



Published in final edited form as:

*Biochemistry*. 2007 January 23; 46(3): 696–708. doi:10.1021/bi061460t.

## Structural and Functional Characterization of the Aryl Hydrocarbon Receptor Ligand Binding Domain by Homology Modeling and Mutational Analysis<sup>†</sup>

Alessandro Pandini<sup>‡</sup>, Michael S. Denison<sup>§</sup>, Yujuan Song<sup>§</sup>, Anatoly A. Soshilov<sup>§</sup>, and Laura Bonati<sup>\*,‡</sup>

<sup>‡</sup>Dipartimento di Scienze dell'Ambiente e del Territorio, Università degli Studi di Milano-Bicocca, Piazza della Scienza, 1, 20126 Milano, Italy

<sup>§</sup>Department of Environmental Toxicology, Meyer Hall, University of California, Davis, California 95616

### Abstract

The aryl hydrocarbon receptor (AhR) is a ligand-dependent transcription factor that is activated by a structurally diverse array of synthetic and natural chemicals, including toxic halogenated aromatic hydrocarbons such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). Analysis of the molecular events occurring in the AhR ligand binding and activation processes requires structural information on the AhR Per-Arnt-Sim (PAS) B-containing ligand binding domain, for which no experimentally determined structure has been reported. With the availability of extensive structural information on homologous PAS-containing proteins, a reliable model of the mouse AhR PAS B domain was developed by comparative modeling techniques. The PAS domain structures of the functionally related hypoxia-inducible factor 2 $\alpha$  (HIF-2 $\alpha$ ) and AhR nuclear translocator (ARNT) proteins, which exhibit the highest degree of sequence identity and similarity with AhR, were chosen to develop a two-template model. To confirm the features of the modeled domain, the effects of point mutations in selected residue positions on both TCDD binding to the AhR and TCDD-dependent transformation and DNA binding were analyzed. Mutagenesis and functional analysis results are consistent with the proposed model and confirm that the cavity modeled in the interior of the domain is indeed involved in ligand binding. Moreover, the physicochemical characteristics of some residues and of their mutants, along with the effects of mutagenesis on TCDD and DNA binding, also suggest some key features that are required for ligand binding and activation of mAhR at a molecular level, thus providing a framework for further studies.

---

The aryl hydrocarbon receptor (AhR)<sup>1</sup> is a basic helix–loop–helix (bHLH), PAS- (Per-Arnt-Sim-) containing transcription factor which is present in numerous species and tissues and activates gene expression in a ligand-dependent manner (1–3). While the AhR can bind and be activated by a large number of structurally diverse chemicals (4–6), the highest affinity ligands include halogenated aromatic hydrocarbons (HAHs), such as 2,3,7,8-

---

<sup>†</sup>This research was supported by the National Institute of Environmental Health Sciences (ES07685 and ES05707) and the California Agricultural Experiment Station.

© 2007 American Chemical Society

\*To whom correspondence should be addressed. Phone: (++39) 02 6448 2821. Fax: (++39) 02 6448 2839. [laura.bonati@unimib.it](mailto:laura.bonati@unimib.it).

#### SUPPORTING INFORMATION AVAILABLE

One figure showing sequence alignment of the mouse *b1* allele (mAhR C57BL/6J), *d* allele (mAhR DBA/2J), and human AhRs. This material is available free of charge via the Internet at <http://pubs.acs.org>.

tetrachlorodibenzo-*p*-dioxin (TCDD, dioxin), and polycyclic aromatic hydrocarbons (PAHs), both widespread classes of environmental contaminants (4,7). Mechanistically, the inducing chemical diffuses across the plasma membrane and binds to the cytosolic AhR which exists as a multiprotein complex containing two molecules of hsp90 (a heat shock protein of 90 kDa), the X-associated protein 2 (XAP2), and the cochaperone p23 (8). Following ligand binding, the AhR is presumed to undergo a conformational change (9), exposing an N-terminal nuclear localization sequence that leads to translocation of the liganded AhR complex into the nucleus (10). Dissociation of the AhR from the protein complex and its dimerization with ARNT (AhR nuclear translocator) convert the AhR complex into its high-affinity DNA binding form (11). Binding of the ligand–AhR–ARNT complex to its specific DNA recognition site, the dioxin responsive element (DRE), leads to an increase in transcription of the adjacent gene (2,3,12).

Among the various domains of the AhR that exhibit distinct functional activities (1,13–15), we have primarily focused our attention on that of PAS B, one of the two structural repeats (PAS A and PAS B) within the PAS domain, that is involved in both ligand and hsp90 binding. In contrast to the significant amount of information available regarding AhR ligands, essentially nothing is known about the AhR ligand binding domain (LBD) itself. Full ligand binding activity and specificity of the mouse AhR are reportedly contained within a small fragment (residues 230–421), and several naturally occurring mutations within this region reduce AhR ligand binding affinity (13,16,17). Moreover, deletion of residues 287–421 results in a ligand-independent constitutively active AhR consistent the role of the AhR LBD–hsp90 complex in regulating AhR functionality (18).

Analysis of the molecular events that result from AhR ligand binding requires detailed structural information; however, no X-ray or NMR-determined structure of ligand-bound or unliganded AhR has been reported. Moreover, structural information on homologous PAS-containing proteins that could be effectively used for comparative modeling purposes was not available until 1998. Once the first crystal structures of distant homologous proteins belonging to the PAS superfamily became available [the bacterial photoactive yellow protein, PYP (19,20), the PAS domain of the human potassium channel HERG (21), and the heme binding domain of the bacterial O<sub>2</sub> sensing FixL protein (22)], we developed the first theoretical model for the LBD of the mouse AhR (mAHR) by applying homology modeling techniques (23). Despite the low level of sequence similarity, the structures of those three PAS domains showed highly conserved structural characteristics. At the time, our analysis suggested FixL as the best template for homology modeling. The resulting model showed consistency with available experimental data, thus providing an initial framework to make hypotheses on the LBD characteristics and functionality of the AhR (6,24). The subsequent determination of the X-ray structure of the PAS domain of the fern photoreceptor Phy3 (25) improved the knowledge on this superfamily and highlighted the presence of different arrangements of the secondary structure elements in the known PAS structures. This insight also provided an avenue for revising the first mAHR LBD model (26), but a more reliable proposal for the structural features of the AhR LBD could not be obtained until the structures of homologous proteins with higher sequence identity to the AhR became available.

---

<sup>1</sup>Abbreviations: AhR, aryl hydrocarbon receptor; bHLH, basic helix–loop–helix; PAS, Per-Arnt-Sim; LBD, ligand binding domain; HAH, halogenated aromatic hydrocarbon; PAH, polycyclic aromatic hydrocarbon; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; TCDF, 2,3,7,8-tetrachlorodibenzofuran; hsp90, heat shock protein of 90 kDa; XAP2, X-associated protein 2; PYP, photoactive yellow protein; ARNT, AhR nuclear translocator; hPASK, human PAS kinase; HIF-2 $\alpha$ , hypoxia-inducible factor 2 $\alpha$ ; NcoA, nuclear coactivator 1; dPER, *Drosophila* clock protein PERIOD; FMN, flavin mononucleotide; DRE, dioxin responsive element; HRE, hypoxia responsive element.

In the last 4 years an impressive growth of structural and functional knowledge on PAS domains has been obtained. During that time, a number of PAS structures were derived by NMR spectroscopy or X-ray diffraction, including the N-terminal PAS domain of human PAS kinase, hPASK (27), the C-terminal PAS B domain of human hypoxia-inducible factor 2 $\alpha$ , HIF-2 $\alpha$  (28), the PAS B domain of the mouse nuclear coactivator 1, NCoA (29), an N-terminal fragment of the *Drosophila* clock protein PERIOD, dPER, including the two PAS repeats (30), and the PAS B domain of the human ARNT protein (31). Some of these structures deserved particular attention as the full-length proteins share some key functional similarities with AhR. Both HIF-2 $\alpha$  and ARNT, like the AhR, belong to the bHLH/PAS family of transcriptional factors that are key regulators of gene expression networks underlying many essential biological processes (32). In order to become transcriptionally active, AhR and HIF-2 $\alpha$  must directly sense environmental signals (exposure to xenobiotic compounds for AhR and hypoxic conditions for HIF-2 $\alpha$ ), the PAS domains of both proteins associate with hsp90, and in both cases, their heterodimerization with ARNT appears to be required for release of chaperone proteins and conversion of the dimeric complexes into their DNA binding form. Finally, binding of the AhR–ARNT and HIF-2 $\alpha$ –ARNT complexes to their distinct but related specific E-box-like DNA recognition sequences, the DRE and the HRE (hypoxia responsive element), respectively, stimulates expression of adjacent genes. Given these commonalities in response, examination of the PAS B domains of HIF-2 $\alpha$  and ARNT, the most functionally related proteins, reveals that they have the highest degrees of sequence identity and similarity with that of AhR among all the PAS structures reported to date.

Following the new data and intriguing insights into the characteristics of PAS domains, the aim of this paper is the structural and functional characterization of the LBD of mAhR through rational mutational analysis guided by a reliable theoretical model. Therefore, an updated model of the mAhR PAS B domain was built by comparative modeling techniques. To validate our proposed model, site-directed mutagenesis of amino acids in key positions within the modeled LBD was performed, and the effects of these mutations on the mAhR ligand binding and ligand-dependent DNA binding were examined.

## MATERIALS AND METHODS

### PAS Domain Template Structures

Structures of the PAS domains currently available were obtained from the Protein Data Bank (33). The specific X-ray structures selected for the photoreceptors were those representing the dark state, and for FixL we selected the unbound state. When different depositions were available, the structure with the highest resolution was selected for use as the template. In addition, the most representative structure within each of the NMR structure bundles was selected using the NMRCLUST program (34). The PDB identification numbers of the selected PAS structures used in our analysis are presented in Table 1.

### Homology Modeling

Homology modeling techniques were applied to predict the structure of the LBD of the mouse AhR, mAhR (GI: 7304873 in the NCBI sequence database), focusing on a fragment between amino acids 230–421 that was reported to contain the full ligand binding activity and specificity (13). Sequence similarity searches with the AhR LBD were performed using PSI-BLAST (36) against a database of known protein structures with default parameters. The three-dimensional models for the wild-type and mutant mAhR LBDs were generated using MODELLER version 8v1 (37–39), which implements an approach to comparative modeling by satisfying spatial restraints, and these are extracted from the alignment of the target sequence with the multiple template structures. The restraints, which were obtained

empirically from a database of protein structure alignments, and CHARMM energy terms (40) were combined into an objective function. The resulting model was obtained by optimizing the objective function, employing methods of conjugate gradients and molecular dynamics with simulated annealing.

The templates were pairwise structurally aligned according to DALI Lite (41,42), using HIF-2 $\alpha$  as a reference. The sequence alignments were obtained by CLUSTALW (43), and the result was confirmed using the Align-2D command within the MODELLER program (37–39). This generated an alignment of sequences with structures using a variable gap opening penalty that favors gaps in exposed regions and avoids gaps within secondary structure elements. The secondary structure of the AhR LBD was then predicted by PSIPRED (44).

