Available online at www.sciencedirect.com
SCIENCE DIRECT
Matrix Biology 24 (2005) 550 – 559
MATRIX BIOLOGY
www.elsevier.com/locate/matbio

Characterization of human Collagen XVIII promoter 2: Interaction of Sp1, Sp3 and YY1 with the regulatory region and a SNP that increases transcription in hepatocytes

Lucia M. Armelin-Correa [a], Chin J. Lin [b,c], Angela Barbosa [c], Kelly Bagatini [a], Sheila M.B. Winnischofer [d], Mari C. Sogayar [d], Maria Rita Passos-Bueno [a,*]

[a] *Department of Genetics and Evolution Biology , Bioscience Institute, University of São Paulo, São Paulo, Brazil*
[b] *Department of Pathology, Hospital das Clínicas da FMUSP, University of São Paulo, São Paulo, Brazil*
[c] *Endocrinology Division, Hospital das Clínicas da FMUSP, University of São Paulo, São Paulo, Brazil*
[d] *Department of Biochemistry, Chemistry Institute, University of São Paulo, São Paulo, Brazil*

Received 31 May 2005; received in revised form 29 July 2005; accepted 22 August 2005

Abstract

Different levels of Collagen XVIII expression have been associated with several pathological processes such as cancer, liver fibrosis, diabetic retinopathy and Alzheimer's disease. Understanding the transcriptional regulation of Collagen XVIII might elucidate some pathways related to the progression of these diseases. The promoter 2 of *COL18A1* gene is poorly understood and is responsible for the transcription of this gene in several adult tissues such as liver, eyes and brain. This study focused upon characterization of *cis*-regulatory elements interacting with human *COL18A1* promoter 2 and identification of SNPs in this region in different ethnic groups. Our results show that there are five conserved regions (I to V) between human and mouse promoter 2 and that the human *COL18A1* core promoter is located between nucleotides −186 and −21. Sp1 and Sp3 bind to conserved regions I and V, while Sp3 and YY1 interact with region II. We have verified that the SNP at position −700 (T>G) is embedded in two common haplotypes, which have different frequencies between European and African descendents. The allele −700G increases transcription and binding for a still unknown transcription factor. SNP −700 affects Sp3 and YY1 interaction with this region, even though it is not part of these transcription factors' predicted binding sites. Therefore, our results show for the first time that Sp3 and YY1 interact with human *COL18A1* promoter 2, and that nucleotide −700 is part of a binding motif for a still unknown TF that is involved in the expression of this gene in hepatocytes. In addition, we also confirm the involvement of Sp1 in the regulation of this gene.

Keywords: Collagen XVIII; Transcriptional regulation; Functional SNP; Sp1; Sp3; YY1

1. Introduction

Type XVIII collagen is a member of the MULTIPLEXIN collagen family characterized by polypeptide chains with multiple triple-helical domains separated and flanked by non-triple-helical regions (Oh et al., 1994; Rehn and Pihlajaniemi, 1994). This collagen, a homotrimer molecule made of three α1 chains, is a heparan sulfate proteoglycan constituent of epithelial and endothelial basement membranes of a wide variety of tissues (Saarela et al., 1998a). Human type XVIII collagen, encoded by the *COL18A1* gene, has three isoforms, which differ in their N-terminal non-collagenous domain (NC11), namely: NC11-303, NC11-493 and NC11-728. The *COL18A1* gene spans 105 kbp and contains 43 exons and 2 putative promoter regions, namely: promoter 1 located 5′ from exon 1, and promoter 2 within intron 2 (Saarela et al., 1998b; Elamaa et al., 2003). The NC11-303 isoform is associated with promoter 1 and the other two isoforms have their transcription regulated by the promoter 2. Variant NC11-728 is expressed in

Abbreviations: SNP, Single Nucleotide Polymorphism; Sp, Specificity proteins; YY1, Yin and Yang 1; C/EBP, CCAAT/Enhancer-Binding Protein; HNF3, Hepatocyte Nuclear Factor 3; TF, Transcription factor; PCR, Polymerase Chain Reaction; EMSA, Eletrophoretic Mobility Shift Assay.

* Corresponding author. Centro de Estudos do Genoma Humano, Depto. de Biologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277 sala 200, São Paulo, SP, CEP: 05508-900, Brazil. Tel.: +55 11 3091 9910; fax: +55 11 3091 7419.
E-mail address: passos@ib.usp.br (M.R. Passos-Bueno).

0945-053X/$ - see front matter © 2005 Elsevier B.V./International Society of Matrix Biology. All rights reserved.
doi:10.1016/j.matbio.2005.08.003

various human fetal tissues (Elamaa et al., 2003) and adult brain and retina (Suzuki et al., 2002), while NC11-493 is highly expressed in liver (Saarela et al., 1998b). Liétard et al. (2000) have shown that the mouse *Col18a1* promoter 2 is up regulated by Sp1 and HNF3 and, that it has a silencer-like element that may be regulated by HNF3 and NF1/CTF interactions with this regulatory region, but, so far, data on human *COL18A1* regulatory regions are limited to the fact that there are two distinct promoters (Saarela et al., 1998b; Elamaa et al., 2003).

