Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Identification and characterization of TGFβ-dependent and -independent *cis*-regulatory modules in the *C4ST-1*/*CHST11* locus

C.M. Willis[1,3], J.L. Wrana[2] and M. Klüppel[1,3]

[1]Department of Pediatrics, Integrated Graduate Program,
Northwestern University, Chicago, IL, USA
[2]Programme in Molecular Biology and Cancer,
Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada
[3]Department of Pediatrics, Feinberg School of Medicine,
Northwestern University, Human Molecular Genetics Program,
Children's Memorial Research Center, Chicago, IL, USA

Corresponding author: M. Klüppel
E-mail: m-kluppel@northwestern.edu

Genet. Mol. Res. 8 (4): 1331-1343 (2009)
Received August 17, 2009
Accepted September 4, 2009
Published November 3, 2009

ABSTRACT. Chondroitin-4-sulfotransferase-1(C4ST-1)/carbohydrate sulfotransferase 11 (CHST11) is a Golgi-bound enzyme involved in the biosynthesis of the glycosaminoglycan chondroitin sulfate. The sulfation pattern of chondroitin is tightly regulated during development, injury and disease, with the temporal and spatial expression of chondroitin sulfotransferase genes believed to be a crucial determinant of the fine balance of chondroitin sulfation. We have previously identified mouse *C4st-1* as a target gene of ligands of the TGFβ superfamily of growth factors, which could positively regulate *C4st-1* expression in a number of cell types. We have also shown that a gene trap loss-of-function mutation in *C4st-1* leads to severe skeletal abnormalities during mouse embryogenesis. In addition, we described a highly specific temporal and spatial expression pattern of *C4st-1* during mouse embryogenesis. However, the transcriptional regulatory mechanisms that control *C4st-1* gene expression remain unexplored. In order to gain knowledge on the transcriptional regulation of *C4ST-1*, we used a bioinformatical approach to identify conserved putative long-range *cis*-regulatory modules in a region of

120 kb spanning the 5' end of the *C4ST-1* gene. Luciferase reporter assays in human HEK293T and mouse NmuMG cells identified a functional *C4ST-1* promoter, as well as a number of *cis*-regulatory modules able to positively and negatively regulate *C4ST-1* expression. Moreover, we identified TGFβ-responsive regulatory modules that can function in a cell type-specific fashion. Taken together, our results identify TGFβ-dependent and -independent *cis*-regulatory modules of the *C4ST-1* gene.

Key words: C4ST-1; CHST11; Chondroitin sulfate; Gene expression; TGFβ; *cis*-regulatory modules

INTRODUCTION

Chondroitin-4-sulfotransferase-1 (C4ST-1) / carbohydrate sulfotransferase 11 (CHST11) is a Golgi-based enzyme involved in the biosynthesis of the glycosaminoglycans chondroitin sulfate (CS) and dermatan sulfate (Habuchi, 2000; Hiraoka et al., 2001). C4ST-1 is anchored in the Golgi-membrane and utilizes its catalytic intra-Golgi domain to transfer sulfate groups to the carbon-4 position of the GalNAc sugar residue in elongating CS chains attached to core proteins (Habuchi, 2000). These CS-proteoglycans are subsequently transported to cell membrane or extracellular matrix and have been shown to fulfill important biological functions, including maintenance of the extracellular matrix and participation in growth factor signaling pathways (Carulli et al., 2005; Hovanessian, 2006; Rolls and Schwartz, 2006; de Wit and Verhaagen, 2007).

Different chondroitin sulfotransferases mediate sulfation of different carbon residues of the disaccharide units, resulting in a delicately tuned balance of chondroitin sulfation (Habuchi, 2000; Kusche-Gullberg and Kjellen, 2003). Differentially sulfated CS forms include C4S (sulfation of carbon-4 of GalNAc), C6S (sulfation of carbon-6 of GalNAc) as well as the oversulfated forms CS-D and CS-E. CS-B has a different sugar composition and is also described as dermatan sulfate (Kusche-Gullberg and Kjellen, 2003). The sulfation pattern of chondroitin has been shown to be tightly regulated during development, injury and disease, with the temporal and spatial expression of chondroitin sulfotransferases as a crucial determinant of the fine balance of chondroitin sulfation. For example, levels of CS-D, as well as expression of uronyl 2-O-sulfotransferase, a sulfotransferase involved in the biosynthesis of CS-D, increase during development of the cerebellum, whereas levels of CS-E and expression of N-acetylgalactosamine 4-sulfate 6-O-sulfotransferase (GalNAc4S-6ST) decrease during the same developmental process (Ishii and Maeda, 2008). Synthesis of C6S and expression of chondroitin 6-sulfotransferase-1 (C6ST-1), but not other chondroitin sulfotransferases, are up-regulated in the CNS after injury (Properzi et al., 2003). Another study showed that as bone marrow cells differentiate to bone marrow-derived mast cells, expression levels of the chondroitin sulfotransferases C4ST-1 and GalNAc4S-6ST increased, whereas C6ST-1 expression decreased (Ohtake et al., 2008). These examples suggest a critical importance of carefully regulated expression of chondroitin sulfotransferase genes during mammalian development, injury and disease.

