

# Analysis of hepatocyte nuclear factor-3 $\beta$ protein domains required for transcriptional activation and nuclear targeting

Xiaobing Qian<sup>+</sup> and Robert H. Costa<sup>\*</sup>

Department of Biochemistry, College of Medicine, University of Illinois at Chicago, 1819 W. Polk Street, Chicago, IL 60612-7334, USA

Received December 8, 1994; Revised and Accepted February 22, 1995

## ABSTRACT

**Three distinct hepatocyte nuclear factor 3 (HNF-3) proteins ( $\alpha$ ,  $\beta$  and  $\gamma$ ) regulate transcription of the transthyretin (TTR) and numerous other liver-specific genes. The HNF-3 proteins bind DNA via a homologous winged helix motif common to a number of developmental regulatory proteins including the *Drosophila* homeotic fork head (*fkh*) protein. The mammalian HNF-3/*fkh* family consists of at least thirty distinct members and is expressed in a variety of different cellular lineages. In addition to the winged helix motif, several HNF-3/*fkh* family members also share homology within transcriptional activation region II and III sequences. In the present study we further define the sequence boundaries of the HNF-3 $\beta$  N-terminal transcriptional activation domain to extend from amino acids 14 to 93 and include conserved region IV and V sequences. We also demonstrate that activity of the HNF-3 N-terminal domain was diminished by mutations which altered a putative  $\alpha$ -helical structure located between amino acid residues 14 and 19. However, transcriptional activity was not affected by mutations which eliminated two conserved casein kinase I sites or increased the number of acidic amino acid residues in the N-terminal domain. Furthermore, we determined that the nuclear localization signal overlaps with the winged helix DNA-binding motif. These results suggest that conserved sequences within the winged helix motif of the HNF-3/*fkh* family may be involved not only in DNA recognition, but also in nuclear targeting.**

## INTRODUCTION

Cellular differentiation results in transcriptional induction of distinct sets of tissue-specific genes whose expression is critical for organ function. Deciphering transcriptional control mechanisms is thus essential to understanding differentiation and

development. Toward this goal, functional dissection of numerous hepatocyte-specific promoter and enhancer regions has revealed that they are structurally complex, consisting of multiple DNA binding sites recognized by distinct families of liver-enriched transcription factors (1). The combinatorial action of these factors on multiple DNA sites is required for the activation of transcription and plays a role in maintaining hepatocyte-specific gene expression. Structural similarities in the DNA-binding and/or dimerization domains define five distinct families of liver-enriched factors: the hepatocyte nuclear factor 3 (HNF-3 $\alpha$ ,  $\beta$  and  $\gamma$ ) proteins (2,3) that use the winged helix DNA-binding motif (4), the steroid hormone receptor family members HNF-4 (5) and ApoAI regulatory protein-1 (ARP-1; 6) that use the zinc finger motif; the POU-homeodomain containing HNF-1 $\alpha$  and  $\beta$  (7,8); the basic leucine zipper (bZIP) family members C/EBP $\alpha$  (9), C/EBP $\beta$  and C/EBP $\delta$  (10-15) and the PAR bZip subfamily DBP (16), VBP/TEF (17,18) and HLF (19,20).

The HNF-3 proteins were originally identified as factors mediating liver-specific transcription of the transthyretin [TTR; (21,22)] gene and were later shown to control the expression of numerous hepatocyte-specific genes (1,23). They bind to the consensus DNA binding site A(A/T)TRTT(G/T)RYTY as monomers through a 100 amino acid winged helix DNA-binding domain (4,24). *In situ* hybridization studies of stage-specific embryos suggest that the HNF-3 proteins function in endodermal determination as well as in the formation of the neural tube and notochordal mesoderm (25-28). In support of the importance of HNF-3 expression in these developing structures, HNF-3 $\beta$  mutant embryos lack properly formed node and notochord, leading to defects in neurotube and somite organization (29,30). Although these mutants are not disrupted in definitive endoderm, they fail to form gut endoderm which give rise to organs such as the liver, lung, pancreas and intestine. HNF-3 also participates with HNF-4 in the hierarchical transcriptional activation of HNF-1 in the hepatocyte, verifying its role in cellular differentiation (31). Furthermore, HNF-3 $\alpha$  induction is critical for differentiation of H2.35 hepatocyte cells (25,32,33) as well as retinoic acid mediated differentiation of F9 embryonic carcinoma cells (34). Clearly, deciphering the mechanism by which HNF-3

<sup>\*</sup> To whom correspondence should be addressed

<sup>+</sup>Present address: Gladstone Institute of Cardiovascular Disease, Cardiovascular Research Institute, University of California, San Francisco, CA 94141-9100, USA

proteins activate transcription is important for understanding their role in differentiation.

Mammalian HNF-3 (2,3) and the *Drosophila* homeotic gene *fork head* (*fkh*) (35) are prototypes of a family of transcription factors sharing homology in their winged helix DNA-binding domain (4,23). Currently, the HNF-3/*fkh* family consists of more than 30 distinct genes that are expressed in a diverse group of mammalian cell lineages and possess distinct DNA binding specificities (23–25,28,36–45). Furthermore, we previously defined an HNF-3 $\beta$  transcriptional activation domain at the C-terminus (amino acids 361–458) containing the conserved region II and III sequences as well as a second activation domain at the HNF-3 $\beta$  N-terminus which includes conserved region IV sequences (46).

In this study we precisely define the N-terminal and C-terminal sequence boundaries of the HNF-3 $\beta$  N-terminal transcriptional activation domain. Analysis of site-directed mutants demonstrates that activity of the HNF-3 $\beta$  N-terminal domain is not influenced by increasing the number of acidic residues, but may depend on maintenance of a putative  $\alpha$ -helical protein structure. We also show that the nuclear localization signal overlaps with the winged helix DNA-binding domain, suggesting a dual role for this protein motif. We propose that sequence constraints of the winged helix motif may be required to preserve both DNA binding activity and nuclear targeting function during the evolution of the HNF-3/*fkh* family.

## MATERIALS AND METHODS

### Transfections, gel mobility shift and CAT assays, and construction of HNF-3 $\beta$ mutants

Human hepatoma HepG2 cells (47) were grown and transfected by calcium phosphate precipitation as described previously (46). Cytoplasmic extracts were assayed for chloramphenicol acetyltransferase (CAT) enzyme levels 48 h after HepG2 transfection of HNF-3 $\beta$  expression plasmids (4  $\mu$ g; described below) with the 4X HNF-3-TATA-CAT reporter plasmid (40  $\mu$ g; 46). CAT enzyme levels were determined in cytoplasmic extracts using [<sup>14</sup>C]chloramphenicol (ICN) and *n*-butyryl coenzyme A (Pharmacia), followed by xylene extraction of *n*-butyryl chloramphenicol and determination of product formation via liquid scintillation counting (Promega Technical Bulletin, 1994). The CMV driven  $\beta$ -galactosidase plasmid (1  $\mu$ g) was included in each transfection to normalize extracts for differences in transfection efficiency as described previously (46).

