Functional Characterization of DNA-binding Domains of the Subunits of the Heterodimeric Aryl Hydrocarbon Receptor Complex Imputing Novel and Canonical Basic Helix-Loop-Helix Protein-DNA Interactions*

(Received for publication, December 29, 1995, and in revised form, January 31, 1996)

Steven G. Bacsi and Oliver Hankinson‡
From the Department of Pathology and Laboratory Medicine and the Jonsson Comprehensive Cancer Center, Medical School, University of California, Los Angeles California 90095-1781

The aryl hydrocarbon receptor (AHR) and the aryl hydrocarbon receptor nuclear translocator (ARNT) belong to a novel subclass of basic helix-loop-helix transcription factors. The AHR-ARNT heterodimer binds to the xenobiotic responsive element (XRE). Substitution of each of four amino acids in the basic region of ARNT with alanine severely diminishes or abolishes XRE binding, intimating that these amino acids contact DNA bases. Three of these amino acids are conserved among basic helix-loop-helix proteins, and the corresponding amino acids of Max and USF are known to contact DNA bases. Alanine scanning mutagenesis of the basic domain of AHR and substitution with conservative amino acids at particular positions in this domain and in a more amino-proximal AHR segment previously shown to be required for XRE binding (Fukunaga, B. N., and Hankinson, O. (1996) J. Biol. Chem. 271, 3743–3749) demonstrate that the most carboxyl-proximal amino acid position of the basic domain and a position within the amino-proximal segment are intolerant to amino acid substitution with regard to XRE binding, suggesting that these two amino acids make base contacts. Amino acid positions in these AHR regions and in the ARNT basic region less adversely affected by substitution are also identified. The amino acids at these positions may contact the phosphodiester backbone. The apparent bipartite nature of the DNA binding region of AHR and the identity of those of its amino acids that apparently make DNA contacts impuete a novel protein-DNA binding behavior for AHR.

The AHR mediates carcinogenesis by certain environmental pollutants, including the halogenated aromatic hydrocarbon, TCDD, and the polycyclic aromatic hydrocarbon, benzo(a)pyrene (reviewed in Ref. 1). Prior to binding of ligand, AHR is located in the cytoplasm as part of a complex that has a molecular mass of about 280 kDa. This complex is comprised of AHR, two molecules of the 90-kDa heat shock protein, and possibly other proteins (2–4). After binding ligand, AHR dissociates from the above complex and translocates to the nucleus where it heterodimerizes with ARNT. The heterodimer of AHR and ARNT constitutes a transcription factor referred to as the transformed AHR complex, which stimulates the synthesis of the CYP1A1 protein and several other proteins involved in xenobiotic metabolism (5–8). Induction of CYP1A1 is regulated exclusively at the transcriptional level (9, 10). Activation of transcription occurs through interaction of the transformed AHR complex with several copies of short sequences, termed xenobiotic responsive elements (XREs) or dioxin-responsive elements, located within the 5′-flanking region of the CYP1A1 gene.

The AHR and ARNT proteins both contain bHLH motifs toward their amino termini (11–13). Additionally, an approximately 300-amino acid PAS homology region is located more centrally in both proteins. PAS regions are also found in the Drosophila proteins PER and SIM (14) and the mammalian hypoxia-inducible factor 1α (15). PAS regions mediate homodimerization of PER and heterodimerization of PER with SIM (16) and are necessary for heterodimerization of AHR and ARNT (17, 18). As well as possessing the PAS homology region, the AHR-ARNT heterodimer differs from other bHLH bearing transcription factors in at least two other ways: (i) AHR activity is ligand activated and (ii) unlike most other bHLH-bearing transcription factors, whose DNA recognition sequence is the E-box sequence, CANNTG (19), the AHR-ARNT heterodimer recognizes an asymmetrical XRE sequence that only partially resembles the E-box. The consensus core XRE sequence is 5′-TNGCGTG-3′ (20–22). We previously determined the orientation of the AHR-ARNT heterodimer on the asymmetric XRE sequence by UV light covalent cross-linking analysis. ARNT contacts the thymidine in the XRE core (5′-CGTG-3′), whereas AHR binds 5′-proximal to this (23).

x-ray crystallographic analysis of four bHLH transcription factors (homodimers of Max, USF, MyoD, and E47) bound to their cognate DNA sequences has shown that protein-DNA interactions occur directly through the basic domain of each monomer (which manifests as an α-helical extension of the subunit's helix 1), whereas the HLH domain mediates intermolecular dimerization (24–27). We previously demonstrated corresponding requirements for the equivalent regions of AHR and ARNT by deletion analysis (17, 28).

The two central nucleotides of the E-box and to a lesser extent the identity of the nucleotides flanking the invariant 5′-CA and TG-3′ residues, dictate to which bHLH proteins the E-box will bind. (29–31). Specificity for the central two nucleo-
otides (typically CG or GC) is determined primarily by the identity of the amino acid located one helical turn toward the carboxyl terminus from a conserved glutamic acid residue in the basic regions of the bHLH proteins (found at position 36 of Max, see Fig. 3). When arginine is found at this position (as for example, in Max, USF, and TFEβ), CG is preferred. When a smaller, nonpolar residues is present (as for example in AP4, MyoD and E12), GC is preferred (19).

