift differences between the zinc fingers of ADR1z' and single finger peptides are shown. Very little chemical shift perturbation is seen in the longer ADR1z' context. What differences exist are limited to the ends of the fingers, where the single finger peptides will exhibit end effects. This implies that the zinc fingers of ADR1z' are very similar in structure to the single fingers. The comparisons in this and the following two figures are based on the chemical shifts of wild-type single finger peptides ADR1b and ADR1a (minor form), except for residues Y132 and C140. Because the ADR1a assignments for these residues were in doubt [Bernstein, 1994 #5], assignments from an ^{15}N-labeled proteolytic fragment of ADR1z' that extends from H118 to E161 were substituted (M. Heese and P. Bowers, personal communication).

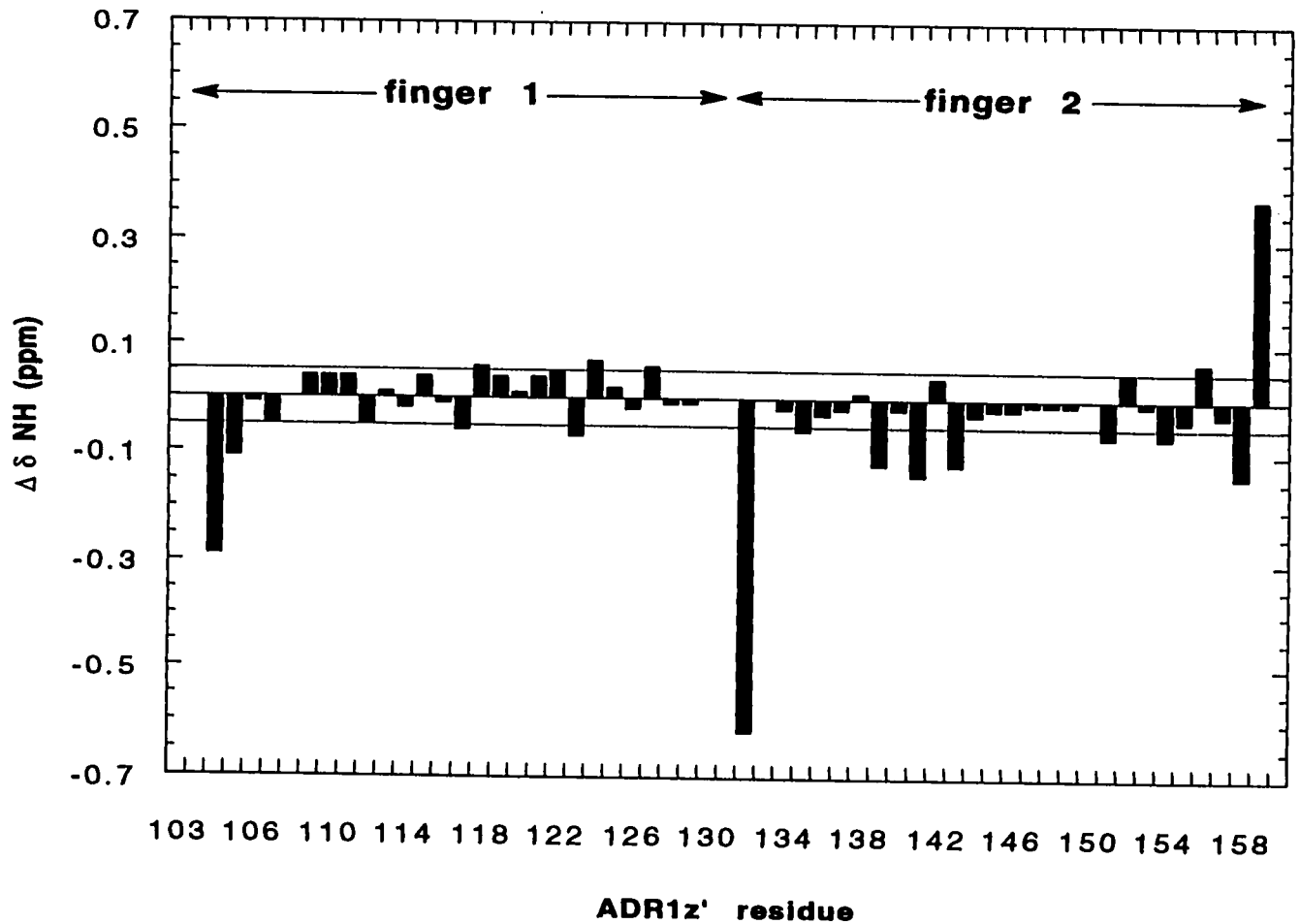

finger 1
finger 2
0.7
0.5
0.3
0.1
-0.1
-0.3
-0.5
-0.7
Δ δ NH (ppm)
103 106 110 114 118 122 126 130 134 138 142 146 150 154 158
ADR1z' residue

Figure 42. ADR1z' Chemical Shifts Compared to Single Finger Shifts: H^{α}.

Differences of H^{α} chemical shift between the zinc fingers of ADR1z' and the single finger ADR1b and ADR1a peptides are plotted with similar results to Figure 41.

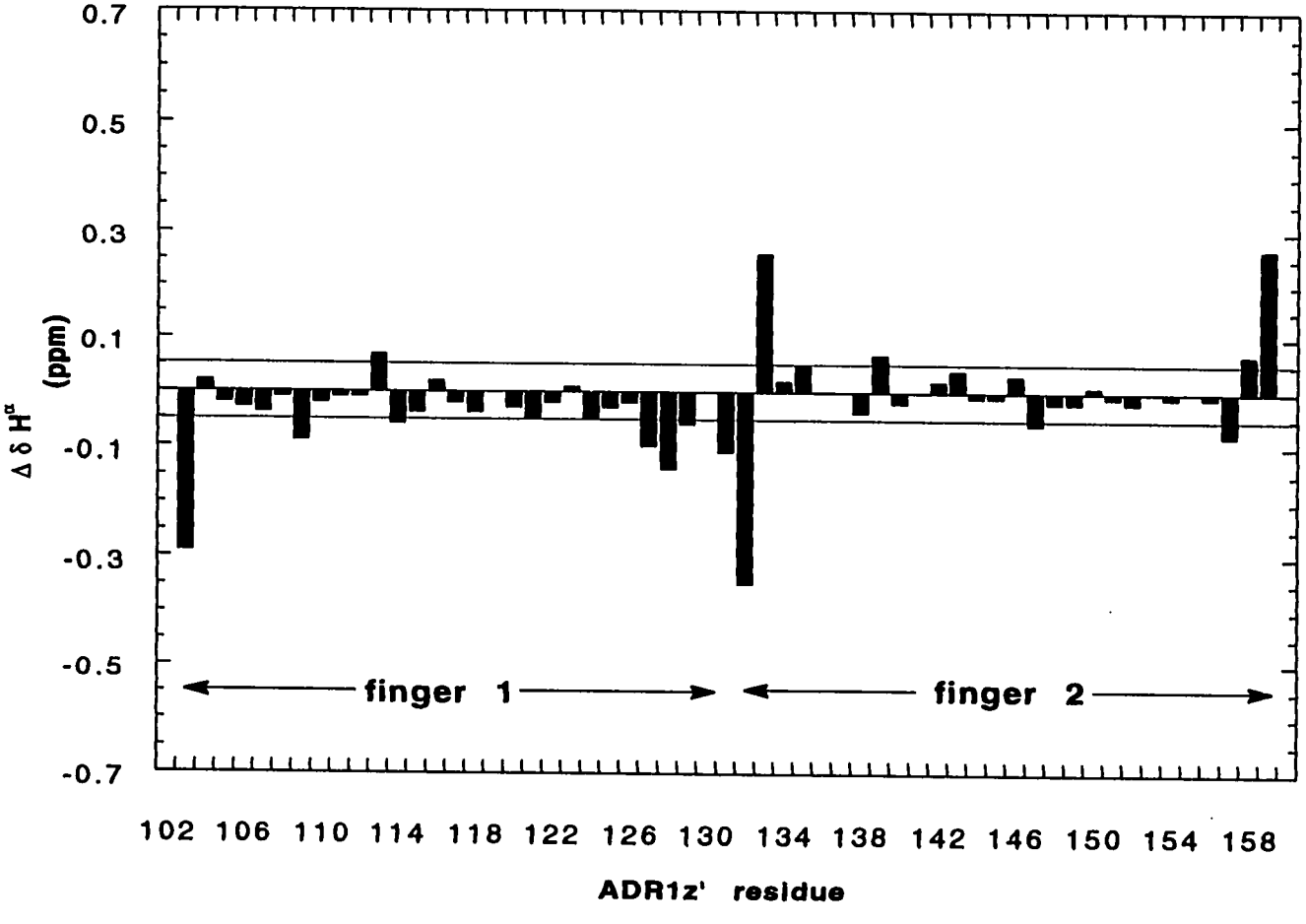
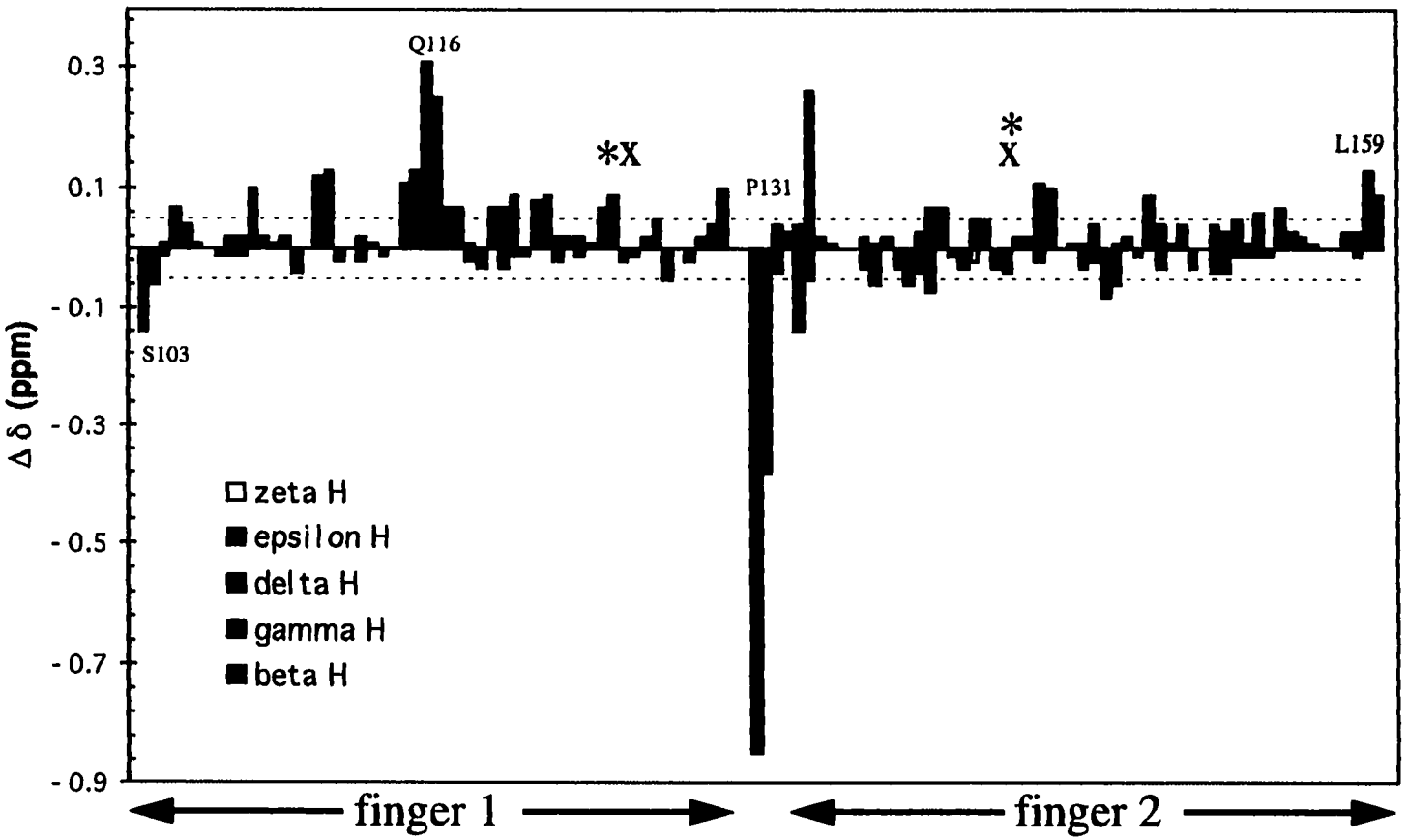

0.7
0.5
0.3
0.1
-0.1
-0.3
-0.5
-0.7
Δ δ Hα (ppm)
finger 1
finger 2
102 106 110 114 118 122 126 130 134 138 142 146 150 154 158
ADR1z' residue

Figure 43. ADR1z' Chemical Shifts Compared to Single Finger Shifts: Sidechains.

ADR1z' zinc finger sidechain proton chemical shifts are essentially identical to those of the single finger ADR1b and ADR1a peptides. This is further indication that the zinc fingers of ADR1z' are identical in structure to the single fingers. It is also good indication that the fingers of ADR1z' do not interact with each other in the absence of DNA. The positions where the MBP-1 zinc fingers interact are indicated by X; those where the Zif268 zinc fingers interact are indicated by asterisks. ADR1z' is not perturbed from single finger shifts in these regions.


Q116
*X
P131
*
X
L159
S103
0.3
0.1
- 0.1
- 0.3
- 0.5
- 0.7
- 0.9
Δ δ (ppm)
zeta H
epsilon H
delta H
gamma H
beta H
finger 1
finger 2

residues no backbone or sidechain chemical shift difference greater than 0.11 ppm is observed, except for β and γ proton resonances of Q116 in the N-terminal finger, which will be discussed later (Figure 43).

B. INTERFINGER INTERACTIONS

Limited interactions between adjacent zinc fingers in tandem arrays have been observed in several cases. In the Zif268 zinc finger-DNA cocrystal structure, the side chain of an arginine equivalent to ADR1 R124 hydrogen bonds to the backbone carbonyl of the equivalent of T142 in each finger-finger interaction (Figure 44) (20). The two zinc fingers of human enhancer binding protein MBP-1 also appear to interact in solution even in the absence of DNA (23). While the resulting relative finger orientation differs from that in Zif268, the participating residues fall at similar positions. A threonine at the equivalent of ADR1 S125 forms a putative hydrogen bond with the highly constrained sidechain of a lysine at the equivalent of T142. In contrast, there appears to be no interaction between adjacent fingers in yeast SWI5 double finger constructs in solution, based on chemical shifts and NOE patterns (24).

Figure 44. Interfinger interactions observed in Zif268.

The Zif268 cocrystal structure is shown, with the protein represented by ribbons and the DNA backbone a simple black trace (20). Residues from the fingertips of fingers two and three and the helices of the preceding fingers interact; these sidechains are shown as space-filling models.

ADR1z' offers a second example in which no interfinger interaction occurs in the absence of DNA. Discounting end effects, only the sidechain of Q116 is significantly perturbed in the N-terminal finger, while no sidechain resonances are significantly perturbed in the C-terminal finger of the construct. Specifically, no sidechain perturbations are observed near the locations of residues shown to interact in Zif268 or MBP-1 (Figure 43). While it is possible that these residues participate in interactions when bound to DNA, it appears that they do not in the free polypeptide.

C. ZINC FINGER STRUCTURE

The coincidence of ADR1z' and ADR1b chemical shifts implies that the NMR structure determined for the isolated N-terminal finger of ADR1 is a good model for the finger in the context of the longer construct (12). The chemical shift differences for Q116 are unexplained, but they are not matched by complementary perturbations elsewhere in the finger (Figure 43). It is possible the differences arise due to deamidation of glutamine to glutamate in the single finger peptide, which was heated during folding (12). Such an alteration may not have been detectable in the mass determination conducted in the early 1990's on the ADR1b sample used for assignment purposes.

The similarity of the single finger chemical shifts to those of ADR1z' further indicates that in the free ADR1z' polypeptide, the N-terminus does not interact with the canonical zinc fingers. In SWI5 and Tramtrack, the first zinc finger folds in the presence of additional N-terminal residues which form a third strand to the β sheet (22, 26). In ADR1z', the lack of chemical shift perturbations of the N-terminal finger in the presence of flanking residues suggests that no such interaction occurs in this construct.

The C-terminal finger of ADR1z' also appears to be nearly identical to the corresponding single finger ADR1a peptide. No structure has been calculated for this peptide, however, due to spectra complicated by extensive peak doubling from a minor conformational form (60). An NMR structure does exist for the double proline to alanine mutant, PAPA, which suffers less from peak doubling (13). These proline mutations are not neutral, however, as the P131 to alanine change diminishes DNA binding (54). Detailed structural studies of ADR1z' will provide the first opportunity to understand the structure of the fully wild-type C-terminal finger.

II. Carbon Chemical Shifts and Secondary Structure of ADR1z'

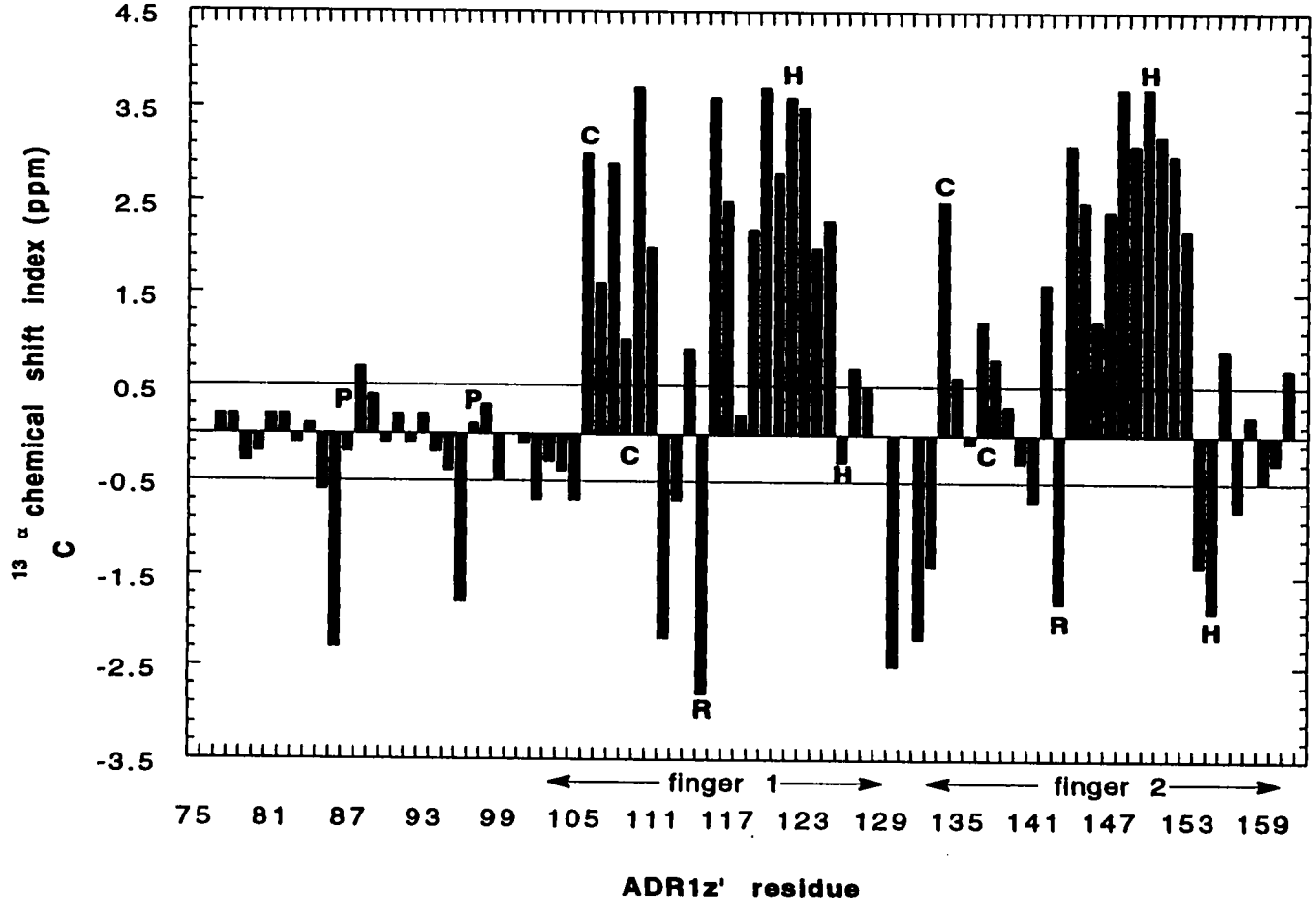
A. HELIX

One simple measure of secondary structure in proteins is the $^{13}C^{\alpha}$ chemical shift index (75). Systematic downfield chemical shift differences from random coil values are observed for C^{α} resonances in stretches of alpha helix, while analogous upfield shifts are found in extended structures, including beta strands.

The difference of $^{13}C^{\alpha}$ chemical shift from random coil for ADR1z' is shown in Figure 45. Strong downfield shifts indicative of alpha helix are observed for residues Q116 to S125 of finger one, and for the analogous R144 to K153 of finger two. The calculated structure of the single finger peptide ADR1b shows a helix extending from Q116 to H126, in good agreement with the secondary structure of ADR1z' (12). In addition, the single finger structure found R115 in an extended conformation and the backbone dihedral angles of A114 in right-handed helical space, in agreement with the respective upfield and downfield alpha carbon chemical shift perturbations observed for these positions in each finger. This arrangement of residues forms a ninety-degree turn in the peptide backbone, linking the Cys-Cys loop to the C-terminal helix.

Figure 45. Alpha Carbon Chemical Shift Index for ADR1z'.

The difference between alpha carbon chemical shifts and random coil values is plotted for all residues of ADR1z'. The positive (downfield) differences are indicative of helical secondary structure; negative (upfield) differences are indicative of extended structure.


4.5
3.5
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
$^{13}\alpha$ chemical shift index (ppm)
C
P
P
C
C
H
R
H
C
C
R
H
H
finger 1
finger 2
75 81 87 93 99 105 111 117 123 129 135 141 147 153 159
ADR1z' residue

Carbon chemical shift assignments have been published for three zinc fingers of TFIIIA, and a chemical shift index analysis was done (76). The pattern of carbon chemical shifts in the C-terminal region of each finger of ADR1z' is quite similar to that seen for the fingers of TFIIIA. Each helix begins at the identical position, preceded by a markedly upfield-shifted alpha carbon. In each case the helical pattern of strong downfield shifts appears to break just prior to the last histidine ligand.

The break from helical chemical shift at the final histidine ligand is unexpected, as the helix extends through this residue in both the structure of ADR1b (12) and of PAPA (13). However, it is unclear whether the random coil histidine C^{α} chemical shift used in the calculation accurately reflects the shift of histidine coordinated to a zinc ion. The C^{α} shift of histidine is known to vary with protonation state (77); in the absence of a better option, the value used in this study is that of a protonated histidine, to match the analysis for TFIIIA.

B. CYS-CYS LOOP

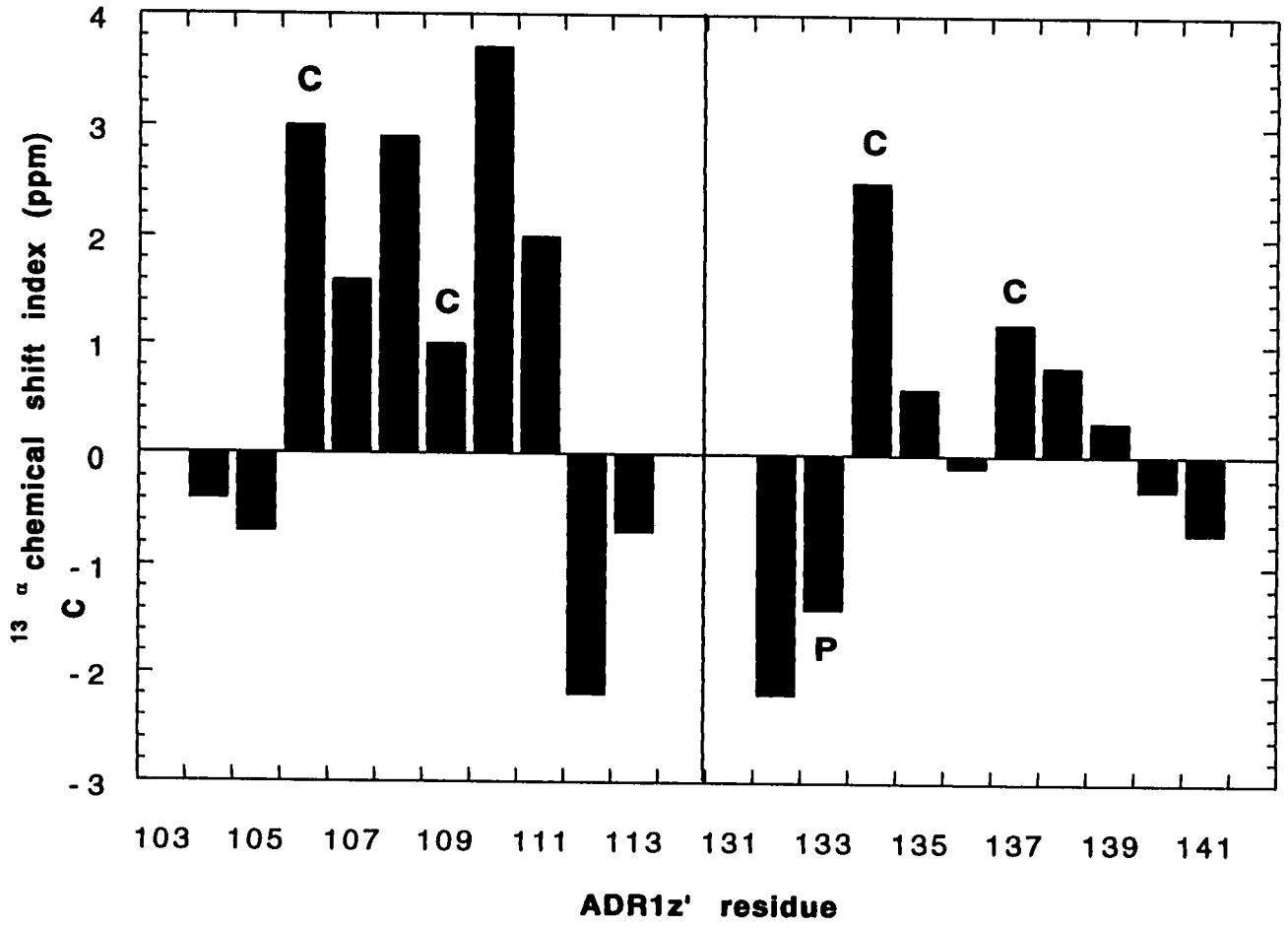
As in the fingers of TFIIIA, the cysteine ligands of ADR1z' show pronounced downfield shifts from the random coil C^{α} value, caused by zinc coordination (Figure 45) (76). Unfortunately, further

comparisons in this region are inappropriate, as the TFIIIA fingers have four residues between cysteines while ADR1 fingers possess only two intervening residues. In the calculated structure of ADR1b, residues F104, V105, A112, and F113 comprise a classical β-sheet (12). In agreement with this result, in ADR1z' the C^{α} of each residue experiences an upfield perturbation from random coil indicative of extended structure. The analogous residues in finger 2 also exhibit upfield shifts, although that of Y132 is difficult to interpret as this residue precedes a proline; residues immediately N-terminal to prolines exhibit upfield chemical shift perturbations even in a random coil setting (78). When a 2.1 ppm upfield shift due to the neighboring proline is accounted for, a small upfield shift remains for Y132.

In the structure of ADR1b, residues C106, T110 and R111 participate in a β-bulge, while residues E107, V108 and C109 form a reverse turn (12). In ADR1z', each of these residues, including the cysteines mentioned previously, exhibits varying but large downfield shift from random coil (Figure 46). No structure is available for wild-type ADR1a, but the $^{13}C^{\alpha}$ chemical shift pattern for finger 2 of ADR1z' suggests that the turn region of this finger will have a very different structure. The residues between and just after the cysteines are shifted upfield of those in the first finger by an average of 2 ppm (Figure 46). This region is preceded by two

Figure 46. Alpha Carbon Chemical Shifts for the Turn Region of the ADR1z' Zinc Fingers

Selected data from Figure 45 are plotted here to highlight differences in the Cys-Cys loops of finger 1 and finger 2 of ADR1z'.


4
3
2
1
0
-1
-2
-3
$^{13}C^{\alpha}$ chemical shift index (ppm)
C
C
C
C
P
103
105
107
109
111
113
131
133
135
137
139
141
ADR1z' residue

prolines, P133 and P131, which are not present in ADR1b. It is possible that one or both of these prolines cause the N-terminal region of this finger to adopt a different conformation. P131 is less likely an agent of profound change in zinc finger structure, as two of the fingers of Zif268 have proline at this position, and fold into structures much like that of ADR1b (12, 20).

A structure of the C-terminal finger of ADR1 in which both P131 and P133 are mutated to alanine has been solved (13). This structure includes a β-bulge and reverse turn similar to ADR1b. In the mutant structure, as well as in the CX_2C zinc fingers of Zif268, the residue analogous to P133 of ADR1a adopts a backbone conformation inaccessible to proline (13). It was concluded that backbone rearrangements would be required to accommodate the wild-type proline at ADR1 residue 133; the results of the carbon chemical shift analysis of ADR1z' indicate that a rearrangement in the region of G135-R139 does indeed occur.

C. PROLINES

In single finger peptides, the C-terminal finger of ADR1 exhibits spectral doubling due to multiple conformations (60). It has long been a question whether this doubling could be due to *cis-trans* isomerization of the two prolines present in finger 2 but not

in finger 1. No evidence for *cis*-proline was ever found in single finger NOE spectra. ADR1z' exhibits only a single set of lines, but the chemical shifts for finger 2 are overall somewhat more similar to the minor conformation of ADR1a than to the major conformation. However, at Y132, exactly between the two prolines, ADR1z' chemical shifts are more similar to the ADR1a major form.

The question remains for ADR1z' whether its prolines are *trans* or *cis*. Based on carbon chemical shifts, there is no evidence that any of the four prolines in ADR1z' is *cis*-proline. *trans* and *cis*-proline differ in their beta and gamma carbon chemical shifts, with *trans* having random coil values of 30.6 ppm for C^{β} and 25.6 ppm for C^{γ}, while *cis* has values of 33.0 and 23.3 ppm (77). For ADR1z', P131 and P133 exhibit C^{β},C^{γ} shifts of 30.5, 25.1 and 30.5, 25.0 ppm respectively. These fall much closer to the shifts expected for *trans* proline. In addition, these residues fall in a region that is most likely somewhat extended, based on analogy to ADR1b; if that were the case, the $^{13}C_{\beta}$ shifts would be expected to be perturbed downfield from random coil if perturbed at all (75), rendering the observed values even less likely for *cis* beta carbons.

D. LINKER

The role of the inter-finger linker in DNA binding is a subject of intense interest. Mutagenesis experiments indicate that altering the highly conserved linker can diminish DNA binding (27). In the existing zinc finger-DNA cocrystal structures, however, the linker makes no significant direct contacts to DNA or to the rest of the protein (20, 25, 26, 28, 29). In the only structure solved to 1.6 Å resolution, water-mediated contacts were detected between the conserved lysine in each linker of Zif268 and phosphate groups on the less-contacted DNA strand. This may not be a general rule, however, as the linker in the Tramtrack cocrystal structure appears to be somewhat disordered, with high temperature factors (26). In the Zif268 and GLI structures the linkers have similar, well-defined backbone conformations that were suggested to influence the exact relative orientation of the fingers (20, 28). Because the structures of these proteins are not known in the absence of DNA, it is not clear whether such order precedes or depends upon DNA binding.

In the TFIIIA three-finger construct, $^{13}C^{\alpha}$ chemical shifts of the linkers were slightly upfield of random coil, suggesting an extended structure even in the absence of DNA (76). This supposition was supported by the pattern of NOEs as well. In ADR1z', linker residues T127 and N128 are shifted somewhat

downfield from random coil, and residue E129 has a random coil shift (Figure 45). Interpretation of the shift of K130 is complicated by the fact that it will be shifted upfield by the presence of the following proline (78). Correction for the neighboring proline leaves a residual 0.5 ppm upfield perturbation.

Thus there is no regular pattern of secondary structure in the linker of ADR1 like that seen in TFIIIA; all shifts lie close to random coil. It has already been established that the zinc fingers themselves do not interact in free ADR1z'; it is also unlikely that they are maintained in any set relative orientation by the linker. Thus it seems that neither fixed orientation of the fingers nor rigid structure of the linker is an absolute prerequisite for DNA binding by tandem Cys_2-His_2 zinc fingers.

E. N-TERMINUS

ADR1 requires sequence N-terminal to the zinc fingers to bind DNA with high affinity (53). The ADR1z' construct includes the minimal flanking region required, from residue 84 to 103. As described above, it is clear from the unperturbed ^1H chemical shifts of the zinc fingers in free ADR1z' that the N-terminal flanking residues do not interact with the fingers themselves. It is further clear from the $^{13}C\alpha$ chemical shift profile of the N-

terminus that this region of the polypeptide is essentially random coil in the absence of DNA (Figure 45). The only significant perturbation from random coil chemical shifts in this region are induced by prolines, which cause upfield shifts of about 2 ppm in the residues preceding them, even in an otherwise random coil context (78). ^{15}N linewidth measurements also support the view that the N-terminus of ADR1z' is more mobile than the zinc fingers, as it displays linewidths an average of 3 Hz narrower than those of the rest of the polypeptide (Figure 47). The random coil nature of the N-terminal region of ADR1z' must be interpreted with caution, however. It is not only possible that this region acquires structure in the presence of DNA, but it may also have a defined structure in the full-length protein that is not stable in the minimal ADR1z' construct.

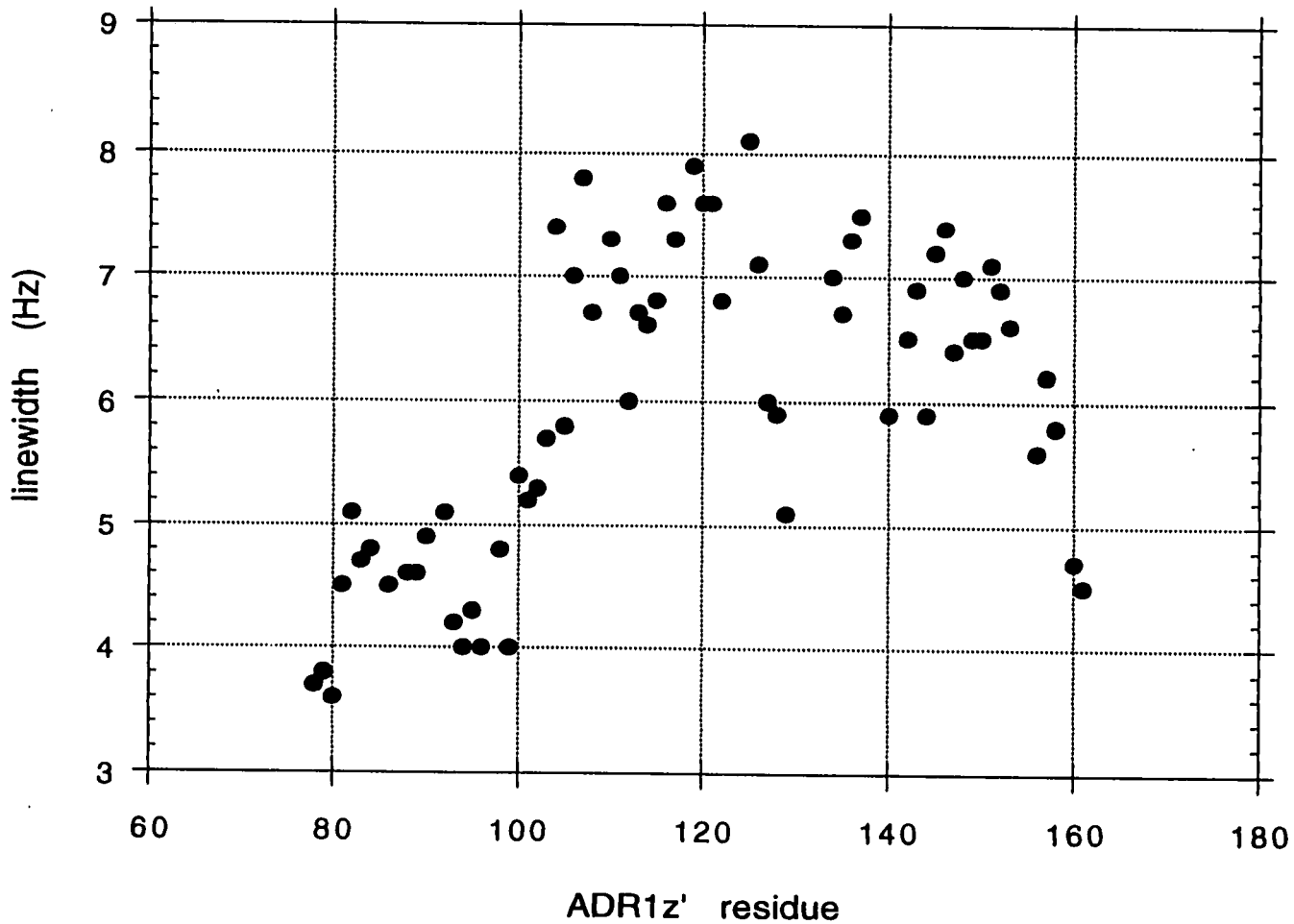
III. Further Evidence for Secondary Structure in ADR1z'

A. NOE CONNECTIVITIES

Amino acid residues involved in regular secondary structure in folded proteins exhibit characteristic patterns of short-range NOEs (61). In alpha helices, strong NOE crosspeaks arise between the NH of residue i and the NH of residue i+1, between the H^{α} of residue i and the NH of residue i+3, and between the H^{α} of residue i

Figure 47. Linewidths of ADR1z'.

