Molecular cloning of matrix Gla protein: Implications for substrate recognition by the vitamin K-dependent \(\gamma\)-carboxylase

\((\gamma\)-carboxyglutamatic acid/bone/protein carboxylation\)

**PAUL A. PRICE**, **JAMES D. FRASER**, and **GABRIELLE METZ-VIRCA**

Department of Biology, University of California at San Diego, La Jolla, CA 92039

Communicated by Russell F. Doolittle, August 14, 1987

**ABSTRACT** Matrix Gla protein (MGP), a low molecular weight protein found in bone, dentin, and cartilage, contains 5 residues of the vitamin K-dependent amino acid \(\gamma\)-carboxyglutamic acid (Gla). We have used antibodies raised against MGP and oligonucleotide probes to screen a \(\lambda gt11\) cDNA library constructed from the rat osteosarcoma cell line (line ROS 17/2) that had been pre-treated with 1a,25-dihydroxyvitamin D\(_3\). By sequencing several cloned cDNAs, we established a 523-base-pair sequence that predicts an 84-residue mature MGP and a 19-residue hydrophobic signal peptide. The 84-residue mature rat MGP predicted from the cDNA sequence has an additional 5 residues at its C terminus (--Arg--Arg--Gly--Gla--Lys) not seen in the sequence of MGP isolated from bovine bone. The structure of rat MGP provides insight into the mechanisms by which the vitamin K-dependent \(\gamma\)-carboxylase recognizes substrate. The present study shows that MGP, unlike other vitamin K-dependent proteins, lacks a propeptide. The absence of an MGP propeptide demonstrates that \(\gamma\)-carboxylation and secretion of vitamin K-dependent proteins need not be linked to the presence of a propeptide or to its proteolytic removal. The propeptides of other vitamin K-dependent proteins are structurally homologous, and there is evidence that this homologous propeptide domain is important to substrate recognition by the \(\gamma\)-carboxylase. Mature MGP has a sequence segment (residues 15–30) that is homologous to the propeptide of other vitamin K-dependent proteins and probably serves the same role in \(\gamma\)-carboxylase recognition. Rat MGP also has a second sequence that has recently been identified in all known vitamin K-dependent vertebrate proteins, the invariant unit Glu–Xaa–Xaa–Xaa–Glu–Xaa–Cys (EXXXEXC). Since the glutamic residues in this unit are sites of \(\gamma\)-carboxylation, it has been suggested that the EXXXEXC unit could allow the \(\gamma\)-carboxylase to discriminate between substrate and product. The demonstration that two structures common to vitamin K-dependent proteins, the homologous propeptide domain and the invariant EXXXEXC unit, are in mature MGP indicates that des-\(\gamma\)-carboxy-MGP should be an excellent in vitro \(\gamma\)-carboxylase substrate for analysis of mechanisms involved in substrate recognition and product dissociation.

Matrix Gla protein (MGP) is a 79-residue protein, found in bone, dentin, and cartilage, that contains 5 residues of the vitamin K-dependent amino acid, \(\gamma\)-carboxyglutamatic acid (Gla) (1–3). The C-terminal domain of bovine MGP (residues 31–79) has 20% sequence identity to the 49-residue structure of bone Gla protein (BGP), the only other vitamin K-dependent protein isolated from bone. Although this degree of sequence identity is at the threshold of statistical significance, the fact that specific amino acid residues conserved in BGP from all species examined to date are also found in MGP indicates that the C-terminal domain of MGP and the secreted form of BGP arose from a common ancestor by gene duplication and subsequent divergent evolution (2). MGP associates with the organic matrix of cartilage and bone and is virtually water-insoluble (<10 \(\mu\)g/ml) in the absence of denaturants, a property that is noteworthy given its high percentage of hydrophilic amino acids and its small size (1–4). Although the function of MGP is unknown, vitamin K-deficient rats develop excessive mineralization of growth-plate cartilage, which indicates that a vitamin K-dependent cartilage protein, possibly MGP, normally prevents cartilage mineralization (5). In clonal osteoblastic cells, MGP production is regulated by 1a,25-dihydroxyvitamin D\(_3\) (1,25-(OH)\(_2\))D\(_3\) (6, 7). In the present study, we have undertaken the molecular cloning of rat MGP in order to generate probes for further investigation of its biosynthesis and hormonal regulation and in order to obtain the complete amino acid sequence of the initial translation product. The results identified the probable structures involved in substrate recognition by the \(\gamma\)-carboxylase and the relationship between the \(\gamma\)-carboxylase binding site and the location of glutamic residues that are modified.

