Fungi and animals may share a common ancestor to nuclear receptors

Chris Phelps*,†, Valentina Gburcik*, Elena Suslova*, Peter Dudek*, Fedor Forafonov*, Nathalie Bot*, Morag MacLean*, Richard J. Fagan*, and Didier Picard*‡

1Département de Biologie Cellulaire, Université de Genève, Sciences III, 30 quai Ernest-Ansermet, 1211 Genève 4, Switzerland; and *Inpharmatica Ltd., 60 Charlotte Street, London W1T 2NU, United Kingdom

Edited by Michael G. Rosenfeld, University of California at San Diego, La Jolla, CA, and approved March 2, 2006 (received for review November 21, 2005)

Nuclear receptors (NRs) are a large family of transcription factors. One hallmark of this family is the ligand-binding domain (LBD), for its primary structure, sequence, and regulatory function. To date, NRs have been found exclusively in animals and sponges, which has led to the generally accepted notion that they arose with them. We have overcome the limitations of primary sequence searches by combining sequence profile searches with structural predictions at a genomic scale, and have discovered that the heterodimeric transcription factors Oaf1/Pip2 of the budding yeast Saccharomyces cerevisiae contain putative LBDs resembling those of animal NRs. Although the Oaf1/Pip2 LBDs are embedded in an entirely different architecture, the regulation and function of these transcription factors are strikingly similar to those of the mammalian NR heterodimer peroxisome proliferator-activated receptor α (PPARα)/RXR. We demonstrate that the induction of Oaf1/Pip2 activity by the fatty acid oleate depends on oleate’s direct binding to the Oaf1 LBD. The alteration of two amino acids in the predicted ligand-binding pocket of Oaf1 abolishes both ligand binding and the transcriptional response. Hence, LBDs may have arisen as allosteric switches, for example, to respond to nutritional and metabolic ligands, before the animal and fungal lineages diverged.

The nuclear receptor (NR) superfamily is characterized by two unique domains. Almost all members contain a very highly conserved DNA-binding domain (DBD) consisting of two Cys4 zinc fingers, as well as a somewhat less conserved ligand-binding domain (LBD) comprising ~250 aa at the C terminus. These building blocks confer the potential to act as both an intracellular receptor and a ligand-regulated transcription factor. Although only a minority of NRs have known ligands, it appears that LBDs evolved as allosteric switches to control NR activities as transcription factors (1–3).

Homologous sequences have been identified in a large number of species by performing searches with DBD and/or LBD sequences (3, 4). Because only animals and sponges have recognizable NR sequences, it has been concluded that NRs evolved in a common animal or urmetazoan ancestor (5, 6). However, because these primary sequence searches were all based on one of the available BLAST algorithms (7, 8), more distant homologs with poor primary sequence similarity could not have been identified. We have now combined such searches with structural predictions and have uncovered additional potential homologs in yeast. Our results challenge the assumption that NRs are an animal-specific family of transcription factors and argue that allosteric regulation by NR LBDs is more ancient than animals.

