The Carboxyl-terminal Domain of Phosphophoryn Contains Unique Extended Triplet Amino Acid Repeat Sequences Forming Ordered Carboxyl-Phosphate Interaction Ridges That May Be Essential in the Biomineralization Process*

(Received for publication, August 23, 1996, and in revised form, October 1, 1996)

Anne George, Leslie Bannon, Boris Sabsay, Jerry W. Dillon, James Malone, and Arthur Veis‡
From Northwestern University, Chicago, Illinois 60611

Nancy A. Jenkins§, Debra J. Gilbert§, and Neal G. Copeland§
From the Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI, National Institutes of Health, Frederick Cancer Research and Development Center, Frederick, Maryland 21702

Phosphophoryns (PPs), a family of Asp and Ser(P)-rich dentin proteins, are considered to be archetypal regulators of several aspects of extracellular matrix (ECM) biomineralization. We have cloned a rat incisor PP gene, Dmp2, from our odontoblast cDNA library and localized it to mouse chromosome 5q21 within 2 centimorgans of Dmp1, another tooth-specific ECM protein. The carboxyl-terminal region of Dmp2 protein (69 residue % Ser, 31 residue % Asp) is divided into two domains, one with unique repetitive blocks of [DS]n, the other with [SD]n. Conformational analysis shows the phosphorylated form of the [DS'S'], repeats to have a unique structure with well defined ridges of phosphates and carboxyls available for counter ion binding. The [SD]n domains have different phosphate and carboxylate interaction edges and thus different calcium ion and apatite surface binding properties. These two domains and the colocalization of Dmp1 and Dmp2 genes at a position equivalent to the dentinogenesis imperfecta type II location on human 4q21 all suggest that the PPs are indeed involved in some aspect of ECM mineralization.

The phosphophoryns (PPs)1 (1) are a family of highly phosphorylated proteins with a unique composition and amino acid sequence. Found as the major noncollagenous proteins of the mineralized dentin matrix, the PPs have been considered to be the archetype of macromolecules that might regulate biomineralization (2, 3) by binding to the matrix of structural proteins (4), nucleating mineralization (5, 6), and controlling crystal growth (7). To understand how the PPs might carry out these diverse functions, it is crucial to know their amino acid sequences and structures. The PP from various species contain from 35 to 45 residue % aspartic acid and 40 to 55 residue % serine, of which as many as 90% of the Ser residues are phosphorylated (8, 9). The difficulties of Edman sequencing of phosphoserine-rich proteins have limited the information attainable by conventional sequencing efforts (10–12) to a few internal acidic domains and a short amino-terminal sequence. Recently, chemical sequencing using means to derivatize the phosphoserines has provided a longer NH2-terminal sequence (13). Sequencing via cloning of a PP cDNA has also been difficult. A putative PP cDNA was described (14) several years ago, but no sequence data have been reported. Nucleotide probes have generally been unsuccessful because of the high redundancy of the Ser codons (15). Antibody screening of cDNA libraries was not fruitful because of the difficulties of producing high titer PP antibodies (16, 17).

We have recently taken a new approach to the production of an antibody to a highly purified, dephosphorylated phosphophoryn (dPP) and, as described below, have been successful in using that antibody to identify several cDNAs in our established rat incisor odontoblast cDNA library (15, 18). Sequencing of the clones obtained thus far yielded the expected Asp-Ser-rich composition but, more importantly, revealed both an unexpected, unique “triplet cassette” repeat motif as well as a “doublet” repeat motif. These might account for several of the biomineralization regulatory functions of the phosphophoryns. We report here the nature of these unique sequence blocks. Furthermore, the identification of a PP cDNA clone has allowed the chromosomal localization of the corresponding PP gene. As shown below, the PP gene is colocalized with the gene for a mineralization disorder, dentinogenesis imperfecta, type II, correcting an erroneous assignment (14).

