Communication

Sequence Determination of an Extremely Acidic Rat Dentin Phosphoprotein*

(Received for publication, May 17, 1996, and in revised form, July 9, 1996)

Helena H. Ritchie‡ and Lee-Ho Wang§

From the ‡Department of Pediatrics, the University of Iowa, Iowa City, Iowa 52242 and the §Division of Hematology, the Department of Internal Medicine, the University of Texas Medical School, Houston, Texas 77030

The mineralization process associated with the conversion of predentin to dentin is believed to be initiated and controlled by a set of acidic regulatory noncollagenous proteins (NCPs) which include phosphophoryn, the major NCP in dentin. Phosphophoryn binds tightly to collagen and is believed to initiate the formation of apatite crystals which play a central role in the mineralization process. During the process of analyzing the 3′ end of an odontoblast-specific cDNA which codes for dentin sialoprotein (Ritchie, H. H., Hou, H., Veis, A., and Butler, W. T. (1994) J. Biol. Chem. 269, 3698–3702), we discovered a 801-base pair open reading frame. This downstream open reading frame encodes a putative leader sequence and a very acidic mature protein sequence having a deduced amino acid composition containing high percentages of both Ser (43%) and Asp (31%) residues which closely coincides with the amino acid composition of phosphophoryns from human, bovine, rat, and rabbit (i.e. Asp (30–40%) and Ser (38–50%)). This newly identified cDNA therefore encodes a protein with characteristics similar to phosphophoryn. Here we present the cDNA sequence, the deduced amino acid sequence, and the prospective Ser residue-specific casein kinase I and II phosphorylation sites for this putative phosphophoryn.

The calcification process that accompanies the transition of predentin to dentin is poorly understood, due in part to the difficulties in isolating and characterizing unique sets of extracellular matrix molecules that contribute to this complex process (1–4). Phosphophoryn, the most abundant noncollagenous protein in dentin, is secreted by odontoblasts through odontoblastic processes and appears at the mineralization front within a short time after labeling with [33P]phosphate (5, 6). Phosphophoryn's ability to target mineralization and participate in the formation of apatite crystals. For example, Linde and co-workers (9) have demonstrated that when phosphophoryn is immobilized on a stable support and incubated in physiological solutions of calcium and phosphate, phosphophoryn induced the formation of hydroxyapatite (HAP). Studies by the same group (10) and by Boskey et al. (11) also suggested a dual role for phosphophoryn as both an initiator of HAP formation at low phosphophoryn concentrations and as an inhibitor of HAP formation at higher phosphophoryn concentrations.

Phosphophoryn is also believed to have a specific affinity for collagen (2, 12, 13) which comprises as much as 80% of the protein in dentin. Furthermore, phosphophoryn was found to be specifically associated with the “e” band of collagen (14). This site-specific protein–protein interaction, coupled with phosphophoryn’s ability to initiate or inhibit HAP formation when calcium is present, has lead to the currently accepted view that phosphophoryn plays a central role in the mineralization process by virtue of its ability to target mineralization to selected sites as well as to couple mineralization, temporally, with organ development.

While only several N-terminal residues and small proteolytically obtained segments of internal amino acid sequences of phosphophoryn are currently known, the complete amino acid sequence of phosphophoryn has not as yet been reported. During the process of analyzing the 3′ end of dentin sialoprotein (DSP) cDNA, another recently cloned dentin-specific protein (15), an open reading frame with a size of 801 bp was revealed. This open reading frame was found to encode a putative leader sequence and a deduced very acidic mature protein sequence with an amino acid composition comprised primarily of Ser (43%) and Asp (31%) residues which coincides with the amino acid composition of phosphophoryns from human, bovine, rat, and rabbit (i.e. Asp (30–40%) and Ser (38–50%)). Here we present the first reported cDNA sequence, the deduced amino acid sequence, and the postulated Ser residue-specific casein kinase I and II phosphorylation sites for this putative phosphophoryn.

EXPERIMENTAL PROCEDURES

RNA Preparation and Reverse Transcription-PCR—The total RNA was extracted from adult rat incisors using RNAzol™ (Biotech Laboratories, Inc., Houston, TX). A cDNA pool was synthesized from total RNA using an oligo(dT) primer and reverse transcriptase. This cDNA pool was then denatured at 95°C for 5 min and amplified with the primer set comprising an oligonucleotide corresponding to rat DSP cDNA nucleotide sequence 1054–1069 (15) and a poly(dT) primer. PCR was then performed as follows: denaturation (1 min at 94°C), reannealing (1 min at 56°C), and amplification (3 min at 65°C), for 40 cycles. DNA Sequencing—The PCR products recognized by the 3′ end rat DSP probe were subcloned into TA vectors using standard techniques (22). Following company procedures, Erase-A-Base Kit (Promega, WI) was used to generate unidirectional deletions of the 2-kb insert for DNA sequencing. DNA was sequenced according to Sanger et al. (23).

