Published in final edited form as:

J Med Chem. 2009 September 24; 52(18): 5635–5641. doi:10.1021/jm900199u.

Modeling of the Aryl Hydrocarbon Receptor (AhR) ligand binding domain and its utility in virtual ligand screening to predict new AhR ligands

William Bisson[#], Daniel Koch[#], Edmond O'Donnell, Sammy M. Khalil, Nancy Kerkvliet, Robert Tanguay, Ruben Abagyan, and Siva Kumar Kolluri^{*}

Cancer Biology Laboratory, Department of Environmental and Molecular Toxicology, Environmental Health Sciences Center, Oregon State University, Corvallis, OR 97331

[‡]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037

Abstract

The Aryl Hydrocarbon Receptor (AhR) is a ligand-activated transcription factor; the AhR Per-AhR/Arnt-Sim (PAS) domain binds ligands. We developed homology models of the AhR PAS domain to characterize previously observed intra- and inter-species differences in ligand binding using Molecular Docking. *In silico* structure-based virtual ligand screening using our model resulted in the identification of pinocembrin and 5-hydroxy-7-methoxyflavone, which promoted nuclear translocation and transcriptional activation of AhR and AhR-dependent induction of endogenous target genes.

The aryl hydrocarbon receptor (AhR) is a ligand activated member of the basic-helix-loop-helix (bHLH) family of transcription factors.^{1–3} The AhR is activated by a variety of compounds, both synthetic and natural, including halogenated aromatic hydrocarbons such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), and mediates their biological activity.^{1–3} The AhR is a cytosolic transcription factor bound to several co-chaperones. Upon ligand binding, the AhR translocates from the cytoplasm to the nucleus and regulates genes, including several drug metabolizing enzymes that can influence the therapeutic activity of a number of compounds.^{1–3} The AhR regulates proliferation and differentiation of cells.^{4–6} The AhR also induces immunosuppressive regulatory T cells with therapeutic implications in hyperimmune disorders.^{7–9} The PAS (Per-AhR/Arnt-Sim) domain of the AhR is the ligand binding domain (LBD).^{1–3} In the present study, homology models of the AhR-LBD were built to study the inter and intra-species differences in ligand binding and to identify AhR ligands by virtual ligand screening.

To construct the homology models, a multiple sequence alignment of the AhR-LBD domain from mouse, human, rat, guinea pig, rabbit and zebrafish (Figure 1a) was performed. The sequence alignment produced sequence identities of greater than 85% between the different species, with the exception of zebrafish isoform 2, which shared a consensus sequence identity of approximately 70% (Table 1). The AhR-LBD of mouse and rat had the highest sequence identity at nearly 98% (Table 1), suggesting a common evolutionary path and a recent species splitting as depicted by the evolutionary tree in Figure 1b. The mouse AhR-

^{*}To whom correspondence should be addressed: Cancer Biology Laboratory, Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR 97331, USA Phone: 001-541-737-1799, Siva.Kolluri@oregonstate.edu.

[#]Equally contributed to this work

Supporting Information is available on Journal of Medicinal Chemistry web site.

LBD residues His285, Cys327, Ile332, Met334, Ala375 and Gln377, which have been characterized by mutagenesis studies and are known to influence TCDD binding,¹⁰ are highlighted in red in Figure 1a. These residues are well conserved with the exception of Met337 in guinea pig and Val381 in human AhR. These amino acid residues are replaced with Ile and Ala in all other species (Figure 1a).

Several PAS domain structures are currently available in the Protein Data Bank (Pdb). The sequence similarity between the structures of the PAS domain containing-protein family is low; however, within the family there is a high structural conservation of the α and β folds. We selected the PAS domain structure of hypoxia-inducible factor 2 α (Hif-2 α) (pdb 1p97)¹¹ for generating a homology model of the mouse AhR-LBD based on its closest similarity in this region. The model was built using the primary amino acid sequence [241–384] of the AhR-PASB domain, producing a sequence identity of 30% from the alignment with Hif-2 α . The backbone geometry was preserved while loops were rebuilt using a database search for matching loop geometries.¹² The side chain conformations of amino acids different in the alignment were globally optimized using the biased probability monte carlo procedure.¹³ The human, rat, zebrafish 1 and 2 AhR-LBD homology models were built starting from the mouse model and using the respective sequences of the other species homologous to that of the mouse primary sequence [241–384].

