The EMBO Journal vol.5 no.8 pp.1885-1890, 1986

Isolation of the human gene for bone gla protein utilizing mouse and rat cDNA clones

A.J.Celeste, V.Rosen, J.L.Buecker, R.Kriz, E.A.Wang and J.M.Wozney

Genetics Institute, Inc., 87 Cambridge Park Drive, Cambridge, MA 02140, USA

Communicated by R.Kamen

cDNAs which encode bone gla protein (BGP), an abundant γ-carboxylated protein of bone, have been cloned from rat and mouse osteosarcoma cell lines. DNA sequence analysis indicates that the cDNAs code for both the 50 (rat) or 46 (mouse) amino acids of the mature proteins and a 49 amino acid leader peptide. The leader peptide of each BGP includes the expected hydrophobic signal sequence and an apparent pro sequence. Although there is no homology between the mature forms of BGP and the γ-carboxylated clotting factors, we note that there is some homology between their leader peptides. These cDNAs have been used to examine the modulation of BGP mRNA levels by osteoblastic cells in response to hormones. The cDNAs have also allowed isolation of the human BGP gene; analysis of this gene indicates the presence of four exons. Comparison of the exon structure of the BGP gene and the Factor IX (a γ-carboxylated clotting factor) gene suggests that the exons encoding the part of the leader peptides presumably directing γ-carboxylation arose from a common ancestral sequence.

Key words: bone gla protein/human gene/γ-carboxylation

Introduction

Bone gla protein (BGP; osteocalcin) constitutes 1-2% of the total protein of bone. Prior to mineralization, BGP is found in bone in only trace amounts but levels of BGP increase markedly concomitant with periods of intense skeletal growth (Hauschka *et al.*, 1983). The presence of BGP during mineralization correlates well with the known ability of BGP to bind calcium and hydroxyapatite (Poser and Price, 1979). BGP synthesis has been demonstrated *in vitro* in bone organ culture (Nishimoto and Price, 1979), dissociated bone cells (Beresford *et al.*, 1984), and in osteoblastic osteosarcoma cell lines such as the rat line ROS 17/2 (Nishimoto and Price, 1980). 1,25-dihydroxy vitamin D_3 ($1,25(OH)_2D_3$), a major calcium regulating hormone, stimulates BGP synthesis in these *in vitro* systems (Price and Baukol, 1980; Lian *et al.*, 1985). These data suggest that BGP may be a marker for bone matrix turnover.

BGP is a small (6000 daltons) polypeptide which is highly conserved among all vertebrate species. Common to all known BGPs are the three γ-carboxyglutamate (gla) residues required for calcium binding and also the two cysteine residues which form an intramolecular disulphide bond. The three gla residues are the result of a vitamin K-dependent post-translational carboxylation of glutamic acid residues at positions 17, 21 and 24 (Poser *et al.*, 1980).

In the present study we describe the isolation from rat and mouse osteosarcoma cell lines of cDNA clones encoding BGP. Using these clones we have determined the amino acid sequence of both BGPs and their precursors. The leader peptides of the proBGP molecules show some sequence homology to those of the clotting factor family of proteins which also contain γ-carboxyglutamic acid. These cDNA clones have also allowed us to isolate the human gene for BGP and to determine its exon structure. Now that direct measurement and localization of BGP is possible, the role of BGP and its interrelationship with various growth factors and hormones involved in bone homeostasis can be studied.

Results

Isolation and sequence analysis of rat cDNA clones

As BGP production by ROS 17/2 cell lines has been extensively characterized (Price and Baukol, 1980; Pan and Price, 1984), $1,25(OH)_2D_3$-induced ROS cells provide an obvious source of BGP mRNA. Two long oligonucleotide probes were designed: a 35mer based on the known partial amino acid sequence of rat BGP (residues 7-18) and a 50mer based on a region of the BGP amino acid sequence where the bovine and human sequences are known to be identical (residues 26-42). The number of oligonucleotide sequences in each pool was reduced (the 35mer is a unique sequence and the 50mer is a pool of eight sequences) based on eukaryotic codon usage, the relative stability of G:T versus A:C base pairs, and the infrequency of the dinucleotide CpG in eukaryotic genes.

To test the usefulness of these probes in detecting BGP mRNA, a modified Northern analysis was performed (see Materials and methods). ROS cell and rat liver polyadenylated RNAs were electrophoresed in agarose gels and subsequently hybridized to the two probes. While each probe hybridized to multiple RNA species, both probes detected only one common band unique to ROS cells which corresponded to an RNA of 0.4-0.8 kb (data not shown). Consequently, a λgt10 cDNA library was constructed from ROS cell mRNA and 50 000 independent clones were screened with the two probes. Six duplicate positives were obtained, and one of these, R22, was subcloned into M13. Sequence analysis clearly showed that it encoded rat BGP. Subsequently two longer clones (R21 and R27) were completely sequenced; they were found to be identical to R22 over the corresponding region. In addition, they extend approximately 50 bp in the 5′ direction.