In a gene trap approach, we have previously identified *C4st-1* as a target gene of the TGFβ-like growth factors, including TGFβ, BMP2 and Activin (Klüppel et al., 2002). These factors were able to positively regulate *C4st-1* expression in undifferentiated mouse ES cells, dif-

ferentiating embryoid bodies, as well as mouse embryonic lung explants (Klüppel et al., 2002). Phenotypic analysis of mice homozygous for this gene trap loss-of-function mutation in the *C4st-1* gene developed severe skeletal and craniofacial abnormalities and died shortly after birth (Klüppel et al., 2005). We have also shown that *C4st-1* displays a highly specific temporal and spatial expression pattern during mouse embryogenesis. We have described expression in the apical ectodermal ridge of the developing limbs, and subsequently in the cartilage growth plate, in the ventral neural tube, brain, the developing heart, as well as developing liver, kidney, hair follicles, and mammary glands during embryogenesis (Klüppel et al., 2002). C4ST-1 has recently been shown to be involved in the biosynthesis of CS-E (Uyama et al., 2006; Ohtake et al., 2008), a highly sulfated CS with a putative role in mouse lung cancer metastasis (Nadanaka et al., 2008). Despite the importance of a tightly controlled regulation of chondroitin sulfotransferase genes, the transcriptional regulatory mechanisms controlling the expression of these genes are unknown.

It has been shown that in higher eukaryotes, *cis*-regulatory information is organized into modular units called '*cis*-regulatory modules' (CRM) of a few hundred base pairs. A common feature of these *cis*-regulatory modules is the presence of multiple binding sites for multiple transcription factors (Berman et al., 2002).

Here, we investigated the transcriptional regulation of the *C4ST-1* gene. We present evidence of putative CRM in 120 kb surrounding the predicted transcriptional start site of *C4ST-1* that is conserved between the mouse and human *C4ST-1* loci. We identify a functional *C4ST-1* promoter, as well as a number of regulatory elements able to positively and negatively regulate *C4ST-1* expression. Moreover, we identify CRM able to mediate the responsiveness of *C4ST-1* expression to stimulation with TGFβ. Together, our results identify regulatory elements controlling *C4ST-1* expression and responsiveness to TGFβ signaling. This is the first study to shed light on the transcriptional regulation of mammalian chondroitin sulfotransferase genes.

MATERIAL AND METHODS

Bioinformatical analysis of the *C4ST-1* locus

A region of 120 kb of genomic sequence surrounding the 5' end of exon 1 (-100 to +20 kb) of mouse and human *CHST11*/*C4ST-1* genes was identified in ENSEMBL (http://www.ensembl.org/index.html). Sequence homologies and conserved transcription factor binding sites between the mouse and human genomic sequences were analyzed with CONSITE (http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite/), an online bioinformatics tool, with sequence homology cut-off settings at 75%, and transcription factor binding site homology cut-off at 80%.

Amplification and cloning of identified CRM

Genomic sequences containing the regions identified above were amplified by polymerase chain reaction (PCR) from genomic DNA isolated from human primary skin fibroblasts (ATCC), using iProof-DNA Polymerase (BioRad), according to manufacturer instructions. Primer pairs were as follows (all 5' to 3'): -82 kb: CCGCTCGAGGTTTG CTATTTTCCAGAGTCCATTGAGTGAGC; CCGCTCGAGCTAGCACCTCTGTTGC TAAAGGCTCTTTGC. -55 kb: CCGCTCGAGGATGACCAACTAGCCAAGCTGCCCCG; CCGCTCGAGGTCACAGGCCTAACAAACTGGTTGAATGGC. -3.5 kb: CTCTCTTGCT

GTCTTTTTCTCCACC; GACGATCTTGTAGAATTTATAGCAGGG. -0.5 kb: GGGGAAG GGAACTGAGGTTGCC; CCCCAAAGGGTAGAGCATCCTCC. +0.1 kb: CCGCTCGAGCT CCGATCCTCCCTCTGAGCCTTGC; CCGCTCGAGCCTTCCCTTCTCCTCCCCATCCCG. +2 kb: CCGCTCGAGGAGTTGCCTTTCTCAGCGTAGTCACAGC; CCGCTCGAGACAC CACCAGTCCCCTGGCGCTC. +14 kb: CCGCTCGAGGCTTAGAGAGGTGAAGGAGC CTGCC; CCGCTCGAGCCAAGAGAGGATGAGCTCACCAGGG.

All primer pairs used contained additional *Xho*I restriction sites at the 5' end to facilitate cloning into pGL3P-E1B, with the exception of the primers used to amplify the -0.5 kb putative C4ST-1 promoter, which contained *Bgl*II (forward primer) and *Hin*dIII (reverse primer) in order to replace the E1B promoter in the pGL3P-E1B construct, and the primers for the -3.5-kb region, which was cloned blunt-end into the *Sma*I site of pGL3P-E1B.

Luciferase assays

HEK293T (human kidney epithelial) and NmuMG (normal murine mammary gland epithelial) cells were transiently transfected with the pGL3P plasmids containing either E1B promoter or the identified C4ST-1 CRM, or the 3TP-luciferase reporter plasmid, along with a pCMV5-Renilla luciferase construct as a transfection control, using Lipofectamine 2000 (Invitrogen), according to manufacturer instructions. Luciferase assays were performed using the Dual-Glo Luciferase Assay System (Promega) or the Firefly & Renilla Luciferase Assay Kit (Biotium), according to manufacturer instructions.

TGFβ stimulation

Four hours after transient transfection, cells were treated with recombinant TGFβ1 (4 ng/mL final concentration; R&D Systems) in DMEM supplemented with 0.1% FBS. Cells were incubated for 36 h at 37°C, followed by luciferase assays.

