Crystal Structure of the HNF4α Ligand Binding Domain in Complex with Endogenous Fatty Acid Ligand*

Received for publication, July 23, 2002, and in revised form, August 20, 2002
Published, JBC Papers in Press, August 21, 2002, DOI 10.1074/jbc.C200420200

Sirano Dhe-Paganon‡, Karen Duda‡, Melanie Iwamoto, Young-In Chi‡, and Steven E. Shoelson

From the Joslin Diabetes Center & Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215

HNF4α is an orphan member of the nuclear receptor family with prominent functions in liver, gut, kidney and pancreatic β cells. We have solved the x-ray crystal structure of the HNF4α ligand binding domain, which adopts a canonical fold. Two conformational states are present within each homodimer: an open form with α helix 12 (α12) extended and collinear with α10 and a closed form with α12 folded against the body of the domain. Although the protein was crystallized without added ligands, the ligand binding pockets of both closed and open forms contain fatty acids. The carboxylic acid headgroup of the fatty acid ion pairs with the guanidinium group of Arg226 at one end of the ligand binding pocket, while the aliphatic chain fills a long, narrow channel that is lined with hydrophobic residues. These findings suggest that fatty acids are endogenous ligands for HNF4α and establish a framework for understanding how HNF4α activity is enhanced by ligand binding and diminished by MODY1 mutations.

Nuclear receptors are ligand-activated transcription factors that regulate such diverse physiological processes as reproduction, development, and metabolism. Ligand binding induces conformational changes that coordinately dissociates corepressors and recruits coactivators to enhance transcriptional activity (1, 2). Physiologically relevant ligands, which are known for less than half of the 48 nuclear receptors encoded by the human genome, include the steroid hormones, retinoids, thyroid hormone, vitamin D₃, and fatty acids. HNF4α is considered to be an orphan member of the nuclear receptor superfamily because its endogenous ligand is not known (3). Originally identified in liver, HNF4α is also present in kidney, gut, and pancreatic islets. It functions in transcriptional cascades, downstream of TGF-β-activated SMAD signaling and GATA-2 and HNF3/forkhead family transcription factors and upstream of the HNF1α/POU homeodomain transcription factor. In addition to HNF1α, the many target genes for HNF4α in liver include coagulation factors and proteins involved in lipid and cholesterol metabolism and transport. Selective targets present in kidney and intestine include cholesteryl ester transfer protein and intestinal fatty acid binding protein, respectively. Targeted disruption of Hnf4α in mice leads to defective gastrulation, underscoring an important role for HNF4α in the developing gut (4). In humans, the clinically apparent phenotype associated with Hnf4α mutations is a Mendelian form of diabetes mellitus associated with abnormalities in lipoprotein and lipid concentrations (5, 6).

Positional cloning methods revealed the relationship between HNF4α and an atypical form of diabetes referred to as maturity onset diabetes of the young (MODY1) (5). Mutations in at least six distinct genes have been linked to MODY (7). Although MODY1 is rare (<20 families have been identified) (5, 8, 9), the associated phenotype is instructive. Like other forms of MODY, MODY1 is characterized by autosomal-dominant inheritance, early onset (<25 years of age), and abnormal pancreatic β cell function. Because reduced β cell function leads to insulin deficiency, it is reasonable to conclude that HNF4α activity is tightly coupled with insulin synthesis and secretion. Since nuclear receptors are an established and valuable class of drug targets, it follows that selective agonists for HNF4α might be useful as new therapeutic strategies for improving insulin secretion in MODY patients and potentially others with more prevalent forms of type 2 diabetes. To learn more about how HNF4α functions in health and disease, and as an important step in drug discovery, we have crystallized the ligand binding domain (LBD) of HNF4α and solved its three-dimensional structure. HNF4α forms an antiparallel, three-layered α helical sandwich akin to other nuclear receptor LBDs. The serendipitous presence of a tightly associated fatty acid in the binding pocket reveals potential mechanisms for ligand binding.