**EXPERIMENTAL PROCEDURES**

**Antibody Screening.** Antibodies raised against bovine MGP (4) were purified by affinity adsorption to rat MGP-Sepharose 4B, elution with 4 M guanidine hydrochloride, and dialysis against phosphate-buffered saline at 4°C. This antibody preparation, which crossreacts with rat MGP (4), was used at a protein concentration of 1.5 \(\mu\)g/ml to screen our rat osteosarcoma (ROS 17/2) \(\lambda gt11\) cDNA library (8) with the Protoblot immunoscreening system (Promega Biotec, Madison, WI). This library was prepared with mRNA from ROS 17/2 cells that had been treated with 1,25-(OH)\(_2\))D\(_3\) at 1 ng/ml for 54 hr (8), a treatment time that we have shown is sufficient to induce MGP expression by ROS 17/2 cells (7). Positive clones were plaque-purified, subcloned into plasmid pUC8, and sequenced by the dideoxynucleotide chain-termination method on double-stranded DNA (9).

**Screening with Labeled Probes.** Additional MGP clones were obtained by screening (8) our ROS 17/2 library with MGP-1 DNA, the largest insert identified by screening with antibody; labeling by nick-translation, nitrocellulose filter preparation, hybridization, and washing have been described (8, 10, 11). In a separate approach to MGP cloning, we also probed our ROS 17/2 \(\lambda gt11\) cDNA library with oligonucleotides corresponding to segments of the N-terminal sequence of rat MGP. MGP was isolated from rat bone (1, 4) and

**Abbreviations:** MGP, matrix Gla protein; BGP, bone Gla protein; 1,25-(OH)\(_2\))D\(_3\), 1a,25-dihydroxyvitamin D\(_3\),

*To whom reprint requests should be addressed.

1This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beraneck, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. 303026).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
subjected to N-terminal sequence analysis for 33 steps as described (2). The N-terminal sequence of rat MGP is Tyr-Gla-Ser-His-Glu-Ser-Met-Glu-Ser-Tyr-Glu-Val-Ser-Pro-Phe-Thr-Asn-Arg-Arg-Asn-Ala-Asn-Thr-Phe-Ile-Ser-Pro-Gln-Gln-Arg-Trp-His-Ala-. The following oligonucleotide probes were synthesized by the University of California at San Diego Oligonucleotide Synthesis Facility and end-labeled (11): 17A, d(GTYGTYGCNACCCTTGC); 17B, d(GTYGTYTCAYACCGTRCG); 14C, d(GTTRCTYTCTTACCT); and 14D, d(GTRCTYGAGNTACTT). Probes 17A and B correspond to residues 28–33 in rat MGP, and probes 14A and B correspond to residues 4–8. Y refers to C and T, R to G and A, and N to all four bases, each at equimolar levels. Clone λMGP-6 was purified based on its hybridization with 17B and 14C. The hybridization of these probes rather than 17A and 14D is in agreement with the final cDNA sequence (Fig. 3).