Each of the AhR ligand binding domains were built using the homology modeling protocol described in detail by Cardozo et al.¹⁴ The initial model was built by using the template backbone and side-chains in their most likely conformation. Next, the side chains were globally optimized using the modified ECEPP/3 energy function described by Abagyan et al.¹⁵ This energy function includes the following energy terms: van der Waals, hydrogen bonding, torsion energy, distance dependent electrostatic term, disulfide bond constraints, side-chain entropy, and the implicit solvent accessible surface dependent solvation term.¹⁴ After the side chain placement the whole model was energy optimized using all the above energy terms but with strong positional restraints. The refined homology models were then geometrically corrected such that all of the residues were in sterically allowed regions of the Ramachandran plot. Following development and refinement of the model, we first analyzed the mouse AhR-LBD homology model to characterize the ligand binding pocket. Using ICM Pocket Finder (Molsoft ICM v3.5-1p),¹⁶ a binding pocket was identified (volume of 201.5 Å³, area of 178.3 Å², and a radius of 3.637 Å) in the region of the protein surrounded by β -sheets H β and I β (Ala375 and Gln377) at the bottom, by the helix-connector F α (Cys327 and Met334) on the left, by A β on the right (His285), and at the top by the loop connecting A β and B β (Figure 2a). In the helix connector F α , the backbone NH of Ile332 established an intra H-bond with the backbone carbonyl oxygen of Ala328, keeping the helical secondary structure stable. Replacing Ile332 with a Proline in our mouse AhR-LBD homology model resulted in a loss of intra H-bond interactions. These results suggested that the decrease of TCDD binding in the I332P mutant¹⁰ is due to an indirect effect caused by the disruption of the F α helix connector in the vicinity of the binding pocket region. Thus, our AhR-LBD homology model provides a structural explanation for the decreased TCDD binding observed with the I332P substitution in the AhR.¹⁰

We should note that while other groups have attempted modeling of the AhR-LBD,^{10,17} there are noteworthy differences between these models and our characterized binding pocket. The ligand binding pocket in our model is more compact, surrounded mainly by the residues Cys327, Gln377, His285, Met334 and Ala375. The side chains of these amino acids are all pointing towards the binding pocket, indicating structurally their direct involvement in TCDD binding (Figure 2a). The model by Pandini et al.¹⁰ is more spread, involving the region between residues Gln377 and His285 surrounded by D α and E α helices in addition to the area characterized in our model. Similar to Pandini et al., we observed that the side chain

of residues Thr311 and Lys350 point outside the modeled LBD and do not affect AhR ligand binding.¹⁰ The side chain of residues Ile319 and His320 also pointed towards the pocket cavity, confirming their key role in ligand binding specificity.¹⁷

To study TCDD binding capacity across species and to provide a structural explanation for the biological data reported in the literature,¹⁻³ TCDD was docked into mouse, human and zebrafish isoform 2 AhR-LBD models. TCDD docked similarly in all three species binding pockets (Figure 2b–d) and the binding energy calculations (see methodology) were in agreement with previous inter-species biological characterizations.¹⁻³ Lower binding energy values indicate a higher binding affinity of a particular ligand in our model. The values obtained for TCDD binding in mouse and zebrafish isoform 2 were -3.796 kcal/mol and -3.972 kcal/mol, respectively, confirming that TCDD has similarly high binding affinity in mice and zebrafish.¹⁸ In contrast, the value obtained with the human model was -2.33 kcal/mol, which was significantly higher than for the mouse and zebrafish models, indicating a decreased binding affinity compared with the other two species investigated. TCDD docking orientation was similar in all three species. Further, the residues that were important for ligand binding and facing the ligand pocket cavity were conserved in mouse and zebrafish isoform 2. In the human AhR, one of the residues, corresponding to mouse Ala375 was replaced by a Valine; the presence of this bulkier residue would likely sterically impair TCDD binding ability.

The mouse AhR-LBD homology model was used to structurally characterize some of the intraspecies experimental biological data available for the AhR. The C57BL/6J mouse strain exhibits a 10-fold higher TCDD susceptibility compared to the DBA/2 mouse strain.¹⁻³ Both the mouse DBA/2 AhRLBD and human AhR-LBD share the A375V substitution (Figure 1a). Hence, the effect of this amino acid substitution on TCDD docking was analyzed in our mouse homology model by mutating A375 into a Valine or a Leucine. In the WT and A375V model TCDD was able to dock into the binding pocket, whereas in the A375L model TCDD failed to dock completely. Thus, the residue at position 375 located in the middle of the ligand binding area appears to directly impact TCDD binding. The replacement of the amino acid Alanine with bulkier residues like Valine (as present in mouse DBA/2 and human sequences) or Leucine decreased the volume of the binding pocket ($195 \text{ \AA}^3$ (WT), $186.1 \text{ \AA}^3$ (A375V) and $148.6 \text{ \AA}^3$ (A375L)) (Figure 2e) with the consequence of partially (Valine) or totally (Leucine) sterically blocking TCDD binding. By changing the tautomerization of one of the key residue for binding His285 in both WT ($194.4 \text{ \AA}^3$) and the A375V mutant ($185.9 \text{ \AA}^3$) models and by re-docking TCDD, the binding energy calculations generated values of -3.617 kcal/mol and -3.413 kcal/mol, respectively. Decreased volume of the binding pocket might result in weaker binding energy that contributes to decreased affinity of TCDD to AhR from human or mouse DBA/2 strain. Indeed, substitution of Alanine at 375 with Leucine has been reported to affect TCDD binding.¹⁰

