1999

CHARACTERIZATION OF A PUTATIVE TRANSCRIPTION FACTOR

Maowen Hu
University of Rhode Island

CHARACTERIZATION OF A PUTATIVE TRANSCRIPTION FACTOR

BY

MAOWEN HU

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

BIOMEDICAL SCIENCES

UNIVERSITY OF RHODE ISLAND

1999

MASTER OF SCIENCE THESIS

OF

MAOWEN HU

APPROVED:

Thesis Committee

Major Professor _____

_____

_____

_____

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

1999

ABSTRACT

Basic helix-loop-helix (bHLH) proteins belong to a large family of transcription factors that are known to play important roles in cell proliferation, differentiation and oncogenesis. These proteins are structurally featured by a bHLH motif, which is responsible for protein dimerization and sequence-specific DNA binding (e.g., E-box). Recently we isolated a cDNA from a human liver library by a gene trapping method. Based on the Kozak rule, this cDNA encodes a protein with 415 amino acids, which is hereafter designated as CCAF. The objective of this thesis is to establish the molecular mass of this protein and to test the hypothesis that CCAF is a transcriptional modulator involving the regulation of cell cycle events.

To establish the molecular mass, CCAF was in vitro translated with TNT reticulocyte lysate and analyzed by autoradiography. Addition of the CCAF cDNA to the reaction mixture yielded a single product with a molecular weight of 52 kDa. This mass is consistent with the estimated weight and suggests that the Kozak favorable sequence indeed harbors the codon for translation initiation. In order to determine whether CCAF undergoes posttranslational modifications, immunochemical experiments were performed. An antibody was raised against a peptide derived from CCAF and subjected to affinity chromatography. This antibody detected a 52-kDa protein in the CCAF cDNA transfected cells but not in the control cells. These results further support the notion that the functional CCAF is a 52-kDa protein and undergoes little post-translational modifications.

To determine the expression of CCAF in different cell growing states, DLD cells derived from colon carcinomas were seeded at different densities. Likewise, this antibody detected the 52-kDa protein in the cells plated at all densities. However, proliferating cells expressed higher levels (~3-5 folds) than the quiescent cells. These studies were further extended to human colon carcinomas. Both Northern and Western blotting analyses detected abundant expression of CCAF in the carcinomas but not in the nearby normal tissues. These findings suggest that CCAF involves the regulation of cell cycle events and contributes to the oncogenic pathogenesis.

To determine the activity of CCAF in transcription regulation, transient cotransfection experiments were conducted with an E-box reporter. CCAF alone caused little change on the reporter enzyme activity. However, CCAF antagonized by 30% the transactivation activity conferred by E47, a bHLH protein that is known to transactivate E-box reporter. The antagonism on E47-mediated transactivation activity and the differential expression relating to cell growing and oncogenic states support the hypothesis that CCAF is a transcriptional modulator that involves the regulation of cell cycle events and plays a role in oncogenic pathology.

ACKNOWLEDGEMENT

I would first like to express my sincere gratitude to my major professor Dr. Bingfang Yan for providing me with the guidance, support, resources and encouragement necessary for completion of my thesis. I would also like to express my sincere appreciation to my committee member, Dr. Jennifer Specker, Dr. Clinton Chichester and Dr. Joan Lausier for their time, their concern and their valuable suggestions.

I would sincerely thank Dr. Raymond Panzica, Dr. John Babson, Dr. Joseph Matoney, Dr. Bonsup Cho, Dr. Shaikh and Ms. Lynn Matoney for their concern and help in the completion of my thesis. Special thanks to Dr. Nancy Thomas for conducting northern blot, to Dr. Lan Liu for performing in vitro translation, and to Dr. Dongsheng Bu for scanning figures in the thesis. I would also like to thank all the members in Dr. Yan's research group and the following graduate students, Huifang Zhang, Weifeng Tang, He Zhang, Connie, Guangzu Wang, Brian Felice, Bo Liu for their friendship and readiness to help.

I dedicate this thesis to my family in China who always inspire and encourage me to be successful. I also dedicate this thesis to Lynn Matoney and her family who always support me and let me feel warm and peace like at home.

PREFACE

This thesis was written in accordance with the thesis format.

LIST OF TABLES PAGE

LIST OF FIGURES PAGE

LIST OF FIGURES PAGE

ABSTRACT

Basic helix-loop-helix (bHLH) proteins belong to a large family of transcription factors that are known to play important roles in cell proliferation, differentiation and oncogenesis. These proteins are structurally featured by a bHLH motif, which is responsible for protein dimerization and sequence-specific DNA binding (e.g., E-box). Recently we isolated a cDNA from a human liver library by a gene trapping method. Based on the Kozak rule, this cDNA encodes a protein with 415 amino acids, which is hereafter designated as CCAF. The objective of this thesis is to establish the molecular mass of this protein and to test the hypothesis that CCAF is a transcriptional modulator involving the regulation of cell cycle events.

To establish the molecular mass, CCAF was in vitro translated with TNT reticulocyte lysate and analyzed by autoradiography. Addition of the CCAF cDNA to the reaction mixture yielded a single product with a molecular weight of 52 kDa. This mass is consistent with the estimated weight and suggests that the Kozak favorable sequence indeed harbors the codon for translation initiation. In order to determine whether CCAF undergoes posttranslational modifications, immunochemical experiments were performed. An antibody was raised against a peptide derived from CCAF and subjected to affinity chromatography. This antibody detected a 52-kDa protein in the CCAF cDNA transfected cells but not in the control cells. These results further support the notion that the functional CCAF is a 52-kDa protein and undergoes little post-translational modifications.

To determine the expression of CCAF in different cell growing states, DLD cells derived from colon carcinomas were seeded at different densities. Likewise, this antibody detected the 52-kDa protein in the cells plated at all densities. However, proliferating cells expressed higher levels (~3-5 folds) than the quiescent cells. These studies were further extended to human colon carcinomas. Both Northern and Western blotting analyses detected abundant expression of CCAF in the carcinomas but not in the nearby normal tissues. These findings suggest that CCAF involves the regulation of cell cycle events and contributes to the oncogenic pathogenesis.

To determine the activity of CCAF in transcription regulation, transient cotransfection experiments were conducted with an E-box reporter. CCAF alone caused little change on the reporter enzyme activity. However, CCAF antagonized by 30% the transactivation activity conferred by E47, a bHLH protein that is known to transactivate E-box reporter. The antagonism on E47-mediated transactivation activity and the differential expression relating to cell growing and oncogenic states support the hypothesis that CCAF is a transcriptional modulator that involves the regulation of cell cycle events and plays a role in oncogenic pathology.

INTRODUCTION

Transcription factors are proteins that initiate and modulate transcription rate by interacting with specific DNA recognition sequences in the target genes. As shown in Fig. 1, these DNA-binding transcription factors are structurally classified into four major classes: Helix-turn-helix homeodomain (e.g. PBX1), C_2H_2 zinc finger (e.g., Sp1), Helix-loop-helix (e.g., c-*myc*) and Leucine zipper (e.g., *c-fos* and *c-jun*). Many other transcription regulators fall into none of these classes. Therefore, a better classification system needs to be developed.

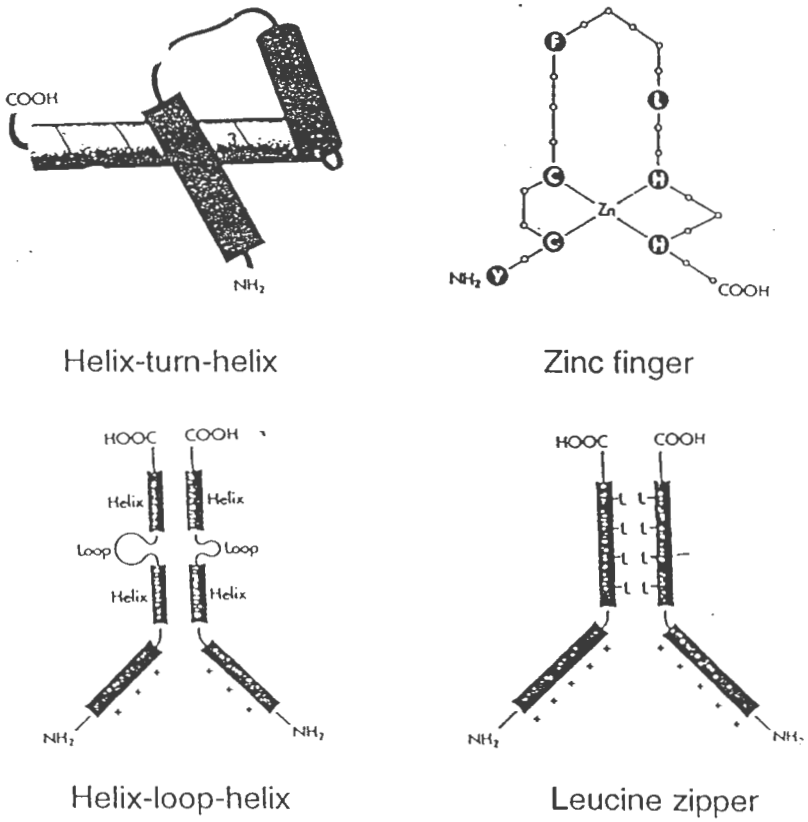


Figure 1. Classification of transcription factors

Basic helix-loop-helix (bHLH) proteins belong to a family of well-characterized transcription factors. They play important roles in the control of cell proliferation and differentiation that are involved in organ development and oncogenesis. For example, they function in myogenesis (MyoD/E47), neurogenesis (NeuroD, Achaete-scute/Daughterless), tumorgenesis (Myc/Max) and sex determination (E12/Da/ESC) (Little et al., 1998). In addition, these proteins are also involved in immunoglobulin gene regulation (TFEC/TFE3) (Zhao et al., 1993), phospholipid and pigment metabolism (Ino2/Delila) (Nikoloff et al., 1992, Goodrich et al., 1992), and xenobiotic response (AHR/ARNT) (Hirose et al., 1996).

I. Common Features of bHLH Proteins

1. Structural characteristics of bHLH proteins

bHLH proteins are distinguished by their bHLH domain, which was first identified in an immunoglobulin enhancer-binding polypeptide and several other proteins (e.g. Daughterless, MyoD and *myc*) (Murre et al., 1989). The bHLH domain is divided into two functional subdomains: a Helix-loop-helix (HLH) subdomain and an adjacent basic region. The HLH subdomain consists of two short amphipathic α-helices separated by a non-conserved loop with various lengths and primarily mediates the dimerization between HLH proteins (Voronova et al., 1990). The basic region consists of a cluster of 10-20 amino acids rich in lysine and arginine residues and is responsible for sequence-specific DNA binding (Burley, 1994).

In addition to the bHLH domain, some proteins contain other structures that are functionally important (Fig. 2). The *myc* oncoproteins contain a leucine zipper (LZ) motif that is responsible for dimerization (Penn et al., 1990). *Drosophila* hairy, E(spl) and the mammalian homologues (e.g., HES) contain an orange domain, a *Drosophila* C-terminal binding protein (dCtBP) motif (PLSLV) and a Groucho motif (WRPW). The orange domain mediates transcription repression (Dawson et al., 1995). PLSLV and WRPW motifs mediate the recruitment of transcription corepressors dCtBP and Groucho, respectively (Poortinga et al., 1998, Sewalt et al., 1999, Zhang et al., 1999). A group of the bHLH proteins (e.g., AHR and ARNT) contains a PAS domain that is responsible for dimerization between PAS proteins, xenobiotic binding and interaction with non-PAS proteins (Lindebro et al., 1995, Gradin et al., 1996). A conserved domain C in E proteins is also required for *in vivo* dimerization (Goldfarb et al., 1998).

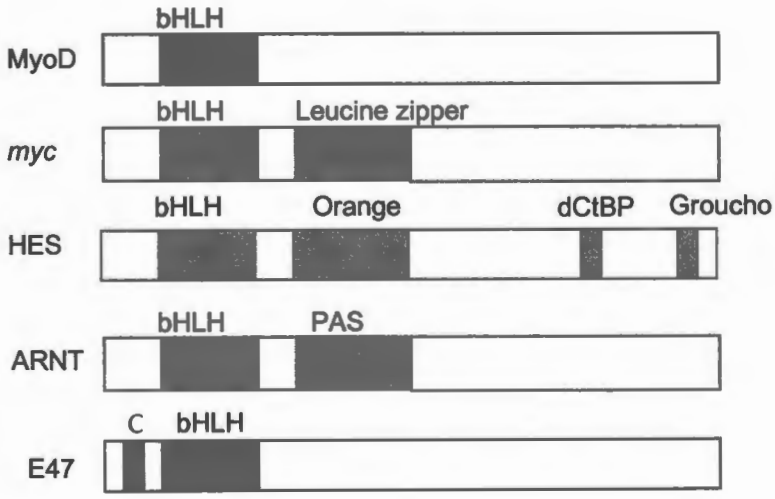


Figure 2. Functional domains of bHLH proteins

2. Classification of bHLH proteins

bHLH proteins are widely distributed in mammalian species, *Drosophila*, yeast and plants. More than 250 bHLH proteins have been identified in mammals and this number is increasingly expanding. Three systems based on different criteria are currently used for the classification of these proteins. Each of these will be discussed individually.

