Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6109–6113, September 1985
Biochemistry

The propeptide of rat bone γ-carboxyglutamic acid protein shares homology with other vitamin K-dependent protein precursors

(γ-carboxylation/rat osteosarcoma cells/phage λgt11 expression library/cDNA sequence/bone matrix protein)

LYDIA C. PAN AND PAUL A. PRICE

Department of Biology, University of California at San Diego, La Jolla, CA 92093

Communicated by Andrew A. Benson, May 28, 1985

ABSTRACT **The molecular cloning of bone γ-carboxyglutamic acid (Gla) protein (BGP; osteocalcin) was accomplished by constructing a phage λgt11 cDNA library from the rat osteosarcoma cell line ROS 17/2 and screening this library with antibodies raised against BGP from rat bone. By sequencing several cloned cDNAs, we have established a 489-base-pair sequence that predicts a mature BGP of 50 amino acid residues with an NH_2-terminal extension of 49 residues. The leader peptide consists of a hydrophobic signal peptide followed by a basic propeptide of 26 or 27 residues that is cleaved after an Arg-Arg dipeptide prior to secretion from the cell. Mature rat BGP is extremely homologous to BGPs from other species except for its COOH-terminal sequence. A stretch of 9 residues proximal to the NH_2 terminus of secreted BGP is strikingly similar to the corresponding regions in known propeptides of the γ-carboxyglutamic acid-containing blood coagulation factors. We suggest that this common structural feature may be involved in the posttranslational targeting of these polypeptides for vitamin K-dependent γ-carboxylation.**

Bone γ-carboxyglutamic acid (Gla) protein (BGP) is a small, highly conserved protein associated with the mineralized matrix of bone (1). Its interaction with synthetic hydroxyapatite *in vitro* is absolutely dependent on three residues of γ-carboxyglutamic acid (2), the vitamin K-dependent amino acid formed posttranslationally from glutamic acid. Mineralization defects develop in vitamin K-deficient rats that produce a non-γ-carboxylated BGP incapable of binding to bone (3, 4). BGP production is developmentally regulated (5) and is stimulated by the calcitropic steroid hormone 1,25-dihydroxyvitamin D_3 (6, 7). We undertook the molecular cloning of BGP in order to generate probes for further investigation of its synthesis and regulation. The present report describes the construction of an osteosarcoma cDNA expression library and determination of the complete nucleotide sequence of rat BGP cDNA. Of particular interest is the information obtained about the structure of the precursor polypeptide that may provide a clue to the mechanism by which the vitamin K-dependent γ-carboxylating system recognizes specific proteins for modification.

MATERIALS AND METHODS

RNA Isolation. RNA was isolated from both osteosarcoma cells and rat calvariae by the guanidinium thiocyanate/CsCl method (8). The rat osteosarcoma cell line ROS 17/2 (9) was cultured as described (10). Confluent cultures at a density of $1–2 \times 10^5$ cells per cm^2 were treated with daily changes of serum-free medium containing 1 ng of 1,25-dihydroxyvitamin D_3 (Hoffmann–La Roche) per ml for a total of 54 hr. Approximately 6 hr after the last media change, cells were lysed directly on the plate with the guanidinium thiocyanate solution. Calvariae of weanling Sprague–Dawley rats (Simonsen Laboratories, Gilroy, CA) were dissected free of adhering membranes, frozen in liquid nitrogen, and pulverized with a chilled mortar and pestle. The frozen powder was transferred immediately to the lysis solution. Typical yields of total RNA were 2 mg per 10^8 ROS 17/2 cells and 0.6 mg per 40 calvariae. Polyadenylylated RNA was selected by fractionation over oligo(dT)-cellulose (11).

Construction of an Osteosarcoma cDNA Library. The λgt11 system of Young and Davis (12, 13) was used to clone ROS 17/2 cDNA. Double-stranded cDNA was synthesized and prepared for insertion into the *Eco*RI site of λgt11 by the procedure of de Wet *et al.* (14); 2 μg of cDNA was ligated to 15 μg of dephosphorylated vector DNA, then one-eighth of this mixture was packaged by the instructions provided with the phage λ packaging extract (Bethesda Research Laboratories). The library was amplified as plate lysates on *Escherichia coli* strain Y1088.