Northern Blot Analysis—The total RNA from rat incisor was electro-phoresed using a 1.2% agarose gel containing 2.2 M formaldehyde (22). RNA was transferred onto a nitrocellulose paper and hybridized overnight with a 32P-labeled putative rat phosphophoryn probe at 42°C in 50% formamide and 6 × SSC. The filter was washed with 2 × SSC and

* This work was supported by National Institutes of Health Grant DE11442–01 (to H. H. R.). 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.

The nucelotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U63111.

§ To whom correspondence should be addressed: BB135, Dept. of Pediatrics, the University of Iowa, Iowa City, IA 52242. Tel.: 319-353-2055; Fax: 319-335-4489; E-mail: drichtie@tic-pot.cic.uiowa.edu.

21695
Rat Dentin Phosphophoryn

**Fig. 1.** The DNA sequence of the open reading frame for putative phosphophoryn and its deduced amino acid sequence. The arrow shows the translation start site (ATG; code for putative initiation codon Met) located at nucleotide position 43. Brackets enclose the 27-amino acid signal peptide sequence. Underlined 4-amino acid sequence is identical to the reported DSP sequence.

**Table I.** Amino acid composition of deduced rat phosphophoryn versus native rat, rabbit, bovine, and human phosphophoryns.
Rat Dentin Phosphophoryn

Fig. 2. A, the potential casein kinase I (*1) and II (*2) phosphorylation sites for the putative phosphophoryn. *1a, primary target site (i.e. (Asp/Glu)-X-Ser) for casein kinase I (29, 36). *1b–*1r, secondary, tertiary (and so forth) sites (i.e. Ser(P)-X-Ser) for casein kinase I (35); *2a, primary target site (i.e. (Ser/Thr)-X-Ser) for casein kinase II (29, 39–41); *2b–*2r, secondary, tertiary (and so forth) sites for casein kinase II (i.e. (Ser/Thr)-X-Ser(P)). B, the acidic residues and potential acidic regions generated by casein kinases I and II are depicted in boxes. Many of these acidic regions consist of (DD)\_n, (pSpS)\_n, (pSD)\_n, units (where \(n = 1, 2, \) or 3). Several significant acidic domains containing for example 21 residues (96–116), immediately followed by a series of 2 residues (i.e. mainly pSD) repeats (118–155) and 24 residues (176–199) are particularly evident.

X-Ser. When this secondary target site is phosphorylated, it then determines the tertiary Ser site for casein kinase I and so forth. For example, Ser\(^{103}\) is a primary target for casein kinase I. Based on this phosphorylation mechanism, once Ser\(^{103}\) is phosphorylated, it can trigger phosphorylation of the following 17 target Ser sites ending at Ser\(^{151}\) (1a through 1r) (see Fig. 2). When all those subsequent secondary, tertiary . . . etc. sites are included, a total of 66 potential phosphorylation sites may be available for casein kinase I.

Casein Kinase II Sites—This rat phosphophoryn contains 23 potential primary casein kinase II sites ((Ser/Thr)-X-(Glu/Asp)). Many of these sites overlap with the sites for casein kinase I, and 13 sites are specific for casein kinase II. It is plausible that the overlapping sites ensure the phosphorylation of specific phosphophoryn domains that may be crucial to collagen and/or calcium binding activities during dentinogenesis (2, 4, 7, 9, 11–13) (see Fig. 2). For example, Ser\(^{154}\), once phosphorylated, becomes the primary target site of casein kinase II and enables the phosphorylation of Ser\(^{151}\). Following this mechanism, the casein kinase II could then phosphorylate the following 17 serines (spaced every two amino acids upstream from Ser\(^{155}\)) ending at Ser\(^{213}\) (2a–2r in Fig. 2). Overlapping kinase I and II activities within this Ser-rich domain (103–154) could therefore provide a "safeguard" mechanism to ensure the phosphorylation of this particular domain. There are, in total, 55 potential casein kinase II sites of which 37 sites overlap with casein kinase I.