2.1 Based on tissue distribution and dimerization

In this classification system, bHLH proteins in mammals are grouped into two classes based on their expression pattern (Table 1). Class I, also known as E-protein family, consists of bHLH proteins encoded by E2A, HEB and E2-2 genes (Little et al., 1998). E2A encodes E12 and E47, which are produced by alternative splicing (Sun et al., 1991). Class I proteins are ubiquitously expressed and capable of forming transcriptionally active homo and/or heterodimers. Class II proteins display a tissue- or cell-restricted expression pattern. For example, myogenic bHLH proteins (e.g., MyoD) present specifically in muscle cell lineage while neurogenic bHLH proteins (e.g., NeuroD) present mostly in neural system. Class II proteins can exist as homodimers, but the active form is predominantly the heterodimers formed with class I proteins (Murre et al., 1994, Little et al., 1998).

Table 1. Classification based on tissue distribution and dimerization

Class	Example	Tissue distribution	Homodimer	Heterodimer
I	E12, E47, HEB, E2A	Universal	Yes	Yes
II	MyoD and NeuroD	Restricted	No activity	Yes

2.2 Based on structural domains

In this classification system, bHLH proteins are grouped into three classes based on their structural domains (Table 2). Class I includes bHLH proteins (e.g., E47 and MyoD) that have a basic region adjacent to the N-terminus of the HLH domain. Class II includes bHLHZ proteins (e.g., *myc* and Max) that contain an additional LZ dimerization domain immediately C-terminal to the bHLH domain. Class III includes HLH proteins that lack the DNA-binding domain due to the loss of the basic region (e.g., Id). These proteins act as negative regulators by effectively forming inactive, non-DNA-binding heterodimers with other bHLH proteins. Although mammalian homologues of E(spl) such as HES have an intact basic region, they also belong to this class because there is a proline residue in their basic region, leading to a preference for an N-box instead of an E-box site. The binding to the N-box results in transcription repression.

Table 2. Classification based on structural features

Class	Family	Proteins
I, bHLH proteins	Ubiquitous	E12/E47, HEB
	Neurogenic	NueroD, MATH and Mash
	Myogenic	MyoD, Myogenin, Myf5, MRF4
	Haemato- poeitic	Ly1, Ly2, Tal1, Tal2, Hen1 and Hen2
	Others	Arnt1, Arnt2 and Ah receptor
II, bHLHZ proteins	Myc	c-Myc, N-Myc, L-Myc and v-Myc
	Max	Max
	Mad	Mad1, Mxi1, Mad3 and Mad4
	AP4	AP4
	USF	USF1 and USF2
	Ig enhancer binding	TFE3, TFEB, TFEC
III, HLH proteins	Mammalian homologues of (E(spl))	Hes1, Hes2, Hes3, Hes4 and Hes5
	Id	Id1, Id2, Id3, and Id4

2.3 *Based on evolutionary relationship*

This classification system was recommended recently based on the evolutionary relationship among bHLH proteins (Atchley et al., 1997). In this classification, the

phylogenetic analysis of amino acid sequences is used to describe the patterns of evolutionary change within the motif and define the evolutionary lineages. These evolutionary lineages are well-known functional groups of proteins that can be further arranged into five classes based on the DNA binding (E-box), the amino acid patterns in the basic region, and the presence or absence of a LZ (Table 3). The hypothesized ancestral amino acid sequence for the bHLH transcription family is given together with the ancestral sequences of the subclasses.

Table 3 Classification of bHLH proteins based on evolutionary relationship

Class	Protein	E-box	LZ	aa pattern in basic region
A	E12, MyoD, NeuroD	CAGCTG	No	5-8-13 (xRx)
B	Myc, ARNT, HES	CACGTG	Yes/No	5-8-13 (BxR)
C	sim	CACGTG	No	Not consistent
D	Id	No	No	No basic region
?	AP-4	CAGCTG	Yes	5-8-13 (xKx)

Note: aa, amino acid; 5-8-13, aa position; x, any aa; R, Arginine;

B, any basic aa; K, lysine.

3. Transcription regulation of bHLH proteins

3.1 Dimerization

bHLH proteins act as transcription modulators, and the dimerization is the prerequisite for their function. For some bHLH proteins such as the ubiquitously expressed E protein family (e.g., E12, E47), the active forms are either homo- or heterodimers. Other bHLH proteins such as the myogenic MyoD family can exist as homodimers, but the active dimers are the heterodimers. Consistent with this finding, the antisense inhibition on E12 expression blocks the muscle-specific gene expression induced by MyoD (French et al., 1991). bHLH proteins exhibit different biological functions by heterodimerizing with different partners. For example, heterodimerization of Da with members of the Achaete-Scute class leads to the formation of neuronal precursors (Cabrera et al., 1991). On the other hand, heterodimerization of Da with Atonal protein leads to the formation of different, nonoverlapping sense organs and photoreceptors (Jarman et al., 1994).

3.2 DNA binding

After dimerization, bHLH proteins usually bind to the *cis*-acting DNA elements present in the target genes, resulting in the change of the gene expression. These DNA elements contain core sequences CANNTG or CACNAG. CANNTG, commonly known as the E-box, was first identified in the immunoglobulin heavy-chain (IgH) intronic enhancer and has been found in a large number of pancreatic-, lymphoid-, and muscle-specific promoter and enhancer elements (Little et al., 1998). CACNAG, commonly referred to as N-box, is present in the

promoter of genes such as HES gene. Most bHLH proteins (e.g., MyoD, E12) bind as dimers to E-box (Hsu et al, 1994), while some bHLH proteins (e.g., Hairy and HES) prefer to binding to the N-box (Dawson et al., 1995). The binding preference is specified by the sequence in the basic region of bHLH proteins. Generally, the proline-containing basic region has a higher affinity toward the N-box, whereas the basic region without a proline preferentially recognizes the E-box. Some bHLH heterodimers recognize different core DNA sequences rather than E- or N-box. For example, the AHR-ARNT complex usually binds to the dioxin response element TNGCGTG (Bacsi et al., 1995).

3.3 Transcriptional regulation

Binding to specific DNA elements by bHLH proteins leads to transcriptional activation or repression. MyoD and its related myogenic bHLH proteins, for example, bind to the E-box and activate transcription of myogenic genes (Weintraub et al., 1991). *Drosophila* Hairy and its related protein E(spl) and HES, however, bind to the N-box and inhibit the transcription of neurogenic genes (Dawson et al., 1995). This transcription repression process is outlined in Fig. 3. The N-box binding of transcription repressive proteins such as Hairy results in an orange domain-mediated repression on the E-box binding proteins such as Da/Scute. This binding also leads to a recruitment of Groucho, which is a transcription corepressor. Other than the N-box binding, *Drosophila* Hairy and its related proteins have been shown to form non-functional heterodimers with E-box binding bHLH transactivation proteins (Sasai et al., 1992). Therefore,

transcriptional repression of bHLH proteins is achieved in two distinct manners, binding to the N-box and/or titrating other bHLH proteins.

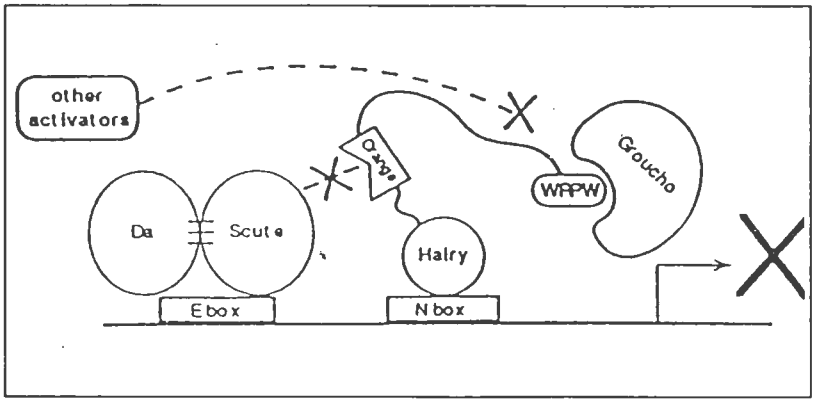


Figure 3. Model for N-box binding mediated transcription repression (Dawson et al., 1995)

Both transcription activation and repression mediated by bHLH proteins are essential for organ development. Lack of either mechanisms results in developmental defects. For example, Mash1 promotes the neuronal differentiation. The absence of Mash1 in mice results in death at birth accompanied by the loss of olfactory and autonomic neurons (Guillemot et al., 1993). HES proteins, however, suppress the neuronal differentiation. The absence of HES proteins (e.g., HES1) accelerates neuronal differentiation, resulting in severe defects such as anencephaly and eye anomalies (Ishibashi et al., 1995).

4. Other features of bHLH proteins

In addition to the characteristics described above, bHLH proteins also have other important properties in the organ developmental processes. Some bHLH proteins compensate functionally for each other and are subjected to auto- and cross-regulation. For example: 1. Myf5, a myogenic bHLH protein, was initially found indispensable for normal rib cage development. In a later experiment, however, the insertion of the myogenin gene, a homologue of myf5, into the myf5 locus (simultaneously disrupting myf5 function) was found to give rise to mice with a normal rib cage (Wang, et al., 1996). 2. In vertebrate myogenesis, MyoD, Myf5 and myogenin all up-regulate their own expression. They are also able to regulate the expression of others (Fig. 4). MyoD is required for the expression of myogenin. Myf5 induces MyoD, and Id is inactivated by MyoD. A similar complex network also occurs in bHLH proteins involved in neurogenesis and sex determination.

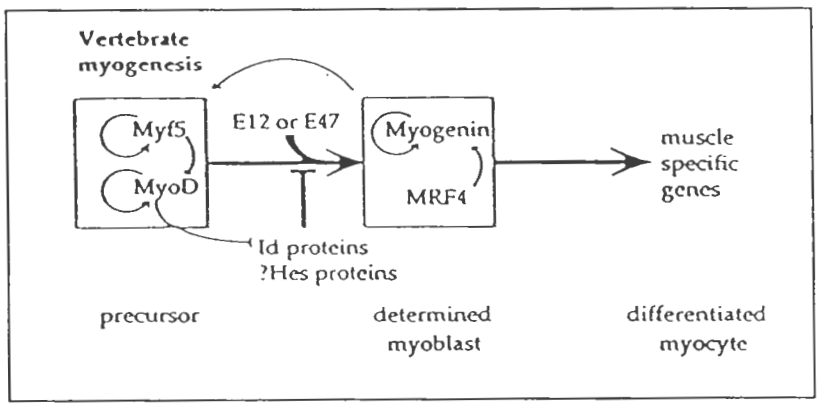


Figure 4. Networks of bHLH proteins involved in vertebrate myogenesis (Little et al., 1998)

II. Significance of bHLH Proteins in Oncogenesis

The bHLH proteins play important roles in the control of cellular proliferation and differentiation in various lineages, from invertebrates to mammals. An imbalance between the cell proliferation and differentiation caused by bHLH proteins may have oncogenic significance. Generally, proliferation-promoting action is oncogenic, while differentiation-promoting action is tumor-suppressing. Myc proteins, for example, are known to promote cell proliferation and inhibit differentiation (Penn et al., 1990, Chin et al., 1995). Several members of the myc family in cooperation with an activated ras oncogene have transformed primary rat embryonic cells in culture (Land et al., 1986). Transgenic mice with enforced c-myc expression also exhibit a significantly higher incidence of malignancy than control mice (Davis et al., 1993).

E proteins, on the other hand, promote cell differentiation. The E2A-null mice (the absence of E2A protein) develop T-cell tumors (Yan et al., 1997). Tal1, a putative oncogene originally identified through its involvement in T-cell acute lymphoblastic leukemia (T-ALL), does not form homodimers but dimerize tightly with E proteins. Such heterodimers confer little transactivation activity (Steven et al., 1998). Sequestration of E proteins by Tal1 oncoprotein likely results in cell de-differentiation and induces oncogenic pathogenesis.

The oncogenic mechanisms by bHLH proteins remain to be established. The deregulation of gene expression by these proteins plays a major role in tumor formation. For example, the p53 tumor suppressor gene contains an essential CACGTG motif within the promoter region. The ectopic c-*myc* can specifically bind to this motif and activate the expression of the mutant p53, leading to oncogenic transformation (Popescu et al., 1998). Other than the gene regulation, a protein-protein interaction is also found responsible for the oncogenesis of some bHLH proteins. For example, in the presence of overexpressed LMO1, the enforced expression of an amino-terminal truncated Tal1, which lacks the transactivation domain, leads to aggressive T-cell malignancies in transgenic mice (Aplan et al., 1993). In this case, Tal1 is not acting by transactivation of the target genes, but acting through a protein-protein interaction.

III. Statement of Purpose

A full-length cDNA was recently isolated in our lab from a human liver library by a gene trapping method. The sequence alignment reveals that this cDNA encodes a protein highly similar to bHLH proteins such as human DEC1 (~92%), and rat SHARP (~80%) and mouse strate B, particularly in the region encoding functional structures such as bHLH domain (Fig. 5). DEC1 is a Bt2cAMP inducible bHLH protein that may function as a transcription regulator in chondrogenesis (Shen et al., 1997). SHARP proteins are mammalian E(spl) and hairy-related bHLH proteins that play essential roles in neurogenesis (Rossner et al., 1997).