Site-directed mutations of the N-terminal transactivation domain of HNF-3 $\beta$  protein were created from the CMV HNF-3 $\beta$  expression plasmid (46) using polymerase chain reaction (PCR)-mediated mutagenesis (PCR; 48). This method involves two sequential PCR reactions: the first generates two PCR products which completely overlap on one end and defines the site of mutagenesis; in the second reaction, the ends of the PCR products will hybridize and prime synthesis of the intact product which is amplified by the outside primers. Our first PCR employed combinations of either CMV 5' sense and HNF-3 $\beta$  antisense mutant primers or HNF-3 $\beta$  sense mutant primer and the HNF-3 $\beta$ .86-81 antisense primer, and our second PCR amplified the final product with the CMV 5' sense and HNF-3 $\beta$ .86-81 antisense outside primers. The resulting mutant HNF-3 $\beta$  PCR product was digested with *Eco*RI and *Sst*II (amino acids 1–45) and ligated to *Eco*RI/*Sst*II digested CMV plasmid containing HNF-3 $\beta$  cDNA (coding amino acids 45–458). Double CKI

mutants were made using the HNF-3 $\beta$  mutant 16-17AA as a template for PCR mutagenesis. All site-directed mutations were verified by dideoxy DNA sequencing using Sequenase DNA polymerase (US Biochemicals). The following mutagenesis oligonucleotides were used:

MUT.16-17 Sense: 5'-GAGCCTGAGGGCTACG(AC)TG(AC)CGTGAGCAA-CATGAAC; (CKI) Antisense: 5'-GTTCATGTTGCTCACG(GT)CA(GT)CGTAG-CCCTCAGGCTC; MUT.26-27 Sense: 5'-GAGCCATCCGACTGGG(AC)CG-(AC)CTACTACGCGGAGCCT; (CKI) Antisense: 5'-AGGCTCCGCGTAGTAG-(GT)CG(GT)CCAGTCGGATGGCTC; MUT.18-19 Sense: 5'-TCCGACTGGA-GCAGC(CG)CC(CG)CCGCGGAGCCTGAGGGC; (YY) Antisense: 5'-GCCCTCAGGCTCCGCGG(CG)GG(GC)GCTGCTCCAGTCGGA; HNF-3 $\beta$ .86-81 antisense: 5'-GGCGCCCGCGCCCGGGA

Construction of N-terminal HNF-3 $\beta$  deletions was performed using PCR and specific sense primers which introduced a 5' *Bam*HI site (see below) and the antisense HNF-3 $\beta$ .257-252 primer. The resulting PCR product was joined with the entire cDNA at a unique *Bgl*III site and cloned into the CMV AUG+1 expression vector (46). Two *Bam*HI containing N-terminal fragments were made by PCR amplification using the CMV primer and an antisense primer spanning HNF-3 $\beta$  amino acids 48–55 or 86–93. Internal deletions were made by fusing the two N-terminal fragments to the HNF-3 $\beta$  153–458 sequences employing a common *Bam*HI site. The following primers were used:

HNF-3 $\beta$ .*Bam*HI.20-25: 5'-CGCGGATCCGCGGAGCCTGAGGGCTAC  
HNF-3 $\beta$ .*Bam*HI.28-33: 5'-CTCGGATCCGTGAGCAACATGAACGCC  
HNF-3 $\beta$ .*Bam*HI.54-59: 5'-CGCGGATCCAGTGGTTCCGGCAACATG  
HNF-3 $\beta$ .*Bam*HI.153-158: 5'-CGCGGATCCCGCAGCTACACTACGCC  
HNF-3 $\beta$ .*Bam*HI.93-86: 5'-CGCGGATCCGCTCATGCCGCCATGGC  
HNF-3 $\beta$ .*Bam*HI.55-48: 5'-CGCGGATCCACTGCCATTGACGCCG  
HNF-3 $\beta$ .257-252: 5'-CTCACACTTGAAGCGCTT

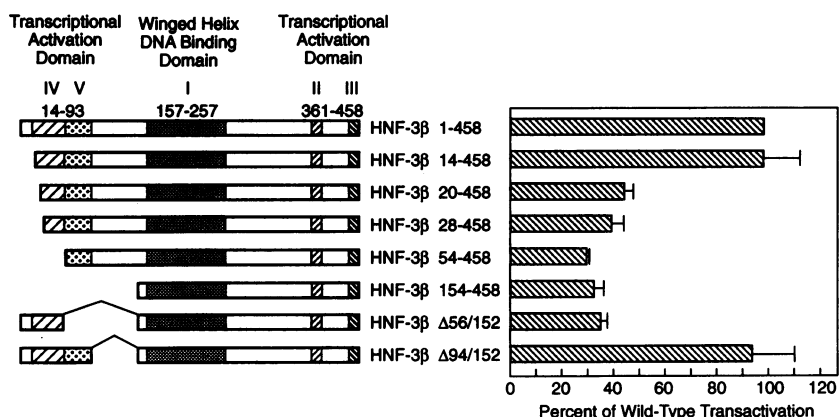
### Indirect immunofluorescence and $\beta$ -galactosidase staining

HepG2 cells grown on glass coverslips were transfected with expression plasmids containing N-terminal and C-terminal HNF-3 $\beta$  deletions (46) and immunofluorescence microscopy was performed as described (49). The CMV promoter driven  $\beta$ -galactosidase–HNF-3 $\beta$  plasmids were constructed from the HNF-3 $\beta$  153-458 N-terminal deletion using CMV sense primer and the antisense primers HNF-3 $\beta$ .*Sal*I.257-252 and HNF-3 $\beta$ .*Sal*I.169-164. These PCR products created a unique *Sal*I site which was used to make an in frame fusion with the N-terminal sequences of the  $\beta$ -galactosidase gene. The SV40 T-antigen nuclear localization signal fused to the  $\beta$ -galactosidase gene was a generous gift from Terry Van Dyke, University of Pittsburgh. HepG2 Cells were transfected with  $\beta$ -galactosidase–HNF-3 $\beta$  fusion protein and  $\beta$ -galactosidase activity was detected on fixed cells as described (50). The sequences of the HNF-3 $\beta$  primers are:

HNF-3 $\beta$ .*Sal*I.257-252: 5'-CTCGTCCGACTCACACTTGAAGCGCTTC  
HNF-3 $\beta$ .*Sal*I.169-164: 5'-CTCGTCCGACGTGATGAGCGAGATG

### Two dimensional NEPHGE/SDS-PAGE electrophoresis separation of proteins

Nuclear proteins from rat liver were analyzed by 2-D non-equilibrium pH gradient gel electrophoresis (NEPHGE)/SDS-PAGE as described by O'Farrell *et al.* (51). After 2-D electrophoresis, proteins were electrotransferred to nitrocellulose,



**Figure 1.** HNF-3 $\beta$  deletion mutants define the N-terminal activation domain. Summary of transcriptional activities produced by various HNF-3 $\beta$  deletions tested in HepG2 cell cotransfection assay (46). Schematically shown are wild-type and deleted HNF-3 $\beta$  proteins as well as conserved sequences which include: region I (winged helix DNA-binding domain) and region II, III (transcriptional activation domain) and regions IV and V whose involvement in transcriptional activation is demonstrated in this figure. The caret indicates two internal deletions of the N-terminus (HNF-3 $\beta$   $\Delta$ 56/152 and  $\Delta$ 94/152). The bar-graph shows the transcriptional activity of HNF-3 $\beta$  deletion mutants determined by normalized CAT assays derived from the average of at least three independent cotransfection experiments and expressed as percent of HNF-3 $\beta$  wild-type transcriptional activity.

incubated with 1:500 dilution of HNF-3 $\beta$  antisera (34) and then visualized with an alkaline phosphatase (AP)-conjugated secondary antibody (1:8000 dilution) and stained for enzymatic activity as described by the manufacturer (Boehringer Mannheim).