R21 and R27 contain 480 residues of identical nucleotide sequence, which along with the derived amino acid sequence are presented as Figure 1. These cDNA clones encode the entire mature BGP molecule of 50 residues (numbered 1-50 in Figure 1). The derived sequence is in agreement with the partial protein sequence published for rat BGP (Otawara *et al.*, 1981) except that residues 4 and 28 are both asparagines rather than aspartates. These observed differences are unlikely to be due to cloning artifacts as two independently isolated cDNAs were sequenced. The differences are more likely to be due to amino acid sequencing errors or variation between rat strains. It is interesting to note that the 50mer probe used to screen the ROS λgt10 library

```
                                                                    GAACAGACAAGTCCCACACAGCAACTCGGTGCAGACCTAGCAGACACC   48

-49                                  -40                                  -30
MET Arg Thr Leu Ser Leu Leu Thr Leu Leu Ala Leu Thr Ala Phe Cys Leu Ser Asp Leu
ATG AGG ACC CTC TCT CTG CTC ACT CTG CTG GCC CTG ACT GCA TTC TGC CTC TCT GAC CTG  108

                                     -20                                  -10
Ala Gly Ala Lys Pro Ser Asp Ser Glu Ser Asp Lys Ala Phe Met Ser Lys Gln Glu Gly
GCA GGT GCA AAG CCC AGC GAC TCT GAG TCT GAC AAA GCC TTC ATG TCC AAG CAG GAG GGC  168

                                  -1   1                                   10
Ser Lys Val Val Asn Arg Leu Arg Arg Tyr Leu Asn Asn Gly Leu Gly Ala Pro Ala Pro
AGT AAG GTG GTG AAT AGA CTC CGG CGC TAC CTC AAC AAT GGA CTT GGA GCC CCA GCC CCC  228

                                  20                                   30
Tyr Pro Asp Pro Leu Glu Pro His Arg Glu Val Cys Glu Leu Asn Pro Asn Cys Asp Glu
TAC CCA GAT CCC CTG GAG CCT CAC AGG GAG GTG TGT GAG CTC AAC CCC AAT TGT GAC GAG  288

                                  40                                   50
Leu Ala Asp His Ile Gly Phe Gln Asp Ala Tyr Lys Arg Ile Tyr Gly Thr Thr Val ***
CTA GCG GAC CAC ATT GGC TTC CAG GAC GCC TAC AAG CGC ATC TAT GGC ACC ACC GTT TAG  348

GGCATGTGTTGCCCTGGAGCCCAACGGCAGCTTCAGCTTTGGCTACTCTCCAGGACTCGACCCTCCCTGTTCCCTCTCT  427

CTGCCTCGAAAGTATGGACGGCACAGCTGCTCCAAAATAAAGTCCAGATGAGG-poly(A)                      480
```

Fig. 1. Nucleotide sequence of cDNA clones encoding rat BGP. The coding sequence is indicated by the derived amino acid sequence above the DNA sequence. Numbering of amino acid residues is with respect to residue 1 being the amino-terminus of the mature protein. Numbering of nucleotides is shown on the right.

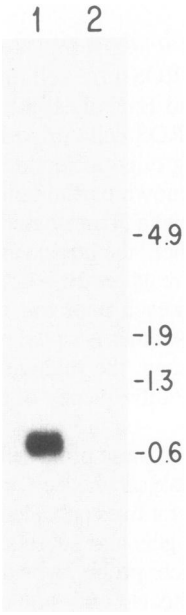

Fig. 2. Detection of BGP mRNA in ROS 17/2 cells. Polyadenylated RNAs from ROS cells (lane 1) and rat liver (lane 2) were electrophoresed, blotted, and hybridized to the rat BGP clone R22. The positions of markers are indicated on the right; sizes are in kb. No additional bands were detected with exposure times 20-fold longer.

was designed on the basis of two incorrect amino acids (asp instead of asn at 28 due to a protein sequencing error and glu instead of asp at 40 due to a rat/human difference) yet still could be used to isolate BGP clones.

The in-frame methionine codon at position −49 is followed by the putative signal sequence necessary for transmembrane transfer of the protein into the rough endoplasmic reticulum. Residues −47 to −33 constitute a region of relative hydrophobicity followed by a more hydrophilic region containing many charged residues. Thus it appears that rat BGP is synthesized as a precursor of 99 amino acids.

There are 132 nucleotides of 3′ non-coding sequence following the UAG stop codon. This region contains the canonical AAUAAA polyadenylation signal located 13 nucleotides 5′ to the poly(A) tail (see Figure 1). Combined with at least 48 nucleotides of 5′ non-coding and 297 nucleotides of coding sequence,

```
                                                                    GAACAGACAAGTCCCACACAGCAGCTTGGTGCACACCTAGCAGACACC   48

-49                                  -40                                  -30
MET Arg Thr Leu Ser Leu Leu Thr Leu Leu Ala Leu Ala Ala Leu Cys Leu Ser Asp Leu
ATG AGG ACC CTC TCT CTG CTC ACT CTG CTG GCC CTG GCT GCG CTC TGT CTC TCT GAC CTC  108

                                     -20                                  -10
Thr Asp Ala Lys Pro Ser Gly Pro Glu Ser Asp Lys Ala Phe Met Ser Lys Gln Glu Gly
ACA GAT GCC AAG CCC AGC GGC CCT GAG TCT GAC AAA GCC TTC ATG TCC AAG CAG GAG GGC  168

                                  -1   1                                   10
Asn Lys Val Val Asn Arg Leu Arg Arg Tyr Leu Gly Ala Ser Val Pro Ser Pro Asp Pro
AAT AAG GTA GTG AAC AGA CTC CGG CGC TAC CTT GGA GCC TCA GTC CCC AGC CCA GAT CCC  228

                                  20                                   30
Leu Glu Pro Thr Arg Glu Gln Cys Glu Leu Asn Pro Ala Cys Asp Glu Leu Ser Asp Gln
CTG GAG CCC ACC CGG GAG CAG TGT GAG CTT AAC CCT GCT TGT GAC GAG CTA TCA GAC CAG  288

                                  40                   46
Tyr Gly Leu Lys Thr Ala Tyr Lys Arg Ile Tyr Gly Ile Thr Ile ***
TAT GGC TTG AAG ACC GCC TAC AAA CGC ATC TAC GGT ATC ACT ATT TAGGACCTGTGCTGCCCCTA  352

AAGCCAAACTCTGGCAGCTCGGCTTTGGCTGCTCTCCGGGACTTGATCCTCCCTGTCCTCTCTCTCTGCCCTGCAAGTA  431

TGGATGTCACAGCAGCTCCAAAATAAAGTTCAGATGAGG-poly(A)                                    470
```

Fig. 3. Nucleotide sequence of cDNA clones encoding mouse BGP. Numbering is as described for Figure 1.