The quality of the models was assessed by different validation methods. General features were evaluated on the basis of the MODELLER's ENERGY scores and violation restraint lists while detailed reliability indexes were obtained by the PROCHECK program (45). Moreover, the PROSA  $z$ -score was calculated using PROSAAII (46) and employed as a quality index, with a more negative  $z$ -score indicating a better structural model. Each  $z$ -score was then normalized using the natural logarithm of the sequence length, and an estimate of the probability that the model was reliable (pG value) was derived by comparison with expected normalized  $Q$ -scores for known structures of the same length (47). Final three-dimensional visualization and images of the AhR LBD structure were generated using PyMOL (48).

### Characterization of Structural Pockets and Cavities

Identification and characterization of surface pockets and internal cavities in template and model structures were performed by the CASTp server (49). The program allows identification and calculation of Connolly's molecular surface and volume (50) for all pockets and cavities in a protein structure. It ranks the cavities by size, where the largest one is usually the binding site (49). The representations of the surface including the largest available cavity were produced with PyMOL (48).

### Molecular Dynamics Simulations

The modeled structures of the wild-type and I332P mutant mAhR LBDs were subjected to molecular dynamics (MD) simulations. These were performed with GROMACS 3.3 (51,52), by using the GROMOS96 43a2 version of the GROMOS force field as available in the GROMACS package. Structures were solvated in SPC water using cubic boxes and simulated with periodic boundary conditions. The dimensions of the box were set to allow at least 0.8 nm between protein and box faces on each side. Solvent was relaxed with 5 ps MD simulation, keeping protein degrees of freedom restrained. After addition of ions to neutralize the systems, a short minimization with steepest descent was performed up to convergence on maximum force lower than 1000 kJ/(mol·nm). The resulting systems were employed as starting points for simulations. These were carried out in the NPT ensemble for 1 ns. Different random seeds were employed to generate different starting velocities from a Maxwellian distribution at 300 K. For long-range electrostatic interactions, the particle mesh Ewald summation method (53) was employed to gain a more accurate description. Van der Waals interactions were described by a 6–12 Lennard-Jones potential with distance cutoff at 0.9 nm; neighbor lists were employed with a list cutoff of 0.9 nm and update frequency every 10 steps.

Protein and solvent were independently coupled with a thermal bath by a Berendsen thermostat at 300 K and a coupling period of 0.1 ps. The internal degrees of freedom of

water molecules were constrained by the SHAKE algorithm (54), and all bond distances in the proteins were constrained by the LINCS algorithm (55).

All of the analyses on trajectory data were performed with GROMACS. The centroid structure of the largest cluster in I332P trajectory was employed for representation in Figure 7.

### **Mutagenesis and Ah Receptor Functional Analysis. (A) In Vitro Protein Expression and Mutagenesis**

The mouse AhR expression plasmid pcDNA3/ $\beta$ AhR, obtained from Dr. Oliver Hankinson (University of California, Los Angeles), and the mouse ARNT expression plasmid pcDNA3.1-mARNT (56) were used as templates for expression of the AhR and ARNT. AhR and ARNT were synthesized *in vitro* using the TNT Quick coupled transcription/translation rabbit reticulocyte lysate system (Promega) as previously described (56). Mutation of selected amino acids within the AhR to alanine (R282, H285, T311, C327, M334, E339, K350, and Q377), proline (I332), valine or leucine (A375), or glutamic acid (M334) was carried out using the QuickChange mutagenesis technique (Stratagene). Translation grade L- $^{35}\text{S}$ -methionine ( $>400$  Ci/mmol) was purchased from MP Biomedical (Solon, OH).  $^{35}\text{S}$ -Radiolabeled wild-type and mutant mAhRs were synthesized in separate reactions *in vitro*, denatured, and subjected to SDS-polyacrylamide gel electrophoresis, and the level of expression of each protein was determined by autoradiography of the dried gel.

### **(B) Synthetic Oligonucleotides and Gel Retardation Analysis**

A complementary pair of synthetic oligonucleotides containing the sequences 5'-GATCTGGCTCTTCTCACG-CAACTCCG-3' and 5'-GATCCGGAGTTGCGTGAGAA-GAGCCA-3' (corresponding to the AhR binding site of DRE3 and designated as the DRE oligonucleotide) was synthesized, purified, annealed, and radiolabeled with [ $\gamma$ - $^{32}\text{P}$ ]-ATP [6000 Ci/mmol (Amersham)] as described (57). Aliquots of *in vitro* synthesized mouse ARNT and wt AhR or mutant AhR (1.5  $\mu\text{L}$  each) were combined with 7  $\mu\text{L}$  of MEDG buffer [25 mM MOPES, pH 7.5, 1 mM EDTA, 1 mM DTT, and 10% (v/v) glycerol] and incubated in the presence of 20 nM TCDD or DMSO solvent control for 1.5 h at room temperature. Aliquots of the incubation reaction were mixed with [ $^{32}\text{P}$ ]DRE oligonucleotide, and gel retardation analysis was carried out as previously described (56,57). Protein-DNA complexes in the dried gels were visualized, and the amount of  $^{32}\text{P}$ -labeled DRE present in the TCDD-induced protein-DNA complex was quantitated by phosphorimager analysis (Molecular Dynamics). The amount of radioactivity present in the induced protein-DNA complex in the TCDD-treated lane (the TCDD-AhR-ARNT-DRE complex) minus that present in the same position in the DMSO-treated sample lane represented the amount of specific TCDD-inducible AhR-ARNT-[ $^{32}\text{P}$ ]DRE complex. Given that multiple gel retardation assays had to be run to analyze the DNA binding activity of mutant AhRs, mutant DNA binding results were normalized to results obtained with wt AhR in each experiment and expressed as a percentage of the amount of TCDD-inducible wt AhR DNA binding.

### **(C) Hydroxyapatite [ $^3\text{H}$ ]TCDD Binding Assay**

[ $^3\text{H}$ ]TCDD (specific activity 14.5 Ci/mmol) and 2,3,7,8-tetrachlorodibenzofuran (TCDF) were provided by Dr. S. Safe (Texas A&M University). An aliquot (25  $\mu\text{L}$ ) of *in vitro* synthesized AhR reaction mixture was diluted to 100  $\mu\text{L}$  with MEDG buffer and incubated with 2 nM [ $^3\text{H}$ ]TCDD in the absence or presence of 200 nM TCDF for 2 h at room temperature. In selected experiments *in vitro* synthesized wild-type and mutant AhRs were incubated with 20 nM [ $^3\text{H}$ ]TCDD as described above. Binding of [ $^3\text{H}$ ]TCDD to the AhR was determined using the hydroxyapatite binding assay as described (57). The specific



binding of [<sup>3</sup>H]TCDD to the wild-type and mutant AhRs was computed by subtracting the amount of [<sup>3</sup>H]TCDD bound in the presence of TCDF from the amount of [<sup>3</sup>H]TCDD bound in the absence of competitor. In experiments with 20 nM [<sup>3</sup>H]TCDD, specific binding was determined by subtracting the amount of [<sup>3</sup>H]-TCDD bound to an identical concentration of unprogramed lysate from the total amount of binding to lysate containing *in vitro* expressed AhR. The amount of [<sup>3</sup>H]TCDD specific binding to each mutant AhR was expressed as a percent of the total [<sup>3</sup>H]TCDD specific binding to wt AhR.

## RESULTS AND DISCUSSION

### Choice of the Template Structures and Alignment

To select the best template structures for modeling the AhR PAS B LBD, we considered all of the available PAS domain structures at that time (Table 1, Figure 1). Besides the high ductility in signal responses developed by the PAS domains (59), these show a high structural conservation of the  $\alpha$  and  $\beta$  folds. This includes a five-stranded  $\beta$ -sheet and a long central helix (helical connector) generally linked to a bulge of three small helices. Following the nomenclature established for FixL (22) and generally adopted for PAS structures in the literature (Figure 2a), the N-terminal  $\beta$ -strands are referred to as A $\beta$  and B $\beta$ , followed by three small helices (C $\alpha$ , D $\alpha$ , and E $\alpha$ ), the helical connector (F $\alpha$ ), and the three C-terminal strands of the  $\beta$ -sheet (G $\beta$ , H $\beta$ , and I $\beta$ ).

It can be observed in the different PAS domain structures in Figure 1 that the helical connector is displaced from the  $\beta$ -sheet in different ways, designing, in some cases, a cavity suitable for arranging different kinds of cofactors. While PYP includes the covalently bound 4-hydroxycinnamic acid within the N-terminal  $\alpha$ -helical cap (including C $\alpha$ , D $\alpha$ , and E $\alpha$ ), in FixL the heme cofactor lies in the center of the domain and points to a wide entrance offered by the F $\alpha$  helical connector, the following G $\beta$  strand, and their interconnecting loop (the FG loop). In contrast, in Phy3, the flavin mononucleotide (FMN) cofactor is noncovalently bound in the middle of the domain and protrudes from it on the opposite side of the helical connector (as compared to heme in FixL), which results in greater space between the helical connector (F $\alpha$ ) and the C $\alpha$ , D $\alpha$ , E $\alpha$  group. Among the PAS domains that do not bind cofactors, the hPASK secondary structure arrangement resembles that of FixL, with the only difference being a more extended FG loop between a shortened helical connector and the G $\beta$  strand, while the arrangements of HIF-2 $\alpha$ , ARNT, dPER, and HERG structures are similar to that observed in Phy3. In NCoA, the helical connector position resembles those of the latter group of domains, but the length and the arrangement of the other helices are different.

On the basis of these observations, it emerges that the choice of the template structures is crucial for modeling the AhR PAS B LBD. The overall fold characteristics could be reproduced independently from the choice of the template; however, the length of the connecting loops and the resulting relative positions of the helices with respect to the  $\beta$ -scaffold, as well as the presence and the location of a binding pocket, would be modeled in significantly different ways depending on the structure selected as the template.

The two most critical issues in homology modeling are the degree of similarity between the target sequence and the templates and the reliability of the alignment, two aspects that are intrinsically interconnected (60). By applying a recursive PSI-BLAST search of the mAHR LBD sequence (residues 230–421) against the PDB database, the only sequence producing a significant alignment in the first cycle was that of the HIF-2 $\alpha$  PAS domain, while statistically significant homologies with the HIF-2 $\alpha$  ( $E$ -value =  $2 \times 10^{-46}$ ), dPER ( $E$ -value =  $9 \times 10^{-41}$ ), and ARNT ( $E$ -value =  $5 \times 10^{-39}$ ) PAS domains were detected after three iterative cycles. On the basis of these results, HIF-2 $\alpha$  appeared to be the more optimal reference sequence/structure for initial AhR alignment.

Known AhR sequences were first aligned internally and then aligned to the HIF-2 $\alpha$  template using CLUSTALW, and this final alignment was in complete agreement with that generated using the Align-2D command within MODELLER (data not shown). A global alignment was subsequently generated by structural alignment of each of the known PAS templates with the PAS B HIF-2 $\alpha$  by DALI, and the pairwise sequence identities and similarities between the target mAHR LBD and nine PAS domains are presented in Table 2. The highest sequence identity and similarity to the mAHR was that of HIF-2 $\alpha$  (31.1% and 62.1%, respectively), followed by ARNT (21.2% and 53.8%, respectively), whereas pairwise identities with all other sequences were below 20% and similarities below 50%. These data are in agreement with the knowledge about functional similarities across the PAS superfamily, suggesting HIF-2 $\alpha$  and ARNT as the most informative template structures.