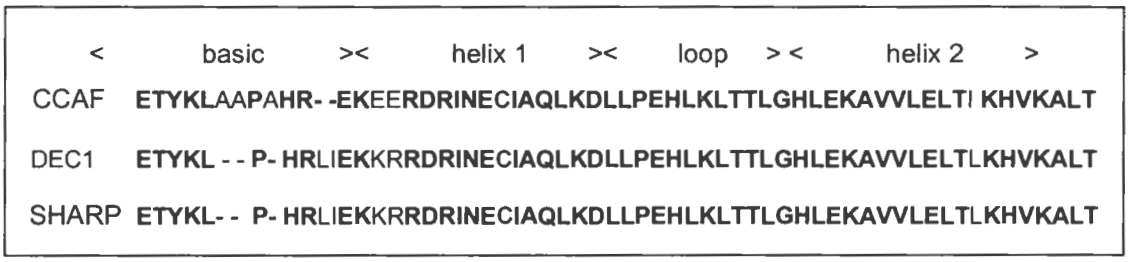


Figure 5. Comparison of CCAF bHLH domain with other bHLH proteins

Based on the Kozak rule, this cDNA encodes a 415 amino acid protein termed colon cancer-associated factor (CCAF), from the first starting AUG codon, which is harbored by a Kozak favorable sequence. **The purpose of this thesis is to determine that the translation is indeed started from the first AUG codon and to establish the molecular weight of CCAF.** CCAF was *in vitro* translated with TNT reticulocyte lysate. The molecular weight of the synthesized CCAF was determined via SDS-polyacrylamide gel electrophoresis (SDS-PAGE) followed by autoradiography. In order to further determine whether CCAF undergoes posttranslational modifications, *in vivo* COS7 cell transient transfection was performed. The CCAF cDNA-transfected cell lysates were subjected to SDS-PAGE and then immunochemically detected by an antibody, which was raised against a peptide derived from CCAF and purified via affinity chromatography. This antibody was also used to determine the cellular localization of CCAF.

bHLH proteins are known to play important roles in the control of cellular proliferation and differentiation. **The experiments in this thesis were designed to determine the expression of CCAF in different cell growth states.** DLD cells derived from colon carcinomas were seeded at different densities, inducing

different cell growth states. The CCAF expression in each state was detected by the antibody prepared against CCAF. DLD cells were also seeded at a certain density, and after reaching confluence, cells were maintained in the same medium or changed to 0.25% medium for 4 additional days. The expression of CCAF was detected at each day point by the same antibody.

The deregulated or ectopic expression of some bHLH proteins (e.g., myc and Tal1) is known closely related to the oncogenic pathogenesis. **The experiments in this thesis were designed to determine the expression of CCAF in human carcinomas and the normal nearby tissues.** Northern blot probed with radioactive-labeled CCAF cDNA was performed to detect the CCAF gene expression in cancer and nearby normal tissues. The same antibody against CCAF was also used to detect the CCAF protein expression in these tissues.

In addition to the bHLH domain, CCAF contains an orange domain and a modified dCtBP motif (PLSLV), which are present in the transcription repressive bHLH proteins such as *Drosophila* Hairy and mammalian HES (Fig. 6). **The purpose of the studies in this thesis was to test the hypothesis that CCAF is a transcription modulator**. To determine the activity of CCAF in transcription regulation, transient cotransfection assays were conducted with an E-box reporter in COS7 cells. The transcriptional modulation activity of CCAF was

determined by measuring the E-box mediated enzyme expression with a dual-luciferase assay.

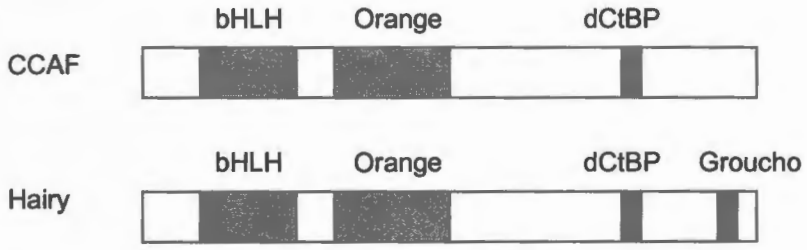


Figure 6. Function domains of CCAF and bHLH repression protein

Overall, the objectives of this thesis are to establish the molecular mass of this protein and to test the hypothesis that CCAF is a transcriptional modulator involving the regulation of cell cycle events. The results in this thesis will provide a fundamental basis for further characterization of CCAF such as the protein-protein interaction and transactivation activity, and therefore expand our understanding of the physiological and pathological significance of bHLH proteins.

Table 4. Abbreviation of bHLH proteins

Ah	aryl hydrocarbon
AP4	activator protein 4
Arnt	Ah receptor nuclear translocation protein
Da	Daughterless protein (*Drosophila*)
E-box	Ephrussi box, consensus sequence CANNTG
E(spl)	Enhancer of split protein (*Drosophila*)
HEB	Hela Ebox binding factor
HES	Hairy- and Enhancer of split-related mammalian proteins
Id	inhibitor of differentiation protein
Hen	Helix-loop-helix encoded in neuroblastoma protein
Ly1	lymphoid leukaemia protein
Mad	Max-associated dimer protein
Mash	mammalian Achaete-Scute homologue protein
Max	Myc-associated x protein
Maxi1	Max interactor 1 protein
MRF4	muscle regulatory factor 1
Myf	myogenic factor
MyoD	myogenic determination factor
N-box	consensus sequence CACNAG
Tal	T cell acute leukemia protein
TFE3	transcription factor E3 binding
TFEB	transcription factor enhancer binding
TFEC	transcription factor C
USF	upstream stimulatory factor

MATERIALS AND METHODS

Plasmid construction

E47 and E-box-luciferase reporter plasmids were kindly provided by Dr. Jinming Chiu (Ohio State University, Ohio). CCAF cDNA was cloned from human liver cDNA library in our lab with a Gene Trapping kit (GIBCO BRL, Grand Island, NY). The plasmid DNA of CCAF was isolated with a QIAprep Spin Miniprep kit (QIAGEN Inc., Valencia, CA). The partial sequencing was determined with a Sequi Therm EXCEL II DNA Sequencing Kit (Epicentre Technologies, Madison, WI) by T7 or SP6 primer. The entire sequence was characterized by Genemed Synthesis Inc. (South San Franciscon, CA).

Transformation and endotoxin-free plasmid extraction

The plasmid was incubated with JM109 bacteria for 20 min on ice, followed by a heat-shock for 45-50 sec at exactly 42°C in a water bath. After 1.5 hr of incubation in SOC medium at 37°C with shaking, the transformed bacteria grew overnight at 37°C in 1.5 % agar plates (100 μg/ml ampicilin). Positive colonies with E47 or E-box-luciferase cDNA insertion were identified with restriction enzymes and analyzed by 1% agarose gel electrophoresis.

CCAF, E47 and E-box-luciferase reporter plasmids for *in vitro* translation and *in vivo* transfection were isolated with an EndoFree Plasmid Maxi kit (QIAGEN, VALENCIA, CA). The individual colonies were incubated in 2-5 ml selective LB medium (100 μg/ml ampicilin) at 37°C for 8 hr with vigorous shaking (~300 rpm).

They were then diluted at 1:1000 into 100 ml selective LB medium and grew under the same conditions for 12~16 hr. Bacteria were harvested by centrifugation for 10 min at 4 °C, 6000 rpm (JOUAN, Winchester, VA), and plasmid DNAs were isolated according to the instruction of the manufacturer. The concentration of plasmid DNA was determined by measuring optical density (OD) with an UV spectrometer at a wavelength of 260 nm (1 OD = 50 μg/ml plasmid DNA).

In vitro translation

CCAF was *in vitro* synthesized via a TNT-coupled reticulocyte lysate system (PROMEGA, Madison, WI) following the instruction by the manufacturer. Briefly, CCAF plasmids (0.5 μg) were added to 25 μl of reaction mixture containing reticulocyte lysate, amino acid mixture without methionine, ^{35}S-labeled methionine and SP6 polymerase. After 60 min of incubation at 30°C, the reaction mixture was chased by adding 625 μM unlabelled methionine (0.5 μl) for 30 min. To determine the molecular weight of the synthesized CCAF, the mixture (5 μl) was denatured in 20 μl of 1X SDS-PAGE sample buffer (31.25 mM Tris-HCl at pH6.8, 5% glycerol, 0.025% bromophenol blue and 2.5% 2-mercaptoethanol) at 80°C for 5 min and size-separated via SDS-PAGE. The gel was washed with solution A (50% methanol and 10% acetic acid) for 30 min and then solution B (7% methanol, 7% acetic acid and 1% glycerol) for 5 min. CCAF was detected by exposing the vacuum-dried gel to an X-ray film.

Cell culture

The COS7 cell line was provided by Dr. Hixon (Rhode Island Hospital, RI). The DLD colon cancer cell line was provided by Dr. Chichester (University of Rhode Island, RI). Unless otherwise indicted, all the reagents for cell culture were purchased from GIBCOL BRL (Grand Island, NY). COS7 cells grow in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS), 1 mM pyruvate, 1x nonessential amino acids, 2 mM L-glutamine and 50 μg/ml gentamicin. DLD cells grow in RPMI 1640 medium supplemented with 10% FCS, 2 mM L-glutamine, 1 x nonessential amino acids and 25 μg/ml gentamicin. Both cell lines were maintained at 37°C, 5% CO_2, 95% O_2 and 100% humidity. After reaching confluence, cells were split by digestion with 0.5% trypsine for 4 to 5 min at 37°C. The number of cells was counted via a hemotocytemeter under a reverse-controversial microscopy after 0.4% typan blue staining.

Transient transfection

CCAF endotoxin-free plasmids were transiently transfected into COS7 cells via lipofectamine, a cationic lipid as described by the manufacturer (Gibco BRL Gaithersburg, MD). COS7 cells were seeded in 12-well plates (CORNING, Corning, NY) in RPMI 1640 medium supplemented with 10% FCS at a density of $1x10^5$ cells per well. On day 2, cells achieved 60-80% confluence. Prior to transfection, plasmid DNA (0.2 or 0.4 μg) was mixed with lipofectamine (2 μl) in 80 μl of transfection DMEM (without FCS and gentamicin) at room temperature

for 45 min, and then diluted to 320 ml with the same medium. Cells were exposed to the transfection mixture at room temperature for 4 to 5 hr, and switched to normal DMEM. The transfected cells were incubated for an additional 48 hr and harvested in lysis buffer (1% SDS in Tris-HCl buffer, pH 6.8, 150 μl per well). Protein concentration of cell lysate was assayed with a BCA micro protein assay kit (PIERCE, Rockford, IL) at a wavelength of 570 nm via a MRX microplate reader (DYNEX, Chantily, VA). A dilution of sample at 1:40 was usually required before performing the protein assay.

Cotransfection and dual luciferase assay

To determine the activity of CCAF on gene transcription, E-box-luciferase reporter plasmids (160 ng) were cotransfected with CCAF plasmids (100 ng) to COS7 cells in the presence of lipofectamine following the same procedure as described above. E47 plasmids (100 ng) were transfected as a positive control. pRL-TK plasmids (32 ng, Promega, Madison, WI) encoding *Renilla* luciferase served as a background control. Transfected cells were incubated for 48 hr, rinsed with sodium phosphate buffer twice and lysed in 250 μl of passive lysis buffer at room temperature for 15 min on a rocking platform. The reporter enzyme activities were determined with a Dual-luciferase Reporter Assay System via a TD-20/20 luminometer (Promega, Maison, WI). The assay program was set up to a 2-second delay and a 10-second integration. This assay system contains two substrates, which give rise to two kinds of luminescence signals. The *Firefly* luminescence signal, which represented the reporter gene activity,

was initiated by mixing cell lysate (20 μl) with substrate Luciferase Assay Reagent II (LARII, 100 μl). The *Renilla* luminescence signal, which represented the background luciferase activity, was detected by sequentially adding Stop & Glu reagent (100 μl) to the same mixture. The induction of reporter enzyme activity was designated as the ratio of *Firefly* luminescence signal over the *Renilla* luminescence signal. In this experiment, each group was tested in triplicate and repeated once.

SDS-Polyacrylamide gel electrophoresis

SDS-PAGE was conducted with Minimum Gel Apparatus (BioRad, Hercules, CA). The gel contained two parts: 4% polyacrylamide staking gel (0.25 M Tris-HCl, pH6.8, 1% SDS) on the top and 7.5% polyacrylamide separating gel (0.375 M Tris-HCl, pH8.8, 1% SDS) at the bottom. Cell lysate containing 15 μg protein was denatured in equal volume of 2 x SDS-PAGE sample buffer (62.5 mM Tris-HCl at pH 6.8, 2% SDS, 10% glycerol, 0.025% bromphenol blue, 5% 2-mercptoethanol) at 95°C for 5 min and centrifuged at 11,000 rpm for 10 min. The supernatant was separated in the stacking gel at 43 V, 16 mA for about 30 min and in the separating gel at 120 V, 26 mA for about 1.5 hr.