Real-time-RT-PCR

RNA was prepared from cell cultures using Trizol reagent, according to manufacturer instructions; 4 µg RNA was subsequently reverse transcribed using M-MLV-Reverse Transcriptase (Promega) in a total volume of 40 µL according to manufacturer instructions. One- to 3-µL RT reactions were then amplified using a 2X SYBR-Green reaction mix (ABI) in a 7500 Fast Real-Time PCR System (ABI) using primer pairs specific for C4ST-1 and HPRT (as a reference).

Statistical analysis

A paired *t*-test was used to obtain the two-tailed P value to determine the statistical significance of results (http://www.graphpad.com/quickcalcs/ttest1.cfm).

RESULTS

Identification of regions with conserved transcription factor binding sites in the mammalian *C4ST-1* locus

The coding region of the human *C4ST-1* locus spans approximately 300 kb (nucleotides 104,850,776-105,151,883, Ensembl Genebuild) on human chromosome 12q23.3. In order

to identify putative transcriptional regulatory elements of the *C4ST-1* gene, we analyzed a 120-kb region (from -100 to +20 kb in relation to the 5' end of the non-coding first exon of the *C4ST-1* gene) for sequence homologies between human and mouse *C4ST-1* loci, using CONSITE, an online bioinformatics tool (Figure 1). Sequence similarities and homologies were apparent in a number of regions, and were particularly concentrated around the putative transcriptional start site and exons 1 and 2 (labeled '0 kb' in Figure 1A). In order to analyze these homologies in more detail, we investigated whether mouse and human loci contained regions with conserved transcription factor (TF) binding sites using the CONSITE bioinformatical tool (Figure 1B). Indeed, in addition to 13 regions containing five or less conserved TF binding sites, we identified seven regions that contained nine or more conserved TF binding sites (Figure 1B), often containing multiple binding sites for multiple transcription factors (data not shown), which has previously been defined as CRM (Berman et al., 2002). The seven regions identified here were located at approximately -82, -55, -3.5, -0.5, +0.1, +2, and +14 kb relative to the start of *C4ST-1* exon 1 (Figure 1B,C). Figure 1C shows a schematic map of the 5' end of the *C4ST-1* locus, including exons 1 and 2, as well as the seven conserved putative CRM in the *C4ST-1* locus. Table 1 summarizes the identified elements, genomic location, length, % conservation between mouse and human sequences, as well as the number of conserved TF binding sites.

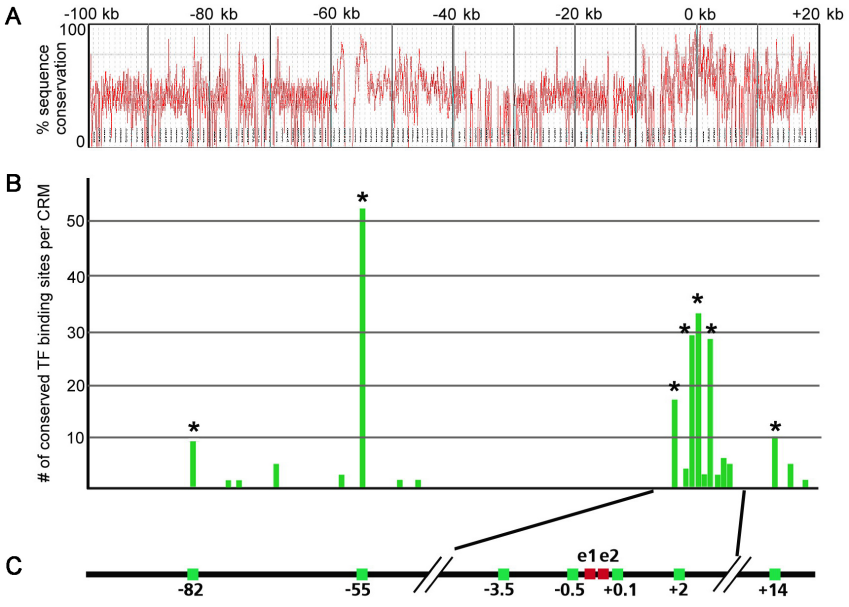

Figure 1. Bioinformatical analysis of putative *cis*-regulatory modules (CRM) in the *C4ST-1* locus. **A.** Schematic representation of sequence conservation between the human and mouse *C4ST-1* loci encompassing 120 kb surrounding the first exon of the *C4ST-1* gene (-100 to +20 kb). **B.** Conserved putative CRM and number of conserved transcription factor (TF) binding sites in the same region depicted in *A*. Green bars indicate the location and number of conserved TF binding sites in putative CRM. Stars indicate regions that were used for further analysis. **C.** Schematic map of the genomic region from -120 to +20 kb; the location of putative CRM at -82, -55, -3.5, -0.5, +0.1, +2, and +14 kb are indicated by green squares; exons 1 and 2 are marked by red squares.

Table 1. Putative *cis*-regulatory modules (CRM) in the *C4ST-1* locus.

