Cloning and Characterization of Rat Dentin Matrix Protein 1 (DMP1) Gene and Its 5'-Upstream Region*

Seetha R. Thotakura‡, Narayan Karthikeyan‡, Taryn Smith, Kai Liu, and Anne George§

From the Department of Oral Biology, University of Illinois at Chicago, Chicago, Illinois 60612

Rat dentin matrix protein 1 (DMP1) is a highly acidic 58-kDa phosphoprotein, and DMP1 was the first gene to be cloned from the mineralized dentin matrix. It exists as a highly phosphorylated protein with a pI of 3 in the dentin matrix and, in that state, might have an important role in the mineralization process. The spatio-temporal distribution during development indicates that the expression of this gene is tightly regulated in the odontoblasts. It is now known that DMP1 is not unique to dentin but is present in other mineralized tissues like long bone, calvaria, and ameloblasts. To study the transcriptional regulation and the function of DMP1 in these tissues, a genomic clone with a functional promoter, introns, and exons was isolated. Sequence analysis showed that the rat DMP1 gene is comprised of six exons and five introns and spans ~13 kilobases (kb). Exon 1 contains the 5'-untranslated sequences. Exon 2 encodes a total of 18 amino acids including the 16 amino acids of the signal sequence. Exons 3–5 encode 16, 11, and 15 amino acids, respectively. Exon 6 contains 1.3 kb of the coding sequence with the RGD domain, stop codon, and the 3'-untranslated region (1.1 kb). We have mapped two transcription start sites within the DMP1 promoter that are 280 and 321 base pairs, respectively, from the ATG start codon. The location of functional elements within the 5'-upstream DMP1 DNA fragment was determined by cloning it into a luciferase reporter gene. Transient transfection and luciferase assays revealed that the 3 kb fragment has the ability to drive the luciferase gene. However, this promoter activity was restricted to MC3T3-E1 cells (an osteoblast cell lineage). The promoter was silent in Chinese hamster ovary cells (an epithelial cell lineage), indicating the necessity of tissue-specific factors to drive the transcription.

Tooth formation is regulated by temporally and spatially restricted reciprocal interactions between the epithelium and the mesenchyme, and this signal traverses back and forth until tooth development is complete (1–3). One type of cells resulting from this process is the odontoblasts, which are the principle cells that synthesize the dentin matrix. Mature odontoblasts are responsible not only for the secretion of the collagen-rich extracellular matrix but also for the secretion of noncollagenous proteins that are responsible for initiating the mineralization cascade. In dentinogenesis, a number of highly controlled extracellular and intracellular events are responsible for the formation of the well-defined mineralized tissue. The noncollagenous proteins produced by the single layer of odontoblasts may control these events (4–6). One of the noncollagenous proteins is dentin matrix protein 1 (DMP1) (7, 8). Based on our initial data, DMP1 was thought to be dentin-specific, but later, its expression was seen in calvaria and long bone (9–11). DMP1 is an acidic protein, rich in aspartic acid, glutamic acid, and serine residues. Fifty-two percent of these serines can be phosphorylated by casein kinase I- and II-like kinases. Upon phosphorylation, DMP1 could be involved in the mineralization process along with other noncollagenous phosphorylated proteins. DMP1 also has a RGD site in its cDNA (7). We have recently reported that the RGD sequence in DMP1 functions as a cell attachment domain (12).

Little is known about the mechanism regulating the ontogeny and restricted tissue-specific expression of DMP1 gene. The cDNA sequence for DMP1 has been determined in various species, implicating the evolutionary pathway of the DMP1 gene. Cloning of DMP1 in these species has revealed the conservation of the acidic residues and serines that are strategically positioned for phosphorylation (13–15).

To understand the tissue-specific mechanisms behind the DMP1 transcription, we have attempted to characterize the promoter of the DMP1 gene. The regulatory factors that are necessary for the transcriptional control of the DMP1 gene may play a role in their tissue specificity. These factors may be cis-acting, such as promoters and enhancers, or trans-acting, such as various DNA binding factors. The differentiation of odontoblasts and secretion of the dentin matrix involve specific genes being expressed at the right time in a temporally-spatial manner (16). To date, very little is known about the recognition elements and the DNA-binding proteins that regulate the transcription of this gene.