It is likely that the first models we derived for the mAHR LBD, based on the only available PAS structures at that time [with FixL (23) and Phy3 (26) as templates], were not optimal due to the low degree of sequence similarity with the templates and the consequential uncertainty in the alignment of some regions. While we were confident about the conserved fold characteristics of the initial modeled mAHR LBD, it was clear that the resolution of the models did not allow us to reliably detect more subtle details such as loop lengths and secondary structure arrangements. Accordingly, additional experimental information was needed to further refine the homology models for use in development of structural hypotheses.

With the availability of more PAS structure information, in combination with the significantly higher degree of sequence identity and similarity exhibited by HIF-2 $\alpha$  and to a lesser extent by ARNT, the possibilities of alignment uncertainties are reduced. Analysis of the global sequence alignment of the mAHR LBD with HIF-2 $\alpha$  and ARNT PAS B (Figure 2b) reveals few insertions or deletions among these domains within the aligned region (mAHR residues 278–384). The only region of the mAHR PAS B that shows slight variability is in the DE loop, which contains a glycine insertion, and in the HI loop, where a two-residue deletion is observed. A single residue deletion in the same position in the FG loops of both mAHR and HIF-2 $\alpha$  with respect to ARNT is also revealed by our alignment. As a consequence, it was possible to define a unique optimal sequence alignment of mAHR based on the HIF-2 $\alpha$  template, and it is conceivable that this coincides with a structural-based alignment. Moreover, no explicit effort was needed to undertake loop modeling.

Interesting elements also emerge from the analysis of similar residues among mAHR, HIF-2 $\alpha$ , and ARNT as well as from comparison of the secondary structures of these templates with those predicted for mAHR by PSIPRED (see Figure 2b). While the majority of similar residues are shared by all three sequences, in some cases the inclusion of the ARNT sequence provides some additional information to support the alignment. This is mainly observed in the region from the middle of the helical connector to the end of the C-terminal  $\beta$ -strands. Additionally, differences in the alignment among the predicted secondary structures of mAHR and those of the templates occur in three places: the length of the B $\beta$  strand of mAHR seems to be in better agreement with that of ARNT, E $\alpha$  is more similar to that of HIF-2 $\alpha$ , and the FG connection appears to differ slightly from both templates. As a consequence, it is conceivable that the inclusion of additional template structures besides HIF-2 $\alpha$  in the modeling procedure, and in particular that of ARNT, could help to further improve and refine the model.

### Modeling and Model Validation

On the basis of the above observations, three different mAHR PAS B LBD models were developed and used to test the influence of choice of the templates on the quality of the modeled structure (Figure 3a). The first model used only the HIF-2R template since it had

the greatest sequence similarity (mod\_HIF-2 $\alpha$ ); the second used both HIF-2 $\alpha$  and ARNT templates (mod\_HIF/ARNT); the third used eight of the PAS structures reported in Figure 1 (mod\_8templates); PYP was excluded because of its very low degree of sequence similarity. To test each possible structure, 100 individual models were derived by MODELLER from random generation of the starting structure, and the representative model was selected by the lowest value of the objective function. The quality of each final model was evaluated by MODELLER's ENERGY command, to verify if the model satisfies most restraints used to calculate it, and additionally by the PROCHECK and the PROSAIL programs.

A limited number of violations of the MODELLER stereochemical restraints were observed for all of the models; these violations cannot be avoided given the medium–low degree of similarity of mAhR with the templates included. Also, the models passed all criteria implemented in PROCHECK: ~87% of residues reside in the “most favored” areas of the Ramachandran plot (90% for structures solved at  $\leq 2.0$  Å resolution), with only one residue (HIS320) scored in “disallowed” regions within the first two models; the overall *G*-factors, measuring stereochemical quality, range from  $-0.2$  to  $-0.1$  [from  $-0.5$  to  $0.3$  for structures solved at  $1.5$  Å resolution (61)]. The values of the PROSA *z*-score for the three tests are reported in Table 3. On the basis of the *pG* value (47), the threshold value for the *z*-score associated with a good quality model for a sequence length of 107 amino acid residues is  $-4.05$ , and models with higher *z*-scores were considered poor models. The average *z*-score for the 100 models generated by MODELLER and the score for the model with the best objective function indicated that mod\_HIF/ARNT and mod\_8templates were good quality models. Conversely, the values obtained for the mod\_HIF-2 $\alpha$  were near to the limit of acceptability. On the basis of this evaluation the reliability of both of the multitemplate models was very similar, but slightly better than the model based on HIF-2 $\alpha$  alone. On the other hand, the three models show very similar fold features, as shown by the root mean square distance (RMSD) between the positions of CR:  $0.34$  and  $0.76$  Å between the mod\_HIF-2 $\alpha$  and the mod\_HIF/ARNT and mod\_8templates, respectively, and  $0.69$  Å between the two multitemplate models. To put these differences into perspective, it has to be considered that the  $1$  Å accuracy of main-chain atom positions corresponds to X-ray structures defined at a resolution of about  $2.5$  Å and that differences between the highly refined X-ray and NMR structures of the same protein also tend to be about  $1$  Å (61). As can be evidenced by the visual comparison of the three modeled structures in Figure 3a, the slight variability among the models mainly involves the length and stereochemistry of the B $\beta$  and G $\beta$  strands and the arrangement of some connecting loops.

Due to the low reliability based on the evaluation by PROSAIL, we excluded the model based only on the HIF-2 $\alpha$  structure for the following analyses, and considering that only marginal differences were found between the model based on the combined HIF-2 $\alpha$ /ARNT structures and that obtained using the eight PAS domain structures, we concentrated on the HIF-2 $\alpha$ /ARNT two-template model (mod\_HIF/ARNT). This model is shown in different orientations and representations in Figure 3b.

The analysis of structural pockets and cavities, performed by the CASTp server, indicated the presence of a buried cavity in the core of the modeled domain, with a volume of  $496$  Å<sup>3</sup>, which falls in the observed range ( $100$ – $800$  Å<sup>3</sup>) for protein binding pockets or cavities (62). As shown in Figure 3b, the cavity, represented by the molecular surface including the available volume, is placed just in the middle of the domain, delimited by the  $\beta$ -sheet and flanked by the helical connector, the D $\alpha$  and E $\alpha$  helices, and the connecting loops.

Interesting observations emerge from the comparison of this result with those obtained by the same analysis for all of the PAS domain structures considered (Table 1 and Figure 1). In agreement with the presence of the heme and FMN cofactors, in both FixL and Phy3 PAS



domains a large pocket is found (985 and 716 Å<sup>3</sup>, respectively). In both cases this presents a wide-mouth opening that, accordingly to the cofactor position, lies between the helical connector and the following Gβ strand, in FixL, and between the helical connector and the Cα, Dα, Eα group, in Phy3. On the contrary, in the HIF-2α and ARNT PAS structures only small cavities with volumes smaller than 100 Å<sup>3</sup> were found. These results confirm that in the HIF-2α and ARNT template structures the domain interior is well packed, whereas the modeled mAhR PAS B domain has enough internal space available for ligand binding. Interestingly, very similar three-dimensional arrangements of the main chains in the mAhR model and in the templates generate very different internal spaces, depending on the different volume occupied by the side chains of not conserved residues. In particular, among the boundary residues of the mAhR cavity, four residues (P291, I319, C327, L347) have considerably smaller side chains than the ones in corresponding positions of HIF-2α (see Figure 2b). Comparing the mAhR and HIF-2α structures, it is clear that the difference in the empty internal space of the two domains is mainly due to those residues.

On the basis of the above observations and the fact that the PAS structures used for modeling were not resolved with bound hsp90, it can be concluded that the proposed mAhR LBD model can be viewed as more appropriate to describe the structural features of the ligand-bound form of mAhR. It is conceivable that the buried cavity can be made available for binding thanks to conformational changes in flexible regions of the domain allowing a “mouth opening” for ligand approach.

Because the model represents the activated state of the mAhR LBD, the domain is expected to describe also a conformation in which hsp90 has been already partially displaced. This supports the assumption that modeling the PAS B domain does not necessarily require inclusion of information from the interacting patches with the chaperone protein. Furthermore, the high conservation of the PAS fold architecture increases the reliability of independently modeling the single domain.

### Experimental Testing and Model Validation

While the above analysis provides us with a model of the mAhR LBD, validation of such a model requires experimental confirmation. Accordingly, site-directed mutagenesis and AhR functional analysis were used to test and confirm the features of the whole modeled domain as well as the location of the binding cavity. Some specific residues (Table 4) were selected to examine the effects of point mutations on both [<sup>3</sup>H]TCDD binding to the AhR and TCDD-dependent transformation and DNA binding. The rationale for these choices is readily apparent by examining the residue positions in the modeled three-dimensional structure as shown in Figure 4. Some of them, namely, Arg282, Thr311, Glu339, and Lys350 (shown in blue), have side chains pointing outside the modeled LBD, and it is expected that their mutation does not affect ligand binding. Conversely, His285, Cys327, Met334, Ala375, and Gln377 (shown in purple) were selected among the boundary residues of the cavity identified in the LBD with side chains pointing inward, to validate it as the active site in the LBD. The role of each of these residues in ligand binding could be related to the side-chain steric effects on the size and shape of the cavity or to particular stereoelectronic requirements useful for stabilizing the ligand association. To examine this aspect, different types of amino acid substitutions were inserted at a targeted amino acid (Met334 and Ala375). Finally, mutation of Ile332 (shown in yellow) to proline was planned in order to test the structural role of the helical connector, as it is likely that the inserted proline would act to break the helical arrangement, thus causing a significant structural change in the overall fold of the domain.

Gel retardation analysis of in vitro synthesized wt AhR and ARNT incubated with TCDD resulted in the formation of an inducible protein–DNA complex compared to a sample

incubated with DMSO carrier solvent (Figure 5, lanes 1 and 2). This inducible protein-DNA complex represents the TCDD–AhR–ARNT–[<sup>32</sup>P]DNA (56), and it is not observed when DNA binding reactions contained either unprogrammed lysate, in vitro expressed AhR alone, or ARNT alone (data not shown). Analysis of the ability of each in vitro synthesized mutant mAhR to transform and bind to [<sup>32</sup>P]-DRE-containing DNA in a TCDD-inducible manner is shown in Figure 5, and quantitation of the amount of TCDD-inducible DNA binding by each mutant mAhR was determined by phosphorimager analysis of multiple gel retardation assays with each mAhR and presented in Table 4. Since each mAhR was synthesized in reticulocyte lysate at similar levels [based on comparable levels of expression of in vitro synthesized [<sup>35</sup>S]-wt and mutant mAhRs (Figure 6)], the observed differences in DNA binding are not simply due to differences in expression levels of each mAhR. Gel retardation analysis revealed that while some mutations had no significant effect on TCDD-inducible DNA binding (R282A, T311A, and K350A), some mutations reduced AhR DNA binding (C327A, M334A, E339A, A375V, and Q377A) and others completely eliminated DNA binding (H285A, I332P, M334E, and A375L). Interestingly, while the amount of transformation and DNA binding of mAhR containing the R282A mutation was greater than that of wt AhR in several gel retardation analyses, it was not significantly different when all DNA binding data were combined.