The basic domain of ARNT conforms well to the consensus polypeptide sequence for this submotif, whereas the basic domain of AHR (which we refer to as its nominal basic domain) conforms poorly, consistent with the observation that the DNA recognition sequence for AHR is noncanonical for this class of transcription factors. In this paper, we have studied the basic domain of ARNT and the equivalent region of AHR, as well as flanking regions, in order to characterize the protein-DNA interactions of the AHR-ARNT-XRE complex. Using an alanine scanning mutagenesis approach, we tested individual residues for XRE binding by performing EMSA. The alanine scanning mutagenesis approach was used, because (i) alanines confer α-helical secondary structure in polypeptide chains and α-helical secondary structure is manifested by the basic regions of solved bHLH proteins as they contact DNA and (ii) alanine lacks a reactive side group. Our results indicate that the mode of interaction of the basic domain of ARNT with DNA resembles that of other bHLH proteins. However, we provide evidence that AHR exhibits a pattern of DNA interaction that is novel for bHLH proteins. We find that XRE binding requires only the carboxyl-proximal portion of the nominal basic region of AHR. We previously demonstrated that DNA binding by AHR also requires a block of amino acids amino-terminal to the nominal basic domain (18). By characterizing both alanine substitution mutations and certain conservative amino acid substitution mutations in both the above amino-terminal block and in the nominal basic region of AHR, we discriminate between amino acids positions that are less tolerant to substitution with regard to XRE binding and those which are more tolerant. The former may contact bases in the XRE, and the latter may contact the phosphodiestere backbone. By peptide analysis and through informative mutant AHR proteins, we impute a non-α-helical structure for the extended putative DNA-binding region of AHR, further indicating the unique nature of the DNA-binding domain of this bHLH protein.

MATERIALS AND METHODS

Generation of ARNT Alanine Scanning Mutants—All ARNT constructs are presented in Table I. The full-length ARNT construct pCDNA3/ARNT was used for the generation of all ARNT mutants. ARNT alanine scanning mutants A(86–89), A(92, 95–97), A(103–106), R92, N93A, H94A, E98A, R99A, R100A, R101A, and R102A were generated by the overlap extension method of PCR (32) using UITma DNA polymerase (Perkin-Elmer). Briefly, oligonucleotides were synthesized to provide internal 5′ and internal 3′ primers containing complementary alanine codon substitutions in their 5′ regions. These primers were used in conjunction with external 3′ and external 5′ primers, respectively, in a primary PCR reaction. pCDNA3/ARNT was used as the template in the primary reactions resulting in alanine codon-substituted 5′ and 3′ PCR products. These products were used as templates for a second round of PCR in which the 5′ and 3′ PCR products linked together to form a full-length alanine codon-substituted product, which upon amplification with only the external 5′ and external 3′ primers results in an alanine codon-substituted product. The external 5′ primer corresponded to pCDNA3 bases 839–871. The external 3′ primer corresponded to the complement of ARNT bases 691–705. PCR products from the second round of amplification were digested with the restriction enzymes HindIII and Apal and subsequently ligated with the 899-base pair ApalI-KpnI and 6443-base pair HindIII-KpnI fragments of pCDNA3/ARNT to produce the alanine-substituted forms of the ARNT protein. Amino-terminally truncated ARNT mutants NAB9 and NA85 were generated using standard PCR reactions and primers that generated the desired truncation. PCR products amplified from pCDNA3/ARNT were digested with HindIII and KpnI and ligated with pCDNA3/ARNT fragments as described above for generation of alanine scanning conserving mutants. All ARNT clones were sequenced to verify the mutation.

Generation of AHR Mutants—AHR constructs are presented in Tables II and III. The full-length AHR construct pCDNA3/ARHR described for the generation of AHR (18) was used for the generation of all AHR alanine scanning and conservative amino acid substitution mutants (A27–29), A(40, 42), E28A, G29A, 130A, K31A, S32A, N33A, P34A, S35A, K36A, R37A, H38A, R39A, D40A, R41A, L42A, N43A, Y9W, Y9S, R14K, H38N, and R39K, were generated by the overlap extension method of PCR using UITma DNA polymerase. The protocol for generating alanine scanning AHR alanine mutants was the same as for ARNT. Differences: (i) the internal primers were designed to generate either the desired alanine codon(s) or the desired conservative substitution codon; (ii) the template for the primary PCR reactions was pCDNA3/ARHR; and (iii) a different external 3′ primer was used in the primary and secondary PCR reactions. PCR products from the second round of amplification were digested with the restriction enzymes HindII and Bpu1102I and subsequently ligated into similarly digested pCDNA3/ARHR from which the corresponding HindII-Bpu1102I fragment had been removed. AHR deletion mutants Δ17–26, Δ17–32, and Δ18–32 were generated by PCR using pCDNA3/ARHR and primers designed to produce the desired deleted form upon amplification. Each AHR mutant clone was sequenced to verify the mutation.

Transcription—All Transcription—All transcription constructs, including the AHR and ARNT parent clones as well as their mutant derivatives, represent cDNA forms of each gene contained within the pCDNA3 expression vector in the appropriate orientation for in vitro transcription from the T7 promoter. Constructs were expressed in the TNT T7 coupled reticulocyte system in the presence or the absence of [35S]methionine (final concentration, 1 μCi/m; specific activity, >1,000 Ci/mmol; Amersham Corp.) according to the protocol provided by the supplier (Promega Biotech). Reactions were incubated at 30 °C for at least 90 min. Expression of each construct was assayed by SDS-polycrylamide gel electrophoresis of an aliquot from an incubation performed in the presence of [35S]methionine. After drying of the gel, quantitation of protein expression was performed with the aid of an AMBIS Radioanalytic Imaging System (AMBIS Inc.), which will henceforth be referred to as β-scanning.

