ABSTRACT

WAN, CHENG. Monogenic Diabetes and Regulation of the Transcription Factor, HNF1\(\alpha\). (Under the direction of Robert B. Rose.)

The purpose of this research is to explore the mechanism of MODY3 disease and role of HNF1\(\alpha\) in MODY3. MODY3, a monogenic diabetes, is caused by mutations in the gene encoding the transcription factor hepatocyte nuclear factor-1alpha (HNF1\(\alpha\)) which plays an important role on glucose homeostasis in pancreatic islets. MODY3 is inherited in an autosomal dominant pattern since mutations in only one allele of the HNF1\(\alpha\) gene impair pancreatic beta-cell function. However, the molecular mechanism for the dominant phenotype of MODY3 remains unclear. By transfection assays, we characterize a MODY3 mutation outside of the dimerization domain that does not heterodimerize with wild type HNF1\(\alpha\) but causes a dominant negative effect by competing for DNA binding. We propose that heterodimerization between mutant and wild type HNF1\(\alpha\) may not occur due to stabilization of HNF1\(\alpha\) dimers by the coactivator DCoH (Dimerization Cofactor of HNF1\(\alpha\)).

This study has implications for how MODY3 mutations may affect HNF1\(\alpha\) activity in patients. On the second part of our studies, we solved the three dimensional protein structure of Erh (Enhancer of Rudimentary Homolog), which is a ubiquitously expressed transcriptional co-regulator and is highly conserved among eukaryotes, from humans to plants to protozoa. Erh was demonstrated to be interaction partner of DCoH and repress HNF1\(\alpha\) activity through binding the coactivator DCoH. However, in our studies, no detectable interaction between DCoH and Erh was found, suggesting that another protein is probably involved to bridge DCoH and Erh. Identification interaction partners of DCoH and
characterization of the binding details will significantly improve our understanding of the role of DCoH on the transcriptional regulation of HNF1α
Monogenic Diabetes and Regulation of the Transcription Factor, HNF1α

by
Cheng Wan

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Biochemistry

Raleigh, North Carolina

2008

APPROVED BY:

_________________________            _________________________
William L. Miller                                  Robert C. Smart

__________________________          _________________________
Carla Mattos                                          Robert B. Rose
Chair of Advisory Committee
DEDICATION

I dedicate this thesis to my wife, Min Wu, and my new baby, Ethan Z. Wan
BIOGRAPHY

Cheng Wan, raised in China, Handan, graduated from Dalian Medical University in 1997. He got his master degree (Biochemistry) from Jilin University in 2000. He married Min in August, 2001. Then, he entered NC State University to study his Ph.D degree (Biochemistry).
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................................ vi
LIST OF FIGURES ........................................................................................................................ vii
CHAPTER 1: GENERAL INTRODUCTION ......................................................................................1
  1.1 INTRODUCTION TO DIABETES ......................................................................................2
  1.2 MATURITY-ONSET DIABETES OF THE YOUNG (MODY) ......................................................2
  1.3 HNF1α ..............................................................................................................................4
  1.4 MODY3 ..................................................................................................................................6
  1.5 HNF1α TARGET GENES ........................................................................................................8
  1.6 HNF TRANSCRIPTIONAL NETWORK AND TISSUE-SPECIFIC GENE EXPRESSION .........................................................................................................................10
  1.7 COOPERATIVE ACTIVATION ON TARGET GENES BY HNF TRANSCRIPTIONAL FACTORS AND OTHER REGULATORS ........................................................................................13
  1.8 POSSIBLE MECHANISMS IN THE DEVELOPMENT OF MODY3 ...........................................15
REFERENCES ..................................................................................................................................31
CHAPTER 2: THE MECHANISM OF DOMINANT NEGATIVE INHIBITION OF WILD TYPE HNF1α BY MODY3 MUTANTS ..................................................................................................................43
  2.1 ABSTRACT ..........................................................................................................................44
  2.2 INTRODUCTION ..................................................................................................................45
  2.3 MATERIALS AND METHODS ..............................................................................................48
    2.3.1 Plasmid construction and site-directed mutagenesis .........................................................48
    2.3.2 Cell culture and transfection ..........................................................................................49
    2.3.3 Preparation of nuclear extraction .....................................................................................50
    2.3.4 Western blot analysis .....................................................................................................50
  2.4 RESULTS ................................................................................................................................51
    2.4.1 Endogenous HNF1α activity detected in β-TC3 cells .......................................................51
    2.4.2 Transfected full length HNF1α has much higher activity in β-TC3 cells, compared to the endogenous HNF1α ........................................................................................................53
    2.4.3 S142F mutation in DNA binding domain disrupts transactivation activity of HNF1α ..........................................................................................................................53
    2.4.4 Transfected HNF1α (1-280) inhibits endogenous HNF1α activity .......................................54
    2.4.5 HNF1α (1-280)_S142F doesn’t inhibit endogenous HNF1α activity ...................................56
    2.4.6 Investigation the dominant negative mechanism by using ROP chimeras .......................56
    2.4.7 Transfected ROP-HNF1α (32-628) fusion has high activity ...............................................59
    2.4.8 Transfected ROP-HNF1α (32-280), not ROP-HNF1α (32-280)_S142F, inhibits endogenous HNF1α activity .................................................................................................59
  2.5 DISCUSSION .........................................................................................................................60
REFERENCES ..................................................................................................................................82
CHAPTER 3: STRUCTURE OF THE CONSERVED TRANSCRIPTIONAL REPRESSOR ERH (ENHANCER OF RUDIMENTARY HOMOLOG) .................................................................................................88
  3.1 ABSTRACT ..........................................................................................................................89
  3.2 INTRODUCTION ..................................................................................................................90
  3.2.1 Plasmid construction and site-directed mutagenesis ..........................................................90
  3.2.2 Cell culture and transfection ..........................................................................................90
  3.2.3 Preparation of nuclear extraction .....................................................................................91
  3.2.4 Western blot analysis .....................................................................................................91
LIST OF TABLES

Table 1.1  Six subtype MODYs ..........................................................................................20
Table 1.2  Biological function of MODY3 mutations ..........................................................21
Table 2.1  Reporter controls illustrating the processing of data from transfection assay and BCA assay .................................................................65
Table 3.1  Data collection, phasing and refinement statistics .............................................106
LIST OF FIGURES

Figure 1.1    Scheme of the hepatocyte nuclear factor 1α (HNF1α) transcription factor ...22
Figure 1.2    Three dimensional structure of dimerization domain of HNF1α and DCoH ...23
Figure 1.3    The stable interface between specific POU domain and homeodomain ..........25
Figure 1.4    Summary of mutations in the HNF-1α gene ....................................................27
Figure 1.5    Proposed model for a hierarchical transcriptional network in pancreatic islets .................................................................................................................28
Figure 1.6    The two promoters P1 (HNF4α1) and P2 (HNF4α7) of HNF4α are differentially utilized in hepatocytes and pancreatic islets .....................29
Figure 2.1    The schematic representation of MODY3 mutants used for transfection assays ..................................................................................................................67
Figure 2.2    Activity of endogenous HNF1α..............................................................................68
Figure 2.3    Activation of the pGL3-RA reporter by transfected HNF1α...............................69
Figure 2.4    A mutation in the DNA binding domain abolishes HNF1α activity.................70
Figure 2.5    HNF1α (1-280) interferes with endogenous HNF1α activity............................72
Figure 2.6    A mutation in the DNA binding domain of HNF1α (1-280) interferes with its dominant negative activity.................................................................73
Figure 2.7    The schematic representation of ROP chimera mutants used for transfection assays ..................................................................................................................75
Figure 2.8    Structural comparison between ROP and the dimerization domain of HNF1α76
Figure 2.9    The Rop-HNF1α fusion protein is functionally active ........................................77
Figure 2.10   Rop-HNF1α is less active than wild type HNF1α..............................................78
Figure 2.11   A mutation in the DNA binding domain of Rop-HNF1α (32-280) interferes with its dominant negative activity .......................................................79
Figure 2.12   DCoH restricts the formation of HNF1α heterodimers ....................................80
CHAPTER 1

GENERAL INTRODUCTION

BY

CHENG WAN AND ROBERT B. ROSE
1.1 INTRODUCTION TO DIABETES

Diabetes mellitus is a metabolic disorder characterized by hyperglycemia resulting from abnormal insulin secretion, or insulin resistance, or both. According to the World Health Organization, in 2006, there were 171 million diabetes patients in world and this number will be doubled in year 2030. In 2005 there are 20.8 million people in the United State who have diabetes and about 41 million people who would be diagnosed with pre-diabetes.

Basically, there are two main types of diabetes. Type 1 diabetes is juvenile onset diabetes, and has been called insulin dependent diabetes mellitus (IDDM). In this type of diabetes, the patient's own immune system attacks the pancreas, leading to absolute insulin deficiency so that treatment with insulin is necessary. Type II diabetes is adult onset diabetes, and has been called non-insulin dependent diabetes mellitus (NIDDM). Type II diabetes is mainly characterized by insulin resistance which occurs when body cells don’t respond to high level of insulin, therefore excess glucose builds up in the blood. While the exact mechanisms of insulin resistance remains unclear, the possible causes may be attributed to diminished or ineffective insulin receptors on target cells, problems in the translocation of glucose in response to insulin by transport proteins, and mitochondrial dysfunction. Most people worldwide or American who are diagnosed with diabetes have type II diabetes.

1.2 MATURITY-ONSET DIABETES OF THE YOUNG (MODY)

Maturity-onset diabetes of the young (MODY) is a hereditary monogenic form of
diabetes responsible for 2-5% of type II diabetes (Fajans et al. 2001). So far, six subtypes of MODY have been identified by genetic studies, linked to heterozygous mutations in six genes (Table 3.1). These genes encode the Hepatocyte Nuclear Factor 4α (HNF4α/MODY1), enzyme glucokinase (GCK/MODY2), Hepatocyte Nuclear Factor 1α (HNF1α/MODY3), Pancreas Duodenum Homeobox-1 (PDX1/MODY4), Hepatocyte Nuclear Factor 1β (HNF1β/MODY5), and NeuroD1/Beta2 (MODY6). All subtypes of MODY are characterized by non-ketotic hyperglycemia, which usually develops during childhood, adolescence or young adulthood, and is associated with primary insulin secretion defects (Velho and Froguel 1998 and Froguel et al. 1999). Except for glucokinase which senses the circulating concentrations of glucose, all other MODY genes encode cell specific transcription factors. These transcription factors play crucial roles in maintaining normal pancreatic β-cell function.

The severity of the different subtypes of MODY varies. Mutations in genes encoding transcription factors exhibit more severe clinical phenotype, while mutations in the glucokinase gene cause the mildest form of MODY. The relative prevalence of different subtype of MODY has shown to vary greatly in studies of MODY families from different populations (Chevre et al. 1998; Frayling et al. 1997; Lindner et al.1999). Mutations in glucokinase (MODY2) represents from 8-63% of cases and mutations in HNF1α (MODY3) from 21-64% of cases. The other MODY subtypes are rare disorders in all these populations, having been described only in a few families. MODY has been considered an attractive model for genetic studies because of its well-defined monogenic mode of inheritance and its
genetic, metabolic, and clinical heterogeneity. Studies on MODY genes will bring insights into mechanisms of β-cell dysfunctions.

Since the MODY genes play crucial biochemical roles in the pancreatic β cell, they were considered as strong candidate genes for type 2 diabetes in the general population. The connection between mutations in glucokinase (Hattersley et al. 1992), HNF4α (Hani et al. 1998), and PDX1 (Hani et al. 1999 and Macfarlane et al. 2000) and late-onset type 2 diabetes were observed in a few families. In addition, a genome scanning in 26 finish families indicates that MODY3 may be also related to type 2 diabetes with low insulin secretion. However, a larger set of genetic linkage analyses studies suggested that MODY genes contribute modestly to the clinical features of diabetes (Permutt et al. 1992 and Zouali et al. 1993).

1.3 HNF1α

HNF1α is expressed in the liver, kidney, intestine, stomach, and pancreas and plays a crucial role in regulating transcription of numerous hepatic and pancreatic genes (Baumhueter et al. 1989; Baumhueter et al. 1990; Serfas et al. 1993; Boudreau et al. 2002). In humans, mutations in HNF1α are linked to the occurrence of MODY3. Mice with homozygous inactivation of the HNF1α gene exhibits Laron dwarfism and non-insulin-dependent diabetes mellitus (Lee et al. 1998). The HNF1α gene comprises 10 exons which span ~23kb and encodes a transcription factor which has three functional
domains: N-terminal dimerization domain (amino acids 1-32), DNA binding domain (amino acids 33-280), and COOH-terminal transactivation domain (amino acids 281-631) (Figure 1.1). HNF1α is unique among POU-homeodomain proteins because it forms stable dimers in solution. The dimerization domain of HNF1α consists of the first 32 amino acids and appears to be essential for HNF1α function, since deletion of the dimerization domain reduces HNF1α DNA binding. A small protein of 11KDa, called Dimerization Cofactor of HNF1α (DCoH), displays a restricted tissue distribution and can bind the dimerization domain of HNF1α to stabilize HNF1α homodimers and activate transcriptional activity of HNF1α (Mendel et al. 1991a). Crystallographic structural studies have demonstrated that the DCoH and HNF1α complex comprises a dimer of dimers in which the α-helical HNF1α dimerization domain binds to a α-helical cap on the DCoH dimer that overlies a molecular saddle composed of antiparallel β–sheets (Figure 1.2) (Rose et al. 2000). DCoH has also been shown to be involved in recycling of tetrahydrobiopterin, a cofactor of aromatic amino acid hydroxylases (Rebrin et al. 1995).

The DNA binding domain of HNF1α contains a POU-specific domain (amino acids 100-172) and an atypical homeodomain (amino acids 199-278). Although the POU-specific domain of HNF1α has a weak sequence homology with the typical POU domains, the homologous protein fold and mode of DNA recognition shown by three dimensional structure clearly establish HNF1α as a new member of the POU domain family of transcription factors that exert critical functions during development (Chi et al. 2002). The
homeodomain (85 amino acids) of HNF1α is larger than the typical homeodomain (61 amino acids), consisting of four helices with a loop of 16 amino acids, instead of the usual three, between the helices 2 and 3 (Ceska et al. 1993, Leiting et al. 1993, and Schott et al. 1997). The three dimensional structure of the POU-homeodomain of HNF1α indicates that the POU-specific domain interacts with the 16 amino acid loop of the homeodomain to create a stable interface (Chi et al. 2002) (Figure 1.3). This may explain why the POU-specific domain is required for orientational recognition of homeodomain DNA binding (Tomei et al. 1992).