We next used homology modeling to study AhR in zebrafish. Interestingly, zebrafish have three AhR isoforms, namely, AhR1a, AhR1b and AhR2, with AhR2 being the dominant isoform that binds TCDD, whereas AhR1a (zfAhR1a) exhibits impaired TCDD binding.^{18,19} The sequence identity between zfAhR1a and zfAhR2-LBD is 65%. The amino acid residues His285, Ala375 and Gln377 (Figure 2f), which are highly conserved in most species (Figure 1a) and contribute to ligand binding are present in zfAhR2; however, in zfAhR1a, these amino acids are substituted with Threonine, Histidine, and Tyrosine, respectively (Figure 2g). Thus, the difference between zfAhR1a and zfAhR2 in TCDD binding may be attributed to these dissimilarities in the ligand binding pocket region. To understand the differences in the ligand binding pockets of zfAhR1a and zfAhR2, homology models for the zfAhR1a and zfAhR2-LBD were built and energetically minimized using the

ICM method. TCDD docking was attempted with both zebrafish AhR isoforms models; however, TCDD was able to dock only into zfAhR2-LBD. By analyzing the binding pockets of zfAhR1a and zfAhR2 identified with ICM Pocket Finder,¹⁶ (Figure 2f and 2g) zfAhR1a was found to have a decreased binding area that may contribute to its lower affinity for TCDD. Substitution of zfAhR2 residues Gln388 and His296 with a Histidine and Tyrosine respectively (as present in zfAhR1a) modified the type of electrostatic interactions established between the protein and TCDD, resulting in decreased binding affinity. Thus, these specific amino acid replacements (Ala386Thr, Gln388His and His296Tyr) in the zfAhR1a-LBD appear to play a direct role in the considerable decrease of TCDD binding by modifying the electrostatic environment of the binding pocket region and the volume of the binding pocket itself. Indeed, electrostatic and van der Waals forces are the driving force in the total binding energy of many halogenated aromatic hydrocarbons towards the AhR.²⁰

In addition to characterization of the prototypical AhR ligand TCDD, we also used our homology models to study the binding of structurally diverse molecules to the AhR. Recently, formylindolo [3,2-b]carbazole (FICZ) and 2-(1'H-indole-3'-carbonyl)-thiazole-4-carboxylic acid methyl ester (ITE) have been described as potent AhR agonists both *in vitro* and *in vivo*.^{8,9} These two ligands were docked into our human and mouse AhR-LBD models to investigate any inter-species differences in their respective binding modes. The two molecules docked into both the human and mouse binding pockets in a similar orientation (Figure 3a–d), establishing H-bond interactions with protein residues facing the binding cavity. FICZ made two H-bonds between the formyl carbonyl oxygen and the side chain of Ser359 (mouse) and Ser 365 (human) and between the nitrogen of the carbazole group and the side chain of Gln377 (mouse) and Gln383 (human) (Figure 3b and 3d). On the other hand, ITE made a single H-bond between its 3'-carbonyl group and the side chain of Ser359 (mouse) and Ser365 (human) (Figure 3a and 3c). The binding affinity of FICZ and ITE is higher for mAHR than hAhR as estimated by the binding energy calculations of the ligand and protein complexes. The binding energies for the ligand-mAhRLBD complex formation for FICZ and ITE were -4.714 kcal/mol and -4.214 kcal/mol respectively. The estimated values for hAhR-LBD-ligand binding were -4.12 kcal/mol for FICZ and -3.46 kcal/mol for ITE. The H-bond pattern established between FICZ and ITE with mouse Ser359 and human Ser365, which are not present for TCDD, may play an important role in increasing AhR binding of these two ligands.

