Biochemical and Biophysical Research Communications 413 (2011) 176–181

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Human aryl-hydrocarbon receptor and its interaction with dioxin and physiological ligands investigated by molecular modelling and docking simulations

Maria Salzano [a], Anna Marabotti [b], Luciano Milanesi [b], Angelo Facchiano [a,*]

[a] Institute of Food Science (ISA), CNR, Via Roma 64, 83100 Avellino, Italy
[b] Institute of Biomedical Technologies (ITB), CNR, Via F.lli Cervi 93, 20090 Segrate (MI), Italy

ARTICLE INFO

Article history:
Received 22 July 2011
Available online 17 August 2011

Keywords:
Aryl-hydrocarbon receptor
Dioxin
Xenobiotics
Docking simulations

ABSTRACT

Molecular structure of the ligand binding domain of hAhR has been modelled by homology modelling techniques and used for docking simulations with dioxin and nine more xenobiotics and endogenous ligands. The study evidences that different sites may bind these ligands, whereas only one binding site has been previously indicated by past studies on the mouse homologous receptor. The differences in the sequence of mouse and human AhR ligand binding domain may explain this observation, being most of them in the additional sites observed. Preferences of the evaluated ligands for the different sites are reported and discussed in view of their functional role.

1. Introduction

Dioxins are a group of chlorinated heteroaromatic polyciclic organic chemicals, and the term usually includes polychlorinated dibenzodioxins and polychlorinated dibenzofurans. Some of them have harmful characteristics depending on the number and structural position of chlorine atoms [1]. 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD or dioxin), the most toxic member of the group, with four chlorine atoms, is considered one of the most toxic compounds ever released into the environment [2,3]. Immune system toxicity and dysfunction are some of the most consistent features observed in all animal species following exposure to dioxins and related chemicals. In humans, TCDD produces a broad spectrum of effects at very low concentrations, leading to indicate TCDD as an "environmental hormone". At non-lethal doses, reproductive and developmental effects, hepato-carcinogenesis, tumour promotion, and immune suppression are observed [3,4].

The link between chemicals-induced toxicity and clinical effects on human health is the activation of aryl-hydrocarbon receptor (AhR), a cytosolic transcription factor that, in its latent unliganded state, forms complexes with HSP90, p23 and XAP2. Upon ligand binding, AhR translocates to the nucleus, where it complexes with its hetero-dimerization partner, the AhR Nuclear Translocator (ARNT), to modulate expression of AhR target genes containing specific DNA enhancer sequences, known as AhR responsive elements (AhREs) [5,6]. AhR is a member of the bHLH-PAS (basic helix-loop-helix Per ARNT Sim) protein family found in organisms as diverse as *Caenorhabditis elegans*, *Drosophila melanogaster* and mammals. bHLH-PAS proteins are biological sensors for a variety of stimuli, controlling neurogenesis, vascularization, circadian rhythms, metabolism and stress responses to hypoxia, among others [7,8]. AhR is highly conserved in evolution and is present in many cell types, albeit at different abundance [9,10]. The selective forces that led to the high degree of conservation of the AhR amino acid sequence are unknown and its physiological function(s) are still being elucidated. The search for *bona fide* endogenous ligands that regulate AhR transcriptional activity under physiological conditions has had limited success, too.

Several efforts were made to model the ligand binding domain (LBD) of AhR in mouse [11,14]. In all these studies, the attention was focused on PAS B domain, one of the two structural repeats (PAS A and PAS B) within the PAS domain, that contains the primary HSP90 and also the ligands binding site [10,13,15]. Studies examining the ability of low- and high-affinity ligands to bind to AhRs with mutations in the ligand-binding pocket, have provided some evidences for differential interactions of ligands with the AhR LBD and for amino acids involved in the binding of ligands, thus identifying a specific binding site.

Abbreviations: AhR, aryl-hydrocarbon receptor; ARNT, AhR Nuclear Translocator; AhREs, AhR responsive elements; bHLH-PAS, basic helix-loop-helix Per ARNT Sim; hAhR, human AhR; LBD, ligand binding domain; mAhR, mouse AhR; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; dFICZ, 6,12-diformylindolo(3,2-b)carbazole; FICZ, 6-formylindolo(3,2-b)carbazole; PentaCB, 3,3′,4,4′,5′-Pentachlorobiphenyl.

* Corresponding author. Fax: +39 0825 781585.
E-mail addresses: maria.salzano@isa.cnr.it (M. Salzano), anna.marabotti@itb.cnr.it (A. Marabotti), luciano.milanesi@itb.cnr.it (L. Milanesi), angelo.facchiano@isa.cnr.it (A. Facchiano).

0006-291X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2011.08.039

The human AhR (hAhR) has high affinity to a distinct subset of ligands, structurally divergent from typical exogenous AhR ligands like TCDD, including sterols, indigoids, heme metabolites, tetrapyrroles such as bilirubin, arachidonic acid metabolites and dietary components [16–18]. hAhR ligand binding domain (hAhR–LBD) is most structurally analogous to the mouse AhR (mAhR)[d] allele ligand binding domain and the approximately 10-fold lower affinity for TCDD, compared to mAhR, has been attributed to the amino acid residue Valine 381 (Alanine 375 in mAhR) in the ligand binding domain of the hAhR [19,20]. Therefore, the hAhR–LBD may be able to bind different ligands than mAhR–LBD.

In this work, our purpose was to investigate the molecular and biochemical aspects of the interaction between hAhR–LBD and the most relevant known AhR ligands, first of all TCDD. The aim was to gain insights about the binding site and the possible preference for the ligands, in the view of obtaining more information about the toxic effect of TCDD and its interference with physiological ligands.