Both the dimerization domain and the DNA binding domain are highly conserved between HNF1α and its close homologue HNF1β, distinguishing them from other flexible POU-homeodomain factors (Chi et al. 2002). HNF1α and HNF1α can form heterodimers and bind to the same DNA sequences (Mendel et al. 1991b).

The COOH-terminal transcriptional domain of HNF1α has two distinct small regions: ADI (amino acids 546–628/Serine-rich) and ADII (amino acids 281–318/Proline-rich). They are required for efficient induction of hepatic and pancreatic genes (Nicosia et al. 1990).

1.4 MODY3

MODY3 is associated with heterozygous mutations in the HNF1α gene. More than 120 different MODY3 mutations have been identified in various populations so far (McCarthy et al. 2002). They are located throughout the entire three functional domains as
well as in the promoter region of the HNF1α gene. Similar to late age onset NIDDM, patients with MODY3 usually progressively develop deterioration of insulin secretion and impaired glucose tolerance and often progress to insulin requirement (Vaxillaire et al. 2006). However, unlike NIDDM, MODY3 is primarily associated with a low prevalence of obesity, dyslipidaemia and arterial hypertension (Velho and Froguel 1998). In addition to its effects on beta cell function, a deficiency of HNF1α affects the kidneys and genital system. Patients with MODY3 have decreased renal reabsorption of glucose and excrete glucose in the urine (glycosuria) (Pontoglio et al. 1996). Interestingly, although HNF1α is also expressed in the liver, liver function is unaffected in MODY3 patients. In contrast to the phenotypes of MODY3 patients who have a single mutated HNF1α allele, insulin responses to glucose and arginine in mice with heterozygous HNF1α mutations (HNF1α (+/-)) is normal, (Pontoglio et al. 1996, Pontoglio et al. 1998, Dukes et al. 1998, and Pontoglio et al. 2000). However, homozygous HNF1α (-/-) mice exhibit hepatic, pancreatic, and renal dysfunction that significantly affect the growth and life spans of the mutant animals, although they do not die as embryos like HNF1β (-/-) mice do (Coffinier et al. 1999 and Barbacci et al. 1999).

More than 120 HNF1α mutations have been identified including missense, frameshift, and nonsense mutations (Figure 1.4). Two-third of the MODY3 mutants are missense mutants that result in a single amino acid change in full-length protein. Other mutations include truncations either by a point mutation leading to a stop codon (nonsense mutation) or a small (≤5bp) nucleotide deletion/insertion resulting in a frameshift. Many frameshifts are
located within repetitive sequences of the HNF1α gene. The truncations in the HNF1α gene are predominantly in the COOH-terminal transactivation domain, while the missense mutations are spread over the entire gene, but are mostly located in the highly conserved DNA binding domain (Figure 1.4). Most of the MODY3 mutations are rare, but some have been reported in more than one family. Especially, P29fsinsC (insertion of a C in the polyC of exon 4 leading to frameshift), leading to a mutant truncated protein of 315 amino acids, was reported in at least 22 distinct families throughout the world. Another frequent mutation is P379fsdelCT/T/C that was reported in 7 families. Some Arg mutations (R131, R159, R200, R229, R272) caused by the well-known deamination of methylcytosin in CpG dinucleotides were also found in different families. All of the Arg mutations are located in the DNA binding domain. In addition, at least 6 MODY3 mutations are reported in the promoter region, with one mutation (58A>C) in the HNF4α binding site (Gragnoli et al. 1997)

1.5 HNF1α TARGET GENES

HNF1α was originally identified as a liver-enriched transcription factor that functions in the regulation of several liver-specific genes including albumin, α-antitrypsin, β-fibrinogen, pyruvate kinase, transthyretin, and aldolase B (Baumhueter et al. 1990 and Baumhueter et al. 1989). Later, HNF1α was found not to be liver-restricted. By now, more than 300 genes have been identified that may be regulated by HNF1α, distributed over hepatocytes, pancreatic β-cells, and kidney, intestine, and stomach cells (Tronche et al. 1997;
Wang et al. 2000; Odom et al. 2004). By using chromatin immunoprecipitation combined with promoter microarrays, Odom et al. (2004) that identified HNF1α is bound to 222 target genes in hepatocytes and 106 in pancreatic islets with only 35 overlapping genes. These results suggest that HNF1α plays a central role in maintaining normal hepatic function, including carbohydrate synthesis and storage, lipid metabolism (synthesis of cholesterol and apolipoproteins), detoxification (synthesis of cytochrome P450 monooxygenases), and synthesis of serum proteins (albumin, complements, and coagulation factors). In pancreatic islets, HNF1α binds to genes encoding proteins in insulin synthesis and secretion and associated pathway. Different target genes of HNF1α in hepatocytes and pancreatic cells suggest that HNF1α has different functions in these two different types of cells.

Pontoglio et al. (1996) have shown that most HNF1α target genes in hepatocytes are only weakly affected in HNF1α (-/-) mice by analyzing transcription rates of hepatic genes and steady-state levels of mRNA species including the albumin, α1-antitrypsin, β-fibrinogen, pyruvate kinase, tyrosine amminotransferase, aldolase B, glutathione S-transferase (GST), fructose diphosphatase, transthyretine, α-fetoprotein, protein C, phenylalanine hydroxylase, and phosphoenolpyruvate carboxy-kinase (PEPCK). One exception in their analysis is the gene encoding phenylalanine hydroxylase (PAH), which is totally inactive in HNF1α deficient mice. The lack of HNF1α prevents chromatin remodeling in the transcriptional control regions of PAH gene (Parrizas et al. 2001). Another HNF1α dependent hepatic gene is glucose-6-phosphatase transporter gene, reported by Hiraiwa et al. (2000). However, a
large set of pancreatic genes, including insulin, glucose transporter-2, 3-hydroxy-3-methylglutaryl coenzyme A reductase, mitochondrial 2-oxoglutarate dehydrogenase (OGDH), and L-pyruvate kinase, were reported to be affected by the inactivation of HNF1α. This difference suggests that, in hepatocytes but not pancreatic cells, the role of HNF1α may be compensated by other proteins, e.g. HNF1β. In fact, HNF1α and HNF1β are co-expressed in liver, kidney, and pancreas, although with different spatial-temporal patterns (Ott et al. 1991 and Lazzaro et al. 1992). The onset of HNF1α and HNF1β expressions are different during embryonic development. In all species analyzed, the expression of HNF1β always precedes that of HNF1α during embryogenesis. Consistent with this differential temporal expression pattern, inactivation of the HNF1β gene in the mouse leads to early embryonic lethality at day 7.5 (Barbacci et al. 1999 and Coffinier et al. 1999), which has precluded any analysis of HNF1β function in later stages of development.

1.6 HNF TRANSCRIPTIONAL NETWORK AND TISSUE-SPECIFIC GENE EXPRESSION

The phenotype of each individual cell type is regulated and defined by the expression of specific groups of genes. As mentioned above, five of six MODY genes are transcriptional factors. They consist of a precise genetic network and play a crucial role in regulating the development and function of liver, pancreatic islets, and other tissue (Figure 1.5). Recently, by using Chip combined with promoter microarray, more than 2000 hepatic and pancreatic
genes were found under the control of HNF1α, HNF6, and HNF4α. Despite the complexity of this regulatory network, several motifs are identified, including an HNF1α/HNF4α feedback loop, HNF6/HNF4α feedforward loop, and HNF4α autoregulation loop (Odom et al. 2004).

The regulation between HNF1α and HNF4α are well documented. Several labs distinctly demonstrated that HNF1α and HNF4α occupy one another’s promoter in both hepatocytes and islets by using transient transfection assays, mutational analyses, or a transgenic mouse model (Stoffel et al. 1997; Kuo et al. 1992; Li et al. 2000; Boj et al. 2001; Thomas et al. 2001). This capacity for feedback control provides a stable mechanism of gene expression maintained by a bi-stable system that can switch between two alternate states. HNF1α has two promoter regions, respectively located between positions -82 and -40 and between -118 and -8. The first promoter contains a single site for HNF4α binding (Tian and Schibler 1991); the second promoter contains both HNF4α and HNF1α binding sites (Miura and Tanaka, 1993). Interestingly, the HNF4α gene also has two different promoters (P1 and P2), with promoter P2 located around 45.6 kb upstream of P1. Two different HNF4α isoforms, HNF4α1 and HNF4α7, are transcribed by P1 and P2 promoters, respectively, and differ only by the amino acids encoded by their first exon. These two promoters are differentially utilized in hepatocytes and pancreatic islets (Figure 1.6). By using chromatin immunoprecipitation analysis and HNF1α deficient mice, promoter P1 was shown predominantly to be occupied by HNF1α in hepatocytes, while promoter P2 is exclusively
used in pancreatic islets (Boj et al. 2001). Meanwhile, the studies by Thomas et al. (2001) similarly demonstrated that the P2 promoter is the major regulatory site for islet-specific HNF4α transcription, although it is also used by HNF1α in hepatocytes. This tissue-specific occupancy by HNF1α of two HNF4α promoters suggests that the regulatory network and transcriptional machinery in liver and pancreatic islets may be different.

Another important transcription factor in the regulatory network is PDX1, expression of which in pancreatic islets is regulated by HNF1α and HNF4α as well as by the forkhead transcription factor HNF3β (FoxA2) (Ben-Shushan et al. 2001; Shih et al. 2001; Wu et al. 1997; Gerrish et al. 2000). PDX1 is a master regulator of both pancreatic development and the differentiation of progenitor cells into the β-cell phenotype, with its expression maintained throughout development in islet cells (Stoffer et al. 1997; Habener et al. 1998; Guz et al. 1995). PDX1 regulates transcription of the genes associated with cell identity, including insulin (Ohlsson et al. 1993 and Peers et al. 1995), glucokinase (Watada et al. 1996), islet amyloid polypeptide (Bretherton-Watt et al. 1996), and glucose transporter type 2 (Watada et al. 1996). Activated by upstream regulator HNF3β, PDX1 expression is first detected on E8.0 (embryonic day 8.0 of mouse development) in the dorsal endoderm of the foregut/midgut junction (Wu et al. 1997; Gerrish et al. 2000). On E9.5, PDX1 expression is seen in both ventral and dorsal pancreatic buds and then reaches the highest level until E10.5 (Edlund 1998). The in vitro studies demonstrated that HNF1α can bind to an evolutionarily conserved HNF1α binding site in the PDX1 promoter and activate its transcription
(Ben-shushan et al. 2001). Furthermore, PDX1 expression levels were reduced in islets of HNF1α (-/-) mice, indicating that HNF1α is required for normal PDX1 expression in pancreatic β cells.

HNF1α is not required to set up the early transcriptional program in β cell differentiation. However, HNF1α is expressed in many cell types of developing pancreas after E13.5, and required for maintaining the expression of HNF4α, HNF3γ, and GLUT2 after E14. This was verified by the fact that expression levels of HNF4α, HNF3γ, and GLUT2 are progressively reduced in HNF1α (-/-) mice after E14 of embryonic development, but expression of these genes are independent of HNF1α expression in the early developmental stages.

A detailed molecular understanding of the transcriptional network of pancreatic β cell will provide us a comprehensive insight into the mechanisms of different MODYs and also contribute to the effort to regenerate missing or malfunctioning β cells in type II diabetes and transplantation of artificial generated β cells in type 1 diabetic patients.

1.7 COOPERATIVE ACTIVATION ON TARGET GENES BY HNF TRANSCRIPTIONAL FACTORS AND OTHER REGULATORS

In hepatocytes or pancreatic cells, genetic regulatory networks are built up with multiple regulators that synergize with HNF transcriptional factors to activate downstream target genes. Numerous co-regulators of HNF1α and HNF4α have been identified, including
the coactivators p300/CBP, and p/CAF, and the transcription factors GATAs, FoxA2, Cdx2. These form efficient transcriptional regulation complexes, also termed enhanceosome. For instance, the liver fatty acid binding protein gene (Fabpl) is activated up to 60 times greater by HNF1α in the presence of the other factors than in their absence (Divine et al. 2003). The mechanism for synergistic activation remains unknown. However, it has been proposed that the increased efficiency of assembly of a competent RNA polymerase II initiation complex could be a cause (Merika and Thanos 2001). In addition, protein-protein interactions have been reported between HNF1α and some regulators, like p300/CBP, p/CAF, CDX2 or GATA5 (Heinemeyer 1998; Krasinski 2001).

Since MODY3 mutations are located throughout the full-length of HNF1α, some MODY3 mutations might affect the cooperation between HNF1α and other regulators. It has been reported that, in HepG2 cells, MODY3 mutants R131Q and Y122C exhibit complete loss of synergistic activation Fabpl gene in concert with the other regulators. Nevertheless, when HNF1α mutants R131Q or Y122C were transfected alone without other factors, complete (R131Q) or 60% (Y122C) wild-type HNF1α activity was observed (Divine et al. 2003). Authors suggested that this loss of transcriptional synergy could be an explanation for reduced transcriptional activity of MODY3 mutants. Similarly, HNF1α also was reported to recruit p300/CBP and P/CAF in cells (Soutoglou et al. 2001). Transcriptional synergy was observed among these three proteins. HNF1α is also required to recruit the HAT activities of coactivator p300/CBP and P/CAF. Interestingly, two MODY3 mutations (P447L and P519L)
abolish this synergy, illustrating that disruption of synergy of HNF1α with other proteins is one mechanism of MODY3.

1.8 POSSIBLE MECHANISMS IN THE DEVELOPMENT OF MODY3

MODY3 mutations exhibit autosomal dominant inheritance, which means MODY3 can be passed down through families, and a single MODY3 allele from one parent is capable of causing disease. But why is one abnormal allele of the MODY3 linked gene, *hnf1α*, enough to result in disease? As we can imagine, one defective allele of *hnf1α* will cause decreased levels of HNF1α protein in cells, which may be the cause of MODY3. So, why do decreased levels of HNF1α protein cause MODY3? A good explanation for this question is that the transcription factor HNF1α plays a crucial role in cells, and there is no redundant HNF1α or other proteins to compensate for the loss of HNF1α in cells. In fact, many transcription factors are short-lived proteins. They are generated in certain tissues at certain developmental stages. After completion of their tasks, they are usually degraded by other mechanisms. On the other hand, some transcription factors are redundant in cells. For instance, there are multiple bZIP and forkhead transcription factors in cells. Although specificity is required in a certain situation, overlapping functions of these proteins are common. So altered levels of these proteins usually doesn’t cause demonstrable phenotypes.