To further validate and apply our AhR-LBD homology model, a database of 498 natural compounds was docked into mouse AhR-LBD model binding pocket for the purpose of identifying new ligands of the AhR by virtual ligand screening (VLS).²¹ After the docking results were obtained we arbitrarily set a binding energy cutoff of $\delta -2.5$ kcal/mol, which resulted in the selection of 25 compounds that docked into the binding pocket in an allowable orientation; these 25 compounds were then tested in cell based experiments. The list of the 25 docked structures and their respective AhR transcriptional activation results are provided in Supplemental Table 1. Six of the 25 compounds were able to activate AhR transcription and at least three of the non-activating compounds strongly antagonized TCDD-induced AhR transcription (Table 2 and data not shown). The top two AhR activators that we identified by VLS were 5-hydroxy-7-methoxyflavone and pinocembrin, both of which are flavonoids. Both compounds strongly activated the AhR transcription in a dose-dependent manner (Figure 4a). The docking orientation of the two potent AhR activators identified was similar, with the 2-phenyl ring allocated in the small hydrophobic cleft surrounded by residues Phe281, Leu309 and Leu347 (Figure 4b). The 1,4-benzopyrone structure of these two flavonoids bound instead in the middle of the binding pocket surrounded by residues Cys327, His285, Met334 and Ala375 (Figure 4b). Both ligands shared H-bond interaction between the carbonyl group of the benzopyrone template and the side chain of residue Ser359 (Figure 4b). Furthermore, 5-hydroxy-7-methoxyflavone

established a second Hbond between its hydroxyl group adjacent to the carbonyl group and the side chain of Ser359 (Figure 4b). We further tested the ability of 5-hydroxy-7-methoxyflavone and pinocembrin to activate endogenous target genes of AhR.²² Both the compounds activated *CYP1A1* and *NADPH quinone reductase* in an AhR dependent manner (Figure 4c). Furthermore, both the compounds promoted nuclear translocation of endogenous AhR from cytosol after one hour of exposure, confirming their status as AhR agonists (Figure 5). Out of the different flavones predicted to bind to AhR ligand binding pocket by the virtual ligand screen, 5-OH-7-methoxyflavone strongly induced AhR transcription, whereas 3-OH-7-methoxyflavone modestly activated AhR transcription. 6-OH-7-methoxyflavone did not activate AhR, suggesting that 6-OH-7-methoxyflavone binds to AhR, but may not bring AhR in to a transcriptionally active form. We tested whether this compound can act as an AhR transcriptional antagonist. Indeed, co-treatment with 6-OH-7-methoxyflavone strongly inhibited both TCDD- and 5-OH-7-methoxyflavone-induced AhR transcription (Supplemental Figure 1). There is precedence for differential effects of different flavones on AhR activity.^{23,24,25} β -naphthoflavene is a strong AhR agonist, whereas α -naphthoflavone is a partial agonist/antagonist of AhR transcription.²⁶ By comparing the binding pattern of 7-methoxyflavone derivatives to AhR ligand binding pocket, we were unable to distinguish agonists from potential antagonists.

In summary, homology models of AhR-LBD in the agonist bound conformation were built using the ICM method.^{14,15} Using these models, differences in ligand binding across various species was structurally characterized. The mouse AhR-LBD homology model was also utilized successfully to identify new AhR agonists by *insilico* structure based VLS.²¹

Experimental Section

Multiple Sequence Alignment

The AhR-LBD sequences in FASTA format for mouse, human, rat, guinea pig, rabbit and zebrafish were retrieved from NCBI database. Multiple sequence alignment was performed online with the ClustalW program.²⁷

Homology Modeling

We used the Nuclear Magnetic Resonance (NMR) structure of the human PAS domain of the hypoxia-inducible factor 2 α (HIF-2 α)¹¹ available in the Protein Data Bank (Pdb) 1P97 as the 3D coordinate template for the homology modeling of mouse, human, and zebrafish isoforms 1 and 2 AhR-LBD. All models were energetically refined using the internal coordinate space with Molsoft ICM v3.5-1p.^{14,15}

Molecular Docking

The receptor model is represented by five types of interaction potentials, namely, (i) the van der Waals potential for a hydrogen atom probe; (ii) the van der Waals potential for a heavy-atom probe (generic carbon of 1.7 Å radius; (iii) an optimized electrostatic term; (iv) hydrophobic terms; and (v) loan-pair-based potential, which reflects directional preferences in hydrogen bonding. The energy terms were based on the all-atom vacuum force field ECEPP/3 with appended terms from the Merck Molecular Force Field to account for solvation free energy and entropic contribution.¹⁵ Modified inter-molecular terms such as soft van der Waals and hydrogen-bonding as well as a hydrophobic term are added. Conformational sampling was based on the biased probability Monte Carlo (BPMC) procedure, which randomly selects a conformation in the internal coordinate space and then makes a step to a new random position independent of the previous one according to a predefined continuous probability distribution. It has also been shown previously that after each random step, full local minimization greatly improves the efficiency of the procedure.

