Genomics 96 (2010) 342–351

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Conserved proximal promoter elements control repulsive guidance molecule c/hemojuvelin (Hfe2) gene transcription in skeletal muscle

Christopher J. Severyn, Peter Rotwein *

Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, OR 97239-3098, USA

ARTICLE INFO

Article history:
Received 12 June 2010
Accepted 15 September 2010
Available online 19 September 2010

Keywords:
Repulsive guidance molecule (RGM)
RGMc
Hemojuvelin
Juvenile hemochromatosis
Transcriptional regulation
Molecular evolution

ABSTRACT

Repulsive guidance molecule c (RGMc; gene symbol: Hfe2) plays a critical role in iron metabolism. Inactivating mutations cause juvenile hemochromatosis, a severe iron overload disorder. Understanding mechanisms controlling RGMc biosynthesis has been hampered by minimal information about the RGMc gene. Here we define the structure, examine the evolution, and establish mechanisms of regulation of the mouse RGMc gene. RGMc is a 4-exon gene that undergoes alternative RNA splicing to yield 3 mRNAs with 5′ different untranslated regions. Gene transcription is induced during myoblast differentiation, producing all 3 mRNAs. We identify 3 critical promoter elements responsible for transcriptional activation in skeletal muscle, comprising paired E-boxes, a putative Stat and/or Ets element, and a MEF2 site, and muscle transcription factors myogenin and MEF2C stimulate RGMc promoter function in non-muscle cells. As these elements are conserved in RGMc genes from multiple species, our results suggest that RGMc has been a muscle-enriched gene throughout its evolutionary history.

1. Introduction

Iron-related metabolic and hematologic disorders affect millions of individuals worldwide. Iron plays a critical role in numerous cellular processes ranging from oxygen exchange and energy metabolism [1] to nucleic acid synthesis and DNA repair [2], yet too much or too little iron can cause severe tissue and organ damage [3]. As a result, iron levels are tightly regulated in humans and other mammalian species [4], with primary control being exerted at the level of absorption from the small intestine [4,5]. Hemojuvelin (HJV) or repulsive guidance molecule c (RGMc; approved gene symbol: Hfe2) is a recently identified gene that was initially linked to systemic iron metabolism by the discovery that mutations in humans caused the rapidly progressive and severe iron overload disorder, juvenile hemochromatosis (JH) [6,7]. This relationship was strengthened when mice engineered to lack RGMc also developed iron overload [8,9]. As RGMc/HJV appears to indirectly regulate the expression of hepcidin (Hamp) [10–12], a peptide hormone made in the liver that negatively controls intestinal iron absorption in the duodenum [5,13], it is thus a component of a homeostatic pathway that regulates iron uptake [3,8–10,12].

RGMc was discovered not only as a gene mutated in JH [10], but also was identified as a novel transcript expressed during skeletal muscle differentiation [14], and as a member of a conserved three-gene family that receives its name from the axonal guidance molecule RGMa [15–18]. Unlike RGMa or RGMb, RGMc is not expressed in the central nervous system, but rather is produced by striated muscle and by hepatocytes in the liver [14,15,18]. During development RGMc transcripts are expressed first in the somites in both mice and zebrafish [14,17,19], and then later in skeletal muscle, as well as in the embryonic heart and liver [9,14]. This unique pattern of RGMc expression in striated myocytes and hepatocytes is maintained in the adult (Fig. 1A). To date, the responsible molecular mechanisms for tissue-specific gene expression have not been elucidated, and very little is known about RGMc gene regulation in any species, as no promoter has been characterized. Here we define the structure of the mouse RGMc gene and identify the DNA elements responsible for its high-level gene transcription in skeletal muscle. Further analysis reveals that these cis-acting muscle-specifying DNA elements are highly conserved in RGMc genes from multiple mammalian species, supporting the hypothesis that RGMc has been a muscle-enriched gene throughout its evolutionary history.

Abbreviations: Ad, adenovirus; BAC, Bacterial Artificial Chromosome; BCA, bicinchoninic acid; bHLH, basic helix-loop-helix; DM, Differentiation Media; DMEM, Dulbecco's modified Eagle's medium; DRB, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole; E, embryonic day; EST, expressed sequence tag; ETS, E26 avian retrovirus Transformation-Specific; FBS, fetal bovine serum; HJV, hemojuvelin; JH, juvenile hemochromatosis; MADS, MCM1 (mini-chromosome maintenance)–agamous–deficiens–serum response factor; MCK, creatine kinase-muscle; MEF2, myocyte enhancer factor 2; MHC, myosin heavy chain (myh3); MyoD, myogenic differentiation-1; Myog., myogenin; NCS, newborn calf serum; RACE, rapid amplification of cDNA ends; RGM, Repulsive Guidance Molecule; RLU, relative luciferase units; RT-PCR, reverse transcription polymerase chain reaction; STAT, signal transducers and activators of transcription; TFBS, transcription factor binding sites; TSS, transcription start site; UTR, untranslated region.

