Mutations in the zinc fingers of ADR1 that change the specificity of DNA binding and transactivation.

S K Thukral, M L Morrison and E T Young
Mutations in the Zinc Fingers of ADR1 That Change the Specificity of DNA Binding and Transactivation

SUSHIL K. THUKRAL,† MICHAEL L. MORRISON, AND ELTON T. YOUNG*
Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received 26 December 1991/Accepted 19 March 1992

ADR1 is a yeast transcription factor that contains two zinc fingers of the Cys-2-His-2 (C2H2) class. Mutations that change the specificity of DNA binding of ADR1 to its target site, upstream activation sequence 1 (UAS1), have been identified at three positions in the first zinc finger. Mutations Arg-115 to Gln, His-118 to Thr, and Arg-121 to Asn led to new specificities of DNA binding at adjacent positions 10, 9, and 8 (3'-GAG-5') in UAS1. Arg-115 is at the finger tip, and His-118 and Arg-121 are at positions 3 and 6, respectively, in the alpha helix of finger 1. One double mutant displayed the binding specificity expected from the properties of its constituent new-specificity mutations. Mutations in the second finger that allowed its binding site to be identified through loss-of-contact phenotypes were made. These mutations imply a tail-to-tail orientation of the two ADR1 monomers on their adjacent binding sites. Finger 1 is aligned on UAS1 in an amino-to-carboxy-terminal orientation along the guanine-rich strand in a 3'-to-5' direction. One of the ADR1 mutants was functional in vivo with both its cognate binding site and wild-type UAS1, but the other two mutants were defective in transactivation despite their ability to bind with high affinity to their cognate binding sites.

Zinc fingers constitute one of the major classes of DNA-binding motifs in eukaryotes (see references 2 and 15 for reviews). The structural elements of a zinc finger of the type discovered in TFIIIA (Cys-2-His-2 [C2H2]; Fig. 1) (5, 18) were predicted on the basis of the structures of known proteins (1, 6, 11). The crystal structure of the three fingers of Zif268 complexed with DNA directly demonstrated the basic mechanism of interaction of C2H2 zinc fingers with DNA (23). Two amino acids, one just amino terminal to and the other within the alpha helix, interact with two bases in the major groove. DNA-binding studies of mutants in which the binding specificity of one zinc finger protein (Krox20) was converted to that of another (SP1) indicated that the residue just preceding the alpha helix was important for determining the specificity of interaction (19). The phenotypes of mutants within the zinc fingers of ADR1 were also consistent with three positions being important for DNA binding but indicated that the relative importance varied greatly both within and between fingers (31). Since the specificity of interaction between proteins and DNA is essential for transcriptional fidelity, it is important to identify all of the amino acids which contribute to this property both in vitro and in vivo. The specificity usually resides in the interaction between amino acid side chains and functional groups of the DNA bases, which lie mainly in the major groove (9, 20, 27). Fingers of the C2H2 type appear to conform to this expectation (23), but only one position has been directly demonstrated to be a specificity determinant (19).

ADR1 is a transcription factor in the yeast Saccharomyces cerevisiae. It is required for derepression of ADH2, the structural gene for the glucose-repressible alcohol dehydrogenase isozyme, and for derepression of several peroxisomal enzymes (26). Its ability to activate ADH2 expression, but not its ability to bind DNA, is regulated posttranslationally (4, 8, 28). Phosphorylation by protein kinase A influences the activation process in an unknown way (7, 28).

Two monomers of ADR1 bind to a perfect palindrome, upstream activation sequence 1 (UAS1) (10, 29), the ADR1-dependent upstream activation sequence in the ADH2 promoter (25, 32), using two C2H2 zinc fingers and an adjacent amino-terminal region (3, 29, 30). Alanine scanning mutagenesis of the two zinc fingers of ADR1 showed that R115, H118, and R121 in the first finger and R143 and D145 in the second were essential for high-affinity DNA binding and for transactivation (31). These residue positions are the same as those shown to be in close contact with DNA in the Zif268-DNA crystal structure (23).