A single abnormal allele can cause the loss of partial or full function of proteins by decreasing the amount of wild type proteins. This mechanism is called haploinsufficiency.
Most of the reported MODY3 mutations have been attributed to haploinsufficiency, and there are few cases where a dominant negative effect has been seen in transfection assays. The definition of a dominant negative effect is the inhibition of the function of a wild type gene product by an inhibitory variant of the same product. In general, therefore, dominant negative mutant proteins will retain intact and functional domains of the parent wild type protein, but must have the complement of non-functional domains by either missing or alteration of amino acids (Herskowitz 1987). In Herskowitz’s paper, he gave two possibilities for the mechanism of a dominant negative effect: (1) if a protein is multimeric, a variant capable of interacting with wild-type polypeptide chains but otherwise defective will be inhibitory if it causes the formation of non-functional multimers; (2) the activity of a wild type protein is limited by the availability of substrate, then a variant capable of binding substrate but not of performing proper functions could be inhibitory. In the second scenario, wild type proteins are out-competed from the binding of substrate by non-functional variants. Since we are trying to distinguish these two possible mechanisms in the development of MODY3, here, we name the first possible mechanism as “heteromerization dominant negative inhibition”, and the second one as “out-competing dominant negative inhibition”.

It has been proposed that the mutant proteins generated by MODY3 allele could heterodimerize with wild type HNF1α proteins and interfere with its activity to result in a dominant negative effect, while the mutant proteins which could not heterodimerize with wild type HNF1α may just lose partial or full activity to cause haploinsufficiency (Ellard
However, this proposed hypothesis is only confirmed by some studies but not all. The conflicting data were demonstrated, even on the same MODY3 mutations, by transient transfection assays. For instance, P291fsinsC, the most common MODY3 mutation, was suggested to be a dominant negative mutation by several studies (Yamagata et al. 1998; Wang et al. 1998; Vaxillaire et al. 1999), whereas, another laboratory showed that P291fsinsC does not inhibit wild type activity in a dominant negative manner suggesting it inhibited through a haploinsufficiency mechanism (Thomas et al. 2002). Yaniv’s lab (Vaxillaire et al. 1999) also showed data contradictory to the proposed hypothesis. In their studies, 10 MODY3 mutations located at either DNA binding domain or transactivation domain were analyzed. These mutations included missense (Y122C, R131Q, R159Q, K205Q, R272H, and P447L), nonsense (R171X), and frameshift (P291fsinsC and T547E548fsdelTG) mutations. The authors demonstrated that HNF1α with MODY3 mutations in the DNA binding domain, like Y112C, R131Q, and S142F, lose partial or full transcriptional activity but don’t exhibit dominant negative inhibition of wild type HNF1α. However, these MODY3 mutants contain intact dimerization domains and are presumably able to heterodimerize with wild type HNFα. According to the previously proposed hypothesis, therefore, they should have dominant negative effects on wild type HNF1α.

In table 1.2, I summarize the functional data of 31 MODY3 mutations that have been analyzed in 12 previously published papers. Seventeen of the mutations were found to have a reduced transcriptional activity, while nine have complete loss-of-function. These 26
mutations were shown to have no effect on wild type HNF1α. The other five MODY3 mutations were reported to be dominant negative inhibitors of wild type HNF1α. Among these 31 MODY3 mutations, three (P447L, P291fsinsC, and T547E548fsdelTG) were reported to have contradictory functional properties by two different laboratories. Most of the analyzed MODY3 mutations cause loss of function. Moreover, the phenotype does not appear to be more severe in cases where a dominant negative effect is observed (Frayling et al. 2001). 15 MODY3 mutations in the DNA binding domain and 7 in the transactivation domain were reported to cause haploinsufficiency, which can’t be explained by the previously proposed hypothesis, since these mutations all have intact dimerization domains and are supposed to interfere with wild type HNF1α activity.

These contradictory results implied that MODY3 probably is caused by unknown mechanisms different from the previously proposed hypothesis. In addition, usage of different cell lines in different laboratories might explain the contradictory data. For example, Yaniv’s (Vaxillaire et al. 1999) and Ryffel’s (Thomas et al. 2002) laboratories performed transient transfection assays by using C33 human epithelial cervical carcinoma and HeLa cells, both of which lack endogenous HNF1α expression.; Matsuzawa’s laboratory performed transient transfection assays by using MIN6 cells, a mouse β-cell line, which has endogenous HNF1α. Another problem in most experimental systems is that higher amounts of MODY3 mutants may be transfected than would be expected in cells. For example, in Yamagata’s studies, 500 ng of P291fsinsC RIP vector was transfected into MIN6 cells which only
decreased the transcriptional activity of wild type HNF1α by half. In addition, some coactivators, like DCoH, may play important roles in the development of MODY3. Thus, the transient transfection assays performed in the absence of them may not elicit the proper mechanism for MODY3 disease.

To investigate the previously proposed hypothesis and the role of the dimerization domain of HNF1α in MODY3 disease, we are performing a series of transient transcription assays to test effects of different HNF1α variants on wild type HNF1α. In these transfection assays, we are using β-TC3 cells that have endogenous HNF1α and DCoH, which enable us to explore the regulation of HNF1α by DCoH. Knowledge of this regulation will help us clarify the roles of other proteins involved, like ERH (enhancer of rudimentary homolog) that was reported to be an interaction partner of DCoH and repress the enhancement of HNF1α transactivity by DCoH.
Table 1.1: Six subtype MODYs

<table>
<thead>
<tr>
<th>Classification</th>
<th>Defect</th>
<th>Chromosomal location</th>
<th>Severity</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODY1</td>
<td>HNF4 alpha</td>
<td>20q12-q13.1</td>
<td>+++</td>
<td>Uncommon</td>
</tr>
<tr>
<td>MODY2</td>
<td>Glucokinase</td>
<td>7p15.3-p15.1</td>
<td>+</td>
<td>Common</td>
</tr>
<tr>
<td>MODY3</td>
<td>HNF1 alpha</td>
<td>12q24.3</td>
<td>+++</td>
<td>Common</td>
</tr>
<tr>
<td>MODY4</td>
<td>PDX-1</td>
<td>13q12.1</td>
<td>++</td>
<td>Rare</td>
</tr>
<tr>
<td>MODY5</td>
<td>HNF1 beta</td>
<td>17cenq21.3</td>
<td>+++</td>
<td>Rare</td>
</tr>
<tr>
<td>MODY6</td>
<td>NeuroD1</td>
<td>2q32</td>
<td>+++</td>
<td>Rare</td>
</tr>
</tbody>
</table>
Table 1.2: Biological function of MODY3 mutations

<table>
<thead>
<tr>
<th>Effect</th>
<th>Missense</th>
<th>Truncations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced activity</td>
<td>L12H\textsuperscript{1,2,3,4,5}</td>
<td>P291fsinsC\textsuperscript{7}</td>
</tr>
<tr>
<td></td>
<td>L107I\textsuperscript{8}</td>
<td>P379fsdelCT\textsuperscript{1,3}</td>
</tr>
<tr>
<td></td>
<td>Y122C\textsuperscript{6}</td>
<td>T547E548fsdelTG\textsuperscript{7}</td>
</tr>
<tr>
<td></td>
<td>R131Q\textsuperscript{6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R159Q\textsuperscript{2,6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G191D\textsuperscript{1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R203C\textsuperscript{2}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R205Q\textsuperscript{6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R229Q\textsuperscript{7}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R263C\textsuperscript{1,3}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R272H\textsuperscript{6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G415R\textsuperscript{9}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P447L\textsuperscript{7}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T620I\textsuperscript{7}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P291fsinsC\textsuperscript{7}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P379fsdelCT\textsuperscript{1,3}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T547E548fsdelTG\textsuperscript{7}</td>
<td></td>
</tr>
<tr>
<td>Complete loss of activity</td>
<td>G20R\textsuperscript{4,5}</td>
<td>Q7X7</td>
</tr>
<tr>
<td>(these presumably have no effect</td>
<td>ΔN127\textsuperscript{7}</td>
<td>R171X\textsuperscript{7}</td>
</tr>
<tr>
<td>on wild type activity)</td>
<td>P112L\textsuperscript{12}</td>
<td>R229X\textsuperscript{2}</td>
</tr>
<tr>
<td></td>
<td>S142F\textsuperscript{6}</td>
<td>L584S585fsinsTC\textsuperscript{1,3,7}</td>
</tr>
<tr>
<td></td>
<td>Q446X\textsuperscript{12}</td>
<td></td>
</tr>
<tr>
<td>Dominant negative</td>
<td>K158N\textsuperscript{2}</td>
<td>P291fsinsC\textsuperscript{6,10,11}</td>
</tr>
<tr>
<td></td>
<td>R272C\textsuperscript{9}</td>
<td>T547E548fsdelTG\textsuperscript{6}</td>
</tr>
<tr>
<td></td>
<td>P447L\textsuperscript{6}</td>
<td></td>
</tr>
</tbody>
</table>

This table is modified from table 2 in Ryffel’s review (2001) and complemented by recent data. These mutations are from \textsuperscript{1}Yang et al. (1999), \textsuperscript{2}Yamada et al. (1999), \textsuperscript{3}Okita et al. (1999), \textsuperscript{4}Rose et al. (2000), \textsuperscript{5}Hua et al. (2000), \textsuperscript{6}Vaxillaire et al. (1999), \textsuperscript{7}Thomas et al. (2002), \textsuperscript{8}Cervin et al. (2002), \textsuperscript{9}Yoshiuchi et al. (1999), \textsuperscript{10}Wang et al. (1998), \textsuperscript{11}Yamagata et al. (1998), and \textsuperscript{12}lise et al. (2003).
Figure 1.1: Scheme of the hepatocyte nuclear factor 1α (HNF1α) transcription factor. The various functional domains are indicated. 10 exons of HNF1α gene are shown on the top. The correspondence between 10 exons and functional domains are indicated. AD II: Activation Domain II (amino acids 281–318/Pro-rich); ADI: Activation Domain I (amino acids 546–628/Ser-rich).
**Figure 1.2:** Three dimensional structure of dimerization domain of HNF1α and DCoH. (A) Three dimensional structure of dimerization domain of HNF1α (Rose *et al.* 2000b). The HNF1α dimer folds into an antiparallel four-helix bundle. Each monomer consists of two helices: a longer N-terminal helix (residues 4-18) and a shorter C-terminal helix (residues 22-30), connected by a short loop (residues 19-22). (B) DcoH binds Dimerization domain of HNF1α (Rose *et al.* 2000a) The DCoH dimer (red and slate) binds the HNF1α dimer (cyan and purple) with two-helix of DCoH in contact with two-helix of HNF1α.
Figure 1.3: The stable interface between specific POU domain and homeodomain (Chi et al. 2002). (A) The significant contact between specific POU domain (red) and homeodoamain (cyan). The unique HNF1α homeodomain insertion interacts at this interface with specific POU domain residues. The amino acids involved in the interaction are shown as ball-and-stick models (B) Structure of the HNF1α/DNA Complex. Two protein molecules are bound to dsDNA. The five α helices of each specific-POU domain and three α helices in each homeodomain domain are labeled Pα1–5 and Hα1–3, respectively (Chi et al. 2002)
**Figure 1.4:** Summary of mutations in the HNF1α gene. Human HNF1α MODY3 mutations are spread over the three functional domains. Frame shift mutations generating truncated proteins are indicated with bold letters. Missense mutations are shown with italics letters. The frame shift mutations are mainly located in the transactivation domain, while the missense mutations are evenly distributed in the three functional domains.
Figure 1.5: Proposed model for a hierarchical transcriptional network in pancreatic islets.

This diagram shows the transcriptional regulation of genes essential for pancreatic islet function by PDX1 and hepatocyte nuclear factors. Arrows show positive regulation. The numbers on the top (E8.0 –E15.5) indicate the approximate onset of gene expression during embryonic development.
Figure 1.6: The two promoters P1 (HNF4α1) and P2 (HNF4α7) of HNF4α are differentially utilized in hepatocytes and pancreatic islets. The upper panel is hepatocyte, in which both P1 and P2 promoters are transcribed to generate two HNF4α isoforms: HNF4α1 and HNF4α7. The lower panel is pancreatic cells, in which P2 promoter is predominantly transcribed to generate HNF4α7. The dash arrows show the regulation of promoters by HNF1α, HNF4α1, or HNF4α7.
REFERENCES:


11. Chi, Y.I., Frantz, J.D., Oh, B.C., Hansen, L., Dhe-Paganon, S., and Shoelson, S.E.


cause of maturity-onset diabetes of the young in the U.K. Diabetes 46, 720-725.


missense mutation in hepatocyte nuclear factor-4 alpha, resulting in a reduced


Physiol 276, G1181-1194.

Differential activation of intestinal gene promoters: functional interactions between
GATA-5 and HNF-1 alpha. Am J Physiol Gastrointest Liver Physiol 281, G69-84.

32. Kuo, C.J., Conley, P.B., Chen, L., Sladek, F.M., Darnell, J.E., Jr., and Crabtree, G.R.

LFB1 and LFB3 homeoproteins are sequentially expressed during kidney
development. Development 114, 469-479.

non-insulin-dependent diabetes mellitus in the Hnf-1alpha knockout mouse. Mol Cell
Biol 18, 3059-3068.

35. Leiting, B., De Francesco, R., Tomei, L., Cortese, R., Otting, G., and Wuthrich, K.
(1993). The three-dimensional NMR-solution structure of the polypeptide fragment
195-286 of the LFB1/HNF1 transcription factor from rat liver comprises a nonclassical
homeodomain. Embo J 12, 1797-1803.

requires the transcription factor HNF-4alpha. Genes Dev 14, 464-474.

37. Lindner, T.H., Cockburn, B.N., and Bell, G.I. (1999). Molecular genetics of MODY in
Germany. Diabetologia 42, 121-123.


are essential elements of the tripartite DNA binding structure of LFB1. Cell 61, 1225-1236.


57. Shih, D.Q., Screenan, S., Munoz, K.N., Philipson, L., Pontoglio, M., Yaniv, M.,


hepatocyte nuclear factor 1 may involve hepatocyte nuclear factor 4. Genes Dev 5, 2225-2234.


gene transcription and impaired metabolism-secretion coupling in a pancreatic beta-cell line. Embo J 17, 6701-6713.