EXPERIMENTAL PROCEDURES

PP and Antibody Preparation—PP was extracted from rat incisors by the procedure of Rahima and Veis (9). The high M, PP components were isolated by CaCl2 precipitation from the crude extract, followed by chromatography on a Bio-Gel DEAE-5 PW preparative column. Analytical gel electrophoresis showed the presence of three bands, with molecular mass 95, 90, and 85 kDa. Preparative gel electrophoresis followed by electrodialysis was used for isolation of the 95-kDa protein band. The PP95K was dephosphorylated with potato acid phosphatase (19). The dephosphorylated PP95K (dPP95K) was isolated by a second round of electrophoresis followed by electrodialysis, was coupled to keyhole limpet hemocyanin (KLH) using carbodiimide. The KLH-dPP95K conjugate was then injected into a goat. Booster injections followed at 16 and 25 days, and the immune serum was collected at day 42. The antibody was concentrated by ammonium sulfate precipitation and...
then passed over an affinity column of dPP$_{95k}$, coupled to Sepharose 4B. The purified antibody was eluted.

**Library Screening, Nucleotide Sequencing, Northern Analysis, and Tissue and Chromosomal Localization**—Our aig11 rat odontoblast expression library (15) was amplified on LB/agar plates and screened using standard procedures. The translation product is expressed as a β-gal-cDNA fusion protein when the lacZ gene is induced with isopropyl-1-thio-β-D-galactopyranoside. The fusion proteins were adsorbed to Hybond nylon membranes, washed for 30 min in 50 mM Tris-buffered saline (TBS), 0.05% Tween 20 (TTBS), drained, and blocked at 4°C in 4% casein solution. The stock anti-dPP$_{95k}$ antibody was added at 1:100 dilution in TBS, drained, and incubated for 3 h with peroxidase-conjugated rabbit anti-goat IgG at 1:1000. After washing in TTBS, the membranes were developed with 4-chloro-naptholin methanol/TBS. Positive clones were identified, collected by stab, and stored in suspension media at 4°C. The clones were amplified on polymerase chain reaction using aig11 forward and reverse primers and subcloned into pGEM-T plasmid vector (Promega). The Sequenase 2.0 (U.S. Biochemical) double stranded DNA sequencing protocol was used. Electrophoresis was on 6% polyacrylamide gels according to standard protocols.

An Invitrogen mRNA isolation kit was used to obtain rat incisor odontoblast pulp, heart, liver, and muscle mRNA. For each tissue 1 μg mRNA was electrophoresed on 1% formaldehyde gels, transferred to Hybond N, and fixed by baking at 80°C. The membranes were prehybridized and hybridized by the Stratagene Quickhybridization method. A [32P]dCTP-labeled probe was obtained using random priming of clone 2PP with a nick translation labeling kit (Boehringer Mannheim). After hybridization the filters were washed at 65°C with 1 × SSC, 0.1% SDS, and then autoradiographed.

The chromosomal location of the mouse PP gene, designated as Dmp2, dentin matrix protein 2 (the symbol has been used to designate another protein), was determined by interspecific backcross analysis using progeny derived from matings of C57BL/6J x Mus spretus. Several strains were used in this backcross mapping (20). The clones were amplified by polymerase chain reaction using the anti-dPP$_{95k}$ antibody and hybridized by the Stratagene Quickhybridization method. A [32P]dCTP-labeled probe was obtained using random priming of clone 2PP with a nick translation labeling kit. After hybridization the filters were washed at 65°C with 1 × SSC, 0.1% SDS, and then autoradiographed.

RESULTS AND DISCUSSION

**Structure Analysis of Model Peptides**—The phosphophoryn family was modeled by energy minimization using the DISCOVER CVFF repetitive building program (BIOVYM software) on a Silicon Graphics R-4000, XZ graphics workstation.

**RESULTS AND DISCUSSION**

**PP Sequence and Domain Structure**—The acid phosphatase was ~95% effective in dephosphorylating the PP$_{95k}$. The dPP$_{95k}$ migrated as a single sharp blue Stains All stained band on the preparatory gels and was easily eluted. The KLH-dPP$_{95k}$ conjugate enhanced the immunogenic response to dPP$_{95k}$, and the anti-dPP$_{95k}$ was reactive with both dPP$_{95k}$ and PP$_{95k}$. The anti-dPP$_{95k}$ was readily purified from an ammonium sulfate precipitate of the crude serum by passage over a PP$_{95k}$ Sepharose 4B affinity column. At a 1:100 dilution of the antiseraum, it was possible to detect as little as 0.1 nmol of PP$_{95k}$ on slot blots.