To characterize the DMP1 gene and study the mechanisms of tissue-specific regulation, we have isolated a genomic clone containing all the introns as well as a ~3-kb segment of the 5'-flanking region that includes a functional promoter. DMP1 promoter deletion constructs were also made, and their consequent effects on the expression of a reporter gene (luciferase) were analyzed. To address the question of tissue specificity, we have studied the promoter activity profile in transiently transfected MC3T3-E1, a murine calvaria-derived osteoblast cell line, and in an epithelial cell lineage, CHO. Our ultimate goal is to identify specific transcription factors within the odontoblasts/osteoblasts that are responsible for driving the expression of DMP1.

* This work was supported by grants from NSF-MCB-9506111, and NIH-DE11657 (to A. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked ‘advertisement’ in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
‡ These authors contributed equally to this work.
§ To whom correspondence should be addressed: Dept. of Oral Biology, University of Illinois at Chicago (MC/690), 801 S. Paulina, Chicago, IL 60612. Tel.: (312) 413-0738; Fax: (312) 996-6044; E-mail: anneg@uic.edu.

1 The abbreviations used are: DMP1, rat dentin matrix protein 1; kb, kilobase(s); bp, base pair(s); CHO, Chinese hamster ovary; PCR, polymerase chain reaction.
ST1 to ST5 primers were used in combination with ST0/BglII for the PCR-mediated deletions within the DMP1 promoter. A BgIII site was incorporated into the ST0/BglIII primer for easy cloning. CP5’ and CP3’ were used for the amplification of the core promoter. ST2-Rev and ST3-Rev were used in the amplification of enhancer fragments in combination with either ST1 or ST2, and denoted as K1, K2, and K3.

<table>
<thead>
<tr>
<th>Name of the primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST1</td>
<td>-2344GTTCTGCGGCTTGGAG2330</td>
</tr>
<tr>
<td>ST2</td>
<td>-1902GACCTAGGGG1689</td>
</tr>
<tr>
<td>ST3</td>
<td>-1460CGCAGTCAAAACAC1397</td>
</tr>
<tr>
<td>ST4</td>
<td>-1775AGCCATCTTTACC759</td>
</tr>
<tr>
<td>ST5</td>
<td>-2176GCGCTGGAATAAAGCT</td>
</tr>
<tr>
<td>ST0/BglII</td>
<td>tetaatacaaaaggtccc217</td>
</tr>
<tr>
<td>CP5’</td>
<td>-2284GAGGATGAAATTCGCC271</td>
</tr>
<tr>
<td>CP3’</td>
<td>-2367CAGAATGGGGCTCTCTC271</td>
</tr>
<tr>
<td>ST2-REV</td>
<td>-1801CCCACCATAGTG1902</td>
</tr>
<tr>
<td>ST3-REV</td>
<td>-1397AGTTGTTAGCTCCG1408</td>
</tr>
</tbody>
</table>

**Table I** List of primers used in experiments

**Materials and Methods**

**Southern Blot Analysis**—Rat tail genomic DNA was isolated by following the procedure of Sambrook et al. (17). DNA (20 μg) was digested with different restriction enzymes, and the digested DNA fragments were separated on a 0.8% agarose gel and processed for Southern hybridization as described previously (18). Full-length rat DMP1 cDNA was used as a probe.

**Promoter Isolation and Analysis**—Rat DMP1 promoter was isolated by using the Rat Geno- walk kit from Clontech. PCR reactions were carried out with a gene-specific primer and an adapter primer. Gene-specific primer was made complementary to nucleotides 51–75, upstream of the translation codon ATG (5’-TGGGTCAAAAGCTCCG-3’) of the DMP1 cDNA. The PCR amplicons were cloned in to TA vector and sequenced by automated sequencing. More gene-specific primers were made, and PCR was carried until 3 kb of 5’-flanking sequences were obtained.