In the ICM-VLS (Molsoft ICM v3.5-1p) screening procedure the ligand scoring was optimized to obtain maximal separation between binders and non-binders.^{28,29} Each compound was assigned a score according to its fit within the receptor, which accounts for continuum and discrete electrostatics, hydrophobicity, and entropy parameters.^{15,28,29}

Binding energy calculations

The free binding energy of a ligand was estimated according to the protocol described.³⁰ In this method, the estimate consists of three essential terms and an additional constant term: $\Delta G_{\text{binding}} = \Delta G_{\text{H}} + \Delta G_{\text{EL}} + \Delta G_{\text{S}} + C$. Every term (except for the constant) is a difference between the corresponding energy in bound and unbound states. The hydrophobic, or cavity term ΔG_{H} , accounts for the change in solvent accessible area upon complexation is calculated as the solvent accessible area, calculated with a 1.4 Å water probe, and multiplied by 30 cal/Å². The electrostatic term ΔG_{EL} , was composed of the coulombic interactions and electrostatic energy of solvation estimated with the continuum dielectric model using the boundary element method³¹ with MMFF94 partial charges, a protein dielectric constant of eight, water dielectric constant of 80 and probe radius of 1.4 Å. The entropic term was a sum of the side chain and ligand entropic terms. The contribution of each rotatable bond frozen upon binding was estimated as 0.5 kcal/mole. The constant value of 3 kcal/mole was used.

Library of compounds

A database of 498 natural compounds from TimTec (Newark, DE) was used for in silico VLS. Twenty five compounds identified by VLS from TimTec were dissolved in DMSO and tested at a final concentration of 20 μM. The structures and ≥ 95% purity of the tested natural compounds were confirmed by TimTec using NMR.

Cell Culture

All cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with L-glutamine (Mediatech Inc., VA) supplemented with 10% FBS (Tissue Culture Biologicals, CA), 100 IU/mL penicillin, and 100 μg/mL streptomycin (Mediatech Inc., VA) in a humidified 5% CO₂ atmosphere. Cells were typically passaged every three days at a dilution of 1:5.

Semi-quantitative Polymerase Chain Reaction (PCR)

For analysis of AhR target gene induction, a pair of Hepa1c1c7 derived cell lines were used, one with significantly reduced AhR expression (C12) and the same cell line stably expressing a WT AhR construct (Hepa1 C12+AhR)³, as described in the supplemental information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

AhR	Aryl Hydrocarbon Receptor
TCDD	2,3,7,8-tetrachlorodibenzo- <i>p</i> -dioxin
PAS	Per-AhR/Arnt-Sim
LBD	ligand binding domain
bHLH	basic-helix-loop-helix
ICM	Internal Coordinate Mechanism

HIF-2α	hypoxia-inducible factor 2 α
FICZ	6-formylindolo [3,2-b] carbazole
ITE	2-(1'H-indole-3'-carbonyl)-thiazole-4-carboxylic acid methyl ester
Pdb	Protein Data Bank
NMR	Nuclear Magnetic Resonance
DMEM	Dulbecco's Modified Eagle Medium
PCR	Polymerase Chain Reaction.

Acknowledgments

We are grateful to Dr. Michael Denison for providing the Hepa1 cells. We thank Christopher Sullivan, Center for Genome Research and Biocomputing at Oregon State University (OSU) for excellent software and hardware support and Duc Nguyen for laboratory assistance. This work was supported in part by the startup funds to SKK from the Department of Environmental and Molecular Toxicology, OSU Research office and grants from the Department of Defense-Breast Cancer research Program, Medical Research Foundation of Oregon and National Institute of Environmental Health Sciences to the Environmental Health Sciences Center at OSU and its core facilities. DK and EOD were supported by the NIEHS predoctoral training grant.