Dimerization of Mutant ARNT and AHR Proteins with Wild Type Partners—Full-length ARNT or its mutant derivatives were synthesized in vitro as described above, in the presence of [35S]methionine, whereas AHR and its mutant derivatives were synthesized in the absence of isotope. Radiolabeled ARNT or a radiolabeled-mutant ARNT was mixed with an equimolar amount of AHR mutant derivative or AHR, respectively. The mixed proteins were incubated with 10 nM TCDD (dissolved in Me2SO to a final concentration of 0.2% Me2SO) or solvent alone for 1.5 h at room temperature. Following this incubation, the mixture was adjusted to 25 mM Hepes, 1.2 mM EDTA, 10% glycerol, 200 mM NaCl, 0.1% Nonidet P-40, pH 7.4 (immunoprecipitation buffer). Affinity purified polyclonal antibody to a peptide corresponding to amino acids 12–31 of AHR (7) or the corresponding preimmunized IgG fraction was then added, and the mixture was incubated at room temperature for 1 h. The resultant immune complexes were precipitated with protein A-Sepharose CL-4B beads (Pharmacia Biotech Inc.) for 1 h at room temperature. The immunoprecipitates were washed four times with immunoprecipitation buffer, whereas the supernatant fractions were precipitated with three volumes of acetone. Before subjecting each fraction to 7.5% SDS-polyacrylamide gel electrophoresis, they were boiled in SDS sample buffer (33). After drying, the gels were exposed to x-ray film and analyzed by β-scanning to quantitate the dimerization of radioactive in each immunoprecipitate. The relative TDCC-induced dimerization capacity of each ARNT or AHR mutant was calculated as a percentage of the amount of wild type ARNT or AHR in the same experiment. Mean values were calculated from three independent experiments.

Wild-type—Wild-type ARNT and AHR or their mutant derivatives were synthesized in vitro in the absence of isotope and mixed in equimolar ratios followed by incubation in the presence or the absence of 10 nM TCDD (in Me2SO to a final concentration of 0.2% Me2SO) or solvent alone at room temperature for 1.5 h. Each reaction was then adjusted to 25 mM Hepes, pH 7.5, 200 mM KCl, 10 mM dithiothreitol, 10% glycerol, 5 mM EDTA, 50 μM of poly(dI-dC)-(dI-dC) (Pharmacia) and incubated an additional 20 min at room temperature. The reactions were then incubated another 20 min in the presence of a 32P-labeled double-stranded synthetic oligonucleotide containing mouse XRE1 (7).
ARNT and AHR, TCDD, and 32P-labeled XRE results in a previously demonstrated that incubation of residues were substituted with alanine. Using EMSA, we pre-
tostudytheXREbindingroleofeachaminoacidresiduewithin of bHLH Proteins with the E-box Subclass CACGTG—

**RESULTS**

ARNT's Interaction with the XRE Resembles the Interaction

Protein-DNA complexes were examined by electrophoretic mobility shift analysis (EMSA) using 4.5% non-denaturing polyacrylamide gels in 1 × HTE buffer (200 mM HEPES, 100 mM Tris, 5 mM EDTA, pH 8.0). After drying, the gels were exposed to x-ray film and then analyzed by β-scanning to quantitate AHR-ARNT-XRE complexes. The value for each mutant AHR or ARNT protein was calculated as a percentage of the amount obtained with the corresponding wild type protein in the same experiment. Mean values were calculated from three independent experiments.

| Clone | Dimmerization | XRE Binding | ARNT A.A. Sequence from residue 86-106 |
|-------|---------------|-------------|---------------------------------------|
| ARNT  | 100%          | 100%        | KERL[AIRENSIEIERRR] YNMT               |
| N489  | 49 ± 5*       | 35 ± 3*     | M[AIRENSIEIERRR] YNMT                 |
| N685  | 47 ± 7*       | 28 ± 3*     | MKERL[AIRENSIEIERRR] YNMT            |
| A(86-89) | 97 ± 6       | AAAA[AIRENSIEIERRR] YNMT |
| A(92,95-97) | 102 ± 9     | KERL[AARHAAAAAER] YNMT              |
| A(103-106) | 115 ± 9     | KERL[AIRENSIEIERRR] AAAA            |
| R91A  | 95 ± 3        | 46 ± 6*     | KERL[AIRENSIEIERRR] YNMT               |
| N91A  | 98 ± 8        | 75 ± 2*     | KERL[AARHAAAAAER] YNMT               |
| H94A  | 111 ±11       | 6 ± 1*      | KERL[AENRASEIERRR] YNMT               |
| R99A  | 104 ± 7       | 1 ± 1*      | KERL[AIRENSIEIERRR] YNMT             |
| R99A  | 86 ± 5        | 40 ± 1*     | KERL[AENRASEIERRR] YNMT               |
| R100A | 98 ±13        | KERL[AIRENSIEIARRR] YNMT            |
| R101A | 86 ± 4        | 9 ± 2*      | KERL[AIRENSIEIERRR] YNMT             |
| R102A | 82 ± 6        | <1%         | KERL[AIRENSIEIERRRA] YNMT            |

*: Significantly less than ARNT (p<0.05)  
µ- Severely Diminished DNA interaction  
0 - Moderately Diminished DNA interaction

These results indicate that the amino acids that are substituted in A(92, 95–97) do not contact DNA and that amino acids immediately amino-terminal (A(86–89)) or carboxyl-terminal (A(103–106)) to the basic region are also unlikely to contact DNA.

Four single alanine substitution mutant ARNT proteins (H94A, E98A, R101A, and R102A) exhibited markedly reduced XRE binding of between 0 and 12% of that observed for wild type ARNT (p<0.05), indicating that critical XRE-protein contacts occur at these positions. Three other single alanine substitution ARNT mutants (R91A, N93A, and R99A) exhibited moderately diminished complex formation (between 40 and 75% of the wild type ARNT value (p < 0.05)), suggesting that these positions may also be involved in contacting the XRE. ARNT mutant R100A bound the XRE as efficiently as wild type ARNT. The conserved arginine residue at position 100 is therefore probably not involved in XRE-protein contact, as has been shown to be the case by x-ray crystallography for the corresponding arginine residue of other bHLH proteins contacting the E-box (see Fig. 3). These findings are summarized in Table I and Fig. 3.