CHAPTER 2

THE MECHANISM OF DOMINANT NEGATIVE INHIBITION OF WILD TYPE HNF1α BY MODY3 MUTANTS

BY

CHENG WAN AND ROBERT ROSE
2.1 ABSTRACT:

The type 3 form of maturity-onset diabetes of the young (MODY3) is caused by mutations in the gene encoding the transcription factor hepatocyte nuclear factor-1alpha (HNF1α) which plays an important role in glucose homeostasis in pancreatic islets. MODY3 is inherited in an autosomal dominant pattern since mutations in only one allele of the HNF-1α gene impair pancreatic beta-cell function. However, the molecular mechanism for the dominant phenotype of MODY3 remains unclear. It has been proposed that the MODY3 mutations with an intact dimerization domain, which are able to heterodimerize with wild type HNF1α and interfere with its activity, may result in a dominant negative effect, while the mutant proteins which could not heterodimerize with wild type HNF1α may just lose partial or full activity to cause haploinsufficiency. In the current study we address whether dominant negative inhibition can be caused by heterodimerization between wild type HNF1α and MODY3 mutations. We characterize a MODY3 mutation outside of the dimerization domain that does not heterodimerize with wild type HNF1α but causes a dominant negative effect by competing for DNA binding. We propose that heterodimerization between mutant and wild type HNF1α may not occur due to stabilization of HNF1α dimers by the coactivator DCoH. This study has implications for how MODY3 mutations may affect HNF1α activity in patients.
2.2 INTRODUCTION

Maturity-onset diabetes of the young (MODY) is a monogenic form of type 2 diabetes characterized by early age of onset and autosomal dominant transmission and is usually not associated with insulin resistance (Fajans et al. 2001, Frayling et al. 2001). To date, 6 MODY genes have been identified. MODY2 is caused by mutations in the glucokinase gene, which determines the level of glucose capture in beta cells (Froguel et al. 1993). Glucose initiates glycolysis to signal insulin secretion. The remaining MODY subtypes are the result of mutations in genes coding for transcription factors, which are Hepatocyte Nuclear Factor 4α (HNF4α/MODY1), Hepatocyte Nuclear Factor 1α (HNF1α/MODY3), Pancreas Duodenum Homeobox-1 (PDX1/MODY4), Hepatocyte Nuclear Factor 1β (HNF1β/MODY5), and NeuroD1/Beta2 (MODY6) (Fajans et al. 2001, Frayling et al. 2001, Malecki et al. 1999).

MODY3 is caused by mutations in a single allele of the HNF1α gene, accounting for 20–65% of all MODY subtypes. So far, more than 120 different MODY3 mutations have been identified in various populations (McCarthy et al. 2002). They are located throughout the full length of HNF1α, which is composed of 3 functional domains (Figure 2.1): the N-terminal dimerization domain, DNA binding domain, and C-terminal transactivation domain, as well as in the promoter region. HNF-1α is expressed in liver, kidney, intestine, and pancreatic islets (Blumenfeld et al. 1991), regulating the expression of target genes as a
homodimer or as a heterodimer with the structurally and functionally related transcription factor HNF1β (Cereghini, 1996). Recently, several studies showed that HNF1α is involved in the regulation of genes encoding proteins in insulin synthesis and secretion and associated pathways, such as L-type pyruvate kinase (PKL), glucose transporter (GLUT)2, glucokinase, and aldolase-B (Baumhueter et al. 1990, Baumhueter et al. 1989, and Wang et al. 2000). Therefore, MODY3 patients usually have impaired insulin secretion in response to glucose, resulting from impaired glucose utilization and mitochondrial oxidation of metabolic substrates.

In contrast to the phenotypes of MODY3 patients who have a single mutated HNF1α allele, insulin responses to glucose and arginine in heterozygous knockout mice lacking one copy of HNF1α (HNF1α (+/-)) is normal, (Pontoglio et al. 1996, Pontoglio et al. 1998, Dukes et al. 1998, and Pontoglio et al. 2000). However, homozygous HNF1α (-/-) mice exhibit hepatic, pancreatic, and renal dysfunction that significantly affect the growth and life spans of the mutant animals, although they do not die as embryos like HNF1β (-/-) mice do (Coffinier et al. 1999 and Barbacci et al. 1999).

Although it is known that HNF1α plays a crucial role in glucose homeostasis in pancreatic islets, how the heterozygous MODY3 mutations in the HNF1α gene affect its overall activity remains to be determined. Mutations can cause disease by loss-of-function or dominant negative mechanisms. Loss-of-function mutations cause phenotypes similar to
those that would be caused by mRNA null mutations for the allele. Dominant negative mutations arise when the mutant allele interferes with the function of the remaining wild-type allele, causing a >50% loss of function. This can occur when the gene product forms active complexes with itself (and possibly with other proteins). A common assumption suggests that HNF1α mutants with an intact dimerization domain impair pancreatic β–cell functions in a dominant negative manner by forming nonproductive dimers with the wild-type protein. On the other hand, the HNF1α mutants with mutations in the dimerization domain would have reduced stability or decreased transactivation, thereby eliciting haploinsufficiency. This hypothesis was supported by several laboratories (Yamagata et al. 1998; Wang et al. 1998). However, contradictory data were also reported. For instance, it was shown that some HNF1α mutants with an intact dimerization domain, like Y112C, R131Q, or S142F, do not interfere with wild type HNF1α activity in cell cultures, suggesting these mutations disrupt activity of the wild type HNF1α allele through haploinsufficiency (Vaxillaire et al. 1999).

To clarify the role of heterodimerization in allowing MODY3 mutations to interfere with wild type HNF1α activity, we investigated in detail the functional properties of different HNF1α mutants in the mouse pancreatic beta cell line β-TC3 that express endogenous HNF1α. In this study, we demonstrate that competition for DNA binding of wild-type HNF1α, not heterodimerization between mutant and wild-type protein, can cause a dominant negative effect on HNF1α activity. A similar competition mechanism may explain other
MODY3 mutations.

2.3 MATERIALS AND METHODS

2.3.1 Plasmid construction and site-directed mutagenesis

Full-length HNF1α and its truncated derivatives were subcloned into the CMV-driven vector pBat12 [a derivative of pBluescript KS (Stratagene), a gift from MS German, University of California San Francisco] by using the vector qeHNF1α (a gift from GR Crabtree, Stanford University) as the template. The rop gene was obtained by PCR from E.coli lysate, and then subcloned into the pBat12 vector by using AvrII and HindIII restriction enzymes. Different lengths of hnf1α gene fragments were amplified by PCR and subcloned into ROP-pBat12 vectors by using EcoRI and HindIII restriction enzymes to generate ROP-HNF1α (32-628)-pBat12 and ROP-HNF1α (32-280)-pBat12 vectors. The serine to phenylalanine mutation at amino acid 142 in the hnf1α gene was constructed in HNF1α (1-628)-pBat12, HNF1α (1-280)-pBat12, and ROP-HNF1α (32-280)-pBat12 vectors using QuickChangeTM site-directed mutagenesis kit (using the primer 5’-CCGACTGA CGCGCACGCTCA-3’, the mutated nucleotides are underlined) (Stratagene #200515). A reporter gene construct, pGL3-RA, containing the promoter of the rat albumin gene (nucleotides -170 to +5), cloned into the Smal site of the firefly luciferase reporter vector pGL3-Basic (Promega #E1751) was kindly provided by Professor Graeme Bell (Howard Hughes Medical Institute, University of Chicago, IL). pGL3-DRA vector was
generated using QuickChange™ site-directed mutagenesis kit (5’-GGAATGGACGT TGGCAACTCA-3’), so that the HNF1α consensus binding site GTTAATNATTAGC is deleted. The sequences of all constructs were confirmed by automated DNA sequencing.

2.3.2 Cell culture and transfection

β-TC3 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) (Fisher) supplemented with 2.5% (v/v) fetal bovine serum (Hyclone), 15% (v/v) horse serum (Hyclone), 100 units/ml penicillin and 100 units/ml streptomycin, and incubated at 37°C in a humidified incubator with 5% CO2/95% air. The transient transfection assays were performed by using Fugene 6 Kits (Roche #11814443001) according to the manufacturer’s protocol. Briefly, cells were plated on a 96-well plate (2x10^4 cells/well). After reaching 50% ~ 70% confluence (approximately 24 hours), transfections were performed with 50 ng pGL3-RA reporter vector and different amounts of expression vectors using Fugene 6 with a DNA/lipid ratio of 3:1. All transfections contained the same total amount of DNA (500 ng), using pBat12 as an empty vector. After 48 hours, cells were harvested using 25 microliters of reporter lysis buffer (Promega #E4030). In order to normalize samples for the total number of cells present, total protein concentration was measured from 10 microliters of cell lysates by the BCA assay (Pierce #23225). The standard curve was calculated from serial concentrations of BSA protein. The luciferase activity of 10 microliters of cell lysates was measured using a Perkin Elmer Victor Light 1420 luminometer. All transfection experiments
were repeated at least three times.

2.3.3 Preparation of nuclear extraction

Nuclear protein extracts from β-TC3 cells were prepared as reported (Andrews and Faller 1991). Briefly, β-TC3 cells were washed with PBS twice. Then low salt lysis buffer (10 mM HEPES, pH 7.9; 10 mM KCl; 0.1 mM EDTA) with protease inhibitor cocktail (leupeptin, PMSF, aprotinin, pepstatin, and DTT) was added to cells. The crude nuclei pellet from the cell lysate was separated after centrifuging at 15,000x g for 3 minutes and resuspended in high salt lysis buffer (20 mM HEPES, pH 7.9; 0.4 M NaCl; 1 mM EDTA; 10% Glycerol) with the protease inhibitors described above. The supernatant (nuclear protein extracts) was separated after centrifuging at 20,000x g for 5 minutes.

2.3.4 Western blot analysis

β-TC3 cell nuclear extracts were prepared as described above followed by boiling of samples in 2x SDS sample buffer. Equal volumes of total samples were separated on a 12% SDS-PAGE gel. Then the proteins were transferred to pre-wet nitrocellulose membrane according to the manufacturer's instructions (BioRad Mini Trans-Blot Cell). The membrane was blocked with 1 x PBS/0.1% Tween 20/5% skim milk for 2-4 hours at room temperature, or overnight at 4 °C. Protein was probed by a monoclonal anti-HNF1α antibody (1:2000 dilution; Genetex, # GTX12064). The transferred protein was incubated with an anti-mouse
IgG antibody conjugated with horseradish peroxidase for 1 hour. All washes were in 1 x PBS/0.1% Tween 20 buffer for 5 min. The proteins were then detected using enhanced chemiluminescent (ECL) detecting reagents and autoradiography (Amersham Pharmacia #28-9068-35).

2.4 RESULTS

2.4.1 Endogenous HNF1α activity detected in β-TC3 cells.

The endogenous HNF1α activity in β–TC3 cells was examined by using a transient transfection assay. Beta-TC3 cells were transfected with 50 ng of pGL3-RA luciferase reporter vector. pGL3-RA vector contains nucleotides -170 to +5 of the rat albumin promoter, which has a consensus HNF1α binding site. The cells were harvested 24 hours post-infection. 10µl of the harvested cells were removed to normalize for variations in cell number by using a BCA protein assay. Luciferase enzyme activity was expressed in terms of relative light units /mg of cell protein determined by the BCA assay (RLU/mg). Table 2.1 show in detail how the data is processed. The β-TC3 cells transfected with pGL3-RA vector has approximately 75 times higher transactivation activity than untransfected β-TC3 cells (Figure 2.2). When β-TC3 cells were transfected with 50 ng, 100 ng, 200 ng, or 400 ng of empty pBat12 vectors, the luciferase counts are similar. No significant difference in luciferase counts were detected between β-TC3 cells transfected with or without pBat12
vector. So we used untransfected β-TC3 cells as a baseline in the following transfection assays. All luciferase reporter assays were repeated at least three times.

To eliminate the possibility that the stimulated activity observed with the pGL3-RA vector is caused by other transcription factors binding to the albumin promoter, we also transfected β-TC3 cells with a pGL3-DRA vector in which the HNF1α binding site in the albumin promoter sequence was deleted. β-TC3 cells transfected with pGL3-DRA have 3.8 times lower transactivation activity than with pGL3-RA (Figure 2.2). This difference is most likely due to the activity of HNF1α bound to pGL3-RA. However, we are still unable to rule out the possibility that other transcription factors in β-TC3 cells, like PDX1 which binds to the consensus HNF1α binding site with high affinity, transactivate the reporter gene. To investigate this question, we are planning to use siRNA to reduce the HNF1α protein in β-TC3 cells. Reduced transactivation activity would be expected in knockdown cells transfected with pGL3-RA if HNF1α is the main cause for the increased transactivation activity in β-TC3 cells transfected with pGL3-RA. Transfection of siRNA for HNF1α was reported by Dohda et al. (2004). The siRNA was designed based on the mRNA sequence [HNF1α (sense), 5’-AGUCUCGAUGACACCGUGGdTdT-3’; HNF1α (antisense), 5’-CCACGGUGUCAUCGAGUCdTdT-3’]. Following transfection of siRNA, the HNF1α protein concentration in the cell will be determined by Western blotting nuclear extracts. Endogenous HNF1α mRNA levels will also be quantified by real time RT-PCR to confirm the results of Western blotting. The real-time quantitative PCR will be performed by using
SYBR Green PCR core kit reagents (Bio-Rad #172-5850). The forward and reverse primers will be 5’-GTACGGTACCAGACAGTCTGCA A-3’ and 5’-TGGTGTAAGGCCGACAG A-3’ (from 120aa to 350aa). First-strand cDNA will be prepared with Super-Script II kit (Invitrogen #11904-018) using either oligo(dT)\textsubscript{12–18} or the HNF1\textalpha gene-specific probe (5’-CCAGCAGCGAGGTCCCTTAGTCAC-3’).

2.4.2 Transfected full length HNF1\textalpha has much higher activity in β-TC3 cells, compared to the endogenous HNF1\textalpha.

To examine transactivation activity by exogenous HNF1\textalpha in β-TC3 cells, various amounts of the pBat12 vector expressing HNF1\textalpha was transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vectors. The total amount of transfected plasmid was kept constant at 500 ng using empty pBAT12 vector in addition to the reporter and HNF1\textalpha vectors. With more transfected HNF1\textalpha, higher activities of the reporter luciferase gene were detected (Figure 2.3).

2.4.3 S142F mutation in DNA binding domain disrupts transactivation activity of HNF1\textalpha.

The MODY3 mutation S142F is located in the DNA binding domain of HNF1\textalpha. Mutant S142F was previously demonstrated to lack any DNA binding activity in electrophoretic mobility shift assays (Vaxillaire et al. 1999). In the future we will confirm this
result by EMSA. Co-transfection of mutant S142F and wild type HNF1α into C33 cells was shown not to inhibit wild type activity in a transient transfection assay (Vaxillaire et al. 1999). This result apparently conflicts with the hypothesis that MODY3 mutants of HNF1α with an intact dimerization domain would heterodimerize with wild type HNF1α and inhibit its activity in a dominant negative manner. To investigate how mutant S142F affects wild type HNF1α activity, we co-transfected various amounts of full-length HNF1α with the S142F mutation as well as 50 ng of pGL3-RA reporter vector into β-TC3 cells that expresses endogenous HNF1α. The transactivation activities by the exogenous HNF1α observed previously are completely lost, suggesting DNA binding ability is indispensable for HNF1α transcriptional activity (Figure 2.4). However, the activity from the endogenous HNF1α was retained (Figure 2.4). Even after transfecting up to 400 ng of HNF1α mutant S142F, no dominant negative inhibition of endogenous wild type HNF1α in β-TC3 cells was detected (Figure 2.4). These experiments suggest that the transfected HNF1α_S142F mutant does not dimerize with the endogenous wild type HNF1α.