Antibody Screening. Antibodies raised in rabbits against rat BGP (15) were purified by affinity adsorption to rat BGP-Sepharose 4B, elution with 4 M guanidium hydrochloride, and dialysis into 5 mM ammonium bicarbonate. Removal of remaining coliform antibodies was accomplished by incubating this fraction with a bacterial lysate affinity material (14) and recovering the unbound antibody. This antibody preparation was used at 3 μg/ml to screen the osteosarcoma library with a chromogenic immunodetection protocol (14).

Preparation and Hybridization of cDNA Probes. Large amounts of cDNA were conveniently obtained by subcloning the desired restriction fragments into pUC plasmids (16). Restriction fragments containing cDNA inserts were labeled with [α-^{32}P]dATP (400 Ci/mmol; 1 Ci = 37 GBq) by nick-translation (18) to a specific activity of at least 2×10^7 cpm/μg. Hybridization to DNA or RNA immobilized on nitrocellulose (BA85, Schleicher & Schuell) was performed with $1–2 \times 10^5$ cpm of DNA probe per ml of 50% formamide containing 5× NaCl/P_i/EDTA (1× = 0.18 M NaCl/10 mM Na phosphate, pH 7.7/1 mM EDTA), and heparin at 0.1 mg/ml (19) after prehybridization for 1 hr in the same buffer minus the probe. Filters were washed in 0.1× NaCl/P_i/EDTA containing 0.1% $NaDodSO_4$ at 37°C. Kodak XAR-5 film was used for all autoradiography.

RNA Gel Blots. RNA was electrophoresed on 1.3% agarose gels containing 2.2 M formaldehyde in a morpholinopropanesulfonic acid buffer (20). RNA was blotted to nitrocellulose (21) and hybridized as usual. Autoradiography was performed at −70°C with DuPont Cronex Lightning Plus screens. After hybridization analysis, RNA size markers were stained on the filter with methylene blue (21).

Deadenylylation of mRNA. To remove poly(A) tracts from mRNA, 1.5 μg of polyadenylylated RNA was annealed with 1.5 μg of oligo(dT) at room temperature, then digested with

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.

Abbreviations: Gla (in sequences), γ-carboxyglutamic acid; BGP, bone Gla protein, osteocalcin.

1 unit of *E. coli* ribonuclease H (Boehringer Mannheim) at 30°C for 1 hr (22).

Plaque Hybridization. Phage grown on *E. coli* strain Y1088 were screened with ^{32}P-labeled cDNA at densities of no more than 2×10^4 plaques per 85-mm plate. A nitrocellulose filter was laid over the plate for 1–2 min, then prepared for hybridization by the method of Benton and Davis (23). Positive signals were generally detectable after 18-hr autoradiography without intensifying screens.

DNA Sequence Analysis. DNA restriction fragments were cloned into the M13 vectors for sequencing by the dideoxynucleotide chain-termination method (26). Analysis of nucleotide sequences was performed with the aid of Staden programs (28, 29) modified for use on a VAX 11/780 computer (Digital Equipment, Maynard, MA) by D. W. Smith in our department.

Primer Extension. An oligonucleotide primer having the sequence d(GCACCTGCCAGGTCAGAGAGGC) complementary to the 5′ end of the coding strand of BGP7 was synthesized by the Peptide Oligonucleotide Synthesis Facility (Chemistry Department, University of California, San Diego). The primer was labeled with [γ-^{32}P]ATP (5000 Ci/mmol) by T4 polynucleotide kinase (Promega Biotec, Madison, WI) to a specific activity of 2.5×10^8 cpm/μg (30). Ten nanograms of end-labeled primer was annealed to 10–20 μg of polyadenylylated ROS 17/2 RNA by heating the mixture to 95°C and then cooling it slowly to 55°C. After addition of unlabeled nucleotides and 2 units of avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim), the reaction was incubated at 37°C for 10 min followed by 42°C for 1 hr. Primer extension products were analyzed on DNA sequencing gels. To determine the sequence of the primer extension product, one of the following dideoxynucleotides was included in each of four reactions: 0.5 mM ddATP, 0.5 mM ddCTP, 0.8 mM ddGTP, or 2.0 mM ddTTP.