A PCR reaction was also carried with the rat genomic DNA in the 3’ direction with primers (forward and reverse) at nucleotides 361–378 (5’-TGGGTCAAAAGCTCCG-3’) of the DMP1 cDNA. The PCR amplicons were cloned into TA vector and sequenced by automated sequencing.

**Primer Extension Analysis**—A synthetic oligonucleotide of 21-mer (5’-GTCAGGTTCTCCCAGAGG-3’) with a sequence complementary to nucleotides 83–64 (upstream of the ATG start codon) of the DMP1 published cDNA (7) was used for primer extension analysis to determine the transcriptional start site. Total RNA was isolated from MC3T3-E1 cells (which express DMP1) by using the Trizol reagent (Life Technologies, Inc.). Yeast tRNA was used as a negative control. Primer extension analysis was carried according to Sambrook et al. (17). Briefly, a γ-32P-end-labeled primer was annealed to 20 μg of total RNA at 55 °C for 18 h. The primer-RNA complex was precipitated and resuspended with diethyl pyrocarbonate-treated water. Extension of the annealed primer was carried out using Superscript II (Life Technologies, Inc.) at 42 °C for 90 min. The cDNA was precipitated in the presence of sodium acetate and analyzed on a 6% polyacrylamide/urea gel with a sequencing reaction as a ladder. The gel was dried onto a Whatman 3MM paper and exposed to Kodak x-ray film.

**Construction of Plasmids for Transfection Assays**—A 3-kb fragment encompassing the upstream sequences of DMP1 was subcloned into a promoterless luciferase pGL3 luciferase basic vector (Promega). The pGL3-basic vector lacks eukaryotic promoter and enhancer sequences, and expression of luciferase activity in cells transfected with this plasmid depends solely on the putative regulatory DMP1 sequences that were inserted upstream to the luciferase gene.

Systematic nested deletions were made by PCR amplification using the primers listed in Table I. The PCR products were initially cloned into a TA vector (Invitrogen) and further subcloned into pGL3-basic vector (Promega).

For the enhancer experiments, the PCR-amplified DNA fragments were cloned into TA vector as mentioned above and further subcloned into pGL3-promoter (SV40 promoter). PGL3-promoter vector has a SV40 promoter to drive the luciferase gene. When compared with the control, any effect on the luciferase activity will be due to the DNA fragment cloned upstream to the SV40 promoter. All the constructs were verified for orientation by partial sequencing or enzymatic digestion.

**Cell Culture and Transfections**—CHO cells, an epithelial cell line, and MC3T3 cells (mouse calvarial 3T3-like cells), an osteoblast cell line, were grown in F-12 and Dulbecco’s modified Eagle’s medium, respectively (Life Technologies, Inc.) supplemented with 10% fetal bovine serum. For transfectional analysis, the various chimeric constructs were transfected into subconfluent cells (60–70% confluence) using the Superfect transfection reagent (Qiagen) as described by the manufacturer. The DNA-Superfectamine mixture was added to the cells, cells were incubated for 2 h, and the medium was replaced by regular medium. The cells were kept in culture for 24 h, and then the cells were lysed and clarified by centrifugation, and the luciferase assay was performed as described below. The amount of protein recovered after lysis was determined by the Bradford assay (Bio-Rad). The pRLSV-40 plasmid was co-transfected with all the transfections to normalize the variations in transfection efficiency. pRLSV-40 encodes Renilla luciferase, and its activity can be distinguished from that of the firefly luciferase encoded in pGL3 using the dual-luciferase assay system (Promega). All transfections were carried out in triplicates.

**Dual-Luciferase Activity Assay**—Activities of the firefly luciferase and Renilla luciferase in a single sample were measured using the dual-luciferase reporter assay (DLRTM) system (Promega) according to manufacturer’s instructions. In short, 50 μg of protein was dispensed into a 96-well plate and placed in a luminometer equipped with two dispensers. The first dispenser added the firefly luciferase-activating reagent, Luciferase Assay Reagent I, and luminescence was recorded. The second dispenser then added the Renilla luciferase-activating reagent, Stop and Glo Reagent, which simultaneously quenched firefly luminescence and activated the Renilla luciferase. The light emission was recorded using a Dynex luminometer. Variations in transfection efficiency were normalized by dividing the measurement for the firefly luciferase activity by that for the Renilla luciferase activity.

