Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 352 (2005) 45 – 56

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Comparative genomics of the *Hlx* homeobox gene and protein: Conservation of structure and expression from fish to mammals

Michael D. Bates[a,b,*], James M. Wells[b], Byrappa Venkatesh[c]

[a]*Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA*
[b]*Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA*
[c]*Institute of Molecular and Cell Biology, 61 Biopolis Drive, Singapore 138673*

Received 18 November 2004; received in revised form 8 February 2005; accepted 1 March 2005

Received by S. Yokoyama

Abstract

Hlx is a homeobox transcription factor gene that is expressed in intestinal and hepatic mesenchyme of the developing mouse embryo and is essential for normal intestinal and hepatic development. Because of the morphological and molecular similarities in the development of the digestive system across species, we hypothesized that the *Hlx* gene and protein sequences and expression patterns would be conserved among vertebrates. Comparison of the *Hlx* gene orthologues of human, chimpanzee, mouse, rat, pufferfish (Fugu) and zebrafish demonstrates that these six genes share an identical organization with four exons and three introns. Comparison of the inferred Hlx protein sequences from these and three additional species (chick, Spanish ribbed newt and rainbow trout) reveals significant sequence identity, with identical homeodomains. The expression of *Hlx* in the mesenchyme of developing chick embryos is highly similar to that of mouse. Fugu *Hlx* is expressed in a tissue-specific manner that is similar though not identical to that of mouse, suggesting a conservation of Hlx function between mammals and birds. The mammalian and fish *Hlx* genes share a putative 5′ upstream enhancer as well as an inverted repeat containing CCAAT boxes on opposite strands that we have previously shown to be important for mouse *Hlx* gene expression. These results suggest that the function of Hlx and the mechanisms regulating its expression are highly conserved in mammals, birds, amphibians and fish.

Keywords: Transcription factor; Developmental gene expression regulation; Genome; Mouse; *Takifugu*

1. Introduction

Among the important participants in developmental programming are members of the family of transcription factors encoded by the homeobox genes. Initially identified in *Drosophila*, homeodomain transcription factors share a well-conserved, 60-amino acid DNA-binding homeodomain. The various homeodomain proteins are expressed in discrete cell types at specific times during development and in adult tissues. They often occupy high-level positions in the genetic hierarchy of development, in that the expression of a homeobox gene initiates a genetic pathway or cascade that regulates cell differentiation and/or proliferation. Mutations of homeobox genes cause or are associated with morphological and developmental anomalies in *Drosophila*, mice and humans. The discovery of such master regulatory transcription factors in development raises questions about how their expression is controlled (upstream regulation) and what genes the encoded transcription factors in turn regulate (downstream targets). Inter-species comparisons of gene and

Abbreviations: kb, kilobases; bp, base pairs; min, minutes; E, embryonic day; RT-PCR, reverse transcription–polymerase chain reaction; Mb, megabases.

* Corresponding author. Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA. Tel.: +1 513 636 4415; fax: +1 513 636 5581.

E-mail address: michael.bates@cchmc.org (M.D. Bates).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.03.021

protein sequences and expression patterns can shed important light on whether specific developmental processes are conserved across species.

Hlx is a homeobox gene that was originally isolated from a mouse pre-B-lymphocyte cell line (Allen et al., 1991). It has homology to the *Drosophila* homeobox gene *H2.0* (Barad et al., 1988) that is expressed in visceral mesenchyme but is not essential for gut morphogenesis (Barad et al., 1991). The mouse and human Hlx proteins are highly conserved, and their homeodomains are 100% identical (Kennedy et al., 1994). In adult mice, *Hlx* is expressed most abundantly in lung, heart, skeletal muscle and hematopoietic tissues and cells, and less abundantly in intestine (small and large), liver, uterus and ovaries; there was also trace expression detected in stomach, brain, kidney and testes by RNase protection assay (Allen et al., 1991). In mouse embryos, *Hlx* is most prominently expressed in mesodermal tissues, in particular in the midgut and hindgut mesenchyme of developing liver, gallbladder and intestine, where it is first detected by in situ hybridization at embryonic day (E) 9.5 (Lints et al., 1996). Targeted disruption of the mouse *Hlx* gene results in an autosomal-recessive embryonic lethal phenotype in which the intestine and liver fail to grow, although this mutation has no effect on the ability of embryonic hematopoietic cells to reconstitute the hematopoietic systems of irradiated mice (Hentsch et al., 1996). Hlx has more recently been shown to interact with the T-box transcription factor T-bet to induce helper T-lymphocyte type 1-specific gene expression (Mullen et al., 2002; Zheng et al., 2004). Thus, the Hlx transcription factor plays diverse roles in a variety of tissues.