RESULTS

Isolation of cDNA Clones by Antibody Screening. The number of recombinants in our ROS 17/2 cDNA library was estimated to be *ca.* 4×10^6 by the titer of infectious phage produced by *in vitro* packaging. After amplification, >90% of the phage formed colorless plaques when grown in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, a chromogenic substrate for the β-galactosidase produced by λgt11.

The initial screening of 2×10^5 recombinants with rat BGP antibodies yielded six positive plaques. These positive signals could be completely abolished by preadsorbing the first antibody with excess rat BGP. After plaque purification, the insert from one clone designated λBGP4 was excised with *Eco*RI, subcloned into M13 mp9 and sequenced. The 314-base-pair insert of λBGP4 contains one long open-reading frame coding for a polypeptide that matches the partial sequence of BGP from rat cortical bone (31) at 28 out of 30 positions. This open reading frame extends to the 5′ end of the insert without revealing any ATG triplet that might act as a translational start site. In the 3′ direction, this reading frame contains a TAG termination codon followed by 135 nucleotides terminating in a run of 8 adenosine residues. We concluded that BGP4 is a partial cDNA containing the 3′ end but not the 5′ end of the mRNA sequence.

Analysis of BGP mRNA. In order to estimate the size of full-length BGP cDNA, RNAs from rat calvariae (Fig. 1, lane 1) and ROS 17/2 cells (Fig. 1, lane 2) were analyzed by gel blots hybridized with ^{32}P-labeled BGP4 DNA. The probe recognized a similar broad band or bands of 580–660 nucleotides in both RNA samples, indicating that the BGP cDNA we had cloned from ROS 17/2 was indeed representative of BGP message in normal bone. Similar analysis of deadenylylated ROS 17/2 RNA yielded a single discrete band of 540 nucleotides (Fig. 1, lane 3). No hybridization was detected to RNA from ROS 25/1, an osteosarcoma cell line that does not synthesize BGP (10) (Fig. 1, lane 4).

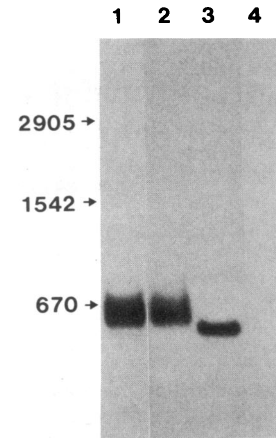

FIG. 1. Gel blot analysis of BGP mRNA. All mRNA samples were purified by a single round of oligo(dT)-cellulose chromatography. The amount of RNA loaded and duration of autoradiographic exposure were varied to achieve bands of comparable intensity. Lanes: 1, 15 μg of rat calvarial RNA (48-hr exposure); 2, 1.5 μg of ROS 17/2 RNA (100-hr exposure); 3, 1.5 μg of ROS 17/2 RNA deadenylylated (100-hr exposure); 4, 10 μg of ROS 25/1 RNA (100-hr exposure). Marker RNA sizes for 23S (32) and 16S (33) *E. coli* ribosomal RNA and rabbit globin mRNA (22) were based on sequence data.

Nucleotide Sequence of Complete BGP cDNA. Since none of the original six clones appeared to contain a full-length cDNA insert, ^{32}P-labeled BGP4 DNA was used to identify additional BGP clones in the library by plaque hybridization. This screening yielded a large number of positive signals. Several of the larger inserts were compared to BGP4 by restriction mapping in order to determine whether they would extend the known sequence in the 5′ direction. The cDNA inserts that were eventually sequenced are diagrammed in Fig. 2. The extreme 5′ end of the sequence was obtained by primer extension of a synthetic oligonucleotide on ROS 17/2 RNA.