While the DNA binding results clearly demonstrate that a variety of mutations within the AhR PAS B LBD adversely affect the ability of TCDD to stimulate AhR DNA binding, gel retardation analysis does not identify the actual mechanistic effect of the mutation. Mutation of a key amino acid could reduce the amount of AhR DNA binding by adversely affecting the ability of TCDD to bind within to the AhR LBD and/or to make key contacts within the ligand binding cavity necessary to stimulate ligand-dependent transformation of the AhR into its DNA binding form (i.e., ligand-dependent release of hsp90 from the AhR and its subsequent dimerization with ARNT). In order to attempt to differentiate between these possibilities, we examined the ability of [<sup>3</sup>H]-TCDD to specifically bind to each of the in vitro expressed mAhRs. As observed in the gel retardation results, mutation of the AhR LBD produced a similar range of effects on [<sup>3</sup>H]-TCDD specific binding (Table 4). As expected, no decrease in [<sup>3</sup>H]TCDD specific binding was observed with the mutant AhRs that exhibited wild-type AhR TCDD-inducible DNA binding (i.e., R282A, T311A, and K350A). The loss of AhR DNA binding activity with the H285A, I332P, M334E, and A375L mutations correlated well with the lack of [<sup>3</sup>H]TCDD specific binding to these mutant AhRs. These results are consistent with a role for these amino acids in binding of TCDD within the AhR LBD. The decrease in [<sup>3</sup>H]TCDD specific binding to mutants C327A, M334A, and Q377A also correlated well with the reduced DNA binding activity of these mutants and supports a role for these residues in TCDD binding to and transformation of the AhR. In these cases, the decrease in ligand binding is presumed to be a consequence of reduced TCDD binding affinity; however, this remains to be confirmed. Our results also revealed that two mutations (A375V and E339A) display some unusual characteristics (Table 4). While *in vitro* expressed mAhR containing the A375V mutation exhibits no [<sup>3</sup>H]TCDD specific binding, it binds to DNA in a TCDD-dependent manner, albeit to ~40% of that of wt AhR. This apparent discrepancy may result from several technical issues with regard to the ligand and DNA binding assays. First, the [<sup>3</sup>H]-TCDD ligand binding assay is much less sensitive than that of the gel retardation assay. The ligand binding assay uses [<sup>3</sup>H]TCDD with a specific activity 14.5 Ci/mmol, while the DNA binding assay uses [<sup>32</sup>P]DNA with a specific activity of ~5000 Ci/mmol, and as such, it would be easier to see [<sup>32</sup>P]DNA binding to TCDD–AhR complexes compared to [<sup>3</sup>H]TCDD binding to the AhR. Second, and more likely, is that previous studies have not only demonstrated that mutation of alanine 375 to valine reduces TCDD binding affinity by about 10-fold (from ~1 to ~10 nM), but also demonstrated that reagents typically used in AhR ligand binding assays (i.e., Tween 80 detergent in the HAP assay and charcoal in other AhR assays) can strip ligand off

of low-affinity AhRs (57). Since the AhR from species that contain the A375V mutation naturally (i.e., human and DBA mice) can still function normally in a TCDD-dependent manner, albeit higher TCDD concentrations are required (6,16,17,24,63), it must still bind ligand. Given the documented lower affinity of the A375V mutation, we repeated the binding assay with this mutant AhR using 20 nM instead of 2 nM [<sup>3</sup>H]TCDD (Table 4, 20 nM [<sup>3</sup>H]TCDD column). In addition, several other mutant AhRs which did not bind [<sup>3</sup>H]TCDD in the above experiments (i.e., H285M, I332P, M334E, and A375L) were also examined to confirm their lack of ligand binding activity. The increased concentration of [<sup>3</sup>H]TCDD in these assays allowed detection of [<sup>3</sup>H]TCDD specific binding to the A375V mutant AhR (17% of wild-type AhR), and the reduced amount of specific binding to this mutant AhR likely still results from some ligand stripping by the Tween 80 washing steps. These secondary ligand binding experiments also confirmed the inability of the other mutant AhRs to specifically bind [<sup>3</sup>H]-TCDD. Taken together, these mAHR ligand binding assay results are now in agreement with the ligand-induced DNA binding data. In contrast to the above results, the E339A mutation is interesting in that it reduced AhR transformation and DNA binding by 50% with no significant decrease in ligand binding (although a trend of lower binding was observed). These results could suggest that the E339A mutation exerts a more specific effect on ligand-dependent transformation events.

Overall, the mutagenesis and functional analysis results are consistent with the proposed model of the mAHR LBD (mod\_HIF/ARNT) and confirm that the cavity modeled in the interior of the domain is indeed involved in ligand binding.

In fact, mutation of those residues that point outside the modeled LBD pocket (Arg282, Thr311, Glu339 and Lys350, shown in blue in Figure 4) to alanine do not affect AhR TCDD binding and, with the exception of Glu339, AhR transformation/DNA binding. It is conceivable that Glu339, whose side chain points out of the pocket toward the back of the  $\beta$ -sheet, is indeed not involved in ligand binding but could affect mAHR transformation by altering ligand-dependent effects on the interaction of AhR with partner proteins such as hsp90 or ARNT. This point deserves further in-depth analyses to elucidate the specific sites for protein-protein interaction on the LBD external surface.

In contrast, those residues whose mutation adversely affects TCDD and/or DNA binding point into the modeled cavity (His285, Cys327, Met334, Ala375, and Gln377, shown in purple in Figure 4) or lie in the helical connector that flanks the cavity (I332). Homology modeling of these mutants, performed on the basis of the same template structures, alignment, and modeling procedure used for the wild-type mAHR (see Materials and Methods section), indicated that only the I332P mutation has structural effects on the LBD. As expected, substitution of isoleucine with proline in that position destabilized the helical connector due to the loss of a hydrogen bond between the nitrogen (on Ile332) and oxygen (on Ala328) backbone atoms. The structural effects of this change were evidenced by performing a 1 ns molecular dynamics simulation (see Materials and Methods section) on the modeled I332P mutant. After about 300 ps of simulation the kink of the helix around the E329 position, with insertion of a turn, was evidenced by the DSSP analysis (58). Moreover, perturbation of the helix N-terminal capping as well as of the orientations of the side chains from the helix N-terminal to P332 was observed (Figure 7). One of the consequences of these structural modifications was the reduction in the internal cavity volume of about 100 Å<sup>3</sup>. It has to be noted that a MD simulation performed on the modeled wild-type mAHR LBD, with the same computational protocol as for I332P, indicated that, in that case, the tertiary and secondary structure of the domain remained stable throughout the simulation. It is therefore conceivable that the complete elimination of TCDD binding and TCDD-inducible AhR DNA binding observed for the I332P mutant is associated to the loss of LBD structural features required for ligand binding.

In contrast, analysis of the models generated for the point mutants in the His285, Cys327, Met334, Ala375, and Gln377 positions indicated that those mutations did not modify either the overall fold nor the secondary structure elements of the LBD. Also the side-chain conformations of residues lying in the modeled cavity and those in the surrounding of the mutated residue were unaffected in all cases. Therefore, the effects of those mutations on TCDD binding appear to be related to the removal of specific molecular requirements for ligand recognition.

Interesting mutations are those of alanine 375 to two residues (valine and leucine) that maintain its hydrophobicity but present increasingly longer side chains. Previous results identified the side-chain size in the Ala375 position of the mAHR as critical for ligand binding activity, since the valine substitution of this residue that is found naturally in the mouse AhRd allele and in the human AhR significantly reduced ligand binding (16,17,24). As expected, mutation of Ala375 to valine dramatically reduced [<sup>3</sup>H]TCDD specific binding, and ligand binding was eliminated when Ala375 was mutated to leucine (Table 4). The position of this residue is directly in the center of the modeled LBD cavity, and decreased TCDD binding must be associated with increased steric hindrance by the amino acid side chain. Thus, TCDD must bind relatively close to this residue within the cavity for it to reduce TCDD binding.

The His285 residue is also very interesting in that its mutation to alanine completely eliminated TCDD binding and TCDD-inducible AhR DNA binding (Table 4). It should be noted that the His285 side chain lies very close to that of Ala375 whose mutation to valine or leucine reduces ligand binding due to steric hindrance. The proximity of this side chain to Ala375 may support the hypothesis that His285 plays a key role in the interaction of the AhR LBD with TCDD. Moreover for histidine, which is unprotonated at the physiological pH particularly when it occurs in buried cavities and has the properties of a polar aromatic residue,  $\pi$ - $\pi$  interactions with aromatic partners are frequently observed in protein systems (64). It is therefore conceivable that His285 plays a role in stabilizing interactions with the AhR aromatic ligands. Binding of TCDD in the pocket may be particularly strengthened by the high polarizability of the electron distribution of this ligand along the prime molecular axis (65) that improves dispersion interactions. In addition, this molecular association may be stabilized by a substantial electrostatic component given the presence of lateral electron-withdrawing chlorine atoms in the TCDD that generates an electronic charge depletion in the central part of the molecule including the aromatic rings (66).

On the other hand, the reduction in TCDD binding to mAHRs containing the C327A, M334A, and Q377A mutations (Table 4) suggests a hypothesis that stabilization of the polar aromatic ligand could result from a network of weak interactions instead of a specific one. In fact, sulfur-arene interactions involving methionine and cysteine side chains, mainly due to dispersion forces between the sulfur atom and the  $\pi$  surface, are frequently recognized in proteins as well as weak N-H/ $\pi$  hydrogen bonds (67). The decreased contribution of each of these interactions to the TCDD stabilization, as a result of their mutation to alanine, may considerably weaken the binding. Strong support for this hypothesis is given by mutation of Met334 to glutamic acid that completely eliminated AhR ligand binding and ligand-dependent DNA binding. It is conceivable that the introduction of a side chain carrying a net negative charge in the molecular environment of the mAHR binding cavity may strongly perturb the existing network of weak interactions.

Finally, it is also informative to include our preliminary models of the mAHR LBD (23,26) in the analysis of the relationship between the mutations and their effect on AhR functionality. While some of the mutagenesis results are also consistent with the LBD features of both our previous models, some mutations produce effects that cannot be

rationalized on the basis of those structural proposals. In particular, Arg282 and Glu339, whose mutations to alanine had no effect on ligand binding and that accordingly reside on the domain surface in the HIF/ARNT model, lie in key positions within the domain modeled on the basis of the FixL template (23). Additionally, in the same model, His285, whose mutation to alanine strongly affects both TCDD and DNA binding, points outside the modeled cavity. Also, in the model based on Phy3 (26), due to some misalignment errors in the N-terminal region, the positions of two residues disagree with what is indicated by the present mutagenesis data: Arg282 was modeled on the inside of the cavity, while His285 was modeled on the external domain surface. These above considerations further support the higher reliability of the HIF/ARNT model proposed here and confirm that it provides us with a significantly improved model to study the AhR LBD features.

Not only is the information derived from mutagenesis experiments consistent with the structural proposal derived from the homology model based on the HIF-2 $\alpha$  and ARNT template structures, but the analysis of experimental results in the framework of the model also highlights a list of residues contributing to the particular binding affinity of this domain to aromatic compounds.

## CONCLUSIONS

The availability of new structural and functional information on PAS domains has allowed the development of an updated and reliable model of the mAHR PAS B domain by homology modeling techniques. While the overall fold features of the domain appeared to be reproduced independently of the choice of the template structure for modeling, comparison of different mono- and multitemplate models demonstrated the crucial influence of this choice on the secondary structure arrangement that determines the features of the mAHR binding cavity. NMR PAS structures of the HIF-2 $\alpha$  and ARNT proteins were selected as the most reliable templates due to their higher degree of sequence identity and similarity with the mAHR PAS B and to the functional similarities of the full-length proteins.