The structures of the two zinc fingers of ADR1 have been determined by nuclear magnetic resonance spectroscopy (14, 22). The side chains of amino acids implicated by mutagenesis to be involved in DNA binding are on the finger tips and the faces of alpha helices in solvent-exposed positions, as are equivalent residues in the Zif268-DNA complex. A complete structure-function analysis of the zinc fingers of ADR1 requires the analysis of functionally defective fingers (21) as well as that of fingers which have altered their specificity of DNA binding. To determine the origins of specificity in DNA binding by ADR1, we have identified mutations that lead to novel specificities of DNA binding in vitro and in vivo. These mutations are at positions −1, 3, and 6 of the alpha helix of finger 1, consistent with the crystal structure of the Zif268-DNA complex (23).

MATERIALS AND METHODS

Yeast and bacterial strains. The yeast strain used for expressing the ADR1 mutants was S. cerevisiae SHH35 MATa adh1Δ1 adrlΔ1::LEU2 adh3 trp1 ura3 leu2. The bacterial strain used to express 17-229ADR1 was Escherichia coli MC1061 (16). Cells were grown and extracts were prepared as described previously (31).

DNA-binding assays and enzyme assays. The assays were performed as described previously (31).

Oligonucleotides. The oligonucleotides used to determine
VOL.

oligonucleotide, with from chain. The numbers polypeptide fingers are AAGTTGGAGATGCCCGGTGTTCCGGCAGAGGAGAG the binding site preferences for ADR1 were synthesized from an oligonucleotide containing 53 nucleotides and an 18-mer primer complementary to the 3' end of the longer oligonucleotide, with a 4-bp overhang for cloning purposes (see below). Their sequences are 5'-GTCTCCCAACTAT AAGTTGGAGATGCCCGGTGTTCCGGCAGAGGAGAG GTAC-3' and 5'-CTCTCCTCGCCGAACACTAA-3' (US1 is underlined). The binding site thus generated also contains US2 (32). After extension using a radioactive primer, the full-length probe was purified by electrophoresis through a 12% polyacrylamide gel and eluted either by electrophoresis or by soaking overnight in buffer. The mutant oligonucleotides contained symmetric changes at each of the 11 positions in each half of US1 to the other three possible bases. They are referred to by the wild-type (wt) base and its position in the 3' binding site, followed by the new base. The center of dyad symmetry is between positions 1' and 1. For example, the mutant oligonucleotide in which the G at position 10 is replaced by an A and the C at position 10' is replaced by a T would be called G10A. Three mutant oligonucleotides that contain four changes from the wt sequence were made: NS1:G10A, A9C; NS2:G10A, G8A; and NS3:A9C, G8A.

Mutagenesis. Cassette mutagenesis of residue positions corresponding to R115, H118, R121, and R143 was performed as follows. Four oligonucleotides containing mixtures of four bases corresponding to codon positions 115, 118, and 121 were synthesized. These synthetic oligonucleotides and the single-stranded uracil-containing template were then used to synthesize the second strand. Thirty plaques from the transformations from each of the four reactions were picked. Mutations were screened by sequencing single-stranded DNA obtained from these plaques. Mutations were cloned into the E. coli expression vector pCQ17-229NF as described earlier (31). Mutations at position 143 were made similarly but with a restricted set of changes.

Double ADR1 mutants. Uracil-containing M13 template DNA was prepared from M13 clones containing single mutations R115Q and R121N. The template containing mutation R115Q was used to perform two sets of mutagenesis using oligonucleotides containing mutations to Thr at position 118 and to Asn at position 121, respectively, resulting in two double mutants of ADR1: R115Q-H118T and R115Q-R121N. Template containing R121N was used with an oligonucleotide containing mutated codon Thr at position 118 to double mutant H118T-R121N. Mutations were confirmed by sequencing and cloned as described above.

Plasmids. ADR1 mutations R115Q, H118T, and R121N were cloned into yeast vector YEpADR1 1-16,17-642Z by switching the BamHI-SphI ADR1 fragment from M13 clones as described earlier (31). These mutations were also cloned into a 2 µm TRP1 yeast episomal vector containing full-length ADR1 under the control of the ADH1 promoter, pKD39 (8a, 28). The ADR1 fragments containing the mutations were switched from YEpADR1 1-16,17-642Z as SalI-SphI fragments into pKD39. These plasmids are referred to as YEp1ADR1-115Q, YEp1ADR1-1183(R115Q), and YEp1ADR1-1213(R115Q, H118T, and R121N).