CRM	Genomic location on human chromosome 12	Length	% Sequence conservation	No. of conserved transcription factor sites
-82 kb	104768318 - 104768395	78 bp	79	9
-55 kb	104799575 - 104799803	229 bp	80	52
-3.5 kb	104847315 - 104847508	194 bp	82	17
-0.5 kb	104849581 - 104849944	364 bp	69	29
+0.1 kb	104851072 - 104851436	365 bp	86	33
+2 kb	104852489 - 104853126	637 bp	70	28
+14 kb	104864629 - 104864695	67 bp	84	10

Functional analysis of the identified putative *C4ST-1 cis*-regulatory modules

We next wanted to determine if the conserved putative CRM identified above possess transcriptional regulatory activity. First, we investigated whether the -0.5-kb region, located just upstream of the first exon, and the only region that contained multiple conserved TATA box-binding protein (TBP) sequences (data not shown), a sequence frequently found in gene promoters, is a functional promoter for the *C4ST-1* gene. A 509-bp genomic fragment containing the conserved human 364-bp region from human primary fibroblast genomic DNA was amplified by PCR, and cloned this fragment into a pGL3P luciferase reporter vector (pGL3P-C4ST-1pr), from which we had removed the minimal E1B promoter. We then asked whether the -0.5-kb region has promoter activity; i.e., can drive expression of the firefly luciferase reporter gene in the absence of any other promoters or regulatory sequences. The pGL3P-C4ST-1pr and pGL3P-E1B (in which luciferase expression is driven by the minimal E1B promoter) vectors, together with a pCMV5-Renilla luciferase control plasmid, were transiently transfected into human HEK293T and mouse NmuMG cells, and luciferase activity was measured after 36 h (Figure 2). These experiments revealed that the -0.5-kb region had very strong promoter activity in both HEK293T and NmuMG cells, with promoter activity increased 150- to 220-fold when compared to the E1B promoter (Figure 2).

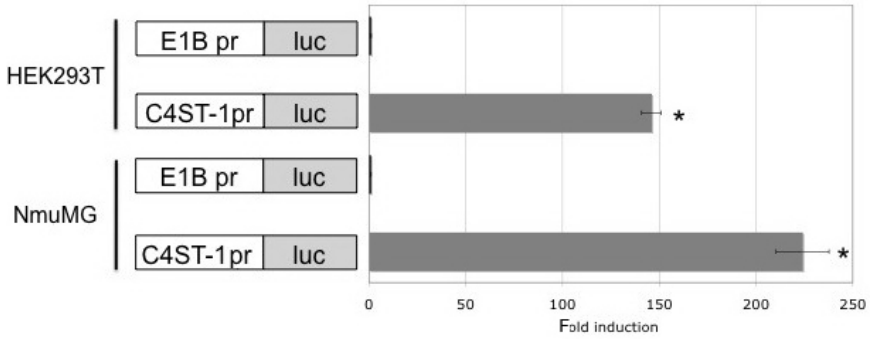

Figure 2. The *C4ST-1* -0.5 kb *cis*-regulatory module (CRM) functions as a promoter. Transient transfection of E1B-luciferase and (-0.5 kb)-luciferase constructs revealed promoter activity of the *C4ST-1* -0.5-kb region in HEK293T, and NmuMG cells. Promoter activity of the -0.5-kb CRM was increased approximately 150- to 220-fold when compared to the minimal E1B promoter (*$P < 0.05$).

Next, we wanted to determine whether the remaining putative CRM have transcriptional regulatory functions. For this, we amplified the -82, -55, -3.5, +0.1, +2, and +14 kb modules from human primary fibroblast genomic DNA by PCR, cloned the fragments in forward orientation in front of a minimal E1B promoter in the pGL3P vector to create pGL3P(-82 kb)-E1B, pGL3P(-55 kb)-E1B, pGL3P(-3.5 kb)-E1B, pGL3P(+0.1 kb)-E1B, pGL3P(+2 kb)-E1B, and pGL3P(+14 kb)-E1B. We then asked whether the presence of these elements can affect E1B-driven luciferase expression and activity. Constructs were again transiently transfected into human HEK293T and mouse NmuMG cells, together with a pCMV5-Renilla luciferase control plasmid, and luciferase activities were measured after 36 h (Figure 3). In HEK293T cells, we observed a 2-fold repression of luciferase activity by the -55 and +0.1 kb modules, whereas the +2 and +14 kb modules were able to induce luciferase activity approximately 2.5-fold. The -3.5-kb module caused an approximately 9-fold induction of luciferase expression (Figure 3). The -82 kb module did not have any regulatory function in this assay. Results from NmuMG cells were similar to those described for HEK293T cell; the -55 and +0.1 kb modules repressed luciferase activity about 2-fold, whereas the -3.5, +2 and +14 kb modules induced luciferase activity. The -82 kb module again did not affect luciferase activity (Figure 3). The activities of the -3.5 and +14 kb CRM in HEK293T, and the -55, -3.5, +2, and +14 kb CRM in NmuMG cells proved to be statistically relevant, whereas the results for the -55, +0.1, and +2 kb CRM in HEK293T, and the +0.1 kb CRM in NmuMG cells were associated with P values between 0.51 and 0.65, and thus just fell short of statistical significance. These results demonstrate that our bioinformatical approach was successful in identifying functional positive and negative *cis*-regulatory modules in the *C4ST-1* locus.