The effects of point mutations in selected key residue positions on both [<sup>3</sup>H]TCDD binding and TCDD-dependent transformation and DNA binding here analyzed confirmed the proposed structural features of the mAHR LBD and revealed some inconsistencies with previous models, thus highlighting the significant improvement given by the newly available PAS template structures. Our experimental results confirmed the role of the largest structural cavity identified in the modeled mAHR LBD as the site involved in ligand binding. In fact, while mutation of those residues that point outside the domain does not affect AhR TCDD binding, mutation of residues lying in this cavity reduces or eliminates TCDD binding and TCDD-inducible DNA binding. The identification of a buried cavity within the core of the mAHR PAS B domain with enough internal space available for ligands highlights the different functional role of this domain with respect to the HIF-2 $\alpha$  and ARNT PAS structures, whose domain interior is well packed. Moreover, our results are consistent with the proposed model being more representative of the ligand-bound form of the mAHR rather than the unliganded form of the AhR.

Some specific structural and chemical requirements for ligand binding were also highlighted by analyzing the mutagenesis results in the framework of the three-dimensional model. The dramatic effects of mutating a residue in the central helical connector (Ile332) to the helix-breaking proline demonstrated the role of this structural element in maintaining the overall fold of the domain and the topology of the cavity. The hypothesis that TCDD may bind relatively close to Ala375, in the center of the modeled LBD cavity, was supported by the adverse effect on TCDD specific binding and ligand-dependent DNA binding to mutants with increased steric hindrance in that position (A375V and A375L). These results also



provide an explanation for the lower ligand affinity of the AhR present in human and some mouse strains, since the presence of a valine residue in this position in the AhR in these species reduces the accessibility of the ligand within the LBD cavity. The closeness of the His285 side chain to that of Ala375 in the proposed model, along with its electronic characteristics that suggest its involvement in the TCDD stabilization, is in complete agreement with the observed loss of AhR TCDD binding and TCDD-inducible DNA binding activity due to its mutation to alanine. Moreover, partial reduction of TCDD and DNA binding associated with the mutation of some hydrophobic or polar residues in the boundary of the modeled cavity suggested that a network of weak interactions involving these residues is the molecular determinant for stabilization of TCDD binding within the LBD. The dramatic effects of introducing a charged residue (M334E mutant) within this environment confirmed this hypothesis.

In contrast to the results where the effects on ligand binding and ligand-inducible DNA binding activity are correlated and are most likely linked, a divergence in these functional activities was observed with the E339A mutation. Mutation of E339, whose side chain points outside the binding cavity toward the back of the  $\beta$ -sheet, resulted in reduced ligand-dependent AhR DNA binding but no significant decrease of ligand [ $^3\text{H}$ ]TCDD binding. These results suggest this region as a putative site for protein–protein interactions that could be involved in the ligand-dependent mAhR transformation. Further analysis of this possibility needs to be performed.

In conclusion, the agreement between this first set of site-directed mutagenesis experiments and modeling results not only confirms and validates the structural features of this improved homology model of the mAhR PAS B LBD but it also provides a framework for developing and testing further hypotheses on the key events involved in the mechanisms of ligand binding and ligand-dependent activation of the AhR.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Dr. Steve Safe (Texas A&M University) for the [ $^3\text{H}$ ]TCDD and TCDF, Dr. Oliver Hankinson (UCLA) and Dr. James P. Whitlock, Jr., for mAhR and mARNT expression vectors, and Dr. Kevin H. Gardner for providing us with the NMR structure of the ARNT PAS B domain prior to publication.

## REFERENCES

1. Schmidt JV, Bradfield CA. Ah receptor signaling pathways. *Annu. Rev. Cell DeV. Biol* 1996;12:55–89. [PubMed: 8970722]
2. Denison, MS.; Elferink, CF.; Phelan, D. The Ah receptor signal transduction pathway. In: Denison, MS.; Helferich, WG., editors. *Toxicant-Receptor Interactions in the Modulation of Signal Transduction and Gene Expression*. Philadelphia, PA: Taylor and Francis; 1998. p. 3-33.
3. Ma Q. Induction of CYP1A1. The AhR/DRE paradigm: Transcription, receptor regulation, and expanding biological roles. *Curr. Drug Metab* 2001;2:149–164. [PubMed: 11469723]
4. Denison, MS.; Seidel, SD.; Rogers, WJ.; Ziccardi, M.; Winter, GM.; Heath-Pagliuso, S. Natural and synthetic ligands for the Ah receptor. In: Puga, A.; Wallace, KB., editors. *Molecular Biology Approaches to Toxicology*. Philadelphia, PA: Taylor & Francis; 1998. p. 393-410.
5. Denison MS, Heath-Pagliuso S. The Ah receptor: a regulator of the biochemical and toxicological actions of structurally diverse chemicals. *Bull. EnViron. Contam. Toxicol* 1998;61:557–568. [PubMed: 9841714]

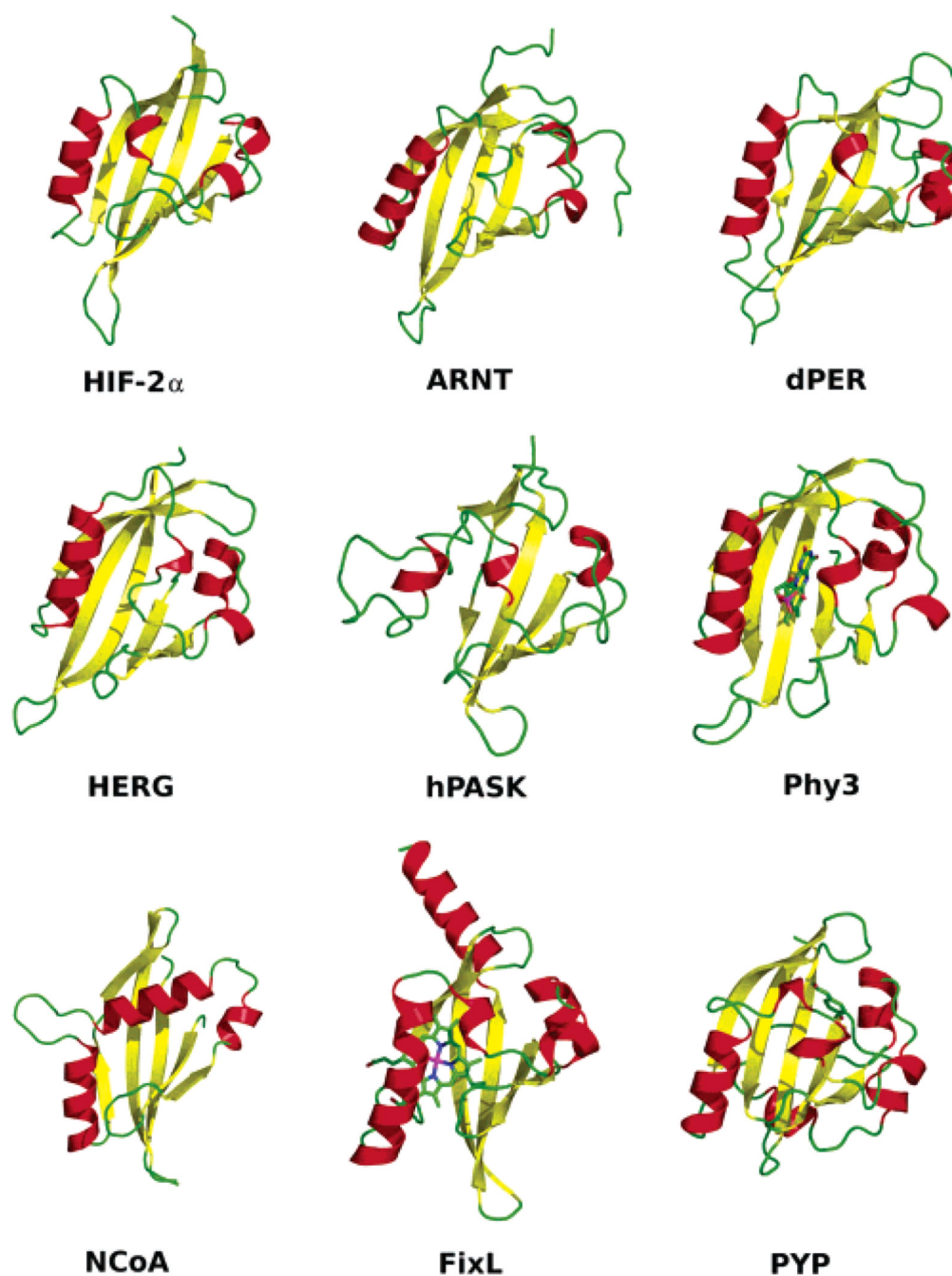
6. Denison MS, Pandini A, Nagy SR, Baldwin EP, Bonati L. Ligand binding and activation of the Ah receptor. *Chem.-Biol. Interact* 2002;141:3–24. [PubMed: 12213382]
7. Safe S. Polychlorinated biphenyls (PCBs), dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs), and related compounds: environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). *Crit. Rev. Toxicol* 1990;21:51–88. [PubMed: 2124811]
8. Petrusis JR, Perdew GH. The role of chaperone proteins in the aryl hydrocarbon receptor core complex. *Chem.-Biol. Interact* 2002;141:25–40. [PubMed: 12213383]
9. Henry EC, Gasiewicz TA. Agonist but not antagonist ligands induce conformational change in the mouse aryl hydrocarbon receptor as detected by parital proteolysis. *Mol. Pharmacol* 2003;63:392–400. [PubMed: 12527811]
10. Hord NG, Perdew GH. Physicochemical and immunochemical analysis of aryl hydrocarbon receptor nuclear translocator: characterization of two monoclonal antibodies to the aryl hydrocarbon receptor nuclear translocator. *Mol. Pharmacol* 1994;46:618–624. [PubMed: 7526149]
11. Hankinson O. The aryl hydrocarbon receptor complex. *Annu. Rev. Pharmacol. Toxicol* 1995;35:307–340. [PubMed: 7598497]
12. Whitlock JP Jr. Induction of cytochrome P4501A1. *Annu. Rev. Pharmacol. Toxicol* 1999;39:103–125. [PubMed: 10331078]
13. Coumailleau P, Poellinger L, Gustafsson J-A, Whitelaw ML. Definition of a minimal domain of the dioxin receptor that is associated with Hsp90 and maintains wild type ligand binding affinity and specificity. *J. Biol. Chem* 1995;270:25291–25300. [PubMed: 7559670]
14. Fukunaga BN, Probst MR, Reisz-Porszasz S, Hankinson O. Identification of functional domains of the aryl hydrocarbon receptor. *J. Biol. Chem* 1995;270:29270–29278. [PubMed: 7493958]
15. Perdew GH, Bradfield CA. Mapping the 90 kDa heat shock protein binding region of the Ah receptor. *Biochem. Mol. Biol. Int* 1996;39:589–593. [PubMed: 8828811]
16. Ema M, Ohe N, Suzuki M, Mimura J, Sogawa K, Ikawa S, Fujii-Kuriyama Y. Dioxin binding activities of polymorphic forms of mouse and human aryl hydrocarbon receptors. *J. Biol. Chem* 1994;269:27337–27343. [PubMed: 7961644]
17. Poland A, Palen D, Glover E. Analysis of the four alleles of the murine aryl hydrocarbon receptor. *Mol. Pharmacol* 1994;46:915–921. [PubMed: 7969080]
18. McGuire J, Okamoto K, Whitelaw ML, Tanaka H, Poellinger L. Definition of a dioxin receptor mutant that is a constitutive activator of transcription: delineation of overlapping repression and ligand binding functions within the PAS domain. *J. Biol. Chem* 2001;276:41841–41849. [PubMed: 11551926]
19. Borgstahl GE, Williams DR, Getzoff ED. 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* 1995;34:6278–6287. [PubMed: 7756254]
20. Dux P, Rubinstenn G, Vuister GW, Boelens R, Mulder FA, Hard K, Hoff WD, Kroon AR, Crielaard W, Hellingwerf KJ, Kaptein R. Solution structure and backbone dynamics of the photoactive yellow protein. *Biochemistry* 1998;37:12689–12699. [PubMed: 9737845]
21. Cabral JHM, Lee A, Cohen SL, Chait BT, Li M, Mackinnon R. Crystal structure and functional analysis of the HERG potassium channel N terminus: a eucaryotic PAS domain. *Cell* 1998;95:649–655. [PubMed: 9845367]
22. Gong W, Hao B, Mansy SS, Gonzalez G, Gilles-Gonzalez MA, Chan MK. Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. *Proc. Natl. Acad. Sci. U.S.A* 1998;95:15177–15182. [PubMed: 9860942]
23. Procopio M, Lahm A, Tramontano A, Bonati L, Pitea D. A model for recognition of polychlorinated dibenzo-*p*-dioxins by the aryl hydrocarbon receptor. *Eur. J. Biochem* 2002;269:13–18. [PubMed: 11784293]
24. Murray IA, Reen RK, Leathery N, Ramadoss P, Bonati L, Gonzalez FJ, Peters JM, Perdew GH. Evidence that ligand binding is a key determinant of Ah receptor-mediated transcriptional activity. *Arch. Biochem. Biophys* 2005;442:59–71. [PubMed: 16137638]