This portion of the PP sequence can be divided in two domains, from −245 to −75, and from −74 to −1. The first domain of 171 residues contains the [DSS]$_{n=2}$ repeats, with n as large as 14, and only 4 residues different from Asp or Ser. The most carboxyl-terminal domain, from −74, has no extended DSS domains and contains all of the other non-Asp or -Ser residues. As noted by the underlines, there are five [SD] repeat sequences, [SD]$_{m}$, with m = 2 or 3. These data are in agreement with the conclusions of Sabsay et al. (11), who examined peptides released by enzymatic digestion and partial acid hydrolysis of intact phosphorylated PP. They had concluded that the end regions had to contain sequences containing the majority of the non-Ser and -Asp residues and flanked a central core composed almost completely of Ser and Asp (Asn).

The existence of these two sequence domains is evidence of the structural conservation. Because the PP is about 90% phosphorylated as isolated from dentin, the (DSS)$_{m}$ domain sequence structure was modeled as a completely Ser
Phosphorylated polypeptide, \( (\text{DS}^*\text{S}^*)_{12} \) by energy minimization on a Silicon Graphics R-4000, X/Z graphics workstation using the DISCOVER CVFF force field and automatic set parameters for 10,000 iterations, until the system energy converged at \(-207\) kcal/mol. The CVFF repetitive build program (25) is a generalized force field with parameters specifically designed to minimize polypeptide conformations and is part of the IN-SIGHT II 950 molecule modeling system from BIOSYM/MSI (San Diego, CA). The potential electrostatic repulsions of the charged phosphate and carboxylate groups were taken into account, along with the other backbone and side chain interactions. The most probable structure is a nonplanar, folded, modified trans-extended backbone chain as depicted in Fig. 2A. Three features are evident. Because of the trans-peptide backbone, the adjacent \( \text{S}^* \) phosphates in each triplet are on opposite edges of the chain, creating two ridges of phosphates. These ridges can be seen best in the end view in Fig. 2B. Within the repetitive [DSS] domains, the most important feature is the [SDS] repeat, which creates pairs of phosphates along the same edge of the extended chain, separated by a single amino acid residue (Fig. 2A). The pairs of phosphates alternate in direction along the backbone. The phosphates in each pair are separated by 5.1 Å, but the pairs are separated

---

**Fig. 2.** The energy minimized structure of a proposed fully phosphorylated \( (\text{DS}^*\text{S}^*)_{12} \) domain of 2PP in the absence of counterion binding. **A**, a lateral view of the extended chain through three \( \text{DS}^*\text{S}^* \) repeats, showing the folding or puckering of the backbone and the relative dispositions of the phosphate (purple) and carboxyl (green and red) groups relative to the backbone. The alignment of the phosphoserines in clusters along interaction edges offer the potential for calcium binding and mineral orientation. Note that the adjacent phosphates on the same edge reside on adjacent triplets, whereas the carboxyl group closest to each pair of phosphates is in the next triplet. If one numbered a sequence \( \text{D}^*\text{S}^*\text{D}^*\text{S}^*\text{D}^* \), the carboxyl of \( \text{D}^* \) would be closest to the phosphates on \( \text{S}^* \) and \( \text{S}^* \). Thus, the creation of the near neighbor ionic concentration of two phosphates and one carboxylate requires the components of three successive triplets. **B**, an end view of the minimized structure emphasizes the formation of the ionic ridges along the peptide backbone, with the parallel and close interaction between carboxyl and phosphate ridges. **C**, the minimized structure of the \( \text{S}^*\text{DS}^*\text{DS}^*\text{D} \) domain approximates a trans-extended chain. The carboxyls and phosphates line up on opposite sides of the chain, presenting very different ionic distributions than that along the \( \text{DS}^*\text{S}^* \) repeats.
along the same edge by about 18 Å. Likewise, the carboxylate edges are very specifically structured and spaced along the peptide backbone. The repetitive DSS motif aligns a carboxyl of the next triplet with each pair of phosphates in the combined phosphate-carboxylate ridge (Fig. 2B). It will be important to investigate the potential calcium ion binding properties of such domains.

On the other hand, the [SD]m sequences are quite different. These have only two ionic edges in the trans-extended chain conformation, with all of the phosphates on one edge and the carboxylates on the other (Fig. 2C).

Tissue and Chromosomal Localization—At the biochemical level, the PPs have been found only in mineralized dentin. To verify that the 2PP was expressed only in the teeth, a 2PP probe was used in Northern analysis to examine the RNA from a variety of tissues. The probe detected a single RNA band (Fig. 3) at ~7–7.5 kb in the odontoblast pulp mRNA, but none of the other tissue mRNAs, including brain, liver, and bone, showed any reactivity (data not shown), indicating that 2PP message level was below detection limits (data not shown).