Overlapping sequences were consolidated into the 489-

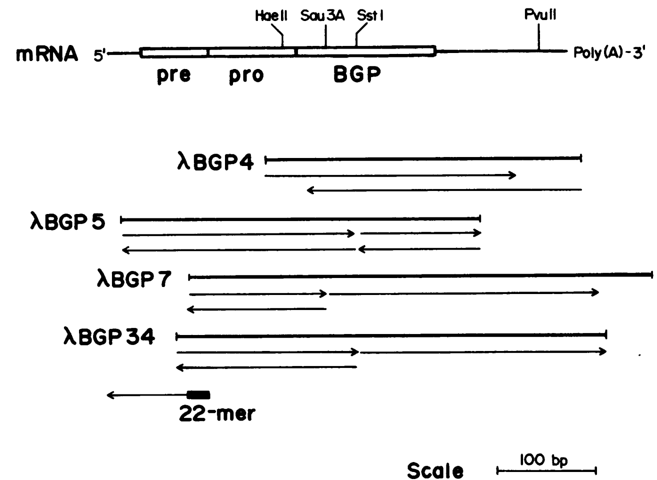

FIG. 2. Strategy for determining the nucleotide sequence of BGP cDNA. Key restriction sites used in mapping and sequencing studies are shown. Inserts of λgt11 cDNA clones that were sequenced are depicted by heavy lines. Arrows denote the direction and extent of sequencing. The solid rectangle labeled "22-mer" represents the synthetic oligonucleotide used for primer extension.

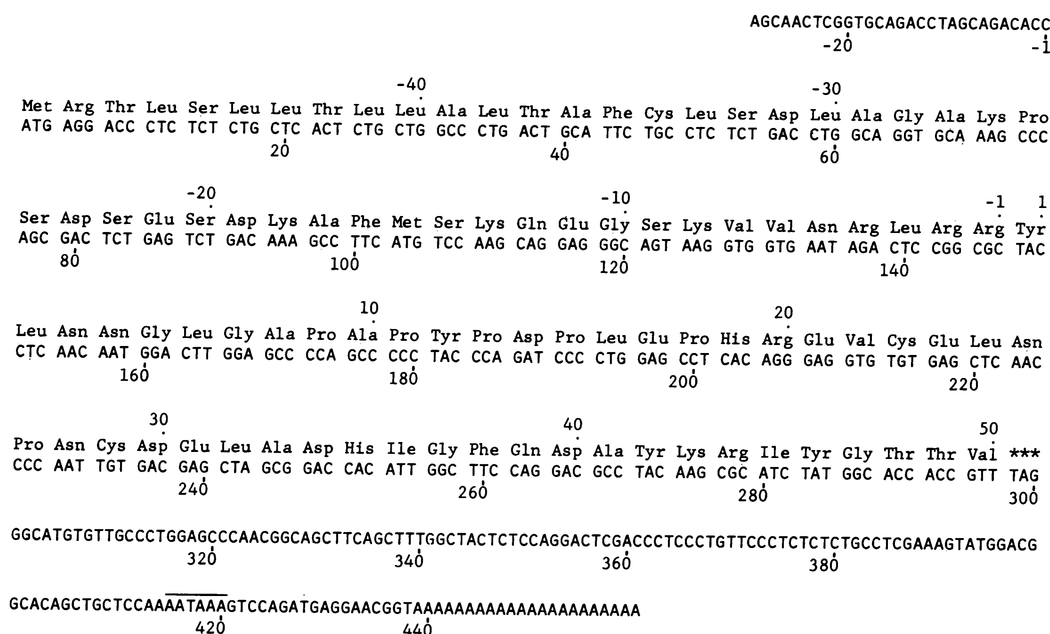

FIG. 3. Complete nucleotide sequence of BGP cDNA and sequence of its predicted polypeptide. Numerical positions in the nucleotide and amino acid sequences are notated below and above each line, respectively. Numbering of the nucleotides in the cDNA sequence begins at the start of the protein coding region. The protein sequence is numbered with respect to the first residue of the mature protein. The termination codon following the protein coding region is indicated by three asterisks. A bar is drawn above the polyadenylylation signal.

nucleotide sequence shown in Fig. 3. If translation proceeds from the first ATG triplet (nucleotide positions 1–3) in the longest open-reading frame, then the COOH-terminal half of the 99-residue predicted polypeptide bears a strong resemblance to the highly conserved primary structure of other mammalian BGPs (34). The coding region is terminated by a single TAG triplet (positions 298–300) commencing the 3′ untranslated region. An AATAAA polyadenylylation signal (35) is found beginning at position 415. In two of three clones sequenced through this region, the poly(A) tail begins 18 bp downstream from this signal at position 439 as shown in Fig. 3. One cDNA insert (BGP4) differs slightly in having a terminal stretch of 8 adenosine residues beginning at position 433.