References

1. Beischlag TV, Morales JL, Hollingshead BD, Perdew GH. The Aryl Hydrocarbon Receptor Complex and the control of gene expression. *Crit Rev Eukaryot Gene Expr.* 2008; 18:207–250. [PubMed: 18540824]
2. Nguyen LP, Bradfield CA. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem Res Toxicol.* 2008; 21:102–116. [PubMed: 18076143]
3. Hankinson O. The aryl hydrocarbon receptor complex. *Annu Rev Pharmacol Toxicol.* 1995; 35:307–340. [PubMed: 7598497]
4. Kolluri SK, Weiss C, Koff A, Goettlicher M. Cell Cycle control by dioxins is mediated by the induction of the p27Kip1 inhibitor. *Genes Dev.* 1999; 13:1742–1753. [PubMed: 10398686]
5. Mitchell KA, Lockhart CA, Huang G, Elferink CJ. Sustained aryl hydrocarbon receptor activity attenuates liver regeneration. *Mol Pharmacol.* 2006; 70:163–170. [PubMed: 16636136]
6. Ma C, Marlowe JL, Puga A. The aryl hydrocarbon receptor at the crossroads of multiple signaling pathways. *EXS.* 2009; 99:231–257. [PubMed: 19157064]
7. Funatake CJ, Marshall NB, Steppan LB, Mourich DV, Kerkvliet NI. Cutting edge: activation of the aryl hydrocarbon receptor by 2,3,7,8-tetrachlorodibenzo-p-dioxin generates a population of CD4+ CD25+ cells with characteristics of regulatory T cells. *J Immunol.* 2005; 175:4184–4188. [PubMed: 16177056]
8. Quintana FJ, Basso AS, Iglesias AH, Korn T, Farez MF, Bettelli E, Caccamo M, Oukka M, Weiner HL. Control of T_{reg} and T_H17 cell differentiation by the aryl hydrocarbon receptor. *Nature.* 2008; 453:65–71. [PubMed: 18362915]
9. Veldhoen M, Hirota K, Westendorf AM, Buer J, Dumoutier L, Renauld J-C, Stockinger B. The aryl hydrocarbon receptor links T_H17-cell-mediated autoimmunity to environmental toxins. *Nature.* 2008; 453:106–109. [PubMed: 18362914]
10. Pandini A, Denison MS, Song Y, Soshilov AA, Bonati L. Structural and functional characterization of the Aryl Hydrocarbon Receptor Ligand Binding Domain by Homology Modeling and Mutational Analysis. *Biochemistry.* 2007; 46:696–708. [PubMed: 17223691]
11. Erbel PJ, Card PB, Karakazu 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]

12. Abagyan R, Batalov S, Cardozo T, Totrov M, Webber J, Zhou Y. Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search. *Proteins*. 1997 Suppl 1:29–37. [PubMed: 9485492]
13. Abagyan RA, Totrov MM. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* 1994; 235:983–1002. [PubMed: 8289329]
14. Cardozo T, Totrov M, Abagyan R. Homology modeling by the ICM method. *Proteins*. 1995; 23:403–414. [PubMed: 8710833]
15. Abagyan RA, Totrov MM, Kuznetsov DA. ICM: A New Method For Protein Modeling and Design: Applications to Docking and Structure Prediction From The Distorted Native Conformation. *J. Comp. Chem.* 1994; 15:488–506.
16. An J, Totrov M, Abagyan R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics*. 2005; 4:752–761. [PubMed: 15757999]
17. Goryo K, Suzuki A, Del Carpio C, Siizaki K, Kuriyama E, Mikami Y, Kinoshita K, Yasumoto K-I, Rannug A, Miyamoto A, Fujii-Kuriyama Y, Sogawa K. Identification of amino acid residues in the Ah receptor involved in ligand binding. *Biochem Biophys Res Commun*. 2007; 354:396–402. [PubMed: 17227672]
18. Tanguay RL, Abnet CC, Heidemann W, Peterson RE. Cloning and characterization of the zebrafish (*Danio rerio*) aryl hydrocarbon receptor. *Biochim. Biophys. Acta*. 1999; 1444:35–48. [PubMed: 9931422]
19. Andreasen EA, Hahn ME, Heidemann W, Peterson RE, Tanguay RL. The zebrafish (*Danio rerio*) aryl hydrocarbon receptor type 1 is a novel vertebrate receptor. *Mol. Pharmacol.* 2002; 62:234–249. [PubMed: 12130674]
20. Zhao Y-Y, Tao F-M, Zeng EY. Theoretical study of the quantitative structure-activity relationships for the toxicity of dibenzo-*p*-dioxins. *Chemosphere*. 2008; 73:86–91. [PubMed: 18589475]
21. Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. *J Med Chem*. 2003; 46:3045–3059. [PubMed: 12825943]
22. Kolluri SK, Balduf C, Hofmann M, Goettlicher M. Novel target genes of the Ah (Dioxin) receptor: transcriptional induction of N-myristoyltransferase 2. *Cancer Res*. 2001; 61:8534–8539. [PubMed: 11731439]
23. Gasiewicz TA, Kende AS, Rucci G, Whitney B, Willey JJ. Analysis of structural requirements for Ah receptor antagonist activity: Ellipticines, flavones, and related compounds. *Biochem Pharmacol.* 1996; 52:1787–1803. [PubMed: 8986142]
24. Lu YF, Santostefano M, Cunningham BD, Threadgill MD, Safe S. Substituted flavones as aryl hydrocarbon (Ah) receptor agonists and antagonists. *Biochem Pharmacol.* 1996; 51:1077–1087. [PubMed: 8866830]
25. Puppala D, Gairola CG, Swanson HI. Identification of kaempferol as an inhibitor of cigarette smoke-induced activation of the aryl hydrocarbon receptor and cell transformation. *Carcinogenesis*. 2007; 28:639–647. [PubMed: 17012224]
26. Wilhelmsson A, Whitelaw ML, Gustafsson JA, Poellinger L. Agonistic and antagonistic effects of alpha-naphthoflavone on dioxin receptor function. Role of the basic region helix-loop-helix dioxin receptor partner factor Arnt. *J Biol Chem*. 1994; 269:19028–19033. [PubMed: 8034660]
27. <http://align.genome.jp/>
28. Totrov, M.; Abagyan, R. Protein-Ligand Docking as an Energy Optimization Problem. In: Raffa, RB., editor. *Drug-Receptor Thermodynamics: Introduction and Experimental Applications*. New York: John Wiley & Sons; 2001. p. 603-624.
29. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins*. 1997 Suppl. 1:215–220. [PubMed: 9485515]
30. Schapira M, Totrov M, Abagyan R. Prediction of the binding energy for small molecules, peptides and proteins. *J. Mol. Recognition*. 1999; 12:177–190.
31. Totrov M, Abagyan R. Rapid boundary element solvation electrostatics calculations in folding simulations: Successful folding of a 23-residue peptide. *Biopolymers*. 2001; 60:124–133. [PubMed: 11455546]