#### Metabolic labeling of cells, immunoprecipitation and phosphoamino acid analysis

HepG2 cells were transfected with 20  $\mu$ g of CMV HNF-3 $\alpha$  or HNF-3 $\beta$  expression plasmids and 40 h later, biosynthetic labeling of cellular protein was performed. For  $^{35}$ S-metabolic labeling, cells were washed twice with Tris-buffered saline, placed in methionine-free DMEM containing 7% fetal calf serum for 1 h, and then labeled for 4 h in methionine-free medium containing 100  $\mu$ Ci/ml [ $^{35}$ S]methionine (*trans*  $^{35}$ S-label; ICN). For  $^{32}$ P-labeling, cells were incubated in MEM lacking sodium phosphate (GIBCO) for 1 h and then cultured in fresh phosphate-free MEM containing 500  $\mu$ Ci/ml [ $^{32}$ P]orthophosphate (ICN) for 4 h.

Cells were rinsed twice with ice-cold PBS and lysed in 1 ml of RIPA buffer which consists of 150 mM NaCl, 50 mM Tris-HCl, pH 8.0, 1% NP-40, 0.5% deoxycholic acid, 0.1% SDS, 1 mM DTT, protease inhibitor cocktail (1  $\mu$ g/ml leupeptin and pepstatin A, and 0.5 mM phenylmethylsulfonyl fluoride) and phosphatase inhibitor cocktail (100 mM sodium fluoride, 2 mM sodium vanadate and 20 mM sodium pyrophosphate). HNF-3 $\alpha$  or HNF-3 $\beta$  proteins were immunoprecipitated using 50  $\mu$ l of Protein A-agarose beads (Bethesda Research Labs) and 5  $\mu$ l of affinity-purified antibodies specific to HNF-3 $\alpha$  or HNF-3 $\beta$ , and then resolved by electrophoresis on 9% SDS-PAGE gels, followed by autoradiography. The immune precipitated phosphorylated HNF-3 protein was fractionated by SDS-PAGE and electrotransferred onto PVDF membrane and then excised after visualization by autoradiography. Phosphoamino acid analysis was performed as described (52).

#### *In vitro* kinase assays

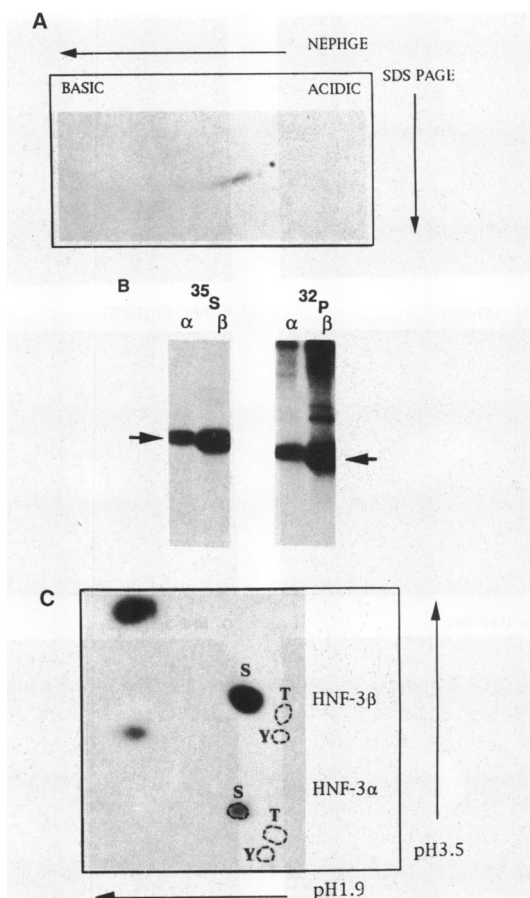
Glutathione-S-transferase (GST) fusion proteins containing either wild-type (7-86) or CKI mutant HNF-3 $\beta$  sequences (see above) were induced for synthesis in *E.coli* DH5 $\alpha$  cells with 1 mM

isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) for 3-5 h from the pGEX2T expression vector (Pharmacia). GST fusion proteins were affinity purified from extracts via glutathione agarose chromatography as described previously (24). Purified casein kinase I (CKI; gift from M. Tao and T. Wei) was used to phosphorylate 50 pmol of GST fusion proteins containing amino acids 7-86 of HNF-3 $\beta$  (wild-type or CKI mutants) as described (53).

## RESULTS

### Transcriptional activity of the HNF-3 $\beta$ N-terminus requires conserved sequences in regions IV and V

Although we had extensively analyzed the HNF-3 $\beta$  C-terminal transcriptional activation domain (46), the sequence requirement for activity of the HNF-3 $\beta$  N-terminal domain warranted further investigation. We prepared HNF-3 $\beta$  expression constructs for synthesis of polypeptides truncated at the N-terminus and examined their transcriptional activity using an HNF-3 dependent transactivation assay in HepG2 cells (46). Normalized mutant activation levels presented in Figure 1 were determined from the average of three separate experiments. Consistent with previous studies (46), deletion of the first 14 N-terminal amino acids had no effect on activity and elicited 15-fold activation of the HNF-3 reporter construct (Fig. 1). However, removal of an additional six N-terminus amino acids caused a 54% reduction in transactivation, and removal of all region IV sequences resulted in a further 10% decrease (Fig. 1, HNF-3 $\beta$  20-458, 54-458). No further decreases were observed with HNF-3 $\beta$  constructs containing more extensive N-terminal deletions that retained the winged helix DNA-binding domain (Fig. 1, HNF-3 $\beta$  153-458). In order to define C-terminal boundaries of this domain, we constructed two internal deletions based on the HNF-3 homology searches: the first HNF-3 $\beta$  mutant construct removed sequences 56-152 and the second HNF-3 $\beta$  mutant deleted residues 94-152. Surprisingly, addition of conserved region IV sequences (amino acids 1-55) to the HNF-3 $\beta$  153-458 deletion did not enhance its activity (Fig. 1, HNF-3 $\beta$   $\Delta$ 56/152). In contrast, replacement of sequences (amino acids 1-93) encompassing both region IV and a second conserved region V restored wild-type activity to the