MATERIALS AND METHODS

Protein Production and Crystallization—DNA encoding the LBD of rat HNF4α (residues 133–382) was subcloned into the pET 28a vector (Novagen) by PCR. Protein was expressed in Escherichia coli BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novagen) by PCR. Protein was expressed in BL21(DE3) (Invitrogen), grown either in LB or synthetic medium containing L-selenomethionine (SeMet), and isolated from bacterial lysates (Novag...
Table I

| Data and refinement statistics |
|--------------------------------|
| Crystal/derivative             |
| Native                         | Selenomethionine |
| Wavelength (Å)                 | 0.979            | 0.9587 |
| Resolution (Å)                 | 20–2.8           | 50–3.0 |
| Unique reflections             | 22,501           | 22,911 |
| Redundancy                     | 6.91             | 6.83   |
| Completeness*                  | 92.3 (78.0)      | 91.6 (73.2) |
| I(0)/Ib* (Å<sup>2</sup>)      | 15.84 (2.82)     | 15.41 (2.78) |
| R<sub>free</sub> (%)          | 11.3 (28.9)      | 12.5 (29.2) |
| Phasing power:                | 1.28/0.99        | 1.16/0.90 |
| Anomalous                     | 0.99             | 0.96 |

Refinement statistics

| Resolution                      | 20–2.8 Å         |
| Reflections                     | 29,727           |
| R-factor (%)                   | 23.2             |
| R<sub>free</sub> (%)          | 28.2             |
| Non-hydrogen atoms per asymmetric unit: |
| Protein                        | 7050             |
| Water                          | 74               |
| <B>                           | 55.1 Å<sup>2</sup> |

Root mean square deviation: bond length/bond angles 0.0078 Å/1.35°

a Data in parentheses are from the highest resolution shell.

b R<sub>free</sub> = Σ|I<sub>obs</sub> – |<I>|Σ<sub>|<I>|, calculated for all data.

c Centric/centric.

Results and Discussion

Domain Architecture and Structural Homology—The amino- and carboxyl-terminal boundaries of the LBD of HNF4α were delineated from sequence comparisons with other members of the nuclear receptor family. Although we attempted to express longer versions, including the entire LBD plus F-domain (133–455), these degraded to shorter fragments. Since susceptibility to proteolysis suggested that the F-domain is at least partially disordered, we expressed and crystallized the LBD domain (133–382) by itself. Crystals belonging to the tetragonal space group P4<sub>2</sub>2<sub>1</sub>2 (unit cell: a = b = 102.3, c = 227.7 Å) reached maximal dimensions of 0.3 × 0.2 × 0.2 mm in sitting drops within 5 days. The structure was solved using MAD phasing of selenomethionine-substituted protein (Fig. 1).

Four protein molecules are assembled in the asymmetric unit as a pair of identical homodimers. Each LBD of HNF4α contains 9–10 α helices and two β strands that adopt the helical sandwich motif common to the LBDs of nuclear receptors (Fig. 1). 194 core α carbons of the ligand binding domains from HNF4α and RXRα (Protein Data Bank number 1FBY) (16), a close structural relative, superimpose with a root mean square deviation of 1.26 Å. Following conventional nomenclature, α helix 2 (α2), which is variably present in nuclear receptor LBDs, is absent in HNF4α, α6 consists of a single α helical turn, and α10 and α11 are contiguous. Pro<sup>333</sup> forces a break in α10. Acidic residues Glu<sup>263</sup> and Asp<sup>262</sup> create a bulge in α5 that may be important for dimerization (17). Electron density is apparent for all residues except those at the amino (133–140) and carboxyl (368–382) termini of the domain and within the loop between α helices 1 and 3 (the 1/3 loop, residues 157–165).

Fatty Acids in the Ligand Binding Pocket of HNF4α—Although we did not add potential ligands to our crystallization trials, the initial electron density maps showed density in the ligand binding pocket of the domain that was not accounted for by the protein. The excess electron density, present in all molecules of the asymmetric unit, resembled a small molecule associated with the guanidinium group of Arg<sup>226</sup>. Matrix-assisted laser desorption ionization and time of flight mass spectrometric analyses of HNF4α revealed, in addition to the expected parent ion for the expressed protein, the presence of non-covalently associated saturated and monounsaturated fatty acids (data not shown). To confirm the identities of the lipids and quantify relative amounts, purified HNF4α solutions were submitted for fatty acid analysis by GC. Recombinant rat HNF4α (133–382) contained a mixture of fatty acids, including 16:0, 17:0 cyclo, 18:1ω7c, and 14:0 (Fig. 2). Similar sets of fatty acids were also found in lipids from HNF4α expressed in yeast and rat liver.