A					
				<u>Aβ</u>	
M	241-HGQNKKGKDG	ALLPPQLALF	AIATPLQPPS	ILEIRTKNFI	FRTHHKLDFT-290
R	245-HGQNKKGKDG	ALLPPQLALF	AIATPLQPPS	ILEIRTKNFI	FRTHHKLDFT-294
H	247-HGQKKKGKDG	SILPPQLALF	AIATPLQPPS	ILEIRTKNFI	FRTHHKLDFT-296
RB	245-HGQNKKGKDG	SLLPPQLALF	AIATPLQPPS	ILEIRTKNFI	FRTHHKLDFT-294
GP	246-HGQNKKGKDG	SVLPPQLALF	AIATPLQSPS	ILELRITKNFL	FRTHHKLDFT-295
Z2	252-HGQNKLAEDG	TLAHPQLALF	IIATPLQPPS	ILEIRSKTLL	FQTHHKLDFT-301
		<u>Bβ</u>	<u>Cα</u>	<u>Dα</u>	<u>Eα</u>
M	291-PIGCDAGKQL	ILGYTEVELC	TRGSGYQFIH	AADILHCAES	HIRMIKTGES-340
R	295-PIGCDAGKQL	ILGYTEVELC	NKGSGYQFIH	AADMLHCAES	HIRMIKTGES-344
H	297-PIGCDAGKRI	VLGYTEAELC	TRGSGYQFIH	AADMLYCAES	HIRMIKTGES-346
RB	295-PTGCDAGKQI	VLGYTEAELC	MRGSGYQFIH	AADMLYCAES	HIRMIKTGES-344
GP	296-PIGCDAGKQI	VLGYTEAELC	AKGSGYQFIH	AGDMLHCAEY	HIRMIKTGES-345
Z2	302-PMGIDTRGKV	VLGYTEIELC	MRGSGYQFIH	AADMMYCADN	HIRMIKTGES-351
		<u>Gβ</u>	<u>Hβ</u>	<u>Iβ</u>	
M	341-GMTVFRLLAK	HSRWRWVQSN	ARLIYRNGRP	DYIIATQRPL	TDEE-384
R	345-GMTVFRLLAK	HSRWRWVQSN	ARLIYRNGRP	DYIIATQRPL	TDEE-388
H	347-GMIVFRLLTK	NNRWTWVQSN	ARLLYKNGRP	DYIIATQRPL	TDEE-390
RB	345-GLAVFRLLTK	DNRWAWVQSN	ARFIYKNGRP	DFIIATQRPL	TDEE-388
GP	346-GMTVFRLLTK	HNRWIWMQSN	ARVVYKNGRP	DYIIATQRAL	TDEE-389
Z2	352-GLTVFRLLSK	GGTWIIVQAN	ARLVYKAGRP	DFIIARQRAL	TNEE-395

M, mouse; R, rat; H, human; RB, rabbit; GP, guinea pig; Z2, zebrafish isoform 2.

Figure 1.

(A) Interspecies multiple sequence alignment of the AhR-LBD performed with the program ClustalW²⁸. (B) Evolutionary tree of different species.

Figure 2.