25. Crosson S, Moffat K. Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction. *Proc. Natl. Acad. Sci. U.S.A* 2001;98:2995–3000. [PubMed: 11248020]
26. Bonati L, Pandini A, Pitea D, Song Y, Denison MS. New insights into the structure of the mAhR ligand binding domain. *Organohalogen Compd* 2002;59:429–432.
27. Amezcua C, Harper S, Rutter J, Gardner K. Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation. *Structure (Cambridge)* 2002;10:1349–1361.
28. Erbel PJ, Card PB, Karakuzu O, Bruick RK, Gardner KH. Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor. *Proc. Natl. Acad. Sci. U.S.A* 2003;100:15504–15509. [PubMed: 14668441]
29. Razeto A, Ramakrishnan V, Litterst CM, Giller K, Griesinger C, Carlomagno T, Lakomek N, Heimburg T, Lodrini M, Pfitzner E, Becker S. Structure of the NCoA-1/Src-1 PAS-B domain bound to the LXXLL motif of the Stat6 transactivation domain. *J. Mol. Biol* 2004;336:319–329. [PubMed: 14757047]
30. Yildiz O, Doi M, Yujnovsky I, Cardone L, Berndt A, Hennig S, Schulze S, Urbanke C, Sassone-Corsi P, Wolf E. Crystal structure and interactions of the PAS repeat region of the *Drosophila* clock protein period. *Mol. Cell* 2005;17:69–82. [PubMed: 15629718]
31. Card PB, Erbel PJ, Gardner KH. Structural basis of Arnt PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization. *J. Mol. Biol* 2005;353:664–678. [PubMed: 16181639]
32. Kewley RJ, Whitelaw ML, Chapman-Smith A. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int. J. Biochem. Cell Biol* 2004;36:189–204. [PubMed: 14643885]
33. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242. [PubMed: 10592235]
34. Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng* 1996;9:1063–1065. [PubMed: 8961360]
35. Getzoff ED, Gutwin KN, Genick UK. Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation. *Nat. Struct. Biol* 2003;10:663–668. [PubMed: 12872160]
36. Altschul SF, Madden TL, Scha ffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402. [PubMed: 9254694]
37. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol* 1993;234:779–815. [PubMed: 8254673]
38. Marti-Renom MA, Stuart A, Fiser A, Sa nchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct* 2000;29:291–325. [PubMed: 10940251]
39. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Sci* 2000;9:1753–1773. [PubMed: 11045621]
40. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 1998;102:3586–3616.
41. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol* 1993;233:123–138. [PubMed: 8377180]
42. Holm L, Park J. DaliLite workbench for protein structure comparison. *Bioinformatics* 2000;16:566–567. [PubMed: 10980157]
43. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]

44. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol* 1999;292:195–202. [PubMed: 10493868]
45. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr* 1993;26:283–291.
46. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins* 1993;17:355–362. [PubMed: 8108378]
47. Sanchez R, Sali A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci. U.S.A* 1998;95:13597–13602. [PubMed: 9811845]
48. DeLano, WL. The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific; 2002.
49. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 2006;34:W116–W118. [PubMed: 16844972]
50. Connolly ML. Analytical molecular surface calculation. *J. Appl. Crystallogr* 1983;16:548–558.
51. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun* 1995;91:43–56.
52. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model* 2001;7:306–317.
53. Darden T, York D, Pedersen L. Particle mesh Ewald—an n-log(n) method for Ewald sums in large systems. *J. Chem. Phys* 1993;98:10089–10092.
54. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys* 1977;23:327–341.
55. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem* 1997;18:1463–1472.
56. Rushing SR, Denison MS. The silencing mediator of retinoic acid and thyroid hormone receptors can interact with the aryl hydrocarbon receptor but fails to repress Ah receptor-dependent gene expression. *Arch. Biochem. Biophys* 2002;403:189–201. [PubMed: 12139968]
57. Denison, MS.; Rogers, JM.; Rushing, SR.; Jones, CL.; Tetangco, SC.; Heath-Pagliuso, S. Analysis of the Ah receptor signal transduction pathway. In: Maines, M.; Costa, LG.; Reed, DJ.; Sassa, S.; Sipes, IG., editors. *Current Protocols in Toxicology*. New York: John Wiley and Sons; 2002. p. 4.8.1–4.8.45.
58. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–2637. [PubMed: 6667333]
59. Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev* 1999;63:479–506. [PubMed: 10357859]
60. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* 1986;5:823–826. [PubMed: 3709526]
61. Sali A, Pottterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. *Proteins* 1995;23:318–326. [PubMed: 8710825]
62. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 1998;7:1884–1897. [PubMed: 9761470]
63. Harper PA, Giannone JV, Okey AB, Denison MS. In vitro transformation of the human Ah receptor and its binding to a dioxin response element. *Mol. Pharmacol* 1992;42:603–612. [PubMed: 1331752]
64. Cauet E, Rooman M, Wintjens R, Lievin J, Biot C. Histidine-aromatic interactions in proteins and protein-ligand complexes: quantum chemical study of X-ray and model structures. *J. Chem. Theory Comput* 2005;1:472–483.
65. Fraschini E, Bonati L, Pitea D. Molecular polar-izability as a tool for understanding the binding properties of polychlorinated dibenzo-*p*-dioxins: definition of a reliable computational procedure. *J. Phys. Chem* 1996;100:10564–10569.

66. Bonati L, Fraschini E, Lasagni M, Palma Modoni E, Pitea D. A hypothesis on the mechanism of PCDD biological activity based on molecular electrostatic potential modeling. Part 2. *J. Mol. Struct. (Theochem)* 1995;340:83–95.
67. Meyer EA, Castellano RK, Diederich F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed* 2003;42:1210–1250.