RESULTS

Isolation of cDNA Clones by Antibody Screening. The initial screening of $5 \times 10^5$ recombinants from our ROS 17/2 cDNA library with rat MGP antibodies yielded 3 positive plaques. After plaque purification, the clone containing the largest insert, designated λMGP-1, was excised with restriction endonuclease EcoRI, subcloned into pUC8, and sequenced. To better establish the sequence in the middle of this insert, MGP-1 was also cleaved with SnaI and the two fragments were subcloned into pUC8 and sequenced. The 495-base-pair insert of λMGP-1 contains one long open reading frame coding for a polypeptide that exactly corresponds to the 33 N-terminal residues of rat bone MGP, which had been determined by amino acid sequence analysis (see Experimental Procedures) and matches the amino acid sequence of MGP purified from bovine bone (2) in 66 out of 79 positions.

The open reading frame extends to the 5' end without revealing any ATG triplet that might act as a translational start site. In the 3' direction, the open reading frame contains a TAA termination codon followed by 197 nucleotides.

Analysis of MGP mRNA. In order to estimate the size of the full-length MGP mRNA, RNAs from rat calvaria (Fig. 1, lane A) and ROS 17/2 cells treated for 6 days with 1,25-(OH)$_2$D$_3$ (lane B) were analyzed by gel electrophoresis followed by blot hybridization with $^{32}$P-labeled MGP-1 DNA. The probe recognized a single broad band at 660–740 nucleotides in both RNA samples, indicating that the MGP cDNA we had cloned from ROS 17/2 was indeed representative of MGP message in normal bone but was not full-length. Analysis of deaminylated ROS 17/2 RNA yielded a single discrete band at 570–580 nucleotides (Fig. 1, lane C). No hybridization was detected to RNA from ROS 17/2 cells that had not been treated with 1,25-(OH)$_2$D$_3$ (data not shown), a finding consistent with the absolute dependence of MGP synthesis in these cells on treatment with 1,25-(OH)$_2$D$_3$ for at least 48 hr (7).

Nucleotide Sequence of Complete MGP cDNA. Since none of the original 3 clones contained a full-length cDNA insert, $^{32}$P-labeled MGP-1 DNA was used to identify additional MGP clones in the library by plaque hybridization. This screening yielded 11 positive clones, none of which proved longer than MGP-1. Most inserts were subcloned into pUC8 and sequenced; one of these, designated λMGP10, is shown in Fig. 2. Independent screening of $3 \times 10^6$ recombinants from our ROS 17/2 cDNA library with two oligonucleotide probes (Fig. 1) based on the N-terminal sequence of rat MGP identified 1 plaque positive for both probes; this plaque also hybridized with $^{32}$P-labeled MGP-1. The insert from this clone, designated λMGP-6, was excised with EcoRI, subcloned into pUC8, and sequenced. The cDNA inserts whose sequences were used to obtain the cDNA structure of MGP are diagrammed in Fig. 2. The sequences of all other clones examined agreed with this structure.

Overlapping sequences were consolidated into the 521-nucleotide sequence shown in Fig. 3. The first ATG triplet (nucleotide positions 1–3) in the longest open reading frame has been designated the translational start site because the sequence surrounding this triplet matches the proposed consensus sequence for initiation of eukaryotic translation at 6 of 7 bases (13). The C-terminal region of the 103-residue-predicted polypeptide (corresponding to nucleotide positions 58–309) has 83% sequence identity with the amino acid sequence of human MGP.

![Fig. 2. Strategy for determining the nucleotide sequence of MGP cDNA. Inserts of the λgt11 cDNA clones that were sequenced are depicted by heavy lines. Horizontal arrows show the direction and extent of sequencing. Vertical arrows indicate SnaI restriction site. bp, Base pairs.](https://example.com/figure2.png)
sequence of bovine MGP (2) and, in the first 33 amino acid residues of this region, is identical to the N-terminal sequence of MGP purified from rat bone. The coding region is terminated by a single TAA triplet (positions 310–312). An AAATAAA polyadenylation signal (14) is found beginning at position 497.