Cloning oligonucleotides into reporter plasmid pLG667Z. For cloning of oligonucleotides containing mutations G10A, A9C, and G8A, pLG669Z (12) was modified. pLG669Z was cut with SmaI, and Kpn1 linkers (10-mer; New England Biolabs) were inserted. The resultant vector is named pLG(K)669Z. The top strand of the above-mentioned oligonucleotides was chemically synthesized. Annealing of two strands leaves a Kpn1 overhang on the lower strand, and the other end of the oligonucleotide is blunt. pLG(K)669Z was prepared for cloning of these oligonucleotides by cutting with XhoI, filling in with the Klenow fragment of DNA polymerase I, and then digesting with Kpn1. Cloning was confirmed by double-stranded sequencing. These plasmids were named pLG(K)669Z(G10A), pLG(K)669Z(A9C), and pLG(K)669Z(G8A). A wt oligonucleotide was cloned into the same vector by annealing two single-stranded oligonucleotides that created blunt ends. This oligonucleotide is UAS1-WT (31). It was cloned into the same vector which had been cut with both SmaI and XhoI and treated with the Klenow fragment of polymerase I. DNA sequencing was used to identify clones which had the oligonucleotide in the normal (as in the ADH2 promoter) or in the reverse orientation. These plasmids are named pLG(K)669Z(UAS1-WT) and pLG(K)669Z(UAS1-WTR), respectively.

RESULTS

Identification of amino acid-base pair contacts. Mutants altered at positions 115, 118, 121, 143, and 145 were used to identify base pairs that were likely contact sites in UAS1 for the two fingers of ADR1 (Fig. 1). Alanine mutants were used first, since a loss-of-contact phenotype can be revealed by testing their binding activity with mutant binding sites (9). Each base in the palindromic binding site, UAS1, was changed to the three others to generate a collection of 33 mutant binding sites, each containing two symmetric changes. The ability of wt ADR1 to bind the 33 mutant probes was compared with its ability to bind the wt probe. Four positions, positions 7 to 10, were critical for binding to wt ADR1; the others could be altered without abolishing binding by ADR1 (Fig. 2 and unpublished data).

Next, Ala mutants were tested with the probes containing substitutions at important positions. Mutants R115A and H118A lost the ability to distinguish base changes at positions 10 and 9, respectively, as opposed to wt ADR1, while retaining the ability to distinguish alterations at other posi-
FIG. 2. Loss-of-contact phenotype of ADR1 mutant proteins. Gel shift assays were performed with 0.02 mg of E. coli extract protein and the probes indicated. The double-stranded probes were made by annealing a 53-mer and an 18-mer complementary to the 3' end of the longer oligonucleotide. Their sequences are 5'-CTCTCC TCTGCCTCAACA-3' and 5'-GTCTCCACTTAAGTGGGA GATGCCGGTTGTCGGCAGAGGAGGTAC-3' (UAS1 is underlined). Each mutant 53-mer had symmetric mutations at two positions, 1 and 1' (T and A adjacent to the center of symmetry), with respect to the center of dyad symmetry indicated by the period in UAS1. After extension using a radioactive primer, the full-length probe was gel purified. Western immunoblotting performed on the extracts demonstrated that they contained the same amount of ADR1 protein. F, CII, and CIII correspond to free probe and to protein-DNA complexes containing one and two molecules of ADR1, respectively (29).

Mutations in UAS1. R115A also showed a less significant alteration from wt discrimination at bp 9.

Mutant D145A distinguished mutant sites differently from wt at positions 6, 7, and 8. The effect was most clear at position 8. D145A bound UAS1 containing an AT base pair nearly as well as it bound the wt site which contains a GC base pair. At bp 7, weak binding to all mutant sites was discernible. Mutants R121A and R143A bound too poorly to be used in this approach.

Cassette mutagenesis was performed at each of the three positions in finger 1 that had been identified as likely to make important DNA contacts, positions 115, 118, and 121, and at position 143 in finger 2 (31). The results of the cassette mutagenesis at positions 115, 118, and 121 are shown in Fig. 3. Only at position 115 was a completely functional replacement obtained. Lys at position 115 bound to the wt probe as well as did Arg. Lys was not a functional substitute for R121 or R143. At each position, substitutions that were either partially functional or nonfunctional were identified.