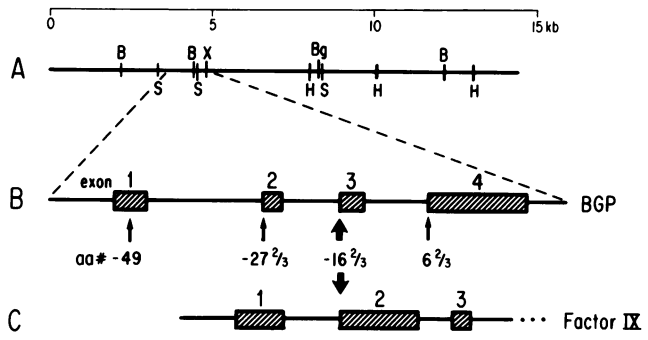

Fig. 4. Characterization of the human BGP gene. **A.** restriction map of a recombinant bacteriophage containing the human BGP gene. Abbreviations are: H, *Hin*dIII; B, *Bam*HI; Bg, *Bgl*II; S, *Sac*I; X, *Xba*I. **B.** Schematic of the BGP gene derived from the DNA sequence between the the *Sac*I and *Xba*I sites. Exons are indicated as boxes; amino acid residue numbers are indicated below. The amino acid numbers at the 5′ ends of exons 2, 3, and 4 indicate the actual positions within the corresponding codons that introns occur, and are therefore non-integral. **C.** Schematic of the 5′ portion of the human factor IX gene.

an mRNA of at least 480 bases is predicted exclusive of the poly(A) tail. This is consistent with an mRNA size of approximately 650 b detected in ROS cell RNA by Northern analysis using the cDNA clone R22 as a probe (Figure 2).

Isolation and sequence analysis of mouse BGP clones

RNA isolated from a murine osteosarcoma cell line was examined for BGP sequences by Northern analysis using the rat cDNA clone as a probe. A band with identical mobility was found in both the ROS cell RNA and murine RNA lanes, although the murine band was of reduced intensity (data not shown). 70 000 plaques of a λgt10 cDNA library constructed from RNA derived from this cell line were screened with the rat cDNA clone as described in Materials and methods. Nine positives were obtained.

DNA sequence analysis of three of these clones yielded the nucleotide and derived amino acid sequence shown in Figure 3. Comparison with the rat sequence indicates that the mouse clones encode a homologous 49 residue leader peptide and the 46 residues of the mature BGP protein. The mature protein is four residues smaller than the rat (50 residues); this is accounted for by a single deletion within the coding sequence (see Discussion). Obvious homology with the rat cDNA clones extends from 48 nucleotides 5′ to the AUG codon through the coding sequence and the 3′ non-coding sequence. Of six clones sequenced, five

had the poly(A) at the position indicated in the figure; one clone had an additional AAGTG preceding the poly(A).