* Corresponding author. Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Mail code L224, 3181 SW Sam Jackson Park Road, Portland, OR 97239-3098, USA. Fax: +1 503 494 8393.
E-mail address: rotweinp@ohsu.edu (P. Rotwein).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.09.001

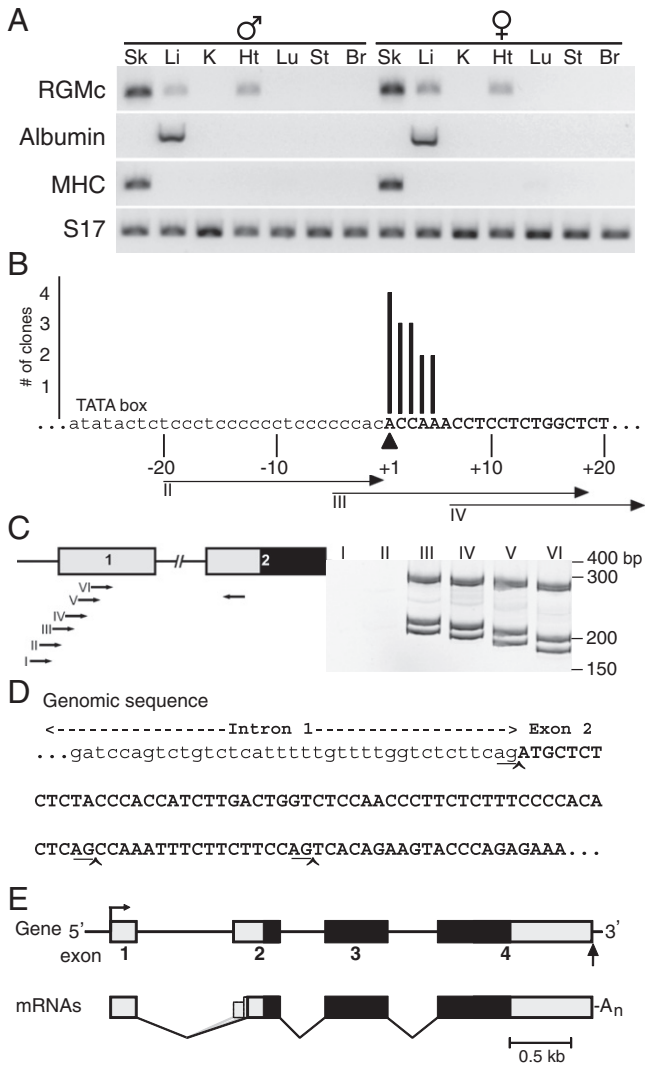

Fig. 1. Establishing mouse RGMc gene structure. A. RGMc mRNA is expressed in striated muscle and in the liver. Results of RT-PCR experiments for RGMc, albumin, skeletal muscle myosin heavy chain polypeptide 3 (MHC), and S17 mRNAs using RNA from tissues (Sk, skeletal muscle (gastrocnemius); Li, liver; K, kidney; Ht, heart; Lu, lung; St, stomach; Br, brain) of adult male (left) and female (right) mice. B. Mapping the 5′ end of the mouse RGMc gene by 5′ RACE using mouse skeletal muscle RNA. The number of clones is graphed on the *y-axis* above the corresponding location of the 5′ residue on the x-*axis*. The putative transcription start site is denoted as +1 (*arrowhead*), with exon 1 in *upper case* letters. A potential TATA box is labeled, and primers II - IV used in (C) are indicated below the sequence. C. Mapping the 5′ end of the mouse RGMc gene by RT-PCR with cDNA from mouse skeletal muscle RNA and overlapping PCR primers located in different parts of RGMc exon 1, as seen on the gene map to the left (see Supplemental Table 1 for DNA sequences of primers). Exons 1 and 2 are depicted as boxes, with the 5′ UTR in *gray* and the protein coding region in *black*, and introns and flanking DNA as horizontal lines. Results are seen to the right, and molecular weight markers are indicated (see Supplemental Fig. 1A for results with heart and liver RNA). In addition to mapping the 5′ end of exon 1, the results also show that alternative RNA splicing occurs between exons 1 and 2. D. DNA sequence of the junction between intron 1 and exon 2 of the mouse RGMc gene. Exon 2 is in *upper case* letters; the locations of alternative RNA splicing are noted by *chevrons*, with the –AG splice-acceptor residues *underlined*. E. Organization of the mouse RGMc gene and mRNAs. The gene contains 4 exons (*boxes*) and three introns (*thin lines*). The transcription start site is denoted as a *bent arrow*, and the polyadenylation site as a *vertical arrow*. The three RGMc mRNAs are diagramed below, and result from use of alternative splice acceptor sites at the 5′ end of exon 2.