The chromosomal backcross mapping results indicated that the Dmp2 gene is located in the central region of mouse chromosome 5, linked to Fgf5, Dmp1, and Adrbk2. Although 111 mice were analyzed for every marker and are shown in the segregation analysis (Fig. 4), up to 191 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data (24). The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci, and the most likely gene order is: centromere-Fgf5-4/191-Dmp2-0/150-Dmp1-2/114-Adrbk2. The recombination frequencies (expressed as genetic distances in centimorgans ± standard error) are Fgf5-2.1 ± 1.0, Dmp1-1.8 ± 1.2 Adrbk2. No recombinants were detected between Dmp1 and Dmp2 in 150 animals typed in common, suggesting that the two loci are within 2.0 centimorgans of each other (upper 95% confidence limit).

Our interspecific map of chromosome 5 has been compared with a composite mouse linkage map that reports the map location of many uncloned mouse mutations (Mouse Genome Database, The Jackson Laboratory, Bar Harbor, ME). The central region of mouse chromosome 5 shares homology with human chromosomes 4q and 22q (summarized in Fig. 4). In particular Fgf5 has been placed on human chromosome 4q21 and Adrbk2 has been linked to human 22q11. The co-localization of Dmp1 and Dmp2 in mouse suggests their colocalization on human chromosome 4q21 as well.

Alpin et al. (26) have confirmed that Dmp1 is on chromosome 4q21 in the human. Thus the above data demonstrate that the clone 2PP represents a tooth (dentin pulp)-specific mRNA, whose gene, Dmp2, probably co-localizes on human chromosome 4q21 with the gene for another tooth dentin-specific protein, Dmp1 (23). The inherited genetic disorder of dentin mineralization, dentinogenesis imperfecta, type II, has also been linked to human chromosome 4q21 (27). Bone sialoprotein and osteopontin have also been localized to this region (28), but neither has shown any defect relative to dentinogenesis imperfecta, type II (29). These data show that the prior report (30) that the PP gene was not on 4q21 and not related to dentinogenesis imperfecta, type II, was in error because of misidentification of the PP gene.

Potential Function of PP—The unique nature of the phosphophoryns and many studies showing that in vitro they can modulate phosphate crystal nucleation and regulate crystal growth (5–7) have led to numerous discussions of the potential role of the phosphophoryns in vivo and to the discussion of this group of molecules as models for protein modulations of biomineralization (2, 31). In spite of the monotony of Ser and Asp composition, the present work demonstrates that the phos-
Phosphophoryn molecule clearly has Ser- and Asp-rich domains of different structure. The unique [DSS]$_n$ triplet repeats create domains with very specific yet unique patterns of phosphates and carboxylates. These distinctive ionic patterns may be involved in either the site-specific binding of phosphophoryn to collagen fibrils (32), nucleation of crystallization (6), or the regulation of the crystal growth patterns (7). Thus, it will now be important to determine, with recombinant constructs of these domains, their interaction with calcium phosphate crystals and crystallization. Further, as sequencing through the NH$_2$-terminal domains is completed, these and other constructs of the domains of the PP can be explored with regard to the mechanism of their phosphorylation. Work is already in progress to verify the relationship between tooth mineralization disorders and PP content and expression. These studies should ultimately reveal the function of this archetypal mineralized tissue-related molecule.