Isolation and characterization of the human BGP gene

A previously constructed and characterized human genomic DNA library (Toole *et al.*, 1984) was screened with R22, also under conditions of reduced stringency. Four of the strongest hybridizing phage were then replated for secondary screening. Three nitrocellulose replicas were made, and the secondaries screened with R22 and the two oligonucleotide probes used to isolate the original rat cDNA clone. One clone, found to hybridize to all three probes, was plaque purified for further characterization.

```
ACGGGGCTGACAGTAGAAATCACAGGCTGTGAGACAGCTGGAGCCCAGCTCTGCTTGAACCTATTTTAGGTCTCTGATCC   80

CCGCTTCCTCTTTAGACTCCCCTAGAGCTCAGCCAGTGCTCAACCTGAGGCTGGGGGTCTCTGAGGAAGAGTGAGTTGGA  160

GCTGAGGGGTCTGGGGCTGTCCCCTGAGAGAGGGGCCAGAGGCAGTGTCAAGAGCCGGGCAGTCTGATTGTGGCTCACCC  240

TCCATCACTCCCAGGGGCCCCTGGCCCAGCAGCCGCAGCTCCCAACCACAATATCCTCTGGGGTTTGGCCTACGGAGCTG  320

GGGCGGATGACCCCCAAATAGCCCTGGCAGATTCCCCCTAGACCCGCCCGCACCATGGTCAGGCATGCCCCTCCTCATCG  400

CTGGGCACAGCCCAGAGGGTATAAACAGTGCTGGAGGCTGGCGGGGCAGGCCAGCTGAGTCCTGAGCAGCAGCCCAGCGC  480

                Met Arg Ala Leu Thr Leu Leu Ala Leu Leu Ala Leu Ala Ala Leu Cys
AGCCACCGAGACACC ATG AGA GCC CTC ACA CTC CTC GCC CTA TTG GCC CTG GCC GCA CTT TGC    543

Ile Ala Gly Gln Ala G
ATC GCT GGC CAG GCA G GTGAGTGCCCCCACCTCCCCTCAGGCCGCATTGCAGTGGGGGCTGAGAGGAGGAAGCA  617

CCATGGCCCACCTCTTCTCACCCCTTTGGCTGGCAGTCCCTTTGCAGTCTAACCACCTTGTTGCAGGCTCAATCCATTTG  697

CCCCAGCTCTGCCCTTGCAGAGGGAGAGGAGGGAAGAGCAAGCTGCCCGAGACGCAGGGGAAGGAGGATGAGGGCCCTGG  777

                                        ly Ala Lys Pro Ser Gly Ala Glu Ser Ser
GGATGAGCTGGGGTGAACCAGGCTCCCTTTCCTTTGCAG GT GCG AAG CCC AGC GGT GCA GAG TCC AGC      845

Lys A
AAA G GTGCAGGTATGAGGATGGACCTGATGGGTTCCTGGACCCTCCCCTCTCACCCTGGTCCCTCAGTCTCATTCCCC  923

CACTCCTGCCACCTCCTGTCTGGCCATCAGGAAGGCCAGCCTGCTCCCCACCTGATCCTCCCAAACCCAGAGCCACCTGA 1003

                la Phe Val Ser Lys Gln Glu Gly Ser Glu Val Val Lys Arg
TGCCTGCCCCCTCTGCTCCACAG CC TTT GTG TCC AAG CAG GAG GGC AGC GAG GTA GTG AAG AGA     1066

Pro Arg Arg Tyr Leu Tyr Gln Trp Leu Gl
CCC AGG CGC TAC CTG TAT CAA TGG CTG GG GTGAGAGAAAAGGCAGAGCTGGGCCAAGGCCCTGCCTCTCC 1136

GGGATGGTCTGTGGGGGAGCTGCAGCAGGGAGTGGCCTCTCTGGGTTGTGGTGGGGGTACAGGCAGCCTGCCCTGGTGGG 1216

CACCCTGGAGCCCCATGTGTAGGGAGAGGAGGGATGGGCATTTTGCACGGGGGCTGATGCCACCACGTCGGGTGTCTCAG 1296

y Ala Pro Val Pro Tyr Pro Asp Pro Leu Glu Pro Arg Arg Glu Val Cys Glu Leu Asn
A GCC CCA GTC CCC TAC CCG GAT CCC CTG GAG CCC AGG AGG GAG GTG TGT GAG CTC AAT     1354

Pro Asp Cys Asp Glu Leu Ala Asp His Ile Gly Phe Gln Glu Ala Tyr Arg Arg Phe Tyr
CCG GAC TGT GAC GAG TTG GCT GAC CAC ATC GGC TTT CAG GAG GCC TAT CGG CGC TTC TAC  1414

Gly Pro Val ***
GGC CCG GTC TAGGGTGTCGCTCTGCTGGCCTGGCCGGCAACCCCAGTTCTGCTCCTCTCCAGGCACCCTTCTTTCCT 1491

CTTCCCCTTGCCCTTGCCCTGACCTCCCAGCCCTATGGATGTGGGGTCCCCATCATCCCAGCTGCTCCCAAATAAACTCC 1571

AGAAGAGGAATCTGTGGGCCTGTGAGTCTGTCCAGTTTATGGAGTGTGGGAGGGAGGTGTCAGGAGGATGGGGGTGAGGA 1651

GGTTTTACCTTCTTCAGTTCTAGA                                                          1675
```

Fig. 5. Nucleotide sequence of the human BGP gene. Coding regions have the derived amino acid sequence printed above. The presumptive promoter sequence TATAAA is underlined (nucleotides 420–425); the polyadenylation signal AATAAA is also underlined (nucleotides 1562–1567).

A restriction map of this phage is shown in Figure 4a. Southern blot analysis of this clone indicated that all hybridization to R22 and to the oligonucleotide probes was between the *Sac*I site at 3.1 kb and the *Xba*I site at 4.8 kb. Consequently, the region contained between these sites was sequenced and the entire human BGP gene localized within it. This sequence is presented as Figure 5. Clear homology between the human BGP gene and the entire sequences of the rat and mouse cDNA clones is apparent. Four blocks of homology are evident, dividing the genomic sequence into four exons. The expected splice donor and acceptor sequences appropriately flank each exon. Some homology extends through the 3′ non-coding region terminating 14 nucleotides past the AATAAA polyadenylation signal. In the 5′ direction, homology ceases approximately 50 bases before the AUG initiator codon.

Regulation of BGP mRNA levels in osteoblasts

The synthesis and regulation of BGP mRNA in osteoblasts and osteoblast-like cell lines was investigated using *in situ* hybridization techniques. ROS cells contained measurable levels of BGP mRNA prior to hormone treatment. Addition of $1,25(OH)_2D_3$ at 10^{-9} M for 24 h resulted in a 7-fold induction of BGP mRNA (Table I). This induction was specific for $1,25(OH)_2D_3$ as incubation of ROS cells with parathyroid hormone (PTH), epidermal growth factor (EGF), or insulin did not increase BGP mRNA. BGP mRNA induction was also concentration dependent (data not shown).

Murine osteoblastic osteosarcoma (MOS) cell lines varied in the level of constitutive synthesis of BGP mRNA. MOS-A expressed very high levels of BGP mRNA, equivalent to ROS cell levels after $1,25(OH)_2D_3$ induction. This level was unchanged by vitamin D (Table I) or by other hormone treatments (data not shown). In contrast, MOS-C appeared to synthesize no BGP mRNA or a level undetected by our assay. No hormone treatment tried induced BGP mRNA synthesis in these cells. MOS-D had BGP mRNA levels slightly higher than ROS 17/2 and this level was not increased by exposure to $1,25(OH)_2D_3$, PTH, or insulin. MOS-B had a BGP mRNA level equivalent to MOS-D, and it could be slightly induced by $1,25(OH)_2D_3$ but not by the other hormones tested.