Fig. 1. Structure of the HNF4α ligand binding domain. A, a ribbon diagram of the homodimer reveals a "LBD" fold. Helices, shaded turquoise (α1–α8), yellow (α10), or red (α12), reveal two distinct conformations: open on the left and closed on the right. β Strands are colored green, and the fatty acids in the ligand binding pockets are colored magenta. B, the ligand and elements of secondary structure are colored similarly in the schematic representations, with open and closed conformations in the same orientations for comparison.
acids bound a longer form of rat HNF4α containing the entire carboxyl-terminal F domain (residues 133–455, data not shown) and the LBD of human HNF4α (residues 140–382), which differs from the rat domain at seven positions (Fig. 2). These findings demonstrate that both human and rat proteins associate spontaneously with endogenous fatty acids and that the F domain does not influence binding.

To determine whether the LBDs select specific fatty acids from the larger pool, we subjected E. coli BL21 cells to similar GC analyses (Fig. 2). The bacterial fatty acids present in greatest abundance also bound HNF4α, although there appears to be a degree of selectivity. For example, the fatty acids 16:0, 18:1ω7c, 16:1ω7c/15 iso 20:0, 15 iso 30H, and 19:0 cycle ω8c are present in greater relative abundance in the bacteria than associated with human or rat HNF4α (Fig. 2) (16:1 iso and 14:0 3OH, and 19:0 cycle ω8c are not distinguished by GC). In contrast, fatty acids 17:0 cyclo, 16:1 ω7c/15 iso 20:0, 15 iso 30H, and 10:0 were bound to the protein in higher relative abundance than their presence in E. coli. Other fatty acids such as 18:1ω7c and 14:0 were neither selected for nor against (Fig. 2). Therefore, although a wide range of fatty acids were found within its ligand binding pocket, HNF4α appears to exhibit selectivity toward a subset of those found in bacteria. Future experiments with mammalian tissues are required to determine which fatty acids bind under relevant conditions and whether the capacity of HNF4α to serve as a fatty acid sensor has physiological or pathological ramifications. The fact that the fatty acids were persistently bound even after a multistep purification indicated that they were tightly associated. This was borne out by additional attempts to strip the lipids from the protein, as the fatty acids remained associated after dialysis and gel-filtration chromatography (data not shown).

The serendipitous presence of fatty acids provided an opportunity to analyze how ligands bind to HNF4α. The lipid carboxyl group was readily identified next to the side chain of Arg226 and the methylene chain could be built to C12. Density beyond that point was weak, indicating that the remainder of the chain is either flexible or occupies multiple conformations. The cavity has an internal volume of 370 Å³, calculated by the program VMD (18), which is almost entirely occupied by fatty acid. While this volume is within the normal range for nuclear receptor LBDs, the elongated, relatively linear shape of the pocket is atypical. Both oxygens of the fatty acid head group are ion-paired with the guanidinium group of Arg226, and one oxygen forms additional hydrogen bonds with the backbone NH of Gly237 and side chain OH of Ser181 (Fig. 3, A and B). Residues that line the internal surface of the ligand binding pocket are largely hydrophobic, including those arising from α3 (Ile175, Val178, Cys179, Met342, α5 (Leu219, Leu220, Ala223), β1 (Leu236), β2 (Val242), loop 2/7 (Leu249), α7 (Met252, Val255, Ser256, Ile260), and α10 (Met242, Gln245, Ile246, Ile249) (Fig. 3B). In the “closed” conformation α12 of HNF4α does not contact the ligand, in contrast with other nuclear receptors where α12 in the canonical agonist-bound conformation often does contact the agonist.