Mutants identified by cassette mutagenesis as retaining some binding activity with wt UAS1 were tested for binding to mutant UAS1 oligonucleotides. Mutants R115Q, H118T, and R121N provided clear indications of new specificity phenotypes (Fig. 4A). Mutant R115Q preferentially recognized an altered UAS1 containing an AT base pair at position 10 rather than the wt GC base pair. This mutant protein also displayed a new specificity at position 9. It recognized a site with either T, A, or G at this position, whereas wt ADR1 recognized only A or G at this position. The results from both the loss-of-contact and new-specificity approaches are consistent with the interpretation that the side chain of R115 contacts bp 10, and perhaps bp 9, in UAS1 and that Arg and Lys are functionally equivalent at this position.

Mutant H118T showed a change in the ability to discriminate the base pair at position 9. Binding to the mutant site A9C was significantly higher than was binding to the wt site. It also bound the mutant site A9G about as well as it bound the wt site. Binding specificities of wt and H118 mutants were identical at all other positions in UAS1.

Mutant R121N also displayed a new specificity phenotype. It bound the variant site G6A much better than it bound wt UAS1. The R121N mutant also displayed detectable binding activity to UAS1 mutants G8T, G8C, G7C, and G7A. wt ADR1 did not bind to any variants of position 7 or 8, wt and R121N had identical preferences at other positions.

None of the three R143 mutants resulted in a new specificity phenotype (Fig. 4B). However, a loss-of-contact phenotype of R143K with respect to position 7 suggests that the side chain of R143 contacts position G7 in UAS1.

The results for the loss-of-contact and new-specificity mutants are summarized in Table 1.

Interdependence of residues within the recognition helix. To test the functional independence of positions 115, 118, and 121 in finger 1, three double mutants were made: R115Q-H118T, R115Q-R121N, and H118T-R121N. Their abilities to bind to wt, three single-mutant (G10A, A9C, and G8A), and three double-mutant sites were tested. The double-mutant
sites contained combinations of the single mutations which alone gave specific binding to the ADR1 mutant proteins. wt and single-mutant proteins behaved as expected from the results described above and in addition did not bind or bound very poorly to the double-mutant probes (Fig. 5A to C and G).

Double mutant R115Q-H118T preferentially bound to the double-mutant site, G10A-A9C, compared with wt or single-mutant sites (Fig. 5D). Double mutant R115Q-R121N showed no binding to any sites (Fig. 5E). Double mutant H118T-R121N had low but detectable binding activity to wt and single-mutant sites A9C and G8A and also to the double-mutant site A9C-G8A (Fig. 5F). Thus, only the double mutant R115Q-H118T had binding properties that were predictable from the properties of its constituent mutations.

A mutant zinc finger protein activates transcription from a site not recognized by the wt protein. New-specificity ADR1 mutants R115Q, H118T, and R121N were introduced into yeast cells containing a lacZ reporter gene with a single UAS1 element in its promoter (12, 32). Each mutant was tested with both wt UAS1 and the mutant UAS1 to which it bound preferentially in vitro.

Expression of the reporter gene in the doubly transformed strains is shown in Table 2. As expected, expression was ADR1 dependent. wt ADR1 increased the activity more than 400- and about 30-fold in the presence and absence of glucose, respectively, when the reporter gene was driven by wt UAS1. The wt protein was unable to activate expression of any of the reporter genes containing UAS1 mutations, as expected, since wt protein was unable to bind to these sequences.

The H118T mutant protein activated expression of reporter genes containing either wt UAS1 or the A9C mutant. It stimulated expression from the reporter gene containing wt UAS1 only during derepressed growth conditions when
ADRI is fully activated. It activated the reporter gene containing its cognate UAS1 mutation as well as wt ADRI activated a reporter gene containing a wt UAS1 element. The activity of the H118T mutant in vivo thus parallels the in vitro DNA-binding activity: it bound significantly less well to wt UAS1 than it did to the A9C mutant UAS1.

Surprisingly, neither of the other mutant proteins, R115Q and R121N, was able to activate its cognate reporter gene efficiently. R115Q activated expression about fivefold during both repressed and derepressed growth conditions. R121N was unable to activate expression above the background level. These results were unexpected, since both mutant proteins bound with high affinity to their cognate UAS1 elements and Western immunoblot analysis showed that the proteins were present at wt levels in yeast extracts (Fig. 4A and unpublished data).