DISCUSSION

Frequency of BGP cDNA. We anticipated a low abundance of BGP mRNA and designed our cloning strategy accordingly. ROS 17/2 was the highest BGP producer of numerous cell lines tested; however, its BGP output represents <0.01% of total protein synthesis. Cells were treated with 1,25-dihydroxyvitamin D_3 to specifically stimulate BGP synthesis (6) prior to isolation of RNA. The λgt11 system was chosen because of the high efficiency with which recombinant phage can be generated, also because this is an expression vector that can be screened at high density with antibody probes.

The ROS 17/2 cDNA library we have generated contains 4×10^6 phage of which >90% harbor inserts based on inactivation of the *lacZ* gene that traverses the cloning site. If the cloning procedure did not introduce a bias against specific messages, then even very rare mRNAs should be represented in our library.

The frequencies of BGP cDNA detected by antibody screening and plaque hybridization were 0.003% and 0.01% respectively. Antibody screening detects only the fraction of cDNA inserts that are ligated in the correct reading frame and orientation for expression, although an occasional frame-shifted insert will express at detectable levels (12). The plaque hybridization frequency of 0.01% is apt to be a more accurate estimate of the representation of BGP cDNA in the library and concurs with predictions of message frequency based on rates of BGP synthesis.

Comparison of cDNA Clones to mRNA. The size of BGP mRNA given by denaturing gel electrophoresis is 580–660 nucleotides. This heterogeneity can be entirely accounted for by variability in the length of poly(A) tails. Thus, BGP mRNA appears to be a unique species of 540 nucleotides excluding poly(A). Our current sequence of 466 nucleotides excluding poly(A) contains the entire protein coding region but appears to be missing 70 or so nucleotides from the 5′ untranslated region. We are pursuing the isolation of genomic clones in order to characterize the 5′ end of the message by S1 nuclease mapping and additional sequencing studies.

A possible heterogeneity was observed among cloned sequences corresponding to the 3′ end of BGP mRNA. Although three inserts all exhibit single AATAAA polyadenylylation signals at identical sequence locations, only two have poly(A) tracts beginning at position 439 as shown in Fig. 3. The third insert has a short terminal run of adenosine residues beginning at position 433. It is not clear whether this third clone contains an alternative site for poly(A) addition or a cloning artifact caused by reverse transcription of a partially mismatched oligo(dT) primer.

Processing of the Predicted Polypeptide. The cDNA sequence predicts a primary translation product of 99 amino acid residues that corresponds to a 50-residue secreted BGP with a 49-residue NH_2-terminal extension. The first ATG triplet has been designated the translational start site based on resemblance of its surrounding sequence ACACCATG to the proposed consensus sequence for eukaryotic translational initiation CCACCATG (37). The second amino acid residue, arginine, is followed by a stretch of 16 hydrophobic residues that presumably form the transmembrane core of a signal peptide (38). Signal cleavage is predicted to occur after glycine at position −28 or alanine at position −27 (39). This first processing step is probably cotranslational because no polypeptide corresponding in size to the 99-residue species has ever been detected. The remaining sequence of 76 or 77 residues has a calculated M_r of 9120–9240, which agrees well with a physical size estimate for the larger intracellular precursor to BGP (10). A second proteolytic processing step

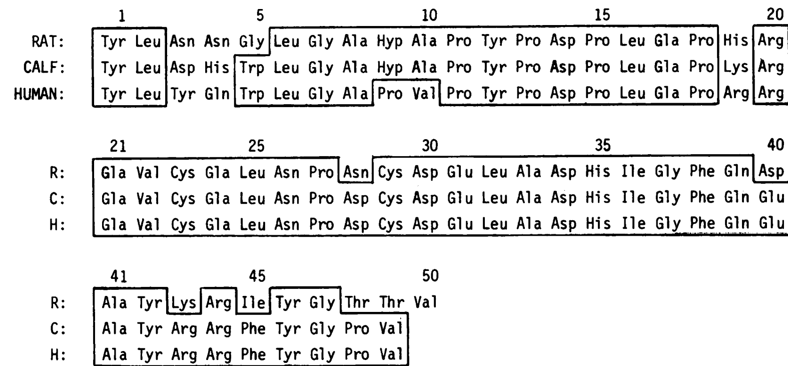

FIG. 4. Comparison of mammalian BGP sequences. Secreted proteins only are compared. Calf and human sequences were taken from ref. 34. Identical residues are boxed.

is required to generate the 50-residue mature protein by cleavage following a pair of basic residues. It appears that this second cleavage occurs prior to the secretion of BGP based on the localization of the larger form exclusively to the intracellular compartment, whereas the mature form was found both inside and outside cells (10). Since this intracellular two-step processing is analogous to the biosynthesis of other secreted proteins such as parathyroid hormone (40) and serum albumin (41), we propose the corresponding designations of prepro-BGP for the 99-residue primary translation product and pro-BGP for the 76- or 77-residue precursor.