Results

Bioinformatic Discovery of LBD Candidates. We have applied the Genome Thresher algorithm (9–11), which uses a threading approach to protein structure prediction, at a genomic scale. This method allows the identification of proteins that are predicted to have a similar structural fold, although they may not share significant sequence homology. The structural similarity requirement enables the identification of very distant members of a functional family with a confidence that cannot be achieved by using sequence similarity alone. A combination of sequence similarity that is limited to a few functionally important residues with an overall conservation of the structural fold has been used successfully as a criterion to assign the proper function to distantly related sequences (12–14). In the course of attempting to identify all proteins predicted to have the NR LBD fold (1, 2), we discovered two proteins from Saccharomyces cerevisiae, Oaf1 and Pip2 (Fig. 1), in the dataset. We predict two known structural folds in each of the proteins: (i) an N-terminal Zn2Cys6 DBD and (ii) an LBD fold C-terminal to the DBD (Fig. 1 A). As expected, the DBD prediction is of high confidence for both proteins because they exhibit substantial sequence identity to the Zn2Cys6 DBD family of proteins. The LBD predictions are of a lower confidence, reflecting the substantial divergence of these domains. The best LBD structural assignments for Oaf1 and Pip2 are to the LBD of human estrogen receptor β (Protein Data Bank ID code 1QKM) and the LBD of retinoic acid-related orphan receptor β (RORβ) (Protein Data Bank ID code 1K4W) at 78% and 68% confidence, respectively. The sequence identities to these mammalian proteins are 12% and 13% for Oaf1 and Pip2, respectively. Given the large number of NR LBD structures in public databases, there were many more assignments with various degrees of confidence for both proteins (data not shown). Fig. 1 B shows a structural alignment of Oaf1 and Pip2 with the LBDs of retinoid X receptor α (RXRα) and peroxisome proliferator-activated receptor γ (PPARγ). On the basis of position-specific scoring matrices (15), the predicted secondary structures of the Oaf1 and Pip2 LBDs align well with the known secondary structure features of RXRα and PPARγ. Based on a partial sequence alignment with the peroxisomal membrane protein Pex11 from yeast, it had previously been speculated that animal LBDs arose from an older protein (16). However, no functional support was provided, and a substantial portion of the LBD was not aligned to Pex11, which may explain why we could not confirm this hit with our own bioinformatic tools. The relationship of LBDs to Pex11, although intriguing, remains to be further explored.

Spectrum of in Vivo Responses of Oaf1/Pip2 to Fatty Acids. Oaf1 and Pip2 are essential heterodimeric transcription factors for the utilization of the fatty acid oleate as an external carbon source (17–20). In the absence of glucose, the preferred carbon source...
of *S. cerevisiae*, oleate induces the expression of the entire panel of genes required for uptake and breakdown of fatty acids in an OAF1/PIP2-dependent fashion (18–21). Moreover, the transcriptional activity of Oaf1 fused to a heterologous DBD is stimulated by oleate (22). In conjunction with our structural predictions, these findings suggested that oleate might be a ligand and direct activator of Oaf1/PIP2.

To test this hypothesis, we began by further characterizing the in vivo response to fatty acids. To facilitate our analysis, we constructed a luciferase reporter gene whose expression is under the control of the oleate response element from the FOX3 gene (23) upstream of a minimal promoter. Moreover, we performed these experiments under conditions in which the Oaf1/PIP2-dependent activating effects of fatty acids could be separated from their nutritional contribution as a carbon source. Hence, cells were cultured in the presence of a mixture of carbon sources (raffinose, glycerol, and ethanol) that ensure cell growth irrespective of the addition of fatty acids and that, unlike glucose, do not elicit catabolite repression. As expected, the C18 fatty acid oleate induces the response robustly (Fig. 2A). Half-maximal induction (EC50) is achieved with 0.24 mM oleate. A survey of a panel of various saturated and unsaturated fatty acids revealed that 9-unsaturated fatty acids are stronger activators than saturated ones (Fig. 2B). Linoleate, a double-unsaturated C18 fatty acid, produces an even stronger transcriptional response than oleate, whereas the saturated C18 fatty acid stearate produces no response at all. The shorter saturated fatty acid palmitate is able to elicit a weak response. Taken together with the previously reported observation that the saturated C12 fatty acid laurate activates Oaf1/PIP2 target genes (24), these data suggest that short and compact fatty acids may be preferred as activators. Despite the strong activating potential of several different fatty acids, oleate was used in all subsequent experiments because it is one of the most abundant physiological fatty acids in yeast membranes (25), and because it has traditionally been used to induce the responses to unsaturated fatty acids, including utilization as a carbon source and peroxisome proliferation (26).
Oleate Itself Induces Oaf1/Pip2 Activity. Next, we determined whether oleate first needs to be metabolized to activate this pathway. An earlier study (24) had suggested that intact peroxisomes may not be needed for the response, but the requirement for the metabolism of unsaturated fatty acids had not been examined directly. Therefore, the response of the reporter gene was tested in a panel of mutant strains that are all defective in oleate utilization but are comparably competent for the induction of a galactose-inducible luciferase reporter gene (data not shown). As expected, the oleate response depends on both OAF1 and PIP2. In contrast, induction still works in Δpxa1 and Δec11 strains that are defective in two key steps for β-oxidative metabolism of fatty acids: peroxisomal import and enoyl isomerization (Fig. 2C). The Δfaa1/Δfaa4 double-mutant strain is defective in fatty acid uptake and activation to CoA derivatives, the first enzymatic step toward metabolism, because it lacks the two main acyl-CoA synthetases for C12–C18 fatty acids (27, 28). It is conceivable that the levels of endogenously synthesized oleate in these mutant cells are elevated because they are trying to cope with their inability to take up fatty acids. Indeed, we found that there is no further induction by added oleate beyond a dramatically elevated basal activity. This finding also suggests that acyl-CoA synthetases may contribute to keeping the basal activity of Oaf1/Pip2 low by controlling the concentration of unconjugated oleate. At this point, the oleate response profile was compatible with the idea that oleate acts directly as a ligand and activator of the Oaf1/Pip2 heterodimer.