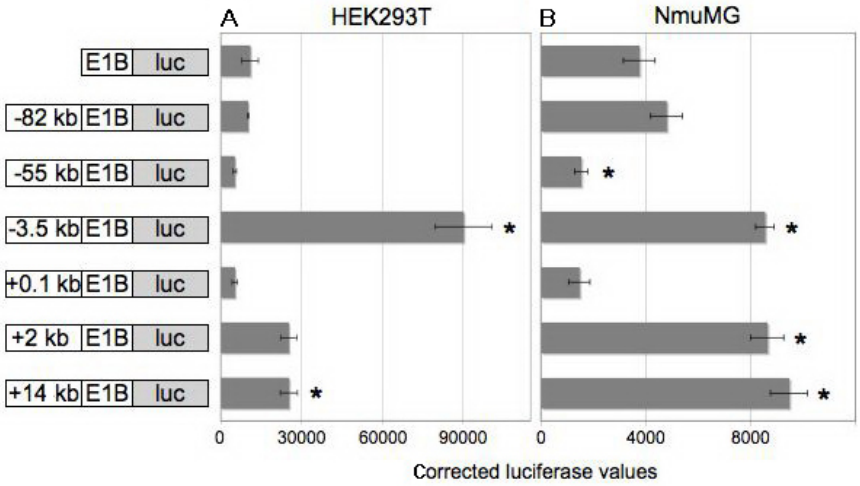

Figure 3. Transcriptional regulatory functions of *C4ST-1 cis*-regulatory modules. Transient transfection of E1B-luciferase, (-82 kb)-E1B-luciferase, (-55 kb)-E1B-luciferase, (-3.5 kb)-E1B-luciferase, (+0.1 kb)-E1B-luciferase, (+2 kb)-E1B-luciferase, (+14 kb)-E1B-luciferase constructs into HEK293T (A), and NmuMG (B) cells. **A.** In HEK293T cells, the *C4ST-1* -3.5 and +14 kb elements significantly stimulated E1B-driven luciferase expression and activity (*P < 0.05). In addition, the -55 and +0.1 kb elements repressed, and the +2 kb element increased luciferase expression although this just fell short of statistical significance. The -82 kb element did not alter luciferase activity. **B.** In NmuMG cells, the -3.5, +2 and +14 kb elements increased E1B-driven luciferase expression, while the -55 kb element was able to negatively regulate expression (*P < 0.05). In addition, the +0.1 kb element was able to repress E1B-driven expression, but this again did not quite reach statistical significance. The -82 kb element did not alter luciferase activity.

Responsiveness of *C4ST-1* expression to TGFβ stimulation in HEK293T and NmuMG cells

Since we have previously shown that *C4ST-1* expression is positively regulated by TGFβ signaling in a number of cell types and tissues (Klüppel et al., 2002), we next wanted to determine if TGFβ treatment could also induce *C4ST-1* expression in HEK293T and NmuMG cells. For this, we first tested whether these cells, which have previously been shown to have functional TGFβ signaling, are also responsive to TGFβ in our experimental setup. For this, cells were transfected with the TGFβ-responsive 3TP-luciferase reporter construct. Indeed, TGFβ treatment of both transfected HEK293T and NmuMG cells for 36 h led to a 2.5- and 7.3-fold increase in luciferase activity, respectively (Figure 4A), demonstrating that our TGFβ treatments are functionally activating the TGFβ signaling pathway in these cell types. Next, we wanted to determine whether TGFβ treatment could induce *C4ST-1* expression in these cells. Cells were treated with TGFβ for 36 h, followed by preparation of RNA, reverse transcription and real-time PCR to quantify levels of *C4ST-1* expression. Our results showed that TGFβ stimulation could induce *C4ST-1* expression in both HEK293T and NmuMG cells (Figure 4B). Thus, these experiments established an experimental model in which we can test whether the *C4ST-1 cis*-regulatory modules identified above could mediate the positive regulation of *C4ST-1* by TGFβ signaling.

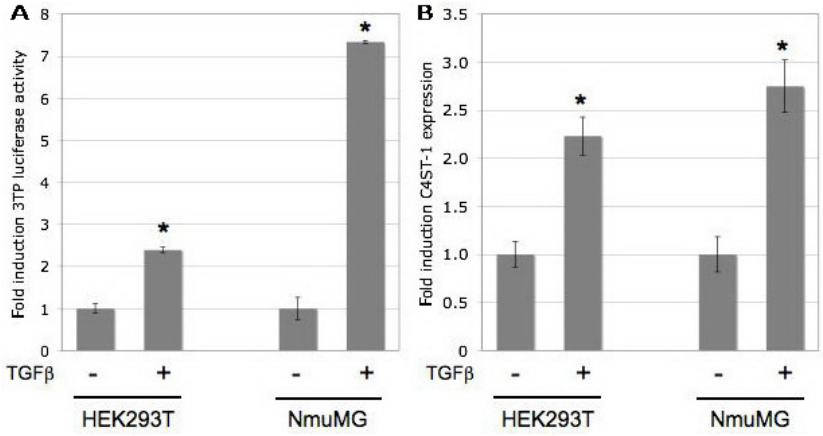

Figure 4. Active TGFβ signaling induces *C4ST-1* expression in HEK293T and NmuMG cells. **A.** Transient transfection of a 3TP-luciferase reporter construct into HEK293T and NmuMG cells in order to establish functional TGFβ signaling. Treatment with TGFβ lead to a significant increase in luciferase activity in both cell types (*P < 0.05), demonstrating functional TGFβ signaling. **B.** Real-time-RT-PCR analysis of TGFβ signaling-mediated induction of *C4ST-1* expression in HEK293T and NmuMG cells. Treatment of HEK293T and NmuMG cells with TGFβ caused a significant increase in *C4ST-1* expression (*P < 0.05).

TGFβ-responsiveness of specific *C4ST-1 cis*-regulatory modules

In order to evaluate whether the *C4ST-1* CRM described above could mediate the effect of TGFβ stimulation on *C4ST-1* expression, we transiently transfected either the parental E1B-luciferase vector, the pGL3P-C4ST-1pr vector, or the pGL3P(-82 kb)-E1B, pGL3P(-55

kb)-E1B, pGL3P(-3.5 kb)-E1B, pGL3P(+0.1 kb)-E1B, pGL3P(+2 kb)-E1B, and pGL3P(+14 kb)-E1B vectors, all with pCMV5-Renilla luciferase as transfection control, into HEK293T and NmuMG cells and asked whether treatment with TGFβ for 36 h could affect luciferase activities of these constructs (Figure 5). In HEK293T cells, treatment with TGFβ lead to a significant increase of luciferase activity when driven by the +14 kb CRM (Figure 5A). In contrast, several elements appeared responsive to TGFβ stimulation in NmuMG cells (Figure 5B). Specifically, the -55, -3.5, and +2 kb *cis*-regulatory modules responded to TGFβ treatment with an increase in luciferase activity (Figure 5B). These results suggest that several of the *cis*-regulatory modules identified above could integrate TGFβ signals in a cell-type specific fashion to affect *C4ST-1* expression.