Coomassie blue assay

Transfected cell lysate was subjected to SDS-PAGE as described previously. The gel was stained for half an hour with 0.1% Coomassie blue R-250 in a fixative solution (40% MeOH, 10% HOAc), destained with 40% MeOH/10%

HOAc (1 to 3 hr) to remove the background and then vacuum-dried on a gel drier. The expression pattern of total proteins in transfected COS7 cells was visualized by Coomassie blue staining.

Antibody purification

A polyclonal antibody specific to the peptide (CSQALKPIPPLNLETKD) derived from C-terminus of CCAF was raised in New Zealand White rabbits in our lab. To diminish the nonspecific binding, the antibody was purified by immunoaffinity chromatography. This was conducted as described by Harlow and Lane (Harlow et al, 1988). First, the peptide (1 mg) was covalently coupled to 2 ml of SulfoLink gel (PIERCE, Rockford, IL) via incubation at room temperature for 30 min in a PD-10 column (Pharmacia Biotech, Sweden). 50 mM cysteine (2 ml) was then added to block the nonspecific binding sites. Prior to applying the antiserum, this antigen-coupled gel column was sequentially pretreated with 20 ml of Tris (pH7.5), 20 ml of 100 mM glycine (pH 2.5) and 20 ml of 100 mM triethylamine (pH11.5, fresh). The antiserum (5 ml) was 1:1 diluted with 10 mM Tris (pH7.5) and repetitively applied to the column three times to ensure the complete binding. The column was washed with 40 ml of 10 mM Tris (pH7.5) followed by 40 ml of 500 mM NaCl in 10 mM Tris (pH7.5). The antigen specific antibody was then eluted from the column with 20 ml of 100 mM glycine (pH2.5), and collected in a tube containing 5 ml of 1M Tris-HCl (pH 8.0). The elution was dialyzed in a tubin against PBS (containing 0.02% sodium azide) at 4°C overnight with stirring, and then aliquoted and stored at -20°C.

Western immunoblotting

The transfected cell lysate was subjected to SDS-PAGE as described previously. For immunoblot analysis, the separated proteins were transferred to a nitrocellulose membrane by electroblotting for 1 hr 20 min at 36 V, 150 mA in transfer buffer (25 mM Tris, 190 mM glycine, and 20% methanol). The membrane was then blocked in 5% non-fat dry milk in TBST (10 mM Tris 8.0, 150 mM NaCl, 0.05% Tween 20) for half an hour, and probed by the purified antibody at 1:100 dilution for 1 hr. The membrane was then washed three times (5-10 min/each) with TBST, and incubated with an alkaline phosphatase-conjugated anti-rabbit IgG at 1:5000 dilution for at least 0.5 hr. Immunoblots were developed by color precipitation catalyzed by phosphatase in the presence of chromogenic substrates, 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and Nitro blue tetrazolium (NBT) in alkline phosphatase buffer (0.1M Tris, o.1M NaCl, and 5mM $MgCl_2$, pH9.5).

CCAF expression at different cell growth states

In this experiment, different cell growth states were induced by three individual methods: (1) DLD colon cancer cells were seeded in 12-well plates at various densities (1-8 x 10^5 cells per well) in RPMI 1640 with 10% FCS. The medium was changed daily after seeding. Cells were harvested 3 days after seeding. (2) DLD cells were seeded in 12-well plates at a cell density of $2.5x10^5$ per well in RPMI 1640 medium with 10% FCS. After achieving confluence (about 3 days), cells were maintained in the same medium for additional 4 days. Cells were

harvested at each of the 7-day points. (3) DLD cells were treated in the same way as (2) except maintaining confluent cells in 0.25% serum for an additional 4 days. All the cells were harvested in the same lysis buffer. Protein concentration in cell lysate was determined with a BCA micro protein assay kit, and each sample containing 15 μg protein was immunoblotted with the antibody against CCAF as described previously.

CCAF gene expression in human colon carcinomas

Total RNAs from human colon cancer and normal tissues were isolated with a TRI REAGENT kit following the instruction of the manufacturer (Sigma, St. Louis, MO). The RNA concentration was determined via UV absorbance at 260 nm, with 1 OD unit equal to 40 μg/ml RNA. Samples (20 μg RNA) was fractionated via electrophoresis in a 1.2% agarose gel containing 2.2 M formaldehyde, and electroblotted to a Nytran nylon membrane (Schleicher & Schuell, Keene, NH) with a vacuum-blotting system (Pharmacia, Piscataway, NJ). The membrane was hybridized with a ^{32}P-labelled CCAF cDNA probe (~2 x 10^6 cpm) in 5 ml of hybridization buffer (250 mM sodium phosphate buffer, pH7.2, 1 mM EDTA, 7% SDS and 0.01% bovine serum albumin) overnight at 68°C in a rotating oven (Robbins Scientific, Clayview, CA). After hybridization, the membrane was washed at 68°C for 1 hr in hybridization buffer (2 x 30 min) and for 1 hr in 40 mM sodium phosphate buffer (pH 7.2, 1 mM EDTA and 1% SDS) (2 x 30 min). An X-ray film (Sigma Chemical Co, St, Louis, MO) was exposed to the membrane with intensifying screen until an appropriate autoradiography was obtained.

CCAF protein expression in human colon carcinomas

To compare the CCAF protein expression in human colon cancer to normal tissues, homogenates of each sample (15 μg protein), which were provided by Dr. Nancy Thomas (Rhode Island Hospital, RI), were size-separated by SDS-PAGE and immunodetected with the antibody against CCAF as described earlier.

Determination of CCAF localization

DLD cells were cultured in 12-well plates in RPMI medium with 10% FCS. When reaching confluence, cells were rinsed twice with ice-cold PBS (1 ml/well), and detached by a rubber policeman. The cells were collected and centrifuged for 5 min at 2000 rpm, 4°C. The cell pellet was mixed with 100 μl of NP-40 lysis buffer A (10 mM Tris, pH7.4, 10 mM NaCl, 3 mM MgCl2 and 0.5% NP-40) on ice for 5 min followed by centrifugation at 1000rpm, 4°C for 5 min. The supernatant containing cytosol proteins were stored in 0.5 ml plastic tubes while the nuclear proteins in the pellet were redissolved in 70 μl of 1% SDS lysis buffer. All the samples prepared above were separated via SDS-PAGE followed by immunoblotting with antibody against CCSF as described previously.

RESULTS

CCAF is a 52-KDa protein.

To establish the molecular weight, CCAF was in vitro translated with TNT reticulocyte lysate and analyzed by autoradiography. Addition of the CCAF cDNA to the reaction mixture yielded a single product with a molecular weight of 52-kDa (Fig. 7). This mass is consistent with the estimated weight. In order to determine whether CCAF undergoes posttranslational modifications, the CCAF cDNA-transfected COS7 cell lysates were size-separated via SDS-PAGE and analyzed by Coomassie blue assay and Western immunoblotting, respectively. As determined by Coomassie blue assay, an extra band with a molecular weight of 52-kDa was detected in the CCAF-transfected cells, and the level of the protein expression increased with the increase of the CCAF plasmid concentration (Fig. 8, lane 2 and 3). An antibody was raised against a peptide derived from CCAF and subjected to affinity chromatography. This antibody detected the 52-kDa protein in the CCAF cDNA-transfected cells but not in the control cells (Fig. 9). Thus, three independent assays all demonstrated that the isolated full-length cDNA encodes a protein with a molecular weight of 52-kDa.

CCAF expression is related to the cell growth states.

To determine the CCAF expression in different cell growth states, DLD colon cancer cells were seeded at different densities, and the cell lysates were fractionated via SDS-PAGE and then analyzed by western immunoblotting. As shown in Fig. 10, the antibody against CCAF detected a 52-kDa protein in DLD

cells at all seeding densities. But the expression of this protein increased with the plating densities. The level of this protein in the cells that are seeded at high density was ~3 to 5-fold higher (lane 3, 4 and 5) than that in the cells that are seeded at low density (lane 1 and 2).

To further confirm this expression pattern, DLD cells were seeded at a certain density, and after reaching confluence, cells were maintained in the same medium or changed to 0.25% medium for 4 additional days. The expression of CCAF was detected at each day point by the same antibody against CCAF. As shown in Fig. 11, the CCAF expression was hardly detectable before cell reaching confluence (lane1 and 2), while after confluence, the level of CCAF protein increased about 3-5 folds (lane 3-6). The level of CCAF expression was also positively related to the starvation days (Fig. 12).

CCAF is abundantly expressed in human colon carcinomas but not in the nearby normal tissues.

Northern blot with the radiolabeled full-length cDNA screened the CCAF mRNA expression in various human carcinomas. Except kidney and ovary cancer, CCAF gene expression was significantly high in some carcinomas such as lung and breast when compared to that in normal tissues (Data not shown). Particularly, without exception, this cDNA detected abundant expression of CCAF in 5 individual colon cancer cases but not in the nearby normal tissues (Fig. 13). The same results were also obtained by using immunoblotting to

detect the CCAF protein expression in colon carcinomas. As shown in Fig. 14, the antibody specific to CCAF recognized a strong band with a molecular weight of 52-kDa in samples from 3 individual colon cancer patients (Lane 3, 4 and 5). In contrast, no bands were detected by the same antibody in normal nearby tissues (Lane 6 and 7).

CCAF is localized in cell nucleus.

DLD cytosol proteins were separated from nuclear proteins, and both were subjected to Western immunoblotting analysis. As shown in Fig. 15, the antibody against CCAF recognized a protein with a molecular weight of 52-kDa in nucleus but not in cytosol, suggesting that the CCAF protein is a nuclear protein.

CCAF inhibits the transactivation activity of E47 on an E-box reporter.

To test the activity of CCAF on the gene regulation, CCAF plasmids were cotransfected with an E-box luciferase reporter and the induction of reporter enzyme was determined via a dual-luciferase assay system. As shown in Fig. 16, E47, a bHLH protein that is known to transactivate E-box reporter, promoted a 380-fold increase of luciferase activity. In contrast, CCAF caused little activation on luciferase expression by itself, but partially antagonized by 30% the enzyme induction conferred by E47.

Figure 7. *In vitro* translation

CCAF was synthesized *in vitro* via a TNT-coupled reticulocyte lysate system. CCAF plasmids (0.5 μg) were added to 25 μl of reaction mixture containing reticulocyte lysate, amino acid mixture without methionine, ^{35}S-labeled methionine and SP6 polymerase. After 60 min of incubation at 30°C, the reaction mixture was chased by adding 625 μM unlabelled methionine (0.5 μl) for 30 min. To determine the molecular of synthesized CCAF, the mixture (5 μl) was denatured in 20 μl of 1X SDS-PAGE sample buffer (31.25 mM Tris-HCl at pH6.8, 5% glycerol, 0.025% bromophenol blue and 2.5% 2-mercaptoethanol) at 80°C for 5 min and size-separated via SDS-PAGE. The gel was washed with solution A (50% methanol and 10% acetic acid) for 30 min and then solution B (7% methanol, 7% acetic acid and 1% glycerol) for 5 min. The synthesized CCAF was detected by autoradiography.

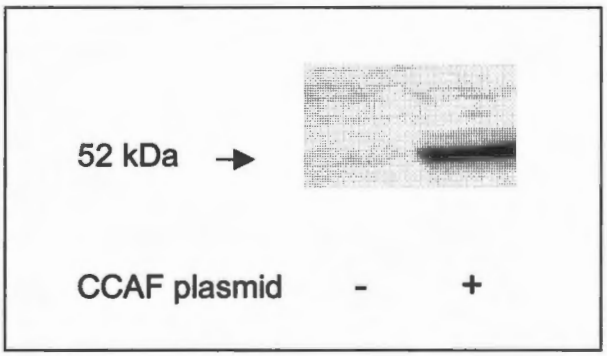

52 kDa
CCAF plasmid
-
+

Figure 8. Coomassie Blue Assay

Transfected cell lysate was subjected to SDS-PAGE. The gel was stained for half an hour with 0.1% Coomassie blue R-250 in a fixative solution (40% MeOH, 10% HOAc), destained with 40% MeOH/10% HOAc (1 to 3 hr) to remove the background and vacuum-dried on a gel drier.

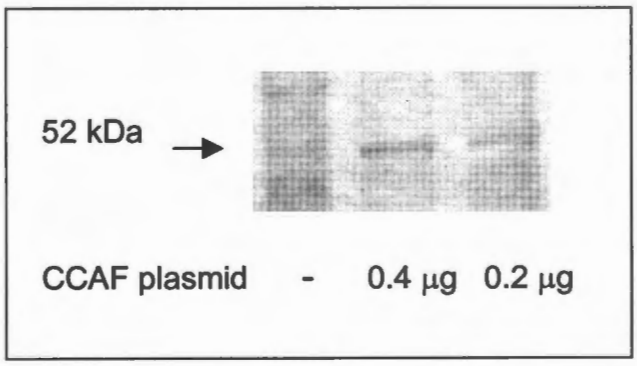

52 kDa
CCAF plasmid - 0.4 μg 0.2 μg

Figure 9. Western immunoblotting

The transfected cell lysate was subjected to SDS-PAGE. The separated proteins were transferred to a nitrocellulose membrane by electroblotting for 1 hr 20 min at 36 V, 150 mA in transfer buffer (25 mM Tris, 190 mM glycine, and 20% methanol). The membrane was then blocked in 5% non-fat dry milk in TBST (10 mM Tris 8.0, 150 mM NaCl, 0.05% Tween 20) for half an hour, and probed by the purified antibody at 1:100 dilution for 1 hr. The membrane was washed three times (5-10 min/each) with TBST, and incubated with an alkaline phosphatase-conjugated anti-rabbit IgG at 1:5000 dilution for at least 0.5 hr. Immunoblots were developed by color precipitation catalyzed by phosphatase in the presence of chromogenic substrates, 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and Nitro blue tetrazolium (NBT) in alkline phosphatase buffer (0.1M Tris, o.1M NaCl, and 5mM $MgCl_2$, pH9.5).