Primary mouse calvarial cells produced by sequential collagenase digestion showed BGP mRNA levels consistent with the dominant phenotype described for the enriched populations. Fibroblast-enriched population 1 had no measurable BGP mRNA and was not induced to make BGP message by $1,25(OH)_2D_3$ or other hormones (Table I). Populations 2 and 3 both had measurable BGP mRNA levels and both populations responded

Table I. Measurement of BGP mRNA by *in situ* hybridization

mRNA source	BGP mRNA (c.p.m.)		Induction (+)/(−) $1,25(OH)_2D_3$
	(−) $1,25(OH)_2D_3$	(+) $1,25(OH)_2D_3$	
ROS 17/2	435 (8)	3306 (6)[a]	7.6
MOS A	2970 (6)	3010 (11)	–
MOS B	647 (7)	989 (3)[a]	1.5
MOS C	−7	18	–
MOS D	622 (8)	637 (11)	–
1° bone cells 1	10	13	–
2	638 (4)	1786 (5)[a]	2.8
3	386 (10)	656 (6)[a]	1.7

Cells were plated in 96-well dishes, fixed, hybridized, and counted as described in Materials and methods. Each number represents an average of six wells; the percent error is given after each number in parentheses. Non-specific background binding of probe to cells has been subtracted. 10^{-9} M $1,25(OH)_2D_3$-treated versus untreated cells were compared with Student's *t*-test.
[a] $P < 0.01$.

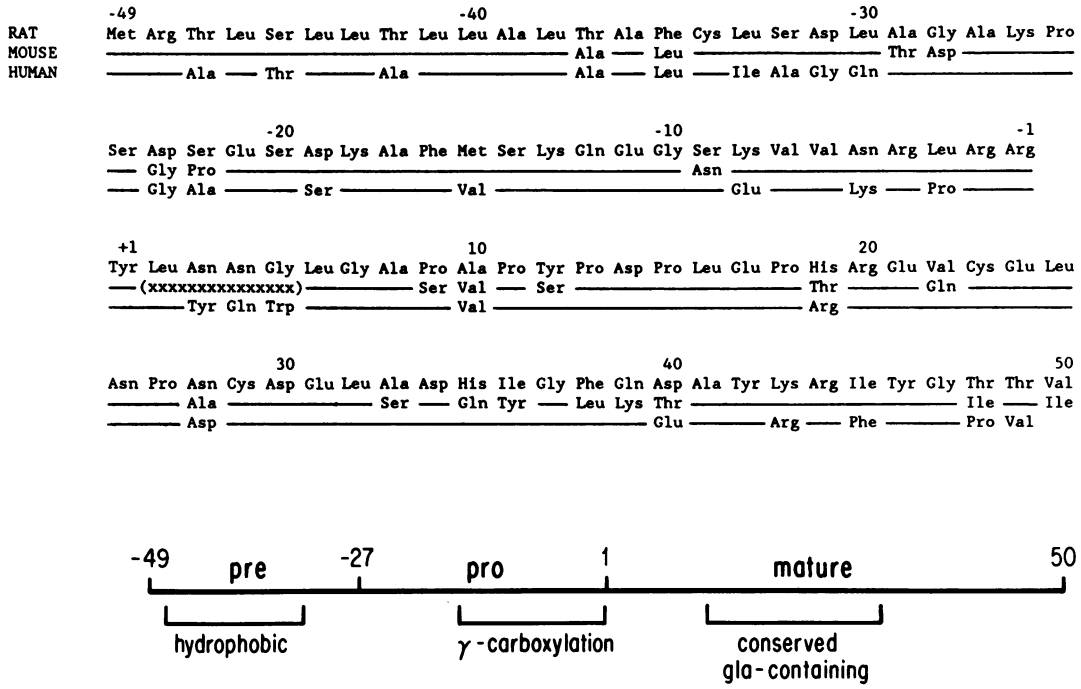

Fig. 6. Comparison of rat, mouse, and human preproBGP sequences. All sequences are derived from DNA sequences. Amino acids which are different in mouse and human relative to rat are noted; a line indicates an identical residue to rat. Numbering is with respect to 1 being the amino-terminus of the mature protein. Below the sequences is a schematic representation of the preproBGP molecule. The hydrophobic signal sequence, the region of homology to the γ-carboxylated clotting factors, and the highly conserved gla-containing domain are indicated.

	-20								-10										-1
RAT BGP	Glu	Ser	Asp	Lys	Ala	Phe	Met	Ser	Lys	Gln	Glu	Gly	Ser	Lys	Val	Val	Asn	Arg	Leu	Arg	Arg
MOUSE BGP	Glu	Ser	Asp	Lys	Ala	Phe	Met	Ser	Lys	Gln	Glu	Gly	Asn	Lys	Val	Val	Asn	Arg	Leu	Arg	Arg
HUMAN BGP	Glu	Ser	Ser	Lys	Ala	Phe	Val	Ser	Lys	Gln	Glu	Gly	Ser	Glu	Val	Val	Lys	Arg	Pro	Arg	Arg
BOVINE PROTHROMBIN	His	Ser	Gln	His	Val	Phe	Leu	Ala	His	Gln	Gln	Ala	Ser	Ser	Leu	Leu	Gln	Arg	Ala	Arg	Arg
BOVINE FACTOR X	Pro	Ala	Gly	Ser	Val	Phe	Leu	Ala	Arg	Asp	Gln	Ala	His	Arg	Val	Leu	Gln	Arg	Ala	Arg	Arg
HUMAN FACTOR IX	Ala	Glu	Cys	Thr	Val	Phe	Leu	Asp	His	Glu	Asn	Ala	Asn	Lys	Ile	Leu	Asn	Arg	Pro	Lys	Arg
BOVINE PROTEIN C	Pro	Pro	Asp	Ser	Val	Phe	Ser	Ser	Ser	Gln	Arg	Ala	His	Gln	Val	Leu	Arg	Ile	Arg	Lys	Arg