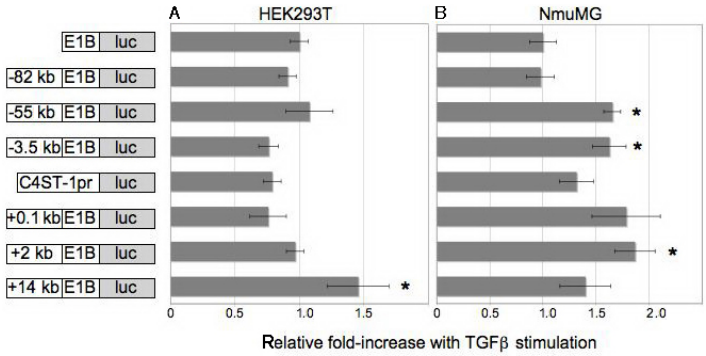

Figure 5. Responsiveness of *C4ST-1 cis*-regulatory modules (CRM) to TGFβ signaling. Transient transfection of E1B-luciferase and *C4ST-1* CRM-luciferase constructs to evaluate responsiveness of identified CRM to TGFβ stimulation. The ratio of E1B-luciferase minus TGFβ to E1B-luciferase plus TGFβ was set to '1', and the relative fold-change for all constructs was determined. **A.** In HEK293T cells, the +14 kb CRM was able to mediate a significant induction of luciferase activity by TGFβ treatment (*P < 0.05). **B.** In NmuMG cells, the -55, -3.5, and +2 kb CRM were able to mediate a significant induction of luciferase activity by TGFβ signaling (*P < 0.05).

In summary, we identified functional *cis*-regulatory modules in the *C4ST-1* locus that are able to modulate transcriptional activity in both TGFβ-dependent and -independent fashion (Figure 6).

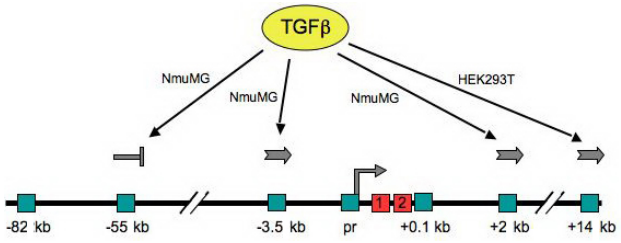

Figure 6. Model of *cis*-regulatory modules (CRM) in the *C4ST-1* locus. We identified seven putative CRM in the *C4ST-1* locus. The -0.5 kb element is a functional promoter (pr) in both HEK293T and NmuMG cells. The -55 kb CRM functions as a negative regulator of expression, whereas the -3.5, +2, and +14 kb CRM are positive regulators. In HEK293T cells, the +14 kb CRM was responsive to TGFβ signaling, whereas the -55, -3.5, and +2 kb elements were responsive to TGFβ stimulation in NmuMG cells.

DISCUSSION

Here, we have identified putative *cis*-regulatory modules in the *C4ST-1* locus using a bioinformatical approach. We established functionality of several of the identified modules in luciferase reporter assays in HEK293T and NmuMG cells. Moreover, we showed that *C4ST-1* expression could be induced by TGFβ stimulation in both cell types, and identified several *cis*-regulatory modules that appear to be able to mediate TGFβ stimulation in a cell type-specific fashion. Together, our results are the first insights into the transcriptional regulation of any chondroitin sulfotransferase genes.

Regulation of *C4ST-1* expression

The sulfation pattern of chondroitin is tightly regulated during development, injury and disease, with the temporal and spatial expression of chondroitin sulfotransferases as a crucial determinant of the fine balance of chondroitin sulfation (Domowicz et al., 2000; Habuchi, 2000; Liu et al., 2006; Akita et al., 2008; Ishii and Maeda, 2008; Ohtake et al., 2008). This suggests a critical importance of carefully regulated expression of chondroitin sulfotransferase genes during mammalian development, injury and disease, in order to produce chondroitin sulfate proteoglycans with specific chondroitin sulfation patterns, enabling these proteoglycans to fulfill specific biological functions. For example, oversulfated CS-D and CS-E side chains of the receptor-type Protein Tyrosine Phosphatase Zeta (PTPzeta) have been shown to facilitate binding of the ligand Pleiotrophin to PTPzeta, and are crucial for the functional role of PTPzeta signaling in neuronal cells (Tanaka et al., 2003; Bao et al., 2005; Maeda et al., 2006). In addition, CS-E has recently been shown to be able to bind the canonical Wnt ligand Wnt3a with high affinity (Nadanaka et al., 2008). Also, CS-E has been shown to play a functional role in Lewis lung cancer metastasis in a mouse model (Li et al., 2008). Since chondroitin 4-sulfation by C4ST-1 represents the necessary first step in the sulfation of CS-E (Uyama et al., 2006; Nadanaka et al., 2008), we propose that the tightly controlled expression of the *C4ST-1* gene might play a crucial role in the biosynthesis of CS-E, and its role in development, cancer and signaling events.