**DISCUSSION**

**Processing of the Predicted Polypeptide.** An unusual feature of the 103-residue primary translation product predicted from the MGP cDNA is the absence of a propeptide. To our knowledge, MGP is the only known example of a vitamin K-dependent protein that lacks a propeptide. The first ATG triplet in MGP is followed by codons for lysine, serine, and a stretch of 16 hydrophobic residues that presumably form the transmembrane core of a signal peptide (15). Signal-peptide cleavage is predicted to occur after the cysteine residue at position −1 (16), and tyrosine at position +1 is the first residue in the N-terminal sequence of mature rat and bovine MGP (Fig. 4).

The primary translation product predicts an 84-residue rat MGP that has an additional 5 amino acids at the C terminus that are not present in the MGP previously isolated from bovine bone and sequenced (Fig. 4). The first two residues of this C-terminal pentapeptide are Arg-Arg, a dibasic sequence that is often a site of proteolytic cleavage in the processing of proteins (17). Although it is not known whether proteolytic cleavage to remove the C-terminal pentapeptide precedes secretion from the cell, a closely spaced doublet has been noted in the protein blot analysis of MGP in guanidine extracts of bone and cartilage from rat and calf (3). If this doublet represents the 84-residue protein predicted from the

![Fig. 3](image-url) Complete nucleotide sequence of MGP cDNA and sequence of its predicted polypeptide. The nucleotide sequence is numbered below each line, with position 1 corresponding to the start of the protein-coding region. The amino acid sequence is numbered above each line, with position 1 corresponding to the N-terminal residue of the protein isolated from bone. The polyadenylation signal is underlined.

![Fig. 4](image-url) Comparison of bovine and rat MGP sequences. The bovine MGP sequence was determined by direct amino acid sequence analysis; the position of each ϒ-carboxyglutamic (Gla) residue in this sequence is therefore known (2). The rat MGP sequence was deduced from the cDNA sequence and the position of the 5 Gla residues in rat MGP is assumed to be the same as in the bovine protein.
cDNA and the 79-residue protein sequenced from bone, the
two proteins must be present at comparable levels in the
extracellular matrix and proteolytic processing probably
occurs after secretion. It will clearly be of great interest to
know the function of proteolytic processing at the C terminus
of MGP and, in particular, whether the extraordinary water-
insolubility of the 79-residue MGP isolated from bone, an
unusual feature of the protein given its high percentage of
hydrophobic amino acids and small size (1-4), is also a
property of the 84-residue protein.

Mechanisms by Which Substrates Are Recognized by the
\( \gamma \)-Carboxylase. Previous work showed that the propeptide
of BGP is homologous to the propeptide of other vitamin
K-dependent proteins (8), and it was postulated that the
homologous propeptide domain is a recognition site for the
\( \gamma \)-carboxylase (8, 18). Although MGP has no propeptide,
there is a region of the mature protein, residues +13 to +30
in Fig. 4, that is homologous to the propeptide of other
vitamin K-dependent proteins. The absence of a propeptide
in MGP demonstrates that the \( \gamma \)-carboxylation and secretion
of a vitamin K-dependent protein need not be linked to
the presence of a propeptide or to its proteolytic removal.