2.4.4 Transfected HNF1α (1-280) inhibits endogenous HNF1α activity.

P291fsinsC is the most common MODY3 mutation, which encodes a mutant truncated protein of 315 amino acids comprising the N-terminal dimerization and DNA binding domain but lacking the transactivation domain. Previously, contradictory data about the mutant P291fsinsC were published by two different laboratories. Transient expression of
HNF1α mutant P291fsinsC in HeLa cells has been shown to inhibit the transactivation of the L pyruvate kinase promoter by wild type HNF1α (Yamagata et al. 1998). In addition, in vitro translated HNF1α mutant P291fsinsC and wild type HNF1α were demonstrated to form heterodimers (Yamagata et al. 1998). This implies that HNF1α mutant P291fsinsC would have a dominant negative effect on the transactivation of wild type HNF1α by heterodimerizing with the wild type protein. However, Vaxillaire et al. showed that mutant P291fsinsC does not inhibit the wild type HNF1α in a dominant negative manner in C33 human epithelial cervical carcinoma cells, which lack endogenous HNF1α (Vaxillaire et al. 1999).

These conflicting results prompted us to investigate this MODY3 mutation. In the current study, a truncated HNF1α, including amino acids from 1 to 280 (HNF1α (1-280)), was generated to mimic the P291fsinsC mutant. The cell line we used for transfection assays is β-TC3, a pancreatic β-cell and would more closely reflect the conditions of real diabetic patients. Various amounts of HNF1α (1-280) vector (cloned into CMV-pBat12 vector) and 50 ng of pGL3-RA reporter vector were co-transfected into β-TC3 cells. Luciferase activities were checked 48 hours after transfection. With more transfected HNF1α vector, lower reporter gene activities were observed, suggesting the inhibition by transfected HNF1α (1-280) could be caused by a dominant negative effect on the endogenous HNF1α (Figure 2.5).
2.4.5 HNF1α (1-280)_S142F doesn’t inhibit endogenous HNF1α activity.

Since mutant HNF1α (1-280) has an intact dimerization domain, it retains the ability to heterodimerize with wild type HNF1α. The inhibition of wild type activity by HNF1α (1-280) could result from this heterodimerization. However, we showed above that full-length HNF1α with the mutation S142F didn’t inhibit wild type HNF1α, although it also contained an intact dimerization domain. In order to investigate the role of heterodimerization, a point mutation S142F was genetically engineered into HNF1α (1-280) to form a new construct, HNF1α (1-280)_S142F. We co-transfected various amounts of HNF1α (1-280)_S142F and 50 ng of pGL3-RA into β-TC3 cells. Surprisingly, the dominant negative inhibition of endogenous wild type activity by HNF1α (1-280) was totally abolished by the S142F mutation in the DNA binding domain (Figure 2.6), although HNF1α (1-280)_S142F also has an intact dimerization domain and is supposed to be capable of heterodimerizing with endogenous HNF1α. Therefore, we speculated that the inhibition of wild type activity by HNF1α (1-280) is not caused by heterodimerization. Another mechanism must be involved in this inhibition in a dominant negative manner. According to Herskowitz (1987), there are two possible mechanisms for dominant negative inhibition: (1) a wild type protein heteromerizes with a mutant with improper function, and (2) the activity of a wild type protein is restricted by the availability of substrate. Based on the experiments above, the inhibition of wild type activity by HNF1α (1-280) observed in this study might not be due to the first mechanism, since no heterodimers are formed. Because the S142F
mutation in the DNA binding domain abolished this inhibition, out-competing wild type HNF1α for DNA binding by HNF1α (1-280) might be the main cause of this dominant negative inhibition.

2.4.6 Investigation the dominant negative mechanism by using ROP chimeras

According to the experiments described above, we speculate that the dominant negative inhibition of wild type activity by HNF1α (1-280) is not caused by heteromerization but by out-competing DNA binding of wild type HNF1α. There are two scenarios for this out-competing hypothesis: (1) HNF1α (1-280) is able to heterodimerize with endogenous wild type HNF1α and bind to DNA, and the heterodimers are inactive. With increasing amount of DNA occupied by heterodimers, wild type HNF1α homodimers are out-competed from DNA binding. By introducing the S142F mutation, heterodimers are prevented from binding to DNA, so the wild type HNF1α activity is restored. (2) It is also possible that no heterodimers are formed when HNF1α (1-280) is transfected into β-TC3 cells. There are only two populations formed in the cells: wild type and mutant homodimers. In this scenario, HNF1α (1-280) homodimers are capable of binding DNA and out-competing the endogenous HNF1α. The S142F mutation in the truncated HNF1α would prevent it from binding DNA, which can be utilized by endogenous wild type HNF1α again. This is supported by the observation that transfected HNF1α (1-280) with mutation S142F doesn’t inhibit endogenous HNF1α activity, which means HNF1α (1-280)_S142F doesn’t heterodimerize with
endogenous HNF1α.

To further investigate whether transfected HNF1α heterodimerizes with the endogenous HNF1α, we compared the activity of HNF1α (1-280) with a ROP-HNF1α fusion protein in which the dimerization domain of HNF1α was substituted with the ROP dimerization domain (Figure 2.7). The ROP protein consists of a dimer of a regular four-alpha-helix bundle, with each monomer comprised almost entirely of two alpha helices (Banner et al. 1987) (Figure 2.8). It was reported that ROP can functionally replace the dimerization domain of HNF1α to form a ROP-HNF1α homodimer, and no heterodimer is formed between a ROP-HNF1α monomer and a wild type HNF1α monomer (Tomei et al. 1992). While the ROP protein itself does not show any detectable DNA binding activity, the chimeric protein, ROP-HNF1α, bound to DNA as tightly as wild type HNF1α (Tomei et al. 1992). We reasoned that ROP-HNF1α dimers would be functional but incapable of dimerizing with wild type HNF1α, allowing us to distinguish the two scenario of the out-competing hypothesis.

We have demonstrated that full length ROP-HNF1α is active in a transient transfection assay (see below). We have not yet verified that ROP-HNF1α cannot dimerize with wild type HNF1α. This will be tested by EMSA using full length and HNF1α (1-280) or ROP-HNF1α (1-280) protein mixed together. When full length HNF1α is mixed with HNF1α (1-280), three different bands should appear, which are HNF1α full length homodimer, HNF1α (1-280) homodimer, and HNF1α full length/ HNF1α (1-280) heterodimer (Mendel et
al. 1991). Since ROP protein doesn’t bind the dimerization domain of HNF1α, only two bands [HNF1α homodimer and ROP-HNF1α (32-280) homodimer] are expected in the protein gel, when full length HNF1α is mixed with ROP-HNF1α (1-280). To demonstrate that ROP-HNF1 is not forming dimers with endogenous HNF1 in vivo, we will transfec
tROP-HNF1α (1-280) fusion vectors into β-TC3 cells. The HNF1α will be immunoprecipitated from nuclear extracts and EMSA will be performed.

2.4.7 Transfected ROP-HNF1α (32-628) fusion has high activity.

We transfected various amounts of ROP-HNF1α (32-628) fusion vectors, along with 50 ng of pGL3-RA luciferase reporter vector, into β-TC3 cells. With more transfected ROP-HNF1α (32-628) fusion vector, higher luciferase activity was observed (Figure 2.9). However, the transcriptional activity by transfected ROP-HNF1α (32-628) was much lower than by transfected wild type HNF1α (Figure 2.10). This result implies that the homodimer of ROP-HNF1α (32-628) has HNF1α transcriptional activity. The lower transactivation activity by transfected ROP-HNF1α (32-628) may be due to the fact that the homodimer of ROP-HNF1α (32-628) is unable to bind DCoH that can stabilize the homodimer of wild type HNF1α and stimulate HNF1α activity (Mendel et al. 1991).

2.4.8 Transfected ROP-HNF1α (32-280), not ROP-HNF1α (32-280)_S142F, inhibits endogenous HNF1α activity.
To further explore the two possibilities in the out-competing hypothesis, we co-transfected various amounts of ROP-HNF1α (32-280) or ROP-HNF1α (32-280)_S142F vectors, along with 50 ng of pGL3-RA reporter vectors into β-TC3 cells. With more transfected ROP-HNF1α (32-280), lower luciferase activity was obtained (Figure 2.11). Although the inhibition is in a dominant negative manner, it must be caused by competition with the wild type HNF1α since ROP-HNF1α (32-280) is unable to heterodimerize with wild type HNF1α. Similarly, this inhibition is completely abolished when the HNF1α (32-280)_S142F mutant was transfected into cells (Figure 2.11), suggesting the inhibition is not due to heterodimerization but out-competing wild type HNF1α for DNA binding by MODY3 mutants. On the other hand, these results also suggest that heterodimerization is not required for the out-competing inhibition. Together with the previous observations that endogenous wild type HNF1α activity is not inhibited by transfected mutant HNF1α_S142F or HNF1α (1-280)_S142F, we conclude that heterodimers (mutant/wild type HNF1α) are not being generated when HNF1α mutants are transfected into β-TC3 cells. To further verify this conclusion, we will perform EMSAs. Beta-TC3 cells will be transfected with HNF1α (1-280). HNF1α will be immunoprecipitated and DNA binding will be measured by EMSA. Formation of heterodimers will be tested by the presence of a shift composed of mixed HNF1α (1-280)/wild type HNF1α dimers.

2.5 DISCUSSION
Previously, at least two laboratories demonstrated a dominant negative effect for the inhibition of wild type HNF1α by MODY3 mutant P291fsinsC. In the current studies we show that no heterodimers are formed in β-TC3 cells when the HNF1α (1-280)-pBat12 vector is transfected, and the dominant negative inhibition is not due to heteromerization but out-competing wild type HNF1α for DNA binding. However, an interesting question has been raised. Why don’t transfected HNF1α mutants containing an intact dimerization domain heterodimerize with wild type HNF1α in β-TC3 cells? This question could be explained by the existence of DCoH, the small protein that binds and stabilizes HNF1α dimers to form a heterotetramer. It was shown that wild type HNF1α doesn’t bind a truncated HNF1α (1-428) mutant to form a heterodimer in the presence of DCoH, while HNF1α heterodimers are formed in the absence of DCoH (Figure 2.12) (Rose et al. 2004). So it is reasonable to imagine that the transfected HNF1α mutant may homodimerize locally, and then bind to DCoH to form a stable heterotetramer. In this scenario, once the HNF1α/DCoH heterotetramer is formed, it would prevent the dissociation of HNF1α monomers and exchanging with the competing mutant HNF1α (Figure 2.13). This hypothesis is consistent with the observation that MODY3 HNF1α mutants can heterodimerize with wild type HNF1α in cells lacking DCoH or in EMSA assays in the absence of DCoH (Yamagata et al. 1998).

To further investigate the role of DCoH, nuclear extraction from β-TC3 cells transfected with ROP-HNF1α (32-280) will be performed. EMSAs will be followed to
examine if the heterodimer of HNF1α/ROP-HNF1α (32-280) exists. If endogenous HNF1α doesn’t heterodimerize with transfected ROP-HNF1α (32-280), only two bands [full length HNF1α homodimer and ROP-HNF1α (32-280)] would be shown in the EMSA. The same nuclear extract sample also will be tested with a DCoH antibody to confirm the binding of DCoH to HNF1α homodimer, not to the ROP-HNF1α (32-280) homodimer.

To make sure all the transfected HNF1α mutants are expressed in the nucleus, we will perform immunolocalization studies to check the subcellular localization of HNF1α mutants. Briefly, β-TG3 cells grown on Falcon Culture slides will be transfected with 0.4 µg of different HNF1α mutants by using Fugene6 Reagent. Transfected cells will be fixed with methanol after 48 h and blocked with PBS containing 5% Normal Donkey Serum for 15 min. The primary antibody, anti-HNF1α, and the secondary antibody, anti-mouse Cy3, will be diluted in PBS + 5% Normal Donkey Serum and incubated for 2 h and 1 h, respectively. Fluorescence will be examined using a fluorescence microscope.

More than 120 MODY3 mutations have been identified located throughout the three functional domains of HNF1α. Interestingly, many mutations were reported located in the DNA binding domain. An unknown mechanism explains this phenomenon. The previous hypothesis proposed that all the MODY3 mutations that are located in the coding region outside of the dimerization domain would have a dominant negative effect on wild type HNF1α, while the MODY3 mutations in the dimerization domain or in the promoter of HNF1α would have a haploinsufficient effect. However, here we demonstrate that
heterodimerization does not occur in beta cells, whereas HNF1α transcriptional activity is inhibited in a dominant negative manner by out-competing wild type HNF1α for DNA binding by MODY3 mutants with intact DNA binding domains. The current study suggests that MODY3 mutations in the activation domain of HNF1α may out-compete wild type activity by binding DNA. MODY mutations in the DNA binding domain may interfere with wild type HNF1α activity by out-competing other essential interactions by HNF1α.