The ability of the wt and mutant proteins to function in vivo was also assessed by measuring ADH2 expression from the endogenous ADH2 gene. Table 3 shows the ADHII enzyme activities in extracts prepared from the strains containing wt and mutant ADRI genes. ADH2 expression was fully constitutive in the strain containing wt ADRI coding sequence as a result of high-level expression. ADH2 expression was weakly constitutive and efficiently derepressed in cells containing the ADRI gene bearing H118T. This new-specificity mutant protein bound to wt UAS1 but to a lesser extent than it bound to its variant binding site containing the A9C mutation (Fig. 3A). The lower level of ADHII activity presumably reflects a reduced level of binding of the H118T mutant protein to the wt UAS1 site in the ADH2 promoter. The mutant R115Q protein allowed very weak constitutive expression of ADH2 and repression to about 7% of the wt level. Mutant R121N was completely inactive. The activation of ADH2 expression by both H118T and R115Q mutant proteins indicates that they can function in the context of an intact ADH2 promoter containing wt UAS1.

**DISCUSSION**

The amino acid-base pair interactions between ADRI and UAS1 shown in Fig. 6 were inferred from change-of-specificity or loss-of-contact mutations in the zinc fingers of ADRI. The conclusions based on the binding behaviors of these mutants confirm and extend the interactions observed in the crystal structure of the Zif268-DNA complex (23). Interaction between the original or substituted amino acid and its respective contact site in UAS1 seems chemically feasible. Several of the interactions are the same as those seen directly in the crystal structure of the Zif268-DNA...
The guanidinium group of R115 could donate H bonds to O6 and N7 of guanine 10. Since Lys can functionally replace Arg at this position in ADR1, positive charge may be more important than the guanidinium group at this position. In order for Lys to H bond with guanine 10, the finger structure would have to move closer to the DNA since the side chain of Lys is shorter than that of Arg. Moreover, there would be a loss of one H bond with the replacement of Arg by Lys, since steriochemically it would be unfavorable for Lys to form two H bonds with O6 and N7 of guanine 10.

The interaction of R121 and R143 with DNA appears to be clearly different from that of R115. Mutation to Ala at either of these positions had a much more profound effect on DNA binding than did mutation to position 115 (31). In addition, neither of these positions could Lys functionally replace Arg, as it did at position 115. This finding can be partially explained for position 143 by invoking the structure that was seen in the Zif268-DNA complex (23). Arg at the analogous position in the three Zif268 fingers interacted with DNA by forming two H bonds between guanidinium group and N7 and O6 of a guanine residue. The guanidinium group also appeared to interact with an Asp carboxylate ion two residues away. The primary amino group of Lys would not be able to perform the same two functions. In ADR1, Arg and Asp in equivalent positions are important for DNA binding, with Arg playing an essential role. Glu can functionally replace Asp at position 145 (31).

The major importance of R121 in finger 1 cannot be explained by analogy to R143, since there is no nearby Asp residue for it to interact with. The failure of Lys, which can functionally replace R115, to functionally replace Arg at position 121 suggests that the two Arg groups interact differently with DNA although each appears to interact with a guanine in UAS1. Possibly, R121 interacts with more than one base pair. The change-of-specificity mutant R121N displayed a different pattern of base discrimination at position 7 as well as at position 8, although the difference was less pronounced at position 7. Since bp 7 is in the subsite expected to interact with finger 2, this observation raises the possibility that each finger does not interact exclusively with a subsite of 3 bp. Alternatively, R115 and R121 may differ in

<table>
<thead>
<tr>
<th>UAS1 element</th>
<th>adrl</th>
<th>wt</th>
<th>R115Q</th>
<th>H118T</th>
<th>R121N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>adrl</td>
<td>wt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>dr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>1</td>
<td>16</td>
<td>430</td>
<td>460</td>
<td>1</td>
</tr>
<tr>
<td>G10A</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>20</td>
<td>ND</td>
</tr>
<tr>
<td>A9C</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>ND</td>
<td>1,031</td>
</tr>
<tr>
<td>G8A</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