Nitrogen linewidths were measured in an ^{15}N-HSQC spectrum of ADR1z'. The N-terminus (76-102) exhibits linewidths on average 3 Hz smaller than those of the zinc fingers, reflecting increased mobility.


linewidth (Hz)
9
8
7
6
5
4
3
60
80
100
120
140
160
180
ADR1z' residue

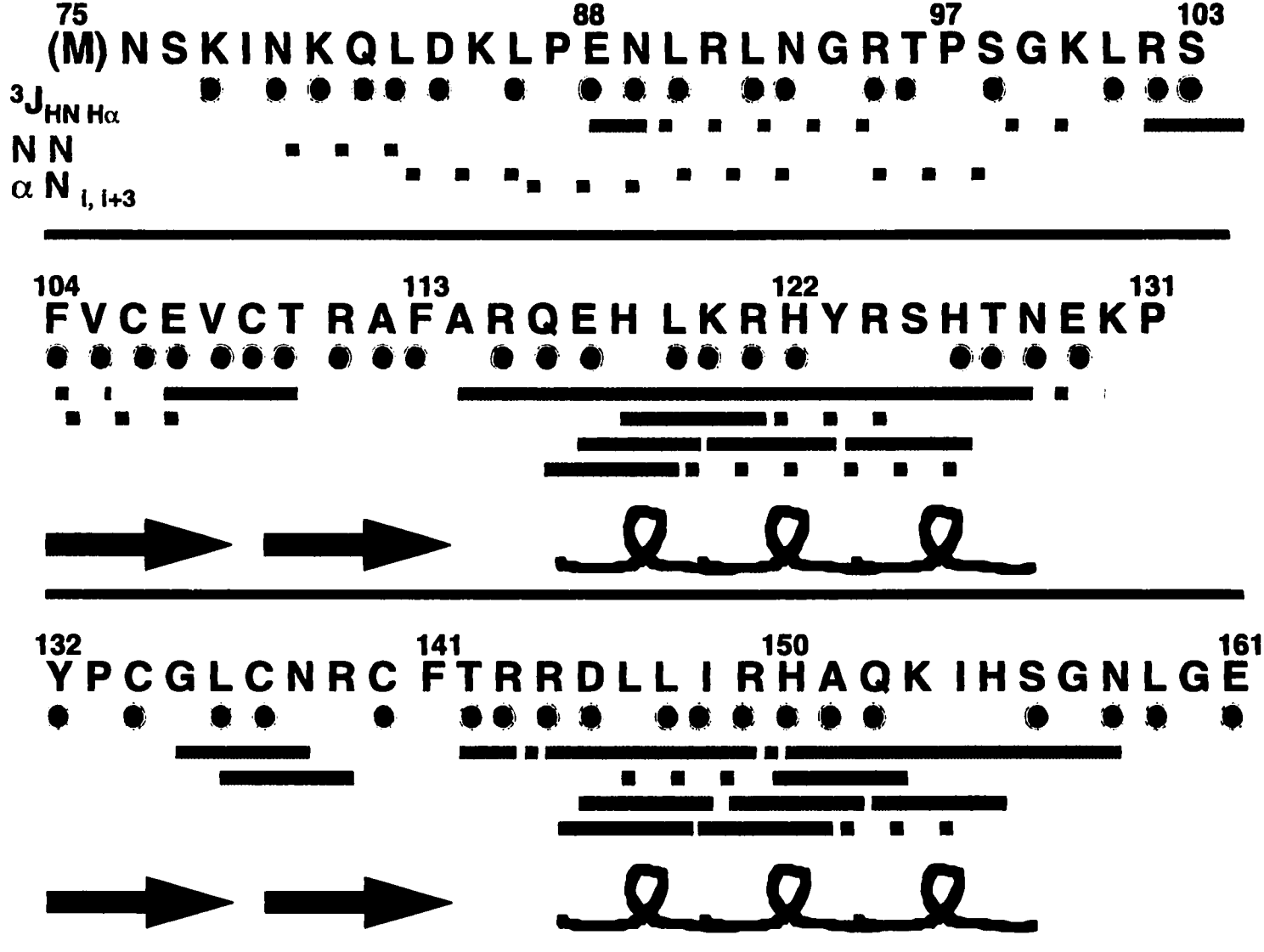
and the H^{β} of residue i+3. 3_{10} helices exhibit similar patterns and also show strong NOES between the H^{α} of residue i and the NH of residue i+2. Extended structures such as the strands in a beta sheet lack strong NOEs of the types described above, but they exhibit very strong NOEs between the H^{α} of residue i and the NH of residue i+1 (61).

The pattern of short-range NOEs for ADR1z' is displayed in Figure 48, providing a second indication of secondary structure in addition to the alpha carbon chemical shifts discussed earlier. Residues Q116 to H126 in finger 1 and R144 to H155 in finger 2 are clearly helical, showing the diagnostic pattern of NN and $\alpha N_{i,i+3}$ NOES. This pattern agrees with alpha carbon chemical shift index data for this polypeptide, and with the single finger ADR1b and PAPA structures (12, 13). The helix of finger 2 appears to be largely alpha helical, while that of finger 1 contains a number of unambiguous $\alpha N_{i,i+2}$ NOEs indicative of 3_{10} helix. A transition to 3_{10} helix was also reported for the analogous single finger peptide, ADR1b (12). This transition is thought to be necessary to bring the more closely spaced histidines into position to coordinate zinc tetrahedrally.

Residues E107 to T110 and G135 to N138 in the Cys-Cys loops display both $\alpha N_{i,i+1}$ and NN NOES and a lack of $\alpha N_{i,i+3}$ NOEs.

Figure 48. Secondary Structure of ADR1z': NOE Connectivities and Coupling Constants.

$^3J_{HNH\alpha}$ coupling constants and d_{NN}, $d_{\alpha Ni,\ i+3}$ NOES are mapped against the ADR1z' sequence. Large couplings greater than 8 Hz are represented by red circles and are indicative of extended structure; small couplings less than 5 Hz indicative of helical structure are represented by blue circles; and intermediate couplings between 5 and 8 Hz, indicative of random coil, by green circles. The strong pattern of NN and $\alpha N_{i,\ i+3}$ NOEs near the C-terminus of each finger is also a hallmark of helical secondary structure. Unambiguous NOEs are represented by solid lines. Where NOEs cannot be unambiguously assigned due to chemical shift overlap, a dashed line is drawn.


75
88
97
103
(M) N S K I N K Q L D K L P E N L R L N G R T P S G K L R S
$^3J_{HN\,H\alpha}$
N N
α N $_{i,\,i+3}$
104
113
122
131
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 P
132
141
150
161
Y P C G L C N R C F T R R D L L I R H A Q K I H S G N L G E

This pattern is consistent with participation in a turn structure at the juncture of the strands of the beta sheet. The remaining residues of the zinc fingers show only $\alpha N_{i,i+1}$ NOES, indicative of extended structure.

The N-terminus of ADR1z' has been shown to be unstructured and mobile by various measures including alpha carbon chemical shift, ^{15}N linewidth, and heteronuclear NOEs (D. E. Hyre, personal communication). The random coil nature of this region results in severe overlap of NH and H^{α} chemical shifts, leading to uncertainty in assignments of short-range NOEs. This difficulty is clearly apparent in Figure 48, in which a large number of ambiguous NOEs are displayed for this region. These ambiguous NOEs cannot be used to diagnose secondary structure. However, no unambiguous NOEs are present in this region that would contradict the earlier conclusion that this region is largely unstructured.

B. COUPLING CONSTANTS

A final method for determining secondary structure in proteins relies on the three-bond coupling between backbone NH and H^{α} protons in each residue. The size of the coupling depends on the dihedral angle between these protons, which in turn reflects the secondary structure adopted by the residue (61). Residues that fall

in helices exhibit small $^3J_{HNH\alpha}$ on the order of 5 Hz or less. Residues that adopt an extended structure have large $^3J_{HNH\alpha}$ greater than 8 Hz. Residues that are conformationally averaged, such as those in unstructured regions, show intermediate $^3J_{HNH\alpha}$ of about 7 Hz (61).

The values of $^3J_{HNH\alpha}$ for ADR1z' are shown in Figure 48. Residues in the previously described helices of the zinc fingers exhibit small coupling constants, as expected. The -1 position of each helix, found to be in an extended structure in ADR1b (12), possesses the anticipated large coupling constant.

In the predicted beta sheet regions of the Cys-Cys loops, residues for which $^3J_{HNH\alpha}$ can be measured, including F104, A112, F113, Y132, and C140, have coupling constants indicative of extended structure. The structural difference between the turns of finger 1 and finger 2 is apparent in the coupling constants. Those in finger 1 are consistent with the type I reverse turn found in ADR1b, with Cys106 having a very small coupling constant, and C109 having a very large coupling constant. The analogous residues in finger 2 both have coupling constants of around 7 Hz, indicating that this loop does not also fold to a classical type I reverse turn.

All residues of the N-terminus of ADR1z' for which coupling constants can be evaluated have values indicative of random coil,

in agreement with all other structural measures. Residues of the interfinger linker also have random coil coupling constants, indicating that the linker does not adopt a fixed structure in the free polypeptide. This result matches the conclusion from alpha carbon chemical shift index data, and supports the notion that the zinc fingers are not held in any fixed orientation prior to DNA binding.

CHAPTER 4. STRUCTURE AND BEHAVIOR OF ADR1z'-DNA

I. Carbon Chemical Shifts and Secondary Structure of Bound ADR1z'

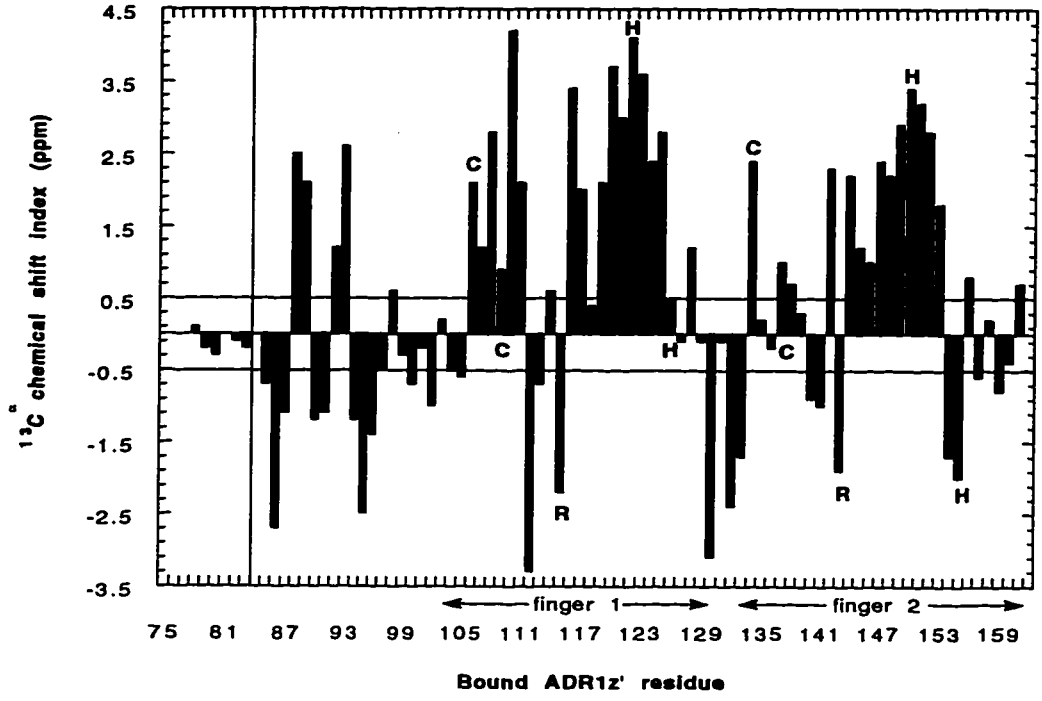
A. ZINC FINGERS

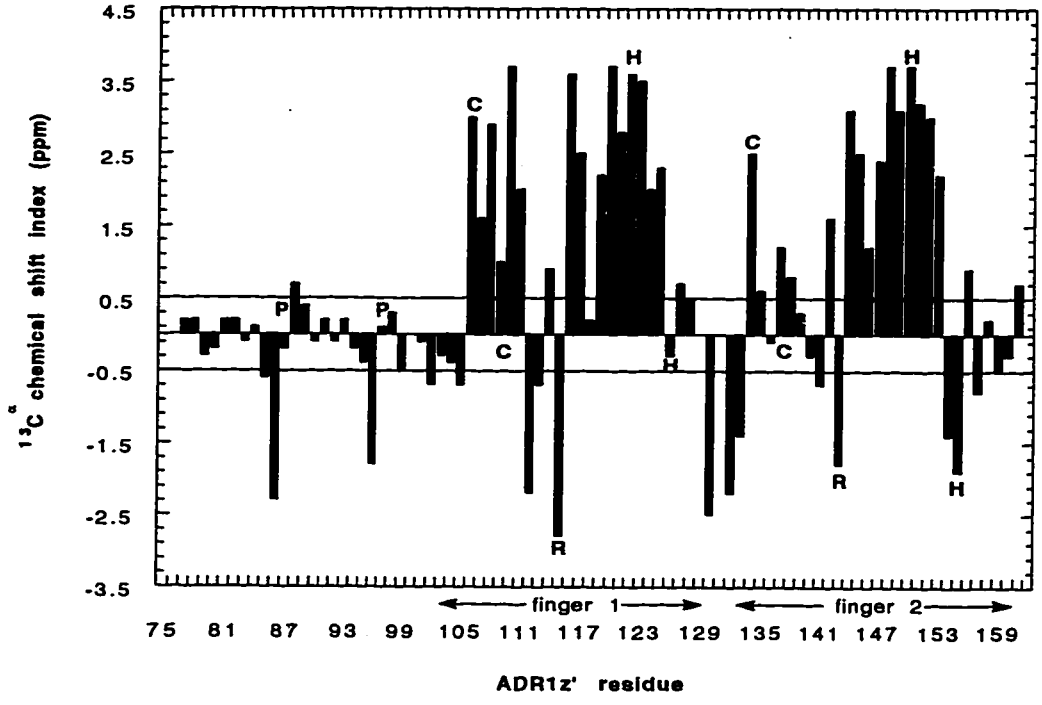
When ADR1z' binds to DNA, a majority of residues experience large perturbations of ^{1}H and ^{15}N chemical shift, as seen in Figure 24. The $^{13}C\alpha$ chemical shift profile of the zinc fingers remains fairly constant, however, suggesting that their structure is not changed much upon DNA binding (Figure 49). The beginning and ending points and overall profile of downfield shifts defining each α-helix is retained upon DNA binding, as is the extended nature of the residue immediately preceding the helix. The pattern of downfield shifts in the Cys-Cys loop of finger 1 is reiterated in the DNA-bound form, and the alternate pattern observed in the same region for finger 2 is retained as well. Both P131 and P133 appear on the basis of beta and gamma carbon shifts to remain in the *trans* conformation.

That the zinc fingers do not alter structure upon DNA binding was expected from comparison of previously determined structures of free and bound fingers. For example, the solution NMR structure

Figure 49. Alpha Carbon Chemical Shift Index for ADR1z'-DNA.

The difference between alpha carbon chemical shifts and random coil values is plotted for all residues of DNA-bound ADR1z'. The positive (downfield) differences are indicative of helical secondary structure; negative (upfield) differences are indicative of extended structure. In the lower panel a similar plot for the free protein (from Figure 45) is shown for comparison.


4.5
3.5
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
$^{13}C^{\alpha}$ chemical shift index (ppm)
C
C
H
H
C
C
R
R
H
finger 1
finger 2
75 81 87 93 99 105 111 117 123 129 135 141 147 153 159
Bound ADR1z' residue


4.5
3.5
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
$^{13}C^{\alpha}$ chemical shift index (ppm)
P
P
C
C
H
H
C
C
R
R
H
finger 1
finger 2
75 81 87 93 99 105 111 117 123 129 135 141 147 153 159
ADR1z' residue

calculated for single finger ADR1b displays a great similarity to the fingers of Zif268 bound to DNA, with mainchain RMSDs of 0.613 and 0.657 Å between ADR1b and Zif268 fingers two and three, respectively (13). Similar RMSD values of 0.655 and 0.644 Å are observed between the Zif268 fingers and the PAPA structure (13). The carbon chemical shift profile of the zinc fingers in ADR1z' both free and bound to DNA supports the picture that the fingers are minimally perturbed upon DNA binding.

B. HELICES

The largest differences in $^{13}C^{\alpha}$ chemical shift in the free versus the bound zinc fingers fall in and near the helices (Figure 49). R115, E117, R144 and D145 all experience changes in chemical shift on the order of 1 ppm upon DNA binding. Mutagenesis experiments predict that R115, H118, and R143 all make contacts to DNA bases in the complex (54). These results complement data from Zif268 and other finger-DNA complexes in which residues at analogous positions make specific contacts (20, 26, 28). In addition, in two of the Zif268 fingers, an aspartate at the position of ADR1 D145 was found to buttress a direct contact by the analogue of R143 (20). A similar sidechain-sidechain interaction was found for the ADR1 PAPA peptide free in solution (13). It is

therefore difficult to interpret whether the limited but significant changes in carbon chemical shift at and near these DNA-contacting residues result from actual changes in secondary structure induced by DNA binding, or from inductive effects on chemical shift environment due to proximity to the DNA itself.

C. N-TERMINUS

The site of the greatest perturbation of $^{13}C^{\alpha}$ chemical shift upon ADR1z' DNA binding falls in the N-terminal flanking sequences required for high-affinity interaction (Figure 49). Residues from E88 through R95 each experience a 1 to 2 ppm perturbation. This is precisely the region in which a suppressor mutation, R91K, falls; this mutation compensates for defects in DNA binding induced by deleterious finger mutations (58). In contrast, residues N-terminal to K85 possess random coil $^{13}C^{\alpha}$ values. These are exactly the residues found to be unnecessary for high affinity binding in mutagenesis experiments (53).

No pattern suggestive of regular secondary structure is apparent in the $^{13}C^{\alpha}$ shifts of the perturbed N-terminal region. It is quite likely that this region is no longer random coil in the presence of DNA. However, the exact nature of the structure in this region is difficult to predict from chemical shift, as this measure

may be sensitive not only to secondary structure, but to proximity to the shift-inducing DNA bases themselves.

II. Further Evidence for the Secondary Structure of Bound ADR1z'

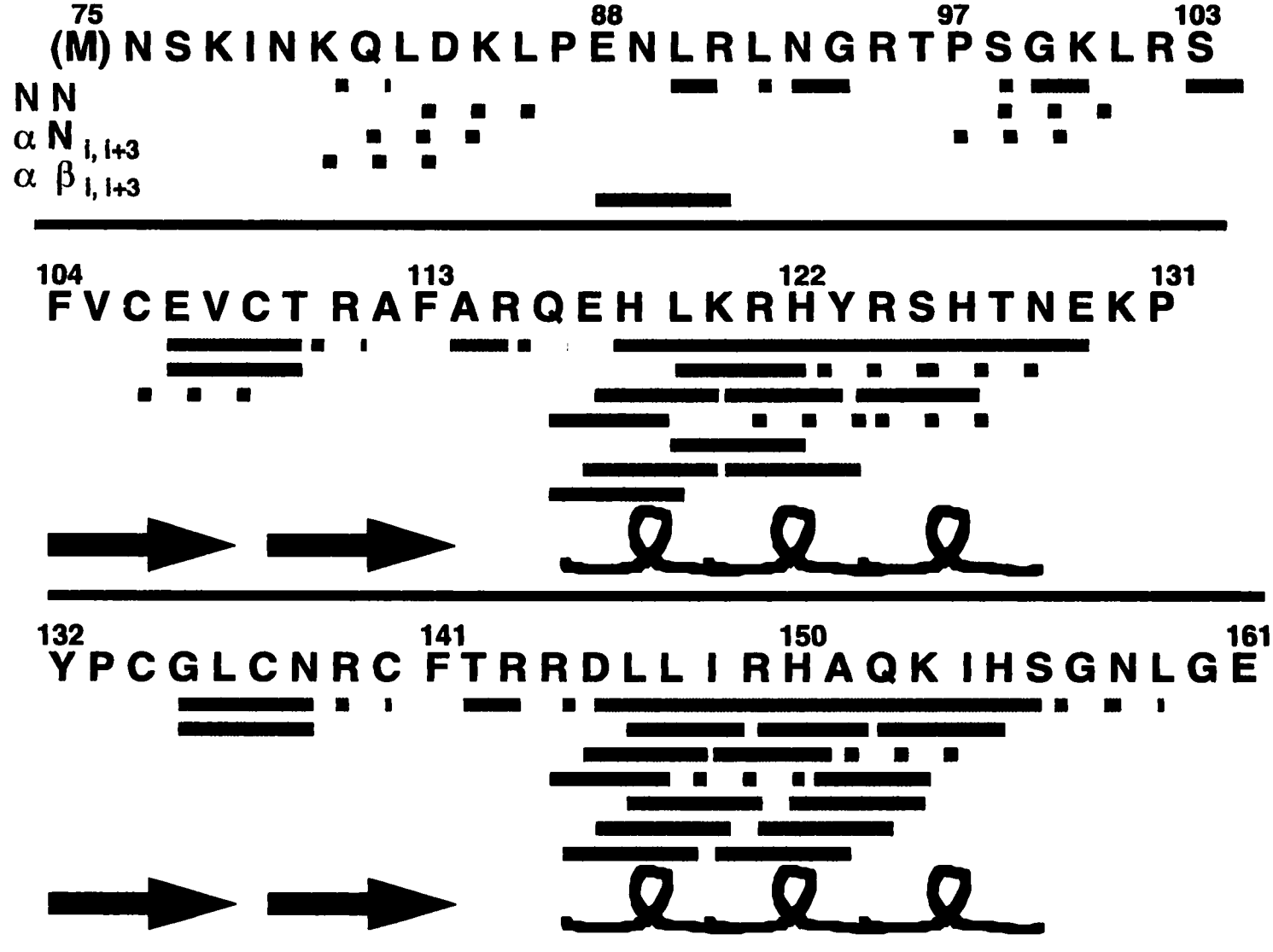
A. NOE CONNECTIVITIES

The pattern of short-range NOE connectivities in DNA-bound ADR1z' supports the result that the structure of the zinc fingers remains largely unchanged (Figure 50). The Cys-Cys loops display short-range NOE patterns essentially identical to those of free ADR1z'. The patterns of NN and $\alpha N_{i,i+3}$ connectivities indicative of helix also fall in identical regions in free and bound ADR1z'; in addition the helix of finger 1 shows $\alpha N_{i,i+2}$ NOEs consistent with partial 3_{10} helix. $\alpha\beta_{i,i+3}$ NOEs are indicative of alpha helix (61). Finger 2 shows these throughout the helical region, while finger 1, likely partly 3_{10} helix, exhibits fewer.

The N-terminus of DNA-bound ADR1z' shows no pattern of NOEs suggestive of regular secondary structure. The distinctive pattern of NN, $\alpha N_{i,i+3}$, and $\alpha\beta_{i,i+3}$ NOEs indicative of helical structure is absent in this region. Previously, residues 70 to 85 of ADR1 were predicted to form an alpha helix, based on sequence

Figure 50. Secondary Structure of ADR1z'-DNA: NOE Connectivities.

d_{NN}, $d_{\alpha Ni,i+3}$, and $d_{\alpha\beta i,\ i+3}$ NOEs are mapped against sequence for DNA-bound ADR1z'. The abundance of these types of NOEs near the C-terminus of each finger is diagnostic for helical structure. The lack of these NOE types in the N-terminus rules out helical structure in that region. Unambiguous NOEs are represented by solid lines. Where NOEs cannot be unambiguously assigned due to chemical shift overlap, a dashed line is drawn.


75
88
97
103
(M) N S K I N K Q L D K L P E N L R L N G R T P S G K L R S
N N
α N i, i+3
α β i, i+3
104
113
122
131
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 P
132
141
150
161
Y P C G L C N R C F T R R D L L I R H A Q K I H S G N L G E

analysis (50). It was noted that such a helix would be amphipathic in nature; furthermore, if residues 85 to 103 adopted a helical structure, that helix would be amphipathic as well. The observed NOEs clearly rule out the possibility that residues 76 to 103 are helical in the minimal DNA binding construct. While there are a large number of $\alpha N_{i,i+1}$ NOEs consistent with extended structure, there are also a handful of NN NOEs that suggest this region is likely not completely extended.

The lack of regular secondary structure in the N-terminus is consistent with the alpha carbon chemical shift index, which also lacked a regular pattern. Coupling constants might help define the nature of the backbone conformation. NMR experiments used to measure $^3J_{HNH\alpha}$ involve a fairly long delay during which the coupling evolves. The nature of the ADR1z'-DNA complex is such that magnetization relaxes quickly during this delay, however, resulting in a loss of signal so severe as to preclude accurate measurement of peaks in the resulting spectra. Definitive description of the irregular course of the N-terminus in the DNA-bound polypeptide must then await a detailed structure calculation based on intra- and intermolecular NOEs in the protein-DNA complex.

B. EXTENSION OF THE CYS-CYS LOOP OF FINGER 1

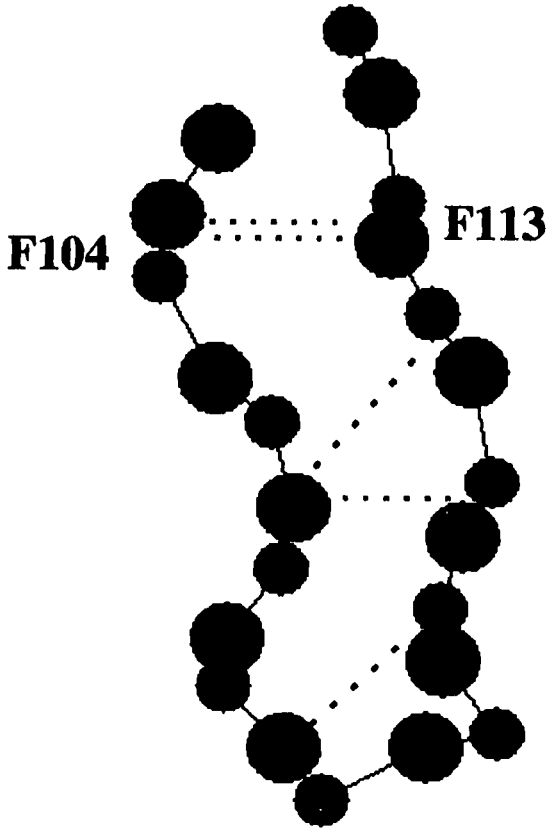
The only apparent change to the ADR1z' zinc finger structure upon DNA binding occurs at the extreme N-terminus of finger 1. Unambiguous cross-strand NOEs confirm that the residues from F104 to F113 adopt a structure like that found in the Cys-Cys loop of ADR1b (12). These NOEs are shown for free and DNA-bound ADR1z' in Figure 51. They include NOEs between F104 NH and F113 NH, between C106 NH and R111 NH, between C106 NH and A112 H^{α}, and between V108 NH and T110 NH.

In the structure of ADR1b, S103 does not extend the Cys-Cys loop to lie across from A114, but rather turns back to lie adjacent to V105 and A112. Given the near-perfect correspondence of finger 1 chemical shift to ADR1b values, it is likely that free ADR1z' adopts this structure as well. However, in the DNA-bound form, S103 occupies a different position. As shown in Figure 51, in bound ADR1z', further cross-strand NOEs exist that place S103 adjacent to A114, extending the Cys-Cys loop. These NOEs occur between S103 NH and A114 H^{α}, between F113 NH and S103 NH, and between A114 NH and R102 H^{α}. The nature of the NOEs is not suggestive of an extension of the classical beta sheet adopted by residues F104, V105, A112, and F113, however. Rather it seems that S103 and R102 fold across F113 and A114, bridging the N-terminal strand of the Cys-Cys loop to the C-terminal strand.

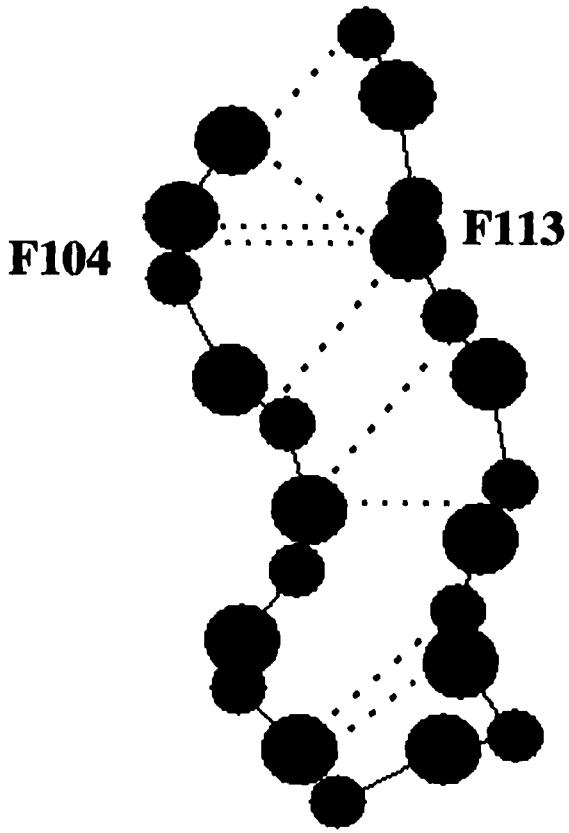
Figure 51. Cross-strand NOEs in the Cys-Cys loop of finger 1 of free and DNA-bound ADR1z'.

Both free and DNA-bound ADR1z' exhibit a number of similar NOEs across the finger 1 Cys-Cys loop that are consistent with the structure observed for ADR1b. The end of the Cys-Cys loop differs between free and DNA-bound ADR1z', however. Extra cross-strand NOEs in the bound protein suggest that S103 and R102 lie adjacent to A114, perhaps crossing over this residue and its strand. In contrast, the ADR1b structure found these residues folded back against the loop to interact with A112 and V105. This apparent extension of the Cys-Cys loop in bound ADR1z' is the only change to the zinc finger structure evident upon DNA binding.

ADR1z'



ADR1z'-DNA



III. ^{1}H Chemical Shift Perturbations upon DNA Binding

A. ZINC FINGERS AND BASE CONTACTS

Although the structure of the zinc fingers does not appear to undergo dramatic changes upon ADR1z' DNA binding, contact with DNA is expected to induce perturbations in ^{1}H chemical shift at the intermolecular interface. In fact, a large number of resonances are affected in the presence of DNA; perturbations of NH, H^{α} and H^{β} are shown in Figures 52, 53 and 54.

The comparisons of chemical shift are made under different conditions of pH and temperature for the free and the bound ADR1z', due to limitations of sample handling. Spectra of ADR1z' were assigned at pH 5.4 and 30 °C to minimize the facile oxidation of this species. Spectra of the more oxidation-resistant ADR1z'-DNA complex were acquired at pH 7.0 and 37 °C, due to improved solubility at high pH and to improved signal from sharper lines at high temperature. The necessity of working under such differing conditions could potentially complicate comparisons of the free and bound ADR1z' polypeptide.

To evaluate the validity of these comparisons, a simple spectrum of the free protein was acquired under identical

Figure 52. Chemical Shift Perturbations upon DNA Binding: NH Shift.

The difference of NH chemical shift between DNA-bound and free ADR1z' is plotted. Dramatic chemical shift perturbations are seen for the N-terminus and finger 1. Finger 2 experiences lesser changes, concentrated at the extreme N-terminus of the helix. Positions thought to make specific contacts to DNA are marked with an asterisk. The perturbations in the N-terminus fall off at exactly the border of the region that was defined by deletion mutagenesis as the minimal DNA binding domain.

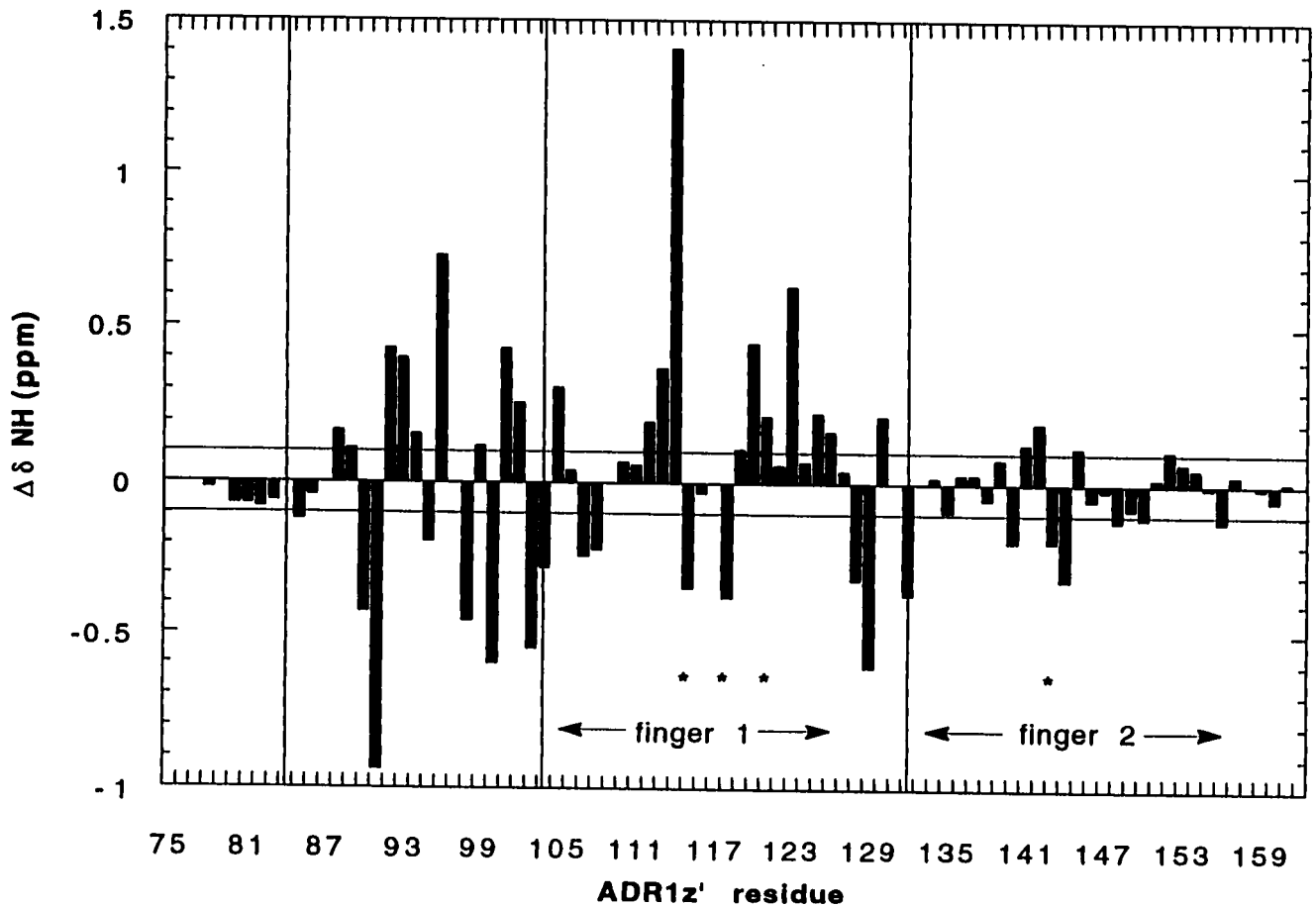

1.5
1
0.5
0
-0.5
- 1
Δδ NH (ppm)
finger 1
finger 2
75 81 87 93 99 105 111 117 123 129 135 141 147 153 159
ADR1z' residue

Figure 53. Chemical Shift Perturbations upon DNA binding: H^{α} Shift.

As in Figure 52, finger 2 is much less perturbed than finger 1 and the N-terminus. The boundary of the severe perturbations in the N-terminus is the same as for NH shift.

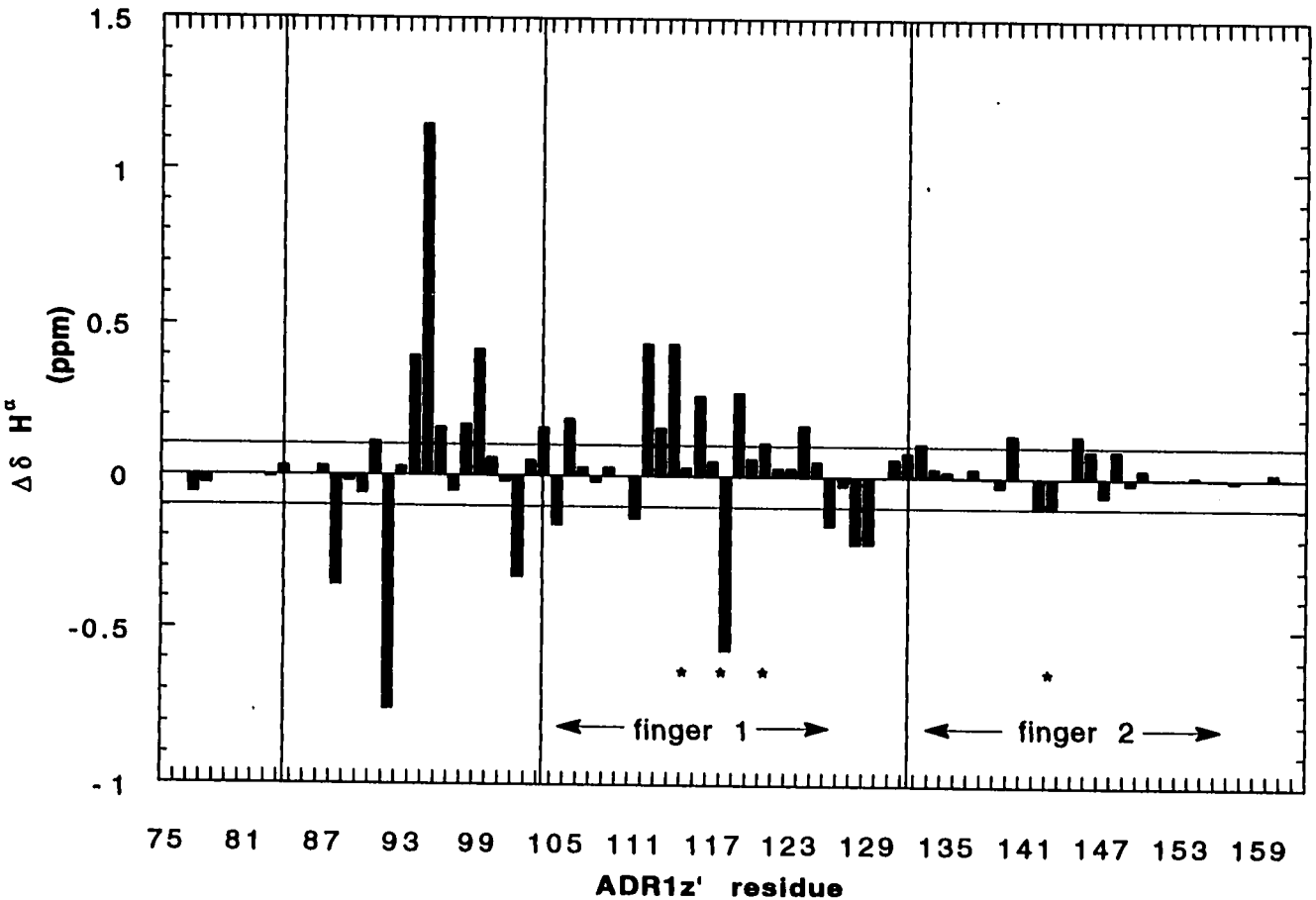
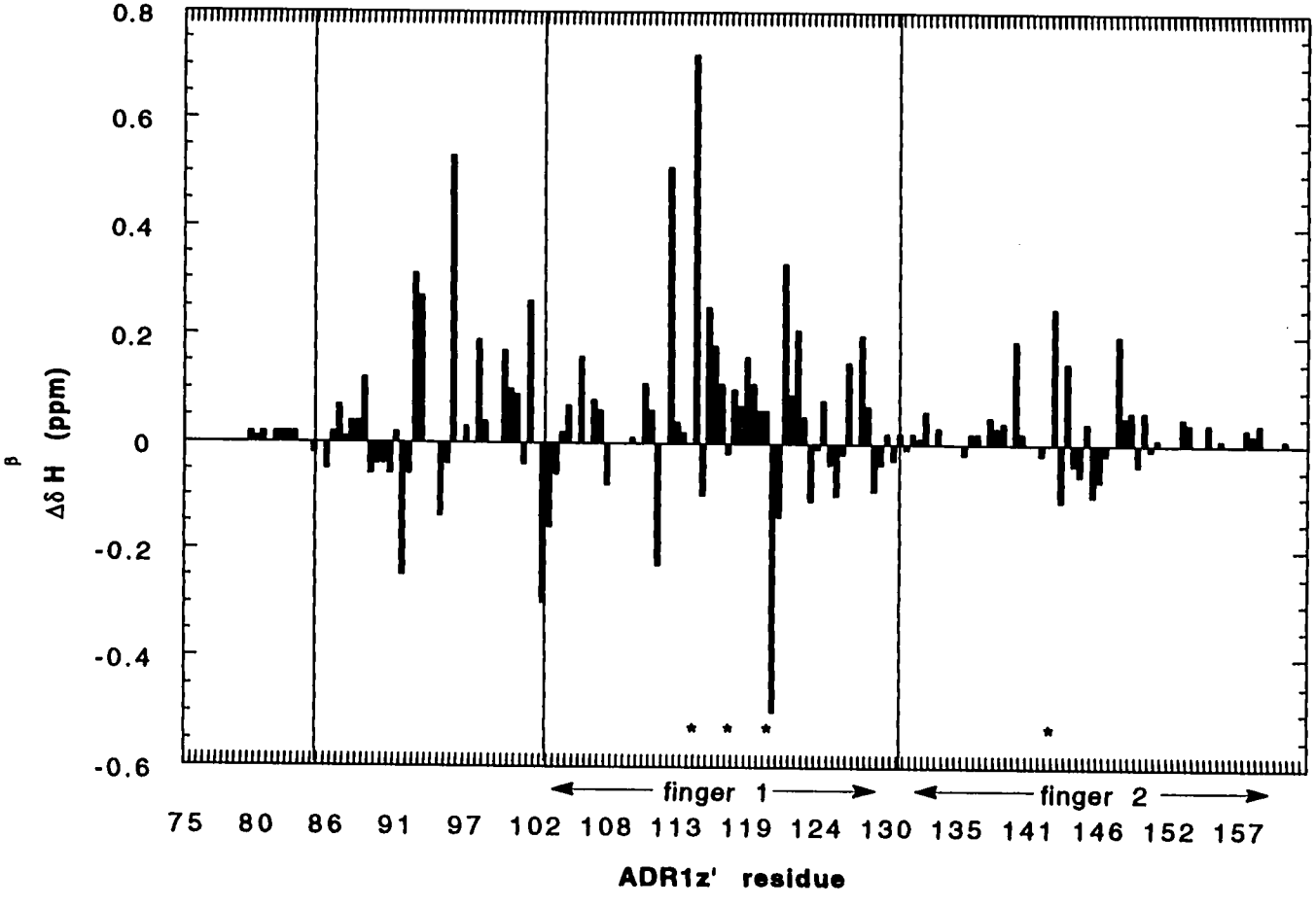

1.5
1
0.5
0
-0.5
-1
Δδ Hα (ppm)
finger 1
finger 2
75 81 87 93 99 105 111 117 123 129 135 141 147 153 159
ADR1z' residue

Figure 54. Chemical Shift Perturbations upon DNA binding: $H\beta$ Shift.