In real MODY3 diabetic situations, the amount of HNF1α mutants probably is limited to out-compete wild type HNF1α, since only one allele is affected in cells. However, one possibility that we are in the process of testing is that MODY3 mutations in the transcriptional domain could affect degradation of HNF1α mutants by the ubiquitin proteasome system and cause accumulation of HNF1α mutants in cells. An interesting connection between ubiquitin-proteasome-mediated proteolysis and transcriptional activation has been demonstrated in the studies of several transcription factors such as Fos, Jun, Myc, Myb, p53, HIF1α, and β-catenin (Treier et al. 1994; Chowdary et al. 1994; Salghetti et al. 1999; Salghetti et al. 2001). For these transcription factors, an overlap between the transcriptional activation domain and sequences targeting ubiquitin–proteasome-mediated degradation was found. It has been proposed that ubiquitination activates transcription, and the ubiquitinated transcription activator is degraded by the proteasome upon the completion of transcription. This intimate linkage between activation and degradation is completely based on the transcriptional activation domain. In the other words, mutations in the
transcriptional activation domain may affect the degradation of transcription factors. In 2004, Park et al. first reported that HNF1α, like other transcriptional factors, can be degraded by the ubiquitin proteasome system (Park et al. 2004). In their experiments, a multiple ubiquitinated HNF1α ladder pattern was shown in immunoprecipitation analysis with an anti-ubiquitin antibody, and the ubiquitinated HNF1α bands were accumulated after the treatment of MG132, a proteasome inhibitor. Taken all the evidences together, we are planning to test whether MODY3 mutations in the transcriptional domain of HNF1α could prevent its degradation from ubiquitin–proteasome- mediated proteolysis and cause its accumulation. Presumably, once HNF1α mutants are accumulated in cells, wild type HNF1α could be out-competed from DNA-binding site. To test this, Beta-TC3 cells will be transfected with wild type HNF1α or HNF1α (1-280) mutant, treated with cycloheximide to inhibit protein synthesis, and the turnover of HNF1α will be followed by Western blot analysis. If the transcriptional domain plays an important role in degradation of HNF1α, cells transfected with HNF1α (1-280) would have a longer half-life than wild type HNF1α. This difference should be reduced by treating cells with the MG132 proteasome inhibitor. This model would explain dominant negative mutations located in the transcriptional domain of HNF1α.
Table 2.1: Reporter controls illustrating the processing of data from transfection assay and BCA assay

<table>
<thead>
<tr>
<th>BCA Data</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
<th>Sample 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>50ng pGL3-RA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abs</td>
<td>0.783</td>
<td>0.755</td>
<td>0.783</td>
<td>0.813</td>
<td>0.807</td>
</tr>
<tr>
<td>Con. (mg/ml)</td>
<td>1.260869</td>
<td>1.209265</td>
<td>1.260869</td>
<td>1.316159</td>
<td>1.305101</td>
</tr>
<tr>
<td>mg in 10µl</td>
<td>0.0126</td>
<td>0.0121</td>
<td>0.0126</td>
<td>0.0132</td>
<td>0.0131</td>
</tr>
<tr>
<td><strong>50ng pGL3-DRA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abs</td>
<td>0.748</td>
<td>0.786</td>
<td>0.83</td>
<td>0.829</td>
<td>0.805</td>
</tr>
<tr>
<td>Con. (mg/ml)</td>
<td>1.196364</td>
<td>1.266398</td>
<td>1.34749</td>
<td>1.345647</td>
<td>1.301415</td>
</tr>
<tr>
<td>mg in 10µl</td>
<td>0.012</td>
<td>0.0127</td>
<td>0.0135</td>
<td>0.0135</td>
<td>0.013</td>
</tr>
<tr>
<td><strong>Untransfected β-TC3 cells</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abs</td>
<td>0.752</td>
<td>0.781</td>
<td>0.757</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Con. (mg/ml)</td>
<td>1.203736</td>
<td>1.257183</td>
<td>1.212951</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mg in 10µl</td>
<td>0.012</td>
<td>0.0126</td>
<td>0.0121</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transfection Data</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
<th>Sample 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>50ng pGL3-RA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLU</td>
<td>73399</td>
<td>71501</td>
<td>68309</td>
<td>71004</td>
<td>71476</td>
</tr>
<tr>
<td>Adjusted RLU</td>
<td>72467</td>
<td>70569</td>
<td>67377</td>
<td>70072</td>
<td>70544</td>
</tr>
<tr>
<td>RLU/mg</td>
<td>5751349</td>
<td>5832149</td>
<td>5347831</td>
<td>5308485</td>
<td>5385038</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5524880</td>
</tr>
<tr>
<td>Std error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>110364.3</td>
</tr>
<tr>
<td><strong>50ng pGL3-DRA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLU</td>
<td>20209</td>
<td>18200</td>
<td>20658</td>
<td>19449</td>
<td>17908</td>
</tr>
<tr>
<td>Adjusted RLU</td>
<td>19277</td>
<td>17268</td>
<td>19726</td>
<td>18517</td>
<td>16976</td>
</tr>
<tr>
<td>RLU/mg</td>
<td>1606417</td>
<td>1359685</td>
<td>1461185</td>
<td>1371630</td>
<td>1305846</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1420953</td>
</tr>
<tr>
<td>Std error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>635488.6</td>
</tr>
<tr>
<td><strong>untransfected β-TC3 cell</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLU</td>
<td>938</td>
<td>1041</td>
<td>817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLU/mg</td>
<td>78166.67</td>
<td>82619.05</td>
<td>67520.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76102.13</td>
</tr>
<tr>
<td>Std error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4479.234</td>
</tr>
</tbody>
</table>
Table 2.1. Continued:

Calculation of HNF1α activity for Figure 2.2. Around $2.0 \times 10^4$ β-TC3 cells were plated per well in a 96-well plate. After 24 hours, 5 wells were transfected with 50 ng pGL3-RA vector; 5 wells were transfected with 50 ng pGL3-DRA vector; 3 wells remained untransfected. After 48 hours of transfection, cells were lysed in 25 µl lysis buffer. 10 µl lysates were subjected to luciferase activity analysis. 10 µl lysates were saved for the BCA assay. The total amount of protein in 10 µl lysates was calculated from the results of BCA assay, shown in upper panel. We used the average RLU of untransfected β-TC3 cells as baseline, which was subtracted from the RLU from cells transfected with 50ng pGL3-RA or 50ng pGL3-DRA. This adjusted luciferase count was divided by the total amount (mg) of protein from corresponding cell lysates. Then the average and standard error were calculated for the 50ng pGL3-RA and 50ng pGL3-DRA group.
Figure 2.1: The schematic representation of MODY3 mutants used for transfection assays. Three functional domains are labeled in different colors (dimerization domain: blue; DNA binding domain: red; activation domain: purple). The sequence number of the amino acids at the ends of each functional domain are indicated in the schematic representation of wild type HNF1α (on the top). MODY3 mutation S142F is shown in yellow color. The truncated HNF1α (1-280) is generated to mimic the MODY3 mutant P291fsinsC.
Figure 2.2: Activity of endogenous HNF1α. None (red) or 50 ng of the pGL3-RA (blue) or pGL3-DRA (purple) vectors were transfected into β-TC3 cells. Luciferase activities were measured by microplate luminometer. The data were expressed as mean values (± standard error) of five experiments. The final activities were reported in terms of RLU/mg total protein.
**Figure 2.3:** Activation of the pGL3-RA reporter by transfected HNF1α. 50, 100, 200, and 400 ng of the pBat12 vectors expressing full length HNF1α were transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vectors. The total amount of DNA for each transfection is 500 ng with different amount of empty pBat12 vector used for balance. Untransfected β-TC3 cells represent cells transfected with no DNA. Luciferase activities were counted by microplate luminometer. The data were expressed as mean values (± standard error) of three experiments. The final activities were reported in terms of RLU/mg total protein.
Figure 2.4: A mutation in the DNA binding domain abolishes HNF1α activity. (A) 50, 100, 200, and 400 ng of the pBat12 vectors expressing wild type HNF1α or full length HNF1α with MODY3 mutation S142F were transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vector. While the wild type HNF1α was active, the S142F mutant was not active. In addition, the S142F mutant did not interfere with the endogenous HNF1α activity. The total amount of DNA for each transfection is 500 ng with different amount of empty pBat12 vector used for balance. Untransfected β-TC3 cells represent cells transfected with no DNA. Luciferase activities were counted by microplate luminometer. The data were expressed as mean values (± standard error) of three experiments. The final activities were reported in terms of RLU/mg total protein. (B) Western blot analysis was performed from each transfected cell by using HNF1α antibody. Cell nuclear extracts were prepared 48 hours after transfection of DNA. Lane 1: untransfected cell (endogenous HNF1α); Lane 2: transfection of 50ng WT HNF1α; Lane 3: transfection of 100ng WT HNF1α; Lane 4: transfection of 200ng WT HNF1α; Lane 5: transfection of 400ng WT HNF1α; Lane 6: transfection of 50ng WT HNF1α with S142F; Lane 7: transfection of 100ng WT HNF1α with S142F; Lane 8: transfection of 200ng WT HNF1α with S142F; Lane 9: transfection of 400ng WT HNF1α with S142F.
**A**

![Bar graph showing RLU/mg for different concentrations of HNF1α and HNF1α(S142F).](image)

**B**

![Image of gel electrophoresis results](image)
**Figure 2.5:** HNF1α (1-280) interferes with endogenous HNF1α activity. 50, 100, 200, and 400 ng of the pBat12 vectors expressing HNF1α (1-280) were transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vector. The total amount of DNA for each transfection is 500 ng with different amount of empty pBat12 vector used for balance. Untransfected β-TC3 cells represent cells transfected with no DNA. Luciferase activities were counted by microplate luminometer. The data were expressed as mean values (± standard error) of three experiments. The final activities were reported in terms of RLU/mg total protein.
**Figure 2.6:** A mutation in the DNA binding domain of HNF1α (1-280) interferes with its dominant negative activity. 50, 100, 200, and 400 ng of the pBat12 vectors expressing HNF1α (1-280) or HNF1α (1-280) with MODY3 mutation S142F were transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vector. While HNF1α (1-280) acted as a dominant negative inhibitor of endogenous HNF1α, the S142F mutation in HNF1α (1-280) did not interfere with endogenous HNF1α activity. The total amount of DNA for each transfection is 500 ng with different amount of empty pBat12 vector used for balance. Untransfected β-TC3 cells represent cells transfected with no DNA. Luciferase activities were counted by microplate luminometer. The data were expressed as mean values (± standard error) of three experiments. The final activities were reported in terms of RLU/mg total protein. (B) Western blot analysis was performed from each transfected cell by using N-terminus of HNF1α antibody. Cell nuclear extracts were prepared 48 hours after transfection of DNA. *Lane 1:* purified HNF1α (1-280) protein; *Lane 2:* transfection of 50ng HNF1α (1-280); *Lane 3:* transfection of 100ng HNF1α (1-280); *Lane 4:* transfection of 200ng HNF1α (1-280); *Lane 5:* transfection of 400ng HNF1α (1-280); *Lane 6:* transfection of 50ng HNF1α (1-280) with S142F; *Lane 7:* transfection of 100ng HNF1α (1-280) with S142F; *Lane 8:* transfection of 200ng HNF1α (1-280) with S142F; *Lane 9:* transfection of 400ng HNF1α (1-280) with S142F.
Figure 2.7: The schematic representation of ROP chimera mutants used for transfection assays. ROP is shown in green. Three functional domains are labeled in different color (dimerization domain: blue; DNA binding domain: red; activation domain: purple). The numbers of the amino acids at the ends of each functional domain are indicated in the schematic representation of wild type HNF1α (on the top). MODY3 mutation S142F is shown in yellow color.
Figure 2.8: Structural comparison between ROP and the dimerization domain of HNF1α (monomer and dimer). ROP protein consists of a dimer of regular four-alpha-helix bundle, with each monomer comprised almost entirely of two alpha helices (Banner et al. 1987). ROP monomer is unable to heterodimerize with HNF1α monomer.
Figure 2.9: The Rop-HNF1α fusion protein is functionally active. 50, 100, 200, and 400 ng of the pBat12 vectors expressing HNF1α (32-628) were transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vector. The total amount of DNA for each transfection is 500 ng with different amount of empty pBat12 vector used for balance. Untransfected β-TC3 cells represent cells transfected with no DNA. Luciferase activities were counted by microplate luminometer. The data were expressed as mean values (± standard error) of three experiments. The final activities were reported in terms of RLU/mg total protein.
**Figure 2.10:** Rop-HNF1α is less active than wild type HNF1α. 50, 100, 200, and 400 ng of the pBat12 vectors expressing HNF1α (32-628) or HNF1α (1-628) were transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vector. The wild type HNF1α might show greater activity than Rop-HNF1α due to its ability to interact with the coactivator DCoH, which binds to the native dimerization domain (reference the DCoH/HNF1α dimerization domain structure). The total amount of DNA for each transfection is 500 ng with different amount of empty pBat12 vector used for balance. Untransfected β-TC3 cells represent cells transfected with no DNA. Luciferase activities were counted by microplate luminometer. The data were expressed as mean values (± standard error) of three experiments. The final activities were reported in terms of RLU/mg total protein.
Figure 2.11: A mutation in the DNA binding domain of Rop-HNF1α (32-280) interferes with its dominant negative activity. 50, 100, 200, and 400 ng of the pBat12 vectors expressing HNF1α (32-280) or HNF1α (32-280) with MODY3 mutation S142F were transfected into β-TC3 cells along with 50 ng pGL3-RA reporter vector. Like HNF1α (1-280), Rop-HNF1α (32-280) acted as a dominant negative of wild type HNF1α activity. The S142F mutation in the DNA binding domain of Rop-HNF1α (32-280) reversed the dominant negative activity. The total amount of DNA for each transfection is 500 ng with different amount of empty pBat12 vector used for balance. Untransfected β-TC3 cells represent cells transfected with no DNA. Luciferase activities were counted by microplate luminometer. The data were expressed as mean values (± standard error) of three experiments. The final activities were reported in terms of RLU/mg total protein.
**Figure 2.12:** DCoH restricts the formation of HNF1α heterodimers. Left panel is an EMSA published by Rose et al. (2004). HNF1α dimers formed in CHO cells were challenged by addition of purified, truncated HNF1α (HNF1α tr). In the absence of DCoH proteins, HNF1α dimers exchanged readily with the challenging HNF1αtr, resulting in a band of intermediate mobility (lane 2). When plasmids containing HNF1α and DCoH were co-transfected into CHO cells, heterodimerization was inhibited. The right panel is diagrammatic representation to illustrate that no heterodimer is formed in the presence of DCoH.
**Figure 2.13:** Model showing local dimerization of HNF1α. Transfected HNF1α mutant may homodimerize locally, and then bind to DCoH to form a stable heterotetramer. Once the heterotetramer is formed, it would prevent the dissociation of HNF1α monomers and exchange with competing mutant HNF1α.
REFERENCES:


vaccines containing primate CpG motifs and fowlpoxvirus vaccines co-expressing IFN\(\gamma\) or IL-12. Vaccine 23, 188-197.


monooxygenase isoforms with particular reference to the CYP2C subfamily. Drug Metab Dispos 32, 333-339.


CHAPTER 3

STRUCTURE OF THE CONSERVED TRANSCRIPTIONAL REPRESSOR ERH
(ENHANCER OF RUDIMENTARY HOMOLOG)

Cheng Wan,§ Wolfram Tempel,|| Zhi-Jie Liu,|| Bi-Cheng Wang|| and Robert B. Rose§*

§ Department of Molecular & Structural Biochemistry, North Carolina State University, Raleigh, NC 27695-7622, USA.

|| Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, 30602-7229, USA.