Oleate Is a Ligand of the Oaf1 LBD. Direct binding of oleate to the hypothesized LBDs was then assessed in vitro by a ligand-binding assay with purified recombinant Oaf1 and Pip2 LBDs (Fig. 3 and Table 1). The Oaf1 LBD binds oleate with a $K_d$ of 16.5 nM, whereas it does not bind the noninducing saturated fatty acid stearate. Moreover, the corresponding Pip2 domain fails to bind oleate. The much higher affinity of the Oaf1 LBD compared with the EC$_{50}$ of the in vivo response may indicate that the oleate concentration that is freely available to Oaf1 inside cells is much lower than what is added to the medium.

Further Genetic Validation of the Structural Prediction. Our results strongly suggested that oleate might stimulate Oaf1 activity by acting as a ligand of the hypothesized LBD. We therefore sought further genetic evidence for the validity of the structural prediction. In animal NRs, helix 11 of the LBD forms the bottom of the ligand-binding pocket (1, 2). Because the topology of the

Table 1. Summary of the ligand-binding assays

<table>
<thead>
<tr>
<th>LBD</th>
<th>Fatty acid</th>
<th>Slope of fitted curve</th>
<th>Affinity, nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oaf1</td>
<td>Oleate</td>
<td>−60.651</td>
<td>16.5 ± 4.5</td>
</tr>
<tr>
<td></td>
<td>Stearate</td>
<td>10.046</td>
<td>ND</td>
</tr>
<tr>
<td>Pip2</td>
<td>Oleate</td>
<td>10.777</td>
<td>ND</td>
</tr>
<tr>
<td>Oaf1 E543A/A544S</td>
<td>Oleate</td>
<td>−10.102</td>
<td>ND (99 ± 150)</td>
</tr>
</tbody>
</table>

*Affinities $(K_d)$ with SEs were calculated from the slope of the fitted curves of Scatchard analyses, such as the analysis shown in Fig. 3. ND, not defined, indicates no binding (positive slopes or large SE).
samples of lanes 1, 2, and 3, respectively. Type and double-mutant Oaf1 and only the FLAG epitope were present for the antibody directed against the FLAG tag. Immunoprecipitated proteins were analyzed in total extracts of the same strains and for function in vivo.