These data support the conclusions drawn from Figures 52 and 53. In addition, especially severe perturbations can be seen at some of the residues thought to make specific contacts (marked by asterisks).


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Δδ H (ppm)
β
finger 1
finger 2
75 80 86 91 97 102 108 113 119 124 130 135 141 146 152 157
ADR1z' residue

conditions to the ADR1z'-DNA assignment spectra. This HSQC spectrum is shown in Figure 55, overlapped with the spectrum of ADR1z' taken at a more typical lower pH and temperature. The spectrum at high pH and temperature suffers slightly from peak doubling and loss caused by oxidation, which is mostly reversed by addition of large excesses of DTT. Besides this effect of oxidation, very little change is noted in the chemical shifts of the protein. The largest proton chemical shift perturbation noted is about 0.7 ppm, just over the apparent proton linewidth of the peaks. Amide NH resonances are exquisitely sensitive to changes in environment; the equivalence of the pH 5.4/22 °C HSQC spectrum with the pH 7.0/37 °C spectrum indicates the validity of comparisons across this pH and temperature range.

Figure 56 presents a classification of alpha proton and sidechain perturbations. Each residue is placed in a category either representing no significant perturbation, with no sidechain or alpha proton resonance experiencing a change of greater than 0.1 ppm; severe perturbation, with an average alpha proton and sidechain alteration of at least 0.2 ppm or an individual alpha proton or sidechain proton change of at least 0.3 ppm; or moderate perturbation, for all cases in between.

Figure 55. HSQC Spectra of ADR1z' at High versus Low pH and Temperature.

^{15}N-HSQC spectra of ADR1z' at pH 7.0, 37 °C (red) and at pH 5.4, 22 °C (green) are overlapped for comparison. It is apparent that the spectrum of ADR1z' is quite insensitive to pH and temperature effects; the largest change in proton chemical shift observed is about 0.07 ppm. This fact makes it feasible to compare work with the free protein at low pH and temperature to work with the DNA-bound protein at high pH and temperature. Extra peaks observed in the red spectrum are due to doubling at the extreme C-terminal residues that is accentuated at high temperature and to slight oxidation of the sample at high pH and temperature.

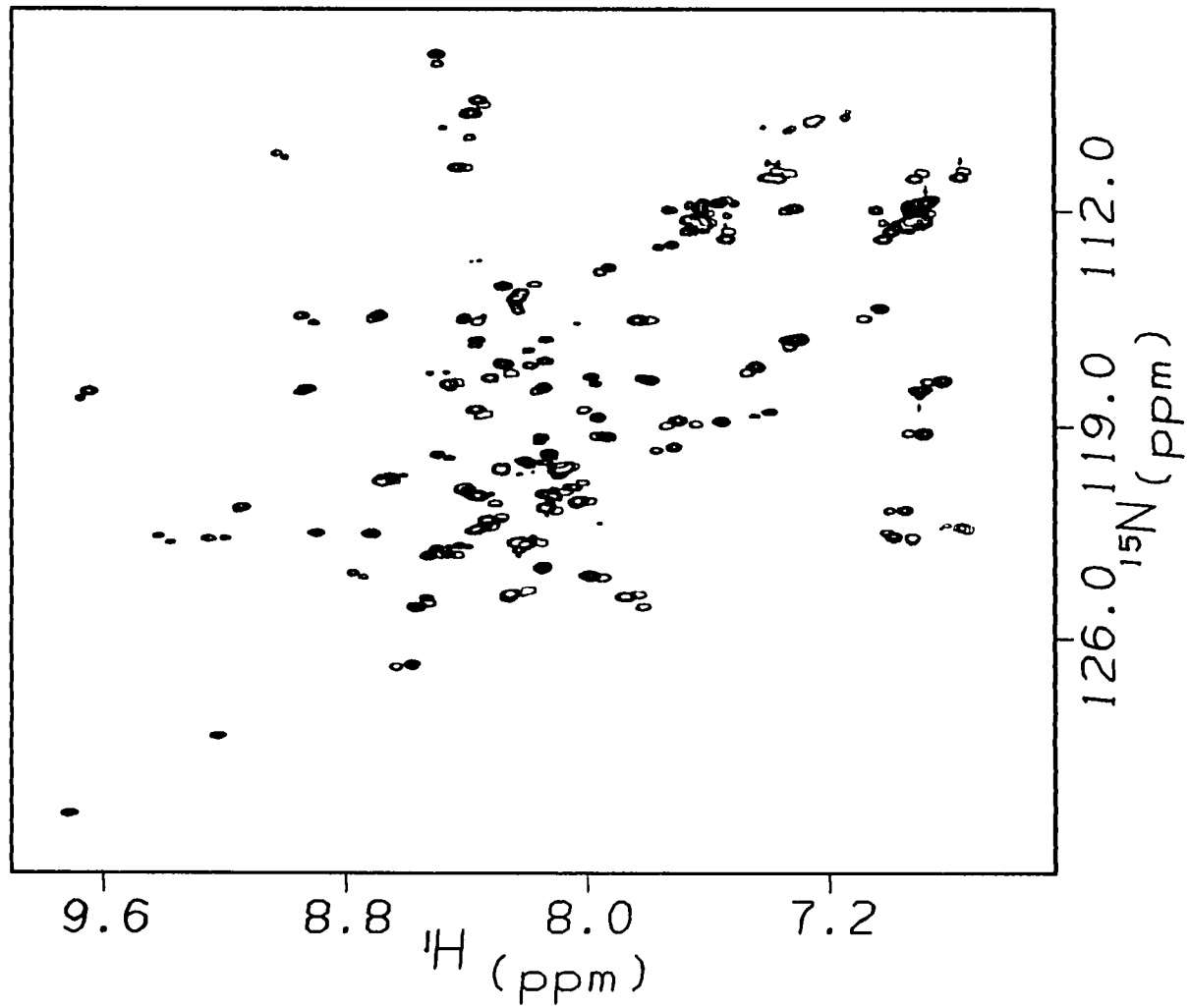
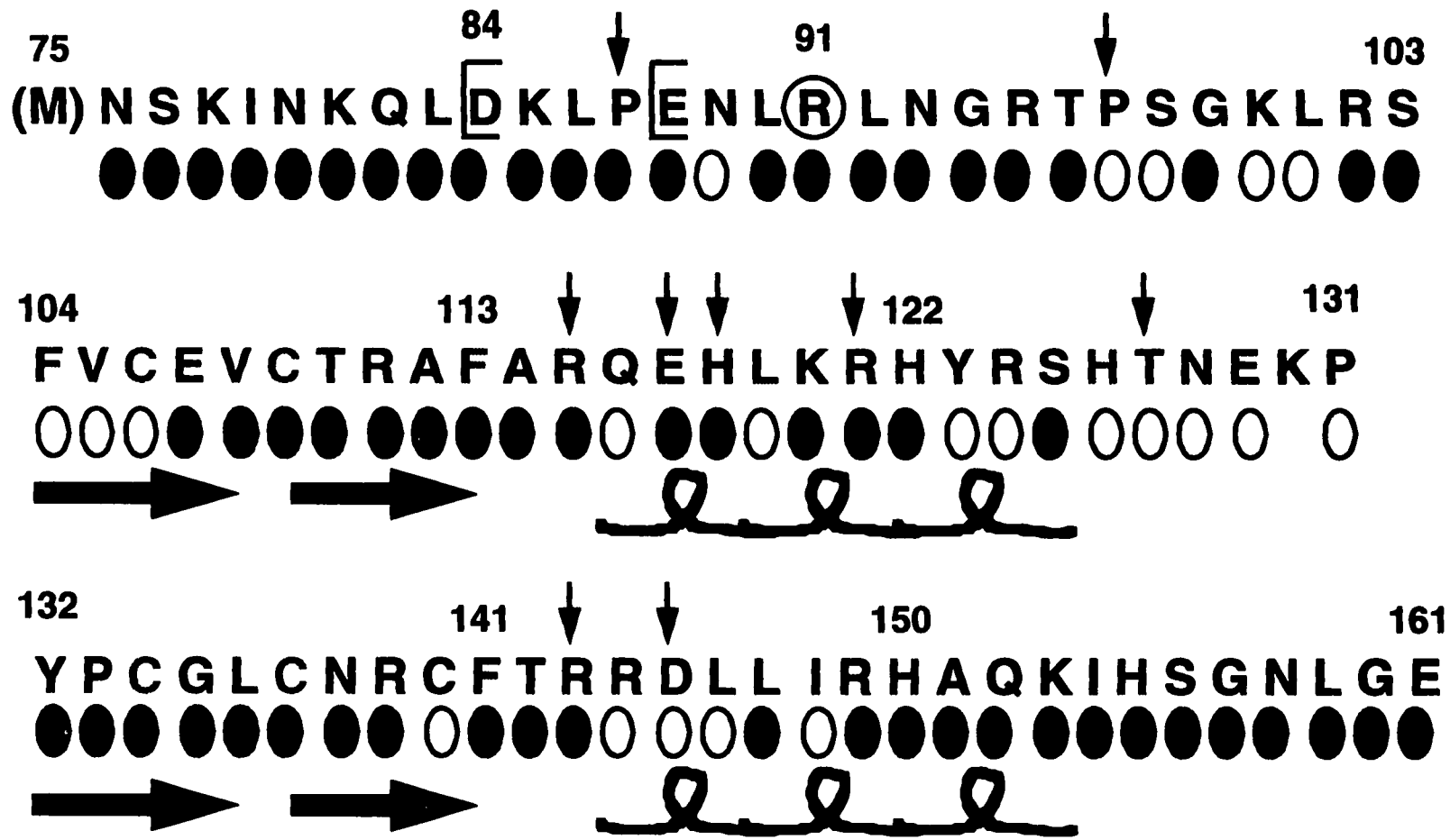

9.6
8.8
8.0
7.2
^1H (ppm)
126.0
119.0
112.0
^{15}N (ppm)

Figure 56. Chemical Shift Perturbations upon DNA Binding: Alpha and Sidechain Protons.

Chemical shift perturbations to the alpha protons and sidechain protons of ADR1z' upon DNA binding are mapped against the sequence. Red circles represent severely perturbed residues, each with an average residue chemical shift change of at least 0.2 ppm or individual resonance perturbations of at least 0.3 ppm. Blue circles represent unperturbed residues, none of which has any alpha or sidechain proton resonance perturbed by more than 0.1 ppm. All other residues are moderately perturbed, and are represented by yellow circles. Brackets mark the extent of the shortest construct that binds DNA with high affinity and the longest construct that does not. R91, the position of a suppressor mutation, is circled. Arrows mark positions where mutations have been isolated that are deleterious to DNA binding; red arrows mark residues thought to make specific contacts to DNA.

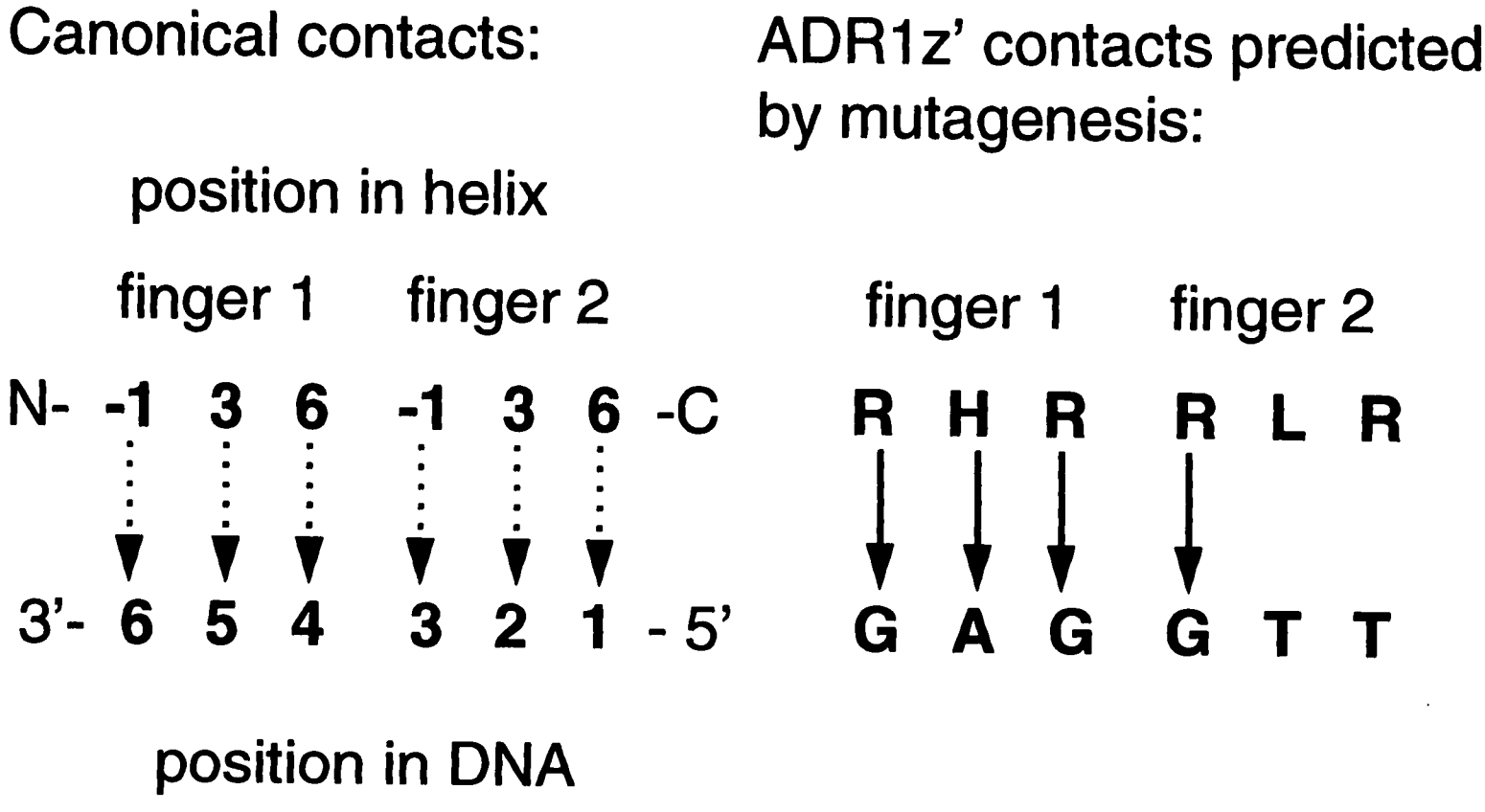

75
84
91
103
(M) N S K I N K Q L D K L P E N L R L N G R T P S G K L R S
104
113
122
131
F V C E V C T R A F A R Q E H L K R H Y R S H T N E K P
132
141
150
161
Y P C G L C N R C F T R R D L L I R H A Q K I H S G N L G E

In the zinc fingers, among the most severely perturbed sidechains are those of R115, H118, R121, and R143. Each of these positions has been implicated in direct contact to DNA bases. In the X-ray crystal structures of zinc finger-DNA complexes available, the majority of base-specific contacts are provided by the sidechains of residues at positions 6, 3, and -1 of the α-helix (20, 26, 28, 29). In ADR1, these positions correspond to residues 121, 118, and 115 of finger 1 and residues 149, 146, and 143 of finger 2 (Figure 57). Mutagenesis of ADR1 showed that mutation of residue 115, 118, 121 or 143 to alanine resulted in dramatically reduced or undetectable DNA binding in EMSA studies and loss of transactivation *in vivo* (54). Furthermore, mutations were created at 115, 118 and 121 that alter the binding site preference of the ADR1 protein, suggesting that these residues are responsible for direct base contacts (56). The ^{1}H chemical shift perturbation map of ADR1z' lends further support to the idea that the sidechains of R115, H118, R121, and R143 make close approach to the DNA at the intermolecular interface.

Other severely perturbed sidechains are those of a zinc ligand, H122, and a conserved arginine, R111. As shown in Figure 56, the sidechain protons of H122 are perturbed upon DNA binding; as will be discussed in a later section, the imidazole ring nitrogens of H122 are also affected. In the Zif268 structure the analogous

Figure 57. Canonical Zinc Finger-DNA Contacts versus ADR1-DNA Contacts.

Residues commonly observed to make specific contacts in zinc finger-DNA cocrystal structures are shown on the left. Specific ADR1 contacts suggested by mutagenesis are shown on the right. While ADR1 makes use of residues at the canonical positions, the pattern of contacts, biased towards finger 1, is unusual.


Canonical contacts:
position in helix
finger 1
finger 2
N- -1 3 6 -1 3 6 -C
3'- 6 5 4 3 2 1 - 5'
position in DNA
ADR1z' contacts predicted by mutagenesis:
finger 1
finger 2
R H R R L R
G A G G T T

histidine and arginine in each finger contact the phosphodiester backbone of the DNA (20). Severe perturbation of R111 and H122 in ADR1z' suggests that similar contacts may be made by finger 1. The ethylation interference pattern for ADR1 binding to UAS1 suggests backbone contacts in the binding subsite of finger 1 that are consistent with a role for R111 and H122. In particular, ethylation of the phosphate immediately 5' to the finger 1 subsite interferes with ADR1 binding (59). In several zinc finger cocrystal structures a similar position is contacted by residues analogous to R111 (20, 26, 28, 29); in one of these structures, this position is also contacted by the analogue to H122 (29).

B. UNUSED CANONICAL CONTACT RESIDUES

Residues L146 and R149 of ADR1 fall at canonical base-contacting positions 3 and 6 of the helix of finger 2 (Figure 57). Mutation of these residues to alanine, however, causes little or no change in the DNA-binding affinity of ADR1 (54). These results suggest that in finger 2, only the -1 position of the helix contributes a canonical DNA contact. This situation stands in contrast to other zinc fingers whose binding residues are known from X-ray structures. In each of these fingers, two of the three canonical binding positions of the helix are typically utilized to

make specific contacts to DNA (20, 26, 28). Finger 2 of ADR1 appears to be unusual in the use of only a single specific contact. This view is strongly supported by the ^{1}H chemical shift perturbation profile of ADR1z', in which L146 is only moderately perturbed, while the sidechain of R149 is unaffected by DNA binding (Figure 56). Although arginine sidechains are ideal for hydrogen bonding to DNA bases, finger 2 appears not to make use of R149 for any sort of contact.

The differences in intermolecular interface with DNA of finger 1 and finger 2 are seen throughout the C-terminal portion of the fingers (Fig. 52-54, 56). Although $^{13}C^{\alpha}$ chemical shifts indicate that the structure of the zinc fingers is probably minimally altered upon DNA binding, ^{1}H chemical shift changes are found throughout the C-terminal half of finger 1. These changes include severe perturbations at the sidechains of residues 115, 118, 121, and 122 and moderate perturbations at most other positions all the way through the last histidine ligand and into the linker. This would seem to indicate full involvement of the helical region of finger 1 with DNA, in line with use of all three canonical positions to make specific base contacts. In contrast, finger 2 experiences severe chemical shift perturbation only for the base contact R143, and moderate perturbations only from R144 through I148. No perturbations are seen for the remainder of the finger, including

the unused canonical contact position, R149. In contrast to finger 1, it seems that only a portion of the helix of finger 2 is in proximity to the DNA, and that close approach may be made only by the N-terminus of the helix. Unlike the similar dispositions of the three fingers of Zif268, each of which make two canonical contacts to DNA (20), the chemical shift data presented here suggest that the fingers of ADR1 approach the DNA with differing orientations.

The minimal contact by finger 2 may include a lack of typical contacts to the phosphodiester backbone. In the Zif268 structure, residues analogous to R111 and H122 of ADR1z' finger 1 and R139 and H150 of finger 2 contact the DNA backbone (20). The finger 1 residues show severe perturbations upon DNA binding, consistent with DNA contact (Figure 56). The finger 2 residues, however, are unperturbed, suggesting that they may not make contact, perhaps due to an unusual finger orientation.

It has been suggested that leucine 146, at position 3 of the helix, may be responsible for the different pattern of contact of finger 2 (13). While this residue does not appear to make an important DNA contact in the wild-type protein (54), mutation at this position can lead to an altered specificity of DNA binding (79). Specifically, the mutation L146H leads to a strong base preference at the UAS1 position predicted to be contacted by helical position 3, where only a weak preference existed before. More surprising,

this mutation also leads to a newfound preference at the base predicted to be contacted by R149, at helical position 6 (79). It appears that the presence of the wild-type leucine at position 3 of the helix prevents interaction of R149 with the DNA, and that this interference can be removed by altering L146. This result was anticipated by a modeling study in which the structure of the ADR1 finger 2 PAPA mutant was superimposed on finger 3 of the Zif268 cocrystal structure (13). The resulting model predicted a steric clash between L146 and the DNA, suggesting that the helix of finger 2 does not approach the major groove in the Zif268 orientation. This model is strongly supported by the ^1H chemical shift perturbation map of ADR1z'-DNA.

The idea that some zinc fingers can make use of their entire helix to contact DNA, while others cannot has been put forward in a general way by Suzuki *et. al.* (80). They have categorized Cys_2-His_2 fingers as either "good" (likely to use the whole helix for contacts) or "bad" (unable to do so), based on the nature of residues at the N-terminal end of the helix. It was suggested that an abundance of short residues in this area would make for a "good" finger, while longer or bulky residues in this area would create a "bad" finger that would make suboptimal contacts. By this criterion, L146 may render finger 2 of ADR1 a "bad" finger; there is now direct evidence

from two separate techniques, mutagenesis and NMR spectroscopy, to demonstrate the minimal contact made by finger 2.

C. UNPERTURBED PORTIONS OF FINGERS

The chemical shift perturbations diagrammed in Figure 56 are also mapped in three dimensions in Figures 58 and 59, where the zinc fingers of ADR1z' have been modeled onto fingers 2 and 3 of the Zif268 cocrystal structure. In the structures of zinc finger-DNA complexes solved thus far, contacts to the DNA arise from one face of the α-helix, while the Cys-Cys loop lies on the "back" of the helix, away from the DNA (20, 26, 28, 29). The chemical shift perturbation map bears out this trend for ADR1, with no perturbation to the sidechains of the Cys-Cys loop of finger 2 and the reverse turn of finger 1, and only moderate perturbation to most of the β-sheet of finger 1 (Figure 58). The severe perturbations that do occur at R102, S103, A112, F113, and A114 and the moderate perturbations at F104 and V105 are likely due to an extension of the interactions across the Cys-Cys loop of finger 1, as discussed earlier (see Figure 51). This change in finger 1 structure likely involves a move of residue S103 from a position near V105 and A112 to a position adjacent to F113 and A114.

Figure 58. Chemical Shift Perturbations upon DNA Binding: 3D Map

The chemical shift perturbaton data of Figure 56 are mapped on a model of the ADR1 zinc fingers bound to DNA. Green is substituted for yellow to represent moderately perturbed residues in this figure, while red represents severely perturbed residues and blue represents unperturbed residues. It is apparent that the most severely perturbed residues lie against the DNA in this model, while the exposed portions of the fingers are less perturbed. The model was generated by substituting ADR1 sequence for Zif268 sequence in the second two fingers of the Zif268 cocrystal structure.

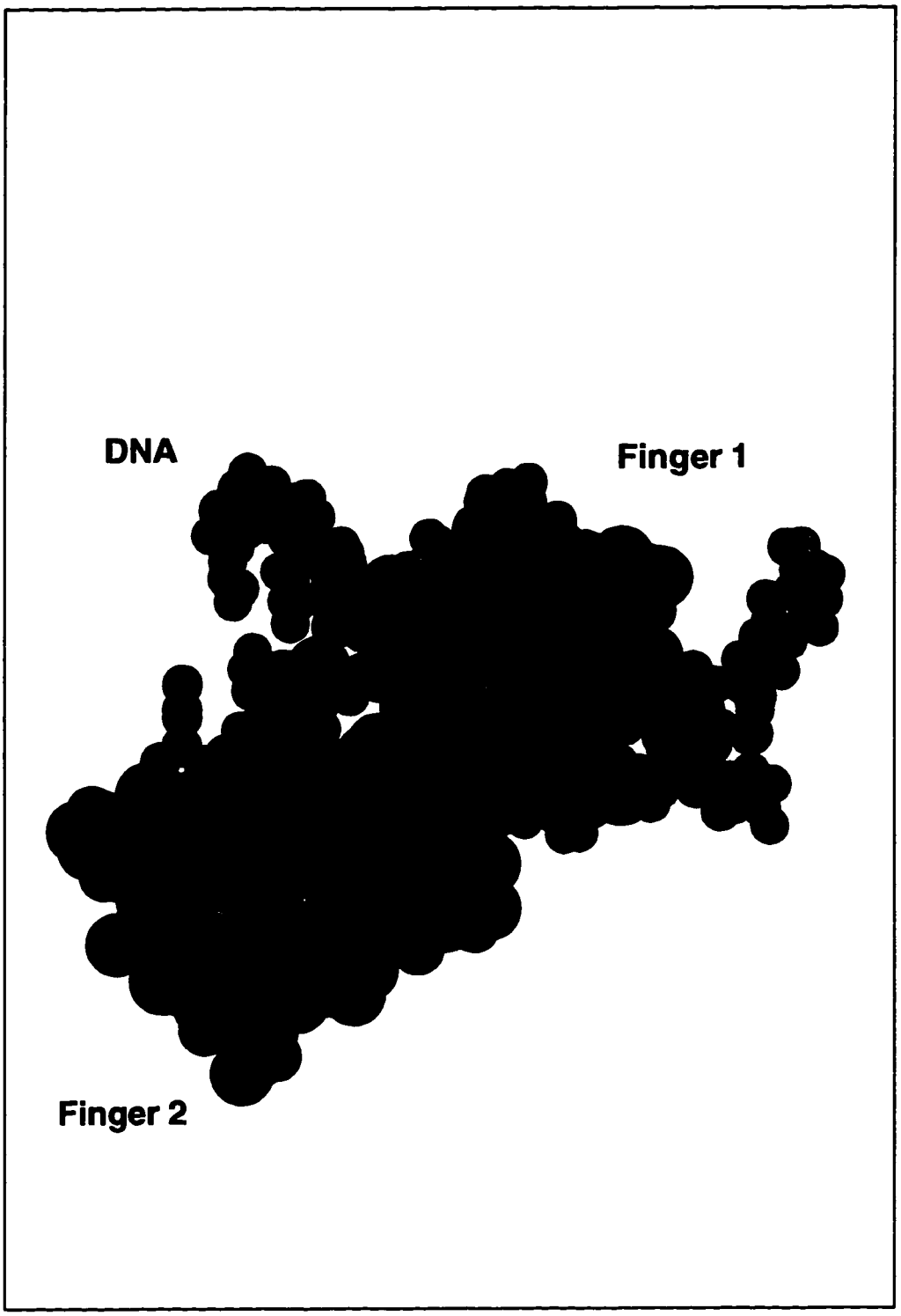
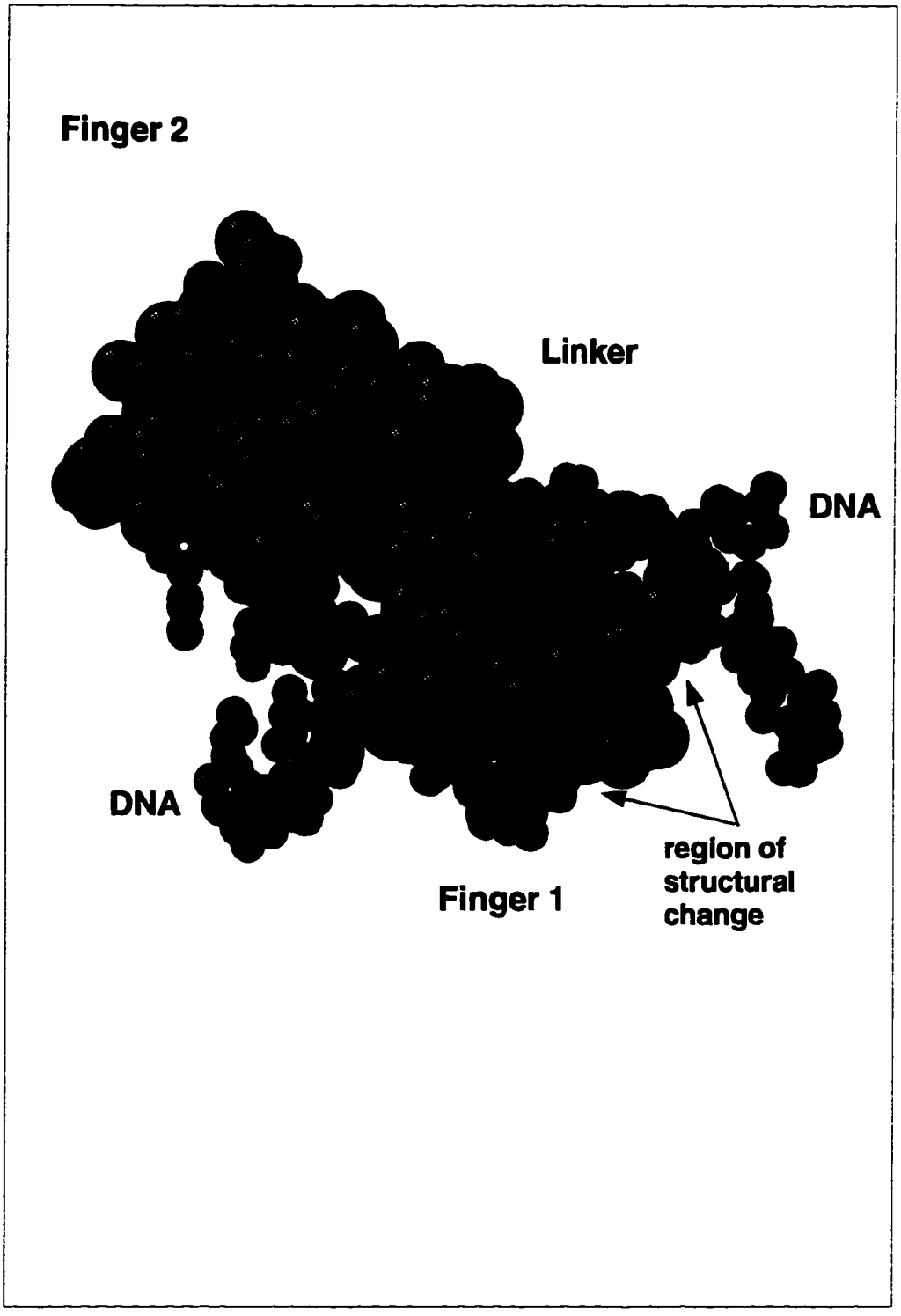

DNA
Finger 1
Finger 2

Figure 59. Chemical Shift Perturbations upon DNA Binding: 3D Map, Rotated 180°.

The model from Figure 58 is rotated 180° to show the exposed surfaces of the fingers. As in Figure 58, the fingers in this model are largely unperturbed. The severely perturbed residues visible are S103, F113 and A114, which fall in the region of Cys-Cys loop that is extended upon DNA binding. The swath of moderately perturbed residues (green) is the linker region between fingers, as shown in Figure 60.

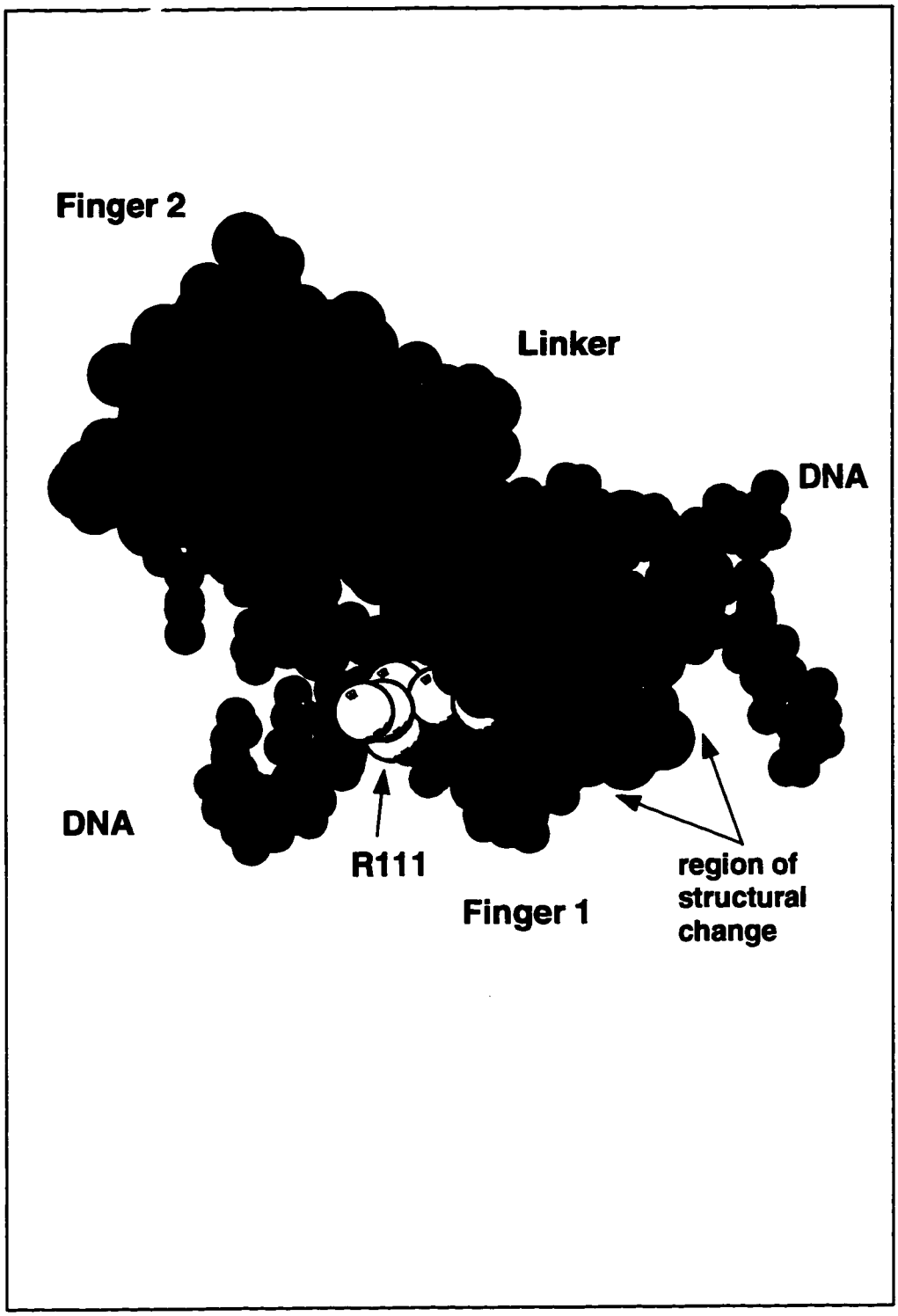

Finger 2
Linker
DNA
DNA
region of
structural
change
Finger 1

A further conclusion from the lack of chemical shift perturbations in the portions of the fingers oriented away from the DNA is that the N-terminus, which is itself severely perturbed, does not interact with these regions. When perturbations in the extended Cys-Cys loop of finger 1, in the interfinger linker, and in predicted DNA backbone contact R111 are accounted for (Figure 60), little of the exposed finger structure is perturbed at all, and none is severely perturbed, as is much of the N-terminus.

It is a formal possibility that the effect of the N-terminus on DNA binding affinity is mediated through interaction of this region with the fingers themselves, allowing them to fold properly, as was the case with the third strand of sheet in SWI5 and Tramtrack, or increasing their ability to bind to DNA. These possibilities seemed unlikely, however, given that each ADR1 finger folds in the absence of flanking sequences to a structure very much like the active binding form observed in zinc finger-DNA complexes (20, 26, 28), and given that in the unbound form, the N-terminal flanking sequences are random coil and make no interaction with the fingers (Figure 43). Now the bound ^{1}H chemical shift map of ADR1z' also tends to rule out an interaction of the N-terminus with the exposed portions of the fingers in the complex, rendering it more likely that the true role of the N-terminus is to interact with the DNA itself and thereby increase binding affinity.

Figure 60. Chemical Shift Perturbations upon DNA Binding: Linker Residues and R111.

The view of Figure 59 is used again, with some changes to the residue color map. As in Figures 56, 58, and 59, red residues are severely perturbed and blue residues are unperturbed. As in Figures 58 and 59, green residues are moderately perturbed. Additionally, linker residues and residues from the extension to the finger 1 Cys-Cys loop are colored magenta. It is clear from a comparison of this figure to Figure 59 that most of the exposed, perturbed residues arise from these two sites of obvious change upon DNA binding. Other portions of the fingers are unperturbed, suggesting that the N-terminus does not contact the exposed face of the fingers. It is possible, however, that the N-terminus does contact R111, which is quite exposed in the model (yellow). This residue is moderately perturbed, and its analogue in Zif268 makes a contact to the phosphate backbone.


Finger 2
Linker
DNA
DNA
R111
Finger 1
region of
structural
change

D. N-TERMINUS

Just as the chemical shift perturbation map for the zinc fingers of DNA-bound ADR1z' closely mirrors the results of mutagenesis experiments, the pattern of chemical shift perturbations for the N-terminal flanking sequence fits well with previous observations. Deletion analysis had shown that truncation of the N-terminus from residue 84 back to residue 88 made the difference between high affinity and non-detectable DNA binding (53). As discussed above, $^{13}C^{\alpha}$ chemical shifts revert to random coil in the ADR1z'-DNA complex in just this critical region, N-terminal to residue 85. The ^{1}H chemical shift map also reflects the importance of the N-terminus from E88 onward, as most residues following this critical point are moderately or severely perturbed, while residues preceding this juncture are not perturbed (Figure 56).