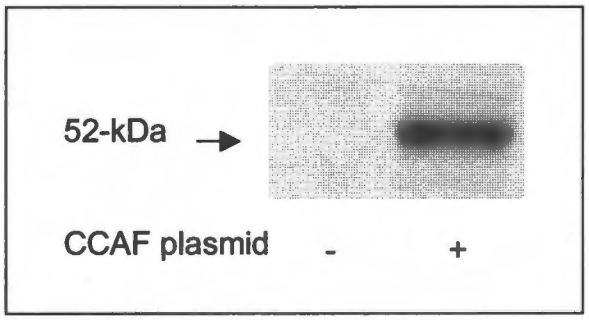

52-kDa
CCAF plasmid
-
+

Figure 10. CCAF expression at different cell densities

DLD colon cancer cells were seeded in 12-well plates at various densities (1-8 x 10^5 cells per well) in RPMI 1640 with 10% FCS. The medium was changed everyday after seeding. On days 3 after seeding, cells were harvested in lysis buffer. Each sample containing 15 μg protein was analyzed with antibody against CCAF. CCAF-transfected COS7 cell lysate is also included in this experiment.

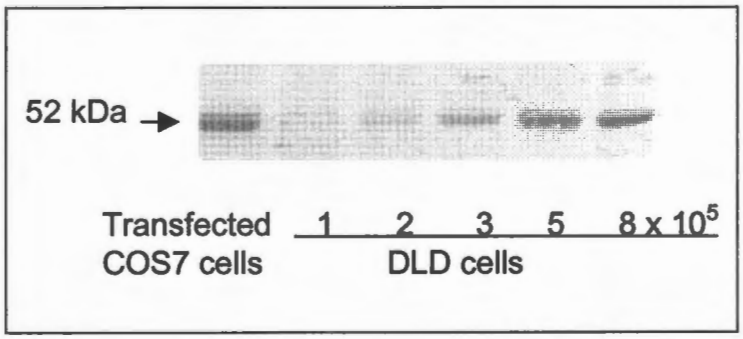

52 kDa
Transfected COS7 cells
1
2
3
5
8 x 10^5
DLD cells

Figure 11. CCAF expression at different cell growth states

(without changing medium)

DLD cells were seeded in 12-well plates at a cell density of $2.5x10^5$ per well in RPMI 1640 medium with 10% FCS. After achieving confluence (about 3 days), cells were maintained in the same media for additional 4 days. At each day point, cells were harvested in lysis buffer. Each sample containing 15 μg protein was analyzed with antibody against CCAF.

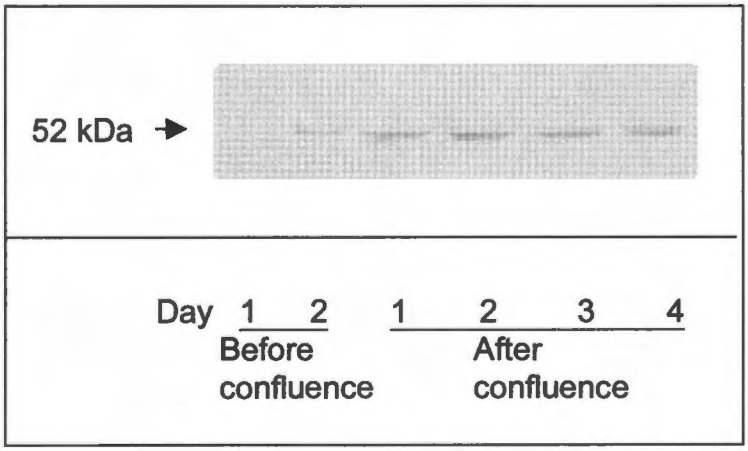

52 kDa
Day 1 2 1 2 3 4
Before confluence
After confluence

Figure 12. CCAF expression at different cell growth states

(changing to 0.25% FCS)

DLD cells were seeded in 12-well plates at a cell density of $2.5x10^5$ per well in RPMI 1640 medium with 10% FCS. After achieving confluence (about 3 days), cells were switched to medium containing 0.25% FCS for additional 4 days. At each day point, cells were harvested in the lysis buffer. Each sample containing 15 μg protein was analyzed by immunoblotting. This figure compared the CCAF expression between two days after starvation.

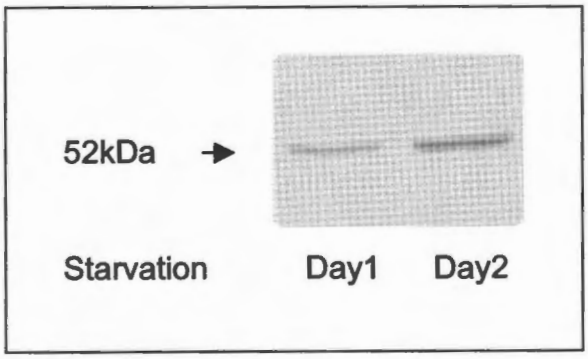

52kDa
Starvation
Day1
Day2

Figure 13. CCAF gene expression in human colon carcinomas

Total RNAs from human colon cancer and normal tissues were isolated with a TRI REAGENT kit. Samples (20 μg RNA) was fractionated via electrophoresis in a 1.2% agarose gel containing 2.2 M formaldehyde, and blotted to a Nytran nylon membrane with a vacuum-blotting system. The membrane was hybridized with a ^{32}P-labelled CCAF cDNA probe (~2 x 10^6 cpm) in 5 ml of hybridization buffer (250 mM sodium phosphate buffer, pH7.2, 1 mM EDTA, 7% SDS and 0.01% bovine serum albumin) overnight at 68°C in a rotating oven. After hybridization, the membrane was washed at 68°C for 1 hr in hybridization buffer (2 x 30min) and for 1 hr in 40 mM sodium phosphate buffer (pH 7.2, 1 mM EDTA and 1% SDS) (2 x 30 min). CCAF gene expression was visualized by autoradiography. The results shown in this figure were obtained from 5 individual patients.

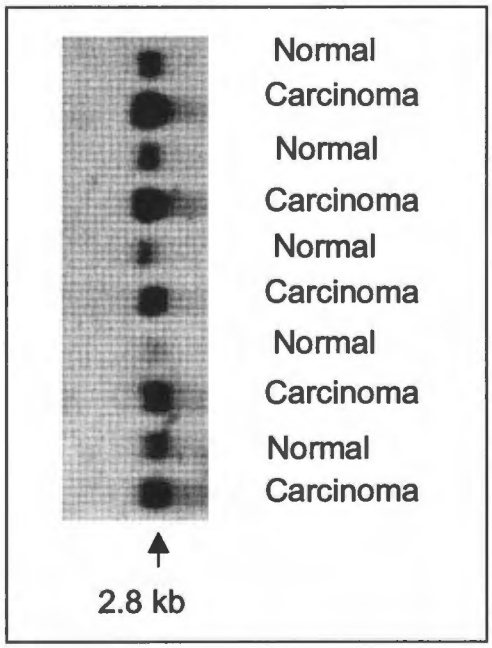

Normal
Carcinoma
Normal
Carcinoma
Normal
Carcinoma
Normal
Carcinoma
Normal
Carcinoma
2.8 kb

Figure 14. CCAF protein expression in human colon carcinomas _

To compare the CCAF protein expression between human colon cancer and normal tissues, homogenates of each sample (15 μg protein) were size-separated by SDS-PAGE and analyzed by immunoblotting.

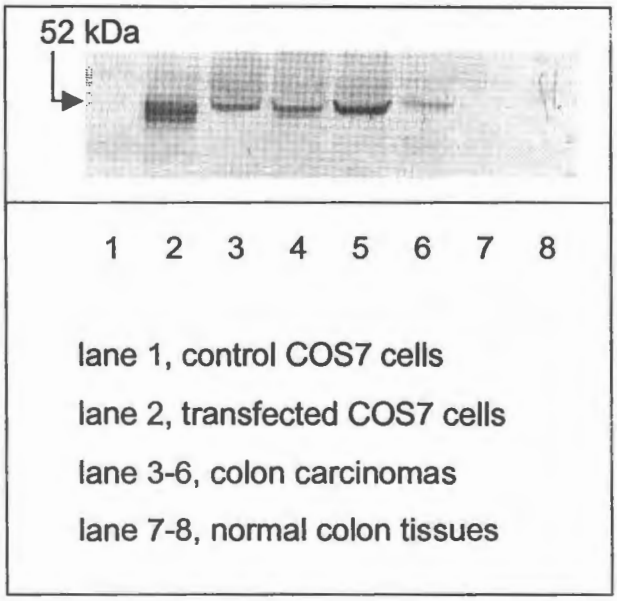

52 kDa
1 2 3 4 5 6 7 8
lane 1, control COS7 cells
lane 2, transfected COS7 cells
lane 3-6, colon carcinomas
lane 7-8, normal colon tissues

Figure 15. CCAF nuclear localization

DLD cells were cultured in 12-well plates in RPMI medium with 10% FCS. When reaching confluence, cells were rinsed twice with ice-cold PBS (1 ml/well), and detached by a rubber policeman. The cells were collected and centrifuged for 5 min at 2000 rpm, 4°C. Cell pellet was mixed with 100 μl of NP-40 lysis buffer A (10 mM Tris, pH7.4, 10 mM NaCl, 3 mM MgCl2 and 0.5% NP-40) on ice for 5 min followed by centrifugation at 1000rpm, 4°C for 5 min. The supernatant containing cytosol proteins were stored in 0.5 ml plastic tubes while the nuclear proteins in the pellet were redissolved in 70 μl of 1% SDS lysis buffer. All the samples prepared above were separated via SDS-PAGE and analyzed by immunoblotting with antibody against CCAF.

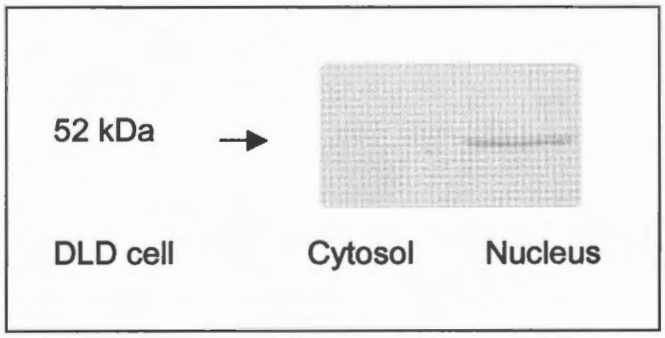
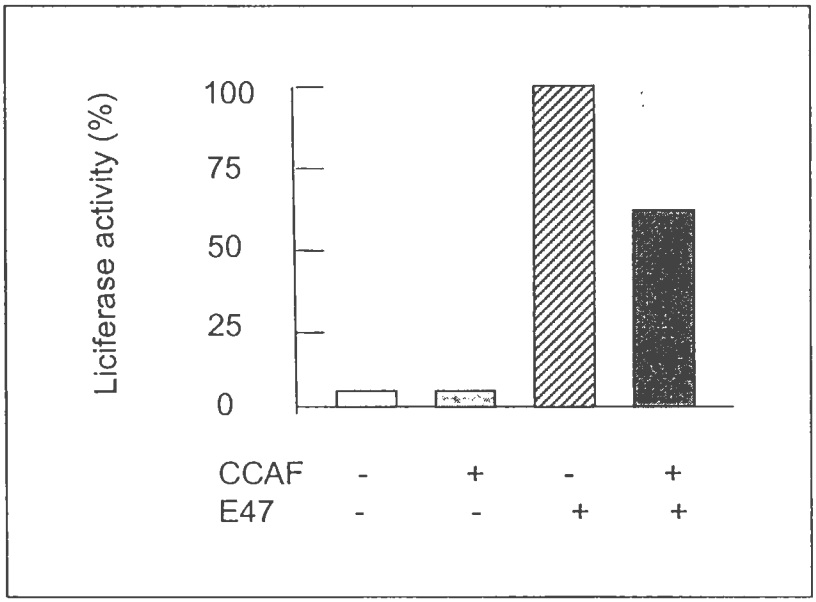

52 kDa
DLD cell
Cytosol
Nucleus

Figure 16. Cotransfection and dual luciferase assay

E-box-luciferase reporter plasmids (160 ng) was cotransfected with CCAF plasmids (100 ng) into COS7 cells in the presence of lipofectamine following the same procedure as described in the Materials and Methods. E47 plasmids (100 ng) were transfected as a positive control. pRL-TK plasmids (32 ng) encoding *Renilla* luciferase served as a background control. Transfected cells were incubated for 48 hr, rinsed with sodium phosphate buffer twice and lysed in 250 μl of passive lysis buffer at room temperature for 15 min on a rocking platform. The reporter enzyme activities were determined with a Dual-luciferase Reporter Assay System via a TD-20/20 luminometer. The *Firefly* luminescence signal was initiated by mixing cell lysate (20 μl) with substrate Luciferase Assay Reagent II (LARII, 100 μl). The *Renilla* luminescence signal was detected by sequentially adding Stop & Glu reagent (100 μl) to the same mixture. The induction of reporter enzyme activity was expressed as the ratio of *Firefly* luminescence signal over *Renilla* luminescence signal. In this experiment, each group was tested in triplicate and repeated once.