This chapter has been accepted for publication in Biochemistry, 2005 Apr 5;44 (13):5017-23

Coordinates and structure factors for erh were deposited in the Worldwide Protein Data Bank under entry code 1W9G.
**ABBREVIATIONS:** erh, enhancer of rudimentary homolog; CE, Combinatorial Extension; CKII, Casein Kinase II; DCoH, Dimerization Cofactor of HNF1α; HNF1α, Hepatocyte Nuclear Factor-1, DROER, Drosophila enhancer of rudimentary; LB, Luria-Bertani media; pdb, protein data bank; SAD, SDS-PAGE, Sodium dodecyl sulphate- Polyacrylamide Gel Electrophoresis; Single wavelength Anomalous Dispersion; SERCAT, SouthEast Regional Collaborative Access Team.
3.1 ABSTRACT:

Erh (Enhancer of Rudimentary Homolog) is a ubiquitously expressed transcriptional co-regulator that is highly conserved among eukaryotes, from humans to plants to protozoa. Functions attributed to erh include enhancement of pyrimidine biosynthesis, a role in cell cycle regulation, and repression of the cell-specific transcription factor HNF1α (Hepatocyte Nuclear Factor-1) through binding the coactivator DCoH (Dimerization Cofactor of HNF1α). No homologous sequences have been identified and little is known about the interactions of erh. In order to further elucidate its function, we determined the crystal structure of erh to 2.0 Å resolution. The erh structure is a novel α +β fold consisting of a four-stranded anti-parallel beta sheet with three amphipathic alpha helices situated on one face of the beta sheet. Structure-based searches of the Protein Data Bank, like sequence based searches, failed to identify evolutionarily-related homologs. We present structural and biochemical evidence that erh functions as a dimer. The dimer interface consists of an 8-stranded anti-parallel beta barrel. Many of the surface residues of erh are conserved, including patches of hydrophobic and charged residues, suggesting protein-protein interaction interfaces. Two putative CKII phosphorylation sites are highly ordered in the structure, and are predicted to disrupt dimerization and protein-protein interactions.
3.2 INTRODUCTION

Erh (Enhancer of rudimentary homolog) is a highly conserved 104-amino acid protein: the frog and human proteins are identical in sequence and differ from zebrafish erh by a single amino acid (Pogge von Strandmann et al., 2001; Gelsthorpe et al., 1997). Expression profile studies indicate that erh is highly expressed in rapidly dividing cells such as carcinoma cell lines, is enriched in ectodermal-derived tissue during development, and is ubiquitously expressed in the adult (Pogge von Strandmann et al., 2001; Isomura et al., 1996; Su et al., 2004; Wojcik et al., 1994). Despite its pervasive presence, little is known of the function of this protein. Erh localizes in the cytoplasm and nucleus (Pogge von Strandmann et al., 2001). It has been characterized as an enhancer and repressor of transcription in a number of different contexts. Erh was initially identified in a genetic screen in Drosophila as a mutation that augmented the phenotype of rudimentary gene mutations, observed as wing truncations (Wojcik et al., 1994). The protein was proposed to be an enhancer of the rudimentary gene, which encodes the first three enzymatic activities of the pyrimidine biosynthetic pathway. Due to the importance of rudimentary during development, it was suggested that the Drosophila enhancer of rudimentary (DROER) protein played a developmental role. A later study demonstrated that erh is a maternal factor in Xenopus that is expressed developmentally in ectodermally-derived tissues such as the eyes, spinal cord and parts of the brain, supporting the development role of erh (Pogge von Strandmann et al.,
This study demonstrated that erh could function as a repressor in a cell-type specific manner when fused to the GAL4 DNA binding domain. Erh was further proposed to regulate the cell cycle, being expressed at higher levels in proliferating cells than non-dividing cells (Gelsthorepe et al., 1997). Erh is phosphorylated by CKII (Casein Kinase II), a cell cycle regulated kinase (Gelsthorepe et al., 1997).

By now, several interaction partners of erh were identified by different laboratoies. The first reported interaction was identified by a yeast two-hybrid screen of the coactivator DCoH (Dimerization Cofactor of HNF1α) (Pogge von Strandmann et al., 2001). In a transient transfection assay, erh appeared to repress the coactivator function of DCoH, preventing it from enhancing the activity of the transcription factor HNF1α (Hepatocyte Nuclear Factor-1). Other identified erh’s interaction partners include FCP1, SPT5, mediator of RNAPII transcription subunit 31 (Med31), and Transducinlike enhancer 1 (TLE1). FCP1 is a phosphatase specific for the carboxyl-terminal domain of the large subunit of the RNA polymerase II (RNAPII) (Amente et al. 2005) that stimulates transcriptional elongation and is required for general transcription and cell viability. Another reported interaction partner of erh is SPT5, a transcription elongation factor that can regulate transcriptional elongation by modulating the activity of RNAPII. Interestingly, both FCP1 and SPT5 are known to interact with, and be methylated by a common partner, PRMT533 (Kwak et al. 2003). Although the consequence of this methylation of FCP1 is unclear, PRMT5 methylation of SPT5 regulates its promoter-binding ability, association with RNAPII, and elongation activity (Kwak et al.
Furthermore, two proteins, Med31 and TLE1, are also identified to be interaction partners of erh in high throughput two-hybrid. TLE1 is transcriptional repressor and also interact with HNF1α (Kwak et al. 2003).

Since multiple interaction partners of erh have been revealed, it is very possible that erh has another unknown role except repressing the enhancement of HNF1α by DCoH. A small interaction network of erh protein based on published reports was summarized by Jin et al. (2007) (Figure 3.1). Through its protein-interaction network, erh has multiple connections with the core apparatus governing RNA transcription, indicating that it is likely to be an important transcription factor.

As noted in the first report of DROER, erh is highly conserved with no close sequence homologs (Wojcik et al., 1994). The lack of sequence similarity has hindered structure predictions. We therefore determined the structure of erh as a step in understanding its function. Although erh is a small protein, the structure indicates that it folds into a unique dimer with no known structural homologs.

3.3 MATERIALS AND METHODS

3.3.1 Sample Preparation

cDNA of human enhancer of rudimentary homolog was obtained from the Mammalian Gene Collection (MGC) through the Lawrence Livermore National Laboratory (http://image.llnl.gov), IMAGE number 3507241. The cDNA was subcloned for
over-expression into pET24b (Novagen) designed as a thrombin-cleavable 6-His-N-terminal fusion. *Escherichia coli* BL21(DE3) (Novagen) transfected cells were grown in LB (Luria-Bertani) media, induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside and harvested after 6 hours. Cell pellets were lysed in 50mM NaH$_2$PO$_4$, 300mM NaCl, 10 mM imidazole, bound to Ni-NTA resin (Ni$^{2+}$ chelated nitrilotriacetate resin) (Amersham Pharmacia), and eluted with 250mM imidazole, 300 NaCl, 1mM β-Mercaptoethanol. The His-tag was removed with thrombin (Sigma) according to the manufacturer's protocol. The protein was further purified by size exclusion chromatography with a Sephacryl S100 column (Amersham Pharmacia) and concentrated to 11 mg/ml using a centrifugal filter device (Millipore).

### 3.3.2 Structure determination

Initial crystallization conditions were identified using a sparse matrix screen (Hampton Research) (Jancarik *et al.* 1991). The final optimized conditions were: 10% isopropanol, 20% PEG4000, 0.1M HEPES (pH 7.5). Prism-shaped crystals grew after 48 hours by the hanging drop vapor diffusion method to about 0.3 mm in the longest dimension. Crystals were frozen in liquid nitrogen after a quick soak in cryoprotectant composed of 5% isopropanol, 20% PEG4000, 5% PEG400, 0.1M HEPES (pH 7.5). Selenomethionine-incorporated protein was generated for phase determination by the Single Anomalous Dispersion (SAD) method. The erh-containing plasmid was expressed in the
methionine auxotroph B834 (DE3) (Novagen) in the presence of selenomethionine (Fisher).
Mass spectrometry indicated 100% selenomethionine substitution of the four methionine
sites of erh.

Data were collected at the Southeast Regional Collaborative Access Team
(SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory
at 0.9793 Å (near the selenium anomalous peak) (Table 3.1). The data were processed
using the ELVES interface (Holton et al., 2004). The data were indexed with MOSFLM
(Leslie et al. 1986), and scaled and merged with SCALA keeping Bijvoet reflections separate
(Evans et al. 1993). F's were calculated with Truncate (French et al. 1978).

The initial phases were determined by the single wavelength anomalous scattering
(SAD) method (Wang et al. 1985). Using the SAD mode of the program SOLVE
(Terwilliger et al. 1999), and diffraction data to 2.8Å resolution, 8 (out of 8 expected)
putative selenium atoms were located in the asymmetric unit (Table 3.1). Initial phases
were improved by density modification with RESOLVE (Terwilliger et al. 2002), setting the
solvent content to 52%, consistent with the presence of a dimer in the asymmetric unit.
Non-crystallographic symmetry operators were obtained from the partial model output by
RESOLVE. Phases were then extended to 2.0Å using the program DM (Cowtan et al.
1999) in a combination of averaging and solvent flattening modes. Further automated
model building with the program ARP/wARP (Perrakis et al. 1999) was followed by manual
model adjustments using XFIT (McRee et al. 1999). Final rounds of refinement were
accomplished using CNS (Brunger et al. 1998) and refmac5 (Potterton et al. 2003; Murshudov et al. 1999) alternated with manual model building using the program O (Jones et al. 1991). CNS was also used to calculate buried surface area.

Erh orthologs were identified with PSI BLAST searching with the human erh sequence (Altschul et al. 1997). Sequences representing 15 distinct eukaryotic genera were identified and aligned with CLUSTAL W (Thompson et al. 1994): Aedes aegypti, Anopheles gambiae, Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Dictyostelium discoideum, Drosophila melanogaster, Echinococcus multilocularis, Homo sapiens, Mus musculus, Oryza sativa, Plasmodium falciparum, Rattus norvegicus, Soybean, Xenopus laevis.

3.3.3 Chemical cross-linking and gel filtration

For glutaraldehyde cross-linking, a 0.04 mg/ml solution of purified erh was incubated with varying concentrations of glutaraldehyde in PBS (phosphate buffered saline) on ice. After one hour, the solution was quenched with the addition of a final concentration of 100 mM glycine at pH 8.0. The mixture was kept on ice for 30 min and analyzed with a 15% SDS-PAGE gel. For gel filtration, the retention time of erh was measured with a Hi-prep 26/60 Sephacryl S100 column (Pharmacia) in PBS and compared with gel filtration standards (Biorad).
3.4 RESULTS AND DISCUSSION

3.4.1 Structure of erh

We report the high-resolution crystal structure of human erh. Although the data could be indexed in the space group C222, the high Rsym values indicated a lower symmetry space group, and the data were processed in space group P2₁. With two monomers in the asymmetric unit, the resulting Matthews coefficient (Vm) was 2.5 Å³/Da and 51% solvent content (Matthews, 1968). The final refined model gave R/R_free values of 0.211/0.259 for all data to 2.0 Å (Table 3.1, Figure 3.2A). Non-crystallographic symmetry restraints were not used during refinement because they increased the R-factor and did not improve the quality of the density. The root mean square deviation between Cα atoms of the two monomers in the final model is 0.26 Å. The final model includes residues 1-102 for each monomer and 104 water molecules. Residues 103 and 104 of each monomer are disordered, as are three residues of the loop between helix 1 and helix 2 of one monomer (Pro48B, Ser49B, Ile50B. B corresponds to the monomer with segid B in the pdb file). This loop is the most mobile region of the protein, with an average B-factor of 46.7 Å² for the Cα atoms (residues 45A-52A, monomer with segid A), compared with an average of 27.5 Å² for the Cα atoms of the overall protein dimer. The side chains of residues Arg42A and B, Asn46B, Ser49A, Ile50A and T51B in the mobile loop, and Lys12A and B, Glu23B, Lys34B, Lys41A and B, Asp75B, Gln77A and B, Gln100A and B, Gln101A and B were disordered and omitted from
the final model.

The erh structure is a novel $\alpha + \beta$ fold consisting of a four-stranded anti-parallel beta sheet with 3 amphipathic alpha helices situated on one face of the beta sheet (Figure 3.2B) (Murzin et al. 1995). The topology of the sheet is 2-1-3-4 (Figure 3.2C) with two alpha helices situated between beta strands 2 and 3. The $\alpha_1$ helix packs against strands 1 and 2 with its axis nearly parallel to the strands, reminiscent of a zinc finger. The axes of the other two helices are situated perpendicular to the beta sheet strands. The final alpha helix is at the C-terminus. The monomer is pseudo-symmetric, with $\beta_1$-$\beta_2$ superimposing on $\beta_3$-$\beta_4$ with an rms deviation of 1.5 Å for the backbone atoms, and with one helix ($\alpha_1$ and $\alpha_3$, respectively) following each set of $\beta$ strands, though in different orientations (Figure 3.2B). There is minimal sequence identity between the two $\beta$-$\beta$-$\alpha$ motifs.

Both faces of the beta sheet contain hydrophobic residues, suggesting the outer beta sheet surface is involved in protein-protein interactions (see Figure 3.5). Erh is a dimer in solution, as indicated by gluteraldehyde cross-linking and size exclusion chromatography (Figure 3.3). The asymmetric unit of the crystal contains the most obvious erh dimer (Figure 3.2D). The dimer interface consists of an 8-stranded anti-parallel beta barrel, with 4 strands contributed from each monomer. The overall topology of the barrel is: 2-1-3-4-4'3'-1'-2' (primed and unprimed strands designate different monomers). The hydrogen bonding network between adjacent dimer-related anti-parallel strands is not extensive: between strand 4 of each monomer, two main-chain hydrogen bonds are formed
(between the carbonyl and nitrogen atoms of Tyr 79 of each monomer). There are no main-chain-main-chain interactions between strands 2 and 2'. Instead the carbonyl oxygen of Arg17 is hydrogen bonded to the side-chain hydroxyl of Thr18', the carbonyl oxygen of Tyr19 may hydrogen bond with the Ne of Arg17', and the carbonyl oxygen of Leu7 interacts with the hydroxyl group of Tyr19' through an ordered water molecule. The primary interactions between monomers result from van der Waals contacts between the hydrophobic side chains of residues Ile5, Leu7, Leu70, and Pro81 and the hydrophobic regions of Arg17 and Tyr19. 1,375 Å² of surface area is buried between monomers in the dimer, less buried surface than expected for a typical homodimer, which has been estimated to be 1,685 Å² (Jones and Thornton 1995). In addition, there is a salt bridge interaction between Arg17 and Asp21', and a hydrogen bond between the side chain hydroxyls of Thr78 and Thr78'.