Although *COL18A1* function is not completely understood, it is well known that its deficiency causes several ocular problems and neuronal cell migration disorders in patients with Knobloch syndrome (Sertié et al., 2000; Suzuki et al., 2002; Kliemann et al., 2003). Type XVIII collagen expression varies in many pathological conditions such as cancer (Musso et al., 2001), liver fibrosis (Musso et al., 1998), Alzheimer's disease (van Horssen et al., 2002) and diabetic retinopathy (Noma et al., 2002). Also, the 184-aminoacid proteolytic carboxyterminus fragment of Collagen XVIII, called endostatin, has been characterized as a potent endogenous inhibitor of angiogenesis (O'Reilly et al., 1997). Interestingly, there is a wide variation in serum levels of endostatin, mainly originated from NC11-493 proteolysis, in a normal human population (Zorick et al., 2001). The characterization of the regulatory mechanisms underlying the *COL18A1* gene expression will help elucidate the variability of Collagen XVIII and endostatin levels both in physical and pathological conditions. Therefore, in the present study we have focused upon characterization of *cis*-regulatory elements interacting with human *COL18A1* promoter 2 and identification of functional SNPs in this region in different ethnic groups.

2. Results

2.1. Functional analysis of the COL18A1 promoter 2

Alignment of approximately 1.5 kb of the human and 3.7 kb of the mouse Collagen XVIII promoter 2 showed five regions with more than 60% identity over a minimum of 40 bp (Fig. 1).

To elucidate the role of these conserved regions in the transcription of *COL18A1* in hepatocytes, six promoter deletion constructs were generated (Fig. 2). The constructs pGL3-1310, representing deletion of the first half of region V, and pGL3-1052, representing deletion of the second half of region V and region IV, showed no significant ($p=0.793$ and 0.966, respectively) differences in transcriptional activity when compared with the largest construct, pGL-1487. The pGL3-555 construct caused a significant ($p<0.001$) increase in transcriptional activity, suggesting the presence of *cis*-repressing elements between nucleotides −1052 and −555. Removal of the sequence spanning −555 to −458, which includes conserved region III, decreased the promoter activity ($p=0.005$). Deletion of nucleotides from −458 to −186 produced an increase in promoter activity, but this increase was not statistically significant ($p=0.332$). These results suggest that *COL18A1* has a core promoter located between nucleotides −186 and −21, an enhancer element between nucleotides −555 and −458 and a repressor element between nucleotides −1052 and −555. The −1487 bp fragment, cloned in the reverse orientation, did not present transcriptional activity (data not shown).

2.2. Nucleoprotein complexes formed by the promoter 2

Since our in silico analysis has identified potential binding sites for Sp1/Sp3, HNF3, C/EBP and YY1 in the *COL18A1* promoter region (Fig. 1), EMSAs were performed using probes representing clusters in regions I, II, III and IV (Fig. 1).

Three specific complexes (A, B and C) were formed with probe −94/−46 (Fig. 3A, lane 6). Anti-Sp1 antibody was able to abrogate complex B (Fig. 3A, lane 7) whereas an antibody directed against Sp3 abrogated complexes A and C (Fig. 3A, lane 8).

The −181/−131 probe generated two specific complexes (A and B) (Fig. 3B, lane 5). Antibody directed against YY1 abrogated complex B (Fig. 3B, lane 7) and antibody against Sp3 abrogated complex A (data not shown). No super-shift or complex abrogation was observed when an antibody against Sp1 was used (Fig 3B, lanes 6).

EMSA analysis with probe −1300/−1268, which corresponds to part of conserved region V with binding sites for Sp1/Sp3 (Fig. 1), yielded at least two specific complexes (A and B) (Fig. 3C, lane 6). Even though anti-Sp1 antibody did not abrogate any complex, it generated a super-shift (Fig. 3C, lane 7), suggesting that Sp1 interacts with this region. Antibody directed against Sp3 abrogated complex B (Fig. 3C, lane 8).

EMSA analysis with probes −517/−470, −1382/−1346 and −1469/−1430 did not show any super-shift or complex abrogation when antibodies against HNF3 and C/EBP were used (data not shown).

Therefore, Sp1 and Sp3 bind to the *COL18A1* promoter 2 from nucleotide −94 to −46, and from −1300 to −1268, while YY1 and Sp3 bind to the region between nucleotides −181 and −131.

2.3. Polymorphisms in the human COL18A1 promoter 2

We have tested the −700 T/G SNP (GenBank acession No. rs755548) in 258 individuals (121 European and 137 African descendents) and observed a significant difference ($p=0.0066$) in genotypic frequencies between the two ethnic groups (Table 1).

Screening of *COL18A1* promoter 2 in DNA samples of 48 European and 46 African descendents revealed also 3 other new SNPs, −251 C/T, −1428 C/T and −1528 G/A. All SNPs are in Hardy–Weinberg equilibrium. Seven putative haplotypes were identified in the two ethnic groups (Table 2). Among the European descendents, the most frequent haplotype is Hap1 (54%), with Hap5 and Hap6 being exclusive to this group. African descendents present Hap2 as the most frequent haplotype (55%) and Hap7 is found only in this group. Hap1 and Hap 2, sequences differ only at the −700 T/G SNP.