The correlation between in vitro ligand binding and the in vivo response strongly supports the conclusion that the S. cerevisiae protein Oaf1 contains a functional domain hitherto thought to be restricted to animals (3, 5). By analogy, it is very likely that the putative LBD of Pip2 is a genuine LBD as well. Its LBD may either bind oleate with much lower affinity or bind an as yet unidentified ligand supporting oleate induction of the Oaf1-Pip2 heterodimer, or it may play a structural role within Pip2 itself or within the heterodimer. Note that only a minority of animal NRs have known ligands and that some NRs cannot bind any ligand at all because their “ligand-binding” pocket is filled with bulky amino acid side chains (29). The structural, molecular, and physiological analogies between the yeast Oaf1/Pip2 and the vertebrate PPAR/RXR heterodimers (30) are striking. In both groups of organisms, these heterodimeric transcription factors bind unsaturated fatty acids such as oleate and activate the metabolic pathway for fatty acids in response to a nutritional input of fatty acids. Surprisingly, their corresponding acyl-CoA esters behave as antagonists of PPAR/RXR, and acyl-CoA synthetase inhibitors result in increased PPAR/RXR activity in mammalian cells (31, 32), which is highly reminiscent of the effects of deleting the FAA1/FAA4 genes on Oaf1/Pip2 activity in yeast. Although we have not determined whether the acyl-CoA esters are able to bind Oaf1/Pip2 as well, it is possible that they contribute directly to lowering the basal activity as antagonists, in the absence of exogenously added fatty acids such as oleate. Thus, although the LBDs are embedded within transcription factors of a different architecture (Fig. 1A), they likely provide a common ancestral function as an allosteric switch.

The mechanistic details of how LBDs regulate the activities of the proteins within which they reside may have diverged considerably. Indeed, although many animal NRs (3) carry within their LBD a ligand-induced transcriptional activation function, referred to as AF-2, the activation domains of Oaf1 and Pip2 residues making up the putative helix 11 of Oaf1 could not be definitively modeled due to the distant nature of the structural relationship, we mutated all residues along the relevant portion of the predicted helix 11 from V537 to A544 (Fig. 4). The single-point mutants E543A and A544S are the only ones that display a diminished response to oleate (Fig. 4B and data not shown). Their induced activities are reduced to 25–50% of wild type. Their basal activities are reduced as well, consistent with the observation that the presence of wild-type Oaf1 (and Pip2) results in appreciable basal activity of the reporter gene, which may be due to endogenously produced, albeit low levels of, oleate. In the absence of any exogenously added oleate, the point mutants E543A and A544S might be too insensitive to respond at all. Remarkably, the response of the double mutant E543A/A544S, which combines both of the partially defective amino acid replacements, is completely defective (Fig. 4B). As shown in Fig. 3B, the ability of the double mutant E543A/A544S to bind oleate in vitro is correspondingly reduced. Thus, changing two amino acids in the predicted helix 11 at the bottom of the LBD abolishes both ligand binding and transcriptional response. These mutants still accumulate to normal levels, and even the defective double mutant is still able to heterodimerize with Pip2 (Fig. 4C and D), indicating structural integrity. This finding is consistent with the fact that Oaf1/Pip2 heterodimers can form and bind DNA in the absence of oleate (17–19).

Discussion

We have incorporated structural predictions in a bioinformatic screen of all available genome sequences for potential homologs of a particular protein domain. Having found candidate homologs for the LBD of NRs, we then identified a ligand and demonstrated that this ligand binds the predicted LBD in vitro and that predicted key residues are important for ligand binding in vitro and for function in vivo.