The overall fold of erh is not represented in the Protein Data Bank (Berman et al. 2000). Sequence analysis by multiple prediction methods, as implemented by the Polish Bioinformatics Web Site Meta-Server (url: http://bioinfo.pl/ (Ginalske et al. 2003), failed to identify close homologs. These included 1d sequence recognition, for example with Blast (Altschul et al. 1997), and 3d fold recognition, for example with 3D-PSSM (kelley et al. 2000), GenTHREADER (Jones 1999; McGuffin et al. 2003) and Pmodeller (Wallner et al. 2003). Searches of known structures with the DALI (http://www.ebi.ac.uk/dali/ (Holm and Sander 1995) and VAST (http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml (Gibrat et al. 1996)) servers did not identify significant structural homologs. We also examined the
The highest scoring alignments identified through searches with the erh structure were large structures that aligned to pieces of erh: VAST, identified photosystem I (pdbid 1jbo) overlapping 40 residues (score 5.7, P-value 0.03); DALI, identified phosphatase 2c (pdbid 1a6q) overlapping 57 amino acids (Z score 3.7). CE identified the α-helical protein retinoblastoma tumor suppressor domain A (pdbid 1A06) and the α+β protein DNA primase/helicase chain A (pdbid 1CR4), overlapping 71-73 residues (Z-score 3.9 with an rmsd of 4.8-4.9 Å). The topology of erh is similar to PsaD (DALI, Z score 3.4, rmsd of 3.5 Å for 58 amino acids), a 138 amino acid protein which is a stromal subdomain of cyanobacterial photosystem I (pdbid 1jb0 chain D). A number of proteins identified through DALI and Meta-Server overlapped the beta-beta-helix motif, with the helix axis oriented parallel to the beta strands. No homologs of erh were identified with the two helices perpendicular to the beta strands. Searching with the monomer or dimer of erh resulted in the same set of structures. Searching with the 8-stranded dimer interface of erh using DALI identified the 7-stranded beta barrel in the ku70 monomer as the closest match (pdbid 1jeq, Z score 2.3). Although the topologies differ and the ku70 strands form a more continuous barrel, the two beta barrels overlap with an rmsd of 3.9 Å for 56 Cα residues.
3.4.2 Evolutionary conservation

Erh is highly conserved among eukaryotes, as has been noted previously (Figure 3.4) (Gelsthorpe et al. 1997). A PSI BLAST search starting with the human erh sequence converged after two cycles and identified erh sequences from 15 genera with an expectation value of $8 \times 10^{-7}$ or less (Aotschul et al. 1997). Figure 3.5A indicates the distribution of the conserved residues on the erh structure. Conserved hydrophobic residues are found in three regions of the protein: core residues within erh monomers, between helices and between the helices and beta sheet; at the dimer interface, between beta sheets of each monomer; and surface residues. Many of the surface residues are conserved within the animal kingdom, from tape worms to humans (Figure 3.5B). A number of conserved hydrophobic residues are solvent-exposed, for example in the cavity generated between the beta-barrel dimer interface and on the surface of the parallel helices 2 and 3. These parallel helices may represent a site for interactions with other proteins.

3.4.3 Phosphorylation sites

The cell-cycle regulator Casein Kinase II (CKII) was shown to phosphorylate Drosophila erh in vitro at two consensus sequences, S/T X X D/E: at Thr18 and Ser24 (Gelsthorpe et al. 1997). Figure 3.5A indicates that Thr18 is located at the beginning of beta strand 2, while Ser24 is located at the beginning of helix 1. Both sites are highly conserved among all erh sequences and are solvent exposed. The acidic residue of the
CKII consensus sequence for Thr18 (Asp/Glu21) is similarly conserved, while the acidic residue for the Ser24 consensus sequence (Glu27) is conserved in the animal kingdom but not in *Dictyostelium*, *Plasmodium*, or plants. Gelsthorpe, et al proposed that phosphorylation at these sites altered the secondary structure of erh (Gelsthorpe *et al*. 1997). Both putative phosphorylation sites are highly ordered in the structure. As described above, the gamma hydroxyl of Thr18 forms a hydrogen bond across the dimer interface with the carbonyl oxygen of Arg17'. The hydroxyl substituents of the dimer-related Thr18 residues are 6 Å apart. Therefore, phosphorylation at this site may disrupt dimerization. Ser24 caps the amino terminus of helix 1 by hydrogen bonding with the backbone nitrogen of Glu27. CKII has been shown to phosphorylate other Ser residues in the N-cap position (Rose *et al*, 2000). Phosphorylation of N-cap sites stabilize the alpha helix (Jones and Thornton 1995; Brownlie *et al*. 1997).

The surface charge potential of erh is not evenly distributed, with positively charged, neutral and negatively charged bands that may influence binding interactions (Figure 3.6). Overall, erh is negatively charged at physiological pH, with a predicted pI of 5.6. While Ser18 is located in a negative region of the surface (Figure 3.6C), Ser24 is situated in a region that is neutral or positively charged. Therefore, phosphorylation of Ser24 might disrupt charge-charge interactions associated with protein binding.

A number of other sites are predicted to be good phosphorylation sites by the NetPhos server, a neural-network-based prediction approach (http://us.expasy.org/tools/. (Andrew *et
al. 2002). Ser47, Thr11, Tyr19, Tyr22 and Tyr92 score well above the threshold (threshold of 0.4, scores of 0.74 or better). Tyr22 and Tyr92 are among five conserved Tyr residues of the animal erh sequences. Human erh contains 8 Tyr residues, all of which are solvent-exposed.

3.4.4 No detectable interaction between DCoH and erh

DCoH is cofactor of HNF1α and able to stabilize the homodimer of HNF1α, and erh was reported to be interaction partner of DCoH and regulate the enhancement of HNF1α by DCoH. To see whether they can form stable complex in vitro, we purified erh and DCoH individually in a high salt buffer. The erh and DCoH were mixed and dialyzed against a low salt buffer. However, no hetero-protein complex was detected after dialysis, indicating the interaction between erh and DCoH does not form stable complex in vitro. We also constructed plasmids pRSF–Histag-erh and pET24b-DCoH and tried to co-express N-terminal His-tagged erh and untagged DCoH in BL21 (DE3) E.coli cells. The purification of possible ERH-DCoH complex from E.coli lysate was carried out with Ni2+ chelating and sizing column. Again, the results indicated that there was no detectable hetero-complex in solution. Finally, to eliminate the interference from His-tag in the possible interaction between erh and DCoH, we are co-express untagged erh and DCoH in BL21 (DE3) E.coli cells, followed by ion-exchange chromatography and sizing column (Figure 3.7). No hetero-complex was detected once again.
3.5 CONCLUSIONS

The structure of erh reveals an as-yet uncharacterized fold, despite the small size of the protein. Structure-based searches of the protein data bank, like previously reported sequence-based searches, failed to identify evolutionarily related proteins that might provide clues to the function of erh.

The structure of the erh monomer consists of a four-stranded anti-parallel beta sheet with three helices on one surface. We present structural and biochemical evidence that erh is a functional dimer. The dimer interface resembles a beta-barrel with a hydrophobic core. The structure highlights likely binding surfaces, including a set of parallel helices with conserved solvent-exposed hydrophobic residues, and a hydrophobic cavity situated at the dimer interface. The charge distribution on the surface of erh is clustered into positive and negative bands. Much of the surface of erh is conserved.

The binding studies showed that no direct interaction between erh and DCoH was detected. It is possible that the affinity between the two proteins was much lower than erh and DCoH self-interactions. It is also possible that erh does not interact with DCoH directly, but their association is bridged by other associated factors not yet identified.
ACKNOWLEDGMENTS

X-ray data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID (or 22-BM) beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. We thank Dr. Michael Goshe for his assistance with mass spectrometry.
Table 3.1: Data collection, phasing and refinement statistics

<table>
<thead>
<tr>
<th>Space group</th>
<th>P2₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit cell lengths (Å)</td>
<td>a = 45.36, b = 62.50, c = 48.45, β = 117.92</td>
</tr>
</tbody>
</table>

**Data collection**

<table>
<thead>
<tr>
<th>Resolution range (Å)</th>
<th>42.9 - 2.84</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest resolution bin (Å)</td>
<td>2.11 - 2.0</td>
</tr>
<tr>
<td>Observations²</td>
<td>33252</td>
</tr>
<tr>
<td>Unique reflections²</td>
<td>5616</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.3 (98.0)</td>
</tr>
<tr>
<td>Rsym(I)¹,² (%)</td>
<td>7.3 (13.1)</td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>22.4</td>
</tr>
<tr>
<td>Ranom²,³ (%)</td>
<td>4</td>
</tr>
<tr>
<td>&lt;I&gt;/&lt;σ(I)&gt;²</td>
<td>13.2 (3.0)</td>
</tr>
<tr>
<td>FOM²,⁴,⁵</td>
<td>0.34</td>
</tr>
<tr>
<td>Anomalous Signal/Noise⁵</td>
<td>0.94</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<table>
<thead>
<tr>
<th>Resolution range (Å)</th>
<th>43.03 - 2.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein atoms in asu</td>
<td>1595</td>
</tr>
<tr>
<td>Water molecules</td>
<td>104</td>
</tr>
<tr>
<td>No. of reflections for Rfree⁶</td>
<td>765</td>
</tr>
<tr>
<td>R-factor/Rfree (%)</td>
<td>21.1/25.9</td>
</tr>
<tr>
<td>average B-factor (Å²)</td>
<td>28.4</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
</tr>
<tr>
<td>bonds (Å)</td>
<td>0.02</td>
</tr>
<tr>
<td>angles (deg)</td>
<td>1.28</td>
</tr>
</tbody>
</table>

¹ Rsym = Σ|I - <I>|/Σ<I>.

² Numbers in parantheses refer to data in the highest resolution shell.

³ Ranom = Σ<I⁺>- <I>/Σ<I>.

⁴ FOM (Figure of Merit)

⁵ From SOLVE

⁶ Based on 5% reflections
Figure 3.1: Recently identified interactions between mammalian erh and other proteins (This figure is originally published in *Proteins: Structure, Function, and Bioinformatics* (2007) 68: 427-437). Protein–protein interactions are shown by a connecting line with an arrowhead on each end. Interactions between enzyme-substrate pairs are indicated by an arrow pointing to the target.
**Figure 3.2:** Structure of erh. Panel A shows a stereodiagram of the final 2.0 Å-resolution 2F_o - F_c electron density map with a stick model for the hydrophobic residues (Ile5, Leu6, Leu7 from each monomer) at the dimer interface (wall-eyed stereo). Panel B shows a ribbon diagram of the erh monomer, colored by residue number from blue (N-terminus) to red (C-terminus). The arrow indicates the pseudo-two-fold axis, superimposing beta-strands 1-2 onto strands 3-4 (numbered). The attached helices (labeled a and b, respectively), are in different orientations. Panel C shows the topology of the monomer, as viewed from the beta sheet: β1-3_{10}-β2-α1-α2-β3-β4-α3. Panel D shows a ribbon diagram of the erh dimer. The monomers (cyan and black) are related by a two-fold axis (arrow) that is almost coincident with the crystallographic x-axis.
**Figure 3.3:** Erh is a dimer in solution. Panel A shows the results of glutaraldehyde cross-linking of erh. Purified erh was reacted with varying concentrations of glutaraldehyde as indicated and analyzed by SDS-PAGE. Molecular weight markers (Precision Plus, Biorad) are shown in the first lane (M), labeled in units of kilodaltons. The arrowheads indicate the monomer and dimer of erh. Panel B shows a chromatogram of erh after gel filtration. The chromatogram of the protein standards (Biorad) were superimposed on the chromatogram of erh for comparison. The molecular weights of the peaks from the standards are marked. The retention time of erh indicates it is a dimer in solution; the molecular weight of an erh monomer is 12kD.
Figure 3.4: Alignment of erh sequences. The erh sequences were identified from a PSI-BLAST search of the non-redundant data base, and aligned with ClustalW. The alignment was drawn with Textshade (Blom et al. 1999) (http://workbench.sdsc.edu) with a similarity threshold of 50% using the default similarities (FYW, IVLM, RK, DE, GA, TS, NQ). The sequence is numbered according to human erh. The secondary structure is depicted below the sequence and was derived from the determined tertiary structure.
**Figure 3.5:** Erh is evolutionarily highly conserved in eukaryotes. Panel A shows a ribbon representation of the human erh dimer colored by conservation among 15 orthologs identified by PSI BLAST (Altschul et al. 1997), from red (least conserved) to magenta to blue (most conserved). One sequence from each identified genus were aligned by CLUSTALW (Thompson et al. 1994). The score of the most highly conserved similarity group at each position was used, according to the similarity groupings: DE, NQ, ND, EQ, KR, AS, YW, IVLM, and ST. Hydrophobic side chains are displayed of residues conserved in 13 or more sequences (hydrophobic residues: VLIFMWY). The sites phosphorylated by casein kinase II, Thr18 and Ser24, are represented by green van der Waals surfaces (The gap in one of the monomers is due to disorder of three residues in the loop between helices 1 and 2). Panel B shows a Conolly surface of erh colored by conservation among animal sequences: orange - hydrophobic, cyan - hydrophilic, grey - not conserved. Conservation is defined as similarity among at least 8 of the 9 animal sequences at each position. '1' and '2' mark hydrophobic cavities formed at the dimer interface (Pogge von Strandmann et al. 2001) or between the parallel helices (Gelsthorpe et al. 1997).
Figure 3.6: The electrostatic potential surface of the erh dimer. Three perpendicular views of the Conolly surface of erh are shown, colored according to electrostatic potential: blue is positive, red is negative, and white is uncharged. An orthogonal coordinate system in each panel indicates the relative orientation of the three views: all rotations are 90° around the y-axis. Panel A shows the top surface is primarily positively charged. Panel B shows the side surface is divided into three zones: positive, uncharged, and negative. The CKII phosphorylation sites at Ser24 (yellow) are situated on the edge of the positive region. Panel C shows the bottom surface is primarily negatively charged. The CKII phosphorylation sites at Thr18 (cyan) are located at the bottom of a deep negative pocket. The electrostatic potential was generated with the APBS plugin of PyMol (version 0.96). Because a number of charged residues were disordered in the X-ray structure, all side chains were added to the model of erh in favorable rotamers before calculating the electrostatic potential.
Figure 3.7: No detectable interaction between erh and DCoH. (Panel A) Untagged erh and DCoH are co-expressed in BL21 (DE3) cells. Proteins were purified by using ion-exchange chromatography. The four peaks were obtained, and then run in polyacrylamide protein gel. The peak of collection #3 (red circled) was shown to contain erh and DCoH. (Panel B) Collection#3 was further purified by sizing column. There were 3 protein peaks obtained. By running protein gel, it is proved that erh and DCoH were separately located in different peaks. Erh exists in collection #31, and DcoH in #28. No complex protein was detected.
REFERENCES:


mapping of a novel human cDNA homologous to DROER, the enhancer of the


for building protein models in electron density maps and the location of errors in these


annotation using structural profiles in the program 3D-PSSM. J Mol Biol 299, 499-520.

Protein Crystallogr. 18, 33-39.


33. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden,