Large perturbations to NH and H^{α} chemical shift are found throughout the N-terminus after residue E88; these could result from the acquisition of secondary structure from random coil and/or from proximity to DNA. In this region, sidechains are also perturbed, including most severely those of residues N93 and T96. As sidechains are less liable to perturbation due to acquisition of

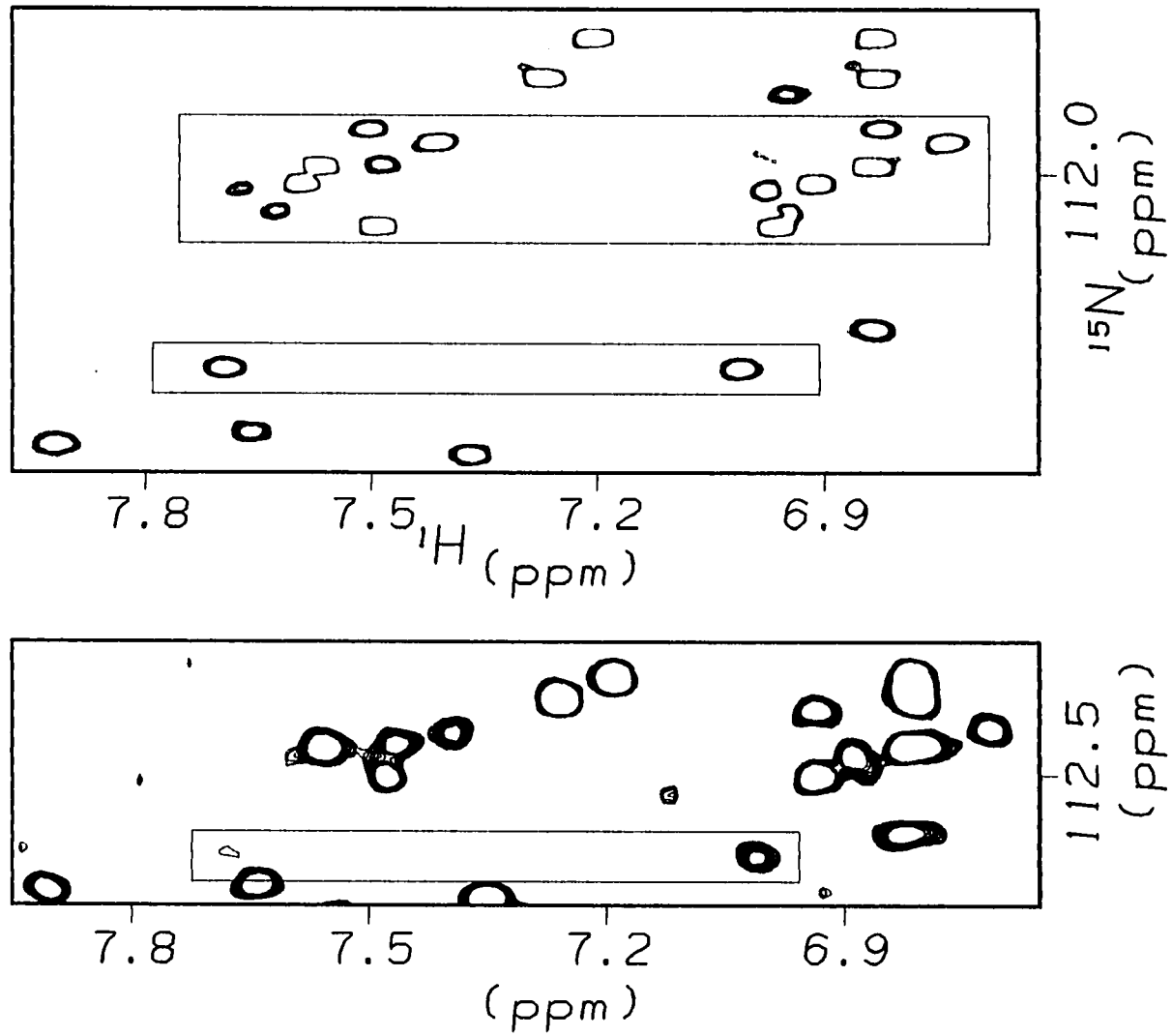
secondary structure than are backbone resonances, these perturbations may indicate involvement in the binding interface. Mutagenesis results suggest the importance of this region as well. Mutation of P87 or P97 to leucine causes a reduction in DNA binding affinity, (53) either through disruption of some necessary structure or through local disruption of a DNA contact.

In addition to the abundant ^1H chemical shift changes in the N-terminus, the sidechain amide nitrogen of N89 is perturbed out of the usual range of chemical shifts for sidechain NH_2s upon DNA binding, by a downfield shift of about 2 ppm (Figure 61). These perturbed peaks are less sharp than those of other sidechain NH_2s. Examination of HSQC experiments taken at high pH also indicate that one of the protons of this sidechain NH_2 group is less liable to exchange with solvent than the other, suggesting differential protection from solvent (Figure 61). Such protection could arise from protein-protein interactions, or from protein-DNA contacts.

N89 and the set of highly perturbed sidechains extending from L92 to T96 fall in exactly the region where a mutation was found that can suppress defects in DNA binding arising from mutations in either zinc finger (58). This suppressor mutation, R91K, also increases DNA binding by about two- to threefold in a wild-type background. Mutation of R91 to glycine decreases DNA binding affinity by less than two-fold. The R91 sidechain does not

Figure 61. The Sidechain NH_2 Group of N89 is Protected from Solvent in ADR1z'-DNA.

The region of an ^{15}N-HSQC of ADR1z'-DNA that contains sidechain amide protons is shown in the top panel. The sidechain amide of N89 is shifted about 2 ppm downfield in nitrogen from the tight grouping of other Asn sidechain amides. A similar spectrum at higher pH (7.7 versus 7.0) is shown for comparison below. At the elevated pH, one of the sidechain amide protons of N89 exchanges very efficiently with solvent and disappears from the spectrum. The other, upfield proton of this sidechain group exchanges less rapidly, and remains in the spectrum. This behavior indicates differential protection of the two N89 sidechain amide protons from solvent.


112.0
^{15}N (ppm)
7.8
7.5
7.2
6.9
^{1}H (ppm)
112.5
(ppm)
7.8
7.5
7.2
6.9
(ppm)

experience a significant chemical shift perturbation upon DNA binding, however, suggesting that the sidechain is not involved in DNA contact in the wild-type construct. Mutation to lysine likely allows the formation of a novel contact that contributes to the overall binding energy. Mutation to glycine may decrease binding energy simply because of the increased entropy of the free polypeptide due to the conformational flexibility of glycine.

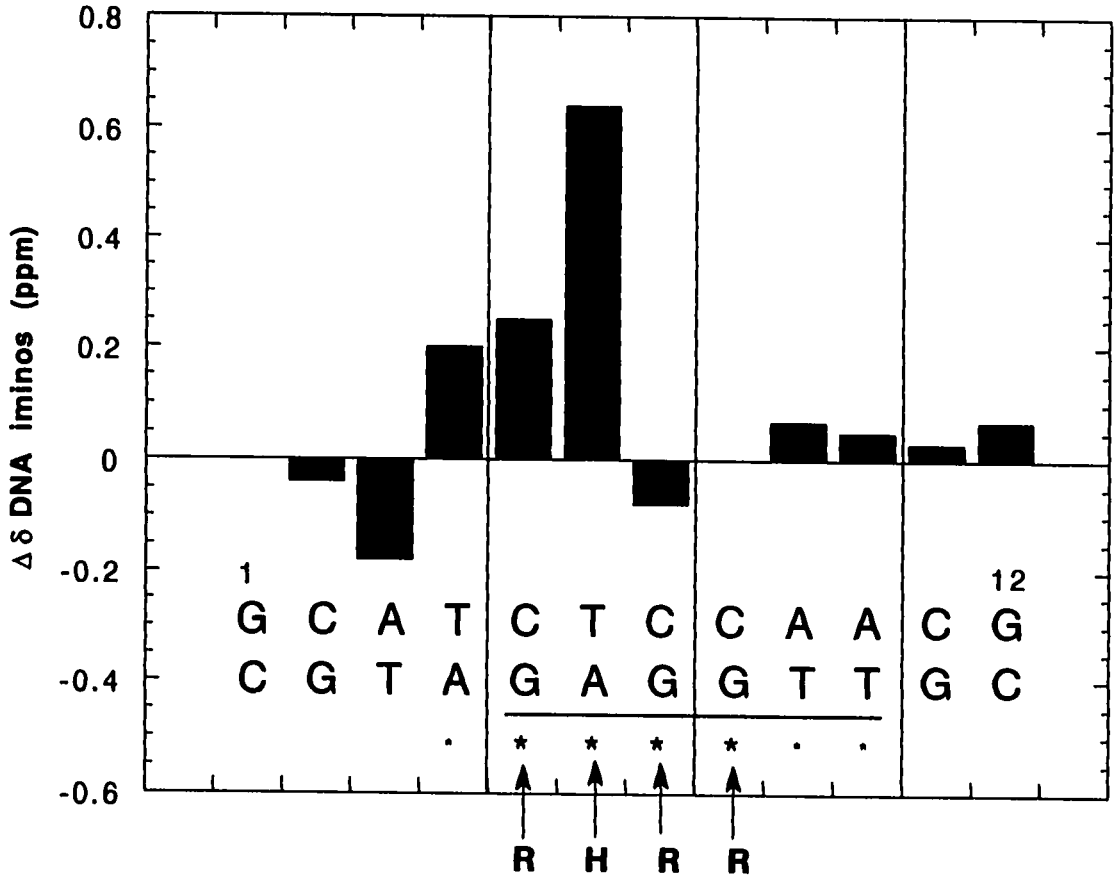
Thus far both mutagenesis results and chemical shift perturbation maps point to the region from P87 to P97 as critical to the DNA-binding contribution of the N-terminus.

IV. DNA Site

Chemical shift perturbations to 14mer DNA upon ADR1z' binding are mapped for the imino protons in Figure 62. The bases that are critical for specific ADR1 binding have been determined by mutagenesis, and are indicated in the figure (59). In X-ray structures of finger-DNA complexes, there is often a one-to-one correspondence of contacts between the -1, 3 and 6 positions of the helix and bases in a triplet in the binding site (Figure 57) (20, 26, 28). Based on this precedent, ADR1 finger 1 might be predicted to bind GAG in the halfsite, using R115, H118 and R121. Finger 2 could be predicted to bind GTT with R143, L146 and R149.

Figure 62. Perturbations to the 14mer DNA Imino Protons upon ADR1z' Binding.

The chemical shift difference between bound and free 14mer imino resonances is plotted versus 14mer sequence. The majority of chemical shift changes fall in the binding subsite for finger 1, consistent with the notion of minimal finger 2 interactions suggested by mutagenesis and protein NMR work. The large downfield perturbation for T6 is consistent with a ring current shift due to binding of this basepair by aromatic His118. Asterisks mark basepairs for which ADR1 exhibits a sequence preference; the larger asterisks indicate a stronger preference.


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Δδ DNA iminos (ppm)
1
12
G C A T C T C C A A C G
C G T A G A G G T T G C
R H R R

Mutagenesis of the protein and DNA site has confirmed the importance of the GAGG sequence, in line with the demonstrated importance of residues 115, 118, 121, and 143 (54). In addition, changes to the preceding and following A/T basepairs are not entirely neutral (Figure 62) (59).

Only a few iminos of the DNA site show significant chemical shift changes upon ADR1z' binding. These could result from contacts by ADR1z', or from perturbation of the DNA structure induced by protein binding. In line with the notion of extensive contacts from finger 1 and very limited contacts from finger 2, the major changes in the DNA occur in the expected GAG binding site for finger 1. Additionally, perturbation of the upstream T confirms that this base pair, while not predicted to make contact with the finger by canonical rules, nevertheless plays a role in the site, as suggested by mutagenesis (59). The lack of perturbation for the final T of the predicted GTT subsite for finger 2 is consistent with the earlier conclusion that R149 does not contact the DNA.

Finally, the largest chemical shift perturbation, a downfield shift of 0.6 ppm, occurs at the T in the A/T basepair predicted to be contacted by H118. This relatively large shift may be due to a ring current effect of the aromatic histidine sidechain. Were the histidine ring to contact the A/T basepair edge-on, the imino

proton would fall in the downfield shift profile of the imidazole ring. The overall imino shift pattern is thus consistent with the notion gleaned from protein shifts that finger 1 is more intimately involved in contacting the DNA than is finger 2.

V. DNA Structure

The DNA in a large number of protein-DNA complexes exhibits a structure that is intermediate to B-form and A-form DNA (81). The major groove in this DNA is as wide as that in B DNA, but also nearly as deep as that in A DNA, and was thus termed $B_{enlarged\text{-}groove}$ DNA. Among the protein-DNA complexes in which the DNA adopts the B_{eg} form are the Zif268, GLI, and Tramtrack zinc finger-DNA complexes (81). Linking number change experiments using Sp1 demonstrate that plasmid DNA bound by Sp1 is underwound; this underwinding is relaxed into supercoils when Sp1 is removed from closed plasmid (82). Such unwinding is consistent with a change from B to B_{eg}-form DNA upon Sp1 binding.

Recently, circular dichroism (CD) experiments were done on the Zif268 zinc finger-DNA complex (25). The DNA signal was observed at wavelengths where protein signal does not interfere; the signal maximum was shifted to lower wavelengths by zinc finger binding. This change in CD spectrum is consistent with the

formation of an enlarged major groove dependent on the presence of zinc finger binding.

Similar CD experiments were done on the UAS1 halfsite in the presence and absence of DNA. The DNA spectrum, shown in Figure 63, also exhibits a shift to lower wavelength upon ADR1z binding. Thus it is likely that the UAS1 adopts the B_{eg} conformation when ADR1 binds it.

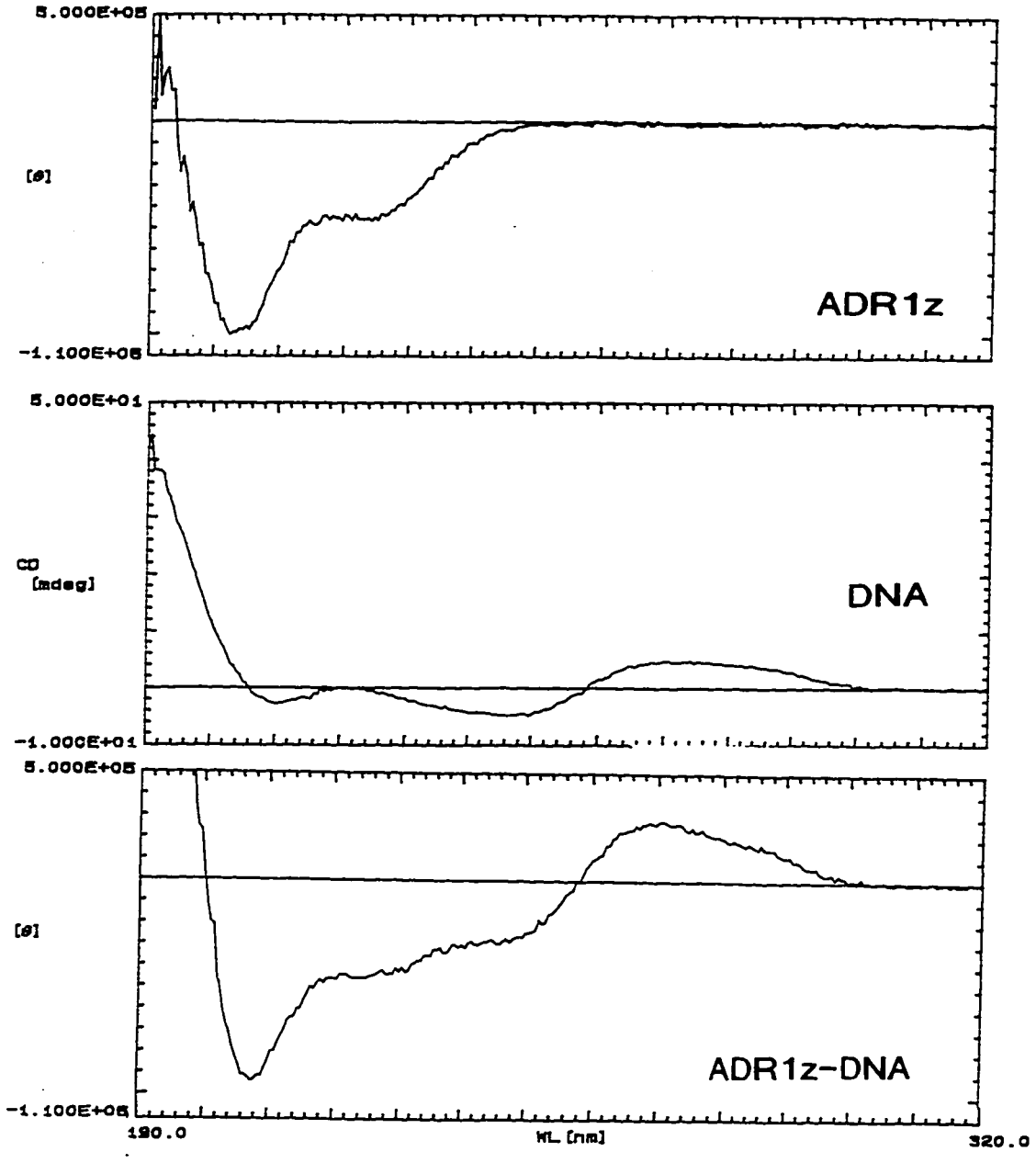
VI. Histidine Sidechain Imidazole Rings

Histidine sidechain imidazole rings can carry either one or two exchangeable protons at the N^{δ} and/or N^{ε} positions. At neutral pH a typical histidine is singly protonated at N^{ε} only. Such a histidine becomes doubly protonated below the typical sidechain pK_a of about 6.5 to 7.

NMR techniques exist that allow determination of the protonation state of histidine sidechains. In the long-range HMQC-J experiment, coupling evolves between imidazole nitrogens and carbon-bound C2 and C4 protons. An analysis of the strength of the couplings and the value of the nitrogen chemical shifts allows determination of the protonation state of the sidechain (83).

Figure 63. CD Spectra of ADR1z'-DNA.

CD spectra are shown for ADR1z, 17mer DNA containing a UAS1 halfsite, and an ADR1z-DNA complex. Samples were prepared by the NMR sample protocol, at pH 5.7 and 40 μM. Comparison of the top and center panels shows that at higher wavelengths only the DNA gives signal. This DNA signal shifts to lower wavelengths upon protein binding, as shown in the bottom panel. This type of shift is a hallmark of the formation of $B_{enlarged\text{-}groove}$ DNA structure.


5.000E+05
[θ]
-1.100E+05
ADR1z
5.000E+01
CD
[mdeg]
-1.000E+01
DNA
5.000E+05
[θ]
-1.100E+05
ADR1z-DNA
190.0
WL [nm]
320.0

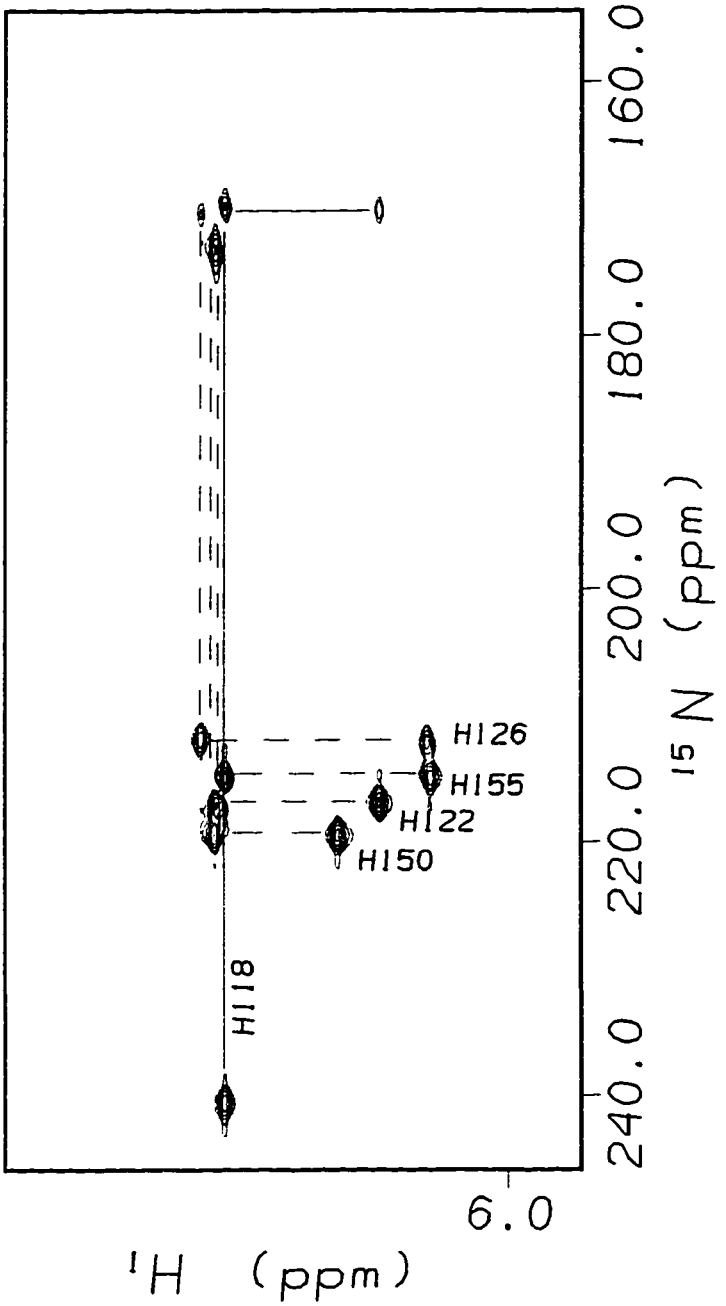
A long-range HMQC-J spectrum of free ADR1z' at pH 5.4 is shown in Figure 64. Four of the histidine residues show a similar, highly unusual tautomer pattern. These residues, H122, H126, H150, and H155, are the zinc ligands. The chemical shifts of the downfield, unprotonated nitrogens are perturbed at least 30 ppm upfield from the typical value of 249 ppm, as observed for the ligand nitrogens in other proteins where histidines coordinate zinc (84). The single proton is carried on the N^{δ} position, as opposed to the typical N^{ε} position. These data are entirely consistent with the use of the N^{ε} of each of these residues to coordinate zinc.

The remaining, quite different tautomer pattern belongs to H118. This pattern is less unusual than those of the zinc ligands, with the single proton carried on the N^{ε} position, and the unprotonated N^{δ} chemical shift a more typical 241 ppm.

The surprising aspect of this pattern is that H118 is not in the charged, doubly protonated form at pH 5.4; the pK_a of H118 appears to be more than 1 ppm below the typical pK_a of histidine residues. It had previously been shown that the zinc ligands are not protonated above pH 5.0 (85). The unusually low pK_a of these residues is likely due to zinc binding; protonation at N^{ε} actually leads to loss of zinc and unfolding of the fingers. The low pK_a of H118 is not as easily explained, as this residue is not a zinc ligand.

Figure 64. Histidine 118 is not Protonated at Low pH in ADR1z'.

The long-range HMQC-J experiment allows determination of the tautomeric form of histidine imidazole rings. The experiment for free ADR1z' at pH 5.4 and 37 °C is shown in this figure. Four histidines exhibit anomalous N_ε chemical shifts; these are the zinc ligands. Each of these histidines displays the peak pattern indicative of unusual single protonation of the imidazole at the N_δ position. This is consistent with previous observations from x-ray crystallography that the N_ε positions are used to coordinate zinc. The single histidine with typical nitrogen chemical shifts is H118. This residue also shows the more usual single protonation at the N_ε position. H118 is not average in every way, however, as its pK_a is quite low. None of the zinc ligands nor H118 is doubly protonated to the charged form at pH 5.5, in contrast to a typical histidine with a pK_a of well above 6.0.


160.0
180.0
200.0
220.0
240.0
^{15}N (ppm)
H126
H155
H122
H150
H118
6.0
^{1}H (ppm)

However, H118 lies near the N-terminus of a helix that is highly basic in nature. In this region there are several positively charged residues, including R115, K120, R121, and R124. These residues lie on approximately the same face of the helix as H118, and may render protonation of the imidazole sidechain less favorable.

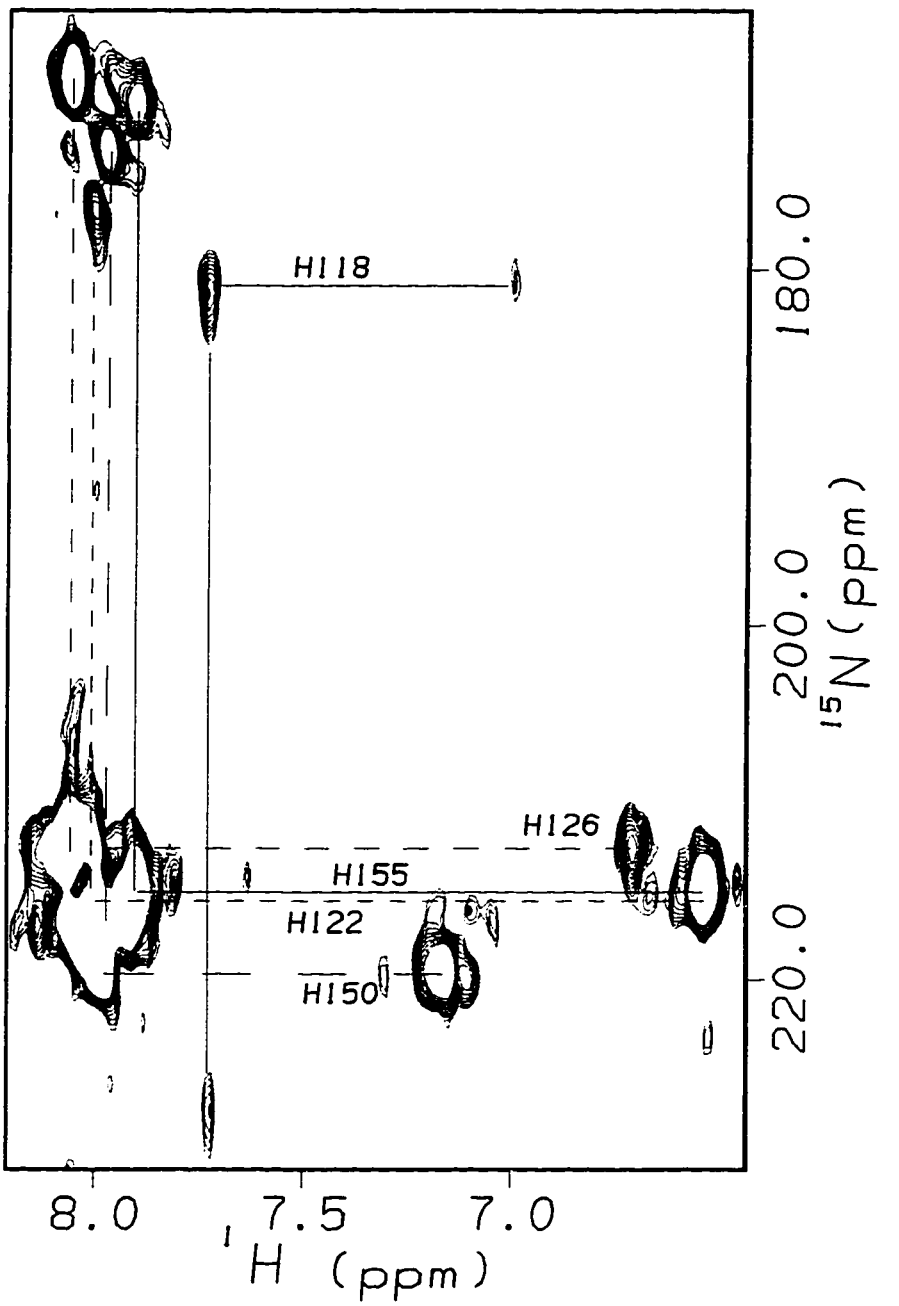
This notion is supported by studies on an H118Y mutant ((12), G. Párraga, report to committee, 1990). NOESY spectra of this mutant indicated interaction between Y118 and the basic R115. Furthermore, the pK_a of the Y118 sidechain hydroxyl was reported to be lower than normal for tyrosine, in analogy to the low H118 pK_a.

A final possible explanation for the low pK_α of H118 is that the unprotonated N^δ acts as an H-bond acceptor. This nitrogen is shifted about 8 ppm upfield of the expected chemical shift of 249 ppm; an upfield perturbation of about 10 ppm is expected for an imidazole nitrogen accepting a proton in an H-bond (86). No such H-bond involving the sidechain of H118 was identified in the single finger ADR1b structure, however. Because the proton chemical shifts of the ADR1z' finger match those of ADR1b exactly, it is likely that no H-bond exists for this sidechain in ADR1z' either.

The long-range HMQC-J spectrum of DNA-bound ADR1z' is shown in Figure 65. The zinc ligands show tautomer patterns very

Figure 65. Long-Range HMQC-J of ADR1z'-DNA.

The HMQC-J spectrum of ADR1z'-DNA at pH 7.7 is shown. As expected, the tautomeric form of the histidine zinc ligands does not change upon DNA binding; the chemical shifts of the imidazole ring protons and nitrogens also remain fairly constant. The tautomer pattern for H118 indicates that this sidechain still bears a proton on the N^{ε} position, consistent with the use of this position to donate a proton in an H-bond with DNA. Also consistent with H118 contact with DNA are the large chemical shift perturbations of the imidazole nitrogens.


H118
H126
H155
H122
H150
180.0
200.0
220.0
^{15}N (ppm)
8.0
7.5
7.0
^{1}H (ppm)

similar to those of the free protein. The structure of these residues is not expected to change much, as they are intimately involved in maintaining the tetrahedral coordination of zinc in each finger. However, overlap of the free and bound spectra shows that one of the zinc ligands does undergo larger chemical shift perturbations upon DNA binding than the others (Figure 66). This residue is H122, which is also severely perturbed at its beta protons. H122 falls at a position analogous to histidines that contact the DNA backbone in the Zif268 cocrystal structure (20).

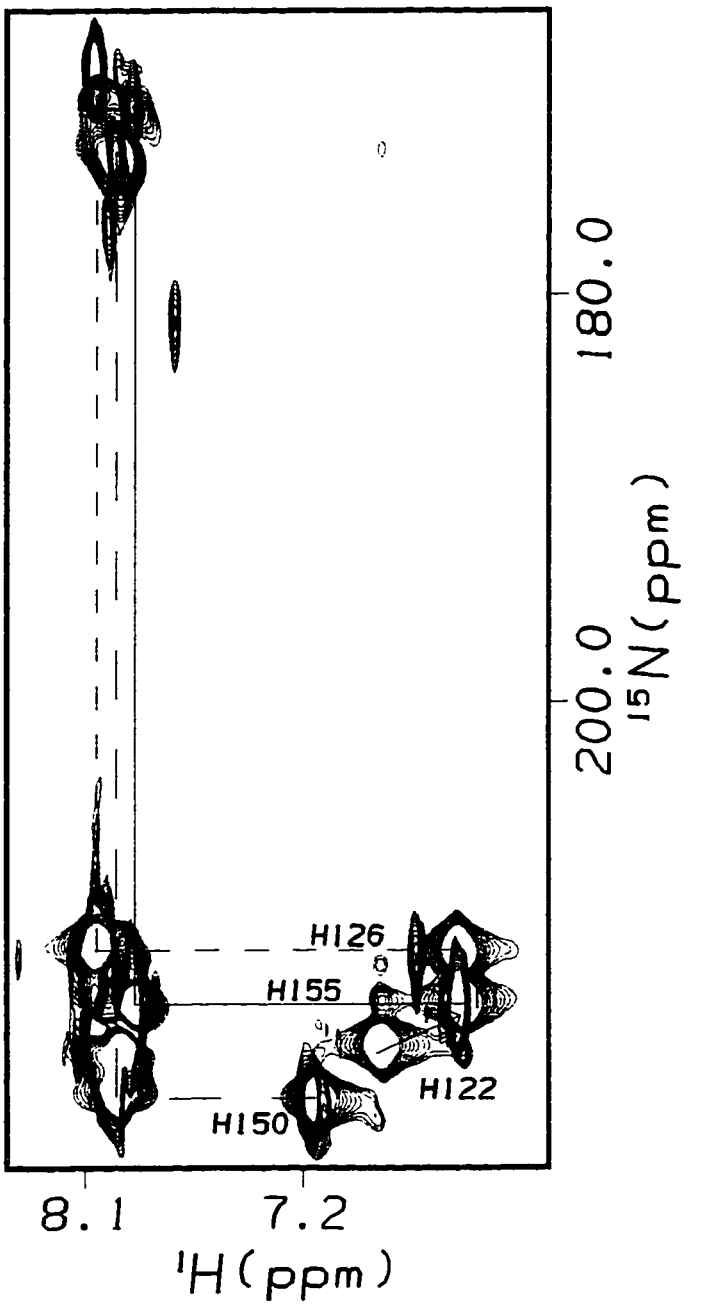
The other zinc ligand in finger 1, H126, also appears to undergo sidechain imidazole chemical shift perturbations (Figure 66), consistent with previous conclusions that the entire helix of finger 1 is perturbed by interaction with DNA. The perturbations to H126 appear to be smaller than those experienced by H122, however; H126 is not predicted to contact DNA directly.

Based on NMR and mutagenesis data discussed previously, neither H150 nor H155 from the suboptimal finger 2 is expected to approach DNA. Consistent with this prediction, the imidazole chemical shifts of these residues appear to be unperturbed upon DNA binding (Figure 66).

While the zinc ligands are affected little by DNA binding, the tautomer pattern for H118 exhibits large differences (Figures 64

Figure 66. Comparison of Free and Bound ADR1z' Tautomer Forms.

Upon ADR1z' DNA binding, one of the zinc ligand histidines undergoes a much larger chemical shift perturbation than the others, as highlighted in the spectrum. This histidine is H122; in Zif268 the analogous residue of each finger contacts the phosphodiester backbone of DNA. No perturbation is observed for His 150 or H155 of finger 2.


H126
H155
H122
H150
180.0
200.0
^{15}N(ppm)
8.1
7.2
^{1}H(ppm)

and 65). The single proton is still carried by N^{ε}, but the sidechain chemical shifts are severely perturbed. The C2 and C4 protons are shifted downfield by 0.10 ppm and upfield by 0.23 ppm respectively. The N^{δ} is shifted upfield by 13 ppm. In contrast, the N^{ε} is shifted downfield by 11 ppm.

Mutagenesis data predict that H118 makes a specific contact to DNA. The severe chemical shift perturbations to the sidechain protons (Figure 56) and imidazole ring nitrogens are consistent with this role and could be due to ring current effects of the DNA bases. In addition, it has previously been shown that nitrogens which act as proton donors in hydrogen bonds experience downfield perturbations of approximately 10 ppm (86). The 11 ppm downfield perturbation of the H118 N^{ε} is thus consistent with a role as a hydrogen bond donor.

UAS1 has been mutagenized to determine base preferences of ADR1 binding (56, 59). These studies showed that both adenine and guanine are acceptable at the position likely contacted by H118. Adenine has both hydrogen bond donors and acceptors in the major groove, but guanine has only acceptors. It is thus likely that H118 acts as a hydrogen bond donor in its specific contact to DNA; this prediction agrees well with the HMQC-J data showing that H118 bears a proton on a downfield shifted N^{ε}.

CHAPTER 5. PROBING FOR THE N-TERMINUS

I. Cobalt Studies

A. STRATEGY

One of the goals of this study is to determine the role of the N-terminal region of the minimal ADR1z' DNA binding domain. It is already known that the N-terminus is essential for DNA binding and that a mutation in this region can suppress defects in binding caused by changes in the zinc fingers. The present NMR study has shown that it is unlikely that the N-terminus interacts with the zinc fingers to increase their DNA binding activity. Therefore the probable role of the N-terminus is to bind DNA directly. The region of the N-terminus most likely to interact with DNA lies between P97 and P87, near the site of the R91K suppressor mutation. The next goal is to determine the orientation of the N-terminus with respect to the UAS1 DNA site.

A simple method for determining proximity by NMR relies on the incorporation of so-called spin labels at fixed sites in molecules. Spin labeled molecules generally possess an unpaired electron, either as a free radical or in a paramagnetic metal such as cobalt. Nuclei near the spin label exhibit broadened resonances

due to increased relaxation rates; nuclei near paramagnetic centers also exhibit chemical shift perturbations. These effects fall off as a function of distance to the spin label, with a $1/r^3$ dependence (87). Thus chemical shift perturbation and/or line broadening in the presence of a spin label is an indication of proximity to the spin label site.

Paramagnetic cobalt(II) can easily be incorporated into the metal binding sites of zinc fingers, including those of ADR1 (85). In the ADR1z'-DNA complex, these fixed cobalt centers can then be used as a probe for proximity of various nuclei, including those in the N-terminus.

B. COBALT PARAMAGNETIC TENSOR

The effect of cobalt on the chemical shift of proximal nuclei is a function of distance from the ion, geometry of ion coordination, and orientation of the nuclei with respect to the metal ligands. The equation that describes the shift perturbation is of the form

$$(\Delta H/H)^{dip} = -\frac{1}{3N}\{\chi_{zz}-\frac{1}{2}(\chi_{xx}+\chi_{yy})\}\{\frac{3\cos^2\theta-1}{r^3}\}-\frac{1}{2N}\{\chi_{xx}-\chi_{yy}\}\{\frac{\sin^2\theta\cos 2\phi}{r^3}\}$$

where χ_{xx}, χ_{yy}, and χ_{zz} are the principal magnetic susceptibilities, r is the distance of a nucleus to the metal ion, and θ and ϕ are polar coordinates that describe the orientation of a nucleus with respect to the principal axes of the susceptibility tensor (87). The position of these principal axes is determined by the geometry of coordination. The function above can be used to generate isoshift surfaces, of which every point would experience an identical chemical shift perturbation. These surfaces are multi-lobed, as shown in Figure 67. Nuclei falling within either of the large lobes experience a downfield chemical shift perturbation, while nuclei falling within the central doughnut-shaped lobe experience an upfield perturbation.

The susceptibility tensor for cobalt bound tetrahedrally into a Cys_2-His_2 zinc finger peptide has been determined empirically (88). Because coordination ligands and geometry are virtually identical from finger to finger, this tensor should be generally applicable to most Cys_2-His_2 zinc fingers. Figure 67 is in fact a depiction of the 0.2 ppm isoshift surface for this tensor, and Figure 68 shows the orientation of the tensor with respect to the zinc finger structure of the peptide used in the tensor determination. The orientation of the finger in the tensor is such

Figure 67. Cobalt Paramagnetic Shift Tensor.

The 0.2 ppm isoshift surface for cobalt-bound Cys_2-His_2 zinc fingers has three lobes. Nuclei up to 23 Å from the cobalt experience a 0.2 ppm paramagnetic shift. Those falling within the large axial lobes experience a downfield shift perturbation; those falling within the central lobe are shifted upfield.