Each ARNT mutant with reduced XRE binding activity was tested for its ability to heterodimerize with AHR, because reduced levels of AHR-ARNT-XRE complex formation could reflect diminished dimerization capacity rather than a direct effect on DNA binding. [35S]methionine-radiolabeled in vitro synthesized mutant or wild type ARNT protein was mixed with an equimolar amount of unlabeled AHR in the presence or the absence of 10 nM TCDD. Following dimerization, the mixtures were immunoprecipitated using AHR antibody. The degree of heterodimerization for each mutant protein with wild type AHR was calculated as a percentage of the amount of wild type ARNT coimmunoprecipitated with wild type AHR. The first six lanes of Fig. 2A represent the controls for the coimmunoprecipitation assay. These reactions were performed with wild type AHR and wild type ARNT proteins incubated in the presence or the absence of TCDD and utilized AHR antibodies or the corresponding preimmune IgG fraction, as indicated. The results of the control reactions showed that TCDD treatment increased the amount of ARNT protein that coimmunoprecipitated with AHR and that very little or no ARNT was coimmunoprecipitated with AHR and that very little or no ARNT was coimmunoprecipitated with AHR and that very little or no ARNT was coimmunoprecipitated with AHR and that very little or no ARNT was coimmunoprecipitated with AHR and that very little or no ARNT was coimmunoprecipitated with AHR and that very little or no ARNT was coimmunoprecipitated with AHR and that very...
noprecipitated by the preimmune IgG fraction. Therefore, the coimmunoprecipitations were efficient, TCDD-inducible, and specific for AHR-ARNT heterodimers. Each ARNT mutant performed as efficiently as wild type ARNT in the heterodimerization assays, demonstrating that the observed reductions in XRE binding ability represent altered DNA binding capabilities rather than decreased formation of AHR-ARNT heterodimers.

Analysis of ARNT Mutants with Deletions Amino-terminal to the Basic Domain—Two ARNT mutant proteins with large amino-terminal deletions were assayed to investigate whether XRE binding requires amino acids located beyond the basic domain in the adjacent amino-terminal portion of ARNT, as previously shown for AHR (18). ARNT mutant N3A9 was constructed by introducing a start (methionine) codon 5′ to residue Ala9, whereas ARNT mutant N3A95 has a methionine start codon 5′ to residue Lys96. Each of the corresponding cDNA were expressed in vitro, producing proteins that expressed at levels equivalent to wild type ARNT and of the expected molecular mass (approximately 77 kDa). These mutants were tested as described for the ARNT alanine mutants in heterodimerization and XRE binding assays. The average results for the mutants, derived from three independent experiments, are shown in Table I. A representative autoradiogram of the heterodimerization analysis is shown in Fig. 2C, and a representative autoradiogram of the XRE binding assay is shown in Fig. 1C. Both amino-terminal deletion ARNT mutants exhibited reduced levels of heterodimerization with ARNT, amounting to approximately 50% of that observed for the wild type ARNT protein (p < 0.05). Their XRE binding activity was somewhat more adversely affected. Because we have shown previously that deletion of the first 70 residues of ARNT results in normal levels of both heterodimerization and XRE binding (17) and because mutant A(86–89) is unaffected in DNA bindings, these results suggest that amino acids 70–85 are required for maximal dimerization and are also possibly involved directly in DNA binding. However, we cannot rule out the possibility that tertiary structural changes associated with these deletion mutants are responsible for the results observed.

AHR-XRE Interactions Differ from Other bHLH Protein-DNA Interactions—A panel of AHR alanine mutants was constructed in order to study the protein-DNA interactions of the nominal basic domain of AHR (residues 24–39) as well as the adjacent amino-terminal region of helix 1 (residues 40–43). EMSA and heterodimerization assays were performed as described above using the mutant AHR proteins in conjunction with wild type ARNT. The average results from three independent experiments are presented in Table II. Representative autoradiograms for EMSA (except for mutant P3A4) are presented in Fig. 1B, and representative results for the heterodimerization assays for those mutants with altered XRE binding are presented in Fig. 2B. A number of AHR single substitution alanine mutants (E28A, G29A, I30A, K31A, S32A, P34A, R37A, and D40A) formed AHR-ARNT-XRE complexes to the same (or greater) degree as wild type AHR. Mutant P3A4 is of particular interest, because nearly all basic-helix-loop-helix proteins lack proline residues within their basic domains. Proline interrupts α-helical structure, and proline 34, being located at such a central position within the AHR basic domain, might be expected to have a significant role. However, our results show that substitution of this proline for alanine affects neither heterodimerization with ARNT nor XRE binding. Two mutants (H38A and R39A) formed complexes at less than 5% of the level observed for wild type AHR. Heterodimerization of these mutants with ARNT was not significantly different from wild type AHR in the case of H38A and was significantly higher than that for wild type AHR in the case of R39A. Thus the loss of complex formation observed with these mutants is not due to decreased heterodimer formation.

Six other AHR alanine single substitution mutants (N33A, S35A, and K36A within the basic domain and R41A, L42A, and N43A within helix 1) generated reduced levels of the XRE complex, ranging from approximately 28 to 69% of the wild type AHR value. None of these mutants heterodimerized any less efficiently with ARNT than did wild type AHR. Addition-
ally, mutant A(40, 42), which contains the substitutions present in mutants D40A and L42A, formed the XRE complex at nearly the same reduced level as observed for the L42A mutant (approximately 38 and 28% of the wild type AHR value, respectively). Contrasting the result observed for this mutant is that of another double alanine substitution mutant, A(27–29), which generated the XRE complex at approximately 40% of the efficiency of the wild type AHR. Substitutions at positions Glu28 and Gly29 individually with alanine in mutants E28A and G29A, respectively, resulted in mutant AHR proteins that formed the XRE complex at undiminished levels. Although A(27–29) carries alterations in the peptide sequence to which the AHR antibodies were raised, it was precipitated by the antibodies as efficiently as AHR and therefore could be tested for its heterodimerization potential with ARNT. A(27–29) heterodimerized with ARNT as efficiently as wild type AHR. The unexpected reduction in XRE binding by A(27–29) compared with mutants E28A and G29A may result from the run of three alanines generated in the former mutant. This may alter the secondary structure of the protein (as discussed later), rendering the mutant less efficient at interacting with the XRE. A summary of the above findings is found at the bottom of Table II.

