Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 10421–10423, November 1993
Review

Review

Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: A family of transcription factors of diverse biologic function

Eseng Lai, Kirk L. Clark†, Stephen K. Burley†, and James E. Darnell, Jr.‡*

**Cell Biology and Genetics Program and Department of Medicine, Memorial Sloan–Kettering Cancer Center, New York, NY 10021; and †Laboratories of Molecular Biophysics, Howard Hughes Medical Institute, and ‡Laboratory of Molecular Cell Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021*

ABSTRACT **A family of transcription factors, first identified as hepatocyte nuclear factors (HNF-3α, -3β, and -3γ) and as a homeotic *Drosophila* mutant, fork head, has been intensively studied for the past 4 years. Important findings have emerged about the structure of the DNA-binding portion of the proteins as well as biologic discoveries about the diversity of the family and its implied role in early development.**

Discovery of the Proteins

The hepatocyte nuclear factor 3 (HNF-3) proteins from rats and mice claimed interest originally because DNA-binding sites for these proteins were required for the hepatocyte-specific expression of two genes expressed at high levels in the liver. HNF-3 proteins were present in nuclear extracts from liver but absent in other tissues—i.e., brain and kidney (1). Because of this limited tissue distribution, these site-specific DNA-binding proteins were candidates to participate in cell-specific gene expression. Therefore one liver protein that formed a prominent, specific DNA–protein complex was purified. Peptide sequences were obtained and a cDNA encoding the HNF-3α gene was obtained (2). The cloning of two other similar cDNAs from genes HNF-3β and HNF-3γ, which were also expressed in liver, followed shortly (3). The highly conserved DNA-binding region that characterized these three proteins (90% identity in amino acids) matched a region of the already known sequence of a *Drosophila* nuclear protein termed fork head. The fork head gene was so named because embryos mutant in the gene were defective in the formation of terminal (end) structures that normally give rise to the anterior and posterior gut, resulting in two spiked head structures (4). Because of the defect in gut development in fork head mutants and because the HNF-3α, -3β, and -3γ mRNAs were found in various organs that originate from the gut endoderm in mice and rats, the genes caught the interest of developmental biologists.

The sporadic reports of individual genes clearly related by sequence to the HNF-3/fork head genes and purposeful searches by systematic low-stringency screens and PCR screens have resulted in the identification of at least seven additional family members in *Drosophila* and more than two dozen additional family members in vertebrates including frogs, rats, mice, and humans. Shown in Table 1 are representative members of this rapidly growing gene family. All of these proteins are characterized by a highly conserved region of 100 amino acids, which has been shown for HNF-3 proteins and several other members of the family to be necessary and sufficient for DNA binding (2).

Two lines of recent experiments have greatly enhanced our knowledge of this protein family. *In situ* hybridization and antibody studies in various tissues of adults and embryos have suggested wide-ranging roles for various members of the family during development and a specific very early developmental role for HNF-3α and -3β and BF-1 in mammals. The three-dimensional structure of the DNA-binding domain has been determined for one of the HNF-3 proteins, revealing a new DNA-binding structure. In this review we will summarize the recent biochemical, structural, and biologic information about this protein family.

Analysis of the Primary Structure and Function

From the recently reported sequences of additional family members, which include 1 yeast protein as well as about 30 other invertebrate and vertebrate proteins, it is clear that this family is quite large. About half of the residues in the 100-amino-acid binding domain are conserved in all family members. Since certain replacements are common, evolutionarily related subgroups within the family can be recognized (13). Despite the strong sequence similarity in the DNA-binding domain, there is a surprising range of DNA sequences recognized by family members. Even among the first three proteins cloned, HNF-3α, -3β, and 3γ, differences in affinities for different oligonucleotides were observed (3). Furthermore, while the HNF-3 proteins are able to bind to a range of sequences with high affinity, one member of the family, BF-1, has been shown to be more selective in its binding site recognition (5). These observations suggest that many of these highly related proteins can serve different functions through subtly distinct interactions with different DNA-binding sites.

A second unsettled issue concerning the structure of the proteins in the family is the function of parts of the molecule other than the DNA-binding domain. HNF-3α, -3β, and 3γ and BF-1 are positive-acting proteins in cotransfection assays (refs. 2 and 3; W. Tao and E.L., unpublished observations). In nonhepatic cells, J. Philippe and V. R. Prezioso (personal communication) found an inhibitory effect of HNF-3α and -3β in cells of pancreatic origin on the promoter of the glucagon gene. Thus the proteins of this gene family, at least in transfection assays, appear to be transcriptional activators in most situations but not exclusively so. Sequence comparisons show that HNF-3α, -3β, and -3γ and fork head have two peptide regions near their carboxyl termini that are quite similar (3, 18). One of these, region II, is also found in at least six additional family members. In three of these, BF-1, slp1 (sloppy paired 1), and slp2, this region is located at the amino terminus. Deletion studies carried out with HNF-3β suggest several regions of the protein that contribute to its transcriptional activation function, including region II (19).