The importance of proper expression of *C4ST-1* is also underscored by our previously published phenotypic analysis of mice homozygous for a gene trap loss-of-function mutation in the *C4st-1* gene. This analysis showed that loss of C4st-1 expression leads to severe skeletal and craniofacial abnormalities and perinatal lethality (Klüppel et al., 2005).

Cis-regulatory modules in the *C4ST-1* locus

C*is*-regulatory information has been shown to be organized into modular units called '*cis*-regulatory modules' (CRM) of a few hundred base pairs. A common feature of these *cis*-regulatory modules is the presence of multiple binding sites for multiple transcription factors (Berman et al., 2002). Indeed, our bioinformatical analysis conserved identified several CRM that shared these features. For example, we identified a functional *C4ST-1* promoter, which contained multiple binding sites for TBP, a protein previously shown to bind eukaryotic promoters (Pugh, 2000; Thomas and Chiang, 2006).

Several of the CRM we identified are located at a considerable distance from the *C4ST-1* promoter. Indeed, distant long-range *cis*-regulatory elements have previously been

shown to play crucial roles in the temporal and spatial regulation of gene expression, and have been shown to come into close proximity and physically interact with promoter sequences through interaction of DNA binding proteins (Sipos and Gyurkovics, 2005; Bartkuhn and Renkawitz, 2008). We hypothesize that some of the long-range CRM we identified in the *C4ST-1* locus will also physically interact with the promoter to facilitate the temporal and spatial transcription of the *C4ST-1* gene. We have previously demonstrated a dynamic temporal and spatial expression pattern of *C4ST-1* during mouse development. In limbs, *C4ST-1* is expressed in the apical ectodermal ridge, and subsequently in the cartilage growth plate of the long bones. Expression was also observed in embryonic kidney, neuronal tissues, heart, mammary glands, and in a subset of hair follicles (Klüppel et al., 2002). It will be interesting to identify the specific roles of the identified CRM in mediating temporal and spatial expression of *C4ST-1*.

We expect that additional, so far unidentified CRM exist in the *C4ST-1* locus, quite possibly further 3' in the locus or at a greater distance to the promoter. We will expand our search for CRM by adding additional regions of the *C4ST-1* locus in our homology search. There is also a possibility that important CRM will be missed by homology searches, for example when homologies are absent due to species-specific functions of elements. It will be interesting to identify CRM by other means, including DNase-hypersensitive site mapping (Nuthall et al., 1999) and identification of regions able to bind acetylated histones (Myers et al., 2001; Agalioti et al., 2002), and analyze their potential physical interactions with the promoter and other CRM using Chromosome Conformation Capture protocols (Ott et al., 2009; van Berkum and Dekker, 2009).

TGFβ and *C4ST-1* expression

We have previously identified *C4st-1* as a target gene of the TGFβ-like growth factors, including TGFβ, BMP2 and Activin (Klüppel et al., 2002). These factors were able to positively regulate *C4st-1* expression in undifferentiated mouse ES cells, differentiating embryoid bodies, as well as mouse embryonic lung explants (Klüppel et al., 2002). Moreover, TGFβ has previously been shown to play important roles in shaping the microenvironment of cells by controlling expression programs of genes involved in the structure and composition of the extracellular matrix (Tabibzadeh, 2002; Verrecchia and Mauviel, 2002). Since expression of C4ST-1 leads to modification of the chondroitin sulfation balance of cell surface and extracellular matrix proteins, the regulation of *C4ST-1* expression by TGFβ will likely contribute to the known role of this signaling pathway to shape the cellular microenvironment. Here we identified several CRM that are responsive to TGFβ treatment, suggesting that these elements fulfill important functions in mediating the effects of TGFβ on *C4ST-1* expression and the CS composition of the cellular microenvironment.

In summary, we have identified functional TGFβ-dependent and -independent *cis*-regulatory modules in the *C4ST-1* locus. Our results represent the first insights into the transcriptional regulation of any chondroitin sulfotransferase gene.

ACKNOWLEDGMENTS

We would like to thank Christoforo Silvestri for introduction to the CONSITE bioinformatical analysis tool. This study was funded by the Canadian Institutes of Health Research

(CIHR) to J.L. Wrana and Children's Memorial Research Center to M. Klüppel. M. Klüppel was also a recipient of a CIHR postdoctoral fellowship. J.L. Wrana is a CIHR investigator and an International Research Scholar of the Howard Hughes Medical Institute.

REFERENCES

Agalioti T, Chen G and Thanos D (2002). Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 111: 381-392.

Akita K, von Holst A, Furukawa Y, Mikami T, et al. (2008). Expression of multiple chondroitin/dermatan sulfotransferases in the neurogenic regions of the embryonic and adult central nervous system implies that complex chondroitin sulfates have a role in neural stem cell maintenance. *Stem Cells* 26: 798-809.

Bao X, Muramatsu T and Sugahara K (2005). Demonstration of the pleiotrophin-binding oligosaccharide sequences isolated from chondroitin sulfate/dermatan sulfate hybrid chains of embryonic pig brains. *J. Biol. Chem.* 280: 35318-35328.

Bartkuhn M and Renkawitz R (2008). Long range chromatin interactions involved in gene regulation. *Biochim. Biophys. Acta* 1783: 2161-2166.

Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, et al. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl. Acad. Sci. U. S. A.* 99: 757-762.

Carulli D, Laabs T, Geller HM and Fawcett JW (2005). Chondroitin sulfate proteoglycans in neural development and regeneration. *Curr. Opin. Neurobiol.* 15: 116-120.

de Wit J and Verhaagen J (2007). Proteoglycans as modulators of axon guidance cue function. *Adv. Exp. Med. Biol.* 600: 73-89.