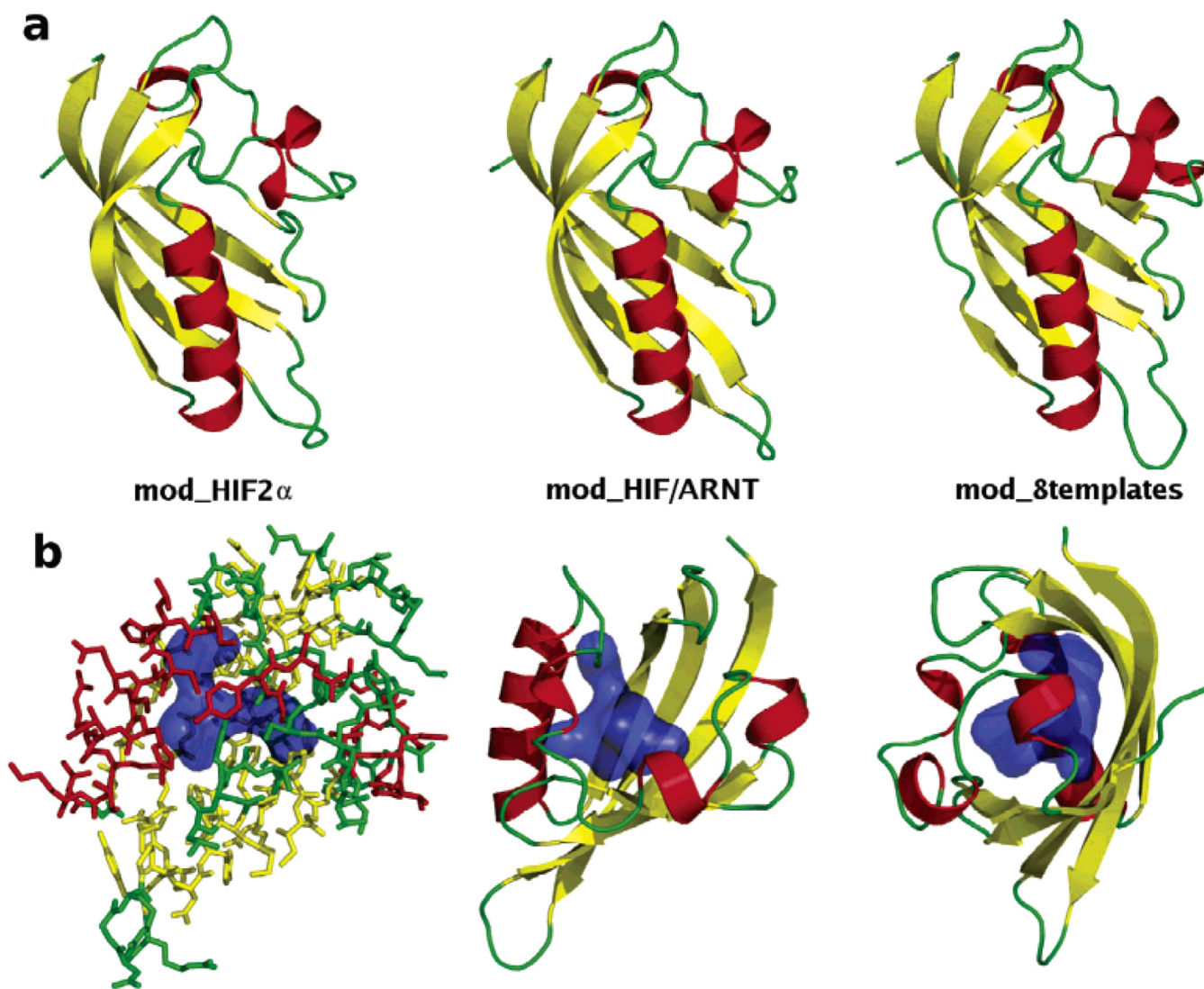


**FIGURE 1.**

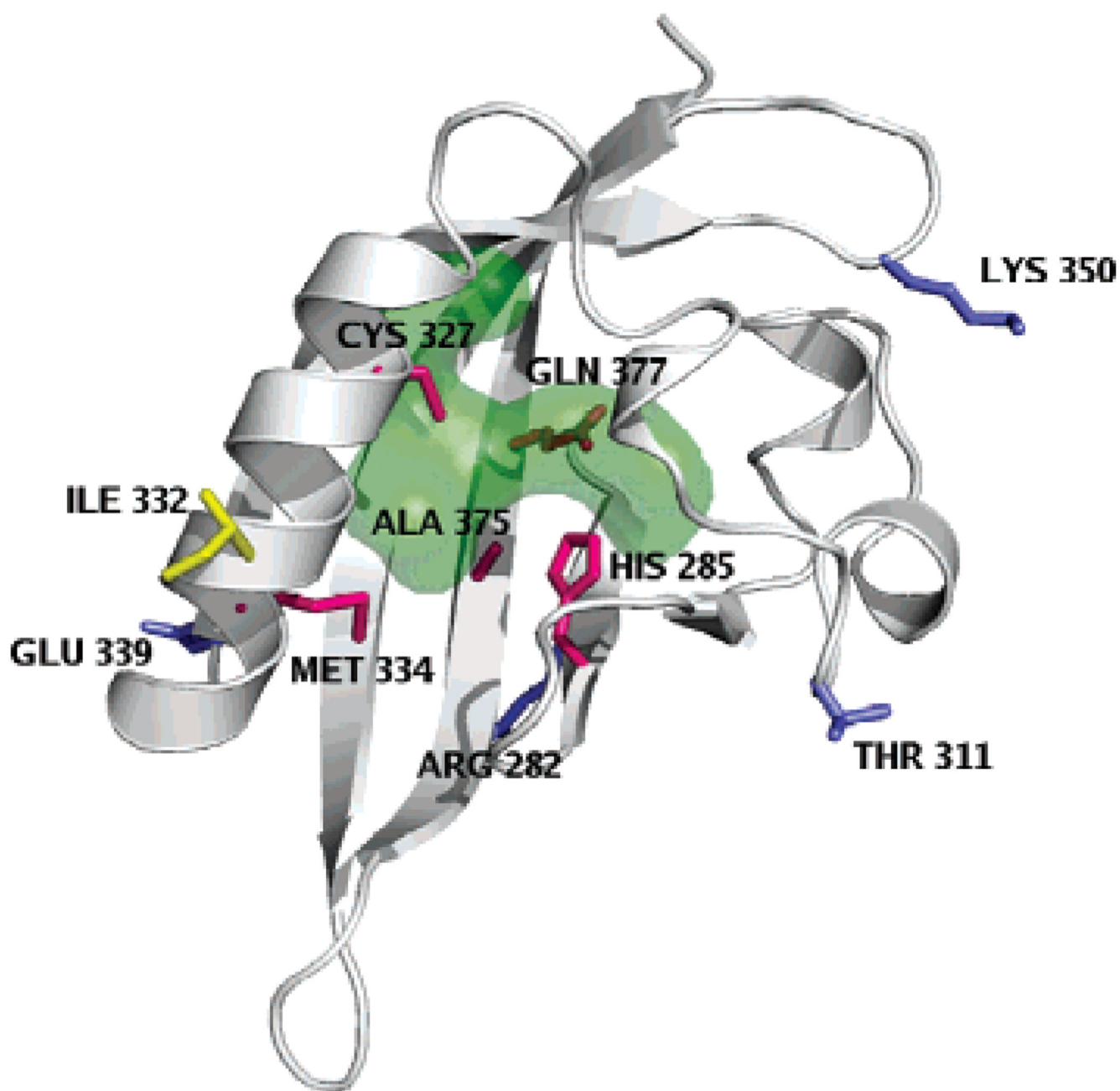
Cartoon representation of the PAS domain structures included in the analysis with their cofactors: HIF-2 $\alpha$  (PDB ID 1P97), ARNT (PDB ID 1X00), dPER (PDB ID 1WA9), HERG (PDB ID 1BYW), hPASK (PDB ID 1LL8), Phy3 (PDB ID 1G28), NCoA (PDB ID 1OJ5), FixL (PDB ID 1DRM), and PYP (PDB ID 1NWZ). Secondary structure attribution was according to the Kabsch and Sander method (58). For FixL and PYP the additional extra-domain elements included in the X-ray structures are also included (a long helix at the C-terminus for FixL and an N-terminal bundle of two helices for PYP).



panel a). Red arrows indicate the boundary residues of the mAhR cavity that have considerably smaller side chains than the corresponding ones in HIF-2 $\alpha$ .

**FIGURE 3.**

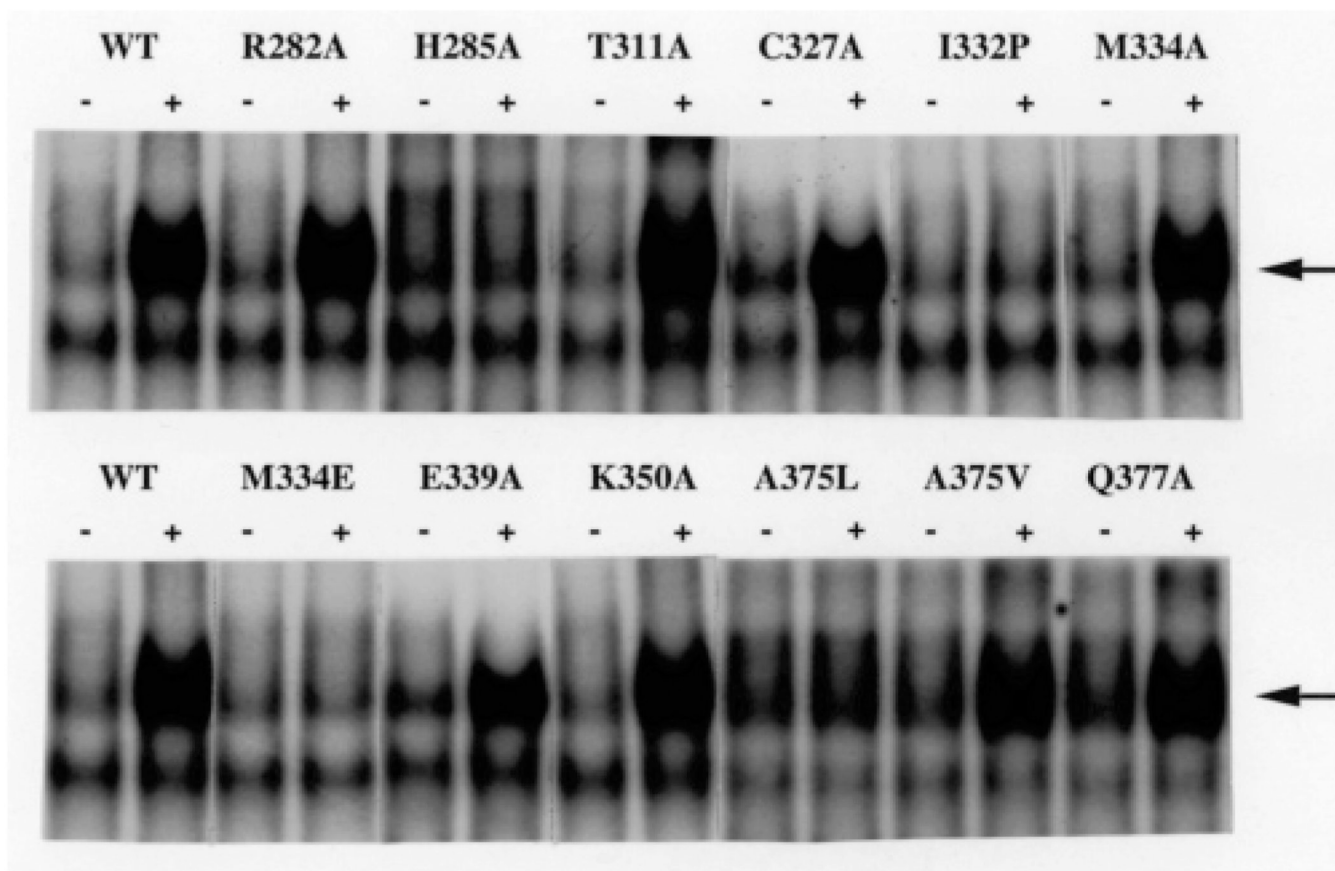
(a) Cartoon representation of the three modeled structures of the mAhR LBD. (b) Stick and cartoon representations of the model based on the HIF-2 $\alpha$  and ARNT template structures (**mod\_HIF/ARNT**) in different orientations, with the molecular surface (in blue) including the available volume in the cavity identified by CASTp. Secondary structure attribution was according to the method of Kabsch and Sander (58).



**FIGURE 4.**

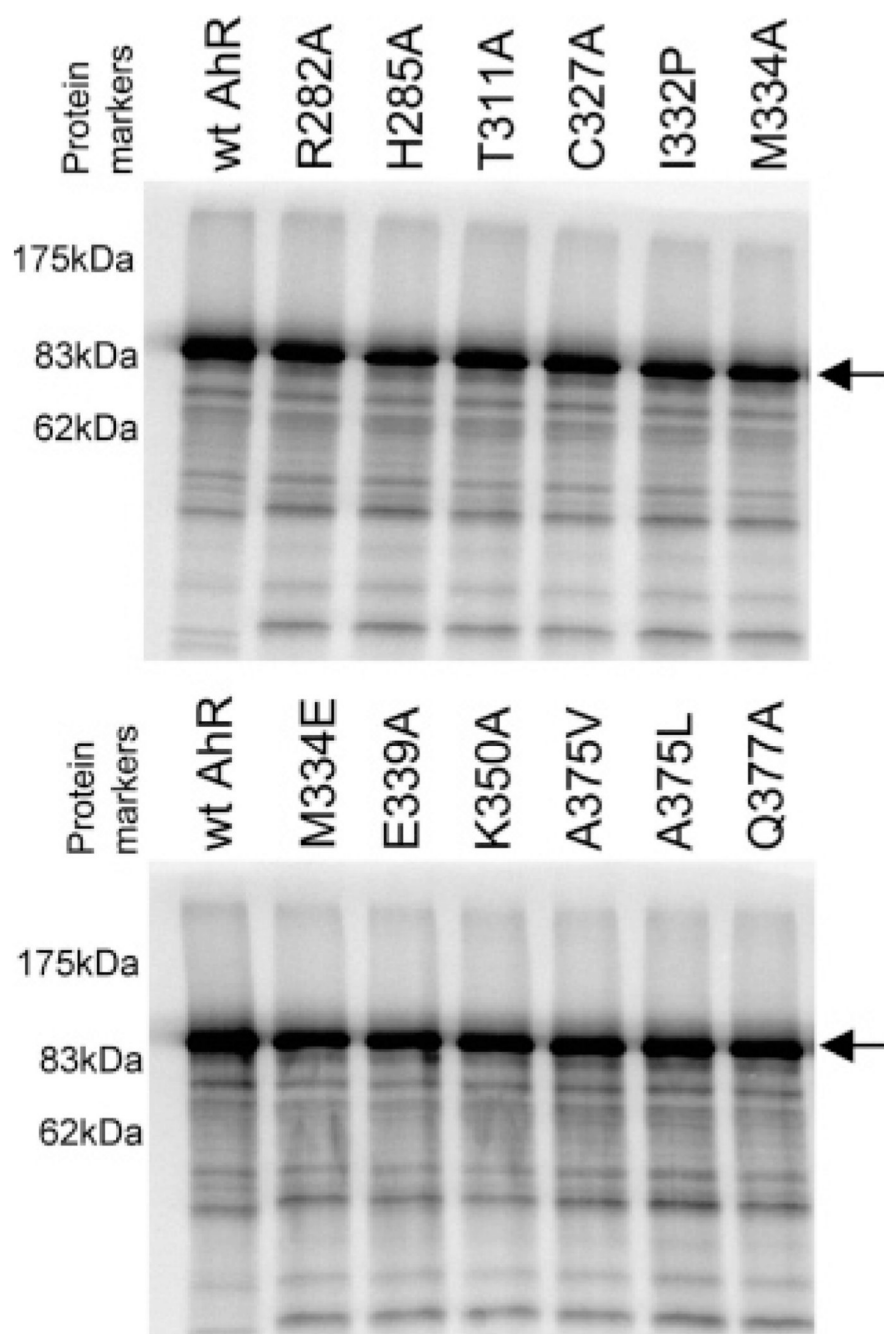
Cartoon representation of the modeled mAhR LBD (mod\_HIF/ARNT) showing selected residues that were mutated. Residues with side chains pointing outside the modeled LBD are shown in blue; boundary residues of the cavity with side chains pointing inside it are shown in purple; Ile332, which is expected to have a structural role, is shown in yellow. The molecular surface (in green) including the available volume in the cavity identified by CASTp is shown.