Since 3D structure of hAhR–LBD is not yet been experimentally determined, we created a model by homology modelling. The choice of ligands was based on the work of Nguyen and Bradfield [17], in which the Authors indicated the major classes of AhR ligands. We selected 10 among the most relevant AhR ligands, and applied docking simulations, to identify the binding site(s), the preference of ligands and possible key amino acids in ligand binding.

2. Methods

The amino acid sequence of hAhR was obtained from Uniprot [21] (Accession No.: P35869). The putative LBD was selected from the sequence in the region between residue 275 and 420. To identify suitable template structures for modelling of hAhR–LBD, we searched the PDB database [22] by BLAST [23] and also applied the fold recognition strategy using the web server GenThreader [24]. The fold recognition search retrieved four structures with high confidence score, all classified as PAS domains. We used these templates for the modelling of the hAhR–LBD. Accordingly to protein modelling procedures already used with success in our laboratory for proteins with low sequence similarity to proteins with experimentally determined 3D structures [25,26], we used a combination of different bioinformatics software, to obtain sequence alignments, secondary structure information for crystallographic models as well as prediction for the sequence of our target protein, to build the 3D model with the highest level of reliability.

Secondary structures on the templates were calculated with DSSP program [27], meanwhile their prediction on target sequence was carried out with Jpred server [28].

The target and the templates sequences were aligned using ClustalW [29], setting default alignment options. Then, the alignment was slightly edited, taking into account the position of secondary structural elements for gap positioning.

3D models of putative hAhR–LBD were built using MODELLER 9v7 version [30]. Several modelling procedures were performed in order to find the most suitable conditions in terms of starting sequence alignment and combination of templates used. Each model created was evaluated for its stereo-chemical quality by PROCHECK program [31], and for the energy profile by PROSAII [32].

CASTp server [33] was used to identify the possible ligand binding pockets within the newly created 3D structures of hAhR–LBD.

Pubchem [34], a freely available database of chemical structures of small organic molecules and information on their biological activities, was used to retrieve all ligand molecules in the Mol2 format. The 3D structures were converted to PDB format using Chimera program [35]. The only exception was 6,12-diformylindolo(3,2-b)carbazole compound (dFICZ), whose 3D structure was built and geometrically optimised, using the Builder module of INSIGHT II (Version 2000.1, Accelrys, Inc., 2000), modifying the structure of 6-formylindolo(3,2-b)carbazole (FICZ) retrieved from PubChem. In Supplementary files (Fig. S1) we show the chemical structures of the selected ligands.

Protein–ligand docking simulations were conducted using AutoDock version 4.0 and ADT Suite [36] to prepare the systems for calculations. For each ligand, 100 docking runs using Lamarckian genetic algorithm with the default parameters were performed, with a maximum of 2.5×10^6 energy evaluations, treating the protein as rigid and the ligand as flexible.

For each ligand type, two sets of docking experiments were carried out: blind docking and focused docking experiments. In the first case, whole hAhR–LBD was included in calculations using a docking box of 126 Å × 126 Å × 126 Å with a spacing of 0.41 Å, centred on the protein. In the second one, a smaller docking box (70 Å × 70 Å × 70 Å with a spacing of 0.375 Å), was specified and approximately centred on the pool of amino acids identified by the blind docking as involved in the interaction with the ligand. In the case of pocket C, the pool of amino acids included also the nine residues identified by previous works based on mutagenesis approaches [12–14,37,38]. Docking poses have been clustered using an RMSD value of 2.0 Å. For each ligand tested, only the conformational clusters most populated and with lower energy binding (E_b) (in some cases, they coincide), were considered for further investigations.

3. Results and discussion

3.1. 3D modelling of hAhR LBD

To identify suitable template structures for modelling of hAhR–LBD, we searched the PDB database by BLAST. Since no significant results were obtained, we applied the fold recognition strategy using the web server GenThreader. The search retrieved the following four structures with high confidence score: crystal structure of human ARNT C-terminal PAS domain (PDB ID: 1X0O [39]); crystal structure of PAS repeat region of the Drosophila clock protein PERIOD (PDB ID: 1WA9 [40]); crystal structure of a PAS domain fragment of mammalian clock protein mPER2 (PDB ID: 3GDI [41]) and crystal structure of a high affinity heterodimer of HIF2α and ARNT C-terminal PAS domains (PDB ID: 3F1P [42]). Being the hAhR–LBD already known as a PAS domain, we considered these results reliable and selected these experimental structures as templates to build the 3D structure of hAhR LBD. The target and the templates sequences were aligned using ClustalW. Some refinements were introduced in the alignment to eliminate the occurrence of gaps within secondary structure elements, leaving unchanged the number of aligned identical residues.

Different models were created with MODELLER 9v7, starting from different alignments obtained by modifying some gap positions, and using all four templates, or alternatively by excluding some of them. The alignment used as starting point is represented in Fig. 1. The best model, selected after evaluation of stereochemical quality and energetics, has in the Ramachandran plot 90.7%, 9.3%, and 0.0% residues with phi/psi angles located in the most favourable regions, the additionally allowed regions, and the disallowed regions, respectively.

Z-score values calculated by PROSA was −4.39. We consider this value indicative of a good quality model, due to the negative value and its similarity to that of templates (ranging from −5.46 to −7.20) for which the structure has been determined for a larger region.

The final model of hAhR–LBD (Fig. 2) folds similarly to the typical PAS domain architecture, consisting of an anti-parallel beta