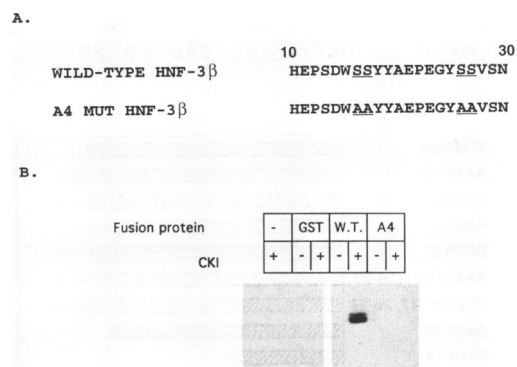


**Figure 2.** HNF-3 proteins are phosphorylated in hepatocytes and in HepG2 cells. (A) Rat liver nuclear extract was fractionated by 2D NEPHGE/SDS-PAGE in the directions indicated by arrows and visualized by anti-HNF-3 $\beta$  immunoblotting (see Materials and Methods). Note that the acidic HNF-3 $\beta$  isoforms migrate more slowly on the gel, which is suggestive of protein phosphorylation. (B) HNF-3 $\alpha$  and HNF-3 $\beta$  proteins are phosphorylated in HepG2 cells. Biosynthetic labelling of HepG2 cells was carried out with either [ $^{35}\text{S}$ ]methionine or [ $^{32}\text{P}$ ]orthophosphate for 4 h, followed by immunoprecipitation of the cell lysate and analysis by SDS-PAGE along with protein molecular weight size markers (see Materials and Methods). The location of the labelled HNF-3 bands are indicated by the arrow. (C) The HNF-3 $\alpha$  and HNF-3 $\beta$  proteins are phosphorylated on serine residues. Phosphoamino acid analysis of the cell lysate and analysis by SDS-PAGE along with protein molecular weight size markers (see Materials and Methods). The location of the labelled HNF-3 bands are indicated by the arrow. (C) The HNF-3 $\alpha$  and HNF-3 $\beta$  proteins are phosphorylated on serine residues. Phosphoamino acid analysis was performed on purified [ $^{32}\text{P}$ ]-labelled HNF-3 $\alpha$  and -3 $\beta$  protein. Dotted circles represent migration positions of standards: S, phosphoserine; T, phosphothreonine; Y, phosphotyrosine.

HNF-3 $\beta$  153–458 deletion mutant (Fig. 1, HNF-3 $\beta$   $\Delta$ 94/152). Taken together, these studies establish that the HNF-3 $\beta$  N-terminal activation domain extends from amino acid residues 14–93 and includes conserved region IV and V sequences.

#### Conserved casein kinase I phosphorylation sites in the HNF-3 $\beta$ N-terminal activation domain are dispensable for function

Because the removal of a potential casein kinase I (CKI; see below) site resulted in reduced activation (Fig. 1, HNF-3 $\beta$  20–458) and the fact that phosphorylation has been shown to regulate activity of transcription factors (54), we investigated whether the HNF-3 $\beta$  protein is phosphorylated. Liver nuclear extracts were separated by 2D NEPHGE/SDS-PAGE and HNF-3 $\beta$  protein was detected by



**Figure 3.** Site-directed mutations abolish CKI phosphorylation of the N-terminus of HNF-3 $\beta$ . (A) Shown are wild type and the alanine substituted, CKI mutant (A4) sequences of amino acids 10–30 of HNF-3 $\beta$  with two potential CKI sites underlined. (B) CKI phosphorylates wild-type but not mutant HNF-3 $\beta$  N-terminus sequences *in vitro*. Affinity purified GST fusion proteins containing CKI mutated (A4) or wild-type (W.T.) HNF-3 $\beta$  N-terminus sequences were used for *in vitro* kinase assays with purified CKI protein (see Materials and Methods).

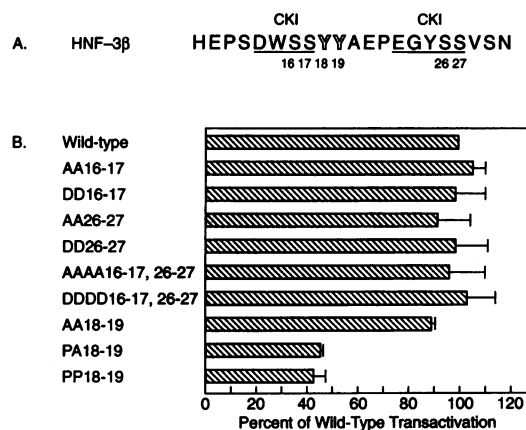
immunoblotting (51). Several slowly migrating acidic HNF-3 $\beta$  isoforms were apparent, thus exhibiting features consistent with protein phosphorylation (Fig. 2A). These differently migrating HNF-3 protein bands are not due to protein degradation as assessed by Western blots and gel shift assays using HNF3-specific antisera (34). To examine HNF-3 phosphorylation, HepG2 cells were metabolically labeled with either  $^{35}\text{S}$ -methionine or  $^{32}\text{P}$ -orthophosphate, and HNF-3 immune precipitates from labeled nuclear extracts were then analyzed by SDS-PAGE (34). Indeed, both HNF-3 $\alpha$  and HNF-3 $\beta$  proteins exhibited abundant  $^{32}\text{P}$ -labeling, confirming that they are phosphorylated in HepG2 cells (Fig. 2B). Phosphoamino acid analysis of the  $^{32}\text{P}$ -labeled HNF-3 proteins determined that phosphorylation was found only on serine residues (Fig. 2C).

Having established that the HNF-3 $\beta$  protein is phosphorylated in HepG2 cells, we sought to test the authenticity of the N-terminal CKI sites. *In vitro* kinase assays were carried out with affinity purified GST-HNF-3 $\beta$  fusion protein (amino acids 14–86) and purified CKI enzyme [Fig. 3A; DWSS and EGYSS, (55)]. Analysis of the kinased wild-type fusion protein (W.T.) by SDS-PAGE revealed that GST-HNF-3 $\beta$  fusion protein was efficiently phosphorylated by CKI protein while no visible phosphorylation of the GST protein was apparent (Fig. 3B). Furthermore, disruption of the CKI sites by replacing critical serine residues with alanine (positions 16, 17, 26 and 27) abolished CKI phosphorylation (Fig. 3B, A4).

To examine the role of the N-terminal CKI sites in HNF-3 $\beta$  transactivation, we created pair-wise substitutions of the serine residues with alanine to disrupt the CKI phosphorylation sites (Fig. 4A). We also replaced the CKI sites with aspartate (D) residues to determine the effect of increasing the number of acidic residues with the HNF-3 $\beta$  N-terminal domain (Fig. 4B). Surprisingly, N-terminal mutations which eliminated CKI sites or increased the number of acidic amino acids did not influence HNF-3 $\beta$  transcriptional activity (Fig. 4B).

#### Proline substitutions in the HNF-3 $\beta$ N-terminus result in decreased transcriptional activation

We next prepared site directed mutations which replaced two tyrosine (Y) residues which are conserved in several HNF-3

