Homology modeling of the AhR ligand binding domain

M. Procopio[#], A. Lahm[§], A. Tramontano[§], L. Bonati[#], D. Pitea[#]

[#] Dipartimento di Scienze dell'Ambiente e del Territorio, Università degli Studi di Milano-Bicocca, Via Emanueli 15, 20126 Milano
[§] Istituto di Ricerche di Biologia Molecolare P. Angeletti, Via Pontina Km 30.600, 00040 Pomezia (Roma)

Introduction

Studies on the biological mechanism of action of PCDDs indicated that biological effects are mediated by the binding to a specific cytoplasmic protein, the aryl hydrocarbon receptor (AhR) [1]. Ligand-induced activation of AhR initiates a process whereby the receptor is transformed into a nuclear transcription factor complex with the protein ARNT (Ah Receptor Nuclear Translocator). The ligand-activated AhR/ARNT heterodimer recognizes the specific core DNA sequence XRE (Xenobiotic Responsive Elements), inducing genes that encodes xenobiotic metabolizing enzymes [2].

Therefore, understanding the PCDD-AhR binding process at a molecular level is a key step for gaining insight into the biological mechanism of action of these compounds.

Till now only indirect studies of the PCDD-AhR interaction have been performed by searching for relationships between ligand properties and biological activity data [3]. Following this approach, in our previous works structure-activity relationship studies were performed, based on some electronic properties (molecular electrostatic potential, molecular polarizability, electron affinity and softness) for a series of PCDDs of varying binding affinities [4]. A further step in modeling the PCDD-AhR interaction is to develop a three-dimensional model for the AhR ligand binding domain and identify the amino acidic residues directly interacting with PCDDs.

It is known that AhR and ARNT belong to the Per-ARNT-Sim (PAS) family of proteins, whose members act as transcriptional activators, or as sensor modules of two-component regulatory systems, or as ion channels [5]. In AhR two PAS domains are present in a ~ 270-residue region encompassing two imperfect repeats of ~ 110 amino acids (PAS-A and PAS-B) separated by a sequence of approximately 50 amino acids. Its functions are, in some cases, to mediate protein-protein interactions and, in other cases such as for AhR, ligand and/or cofactor binding. A minimal ligand binding domain in the AhR of mouse was mapped between amino acids 230 and 397, the region that encompasses the PAS-B repeat [6]. Deletion analysis showed that modifications outside the PAS domain had no effect on ligand binding. Deletion of the PAS-A (aa 121-182) repeat reduced ligand binding to 30% and the PAS-B (aa 259-374) completely abolished binding as did deletion of the complete PAS region [6].

Taking advantage of the recently established Xray-structures of several PAS domains, we have initiated a modeling study of the AhR ligand binding domain (LBD).

Methods

With the goal of building a three-dimensional model for the ligand binding domain of AhR (mouse C57BL/6), the modeling proceeded in five steps: an initial search for homologous sequences, subsequent multiple sequence alignment, secondary structure prediction, alignment of

the predicted secondary structure/multiple sequence alignment against the structural templates; building of a three-dimensional models according to the latter alignment.

Sequence database searches were performed using the LBD of the Ah receptor [6] as query through application of the FASTA [7], BLAST and PSI-BLAST [8] search algorithms on the non-redundant protein and DNA sequence database at the NCBI. In all cases defaults parameters were used.

The target sequence was also submitted to the JPRED web server [9] in order to obtain a secondary structure prediction for the AhR PAS-B domain. JPRED, after an initial sequence database search, generates a non-redundant multiple sequence alignment which serves as input to a number of distinct algorithms for secondary structure prediction: PHD [10], Predator [11], DSC [12], NNSSP [13], Zpred [14]. In a final step a consensus secondary structure prediction is generated from the individual results.

Due to the low level of sequence homology, the relationship between the target sequence (the AhR PAS-B domain) and the possible structural templates was established by manually aligning multiple sequence alignments together with the predicted secondary structure against multiple sequence alignments generated for the structural templates. During this process both the conservation of secondary structure elements and conservation of amino acid character (hydrophilic/hydrophobic) was optimized. All alignments were manipulated using the interactive display program SEAVIEW [15]. Three-dimensional models were subsequently generated with the INSIGHT II molecular modeling system [16].

Results and Discussion

Application of the recursive PSIBLAST database search revealed a number of distant homologies between the AhR ligand binding domain and other PAS proteins, amongst which the hypoxia-inducible factor 1 alpha (HIFA; 28% identity), the period clock protein (PER; 16% identity) and two proteins for which the crystal structure of the respective PAS domains have recently been resolved: the human potassium channel HERG (13% of identity) and the heme binding domain of the bacterial O_2 sensing FixL protein (8% of identity).