Comparison of Rat BGP to Other BGPs. Nucleotides 148–297 in Fig. 3 encode a mature BGP of 50 amino acid residues. Its primary structure as deduced from the cDNA sequence agrees with the NH_2-terminal 30-residue sequence of rat cortical bone BGP (31) except at positions 4 and 28. Both of these discrepancies involve amide-to-acid substitutions, which probably reflect deamidation of the protein during preparation for sequencing. Therefore, the assignment of asparagine to these positions is more likely to be correct. The earlier study of the rat bone protein further revealed that the proline at position 9 is hydroxylated and confirmed the presence of γ-carboxyglutamic acid at positions 17, 21, and 24.

The sequences of rat, calf, and human BGPs are compared in Fig. 4. Rat BGP exhibits 80% and 76% sequence identity to the calf and human proteins, respectively. This homology is strongest in the γ-carboxyglutamic acid-containing central domain. The rat and calf proteins differ at only three positions between residues 6 and 42 inclusively, and these are conservative replacements of histidine for lysine at position 19, asparagine for aspartic acid at position 28, and aspartic acid for glutamic acid at position 40. The evolutionary conservation of primary structure within this domain has been correlated with the ability of all vertebrate BGPs to bind to bone mineral.

Rat BGP differs most noticeably from other BGPs at its COOH terminus. The COOH-terminal hexapeptide Ile-Tyr-Gly-Thr-Thr-Val of the rat protein contrasts with the pentapeptide Phe-Tyr-Gly-Pro-Val of chicken, calf, monkey, and human BGPs (34, 43). This difference can be invoked to explain the unique antigenic properties of rat BGP. Antibodies raised against the rat protein fail to recognize calf or human BGP, whereas antibodies raised against calf BGP do cross-react strongly with the human protein (44). The COOH-terminal tryptic fragment of calf BGP is known to be vital to its antigenic identity (44).

Comparison of Vitamin K-Dependent Protein Precursors. Fig. 5 shows the predicted leader sequences for BGP and four of the vitamin K-dependent blood coagulation factors aligned with respect to the NH_2 terminus of their secreted proteins. All of the leader sequences begin with typical hydrophobic signal peptides, but these show little conservation at the primary sequence level even among the four hepatic proteins. Only three residues are identical in all five sequences: phenylalanine at position −16, arginine at position −4, and

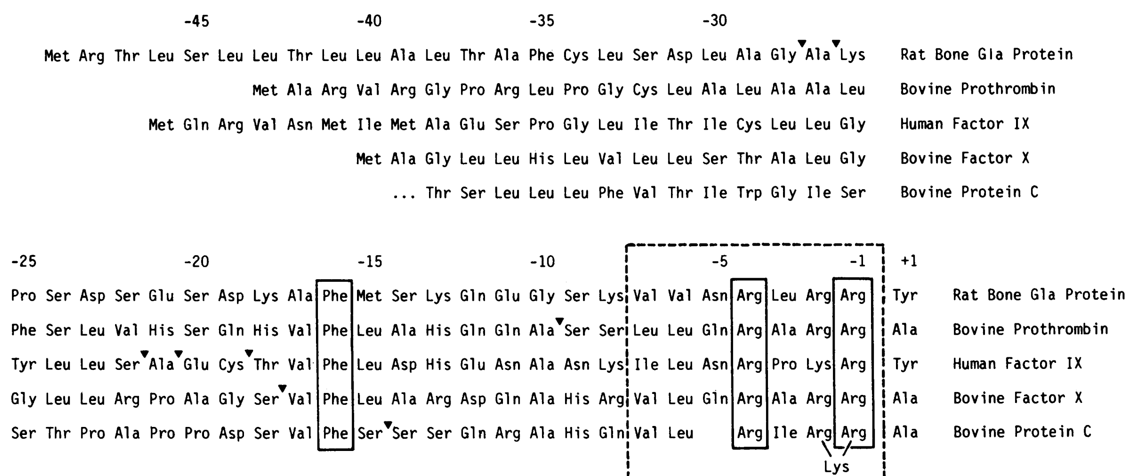

FIG. 5. Comparison of leader sequences of several precursors to vitamin K-dependent proteins. Sequences are numbered backwards from the cleavage site that releases the secreted proteins and aligned to give maximum homology. The region of maximal homology is enclosed within the dashed line. Identical residues are boxed. Inverted triangles (▼) denote probable sites for signal peptide cleavage. Sequences of the coagulation factors were taken from refs. 45–48.