**Results**

**Copy Number**—To determine the copy number for the rat DMP1 gene, Southern analysis of the rat genomic DNA was performed with various restriction enzymes, and a short cDNA probe was used. A single band was detected in each lane, indicating that DMP1 is a single copy gene (data not shown).

**Gene Structure**—A promoter walking technique was used to isolate 3-kb 5’-upstream sequences. All the introns for DMP1 were also isolated from the rat genomic DNA with the Genome-walker technique. The genomic structure for DMP1 consists of six exons and five introns (Fig. 1). The first exon is 95 bp long and encodes most of the 5’-untranslated region. The second exon codes for the remainder of the 5’-untranslated region, the signal peptide and the first two N-terminal amino acids of the secreted protein. The third, fourth, and fifth exons encode for 16, 11, and 15 amino acids, respectively. The sixth and largest exon codes for the rest of the protein. The exon-intron boundaries were determined and belongs to the class 0 type. The sequences at the boundaries conformed closely to the classical GT/AG rule. There are five introns, and the sizes of the introns 1–5 are 3791, 465, 2047, 162, and 1375 bp, respectively.

**Identification of the Transcription Start Site of DMP1 by**
we have made PCR fragments between -2344 and -1902 bp (K1), -1902 and -1397 bp (K2), and -2344 bp and -1397 bp (K3). These DNA fragments were cloned into a pGL3-promoter vector (which had SV40 promoter). Any enhancer activity within these fragments will affect the luciferase activity through the SV40 promoter.

Luciferase assay result of these constructs is shown in Fig. 6. It is evident that the K1 construct has no enhancer activity, whereas K2 has increased luciferase activity (~40-fold), indicating the presence of an enhancer element within this region. The K3 construct, on the other hand, has enhancer activity of about 20-fold with respect to the pGL3/SV40 transcriptional activity (Fig. 6). The enhancer activity of these fragments was restricted to MC3T3-E1 cells. Transfections in CHO cells did not have any significant enhancer activity.

Core Promoter Identification—The minimal sequences within a promoter to drive the normal activity is called the core promoter. To identify the core promoter region, we have PCR-amplified a 300-bp fragment that spans +38 to -288 bp from the transcription start site. This fragment was cloned into the pGL3-basic vector to analyze its activity. Fig. 7 shows that the 300-bp fragment has complete promoter activity. In fact, this core promoter is almost 1-fold higher in activity than the full-length DMP1 promoter. The core promoter identified was equally active in both MC3T3-E1 and CHO cell lines.

**DISCUSSION**

DMP1, the first cloned dentin-related gene, has an overall composition between that of bone phosphoprotein and that of dentin phosphophoryn (7). The expression of DMP1 was found to be completely restricted to polarized odontoblasts and differentiated osteoblasts. To study the mechanisms regulating cell type- and differentiation-specific expression of the DMP1 gene, we have isolated and characterized the DMP1 gene along with its exon-intron boundaries and the upstream promoter region. This is the first report describing the genomic sequence and organization of the rat DMP1 gene.

DMP1 gene is a single copy gene comprised of six exons and five introns. The fifth exon is present in an isoform of DMP1 as shown in Fig. 1 and was first reported by MacDougall et al. (8) in the mouse. This additional exon containing 15 amino acids was found to be spliced out from our original cDNA clone. The functional significance of this alternate splicing has not yet been determined.

There are five introns of varying sizes, with the largest of them being the first intron. All of the introns are of the phase 0 type because they are flanked on either side by coding triplets. It is due to this type of arrangement that the presence of exon shuffling is noticed in one of the isoforms.

The isolated rat DMP1 promoter is characterized by a TATA box and an inverted CCAAT box. In our experiments, we have shown that rat DMP1 is active in the correct orientation. An inverted CCAAT box is also seen in the promoter of bone sialoprotein and chicken osteopontin (19, 20). Analysis of the promoter sequences by the MatInspector program shows that DMP1 has binding sites for other transcription factors such as SP1, AP1, GATA, CREB, MSX1, MYOD, AP4, c-MYC, ETS-1, and so forth and for serum-responsive elements such as ELK1, IRSRE, and so forth. The most common binding sites are shown in Fig. 3.