### Table II

| Clone | XRE Binding | AHR A. A. Sequence from residue 24-43 |
|-------|-------------|--------------------------------------|
| AHR   | 100%        | basic domain helix | AEGIKSNPSKRRH | DRNL |
| A(27-29) | 39 ± 16      | 40 ± 4* | AAAIKSNPSKRRH | DRNL |
| A(40-42) | 33 ± 5       | 38 ± 2* | AEGIKSNPSKRRH | ARAE |
| E28A  | 32 ± 8       | AEGIKSNPSKRRH | DRNL |
| G29A  | 99 ± 6       | ARAIKSNPSKRRH | DRNL |
| D30A  | 113 ± 6      | AEGIKSNPSKRRH | DRNL |
| K31A  | 108 ± 9      | AEGIKSNPSKRRH | DRNL |
| S32A  | 133 ± 7%     | AEGIKSNPSKRRH | DRNL |
| N33A  | 136 ± 16     | 64 ± 1* | AEGIKSNPSKRRH | DRNL |
| P34A  | 102 ± 4      | 91 ± 6  | AEGIKSNPSKRRH | DRNL |
| E55A  | 141 ± 44     | 69 ± 5* | AEGIKSNPSKRRH | DRNL |
| K36A  | 97 ± 2       | 37 ± 4* | AEGIKSNPSKRRH | DRNL |
| R37A  | 79 ± 2       | AEGIKSNPSKRRH | DRNL |
| E38A  | 125 ± 11     | 4 ± 1   | AEGIKSNPSKRRH | DRNL |
| R39A  | 142 ± 94     | 1 ± 1   | AEGIKSNPSKRRH | DRNL |
| D40A  | 118 ± 23     | AEGIKSNPSKRRH | ARNL |
| R41A  | 124 ± 7      | 46 ± 1* | AEGIKSNPSKRRH | DRNL |
| L42A  | 120 ± 8      | 28 ± 6* | AEGIKSNPSKRRH | DRNL |
| D43A  | 104 ± 5      | 33 ± 1* | AEGIKSNPSKRRH | DRNL |

*Significantly greater than AHR (p<0.05)  | *Significantly less than AHR (p<0.05)  | *Moderately diminished DNA interaction

### Figure 2

**A**. Dimerization analysis of AHR and ARNT mutant proteins in combination with normal AHR or ARNT, respectively. The first six lanes of each panel represent control reactions. AHR antibody (Ab) was used in all reactions except in the fifth and sixth control lanes of each panel, where preimmune IgG was used. Equimolar amounts of radiolabeled ARNT or its mutant derivatives were incubated with normal AHR or its mutant derivatives as indicated below. Immunoprecipitated pellets and acetone supernatants were subjected to 7.5% SDS-polyacrylamide gel electrophoresis. AHR, AHR antibody; P1, preimmune IgG; +, no TCDD treatment; TCDD treatment; p, immunoprecipitate; s, supernatant. The positions of the molecular mass markers are indicated on the left in kDa.

**B**. Dimerization analysis of the indicated radiolabeled ARNT alanine scanning mutants with normal AHR. B, dimerization analysis of the indicated AHR alanine scanning mutants with radiolabeled normal ARNT. The eighth lane is derived from the same gel as the first seven lanes. A(40,42) appears to migrate more slowly than the other AHR proteins because of the “smile” affect manifested in lanes at the ends of gels. C, dimerization analysis of additional AHR alanine scanning and conservative substitution mutant proteins with radiolabeled ARNT and dimerization analysis of radiolabeled ARNT amino-terminal deletion mutants with normal AHR.

**C**. Discrimination between Amino Acids of AHR That May Contact Bases in the XRE and Those That May Contact the Phosphate Backbone—In order to study the protein-DNA interactions of AHR with the XRE in greater detail, several additional AHR mutants were generated in which the amino acid substitutions represent conservative changes with regard to side chain and charge. Previous work has suggested that amino acid-DNA base contacts may be distinguished from amino acid-phosphate backbone contacts by analyzing the tolerance of a specific residue for conservative amino acid substitution (34). If both alanine and conservative substitutions result in nearly or completely abolishing protein-DNA complex formation, the residue is likely to be involved in a DNA base contact. If substitution for alanine results in a near or complete abolishment of complex formation, whereas conservative substitution results in close to normal levels of complex formation, the residue in question most probably contacts the DNA phosphate backbone. Four specific AHR positions were chosen for this analysis based on the markedly reduced levels of AHR-ARNT-XRE complex formed when they were each individually substituted with alanine. Two of the substituted positions are located in the nominal basic region (residues His28 and Arg29), which are amino-terminal to the nominal basic region in a region unique to AHR among bHLH proteins that we previously identified as being required for XRE binding and that may directly contact DNA (18). The two positions chosen in the amino-terminal region are the most sensitive with regard to XRE binding among several positions in this region that are adversely affected by alanine substitution. The substituted mutants were tested for XRE
binding and for heterodimerization, as described above. The average results for three independent experiments appear in Table III, and a representative autoradiogram showing EMSA and heterodimerization assays is presented in Figs. 1C and 2C, respectively.