If residues 15-30 in MGP correspond to the propeptide
component of the binding site for the \( \gamma \)-carboxylase, MGP is
the first example of a vitamin K-dependent protein in which
a glucogenic residue on the N-terminal side of this \( \gamma \)-carbox-
ylase binding site is known to be \( \gamma \)-carboxylated. Previous
N-terminal sequence analysis of bovine MGP revealed ap-
proximately 80% \( \gamma \)-carboxylglutamic acid and 20% glutamic
acid at position 2 (2). The other three glutamic residues on the
N-terminal side of this \( \gamma \)-carboxylase binding site were not
\( \gamma \)-carboxylated. In the course of the present studies, we
carried out N-terminal sequence analysis of rat MGP. The
results support the presence of partial \( \gamma \)-carboxylation at
position 2 and no \( \gamma \)-carboxylation at positions 5, 8, and 11.
Based on extrapolated repetitive yields in the N-terminal
sequence of rat MGP, position 2 is 20% glutamic acid and 80%
\( \gamma \)-carboxylglutamic acid. The consistent finding of a \( \gamma \)-
carboxylglutamic residue on the N-terminal side of the \( \gamma 
\)-
carboxylase binding site demonstrates that glucogenic residues
on either side of the binding site have access to the active site
of the \( \gamma \)-carboxylase. It follows from these observations that
the \( \gamma \)-carboxylase binding site in as-yet-uncharacterized
vitamin K-dependent proteins can be anywhere in the mature
protein and that \( \gamma \)-carboxylation of glucogenic residues can
occur to either side of the propeptide component of the
binding site.

A previously unrecognized invariant structure in the \( \gamma 
\)-
carboxylglutamic acid-containing region of vitamin K-dependent
vertebrate proteins is the sequence Glu-Xaa-Xaa-Glu-Xaa-Cys
(EXXXEXC, positions 17-23, Fig. 5). The separation between the
first glucogenic residue in this invariant structure (at +17) and
the previously recognized invariant arginine in the homologous
propeptide domain (at -1) is 15-17 residues in all cases except
mouse BGP, where the separation is 12 residues. If the sequences
of the homologous propeptide domain and the \( \gamma \)-carboxylglutamic acid-containing
region of known vitamin K-dependent vertebrate proteins
are compared (-16 to +30, Fig. 5), the two vitamin K-
dependent bone proteins, BGP and MGP, are each as closely
related to coagulation proteins as to each other. There are 2
additional residues that are invariant in MGP and in all
coaagulation proteins but not in BGP (alanine at -10 and
glutamic acid at +30), and 1 additional residue that is
invariant in BGP and in all coagulation proteins but not in
MGP (leucine at +6). There are, for comparison, only 2
residues that are invariant in BGP and MGP but are not found
in coagulation proteins (arginine at +20 and cysteine at +29).
These sequence relationships are consistent with the hypo-
thesis that the propeptide and \( \gamma \)-carboxylglutamic acid domain
units of known vitamin K-dependent proteins evolved from
a common ancestor by gene duplication and subsequent
divergent evolution and indicate that the two bone proteins