arginine at the propeptide cleavage site. The only salient homology is localized to a portion of the propeptide proximal to the secreted protein. Each sequence shown is characterized by a dibasic or tribasic propeptide cleavage site having arginine at the −1 position. The common arginine at position −4 has already been mentioned. The amino acid occupying position −5 is an amide (asparagine or glutamine) except in protein C. Nonpolar residues are found in positions −3, −6, and −7, while polar residues predominate in positions −8 and −9. The basic nature of this segment is reinforced by the complete absence of acidic residues in positions −1 through −10. No aromatic amino acids are found until the common phenylalanine at position −16.

The similarity between the propeptides is striking, given the lack of homology between the secreted forms of BGP and these coagulation factors, including their γ-carboxylated domains (17). To better assess its significance, we searched the updated "Newat" library of *ca.* 2000 protein sequences (24) for the pentapeptide consensus sequence Asn/Gln-Arg-X-Arg/Lys-Arg, where X can be Gly, Ala, Val, Leu, Ile, Cys, Met, or Pro. No matches in any leader sequences outside of the known vitamin K-dependent protein precursors were found.

The discovery of even limited homology between the propeptides of γ-carboxyglutamic acid-containing proteins as unrelated as BGP and the blood coagulation factors is potentially significant to the question of what targets specific proteins for γ-carboxylation. Several lines of evidence suggest that a common structural feature may be recognized by the γ-carboxylating system and that this putative targeting signal may reside in the propeptide. In studies of substrate specificity utilizing a soluble γ-carboxylase assay, peptides derived from γ-carboxyglutamic acid containing sequences were not utilized preferentially compared to peptides derived from a protein that does not normally contain this residue (25). Some interaction of the enzyme and substrate outside the catalytic site has been suggested by the observation that endogenous substrate can be converted to product and remain associated with the enzyme, while it continues to catalyze the conversion of added peptide substrates (27). Localization of the enzymatic activity to the rough endoplasmic reticulum (36) implies that γ-carboxylation is an early step in the biosynthesis of these proteins. When γ-carboxylation is blocked with warfarin in ROS 17/2 cells, a basic form of the precursor accumulates (42). We speculate that the enzyme complex could have a binding site for the targeting structure apart from its catalytic site. This association would then enable the γ-carboxylase to modify accessible glutamic acid residues in an adjoining segment of the polypeptide. This type of mechanism is supported by the fact that γ-carboxyglutamic acid residues in all known vitamin K-dependent proteins are clustered, for the most part in the NH_2-terminal portion of the secreted polypeptides. The current structural data intimate additional functional tests to identify the putative targeting signal and elucidate its role in γ-carboxylation.

We thank J. R. de Wet for bacterial strains, λgt11, and advice on cDNA cloning and J. D. Fraser for providing rat calvarial RNA. We are grateful to R. F. Doolittle for the computerized searches of his protein sequence library and helpful discussions on the subject of protein sequence homology. This work was supported by U.S. Public Health Service Grant AM25921.