All the mutants dimerized with ARNT as efficiently as wild type AHR. (Although Arg14 contains a substitution within the peptide used to generate the AHR antibodies, mutation at this position does not affect its ability to be precipitated by the AHR antibodies (18).) Substitution of the tyrosine residue at position 9 with either tryptophan or serine markedly reduced XRE binding activity (10 and 8% of the wild type AHR level of complex formation for mutants Y9W and Y9S, respectively). Substitution of lysine for arginine at position 39 adversely affected XRE to nearly the same degree as alanine substitution. These data suggest that DNA base contact probably occurs at Arg39 and also perhaps at Tyr9 (if indeed the latter contacts DNA directly). Substitution of histidine 38 with asparagine affected XRE binding only mildly, suggesting that His38 probably contacts the phosphodiester backbone. Substitution of arginine 14 with lysine moderately affected XRE binding, indicating that this amino acid could make either base or phosphate contact or both (assuming that this amino acid contacts DNA directly).

A Spacer Region Is Required Between the Amino-terminal Block of Basic Amino Acids of AHR Required for XRE Binding and Its Nominal Basic Domain—Three AHR mutants (D17–26, D17–32, and D18–32) with deletions in their amino acid sequence corresponding to the residues between and including those designated in the name of each mutant, were generated to test the function of residues positioned between the nominal basic domain and a block of highly basic amino acids within the amino-terminal region of AHR required for XRE binding. When tested by EMSA, mutants D17–26, D17–32, and D18–32 all showed less than one percent of the level of AHR-ARNT-XRE complex formation compared with wild type AHR (data not shown). Dimerization analysis with ARNT was not possible for these mutants because they were not precipitable by our AHR antibody. However, because analysis of other mutants indicate that this region is not involved in heterodimerization (Table II and Ref. 18), it is reasonable to assume that these deletions adversely affect XRE binding per se.

### Table III

| Close | Dimmerization | XRE Binding | AHR A.A. Sequence from residue 1-39 |
|-------|---------------|--------------|-----------------------------------|
| AHR   | 100%          | 100%         | MSGANITYASRRKRVQKTVKPI|AE|
| Y9A   | 112 ± 3       | <13          | MSGANITYASRRKRVQKTVKPI|AE|
| Y9W   | 94 ± 7        | 10 ± 2       | MSGANITYASRRKRVQKTVKPI|AE|
| Y9S   | 98 ± 4        | 10 ± 2       | MSGANITYASRRKRVQKTVKPI|AE|
| R14A  | 84 ± 9        | 4 ± 1        | MSGANITYASRRKRVQKTVKPI|AE|
| R14K  | 107 ± 10      | 25 ± 2       | MSGANITYASRRKRVQKTVKPI|AE|
| H9A   | 125 ± 11      | 4 ± 1        | MSGANITYASRRKRVQKTVKPI|AE|
| H9N   | 92 ± 8        | 49 ± 5       | MSGANITYASRRKRVQKTVKPI|AE|
| R9A   | 142 ± 96      | 1 ± 1        | MSGANITYASRRKRVQKTVKPI|AE|
| R9K   | 93 ± 9        | 4 ± 1        | MSGANITYASRRKRVQKTVKPI|AE|

* 6. Significantly greater than AHR (p<0.05)
* 7. Significantly less than AHR (p<0.05)
* Alaine and conservative amino acid substitution severely diminishes DNA interaction
* Alaine but not conservative amino acid substitution severely diminishes DNA interaction

### DISCUSSION

bHLH and bHLH leucine zipper proteins govern the expression of critical genes involved in growth control and differentiation through specific activation or repression programs. Regulation of transcription by these proteins involves their interaction with specific DNA recognition sequences in target genes. Most bHLH and bHLH leucine zipper protein-DNA interactions occur at the E-box sequence (5'-CANNTG-3'). The bHLH-PAS proteins, AHR and ARNT, heterodimerize in the presence of an activating ligand and then transcriptionally activate responsive genes, such as CYP1A1. The DNA recognition element for the AHR-ARNT dimer, the XRE (5'-TNGCGTG-3'), is asymmetrical and only resembles an E-box at the underlined nucleotides.

DNA binding by HLH proteins requires formation of homodimers or heterodimers. The HLH protein motif is critical for formation of these dimers. Secondary dimerization domains such as the leucine zipper of bHLH leucine zipper proteins or the PAS domain found in AHR, ARNT, hypoxia-inducible factor 1α, and SIM not only function as additional protein-protein dimerization interfaces but also probably serve as a means to determine the permissible combinations of homo- or heterodimers. In the case of ARNT, evidence is building that it represents one of the ubiquitously expressed bHLH proteins, which through heterodimerization with several other bHLHPAS proteins, is involved in the regulation of multiple genes. For example, the hypoxia-inducible factor 1α is a heterodimer of hypoxia-inducible factor 1α and SIM not only function as additional protein-protein dimerization interfaces but also probably serve as a means to determine the permissible combinations of homo- or heterodimers. In the case of ARNT, evidence is building that it represents one of the ubiquitously expressed bHLH proteins, which through heterodimerization with several other bHLHPAS proteins, is involved in the regulation of multiple genes. For example, the hypoxia-inducible factor 1α is a heterodimer of hypoxia-inducible factor 1α and ARNT. ARNT can also homodimerize in vitro and bind the E-box sequence, CACGTG, and can drive transcription of a reporter gene driven by the E-box sequence (35, 36).