Primer extension studies using total RNA from the rat teeth/MC3T3-E1 cells have identified a major transcription start site at 280 bp upstream from the ATG site. An additional minor start site was also identified at 321 bp upstream of the ATG codon. The functional significance of the additional start site has yet to be analyzed. Similar patterns have been observed in other systems as well (20–23).
To identify cell type-specific transcription of the DMP1 gene, several deletions within the DMP1 promoter were made and analyzed for the ability to drive the luciferase gene. Luciferase assays demonstrated that the highest activity was found in the ST2 construct (with a deletion of 1 kb from the 5'-end of the DMP1 promoter). Construct ST4 (with a deletion of 2 kb from the 5'-end of the DMP1 promoter) showed a sharp drop in luciferase activity. However, the ST4 construct has a 1-fold higher activity than that of the full-length DMP1 promoter. Any further deletion of the ST4 construct showed a complete loss of promoter activity. In contrast, all the deletion constructs except ST4 showed no promoter activity in the CHO cell line. This indicates that the sequences upstream to ST4 in the DMP1 promoter construct probably contain tissue-specific elements that are necessary for the expression of DMP1 promoter in the cell lines tested. However, upon the deletion of these tissue-specific sequences, the ST4 construct was equally active in CHO cells. ST5 construct, on the other hand, failed to drive the luciferase gene in both MC3T3-E1 and CHO cell lines, possibly due to the loss of all transcription factor binding sites.

**FIG. 3.** Complete nucleotide sequence of the 5'-upstream region of the DMP1 gene showing the position of various transcription factor binding sites. The consensus sequences for the binding factors indicate the transcription start site. The core promoter region identified was between 2228 (start) and 1338 bp. The enhancer element identified was between 2190 and 1397 bp.
Overall, these results suggest that there are tissue-specific regulatory elements within the 2 kb at the distal end of the DMP1 promoter. Furthermore, the above results also suggest that the 1-kb fragment at the 3′-end of the DMP1 promoter has the necessary sequences for a minimal promoter. Studies are being conducted to identify some of the tissue-specific factors that can control the DMP1 gene expression.

The core/minimal promoter sequence resides between +38 and −288 bp with respect to transcription start site. The luciferase assay results show that this 300-bp fragment is sufficient to drive the reporter gene. This core promoter is equally active in CHO and MC3T3-E1 cells, indicating that the DMP1 promoter lost its tissue specificity.

Data obtained from the serial deletions indicated the presence of an enhancer element within the DMP1 promoter. A 1-kb deletion at the 5′-end of the DMP1 promoter sharply increased its promoter activity by severalfold. This result prompted us to investigate the presence of any enhancer/repressor in and around this region. Luciferase assay showed the presence of an enhancer at −1902 to −1397 bp of the DMP1 promoter. Furthermore, transfection into CHO cells of the same constructs has no effect on the SV40 promoter activity. These results indicate that the enhancer is cell type-specific.

In summary, our results have identified cell type-specific regions in the distal part of the DMP1 promoter. These findings will serve as a basis for future studies aimed at identifying the regulatory sequences and specific transcription factor(s) within odontoblasts or osteoblasts that are responsible for controlling the tissue-specific and developmental regulation of the DMP1 gene. We are also attempting to identify the tissue-specific enhancer binding factors responsible for the expression of DMP1 in mineralized matrix.

Acknowledgement—We thank Dr. Vani Turumalla for help with primer extension assay.
REFERENCES

Cloning and Characterization of Rat Dentin Matrix Protein 1 (DMP1) Gene and Its 5′-Upstream Region
Seetha R. Thotakura, Narayanan Karthikeyan, Taryn Smith, Kai Liu and Anne George

doi: 10.1074/jbc.275.14.10272

Access the most updated version of this article at http://www.jbc.org/content/275/14/10272

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 19 references, 1 of which can be accessed free at http://www.jbc.org/content/275/14/10272.full.html#ref-list-1