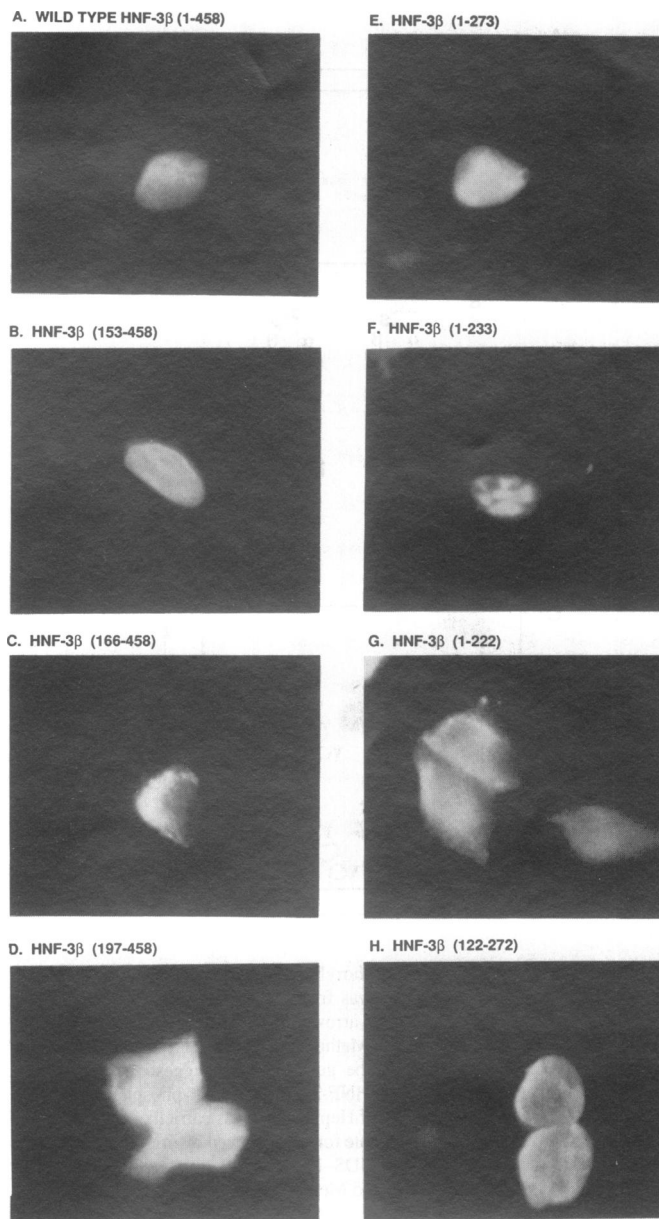


**Figure 4.** Transcriptional activity of site-directed mutants in the HNF-3 $\beta$  N-terminal activation domain. (A) Amino acid sequences of the N-terminal region targeted by site-directed mutagenesis (amino acids 10–30 of HNF-3 $\beta$ ). Two casein kinase I phosphorylation sites are underlined and two conserved tyrosine residues are in outline letters. Numbers indicate the positions of amino acid residues. (B) Activity of the HNF-3 $\beta$  N-terminal domain does not require phosphorylation or acidic residues but relies on the maintenance of  $\alpha$ -helical structure. Transcriptional activities of HNF-3 $\beta$  N-terminal site-directed mutants averaged from at least three independent cotransfection experiments are summarized in bar-graph as a percentage of wild-type HNF-3 $\beta$ . The substitutions are indicated by the amino acid letters (A for alanine, D for aspartate and P for proline) and the numbers correspond to the amino acid position as indicated in panel A.

family members (positions 18 and 19) with either alanine or proline residues (Fig. 4). Proline substitutions are predicted to disrupt the formation of any  $\alpha$ -helical protein structure, while alanine substitutions will enable us to examine whether the tyrosine residue is required for transcriptional activation. Interestingly, alanine substitutions did not reduce HNF-3 $\beta$  transactivation, suggesting that tyrosine residues in positions 18 and 19 are not critical for N-terminal activity (Fig. 4, AA18 and 19). In contrast, replacement of tyrosine residues with proline residues resulted in 60% reduction in HNF-3 $\beta$  transcriptional activity, which is similar to decreases observed when the N-terminal domain is removed (Fig. 4B, PA18-19 and PP18-19). Because comparable nuclear expression levels were exhibited by mutant and wild-type HNF-3 $\beta$  proteins (data not shown), the differences in transactivation are the result of the N-terminal proline substitutions. These results suggest that transcriptional activity of the HNF-3 $\beta$  N-terminal domain requires maintenance of a putative  $\alpha$ -helical structure (see Discussion section).

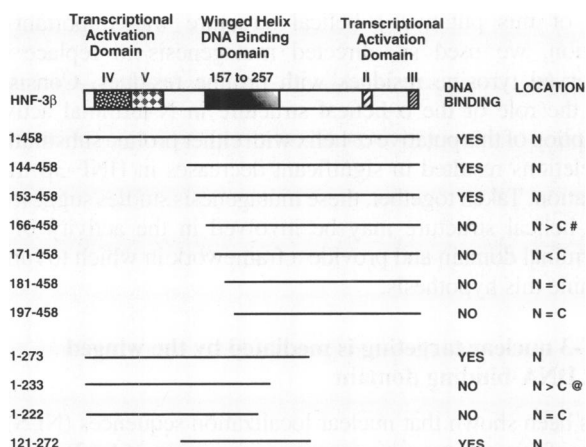
#### The nuclear localization signal overlaps with the HNF-3 $\beta$ winged helix DNA-binding domain

During functional analysis of HNF-3 $\beta$  we observed that deletion mutants which retained the winged helix DNA-binding domain were localized to the nucleus. To further define the nuclear localization signals, we examined the subcellular localization of various HNF-3 $\beta$  deletion mutants via indirect immunofluorescence microscopy (Fig. 5). As expected, deletion of the amino acid sequences outside of the DNA-binding domain had little effect on the nuclear localization of the HNF-3 $\beta$  protein (Fig. 5A, B, E and H). However, removal of 12 N-terminal residues from the HNF-3 $\beta$  DNA-binding domain (153–166) partially impaired nuclear targeting (compare Fig. 5B and C) and further deletions



**Figure 5.** The HNF-3 $\beta$  nuclear localization signal requires sequences on both ends of the winged helix DNA-binding motif. HNF-3 specific antibodies were used for indirect immunofluorescence assay to detect cellular localization of wild-type (A) or N- or C-terminal deleted HNF-3 $\beta$  proteins expressed in HepG2 cells (B–H, numbers represent amino acid residues encoded by the expression constructs). Similar results were also observed in HeLa cells (data not shown). Affinity-purified HNF-3 $\beta$  antibodies specific to the N-terminus (1:100 diluted) were used to detect full-length and C-terminus deleted HNF-3 $\beta$  protein by indirect immunofluorescence (see Materials and Methods). Antibodies made against the DNA-binding domain of HNF-3 $\alpha$  and HNF-3 $\beta$  were used to detect N-terminal deleted HNF-3 $\beta$  proteins. Note that the expression of the HNF-3 mutants in transfected HepG2 cells overwhelms staining obtained from endogenous protein.

completely eliminated nuclear localization (Fig. 5D, HNF-3 $\beta$  197–458). Disruption of the C-terminus of the HNF-3 $\beta$  DNA-binding domain also altered the nuclear staining pattern (Fig. 5F, HNF-3 $\beta$  1–233) and further deletions completely eliminated nuclear targeting (Fig. 5G, HNF-3 $\beta$  1–222). These data provide compelling evidence that the HNF-3 $\beta$  nuclear localization signal



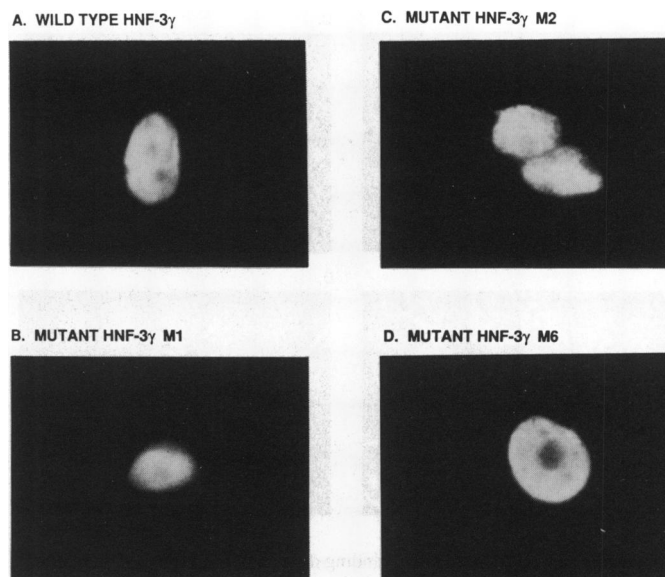
**Figure 6.** Summary of studies defining the nuclear localization signal (NLS) of HNF-3 $\beta$ . Schematically shown is HNF-3 protein as a rectangular box from the N- to C-terminus with shaded areas representing functional domains of HNF-3 (see Fig. 4). HNF-3 $\beta$  deletions used in the NLS study are represented by lines spanning amino acid sequences contained in each mutant. The HNF-3 $\beta$  mutant DNA-binding activity and their localization in the cells are also shown. The symbols indicated the following: N, nuclear localized; C, cytoplasmic localized; N > C, greater staining in the nucleus; N < C, greater staining in the cytoplasm; #, perinuclear staining pattern; @, punctate staining pattern. The HNF-3 $\beta$  144-458, 171-458 and 181-458 N-terminal mutants were analyzed but not shown in Figure 5.