2. Results

2.1. Defining RGMc gene structure

Analyses of genomic databases suggest that RGMc is a 4-exon gene in mice, humans, and several other species [20], but the 5′ end of exon 1 has not been established in any species, and the promoter has not been characterized. We thus mapped the transcription start site for mouse RGMc as a means to first identify and then functionally dissect the promoter. RGMc mRNA is expressed in adult male and female mouse skeletal and cardiac muscle and in the liver with no apparent gender differences in transcript abundance, at least in the C57Bl6 strain (Fig. 1A). By 5′ RACE we mapped the RGMc transcription start site (TSS) in skeletal muscle, and found that the 5′ end of 14/15 independent cDNA clones clustered within a 5-nucleotide region that was located ~25 nucleotides 3′ to a putative TATA box [21] in genomic DNA (Fig. 1B). We obtained similar results by RT-PCR with overlapping exon 1-specific primers and mouse muscle RNA (Fig. 1C), which validated the same TSS with RNA from mouse heart and liver (Supplemental Fig. 1).

The RT-PCR experiments designed to map the 5′ end of RGMc exon 1 used a common primer located in exon 2 (Fig. 1C), and results consistently yielded 3 distinct cDNAs that differed in length by 18 to 77 nucleotides (Fig. 1C and Supplemental Fig. 1). By DNA sequencing, all three classes of cDNAs contained identical parts of RGMc exon 1, but differed in the extent of exon 2 (all sequences matched mouse RGMc genomic DNA, Fig. 1D and Supplemental Fig. 1). We interpret these results to indicate that the mouse RGMc gene undergoes alternative RNA splicing to generate transcripts with varying lengths of exon 2, a hypothesis supported by evidence for splice acceptor sites at each of the three putative junctions between intron 1 and exon 2 (AG nucleotides underlined in Fig. 1D). Additional support comes from an expressed mouse sequence tag in GenBank (**AI196626**), which matches the intermediate-sized version of exon 2. A similar intermediate-sized exon 2 has been identified for human RGMc (EST numbers: **DA762328** and **DA764726**). Also, comparative analyses of 10 mammalian species reveals sufficiently similar genomic DNA sequences to suggest that alternative RNA splicing is a common feature of many RGMc genes (Supplemental Fig. 1). Taken together, our results show that mouse RGMc is a 4-exon gene with a discrete TSS in exon 1 and alternative RNA splicing involving exon 2 that leads to three distinct transcripts that vary in the length of the 5′ untranslated region (Fig. 1E).

2.2. RGMc gene transcription is induced during skeletal muscle differentiation

We next examined RGMc gene expression during skeletal muscle differentiation, using the C2 myoblast line as a model [22–24] (Fig. 2A). RGMc mRNA was detected within 12 hr after onset of C2 cell differentiation, and its abundance increased progressively during the subsequent 60 hr in a pattern similar to myogenin, a critical transcription factor that is expressed early in muscle differentiation [25,26] (Fig. 2B). Accumulation of RGMc mRNA in differentiating C2 myoblasts appeared to be a secondary to induction of RGMc gene transcription, as measured by stimulation of nascent nuclear RGMc RNA beginning at ~8 hr after addition of DM, a pattern that was temporally similar to myogenin gene (MYOG) activation (Fig. 2C). We also examined RGMc mRNA stability in myoblasts after 48 hr of differentiation [27], and found that RGMc is a moderately long-lived mRNA, with a half-life of ~5.2 hrs, more than twice that of myogenin, and nearly four times as long as MyoD (Fig. 2D). Taken together, results in Fig. 2 demonstrate that induction of RGMc gene transcription is a critical regulatory step responsible for accumulation of RGMc mRNA in differentiating muscle cells.

2.3. Analysis of RGMc promoter function in muscle differentiation

To investigate RGMc promoter function in myoblasts, we first showed that a genomic fragment containing ~4.1 kb of 5′ flanking DNA plus 118 nt of exon 1 could stimulate luciferase reporter activity in differentiating C2 myoblasts and in 10T½ mesenchymal stem cells