Crystallographic studies of four homodimeric bHLH proteins complexed with their cognate DNA sequence show that in each case the basic domain of each subunit manifests an α-helical extension of helix 1 as it interacts with DNA through the major groove (24–27). We found a pattern of amino acid residue-XRE interactions for ARNT that are similar to those known for homodimers of the bHLH proteins, Max or USF co-crystallized with their target sequences (Fig. 3). Max and USF both bind the E-box subclass CACGTG, whose half site is identical to the segment of the XRE core sequence believed to be contacted by ARNT (22, 23). We found a strong sensitivity to alanine sub-
stitution for ARNT residues His94, Glu98, Arg101, and Arg102. His94, Glu98, and Arg102 correspond in position to Max basic domain residues His28, Glu32, and Arg36, which each make direct contact with DNA bases within the E-box (see Fig. 3). Our finding of high sensitivity to alanine substitution at these positions in ARNT suggests that these residues make base contacts within the XRE. The spacing of these residues in both Max and ARNT is significant in that it sets the α-helical register such that every fourth residue aligns to the same plane on the α-helix. Thus, all three residues can face into the major groove of DNA, enabling base interaction. Max residue Arg36, corresponding to ARNT Arg101, makes contact only with the phosphodiester backbone. ARNT Arg101 could be involved in a base contact, or it may contact the phosphodiester backbone and be particularly sensitive to alanine substitution. The moderately reduced levels of AHR-ARNT-XRE complex formation resulting from substitution of each of ARNT amino acid residues Arg91, Asn93, and Arg99 with alanine suggests that these residues contact the phosphodiester backbone.

Introduction of a run of four alanine substitutions flanking the amino-terminal border of the basic domain in ARNT alanine mutant A(86–89) had no effect on AHR-ARNT-XRE complex formation. These alterations fall within an alternatively spliced region of ARNT (11, 37) and demonstrate that Phe74 and Leu89, which are located immediately adjacent to the basic region in the two alternatively spliced proteins, are insensitive to alanine substitution, consistent with the observation that both alternatively spliced products bind the XRE.

The putative XRE contacts of AHR are illustrated in Fig. 3. Two adjacent residues at the extreme carboxyl end of the nominal basic domain of AHR (His38 and Arg39) exhibited high sensitivity to alanine substitution. When a conservative substitution to lysine was made for Arg39, XRE binding was impaired strongly, suggesting that Arg39 makes base contact in DNA. Substitution of histidine 38 with asparagine had much less of an adverse affect on XRE binding than did substitution with alanine, suggesting that His38 contacts the phosphodiester backbone. Three other residues within the nominal basic domain (Asn33, Ser35, and Lys36) and three residues near the amino terminus of helix 1 (Arg41, Leu42, and Asn43) exhibited only moderate sensitivity to alanine substitution, suggesting that they may contact the phosphodiester backbone. Residue Ser35 presents the only potential target for phosphorylation within the nominal basic domain. This serine lies within a recognition sequence for phosphorylation by protein kinase C, and the reduced level of complex formation in the alanine mutant S35A could reflect a loss of the phosphorylation target, rather than presence of a phosphodiester contact. However, a recent study suggests that AHR is probably not phosphorylated in the amino-terminal half of the protein (38). We further characterized the AHR-XRE interaction by extending the observations of Fukunaga and Hankinson (18), who demonstrated that amino acids near the amino terminus of AHR, well removed from the nominal basic domain, are required for XRE binding. Two positions, Tyr9 and Arg14, at which alanine substitution abolished complex formation, suggesting possible DNA base contacts, and several other positions that were moderately sensitive to alanine substitution, suggesting possible phosphodiester contacts, were identified. We generated mutants in which we replaced Tyr9 and Arg14 with amino acids conservative with regard to charge and side group. The results from these mutants support the notion that amino acid residue Tyr9 makes base contact but leave open the question as to whether Arg14 contacts base(s) in DNA or the phosphodiester backbone. It is unlikely that tyrosine 9 is phosphorylated, because its site does not conform to any known phosphorylation recognition sequences. (It of course remains possible that the amino-terminal region of AHR does not contact DNA directly but plays some other role in DNA binding, such as directing the nominal basic region into the correct conformation for DNA binding).

A prominent feature both within the nominal basic domain of AHR and in the adjacent amino-terminal region is the presence...
of several proline residues. Proline residues affect the secondary structure of a protein by interrupting α-helices. Insight into the secondary structural nature of the nominal basic domain of AHR was provided by the data from two alanine substitution mutants. Substitution of alanine for the proline residue at position 34 did not affect AHR-ARNT dimer and the XRE, whereas reports such as the AHR-ARNT dimer and the XRE, whereas reports such as the

AHR's amino terminus through helix 1 (residues 1–60) identified no regions predicted α-helical nature for the relevant region of AHR as it associates with DNA. Certain other bHLH proteins contain proline residues in their basic regions and, like AHR, bind recognition sequences divergent from the E-box. These include E2F1, which binds the sequence 5'-GGCCGAAA-3' (39) and some heterodimeric transcription factors, including the Hes family of proteins which recognize the N-box (5'-CAGCAG-3') as well as the E-box (40–42) and enhancer of split, E(spl), which only binds the N-box (43). E2F1 is particularly interesting because it contains a unique α-helical nature for the relevant region of AHR as it associates with DNA.

In summary, we have provided evidence that ARNT interacts with the XRE in a manner highly analogous to several bHLH proteins that recognize the 5'-GTG-3' E-box half site, whereas AHR interacts with DNA in a manner unique among bHLH proteins. The putative DNA-binding region we have identified in AHR spans at least 35 amino acids, from Tyr38 to Asn43 or beyond, and is composed of two highly basic blocks of amino acids separated by a 16-amino acid intervening sequence (18) containing four proline residues. The intervening amino acids may be critical for establishing a precise protein conformation, because deletion of this region resulted in a complete loss of DNA binding. These findings indicate that AHR possesses a novel DNA-binding motif among bHLH proteins. Ultimately, x-ray crystallography should definitively reveal the structure of the DNA-protein interaction between the AHR-ARNT dimer and the XRE, whereas reports such as the present study and that of Fukunaga and Hankinson (18), provide strong indications of DNA-protein contacts and also identify which interactions are essential for XRE binding.

Acknowledgments—We thank Dr. M. Probst for providing the antibodies and for critical reading of the manuscript. We thank Dr. B. Fukunaga for assistance in designing the mutant proteins.