Antibody staining confirms these transcription factors to be localized to the nucleus. In HNF-3α, four basic amino acids near the end of the DNA-binding region and a second amino-terminal region are required for nuclear translocation (W. S. Chen, V. R. Prezioso, and J.E.D., unpublished observations).

Abbreviations: HNF-3, hepatocyte nuclear factor 3; WH, winged helix; BF-1, brain factor 1; XFKH, *Xenopus* fork head; TTR, transthyretin.

Table 1. Expression and function of winged helix (WH) proteins

Gene	Species	Expression pattern and function	Ref(s).
HNF-3α	Rat, mouse	Gut endoderm-derived tissues in adult	2, 3
HNF-3β		HNF-3β in Hensen's node of gastrulating embryo	
HNF-3γ			
BF-1	Rat, mouse	Telencephalon of developing and adult brain	5
ILF	Human	Binds to sites in interleukin promoter and human immunodeficiency virus long terminal repeat	6
XFKH1	*Xenopus*	Dorsal blastopore lip of early gastrulae, induced by activin	7
qin	Chicken	Avian sarcoma virus oncogene, homolog of BF-1	8
Axial	Zebrafish	Developing body axis, induced by activin	9
Fork head	*Drosophila*	Terminal differentiation of embryo, expressed in gut and central nervous system	4
Sloppy paired 1	*Drosophila*	Segmentation of the embryo	10
Sloppy paired 2			
lin-31	*C. elegans*	Specification of cell fate in vulval development	11
Hcm1p	Yeast	Enhances function of mutant calmodulin	12

Additional members of this gene family are described in refs. 13–17. BF-1, brain factor 1; ILF, interleukin binding factor; XFKH1, *Xenopus* fork head 1.

X-Ray Crystallographic Analysis

Given the high degree of sequence conservation within the DNA-binding domain of members of the HNF-3/fork head family, a 115-residue DNA-binding fragment of HNF-3γ was cocrystallized with a duplex oligonucleotide bearing the sequence GACTAAGTCAACC from the HNF-3γ binding site in the transthyretin (TTR) promoter (1). Fig. 1 illustrates the three-dimensional structure, which was determined by multiple isomorphous replacement and refined at 2.5 Å resolution (21). The polypeptide chain binds DNA as a monomer, which is folded into a compact structure with its amino and carboxyl termini on opposite faces of the molecule. Three α-helices (H1, H2, and H3) dominate the amino-terminal half of the protein and form a globular, three-helix cluster with a well-defined hydrophobic core. Between H1 and H2 is a β-strand (S1), which interacts with strands S2 and S3 to form a three-stranded, twisted, antiparallel β-sheet that lies against the DNA. The loop connecting H2 and H3 runs along the phosphodiester backbone and has been denoted T′. Strands S2 and S3 of the β-sheet are connected by a loop (W1), which interacts with the phosphate backbone of DNA. An extended polypeptide chain emerges from the carboxyl terminus of S3 to form another long loop (W2), which makes a variety of contacts, including those with the minor groove, and then disappears in a solvent channel. Analysis of the DNA reveals a bend of about 13°, narrowing the major groove in which H3 is located. HNF-3γ interacts with DNA as a monomer, over a linear distance of about 40 Å along the axis of the double helix. This extensive protein–nucleic acid interaction surface compares favorably with the results of free-radical and DNase I footprinting studies of HNF-3γ bound to the TTR promoter (1, 2). Points of contact between protein and DNA occur on both DNA strands and are distributed throughout the linear sequence of the polypeptide chain. α-Helices H1 and H2 are oriented with their amino termini pointing toward the negatively charged phosphate backbone of the DNA. The principal contact surface is provided by α-helix H3, which lies in the major groove and is flanked by two long loops (W1 and W2) that touch the DNA backbone and base atoms. The second loop (W2) projects an arginine side chain into the minor groove of the DNA making a side chain–base contact. The structure of residues 107–223 of HNF-3γ can be likened to a butterfly with an α-helical thorax and two wing-like loops and has been referred to as the "winged-helix" (WH) motif. Protein–DNA contacts with the nucleic acid backbone include both phosphate and ribose groups. These side chain–backbone contacts are typical of protein–DNA complexes and are thought to make largely nonspecific contributions to DNA binding (reviewed in ref. 22). HNF-3γ makes a combination of direct and water-mediated side chain–base contacts with the sequence **TAAGTCA**, which lies in the middle of the HNF-3γ DNA footprint and corresponds to positions −105 to −99 of the TTR promoter (1). The three bases that interact with helix H3 (indicated in bold type above) also occur within two known high-affinity HNF-3 binding sites found in the α-1-antitrypsin and HNF-1α promoters. The relationship between the WH motif and the helix–turn–helix motif is reviewed in ref. 23.