During the process of analyzing the 3' end of an odontoblast-specific cDNA which codes for dentin sialoprotein (15), we discovered an open reading frame with a size of 801 bp. Our newly discovered cDNA was found to encode a novel rat dentin protein whose characteristics are in solid agreement with the following reported features for phosphophoryn: (i) the predicted NH\(_2\)-terminal amino acid sequence (i.e. Asp-Asp-Pro-Asn) is identical to that derived from protein microsequencing for one form of phosphophoryn (25), (ii) the six NH\(_2\)-terminal amino acid sequence (i.e. Asp-Asp-Pro-Asn-Ser-Ser) obtained from our putative phosphophoryn (Fig. 1) agrees with that reported by Reynolds and co-workers (33) (i.e. Asp-Ser-Pro-Asn-Ser-Pro-Ser) for bovine phosphophoryn, (iii) amino acid sequences of Asp-Ser and Ser-Ser were found interspersed in our putative rat phosphophoryn. Additionally, a sequence of Asp-Ser-
Ser-Ser-Ser was identified in our putative protein. These sequence combinations of Asp-Ser, Asp-Ser-Ser, and Asp-Ser-Ser were also observed in the NH2-terminal 50 residues of bovine phosphophoryn (33), and (iv) its deduced amino acid composition contained high percentages of Ser (43%) and Asp (31%) residues which coincided with the amino acid composition of phosphophoryns from human, bovine, rat, and rabbit (i.e. Asp (~30–40%) and Ser (~38–50%) (2), 16–21).

Phosphophoryn is the most acidic protein so far discovered. The interspersed arrangement of Ser and Asp residues enables phosphophoryn to be an excellent substrate for casein kinases I and II phosphorylation action. The 78% of Ser residues in this protein which can potentially be phosphorylated coincide with the reported 85–87% of Ser(P) in authentic phosphophoryn (21). As discussed previously (see "Results"), many of these phosphorylation reactions occur at secondary and tertiary Ser sites and therefore result from a phosphorylation cascade-type mechanism involving casein kinases I and II operating over similar domains but in opposite directions to ensure complete phosphorylation within these specific Ser-rich domains.

Roach and co-workers (34–36) have reported that threonine can also serve as a substrate for casein kinase I. Based on this information, all 9 threonine residues and an additional 10 serine residues could also be phosphorylated. This mechanism could enable the sequential phosphorylation of a 47-residue acidic patch extending from Ser396 to Ser412. In this case, 88% of the Ser residues (91 out of 103) could potentially become phosphorylated by casein kinases I and II.

The full phosphorylation of the presumed phosphophorylatable serines in phosphophoryn therefore leads to the generation of many acidic patches consisting of (DD)$_n$, (pSpS)$_n$, (pSD)$_n$ repeat units. Furthermore, it was reported that bovine phosphophoryn can undergo a conformational folding in the presence of CdII and a pH-dependent conformational folding (37, 38). These folding experiments, suggesting that bovine phosphophoryn was comprised of (DD)$_n$, (pSpS)$_n$, and (pSD)$_n$ structures arranged into polyelectrolyte cluster regions (37), are in agreement with the predicted (DD)$_n$, potential (pSpS)$_n$, and (pSD)$_n$ acidic patches shown for our putative phosphophoryn (Fig. 2). Taken together, our deduced acidic protein likely represents one form of rat phosphophoryn.

By using a cDNA probe containing only the phosphophoryn DNA sequence beginning from position 208, we determined by Northern blot analysis whether this putative phosphophoryn cDNA was indeed transcribed in tooth germ. Multiple transcripts, sized around 4.6 kb, were detected (Fig. 3). The presence of multiple phosphophoryn transcripts may be due to more than one phosphophoryn gene or due to alternative splicing. However, it is equally likely that these multiple transcripts are due to the use of multiple polyadenylation signals, similar to many other mRNAs encoding extracellular matrix proteins. Further experiments are needed to determine the origin of these multiple transcripts. However, our Northern blot strongly suggests that the putative phosphophoryn mRNA was present in the rat tooth germ total RNA pool. Therefore, the DNA sequence for this novel protein is unlikely to be an artifact. Taken together, we strongly feel that rat tooth germs do actively synthesize the mRNA of this putative phosphophoryn.

The presence of both DSP and phosphophoryn DNA sequences in the PCR product could be due to an artifact generated during reverse transcription-PCR or subsequent cloning processes. However, the possibility of a bicstronic gene could not be excluded. Further work, such as the examination of these two genes at the genomic level, is needed to investigate these possibilities.

Acknowledgment—We thank Dr. David G. Ritchie for helpful discussion and contributions during the preparation of this manuscript.

REFERENCES