Recent genome sequencing efforts have provided the opportunity to gain additional insights into the structure and function of the *Hlx* gene and protein through cross-species comparisons of genomic data. In addition to the human and mouse genomes that have essentially been completely sequenced (for references, see Ureta-Vidal et al., 2003), progress is being made to complete the genomic sequence of a number of other species. Additional gene sequencing efforts have focused on species that are important model systems for biological, developmental or medical studies, such as the worm *Caenorhabditis elegans*, the fruitfly *Drosophila melanogaster*, the mosquito *Anopheles gambiae*, the pufferfish *Takifugu rubripes* (Fugu) and the zebrafish *Danio rerio* (for references, see Ureta-Vidal et al., 2003). The Fugu genome is of particular interest because, although it contains approximately as many genes as mammals, it is one-eighth the size of mammalian genomes (Venkatesh et al., 2000). This smaller size means that regulatory sequences are more tightly spaced in the non-coding regions of the genome.

Because of the morphological and molecular similarities in the development of the digestive system among vertebrates, we hypothesized that the *Hlx* gene and encoded protein sequences and expression pattern would be conserved. In this study, we have taken advantage of new genomic information to compare the gene structure and encoded protein sequences of *Hlx* from human and mouse with those from seven other species: chimpanzee, rat, chick, newt, Fugu, zebrafish and rainbow trout. We find that the *Hlx* gene and protein sequence is strongly conserved among vertebrates, including complete conservation of the DNA-binding homeodomain. Expression in chick and Fugu is highly similar to that previously reported for mouse (Lints et al., 1996). The similarities and differences observed provide insights into *Hlx* gene and protein structure and function.

2. Materials and methods

2.1. DNA and protein sequences

Human and mouse *Hlx* genomic and cDNA sequences have previously been reported (Kennedy et al., 1994; Allen et al., 1991; Bates et al., 2000). Table 1 lists accession

Table 1
Accession numbers and related information for *Hlx* genomic and cDNA sequences from nine vertebrate species and the *Drosophila* gene *H2.0*

Species	GenBank	Ensembl build	Chromosome	Contig	Nucleotides (gene±10 kb)	Ensembl peptide
Human	NM_021958	12.31.1	1	AL445423.13.1.177941	216886181–216911839	ENSP00000259148
Chimp	AADA01226078	1	–	AADA012260781	7994–3141[a]	–
Mouse	AF172318	12.3.1	1	CAAA01218972.1.1.9634, CAAA01110281.1.1.1286	185016742–185042733	ENSMUSP00000040505
Rat	–	12.2.1	13	RNOR01034700	95029139–95003612	ENSRNOP00000003155
Fugu	–	12.2.1		Chr_scaffold_316	137163–114922	SINFRUP00000157884
Zebrafish	–	16.2.1	?	ctg10680	4219–7478[a]	ENSDARP00000004527
Chick	BM489797[b]					
Newt	AF106694[b]					
Trout	[c]					
Drosophila	NM_078764[b]					

[a] Predicted coding region only; contig does not include 10 kb in each direction.
[b] GenBank accession number is for the cDNA sequence.
[c] Rainbow Trout Gene Index tentative consensus sequence TC29346 (see Section 2.1 for details).