Liciferase activity (%)
100
75
50
25
0
CCAF - + - +
E47 - - + +

DISCUSSION

Basic helix-loop-helix (bHLH) proteins belong to a large family of transcription factors that are known to play important roles in cell proliferation, differentiation and oncogenesis. These proteins are structurally featured by a bHLH motif, which is responsible for protein dimerization and sequence-specific DNA binding (e.g., E-box). Recently we isolated a cDNA from a human liver library by a gene trapping method. Sequence alignment indicates that this cDNA encodes a putative bHLH protein termed CCAF. The objective of this thesis is to establish the molecular mass of this protein and to test the hypothesis that CCAF is a transcriptional modulator involving the regulation of cell cycle events.

CCAF is a 52-kDa protein with little posttranslational modifications.

Based on the Kozak rule, the largest protein encoded by the isolated cDNA is 415 amino acids long with a calculated molecular mass of 52-kDa. One of the studies described in this thesis is to determine the molecular weight of CCAF. This is achieved by measuring the mass of CCAF translated *in vitro* and in CCAF cDNA-transfected cells. In both methods, a protein with a molecular weight of 52-kDa is produced. The consistence with the calculated mass suggests that CCAF undergoes little posttranslational modifications.

CCAF is a cell cycle regulator.

The expression of CCAF is detected in different cell growing states induced by either contact inhibition or serum starvation. The proliferating cells express lower

level of CCAF than the growing-arrest cells. This differential expression is closely related to the regulation of cell cycle. Myc proteins are highly expressed in the proliferating cells but less in the differential cells, and these proteins are known to promote the cell proliferation but inhibit the differentiation (Penn et al., 1990, Chin et al., 1995). The low level of CCAF in the proliferating cells suggests that CCAF is a cell cycle regulator, which has little proliferation-promoting activity but plays an important role in the rescuing cells from death (anti-apoptotic).

CCAF has oncogenic significance.

The expression pattern of CCAF in different cell growing states is similar to the oncoprotein junD, which is found to cooperate with ras oncogene in transforming rat embryo fibroblast (Vandel L et al., 1996). This similarity suggests that CCAF has oncogenic significance. Moreover, abundant CCAF is found in human colon carcinomas but not in the nearby tissues. Ectopic and deregulated expression of several bHLH proteins is related to oncogenesis. Tal1 and other closely related bHLH proteins (Tal2 and LYL1), are normally not expressed in T cells, but are constitutively expressed in >60% of T-cell acute lymphoblastic leukemia (T-ALL) (Alphan et al., 1992, Bash et al., 1995). In transgenic mice, overexpression of TAL1 gene in cooperation with a misexpressed LMO1 protein induces aggressive T-cell malignancies (Aplan et al., 1993).

It has been proposed that the progression of a tumor might not only be a function of cell proliferation but also result from inappropriate suppression of apoptosis

(Marx, 1993). The oncogenic nature of some bHLH proteins is known to suppress apoptosis. For example, TAL1 bHLH oncoprotein is recently found antiapoptotic. A Jurkat leukemic T cell subline expressing a C-terminally truncated mutant TAL1 undergoes rapid apoptosis upon medium depletion. Transfection with a wild type of TAL1 reverses this process, suggesting that TAL1 inhibits the apoptotic signaling in the absence of survival factors (Leroy-Viard et al., 1995). Overexpression of TAL1 significantly blocks granulopoietic and monocytic cell apoptosis induced by chemotherapeutic agents (Bernard et al., 1998). The oncogenic pathology of CCAF is likely due to its antiapoptosis.

CCAF is a gene transcription regulator.

Not all the proteins that possess bHLH domains are transcription factors. For example, the calcium binding proteins with the bHLH domain are just components of the calcium-signaling pathway (Marsden et al, 1990). Thus the final purpose is to determine if CCAF is a transcription modulator. Firstly, the nuclear localization provides the evidence that CCAF is a nuclear protein. Then the antagonism of CCAF on E47-mediated transactivation further supports that CCAF is a transcription regulator. The molecular mechanisms for such an antagonism remain to be determined. There are three hypotheses: (1) CCAF binds to the E-box, but the binding doesn't activate transcription; (2) CCAF sequestrates E47 by forming inactive, non-DNA-binding heterodimers; (3) CCAF dimerizes with E47 but the heterodimers are less active than E47 homodimers. E47 is also known a tumor-suppressing protein. The repression on E47 may

have oncogenic significance. This has been confirmed in oncoproteins such as Tal1, which forms heterodimers with E47 (Steven et al., 1998).

In summary, the results in the thesis support our overall hypothesis that CCAF is a transcriptional modulator involving the regulation of cell cycle events, and the expression of CCAF is related to oncogenic pathology. However, further experiments need to be conducted to fully understand CCAF from the structure, the function and its molecular basis. The electrophoretic gel mobility assay (EGMA) can be used to study whether the DNA-binding is required for the transcription repression (E-box and/or N-box transactivation). The gene array can be used to study the target genes regulated by CCAF, and the yeast two-hybrid system can be used to study the CCAF mediated protein-protein interaction. These studies will likely define whether CCAF is a marker for colon tumor diagnosis and whether CCAF is a potential target for the therapeutic intervention.

REFERENCES

Aplan PD, Lombardi DP, Reaman GH, Sather HN, Hammond GD and Kiysch IR (1992) Involvement of the putative hematopoietic transcription factor SCL in T-cell acute lymphoblastic leukemia *Blood* **79**, 1327-1333

Aplan PD, Jones CA, Chervinsky DS, Zhao X, Ellsworth M, Wu C, McGuire EA and Gross KW (1993) An scl gene product lacking the transactivation domain induces biny abnormalities and c0orperates with LMO1 to generate T-cell malignances in transgenic mice *EMBO J.* **16**, 2408-2419

Atchley WR and Fitch WM (1997) A natural classification of the basic helix-loop-helix class of transcription factors *Proc. Natl. Acad. Sci. USA* **94,** 5172-5176

Bacsi SG, Reisz Porszas S and Hankinson O (1995) Orientation of the heterodimeric aryl hydrocarbon (dioxin) receptor complex on its asymmetric DNA recognition sequence *Mol. Pharmacol.* **47**, 432-438

Bash RO, Hall S, Timmons CF, Crist WM, Amylon M, Smith RG and Bear R (1995) Does activation of the TAL1 gene occur in a majority of patients with T-cell acute lymphoblastic leukemia? *Blood* **86,** 666-676

Benezra R, Davis RL, Lockshon D, Turner DL and Weintraub H (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins *Cell* **61,** 49-59

Bernard M, Delabesse E, Novault S, Hermine O and Macintyre EA (1998 Antiapoptic effect of ectopic TAL/SCL expression in a human leukemic T-cell line *Cancer Res.* **58**, 2680-2687

Burley SK (1994) DNA-binding motif from eukaryotic transcription factors *Curr. Opin. Cell Biol.* **4,** 3-11

Cabrera CV and Alonso MC (1991) Transcriptional activation by heterodimers of the achaete-scute and daughterless gene products of *Drosophila EMBO J* **10,** 2965-2973

Chin L. Schreiber Agus N, Pellicer I, Chen K, Lee HW, Dudast M, Cordon Cardo C and DePinho RA (1995) Contrasting roles for Myc and mad proteins in cellular growth and differentiation *Proc. Natl. Acad. Sci. USA* **92,** 8488-8492

Condorelli GL, Tocci A, Botta R, Facchiano F, Testa U, Vitelli L, Valteis M, Croce CM and Peschle C. (1997) Ectopic *TAL-1/SCL* expression in phenotypically normal or leukemic myeloid precursors: Proliferative and antiapoptotic effects coupled with a differentiation blockade *Mol. Cell. Biol.* **17,** 2954-2969

Davis AC, Wims M, Spotts GD, Hann SR and Bradley A (1993) A null-c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice *Genes Dev.* **7,** 671-682

Davis RL, Cheng P-F, Lassar AB, and Weintraub H (1990) The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation *Cell* **60,** 733-746

Dawson SR, Turner DL, Weintraub H and Parkhurst SM (1995) Specificity of the Hairy/Enhancer of split basic helix-loop-helix (bHLH) proteins maps outside the bHLH domain and suggested two separable modes of transcripitonal repression *Mol. Cell. Biol.* **15,** 6923-6931

French BA, Chow KL, Olson EN and Schwartz RJ (1991) Heterodimers of myogenic helix-loop-helix regulatory factors and E12 bind a complex element governing myogenic induction of avian cardiac alpha-actin promoter *Mol. Cell. Biol.* **11**, 2439-2450

Goldfarb AN, Lewandowska K, and Pennell CA (1998) Identification of highly conserved module in E proteins required for in vivo helix-loop-helix dimerization *J. Biol. Chem.* **273,** 2866-2873

Goodrich J, Carpenter R and Coen ES (1992) A common gene regulates pigmentation pattern in diverse plant species *Cell* **68**, 955-964

Guillemot F, Lo L-C, Johnson JE, Auerbach A, Anderson DJ and Joyner AL (1993) Mammalian *achaete-scute* homologue 1 is required for the early development of olfactory and autonomic neurons *Cell* **75,** 463-476

Gupta S, Seth A and Davis RJ (1993) Transactivation of gene expression by Myc is inhibited by mutation at the phosphorylation sites Thr-58 and Ser-62 *Proc. Natl. Acad. Sci. USA* **90,** 3216-3220

Hirose K, Morita M, Ema M, Mimura J, Hamada H, Fujii H, Saijo Y, Gotoh O, Sogawa K and Fujii-Kuriyama Y (1996) cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt) *Mol. Cell Biol.* **16**, 1706-1713

Hsu HL, Wadman I and Baer R (1994) Formation of in vivo complex between the TAL1 and E2A polypeptides of leukemic T cells *Proc. Natl. Acad. Sci. USA* **91,** 5947-5951

Ishibashi M, Ang S-L, Shiota K, Nakanishi S, Kageyama R and Guillemot F (1995) Targeted disruption of mammalian hairy and Enhancer of split homolog-1

(HES1) leads to up-regulation of neuronal helix-loop-helix factors, premature neurogenesis, and sever neural tube defects *Genes Dev.* **9,** 3136-3148

Issack PS and Ziff EB (1998) Altered expression of helix-loop-helix transcriptional regulators and cyclin D1 in Wnt-1-PC12 cells *Cell Growth Differ* **9**, 837-845

Jarman AP, Grell EHAckerman L, Jan LY and Jan YN (1994) Atonal is the proneural gene for Drosophila photoreceptors *Nature* **369**, 398-400

Land H, Chen AC, Morgenstern JP, Parada LF and Weinberg RA (1986) Behavior of myc and ras oncogenes in transformation of rat embryo fibroblasts *Mol. Cell. Biol.* **6,** 1917-1925

Leroy-Viard K, Vinit MA, Lecointe N, Jouault H, Hibner U, Romeo PH and Matieu-Mahul D (1995) Loss of TAL-1 protein activity induces premature apoptosis of Jukat leukemic T cells upon medium depletion *Proc. Nati. Acad. Sci. USA* **14**, 2341-2349

Lindebro MC, Poellinger L and Whitelaw ML (1995) Protein-protein interaction via PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex *EMBO J* **14,** 3528-3539

Little TD and Evan GI (1998) Helix-loop-helix transcription factors, *Oxford University Press*, PP1-48

Luo X and Sawadogo (1996) Antiproliferative properties of the USF family of helix-loop-helix transcription factors Proc. *Natl. Acad. Sci. USA* **93**, 1308-1313

Marx J (1993) Cell death studies yield cancer clues *Science* **259,** 760-761

Marsden BJ, Shaw GS and Sykes BD (1990) Calcium binding proteins. Elucidating the contributions to calcium affinity from an analysis of species variants and peptide fragments *Biochem. Cell Biol.* **68,** 587-601

Murre C, Bain G, Dijk MAv, Engel I, Furnari BA, Massari ME, Matthews JR, Quong MW, Rivera RR and Stuiver MH (1994) Structure and function of helix-loop-helix proteins *Biochem. Biophys. Acta* **1218,** 129-135

Nambu JR, Murre C, McCaw PS and Baltimore D (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc Proteins *Cell* **56,** 777-783

Nikoloff DM, McGraw P and henry SA (1992) The INO2 gene of sachromyces cerevisiae encodes a helix-loop-helix protein that is required for activation of phospholipid synthesis *Nucleic Acids Res.* **20**, 3253-3258