In the HERG K^+ channel, a member of the eag (ether a go go) K^+ channel family, the PAS domain corresponds to the first 135 residues of the cytosolic N-terminus. The crystal structure showed a five-stranded antiparallel β−sheet decorated on two sides by α helices (Fig. 1) [17] and bares extensive structural similarity to the photoactive yellow protein (PYP), a bacterial light-sensing protein, which has been proposed as structural prototype for the three dimensional fold of the PAS domain superfamily [18].

In FixL oxygen binding is achieved through a heme binding domain which controls the activity of a histidine kinase domain. Unlike the frequently observed mainly α-helical classes of heme proteins [19], the dominant feature of the FixL heme domain is a five-stranded antiparallel β-barrel (Fig. 2), again with significant structural homology to PYP.

A structural alignment between HERG and FixL, obtained from the FSSP database [20], showed 15% of amino acid sequence identity and 2.8Å rmsd on superimposed Cα atoms.

To obtain a final multiple alignment establishing a relation between the sequences to be modeled and the structural templates, we combined all collected information in the most consistent way: (i) the alignments based on sequence similarity generated by PSI-BLAST; (ii) the predicted secondary structures of PAS-B, PER and HIFA; (iii) the structural alignments of FixL and HERG obtained from the FSSP database.

The results suggest that, despite low sequence homology with the structural templates FixL and HERG, the models for the AhR PAS-B domain should be accurate enough to allow us to generate plausible hypotheses regarding ligand-binding. Having the models at hand will be particularly useful later on in guiding experiments that test the hypothesis and subsequently, in interpreting the results.

Fig. 1 Crystal structure of the HERG eag domain

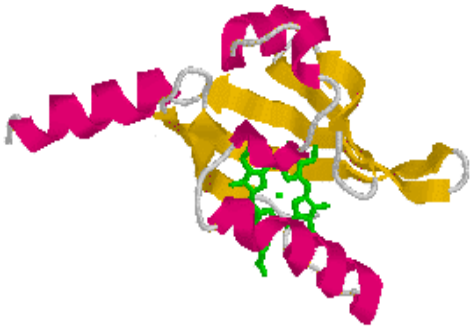

Fig. 2 Crystal structure of the FixL heme binding domain

References

[1] Landers J.P., Bunce N.J., *Biochem J*., 1991, **276,** 273.
[2] Sogawa K., Fujii-Kuriyama Y., *J. Biochem*., 1997, **122**, 1075.
[3] Waller C.L., McKinney J.D., *Chem. Res. Toxicol*., 1995, **8**, 847, and references therein.
[4] a) Bonati L., Fraschini E., Lasagni M., Palma Modoni E., Pitea D., *J. Mol. Stru. THEOCHEM*, 1995, **340**, 83; b) Fraschini E., Bonati L., Pitea D., *J. Phys. Chem.*, 1996, **100**, 10564.
[5] Hahn M.E., *Comparative Biochem. Physiol. Part C*, 1998, **121**, 23.
[6] Fukunaga B.N., Probst M.R., Reisz-Porszasz S., Hankinson O., *J. Biol. Chem*., 1995, **270**, 29270.
[7] Pearson W.R., *Methods Enzimo*l., 1990, **183**, 63.
[8] Altschul S.F., Madden T.L.,. Schäffer A.A, Zhang J., Zhang Z., Miller W., Lipman D.J. *Nucleic Acids Res.*, 1997, **25**, 3389.
[9] Cuff J.A., Clamp M.E., Siddiqui A.S., Finlay M., Barton G.J., *Bioinformatics*, 1998, **14**, 892.
[10] Rost B., Sander C., Schneider R., *J. Mol. Biol*., 1994, **235**, 13.
[11] Frishman D., Argos, P., *Prot. Eng.*, 1996, **9**, 133.
[12] King R.D., Sternberg M.J.E., *Prot Sci*., 1996, **5**, 2298.
[13] Salamov A.A., Solovyev V.V., *J. Mol. Biol*., 1995, **247**, 11.
[14] Zvelebil *et. al*., *J. Mol. Biol*., 1987, **195**, 957.
[15] Galtier N., Gouy M., Gautier C., *Comput. Appl. Biosc*., 1996, **12**, 543.
[16] INSIGHT II, Molecular Simulation Inc., San Diego, CA, USA.
[17] Cabral J.H.M., Lee A., Cohen S.L., Chait B.T., Li M., Mackinnon R., *Cell*, 1998, **95**, 649.

[18] Pellequer J.L., Wager-Smith K.A., Kay S.A., Getzoff E.D., *Proc. Natl. Acad. Sci. U. S. A.,* 1998, **95**, 5884.
[19] Gong W., Hao B. S.S. Mansy, Gonzalez G., Gilles-Gonzalez M.A., Chan M.K. *Proc. Natl. Acad. Sci. U. S. A.,* 1998, **95,** 15177.
[20] Holm L.L., Sander C., *Science*, 1996, **273**, 595.