REFERENCES

1. Dimuzio, M. T., and Veis, A. (1978) Calcif. Tissue Res. **25**, 169–178
2. Heuer, A. H., Fink, D. J., Laria, V. J., Arias, J. L., Calvert, P. D., Kendall, K., Messing, G. L., Blackwell, J., Reike, P. C., Thompson, D. H., Wheeler, A. P., Veis, A., and Caplan, A. I. (1992) *Science* **255**, 1098–1105
3. Falini, G., Albeck, S., Weiner, S., and Addadi, L. (1996) *Science* **271**, 67–69
4. Stetler-Stevenson, W. G., and Veis, A. (1986) Calcif. Tissue Int. **38**, 135–141
5. Veis, A., and Sabsay, B. (1982) in *Biomineralization and Biological Metal Accumulation* (Westbrook, P., and deJong, E. W., eds) pp. 273–284, D. Reidel Publishing Co., Dordrecht, Holland
6. Linde, A., Lussi, A., and Crenshaw, M. A. (1989) *Calcif. Tissue Int.* **44**, 286–295
7. Furedi-Milhofer, H., Moradian-Oldak J., Weiner, S., Veis, A., Mintz, K. P., and Addadi, L. (1994) *Connect. Tissue Res.** **30**, 251–264
8. Stetler-Stevenson, W. G., and Veis, A. (1983) *Biochemistry* **22**, 4326–4335
9. Rahima, M., and Veis, A. (1988) *Calcif. Tissue Int.* **42**, 104–112
10. Butler, W. T., Bhown, M., Dimuzio, M. T., Cotran, W. C., and Linde, A. (1993) *Arch. Biochem. Biophys.* **225**, 178–186
11. Sabsay, B., Stetler-Stevenson, W. G., Lechner, J. H., and Veis, A. (1991) *Biochem. J.* **276**, 699–707
12. Chang, S. R., Chiego, D., Jr., and Clarkson, H. B. (1996) *Calcif. Tissue Int.*, **59**, 149–153
13. Crossley, M. A., Huq, N. L., Kirszbaum, L., and Reynolds, E. C. (1996) *J. Dent. Res.* **75**, 154
14. MacDougall, M., Slavkin, H. C., and Zeichner-David, M. (1992) *Biochem J.* **287**, 651–655
15. George, A., Sabsay, B., Simonian, P. A. L., and Veis, A. (1993) *J. Biol. Chem.* **268**, 12624–12630
16. Rahima, M., Tsay, T.-G., Andujar, M., and Veis, A. (1988) *J. Histochem. Cytochem.* **36**, 153–157
17. Babie, A. M., and Veis, A. (1995) *Connect. Tissue Res.* **31**, 197–209
18. Ritchie, H., Hsu, H., Veis, A., and Butler, W. T., (1994) *J. Biol. Chem.* **269**, 3698–3702
19. Bessey, O. A., Lowry, O. H., and Brock, M. J. (1946) *J. Biol. Chem.* **164**, 321–329
20. Copeland, N. G., and Jenkins, N. A. (1991) *Trends Genet.* **7**, 113–118
21. Jenkins, N. A., Copeland, N. G., Taylor, B. A., and Lee, B. K. (1982) *J. Virol.* **43**, 26–36
22. Benovic, J. L., Nolan, J. A., Ariza, J. L., Stone, W. C., Lohse, M., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., Caron, M. G., and Lefkowitz, R. J. (1991) *J. Biol. Chem.* **266**, 14939–14946
23. George, A., Gui, J., Jenkins, A. N., Gilbert, J. D., Copeland, G. N., and Veis, A. (1994) *J. Histochem. Cytochem.* **42**, 1527–1531
24. Green, E. L. (1981) in *Genetics and Probability in Animal Breeding Experiments*, pp. 77–113, Oxford University Press, New York
25. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., and Hagler, A. T. (1988) *Protein Struct. Funcit. Genet.* 4, 31–47
26. Alpin, H. M., Hirst, K. L., Croby, A. H., and Dixon, M. J. (1995) *Genomics* **30**, 347–349
27. Crosby, A. H., Scherphier-Heddena, T., Wijmenga, C., Altherr, M. R., Murray, J. C., Buetow, K. H., and Dixon, M. J. (1995) *Am. J. Hum. Genet.* **57**, 832–839
28. Goold, R. D., diSilvio, G. L., Xu, H., Lang, D. B., Dadgar, J., Magrane, G. G., Dugaiczyk, A., Smith, K. A., Cox, D. R., Masters, S. B., and Myers, R. M. (1993) *Hum. Mol. Genet.* **2**, 1271–1288
29. Crosby, A. H., Edwards, S. J., Murray, J. C., and Dixon, M. J. (1995) *Genomics* **27**, 155–160
30. MacDougall, M., Zeichner-David, M., Murray, J., Crall, M., Davis, S., and Slavkin, H. C. (1992) *Am. J. Hum. Genet.* **50**, 190–194
31. Lowenstam, H. A., and Weiner, S. (1989) *On Biomineralization*, Oxford University Press, New York
32. Traub, W., Jodaikin, A., Arad, T., Veis, A., and Sabsay, B. (1992) *Matrix* **12**, 197–201