Although HNF4α and RXRα are similar in terms of amino acid sequence, their ligand binding specificities are distinct. This is due in part to Phe213 in RXRα, which points inward and directs bound 9-cis-retinoic acid upward toward α5 (19). The corresponding residue in HNF4α is Ala213, which, presumably because of the decreased bulk of the side chain, allows the fatty acid ligand to curl downwards toward α7. It is thus particularly interesting to note that F313A substitution in RXRα leads to constitutive activation, due to the spontaneous binding of an endogenous fatty acid as we have seen for wild-type HNF4α (19). The fatty acid in the binding pocket of RXRα F313A adopts a U shape, and density is interpretable out to the C18 position. In contrast, the side chains from residues Met342 and Met344 fill the upper portion of the HNF4α ligand binding pocket to prevent the fatty acid from looping around to form a U, and we don’t see density past the C12 position. These two methionine residues of HNF4α are unique among nuclear receptors.
Helix α12 Adopts Two Conformations—The two molecules within each dimer adopt distinct conformations. In the first, referred to as the open state, α12 is fully extended and contiguous and collinear with α10 (Fig. 1). The second molecule of each homodimer adopts a closed conformation, with α12 situated against the body of the LBD. The conformation of the “open” state is fixed by crystal packing between two open forms at an axis of symmetry. In the closed state, residues Leu360, Leu361, Met364, and Leu365 form a hydrophobic face on α12 that interacts with a corresponding hydrophobic surface on the body of the domain created by residues α3 (Met182, Lys183, Leu186, Leu187, Leu188, and Val190) and α4 (Leu211, Ala215, Gly216, and Leu219). Residues Gln362 and Glu363 help to anchor α12 through hydrogen bonds with Lys350 in α10 and Arg322 in α4, respectively. Since fatty acid ligands similarly occupy the ligand binding pockets of molecules in open and closed conformations, the structural rearrangement does not appear to be required for ligand-mediated activation as has been proposed for other nuclear receptors (17, 20, 21). It is possible to lock the domain in its closed and active state, both ligand and coactivator must be bound, but since our structure lacks coactivator, it does not test this hypothesis.

The Dimer Interface—As opposed to many nuclear receptors that form heterodimers, HNF4α functions primarily as a homodimer (22). Consistent with this, the LBD of HNF4α crystallizes as a canonical homodimer with near-perfect 2-fold symmetry about the interface (Fig. 3C). A 1156-A2 surface of each monomer is buried at the dimer interface. A coiled coil interaction between α10 from each molecule dominates the interface, with hydrophobic side chain-side chain interactions (Phe325:Phe325, Leu326:Leu326, Ile328:Leu328, Leu330:Pro333:Pro333, and Trp340:Trp340) along its length (Fig. 3C). Intermolecular salt bridges (Glu363:Lys300, Arg303:Glu327, and Arg322:Asp312) and hydrogen bonds (Gln307:Gln327, Gln311:Gly323, and Gly336:Gln336) appear to contribute to the stability of the interface.

MODY1 Mutations—Of 14 mutations in HNF4α that have been found in patients with the MODY1 form of diabetes (5, 8, 9, 23), three encode substitutions within the LBD (V255M, E276Q, R324H). The side chain of Val255 (in sequence of the V255M substitution (24), the structure suggests a methionine side chain at this position would partially stabilize dimerization.

Conclusions—The LBD of HNF4α is structurally similar to the LBDs of other nuclear receptors. The presence of a fatty acid in its ligand binding pocket is especially noteworthy and should help resolve the controversy surrounding putative ligands for HNF4α. From a structural point of view, the fatty acid appears to be an ideally suited ligand for HNF4α. The carboxylate headgroup of the fatty acid ion pairs with the guanidinium group of Arg226 of the protein to fix the orientation, while the aliphatic portion of the fatty acid occupies a long narrow pocket that is lined with hydrophobic side chains. Considering this mode of binding, it is not obvious how more complex ligands might bind HNF4α. It is thus reasonable to presume that fatty acids or related molecules might be the endogenous ligands for HNF4α. The high apparent affinity of recombinant HNF4α for fatty acids suggests a similar situation for the native protein in mammalian cells, providing a mechanism for HNF4α to function as a biosensor for fatty acids within the cell. The structural features of fatty acid binding to HNF4α provide a useful starting point for identifying potential pharmacophores, but the search may prove all the more challenging due to the kinetics of ligand binding.
Crystal Structure of the HNF4α Ligand Binding Domain in Complex with Endogenous Fatty Acid Ligand
Sirano Dhe-Paganon, Karen Duda, Melanie Iwamoto, Young-In Chi and Steven E. Shoelson

J. Biol. Chem. 2002, 277:37973-37976.
doi: 10.1074/jbc.C200420200 originally published online August 21, 2002

Access the most updated version of this article at doi: 10.1074/jbc.C200420200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 23 references, 7 of which can be accessed free at http://www.jbc.org/content/277/41/37973.full.html#ref-list-1