* An ADR1 expression plasmid (TRP1) and a reporter plasmid (URA3) were introduced into strain SSH35. ADR1 wild-type and mutant proteins were expressed from a multicopy plasmid containing TRP1 and the 2μm origin of replication (8a, 28). The intact ADR1 gene was fused to the ADH1 promoter. Cultures of strain SSH35 were grown in synthetic medium containing 2% glucose and lacking uracil and tryptophan for plasmid selection, centrifuged, and resuspended in fresh synthetic medium containing glucose (repressing conditions [r]) or in complete medium containing ethanol, glycerol, and lactate (2% each) and 0.1% glucose (derepressing conditions [dr]). After further growth to an Amax of about 1 to 2, the cultures were harvested and the cells were assayed as described in Materials and Methods. Enzyme activity is expressed in Miller units (17). ND, not determined.

b UAS1 elements were present in single copy in the CYCI-lacZ reporter gene in plasmid pLG669Z, a URA3 2μm-based plasmid (12).

---

**TABLE 3. Activation of ADH2 expression by ADR1 new-specificity mutants**

<table>
<thead>
<tr>
<th>ADR1 gene</th>
<th>ADHII activity* (U/mg of protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
</tr>
<tr>
<td>adrl</td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td></td>
</tr>
<tr>
<td>R115Q</td>
<td>16</td>
</tr>
<tr>
<td>H118T</td>
<td>53</td>
</tr>
<tr>
<td>R121N</td>
<td>3</td>
</tr>
</tbody>
</table>

* Measured as described in Materials and Methods. Growth of cells and abbreviations are described in Table 2, footnote a.

Description in Table 2, footnote a.

---

**FIG. 6.** Contacts between amino acids in the recognition helices of ADR1 zinc fingers and bases in the purine-rich DNA strand of UAS1. The highlighted base pairs in UAS1 are those implicated by mutagenesis, methylation interference, and change-of-specificity studies to be important contact sites for ADR1. The positions of bp 10 and 10' are shown by superscripts and subscripts, respectively. The numbers adjacent to fingers 1 and 2 correspond to the residue positions of the arginines shown nearest those numbers. Dashed, continuous, and filled triangles between amino acids and base pairs represent the apparent contribution to the binding interaction, ranging from weak to strong, respectively, based on Ala scanning mutagenesis (31).
the ability to be functionally replaced by Lys because of structural constraints in the finger. In order for the shorter Lys side chain to contact G10, the structure would have to move closer to the DNA. Such flexibility might be possible at the finger tip but not in the middle of the alpha helix where R121 resides.

The interaction of H118 with bp 9 in UAS1 is more easily understood. The imidazole nitrogen is capable of either donating or accepting an H bond, depending on its protonation state. The adenine at position 9 in UAS1, which appears to interact with H118, can accept an H bond at N7 in the major groove or donate an H bond from the primary amino group at position 6, also in the major groove. Guanine was a functional substitute for adenine at this position. It can accept an H bond at N7 and also at O6. Since guanine and adenine were equally acceptable at this position, our results favor the interaction of His in its protonated state with the N7 of guanine and adenine. This interpretation of our results is consistent with the model based on crystallographic data that was presented by Pavletich and Pabo (23). If His-118 were rotated 180° about the C9-Cy bond, it could H bond with the O6 of guanine 9 but would not be able to serve as an H-bond donor to adenine 9. A model in which an equivalent His in Zif268 was similarly rotated was equally consistent with the crystallographic data but seems less likely given the interchangeability of adenine and guanine at position 9 in UAS1.

The interactions with DNA of the three new-specificity mutations isolated in this study also seem chemically feasible. Mutants R115Q and R121N, which bind to variant sites G10A and G8A, respectively, have amide side chains that could H bond with the ring N7 group and the amino group on position 6 of adenine, as pointed out by Seeman et al. (24).

The change-of-specificity mutant H118T preferably interacts with the mutant site with a CG base pair at position 9. The wild-type protein had no detectable binding to this mutant site. The Thr hydroxyl group might donate an H bond to the O6 position of guanine in the major groove, although this would be an interaction with a base on the strand opposite that with which the other residues are thought to interact.

In all of these cases, DNA structure may also play a role in determining the specificity of interaction. However, in the Zif268-DNA complex, the DNA was essentially in a B-type helix (23).