Proteins containing this structural motif can be described as winged helix (WH) proteins much as other groups of transcription factors are described as basic leucine zipper (bZIP), helix–loop–helix (HLH), helix–turn–helix, or homeodomain proteins.

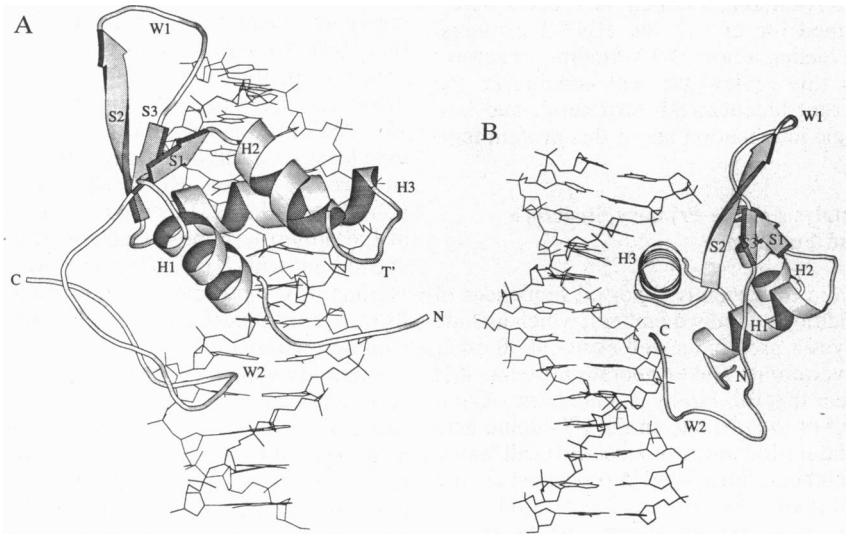

FIG. 1. MOLSCRIPT (20) cartoons of the three-dimensional structure of the DNA-binding domain of HNF-3γ complexed with a 13-base-pair oligonucleotide. α-Helices and β-strands are labeled with H and S, respectively. The short loop between α-helices H2 and H3 is designated T′, and the two long loops are labeled W1 and W2. (*A*) View showing the β-sheet and three α-helices and the amino and carboxyl termini of the protein. (*B*) As for *A*, using a view rotated ≈90° about the vertical to show helix H3 binding in the major groove of DNA.

Diversity in Biologic Function of the WH Proteins

As noted earlier, initial studies suggested that the first members of this family were involved in gut development in *Drosophila* and in mammals. It is now apparent that distinct WH proteins are found in many, if not all, tissues. The expression pattern where examined is usually temporally and spatially restricted, consistent with roles in the development or differentiation of particular groups of cells. Table 1 lists the known functions of some representative family members. The original HNF-3 proteins were isolated from and appear to have important function in the differentiated hepatocytes of adults. However more recent studies (see below) suggest that they play additional roles in early development. Other members of the family also apparently have multiple functions. For example, the sloppy paired locus in *Drosophila* is expressed in distinctly different patterns at different stages in development. The recent finding that the oncogene from a newly characterized chicken sarcoma virus, *qin*, is a homolog of BF-1 along with the observation that BF-1 expression is highest in rapidly proliferating, undifferentiated neuroepithelial cells suggests that some WH proteins may regulate cell proliferation as well as cell differentiation.