Domowicz M, Mangoura D and Schwartz NB (2000). Cell specific-chondroitin sulfate proteoglycan expression during CNS morphogenesis in the chick embryo. *Int. J. Dev. Neurosci.* 18: 629-641.

Habuchi O (2000). Diversity and functions of glycosaminoglycan sulfotransferases. *Biochim. Biophys. Acta* 1474: 115-127.

Hiraoka N, Misra A, Belot F, Hindsgaul O, et al. (2001). Molecular cloning and expression of two distinct human N-acetylgalactosamine 4-O-sulfotransferases that transfer sulfate to GalNAc beta 1→4GlcNAc beta 1→R in both N- and O-glycans. *Glycobiology* 11: 495-504.

Hovanessian AG (2006). Midkine, a cytokine that inhibits HIV infection by binding to the cell surface expressed nucleolin. *Cell Res.* 16: 174-181.

Ishii M and Maeda N (2008). Spatiotemporal expression of chondroitin sulfate sulfotransferases in the postnatal developing mouse cerebellum. *Glycobiology* 18: 602-614.

Klüppel M, Vallis KA and Wrana JL (2002). A high-throughput induction gene trap approach defines C4ST as a target of BMP signaling. *Mech. Dev.* 118: 77-89.

Klüppel M, Wight TN, Chan C, Hinek A, et al. (2005). Maintenance of chondroitin sulfation balance by chondroitin-4-sulfotransferase 1 is required for chondrocyte development and growth factor signaling during cartilage morphogenesis. *Development* 132: 3989-4003.

Kusche-Gullberg M and Kjellen L (2003). Sulfotransferases in glycosaminoglycan biosynthesis. *Curr. Opin. Struct. Biol.* 13: 605-611.

Li F, Ten Dam GB, Murugan S, Yamada S, et al. (2008). Involvement of highly sulfated chondroitin sulfate in the metastasis of the Lewis lung carcinoma cells. *J. Biol. Chem.* 283: 34294-34304.

Liu J, Chau CH, Liu H, Jang BR, et al. (2006). Upregulation of chondroitin 6-sulphotransferase-1 facilitates Schwann cell migration during axonal growth. *J. Cell Sci.* 119: 933-942.

Maeda N, Fukazawa N and Hata T (2006). The binding of chondroitin sulfate to pleiotrophin/heparin-binding growth-associated molecule is regulated by chain length and oversulfated structures. *J. Biol. Chem.* 281: 4894-4902.

Myers FA, Evans DR, Clayton AL, Thorne AW, et al. (2001). Targeted and extended acetylation of histones H4 and H3 at active and inactive genes in chicken embryo erythrocytes. *J. Biol. Chem.* 276: 20197-20205.

Nadanaka S, Ishida M, Ikegami M and Kitagawa H (2008). Chondroitin 4-O-sulfotransferase-1 modulates Wnt-3a signaling through control of E disaccharide expression of chondroitin sulfate. *J. Biol. Chem.* 283: 27333-27343.

Nuthall HN, Moulin DS, Huxley C and Harris A (1999). Analysis of DNase-I-hypersensitive sites at the 3' end of the cystic fibrosis transmembrane conductance regulator gene (CFTR). *Biochem. J.* 341 (Pt 3): 601-611.

Ohtake S, Kondo S, Morisaki T, Matsumura K, et al. (2008). Expression of sulfotransferases involved in the biosynthesis of chondroitin sulfate E in the bone marrow derived mast cells. *Biochim. Biophys. Acta* 1780: 687-695.

Ott CJ, Blackledge NP, Leir SH and Harris A (2009). Novel regulatory mechanisms for the CFTR gene. *Biochem. Soc. Trans.* 37: 843-848.

Properzi F, Asher RA and Fawcett JW (2003). Chondroitin sulphate proteoglycans in the central nervous system: changes and synthesis after injury. *Biochem. Soc. Trans.* 31: 335-336.

Pugh BF (2000). Control of gene expression through regulation of the TATA-binding protein. *Gene* 255: 1-14.

Rolls A and Schwartz M (2006). Chondroitin sulfate proteoglycan and its degradation products in CNS repair. *Adv. Pharmacol.* 53: 357-374.

Sipos L and Gyurkovics H (2005). Long-distance interactions between enhancers and promoters. *FEBS J.* 272: 3253-3259.

Tabibzadeh S (2002). Homeostasis of extracellular matrix by TGF-beta and lefty. *Front Biosci.* 7: d1231-d1246.

Tanaka M, Maeda N, Noda M and Marunouchi T (2003). A chondroitin sulfate proteoglycan PTPzeta /RPTPbeta regulates the morphogenesis of Purkinje cell dendrites in the developing cerebellum. *J. Neurosci.* 23: 2804-2814.

Thomas MC and Chiang CM (2006). The general transcription machinery and general cofactors. *Crit Rev. Biochem. Mol. Biol.* 41: 105-178.

Uyama T, Ishida M, Izumikawa T, Trybala E, et al. (2006). Chondroitin 4-O-sulfotransferase-1 regulates E disaccharide expression of chondroitin sulfate required for herpes simplex virus infectivity. *J. Biol. Chem.* 281: 38668-38674.

van Berkum NL and Dekker J (2009). Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods Mol. Biol.* 567: 189-213.

Verrecchia F and Mauviel A (2002). Transforming growth factor-beta signaling through the Smad pathway: role in extracellular matrix gene expression and regulation. *J. Invest. Dermatol.* 118: 211-215.