A critical point in designing zinc finger proteins with new DNA-binding specificities is the independence of each residue in the recognition helix from neighboring residues. The ability of the R115Q-H118T double finger mutant to bind with high affinity to a double-mutant site indicates that these two positions in the recognition helix can be functionally independent. The same was not true of double mutants in which R121 was substituted by Asn. Thus, only certain combinations of amino acids appear to be tolerated in the recognition helix. This conclusion is also suggested by the highly nonrandom nature of amino acids in the finger tip and recognition helix of zinc finger proteins.

A second critical feature in the construction of designer DNA-binding fingers is the relationship between adjacent subsites and adjacent fingers. Does each finger recognize uniquely the base pairs within its 3-bp subsite, or are there important contributions from interactions of a finger with base pairs in neighboring subsites? If there are interactions between a finger and an adjacent subsite, these interactions would impose a constraint on the combinations of fingers which could be assembled to bind DNA. As described above, one interpretation of the phenotype of the change-of-specificity mutant R121N suggested an interaction of the amino acid at this position with a base pair in the subsite recognized primarily by finger 2, guanine 7. This possibility is raised more strongly by the partial loss-of-contact phenotype of D145A at position 8 in UAS1, a position in the subsite primarily occupied by finger 1 amino acids. In the Zif268-DNA complex, it was noted that an equivalent Asp residue had one of its carboxylate oxygens within H-bonding distance of a neighboring base on the secondary strand (23), that is, with a base in an adjacent subsite. Although it was suggested that because of unfavorable geometry this interaction might not have a role in DNA recognition, our results suggest that in ADR1 such an interaction may be influential.

These studies allow the alignment of the zinc fingers of ADR1 on their DNA-binding sites. Each finger is oriented from N to C terminus of the recognition helix along the DNA in a 3'-to-5' orientation on the G-rich strand, as shown in Fig. 6. This is the same orientation that Zif268 has in its complex with DNA as seen by X-ray crystallography (23) and the same orientation that was deduced for Krox20 on its DNA site (19). Thus, the two monomers of ADR1 bound to UAS1 are oriented in a tail-to-tail orientation.

Determining the binding specificities of additional fingers might eventually allow a recognition code (13, 23) to be formulated. At present, only a few generalities seem to be warranted. The interaction of Arg with guanine seems to dominate the interaction of zinc fingers with DNA.

The importance of Arg-guanine interactions is underscored by the failure of the ADR1 mutants containing change-of-specificity mutations R115Q and R121N to function normally in vivo despite their ability to bind DNA with high affinity in vitro. Why are they not active? A nearby residue, E117, has a role in transactivation at a step after DNA binding (31). One possible explanation for the defective transactivation function of mutants R115Q and R121N is that the function of E117 is impaired in the mutant fingers. This explanation is consistent with the observation that mutants R115A and H118A were much more defective in transactivation than they were in DNA binding (31). Structural analysis of these and other mutant peptides is in progress (12a, 21). If these studies indicate that the mutant peptides have altered structures, this finding might help to explain their defective transactivation.

Can a set of rules governing the interaction between zinc fingers and DNA be devised? The precise binding sites for only a few zinc finger proteins of the C2H2 class are known, together with the fingers which interact with these sites. These finger proteins are Krox20, SP1, and Zif268, and ADR1 (19, 23; this work). The Krox20, SP1, and Zif268 proteins each contain three fingers, but they are very closely related (Table 4). At the positions important for specificity, designated positions -1, 3, and 6 of the recognition helix, they are either RER or R(K)HX. In the RER type, Gln did not contact DNA in the Zif268-DNA complex. In the RHX class, X is a Thr in finger 2 of Zif268 and it did not contact DNA either. The binding sites are likewise very limited in composition, being either GGC for the RER class or GGG and GGT for the RHX class. Only the ADR1 fingers recognized a site with an A in the G-rich strand.