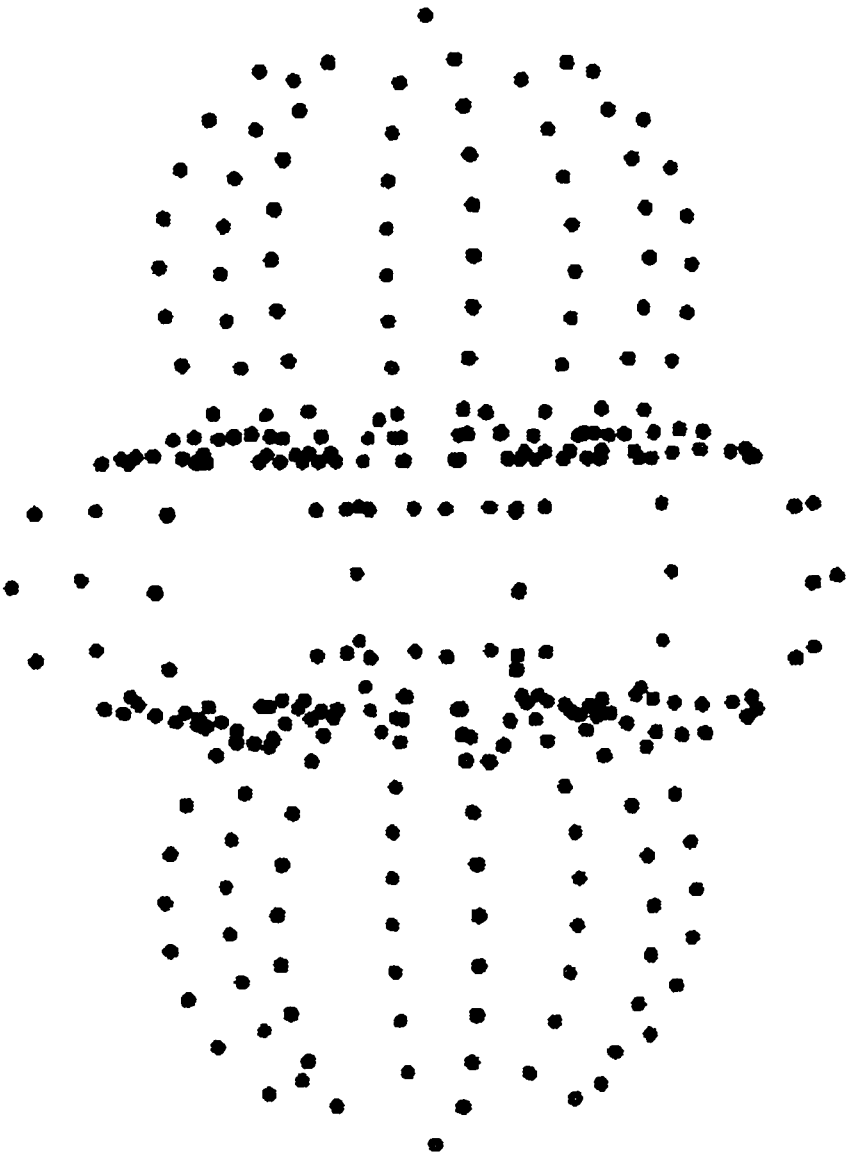
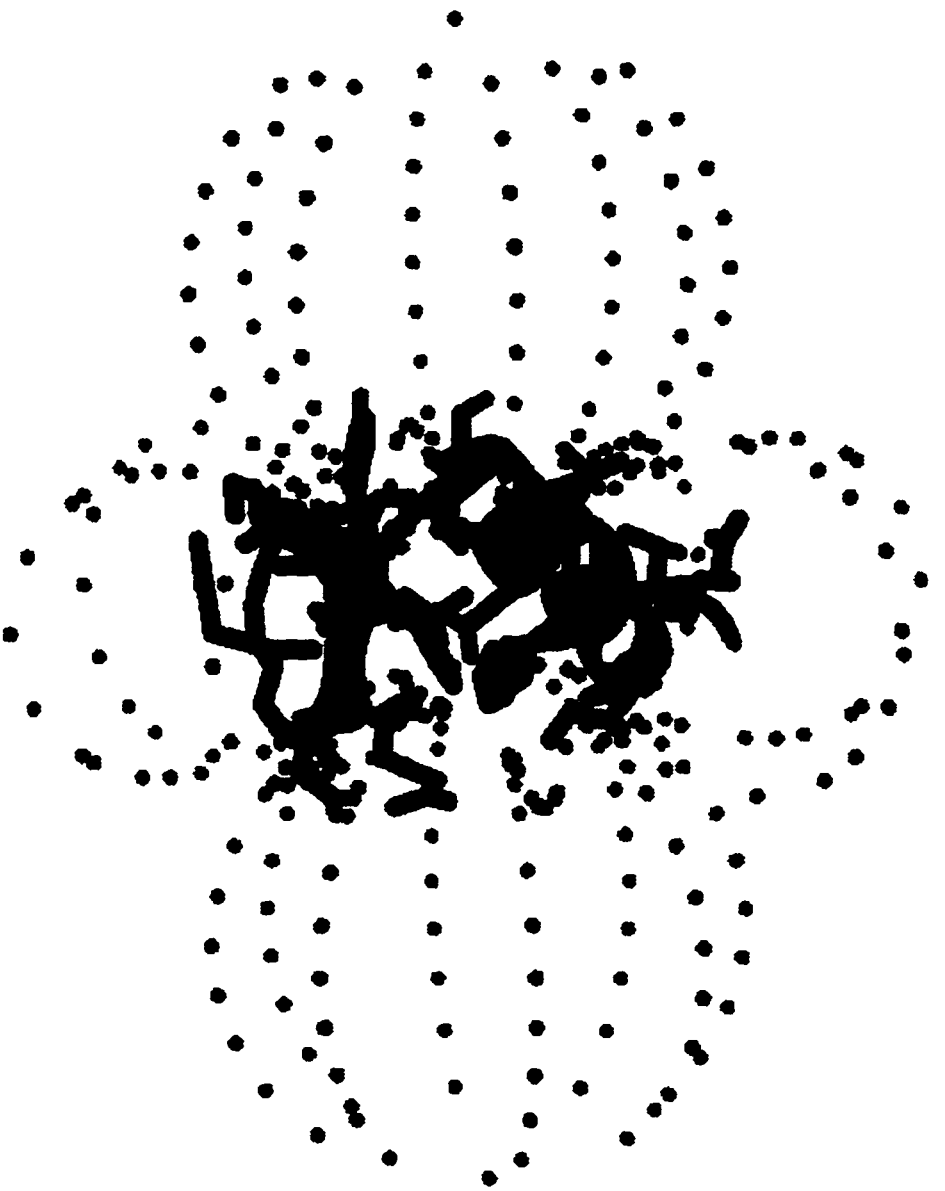


Figure 68. Zinc Finger Oriented in Paramagnetic Tensor.

The isoshift surface of Figure 67 is now oriented properly with respect to the ligand geometry of a Cys_2-His_2 zinc finger (blue). Because most of the finger falls within the central lobe, its resonances should be shifted primarily upfield.



that most finger resonances will experience an upfield chemical shift perturbation due to Co^{++} binding.

C. PARAMAGNETIC SHIFT OF COBALT-BOUND ADR1z'

i. Free ADR1z'

Figure 69 shows an overlap of ^{15}N-HSQC spectra of Zn^{++}-ADR1z' and Co^{++}-ADR1z'. It is readily apparent that most resonances are shifted by cobalt binding, and that most of these shifts are upfield, in agreement with the prediction for the zinc fingers. Closer inspection reveals a number of resonances that are neither shifted nor broadened (Figure 70). These resonances are from the random coil residues N-terminal to L101. Because these residues are mobile, they do not spend any significant amount of time in proximity to the bound cobalt and thus are not perturbed in the free protein.

ii. DNA-Bound ADR1z'

Because the zinc fingers of ADR1z' do not change structure upon DNA binding, the cobalt shift tensor is expected to remain much the same as for the free protein. Inspection of the HSQC

Figure 69. HSQC of Cobalt-Bound ADR1z'.

The ^{15}N-HSQC of Zn^{++}-ADR1z' is plotted in blue, that of Co^{++}-ADR1z' in red. As expected, most resonances of ADR1z' are shifted upfield upon cobalt binding.

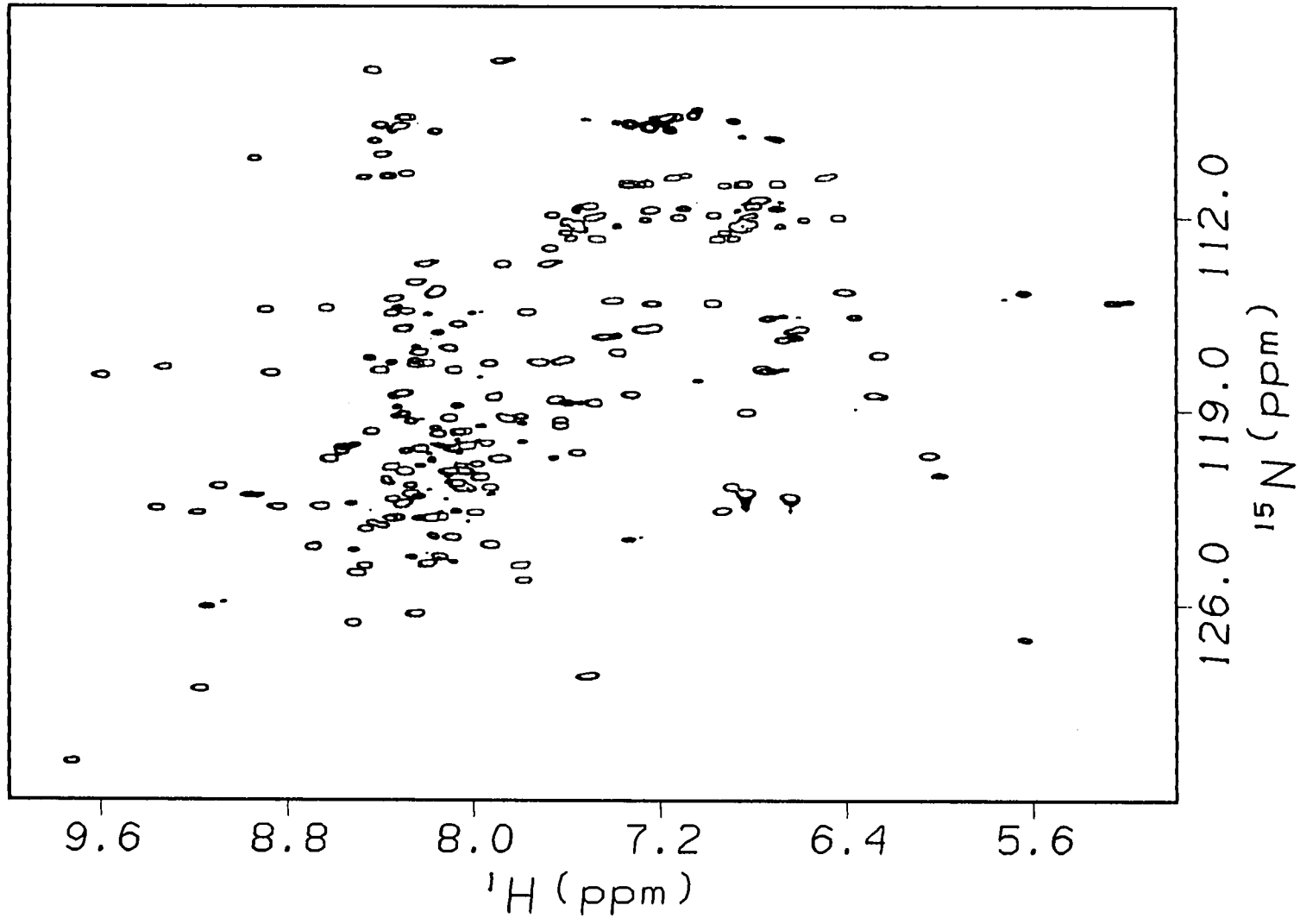
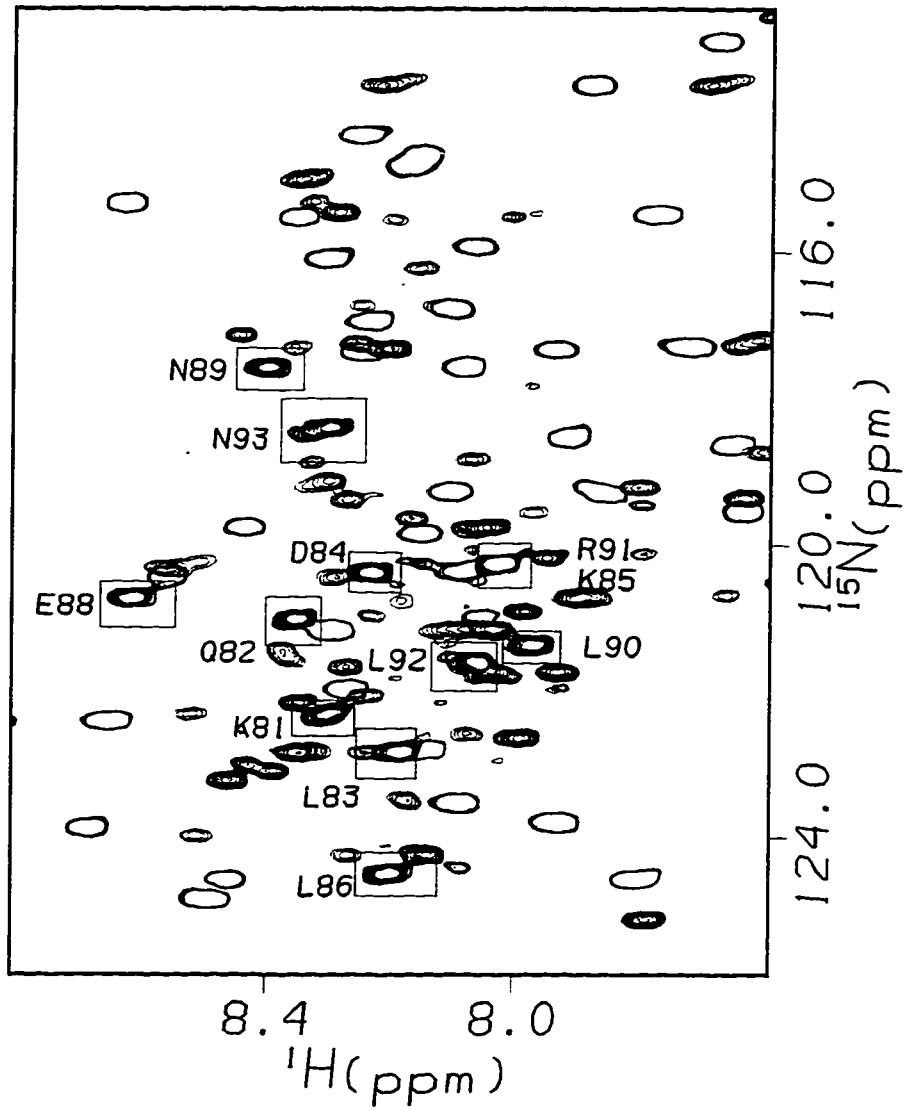

9.6
8.8
8.0
7.2
6.4
5.6
^1H (ppm)
126.0
119.0
112.0
^{15}N (ppm)

Figure 70. HSQC of Cobalt-Bound ADR1z': Close-up on the N-Terminus.

The random coil region of Figure 69 is plotted in more detail. Every resonance from the N-terminus of ADR1z' remains unperturbed in terms of chemical shift and linewidth. This is likely due to the random coil, mobile nature of this region, which does not permit any nucleus to remain long in proximity to the finger-bound cobalt.


N89
N93
D84
R91
K85
E88
Q82
L92
L90
K81
L83
L86
116.0
120.0
124.0
15N(ppm)
8.4
8.0
1H(ppm)

spectrum of Co^{++}-ADR1z-DNA shows that once again, a majority of resonances are shifted upfield, as expected for the zinc fingers (Figure 71).

It has been shown that the N-terminus becomes less mobile upon DNA binding. If this region of the protein lies in fixed proximity to the zinc fingers in the ADR1z'-DNA complex, then its resonances would be expected to be perturbed. If the N-terminus extends away from the fingers, however, its resonances should remain unperturbed.

Inspection of the Co^{++}-ADR1z'-DNA spectrum indicates that for many N-terminal resonances, the former case applies (Figure 72). In particular, the amide resonances of G94, N93, E88, L86, K85, and L83 are clearly perturbed in both the proton and nitrogen dimensions. No peak in the cobalt spectrum lies within 0.2 ppm of these peaks in the zinc spectrum. Other N-terminal resonances may or may not be shifted by this much. Peaks in the cobalt spectrum are indeed found within 0.2 ppm of the zinc peaks for these resonances; these cobalt peaks may arise from the N-terminal resonances in question or may have been shifted from farther away in the spectrum.

The N-terminus is thus unambiguously perturbed in regions as far away as twenty residues from the consensus zinc finger

Figure 71. HSQC of Cobalt-Bound ADR1z'-DNA.

Again, as expected from Figure 68, most resonances experience an upfield chemical shift perturbation in the presence of cobalt (red) as opposed to zinc (green). In the case of ADR1z'-DNA, some of the shifted residues are from the N-terminus, as highlighted in the spectrum.

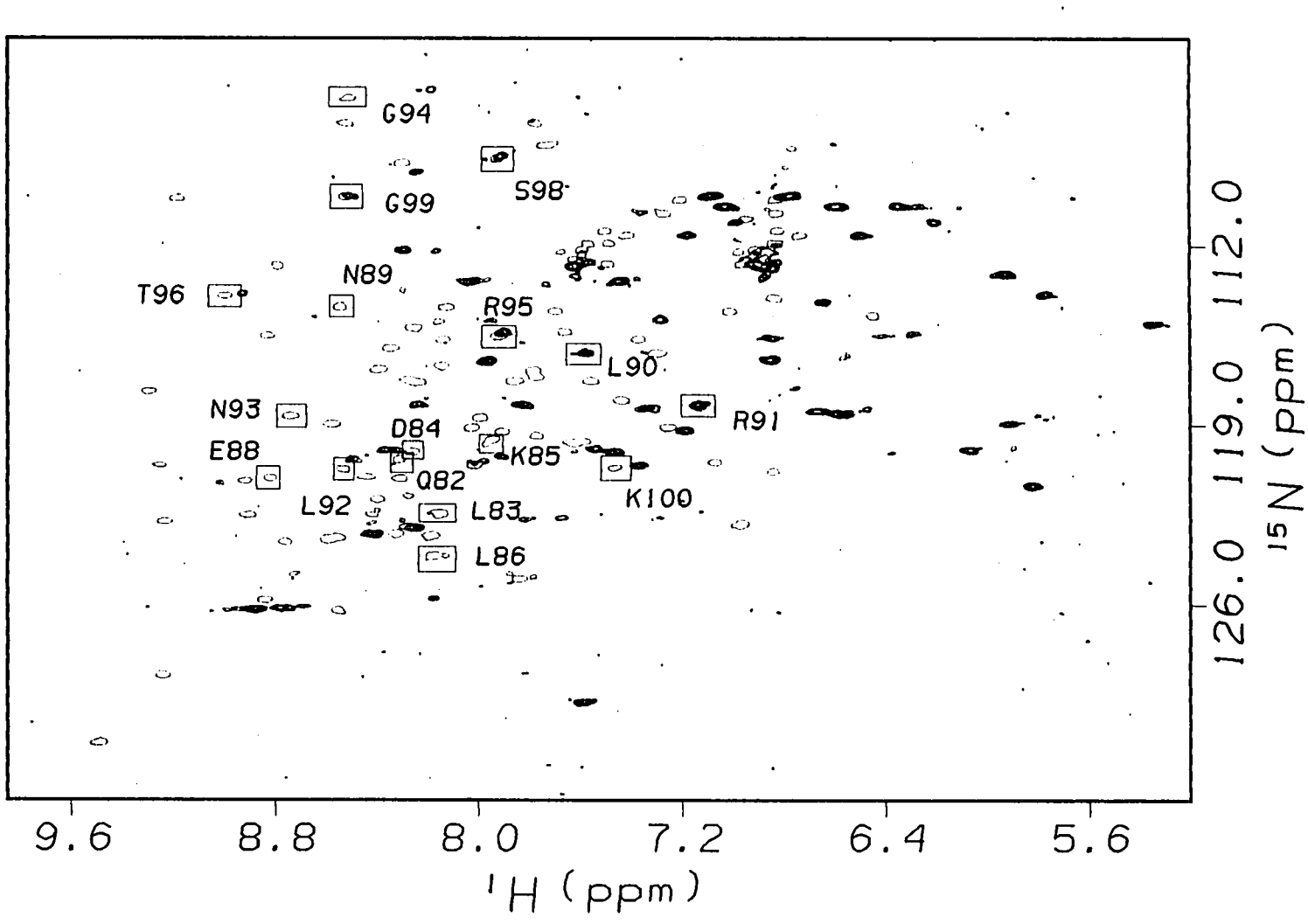
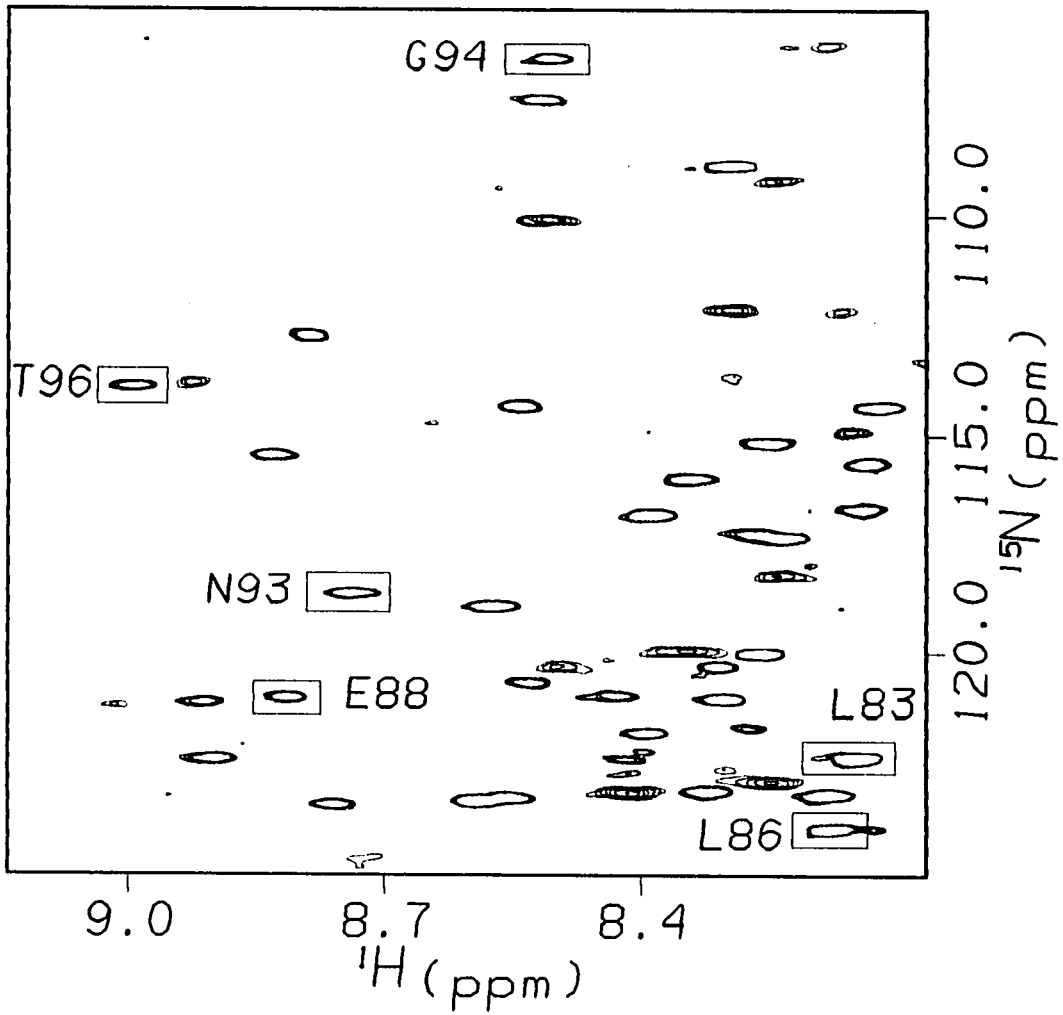

G94
S98
G99
N89
T96
R95
L90
N93
R91
D84
E88
K85
Q82
K100
L92
L83
L86
9.6
8.8
8.0
7.2
6.4
5.6
^{1}H (ppm)
126.0
119.0
112.0
^{15}N (ppm)

Figure 72. HSQC of Cobalt-Bound ADR1z'-DNA: Close-up on the N-Terminus.

The spectra of Figure 71 are plotted to display detail in a region containing several peaks from the N-terminus. It is clear from this figure that a number of residues from the N-terminus must fall within the 0.2 ppm isoshift surface shown in Figure 67. Residues as far N-terminal as L83 experience chemical shift perturbations due to cobalt of greater than 0.2 ppm.


G94
T96
N93
E88
L83
L86
110.0
115.0
120.0
^{15}N (ppm)
9.0
8.7
8.4
^{1}H (ppm)

sequence. The N-terminus must assume a position in the ADR1z'-DNA structure such that these residues fall within the 0.2 ppm isoshift surfaces of the cobalt fingers. This situation is modeled in Figure 73. Here the fingers of ADR1z' are superimposed on fingers 2 and 3 of the Zif268 cocrystal structure. The 0.2 ppm isoshift surfaces for each finger are drawn in the orientation empirically determined for a single finger peptide (88).

It is clear from this figure that were the N-terminus to extend along the DNA away from the zinc fingers, there would be no possibility of entering the 0.2 ppm isoshift surfaces. This holds for binding along either strand or in either of the grooves. The most likely course of the N-terminus along the DNA lies within the UAS1 site itself, along the phosphodiester backbone of the strand contacted by the zinc fingers. This strand passes directly through the isoshift surface of finger 1 in the model (Figure 73).

iii. Evidence for Backbone Contacts by the N-Terminus

The notion that the N-terminus of ADR1z' binds to the DNA phosphodiester backbone is supported by several previous findings. First, the base preferences of ADR1 do not appear to extend beyond the region of DNA contacted specifically by the zinc fingers themselves. This suggests that the N-terminus may make non-

Figure 73. Model of ADR1 Zinc Fingers Bound to DNA: Position of the Cobalt Tensors.

The zinc fingers of ADR1 were modeled on fingers 2 and 3 of the Zif268 cocrystal structure. The 0.2 ppm isoshift surfaces for cobalt-bound zinc fingers are oriented based on the ligand geometry of the fingers. The portion of the Cys-Cys loop of finger 1 from which the N-terminus of ADR1z' arises is highlighted in green. It is clear that were the N-terminus to extend away from the zinc fingers along the DNA, at no point would it enter either 0.2 ppm isoshift lobe. Because residues from the N-terminus are in fact shifted by at least 0.2 ppm by cobalt, the N-terminus must contact the fingers themselves, or traverse the region of DNA also contacted by the fingers. Previous experiments have ruled out the former possibility, leaving it likely that the N-terminus runs along the DNA anti-parallel to the zinc fingers, through the isoshift lobes.

specific contacts only. DNA binding by ADR1 has also been shown to be fairly sensitive to salt concentration, suggesting that a fair portion of the binding energy comes from charged interactions (58).

This result is supported by NMR studies in which ADR1z'-DNA complex formation was shown to be reversible by addition of low levels of salt (Figure 74). Only 375 mM NaCl is required to disrupt the complex and yield a spectrum identical to that of free protein. This salt concentration is relatively low compared to the 450 - 750 mM used in Zif268 cocrystal growth, suggesting that charged interactions play a more prominent role in the ADR1z' complex, perhaps due to the contributions from the N-terminus.

The notion that the N-terminus runs back along the UAS1 site antiparallel to the zinc fingers, towards the center of the palindrome is also consistent with earlier findings. The pattern of ethylation interference by ADR1 constructs determined previously is shown in Figure 75 (59). A construct truncated at residue 75 of the N-terminus apparently makes no contacts to the central positions of the UAS1 palindrome. However, a construct that extends further to residue 17 does make contacts to the DNA backbone at the center of the palindrome (59). This result has two possible interpretations. Less likely is that the presence of the additional N-terminal sequence causes the rest of the protein to

Figure 74. The Formation of ADR1z'-DNA is Reversible by Salt.

Free ADR1z' is shown in 400 mM NaCl in the blue HSQC spectrum. Addition of 375 mM NaCl to an ADR1z'-DNA complex results in apparently complete reversal of complex formation to yield a spectrum identical to free ADR1z' in salt (red).

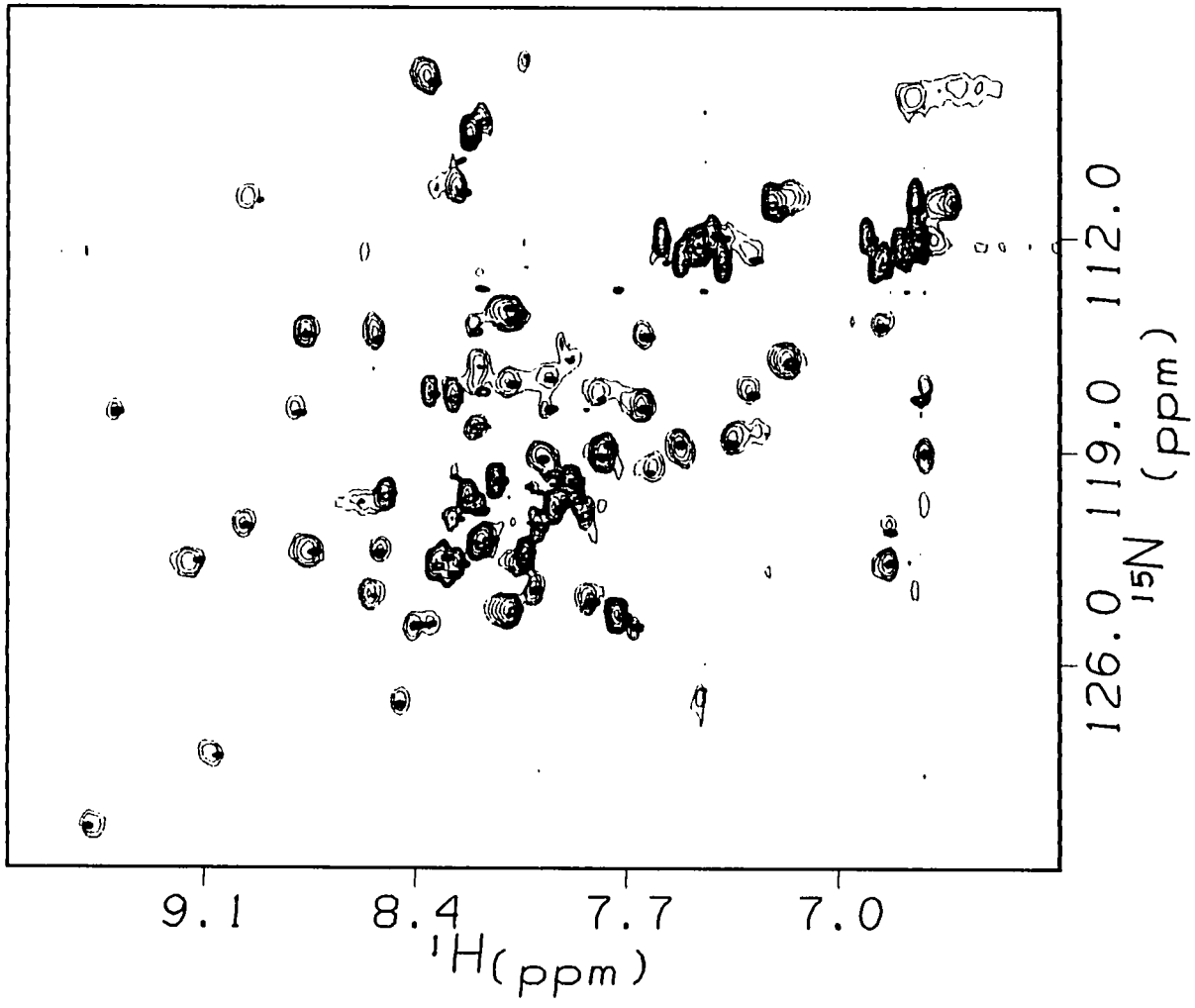
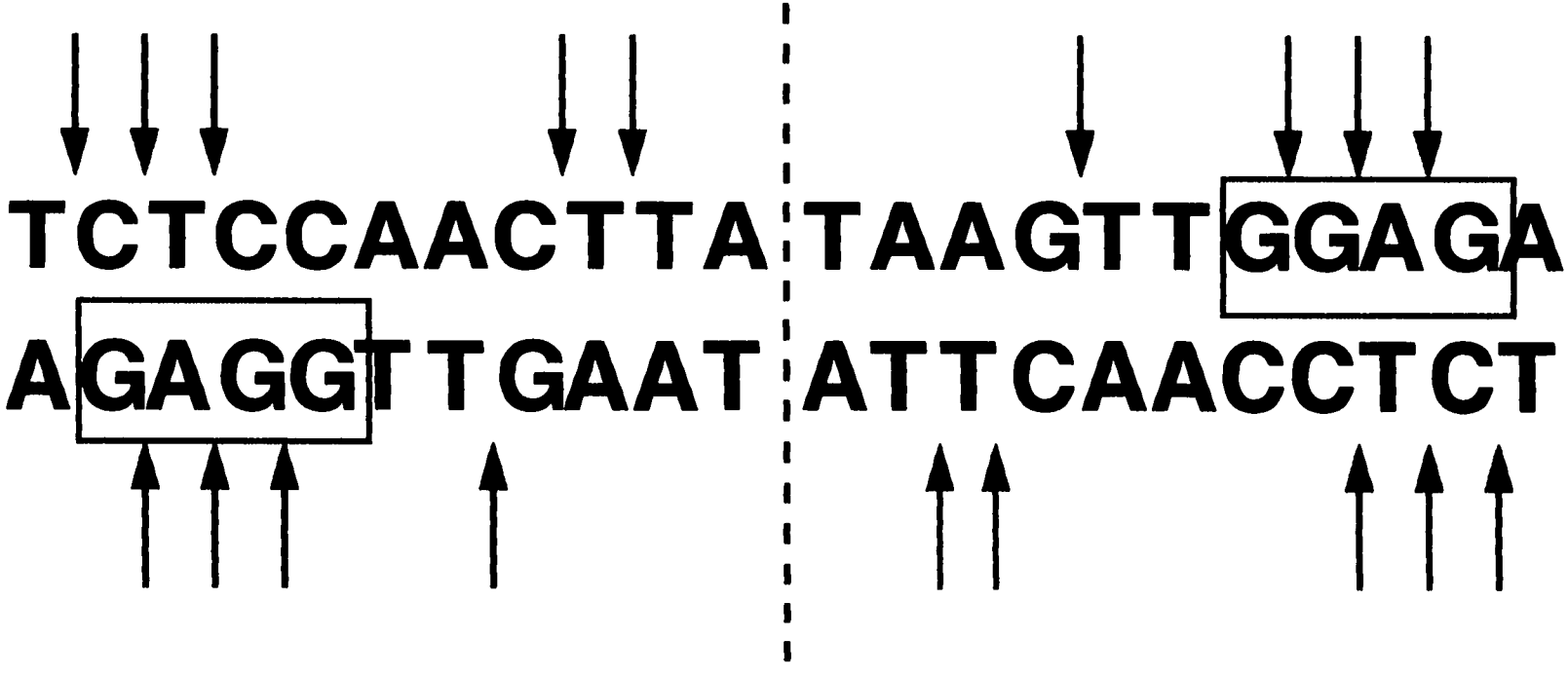

9.1
8.4
7.7
7.0
^1H(ppm)
126.0
119.0
112.0
^{15}N (ppm)

Figure 75. Ethylation Interference Data for UAS1.

The pattern of ethylation interference for ADR1 binding to UAS1 has been determined previously (59). There is a strong pattern in the region of the DNA expected to be contacted specifically by finger 1 and the extreme N-terminus of finger 2. The pattern of interference at the center of the palindrome, however, requires amino acid residues N-terminal to the ADR1z' construct. Some portion of the N-terminus thus likely traverses the DNA between the edge and the center of the palindrome, in agreement with cobalt data. Furthermore, the extremely minimal contact for finger 2 predicted by NMR results would seem to preclude finger 2 being responsible for the 3'-GTT*-3' contact. This contact may instead be made by the N-terminus as it traverses the DNA site bound by the zinc fingers.



→ ADR1 (75-229), ADR1 (17-229)

→ ADR1 (17-229)

make new contacts at the center of UAS1. More likely is that the N-terminus, which necessarily originates from finger 1 at the edge of UAS1, runs back toward the center of UAS1 and there contacts the DNA backbone.

Finally, ADR1 (75-229) appears to make a contact to the DNA backbone at the 5' edge of the 3'-GTT-5' subsite likely contacted by finger 2 (Figure 75) (59). However, given the wholly unperturbed nature of all of finger 2 except for the precise region expected to contact the 3' end of this subsite, it seems likely that finger 2 does not lie near the DNA at the 5' end of the subsite. This opens the possibility that the ethylation interference pattern in this part of the subsite actually arises not from finger 2, but rather from the N-terminus.

iv. Model of ADR1z'-DNA Interaction

Fingers 2 and 3 of the Zif268 cocrystal structure are shown in magenta in Figure 76 (20). It is apparent that both fingers lie close to DNA along their helices. In green is shown a proposed model for the orientation adopted by finger 2 of ADR1z'. All of the NMR evidence suggests that this finger approaches DNA with only the extreme N-terminus of the helix.

Figure 76. Orientation of Zif268 on DNA.

Fingers 2 and 3 of Zif268 bind DNA in a similar orientation, with the helix in the major groove of the DNA. A proposed alternative binding orientation for finger 2 of ADR1 is shown in green. Based on all NMR data, finger 2 makes very suboptimal contact to DNA, concentrated exclusively in the region of the single predicted specfic contact, R143.

Figure 77 is a model of ADR1z' interaction with DNA that incorporates all of the observations of this study. In this model, finger one is oriented like finger two of Zif268, with its entire helix in proximity to DNA. Finger 2 adopts an alternative orientation, with only the extreme N-terminus of the helix near DNA, in agreement with chemical shift perturbation data. The Cys-Cys loop of finger one is extended, and residues at the junction with the N-terminus, S103 and R102, cross over the opposite beta strand, in agreement with NOE data. Based on cobalt data, the N-terminus is modeled to extend along the UAS1 site, back toward the center of the palindrome, contacting the backbone of the strand bound by the zinc fingers. Here it would be in a position to make a backbone contact at 3'-GTT*-5', and it could deliver residues 17-75 to the center of the palindrome to make further backbone contacts in agreement with ethylation interference data. The bound N-terminus is not modeled as adopting any particular regular secondary structure, as observed from NOE data.

Figure 77. Model of ADR1z' Binding to DNA.

The fingers of ADR1z' are modeled on fingers 2 and 3 of the Zif268 cocrystal structure. Finger 1 is oriented like the Zif fingers, but finger 2 is depicted as in Figure 76, tilted away from the major groove. The extension to the Cys-Cys loop of finger 1 is shown, along with the likely cross-over of strands predicted by NOEs. Finally, the N-terminus is sketched as binding along the DNA backbone, on the strand for which the 3'-GTT*-5' ethylation interference occurs. The N-terminus is drawn as running antiparallel to the zinc fingers, back toward the center of the UAS1 palindrome.

II. Nitroxide Spin Labeling

A. STRATEGY

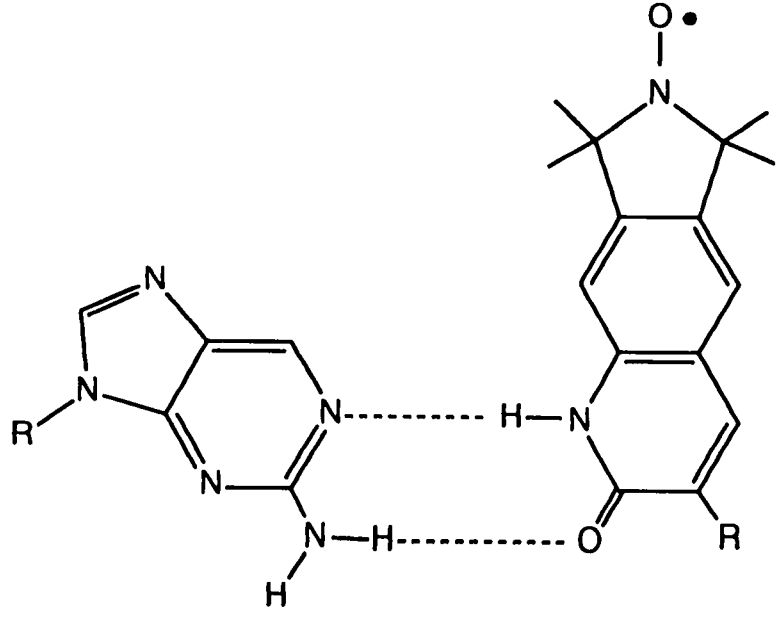
A second way to probe for the location of the N-terminus in an ADR1z'-DNA complex is to spin label the DNA. In this approach a free radical is introduced at a specific location in the DNA, and experiments sensitive to increased relaxation rates are used to test for effects on nearby protons. Spin-label dependent broadening of a resonance implies proximity to the free radical. Because nitroxide free radicals broaden lines but do not perturb chemical shift, affected resonances can be easily assigned to specific regions of the DNA and bound protein.

B. SPIN-LABELED DNA

This spin labeling approach was implemented for the ADR1z'-DNA complex in collaboration with Stephen Alley and Todd Miller in the laboratory of Paul Hopkins at the University of Washington. A nitroxide spin label has been designed in that lab which is essentially a modified thymidine nucleotide (Figure 78) (89). This nitroxide-bearing nucleotide can be used instead of thymidine at

Figure 78. Chemical Structure of Spin-Labeled Base.

The spin-labeled base incorporated into 14mer DNA is a modified thymidine that carries a nitroxide free radical. This base can pair with 2-aminopurine on the opposite strand.