Early Embryonic Expression

In vertebrates, genetic screens of the type used in *Drosophila* or *Caenorhabditis elegans* to identify the earliest acting embryonic genes are not practical, so developmental biologists rely on other clues about function for candidate genes (e.g., expression patterns in early embryogenesis followed by targeted attempts to affect function of the product of such genes). Thus two HNF-3 family members, XFKH1 and pintallavis, were obtained from cDNA libraries of early frog gastrulae and were shown to be expressed in the lip of the blastopore, where cells that form the embryo are first determined (7, 14). XFKH1 expression was shown to be rapidly induced in the presence of cycloheximide by activin treatment of animal caps. Activin has been shown to induce axis formation in animal caps. Injection of pintallavis RNA leads to overexpression of the protein and causes improper axial development. Next from a rat foreplate library came a cDNA clone that exactly matched the HNF-3β sequence except for an amino-terminal extension of six amino acids, which was also found in XFKH1. HNF-3β was then found to be expressed in Henson's node, one of the earliest coherently behaving group of vertebrate embryonic cells through which cells migrate before separating into the three primitive germ layers, the endoderm, ectoderm, and the mesoderm (ref. 15; A. Ruiz i Altaba, V. R. Prezioso, J.E.D., and T. Jessell, unpublished results). Expression of HNF-3β continues in the notochord and in the foreplate cells of the developing neural tube. The primitive endoderm, just ventral to the notochord, also continues to express HNF-3β. Expression of HNF-3α is almost as early as HNF-3β, but HNF-3γ is definitely not present in these early cells. While the detailed developmental pattern of expression during organogenesis of the various HNF-3 family members has not been clarified, it is clear that late in embryogenesis, in newborns, and in adults, there is a differential distribution of HNF-3α, -3β, and -3γ in different cells derived from the primitive endoderm. Finally it has recently been found that in addition to the large number of HNF-3 fork head family members there are multiple mRNAs that arise from at least the HNF-3β and BF-1 genes, probably as a result of differential splicing within the same primary transcript as well as variation in start sites. It seems likely that some or all of these variations may be significant in the course of development.

Summary

In the 4 years since the WH family was initially described, it has been found to be made up of many members expressed in a wide range of tissues. Many are likely to have important developmental roles. The detailed structure of the protein–DNA complex now available will facilitate rapid progress in determining exact functions for family members from both the biologic and structural point of view.

Our work was supported by the American Cancer Society (E.L.), the National Institutes of Health (J.E.D.), and the Howard Hughes Medical Institute (S.K.B.).

1. Costa, R. H., Grayson, D. R. & Darnell, J. E., Jr. (1989) *Mol. Cell. Biol.* **9,** 1415–1425.
2. Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H. & Darnell, J. E., Jr. (1990) *Genes Dev.* **4,** 1427–1436.
3. Lai, E., Prezioso, V. R., Tao, W., Chen, W. S. & Darnell, J. E., Jr. (1991) *Genes Dev.* **5,** 416–427.
4. Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. & Jackle, H. (1989) *Cell* **57,** 645–658.
5. Tao, W. & Lai, E. (1992) *Neuron* **8,** 957–966.
6. Li, C., Lai, C., Sigman, D. S. & Gaynor, R. B. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7739–7743.
7. Dirksen, M. L. & Jamrich, M. (1992) *Genes Dev.* **6,** 599–608.
8. Li, J. & Vogt, P. K. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4490–4494.
9. Strahle, U., Blader, P., Henrique, D. & Ingham, P. (1993) *Genes Dev.* **7,** 1436–1446.
10. Grossniklaus, U., Pearson, R. K. & Gehring, W. J. (1992) *Genes Dev.* **6,** 1030–1051.
11. Miller, L. M., Gallegos, M. E., Morisseau, B. A. & Kim, S. K. (1993) *Genes Dev.* **7,** 933–947.
12. Zhu, G., Muller, E. L. D., Amacher, S. L., Northrop, J. L. & Davis, T. N. (1993) *Mol. Cell. Biol.* **13,** 1779–1787.
13. Clevidence, D., Overdier, D., Tao, W., Qian, X., Pani, L., Lai, E. & Costa, R. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 3948–3952.
14. Ruiz i Altaba, A. & Jessel, T. (1992) *Development* **116,** 81–93.
15. Sasaki, H. & Hogan, B. (1993) *Development* **118,** 47–59.
16. Li, C., Lusis, A., Sparkes, R., Tran, S. & Gaynor, R. (1992) *Genomics* **13,** 658–664.
17. Hacker, U., Grossiniklaus, U., Gehring, W. & Jackle, H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8754–8758.
18. Weigel, D. & Jackle, H. (1990) *Cell* **63,** 455–456.
19. Pani, L., Overdier, D. G., Porcella, A., Qian, X., Lai, E. & Costa, R. H. (1992) *Mol. Cell. Biol.* **12,** 3723–3732.
20. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
21. Clark, K., Halay, E., Lai, E. & Burley, S. (1993) *Nature (London)* **364,** 412–420.
22. Pabo, C. O. & Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61,** 1053–1095.
23. Brennan, R. G. (1993) *Cell* **74,** 773–776.