From this limited set of data, derived from structural studies (23) and change-of-specificity mutants (19; this work), a few generalities are apparent. Positions -1, 3, and 6 in the recognition helix (counting the first Arg as position -1) are important for base discrimination. The amino acid at position 2 may be important as well, but more studies are needed to examine its role in binding specificity. Arg is the
TABLE 4. Comparison of zinc finger recognition helices and their cognate binding sites in DNA

<table>
<thead>
<tr>
<th>Protein*</th>
<th>Recognition helix position*</th>
<th>Binding site (3' to 5')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADR1wt F1</td>
<td>R Q E H L K R</td>
<td>G A/G G</td>
</tr>
<tr>
<td>ADR1/R115Q F1</td>
<td>Q E H L K R</td>
<td>A A G</td>
</tr>
<tr>
<td>ADR1/H118T F1</td>
<td>R Q E T L K R</td>
<td>G C G</td>
</tr>
<tr>
<td>ADR1/R121N F1</td>
<td>R Q E H L K N</td>
<td>G A A</td>
</tr>
<tr>
<td>ADR1 wt F2</td>
<td>R D L L I R</td>
<td>G T T</td>
</tr>
<tr>
<td>Zif268/Krox20 F1</td>
<td>R S D E L T R</td>
<td>G C G</td>
</tr>
<tr>
<td>F3; Krox20/2ER</td>
<td>F2; SP1 F2</td>
<td>R S D E L Q R</td>
</tr>
<tr>
<td>Zif268 F2</td>
<td>R S D H L T T</td>
<td>G G T</td>
</tr>
<tr>
<td>Krox20 F2</td>
<td>R S D H L T T</td>
<td>G G G</td>
</tr>
<tr>
<td>SP1 F3</td>
<td>K T S H L R A G</td>
<td>G G G</td>
</tr>
<tr>
<td>SP1 F3</td>
<td>R S D H L S K</td>
<td>G G G</td>
</tr>
<tr>
<td>Krox20/2HR F2</td>
<td>R S D H L T R</td>
<td>G G G</td>
</tr>
<tr>
<td>Krox20/QR F2</td>
<td>R S D Q L T R</td>
<td>G C G</td>
</tr>
</tbody>
</table>

* F1, F2, and F3, fingers 1, 2, and 3, respectively.

Changes of specificity mutants

most common amino acid at positions -1 and 6 in this class of zinc finger proteins (unpublished data). In the examples studied, Arg always recognizes a guanine in its binding site. In the case of ADR1, Lys could functionally substitute for Arg at position -1 in finger 1 but not in finger 2, nor at position 6 in finger 1. Lys at position -1 in finger 1 of ADR1 also recognizes a guanine. Gln can functionally replace Arg at position -1 in finger 1 of ADR1 but not in finger 2, and when it does so, the mutant protein recognizes a mutant site with A replacing G.

At position 3 of the recognition helix, His can recognize either an adenine or a guanine in the middle position of its binding site. The Krox20 mutant QRQ recognizes GCC and GGG, but the importance of the middle base for its binding is uncertain. In the Zif268 finger of this type, the amino acid at position 3 in the helix, Glu, did not contact DNA. When Thr replaces His at this position in ADR1, GCC is the preferred binding site and A and G are less favored substitutions. The Asn substitution at position 3 was not a functional replacement for His, even when the middle base of the triplet was A (unpublished data). Thus, the ability of Asn to recognize A is position dependent: it can do so at position 6 of the recognition helix but not at position 3.

It is clear from this brief synopsis that further work is needed to fully understand the interactions of finger proteins with DNA. There are different types of interactions between DNA and identical residues situated at different positions within a finger, and there are different interactions between DNA and identical residues situated at analogous positions in adjacent fingers. The data for the double mutants also suggest that there are interactions between residues that themselves interact with DNA. These interactions may be important to allow a functional recognition helix to form. Nevertheless, these data do allow predictions to be made about the sequence of the binding sites for some of the many C2H2 zinc finger proteins that have been described. Conversely, once the binding site for a zinc finger protein of this class has been precisely determined, some predictions are possible about the composition of the recognition helix.

ACKNOWLEDGMENTS

We thank K. Hughes, R. Klevit, J. Herriott, S. Camier, K. Dombek, C. Cheng, M. Donoviel, and J. Saario for valuable comments on the manuscript and encouragement during the course of these experiments. A perceptive reviewer also helped us improve the manuscript.

This research was supported by grant GM26079 from the National Institutes of Health.

REFERENCES

8a. Dombek, K. Personal communication.
12a. Klevit, R. E. Personal communication.