(NLS) requires sequences from both ends of the winged helix DNA-binding domain (see summary Fig. 6).

Because removal of sequences from the HNF-3 $\beta$  winged helix motif also inhibited DNA recognition, we sought to determine whether nuclear localization required DNA-binding activity. We therefore examined nuclear targeting of site-directed mutations within the HNF-3 $\gamma$  winged helix motif (95% amino acid homology) which were impaired in DNA-binding activity (37). In support of the fact that DNA binding activity was not a requisite for nuclear localization, all of the HNF-3 $\gamma$  mutants were efficiently targeted to the nucleus and exhibited properties similar to the wild-type protein (Fig. 7).

#### The HNF-3 $\beta$ winged helix DNA-binding motif is sufficient for nuclear targeting of the $\beta$ -galactosidase enzyme

In order to determine whether the HNF-3 $\beta$  DNA-binding domain is sufficient for nuclear targeting, we made expression constructs which synthesized different  $\beta$ -galactosidase-HNF-3 $\beta$  fusion proteins. These expression constructs were transfected into HepG2 cells, and the cellular distribution of the fusion proteins was examined by histochemical staining of  $\beta$ -galactosidase enzymatic activity. Consistent with the analysis above, the 153-169 N-terminal sequences of the HNF-3 $\beta$  DNA-binding domain did not elicit nuclear targeting of the cytoplasmic  $\beta$ -galactosidase protein (Fig. 8A and B). However, the entire DNA-binding domain of HNF-3 $\beta$  (153-257) was sufficient to direct the  $\beta$ -galactosidase protein to the nucleus (Fig. 8C) and elicited nuclear staining comparable to that of the SV40 T antigen NLS (Fig. 8D). Taken together, these data establish that the HNF-3 NLS resides within the winged helix DNA-binding domain.

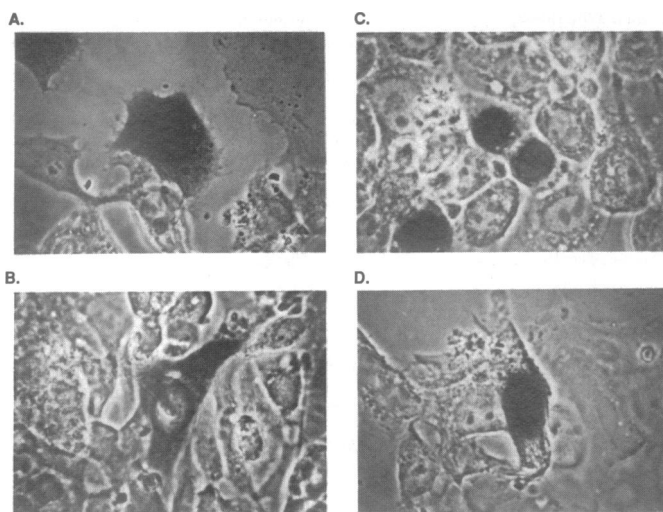


**Figure 7.** Nuclear localization does not require DNA binding activity. Mutated HNF-3 $\gamma$  winged helix motif proteins which are impaired for DNA binding activity (37) were expressed in HepG2 and detected by immunofluorescence microscopy using HNF-3 DNA-binding domain antisera (see Fig. 8 legend). All mutants (B, C and D) exhibited nuclear staining patterns identical to wild-type protein (A).

## DISCUSSION

### Analysis of the conserved HNF-3 $\beta$ N-terminal transcriptional activation domain

Because of the importance of HNF-3 $\beta$  in hepatocyte-specific gene expression, it is critical to decipher mechanisms of HNF-3 mediated transcriptional activation. Previous functional studies of the HNF-3 $\beta$  protein identified a 100 amino acid C-terminal domain which required conserved region II and III sequences for transcriptional activation (46). In the current study we continued this analysis by precisely defining N-terminal residues required for HNF-3 $\beta$  transcriptional activation. We found that HNF-3 $\beta$  activation required sequences between amino acid residues 14 and 93 and comprised of region IV and V sequences which displayed homology with closely related HNF-3/*fkh* family members (Fig. 9). This included mammalian HNF-3 $\alpha$  and HNF-3 $\beta$  (3), *Xenopus* HNF-3 $\beta$ , *pintallavis* (27,56), *XFKH1* (57) and Zebrafish Axial (58) proteins. Furthermore, the N-terminal sequence contains an abundance of hydrophobic amino acids, a feature found in the transcriptional activation domain of the *Drosophila* NTF-1 protein (59). HNF-3/*fkh* sequence comparisons reveal that both methionine and hydrophobic residues exhibit precise alignment throughout this N-terminal domain (Fig. 9). The conservation of these residues may provide a sequence motif mediating association with components of the basal transcriptional machinery. Previous studies demonstrated that activity of the N-terminal domain required the presence of the C-terminal region II and III motif, thus suggesting that these HNF-3 domains interact and cooperate to activate transcription (46). It is interesting to note that the HNF-3/*fkh* proteins aligned in Figure 9 also possess strong homology within the region II and



**Figure 8.** The winged helix DNA-binding domain of HNF-3β is sufficient for nuclear targeting of the β-galactosidase enzyme. CMV expression plasmids synthesizing HNF-3β fusion proteins linked to the N-terminus of β-galactosidase gene were transfected into HepG2 cells and localized by histochemical staining (see Materials and Methods). Shown are the cellular localization patterns of (A) β-galactosidase, (B) β-galactosidase fusion protein containing amino acid residues 153–169 of HNF-3β, (C) β-galactosidase fusion protein containing the DNA-binding domain (amino acid residues 153–257) of HNF-3β and (D) The nuclear localization signal of SV40 T antigen linked to the N-terminus of β-galactosidase protein.

III motif and conservation of sequences in both domains may be required to maintain protein association.

Protein phosphorylation constitutes one mechanism for modulation of transcriptional activation via creation of a region of acidic residues (54). The HNF-3 N-terminal domain contains several conserved serine residues that are potential CKI protein phosphorylation sites, which can be used to regulate its activity. In order to test this hypothesis, we demonstrated that retention of the N-terminal CKI phosphorylation sites was not required for HNF-3β transactivation (Figs 2–4). However, these CKI sites may play a role in other biological responses not yet examined. Furthermore, mutations which increased the number of acidic residues within the N-terminal activation domain did not influence transactivation by HNF-3β protein.