Penn LJZ, Laufer EM and Land H (1990) C-myc: evidence for multiple regulatory functions *Semin. Cancer Biol.* **1**, 69-80

Poortinga G, Watanabe M and Parkhurs SM (1998) Drosophila CtBP: a Hairy-interacting protein required for embryonic segmentation and Hairy-mediated transcriptional repression *EMBO J* **17,** 2067-2078

Popescu RA, Lohri A, de Kant E, Thiede C, Reuter J, Herrmann R, and Rochlits CF (1998) bcl-2 expression is reciprocal to p53 and c-myc expression in metastatic human colorectal cancer *Eur J Cancer* **34,** 1268-1273

Rossner MJ, Dorr J, Gass P, Schwab MH and Nave KA (1997) SHARPs: mammalian Enhancer-of-Split- and Hairy-related proteins coupled to neuronal stimulation *Mol. Cell. Neuro.* **9**, 460-475

Sasai Y, Kageyama R, Tagawa Y, Shigemoto R and Nakanishi S (1992) Two mammalian helix loop helix factors structurally related to *Drosophila hairy* and *enhancer-of split* Genes *Dev* **6,** 2620-2634

Sewalt RG, Gunster MJ, van der Vlag J, Satijn DP and Otte AP (1999) C-terminal binding protein is a transcriptional repressor that interacts with a specific class of vertebrate polycomb proteins *Mol. Cell. Biol.* **19,** 777-87

Shen M, Kawamoto T, Yan W, Nakamasu K, Tamagami M, Koyano Y, Noshiro M and Kato Y (1997) Molecular characterization of the novel basic helix-loop-helix protein DEC1 expressed in differentiated human embryo chondrocytes *Biochem. Biophys. Res. Commun.* **159**, 1006-1011

Steven T and Sun XH (1998) The Tal1 oncoprotein inhibits E47-mediated transcription *J. Biol. Chem.* **273**, 7030-7037

Sun XH and Bltimore D (1991) An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers *Cell* **64**, 459-470

Voronova A and Baltimore D (1990) Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains *Proc. Natl. Acad. Sci. USA* **87**, 4722-4726

Wang YP, Schnegelsberg PN, Dausman J and Jaenisch (1996) Functional redundancy of muscle-specific transcription factors Myf5 and myogenin *Nature*, **379**, 823-825

Weintraub H, Dwarki VJ, Verma I, Davis R, Holenberg S, Snider L, Lassar A and Tapscott SJ (1991) Muscle-specific transcription activation by MyoD *Proc. Natl. Acad. Sci. USA* **86**, 5434-5438

Weintraub H (1993) The MyoD family and myogenesis: redundancy, network, and thresholds *Cell* **75**, 1241-1244

Yan W, Young AZ, Soares VC, Kelley R, Benezra R and Zhuang Y (1997) High incidence of T-cell tumors in E2A-null mice and E2A/Id1 double-knockout mice *Mol. Cell. Biol.* **17**, 7317-7327

Zhang H and Levine M (1999) Groucho and dCtBP mediate separate pathways of transcriptional repression in the Drosophila embryo Proc. *Natl. Acad. Sci. USA* **96,** 535-540

Zhao GQ, Zhao Q, Zhou X, Mattei MG and de Crombrugghe (1993) TFEC, a basic helix-loop-helix protein, forms heterodimers with TFE3 and inhibits TFE3-dependent transcription activation *Mol. Cell Biol.* **13**, 4505-4512

APPENDIX **ISOLATION OF MULTIPLE DISTINCT cDNAs BY A SINGLE CDNA-TRAPPING PROCEDURE**

Maowen Hu and Bingfang Yan

(Published in 1999, *Analytical Biochemistry* **266**, 233-235)

Screening of a cDNA library is primarily conducted with nuclei acid hybridization or antibody staining. These procedures usually involve library plating, colony transferring to a membrane, and clone identification. Because of the high density at which clones are plated, isolation of pure clones requires secondary and multiple screening at decreasing alone densities. In addition, screening of a meaningful number (10^5-10^6) of cloned may take weeks or even months (Sambrook et al., 1989).

Recently a cDNA trapping method was developed (Li et al., 1995). In this method, a complex population of double-stranded phagemid DNA is isolated from a cDNA library and converted to single-stranded DNA (ssDNA). An oligonucleotide complementary to a segment of the target cDNA is synthesized and biotinylated. The biotin-labeled oligonucleotide is then subjected to hybridization with the ssDNAs in solution, and the duplexes are captured on streptavidin-coated paramagnetic bead. The captured ssDNAs are then released from the beads, repaired with a polymerase, and transformed into bacteria. This

method, compared with nucleic acid hybridization and antibody staining, can screen a large number of clones and isolate full-length cDNA within days.

This report describes a modified cDNA trapping procedure, in which several oligonucleotides and ssDNAs from more than one library were included in a single captured reaction. These oligonucleotides were targeted to distinct cDNAs. As a result, full-length cDNAs encoding different proteins were isolated by a single trapping procedure. Using this method, we successfully isolated several cDNAs encoding different human cytochrome P450 enzymes and multiple forms of kinase, which phosphorylate the inositol ring. This modification significantly reduces cost and time, which is particularly useful for isolating multiple forms of cDNAs.

Materials and methods. GeneTrapper cDNA kit, human liver and mouse brain cDNA libraries, and DH10B *Escherichia coli* (*E. coli*) were purchased from GibcoBRL (Gaithersburg, MD). All oligonucleotides were synthesized by Genemed Synthesis Inc. (South San Franciscon, CA), and their sequences are listed in Table 1. Unless otherwise indicated, all other reagents were purchased from Sigma (St. Louis, MO).

The phagemid libraries were grown in 100 ml Luria broth containing 100 μg/ml ampicillin with 1 x 106 independent clones to saturation at 30°C (~16h). The bacterial cells were then centrifuged at 4800g for 15 min at 4°C. The pellet was

resuspended in 10 ml buffer I containing 15 mM Tris-HCl (pH8.0), 10 mM EDTA, RNase A (100 μg/ml), and RNase T1 (1200 units/ml). Bacterial cells were lysed by adding 10 ml buffer II containing 0.2 M NaOH and 1% SDS. The solution was then neutralized by adding 10 ml of ice-cold 7.5 M NH4Oac, mixed gently by inverting the tubes five times, and incubated on ice for 10 min. The supernatant was prepared by centrifuging at 3000g for 15 min at 4C and filtered through cheesecloth. Phagemid DNA was then precipitated by adding an equal amount of cold isopropanol and centrifuging at 5000g for 15 min at 4C. The DNA pellet was resuspended in 1 ml buffer I and centrifuged at 12,000g for 1 min at 4C. The DNA sample was incubated at 37C for 15 min and extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) four times. The phagemid DNA in the aqueous phase was mixed with an equal volume of isopropanol followed by centrifugation at 14,000g for 15 min at 4°C. The pellet was then washed once with 500 μl of 70% ethanol and left to air dry for 15 min at room temperature. The DNA was then dissolved in 200 μl TE buffer [10 mM Tris-HCl (pH8.0), 1 mM EDTA] and stored at –29°C until use.

Two groups with a total of six oligonucleotides were synthesized (Table1): one group (P450 group) was targeted to human cytochrome P4502C8, P4502E1, and P45 reductase, respectively; and the other group (kinase group) was targeted to mouse type I phosphatidylinoditol 4-phosphate 5-kinases (PI4P5 kinase), α and β and human inositol 1,3,4-trisphosphate 5/6-kinase [Ins (1,3,4) P3 5/6-kinase], respectively. Oligonucleotides were purified by 16% polyacrylamide gel

electrophoresis (Vorndam et al., 1986). The oligonucleotides (1 μg each) in the same group were then mixed and biotinylated in a total volume of 25 μl with biotin-14-dCTP and terminal deoxynucleotidyl transferase as described by the manufacturer.

Double-stranded phagemid DNA was converted to ssDNA by a sequential digestion with Gene II and ExoIII, and the ssDNA was denatured at 95C for 1 min and chilled in ice for 1 min. Hybridization between the ssDNA and biotinylated oligonucleotides was conducted at 37°C for 1 h with the hybridization buffer provided by the manufacturer. One hybridization reaction was conducted with the biotinylated oligonucleotides (20 ng) from the P450 group and the ssDNAs (2.5 μg) from the human liver library, whereas the other hybridization was conducted with the biotinylated oligonucleotides (20 ng) from the kinase group and a mixture of the ssDNAs from both the human liver and mouse brain libraries (1.25 μg each). The hybridized ssDNAs were repaired by a thermostable polymerase. These procedures were conducted as described by the manufacturer.

The repaired captured DNAs (~20% of total) were used to transform DH10B *E. coli* (23 μl) with a Cell Porator electroporation system at a field strength of 16.6 kV/cm and a pulse length of 4 ms (GibcoBRL, Gaithersburg, MD). The electroporated cells were incubated at 37°C overnight. Individual colonies were grown in 5 ml of Lauri broth, the phagemid DNA was isolated with a QIAprep

Spin Miniprep kit (QIAGEN Inc., Santa Clarita, CA) and sequences of the captured DNAs were determined with an EXCEL II DNA sequencing kit from Epicentre Technologies (Madison, WI) as described previously (Yan et al., 1995).

Results and discussion. Oligonucleotides in the P450 group were targeted to human cytochrome P4502C8, P4502E1, and P450 reductase, respectively. They were designed to anneal to the 5'-coding region of each cDNA. The trapping experiment with these oligonucleotides results in the isolation of ~500 individual clones. From those, 32 individual clones were cultured and the corresponding phagemid DNAs were isolated. Sequencing of the 5'-end (~350 bases long) demonstrated that they all contained full-length cDNA inserts. Among them, 6 clones contained the cDNA encoding P4502C8, 4 clones contained the cDNA encoding P4502C18, and 16 clones contained the cDNA encoding P4502E1, as shown in Table 2. Therefore, half of the sequenced clones contained the cDNA encoding P4502E1.

The P4502C subfamily has four members: they are used in this experiment for this subfamily matched the sequence of 2C8 perfectly, but did not match the sequence of 2C9, 2C18, or 2C19 by two bases. As expected, the trapping experiment isolated the cDNA encoding 2C8. However, surprisingly the trapping experiment also isolated the cDNA encoding 2C18 but not 2C9 and 2C19. The reason for this phenomenon was probably due to the relative abundance of the cDNA in this library. Namely, the cDNA encoding 2C18 was more abundant in

this library than those encoding either 2C9 or 2C19. Similarly, a relatively high abundance of 2E1 cDNA in the library was probably the reason that 50% of the clones sequenced in the P450 group contained the cDNA encoding this enzyme. Compared with other P45 enzymes, P4502E1 is not a very abundant enzyme in normal liver (only ~9% of total P450). However, the level of P4502E1 can be significantly increased by environmental factors such as alcohol consumption and disease status such as diabetes. It is likely that the individual whose liver mRNA was used for the constriction of this library was previously exposed to one or more of these factors.

In contrast, the trapping experiment with the oligonucleotides in kinase group yielded only 31 individual clones. As shown in Table 3, 2 of these clones encoded mouse type I PI4/5-kinase α; 6 of them encoded mouse type I PI4/5-kinase β; 4 of them encoded a recently identified type I PI4P5-kinase, termed γ; and 7 of them encoded human Ins (1,3,4) P3 5/6-kinase. The rest of these clones encoded other proteins. Form γ exhibits a high degree of sequence identity (20 of 21 bases) with form β in the region where the oligonucleotide was derived. Thus, the cDNA encoding form γ was probably captured by the oligonucleotide for form β. It should be noted that the human liver and mouse brain libraries were constructed with vectors that have different flanking sequences at the polylinker region, which was used to establish by the polymerase chain reaction which library was the source from which a cDNA was isolated.

Among the cDNAs encoding other proteins in the kinase group, half coded for vitamin D-binding protein (Table 3), and the rest apparently coded for yet unidentified proteins based on a BLST search in Gene Bank. These unidentified proteins may actually share a significant sequence identity with the oligonucleotides used for the trapping experiment (only a ~350-bp sequence in the 5' end was determined), which probably contributed to their isolated. In contrast, the cDNA encoding vitamin D-binding protein shows no significant sequence identity with the oligonucleotides; the reason for its isolation remains to be determined. It was likely that the abundance of the cDNAs coding for the kinases to be isolated in these libraries was so low that an excess of the labeled oligonucleotides was available to hybridize with unrelated cDNAs and yields false positives. This notion if supported by the fact that all of the sequenced clones in the P450 trapping experiment resulted in the isolation of target cDNAs, because P450 enzymes, the primary phase I drug-metabolizing enzymes, are very abundant in the liver (Shephard et al., 1992).

We report here a modified cDNA trapping technique that can accommodate multiple oligonucleotides and more than one library at the same time. This method has all the features of the original method such as screening of a large number of clones within a week, rapid isolation of full-length cDNA, and identification of related sequences. This modified method, compared with the original method, saves time and significantly reduces the cost, as the reagent

used for cDNA trapping are expensive. This is particularly effective for isolating multiple forms of cDNAs.