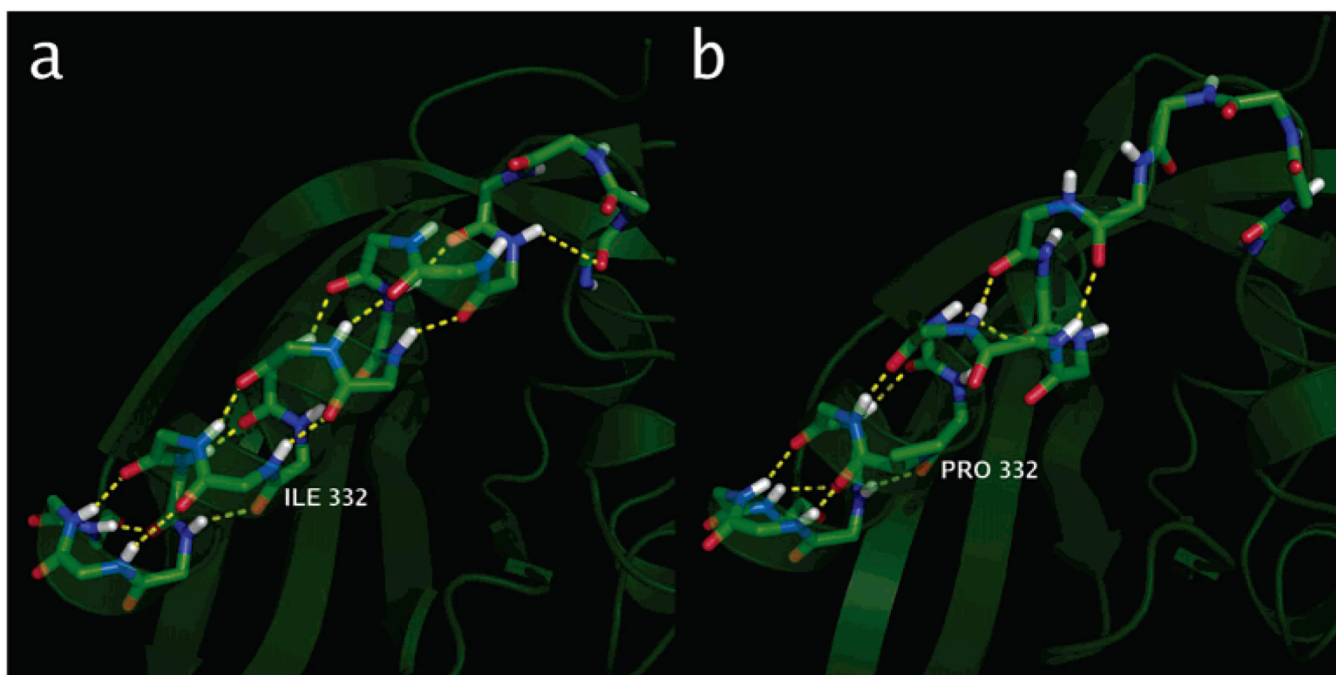
**FIGURE 5.**

Effect of mutation of selected residues within the mAhR LBD on TCDD-dependent AhR DNA binding. *In vitro* expressed wild-type or mutant AhR and wt ARNT were incubated with TCDD, and inducible AhR–ARNT–DRE complex formation was determined by gel retardation analysis as described under Materials and Methods. The positions of the induced AhR–ARNT–DRE complex are indicated by an arrow. Quantitation of the amount of the TCDD–AhR–ARNT–DRE complex was determined by phosphorimager analysis, and the results of multiple receptor preparations and gel retardation analyses ( $n \geq 3$ ) are presented in Table 4.



**FIGURE 6.**

Expression levels of in vitro synthesized wild-type and mutant AhRs. <sup>35</sup>S-Labeled wild-type and mutant AhRs were synthesized in vitro, denatured, and resolved by SDS-polyacrylamide gel electrophoresis and autoradiography of the dried gels as described in Materials and Methods. An arrow shows the bands of the AhR.



**FIGURE 7.** Stick representation of the helical connector main chain in the modeled mAhR LBD subjected to 1 ns MD simulation: (a) wild-type mAhR; (b) I332P mutant. The coloring scheme is according to the atom types. Hydrogen bonds are highlighted in yellow. Cartoon representation of the domain is shown in transparency.

**Table 1**

## PDB Identification Numbers of the PAS Domain Structures Employed

| PDB ID            | protein   | structure method    | ref |
|-------------------|---|---------------------|-----|
| 1P97              | human hypoxia-inducible factor, HIF-2 $\alpha$  | NMR (26 structures) | 28  |
| 1X00              | human aryl hydrocarbon receptor nuclear translocator, ARNT                                | NMR (20 structures) | 31  |
| 1WA9 <sup>a</sup> | <i>Drosophila</i> clock protein PERIOD, dPER  | X-ray diffraction   | 30  |
| 1BYW              | human erg potassium channel, HERG   | X-ray diffraction   | 21  |
| 1LL8              | human PAS kinase, hPASK   | NMR (20 structures) | 27  |
| 1G28 <sup>b</sup> | <i>Adiantum capillus-veneris</i> chimeric phytochrome/<br>phototropin photoreceptor, Phy3 | X-ray diffraction   | 25  |
| 1OJ5              | mouse steroid receptor coactivator 1A, NCoA   | X-ray diffraction   | 29  |
| 1DRM              | <i>Bradyrhizobium japonicum</i> sensor protein FixI, FixL                                 | X-ray diffraction   | 22  |
| 1NWZ              | <i>Ectothiorhodospira halophila</i> photoactive yellow protein, PYP                       | X-ray diffraction   | 35  |

<sup>a</sup>Chain A was employed and only the region encompassing the PAS B domain was included in the alignment. The location of this region was derived by structural alignment with HIF-2 $\alpha$ .

<sup>b</sup>Chain A was employed.

**Table 2**Sequence Identity and Similarity between PAS Domain Templates and the mAhR<sup>a</sup>

| PAS domain     | identity (%) | similarity (%)<br>(BLOSUM 62) | no. of aligned positions<br>(including gaps) |
|----------------|--------------|-------------------------------|--|
| HIF-2 $\alpha$ | 31.1         | 62.1                          | 103  |
| ARNT           | 21.2         | 53.8                          | 104  |
| dPER           | 16.7         | 45.4                          | 108  |
| HERG           | 12.0         | 41.7                          | 108  |
| hPASK          | 11.9         | 41.3                          | 109  |
| Phy3           | 11.1         | 40.7                          | 108  |
| NCoA           | 10.3         | 38.8                          | 116  |
| FixL           | 10.3         | 36.4                          | 107  |
| PYP            | 6.7          | 32.4                          | 105  |

<sup>a</sup>See text for more details.



**Table 3**PROSA  $z$ -Score Values for the Three AhR PAS B Models

| model              | PROSA $z$ -score  |      |            |
|--------------------|-------------------|------|------------|
|                    | mean (100 models) | SD   | best model |
| mod_HIF-2 $\alpha$ | -4.00             | 0.20 | -4.03      |
| mod_HIF/ARNT       | -4.53             | 0.26 | -4.56      |
| mod_8templates     | -4.71             | 0.24 | -4.58      |

**Table 4**

Effect of Mutagenesis of Individual Amino Acids within the mAHR LBD on [<sup>3</sup>H]TCDD Specific Binding and TCDD-Inducible AhR DNA Binding

| mutant           | DRE binding<br>(% of wt AhR) | [ <sup>3</sup> H]TCDD specific binding      |  |
|------------------|------------------------------|---|--|
|                  |                              | 2 nM [ <sup>3</sup> H]TCDD<br>(% of wt AhR) | 20 nM [ <sup>3</sup> H]TCDD<br>(% of wt AhR) |
| wild-type<br>AhR | 100                          | 100 ± 19 <sup>a</sup>                       | 100 ± 3                                      |
| R282A            | 125 ± 22                     | 76 ± 2                                      | ND <sup>c</sup>                              |
| H285A            | -2.3 ± 4 <sup>b</sup>        | 4 ± 7 <sup>b</sup>                          | 3 ± 5 <sup>b</sup>                           |
| T311A            | 97 ± 21                      | 91 ± 15                                     | ND   |
| C327A            | 82 ± 11 <sup>b</sup>         | 59 ± 3 <sup>b</sup>                         | ND   |
| I332P            | -10 ± 10 <sup>b</sup>        | -4 ± 2 <sup>b</sup>                         | 2 ± 1 <sup>b</sup>                           |
| M334A            | 70 ± 9 <sup>b</sup>          | 57 ± 7 <sup>b</sup>                         | ND   |
| M334E            | 3 ± 6 <sup>b</sup>           | -16 ± 12 <sup>b</sup>                       | 3 ± 1 <sup>b</sup>                           |
| E339A            | 53 ± 8 <sup>b</sup>          | 75 ± 12                                     | ND   |
| K350A            | 95 ± 11                      | 94 ± 4                                      | ND   |
| A375V            | 42 ± 12 <sup>b</sup>         | -6 ± 7 <sup>b</sup>                         | 17 ± 1 <sup>b</sup>                          |
| A375L            | 2.8 ± 8 <sup>b</sup>         | -2 ± 17 <sup>b</sup>                        | 1 ± 2 <sup>b</sup>                           |
| Q377A            | 52 ± 3 <sup>b</sup>          | 31 ± 4 <sup>b</sup>                         | ND   |

<sup>a</sup>Values represent the mean ± SD of three to five individual analyses.

<sup>b</sup>Values are significantly different from wild-type mAHR at  $P \leq 0.05$  as determined by Student's *t*-test.

<sup>c</sup>Not determined.