2-aminopurine spin-labeled thymidine analog

any point in DNA oligo synthesis, and it will pair with 2-aminopurine in a complementary strand.

In this study, the spin-labeled base was introduced into the previously described 14mer DNA sequence used in complexes with ADR1z', at a position near the binding site for finger 2 (Figure 79). The hope was that if the N-terminus of ADR1z' turns back from the edge of UAS1 to approach the center of the palindrome, its end will be perturbed by proximity to the spin label. If the N-terminus remains unperturbed, an estimate of minimum distance from the spin label might be made. Nitroxide spin labels have been shown to affect nuclei up to 20 Å away (90).

The presence of the active free radical can be confirmed in 1D proton spectra of the 14mer DNA alone. Spectra of the imino region of 14mer with and without spin label are shown in Figure 80. Broadening of a subset of imino lines is readily apparent when the spin label is present. Those lines most broadened have been assigned by other methods to basepairs lying nearest the nitroxide (see Table IV). Those lines that appear to be unaffected are assigned to basepairs at the opposite end of the DNA from the spin label site. Varying levels of broadening are seen for intermediate basepairs; the slightest detectable broadening in this experiment is seen for an imino expected to be approximately 18 Å away from

Figure 79. Spin-Labeled 14mer DNA Sequence.

The location of the thymidine-nitroxide spin labeled base is shown as "N". The 2-aminopurine on the opposite strand is represented by "2".

14mer

5'- GGCATCTCCAACGC -3'
3'- CCGTAGAGGTTGCG -5'

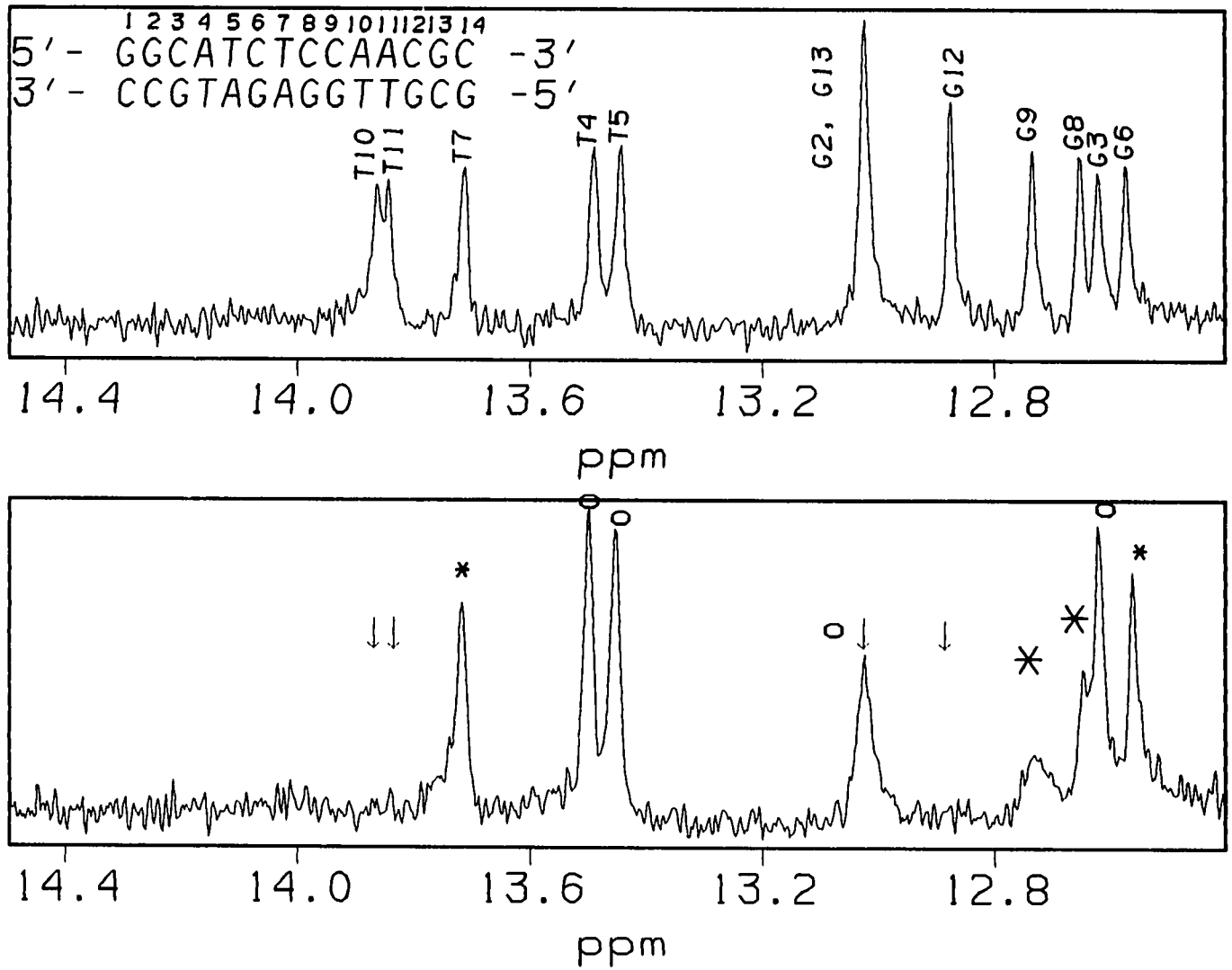
14mer with spin label

5'- GGCATCTCCAA*NGC -3'
3'- CCGTAGAGGTT2CG -5'

finger 1 finger 2

Figure 80. Effect of Nitroxide-Spin Label on DNA Iminos.

The imino region of a 1D proton spectrum of spin-labeled 14mer in water is shown versus the spectrum of unlabeled DNA. Broadening of lines by the spin label is detected at iminos up to 18 Å from the free radical.


1 2 3 4 5 6 7 8 9 10 11 12 13 14
5′- GGCATCTCCAACGC -3′
3′- CCGTAGAGGTTGCG -5′
T10
T11
T7
T4
T5
G2, G13
G12
G9
G8
G3
G6
14.4
14.0
13.6
13.2
12.8
ppm
14.4
14.0
13.6
13.2
12.8
ppm

the nitroxide. This experiment not only establishes the presence of intact free radical, it also confirms assignments for the iminos of the 14mer DNA.

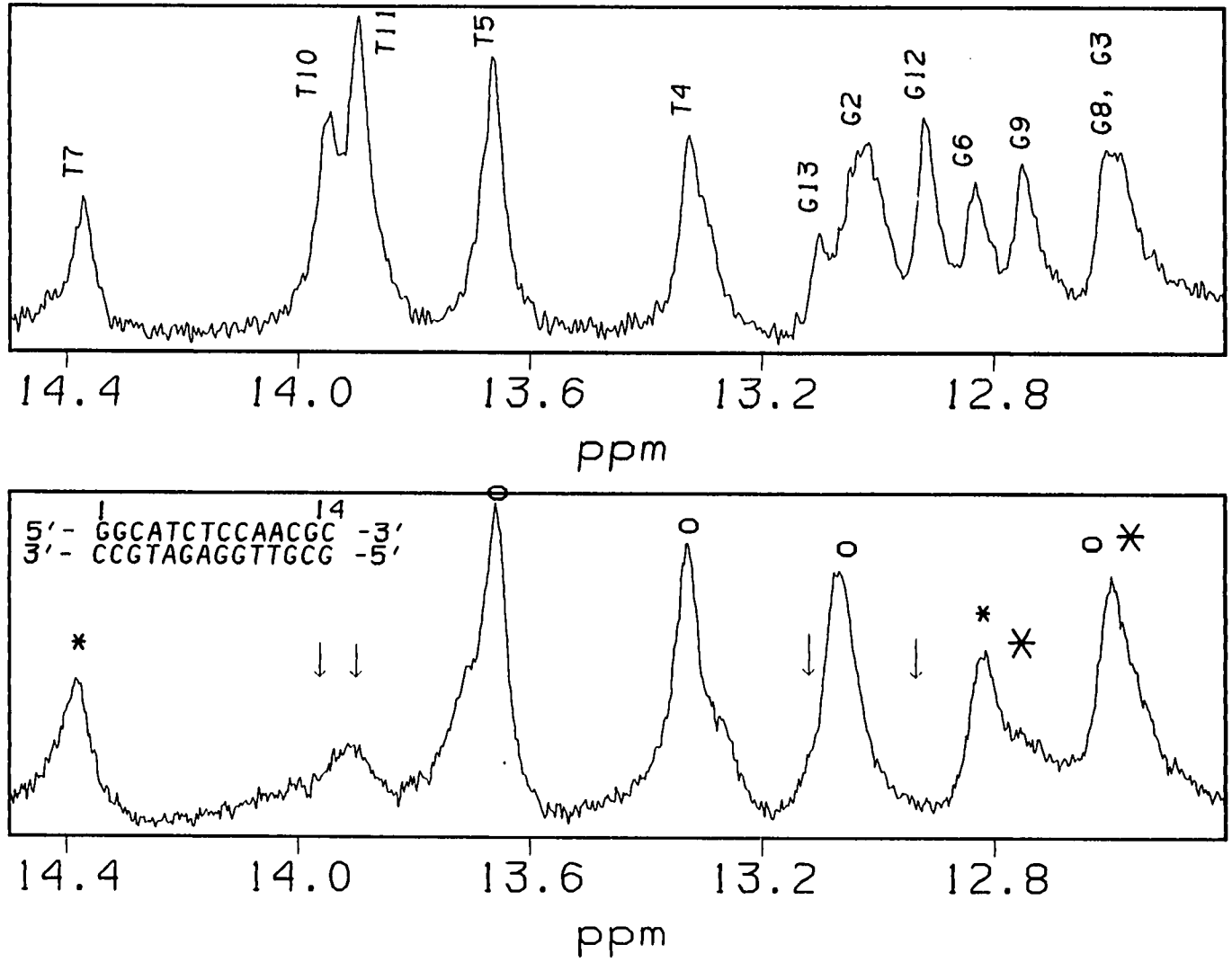
The same sort of line broadening effect can be seen for the 14mer iminos in the context of an ADR1z'-DNA complex (Figure 81). Again, those iminos close to the spin label are severely broadened, while those far away are not. The relative effects correspond nicely with the assignments made earlier by NOESY techniques (see Table IV and Figure 39). The broadening of imino lines in Figure 81 confirms that the free radical remains active in the context of the protein-DNA complex. Furthermore, the chemical shifts of the bound spin-labeled DNA lines are identical to those of the unlabeled DNA; HSQC spectra similarly show no protein chemical shift perturbations due to the presence of spin label in the DNA (data not shown). This indicates that the spin-labeled base, while bulky, does not interfere with formation of the ADR1z'-DNA complex.

C. EFFECT OF SPIN LABEL ON DNA-BOUND ADR1z'

Figures 80 and 81 demonstrated that imino lines from the DNA could be broadened by the nearby spin label. Similarly, nuclei from the protein that lie in proximity to the spin label should

Figure 81. Effect of the Spin Label on ADR1z'-Bound 14mer DNA.

A comparison like that of Figure 80 is made for ADR1z'-DNA, with similar results.


T7
T10
T11
T5
T4
G13
G2
G12
G6
G9
G8, G3
14.4
14.0
13.6
13.2
12.8
ppm
1
14
5'- GGCATCTCCAACGC -3'
3'- CCGTAGAGGTTGCG -5'
14.4
14.0
13.6
13.2
12.8
ppm

experience faster relaxation. This is being tested using a modification of the HSQC spectrum that measures T_2 relaxation rates for backbone amide groups. This experiment is exquisitely sensitive to changes in relaxation rates, and should reveal any alterations caused by the spin label. Reduction of the free radical should eliminate these effects, establishing that they are truly spin label-dependent.

Initial results of these experiments show that only residues from the extreme C-terminus of finger 2 and from the extreme N-terminus of ADR1z' experience any perturbation to their relaxation rates in the presence of the spin label. The control experiment in which the spin label is reduced to see if these effects are reversible upon elimination of the free radical is being analyzed. If the relaxation rates prove to be truly spin label-dependent, these results would support the view of the N-terminus developed in the cobalt experiments. Relaxation changes in the extreme N-terminus of ADR1z' would imply that this region lies near the spin label, close to the center of the UAS1 site. This is exactly what is predicted in the model of ADR1z'-DNA interaction presented in Figure 77.

CHAPTER 6. CONCLUSIONS AND FUTURE DIRECTIONS

I. ADR1z' Zinc Finger Structure

The system presented here represents the first protein-DNA complex to be studied in detail in the Klevit lab. Many of the techniques used, including the triple resonance experiments now possible on our recently upgraded instrument, were applied for the first time in the lab in this thesis work. These special efforts were required by the size of the complex; at 20 kD, ADR1z'-DNA is the largest system assigned in the lab to date. No NMR assignments for other zinc finger-DNA complexes exist in the literature, so this work offers the first opportunity to confirm and extend observations from x-ray crystallography.

The first question we asked was whether the detailed structures obtained for the ADR1 single zinc finger peptides, ADR1b and PAPA, are good models for the fingers in the high-affinity DNA binding construct. The answer appears to be a conditional yes. Based on chemical shift similarities, the zinc fingers of ADR1z' are identical in structure to the wild-type single finger peptides ADR1b and ADR1a; other information suggests that there are some differences from the known structure of the PAPA double-mutant peptide.

In addition to the correspondence of chemical shifts between ADR1z' and ADR1b, short-range and cross-strand NOEs and coupling constants support the view that the first finger of ADR1z' adopts a structure identical to that determined for single finger peptide ADR1b. The structure of finger 1 thus includes a Cys-Cys loop with a classical beta sheet between residues 104, 105, 112, and 113, a beta bulge, and a type I reverse turn. Residues following R115 participate in a helix that extends through the last histidine ligand. In the fingertip, residue A114 falls in the helical region of conformational space while R115 adopts an extended conformation, allowing for the 90° turn in the backbone between sheet and helix that was observed in ADR1b.

The structure of finger 2 appears to be similar to finger 1 in the fingertip region and the beginning of the helix. There is stronger evidence for alpha helical structure throughout the helix in this finger than in finger 1, consistent with the observation that the single finger 1 peptide, ADR1b, undergoes a transition to 3_{10} helix.

There is evidence from dramatically different carbon chemical shifts and coupling constants that the structure of the Cys-Cys loop of ADR1z' finger 2 differs from that of finger 1. A search for interactions across the Cys-Cys loop of finger 1 reveals a number of unambiguous cross-strand NOEs that support the

presence of a classical beta sheet involving residues 104, 105, 112, and 113 (Figure 51); no equivalent NOEs are observed for finger 2, suggesting that this finger lacks regular beta sheet structure. The only unambiguous NOEs across this loop, occurring between Y132 and R139, are completely inconsistent with a sheet structure like that of finger 1.

The difference in Cys-Cys loop structure in fingers 1 and 2 is likely due to the presence of P133 in finger 2. The PAPA double mutant peptide lacks both P131 and P133; this peptide has a structure quite similar to that of finger 1 peptide ADR1b, including the short stretch of classical beta sheet. It is unlikely that P131 disrupts the beta sheet in the wild-type peptide, since other zinc fingers with proline at an equivalent position, including fingers from Zif268, have the same fold as ADR1b and PAPA. In these structures, the residue at a position equivalent to P133 adopts a conformation inaccessible to proline; it was predicted that accommodation of proline would require backbone rearrangement (13). The ADR1z' study clearly indicates that finger 2 is indeed different from finger 1 in the Cys-Cys loop, lacking the classical beta sheet structure. This difference is likely present in the wild-type single finger 2 peptide, ADR1a, as well, which resembles ADR1z' in its chemical shift profile. No structure has been calculated for the ADR1a peptide, however, so a structure

calculation for ADR1z' will be the first opportunity to understand the effect of P133 on zinc finger structure.

II. Zinc Finger Structural Changes upon DNA Binding

This study has shown that there is little change to the zinc fingers upon DNA binding, with the most notable change involving an extension of the Cys-Cys loop of finger 1. Additional NOEs in the bound protein indicate that residues S103 and R102 lie adjacent to and possibly cross over residue A114 in the fingertip. In contrast, in the free protein S103 likely folds back to lie adjacent to A112 and V105 as in the single finger ADR1b structure. The strand crossover in DNA-bound finger 1 may be necessary for positioning of the neighboring N-terminal residues which play a role in DNA binding. In fact, mutation of A114, the apparent crossover point, to valine was one of the first mutations identified as deleterious to ADR1 activity (52).

No inter-finger interactions were detected for the free ADR1z' polypeptide, and the linker region is random coil in the absence of DNA. Thus the zinc fingers themselves do not adopt a fixed orientation with respect to one another prior to DNA binding, suggesting that this is not a necessary step in DNA recognition. The fingers may interact with one another in the DNA-bound form;

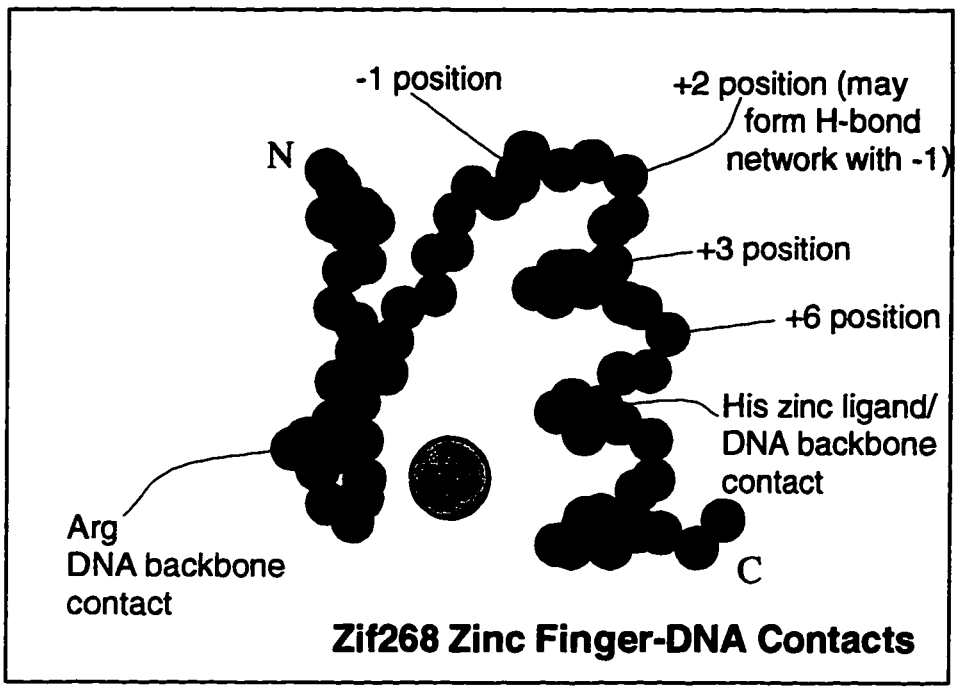
one likely interaction is that seen in the Zif268 cocrystal structure and the MBP-1 double- finger NMR structure. In each case, the residue at a position equivalent to T142 was found to contact the preceding finger. Mutation of T142 to isoleucine does in fact disrupt ADR1 activity, consistent with a role at the interface between fingers. The chemical shift profile of T142 is unchanged upon DNA binding, however, suggesting that no contact is made using this residue. A lack of the typical interaction could result from an unusual finger-finger orientation in the case of the ADR1 zinc fingers; other evidence for an atypical orientation for finger 2 is discussed below.

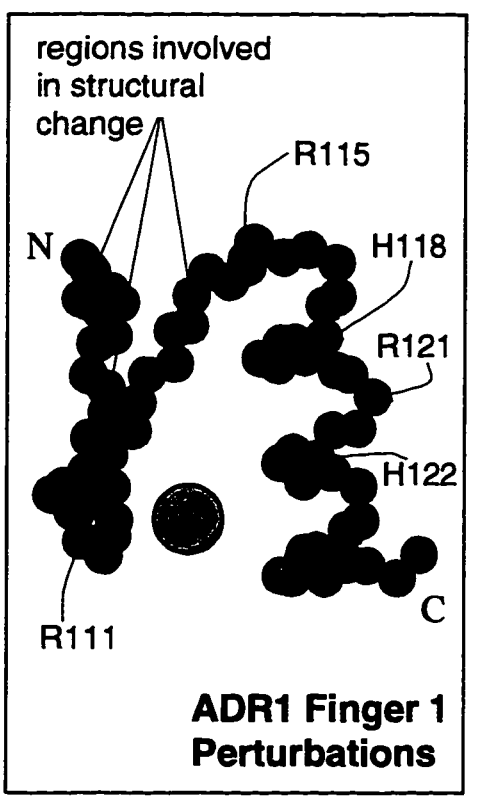
III. Protein-DNA Interface

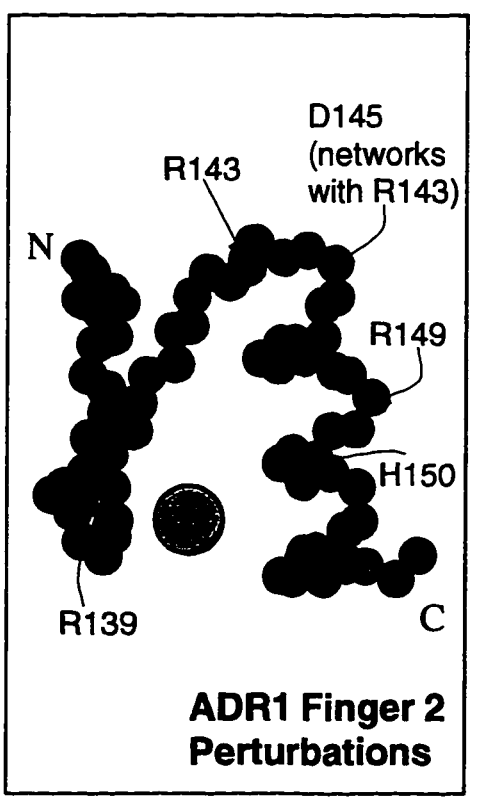
The pattern of zinc finger-DNA contacts suggested by mutagenesis is supported in detail by chemical shift perturbations that map the protein-DNA interface (Figure 82). In finger 1, in addition to residues involved in a structural change, residues R111, R115, H118, R121, and H122 are severely perturbed. R115, H118, and R121 have been shown by change-of-specificity mutagenesis to make base-specific contacts. The only direct DNA backbone contacts made by the Zif268 zinc fingers involve residues analogous to R111 and H122.

Figure 82. Chemical Shift Perturbation Data Compared to Mutagenesis Data

Positions of the Zif268 zinc fingers that contact DNA are mapped in red on the backbone structure of Zif268 finger 2. Positions -1, +3, and +6 of the helix make specific contacts to the DNA bases. The lower panels show the same structure, with chemical shift perturbations observed for ADR1 fingers 1 and 2 mapped at the corresponding Zif268 residues. Positions colored in red experience severe chemical shift perturbation, those in green, moderate perturbations, and those in blue, no significant perturbation (as defined for Figure 56).


-1 position
+2 position (may form H-bond network with -1)
N
+3 position
+6 position
His zinc ligand/ DNA backbone contact
Arg DNA backbone contact
C
Zif268 Zinc Finger-DNA Contacts


regions involved in structural change
R115
N
H118
R121
H122
C
R111
ADR1 Finger 1 Perturbations


D145 (networks with R143)
R143
N
R149
H150
R139
C
ADR1 Finger 2 Perturbations

In contrast, in finger 2 only R143 is severely perturbed. This residue, which falls at a canonical DNA contact position, is essential for DNA binding activity and is predicted to make a base-specific contact. Other residues that might be expected to contact DNA include R149, R139, and H150. None of these residues is significantly perturbed in the ADR1z'-DNA complex. In fact, very little of finger 2 is perturbed upon DNA binding, with changes in chemical shift limited to the extreme N-terminus of the helix, where R143 lies (Figure 82). Thus the zinc fingers of ADR1 appear to bind DNA in different orientations, with finger 1 making full use of its alpha helix to bind DNA while finger 2 makes very limited contact. The extremely suboptimal contact by finger 2 likely explains the necessity of the N-terminus and the requirement for three specific contacts by finger 1 as opposed to the more usual two.

IV. The Role of the N-terminus

The ADR1z' construct contains the entire region N-terminal to the zinc fingers that is required for high affinity DNA binding. This N-terminal region does not appear to adopt an ordered structure in this construct, however, exhibiting all the hallmarks of random coil. This situation may not reflect the behavior of the

N-terminus in the context of native ADR1. It is possible that other domains of the protein are necessary to stabilize structure in this region. In particular, it is known that while residues 17-84 are not necessary for high affinity DNA binding, their presence does increase DNA binding affinity somewhat. These residues may stabilize structure in the necessary region of the N-terminus, from residue 85-102.

While there is little change to the zinc fingers upon DNA binding, the N-terminus becomes structured. There is no evidence from the NMR data that the N-terminus adopts a regular secondary structure, however, and there is direct evidence demonstrating that it does not fold into a helix. The structure in this region is probably mostly extended.

The transition between random coil residues and structured residues in the N-terminus of bound ADR1z' falls at precisely the border of the minimal DNA-binding domain determined by mutagenesis. A pattern of strong chemical shift perturbations in the N-terminus also terminates at exactly this position, clustering in the region of the known suppressor mutation, R91K. These data suggest that the most critical region of the N-terminus is the region near the edge of the minimal DNA-binding domain. The position of the suppressor mutation itself, R91, does not appear to be perturbed upon DNA binding. This suggests that the wild-type

residue does not make a contact to DNA; the mutant lysine sidechain likely forms a novel contact that increases DNA binding affinity.

The N-terminus appears to play its role in DNA binding by making direct contacts to the DNA. While the N-terminus is strongly perturbed upon DNA binding, only the face of the zinc fingers that contacts the DNA is similarly perturbed. Other portions of the fingers are largely unperturbed, suggesting that the N-terminus does not interact with the exposed portions of the fingers. Data from paramagnetic labeling experiments show that the N-terminus lies near the zinc fingers in the ADR1z'-DNA complex. In agreement with data from ethylation interference experiments, the N-terminus likely binds to the same DNA site as the zinc fingers, arising from finger 1 at the edge of the UAS1 site and extending back toward the center of the palindrome.

It is not clear from mutagenesis or NMR data whether the N-terminus contacts the backbone of the DNA or the minor groove. The preponderance of circumstantial evidence suggests that the N-terminus makes some contact to the DNA backbone, likely on the same strand contacted by the zinc fingers. This is a model that remains to be tested in future structural studies of ADR1z'-DNA.

ADR1 is the first zinc finger-based DNA binding domain shown to use a region outside the finger array to contact DNA. Other finger proteins do require additional regions for DNA binding activity, including SWI5 and Tramtrack. In these cases, however, the sequence adjacent to the N-terminal finger folds to become an integral part of that finger's structure; in the case of Tramtrack the resulting third strand of beta sheet does not contact DNA directly. In the case of SWI5, this third strand of beta sheet is preceded by a helix that may contact DNA, but this has not been demonstrated directly. Similarly, while the GAGA protein requires an extensive region N-terminal to its single finger to bind DNA with high affinity, the precise role of this region is not known (91).

ADR1 is not alone in its use of sequences adjacent to a conserved DNA binding motif to contact DNA. This happens frequently for the ubiquitous helix-turn-helix DNA binding motif, for example, with different types of accessory sequences used by the lambda repressor, homeodomains, and winged HTH DNA binding domains. As more and more zinc finger proteins are studied, it is possible that the presence of accessory DNA binding sequences will not be uncommon.

Because zinc fingers fold to stable domains and because arrays of three or more such domains bind DNA with sufficient affinity for structural studies, the obvious first step in structural

characterizations has been to study isolated zinc finger sequences. Of the hundreds of putative Cys_2-His_2 zinc finger proteins identified in eukaryotic genomes, only a handful have been characterized thoroughly in terms of DNA binding activity. Fewer have been characterized structurally, and in only three of these cases have the fingers been studied in a context that extends beyond the zinc finger consensus sequence. It seems likely that as more zinc fingers are analyzed in a larger context of surrounding protein sequence, more variation on the basic zinc finger theme will be identified.

V. Future Directions

Future studies will be conducted by a post-doctoral fellow and two graduate students. Dr. David Hyre is currently using the system and assignments developed in this thesis work to study relaxation and hydrogen exchange in an effort to understand the dynamics of free and bound ADR1z'. Lawrence Schaufler has completed assignments of the free 14mer DNA used in this project and is currently working on assignments for the bound DNA. Peter Bowers is beginning work on various mutants of ADR1z' in an effort to understand their DNA-binding behavior. Calorimetry will also be used to study the DNA binding properties of ADR1z' and longer ADR1

constructs. These quantitative studies should establish whether or not there is any cooperativity to ADR1 DNA binding.

The next step of the structural studies will be a calculation of the structure of the free ADR1z' polypeptide, based on the assignments obtained here and on additional NOESY data. This structure should make clear for the first time the nature of the Cys-Cys loop of finger 2. While detailed structural information has been unavailable for the poorly behaved wild-type single finger peptide ADR1a, qualitative evidence from the present work indicates that this region will be different from the Cys-Cys loop of finger 1 and the corresponding ADR1b peptide.

A final step will be the calculation of a three-dimensional structure of the ADR1z'-DNA complex, based on the system and assignments developed here. This structure will likely rely heavily on intermolecular NOEs to define the precise protein-DNA interface, and therefore the structure determination must await assignment of the bound DNA. All attempts at obtaining coupling constants for the bound protein have failed thus far, due to intrinsic relaxation properties of the complex. Thus it is unlikely that a description of the exact course of the N-terminus will be available until intermolecular NOEs can be identified unambiguously.

CHAPTER 7. METHODS

I. Subcloning

A. STARTING MATERIALS

The polypeptides employed in this study were produced using a bacterial expression system based on the method of Studier (72). Two plasmids were built that yielded polypeptides encompassing either residues 75 to 178 (pMS4) or residues 75 to 161 (pMS8) of ADR1. The starting point for each construction was a plasmid (pUC19 #1) provided by Sushil Thukral of the laboratory of Ted Young at the University of Washington. All vectors used in construction were propagated in *E. coli* strain DH5α in M9 media, using the calcium chloride method to produce competent cells and heat shock to transform them (92). Restriction enzymes and ligases were purchased from Boehringer Mannheim or New England Biolabs.

pUC19 #1 was constructed by Sushil Thukral from Ted Young's plasmid pTY-7, in which the 460 bp. *Hinc* II-*Taq* I fragment of ADR1 that spans the coding sequence for amino acids 18 to 170 was ligated into the pUC19 polylinker at the *Sma* I and *Acc* I sites. The *Sma* I site of the polylinker was destroyed by ligation with the *Hinc* II sticky ends, but the *Taq* I/*Acc* I ligation created a new *Sal* I site at the 3' end of the ADR1 sequence.

This new *Sal* I site and a downstream *Hin*d III site from the pUC19 polylinker were used to insert an oligonucleotide cassette into pTY-7. This oligonucleotide had 5' and 3' overhangs complementary with *Sal* I and *Hin*d III sticky ends respectively and contained the coding sequence for ADR1 residues 170 to 178, followed by a TGA stop codon and a *Cla* I site. Insertion of the cassette into pTY-7 created a pUC19 plasmid that coded for ADR1 18 to 178, followed by a stop codon. The *Sal* I site was destroyed by insertion of the cassette.

The final step in the creation of pUC19 #1 was the insertion into the modified pTY-7 of a sequence from an M13mp18 vector. This vector had been created by ligating the sequence coding for ADR1 amino acids 40 to 150 into the polylinker of M13mp18 between the *Sal* I and *Sph* I sites. The *Sph* I site at codon 150 in ADR1 is natural, but the *Sal* I site had been added to the 5' end of the truncated ADR1 sequence using a *Sal* I linker. Oligonucleotide site directed mutagenesis was then used on the M13mp18 construct to create an *Nde* I site at the codons for ADR1 amino acids 74 and 75. The ATG codon for amino acid 75, a methionine, did not change, but the preceding codon was altered from AAT to CAT. This sequence was then lifted from M13mp18 at the *Sst* I site in the polylinker upstream of the ADR1 sequence and at the *Sph* I site at the codon for amino acid 150. This fragment was ligated into the

modified pTY-7 vector, using the *Sac* I site in the remaining upstream polylinker and the *Sph* I site at the codon for ADR1 amino acid 150.

The resulting pUC19 #1 plasmid supplied by S. Thukral carries the pUC19 sequence, with the upstream pUC19 *Eco*R I and *Sac* I sites. These are followed by the polylinker of M13mp18, through the *Sal* I site. Then the sequence of ADR1 follows, starting with the codon for amino acid 40 and ending with the codon for amino acid 178, which is followed by a stop codon, a *Cla* I site, and the *Hin*d III site from the pUC19 polylinker. The ADR1 sequence contains one mutation which lies at the codon for amino acid 74 and arose from the creation of the *Nde* I site. Residue 74 is not included in either expressed polypeptide, however, since both start at methionine 75.

B. pMS4 FOR ADR1z (ADR1 75-178) EXPRESSION

The plasmid pUC19 #1 described above was used as the starting point for the construction of pMS4, a construct for overexpression of ADR1z (75-178). The vector used for overexpression is the Studier-type plasmid (72) pET11a (Novagen). This vector allows insertion of an expression sequence between unique *Nde* I and *Bam*H I sites. pUC19 #1 contains an *Nde* I site at

the position of codon 75, but it does not contain a convenient 3' *Bam*H I site. For this reason, the *Bam*H I site from the plasmid pATH-3 was picked up in the following manner. Both pUC19 #1 and pATH-3 were subjected to *Hin*d III/*Sca* I digestion, and the fragments were isolated on a low melting point agarose gel. The small pUC19 #1 fragment and the large pATH-3 fragment were recovered and ligated to give pMS5. This places a *Bam*H I site from the pATH-3 polylinker just 3' to the coding sequence for ADR1. Cutting with *Nde* I and *Bam*H I, isolating the fragment of appropriate size, and ligating into the pET11a vector yielded pMS4 (named before the strategy for its construction was formulated, and thus before the design of its progenitor, pMS5).

C. pMS8 FOR ADR1z' (ADR1 75-161) EXPRESSION

Plasmid pMS5 described above was used as the starting point for construction of pMS8, a construct for overexpression of ADR1z' (75-161). An oligonucleotide cassette was inserted into pMS5 between the *Sph* I and *Hin*d III sites. This oligonucleotide contained 5' and 3' overhangs complementary to *Sph* I and *Hin*d III sticky ends, respectively. It contained the coding sequence for ADR1 residues 150 to 161, followed by a TAA stop codon. This oligonucleotide was phosphorylated and ligated into pMS5, to give

pMS6. pMS6 contains coding sequence for ADR1 residues 40 to 161, followed by a stop codon, with an *Nde* I site at codon 74 and 75.

The *Sal* I/*Hin*d III fragment of pMS6 was isolated and ligated into the polylinker of pBluescript II KS +/- (Stratagene) for sequencing. The resulting plasmid is named pMS7. The sequence was read from both the 5' and 3' ends using the T3 and SK primers (Stratagene). The sequence of the inserted cassette was confirmed, including the in-frame 5' junction with the ADR1 sequence at codon 150, the presence of the new in-frame stop codon after codon 161, and the presence of the point mutation at codon 74 in the *Nde* I site. Another mutation lies at the codon for amino acid 77 and arose spontaneously. However, the mutation is silent and changes the wild-type serine codon UCC to the serine codon UCG.

To create pMS8 for ADR1 75 to 161 expression, pMS7 was cut at the unique *Nde* I and *Bam*H I sites, and the appropriate fragment was ligated into pET11a, as for pMS4 .

D. OVEREXPRESSION SYSTEM

For overexpression of ADR1z and ADR1z', plasmids pMS4 and pMS8, respectively, were placed in *E. coli* strain BL21/DE3. This strain carries a lambda prophage containing the gene for T7 polymerase under the control of an inducible *lac* promoter (72).

Upon addition of IPTG to the culture, repression of T7 polymerase expression is lifted, and T7 polymerase is produced. This polymerase in turn initiates transcription exclusively from the pET11a T7 promoter, driving expression of the ADR1 insert. The resulting transcripts are translated, beginning with the ATG start codon that exists as part of the *Nde* I insertion site and ending with the engineered stop codon following codon 178 or 161. Because the start codon used is a part of the wild-type ADR1 sequence, the resulting expressed protein contains only native sequence.

II. Purification

A. PROTEIN

i. Media and Isotopic Labeling

In all cases the expression strains were grown on minimal medium. The basic medium contains M9 salts including 1 g/L NH_4Cl; 4 g/L glucose; 0.002% thiamine; 1 mM $MgSO_4$; 0.1 mM $CaCl_2$; and 100 μg/mL sodium ampicillin (92).

In cases where uniform ^{15}N labeling was required, $^{15}NH_4Cl$ was substituted for the unlabeled equivalent. All stable isotopes

were obtained from Cambridge Isotope Laboratories or Isotec. When both uniform ^{15}N and ^{13}C labeling were required, $^{15}NH_4Cl$ and $^{13}C_6$-glucose were substituted for the unlabeled equivalents. The glucose was added by sterile filtration of a 10% solution as the last step of media preparation, to a final concentration of 2 g/L. This glucose concentration is half the normal value, and it gives more cost-effective double-labeled preps since the yield per gram of glucose is slightly better at this concentration. This strategy was not used for any other labeling, since all other additives must remain at normal concentrations.

In cases where specific ^{13}C labeling was done, unlabeled glucose was used, $^{15}NH_4Cl$ was substituted for the unlabeled equivalent, and the specifically ^{13}C-labeled amino acid was added by sterile filtration as the last step in media preparation. A total of 18 ml of 5.5 mg/ml labeled amino acid solution was added per liter of culture, with 4 ml added at the start of growth, 10 ml at induction, and the final 4 ml at the induction halfway point (74).

When specific ^{15}N labeling was required, unlabeled NH_4Cl was used. Every unlabeled amino acid was added to the medium, except glutamate and the amino acid to be labeled. Solutions of unlabeled amino acids were prepared and filter sterilized at the stock concentration advised (93), except for glutamine, which was made to half of the recommended concentration. However, rather than

adding the recommended 5 ml of each amino acid stock per liter of medium, 10 ml per liter was added at the start of growth, and then an additional 2.5 ml at induction. The specific ^{15}N-labeled amino acid was added at induction by filter sterilization, to a concentration of 80 mg/L (93).

ii. Cell Growth and Induction

Cells were stored as freezer stocks at -70 °C in minimal medium with 10% glycerol. For each protein prep. a colony was selected from a freshly streaked MMA plate containing 100 μg/ml ampicillin (92), that was incubated one day at 34.5 °C. This colony was inoculated into a small starter culture identical to the medium to be used, except that no amino acids were added. The volume of the starter culture was 2% of the final medium volume and was grown in a large (500 ml or larger) culture flask, shaking at 37 °C , for around twenty hours.

The final medium was inoculated with this overnight culture at 20 ml per liter of culture, yielding a starting OD at 600 nm of about 0.01 to 0.04 with a 1 cm path length. The large cultures were grown 500 ml per 2 L flask for proper aeration. The cultures were grown shaking, at 37 °C. Typical preparations involved two to three liters of culture.

When the cultures reached an OD reading of about 0.25 at 600 nm, zinc chloride was added through a sterile filter to a final concentration of 200 μM. At this point the cultures were near log phase, doubling about once every hour. When the cultures reached an OD reading of about 0.5 at 600 nm, just past log phase, IPTG was added by sterile filter to a final concentration of 1 mM. Induction continued for ten hours, except in the case of labeling with ^{15}N amino acids, where induction proceeded only three hours to minimize metabolic scrambling of the label. Cultures allowed to induce for ten hours typically reached an OD at 600 nm of about 1.7.

iii. Cell Harvest and Lysis

Cells were harvested by centrifugation at 3000 RPM in swinging bucket rotors, for 30 minutes at 4 °C. Cells were resuspended in about 40 ml harvest buffer, consisting of 50 mM HEPES pH 7.6, and 50 mM NaCl. The cells were respun at 10,000 RPM for 10 minutes at 4 °C and decanted. The cell pellet was stored at -70 °C.