Use of a secondary structure computer program (60) enabled us to predict the formation of an α-helix within N-terminal residues 14–19 (DWSSYY), which are conserved in several HNF-3 family members (Fig. 9). To examine whether mainten-

ance of this putative α-helical structure was important for function, we used site-directed mutagenesis to replace two conserved tyrosine residues with proline residues. Consistent with the role of the α-helical structure in N-terminal activity, disruption of this putative α-helix with either proline substitutions or deletions resulted in significant decreases in HNF-3β trans-activation. Taken together, these mutagenesis studies suggest that an α-helical structure may be involved in the activity of the N-terminal domain and provide a framework in which to further examine this hypothesis.

### HNF-3 nuclear targeting is mediated by the winged helix DNA-binding domain

It has been shown that nuclear localization sequences (NLS) do not conform to any one consensus sequence (61,62), but are typically short, usually rich in lysine and arginine residues, and often contain proline. Our data clearly show that nuclear targeting of the HNF-3β protein requires sequences at both ends of the winged helix domain, yet nuclear localization does not need to retain DNA binding activity (Figs 5–8). Examples of complex nuclear localization signals have been observed in a number of nuclear proteins as well (63). Furthermore, we show that the HNF-3β winged helix DNA-binding domain is sufficient for nuclear targeting of cytoplasmic β-galactosidase protein (Fig. 8). This is consistent with other examples of NLSs which colocalize with the DNA-binding domains of bZip proteins (64) and the homeodomain of MATα2 (65). In the case of GATA-3 transcription factor, the NLS resides in an N-terminal zinc finger motif which is situated in close proximity to a second zinc finger motif that mediates specific DNA recognition (66).

The existence of the NLS in the winged helix motif suggests that other members of the HNF-3/*fkh* family may conserve certain residues in their DNA-binding domains for nuclear translocation. The HNF-3β C-terminal region of the winged helix motif consists of basic amino acid residues (LRRQKRFKC), a feature in common with other NLS (61,64). Although the LRR residues are involved in a base-specific contact in the minor groove (4), the other basic amino acid residues are also conserved throughout the HNF-3/*fkh* family. Furthermore, the NLS found at the N-terminus of the winged helix motif (KPPYSYISLITMAIQQ) comprises one of the most conserved regions of this motif and constitutes an α-helix that precedes the DNA recognition helix–turn–helix portion of this domain. Conservation of sequences at both ends of the DNA-binding domain may function to provide specific interaction with proteins involved in nuclear translocation (67). These studies therefore suggest that evolution of the winged helix

	Region IV	Region V	
	14	52	93
<b>HNF-3Beta</b>	DWSSYYAEP-EGYSS--VSNMNASLG-MNQMNTYMSMSAAAMGSGSGNMSAGS-MNMSYVVGAGMSPSLAGMSPGAGAMAGMSG		100%
<b>Axial</b>	DWSTTYGEP-ECYTS--VSNMNTGLG-MNSMNTYMTMSG---MSSTANMTAANTMMS-YVNTGMSPSMTGMSPGTGAMAGMGA		79%
<b>XHNF-3Beta</b>	DWSSYYGEA-EAYSS--VGNMIDAGLS-MNPMNTYMSMSA---MSTSAWMTASS-MNMS-YVNTGMSPSLTGMSPGTGAMTGMGT		75%
<b>HNF-3Alpha</b>	DWNSYYADTQEAYSVPVSNMNSGLGSMNSMNTYMTMT---MTTSGNMTPAS-FNMS-YANPGLGGRLSP-----GAVAGIPG		60%
<b>Pintallavis</b>	DWNTMYQEN-EMYSG--IHNMTNVLPS----NSFLPND---VST---VTT---SMP-YMSNGLPGPVVTSIQGNIGSLGSMPTQ		41%
<b>XFKH1</b>	DWNTMYQEN-EIYSG--IHNMTNGLPS----NSFLPTD---VPT---VTS---SMT-YMSNGLPGPVASIQGNLGLSLSMTQ		41%
<b>CONSENSUS</b>	DWSSYYAE--E-YSS--VSNMN-GLG-MN-MNTYMSMS----MSTSGNMTA-S-MNMS-YVNTGMSPSLAGMSPG-GAMAGM--		
	NTM GD TG IH T P SFLT N VTSTA VS N FS T MSN LPPVSSIQGN SLGS		
			T

**Figure 9.** Sequence alignment of HNF-3/*fkh* N-terminal domains. Comparison of the N-terminal amino acid sequence from mammalian HNF-3β and HNF-3α (2,3), *Xenopus* HNF-3β (27), *Pintallavis* (56) and *XFKH1* (57) and Zebrafish Axial (58) proteins. Shown below is an N-terminal consensus sequence in which the alignment of conserved residues is highlighted in bold letters.

motif may involve sequence constraints required to maintain both nuclear targeting and DNA binding function.

## ACKNOWLEDGEMENTS

We thank Pradip Raychaudhuri, and members of our laboratory Uzma Samadani, Derek Clevidence, Richard Peterson and David Overdier for critically reviewing the manuscript and for helpful discussions. We are grateful to K. Colley for her assistance during the nuclear targeting studies. We also would like to thank M. Tao and T. Wei for their gracious gift of CKI enzyme and for their advice on the phosphorylation studies. Secondary structure prediction of the N-terminal domain was performed using the Predict Protein program (60) which is available through e-mail predictprotein@embl-heidelberg.de. This work was supported by Public Health Service grant R01 GM43241 from the National Institute of General Medical Sciences. RHC is an Established Investigator of the American Heart Association/Bristol-Myers Squibb.