Fig. 7. Comparison of leader sequences of BGPs and γ-carboxylated clotting factors, all derived from DNA sequences. For simplicity, only the sequences of bovine prothrombin (MacGillivray and Davie, 1984), bovine factor X (Fung *et al.*, 1984), human factor IX (Anson *et al.*, 1984), and bovine protein C (Long *et al.*, 1984) have been included; the sequences of human prothrombin (Friezner-Degen *et al.*, 1983) and human factor X (Fung *et al.*, 1985) are also available. Residues which are identical to the BGPs are boxed.

to $1,25(OH)_2D_3$ treatment by induction of the mRNA. Population 2, highly osteoblast enriched, showed greater induction of BGP mRNA with $1,25(OH)_2D_3$ than did osteoblast/osteoclast mixed population 3. PTH and insulin treatments did not affect BGP mRNA levels in these populations.

Discussion

The amino acid sequences predicted from the recombinant clones encoding rat, mouse, and human BGPs yield insights into the structure and biochemistry of the BGP molecule. A comparison of these sequences is shown in Figure 6. All three BGPs are apparently synthesized as preproproteins and proteolytically cleaved to yield the mature BGP molecules. The hydrophobic region of the prepropeptide (residues −47 to −33) presumably targets the peptide for secretion. The signal peptidase probably cleaves the molecule carboxy-terminal to the alanine residue (−27) leaving a 26 residue propeptide attached to the mature form (von Heijne, 1984). This cleavage site is consistent with the size of BGP reported in previous studies on the intracellular production of BGP by ROS cells (Nishimoto and Price, 1980). The pro sequence ends with the two basic arg−arg residues. Dibasic sequences are also seen in the propeptide forms of various peptide hormones and clotting factors known to require proteolytic activation. Cleavage of the pro sequence to produce mature BGP would occur in a step subsequent to cleavage by the signal peptidase.

Comparison of the sequences of the three mature BGPs indicates that the central portion of the molecule is highly conserved between species. Conservation in this region is expected as it encompasses the three glutamic acid residues at positions 17, 21 and 24 which are γ-carboxylated and the cysteine residues at 25 and 29. Outside of this region there is a surprising lack of homology between the BGPs. Both the rat and mouse polypeptides contain an additional amino acid residue at the carboxy-terminus relative to the human (and bovine) BGP sequences. The mouse sequence also contains an obvious deletion of four residues near the amino-terminus. In spite of the fact that the rat and mouse sequences are 89% homologous at the nucleotide level, including the 5′ and 3′ untranslated regions, the mature BGP polypeptides are only 70% homologous (discounting the deletion).

The prepropeptide regions of BGPs from the three species show greater conservation than the mature proteins. The hydrophobic secretory signal sequence is highly conserved, as is the region between residues −21 to −1 (see Figure 6). This region shows some homology to the leader sequences of the vitamin K-

dependent clotting factors as indicated in Figure 7. Since γ-carboxylation is common to both BGP and the vitamin K-dependent clotting factors, it is plausible that this shared region of leader sequence is recognized by the carboxylating enzyme or directs the polypeptide to the correct subcellular environment for γ-carboxylation. The fact that BGP accumulated by ROS cells treated with warfarin (which is a vitamin K antagonist) retains this part of the leader (Pan *et al.*, 1985) supports its role in γ-carboxylation.

It is interesting to note that this region of the leader sequences shows homology to a region of the mature form of another γ-carboxylated protein derived from bone, matrix gla protein or MGP (Price and Williamson, 1985). The segment bounded by the absolutely conserved phe and arg residues at −16 and −1 in Figure 7 is homologous to residues 15 (phe) to 30 (arg) in MGP. As three of the glutamic acid residues which are amino-terminal to this sequence in MGP are not γ-carboxylated, and the four glutamic acid residues following this sequence are γ-carboxylated, it is tempting to speculate that this sequence serves to target MGP for the γ-carboxylase reaction. In MGP, unlike BGP and the clotting factors, this sequence would remain a part of the mature protein. It will be of interest to determine if the leader of MGP contains an additional sequence which is homologous to those in Figure 7.

The human gene encoding BGP has been isolated and characterized by homology to the rat and mouse cDNA clones. At the 3′ end, a short region of very strong homology (27 out of 30 nucleotides) surrounds the AATAAA polyadenylation signal. The human gene shows no homology to either the extra five nucleotides preceding the poly(A) in the one mouse cDNA clone described in this paper, or to the additional four nucleotides in two rat cDNA clones described previously (Pan and Price, 1985). Whether these clones represent different polyadenylation sites or the differences represent cloning artifacts is unclear. Several 'G/T clusters', which have been implicated in 3′ processing of mRNAs (Birnstiel *et al.*, 1985), are evident 3′ to the polyadenylation site. At the 5′ end of the gene, the consensus 'promoter' sequence TATAAA occurs 25 nucleotides 5′ to the beginning of the area of homology between the human gene and the rat and mouse cDNA clones.

The relatively small preproBGP coding sequence is divided into four exons (Figure 4b). The hydrophobic leader sequence is encoded by exon 1. Exon 2 is only 33 bp and contains the predicted signal peptidase cleavage site near its 5′ end. Exon 3 encodes the region showing homology to the γ-carboxylated clotting factors. The amino-terminal seven residues of the mature BGP protein are also coded for by exon 3; the four amino acid deletion in the murine BGP sequence occurs in this region. The majority of the mature protein (residues 8−50) is encoded by a single exon, exon 4. One may have expected an intron separating the coding sequences for the highly conserved gla-containing domain and the carboxy-terminal portion of the BGP molecule which is not highly conserved among species.