<table>
<thead>
<tr>
<th>-10</th>
<th>-1</th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>hBGP</td>
<td>KAFVSKQEGSEEVKRR</td>
<td>ILYQLWAGAPVYPDLEPRRVECDLENPDCD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rBGP</td>
<td>KAFMSKQEGSKVNVRLRR</td>
<td>ILYLNNGLGAPAPAPLPPHEVCECDLENPDCD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mBGP</td>
<td>KAFMSKQEGSKVNVRLRR</td>
<td>ILYLGSVSPDPDLPPTECDLENPACD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bBGP</td>
<td>NPIFRGNANSFISPQORWRKAKQERIRLENKPOYNRLFACCDDFKLC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mBGP</td>
<td>PFTNRRNATFISPQORWRKAKQERIRLENKPOYNRLFACCDDFKLC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hPT</td>
<td>HVFLAPQQRSSLLOVRRIJAFTLEÉÉVRGKLÉERCVÉETCSYÉÉAFÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bPT</td>
<td>HVFLAHOQASSSLLOVARRIAKNGLÉÉÉVRGKLÉERCLEÉPCSÉÉÉAFÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hFX</td>
<td>SLIFRéQANNVIPARTRIASLFEÉÉMKGLÉERCEÉÉÉCSYÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bFX</td>
<td>SVFLPRDOMHLVRARRAIJNSLFEÉÉVKGGLÉERCEÉACSLÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hFX</td>
<td>TVFLDNKANILRPKRINSGKLÉÉÉVFGLÉERCMÉÉÉKCSFÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hFVII</td>
<td>RVFQTFDTEAHVGLHRRIAJANFLÉÉÉLRPGSLÉÉÉKCÉÉÉQFÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hPS</td>
<td>ANLSSLKQQASQVLVRARRIRAILNLEÉTKGKLÉERCIÉÉÉLCNKÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bPS</td>
<td>ANFLSRQASHVLRLRRIAJAILNLEÉETKGLÉERCIÉÉÉLCNKÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hPC</td>
<td>SVFSSERAHQVLRIRKRAINFSLÉÉÉLRHSSLÉÉÉCÉÉICDFÉÉAKÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bPC</td>
<td>SVFSSQRAHQVLRIRKRAINFSLÉÉÉLRPQVNÉÉÉCÉÉÍVEFÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bPZ</td>
<td>JAGSYLLÉÉÉLFÉGHKLCÉÉÉCVYÉÉARÉ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Conserved amino acids in the propeptide and \( \gamma \)-carboxylglutamic acid-containing domains of known vertebrate vitamin K-dependent proteins. Amino acid sequence positions are numbered so that 1 corresponds to the first residue of mature BGP. All other sequences have been aligned to give maximum homology. Arrows preceding residue 1 indicate the site of propeptide cleavage; the absence of an arrow in the MGP structure is due to an absence of a propeptide in this protein. Sequences are given in standard one-letter symbols [E represents \( \gamma \)-carboxyglutamic acid (Gla)]. Sequences (from top to bottom): hBGP, human BGP (19); rBGP, rat BGP (8); mBGP, mouse BGP (19); bBGP, bovine MGP (2); rMGP, rat MGP (Fig. 3); hPT, human prothrombin (20); bPT, bovine prothrombin (21); hFX, human factor X (22); bFX, bovine factor X (23); hFVII, human factor IX (24); hFVII, human factor VII (25); hPS, human protein S (26); bPS, bovine protein S (27); hPC, human protein C (28); bPC, bovine protein C (29); bPZ, bovine protein Z (30).
probably diverged from this common ancestor at about the same time as coagulation proteins.

There is evidence that the homologous propeptide and invariant EXXXXXC structures of vitamin K-dependent proteins both play a role in substrate recognition by the γ-carboxylase. Site-directed mutagenesis studies (31) showed that two conserved amino acids in the propeptide region of factor IX (phenylalanine at -16 and alanine at -10; Fig. 5) are indeed critical to recognition by the γ-carboxylase. Although the importance of the EXXXXXC unit has not yet been tested by site-directed mutagenesis, there is indirect evidence to support a role for this unit in γ-carboxylase binding to substrate. A variety of peptides have been synthesized that correspond to the sequences of vitamin K-dependent proteins in γ-carboxyglutamic acid-containing regions but that lack the invariant EXXXXXC unit (32). None of these peptide substrates proved to have $K_a$ values lower than a few millimolar (32-34). In contrast, both intact BGP and the peptide corresponding to residues 13-29 in prothrombin proved to be excellent γ-carboxylase substrates (33, 34) after γ-carboxyglutamic residues were converted to glutamic residues by decarboxylation (35). For both decarboxylated polypeptides, $K_a$ values are 2-3 orders of magnitude lower than for the best known synthetic peptide substrate, Phe-Leu-Glu-Glu-Leu (33, 34). Since both decarboxylated BGP and decarboxylated prothrombin (13-29)-peptide lack the homologous propeptide domain yet retain the EXXXXXC invariant structure, the excellent substrate activities of these decarboxylated polypeptides is strong evidence for a role of the EXXXXXC structure in γ-carboxylase binding to substrate.

We thank Matthew K. Williamson for carrying out the N-terminal sequence analysis of rat MGP. This work was supported by U.S. Public Health Service Grant AM25921.