Fig. 4. Alteration of two amino acids within the predicted ligand-binding pocket abolishes the oleate response of Oaf1. (A) Schematic representation of the Oaf1 LBD, with magnification of a region of helix 11 that was probed by mutagenesis. The Oaf1 sequence is threaded onto the PPAR structure (Protein Data Bank ID code 1PRG). The ribbon representation of helix 11 is broken up around E543 (red) and A544 (purple) to reveal more detail. The C and F boxed in yellow are oleate-binding residues in the RXRα. The residues boxed in black in the PPARγ sequence (and marked with a red dot in the helical representation) represent ligand-binding residues of both PPARα and other NRs. (B) Oleate responses of Oaf1 point mutants. Both wild type and mutants were expressed with a FLAG epitope in a strain. (C) Expression of wild-type and mutant Oaf1 variants. Total extracts of the same strains analyzed in B were immunoblotted with antibodies against the FLAG epitope (Upper) and Hsp90 as an unrelated control protein (Lower). (D) The Oaf1 double mutant still heterodimerizes with Pip2. Association of FLAG-Oaf1 with Pip2-Myc was assessed by a coimmunoprecipitation experiment with an antibody directed against the FLAG tag. Immunoprecipitated proteins were revealed by immunoblotting with antibodies against the FLAG (Upper) and Myc (Lower) epitopes. Pip2-Myc was present in all three extracts, whereas wild-type and double-mutant Oaf1 and only the FLAG epitope were present for the samples of lanes 1, 2, and 3, respectively.
appear to reside exclusively at the very C terminus of the proteins (ref. 22 and data not shown) (Fig. 1A). It is conceivable that allosteric regulation by oleate relieves an inhibitory mechanism that blocks some step after DNA binding of the heterodimer by intra- and/or intermolecular interactions. The presence of auxiliary and inhibitory domains has previously been proposed on the basis of mutagenesis (22) and sequence (33) analyses. These domains partially overlap with the LBDs that we have defined here as a distinct structural and functional unit. Taken together, these findings point to an intricate intramolecular interplay between multiple domains. The same technology that enabled us to discover the Oaf1 and Pip2 LBDs by predicting a similar structural fold, despite poor primary sequence conservation, produced no significant hits beyond animals and fungi. Highly conserved orthologs are present in several other fungi of the phylum ascomycetes, notably of the order

**Materials and Methods**

**Plasmids.** Full-length OAF1 and PIP2 coding sequences were cloned from yeast strain W303. Our Oaf1 sequence differs in three amino acids (R70, Q447, and K588) from that deposited in the Saccharomyces Genome Database (www.yeastgenome.org). The latter sequence, derived from yeast strain S288C, may be deviant because our “variations” are conserved in several other Saccharomyces species. Both wild-type and point-mutant Oaf1 and Pip2, tagged with FLAG and 13 copies of the c-Myc epitope, respectively, were expressed from the TDH3 promoter in CEN/ARS plasmids. For expression in bacteria, sequences encoding the Oaf1 (amino acids 244–573) and Pip2 (amino acids 180–507) LBDs were cloned into vector pET-32EKLIC (Novagen). The luciferase reporter plasmid p2UG-2XORE-Luc was built from plasmid p2UG (35); it is a 2μ-based episome containing two tandem copies of the olate response elements from the FOX3 gene inserted upstream of a minimal CYC1 promoter driving the expression of firefly luciferase.

**Yeast Experiments.** Strain BY4741 was used as the wild-type strain and as a background for all yeast deletion strains except for the ∆sfa1 ∆sfa4 strain YB525 (27). For luciferase assays and the preparation of protein extracts, transformants were grown in rich medium with 1.5% raffinose, 1.5% glycerol, and 1% ethanol as carbon sources. Induction was done with 6 mM oleate overnight, unless indicated otherwise. Luciferase activities were determined in triplicate, essentially as described (36), except that cells were first washed once with water. Luciferase activities were normalized to the optical densities of the cultures.

**Ligand-Binding Assays.** Recombinant LBDs were purified from Escherichia coli as thioredoxin fusion proteins with His6 tags by affinity chromatography. Ligand-binding assays were done with oleate and stearate by using the ADIFAB free fatty acid indicator (Molecular Probes) as a fluorescent probe, as described (37). Briefly, the fluorescence of the modified fatty acid binding protein ADIFAB changes upon binding fatty acids. From this change, the free fatty acid concentration in the test solution at binding equilibrium can be calculated, as can, indirectly, the concentration of the fatty acid bound for a known total amount of the test LBD (for example, Oaf1t). Data from three to four replicates were analyzed by the Scatchard plot method.

We thank Natasha Kralli for having brought together the partners of this study. We thank Gisou van der Goot for help with the Fluorometer, Jeffrey I. Gordon for yeast strains, Katharina Strub for critical comments on the manuscript, Matt Couch for help with the bioinformatics analysis, and Richard Kidd for homology modeling. We are indebted to Pierre-André Briand for technical assistance. Work in the laboratory of D.P. was supported by the Canton de Genève and the Swiss National Science Foundation.