REFERENCES
1. Hankinson, O. (1995) Ann. Rev. Pharmacol. Toxicol. 35, 307–340
2. Perdew, G. H. (1988) J. Biol. Chem. 263, 13802–13805
3. Denis, M., Cuthill, S., Wilkstrom, A. C., Poellinger, L., and Gustafsson, J. K. (1988) Biochem. Biophys. Res. Commun. 155, 801–807
4. Chen, H.-S., and Perdew, G. H. (1994) J. Biol. Chem. 269, 27554–27558
5. Reyes, H., Reisz-Porszasz S., and Hankinson, O. (1992) Science 256, 1193–1195
6. Whitelaw, M. L., Gottlicher, M., Gustafsson, J. A., and Poellinger, L. (1993) EMBO J. 12, 4169–4179
7. Probst, M. R., Reisz-Porszasz S., Agbunog, R. V., Ong, M. S., and Hankinson, O. (1993) Mol. Pharmacol. 44, 511–518
8. Pollen, R. S., Sattler, C. A., and Poland, A. (1994) Mol. Pharmacol. 45, 428–438
9. Israel, D. I., and Whitlock, J. P., Jr. (1984) J. Biol. Chem. 259, 5400–5402
10. Gonzalez, F. J., Tukey, R. H., and Nobert, D. W. (1984) Mol. Pharmacol. 26, 117–121
11. Hoffman, E. C., Reyes, H., Chu, F.-F., Sander, F., Conley, L. H., Brooks, B. A., and Hankinson, O. (1991) Science 252, 954–958
12. Enna, M., Sogawa, K., Watanabe, N., Chujoh, Y., Matsushita, N., Gotoh, O., Funae, Y., and Fuji-Kuriyama, Y. (1992) Biochem. Biophys. Res. Commun. 184, 246–253
13. Burbach, K. M., Poland, and Arnaud, M. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 8185–8189
14. Nambo, J. R., Lewis, J. O., Wharton, K. A., and Crews, S. T. (1991) Cell 67, 1157–1167
15. Wang, G. L., J. Jang, B., Rue, E. A., and Semenza, G. L. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 5510–5514
16. Huang, Z. J., Edery, I., and Rosbash, M. (1993) Nature 364, 259–262
17. Reisz-Porszasz, S., Probst, M. R., Fukunaga, B. N., and Hankinson, O. (1994) Mol. Cell. Biol. 14, 6075–6086
18. Fukunaga, B. N., and Hankinson, O. (1996) J. Biol. Chem. 271, 3743–3749
19. Littlewood, T., and Evan, G. (1994) Transcription factors 2: Helix-loop-Helix, Academic Press, London
20. Yao, E. F., and Denison, M. S. (1992) Proc. Natl. Acad. Sci. U.S.A. 92, 3577–3581
21. Lusska, A., Shen, E., and Whitlock, J. P. (1993) J. Biol. Chem. 268, 6575–6580
22. Swanson, H. I., Chan, W. K., and Bradfield, C. A. (1995) J. Biol. Chem. 270, 26292–26302
23. Baci, S. G., Reisz-Porszasz S., and Hankinson, O. (1995) Mol. Pharmacol. 47, 43248
24. Ferré-D’Amáre, A. R., Pendegast, G. C., Ziff, E. B., and Burley, S. K. (1993) Nature 366, 38–45
25. Ferré-D’Amáre, A. R., Pogonpec, P., Roeder, R. G., and Burley, S. K. (1994) EMBO J. 13, 180–189
26. Ma, P. C. M., Roul, M. A., Weintraub, H., and Pabo, O. C. (1994) Cell 77, 451–459
27. Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S. C. (1994) Genes & Dev. 8, 970–980
28. Fukunaga, B. N., Probst, M. R., Reisz-Porszasz, S., and Hankinson, O. (1995) J. Biol. Chem. 270, 29270–29278
29. Murre, C., McCaw, P. S., and Baltimore, D. (1989) Cell 56, 777–783
30. Fisher, D. E., Carr, C. S., Parent, L. A., and Sharp, P. A. (1991) Genes & Dev. 5, 2342–2352
31. Halazonetis, T. D., and Kandi, A. N. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 6162–6166
32. Higuchi, R., Krummel, B., and Sasaki, R. K. (1988) Nucleic Acids Res. 16, 7351–7367
33. Laemmli, U. K. (1970) Nature 227, 680–685
34. Fisher, D. E., Parent, L. A., and Sharp, P. A. (1993) Cell 72, 467–476
35. Sogawa, K., Nakano, R., Kobayashi, A., Kikuchi, Y., Ohe, N., Matsushita, N., and Fuji-Kuriyama, Y. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 1386–1940
36. Antonsson, C., Arulampalam, V., Whitelaw, M. L., Pettersson, S., and Poellinger, L. (1995) J. Biol. Chem. 270, 13986–13972
37. Li, H., Dong, L., and Whitlock, J. P., Jr. (1994) J. Biol. Chem. 269, 28098–28105
38. Mahon, M. J., and Gasiwicz, T. A. (1995) Arch. Biochem. Biophys. 318, 166–174
39. Nevin, R. J. (1992) Science 258, 424–429
40. Sasaki, Y., Kageyama, R., Tagawa, Y., Shigemoto, R., and Nakashii, S. (1992) Genes & Dev. 6, 2620–2634
41. Ishibashi, M., Sasaki, Y., Nakashii, S., and Kageyama, R. (1993) Eur. J. Biochem. 215, 645–652
42. Akazawa, C., Sasaki, Y., Nakashii, S., and Kageyama, R. (1992) J. Biol. Chem. 267, 21879–21885
43. Tietze, K., Deller, N., and Knust, E. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 6152–6156
44. Jordan, K. L., Haas, A. R., Logan, T. J., and Hall, D. J. (1994) Oncogene 9, 1177–1185