Table 5. Sequence of oligonucleotides for cDNA trapping

Enzyme	Sequence	Reference
Human cytochrome P4502C8	5'-gtc tat ggt cct gtg ttc acc-3'	5
Human cytochrome P4502E1	5'-gtg gtg atg cac ggc tac aag-3'	11
Human cytochrome P450 reductase	5'-cag cat gac ggc cat gat tct-3'	12
Mouse PI-4-phosphat 5-kinase α	5'-caa tgg gag gca ttc cag cta-3'	8
Mouse PI-4-phosphate 5-kinase β	5'-caa gac cta tgc acc tgt tgc-3'	9
Inositol 1,3,4-triphosphate 5/6-kinase	5'-atc atc cac aag ctg act gac-3'	10

Table 6. cDNA trapping of human cytochrome P450 enzymes and reductase

Enzyme	Number of positive clones	% of clones sequenced
P4502C8	6	19
P4502C18	4	13
P4502E1	16	50
P450 reductase	6	18

Table 7. cDNA trapping of PI-4-phospate 5-kinse and inositol 1,3,4-triphosphate 5/6-kinase

Enzyme	Number of positive clones	% of clones sequenced
Mouse PI-4-phosphat 5-kinase α[a]	2	6
Mouse PI-4-phosphate 5-kinase β	6	20
Mouse PI-4-phosphate 5-kinase γ	4	13
Inositol 1,3,4-triphosphate 5/6-kinase	7	23
Others	12	38

BIBLIOGRAPHY

Aplan PD, Lombardi DP, Reaman GH, Sather HN, Hammond GD and Kiysch IR (1992) Involvement of the putative hematopoietic transcription factor SCL in T-cell acute lymphoblastic leukemia *Blood* **79**, 1327-1333

Aplan PD, Jones CA, Chervinsky DS, Zhao X, Ellsworth M, Wu C, McGuire EA and Gross KW (1993) An scl gene product lacking the transactivation domain induces biny abnormalities and c0orperates with LMO1 to generate T-cell malignances in transgenic mice *European Molecular Biology organization Journal* **16**, 2408-2419

Atchley WR and Fitch WM (1997) A natural classification of the basic helix-loop-helix class of transcription factors *Proceedings of the National Academy of Science, USA* **94,** 5172-5176

Bacsi SG, Reisz Porszas S and Hankinson O (1995) Orientation of the heterodimeric aryl hydrocarbon (dioxin) receptor complex on its asymmetric DNA recognition sequence *Molecular Pharmacology* **47**, 432-438

Bash RO, Hall S, Timmons CF, Crist WM, Amylon M, Smith RG and Bear R (1995) Does activation of the TAL1 gene occur in a majority of patients with T-cell acute lymphoblastic leukemia? *Blood* **86,** 666-676

Benezra R, Davis RL, Lockshon D, Turner DL and Weintraub H (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins *Cell* **61**, 49-59

Bernard M, Delabesse E, Novault S, Hermine O and Macintyre EA (1998 Antiapoptic effect of ectopic TAL/SCL expression in a human leukemic T-cell line *Cancer Research* **58**, 2680-2687

Burley SK (1994) DNA-binding motif from eukaryotic transcription factors *Current Opinion in Cell Biology* **4**, 3-11

Cabrera CV and Alonso MC (1991) Transcriptional activation by heterodimers of the achaete-scute and daughterless gene products of *Drosophila European Molecular Biology organization Journal* **10,** 2965-2973

Chin L. Schreiber Agus N, Pellicer I, Chen K, Lee HW, Dudast M, Cordon Cardo C and DePinho RA (1995) Contrasting roles for Myc and mad proteins in cellular growth and differentiation *Proceedings of the National Academy of Science, USA* **92,** 8488-8492

Condorelli GL, Tocci A, Botta R, Facchiano F, Testa U, Vitelli L, Valteis M, Croce CM and Peschle C. (1997) Ectopic *TAL-1/SCL* expression in phenotypically normal or leukemic myeloid precursors: Proliferative and antiapoptotic effects

coupled with a differentiation blockade *Molecular and Cellular Biology* **17,** 2954-2969

Davis AC, Wims M, Spotts GD, Hann SR and Bradley A (1993) A null-c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice *Genes and Development* **7,** 671-82

Davis RL, Cheng P-F, Lassar AB, and Weintraub H (1990) The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation *Cell* **60**, 733-746

Dawson SR, Turner DL, Weintraub H and Parkhurst SM (1995) Specificity of the Hairy/Enhancer of split basic helix-loop-helix (bHLH) proteins maps outside the bHLH domain and suggested two separable modes of transcripitonal repression *Molecular and Cellular Biology* **15**, 6923-6931

French BA, Chow KL, Olson EN and Schwartz RJ (1991) Heterodimers of myogenic helix-loop-helix regulatory factors and E12 bind a complex element governing myogenic induction of avian cardiac alpha-actin promoter *Molecular and Cellular Biology* **11**, 2439-2450

Gelboin HV, Park SS, Yang CS and Gonzalez FJ (1996) *The Journal of Biological Chemistry* **261**, 16689-16697

Goldfarb AN, Lewandowska K, and Pennell CA (1998) Identification of highly conserved module in E proteins required for in vivo helix-loop-helix dimerization *The Journal of Biological Chemistry* **273,** 2866-2873

Goodrich J, Carpenter R and Coen ES (1992) A common gene regulates pigmentation pattern in diverse plant species *Cell* **68**, 955-964

Gred C, Umbenhauer DR, Bellew TM, Bork RW, Srivastava PK, Shinriki N Lloyd RS and Guengerich FP (1988) *Biochemistry* **27**, 6929-6940

Guillemot F, Lo L-C, Johnson JE, Auerbach A, Anderson DJ and Joyner AL (1993) Mammalian *achaete-scute* homologue 1 is required for the early development of olfactory and autonomic neurons *Cell* **75**, 463-476

Hirose K, Morita M, Ema M, Mimura J, Hamada H, Fujii H, Saijo Y, Gotoh O, Sogawa K and Fujii-Kuriyama Y (1996) cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt) Molecular *and Cellular Biology* **16**, 1706-1713

Hsu HL, Wadman I and Baer R (1994) Formation of in vivo complex between the TAL1 and E2A polypeptides of leukemic T cells *Proceedings of the National Academy of Science, USA* **91**, 5947-5951

Ishibashi M, Ang S-L, Shiota K, Nakanishi S, Kageyama R and Guillemot F (1995) Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES1) leads to up-regulation of neuronal helix-loop-helix factors, premature neurogenesis, and sever neural tube defects *Genes and Development* **9,**3136-48

Ishihara H, Shibasaki Y, Kizuki N, Wada T, Katagiri H, Yazaki Y, Asano T and Oka Y (1996) *The Journal of Biological Chemistry* **217**, 11904-11910

Issack PS and Ziff EB (1998) Altered expression of helix-loop-helix transcriptional regulators and cyclin D1 in Wnt-1-PC12 cells *Cell Growth Differentiation* **9**, 837-845

Jarman AP, Grell EHAckerman L, Jan LY and Jan YN (1994) Atonal is the proneural gene for Drosophila photoreceptors *Nature* **369**, 398-400

Land H, Chen AC, Morgenstern JP, Parada LF and Weinberg RA (1986) Behavior of myc and ras oncogenes in transformation of rat embryo fibroblasts *Molecular and Cellular Biology* **6,** 1917-1925

Leroy-Viard K, Vinit MA, Lecointe N, Jouault H, Hibner U, Romeo PH and Matieu-Mahul D (1995) Loss of TAL-1 protein activity induces premature apoptosis of Jukat leukemic T cells upon medium depletion. Loss of TAL-1 protein activity induces premature apoptosis of Jurkat leukemic T cells upon

medium depletion *Proceedings of the National Academy of Science, USA* **14**, 2341-2349

Li WB, Gruber CE, Noon MC, Polayes D, Schmidt B and Jessee J. (1995) *Focus* **17**,15

Lindebro MC, Poellinger L and Whitelaw ML (1995) Protein-protein interaction via PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex *European Molecular Biology organization Journal* **14**, 3528-3539

Little TD and Evan GI (1998) Helix-loop-helix transcription factors, *Oxford University Press*, PP1-48

Luo X and Sawadogo (1996) Antiproliferative properties of the USF family of helix-loop-helix transcription factors *Proceedings of the National Academy of Science, USA* **93**, 1308-1313

Marx J (1993) Cell death studies yield cancer clues *Science* **259**, 760-761

Murre C, Bain G, Dijk MAv, Engel I, Furnari BA, Massari ME, Matthews JR, Quong MW, Rivera RR and Stuiver MH (1994) Structure and function of helix-loop-helix proteins *Biochemica Biophysica Acta* **1218**, 129-135

Nambu JR, Murre C, McCaw PS and Baltimore D (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc Proteins *Cell* **56**, 777-783

Nikoloff DM, McGraw P and henry SA (1992) The INO2 gene of sachromyces cerevisiae encodes a helix-loop-helix protein that is required for activation of phospholipid synthesis *Nucleic Acids Research* **20**, 3253-3258

Penn LJZ, Laufer EM and Land H (1990) C-myc: evidence for multiple regulatory functions *Seminars in Cancer Biology* **1**, 69-80

Poortinga G, Watanabe M and Parkhurs SM (1998) Drosophila CtBP: a Hairy-interacting protein required for embryonic segmentation and Hairy-mediated transcriptional repression *European Molecular Biology Organization Journal* **17**, 2067-2078

Popescu RA, Lohri A, de Kant E, Thiede C, Reuter J, Herrmann R, and Rochlits CF (1998) bcl-2 expression is reciprocal to p53 and c-myc expression in metastatic human colorectal cancer *European Journal of Cancer* **34**, 1268-1273

Romkes m, Faletto MB, Blaisdell JA Raucy JL and Goldstein JA (1991) *Biochemistry* **27**, 3247-3255

Rossner MJ, Dorr J, Gass P, Schwab MH and Nave KA (1997) SHARPs: mammalian Enhancer-of-Split- and Hairy-related proteins coupled to neuronal stimulation *Molecular and Cellular Neuroscience* **9**, 460-475

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular cloning: A laboratory manual, 2nd ed., pp. 8.3-8.82, *Cold spring Harbor Laboratory Press, Cold Spring Harbor, NY*

Sasai Y, Kageyama R, Tagawa Y, Shigemoto R and Nakanishi S (1992) Two mammalian helix loop helix factors structurally related to *Drosophila hairy* and *enhancer-of split* Genes and *Development* **6**, 2620-2634

Sewalt RG, Gunster MJ, van der Vlag J, Satijn DP and Otte AP (1999) C-terminal binding protein is a transcriptional repressor that interacts with a specific class of vertebrate polycomb proteins *Molecular and Cellular Biology* **19,** 777-87

Shen M, Kawamoto T, Yan W, Nakamasu K, Tamagami M, Koyano Y, Noshiro M and Kato Y (1997) Molecular characterization of the novel basic helix-loop-helix protein DEC1 expressed in differentiated human embryo chondrocytes *Biochemical and Biophysical Research Communications* **159**, 1006-1011

Shephard AE, Palmer CAN, Segall HJ and Philips IR (1992) *Archives of Biochemistry and Biophysics* **294**, 168-172

Steven T and Sun XH (1998) The Tal1 oncoprotein inhibits E47-mediated transcription *The journal of Biological Chemistry* **273**, 7030-7037

Sun XH and Bltimore D (1991) An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers *Cell* **64**, 459-470

Vorndam AV and Kerschnner J (1986) *Analytical Biochemistry* **152**, 2 21-225

Voronova A and Baltimore D (1990) Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains *Proceedings of the National Academy of Science, USA* **87,** 4722-4726

Wang YP, Schnegelsberg PN, Dausman J and Jaenisch (1996) Functional redundancy of muscle-specific transcription factors Myf5 and myogenin *Nature* **379**, 823-825

Weintraub H, Dwarki VJ, Verma I, Davis R, Holenberg S, Snider L, Lassar A and Tapscott SJ (1991) Muscle-specific transcription activation by MyoD *Proceedings of the National Academy of Science, USA* **86**, 5434-5438

Weintraub H (1993) The MyoD family and myogenesis: redundancy, network, and thresholds *Cell* **75**, 1241-1244

Wilson M and Majerus PW (1996) Isolation of inositol 1,3,4-trisphosphate 5/6-kinase, cDNA cloning, and expression of the recombinant enzyme *The journal of Biological Chemistry* **271,** 11904-11910

Yan B, Yang D, Bullock P and Parkinson A (1995) Rat Serum Carboxylesterase *The journal of Biological Chemistry* **270**,19128-19134

Yan W, Young AZ, Soares VC, Kelley R, Benezra R and Zhuang Y (1997) High incidence of T-cell tumors in E2A-null mice and E2A/Id1 double-knockout mice *Molecular and Cellular Biology* **17,** 7317-7327

Zhang H and Levine M (1999) Groucho and dCtBP mediate separate pathways of transcriptional repression in the Drosophila embryo *Proceedings of the National Academy of Science, USA* **96,** 535-540

Zhao GQ, Zhao Q, Zhou X, Mattei MG and de Crombrugghe (1993) TFEC, a basic helix-loop-helix protein, forms heterodimers with TFE3 and inhibits TFE3-dependent transcription activation *Molecular and Cellular Biology* **13**, 4505-4512