The homology between the leader sequences of the vitamin K-dependent clotting factors and BGP led us to compare the structures of their genes. If the amino termini of the mature BGP and factor IX proteins are aligned, an intron occurs in the identical position in each gene (in the codon for amino acid residue −17; Anson *et al.*, 1984). This comparison is shown schematically in Figure 4c. The region delineated begins with the leader sequence homologous to BGP and the vitamin K-dependent clotting factors, suggesting that the exons encoding the domains directing γ-carboxylation arose from a common ancestral coding sequence. The corresponding intron in the human prothrombin gene occurs in this same position (Davie *et al.*, 1983); that in the human protein C gene occurs at a similar though not identical position (in the codon for amino acid residue −19; Foster *et al.*, 1985).

The BGP probes described in this report allow *in situ* measurement of BGP mRNA, and investigation of the effects of various hormones and factors on BGP mRNA synthesis. Preliminary experiments outlined in Table I (and unpublished observations) have allowed us to directly compare BGP regulation in model systems used to study bone formation. We have been able to demonstrate that $1,25(OH)_2D_3$ regulates the synthesis of BGP at the transcriptional level in primary populations of osteoblasts and in certain osteoblast-like cell lines. Cell lines presently used as models of osteoblasts *in vivo* are heterogeneous with respect to the amount of BGP mRNA synthesized and their ability to modulate BGP mRNA levels with $1,25(OH)_2D_3$. This may reflect changes in BGP expression during osteoblast differentiation and life cycle. We are presently using BGP cDNA probes to determine when BGP message is first expressed by osteoblasts during bone development and what factors regulate this expression.

While recent investigations have suggested a role for BGP in skeletal formation, mineralization, and bone remodeling, BGP-deficient animals exhibit normal skeletal development and maintenance. By examining a number of osteoblast-like cell lines, we have observed that the ability of these cells to mineralize *in vitro* or *in vivo* does not correlate with their ability to make or regulate BGP synthesis, suggesting BGP has no direct role in mineralization. Since we are now able to monitor and localize BGP mRNA expression during bone development, we may be able to define a role of BGP in bone formation or remodeling.

Materials and methods

RNA isolation

The rat osteosarcoma cell line ROS 17/2 (kindly provided by G.Rodan) was grown in Ham's F12 (Gibco), 28 mM Hepes, 1.5 mM glutamine, 10% heat-inactivated fetal bovine serum (HIFBS), pH 7.6, to confluence followed by induction for 24 h in F12, 2% HIFBS, 10^{-9} M $1,25(OH)_2D_3$ (kindly given by B.Kream). RNA was isolated from these cells and rat liver by the guanidinium isothiocyanate method (Chirgwin *et al.*, 1979). A murine cell line derived from a spontaneous osteosarcoma (kindly provided by N.Teich) was grown in DME, 4 mM glutamine, 10% HIFBS. RNA was isolated by a modification of the NP-40 lysis method (Favaloro *et al.*, 1980).

RNA hybridization analysis

For RNA detection using a nick-translated probe, 3 μg polyadenylated RNA from either vitamin D-induced ROS cells or rat liver were electrophoresed on a 1.0% agarose formaldehyde gel and transferred to nitrocellulose as described (Goldberg, 1980). The blot was then hybridized to 8 × 10^6 c.p.m. of the nick-translated insert from the cDNA clone R22 in 50% formamide, 5 × SSC, 5 × Denhardt's, 100 μg/ml denatured salmon sperm DNA, 0.1% SDS at 42°C. Washes were performed at 50°C with 0.2 × SSC, 0.1% SDS. Two of the mol. wt markers (1.3 and 0.6 kb) were synthesized using the SP6 system (Promega-Biotec) from cloned DNAs of known size.

For oligonucleotide probes, 1 μg of polyadenylated RNA was denatured in 10 mM CH_3HgOH and electrophoresed on a 1.4% agarose gel. The gel was subsequently dehydrated (on a gel dryer) and hybridized overnight at 45°C in standard hybridization buffer (5 × SSC, 0.1% SDS, 5 × Denhardt's, 100 μg/ml denatured salmon sperm DNA) with 5 × 10^6 c.p.m./ml kinased oligonucleotide probe. Washes were performed at 45°C with 5 × SSC, 0.1% SDS.

cDNA library screening

Oligo dT primed cDNA synthesis was performed as described (Toole *et al.*, 1984) except that linkers were removed by agarose gel electrophoresis rather than by CL-4B column chromatography. Phage were plated at a density of 5 × 10^3/150 mm plate. Nitrocellulose filter replicas of the plates were hybridized with 10^6 c.p.m./ml kinased oligonucleotide probes under identical conditions as described in the preceding section. The probes were: (1) TAGGCTTCTTGGAAGCC($^{A}_{G}$)ATGTGGTCAGC($^{G}_{C}$)AGTTCATCACAGTC($^{T}_{G}$)GGGTT and (2) GGCTCCAGGGGGTCAGGGTAGGGGGCAGGGGCACC. Hybridization of

the murine osteosarcoma cDNA library filters with the nick-translated insert for the rat cDNA clone R22 (10^6 c.p.m./ml) was done in standard hybridization buffer at 65°C. Washes were at 65°C with 1 × SSC, 0.1% SDS.

Genomic library screening

8×10^5 phage were plated on 20 plates and screened with the nick-translated R22 probe. Hybridization was performed at 50°C in standard hybridization buffer, and washes at 50°C with 1 × SSC, 0.1% SDS.

DNA sequence analysis

All DNA sequence analysis was performed by the dideoxy chain termination method.