Cells were resuspended for lysis in about 20 ml cold lysis buffer, consisting of 50 mM HEPES pH 7.6, 100 mM KCl, 0.1 mM EGTA, 0.25 mM fresh PMSF and 5 mM fresh DTT. They were broken in two runs through a chilled French Pressure cell, at a setting of

900. The viscous extract was incubated on ice with 100 µl of $MgCl_2$ and approximately 800 Kunitz units of DNAse I until the extract flowed smoothly. The extract was spun at 10,000 RPM for 20 minutes at 4 °C, to remove a pellet of insoluble debris. The pellet was resuspended in a few milliliters of lysis buffer, respun, and the supernatant was combined with the extract. The extract was stored at -70 °C.

iv. Cation Exchange Chromatography

The thawed extract was loaded at 4 °C on a 1.5 x 25 cm column of CM Sephadex C-25 (Pharmacia) equilibrated in low-salt running buffer consisting of 50 mM HEPES pH 7.6, 0.1 mM EGTA, 0.1 mM fresh PMSF, 5 mM fresh DTT, and either 200 or 300 mM NaCl, for ADR1z' or ADR1z respectively. The column was washed in the appropriate low salt buffer until the absorbance of the eluent fell below 0.07 at 275 nm with a 1 cm path. At this point a 600 ml gradient was started, using 300 ml of the low salt buffer and 300 ml of the same buffer containing 2 M NaCl. Eighty 5.0 ml fractions were collected. The polypeptides ADR1z and ADR1z' first appear in fractions having a conductivity of 31 or 27.5 mS respectively, where the 200 mM low salt buffer has a conductivity of 14 mS.

The fractions were analyzed for ADR1 polypeptides on 12% SDS-polyacrylamide gels made with a ratio of 38:2 acrylamide:bis-acrylamide. ADR1z runs approximately even with a 14,400 dalton lysozyme standard (Biorad), and ADR1z' runs slightly farther. The presence of ADR1 in the induced cells and in the fractions was tested in early preps by Western blots probed with a polyclonal antibody to the ADR1 zinc fingers. This antibody was obtained from the Young lab and affinity purified against a LacZ-ADR1 fusion protein bound to a nitrocellulose filter. The specific DNA-binding activity of the fractions was also assayed in early preps using gel shift assays (48). In these tests the cell extract or fractions were incubated with radioactively labeled oligonucleotides containing the ADR1 recognition sequence. The resulting complexes were run out on non-denaturing gels and autoradiograms were taken.

v. Reverse-Phase HPLC Chromatography

Those fractions that contained ADR1 polypeptide were pooled, dialyzed into 0.1% acetic acid in MWCO 3500 Spectrapor tubing (Spectrum) at 4 °C, and lyophilized. The granular, dried protein was suspended in about 16 ml 50 mM tris buffer pH 7.5, and fresh DTT was added to 50 mM. The protein was allowed to reduce at room temperature for two hours. The solution was clarified by adding about 20 µl glacial acetic acid, spun, and loaded on a semi-

preparative C4 reverse-phase HPLC column (Waters) in 0.5 ml aliquots. The column was run in a mix of water/0.1% TFA and acetonitrile/0.1% TFA at 3 ml/min and the detector was set at 212 nm. ADR1z was loaded in 26% CH_3CN and was eluted with a 26% to 36% CH_3CN gradient over 10 minutes. ADR1z' was loaded in 29% CH_3CN and was eluted with a 29% to 35% CH_3CN gradient over 12 minutes. The largest peak was collected. This peak appeared for ADR1z' at about 34% CH_3CN (approximate concentration at column - there is a lag between pumps and detector), and for ADR1z at about 33% CH_3CN. The pool was lyophilized and the resulting powder was stored at -20 °C.

B. DNA

i. Deprotecting Oligos

Single-stranded DNA oligonucleotides were obtained from the DNA Program Project synthesis facility run by the Chemistry Department at the University of Washington. Typically a 10 μmol synthesis was requested of two complementary strands of about 14 to 17 bp. Generally, the oligonucleotides were delivered deprotected in ammonia. When oligos were supplied prior to deprotection, cleavage from the synthesis support resin and deprotection of the base groups was accomplished by incubation

with ammonia according to published protocols (94). This step is critical for the later separation of failure sequences on a sizing column and for protein-binding activity. The most critical aspects of the technique are use of saturated ammonia that is as fresh as possible, and proper sealing of the incubation tubes to prevent escape of ammonia during heating steps. Following deprotection, the oligos are dried on a Speedvac, resuspended, and precipitated twice from 0.5 M NaCl with 3 volumes of dry ice cold ethanol.

ii. Separating Failure Sequences

The failure sequences, including n-1 failures, were separated from the DNA of interest by size exclusion chromatography. A 2.5 cm x 95 cm column of G50 superfine Sephadex gel filtration resin (Pharmacia) was poured and calibrated using blue dextran and xylene cyanol as standards. Great care was taken to ensure that the resin was degassed before pouring; the resin was not stirred, however. In addition, the column was run in thoroughly degassed water only. A single-stranded DNA oligo was brought into about 1 ml of water, and a few drops of NH_4OH were added if needed for solubility. The DNA was loaded directly onto the bed of the column while running. Ninety-five 3.1 ml fractions were collected after the void volume. The fractions were assayed by running 2 or 3 µl on a 20% polyacrylamide gel prepared with 19:1 acrylamide:bis-

acrylamide and 0.6 g/ml urea. The gel was allowed to stand at room temperature for at least two hours before use. Samples were loaded in 750 mM urea containing not more than 0.03% bromophenol blue. The gel was run until the dye was about 1 cm from the bottom (for oligos around 14 to 17 bp). The gels were stained with a Silver Stain Plus kit (Biorad). The DNA appeared as negative bands on a stained background. Those fractions containing pure full-length sequence were pooled and lyophilized. Those containing contamination primarily with n-1 failures were pooled, lyophilized, and reapplied to the G50 column as before.

iii. Annealing Strands and Removing ssDNA

The pure, full-length, single-stranded oligo was quantitated (see below) and combined in equimolar amounts with its complementary sequence in 3 ml annealing buffer consisting of 20 mM sodium phosphate pH 6.8, 400 mM NaCl, and 0.2 mM EDTA. The mixture was placed in a water bath at 82 °C, which was then allowed to cool to room temperature overnight.

The mixture was then applied to a hydroxylapatite column to remove single-stranded DNA. The 1.5 cm diameter column was poured from 10 g Biogel HTP (Biorad) and 2 g cellulose powder (Whatman) swollen in about 60 ml 10 mM sodium phosphate pH 6.8.

It was prerun in 100 ml 400 mM sodium phosphate pH 6.8, and then in 200 ml 10 mM sodium phosphate pH 6.8. The annealing mixture was diluted ten times in 10 mM sodium phosphate pH 6.8 before loading on the column. Fractions (1.2 ml) were collected. The column was washed in pH 6.8, 10 mM phosphate buffer until the absorbance of the fractions at 260 nm came down near baseline. Then a 200 ml 10 to 400 mM pH 6.8 phosphate gradient was run. Those fractions eluting during the gradient which absorbed above about 1.0 at 260 nm with a 1 cm path were pooled. Care was taken that only about 22 fractions were pooled, because adding more caused such a high concentration of phosphate that the sample would not easily desalt.

iv. Desalting

The pooled fractions were lyophilized, and then resuspended in 1.5 ml water. The double-stranded DNA was applied to a 2.5 cm x 95 cm column of degassed Sephadex G10 resin run in degassed water. Fractions were collected (3.2 ml), and those containing DNA by the absorbance at 260 nm were pooled and lyophilized. The powder was stored at -20 °C

III. NMR Samples

A. HANDLING AND QUANTITATING MATERIALS

i. DNA

Double-stranded DNA was routinely stored as a powder at -20 °C. It was quantitated by reading absorbance values at 260 nm. An absorbance of 1.0 with a 1 cm path is expected for a solution containing approximately 50 μg/ml ds DNA. Each basepair in the DNA contributes approximately 660 daltons to its molecular mass. DNA quantities were commonly recorded in terms of ODUs, where one ODU corresponds to the amount of DNA in one ml of a solution with an A_{260} of 1.0 using a 1 cm path.

ii. Protein

Polypeptides were routinely stored as a powder at -20 °C. They were quantitated by taking advantage of the tyrosine absorbance at 275 nm. At this wavelength free tyrosine has a molar extinction coefficient of 1420 $M^{-1}cm^{-1}$ (95). Both ADR1z and ADR1z' have two tyrosine residues, and therefore the molar extinction coefficient for each polypeptide can be approximated as 2840 $M^{-1}cm^{-1}$.

When zinc finger peptides are dissolved in water, three processes can occur: zinc binding, folding, and oxidation. Management of zinc finger peptides requires careful consideration of all three. The peptides as they come off the HPLC are the easiest to handle, because they represent a sort of blank slate. Because the column is run in TFA, the peptides have been exposed to an extremely acidic environment. This has two effects. First, while the DTT is removed on the column, the low pH protects the reduced peptides from oxidation. Second, low pH (< 4.8) will cause protonation of the histidine zinc ligands and unfold the fingers, and extremely low pH is expected to strip zinc off the cysteine ligands as well, forming apopeptide. This fully reduced, unfolded, zinc-free state is a useful starting place for making NMR samples.

It is often convenient to return to this "blank" state when a sample needs to be overhauled, and the simplest way to do this is to reduce and re-HPLC the sample. Dialysis is tricky. It is best accomplished by reducing the protein with DTT and then dialysing into 0.1% acetic acid, since acid conditions favor the solubility of the protein and protect from oxidation. However, exposing the peptide to acid conditions for extended periods of time leaves the unfolded form available for proteolysis. Also, since zinc may diffuse off the partially liganded peptide into large volumes of dialysis buffer, it is difficult to know the final stoichiometry of

zinc in the sample. Adding zinc to the buffer is problematic, as zinc and DTT form an insoluble complex near neutral pH. Using β-mercaptoethanol as the reducing agent prior to dialysis may alleviate this problem during dialysis, but there is likely to be excess zinc in the final product that will interact with any DTT necessarily added to the NMR sample.

Of all the issues faced in handling zinc finger polypeptides, the most challenging is oxidation. There are five cysteine residues in ADR1z and ADR1z', four of which ligand zinc. The unfolded peptide is subject to rampant oxidation to multiple species, and it must be protected from high pH environments prior to the addition of zinc and folding of the peptide. The folded peptide is much more resistant to oxidation, but an intermolecular dimer can still form between non-zinc-ligand cysteines. This dimer can be distinguished from the monomer by analytical HPLC. Again, acid conditions protect against this oxidation, and both β-mercaptoethanol and DTT are effective at reducing the dimer. DTT is the stronger reducing agent, however, and at high temperature and pH (eg. 310K and pH 7) it can maintain a reduced state where βME cannot.

While DTT is the favored reducing agent for ADR1 polypeptides, one must be aware of its affinity for zinc and the insolubility of the resultant complex. Because DTT's zinc affinity is less than that of the ADR1 fingers, zinc is added and the fingers

are folded before any large quantity of DTT is added. Folding of the peptide under these conditions is contingent upon the sample being in a reduced state to begin with, so peptide that was reduced, kept as powder, and then solubilized at low pH is used.

B. FREE PROTEIN SAMPLES

i. Zinc-bound Samples

Both ADR1z and ADR1z' are soluble as free polypeptide up to the maximum concentration tested, 5 mM. However, samples more concentrated than 2 mM have been observed to experience concentration-dependent line broadening of NMR spectra, and are therefore not recommended. Samples were normally prepared at room temperature, but they have been studied in the NMR at up to 310K. Zinc fingers are quite stable to thermal denaturation (Grace Párraga, doctoral thesis, University of Washington) , so this should not represent an upper limit.

Samples were typically prepared from lyophilized polypeptide that was brought up in 0.1% deuterated acetic acid to protect from oxidation. Zinc was added next, from a 100 mM stock of zinc chloride containing just enough added HCl to avoid precipitation of zinc hydroxide. Normally 2.0 equivalents of zinc were added (one

per finger), but polypeptides have been folded with non-stoichiometric addition of 2.2 - 2.5 equivalents of zinc. Excess zinc is not recommended if DTT will be added to the sample. Deuterated tris buffer was added next at 25 to 50 mM, and the sample was adjusted to a pH above 5.2 with deuterated sodium hydroxide. This increase in pH drives the folding of the fingers by allowing unprotonated histidine sidechains to coordinate the zinc ion. The polypeptides prefer acid pH, but they are folded and soluble from pH 5.2 to at least pH 7. The samples were completed with 0.02% sodium azide, to prevent bacterial growth, enough D_2O to reach 10%, and fresh reducing agent (see below). Samples were transferred to 5 mm NMR tubes (Aldrich) and placed under an atmosphere of nitrogen or argon to help prevent air oxidation. The tubes were stored at 4 °C.

Both 1 µl βME (35 mM) or 5 mM DTT worked as reducing agents in NMR samples, but DTT was preferred, because it reduces effectively at high pH and temperature where βME does not. It is critical that DTT stocks are fresh. They are prepared fresh from powder, or stored frozen at 1 M. It is essential that frozen aliquots be large, on the order of several hundred µl. Smaller aliquots freeze and thaw in frost-free freezers and rapidly become air-oxidized and useless. Therefore, the use of cheap protonated DTT is recommended, especially for heteronuclear edited NMR experiments

where large ^1H lines will not interfere. Deuterated DTT is prohibitively expensive and therefore must be made up once and stored in small aliquots; this is disastrous for the reducing power of the agent.

ii. Cobalt-bound Samples

Cobalt-bound samples were prepared much like zinc-bound samples. Because cobalt is paramagnetic and relaxes NMR signals, care was taken to avoid excess cobalt in the NMR samples. Cobalt was quantitated as a complex with terpyridine (96), using 2.88×10^4 as the molar absorbance of the $terpyridine_2Co_2^{++}$ complex at 320 nm. 0.95 equivalents of cobalt were added per zinc finger, and the presence of tetrahedrally coordinated cobalt was confirmed by UV absorbance (85). All cobalt-bound samples were prepared in a glove bag under an atmosphere of argon in the absence of reducing agents. Reducing agents can alter the oxidation state of the metal, and furthermore, DTT competes efficiently with the zinc fingers for cobalt. Zinc fingers bind cobalt with an approximately 10,000-fold reduced affinity compared to zinc (97).

C. FREE DNA SAMPLES

DNA samples were made up from lyophilized material. The ingredients were essentially the same as for the free protein samples, with the following exceptions. No reducing agent was used. Zinc was not added, but 40 μM EDTA was often added to avoid effects from heavy metals. In cases where extra zinc was necessary for other reasons, it was noted that the chemical shifts of peaks from the DNA were perturbed by the presence of excess zinc. While DNA prefers basic pH, the conditions can be changed quite freely with respect to pH, temperature, and concentration.

D. SPIN-LABELED DNA

How got, what is, where, how anneal, argon, don't reduce until end!

E. PROTEIN-DNA COMPLEXES

There are some special considerations when preparing a protein-DNA complex, but the starting point is much the same as for the free samples. Each component of the complex was prepared exactly as an NMR sample of the free macromolecule would be prepared (see above), except that the Tris concentration was often

10 mM. The concentration of each macromolecule was twice the desired complex concentration, in half the desired final volume. The pH was typically 7, since the complex prefers alkaline pH. The pH was set correctly in the component samples, as adjusting the pH of the complex can cause precipitation, especially as lower pH values are approached.

The components of the complex were mixed at room temperature, but NMR spectra of the complex were significantly better at 310K than at room temperature or 303K, and so most NMR work was done at the high temperature. Prior to mixing, the DNA sample was placed in a commercially available pre-siliconized Eppendorf tube. The protein was added to the DNA in aliquots of 4 to 6 μl, mixing gently by tapping the tube between additions. The full addition took about three hours. If all proceeded smoothly, most of the additions visibly dissolved. There was always some amount of cloudiness or the presence of small chunks by the end of the full addition, but abundant precipitates that do not dissolve early in the titration are a sign that all is not well.

The most likely cause of problems is that the protein is not fully reduced. It is absolutely essential that βME or fresh DTT be present in the protein component as described above for free samples. If reducing agents were omitted or were air-oxidized, the highest complex concentration that was achieved was somewhat

less than 0.5 mM. If the protein was properly reduced, complex samples of at least 2.0 mM were possible.

Unfortunately, protein-DNA complex samples of any concentration are not very stable to manipulation. A perfectly clear sample can be caused to precipitate by a change in pH, a change in temperature (especially to higher temperatures), or by the addition of molecules such as DTT or zinc. Some precipitates are minor and can be spun out without loss of too much sample. A precipitate like this often appears after the first exposure to a higher temperature or when fresh DTT is added to an old sample to ensure reducing conditions.

The more severe type of precipitate, caused by oxidation, large changes in pH, or by the addition of excess zinc when DTT is present (or of DTT when excess zinc is present), often leads to loss of too much material for the sample to be useful. In these cases the protein can be recovered by adding 800 mM NaCl to the precipitated sample to dissociate the complex, and then subjecting the sample to a recovery by HPLC. In some cases, addition of guanidine HCl to concentration of 4 M may be necessary to solubilize the precipitate. In all cases, the protein elutes from the column at the same percentage of CH_3CN, using the same gradients, as normal, free protein.

IV. NMR Spectra

A. GENERAL CONDITIONS

Except as noted, all NMR spectra were acquired on either a Bruker AM500 or a DMX500 Spectrometer, the latter being equipped with triple-axis pulsed field gradients and a three-channel probe. Sample volumes were 450 µl in 5 mm tubes. Free polypeptide spectra were generally acquired at pH 5.4, 303 K, and 2 mM. ADR1z'-DNA spectra were generally taken at pH 7.0, 310 K, and 1.9 mM. Temperature, pH, and concentration are described in the text or in figure legends for each experiment when they vary from these typical conditions. Spectra were processed using Felix 2.3 or Felix95 (Hare Associates, Biosym).

B. REFERENCING OF SPECTRA

Proton chemical shift was referenced to external TSP in sample buffer as the zero ppm standard. The position of TSP was measured for various combinations of temperature and pH, and the appropriate value was used for each sample.

^{15}N chemical shifts were referenced indirectly to TSP. The ratio of the ^{15}N zero ppm frequency to the ^{1}H zero ppm frequency was taken to be 0.1013291444 (98).

^{13}C chemical shifts were referenced to external p-dioxane in sample buffer. The position of p-dioxane resonance was taken as 67.8 ppm (77).

C. NMR EXPERIMENTS AND PARAMETERS

Table V presents a listing of relevant experimental parameters for each type of NMR spectrum used in these studies. Included are the carrier frequencies, spectral widths, and matrix sizes for each dimension, special processing manipulations, detection schemes, and water suppression strategy. The descriptions below are therefore limited to literature references for each technique and any special considerations.

i. HSQC

All HSQC experiments were acquired using gradients for coherence selection and water suppression (99). Typically 450 to 512 nitrogen increments were acquired for uniformly or specifically ^{15}N-labeled samples of free or DNA-bound ADR1z'. In

Table V. NMR Parameters

spectrum, nucleus	spectral width (Hz)	carrier (ppm)	increments acquired	matrix size (pts)	linear prediction	detection routine	water suppression
^{15}N-HSQC							
^{1}H	6410	H_2O	2048	1024	0%	TPPI	grad. coher. selection
^{15}N	2000	116.1	512	2048	0%		
3D ^{15}N-TOCSY							
^{1}H	7002	H_2O	1024	512	0%	TPPI	grad. coher. selection
^{1}H	4000	8.56	128	256	100%		
^{15}N	2000	116.1	64	128	100%		
2D ^{15}N-TOCSY							
^{1}H	10000	H_2O	2048	1024	0%	TPPI	presat.
^{15}N	10000	148	512	1024	0%		
3D ^{15}N-NOESY							
^{1}H	4006	8.53	1024	512	0%	STATES	watergate
^{1}H	6410	H_2O	256	512	100%		
^{15}N	1562	116.1	34	64	100%		

Table V. (continued)

spectrum, nucleus	spectral width (Hz)	carrier (ppm)	increments acquired	matrix size (pts)	linear prediction	detection routine	water suppression
HNCA							
^{1}H	4006	8.63	1024	512	0%	TPPI	watergate
^{15}N	2000	116.1	80	256	100%		
^{13}C	3521	52.4	48	128	100%		
HN(CO)CA							
^{1}H	4006	8.63	1024	512	0%	TPPI	watergate
^{15}N	2000	116.1	80	256	100%		
^{13}C	3521	52.4	48	128	100%		
CBCA(CO)NH							
^{1}H	4006	8.56	1024	512	0%	STATES	gradient
^{15}N	2000	116.1	80	256	0%	sensitivity	coherence
^{13}C	5000	43.4	60	128	0%	enhanced	selection
HNCACB							
^{1}H	4006	8.65	1024	512	0%	STATES	gradient
^{15}N	2000	116.1	80	256	0%	sensitivity	coherence
^{13}C	7000	43.4	60	128	0%	enhanced	selection

Table V. (continued)

spectrum, nucleus	spectral width (Hz)	carrier (ppm)	increments acquired	matrix size (pts)	linear prediction	detection routine	water suppression
3D ^{13}C-NOESY							
^{1}H	5121	H_2O	1024	512	0%	STATES	gradient coherence selection
^{13}C	2641	H_2O	54	128	50%		
^{1}H	5000	43.4	220	256	50%		
HCCHTOCSY							
^{1}H	5000	2.96	1024	512	0%	STATES	gradient coherence selection
^{1}H	3501	2.96	236	256	50%		
^{13}C	2641	43.4	58	128	50%		
long range HMQC-J							
^{1}H	6410	H_2O	2048	1024	0%	magnitude	grad. coher. selection
^{15}N	7645	200	96	1024	0%		
J-mod. HSQC							
^{1}H	6410	H_2O	2048	1024	0%	TPPI	trim pulses
^{15}N	950	118.3	128	1024	100%		

the case of linewidth determinations for the free protein, 1200 increments of nitrogen were collected and the nitrogen dimension was processed without apodization.

Differential labeling of residues in specifically ^{15}N-labeled samples was detected by HSQC experiments. Intensities of HSQC crosspeaks from specifically labeled samples were compared to intensities from uniformly labeled samples. The resulting ratios are all less than one, even for the residue type that originally bore the ^{15}N label, due to some remaining cellular synthesis of the amino acid and due to transamination of the labeled amino acid by other, unlabeled amino acids (73). Such transamination also leads to transfer of the ^{15}N label into multiple amino acid types during the process of growth and induction. Because the cellular transaminases have differing amino acid specificities, the transfer is not uniform across all amino acid types. This differential scrambling of label is easily measured by the ratio of crosspeak intensities mentioned above; the ratio should be constant for a given amino acid type within a protein, but it should vary across amino acid types, depending on the efficiency of transamination. All crosspeaks can thus be classified based on their extent of labeling. Once an example of each amino acid type has been assigned by other methods and its extent of labeling determined,

the amino acid types belonging to each class of labeling are evident.

ii. NOESY

A 2D D_2O homonuclear ^1H NOESY with a mixing time of 150 ms was taken on ADR1z' using a standard pulse sequence. A single 3D ^{15}N-edited NOESY-HSQC spectrum of free ADR1z' was also acquired with a mixing time of 150 ms (100). Two such 3D experiments were collected for DNA-bound ADR1z', with mixing times of 75 and 150 ms. In addition, a 75 ms 3D ^{13}C-edited NOESY (101) was collected in D_2O on uniformly $^{13}C,^{15}N$-labeled ADR1z' bound to DNA, to determine $\alpha\beta_{i,i+3}$ short-range NOEs.

iii. TOCSY

2D ^{15}N-edited HSQC-TOCSY and TOCSY-HSQC experiments were acquired for free ADR1z' in water using standard pulse sequences, for the assignment of alpha protons. Mixing times were both 40 and 76 ms. In addition, a standard 72 ms 2D homonuclear ^1H D_2O TOCSY was acquired for assignment of aromatic sidechains. Assignment of alpha protons in ADR1z'-DNA required use of a 3D ^{15}N-edited HSQC-TOCSY. The mixing time in this experiment was 31 ms.

iv. HN(CO)CA, HNCA, HNCACB, CBCA(CO)NH

Constant time HNCA and HN(CO)CA experiments were collected in water on uniformly ^{13}C, ^{15}N-labeled ADR1z' free and bound to DNA (102). The carbon dimension of the HNCA was reversed to match the HN(CO)CA. CBCA(CO)NH and HNCACB experiments were collected in water for ADR1z'-DNA (103).

v. HCCH-TOCSY

HCCH-TOCSY experiments were acquired in D_2O with mixing times of 19.6 ms for ADR1z' and 17.9 and 25 ms for ADR1z'-DNA (104, 105). The two experiments for the complex were added to each other following transformation and prior to analysis. ADR1z' was labeled uniformly with ^{13}C and ^{15}N in each case.

vi. J-modulated HSQC

A series of six J-modulated HSQC experiments were acquired for free ADR1z' for the determination of backbone $^3J_{HNH\alpha}$ coupling constants . Total delays τ_2 were set to 46, 65.2, 84.4, 103.6, 122.8, and 142 ms in successive experiments.

vii. Long-range HMQC-J

2D long-range HMQC-J experiments were acquired for ADR1z' and ADR1z'-DNA to determine the tautomeric forms of histidine sidechains (83). The value of $1/2J_{XH}$ was taken as 22 ms.

V. Crystal Trials

Crystal growth trials were conducted using samples of an ADR1z'-DNA complex prepared identically to NMR samples. Specifically, a 2 mM sample of unlabeled ADR1z' was prepared bound to a 14mer ds DNA sequence containing a single basepair overhang on each end. The samples contained 30 mM β-mercaptoethanol.

Sitting drop wells (Charles Supper Co.) were used for crystal growth. Each reservoir received 0.46 mL of precipitant solution plus 2 microliters of β-mercaptoethanol. 2 uL of this solution was placed in the sitting drop platform, along with 2 ul of the protein-DNA complex solution.

Small, twinned crystals were obtained over the course of several weeks at 23 °C. The largest crystals observed were about

0.2mm on the longest side, but these crystals were quite irregular. More regular crystals were also obtained, but these were at most 0.1mm on the longest side. Crystals were obtained over a range of conditions that center on a precipitant solution of 35% PEG 3000 (Sigma), 100 mM tris, pH 6.8 - 8.3, 100 -200 mM NaCl.

Because of the small size and twinning of the crystals, no further characterization was performed, and it is not known whether the crystals consisted of both protein and DNA. Side-by-side controls containing only protein or only DNA did not yield similar crystals, however. Any future trials should focus on verification of the molecular makeup of the crystals, and on microseeding procedures to avoid the twinning and irregular crystal growth observed in these preliminary trials.

References

1. Miller, J., McLachlan, A.D. and Klug, A. *EMBO J.* **4**, 1609-1614 (1985).

2. Brown, R.S., Sander, C. and Argos, P. *FEBS Lett.* **186**, 271-274 (1985).

3. Hanas, J.S., Hazuda, D.J., Bogenhagen, D.F., Wu, F.Y.-H. and Wu, C.-W. *J. Biol. Chem.* **258**, 14120-14125 (1983).

4. Diakun, G.P., Fairall, L. and Klug, A. *Nature* **324**, 698-699 (1986).

5. Frankel, A.D., Berg, J.M. and Pabo, C.O. *Proc. Natl. Acad. Sci. USA* **84**, 4841-4845 (1987).

6. Berg, J.M. and Shi, Y. *Science* **271**, 1081-1085 (1996).

7. Jacobs, G.H. *EMBO J.* **11**, 4507-4517 (1992).

8. Berg, J.M. *Proc. Natl. Acad. Sci. USA* **85**, 99-102 (1988).

9. Párraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T. and Klevit, R.E. *Science* **241**, 1489-1492 (1988).

10. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E. *Science* **245**, 635-637 (1989).

11. Klevit, R.E., Herriott, J.R. and Horvath, S.J. *Proteins Struct. Funct. Genet.* **7**, 215-226 (1990).

12. Hoffman, R.C., Horvath, S.J. and Klevit, R.E. *Protein Sci.* **2**, 951-965 (1993).

13. Bernstein, B.E., Hoffman, R.C., Horvath, S., Herriott, J.R. and Klevit, R.E. *Biochemistry* **33**, 4460-4470 (1994).

14. Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K. and Gronenborn, A.M. *Biochemistry* **29**, 9324-9334 (1990).

15. Kochoyan, M., Keutmann, H.T. and Weiss, M.A. *Biochemistry* **30**, 9396-9402 (1991).

16. Kochoyan, M., Havel, T.F., Nguyen, D.T., Dahl, C.E., Keutmann, H.T. and Weiss, M.A. *Biochemistry* **30**, 3371-3386 (1991).

17. Berg, J.M. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 405-421 (1990).

18. Kaptein, R. *Curr. Opin. Struct. Biol.* **2**, 109-115 (1992).

19. Berg, J.M. *Curr. Opin. Struct. Biol.* **3**, 11-16 (1993).

20. Pavletich, N.P. and Pabo, C.O. *Science* **252**, 809-817 (1991).

21. Kraulis, P.J. *J. Appl. Crystallogr.* **24**, 946-950 (1991).

22. Neuhaus, D., Nakaseko, Y., Schwabe, J.W.R. and Klug, A. *J. Mol. Biol.* **228**, 637-651 (1992).

23. Omichinski, J.G., Clore, G.M., Robien, M., Sakaguchi, K., Appella, E. and Gronenborn, A.M. *Biochemistry* **31**, 3907-3917 (1992).

24. Nakaseko, Y., Neuhaus, D., Klug, A. and Rhodes, D. *J. Mol. Biol.* **228**, 619-636 (1992).

25. Elrod-Erickson, M., Rould, M.A., Nekludova, L. and Pabo, C.O. *Structure* **4**, 1171-1180 (1996).

26. Fairall, L., Schwabe, J.W.R., Chapman, L., Finch, J.T. and Rhodes, D. *Nature* **366**, 483-487 (1993).

27. Choo, Y. and Klug, A. *Nucleic Acids Res.* **21**, 3341-3346 (1993).

28. Pavletich, N.P. and Pabo, C.O. *Science* **261**, 1701-1707 (1993).

29. Kim, C.A. and Berg, J.M. *Nature Struct Biol* **3**, 940-945 (1996).

30. Dutnall, R.N., Neuhaus, D. and Rhodes, D. *Structure* **4**, 599-611 (1996).

31. Jordan, S.R. and Pabo, C.O. *Science* **242**, 893-899 (1988).

32. Schultz, S.C., Shields, G.C. and Steitz, T.A. *Science* **253**, 1001-1007 (1991).

33. Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K. *Nature* **364**, 412-420 (1993).

34. Liang, H., Mao, X., Olejniczak, E.T., Nettesheim, D.G., Liping, Y., Meadows, R.P., Thompson, C.B. and Fesik, S.W. *Nature Struct Biol* **1**, 871-876 (1994).

35. Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. *Cell* **63**, 579-590 (1990).

36. Lee, M.S., Kliewer, S.A., Provencal, J., Wright, P.E. and Evans, R.M. *Science* **260**, 1117-1121 (1993).

37. Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M. *Science* **261**, 438-446 (1993).

38. Berg, J.M. *Proc. Natl. Acad. Sci. USA* **89**, 11109-11110 (1992).

39. Fairall, L., Harrison, S.D., Travers, A.A. and Rhodes, D. *J. Mol. Biol.* **226**, 349-366 (1992).

40. Klug, A. and Schwabe, J.W.R. *FASEB J.* **9**, 597-604 (1995).

41. Desjarlais, J.R. and Berg, J.M. *Proteins Struct. Funct. Genet.* **12**, 101-104 (1992).

42. Nardelli, J., Gibson, T. and Charnay, P. *Nucleic Acids Res.* **20**, 4137-4144 (1992).

43. Desjarlais, J.R. and Berg, J.M. *Proc. Natl. Acad. Sci. USA* **90**, 2256-2260 (1993).

44. Denis, C.L. and Young, E.T. *Mol. Cell. Biol.* **3**, 360-370 (1983).

45. Simon, M., Adam, G., Rapatz, W., Spevak, W. and Ruis, H. *Mol. Cell. Biol.* **11**, 699-704 (1991).

46. Shuster, J., Yu, J., Cox, D., Chan, R.V.L., Smith, M. and Young, E. *Mol. Cell. Biol.* **6**, 1894-1902 (1986).

47. Hartshorne, T.A., Blumberg, H. and Young, E.T. *Nature* **320**, 283-287 (1986).

48. Eisen, A., Taylor, W.E., Blumberg, H. and Young, E.T. *Mol. Cell. Biol.* **8**, 4552-4556 (1988).

49. Bemis, L.T. and Denis, C.L. *Mol. Cell. Biol.* **8**, 2125-2131 (1988).

50. Thukral, S.K., Tavianini, M.A., Blumberg, H. and Young, E.T. *Mol. Cell. Biol.* **9**, 2360-2369 (1989).

51. Cook, W.J., Chase, D., Audino, D.C. and Denis, C.L. *Mol. Cell. Biol.* **14**, 629-640 (1994).

52. Blumberg, H., Eisen, A., Sledziewski, A., Bader, D. and Young, E.T. *Nature* **328**, 443-445 (1987).

53. Thukral, S.K., Eisen, A. and Young, E.T. *Mol. Cell. Biol.* **11**, 1566-1577 (1991).

54. Thukral, S.K., Morrison, M.L. and Young, E.T. *Proc. Natl. Acad. Sci. USA* **88**, 9188-9192 (1991).

55. Cook, W.J., Mosley, S.P., Audino, D.C., Mullaney, D.L., Rovelli, A., Stewart, G. and Denis, C.L. *J. Biol. Chem.* **269**, 9374-9379 (1994).

56. Thukral, S.K., Morrison, M.L. and Young, E.T. *Mol. Cell. Biol.* **12**, 2784-2792 (1992).

57. Taylor, W.E., Suruki, H.K., Lin, A.H.T., Naraghi-Arani, P., Igarashi, R.Y., Younessian, M., Katkus, P. and Vo, N.V. *Biochemistry* **34**, 3222-3230 (1995).

58. Camier, S., Kacherovsky, N. and Young, E.T. *Mol. Cell. Biol.* **12**, 5758-5767 (1992).

59. Cheng, C., Kacherovsky, N., Dombek, K.M., Camier, S., Thukral, S.K., Rhim, E. and Young, E.T. *Mol. Cell. Biol.* **14**, 3842-3852 (1994).

60. Xu, R.X., Horvath, S.J. and Klevit, R.E. *Biochemistry* **30**, 3365-3371 (1991).

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

62. Clore, G.M. and Gronenborn, A.M. *Protein Sci.* **3**, 372-390 (1994).

63. Clore, G.M. and Gronenborn, A.M. *Methods Enzymol.* **239**, 349-363 (1994).

64. Chuprina, V.P., Rullman, J.A.C., Lamerichs, R.M.J.N., van Boom, J.H., Boelens, R. and Kaptein, R. *J. Mol. Biol.* **234**, 446-462 (1993).

65. Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. *EMBO J.* **9**, 3085-3092 (1990).

66. South, T.L. and Summers, M.F. *Protein Sci.* **2**, 3-19 (1993).

67. Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. *J. Mol. Biol.* **234**, 1084-1097 (1993).

68. Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. *Nature* **376**, 791-795 (1995).

69. Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. *Cell* **81**, 705-714 (1995).

70. Morikawa, S., Ogata, K., Sekikawa, A., Sarai, A., Ishii, S., Nishimura, Y. and Nakamura, H. *J. Biomol. NMR* **6**, 294-305 (1995).

71. Werner, M.H., Clore, G.M., Fisher, C.L., Fisher, R.J., Trinh, L., Shiloach, J. and Gronenborn, A.M. *Cell* **83**, 761-771 (1995).

72. Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. *Methods Enzymol.* **185**, 60-89 (1990).

73. Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. *Methods Enzymol.* **177**, 44-73 (1989).

74. Matsuo, H., Shirakawa, M., Ohkubo, T., Yamazaki, T. and Kyogoku, Y. *J. Biomol. NMR* **1**, 191-204 (1991).

75. Spera, S. and Bax, A. *J. Am. Chem. Soc.* **113**, 5490-5492 (1991).

76. Liao, X., Clemens, K., Cavanagh, J., Tennant, L. and Wright, P.E. *J. Biomol. NMR* **4**, 433-454 (1994).

77. Richarz, R. and Wüthrich, K. *Biopolymers* **17**, 2133-2141 (1978).

78. Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. *J. Biomol. NMR* **5**, 67-81 (1995).

79. Cheng, C. and Young, E.T. *J. Mol. Biol.* **251**, 1-8 (1995).

80. Suzuki, M., Gerstein, M. and Yagi, N. *Nucleic Acids Res.* **22**, 3397-3405 (1994).

81. Nekludova, L. and Pabo, C.O. *Proc. Natl. Acad. Sci. USA* **91**, 6948-6952 (1994).

82. Shi, Y. and Berg, J.M. *Biochemistry* **35**, 3845-3848 (1996).

83. Pelton, J.G., Torchia, D.A., Meadow, N.D. and Roseman, S. *Protein Sci.* **2**, 543-558 (1993).

84. Gooley, P.R., Johnson, B.A., Marcy, A.I., Cuca, G.C., Salowe, S.P., Hagmann, W.K., Esser, C.K. and Springer, J.P. *Biochemistry* **32**, 13098-13108 (1993).

85. Párraga, G., Horvath, S., Hood, L., Young, E.T. and Klevit, R.E. *Proc. Natl. Acad. Sci. USA* **87**, 137-141 (1990).

86. Bachovchin, W.W. *Biochemistry* **25**, 7751-7759 (1986).

87. La Mar, G.N., Horrocks, W.D., Jr. and Holm, R.H., ed. *NMR of Paramagnetic Molecules Principles and Applications* (Academic Press,New York, 1973).

88. Harper, L.V., Amann, B.T., Vinson, V.K. and Berg, J.M. *J. Am. Chem. Soc.* **115**, 2577 (1993).

89. Miller, T.R., Alley, S.C., Reese, A.W., Solomon, M.S., McAllister, W.V., Mailer, C., Robinson, B.H. and Hopkins, P.B. *J. Am. Chem. Soc.* **117**, 9377-9378 (1995).

90. Dwek, R.A. *Eur. J. Biochem.* **53**, 25 (1975).

91. Pedone, P.V., Ghirlando, R., Clore, G.M., Gronenborn, A.M., Felsenfeld, G. and Omichinski, J.G. *Proc. Natl. Acad. Sci. USA* **93**, 2822-2826 (1996).

92. Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

93. Davis, R.W., Botstein, D. and Roth, J.R. *A Manual for Genetic Engineering: Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1980).

94. Caruthers, M.H., Barone, A.D., Beaucage, S.L., Dodds, D.R., Fisher, E.F., McBride, L.J., Matteucci, M., Stabinsky, Z. and Tang, J.Y. *Methods Enzymol.* **154**, 287-313 (1987).

95. Creighton, T.E. *Proteins: Structure and Molecular Properties* (W. H. Freeman and Company, New York, 1984).

96. Storm, M.C. and Dunn, M.F. *Biochemistry* **24**, 1749-1756 (1985).

97. Krizek, B.A., Merkle, D.L. and Berg, J.M. *Inorg. Chem.* **32**, 937-940 (1993).

98. Live, D.H., Davis, D.G., Agosta, W.C. and Cowburn, D. *J. Am. Chem. Soc.* **106**, 1939-1941 (1984).

99. Kay, L.E., Keifer, P. and Saarinen, T. *J. Am. Chem. Soc.* **114**, 10663-10665 (1992).

100. Sklenár, V., Piotto, M., Leppik, R. and Saudek, V. *J. Magn. Reson. Ser. A* **102**, 241-245 (1993).

101. Majumdar, A. and Zuiderweg, E.R.P. *J. Magn. Reson. Ser. B* **102**, 242-244 (1993).

102. Grzesiek, S. and Bax, A. *J. Magn. Reson.* **96**, 432-440 (1992).

103. Muhandiram, D.R. and Kay, L.E. *J. Magn. Reson. Ser. B* **103**, 203-216 (1994).

104. Bax, A., Clore, G.M. and Gronenborn, A.M. *J. Magn. Reson.* **88**, 425-431 (1990).

105. Kay, L.E., Xu, G.-Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.D. *J. Magn. Reson. Ser. B* **101**, 333-337 (1993).