(A) Homology model of the mouse AhR-LBD with the protein backbone displayed as ribbon and colored by secondary structure (ICM v3.5-1n, Molsoft). The residues are displayed as sticks and colored by atom type with the carbon atoms in white. The binding pocket volume is colored in green and was calculated using the program ICM Pocket Finder (ICM v3.5-1n, Molsoft).¹⁶ (B–D) TCDD docking orientation into mouse (B), human (C), zebrafish isoform 2 (D) AhR-LBD binding pocket. The protein backbone is displayed as ribbon and colored by secondary structure. The residues are displayed as sticks and colored by atom type with the carbon atoms in white. TCDD is displayed as sticks and colored by atom type (carbon atoms in yellow). (E) mouse AhR-LBD binding pocket volume is colored

in green and was calculated using the program ICM Pocket Finder.¹⁶ The protein backbone is displayed as ribbon and colored by secondary structure. The residues are displayed as sticks and colored by atom type with the carbon atoms in orange (wild type) and white (A375L mutant). (F–G) binding pocket region identified by ICM Pocket Finder of the zebrafish AhR-LBD of isoform 2¹⁹ (F) and isoform 1a (G) colored in translucent blue and dark blue respectively. The protein backbone is displayed as ribbon and colored by secondary structure. Residues are displayed as sticks and colored by atom type with carbon atoms in white. Docked TCDD is displayed as sticks and colored in red.

Figure 3. ITE (A, C) and FICZ (B, D) docking orientation respectively into mouse (A, B), human (C, D) AhR-LBD homology model. The protein backbone is displayed as ribbon and colored by secondary structure. The residues are displayed as sticks and colored by atom type with the carbon atoms in white. FICZ and ITE are displayed as sticks and colored by atom type with carbon atoms in yellow. H-bonds are represented by white dashed lines.

Figure 4.

(A) Mouse Hepa1 cells transfected with the AhR response element (AhRE)/ xenobiotic response element (XRE)-luciferase reporter were treated with compounds at the indicated concentrations for 6 hours and assayed for reporter gene activity. (B) Docking orientation of 5-hydroxy-7-methoxyflavone and pinocembrin into mouse AhR-LBD binding pocket (ICM v3.5-1n, Molsoft). The protein backbone is displayed as ribbon and colored by secondary structure. The residues are displayed as sticks and colored by atom type with the carbon atoms in green. The ligands are displayed as sticks, colored by atom type with carbon atoms in yellow. H-bonds are represented as black dashed lines. (C) Western blot showing the expression of AhR in the mouse Hepa1 derivative cell line C12, which has a significantly decreased abundance of AhR, and the C12+AhR, wildtype AhR re-expressing cells. The blot shows 30 μ g of whole cell lysate from each line (upper panel). Semi-quantitative RT-PCR for AhR target genes²² following exposure to vehicle (0.1% DMSO), TCDD (1 nM), 5-hydroxy-7-methoxyflavone (20 μ M) and pinocembrin (20 μ M) for 18 hours. GAPDH expression control was used to indicate similar amount of RNA usage (lower panel). CYP1A1 (Cytochrome p450 1A1), and NADPH-QO (NADPH quinone-oxidoreductase). PCR cycle numbers are indicated.

Figure 5.

Nuclear localization of AhR in Hepa1 cells. Cells were treated with the indicated compounds for 60 minutes, fixed and then immunostained with AhR followed by fluoresceinisothiocyanate(FITC)-conjugated secondary antibody. Cells were stained by 4,6-diamino-2-phenylindole (DAPI) to visualize nucleus and were imaged on a Zeiss Axiovert S100TV microscope.

Table 1

Interspecies percentage of sequence identity in the AhR-LBD domain

SPECIES							
	M	R	H	GP	Z2	RB	
SPECIES	M	100	97.91	88.19	84.72	72.22	88.19
	R	97.91	100	87.5	86.11	72.22	88.19
	H	88.19	87.5	100	86.11	72.22	90.97
	GP	84.72	86.11	86.11	100	71.52	86.11
	Z2	72.22	72.22	72.22	71.52	100	75.69
	RB	88.19	88.19	90.97	86.11	75.69	100

M, mouse; R, rat; H, human; GP, guinea pig; Z2, zebrafish isoform 2; RB, rabbit.

Table 2

Activation of AhR transcription by compounds identified by virtual docking into AhR LBD

Compound Name	Structure	Binding Energy	Induction of AhR Transcriptional Activity
2,3,7,8-Tetrachlorodibenzo-p-dioxin		-3.80	+++ 21 fold
5-Hydroxy-7-methoxyflavone		-4.3	+++ 11.3 fold
7-Methoxyisoflavone		-3.75	+ 2.7 fold
6-Methylflavone		-3.52	++ 4.8 fold
3-Hydroxy-6-methylflavone		-3.37	+ 2.5 fold
Pinocembrin (5,7-Dihydroxyflavanone, R-form)		-2.97	+++ 8.0 fold
7,8,2'-Trihydroxyflavone		-2.74	+ 2.8 fold

+++ Very Strong activation (≥ 8 fold);++ Strong activation (>4 fold);+ Activation (<3 fold)