Measurement of BGP mRNA in cells

ROS 17/2 and MOS cells were seeded in 96-well dishes (Dynatech) at 2×10^4 cells/well in the media described above with 10% HIFBS until confluent (~48 h). Medium was then changed to 2% HIFBS. After 24 h, medium was replaced with fresh medium with or without 10^{-9} M $1,25(OH)_2D_3$ and cells were incubated on a rocker platform for 24 h. At the end of the experiment, cells were rinsed with Ca/Mg-free PBS and fixed with 4% paraformaldehyde in PBS for 35 min at 37°C (Lawrence and Singer, 1985). The cells were then rinsed sequentially with PBS and 0.2 M Tris, 0.1 M glycine, pH 7.0 for 15 min each at 37°C. Hybridization was for 14 h at 42°C in 50% formamide, 2 × SSC, 50 mM sodium phosphate pH 6.5, 4 mM vanadyl ribonucleosides, using a ^{32}P-labeled single stranded DNA probe synthesized from an M13 template containing the rat BGP cDNA clone. The cell layers were then washed for 30 min at 42°C with 50% formamide, 0.5 × SSC (twice) followed by 30 min with 0.1 × SSC. Wells were then removed from the plates and counted in a Beckman LS 7800 scintillation counter. Rat embryo muscle fibroblasts, which synthesize no BGP mRNA by Northern blot analysis, were run as a control for non-specific binding of the probe. Primary populations of mouse bone cells were isolated from 2-day-old mouse calvaria using the procedure described by Peck *et al.*, 1964 for rat calvaria. The phenotypes of populations 1, 2, and 3 were determined by measuring the change in Type I collagen synthesis of each population when incubated with 10^{-9} M PTH and $1,25(OH_2)D_3$. It was determined that population 1 was fibroblast enriched while populations 2 and 3 were osteoblast and osteoclast enriched, respectively. For BGP mRNA determination the cells were seeded at 2×10^4/well in BGJ/DME, 10% HIFBS, 4 mM glutamine and treated as described for the ROS and MOS cells above.

References

Anson,D.S., Choo,K.H., Rees,D.J.G., Giannelli,F., Gould,K., Huddleston,J.A. and Brownlee,G.G. (1984) *EMBO J.*, **3**, 1053–1060.

Beresford,J.N., Gallagher,J.A., Poser,J.W. and Russell,R.G.G. (1984) *Metab. Bone Dis. Rel. Res.*, **5**, 229–234.

Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.

Davie,E.W., Degen,S.J.F., Yoshitake,S. and Kurachi,K. (1983) *Dev. Biochem.*, **25**, 45–52.

Favoloro,J., Treisman,R. and Kamen,R. (1980) *Methods Enzymol.*, **65**, 718–749.

Foster,D.C., Yoshitake,S. and Davie,E.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4673–4677.

Friezner-Degen,S.J., MacGillivray,R.T.A. and Davie,E.W. (1983) *Biochemistry*, **22**, 2087–2097.

Fung,M.R., Campbell,R.M. and MacGillivray,R.T.A. (1984) *Nucleic Acids Res.*, **12**, 4481–4492.

Fung,M.R., Hay,C.W. and MacGillivray,R.T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3591–3595.

Goldberg,D.A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5794–5798.

Hauschka,P.V., Frenkel,J., DeMuth,R. and Gundberg,C.M. (1983) *J. Biol. Chem.*, **258**, 176–182.

Lawrence,J.B. and Singer,R.H. (1985) *Nucleic Acids Res.*, **13**, 1777–1799.

Lian,J.B., Coutts,M. and Canalis,E. (1985) *J. Biol. Chem.*, **260**, 8706–8710.

Long,G.L., Belangaje,R.M. and MacGillivray,R.T.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5653–5656.

MacGillivray,R.T.A. and Davie,E.W. (1984) *Biochemistry*, **23**, 1626–1634.

Nishimoto,S.K. and Price,P.A. (1979) *J. Biol. Chem.*, **254**, 437–441.

Nishimoto,S.K. and Price,P.A. (1980) *J. Biol. Chem.*, **255**, 6579–6583.

Otawara,Y., Hosoya,N., Moriuchi,S., Kasai,H. and Okuyama,T. (1981) *Biomed. Res.*, **2**, 442–446.

Pan,L.C. and Price,P.A. (1984) *J. Biol. Chem.*, **259**, 5844–5847.

Pan,L.C. and Price,P.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6109–6113.

Pan,L.C., Williamson,M.K. and Price,P.A. (1985) *J. Biol. Chem.*, **260**, 13398–13401.

Peck,W.A., Birge,S.J.,Jr. and Fedak,S.A. (1964) *Science*, **146**, 1476–1477.

Poser,J.W. and Price,P.A. (1979) *J. Biol. Chem.*, **254**, 431–436.

Poser,J.W., Esch,F.S., Ling,N.C. and Price,P.A. (1980) *J. Biol. Chem.*, **255**, 8685–8691.

Price,P.A. and Baukol,S.A. (1980) *J. Biol. Chem.*, **255**, 11660–11663.

Price,P.A. and Williamson,M.K. (1985) *J. Biol. Chem.*, **260**, 14971–14975.

Toole,J.J., Knopf,J.L., Wozney,J.M., Sultzman,L.A., Buecker,J.L., Pittman,D.D., Kaufman,R.J., Brown,E., Shoemaker,C., Orr,E.C., Amphlett,G.W., Foster,W.B., Coe,M.L., Knutson,G.J., Fass,D.N. and Hewick,R.M. (1984) *Nature*, **312**, 342–347.

von Heijne,G. (1984) *J. Mol. Biol.*, **173**, 243–251.

Received on 3 March 1986; revised on 26 May 1986