## REFERENCES

- Zaret, K. S. (1993) In Tavoloni, N., and Berk, P. D. (eds), *Hepatic Transport and Bile Secretion: Physiology and Pathophysiology*. Raven Press, Ltd, New York, pp. 135–143.
- Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H. and Darnell, J. E., Jr. (1990) *Genes Dev.*, **4**, 1427–1436.
- Lai, E., Prezioso, V. R., Tao, W. F., Chen, W. S. and Darnell, J. E., Jr. (1991) *Genes Dev.*, **5**, 416–427.
- Clark, K. L., Halay, E. D., Lai, E. and Burley, S. K. (1993) *Nature*, **364**, 412–420.
- Sladek, F. M., Zhong, W. M., Lai, E. and Darnell, J. E., Jr. (1990) *Genes Dev.*, **4**, 2353–2365.
- Ladiaz, J. A. and Karathanasis, S. K. (1991) *Science*, **251**, 561–565.
- Baumhueter, S., Mendel, D. B., Conley, P. B., Kuo, C. J., Turk, C., Graves, M. K., Edwards, C. A., Courtois, G. and Crabtree, G. R. (1990) *Genes Dev.*, **4**, 372–379.
- Frain, M., Swart, G., Monaci, P., Nicosia, A., Stampfli, S., Frank, R. and Cortese, R. (1989) *Cell*, **59**, 145–157.
- Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1989) *Science*, **243**, 1681–1688.
- Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T. and Kishimoto, T. (1990) *EMBO J.*, **9**, 1897–1906.
- Cao, Z., Umek, R. M. and McKnight, S. L. (1991) *Genes Dev.*, **5**, 1538–1552.
- Descombes, P., Chojkier, M., Lichtsteiner, S., Falvey, E. and Schibler, U. (1990) *Genes Dev.*, **4**, 1541–1551.
- Kinoshita, S., Akira, S. and Kishimoto, T. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1473–1476.
- Poli, V., Mancini, F. P. and Cortese, R. (1990) *Cell*, **63**, 643–653.
- Williams, S. C., Cantwell, C. A. and Johnson, P. F. (1991) *Genes Dev.*, **5**, 1553–1567.
- Mueller, C. R., Maire, P. and Schibler, U. (1990) *Cell*, **61**, 279–291.
- Drolet, D. W., Scully, K. M., Simmons, D. M., Wegner, M., Chu, K. T., Swanson, L. W. and Rosenfeld, M. G. (1991) *Genes Dev.*, **5**, 1739–1753.
- Iyer, S. V., Davis, D. L., Seal, S. N. and Burch, J. B. E. (1991) *Mol. Cell Biol.*, **11**, 4863–4875.
- Hunger, S. P., Ohyashiki, K., Toyama, K. and Cleary, M. L. (1992) *Genes Dev.*, **6**, 1608–1620.
- Inaba, T., Roberts, W. M., Shapiro, L. H., Jolly, K. W., Raimondi, S. C., Smith, S. D. and Look, A. T. (1992) *Science*, **257**, 531–534.
- Costa, R. H., Grayson, D. R. and Darnell, J., Jr. (1989) *Mol. Cell Biol.*, **9**, 1415–1425.
- Costa, R. H. and Grayson, D. R. (1991) *Nucleic Acids Res.*, **19**, 4139–4145.
- Costa, R. H. (1994) In Tronche, F., and Yaniv, M., (eds.), *Liver Gene Transcription*. R.G. Landes Co., Austin, Texas, pp. 183–205.
- Overdier, D. G., Porcella, A. and Costa, R. H. (1994) *Mol. Cell Biol.*, **14**, 2755–2766.
- Ang, S. L., Wierda, A., Wong, D., Stevens, K. A., Cascio, S., Rossant, J. and Zaret, K. S. (1993) *Development*, **119**, 1301–1315.
- Monaghan, A. P., Kaestner, K. H., Grau, E. and Schutz, G. (1993) *Development*, **119**, 567–578.
- Ruiz i Altaba, A., Prezioso, V. R., Darnell, J. E. and Jessell, T. M. (1993) *Mech. Dev.*, **44**, 91–108.
- Sasaki, H. and Hogan, B. L. (1993) *Development*, **118**, 47–59.
- Ang, S. L. and Rossant, J. (1994) *Cell*, **78**, 561–574.
- Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M. and Darnell, J., Jr. (1994) *Cell*, **78**, 575–588.
- Kuo, C. J., Conley, P. B., Chen, L., Sladek, F. M., Darnell, J., Jr and Crabtree, G. R. (1992) *Nature*, **355**, 457–461.
- DiPersio, C. M., Jackson, D. A. and Zaret, K. S. (1991) *Mol. Cell Biol.*, **11**, 4405–4414.
- Liu, J. K., DiPersio, C. M. and Zaret, K. S. (1991) *Mol. Cell Biol.*, **11**, 773–784.
- Jacob, A., Budhiraja, S., Qian, X., Clevidence, D., Costa, R. H. and Reichel, R. R. (1994) *Nucleic Acids Res.*, **22**, 2126–2133.
- Weigel, D. and Jackle, H. (1990) *Cell*, **63**, 455–456.
- Bassel-Duby, R., Hernandez, M. D., Yand, Q., Rochelle, J. M., Seldin, M. F. and Williams, R. S. (1994) *Mol. Cell Biol.*, **14**, 4596–4605.
- Clevidence, D. E., Overdier, D. G., Tao, W., Qian, X., Pani, L., Lai, E. and Costa, R. H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3948–3952.
- Clevidence, D. E., Overdier, D. G., Peterson, R. S., Porcella, A., Ye, H., Paulson, E. K. and Costa, R. H. (1994) *Dev. Biol.*, **166**, 195–209.
- Hromas, R., Moore, J., Johnston, T., Socha, C. and Klemsz, M. (1993) *Blood*, **81**, 2854–2859.
- Kaestner, K. H., Lee, K. H., Schlondorff, J., Hiemisch, H., Monaghan, A. P. and Schutz, G. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7628–7631.
- Li, C., Lai, C., Sigman, D. S. and Gaynor, R. B. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7739–7743.
- Li, C., Lusa, A. J., Sparkes, R., Tran, S. and Gaynor, R. (1992) *Genomics*, **13**, 658–664.
- Li, C. and Tucker, P. W. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11583–11587.
- Pierrou, S., Hellqvist, M., Samuelsson, L., Enerback, S. and Carlsson, P. (1994) *EMBO J.*, **13**, 5002–5012.
- Tao, W. and Lai, E. (1992) *Neuron*, **8**, 957–966.
- Pani, L., Overdier, D. G., Porcella, A., Qian, X., Lai, E. and Costa, R. H. (1992) *Mol. Cell Biol.*, **12**, 3723–3732.
- Knowles, B., Howe, C. C. and Arden, D. P. (1980) *Science*, **209**, 497–499.
- Ho, A. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R. (1989) *Gene*, **77**, 51–59.
- Colley, K. J., Lee, E. U. and Paulson, J. C. (1992) *J. Biol. Chem.*, **267**, 7784–7793.
- Sanes, J. R., Rubenstein, J. L. and Nicolas, J. F. (1986) *EMBO J.*, **5**, 3133–3142.
- O'Farrell, P. Z., Goodman, H. M. and O'Farrell, P. H. (1977) *Cell*, **12**, 1133–1141.
- Boyle, W. J., van der Geer, P. and Hunter, T. (1991) In Hunter, T., and Sefton, B. M. (eds), *Methods in Enzymology*. Academic Press, New York, Vol. 201, pp. 110–149.
- Wei, T. Q. and Tao, M. (1991) *FEBS Lett.*, **292**, 141–144.
- Hunter, T. and Karin, M. (1992) *Cell*, **70**, 375–387.
- Kennelly, P. J. and Krebs, E. G. (1991) *J. Biol. Chem.*, **266**, 15555–15558.
- Ruiz i Altaba, A. and Jessell, T. M. (1992) *Development*, **116**, 81–93.
- Dirksen, M. L. and Jamrich, M. (1992) *Genes Dev.*, **6**, 599–608.
- Strahle, U., Blader, P., Henrique, D. and Ingham, P. W. (1993) *Genes Dev.*, **7**, 1436–1446.
- Attardi, L. D. and Tjian, R. (1993) *Genes Dev.*, **7**, 1341–1353.
- Rost, B. and Sander, C. (1993) *J. Mol. Biol.*, **232**, 584–599.
- Garcia-Bustos, J., Heitman, J. and Hall, M. N. (1991) *Biochim. Biophys. Acta*, **1071**, 83–101.
- Wagner, P., Kunz, J., Koller, A. and Hall, M. N. (1990) *FEBS Lett.*, **275**, 1–5.
- Richardson, W. D., Roberts, B. L. and Smith, A. E. (1986) *Cell*, **44**, 77–85.
- Dingwall, C. and Laskey, R. A. (1991) *Trends Biochem. Sci.*, **16**, 478–481.
- Hall, M. N., Craik, C. and Hiraoka, Y. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6954–6958.
- Yang, Z., Gu, L., Romeo, P. H., Bories, D., Motohashi, H., Yamamoto, M. and Engel, J. D. (1994) *Mol Cell Biol.*, **14**, 2201–2212.
- Nigg, E. A., Baeuerle, P. A. and Luhrmann, R. (1991) *Cell*, **66**, 15–22.