Appendix A

Zinc Finger Diversity

Zinc finger diversity

Mia Schmiedeskamp and Rachel E Klevit

University of Washington, Seattle, USA

As in past years, research on a wide assortment of metal binding and nucleic acid binding domains has proceeded apace. Structural studies of individual domains and of peptide–nucleic acid complexes, combined with mutagenesis and footprinting, have continued to increase our understanding of the details of zinc finger–nucleic acid interactions.

Current Opinion in Structural Biology 1994, 4:28–35

Introduction

Diversity is the hallmark of the zinc binding and nucleic acid binding motifs known collectively as zinc fingers. The various zinc finger families differ in metal coordination strategy, secondary structure, and modularity, giving rise to distinctly different global folds (Fig. 1). The trend toward variety continued this year, with new reports on a complex of an HIV-1 nucleocapsid protein finger with single-stranded DNA [1••] and the novel structure of a zinc- and DNA-bound domain from the erythroid-specific transcription factor GATA-1 (EryF-1) [2••]. Each of these molecules exhibits nucleic acid recognition strategies that are unique among the zinc fingers. In addition, significant variation within families has been described, specifically for molecules of the Cys_2His_2 and nuclear hormone receptor classes of zinc fingers. This review will focus on a selection of the enormous amount of information on zinc fingers generated in the past year, with an emphasis on the increasing variety discovered both within and among the families.

Retroviral CCHC motifs

The two retroviral-type finger motifs found in HIV-1 nucleocapsid protein have been extensively characterized in terms of zinc binding and solution structure (reviewed by Blake and Summers [3•]). In the past year, South and Summers [1••] used ^1H NMR to explore the way in which the amino-terminal finger interacts with nucleic acid. Using synthetic single-stranded DNA as an RNA mimic they demonstrated a sequence requirement for tight binding and characterized the structure of a complex, revealing features of the protein–nucleic acid interface. NMR studies of finger-DNA titrations indicated that only oligonucleotides containing guanosine induced spectral perturbations diagnostic of complex formation. Therefore the sequence dACGCC, corresponding to a region of the HIV-1 Psi-site, was used for further structural studies of the finger–DNA complex. Three types of information were gleaned from NMR experiments on this complex. First, the structure of the DNA-bound finger did not differ from that of the free finger (see [3•] for a review). Second, the NMR data were consistent with a single-stranded, A-form conformation for the bound oligonucleotide, validating DNA as a substitute for RNA. Finally, the peptide–DNA interaction was defined by a number of intermolecular nuclear Overhauser effect (NOE) crosspeaks.

This information was used to model the interaction between the bound peptide and ideal A-form DNA. Several highly conserved hydrophobic and aromatic residues in the amino-terminal portion of the peptide form a cleft that contacts the nucleic acid. DNA binding in this hydrophobic cleft appears to be mediated by surface complementarity, a potential basic sidechain interaction with the phosphate backbone, and hydrogen bonds to guanosine. One backbone carbonyl oxygen and two backbone amide groups are in positions consistent with hydrogen bonding to the guanosine N1H and O6 atoms. Such interactions may account for the finger's observed specificity for this base. Thus single-stranded DNA recognition by the retroviral finger contrasts with double-stranded DNA recogition by other types of fingers: in the former a specific base reaches into a proteinaceous cleft to contact the polypeptide backbone, whereas in the latter protein sidechains generally extend into the grooves of DNA to contact specific bases.

GATA-1

The carboxy-terminal $CX_2CX_{17}CX_2C$ (where X represents any amino acid) motif of GATA-1 is a new addition to the gallery of metal binding and nucleic acid

Abbreviations

DBD—DNA binding domain; ER—estrogen receptor; GR—glucocorticoid receptor; NOE—nuclear Overhauser effect; RARβ—retinoic acid receptor-β; RXRα—retinoic X receptor-α; TF—transcription factor.

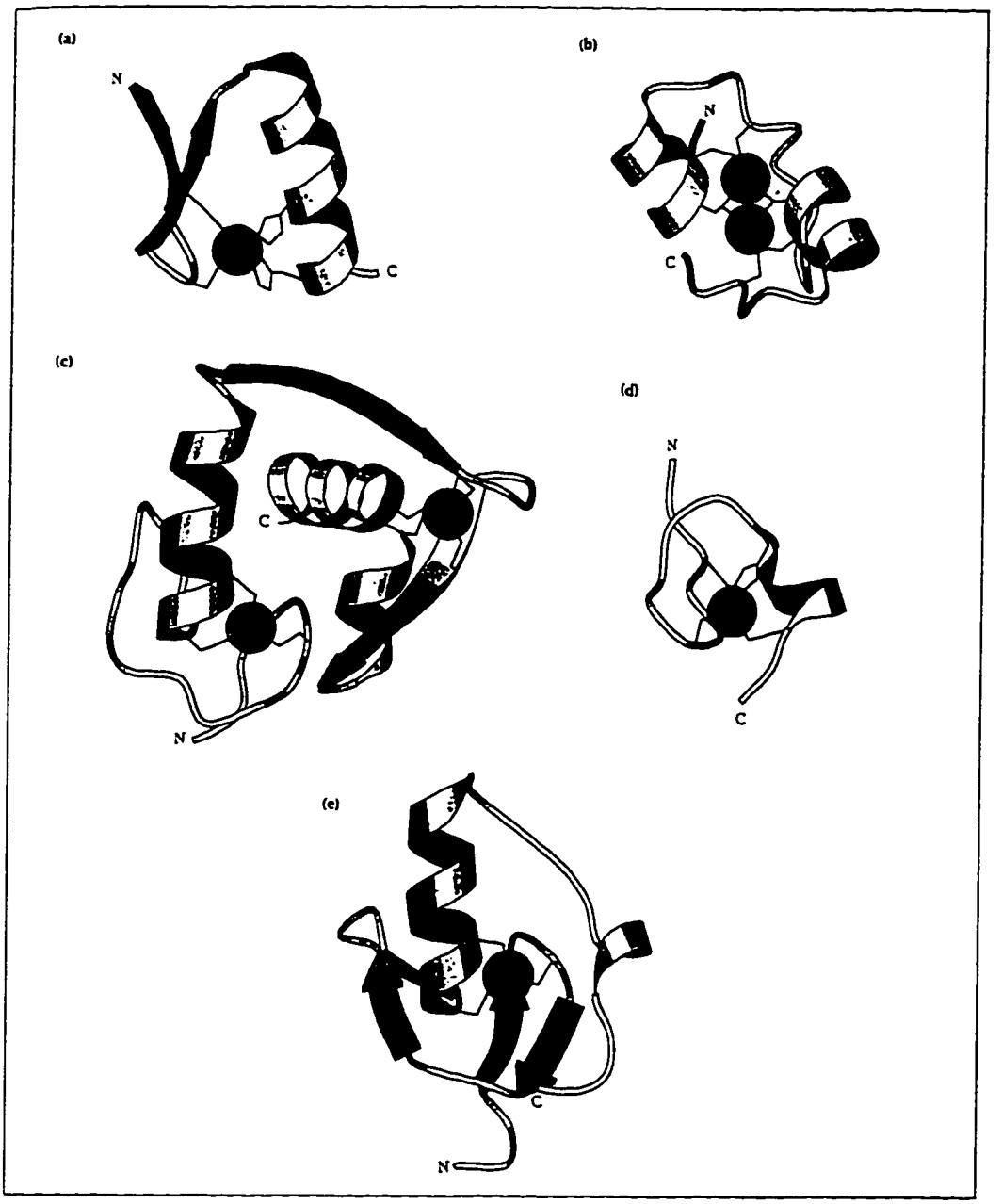


Fig. 1. Structural diversity among zinc-binding and nucleic acid-binding domains. Representative structures from the various zinc finger families are depicted schematically using the program MOLSCRIPT [43]. The sidechains that act as metal ligands are drawn. (a) Zif268 finger 2, a TFIIIA-like Cys_2His_2 zinc finger. (b) The GAL4 Zn_2Cys_6 binuclear cluster. (c) The glucocorticoid receptor DNA-binding domain. (d) The amino-terminal retroviral-type CCHC finger of HIV-1 nucleocapsid protein. (e) The carboxy-terminal metal-binding domain of GATA-1. Parts (a)–(d) were generated using coordinates stored in the Brookhaven Protein Data Bank (accession number 1zaa, 1d66, 1glu, and 1hvn, respectively). Part (e) is modified from [2••].

binding domains. Omichinski *et al.* [4•] showed that a peptide containing this motif plus a number of adjacent residues binds target DNA specifically and with high affinity in a metal-dependent manner. The metal-binding preference of GATA-1 *in vivo* is not known, but Zn^{2+}, Fe^{2+}, and Co^{2+} all support DNA-binding activity *in vitro*. Intriguingly, Fe^{2+} promoted slightly higher levels of DNA binding than did Zn^{2+}, but the significance of Fe^{2+} binding in the context of erythroid cells is unclear.

A specific complex between the zinc-bound form of the peptide and DNA was characterized by Omichinski *et al.* [2••]. The structure of the complex was calculated from many intramolecular and intermolecular restraints obtained from multidimensional heteronuclear NMR. These restraints yielded a topology in which the zinc-binding core of the peptide contacts the major groove of B-form DNA, while the additional carboxy-terminal residues extend around the DNA opposite the core, making specific contacts in the minor groove. The role of the carboxy-terminal arm is consistent with deletion studies [4•], in which shortening the arm by as few as six residues resulted in loss of specific DNA binding. The zinc-binding core consists of two short, irregular, two-stranded β-sheets connected by a loop and followed by an α-helix (see Fig. 1e). Cysteine residues located in the first β-sheet unit and at the amino terminus of the helix coordinate the metal tetrahedrally. The structure of this core is remarkably similar to that of the amino-terminal zinc-binding module of glucocorticoid receptor (GR), with a root mean square difference of only 1.4 Å between the Cα positions of the two structures [2••] (see Fig. 1c and [5•] for a description of GR). However, the helices in the two molecules approach the major groove at different angles, and while the GR helix is buttressed by the carboxy-terminal zinc-binding module, the GATA-1 helix is followed directly by a loop and then the carboxy-terminal arm which contacts the minor groove. Specific contacts to the A/T-rich GATA target DNA are primarily hydrophobic in nature, and they arise mainly from positions in the helix and carboxy-terminal arm that are highly conserved in the GATA-1 family.

The GATA-1 domain is unique among zinc-binding fingers in that it binds double-stranded DNA with high affinity using a single metal core. The nuclear hormone receptor and GAL4-like classes bind as dimers, while the Cys_2His_2 class uses tandem arrays of zinc-bound domains. The ability of the GATA-1 domain to bind with only one core apparently depends on the presence of the carboxy-terminal DNA-contacting arm. This is the first structure to exhibit use of an extended region outside the zinc core to contact DNA specifically, although such arms have been reported for other classes of DNA-binding motifs (reviewed in [6]).

Cys_2His_2 zinc fingers

Structural determinations of a large number of classical Cys_2His_2 domains established the canonical ββα fold of this type of zinc finger (reviewed in [6,7]). Last year, Nakaseko, Neuhaus and colleagues [8•,9•] reported a solution structure of a peptide containing two of the three Cys_2His_2 fingers of the yeast transcription factor SWI5 and showed that it differs in two respects from previous members of its class. Based on a lack of inter-finger NOE crosspeaks and the striking conservation of chemical shift between one- and two-finger peptides, it was concluded that these two fingers of SWI5 do not interact with each other. This behavior contrasts with that of the two fingers of the human enhancer-binding protein MBP-1, where interactions between the carboxyl terminus of one finger and the fingertip of the next were observed [10]. Of course, the fingers of SWI5 may interact when bound to DNA, as was found for the fingers of mouse immediate early protein Zif268 [11]. Structurally, the SWI5 fingers are similar to other Cys_2His_2 fingers, consisting of an irregular β-sheet and α-helix that pack together around zinc. In the case of the amino-terminal finger, however, the irregular sheet consists of three strands as opposed to the canonical two. Surprisingly, folding of finger one appears to be dependent on the presence of the additional strand. Very recently a second example of a three-stranded β-sheet in a Cys_2His_2 finger was discovered in the amino-terminal finger of the *Drosophila* transcription repressor protein Tramtrack by Fairall and colleagues [12•]. The third strand, while essential for DNA-binding activity, makes no direct contact with DNA in a cocrystal structure, suggesting it is somehow important for maintaining finger structure. It remains to be seen why these SWI5 and Tramtrack fingers differ from other known Cys_2His_2 fingers.

In the original Zif268–DNA cocrystal structure, three amino acids at positions just before and within the α-helix were found to make base-specific contacts to DNA [11]. Evidence continues to accumulate implicating these three positions as determinants of Cys_2His_2 finger specificity. Desjarlais and Berg [13•] designed several polypeptides in which these positions matched fingers of known specificity in the context of a multifinger consensus sequence. In each case, the polypeptide displayed essentially the expected sequence preference. The work of Hoffman *et al.* [14•] suggests that an amino acid sidechain at one of these critical positions in yeast transcription factor ADR1 makes a direct DNA contact, as mutations there reduce DNA binding affinity, while causing no significant structural perturbation in the finger. Kriwacki *et al.* [15] and Kuwahara *et al.* [16] presented evidence that fingers 2 and 3 of the transcription factor Sp1 bind with greater affinity than finger 1. This finding is consistent with use of the Zif268 contact positions, as that scheme predicts many more H-bonds for each of fingers 2 and 3 than for finger 1. As pointed out by Berg [17•], the Zif268 base-contact scheme also explains the variety of natural sites recognized by Sp1. Finally, the generality of the

base-contact scheme is supported by a sequence analysis done by Jacobs [18•], in which a large zinc finger database was searched for positions that are generally highly variable among all fingers, yet are conserved in closely related proteins. These positions, predicted to be involved in specific base recognition, include the three primary Zif268 contact positions. Studies such as these seem to indicate that Zif268 and at least its close cousins recognize DNA in a similar manner.

Although it is tempting to apply the simple Zif268 base-recognition scheme to the plethora of uncharacterized zinc fingers, some caution is warranted. With evidence on less closely-related fingers emerging, the situation is becoming more complicated. The five fingers of the human *GLI* oncogene product are a case in point. Pavletich and Pabo [19•] determined an X-ray structure of Co^{2+}-bound GLI fingers cocrystallized with DNA. Because physiologically relevant GLI-binding sites are not yet known, the DNA sequence used was one of several high-affinity sites selected *in vitro* from genomic DNA. Surprisingly, fingers 2 and 3 bind a section of the DNA non-specifically, making only one base contact, while finger 1 does not contact the DNA at all. However, given the variability of this region of DNA among the *in vitro* selected sites, it is difficult to draw general conclusions from these results. On the other hand, fingers 4 and 5 do bind in a sequence-specific manner to a conserved region of the site. It is their interaction with DNA that introduces new complexity into the issue of specificity. While these fingers make some use of the three primary base contact positions identified in Zif268, several of these interactions are unique in terms of the subsite position of the base contacted, the strand contacted, and the nature of the sidechain–base interaction. Furthermore, additional finger residues make specific contacts, including the first, second and fifth α-helical positions. The contacts made by residue 2 of the helix are particularly notable, as use of this position is emerging as a new theme. In the recent Tramtrack cocrystal structure, the amino-terminal finger also uses position 2, in a novel contact in which DNA bending allows a short serine sidechain to approach a thymine [12•]. In addition, position 2 contacts were observed in the original Zif268 complex [11], but they did not appear to be of primary importance. Given the GLI and Tramtrack data, the role of this position in finger–DNA recognition deserves further consideration. Taken as a whole, the recent crystal structure data reiterate several themes established by Zif268, but they also display additional variety. Unless further rules can be deduced from these and future studies, prediction of zinc finger specificity may not be as reliable as once hoped.

Of all the Cys_2His_2 zinc fingers studied, those of the polyfinger DNA- and RNA-binding protein transcription factor (TF)IIIA have remained most elusive. It is clear, however, that they represent prime examples of zinc finger versatility. Recently, the quest to understand TFIIIA has generated new and somewhat contradictory information as to its interactions with nucleic acids. Hayes and Clemens [20•] and Clemens *et al.* [21] conducted studies using nested zinc finger deletion mutants of TFIIIA in DNA footprinting experiments. The data were interpreted to support a previously proposed model [22], in which fingers 1–3 and 7–9 wrap around the ends of the TFIIIA binding site, while fingers 4–6 span the center, crossing the minor groove twice. To avoid potential end effects due to truncation, Del Rio *et al.* [23•] took the alternative approach of structurally disrupting individual fingers of TFIIIA. DNAse I footprinting data obtained with these 'broken finger' mutants are difficult to reconcile with any existing model of TFIIIA–DNA binding, and it will be interesting to see if similar data are obtained with a more delicate footprinting method such as the missing nucleoside assay [22]. Thermodynamic studies of DNA binding by the broken finger mutants indicated that all nine of the fingers make varying, yet significant, contributions to the binding energy, which contrasts with earlier studies involving truncation mutants [24]. It is becoming increasingly obvious that the energetics of TFIIIA–DNA binding are not simple, and serious consideration of factors such as entropy, DNA distortion, and interdependence of finger binding is warranted.

The TFIIIA–5S RNA interaction was studied by Clemens *et al.* [25•] and Theunissen *et al.* [26]. Their data suggest that in the case of RNA binding the central fingers are critical, contrasting with an earlier result suggesting that most of the binding energy comes from the carboxy-terminal fingers [27]. Both new reports indicated that finger 6 has a special role in 5S RNA binding, an idea which is supported by the study of Rollins *et al.* [28] which showed that feedback inhibition of 5S RNA transcription by 5S RNA failed for a broken finger 6 mutant. Finally, Choo and Klug [29•] threw another variable into the TFIIIA–nucleic acid interaction equation by showing that the linker sequence between fingers also plays a role in determining DNA-binding affinity.

Nuclear hormone receptor DNA-binding domains

A recent review by Freedman and Luisi [5•] described features of the nuclear hormone receptor class of zinc fingers, based on studies of the glucocorticoid and estrogen receptor (ER) DNA-binding domains (DBDs). Since then, further information on the issue of site-specific recognition has become available, with a report by Schwabe *et al.* [30•] on the crystal structure of an ER DBD–DNA complex. The bound ER DBD has much the same structure as the bound GR DBD [31]. This fold has also been identified in a well-defined solution structure of free GR DBD [32•], suggesting that little structural rearrangement occurs upon DNA binding. Furthermore, the overall topology of the ER DBD–DNA complex also resembles that found for the GR DBD. Mutagenesis studies previously identified a handful of amino acid positions involved in discriminating the closely related GR and ER binding sites (reviewed in [5•]). In the new crystal structure, a glutamate

residue at one of these critical positions contacts an ER-specific base, just as another such position in the GR DBD has been found to contact a GR-specific base [30]. In addition, several residues which are conserved in the GR and ER play different roles when interacting with the two different recognition sequences. Thus, while the complex of the ER DBD with DNA generally resembles that of the GR DBD, there are multiple differences at the protein–DNA interface.

Two additional members of the family have been characterized by NMR: the retinoic acid receptor-β (RARβ) DBD, solved by Knegtel *et al.* [33•], and the retinoid X receptor-α (RXRα) DBD, solved by Lee *et al.* [34•]. The solution structure of the RARβ DBD resembles the crystal structure of the GR DBD in most respects, with the central helices of the two domains superimposing with a backbone root mean square difference of only 1.1 Å. However, the domains differ somewhat in the carboxy-terminal zinc module; in the RARβ DBD the final α-helix is shorter and the carboxyl terminus is extended and packed against the core. The structure of the RXRα DBD, while not yet as well defined, is also similar to the fold seen in the GR DBD, with the notable exception of an additional helical segment at the carboxy-terminus.

RAR and RXR are particularly interesting because unlike the GR and ER, they recognize hormone response elements organized as direct repeats. RXR can bind as a homodimer or as a heterodimer with one of the other non-steroid nuclear hormone receptors (reviewed in [5•]). These interactions may involve head-to-tail contacts between receptor DNA-binding domains, and modeling suggests that the extreme carboxy-terminal end of one domain might make an asymmetric contact with the carboxy-terminal zinc module of the other [34•,35]. In fact, Towers *et al.* [35] found that mutations in these two regions interfere with the cooperativity and spacing preference of direct repeat binding by vitamin D_3 receptor DBDs. Deletion of the extreme carboxyl terminus of the RXRα DBD studied by Lee *et al.* [34•] also disrupts cooperative binding, and it is notable that this truncation falls in the extra α-helix detected by NMR. Further structural work on non-steroid hormone receptor DBDs bound to direct repeats may answer many questions that remain concerning the basis of the observed orientation and spacing preferences.

GAL4

The structure of the DNA-binding domain of yeast transcription factor GAL4 has been determined by NMR and crystallography and was reviewed last year by Berg [36]. New information includes a further NMR structural study by Shirakawa *et al.* [37] in which ^{13}C assignments were obtained using double- and triple-resonance techniques. Mau *et al.* [38] measured amide proton exchange rates to characterize the relative backbone dynamics of free and DNA-bound GAL4 DBD and found that DNA binding reduces internal structural fluctuations throughout the domain. Reece and Ptashne [39•] have performed sequence swaps with two other family members to address the question of GAL4 DNA binding specificity. They found that nineteen residues immediately following the zinc core confer the homologue-specific preference for site spacing. Future structural studies of GAL4 family members should help provide insight into how these residues accomplish this task.

Other finger-like molecules

Many new types of sequences rich in cysteine and histidine are pulled out of genomes and databases each year; we confine our remarks to those that have been characterized in terms of metal- and/or nucleic-acid binding. The so-called LIM motif, postulated to bind DNA, has been shown to bind two equivalents of zinc, and UV/vis and NMR studies using Co^{2+} and ^{113}Cd indicate tetrahedral metal coordination mainly involving cysteines [40]. The structure of a peptide bearing the zinc-binding motif of methionyl-tRNA synthetase has been determined [41•], but whether this domain plays a direct role in tRNA recognition remains unknown. Chelation of zinc has been shown to interfere with specific DNA binding by the transcription factor Myt1, which bears a $CX_5CX_{12}HX_4C$ motif [42]. It has been reported that a motif found in the RING1 family of genes can bind metal [43]. However, specific DNA binding has not been demonstrated and further metal-binding studies are necessary. Finally, there has been a report that the zinc-binding Cys_4 motif of TFIIS forms a β-sheet structure to which functionally deleterious mutations map (Qian X, Yoon H, Jeon C, Agarwal K, Weiss M, abstract W-PM-H1, 37th annual meeting of the Biophysical Society, Biophys J, 1993, 64:A237), suggesting that like various other DNA-binding motifs [6], some zinc-binding domains may use β-structure to interact with DNA. (See note added in proof.)

Conclusion

In the past few years it has become clear that each of the known zinc finger classes displays unique properties of metal binding, structure, and nucleic acid recognition. This diversity continues to expand, with reports in the past year that a retroviral finger and a GATA-1 domain behave unlike any previously characterized fingers. Variations are increasingly being discovered within the zinc finger families as well. Given the large number and widespread use of these zinc-binding domains, it seems likely that there are further structural refinements and variations we have yet to encounter. On the other hand, some themes have emerged. As Blake and Summers describe [3•], most CXXCXX motifs adopt nearly identical structures in the various classes

of fingers in which they are found. The global similarity of the GATA-1 zinc core to the amino-terminal GR zinc core is striking, especially as they have little sequence similarity. It will be interesting to follow these themes and variations as further members of each class are characterized and additional zinc finger families are discovered.

Note added in proof

Following submission of this article, Quian *et al.* reported an NMR structure of human TFIIS zinc-binding motif mentioned above [45••,46•]. This small domain binds nucleic acid and consists of a three-stranded antiparallel β-sheet. The zinc is coordinated tetrahedrally by four cysteine residues located in a loop preceding the sheet and in the turn connecting the second and third strands. This type of structure is novel among the zinc-binding and nucleic acid-binding domains and thus represents a further striking example of zinc finger diversity.

Acknowledgements

We thank those authors who contributed manuscripts prior to publication and Melissa Starovasnik and Peter Brzović for helpful comments. MS is supported by an NIH Training Grant in Molecular Biophysics (T32 GM08268). REK is supported by an AHA Established Investigatorship and NIH PO1 GM32681.

References and recommended reading

Papers of particular interest, published in the annual period of review have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• South TL, Summers MF: Zinc- and Sequence-Dependent Binding to Nucleic Acids by the N-Terminal Zinc Finger of the HIV-1 Nucleocapsid Protein: NMR Structure of the Complex with the Psi-Site Analog, dACGCC. *Protein Sci* 1993, 2:3–19.

NMR studies demonstrate the importance of guanosine residues to the interaction of a zinc-bound nucleocapsid finger with single-stranded DNA. Two-dimensional ^1H NMR is used to define the bound conformation of peptide and DNA, and the complex is modeled based on observed intermolecular NOEs.

2. •• Omichinski JG, Clore GM, Schaad O, Felsenfeld G, Trainor C, Appella E, Stahl SJ, Gronenborn AM: NMR Structure of a Specific DNA Complex of Zn-Containing DNA Binding Domain of GATA-1. *Science* 1993, 261:438–446.

Multidimensional heteronuclear NMR techniques yield a large number of intra- and intermolecular distance restraints used to calculate the structure of the complex. Features of the complex include an α-helix that lies in the major groove of the DNA and that provides primarily hydrophobic contacts to the bases, and a carboxy-terminal tail that extends into the minor groove opposite the α-helix.

3. • Blake PR, Summers MF: Probing the Unusually Similar Metal Coordination Sites of Retroviral Zinc Fingers and Iron-Sulfur Proteins by Nuclear Magnetic Resonance. *Adv Biophys Chem* 1994, in press.

This review of nucleocapsid zinc fingers highlights some striking similarities among the structures adopted by CXXC motifs in the retroviral fingers, in rubredoxin, and in other classes of zinc fingers.

4. • Omichinski JG, Trainor C, Evans T, Gronenborn AM, Clore GM, Felsenfeld G: A Small Single-"Finger" Peptide from the Erythroid Transcription Factor GATA-1 Binds Specifically to DNA as a Zinc or Iron Complex. *Proc Natl Acad Sci USA* 1993, 90:1676–1680.

Peptides containing one of the finger-like motifs from GATA-1 and an adjacent region, which includes several basic residues, are shown to bind DNA specifically and with high affinity in a metal-dependent manner.

5. • Freedman LP, Luisi BF: On the Mechanism of DNA Binding by Nuclear Hormone Receptors: A Structural and Functional Perspective. *J Cell Biochem* 1993, 51:140–150.

A recent review that addresses multiple aspects of nuclear hormone receptor binding to DNA, using the known structures as a guide.

6. Pabo CO, Sauer RT: Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu Rev Biochem* 1992, 61:1053–95.

7. Kaptein R: Zinc-Finger Structures. *Curr Opin Struct Biol* 1992, 2:109–115.

8. • Nakaseko Y, Neuhaus D, Klug A, Rhodes D: Adjacent Zinc-Finger Motifs in Multiple Zinc-Finger Peptides from SWI5 Form Structurally Independent, Flexibly Linked Domains. *J Mol Biol* 1992, 228:619–636.

An NMR study of a double Cys_2His_2 zinc finger peptide. In contrast to what was reported for the two fingers of MBP-1 [10], no evidence for inter-finger interactions was found. This paper accompanies [9•].

9. • Neuhaus D, Nakaseko Y, Schwabe JWR, Klug A: Solution Structures of Two Zinc-Finger Domains from SWI5 Obtained Using Two-Dimensional ^1H Nuclear Magnetic Resonance Spectroscopy. *J Mol Biol* 1992, 228:637–651.

The structure of a double Cys_2His_2 zinc finger peptide includes a third strand to the irregular β-sheet of the amino-terminal finger. This is the first report of a Cys_2His_2 zinc finger with secondary structure beyond the canonical fold. This paper accompanies [8•].

10. Omichinski JG, Clore GM, Robien M, Sakaguchi K, Appella E, Gronenborn AM: High-Resolution Solution Structure of the Double Cys_2-His_2 Zinc Finger from the Human Enhancer Binding Protein MBP-1. *Biochemistry* 1992, 31:3907–3917.

11. Pavletich NP, Pabo CO: Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science* 1991, 252:809–817.

12. • Fairall L, Schwabe JWR, Chapman L, Finch JT, Rhodes D: The Crystal structure of a Two Zinc-finger Peptide reveals an Extension to the Rules for Zinc-Finger DNA Recognition. *Nature* 1993, 366:483–487.

This cocrystal structure suggests an additional rule for Cys_2His_2 finger specificity beyond the simple Zif268 scheme. It also provides the second example of a three-stranded β-sheet in a Cys_2His_2 finger.

13. • Desjarlais JR, Berg JM: Use of a Zinc-Finger Consensus Sequence Framework and Specificity Rules to Design Specific DNA Binding Proteins. *Proc Natl Acad Sci USA* 1993, 90:2256–2260.

Several polypeptides containing three Cys_2His_2 zinc fingers were designed, combining a consensus sequence for the structural framework with the putative base-contacting residues of fingers of known specificity. These proteins displayed essentially the predicted sequence preferences.

14. • Hoffman RC, Horvath SJ, Klevit RE: Structures of DNA-Binding Mutant Zinc Finger Domains: Implications for DNA Binding. *Protein Sci* 1993, 2:951–965.

The refined NMR structures of a wild-type and two mutant zinc fingers are compared and found to be statistically indistinguishable. Defects in DNA binding caused by the mutations therefore appear to be due to loss of a direct contact provided by the wild-type sidechain.

15. Krwacki RW, Schultz SC, Steitz TA, Caradonna JP: Sequence-Specific Recognition of DNA by Zinc-Finger Pep-

tides Derived from the Transcription Factor Sp1. *Proc Natl Acad Sci USA* 1992, 89:9759–9763.

16. Kuwahara J, Yonezawa A, Futamura M, Sugiura Y: Binding of Transcription Factor Sp1 to GC Box DNA Revealed by Footprinting Analysis: Different Contact of Three Zinc Fingers and Sequence Recognition Mode. *Biochemistry* 1993, 32:5994–6001.

17. • Berg JM: Sp1 and the Subfamily of Zinc Finger Proteins with Guanine-Rich Binding Sites. *Proc Natl Acad Sci USA* 1992, 89:11109–11110.
A brief commentary that ties together results concerning the DNA binding site preferences of Cys_2His_2 finger proteins Sp1, Krox20, Zif268, ADR1 and various altered specificity mutants thereof.

18. • Jacobs GH: Determination of the Base Recognition Positions of Zinc Fingers from Sequence Analysis. *EMBO J* 1992, 11:4507–4517.
Sequence positions that are particularly variable in Cys_2His_2 zinc fingers, but that are highly conserved in fingers from homologous proteins are predicted to be base contact positions. Three such positions coincide with the base contacts found in the Zif268–DNA cocrystal structure.

19. • Pavletich NP, Pabo CO: Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers. *Science* 1993, 261:1701–1707.
This structure demonstrates that not all Cys_2His_2 fingers interact with DNA in the same manner. While some themes first seen for Zif268 are reiterated, several aspects of GLI–DNA recognition are novel.

20. • Hayes JJ, Clemens KR: Locations of Contacts between Individual Zinc Fingers of *Xenopus laevis* Transcription Factor IIIA and the Internal Control Region of a 5S RNA Gene. *Biochemistry* 1992, 31:11600–11605.
Hydroxyl radical footprinting studies of nested zinc finger deletion mutants of TFIIIA are used to support a model in which fingers 1–3 and 7–9 bind the ends of the TFIIIA site, while fingers 4–6 span the center, crossing the minor groove twice.

21. Clemens KR, Liao X, Wolf V, Wright PE, Gottesfeld JM: Definition of the Binding Sites of Individual Zinc Fingers in the Transcription Factor IIIA-5S RNA Gene Complex. *Proc Natl Acad Sci USA* 1992, 89:10822–10826.

22. Hayes JJ, Tullius TD: Structure of the TFIIIA-5S DNA Complex. *J Mol Biol* 1992, 227:407–417.

23. • Del Rio S, Menezes SR, Setzer DR: The Function of Individual Zinc Fingers in Sequence-Specific DNA Recognition by Transcription Factor IIIA. *J Mol Biol* 1993, 233:567–579.
Mutants of TFIIIA disrupted in individual zinc fingers are used to demonstrate the structural independence of the fingers. Binding of the mutants to the 5S RNA gene is evaluated in terms of energetics and DNAse I protection.

24. Liao X, Clemens KR, Tennant L, Wright PE, Gottesfeld JM: Specific Interaction of the First Three Zinc Fingers of TFIIIA with the Internal Control Region of the Xenopus 5S RNA Gene. *J Mol Biol* 1992, 223:857–871.

25. • Clemens KR, Wolf V, McBryant SJ, Zhang P, Liao X, Wright PE, Gottesfeld JM: Molecular Basis for Specific Recognition of Both RNA and DNA by a Zinc Finger Protein. *Science* 1993, 260:530–533.
This report details the affinity for 5S RNA of a series of nested zinc finger deletion mutants of TFIIIA. The authors report that fingers 4–7 are critical for 5S RNA binding, and a model for the TFIIIA–RNA interaction is proposed.

26. Theunissen O, Rudt F, Guddat U, Mentzel H, Pieler T: RNA and DNA Binding Zinc Fingers in Xenopus TFIIIA. *Cell* 1992, 71:679–690.

27. Darby MK, Joho KE: Differential Binding of Zinc Fingers from Xenopus TFIIIA and p43 to 5S RNA and the 5S RNA Gene. *Mol Cell Biol* 1992, 12:3155–3164.

28. Rollins MB, Del Rio S, Galey AL, Setzer DR, Andrews MT: Role of TFIIIA Zinc Fingers in Vivo: Analysis of Single-Finger Function in Developing *Xenopus* Embryos. *Mol Cell Biol* 1993, 13:4776–4783.

29. • Choo Y, Klug A: A Role in DNA Binding for the Linker Sequences of the First Three Zinc Fingers of TFIIIA. *Nucleic Acids Res* 1993, 21:3341–3346.
Substitution of TFIIIA linkers with those from p43, a zinc finger protein that binds 5S RNA but not DNA, caused dramatic decreases in DNA-binding affinity. Mutation of several individual linker residues also had adverse effects.

30. • Schwabe JWR, Chapman L, Finch JT, Rhodes D: The Crystal Structure of the Estrogen Receptor DNA-Binding Domain Bound to DNA: How Receptors Discriminate Between Their Response Elements. *Cell* 1993, 75: 567–579.
The details of the protein–nucleic acid interface are described, including an ER-specific contact that was anticipated by mutagenesis. Additional contacts prove to be somewhat more complicated.

31. Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB: Crystallographic Analysis of the Interaction of the Glucocorticoid Receptor with DNA. *Nature* 1993, 352:497–505.

32. • Baumann H, Paulsen K, Kovács H, Berglund H, Wright APH, Gustafsson J-Å, Härd T: Refined Solution Structure of the Glucocorticoid Receptor DNA-Binding Domain. *Biochemistry*, 1994, in press.
This study shows that the secondary structure of free GR DBD is very similar to that of DNA-bound GR DBD. Based on this well-defined structure, it now appears that little structural rearrangement occurs upon DNA binding, except possibly at the dimerization interface.

33. • Knegtel RMA, Katahira M, Schilthuis JG, Bonvin AMJJ, Boelens R, Eib D, van der Saag PT, Kaptein R: The Solution Structure of the Human Retinoic Acid Receptor-β DNA-Binding Domain. *J Biomolec NMR* 1993, 3:1–17.
The structure of the RARβ DBD is found to be very similar to that of the GR DBD, both in secondary structure and overall tertiary fold. An exception is the extreme carboxy-terminal region, which is extended and well-defined in the RARβ DBD.

34. • Lee MS, Kliewer SA, Provencal J, Wright PE, Evans RM: Structure of the Retinoid X Receptor α DNA Binding Domain: A Helix Required for Homodimeric DNA Binding. *Science* 1993, 260:1117–1121.
The solution structure of the RXRα DBD is similar to that of the GR DBD, with the exception of an additional carboxy-terminal α-helix. Mutagenesis indicates that this carboxy-terminal region contributes to the affinity and cooperativity of RXRα binding to hormone response elements organized as direct repeats.

35. Towers TL, Luisi BF, Asianov A, Freedman LP: DNA Target Selectivity by the Vitamin D_3 Receptor: Mechanism of Dimer Binding to an Asymmetric Repeat Element. *Proc Natl Acad Sci* USA 1993, 90:6310–6314.

36. Berg JM: Zinc-Finger Proteins. *Curr Opin Struct Biol* 1993, 3:11–16.

37. Shirakawa M, Fairbrother WJ, Serikawa Y, Ohkubo T, Kyogoku Y, Wright PE: Assignment of ^1H, ^{15}N and ^{13}C Resonances, Identification of Elements of Secondary Structure and Determination of the Global Fold of the DNA-Binding Domain of GAL4. *Biochemistry* 1993, 32:2144–2153.

38. Mau T, Baleja JD, Wagner G: Effects of DNA Binding and Metal Substitution on the Dynamics of the GAL4 DNA-Binding Domain as Studied by Amide Proton Exchange. *Protein Sci* 1992, 1:1403–1412.

39. • Reece RJ, Ptashne M: Determinants of Binding-Site Specificity Among Yeast C_6 Zinc Cluster Proteins. *Science* 1993, 261:909–911.
In sequence swaps among GAL4, PUT3 and PPR1 the zinc clusters appear to be interchangeable for recognition of CGG elements, while the subsequent nineteen residues appear to dictate spacing requirements.

40. MICHELSEN JW, SCHMEICHEL KL, BECKERLE MC, WINGE DR: The LIM Motif Defines a Specific Zinc-Binding Protein Domain. *Proc Natl Acad Sci USA* 1993, 90:4404–4408.

41. • FOURMY D, DARDEL F, BLANQUET S: Methionyl-tRNA Synthetase Zinc Binding Domain: Three-Dimensional Structure and Homology with Rubredoxin and *gag* Retroviral Proteins. *J Mol Biol* 1993, 231:1078–1089.
In the wake of this paper, the crystal structure of this synthetase requires some reconsideration.

42. KIM JG, HUDSON LD: Novel Member of the Zinc Finger Superfamily: A C_2HC Finger That Recognizes a Glia-Specific Gene. *Mol Cell Biol* 1992, 12:5632–5639.

43. LOVERING R, HANSON IM, BORDEN KLB, MARTIN S, O'REILLY NJ, EVAN GI, RAHMAN D, PAPPIN DJC, TROWSDALE J, FREEMONT PS: Identification and Preliminary Characterization of a Protein Motif Related to the Zinc Finger. *Proc Natl Acad Sci USA* 1993, 90:2112–2116.

44. KRAULIS PJ: MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J Appl Crystallogr* 1991, 24:946–950.

45. •• QUIAN X, GOZAN SN, YOON H, JEON C, AGARWAL K, WEISS MA: Novel Zinc Finger Motif in the Basal Transcription Machinery: Three-Dimensional NMR Studies of the Nucleic Acid Binding Domain of Transcription Elongation Factor TFIIS *Biochemistry* 1993, 32:9944–9959.
A novel zinc binding and nucleic acid binding fold is reported, consisting primarily of a three-stranded antiparallel β-sheet. See also [46•].

46. • QUIAN X, JEON C, YOON H, AGARWAL K, WEISS MA: Structure of a New Nucleic-Acid-Binding motif in Eukaryotic Transcriptional Elongation Factor TFIIS. *Nature* 1993, 365: 277–279.
See [45••].

M Schmiedeskamp and RE Klevit, Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA.

VITA

Mia Ruth Schmiedeskamp was born in Ann Arbor, Michigan on November 21, 1966 to Jay and Jane Schmiedeskamp. She graduated from Ann Arbor Pioneer High School in 1984. Mia attended the University of Michigan as an Alvin and Arvella Bentley Scholar, graduating *summa cum laude* with a B. S. in Chemistry in December of 1988. She began work at the University of Washington Department of Biochemistry in 1989, as an NSF pre-doctoral fellow.