1. Price, P. A. (1983) in *Bone and Mineral Research*, ed. Peck, W. A. (Excerpta Medica, Amsterdam), Annual 1, pp. 157–190.
2. Price, P. A., Williamson, M. K. & Lothringer, J. W. (1981) *J. Biol. Chem.* **256**, 12760–12766.
3. Price, P. A., Williamson, M. K., Haba, T., Dell, R. B. & Jee, W. S. S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7734–7738.
4. Price, P. A. & Sloper, S. A. (1983) *J. Biol. Chem.* **258**, 6004–6007.
5. Price, P. A., Lothringer, J. W., Baukol, S. A. & Reddi, A. H. (1981) *J. Biol. Chem.* **256**, 3781–3784.
6. Price, P. A. & Baukol, S. A. (1980) *J. Biol. Chem.* **255**, 11660–11663.
7. Price, P. A. & Baukol, S. A. (1981) *Biochem. Biophys. Res. Commun.* **99**, 928–935.
8. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
9. Majeska, R. J., Rodan, S. B. & Rodan, G. A. (1980) *Endocrinology* **107**, 1494–1503.
10. Nishimoto, S. K. & Price, P. A. (1980) *J. Biol. Chem.* **255**, 6579–6583.
11. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
12. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
13. Young, R. A. & Davis, R. W. (1983) *Science* **222**, 778–782.
14. de Wet, J. R., Fukushima, H., Dewji, N. N., Wilcox, E., O'Brien, J. S. & Helinski, D. R. (1984) *DNA* **3**, 437–447.
15. Price, P. A., Lothringer, J. W. & Nishimoto, S. K. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2938–2942.
16. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
17. Jackson, C. M. & Nemerson, Y. (1980) *Annu. Rev. Biochem.* **49**, 765–811.
18. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
19. Singh, L. & Jones, K. W. (1984) *Nucleic Acids Res.* **12**, 5627–5638.
20. Seed, B. (1982) in *Genetic Engineering: Principles and Methods*, eds. Setlow, J. K. & Hollaender, A. (Plenum, New York), Vol. 4, pp. 91–102.
21. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 203–206.
22. Vournakis, J. N., Efstratiadis, A. & Kafatos, F. C. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2959–2963.
23. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
24. Doolittle, R. F. (1981) *Science* **214**, 149–159.
25. Suttie, J. W., Lehrman, S. R., Geweke, L. O., Hageman, J. M. & Rich, D. H. (1979) *Biochem. Biophys. Res. Commun.* **86**, 500–507.
26. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
27. de Metz, M., Vermeer, C., Soute, B. A. M., van Scharrenburg, G. J. M., Slotboom, A. J. & Hemker, H. C. (1981) *FEBS Lett.* **123**, 215–218.
28. Staden, R. (1977) *Nucleic Acids Res.* **4**, 4037–4051.
29. Staden, R. (1978) *Nucleic Acids Res.* **5**, 1013–1015.
30. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
31. Otawara, Y., Hosoya, N., Moriuchi, S., Kasai, H. & Okuyama, T. (1981) *Biomed. Res.* **2**, 442–446.
32. Brosius, J., Dull, T. J. & Noller, H. F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 201–204.
33. Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4801–4805.
34. Poser, J. W., Esch, F. S., Ling, N. C. & Price, P. A. (1980) *J. Biol. Chem.* **255**, 8685–8691.
35. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
36. Carlisle, T. L. & Suttie, J. W. (1980) *Biochemistry* **19**, 1161–1167.
37. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 5145–5164.
38. Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H. & Lingappa, R. (1979) in *Secretory Mechanisms, Society for Experimental Biology Symposium*, eds. Hopkins, C. R. & Duncan, C. J. (Cambridge University Press, London), Vol. 33, pp. 9–36.
39. von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243–251.
40. Habener, J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2616–2620.
41. Lawn, R. M., Adelman, J., Bock, S. C., Franke, A. E., Houck, C. M., Najarian, R. C., Seeburg, P. H. & Wion, K. L. (1981) *Nucleic Acids Res.* **9**, 6103–6114.
42. Nishimoto, S. K. & Price, P. A. (1985) *J. Biol. Chem.* **260**, 2832–2836.
43. Hauschka, P. V., Carr, S. A. & Biemann, K. (1982) *Biochemistry* **21**, 638–642.
44. Price, P. A. & Nishimoto, S. K. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2234–2238.
45. MacGillivray, R. T. A. & Davie, E. W. (1984) *Biochemistry* **23**, 1626–1634.
46. Jaye, M., de la Salle, H., Schamber, F., Balland, A., Kohli, V., Findeli, A., Tolstoshev, P. & Lecocq, J.-P. (1983) *Nucleic Acids Res.* **11**, 2325–2335.
47. Fung, M., Campbell, R. M. & MacGillivray, R. T. A. (1984) *Nucleic Acids Res.* **12**, 4481–4492.
48. Long, G. L., Belagaje, R. M. & MacGillivray, R. T. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5653–5656.